

# Experimental Physics 3 Course on Optics and Quantum Mechanics

Frank Cichos

2024-10-20



# **Table of contents**



## Chapter 1

# Experimental Physics 3

Experimental  
Physics 3





## Chapter 2

# Welcome to the Experimental Physics 3 Course!

In this Experimental Physics 3 course, we will explore fundamental experiments and mathematical descriptions related to light propagation, electromagnetic waves, and their material counterpart, matter waves. Specifically, we will focus on:

- Geometrical Optics
- Wave Optics
- Electromagnetic Waves
- Matter Waves and Quantum Mechanics

The fields of optics and quantum mechanics are currently vibrant areas of research, with rapidly evolving optical technologies, high-resolution microscopy, and quantum information science. All of these advancements build upon the foundations that we will address in this course.



# Part I

## Geometrical Optics



# Chapter 3

## Geometrical Optics

In this section, we will explore the fundamental principles that govern how light behaves when it encounters different media and surfaces.

### 3.1 Learning Objectives

By the end of this section, you should be able to:

- Understand and apply the laws of reflection and refraction.
- Analyze image formation by mirrors, lenses, and prisms.
- Describe the working principles of various optical instruments.
- Explain phenomena such as dispersion and imaging errors.

### 3.2 Topics Covered

1. **Reflection** Explore how light reflects off surfaces following the law of reflection.
2. **Refraction and Total Internal Reflection** Understand how light bends when passing through different media and the conditions for total internal reflection.
3. **Mirrors, Prisms and Lenses** Learn about various optical elements and how they form images.
4. **Optical Instruments** Study devices like telescopes and microscopes that utilize mirrors and lenses.
5. **Dispersion** Discover how different wavelengths of light refract differently, leading to phenomena like rainbows.
6. **Imaging Errors** Examine common aberrations in optical systems and methods to correct them.

### 3.3 Introduction

Geometrical optics is an approximate description of light propagation in the limit of infinitely small wavelength, where all wave phenomena like diffraction can be neglected.

Light interacts with materials in predictable ways, allowing us to design optical systems for imaging, magnification, and more.

### 3.4 Assumptions of Geometrical Optics

Geometrical optics provides an approximate description of light behavior and is based on several key assumptions. These assumptions simplify the complex nature of light while still allowing for accurate predictions in many practical scenarios.

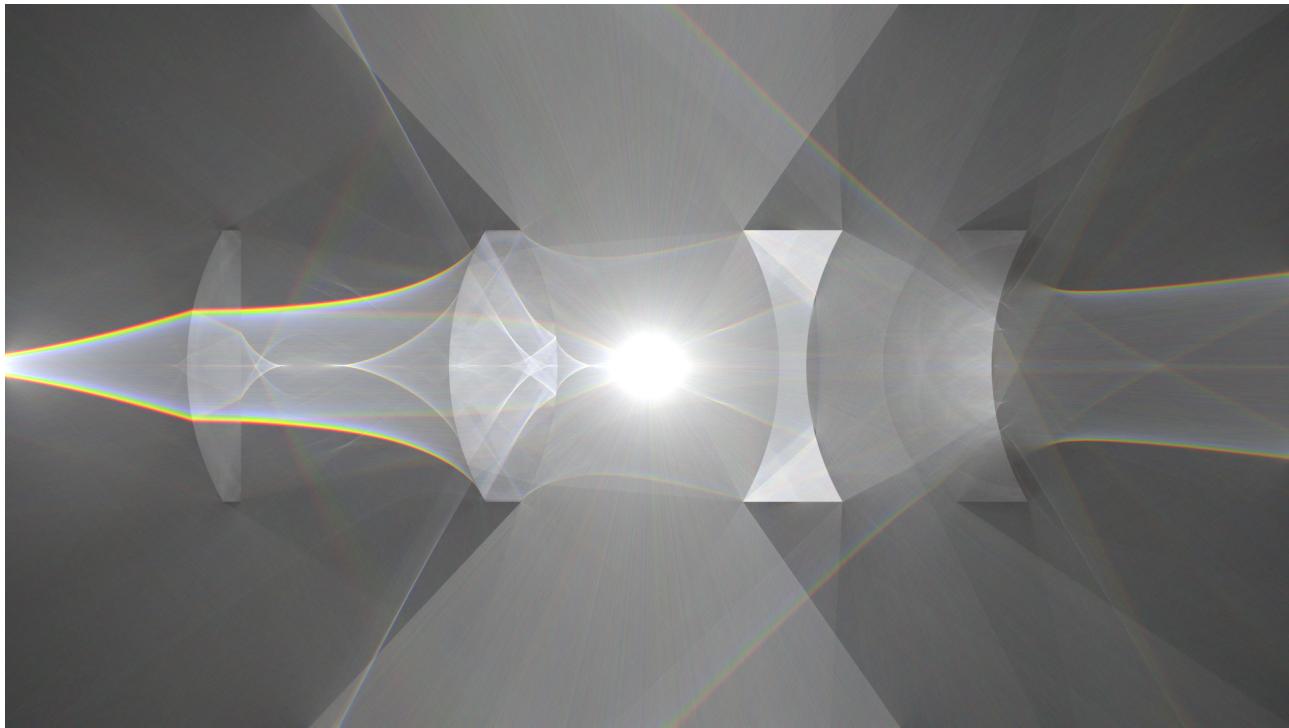


Figure 3.1: Light rays passing through a lens system generated with [Tantalum](#)

### **i** Core Assumptions

#### 1. Light Sources and Detection:

- Light rays emerge from a light source
- Light rays are detected by a detector

#### 2. Light-Matter Interaction:

- Interaction is characterized by a refractive index  $n$
- The speed of light in a medium is given by  $c = c_0/n$ , where  $c_0$  is the speed of light in vacuum
- The speed in vacuum is **299.792.458 m/s** and is connected to the definition of the meter

#### 3. Light Propagation:

- Light propagates in straight line paths (rays) in a homogeneous medium
- Light bends to a curved path in inhomogeneous media with varying refractive index  $n(\mathbf{r})$

#### 4. Behavior at Interfaces:

- Rays may be reflected and refracted at interfaces between media

These assumptions form the foundation for understanding and predicting light behavior in the context of geometrical optics.

# Chapter 4

## Reflection

### Historical Context of Reflection Laws

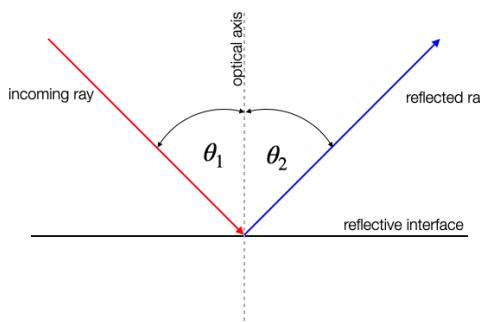
The study of reflection has a rich history dating back to ancient times:

1. **Ancient Greece (300 BCE)**: Euclid, in his work “Catoptrics,” was among the first to formally describe the law of reflection. He stated that the angle of incidence equals the angle of reflection.
2. **Ancient Rome (50 CE)**: Hero of Alexandria expanded on Euclid’s work, applying the principle that light travels along the path of least distance.
3. **Islamic Golden Age (1000 CE)**: Ibn al-Haytham (Alhazen) made significant contributions to optics in his “Book of Optics.” He conducted experiments to verify the law of reflection and explored the properties of spherical and parabolic mirrors.
4. **17th Century**: Fermat’s Principle, formulated by Pierre de Fermat, provided a more general framework for understanding reflection (and refraction) based on the principle of least time.
5. **Modern Era**: The understanding of reflection has been further refined with the development of electromagnetic theory by James Clerk Maxwell in the 19th century and quantum optics in the 20th century.

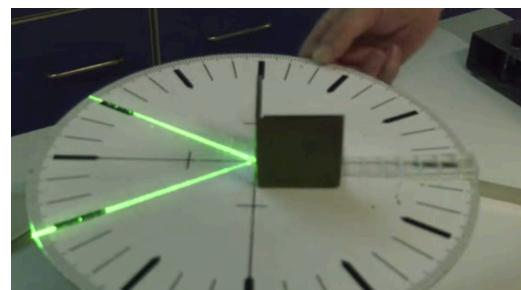
The law of reflection is probably the most simple one. Yet the simplicity gives us the chance to define some basic objects which we will further use for the description of light rays and their propagation.

### Law of Reflection

The sketch below shows the reflection of an incoming light ray (red) on an interface. This incoming light ray has an angle  $\theta_1$  with the axis (dashed line), which is perpendicular to the reflecting surface. As compared to X-ray diffraction, we measure the angle to the normal of the surface and not to the surface itself.



(a) Law of reflection



(b) Experimental Demonstration

Figure 4.1: Figure ?? illustrates the law of reflection, while Figure Figure ?? shows an experimental demonstration.

Figure ?? shows the reflection of an incoming light ray (red) on an interface. This incoming light ray has an angle  $\theta_1$  with the axis (dashed line), which is perpendicular to the reflecting surface. As compared to X-ray diffraction, we measure the angle to the normal of the surface and not to the surface itself.

The law of reflection tells us now, that the outgoing reflected ray is now leaving the surface under an angle  $\theta_2 = \theta_1$ . So both angles are the same for the reflection.

### Law of Reflection

If a ray is incident to a reflecting surface under an angle  $\theta_1$  it will be reflected towards under an angle  $\theta_2 = \theta_1$  to the same side of the surface.

## Fermat's Principle

The law of reflection can be actually obtained from a variational principle saying the light rays propagate along those path on which the propagation time is an extremum. This variational principle is called Fermat's principle.

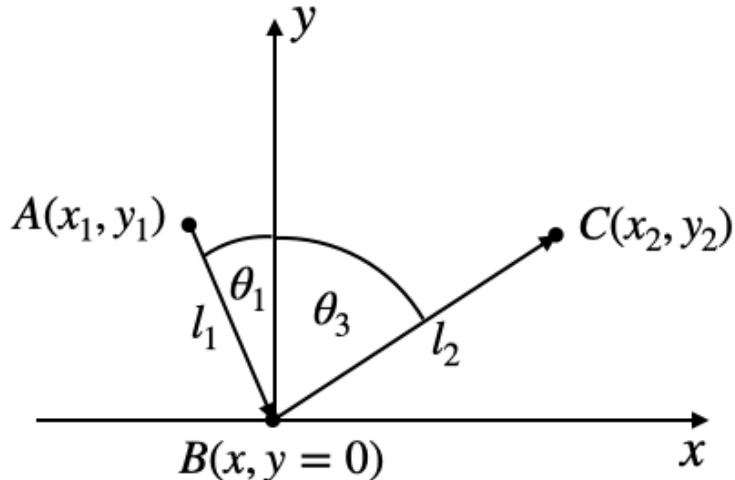


Figure 4.2: Sketch for deriving the law of reflection from Fermat's principle

Consider now a light ray that should travel from point  $A$  to point  $C$  via a point  $B$  on the mirror surface. In general multiple paths are possible such as the one indicated in the above picture. Clearly this path is not satisfying our reflection law formulated above. Fermat's principle now restricts the path length from  $A$  to  $C$  to be the one, which takes the least amount of time.

### Fermat's principle

The path taken by a ray between two given points  $A, B$  is the path that can be traversed in the least time.

*More precise alternative:* A ray going in a certain particular path has the property that if we make a small change in the ray in any manner whatever, say in the location at which it comes to the mirror, or the shape of the curve, or anything, there will be no first-order change in the time; there will be only a second-order change in the time.

So let us consider that constraints to the path length. The total length the light hast to travel via the three points is

$$l = l_1 + l_2 = \sqrt{(x - x_1)^2 + y_1^2} + \sqrt{(x_2 - x)^2 + y_2^2}$$

The time that is required by the light to travel that distance  $l$  is then given by

$$t = \frac{l}{c},$$

where  $c$  is the speed of light in the medium above the mirror. If this time should now be a minimum, we have to take the derivative of the time  $t$  with respect to the position  $x$  on the mirror and set that to zero, i.e.,

$$\frac{dt}{dx} = 0. \quad (4.1)$$

This results in Equation ??

$$\frac{x - x_1}{\sqrt{(x - x_1)^2 + y_1^2}} = \frac{x_2 - x}{\sqrt{(x_2 - x)^2 + y_2^2}},$$

which is actually

$$\frac{x - x_1}{l_1} = \frac{x_2 - x}{l_2}$$

or

$$\sin(\theta_1) = \sin(\theta_2)$$

which finally requires

$$\theta_1 = \theta_2$$

and is our law of reflection. Thus, reflection satisfies Fermat's principle, i.e. the light rays propagate along those path on which the propagation time is an extremum.

### 💡 Principle of Least Action (Hamilton's Principle)

The Principle of Least Action, also known as Hamilton's Principle, is a fundamental concept in classical mechanics. It states that the path taken by a physical system between two states is the one for which the action integral is stationary (usually a minimum).

## Action Integral

The action  $S$  is defined as the integral of the Lagrangian  $L$  over time:

$$S = \int_{t_1}^{t_2} L dt$$

## Lagrangian

The Lagrangian  $L$  is a function that summarizes the dynamics of the system. For a system with generalized coordinates  $q_i$  and velocities  $\dot{q}_i$ , the Lagrangian is typically given by:

$$L = T - V$$

where:

- $T$  is the kinetic energy of the system.
- $V$  is the potential energy of the system.

## Euler-Lagrange Equations

Hamilton's Principle leads to the Euler-Lagrange equations, which are the equations of motion for the system. These equations are derived by requiring that the action  $S$  be stationary with respect to variations in the path  $q_i(t)$ :

$$\delta S = 0$$

This condition leads to the following differential equations:

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_i} \right) - \frac{\partial L}{\partial q_i} = 0$$

These are the Euler-Lagrange equations, and they provide a powerful method for deriving the equations of motion for a wide variety of physical systems.

### Example: Simple Harmonic Oscillator

For a simple harmonic oscillator with mass  $m$  and spring constant  $k$ , the Lagrangian is:

$$L = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2$$

Applying the Euler-Lagrange equation:

$$\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{x}} \right) - \frac{\partial L}{\partial x} = 0$$

we get:

$$m\ddot{x} + kx = 0$$

which is the familiar equation of motion for a simple harmonic oscillator.

Hamilton's Principle and the associated Euler-Lagrange equations are foundational in classical mechanics and have far-reaching implications in other areas of physics, including quantum mechanics and field theory.

# Chapter 5

## Refraction and Total Internal Reflection

### 5.1 Refraction

#### Historical Context of Refraction

The understanding of refraction has evolved over centuries, with contributions from various cultures and scientific traditions. This timeline highlights key milestones in the discovery and formalization of refraction, showcasing how our comprehension of this fundamental optical phenomenon has deepened over time:

1. **Ancient Greece (3rd century BCE):** Euclid noticed that a stick partially submerged in water appears bent. Archimedes studied the refraction of light in water.
2. **Ancient Rome (1st century CE):** Ptolemy conducted experiments on refraction and compiled tables of refraction angles for different media.
3. **Islamic Golden Age (10th-11th centuries):** Ibn Sahl (940-1000) discovered the law of refraction, describing it geometrically. Alhazen (965-1040) studied lenses and the human eye, contributing significantly to optics.
4. **Middle Ages:** Robert Grosseteste (1175-1253) and Roger Bacon (1214-1294) studied refraction and its application to lenses.
5. **Renaissance:** Thomas Harriot (1560-1621) rediscovered the law of refraction but didn't publish his findings.
6. **17th Century:** Willebrord Snellius (1580-1626) derived the mathematical law of refraction (Snell's law) around 1621. René Descartes (1596-1650) independently derived and published the law of refraction in 1637. Pierre de Fermat (1607-1665) derived the law of refraction using his principle of least time.
7. **19th Century:** Augustin-Jean Fresnel (1788-1827) developed the wave theory of light, explaining refraction in terms of changes in wave speed.
8. **20th Century:** The quantum mechanical understanding of light, which emerged in the early 20th century, significantly impacted our view of refraction. Max Planck's work on black body radiation in 1900 and Albert Einstein's explanation of the photoelectric effect in 1905 laid the groundwork for the quantum nature of light. This quantum perspective provided a complementary explanation to the wave theory, describing refraction in terms of photons interacting with the atoms in the medium. While this quantum view offers insights into certain aspects of refraction, it's important to note that both the wave and particle descriptions of light are necessary for a complete understanding of optical phenomena.

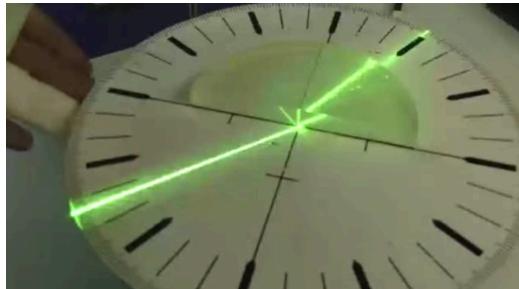
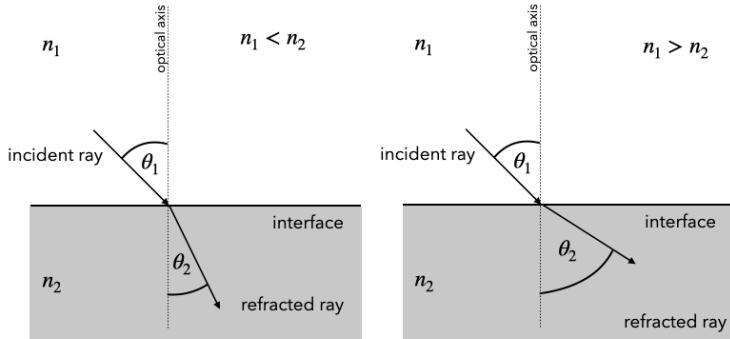
The law of refraction is the second important law of geometrical optics. It relates the refractive index  $n_1$  and angle of incidence  $\theta_1$  on one side of an interface to the refractive index  $n_2$  and angle of refraction  $\theta_2$  on the other side. Both the law of reflection and the law of refraction can be derived from more fundamental principles

such as Fermat's principle of least time and are consistent with the conservation of energy. Their relation to momentum is more complex and involves considering the interaction of light with the medium at an atomic level. These laws provide a mathematical framework for predicting how light behaves when it encounters interfaces between different media, forming the basis for understanding a wide range of optical phenomena and the design of optical devices.

## Refractive Index

The refractive index  $n$  is a material constant representing the factor by which the speed of light is reduced in the medium compared to its speed in vacuum. For most natural materials and visible light, the refractive index is  $n \geq 1$ , as light typically travels slower in media than in vacuum. However, in certain special cases—such as for X-rays in some materials or in engineered metamaterials—the refractive index can be less than 1 or even negative. Understanding these exotic cases requires a deeper exploration of the electromagnetic properties of materials and the origin of the refractive index, which we will address later.

## Snells Law



### Law of Refraction (Snell's Law)

The law of refraction (Snell's law) is given for the above sketch by the equation:

$$n_1 \sin(\theta_1) = n_2 \sin(\theta_2)$$

You can explore the law of refraction using the interactive visualization below. The visualization shows a light ray incident on an interface between two media with different refractive indices. You can adjust the angle of incidence and the refractive index of the first medium to see how the angle of refraction changes according to Snell's law.

Snell's law leads to some general patterns in the behavior of light rays at interfaces, which are worth remembering. Consider these two cases:

1. When light moves from a medium with lower refractive index to one with higher refractive index ( $n_1 < n_2$ ):
  - The refracted ray bends towards the normal (optical axis)
  - The angle of refraction is smaller than the angle of incidence ( $\theta_2 < \theta_1$ )
2. When light moves from a medium with higher refractive index to one with lower refractive index ( $n_1 > n_2$ ):

- The refracted ray bends away from the normal (optical axis)
- The angle of refraction is larger than the angle of incidence ( $\theta_2 > \theta_1$ )

Figure Figure ?? illustrates these principles with three plots showing how the refracted angle changes with the incident angle for two common interface scenarios: glass-to-air and air-to-glass. These plots clearly demonstrate the different behaviors described above.

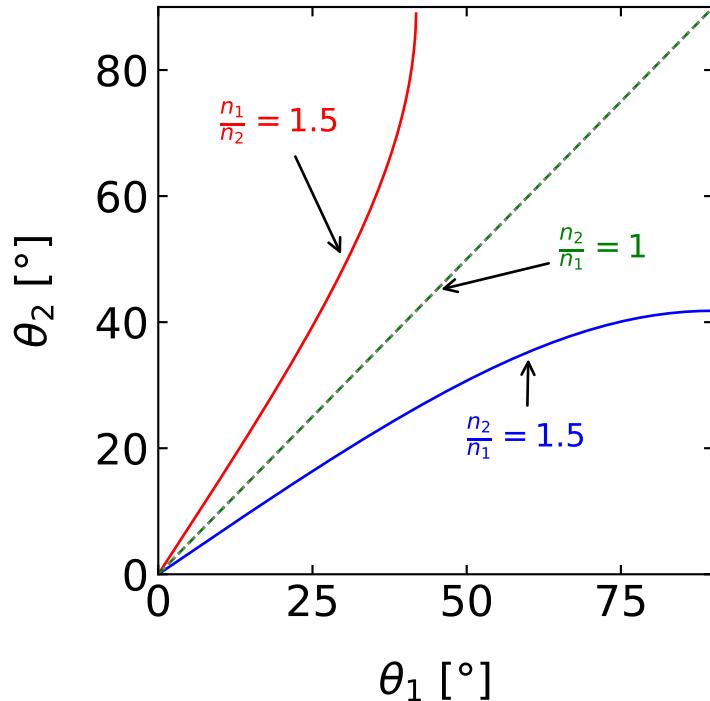


Figure 5.1: Snell's law for different combinations of refractive indices. The plots show the relationship between incident angle ( $\theta_1$ ) and refracted angle ( $\theta_2$ ) for three scenarios: (a) light passing from air to glass, (b) light passing from glass to air, and (c) a comparison of both cases. Note how the curves differ when light moves into a medium with higher refractive index versus a lower refractive index.

## Total Internal Reflection

The above diagram reveals a special case occurring when  $n_1 > n_2$ . Under these conditions, we can increase the incident angle  $\theta_1$  until the outgoing angle reaches  $\theta_2 = \frac{\pi}{2}$ . At this point, the refracted ray would be traveling along the interface between the two media. For any incident angle  $\theta_1$  larger than this critical angle, there is no refracted ray at all; instead, we observe only a reflected ray. This phenomenon, known as **total internal reflection**, occurs despite the fact that the material with refractive index  $n_2$  is completely transparent.

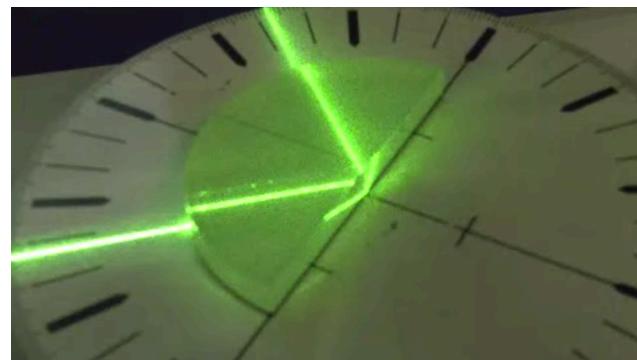
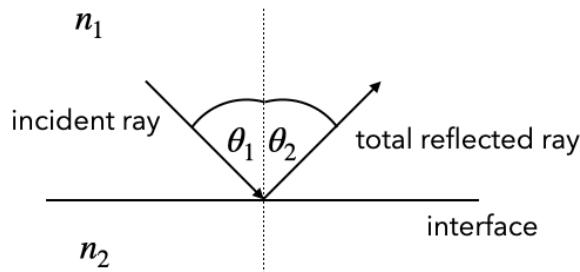
Let's formalize this concept mathematically. Using Snell's law and setting  $\theta_2 = \frac{\pi}{2}$ , we obtain the equation for the critical angle  $\theta_c$ :

$$\theta_1 = \theta_c = \sin^{-1} \left( \frac{n_2}{n_1} \right)$$

Note that the  $\sin^{-1}()$  function requires an argument  $\leq 1$ , which is why this phenomenon only occurs when  $n_2 < n_1$ .

It's important to understand that during total internal reflection, all of the light energy is reflected back into the first medium, hence the term 'total'. However, electromagnetic optics reveals an interesting subtlety: an evanescent wave penetrates a short distance into the second medium, though it doesn't propagate energy across the boundary.

When the incident angle exceeds the critical angle, Snell's law as we've written it no longer applies. Instead, we observe perfect reflection, where the angle of reflection equals the angle of incidence, just as in regular reflection from a mirror. This reflection occurs without any loss of energy to the second medium, making it an extremely efficient process.

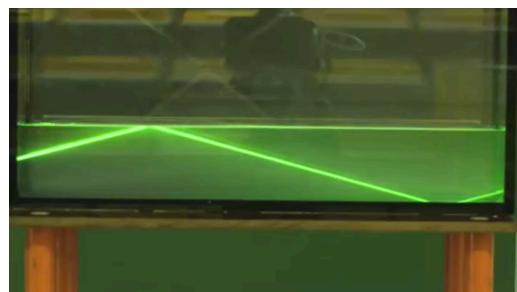
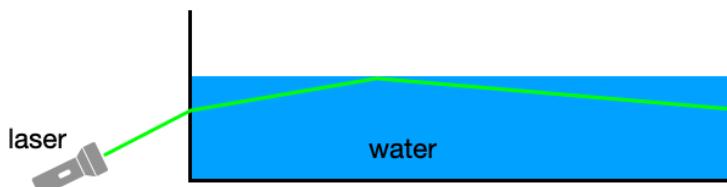


### Total Internal Reflection

Total internal reflection occurs when light is passing from higher refractive index to lower refractive index materials for incidence angle larger than a critical angle

$$\theta_c = \sin^{-1} \left( \frac{n_2}{n_1} \right)$$

We can demonstrate total internal reflection very easily with a water basin, for example, where we couple in light from a laser from the side.



But you could try that yourself also in the bath tub diving below the water surface.

Total internal reflection has numerous practical applications:

1. Fiber optic communications: Light signals can travel long distances with minimal loss through optical fibers.
2. Optical instruments: Prisms in binoculars and telescopes use total internal reflection to manipulate light paths.
3. Gemstones: The sparkle of diamonds is enhanced by total internal reflection trapping light within the stone.
4. Medical endoscopes: Total internal reflection helps guide light through flexible tubes for internal imaging.

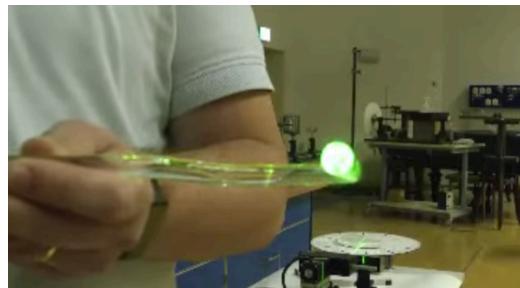
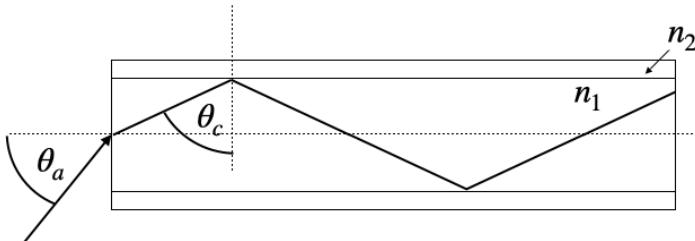
### Optical Fibers and Total Internal Reflection

Total internal reflection plays a crucial role in modern telecommunications, particularly in optical fibers, which are also part of many experimental setups. These fibers are essentially ultra-thin glass wires, ranging in diameter from a few micrometers to several hundred micrometers, designed to transport light over long distances with minimal loss.

The structure of an optical fiber is key to its function:

1. Core: A central glass core with a refractive index  $n_1$
2. Cladding: A surrounding layer with a slightly lower refractive index  $n_2$

This difference in refractive indices is what allows total internal reflection to occur within the fiber.



For light to propagate effectively through the fiber, it must enter at an angle that ensures total internal reflection at the core-cladding interface. This leads to the concept of the acceptance angle,  $\theta_a$ , which is the maximum angle at which light can enter the fiber and still undergo total internal reflection.

To characterize this acceptance angle, optical engineers use a parameter called the **Numerical Aperture (NA)**.

#### Numerical Aperture

The Numerical Aperture of a fiber is defined as the sine of the maximum acceptance angle:

$$NA = \sin(\theta_a) = \sqrt{n_1^2 - n_2^2} \quad (5.1)$$

This equation relates the NA directly to the refractive indices of the core and cladding. The derivation of this formula involves applying Snell's law at the air-fiber interface and at the core-cladding interface, then using the

condition for total internal reflection.

In practice, typical values for the refractive indices might be  $n_1 = 1.475$  for the core and  $n_2 = 1.46$  for the cladding. Plugging these into our equation:

$$NA = \sqrt{1.475^2 - 1.46^2} \approx 0.2 \quad (5.2)$$

This means that light entering the fiber within a cone of about  $11.5^\circ$  ( $\arcsin(0.2)$ ) from the fiber's axis will be transmitted through the fiber via total internal reflection.

The NA is an important parameter in fiber optic design:

1. It determines the light-gathering ability of the fiber.
2. It affects the fiber's bandwidth and its susceptibility to certain types of signal distortion.
3. It influences how easily the fiber can be coupled to light sources and other fibers.

Optical fibers come in various types, each optimized for different applications. Some fibers are designed to transmit light over long distances with minimal loss, while others are engineered for specific wavelengths or to guide light in unusual ways. The figure below shows a few examples of optical fiber types.

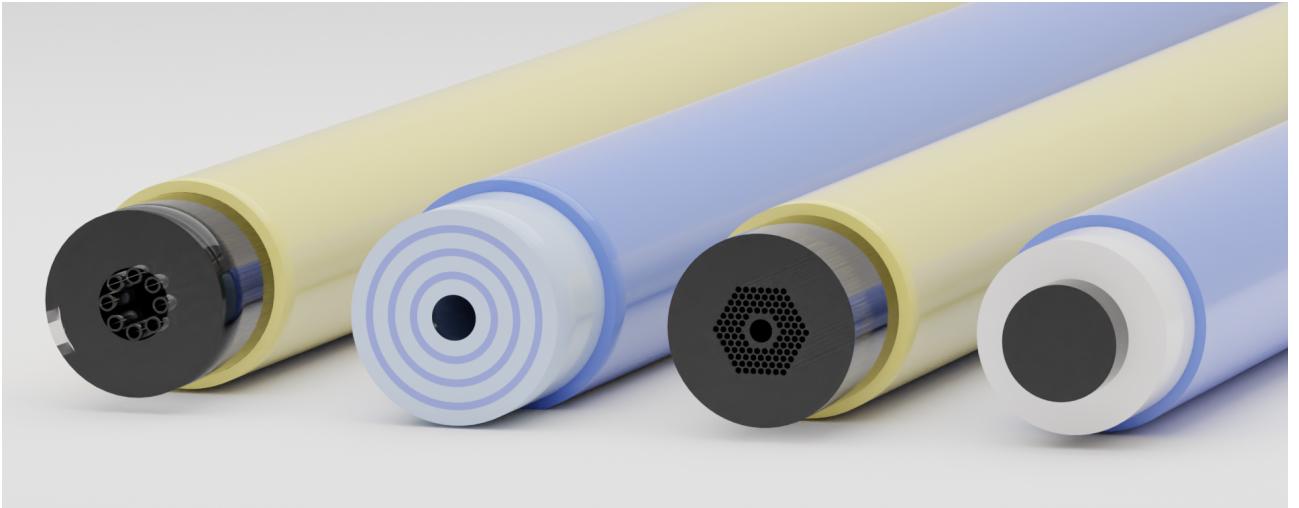


Figure 5.2: Rendering of different optical fibers types (from left to right): Hollow core optical fiber, hollow core bragg fiber, photonic crystal fiber, conventional fiber

### Fermat's Principle for Inhomogeneous Media

While before we have considered Fermat's principle for the special case of refraction and light propagation in a homogeneous medium, we can define a more general version of it corresponding to the following situation also involving an inhomogeneous refractive index  $n(\vec{r})$ .

For this general scenario of light traveling along a path, we can define an optical path length (OPL) as

$$\text{OPL} = \int_A^C n(\mathbf{r}) ds = 0, \quad (5.3)$$

with a varying refractive index  $n(\mathbf{r})$ . Fermat's Principle states that the actual path taken by the light makes the OPL stationary:

$$\delta \left( \int_A^B n(\mathbf{r}) ds \right) = 0$$

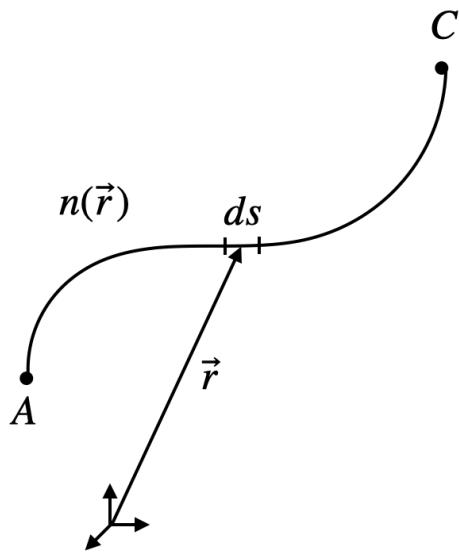


Figure 5.3: Sketch for a general description of Fermat's principle

Using the calculus of variations, this leads to the Euler-Lagrange equation for the path of light. In Cartesian coordinates, if the path is parameterized by  $\mathbf{r}(s) = (x(s), y(s), z(s))$ , the Euler-Lagrange equations become:

$$\frac{d}{ds} \left( n \frac{d\mathbf{r}}{ds} \right) = \nabla n$$

where  $\nabla n$  is the gradient of the refractive index. This equation describes how the light ray bends in response to changes in the refractive index of the medium.

Fermat's Principle is a cornerstone of geometrical optics and has applications in designing optical systems, understanding mirages, and analyzing the behavior of light in various media.

### Fermat's Principle and Snells Law

We would like to apply Fermat's principle to derive Snell's law, which is a more lengthy calculation. To do this, we consider a light ray traveling from point  $A$  in medium 1 (with refractive index  $n_1$ ) to point  $C$  in medium 2 (with refractive index  $n_2$ ), crossing the interface at point  $B$ . Let the coordinates of points  $A$ ,  $B$ , and  $C$  be  $(x_A, y_A)$ ,  $(x_B, y_B)$ , and  $(x_C, y_C)$ , respectively. Assume the interface between the two media is at  $y = y_B$ . The optical path length (OPL) is given by:

$$\delta \int_A^C n(\vec{r}) ds = 0,$$

where  $n(\vec{r})$  is the refractive index at position  $\vec{r}$ , and  $ds$  is an infinitesimal element of the path.

Consider a light ray traveling from point  $A$  in medium 1 (with refractive index  $n_1$ ) to point  $C$  in medium 2 (with refractive index  $n_2$ ), crossing the interface at point  $B$ . Let the coordinates of points  $A$ ,  $B$ , and  $C$  be  $(x_A, y_A)$ ,  $(x_B, y_B)$ , and  $(x_C, y_C)$ , respectively. Assume the interface between the two media is at  $y = y_B$ .

### Optical Path Length

The optical path length (OPL) is given by:

$$\text{OPL} = n_1 \int_A^B ds_1 + n_2 \int_B^C ds_2,$$

where  $ds_1$  and  $ds_2$  are the infinitesimal path lengths in media 1 and 2, respectively.

### Path Lengths

The path lengths  $ds_1$  and  $ds_2$  can be expressed in terms of the coordinates:

$$ds_1 = \sqrt{(dx_1)^2 + (dy_1)^2}, \quad ds_2 = \sqrt{(dx_2)^2 + (dy_2)^2}.$$

Since the interface is at  $y = y_B$ , we have  $dy_1 = y_B - y_A$  and  $dy_2 = y_C - y_B$ . The total optical path length is:

$$\text{OPL} = n_1 \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2} + n_2 \sqrt{(x_C - x_B)^2 + (y_C - y_B)^2}.$$

### Applying Fermat's Principle

To find the stationary path, we take the variation of the OPL with respect to  $x_B$ :

$$\delta \text{OPL} = \delta \left[ n_1 \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2} + n_2 \sqrt{(x_C - x_B)^2 + (y_C - y_B)^2} \right] = 0.$$

Taking the derivative with respect to  $x_B$ :

$$\frac{\partial}{\partial x_B} \left[ n_1 \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2} + n_2 \sqrt{(x_C - x_B)^2 + (y_C - y_B)^2} \right] = 0.$$

### Differentiating

Differentiating each term separately:

$$n_1 \frac{\partial}{\partial x_B} \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2} + n_2 \frac{\partial}{\partial x_B} \sqrt{(x_C - x_B)^2 + (y_C - y_B)^2} = 0.$$

Using the chain rule:

$$n_1 \frac{x_B - x_A}{\sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}} + n_2 \frac{x_B - x_C}{\sqrt{(x_C - x_B)^2 + (y_C - y_B)^2}} = 0.$$

### Simplifying

Let  $\theta_1$  be the angle of incidence and  $\theta_2$  be the angle of refraction. Then:

$$\sin \theta_1 = \frac{x_B - x_A}{\sqrt{(x_B - x_A)^2 + (y_B - y_A)^2}}, \quad \sin \theta_2 = \frac{x_C - x_B}{\sqrt{(x_C - x_B)^2 + (y_C - y_B)^2}}.$$

Substituting these into the equation:

$$n_1 \sin \theta_1 + n_2 \sin \theta_2 = 0.$$

Since  $\sin \theta_2$  is in the opposite direction, we have:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2.$$

This is Snell's law, which describes the relationship between the angles of incidence and refraction when light passes from one medium to another.

### Conclusion

By applying Fermat's principle and taking the variation of the optical path length, we have derived Snell's law:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2.$$

This demonstrates how the principle of least time leads to the well-known law of refraction in optics.

 Example: Light in a Graded-Index Medium

Consider a medium where the refractive index varies with height  $y$  as  $n(y) = n_0(1 - \frac{\alpha^2}{2n_0}y^2)$ . The path of light in such a medium can be found by using Fermat's principle in differential form:

$$\frac{d}{ds} \left( n(\mathbf{r}) \frac{d\mathbf{r}}{ds} \right) = \nabla n(\mathbf{r})$$

Typically, this requires to express the coordinates in terms of a parameter, such as  $x(s)$  and  $y(s)$ , and then solve the differential equation. The solution will give the path of light in the medium. This is difficult and commonly done numerically. In paraxial optics, when the light is propagating roughly in the direction of  $z$ , the differential element  $ds$  can be approximated as  $dz$  since then

$$ds = dz \sqrt{1 + \left( \frac{dy}{dz} \right)^2 + \left( \frac{dx}{dz} \right)^2} \approx dz$$

which yields

$$\frac{d}{dz} \left( n \frac{dx}{dz} \right) \approx \frac{dn}{dx}$$

and

$$\frac{d}{dz} \left( n \frac{dy}{dz} \right) \approx \frac{dn}{dy}$$

This readily yields the path of light in a homogeneous medium, where  $n$  is constant. In this case we have

$$\frac{d^2x}{dz^2} = \frac{d^2y}{dz^2} = 0$$

which is true for a straight line. In a graded-index medium, the path of light can be found by solving the differential equation

$$\frac{d^2y}{dz^2} = -\alpha^2 y$$

which is reminiscent of the equation of motion of a harmonic oscillator. The solution is therefore an oscillating function

$$y(z) = y_0 \cos(\alpha z) + \frac{\theta_0}{\alpha} \sin(\alpha z)$$

where the angle  $\theta_0$  is the initial angle of the light ray with respect to the  $z$  axis. This solution describes the path of light in a graded-index medium.

 Fermat's Principle in Integral and Differential Form

We have described Fermat's principle in an integral form specifying the optical path length  $S$  as

$$OPL = \int n(\mathbf{r}) ds$$

The path length  $ds$  can be given in terms of two coordinates  $x_1$  and  $x_2$  parametrized by  $\lambda$  such that

$$ds = \sqrt{\dot{x}_1^2 + \dot{x}_2^2} d\lambda$$

where  $\dot{x}_1 = \frac{dx_1}{d\lambda}$ . We can therefore write Fermat's principle as

$$\delta OPL = \int \left[ \left( \frac{\partial n}{\partial x_i} \delta x_i \right) \sqrt{\dot{x}_1^2 + \dot{x}_2^2} + n \frac{1}{\sqrt{\dot{x}_1^2 + \dot{x}_2^2}} \dot{x}_i \delta x_i \right] d\lambda = 0$$

To evaluate this integral we would like to integrate by parts. We can write the integrand as

$$u = n \frac{1}{\sqrt{\dot{x}_1^2 + \dot{x}_2^2}} \dot{x}_i$$

and

$$dv = \delta x_i d\lambda$$

We can now calculate  $du$  and  $v$  and obtain

$$du = \frac{d}{d\lambda} \left[ n \frac{1}{\sqrt{\dot{x}_1^2 + \dot{x}_2^2}} \dot{x}_i \right] d\lambda$$

and

$$v = \delta x_i$$

With these expressions we can now apply the integration by parts formula  $\int u dv = uv - \int v du$ , we get:

$$\int n \frac{1}{\sqrt{\dot{x}_1^2 + \dot{x}_2^2}} \dot{x}_i \delta x_i d\lambda = \left[ n \frac{1}{\sqrt{\dot{x}_1^2 + \dot{x}_2^2}} \dot{x}_i \delta x_i \right] \Big|_{\lambda_1}^{\lambda_2} - \int \delta x_i \frac{d}{d\lambda} \left[ n \frac{1}{\sqrt{\dot{x}_1^2 + \dot{x}_2^2}} \dot{x}_i \right] d\lambda$$

This can be substituted back into the original equation to obtain

$$\begin{aligned} \delta OPL &= \int \left[ \left( \frac{\partial n}{\partial x_i} \delta x_i \right) \sqrt{\dot{x}_1^2 + \dot{x}_2^2} \right] d\lambda \\ &\quad + \left[ n \frac{1}{\sqrt{\dot{x}_1^2 + \dot{x}_2^2}} \dot{x}_i \delta x_i \right] \Big|_{\lambda_1}^{\lambda_2} \\ &\quad - \int \delta x_i \frac{d}{d\lambda} \left[ n \frac{1}{\sqrt{\dot{x}_1^2 + \dot{x}_2^2}} \dot{x}_i \right] d\lambda = 0 \end{aligned}$$

After rearranging the terms we get

$$\begin{aligned} \delta OPL &= \int \left\{ \frac{\partial n}{\partial x_i} \sqrt{\dot{x}_1^2 + \dot{x}_2^2} - \frac{d}{d\lambda} \left[ n \frac{1}{\sqrt{\dot{x}_1^2 + \dot{x}_2^2}} \dot{x}_i \right] \right\} \delta x_i d\lambda \\ &\quad + \left[ n \frac{1}{\sqrt{\dot{x}_1^2 + \dot{x}_2^2}} \dot{x}_i \delta x_i \right] \Big|_{\lambda_1}^{\lambda_2} = 0 \end{aligned}$$

and therefore finally

$$\delta OPL = \int \left[ \left( \frac{\partial n}{\partial x_i} \right) \sqrt{\dot{x}_1^2 + \dot{x}_2^2} - \frac{d}{d\lambda} \left( n \frac{1}{\sqrt{\dot{x}_1^2 + \dot{x}_2^2}} \dot{x}_i \right) \right] \delta x_i d\lambda + \text{boundary terms}$$

for which we choose the parameter  $\lambda$  such that the boundary terms vanish.

$$\lambda = s$$

such that

$$\sqrt{\dot{x}_1^2 + \dot{x}_2^2} = 1$$

and finally leads to the Euler-Lagrange equation

$$\left( \frac{\partial n}{\partial x_i} \right) - \frac{d}{ds} (n \dot{x}_i) = 0$$

which is the differential form of the Fermat's principle.



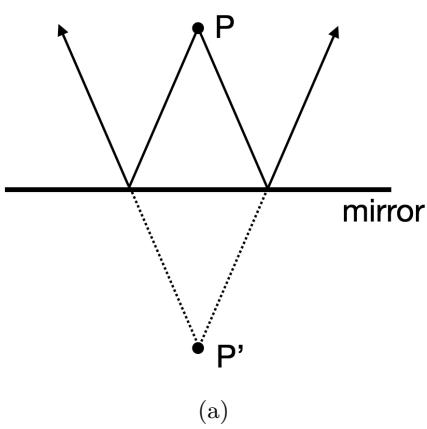
# Chapter 6

## Optical Elements Part I

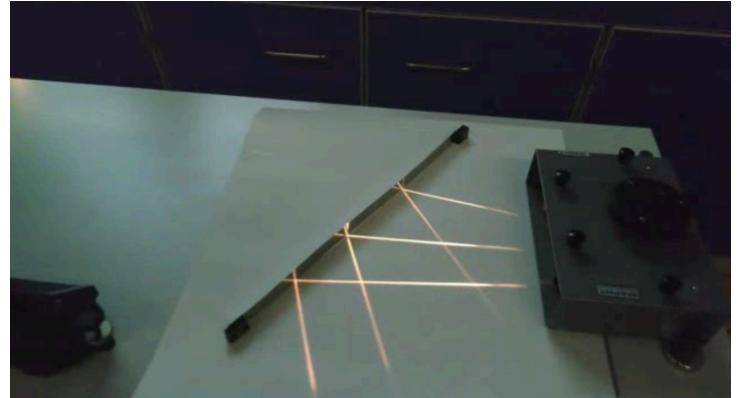
### 6.1 Mirrors

#### Plane Mirrors

When light radiates from a point  $P$  and reflects off a mirror, as shown in the image, the reflected rays diverge but appear to originate from a point  $P'$  located behind the mirror. According to the law of reflection, this image point is positioned at the same distance behind the mirror as the original object point is in front of it. As a result, an observer receiving these reflected rays, such as on their retina, perceives the point as if it were situated behind the mirror, even though no light actually travels behind the mirror surface.



(a)



(b)

Figure 6.1: Image formation on a plane mirror.

When multiple points of an object emit light towards the mirror, this principle applies to each point. As a result, the entire object appears as an image behind the mirror. Since each point of the image is equidistant from the mirror as its corresponding object point, the image has the same size as the object. This leads to the definition of magnification as:

$$M = \frac{h_{\text{image}}}{h_{\text{object}}}$$

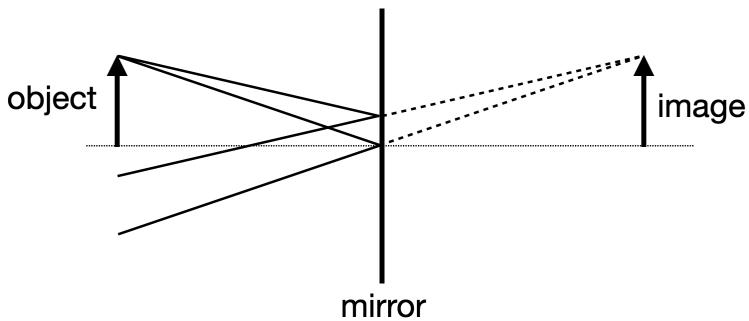


Figure 6.2: Image formation on a plane mirror.

### **i** Virtual Images

A virtual image is an optical illusion where light rays appear to come from a point, but don't actually converge there. Unlike real images, virtual images can't be projected onto a screen. They're commonly seen in plane mirrors, convex mirrors, and when objects are closer to a lens than its focal point. **Remember:** for virtual images, light rays only seem to originate from the image when traced backwards.

### **i** Real Images

A real image forms when light rays actually meet at a point after reflection or refraction. These images can be projected onto a screen because light physically passes through the image location. Real images are often inverted and occur with concave mirrors and lenses when objects are beyond the focal point. **Key point:** real images involve actual convergence of light rays.

## Concave Mirrors

For a concave mirror (where the reflecting surface is on the inside of the spherical curve), applying the law of reflection yields interesting results. Light rays parallel to the optical axis, at a distance  $h$  from it, are reflected towards the axis and intersect it at a specific point  $F$ . Due to the mirror's symmetry, a parallel ray on the opposite side of the axis will also converge to this same point  $F$ .

We may calculate the position of the point  $F$ , e.g. the distance from the mirror surface point  $O$ , by applying the law of reflection. If the spherical mirror surface has a radius  $R$ , then the distance between the center of the sphere  $M$  and the point  $F$  is given by

$$FM = \frac{R}{2 \cos(\alpha)}$$

Therefore, we can also calculate the distance of the mirror surface from the point  $F$ , which results in

$$OF = R \left( 1 - \frac{1}{2 \cos(\alpha)} \right) = f \quad (6.1)$$

This distance is the so-called focal length of the concave mirror  $f$ . For small angle  $\alpha$ , the above equation yields the so called paraxial limit (all angles are small and the rays are close to the optical axis). In this limit we find  $\cos(\alpha) \approx 1$  and the focal length becomes  $f = R/2$ . If we replace the cosine function by  $\cos(\alpha) = \sqrt{1 - \sin^2(\alpha)}$  with  $\sin(\alpha) = h/R$ , we find

$$f = R \left[ 1 - \frac{R}{2\sqrt{R^2 - h^2}} \right] \quad (6.2)$$

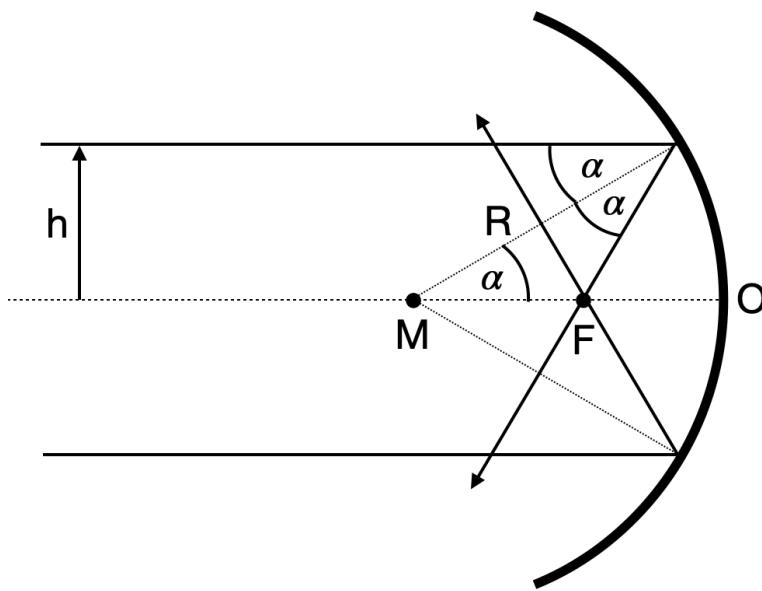


Figure 6.3: Reflection of a parallel ray incident at a height  $h$  from the optical axis on a concave mirror.

This equation is telling us, that the focal distance is not a single value for a concave mirror. The focal distance rather changes with the distance  $h$  from the optical axis. If  $h$  approaches  $R$  the focal length become shorter.

#### Focal Length of a Concave Spherical Mirror

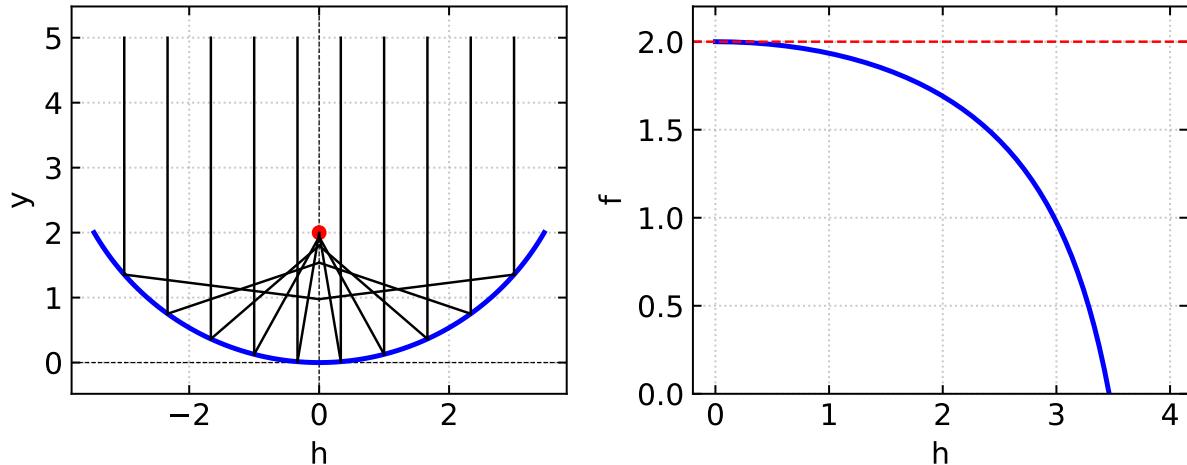


Figure 6.4: Spherical mirror of radius  $R = 4$  reflecting parallel rays, showing spherical aberration and focal distance as a function of the distance from the optical axis  $h$ .

To obtain now an equation which predicts the point at which the reflected ray intersects the optical axis if it emerged at a point  $A$ , we just consider the following sketch.

For this situation, we can write down immediately the following relations

$$\delta = \alpha + \gamma$$

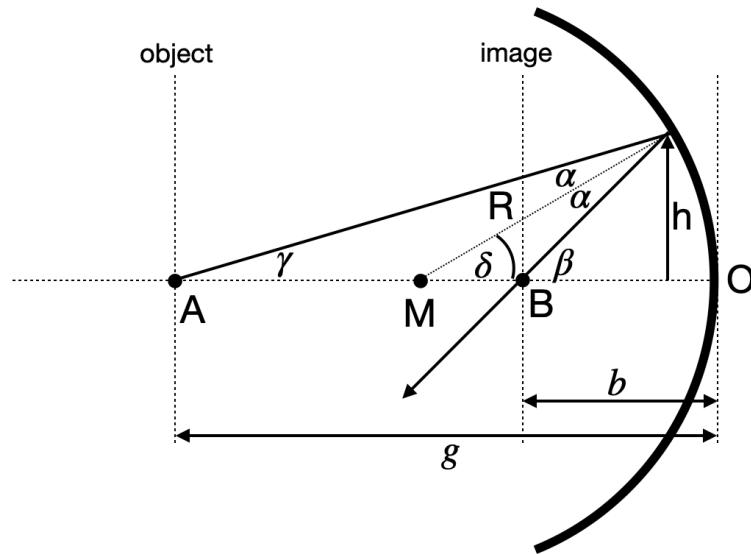


Figure 6.5: Image formation on a concave mirror.

$$\gamma + \beta = 2\delta$$

Further under the assumption of small angles (paraxial approximation) we can write down

$$\tan(\gamma) \approx \gamma = \frac{h}{g}$$

$$\tan(\beta) \approx \beta = \frac{h}{b}$$

$$\sin(\delta) \approx \delta = \frac{h}{R}$$

from which we obtain

$$\frac{h}{g} + \frac{h}{b} = 2\frac{h}{R}$$

and by dividing by  $h$  finally the imaging equation:

$$\frac{1}{g} + \frac{1}{b} = \frac{2}{R} = \frac{1}{f}$$

where we have used the focal length  $f = R/2$ . This equation has some surprising property. It is completely independent of  $h$  and  $\gamma$ . That means all points in a plane at a distance  $g$  are images into a plane at a distance  $b$ . Both planes are therefore called conjugated planes.

### Imaging Equation Concave Mirror

The sum of the inverse object and image distances equals the inverse focal length of the concave mirror.

$$\frac{1}{g} + \frac{1}{b} \approx \frac{1}{f}$$

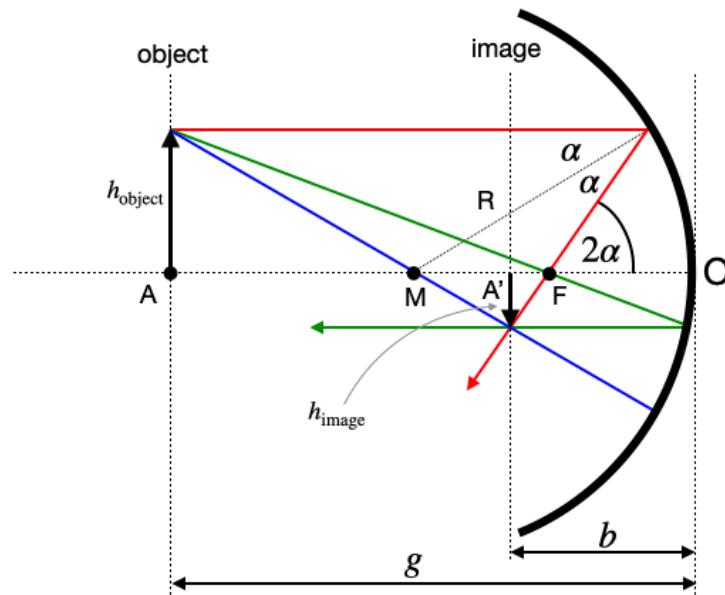


Figure 6.6: Image formation on a concave mirror.

This equation now helps to construct the image of an object in front of a concave mirror and we may define 3 different rays to identify the size of an image  $h_{\text{image}}$  from the size of an object  $h_{\text{object}}$ .

In the diagram above, three key rays are used to construct the image:

1. **Red ray:** Parallel to optical axis → reflects through focal point
2. **Green ray:** Through focal point → reflects parallel to optical axis
3. **Central ray:** Through center of curvature → reflects back along same path

The behavior of these reflected rays determines the nature of the image:

- If the rays intersect on the same side of the mirror as the object, a **real image** forms. This image is inverted, as shown in the sketch.
- If the rays diverge after reflection, they appear to intersect behind the mirror, creating a **virtual image**. This image is upright and located behind the mirror, though no actual ray intersection occurs.

The point where these rays meet (or appear to meet) determines the image size. By drawing a ray from the object's tip through the mirror's center (point O), we can easily determine the image height  $h_{\text{image}}$ . As an exercise, consider how this construction demonstrates that the magnification of a concave mirror is given by

$$\frac{h_{\text{image}}}{h_{\text{object}}} = -\frac{b}{g} = M$$

This ratio indeed represents the magnification  $M$ . The negative sign in the expression reflects an important optical property: for real images formed by concave mirrors, the image is inverted relative to the object. This inversion is mathematically represented by the negative magnification value. Conversely, a positive magnification would indicate an upright image, which occurs with virtual images.

With the help of the imaging equation and the magnification we may in general differentiate between the following general situations:

Object Distance	Image Characteristics	Image Position	Magnification
$g > 2f$	Real, inverted, smaller	Between f and 2f	$ m  < 1$
$g = 2f$	Real, inverted, same size	At 2f	$ m  = 1$

Object Distance	Image Characteristics	Image Position	Magnification
$f < g < 2f$	Real, inverted, larger	Beyond 2f	$ m  > 1$
$g = f$	Image at infinity	At infinity	N/A
$g < f$	Virtual, upright, larger	Behind mirror	$ m  > 1$

### i Parabolic Mirrors Focus Parallel Rays

We would like to show in the following, that a parabolic mirror is a shape which reflects all light rays parallel to the principal axis to a single point, the focus. This is a fundamental property of parabolic mirrors and is used in many optical systems, such as telescopes, satellite dishes, and car headlights. For this purpose, we would like to use Fermat's principle. We examine a light ray originating from a point  $x, y_0$  and travelling parallel to the principal axis. The light ray is reflected at a point  $(x, y)$  on the mirror and travels to the focus at  $(0, p)$ . The light path is therefore consisting out of two linear segments A and B for which we have to calculate the time of travel. The total duration of the light's journey is then:

$$t = t_A + t_B$$

where:

- $t_A$  is the time taken to travel from  $x, y_0$  to the mirror.
- $t_B$  is the time taken to travel from  $(x, y)$  to  $(0, p)$ .

#### Time for Path A

The distance covered in path A is equal to  $y_0 - y$ , where  $y$  represents the y-coordinate of the point where the ray meets the mirror. Consequently, the time taken for the light to traverse path A can be expressed as:

$$t_A = \frac{y_0 - y}{c}$$

In this equation,  $c$  represents the speed of light in the medium.

#### Time for Path B

After reflection, the light ray travels from the point  $(x, y)$  on the mirror's surface to the focal point located at  $(0, p)$ . The length of this segment of the path can be calculated using the distance formula:

$$\sqrt{x^2 + (y - p)^2}$$

Consequently, the time required for the light to traverse path B is expressed as:

$$t_B = \frac{\sqrt{x^2 + (y - p)^2}}{c}$$

#### Total Time

The total time for the light ray's journey is the sum of times for paths A and B:

$$t = \frac{y_0 - y}{v} + \frac{\sqrt{x^2 + (y - p)^2}}{c}$$

According to Fermat's principle, all light rays should take the same time. We can express this by setting the total time equal to a constant  $t_c$ :

$$\frac{y_0 - y}{v} + \frac{\sqrt{x^2 + (y - p)^2}}{v} = t_c$$

For a ray traveling along the y-axis, reflecting at  $(0, 0)$ , the total distance is  $y_0 + p$ . The time for this ray is:

$$\frac{y_0 + p}{c}$$

This gives us  $t_c = \frac{y_0 + p}{c}$ . Substituting into our general equation:

$$\frac{y_0 - y}{v} + \frac{\sqrt{x^2 + (y - p)^2}}{c} = \frac{y_0 + p}{c}$$

Multiplying by  $c$  and rearranging:

$$y_0 - y + \sqrt{x^2 + (y - p)^2} = y_0 + p$$

$$\sqrt{x^2 + (y - p)^2} = y + p$$

Squaring both sides and simplifying:

$$x^2 + (y - p)^2 = (y + p)^2$$

$$x^2 + y^2 - 2py + p^2 = y^2 + 2py + p^2$$

$$x^2 = 4py$$

or

$$y = \frac{1}{4p}x^2$$

This final equation describes a parabola with its focus at  $(0, p)$ . The code below plots a parabolic mirror reflecting parallel rays to the focal point. Yet, I'm cheating a bit here. I'm not calculating the reflected rays, but just plotting them.

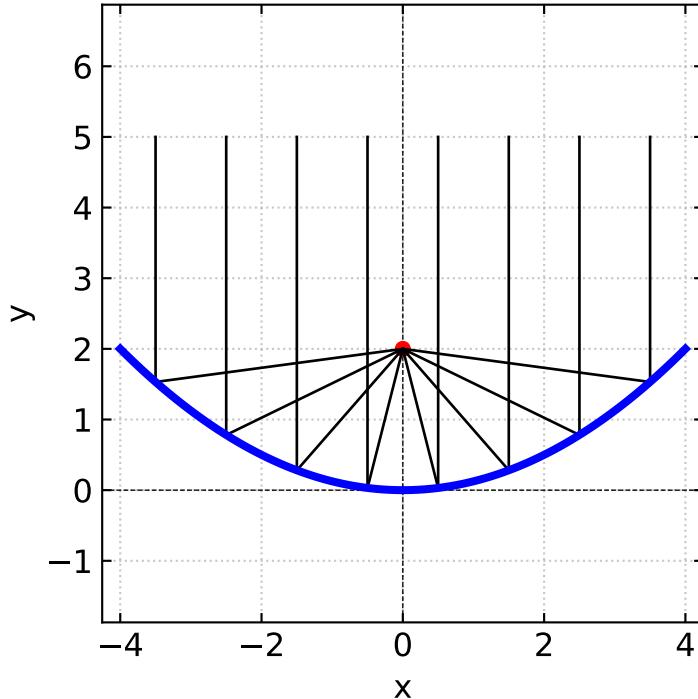


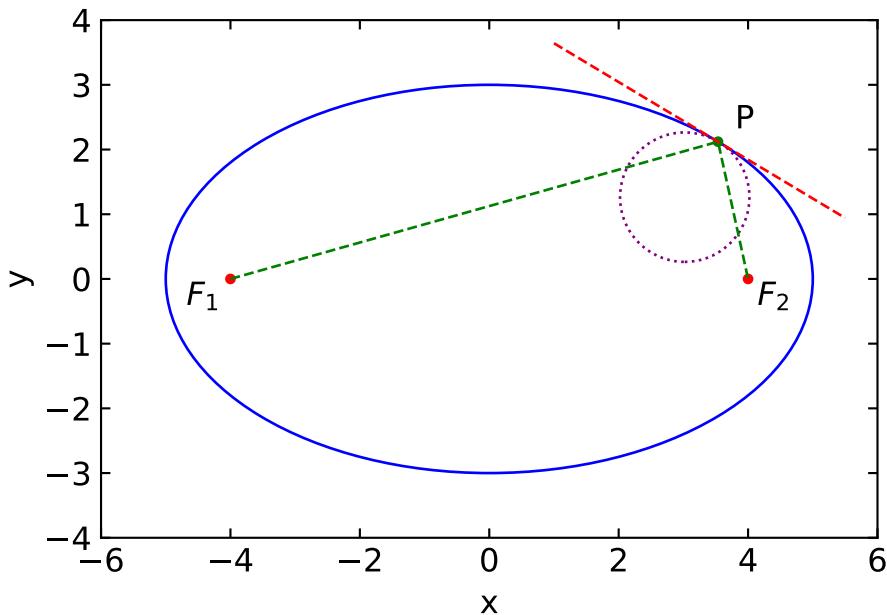
Figure 6.7: Parabolic mirror reflecting parallel rays to focal point

### Elliptical Mirrors and Fermat's Principle

There is one interesting feature about elliptical mirrors: they can focus light from one focal point to the other. This is because the sum of the distances from any point on the ellipse to the two focal points is constant. This property is known as the **ellipse's geometric definition** and you can try that at home with a piece of string and two pins.

We can now apply Fermat's principle to proof that the light reflected from the ellipse travels a path length that is a saddle point. This means that the path length is stationary with respect to small perturbations in the path. Assuming for example that light travels from one focal point by a different path that is reflected from a line which is tangent to the ellipse at the point of reflection, the path length would be longer at any other point than the initial reflection point.

On the other side, if we reflect the ray on a surface that is a circle, which is intersecting the ellipse at the point of reflection, the path length would be shorter at any other point than the initial reflection point. This is a proof that the ellipse is a saddle point.



## Mathematical Description

### Ellipse Definition

Consider an ellipse with semi-major axis  $a$  and semi-minor axis  $b$ , defined by:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

### Focal Points

The focal points are located at  $F_1(-c, 0)$  and  $F_2(c, 0)$ , where:

$$c^2 = a^2 - b^2$$

### Path Length

Let  $P(x_0, y_0)$  be a point on the ellipse. The total path length  $L$  from  $F_1$  to  $F_2$  via  $P$  is:

$$L = |F_1P| + |PF_2| = \sqrt{(x_0 + c)^2 + y_0^2} + \sqrt{(x_0 - c)^2 + y_0^2}$$

**Fermat's Principle**

The path length  $L$  is stationary with respect to small perturbations in  $P$ :

$$\frac{\partial L}{\partial x_0} = 0 \quad \text{and} \quad \frac{\partial L}{\partial y_0} = 0 \quad \text{at the reflection point}$$

**Tangent Line**

The tangent line to the ellipse at  $P(x_0, y_0)$  is given by:

$$\frac{x_0 x}{a^2} + \frac{y_0 y}{b^2} = 1$$

Let  $Q$  be any point on this tangent line different from  $P$ . The path  $F_1 \rightarrow Q \rightarrow F_2$  is longer than  $F_1 \rightarrow P \rightarrow F_2$ :

$$|F_1 Q| + |Q F_2| > |F_1 P| + |P F_2|$$

**Circle of Curvature**

The radius of curvature  $R$  at  $P$  is:

$$R = \frac{(a^2 b^2)^{3/2}}{(b^2 x_0^2 + a^2 y_0^2)^{3/2}}$$

The center of curvature  $C$  is located at:

$$C = P + R \cdot \mathbf{n}$$

where  $\mathbf{n}$  is the unit normal vector at  $P$ .

Let  $Q$  be any point on this circle different from  $P$ . The path  $F_1 \rightarrow Q \rightarrow F_2$  is shorter than  $F_1 \rightarrow P \rightarrow F_2$ :

$$|F_1 Q| + |Q F_2| < |F_1 P| + |P F_2|$$

As a consequence, the path length for the reflection on an ellipse between the two focal points must be a saddle point.



# Chapter 7

## Optical Elements Part II

### 7.1 Prism

Prisms are wedge-shaped optical elements made of a transparent material, such as glass. A special form of such a prism is an isosceles prism with two sides of equal length. The two equal sides enclose an angle  $\gamma$ , known as the apex angle of the prism. When light passes through this prism, it undergoes refraction twice.

First, the incident angle  $\alpha_1$  is changed into a refracted angle  $\beta_1$  as the light enters the prism. This refracted ray then hits the second interface at an angle  $\beta_2$ , leading to a second refraction as it exits the prism at an angle  $\alpha_2$ .

Of particular interest is the total deflection of the incident ray, which is measured by the angle  $\delta$ . This deflection angle represents the difference between the final outgoing angle  $\alpha_2$  and the initial incident angle  $\alpha_1$ .

Understanding how this deflection angle changes based on the prism's properties and the incident angle is crucial in various optical applications. In the following sections, we will explore how to calculate this deflection angle and examine its dependence on different parameters.

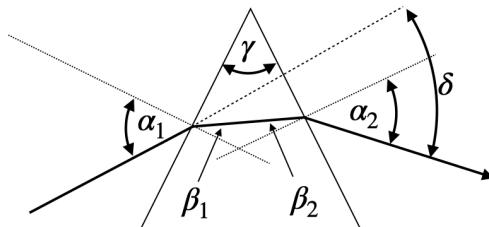


Figure 7.1: Refraction of rays on a prism.

### Deflection angle

We can calculate the deflection angle  $\delta$  from a number of considerations. First consider the following relations between the angles in the prism and Snell's law

$$\beta_1 = \sin^{-1} \left( \frac{n_0}{n_1} \sin(\alpha_1) \right)$$

$$\beta_2 = \gamma - \beta_1$$

$$\alpha_2 = \sin^{-1} \left( \frac{n_1}{n_0} \sin(\beta_2) \right)$$

$$\theta_2 = \alpha_2 - \gamma$$

where  $\theta_2$  is the angle between the incident surface normal and the outgoing ray. The total deflection angle  $\delta$  is then

$$\delta = \alpha_1 - \beta_1 + \alpha_2 - \beta_2$$

or

$$\delta = \alpha_1 + \alpha_2 - \gamma$$

from which we obtain

$$\delta = \alpha_1 + \sin^{-1} \left( \frac{n_1}{n_0} \sin \left[ \gamma - \sin^{-1} \left( \frac{n_0}{n_1} \sin(\alpha_1) \right) \right] \right) - \gamma$$

as the deflection angle.

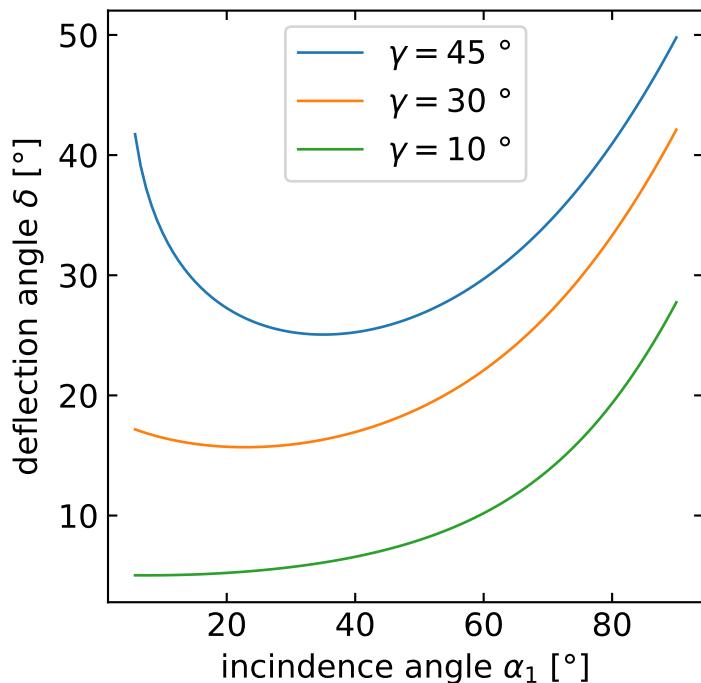


Figure 7.2: Deflection angle as a function of the incidence angle for different prism angles.

### Minimum deflection angle

If we now would like to know how the deflection angle changes with the incident angle  $\alpha_1$ , we calculate the derivative of the deflection angle  $\delta$  with respect to  $\alpha_1$ , i.e.,

$$\frac{d\delta}{d\alpha_1} = 1 + \frac{d\alpha_2}{d\alpha_1}.$$

We are here especially interested in the case, where this change in deflection is reaching a minimum, i.e.,  $d\delta/d\alpha_1 = 0$ . This readily yields

$$d\alpha_2 = -d\alpha_1.$$

This means a change in the incidence angle  $d\alpha_1$  yields an opposite change in the outgoing angle  $-d\alpha_2$ . We may later observe that in the experiment.

As both, the incident and the outgoing angle are related to each other by Snells's law, we may introduce the derivatives of Snell's law for both interfaces, e.g.,

- $\cos(\alpha_1)d\alpha_1 = n \cos(\beta_1)d\beta_1$
- $\cos(\alpha_2)d\alpha_2 = n \cos(\beta_2)d\beta_2$

where  $n$  is the refractive index of the prism material and the material outside is air ( $n_{\text{air}} = 1$ ). Replacing  $\cos(\alpha) = \sqrt{1 - \sin^2(\alpha)}$  and dividing the two previous equations by each other readily yields

$$\frac{1 - \sin^2(\alpha_1)}{1 - \sin^2(\alpha_2)} = \frac{n^2 - \sin^2(\alpha_1)}{n^2 - \sin^2(\alpha_2)}.$$

The latter equation is for  $n \neq 1$  only satisfied if  $\alpha_1 = \alpha_2 = \alpha$ . In this case, the light path through the prism must be symmetric and we may write down the minimum deflection angle  $\delta_{\min}$ :

### Minimum prism deflection

The minimum deflection angle of an isosceles prism with a prism angle  $\gamma$  is given by

$$\delta_{\min} = 2\alpha - \gamma.$$

Given this minimum deflection angle  $\delta_{\min}$  and the properties of the prism, we may also write down Snell's law using  $\sin(\alpha) = n \sin(\beta)$ , which results in

$$\sin\left(\frac{\delta_{\min} + \gamma}{2}\right) = n \sin\left(\frac{\gamma}{2}\right).$$

which indicates the dependence of the deflection in the refractive index  $n$  of the prism material.

## Dispersion

Very important applications now arise from the fact, that the refractive index is a material property, which depends on the color (frequency or wavelength) of light. We do not yet understand the origin of this dependence. The plots below show the wavelength dependence of three different glasses. You may find much more data on the refractive index of different materials in an [online database](#).

The plots have a general feature, which is that the refractive index is largest at small wavelength (blue colors), while it drops continuously with increasing wavelength towards the red (800 nm). If you would characterize the dependence by the slope, i.e.,  $dn/d\lambda$  then all displayed curves show in the visible range

- $\frac{dn}{d\lambda} < 0$ , is called normal dispersion

while

- $\frac{dn}{d\lambda} > 0$ , is called anomalous dispersion

This wavelength dependence of the refractive index will yield a dependence of the deflection angle on the color of light as well. The change of the deflection angle with the refractive index can be calculated to be

$$\frac{d\delta}{dn} = \frac{2 \sin(\gamma/2)}{\sqrt{1 - n^2 \sin^2(\gamma/2)}}$$

together with the relation

$$\frac{d\delta}{d\lambda} = \frac{d\delta}{dn} \frac{dn}{d\lambda}$$

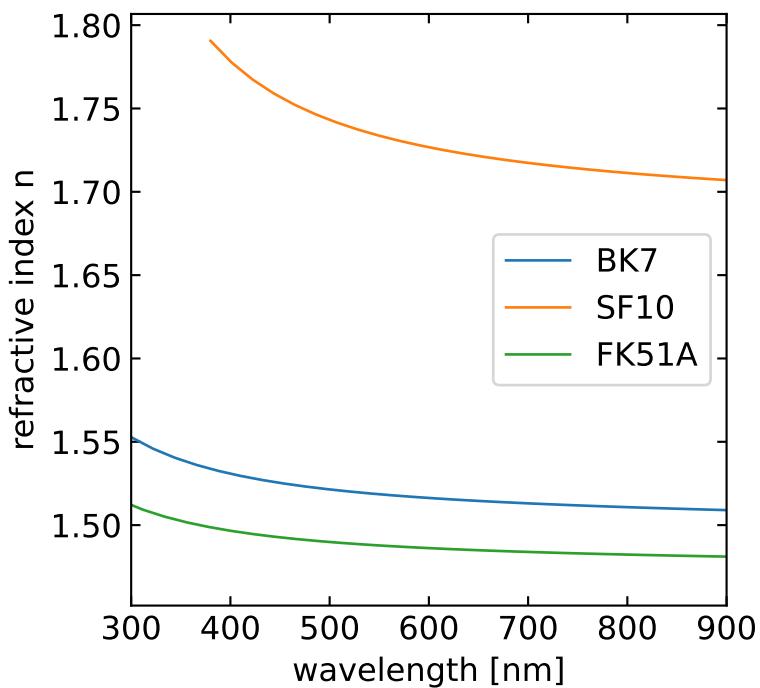


Figure 7.3: Refractive index of different glasses as a function of the wavelength.

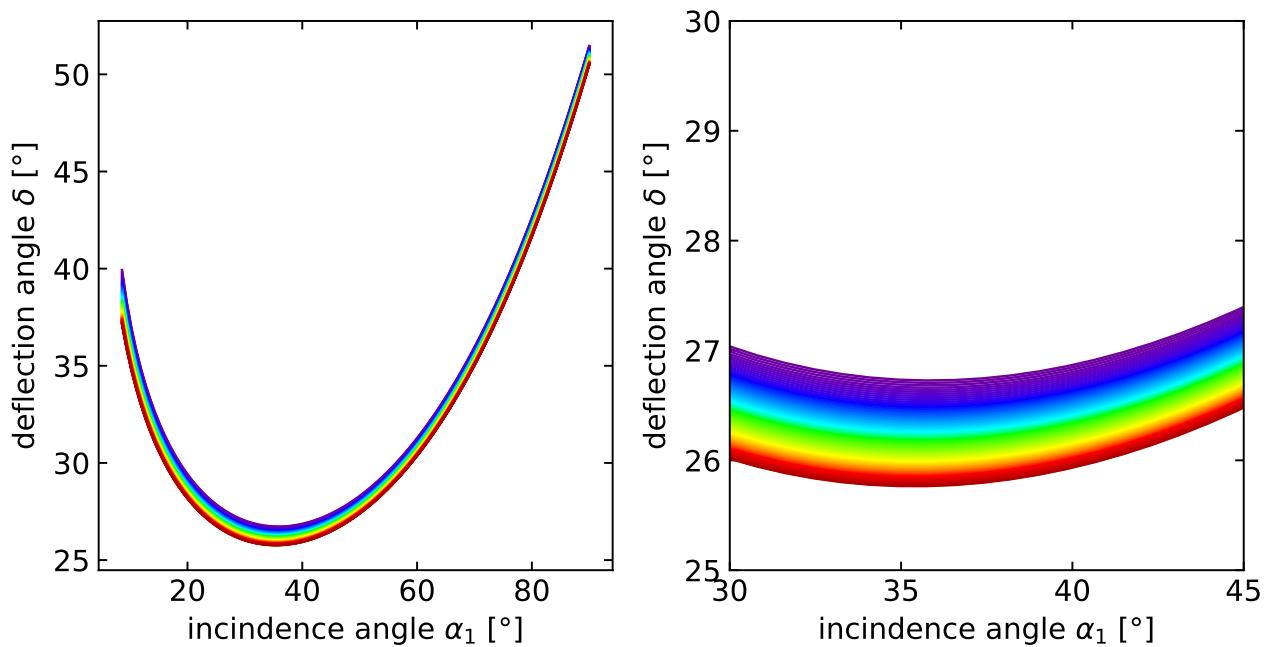


Figure 7.4: Deflection angle as a function of the incidence angle for different wavelengths.

we obtain

$$\frac{d\delta}{d\lambda} = \frac{2 \sin(\gamma/2)}{\sqrt{1 - n^2 \sin^2(\gamma/2)}} \frac{dn}{d\lambda}.$$

The refraction of white light through a prism splits the different colors composing white light spatially into a colored spectrum. In this process, light with the longest wavelength (red) is deflected the least, while light with the shortest wavelength (violet) is deflected the most. This occurs because the refractive index of the prism material varies with wavelength, a phenomenon known as dispersion.

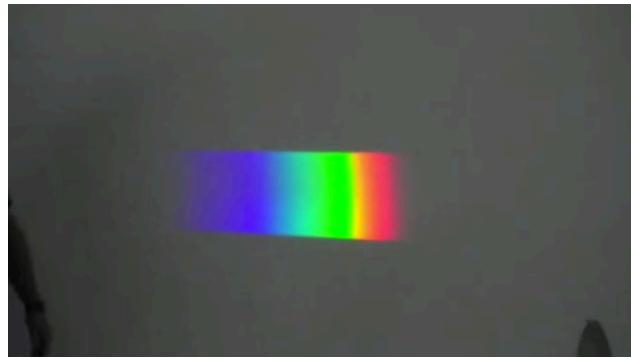
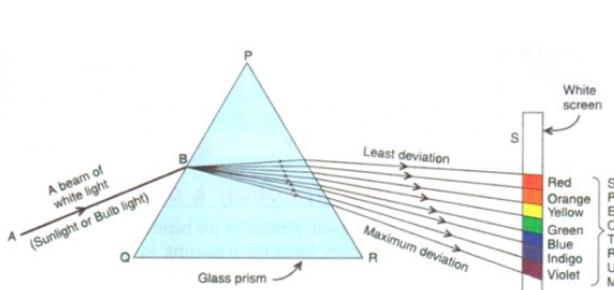


Figure 7.5: Spectrum as created by a prism in the lecture.



(a) Spectrum



(b) Prism

Figure 7.6: Deflection of different wavelengths of light in a prism with normal dispersion.

## Prisms spectrograph

This wavelength-dependent refraction is crucial as it forms the basis for spectroscopy, a powerful analytical technique that measures and records the intensity of light as a function of wavelength. Spectroscopy allows scientists to analyze the composition and properties of matter by examining its interaction with light across different wavelengths.

### DIY prism

If you don't have a prism at home (which most people don't), you can create a simple substitute using a mirror and a basin of water. Here's how:

1. Place a mirror in a basin of water, partially submerged.
2. Shine white light from a flashlight onto the mirror.
3. Observe the reflected and refracted light, paying special attention to the edges.

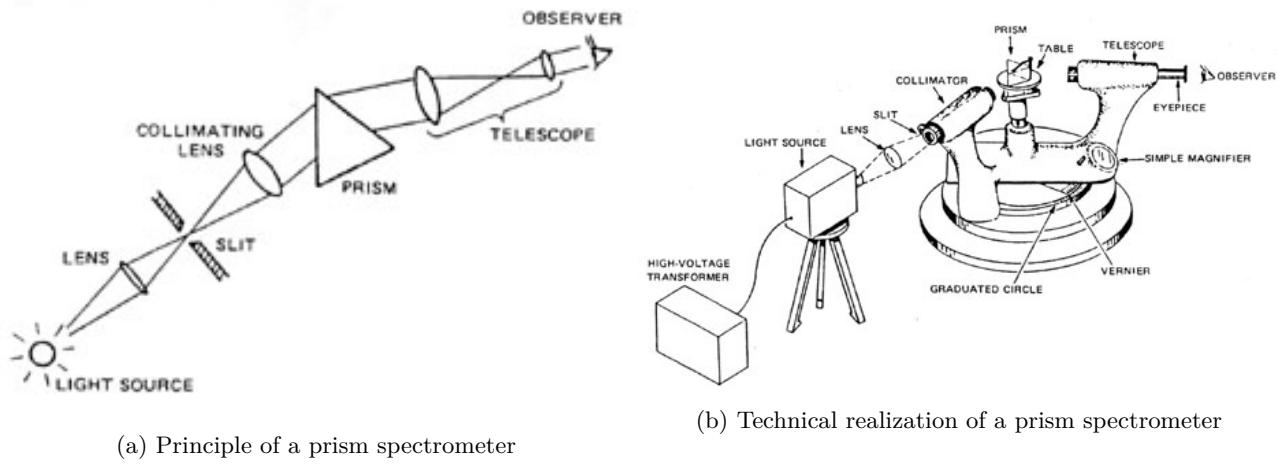


Figure 7.7: Principle and technical realization of a prism spectrometer.

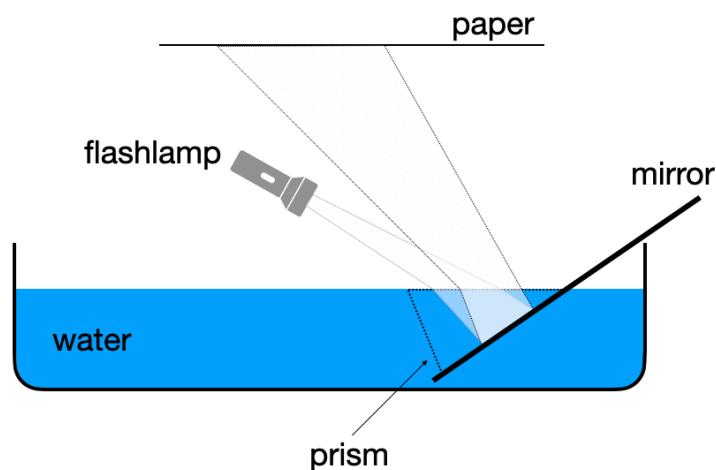


Figure 7.8: Home made water prism.

For better results, you can create a small aperture by making a tiny hole in a piece of black paper and placing it in front of the flashlight.

While the dependence of water's refractive index on wavelength is relatively weak, it's still sufficient to demonstrate the familiar colors of the rainbow. This phenomenon will be referenced later in our discussion.

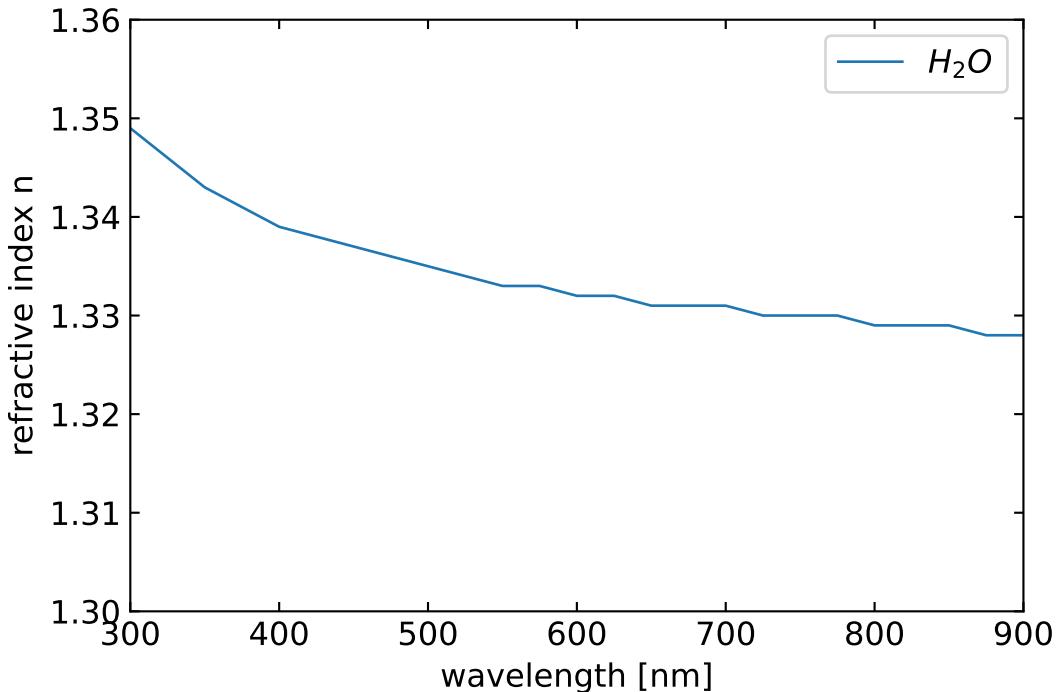


Figure 7.9: Refractive index of water as a function of the wavelength.

### **i Applications of prisms**

Prisms are versatile optical components with a wide range of applications across various fields. Here are some common uses of prisms:

#### **Binoculars and Telescopes:**

Porro prisms in traditional binoculars and roof prisms in modern designs serve to correct image inversion and provide a compact form. These prisms enable a longer optical path within a shorter physical length, enhancing magnification while maintaining portability. This design is crucial for both binoculars and some telescopes, offering users powerful magnification in a handheld device.

#### **Periscopes:**

Right-angle prisms are the key component in periscopes, redirecting light at 90-degree angles. This simple yet effective design allows viewers to see over obstacles or around corners, making periscopes invaluable in submarines and various military applications where direct line of sight is obstructed.

#### **Beam Splitting:**

Cube beamsplitters play a vital role in dividing a single beam of light into two separate beams. This capability is essential in various scientific and medical applications, including interferometry, holography, and optical coherence tomography (OCT). The ability to split light beams precisely opens up numerous possibilities in research and diagnostics.

**Beam Steering:**

Risley prisms, consisting of a pair of rotating wedge prisms, offer precise control over laser beam direction. This technology finds applications in laser scanning, target tracking, and adaptive optics. The ability to steer beams accurately is crucial in fields ranging from military applications to advanced scientific research.

**Digital Projectors:**

Total Internal Reflection (TIR) prisms are a crucial component in Digital Light Processing (DLP) projectors. They direct light from the lamp to the Digital Micromirror Device (DMD) and then to the projection lens, enabling the high-quality image projection that DLP technology is known for.

**Camera Systems:**

In Single-Lens Reflex (SLR) cameras, pentaprism play a critical role in the viewfinder system. They flip the image from the lens to appear upright and correctly oriented in the viewfinder, allowing photographers to accurately compose their shots.

**Laser Systems:**

Brewster prisms find use in laser systems for polarization and wavelength separation. Additionally, dispersing prisms can be employed for wavelength tuning in certain laser setups, providing precise control over the laser's output characteristics.

**Fiber Optic Communications:**

In the realm of telecommunications, prisms are utilized in some fiber optic connectors and switches. They help redirect light between fibers, playing a crucial role in maintaining signal integrity and enabling complex network architectures.

**Solar Energy:**

Fresnel lenses, a specialized type of prism, are employed in concentrated solar power systems. These lenses focus sunlight efficiently, contributing to the development of more effective solar energy collection technologies.

**Head-Up Displays (HUDs):**

Prisms are an integral part of HUD systems in both automotive and aviation contexts. They project crucial information onto the windshield or a combiner glass, allowing drivers or pilots to access important data without taking their eyes off their primary viewpoint.

**Microscopy:**

Nomarski prisms enhance the capabilities of differential interference contrast microscopy. They increase contrast in transparent specimens, enabling scientists to observe details that would be difficult or impossible to see with conventional microscopy techniques.

**Optical Coherence Tomography (OCT):**

In some OCT systems, prisms are employed for sample arm scanning and reference arm delay. This application of prisms contributes to the high-resolution imaging capabilities of OCT, which is particularly valuable in medical diagnostics, especially in ophthalmology.

## Chapter 8

# Optical Elements Part III

### Lenses

The most important optical elements are lenses, which come in many different flavors. They consist of curved surfaces, which most commonly have the shape of a part of a spherical cap. It is, therefore, useful to have a look at the refraction at spherical surfaces.

#### Refraction at spherical surfaces

For our calculations of the refraction at spherical surfaces, we consider the sketch below.

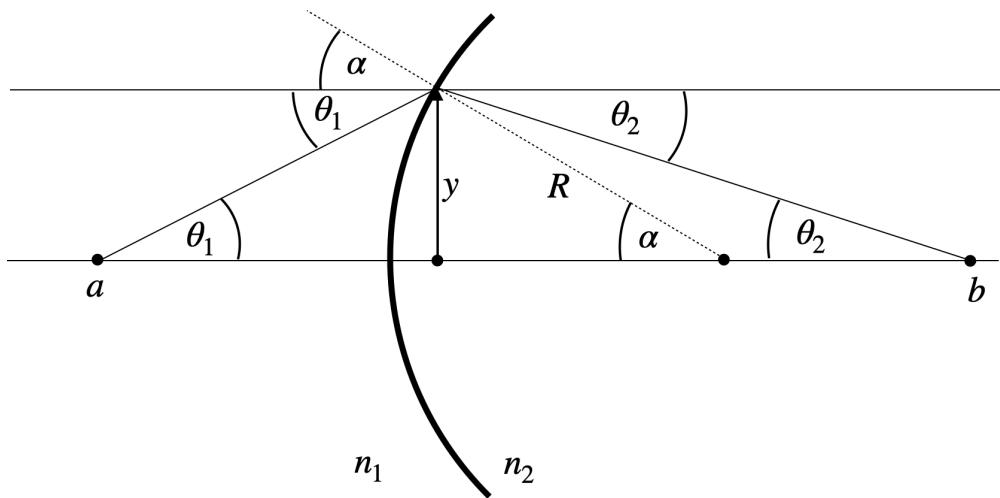


Figure 8.1: Refraction at a curved surface.

To derive an imaging equation for a lens, we aim to calculate the distance  $b$  and angle  $\theta_2$  at which a ray crosses the optical axis, given its origin at distance  $a$  and angle  $\theta_1$ . We begin with Snell's law for the geometry:

$$n_1 \sin(\alpha + \theta_1) = n_2 \sin(\alpha - \theta_2)$$

We define key relationships:

$$\sin(\alpha) = \frac{y}{R}, \quad \tan(\theta_1) = \frac{y}{a}, \quad \tan(\theta_2) = \frac{y}{b}$$

To simplify this, we employ the **paraxial approximation**, which assumes all angles are small. This allows us to use first-order approximations of trigonometric functions, effectively linearizing them:

$$\sin(\theta) \approx \theta + O(\theta^3), \quad \tan(\theta) \approx \theta + O(\theta^3), \quad \cos(\theta) \approx 1 + O(\theta^2)$$

This approach, common in optics, significantly simplifies our calculations while maintaining accuracy for most practical scenarios involving lenses.

With the help of these approximations we can write Snell's law for the curved surface as

$$n_1(\alpha + \theta_1) = n_2(\alpha - \theta_2).$$

With some slight transformation which you will find in the video of the online lecture we obtain, therefore,

$$\theta_2 = \frac{n_2 - n_1}{n_2 R} y - \frac{n_1}{n_2} \theta_1,$$

which is a purely linear equation in  $y$  and  $\theta_1$ .

### Paraxial Approximation

The paraxial approximation is a fundamental simplification in optics that assumes all angles are small. This allows us to use linear approximations for trigonometric functions, significantly simplifying calculations while maintaining accuracy for most practical scenarios involving lenses.

To visualize the validity of this approximation, let's examine two plots:

1. The first plot compares  $\sin(\theta)$  (blue line) with its linear approximation (red dashed line) for angles ranging from 0 to  $\pi/2$  radians.
2. The second plot shows the absolute error between  $\sin(\theta)$  and  $\theta$ .

These plots demonstrate that:

1. For small angles (roughly up to 0.5 radians or about 30 degrees), the approximation is very close to the actual sine function.
  2. The error increases rapidly for larger angles, indicating the limitations of the paraxial approximation.
- In most optical systems, especially those involving lenses, the angles of incident and refracted rays are typically small enough for this approximation to be valid. However, it's important to be aware of its limitations when dealing with wide-angle optical systems or scenarios where precision is critical.

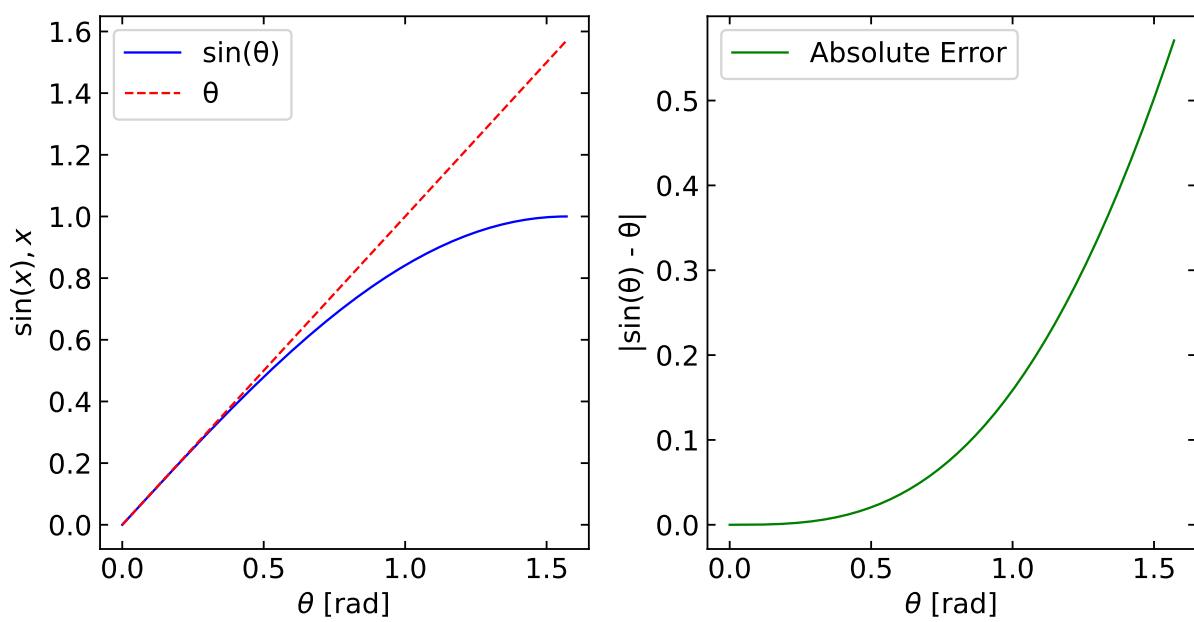


Figure 8.2: Visualization of the paraxial approximation plotting the  $\sin(\theta)$  and the linear approximation  $\theta$  (dashed line) for angles ranging from 0 to  $\pi/2$  radians.

Consider light originating from a point at distance  $y$  from the optical axis. We'll analyze two rays: one traveling parallel to the optical axis and hitting the spherical surface at height  $y$ , and another incident at  $y = 0$ .

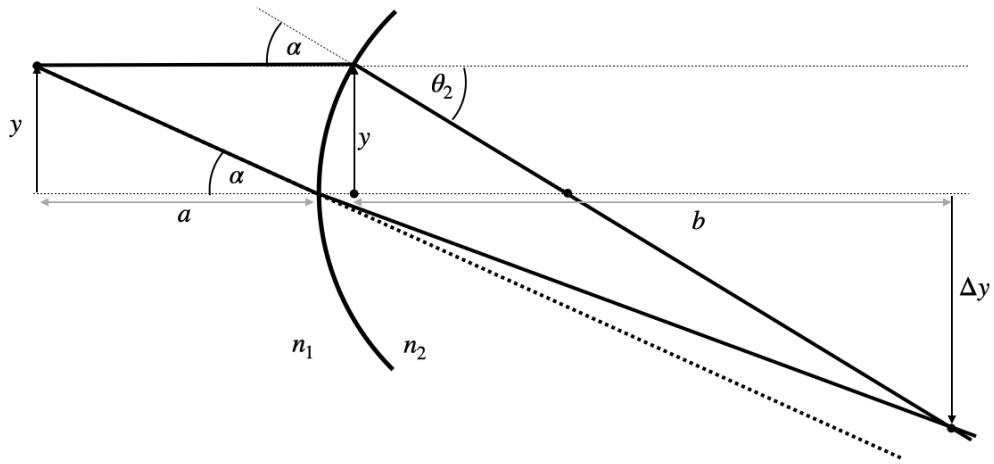


Figure 8.3: Image formation at a curved surface.

Applying our derived formula to these two cases:

For the parallel ray ( $\theta_1 = 0$ ):

$$\theta_2 = \frac{n_2 - n_1}{n_2} \frac{y}{R}$$

$$\theta_2 = \frac{y + \Delta y}{b}$$

Equating these expressions:

$$\frac{y + \Delta y}{b} = \frac{n_2 - n_1}{n_2} \frac{y}{R}$$

For the ray through the center ( $y = 0$ ):

$$n_2 \frac{\Delta y}{b} = n_1 \frac{y}{a}$$

Combining these equations yields the imaging equation for a curved surface:

$$\frac{n_1}{a} + \frac{n_2}{b} = \frac{n_2 - n_1}{R}$$

We can define a new quantity, the **focal length**, which depends only on the properties of the curved surface:

$$f = \frac{n_2}{n_2 - n_1} R$$

### Imaging Equation for Spherical Refracting Surface

The sum of the inverse object and image distances equals the inverse focal length of the spherical refracting surface:

$$\frac{n_1}{a} + \frac{n_2}{b} \approx \frac{n_2}{f}$$

where the focal length of the refracting surface is given by:

$$f = \frac{n_2}{n_2 - n_1} R$$

in the paraxial approximation.

## Thin lens

In our previous calculation we have found a linear relation between the incident angle  $\theta_1$  with the optical axis, the incident height of the ray  $y$  and the outgoing angle  $\theta_2$ :

Analyzing refraction in a lens involves two spherical surfaces. Light initially travels from a medium with refractive index  $n_1$  into the lens material with index  $n_2$ . The first surface's radius,  $R_1$ , is typically positive for a convex surface facing the incident light.

At the second surface, the outgoing angle from the first refraction becomes the incident angle for the second refraction. Here, light travels from  $n_2$  back into  $n_1$ . The radius  $R_2$  of this surface often has a negative value in a converging lens due to its opposite curvature relative to the optical axis.

For thin lenses, where the thickness  $d$  is much smaller than  $R_1$  and  $R_2$  ( $d \ll R_1, R_2$ ), we can simplify our analysis. We assume that the height of the ray at both surfaces is approximately equal ( $y \approx y'$ ), neglecting the displacement inside the lens.

This simplification allows us to treat all refraction as occurring on a single plane at the lens center, known as the **principal plane**. This concept, illustrated by the dashed line in the figure, greatly simplifies optical calculations and ray tracing for thin lenses.

The radii's sign convention (positive for convex surfaces facing incident light, negative for concave) and this two-surface analysis form the basis for the thin lens formula. This formula relates object distance, image distance, and focal length, encapsulating the lens's imaging properties.

The result of the above calculation is leading to the imaging equation for the thin lens.

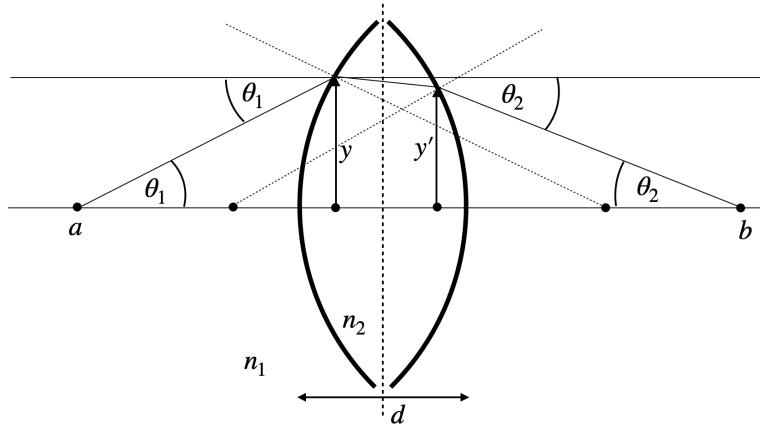


Figure 8.4: Refraction on two spherical surfaces.

### i Imaging Equation for Thin Lens

The sum of the inverse object and image distances equals the inverse focal length of the thin lens:

$$\frac{1}{a} + \frac{1}{b} \approx \frac{n_2 - n_1}{n_1} \left( \frac{1}{R_1} - \frac{1}{R_2} \right) = \frac{1}{f}$$

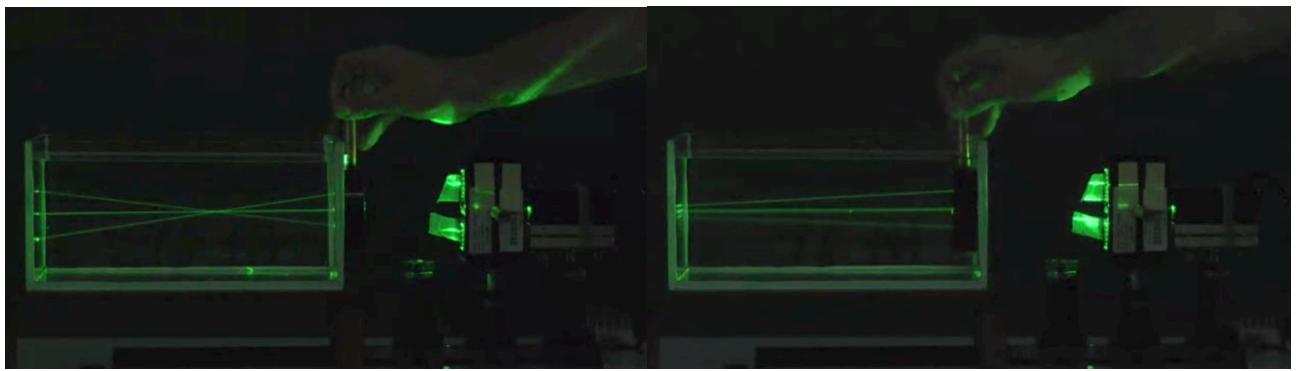
### i Lensmaker equation

The focal length of a thin lens is calculated by the **lensmaker equation**:

$$f = \frac{n_1}{n_2 - n_1} \left( \frac{R_1 R_2}{R_2 - R_1} \right)$$

in the paraxial approximation.

The equation for the focal length has some important consequence. It says that if the difference of the refractive indices inside ( $n_2$ ) and outside  $n_1$  get smaller, the focal length becomes larger and finally infinity. This can be nicely observed by placing a lens outside and inside a water filled basin as shown below.



(a) Lens in air

(b) Lens in water

Figure 8.5: Focusing of parallel rays by a lens in air ( $n_1 = 1$ , left) and in water ( $n_1 = 1.33$ , right). The images clearly show the change in focal length between the two situations.

### Bessel's method to measure the focal length of a lens

The is an interesting way to measure the focal length of a lens. Fix a distance  $D$  between object and screen. Then place a converging lens between them. Due to the reversibility of the light path, the lens will create a sharp image on the screen at two positions, which are separated by a distance  $d$ .

The equation for the focal distance can then be obtained from the

- Lens equation:  $\frac{1}{f} = \frac{1}{a} + \frac{1}{b}$
- Total distance:  $D = a + b$

Where  $f$  is focal length,  $a$  is object distance, and  $b$  is image distance. To obtain the focal distance according to this method, which is called the **Bessel method**, the following steps are taken:

For the first lens position:

$$D = a_1 + b_1$$

For the second lens position:

$$D = a_2 + b_2$$

We can further calculate the distance between the two lens positions:

$$d = a_1 - a_2 = b_2 - b_1$$

and use the imaging equation to find the focal length:

$$\frac{1}{f} = \frac{1}{a_1} + \frac{1}{b_1} = \frac{1}{a_2} + \frac{1}{b_2}$$

Substituting  $b_1 = D - a_1$  and  $b_2 = D - a_2$  we get further

$$\frac{1}{f} = \frac{1}{a_1} + \frac{1}{D - a_1} = \frac{1}{a_2} + \frac{1}{D - a_2}$$

Both euqations can be solved by

$$a_1 = \frac{D + d}{2} \quad \text{and} \quad a_2 = \frac{D - d}{2}$$

If we substitute that back into the imaging equation we obtain

$$\frac{1}{f} = \frac{2}{D} + \frac{2}{d}$$

which can be rearranged to get Bessel's formula:

$$f = \frac{D^2 - d^2}{4D}$$

This method only requires measuring  $D$  (fixed distance) and  $d$  (distance between lens positions). It eliminates the need to know exact object or image distances from the lens, making it more accurate than methods requiring precise distance measurements from the lens.

### Image Construction

Images of objects can be now constructed if we refer to rays which do not emerge from a position on the optical axis only. In this case, we consider three different rays (two are actually enough). If we use as in the case of a concave mirror a central and a parallel ray, we will find a position where all rays cross on the other side. The conversion of the rays is exactly the same as in the case of a spherical mirror. The relation between the position of the object and the image along the optical axis is described by the imaging equation.

Similar to the concave mirror, we may now also find out the image size or the magnification of the lens.

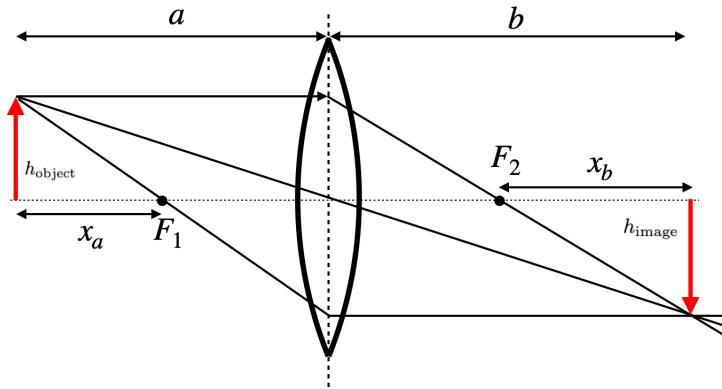


Figure 8.6: Image construction on a thin lens.

### **i** Magnification of a Lens

The magnification is given by:

$$M = \frac{h_{\text{image}}}{h_{\text{object}}} = -\frac{b}{a} = \frac{f}{f-a}$$

where the negative sign is the result of the reverse orientation of the real images created by a lens.

According to our previous consideration  $M < 0$  corresponds to a reversed image, while it is upright as the object for  $M > 0$ . We, therefore, easily see the following:

Object Position	Image Characteristics	Magnification (M)	Image Type
$a < f$	Upright and magnified	$M > 0$	Virtual
$f < a < 2f$	Reversed and magnified	$M < -1$	Real
$a = 2f$	Reversed, same size	$M = -1$	Real
$a > 2f$	Reversed and shrunk	$-1 < M < 0$	Real
$a = f$	Appears at infinity	$M = \infty$	-

The image below illustrates the construction of images in 4 of the above cases for a bi-convex lens with a parallel and a central ray for different object distances.

---

**Fig.:** Image construction on a biconvex lens with a parallel and a central ray for different object distances.

---

### Thick lens

For a thin lens, the displacement of the beam in height ( $y, y'$ ) due to the thickness has been neglected. That means that we can reduce all refracting action of the lens to a single plane, which we call a principle plane. This approximation is (independent of the paraxial approximation) not anymore true for lenses if the displacement  $\Delta$  of the ray as in the image below cannot be neglected. Such lenses are called **thick lenses** and they do not have a single principle plane anymore. In fact, the principle plane splits up into two principle planes at a distance  $h$ .

As indicated in the sketch above, an incident ray which is not deflected can be extended to its intersection with the optical axis at a point, which is a distance  $h_1$  behind the lens surface. This is the location for the first principle plane. The position of the second principle plane at a distance  $h_2$  before the back surface is found for by reversing the ray path. According to that, both principle planes have a distance  $h = d - h_1 + h_2$  (mind the sign of the  $h$ ). Using some mathematical effort, one can show that the same imaging equation as for a thins lens can be used with a new definition of the focal length and taking into account that object and image distances refer to their principle planes.

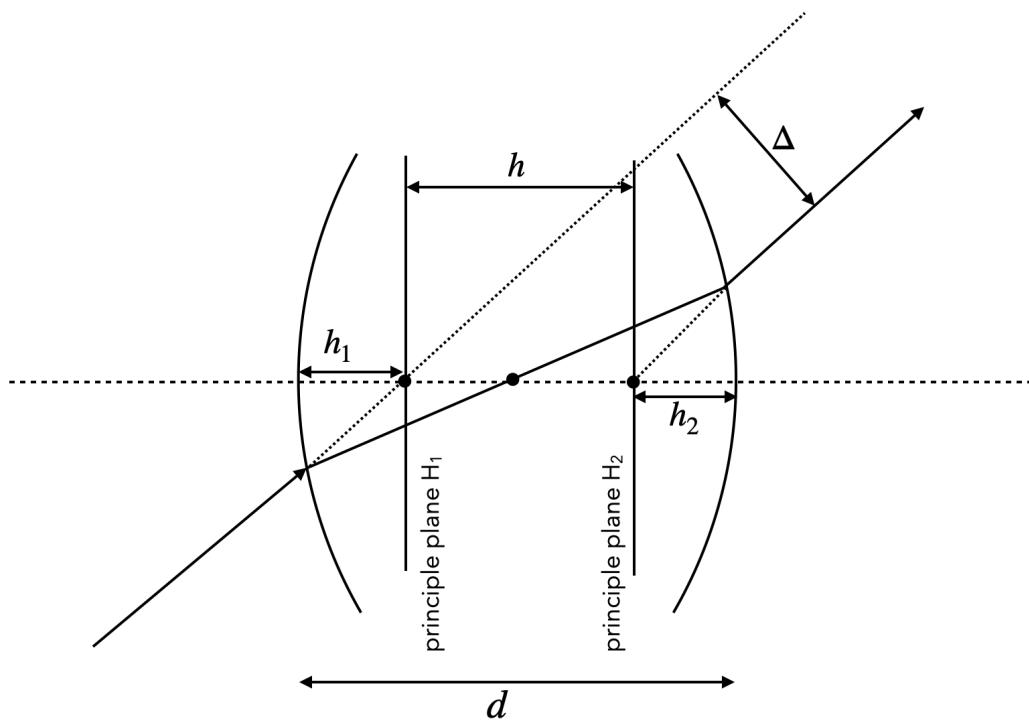


Figure 8.7: Thick lens principal planes.

### Matrix Optics

The above derived equations for a single spherical surface yield a linear relation between the input variables  $y_1$  and  $\theta_1$  and the output variables  $y_2$  and  $\theta_2$ . The linear relation yields a great opportunity to express optical elements in terms of linear transformations (matrices). This is the basis of **matrix optics**. The matrix representation of a lens is given by

$$\begin{pmatrix} y_2 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ \theta_1 \end{pmatrix}$$

where the matrix is called the **ABCD matrix** of the lens. Due to the linearization of Snells law we can write down more generally

$$\begin{pmatrix} y_2 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} y_1 \\ \theta_1 \end{pmatrix}$$

and one can obtain a Matrix for all types of optical elements such as free space of distance  $d$ .

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}$$

Here are some useful matrices for optical elements:

$$\mathbf{M} = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \quad (\text{Free space})$$

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{n_1}{n_2} \end{bmatrix} \quad (\text{Planar interface})$$

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ -\frac{(n_2-n_1)}{n_2 R} & \frac{n_1}{n_2} \end{bmatrix} \quad (\text{Spherical Boundary})$$

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{bmatrix} \quad (\text{Tin Lens})$$

If we have now a system of optical elements, we can multiply the matrices of the individual elements to obtain the matrix of the whole system.

$$\rightarrow \mathbf{M}_1 \rightarrow \mathbf{M}_2 \rightarrow \mathbf{M}_N \rightarrow \mathbf{M} = \mathbf{M}_N \dots \mathbf{M}_2 \mathbf{M}_1.$$

This is a very powerful tool to analyze optical systems.

### **i Thick Lens Focal Length**

We would like to model the imaging properties of a thick lens using the matrix method. The lens is divided into three components:

1. A spherical surface at the front of the lens with radius  $R_1$
2. A region of free space propagation through the lens material of thickness  $d$
3. Another spherical surface at the back of the lens with radius  $R_2$

The matrices for these components are:

$$M_1 = \begin{bmatrix} 1 & 0 \\ -1/f_1 & 1/n \end{bmatrix}, \quad M_2 = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}, \quad M_3 = \begin{bmatrix} 1 & 0 \\ -1/f_2 & n \end{bmatrix}$$

Where -  $f_1$  and  $f_2$  are the focal lengths of the front and back surfaces, which are determined by the radii of curvature  $R_1$  and  $R_2$  of the surfaces -  $d$  is the thickness of the lens, measured along the optical axis

## 8.1 Derivation

For a lens with refractive index  $n$  in air, the focal lengths of the surfaces are:

$$\frac{1}{f_1} = \frac{1-n}{nR_1}, \quad \frac{1}{f_2} = \frac{n-1}{R_2}$$

Where  $R_1$  and  $R_2$  are the radii of curvature of the front and back surfaces.

The total system matrix is then

$$M_{total} = M_3 \cdot M_2 \cdot M_1$$

After multiplication the total matrix is

$$M = \begin{bmatrix} 1 - \frac{d(1-n)}{R_1} & d \\ -\frac{n-1}{R_2} - \frac{(1-n)(n - \frac{dn(n-1)}{R_2})}{R_1 n} & \frac{(n - \frac{dn(n-1)}{R_2})}{n} \end{bmatrix}$$

where the element in the lower left corner is the inverse of the focal length of the thick lens. This can be simplified to the following expression:

$$-\frac{1}{f} = -\frac{1}{f_2} - \frac{n}{f_1} - \frac{dn}{f_1 f_2}$$

Substituting the expressions for  $1/f_1$  and  $1/f_2$ :

$$\frac{1}{f} = \frac{n-1}{R_1} - \frac{n-1}{R_2} + \frac{d(n-1)^2}{nR_1 R_2}$$

Factoring out  $(n-1)$  gives the final expression for the focal length of a thick lens:

$$\frac{1}{f} = (n-1) \left[ \frac{1}{R_1} - \frac{1}{R_2} + \frac{(n-1)d}{R_1 R_2} \right]$$

This is the Lensmaker's equation for a thick lens.

The construction of ray diagrams for thick lenses is similar to that for thin lenses, but the object and image distances are measured from the principal planes. The magnification is also calculated using the distances from the principal planes. **Principal planes are where a thick lens can be treated as an equivalent thin lens. At these planes, the magnification is unity.**

The derivation of the local of the principle planes will be part of the seminar.

$$\frac{1}{f} = \frac{n_1 - n_2}{R_2 n_1} - \frac{n_1 - n_2}{R_1 n_1} + \frac{d(n_1 - n_2)^2}{R_1 R_2 n_1}$$

### **i** Imaging Equation for Thick Lens

The sum of the inverse object and image distances to the principal planes ( $H_1, H_2$ ) equals the inverse focal length of the thick lens:

$$\frac{1}{a} + \frac{1}{b} \approx \frac{1}{f}, \quad \text{where} \quad \frac{1}{f} = n - 1 \left( \frac{1}{R_1} - \frac{1}{R_2} + \frac{(n-1)d}{nR_1 R_2} \right)$$

in the paraxial approximation. The construction of the image on a thick lens is done with the help of two principle planes. The object distance  $a$  and the image distance  $b$  are measured from these principle planes. The location of the two principle planes are found to be

$$h_1 = -\frac{(n-1)fd}{nR_2}$$

$$h_2 = -\frac{(n-1)fd}{nR_1}$$

As compared to the construction of an image on a thin lens, we now have to consider some peculiarities for the thick lens. An incident parallel ray, which turns into a focal ray is now refracted at the second principle plane. The reverse must, therefore, be true for an incident focal ray. This ray is refracted on the first principle plane. The central ray is deflected on both principle planes. It is incident under a certain angle at the first principle plane and outgoing with the same principle angle to the second principle plane. The sketch below summarizes these issues for a thick lens.

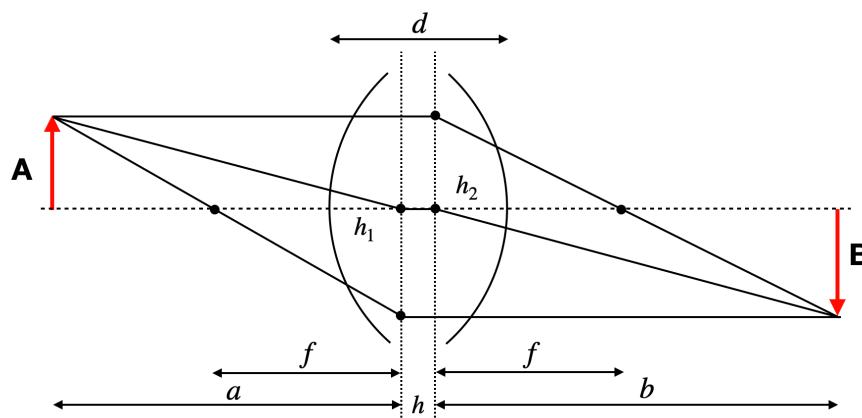


Figure 8.8: Thick lens image construction.

### Lens types

Depending on the radii of curvature and their sign, one can construct different types of lenses that are used in many applications. Modern microscopy lenses, for example, can contain up to 20 different lenses, each with

carefully designed curvatures and materials to correct for various optical aberrations and achieve high-quality imaging.

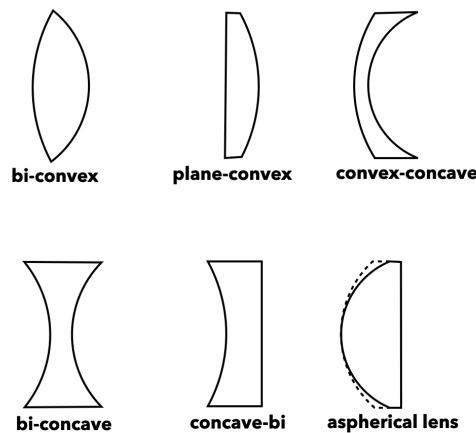


Figure 8.9: Different lens types.

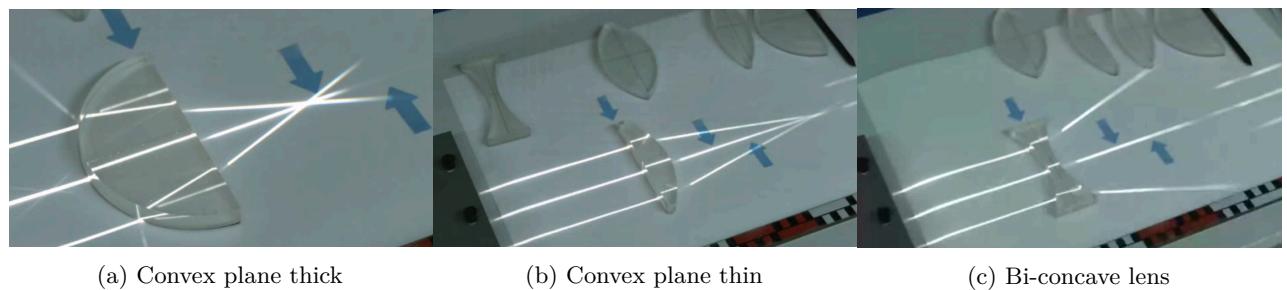


Figure 8.10: Focusing behavior of a few different lens types.



# Chapter 9

## Optical Instruments

### 9.1 The Human Eye

The human eye stands is one of the most remarkable sensory systems. This sophisticated organ combines an array of precisely crafted components—including an adjustable aperture, an adaptive lens, and a highly sensitive photodetector—all interconnected with a neural network capable of rapid and accurate pattern recognition. What's truly astounding is that this entire system operates on mere watts of power.

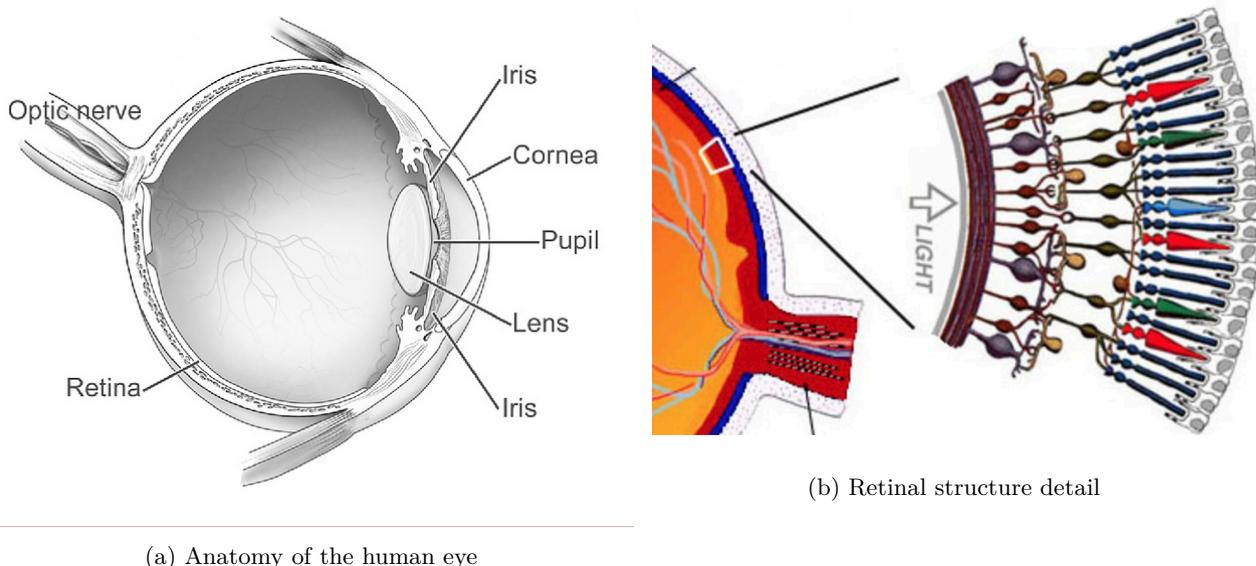


Figure 9.1: Left: Key components of the human eye, including the lens, vitreous body, and retina with its light-sensitive cells. Right: Detailed view of the retina, showing the arrangement and neural connections of rods and cones.

### Key Components and Their Functions

1. **Pupil and Iris:** The pupil, surrounded by the iris, acts as an adjustable aperture. It regulates the amount of light entering the eye and influences the depth of field. In bright conditions, a constricted pupil increases the depth of field, allowing a wider range of distances to be in focus simultaneously.
2. **Lens:** Connected to the ciliary muscles, the lens can change its curvature to adjust focal length, a process known as accommodation. This allows the eye to focus on objects at varying distances.

3. **Vitreous Humor:** This gel-like substance fills the eye cavity, maintaining its shape and contributing to the eye's optical properties.
4. **Retina:** The light-sensitive layer at the back of the eye, containing photoreceptor cells (rods and cones) that convert light into neural signals.

## Photoreceptors: Cones and Rods

The retina contains two types of photoreceptor cells:

1. **Cones:** Responsible for color vision and high acuity in bright light. They are concentrated around the fovea, the area of highest visual acuity. There are about 6 Million cones in the human eye.
2. **Rods:** More sensitive to light but do not distinguish colors, providing vision in low light conditions. There are around 12 Million rods in the human eye.

Rods and cones obtain their function from a chromophore molecule called **retina**, which undergoes a conformational change when exposed to light. This change triggers a cascade of chemical reactions that ultimately lead to the generation of neural signals. The color vision is achieved with the same chromophore molecule that is embedded in slightly different protein structures in cones. This allows cones to be sensitive to different wavelengths of light.

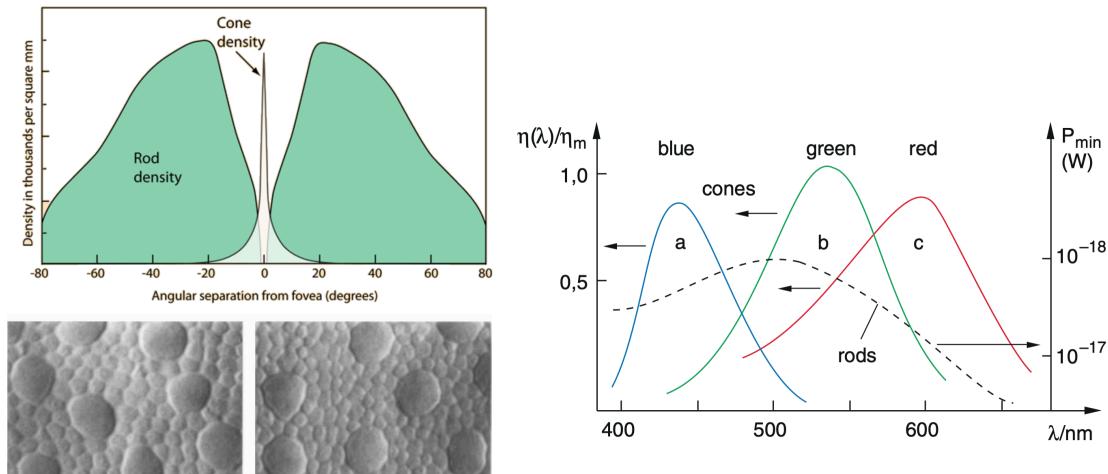


Figure 9.2: Distribution of cones and rods around the fovea, their microscopic structure, and spectral sensitivity.

Cones contain light-sensitive pigments based on retinal molecules, which undergo conformational changes when excited by light, triggering a cascade of chemical processes. There are three types of cones, each sensitive to different wavelengths of light, enabling color vision.

## Visual Acuity and Performance

Visual acuity, often measured using an eye chart, quantifies the eye's ability to resolve fine details. It's typically expressed as a fraction (e.g., 20/20 vision), where the numerator is the test distance and the denominator is the distance at which a person with normal acuity can read the same line.

The human eye's remarkable performance in pattern recognition, depth perception, and adaptability to varying light conditions is achieved through the complex interplay of its optical components and neural processing. This sophisticated system continues to inspire developments in artificial vision systems and optical technologies.

## Optical Properties of the Eye

The eye's optical system is asymmetrical due to the different media it interfaces with (air on one side, vitreous humor on the other). This results in different focal lengths:

- Front focal length:  $f_1 = 17 \text{ mm}$
- Back focal length:  $f_2 = 22 \text{ mm}$

These values can change during accommodation for near vision:

- Close object front focal length:  $f_1 = 14 \text{ mm}$
- Close object back focal length:  $f_2 = 19 \text{ mm}$

The eye's refractive power, measured in diopters (D), is the reciprocal of the focal length in meters. For a relaxed eye:

$$P = \frac{1}{f} = \frac{1}{0.022 \text{ m}} \approx 45.45 \text{ D}$$

During accommodation, this can increase to about 52 D.

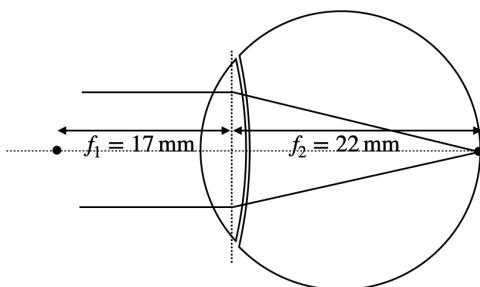


Figure 9.3: Illustration of the eye's focal distances.

## Resolution Limit of the Eye

The resolution of the eye is limited by diffraction and the spacing of photoreceptors. The minimum angle of resolution  $\theta_{\min}$  can be approximated by:

$$\theta_{\min} \approx \frac{1.22\lambda}{D}$$

where  $\lambda$  is the wavelength of light and D is the diameter of the pupil. For a 3 mm pupil and 555 nm light (peak sensitivity), this gives a theoretical resolution of about 1 arc minute.

## Refractive Errors and Visual Correction

The human eye, under normal conditions, focuses images of distant objects onto the retina at the back focal distance of approximately 22 mm. However, various refractive errors can occur due to imperfections in the eye's optical system, primarily the cornea and lens. These errors affect the eye's ability to focus light accurately on the retina, leading to vision problems.

Common refractive errors include:

1. **Myopia (Short-sightedness):** Light from distant objects focuses in front of the retina, causing distant objects to appear blurry while near objects remain clear.
2. **Hyperopia (Far-sightedness):** Light focuses behind the retina, making nearby objects appear blurry while distant objects may remain clear.
3. **Astigmatism:** The cornea or lens isn't perfectly spherical, causing light to focus at multiple points rather than a single sharp point on the retina.

The severity of refractive errors can be quantified using the concept of refractive power. The refractive error R of the eye, measured in diopters (D), is calculated as:

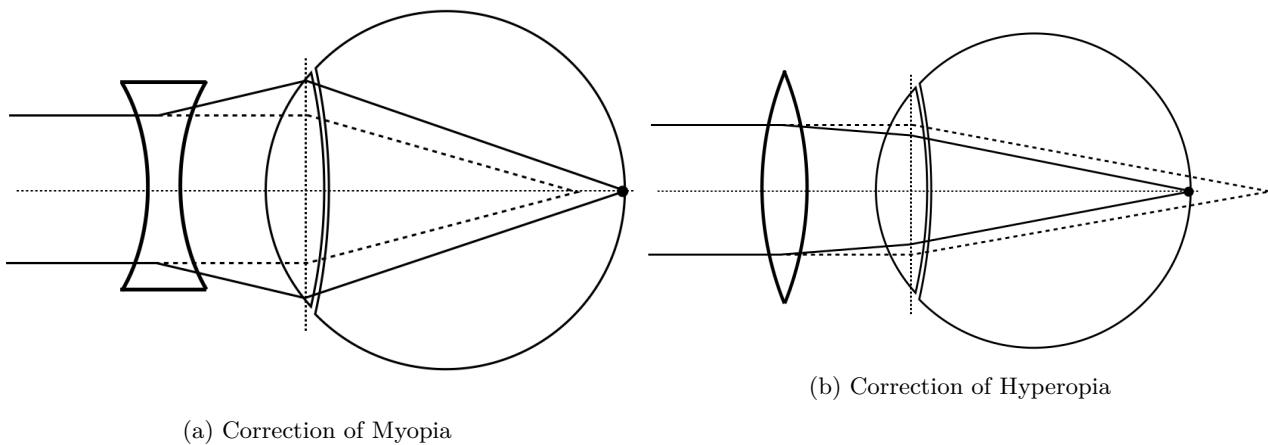


Figure 9.4: Left: Myopia correction using a concave lens. Right: Hyperopia correction using a convex lens.

$$R = \frac{1}{f_{\text{required}}} - \frac{1}{f_{\text{actual}}}$$

where  $f_{\text{required}}$  is the focal length needed for perfect focus, and  $f_{\text{actual}}$  is the eye's actual focal length. This formula helps determine the degree of correction needed for various eye defects.

In a normal, relaxed state, the human eye can observe objects clearly up to a distance of approximately  $s_0 = 25$  cm without additional accommodation of the lens. This distance, known as the **range of clear visual sight**, varies among individuals and is used as a standard in optical calculations. Objects within this range can be observed under a visual angle  $\epsilon_0$ . For small angles, which is typically the case in vision, the angular size  $\epsilon_0$  of an object of height  $h$  at a distance  $s_0$  is approximated by:

$$\epsilon_0 \approx \tan(\epsilon_0) = \frac{h}{s_0}$$

This relationship is fundamental in understanding how objects are perceived and in designing corrective lenses and optical instruments.

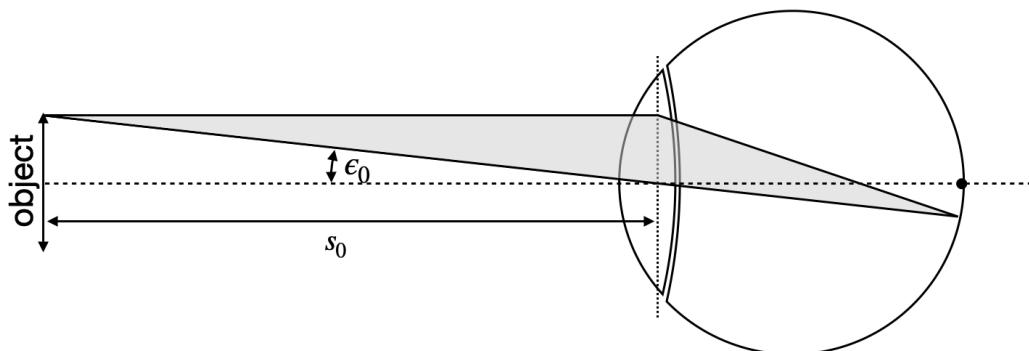


Figure 9.5: Diagram of a relaxed eye focusing on a distant object.

history |grep git

Understanding these concepts is crucial for diagnosing vision problems and designing appropriate corrective measures, whether through eyeglasses, contact lenses, or surgical interventions.

## Magnification

Having discussed the basic structure and function of the human eye, we now turn to how optical instruments can enhance our vision. Instead of calculating the magnification of optical instruments from object and image distances, we introduce a more relevant measure: the **angular magnification**.

The angular magnification,  $V$ , is defined as the ratio of the angle subtended by the image when viewed through the instrument to the angle subtended by the object when viewed with the naked eye at the near point. It is given by:

$$V = \frac{\tan(\epsilon)}{\tan(\epsilon_0)} \approx \frac{\epsilon}{\epsilon_0}$$

where: -  $\epsilon$  is the angle subtended by the image at the eye when viewed through the instrument -  $\epsilon_0$  is the angle subtended by the object when viewed with the naked eye at the near point

This concept is crucial in understanding how optical instruments like telescopes and microscopes enhance our vision. Angular magnification effectively increases the apparent size of objects by increasing the angle at which they are viewed. This measure is particularly useful as the actual image size is often not directly accessible or relevant to the viewer's experience.



# Chapter 10

## Optical Instruments

Optical instruments now combine a number of optical elements or even consist only out of a single one as in the case of the magnifying glass or the eye.

### 10.1 Magnifying Glass

A magnifying glass has several applications. First of all, it allows to see objects with details that would otherwise be too small to be observed with the eye even if the eye lens can accommodate to the distances. Such magnifying glasses are also used in microscopes as the so-called **eye-piece** as we will later see in the section on microscopes.

Consider the sketch below. The sketch shows an object of a size  $A$  which is at a distance of  $s_0$  from the eye. The object makes an angle  $\epsilon_0$  with the optical axis. If we insert now a lens into the space between object and eye and the lens is positioned in a way that it is exactly at a distance  $f$  (the focal distance of the lens) from the object then we are able to observe the object under a different angle  $\epsilon$ .

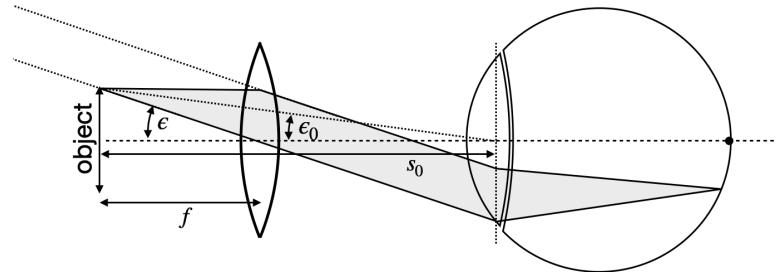


Figure 10.1: Magnifying glass at the focal distance.

The magnification of this magnifying glass can be calculated from the angles  $\epsilon \approx A/f$  and  $\epsilon_0 \approx A/s_0$ :

$$V = \frac{\tan(\epsilon)}{\tan(\epsilon_0)} \approx \frac{\epsilon}{\epsilon_0} = \frac{A}{f} \frac{s_0}{A} = \frac{s_0}{f}.$$

The angular magnification is, thus, just given by the ratio of the clear visual range to the focal distance of the lens. If the focal distance  $f$  becomes much smaller than  $s_0$ , large magnifications are possible.

A second very useful effect is that when the object is placed inside the focal distance from the lens, the eye images a virtual image at infinite distance to the retina (see sketch). This means the eye muscle can stay relaxed when observing the object, while it would otherwise probably have to accommodate to the distance.

Yet, placing the object at exactly the focal distance is rather tedious when holding the magnifying glass by hand. If the object is now placed inside the focal distance of the magnifying glass, we may also calculate a magnification in this case knowing the virtual image size  $B$  created in this case (see sketch below)

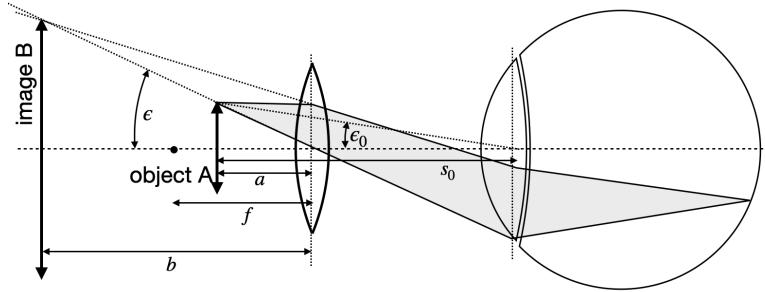


Figure 10.2: Magnifying glass for an object inside the focal range of the lens.

If  $a$  is the distance of the object from the principle plane of the magnifying glass and  $b$  and  $B$  are the distance and the size of the virtual image, respectively, we obtain

$$V = \frac{\tan(\epsilon)}{\tan(\epsilon_0)} \approx \frac{\epsilon}{\epsilon_0} = \frac{B s_0}{b A} = \frac{s_0}{a}.$$

Using the imaging equation

$$\frac{1}{f} = \frac{1}{a} + \frac{1}{b}$$

we may finally arrive at

$$V = \frac{s_0(b - f)}{b f}$$

in this case. If we place the virtual image directly at the clear visual range, i.e.,  $b = -s_0$ , we find

$$V = \frac{s_0}{f} + 1.$$

# Chapter 11

## Optical Instruments

### 11.1 Microscope

#### Historical Context of Microscope Development

Optical Microscopy has a rich history of development, and is a very important tool in the fields of biology, materials science, and nanotechnology. Here are some key milestones in the history of microscopy:

**Ancient Times - 13th Century:** Simple magnifying glasses - The concept of magnification was known to ancient civilizations. - In the 13th century, Italian craftsmen created the first wearable glasses.

**1590: Compound Microscope** - Hans and Zacharias Janssen, Dutch spectacle makers, created the first compound microscope.

**1665: Robert Hooke's "Micrographia"** - Hooke published detailed observations made with his improved compound microscope. - He coined the term "cell" after observing cork tissue.

**1670s: Antonie van Leeuwenhoek's Single-Lens Microscopes** - Developed high-quality single-lens microscopes with up to 270x magnification. - First to observe and describe bacteria, yeast, and other microorganisms.

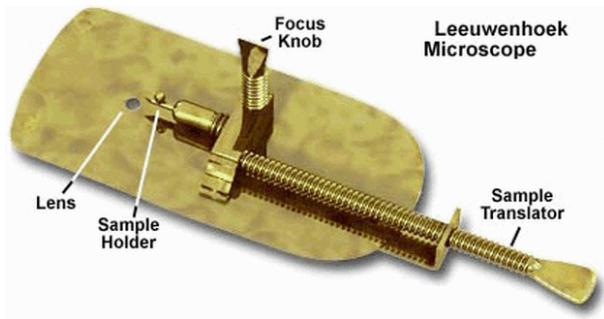


Figure 11.1: Image of Leeuwenhoek's microscope

**18th-19th Centuries: Achromatic Lenses** - Joseph Jackson Lister developed achromatic lenses, reducing chromatic aberration.

**1830s: Ernst Abbe's Theoretical Work** - Formulated the Abbe Sine Condition, crucial for modern lens design.

**Late 19th Century: Oil Immersion Lenses** - Allowed for higher resolution in light microscopy.

**1931: Electron Microscope** - Ernst Ruska and Max Knoll developed the first electron microscope.

**1950s-1960s: Phase Contrast and Fluorescence Microscopy** - Frits Zernike invented phase contrast microscopy. - Development of fluorescence microscopy techniques.

**1981: Scanning Tunneling Microscope** - Gerd Binnig and Heinrich Rohrer invented the STM, allowing

imaging at the atomic level.

**1980s-Present: Digital and Computational Microscopy** - Integration of CCD cameras and digital imaging. - Development of confocal microscopy, super-resolution techniques, and computational methods like ptychography.

In this section we will analyze the optical properties of microscopes from the perspective of geometrical optics which explains image formation. Yet, the key to the performance of a microscope is the understanding provided by wave optics. We will discuss this in a later section. The simplest form of a microscope consists of an objective lens with a focal distance  $f_1$  and a magnifying glass called eye-piece with a focal length  $f_2$ . In this system of two lenses (which are itself systems of lenses in modern microscopes, see below),



Figure 11.2: Fig.: Cut through a microscope objective lens (left) and an eye-piece.

the object is placed at a distance  $f_1 < a_1 < 2f_1$  from the objective lens creating a real and reversed image at a distance  $b_1$  behind the lens. This reversed image is observed by the eye through the eye-piece. The image of the objective lens is thereby adjusted to appear at the focal distance of the eye-piece.

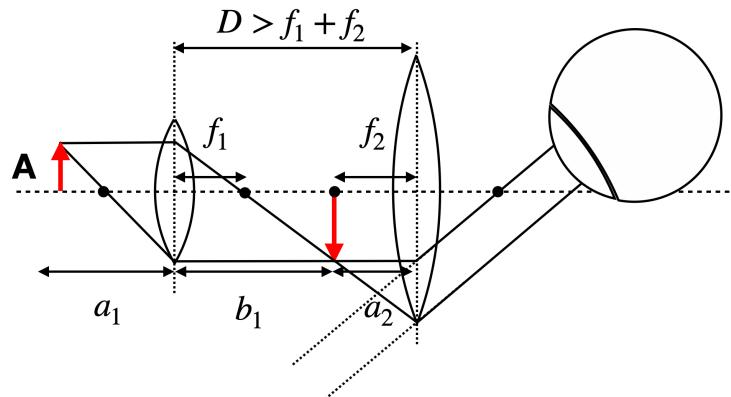


Figure 11.3: Fig.: Sketch of a simple microscope. The strange object on the right is an eye.

For this simple microscope system we may calculate first the intermediate image position  $b_1$ :

$$\frac{1}{f_1} = \frac{1}{a_1} + \frac{1}{b_1}$$

resulting in

$$b_1 = \frac{a_1 f_1}{a_1 - f_1}.$$

If we assume a  $\delta$  to be the distance of the object from the focal point of the objective lens, we even find for  $\delta \rightarrow 0$

$$b_1 = \frac{a_1 f_1}{\delta}.$$

The intermediate image of size  $B_1$  is now imaged by a magnifying glass of focal distance  $f_2$ . According to what we calculated earlier, we have now the observation angle

$$\tan(\epsilon) = \frac{B_1}{f_2} = \frac{Ab_1}{a_1 f_2}.$$

If we observe the object of a size  $A$  and the clear visual distance  $s_0$ , it would cover an angle of

$$\tan(\epsilon_0) = \frac{A}{s_0}$$

and we may obtain the total angular magnification

$$V = \frac{Ab_1 s_0}{Aa_1 f_2} = \frac{b_1 s_0}{a_1 f_2}.$$

If we set the distance between the two lenses to  $D = b_1 + f_2$  and  $a_1 \approx f_1$  then we obtain

$$V = \frac{(D - f_2)s_0}{f_1 f_2}$$

which says that the magnification is the result of the two focal length  $f_1, f_2$ .

### Modern microscopy

While the description above accurately represents the simplest microscope design, contemporary microscopes employ more intricate light paths and generally utilize what is known as **infinity corrected** optics. This system incorporates an objective lens that projects images of objects in the focal plane to infinity. Such an objective lens is invariably paired with a secondary lens, called the **tube lens**. Together, these lenses are engineered to provide a magnification level that is specified on the objective lens housing.

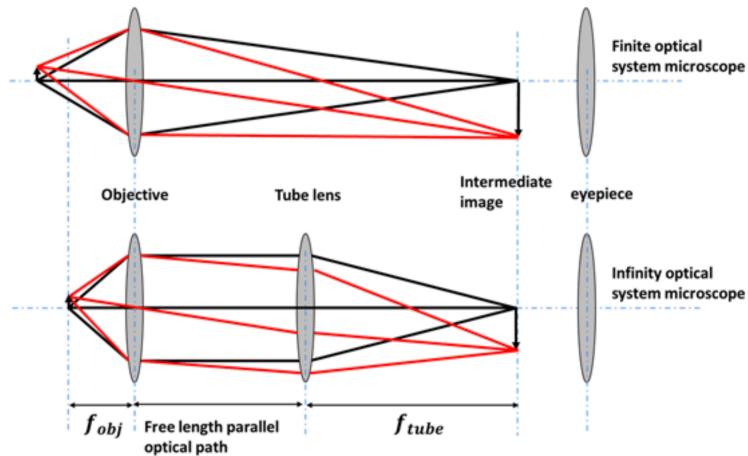


Figure 11.4: Fig.: Infinity optics vs. normal microscopy optics.

Infinity optics allows you to have a free length with a parallel optical path where you can insert optical elements. There is no fixed tube length as in the case sketched above, where the distance of the intermediate image has to be considered. Therefore, it has tremendous technical advantages. Common optical microscopes are further

today coupled to CCD cameras to record images digitally. Yet, an eye-piece may still be available in many cases. The sketch below shows the light path for a simple fluorescence microscope recording fluorescence images with a camera.

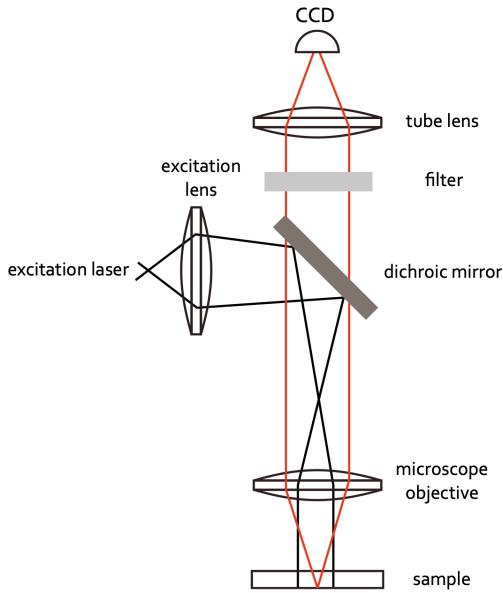
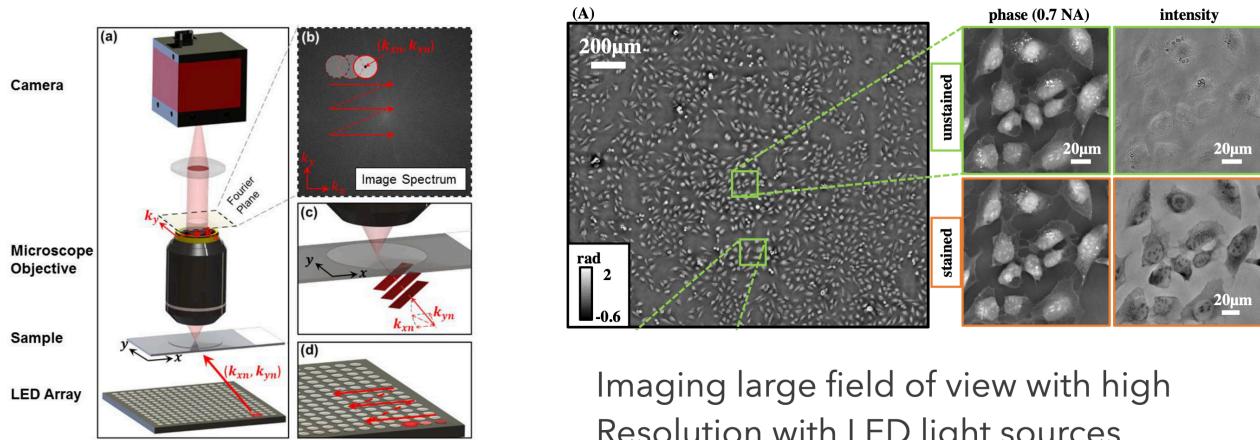


Figure 11.5: Fig.: Simple fluorescence microscope.

### Digital Microscopy

The possibility to digitally record images creates endless possibilities to computationally enhance and combine images. Nowadays the field of optics is one of the fastest developing fields in physics with numerous new techniques appearing every week. In this field of imaging methods of machine learning also play an increasingly important role. While I'm not able to refer to all possible optical microscopy techniques here, I will exemplarily show some data from the Waller group at Berkley using computational methods to enhance the resolution by keeping at the same time a large field of view for imaging. This technique is called **ptychography** and can be understood if you consider Fourier Optics (a field of optics describing light propagation in terms of Fourier transforms).



Imaging large field of view with high Resolution with LED light sources.

Figure 11.6: Fig.: Ptychographic imaging with LED arrays.

There is a massive amount of other techniques with increadible images being generated. Have a look around.

### i Advanced Microscopy Techniques

While traditional light microscopy has been invaluable, it's limited by the diffraction of light, restricting resolution to about 200 nm as will be discussed in a later part of this course. Modern techniques have pushed beyond this limit, revolutionizing our ability to visualize microscopic structures.

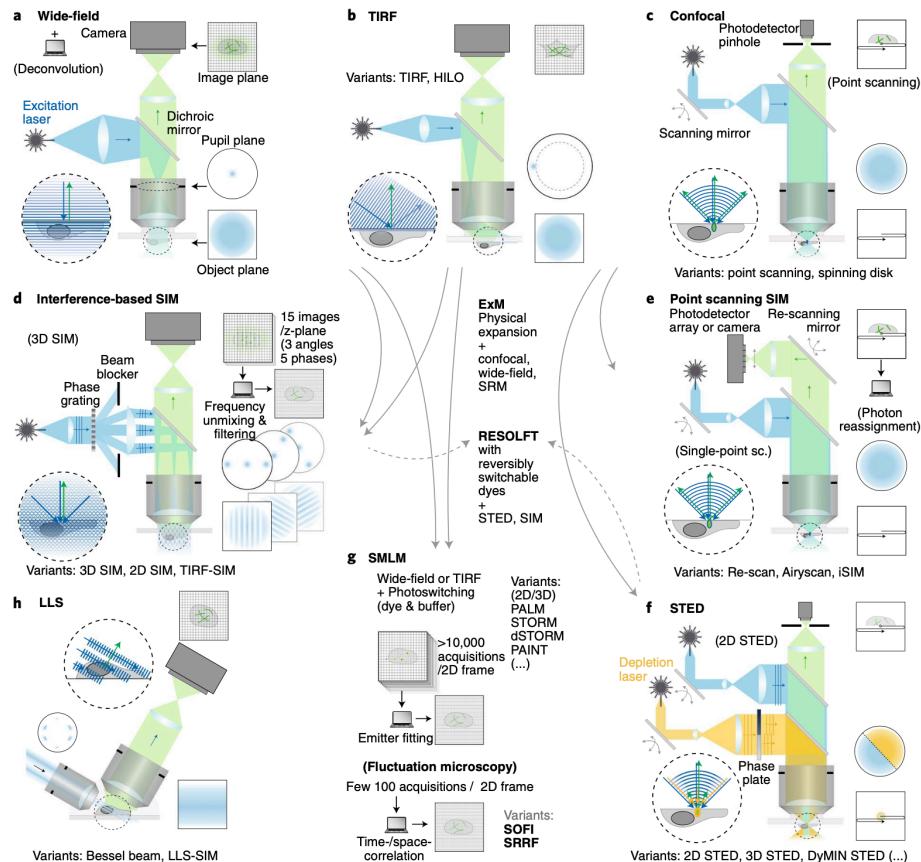


Figure 11.7: Super resolution Imaging Methods Overview (Schermelleh, L. et al. Super-resolution microscopy demystified. Nat. Cell Biol. 21, 72–84 (2019))

### Confocal Microscopy

Confocal microscopy uses point illumination and a pinhole in an optically conjugate plane in front of the detector to eliminate out-of-focus signal.

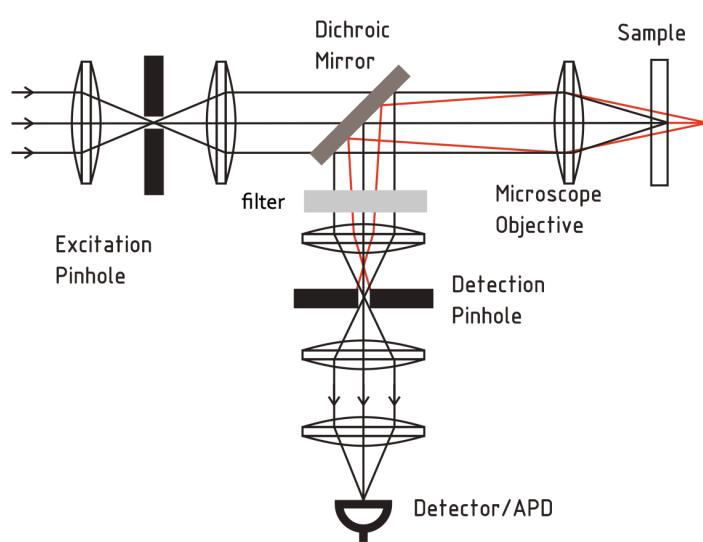


Figure 11.8: Confocal Microscope.

- **Key Features:**
  - Improved optical resolution and contrast
  - Ability to collect serial optical sections from thick specimens
  - Widely used in biological sciences
- **Overcoming Limitations:**
  - Eliminates background information away from the focal plane
  - Allows for 3D reconstruction of samples

## Examples for Super-Resolution Microscopy Techniques

Super-resolution techniques bypass the diffraction limit, achieving resolutions down to tens of nanometers.

1. **Stimulated Emission Depletion (STED) Microscopy**
  - Uses two laser beams: one to excite fluorescent molecules, another to suppress fluorescence around the excitation spot
  - **Overcoming Limitations:** Achieves resolution as fine as 20-50 nm by precisely controlling which fluorophores are allowed to fluoresce
2. **Photoactivated Localization Microscopy (PALM)**
  - Relies on selective activation and sampling of sparse subsets of photoactivatable fluorescent molecules
  - **Overcoming Limitations:** Locates individual molecules with nanometer precision by isolating their signals over time
3. **Stochastic Optical Reconstruction Microscopy (STORM)**
  - Similar to PALM, but uses photoswitchable fluorophores
  - **Overcoming Limitations:** Achieves resolutions of ~20 nm by precisely locating the centers of single fluorescent molecules
4. **Structured Illumination Microscopy (SIM)**
  - Uses patterned illumination to create moiré fringes, which are computationally processed to reconstruct super-resolution images
  - **Overcoming Limitations:** Doubles the resolution of traditional light microscopy to ~100 nm
5. **Superresolution Photothermal Infrared Imaging**
  - Superresolution photothermal infrared imaging is a novel technique that brings the advantages of superresolution microscopy to the infrared regime.
  - **Overcoming Limitations:** Achieves resolutions of ~300 nm for infrared imaging at wavelength of 10 μm by using photothermal lensing effects.

# Chapter 12

## Telescopes

Other than the microscope, the telescope is made to observe distant objects, which would appear under a very small observation angle. This can be achieved with different optical designs. The most common telescopes are refracting telescopes, which use lenses to magnify the angle under which the object is observed. The second type of telescopes are reflecting telescopes, which use mirrors to magnify the angle under which the object is observed.

### Refracting Telescopes

The telescope is therefore made to magnify the angle under which the object is observed. In the same way as a microscope, the telescope consists of two lenses with the focal distances  $f_1, f_2$ .

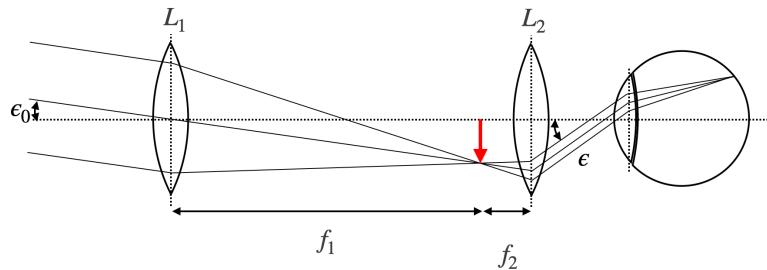


Figure 12.1: Kepler telescope with two biconvex lenses, creating a reversed image of distance objects.

As indicated in the sketch above, the first lens generates an image at the focal length of the first lens. This intermediate image is the magnified by an eye-piece as well acting as a magnifying glass. We may therefore apply the same kind of techniques as earlier for the calculation of the angular magnification. The angle of observation for the object of size  $D$  is given by

$$2\epsilon_0 = \frac{D}{f_1}$$

while the angle of observation through the telescope is given as

$$2\epsilon = \frac{D}{f_2}$$

Correspondingly, the angular magnification is given by

$$V = \frac{\epsilon}{\epsilon_0} = \frac{D f_1}{f_2 D} = \frac{f_1}{f_2}$$

The magnification is therefore given by the ratio of the focal length of the entrance lens and the eye-piece. The above telescope is also termed **astronomical telescope** or **Kepler telescope**, since it has been used for astronomical observations. It creates an image which is reversed.

A telescope with an upright image may be created with the help of a concave lens. This type of telescope is called Galilei telescope and obeys the same magnification formula as above. Due to the fact that a concave lens has a negative focal length, the total magnification will be negative as well being indicative for an upright image.

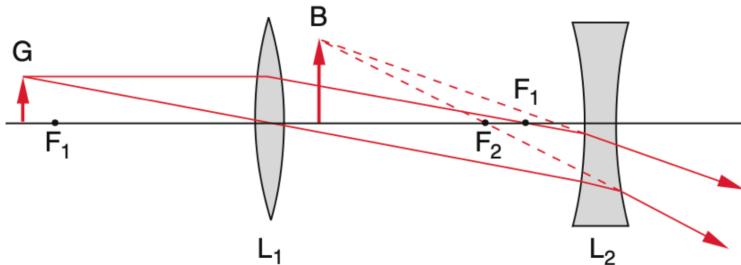


Figure 12.2: Galilei telescope for imaging objects into upright images with the help of a concave and a convex lens.

## Reflecting Telescopes

Modern powerful telescopes also use mirrors instead of refracting optical elements, as reflecting elements with nearly 100 percent reflectivity can be built with a much smaller mass than large glass elements. Such telescopes come in different setups. The one below is a Cassegrain telescope, where a secondary convex mirror is used for imaging the intermediate image to the eye.

## Adaptive Optics

When observing objects from the ground, the atmosphere can distort the image. This is due to the fact that the atmosphere is not homogeneous and the refractive index of the air is changing with time. This leads to a distortion of the image, which can be corrected with the help of adaptive optics. The principle of adaptive optics is to measure the distortion of the image with the help of a laser beam and to correct the image with the help of a deformable mirror. The deformable mirror is a mirror with a number of actuators, which can change the shape of the mirror in order to correct the distortion of the image. The principle of adaptive optics is shown in the figure below.



(a) Reflective optics is commonly used in modern high quality telescopes for the advantage of weight. The image and sketch shows the optics of a so-called Cassegrain telescope.

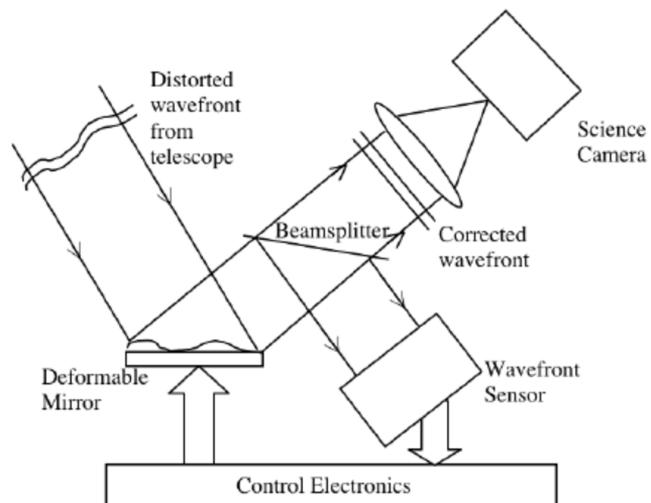


Figure 12.4: Principle of adaptive optics. The image of a star is distorted by the atmosphere. The distortion is measured with the help of a laser beam and corrected with the help of a deformable mirror.

## Space-Based Telescopes

Placing telescopes in space eliminates atmospheric interference completely. The Hubble Space Telescope revolutionized astronomy with its crystal-clear views of the universe.

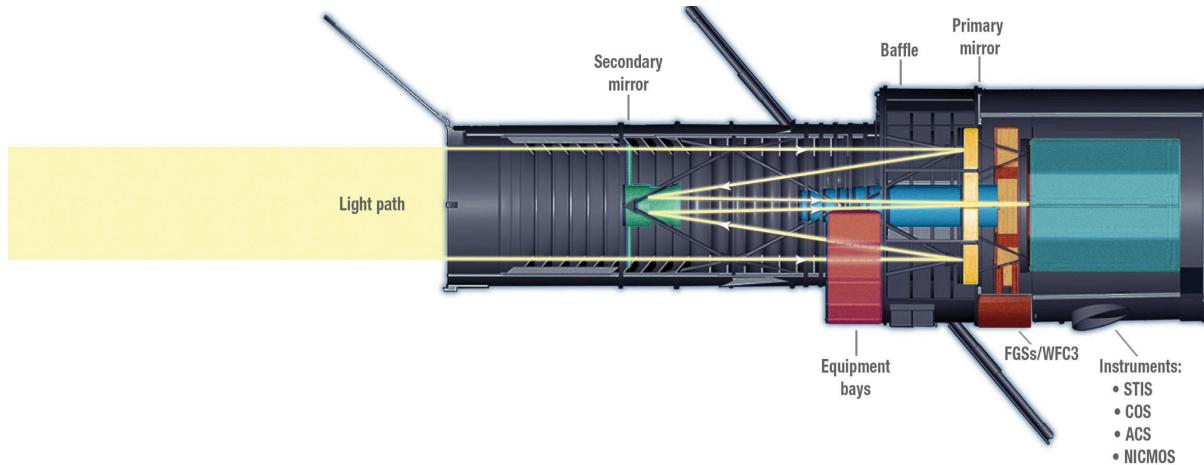


Figure 12.5: Light path in the Hubble space telescope.

Its successor, the James Webb Space Telescope (launched in 2021), operates in the infrared spectrum and can peer even further into space and time. Being above the atmosphere not only provides clearer images but also allows these telescopes to observe wavelengths that are normally blocked by Earth's atmosphere, particularly in the infrared and ultraviolet regions.

# Chapter 13

## Rainbow

As the last topic of the optical elements we would like to have a look at a phenomenon, which has nothing to do with optical elements but is fun and just fits to the topic of dispersion. We will explore the rainbow and in addition a DIY version, the glassbow.

### Historical Background

The rainbow has fascinated humans throughout history, but its first scientific explanation came from René Descartes in 1637. Using geometric optics, he explained how sunlight is reflected and refracted by spherical water droplets to create the bow shape. However, Descartes could not explain the colors. It was Isaac Newton who, through his experiments with prisms and understanding of dispersion, provided the complete explanation of the rainbow's colors in 1672. Newton showed that white sunlight consists of different colors that are refracted at slightly different angles due to dispersion.

### 13.1 Single Drop Analysis

To understand the rainbow we will have first a look at the reflection of rays from a single droplet.

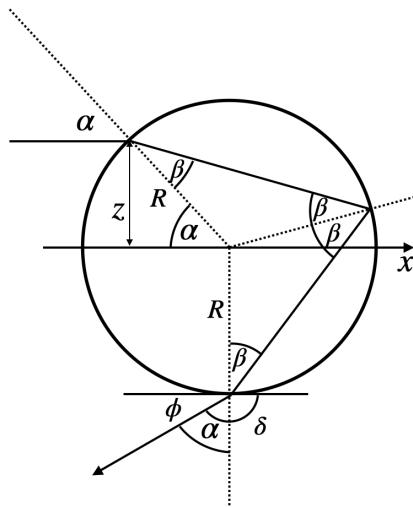


Figure 13.1: Reflection of rays in a single drop

In the sketch above a light ray of white light is entering the droplet under an angle  $\alpha$  to the surface normal on the top. The ray is refracted and enters the droplet under an angle  $\beta$  to the surface normal. The angle can be

calculated from Snell's law

$$n_{\text{air}} \sin(\alpha) = n_{\text{water}} \sin(\beta).$$

Inside the droplet, the ray is now hitting the water/air surface at the backside from which it gets reflected. There, the incident angle is also  $\beta$  and the ray is reflected under an angle  $\beta$  as well. At that point, most of the light will, however, exit the drop on the backside, so that only a small fraction is reflected and traveling further to hit a second time the water/air surface at the angle  $\beta$ . The light refracted out at that point leaves the droplet under an angle  $\alpha$  with the surface normal due to the reversibility of the light path. We are, however, interested in the angle  $\phi$  that the ray makes with the incident direction.

This angle  $\phi$  can be calculated from the above sketch to be

$$\phi = 4\beta - 2\alpha.$$

Since

$$\beta = \sin^{-1} \left( \frac{n_{\text{air}}}{n_{\text{water}}} \sin(\alpha) \right)$$

such that finally

$$\phi = 4 \sin^{-1} \left( \frac{n_{\text{air}}}{n_{\text{water}}} \sin(\alpha) \right) - 2\alpha$$

So let us have a look at this dependence of the deflection angle as a function of the incidence angle.

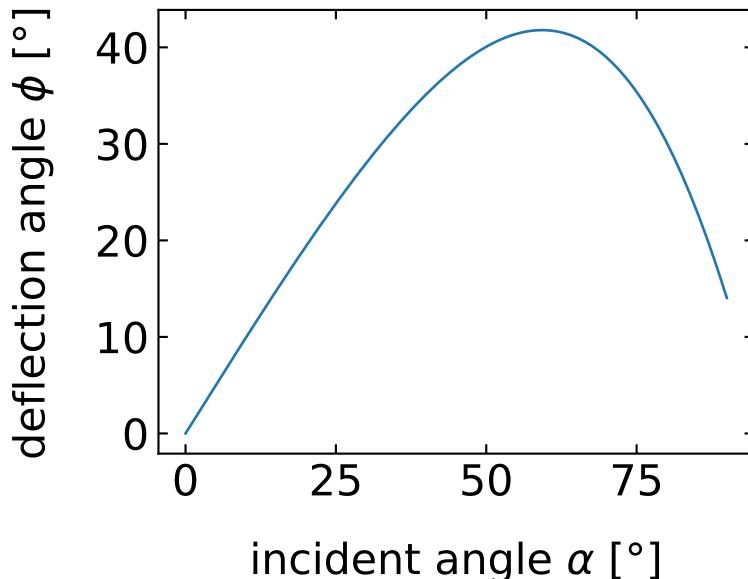


Figure 13.2: Deflection angle of a light ray in a water droplet as a function of the incidence angle. The refractive index of water is taken from the data file `H2O.csv`.

The dependence seems to show a maximum deflection angle at an incidence angle of around  $\alpha = 60^\circ$ . This is an important finding, as the whole appearance of the rainbow depends on that.

Maximum deflection angle 41.78815648670841

## Color Formation and Dispersion

The color of the rainbow is the result of the fact that the maximum deflection angle depends on the color of the light due to the dispersion. Since we have a refraction, reflection and another refraction, the largest maximum deflection angle is observed for red light, while the smallest one appears for blue light. The diagrams below show this result, which is in general true for materials with normal dispersion.

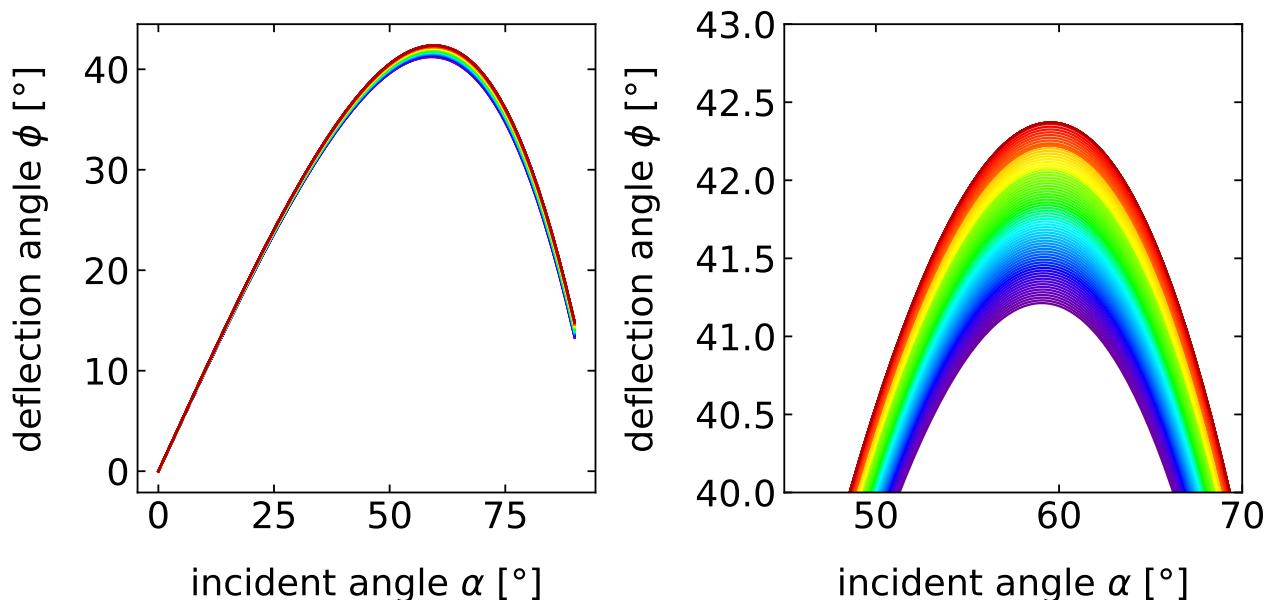


Figure 13.3: Caption

This order of the colors is actually true for all incident angles, which raises the question, why the rainbow is actually colored. The blue color of a certain incidence angle would actually overlap with the green color of a different incidence angle and the red color of an even different incidence angle. If you select a specific outgoing angle under which you observe the rainbow, let us say  $41^\circ$ , then you find under this observation angle all color and, therefore, should observe always white light.

This is actually true if you look at the inside of a rainbow. You clearly recognize that inside the rainbow it is much brighter than outside. Yet when you reach the maximum angle of each color, you have a region, where even for larger angles for the incidence angles, the deflection angle does not change. Thus, if you assume you send in rays at constantly spaced incidence angles, you will have more rays with a deflection angle close to the maximum. The diagram below just counts the number of deflection angles in the different for each color and you clearly see that around the maximum deflection angles for each color you have a strong peak.

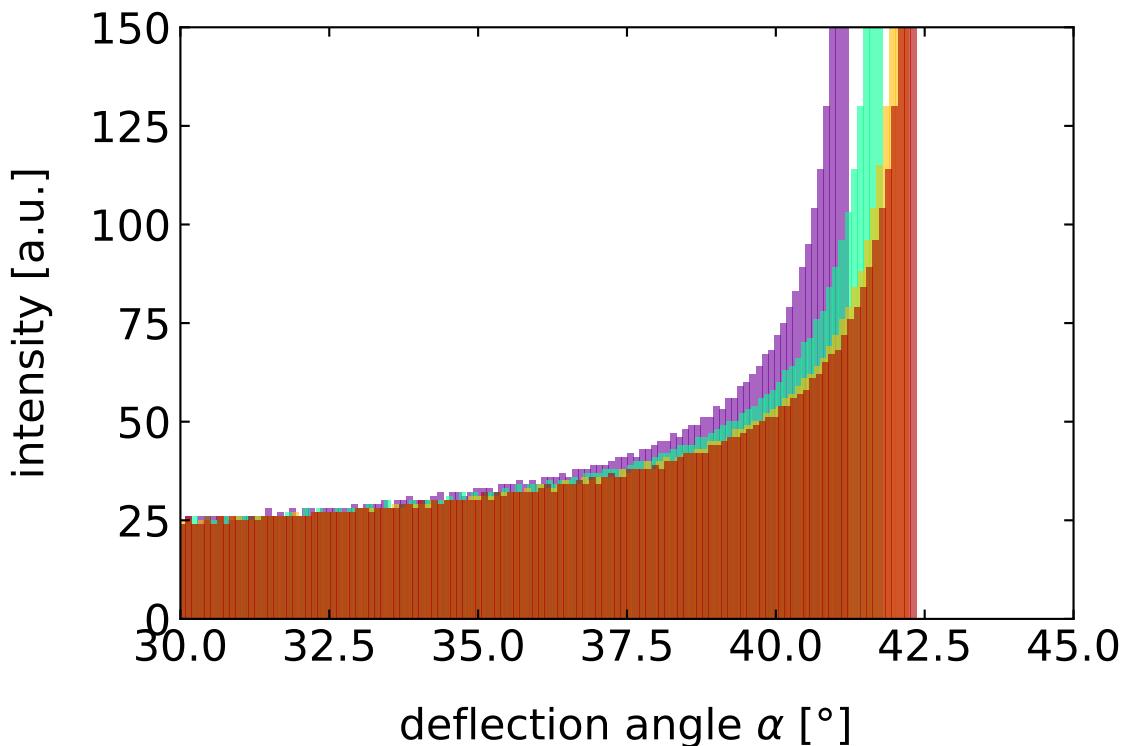


Figure 13.4: Histogram of the deflection angle of light rays in a water droplet for different colors. The refractive index of water is taken from the data file H2O.csv.

Thus, around each droplet on the sky, there is a cone of deflected light reflected back from the sun. On the outside of that cone is red light under an angle of almost  $42^\circ$  while on the inside edge we find blue light and finally white light (see left image below). We just have to connect that to the observer now. This is shown in the right sketch. The rainbow, therefore, results from the fact that we look at different height at different edges of the cone.

## Primary and Secondary Rainbows

The photos below show a remarkable example of both primary and secondary rainbows. Between these two bows, a careful observer will notice a darker region known as Alexander's dark band, first described by Alexander of Aphrodisias in 200 AD. This dark band occurs because no light is scattered back to the observer at angles between approximately  $42^\circ$  (maximum angle of primary rainbow) and  $50^\circ$  (minimum angle of secondary rainbow).

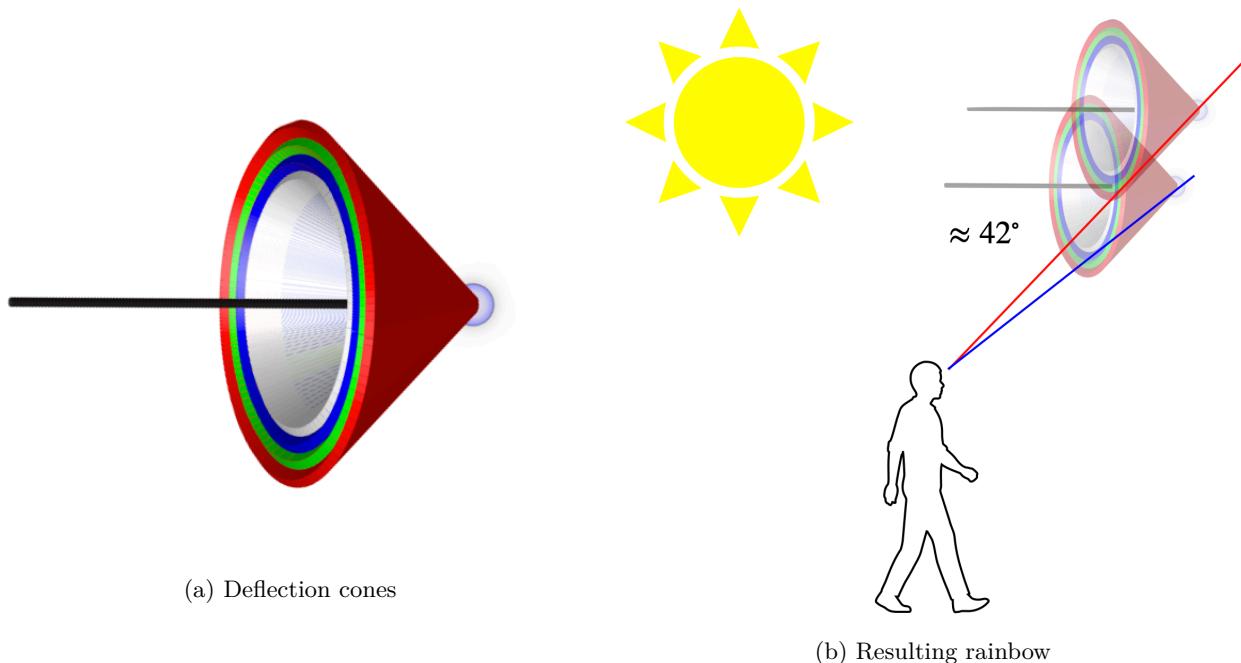


Figure 13.5: Deflection cones of different colors on a single drop in a rainbow (left) and the resulting rainbow as observed from these cones (right).



Figure 13.6: Double rainbow over Grand Canyon. (Frank Cichos)

The primary rainbow results from one internal reflection within each water droplet, while the secondary rainbow involves two internal reflections. This double reflection explains both the reversed color order and the reduced brightness of the secondary rainbow, as each reflection decreases the light intensity.



Figure 13.7: Zoom into the Rainbow photograph showing Alexander's dark band.

## 13.2 Glassbow

A beatiful demonstration of rainbow physics can be created at home using glass beads of about  $200\ \mu m$  diameter (available from our lab). When these beads are placed on a black background and illuminated with a flash lamp, they create what we call a “glassbow” - a rainbow-like pattern produced by glass instead of water droplets.

The main difference between a glassbow and a natural rainbow lies in the observation angle, due to the different refractive index of glass ( $n \approx 1.5$ ) compared to water ( $n \approx 1.33$ ). Using the same deflection angle equation:

$$\phi = 4 \sin^{-1} \left( \frac{n_{\text{air}}}{n_{\text{glass}}} \sin(\alpha) \right) - 2\alpha$$

we can calculate why the glassbow appears at a different angle than its atmospheric counterpart.



Figure 13.8: A glassbow created by light reflection and refraction in microscopic glass beads on a black surface.  
(c) Picture by Axel Märcker



# Chapter 14

## Imaging Errors

During our derivation of the imaging equation for lenses and the lens-maker equation we have been working under the paraxial approximation. This approximation stated, that all rays are close to the optical axis and therefore make only small angles with the surface normals of the curved surfaces of lenses (but also mirrors). If we violate this approximation, i.e. if we use rays, which are incident far from the optical axis or strongly inclined, then we end up with reflections and refraction which do not obey the imaging equation. In addition we have seen that light propagation for different colors is subject to different refractive indices (remember the prism). Thus we will induce aberrations, related to color. According to Seidel, aberration are classified the following way

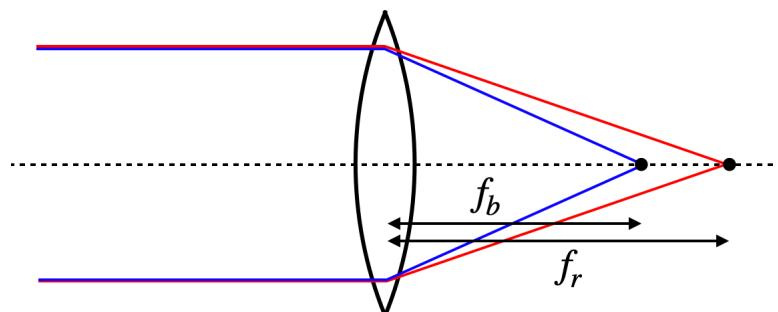
- chromatic aberration
- spherical aberration
- coma
- astigmatism
- field curvature
- field distortion

### 14.1 Chromatic Aberration

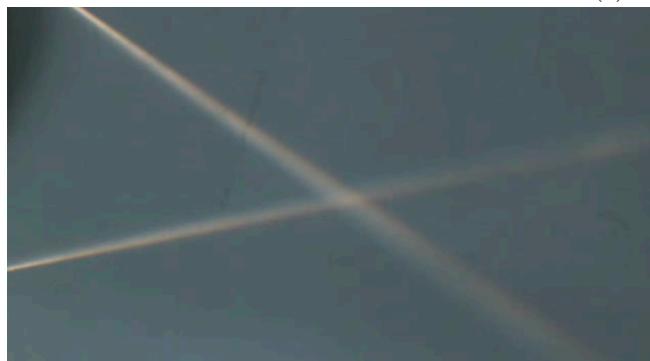
Chromatic Aberration are based on the fact that light of different color has a different speed of propagation and thus also a different refractive index. We experienced that also for the prism, where it was useful to create a spectrograph. Here it is causing colored edges in your image, which you do not want.

As the refractive index for shorter wavelength is typically higher, we expect that the blue color has a shorter focal distance than the red color.

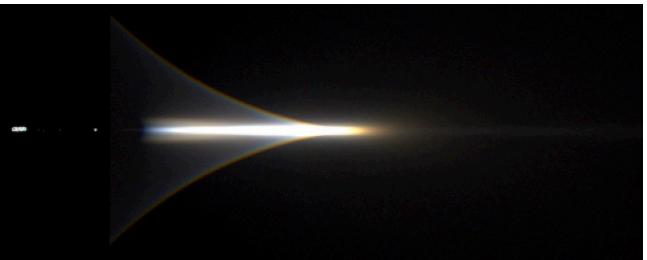
Here is a plot for the variation of the focal distance of a lens with a radius of curvature of 100 mm as a function of the wavelength for three different glasses.



(a) Sketch



(b) Lecture



(c) Rendered

Figure 14.1: Chromatic aberration. Left: Sketch of the chromatic aberration, focusing red light less strong than blue. Middle: Image from the lecture. Right: Rendered image using the refractive index for BK7 glass.

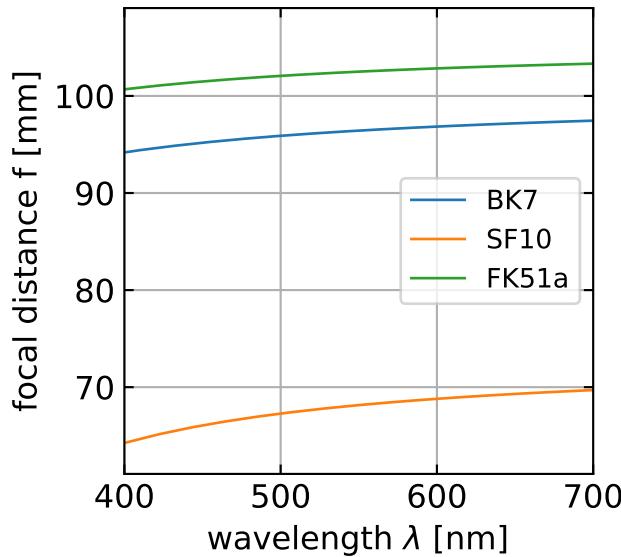


Figure 14.2: Focal distance of a Bk7, SF10 and FK51a lens with a radius of curvature of 100 mm as a function of the wavelength.

Such a chromatic aberrations may be corrected by using a system of two lenses as shown below.

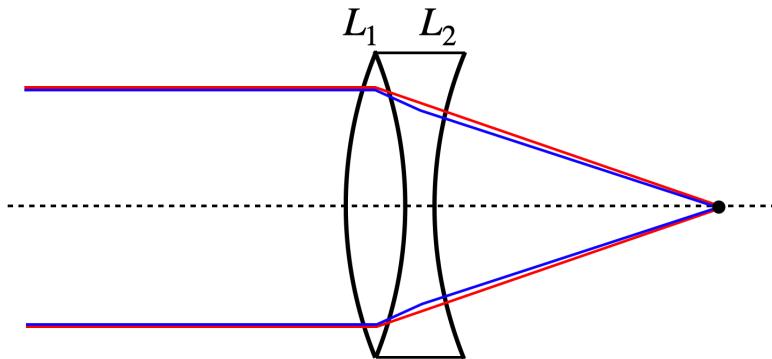


Figure 14.3: Correction of chromatic aberration by a lens doublet with a convex and a concave lens.

Chromatic aberration can be corrected by using an achromatic lens. Achromatic lenses are typically constructed by combining two lenses with different optical properties: a biconvex lens made of crown glass (lower dispersion) bonded to a biconcave lens made of flint glass (higher dispersion). This combination allows for the correction of chromatic aberration. Each lens  $i$  has a focal length according to the lensmaker equation:

$$\frac{1}{f_i} = (n_i - 1)\rho_i$$

where  $\rho_i$  is given by:

$$\rho_i = \frac{(R_{i2} - R_{i1})}{R_{i1}R_{i2}}$$

For a system of two lenses in contact, the total refractive power is:

$$\frac{1}{f} = (n_1 - 1)\rho_1 + (n_2 - 1)\rho_2$$

For color correction, we require equal focusing of red and blue light:

$$(n_{1r} - 1)\rho_1 + (n_{2r} - 1)\rho_2 = (n_{1b} - 1)\rho_1 + (n_{2b} - 1)\rho_2$$

This transforms to:

$$\frac{\rho_1}{\rho_2} = -\frac{n_{2b} - n_{2r}}{n_{1b} - n_{1r}}$$

For the specific geometry shown in the achromat figure where:

- $R_{12} = R_{21}$  (common radius at contact surface)
- $R_{11} = R_1 = -R_{12} = -R_{21}$
- $R_{22} = R_2$

we can express  $\rho_1$  and  $\rho_2$  as:

$$\rho_1 = \frac{2}{R_1} \quad \text{and} \quad \rho_2 = \frac{-1}{R_1} + \frac{1}{R_2}$$

Substituting these into the color correction condition gives us the relationship between  $R_1$  and  $R_2$  needed for an achromatic doublet. Substituting the expressions for  $\rho_1$  and  $\rho_2$  into:

$$\frac{\rho_1}{\rho_2} = -\frac{n_{2b} - n_{2r}}{n_{1b} - n_{1r}}$$

we get:

$$\frac{2/R_1}{-1/R_1 + 1/R_2} = -\frac{n_{2b} - n_{2r}}{n_{1b} - n_{1r}}$$

Let's define the dispersion ratios (typically called V-numbers or Abbe numbers):

$$V_1 = \frac{n_{1r} - 1}{n_{1b} - n_{1r}} \quad \text{and} \quad V_2 = \frac{n_{2r} - 1}{n_{2b} - n_{2r}}$$

Then, after some algebra, the relationship between  $R_1$  and  $R_2$  becomes:

$$\frac{R_2}{R_1} = 1 + 2 \frac{V_2(n_{1r} - 1)}{V_1(n_{2r} - 1)}$$

This equation determines the ratio of radii needed to achieve an achromatic doublet for the chosen glass materials.

### **i** Chromatic Aberration

An optical aberration caused by the wavelength-dependent refractive index of materials, resulting in different colors focusing at different distances from the lens, typically with blue light focusing closer to the lens than red light.

## 14.2 Spherical Aberration

The spherical aberration arises due to the fact that we have always considered a simplification of the angular functions to their first order Taylor series expansion. If the angles of incidence on the spherical surfaces get to large, we cannot do that anymore and need to consider higher order corrections.

The result is that parallel rays which are far from the optical axis are not imaged into the same focal point as the paraxial rays, but to points closer to the lens. You might have all seen such effect also in the case of your empty coffee cup, when the sunlight enters and causes a so-called caustics. This pattern, you observe there is also the result of a spherical aberration. The image below shows the spherical aberration of a lens.

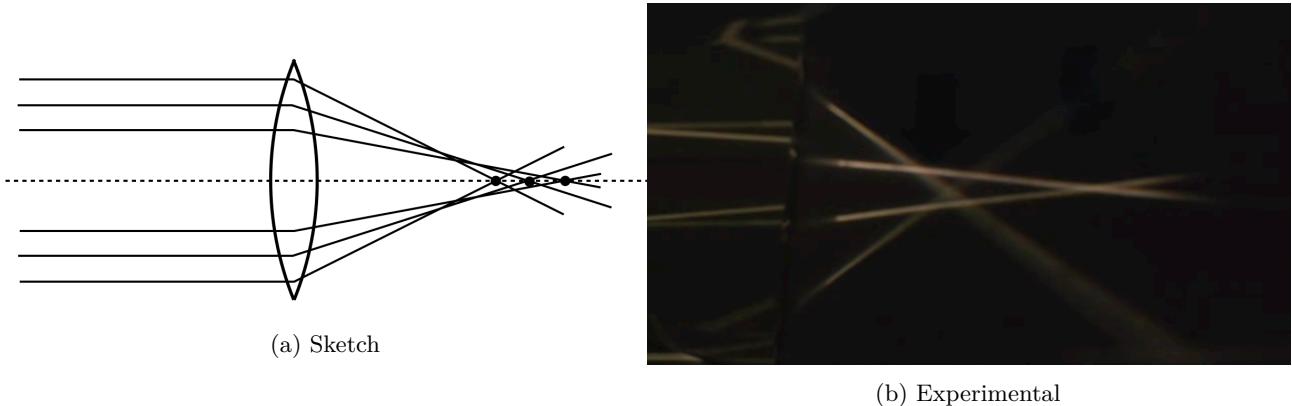


Figure 14.4: Spherical aberration. Left: Schematic illustration showing how parallel rays at different distances from the optical axis focus at different points. Right: Experimental demonstration from the lecture.

To be a bit more quantitative, we would like to reconsider the refraction at a single spherical surface as depicted in the image below.

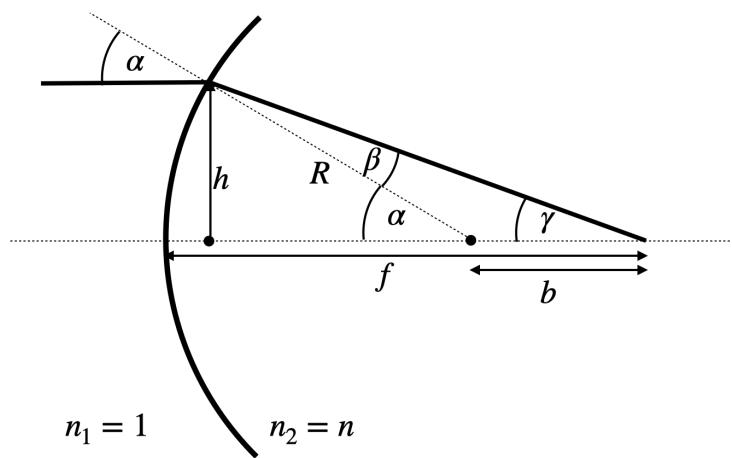


Figure 14.5: Spherical aberration: Theoretical ray tracing showing the focal point variation with incident ray height.

For spherical surfaces, we can derive a more accurate expression for the focal length using the following relations:

$$\sin(\beta) = \frac{\sin(\alpha)}{n}, \quad \sin(\alpha) = \frac{h}{R}, \quad \alpha = \beta + \gamma$$

Using these relations, we obtain  $f = R + b$  and  $b = R \sin(\beta) / \sin(\gamma)$ , which can be transformed into:

$$f = R \left[ 1 + \frac{1}{n \cos(\beta) - \cos(\alpha)} \right]$$

By replacing the cosines with their corresponding expressions, we get:

$$f = R \left[ 1 + \frac{1}{n \sqrt{1 - \frac{h^2}{n^2 R^2}} - \sqrt{1 - \frac{h^2}{R^2}}} \right]$$

Expanding the square roots leads to:

$$f = R \left[ \frac{n}{n-1} - \frac{h^2}{2n(n-1)R^2} \right]$$

This result reveals that the focal length depends on the height  $h$  at which the ray is incident on the spherical surface, similar to the case of concave mirrors. The second term in the square brackets represents this height-dependent contribution, which reduces the focal length when  $h \neq 0$ .

From these relations, we can derive an imaging equation for a single spherical surface:

$$\frac{1}{a} + \frac{n}{b} = \frac{n-1}{R} + h^2 \left[ \frac{1}{2a} \left( \frac{1}{a} + \frac{1}{R} \right)^2 + \frac{n}{2b} \left( \frac{1}{R} - \frac{1}{b} \right)^2 \right]$$

This complex equation for a single surface demonstrates that the image is no longer formed on a plane; instead, its location depends on both  $R$  and  $h$ . This height dependence for a single surface manifests in various image distortions, including field curvature.

### **i** Spherical Aberration

An optical aberration where rays passing through a lens at different distances from the optical axis focus at different points along the axis, with rays through the outer regions of the lens focusing closer to the lens than rays passing near the center.

## 14.3 Field Curvature

The field curvature is related to our calculations of the spherical aberration. We have seen there, that the focal distance depends on the height  $h$  of the rays over the optical axis. This means also means that the image plane is actually not anymore a plane but a curved surface as shown below. The rays incident from point  $A_0$  and  $A_1$  do not meet in the same plane. This plane is even different for meridional and saggital rays. This typically results in the fact, that you may have the center of the image in focus, but not the edges or vice versa.

### **i** Field Curvature

An optical aberration where the image of a flat object is formed on a curved surface rather than a plane, causing the center and edges of the image field to not be simultaneously in focus on a flat detector or screen.

## 14.4 Coma

While our previous discussion focused on rays parallel to the optical axis at varying distances, significant aberrations also occur for rays emanating from off-axis object points (both at finite and infinite distances). One important example of such an aberration is “coma.”

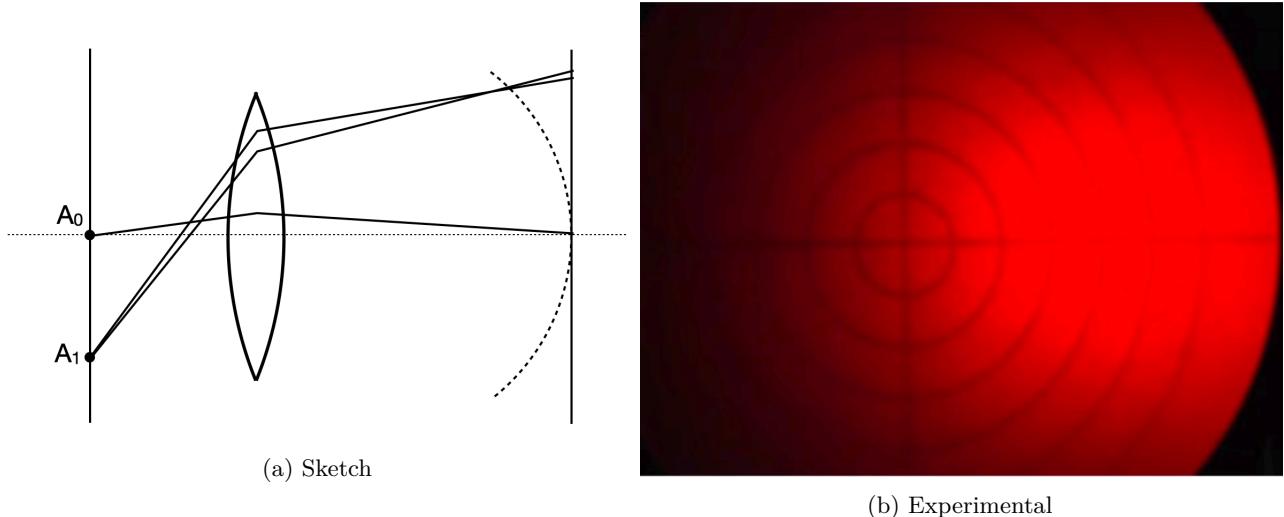


Figure 14.6: Field curvature in optical systems. Left: Schematic diagram showing how a flat object plane is imaged onto a curved image surface (Petzval surface) rather than a flat image plane. This causes different regions of the image to focus at different distances from the lens. Right: Experimental demonstration showing how either the center or the edges of the image can be in focus, but not simultaneously, when using a flat detector. This aberration is particularly noticeable in wide-field imaging systems with simple lenses.

In the case of coma, rays entering the lens at different heights with an angle to the optical axis do not converge to a single point in the image plane. Instead, they create a characteristic comet-shaped intensity distribution, where the light is asymmetrically distributed with a bright head and a diffuse tail pointing radially outward from the optical axis. This asymmetric distribution occurs because rays passing through different zones of the lens experience different effective magnifications, leading to the distinctive comet-like shape that gives this aberration its name.

The severity of coma typically increases with the distance from the optical axis and with larger apertures, making it particularly problematic in wide-field imaging systems or when using large-aperture optics.

### **i** Coma

An optical aberration where rays from an off-axis point source passing through different zones of a lens focus at different positions in the image plane, creating a characteristic comet-shaped intensity distribution with a bright head and a diffuse tail pointing radially outward.

## 14.5 Astigmatism

Astigmatism also occurs when imaging point sources located away from the optical axis. To understand this effect, we can analyze the rays from such a source by separating them into two categories:

1. Rays in the vertical (meridional) plane
2. Rays in the perpendicular (sagittal) plane

Analysis shows that meridional rays focus at a point closer to the lens compared to sagittal rays. This difference in focal positions creates a characteristic pattern in the image: when moving a screen through the focal region, the image of a point source transforms from a horizontal ellipse, through a circular point (at the circle of least confusion), to a vertical ellipse. This variation in image shape occurs because the focal surfaces for the meridional and sagittal rays are curved differently and intersect only at points along the optical axis.

The distance between these two focal surfaces increases with the distance from the optical axis, making astigmatism particularly noticeable for off-axis object points. This aberration is especially significant in systems where

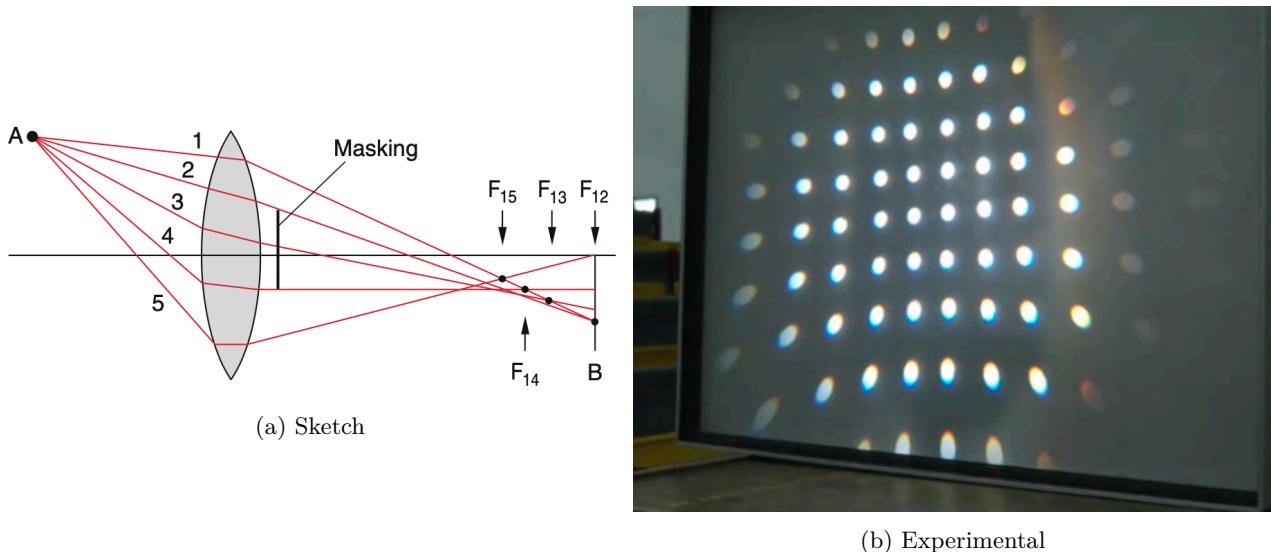


Figure 14.7: Coma aberration in optical systems. Left: Schematic illustration showing how oblique rays entering the lens at different heights focus at different positions in the image plane, creating a characteristic comet-shaped blur (coma). The rays passing through different zones of the lens have different effective focal lengths and magnifications, resulting in the asymmetric image formation. Right: Experimental demonstration from the lecture showing the characteristic comet-like shape of the aberration, where the intensity distribution is asymmetric around the central image point.

the object plane is not perpendicular to the optical axis or when using simple spherical lenses for wide-field imaging.

For an extended image as shown below, this results in the separate focusing of vertical (left) and horizontal lines (right) in the image.

This distortion, i.e. the elliptical shape of the focus has been used advantageously in single molecule microscopy to locate their position along the optical axis, which is typically a challenge for optical microscopy.

### **i** Astigmatism

An optical aberration where rays from an off-axis point source focusing in two perpendicular planes (meridional and sagittal) have different focal lengths, resulting in image points that appear as ellipses oriented either horizontally or vertically, depending on the observation plane.

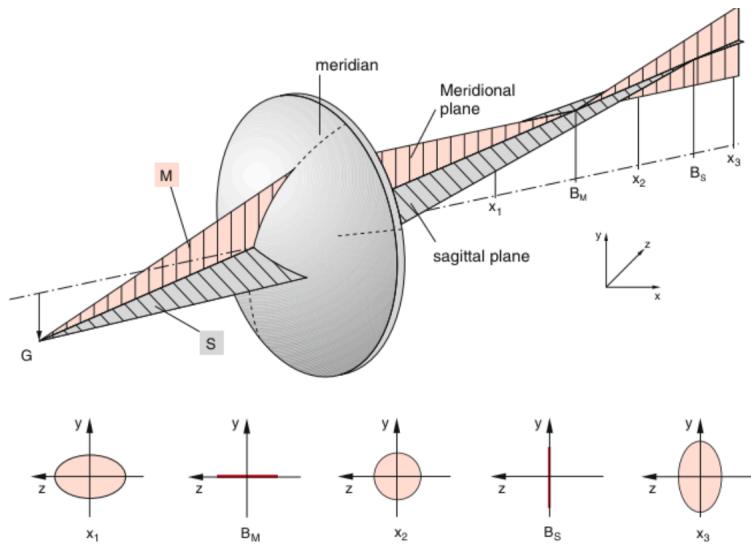
## Distortions

Barrel or cushion shaped distortions in the image are found when inserting apertures in the optical path. This results in the removal of certain ray paths ending up in field distortions.

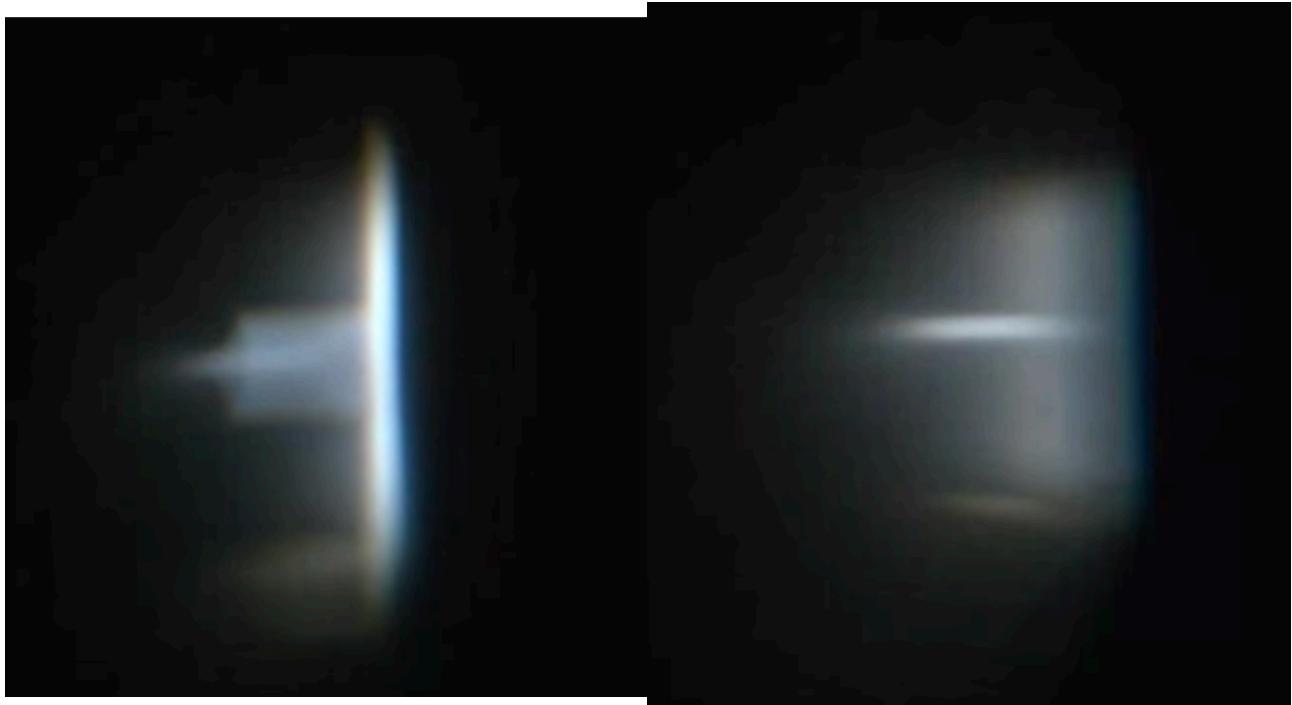
Geometric distortions arise from the position of the aperture relative to the lens and its effect on ray paths. This mechanism can be understood by analyzing how different rays contribute to image formation:

When an aperture is placed behind the lens, rays that pass through it create an image point  $M_1$  that is farther from the optical axis than the ideal image point  $M$  (where  $M$  is determined by the central ray from object point  $A_0$ ). As the object point moves farther from the optical axis, the displacement between  $M_1$  and  $M$  increases proportionally. This progressive displacement transforms a regular grid pattern into a cushion (or pincushion) shape.

Conversely, when the aperture is placed in front of the lens, the opposite effect occurs. The rays that pass through it create an image point closer to the optical axis than the ideal image point, resulting in barrel distortion. The magnitude of this displacement also increases with distance from the optical axis.



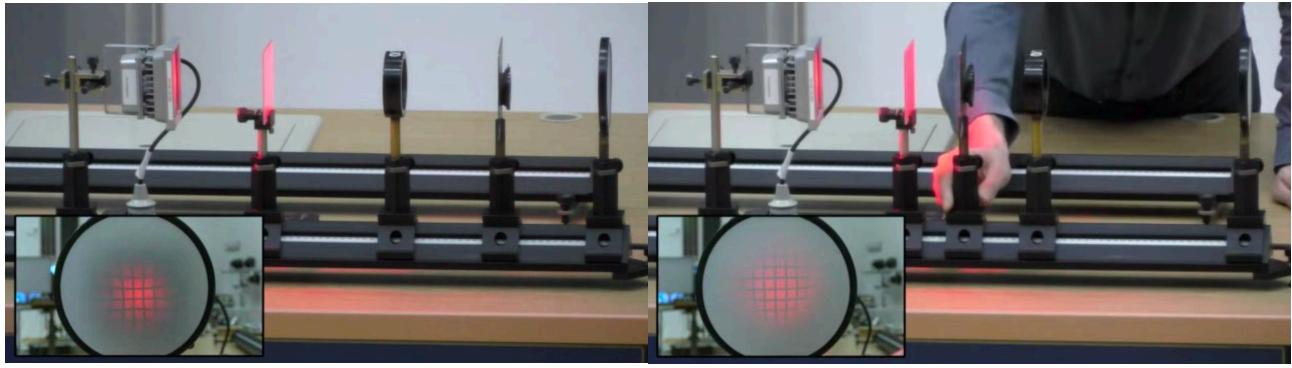
(a) Sketch



(b) Vertical Focus

(c) Horizontal Focus

Figure 14.8: Astigmatism in optical systems. Left: Schematic diagram illustrating how a tilted lens creates different focal planes for rays in different meridional planes. The tangential (vertical) and sagittal (horizontal) rays focus at different distances from the lens. Middle: Experimental image showing the focusing of vertical lines of a test object (letter "F") at one focal position. Right: The same object at a different focal position where horizontal lines are in focus. This demonstrates how astigmatism causes different focal planes for vertical and horizontal features when a lens is tilted relative to the optical axis. The inability to focus both orientations simultaneously is a characteristic feature of astigmatic aberration.



(a) Cushion Distortion

(b) Barrel Distortion

Figure 14.9: Geometric distortions in optical systems. Left: Cushion (or pincushion) distortion, where the magnification increases with distance from the optical axis, causing straight lines to bow inward and creating a cushion-like appearance. Right: Barrel distortion, where the magnification decreases with distance from the optical axis, causing straight lines to bow outward, resembling the shape of a barrel. These distortions maintain image sharpness but alter the geometric shape of the image, particularly noticeable in architectural photography or when imaging regular grid patterns.

These theoretical predictions are confirmed by experimental observations, where:

- A rear aperture produces cushion distortion, causing straight lines to bow inward
- A front aperture produces barrel distortion, causing straight lines to bow outward

The severity of these distortions depends on both the aperture position and the distance of object points from the optical axis.

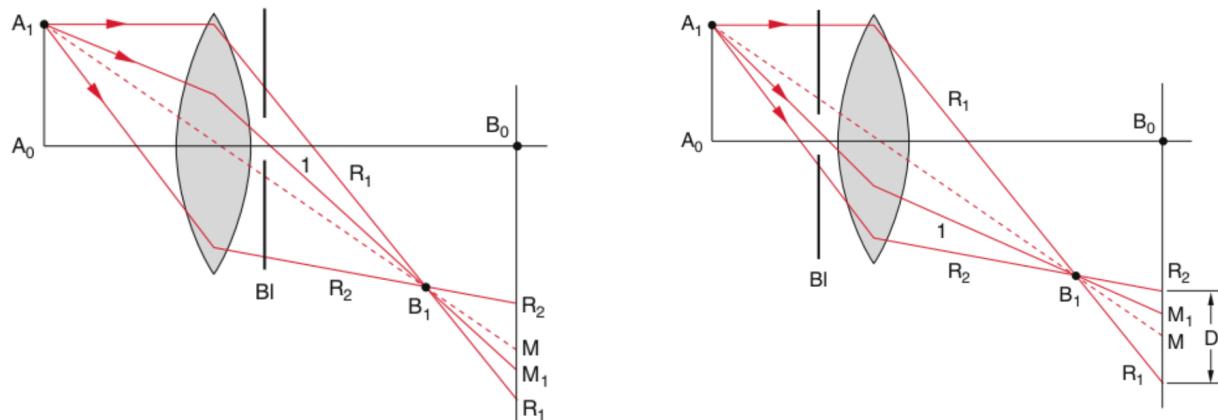


Figure 14.10: Cushion (left) and barrel (right) type of distortions.

### Cushion Distortion

Cushion Distortion (also called Pincushion Distortion) - An optical aberration where straight lines appear to bow inward toward the center of the image, like the sides of a cushion or pincushion. This type of distortion is typically seen in telephoto lenses and makes the center of the image appear to be pinched inward.

### Barrel Distortion

Barrel Distortion - An optical aberration where straight lines appear to bow outward from the center of the image, like a barrel shape. This type of distortion is common in wide-angle lenses and makes the center of the image appear to bulge outward.

### Aberration Characterization and Zernike Polynomials

The Zernike polynomials are a set of orthonormal polynomials that are widely used in optics to describe wavefronts and to characterize optical aberrations. As we did not discuss wavefronts and waves yet, this is a more advanced topic here and only for information. Zernike polynomials are defined over the unit disk and are particularly useful because they are orthogonal under the inner product, which involves integration over the unit circle. This makes them suitable for decomposing a wavefront into a sum of orthogonal modes, each representing a different type of aberration.

The general form of the Zernike polynomials can be expressed in polar coordinates  $(\rho, \phi)$ , where  $\rho$  is the radial distance from the origin (normalized to the unit circle) and  $\phi$  is the azimuthal angle. The Zernike polynomials are defined as:

$$Z_n^m(\rho, \phi) = \begin{cases} R_n^m(\rho) \cos(m\phi) & \text{for } m \geq 0 \\ R_n^{|m|}(\rho) \sin(|m|\phi) & \text{for } m < 0 \end{cases}$$

where  $n$  is a non-negative integer,  $m$  is an integer such that  $n - |m|$  is even and  $0 \leq |m| \leq n$ , and  $R_n^m(\rho)$  is the radial polynomial given by:

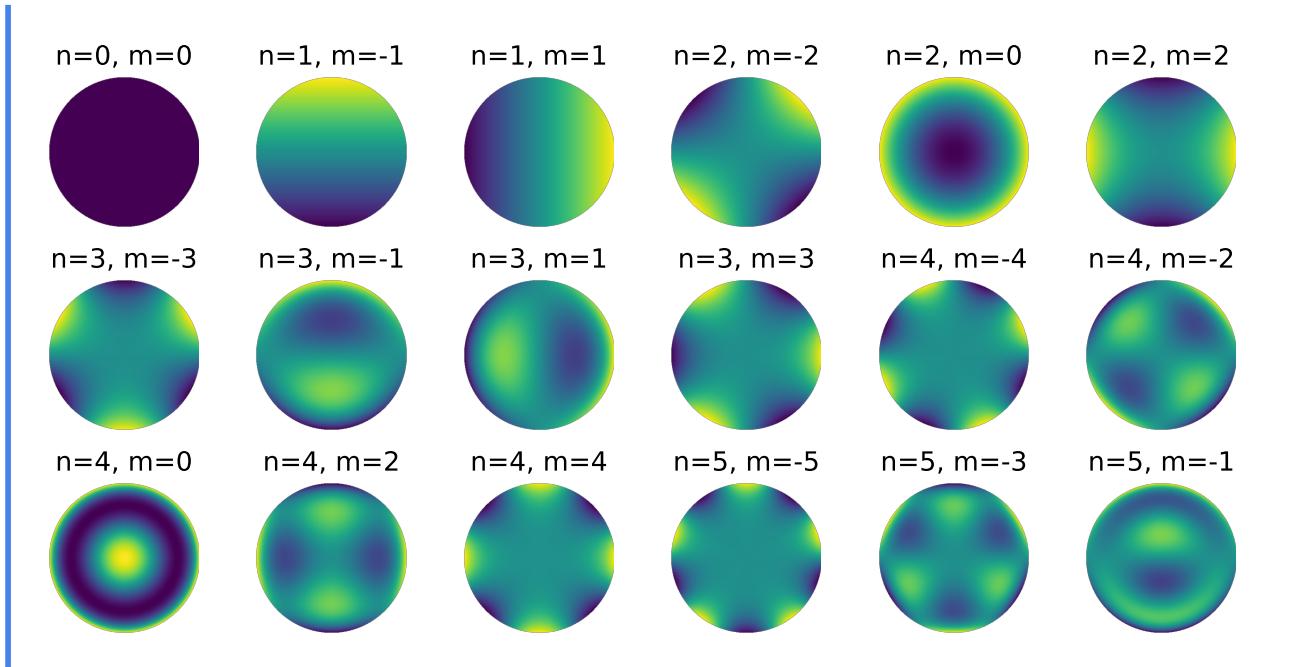
$$R_n^m(\rho) = \sum_{k=0}^{\frac{n-m}{2}} \frac{(-1)^k (n-k)!}{k! (\frac{n+m}{2} - k)! (\frac{n-m}{2} - k)!} \rho^{n-2k}$$

The radial polynomials  $R_n^m(\rho)$  are only dependent on the radial distance  $\rho$ , and they modulate the angular functions  $\cos(m\phi)$  and  $\sin(|m|\phi)$  that describe the azimuthal variation of the wavefront.

The Zernike polynomials are indexed in several ways, with one common method being the Noll index, which provides a single index  $j$  to each polynomial. Another method uses the pair  $(n, m)$  to index the polynomials, where  $n$  indicates the order of the polynomial and  $m$  its azimuthal frequency.

These polynomials are particularly useful in optics and ophthalmology for describing the shape of optical wavefronts and the aberrations of optical systems, including the human eye. They allow for the decomposition of a complex wavefront into simpler, orthogonal components, each corresponding to a specific type of aberration, such as defocus, astigmatism, coma, etc.

The plots below visualize the Zernike Polynomials up to a certain order.



## **Part II**

# **Wave Optics**



# Chapter 15

## Wave Optics

### Historical Development of Scalar Wave Optics

Wave optics represents a fundamental shift in our understanding of light's nature. Here are the key developments in its history:

**Ancient Times - 16th Century: Geometric Optics Dominance** - Ancient civilizations understood basic reflection and refraction. - Focus was primarily on ray-based descriptions of light.

**1660s: Robert Hooke's Wave Theory** - Proposed that light might be a rapid vibrational motion. - Observed interference effects in thin films ("Newton's Rings," though named later).

**1690: Christiaan Huygens' Wave Theory** - Published "Traité de la Lumière" (Treatise on Light). - Introduced the concept of wavefronts. - Developed principle for wave propagation (Huygens' Principle).

**1704: Newton's "Opticks"** - Despite observing interference effects, favored particle theory. - His authority led to wave theory being largely dismissed for nearly a century.

**1801: Thomas Young's Double-Slit Experiment** - Demonstrated interference of light definitively. - Measured wavelengths of different colors. - Introduced the concept of transverse waves.

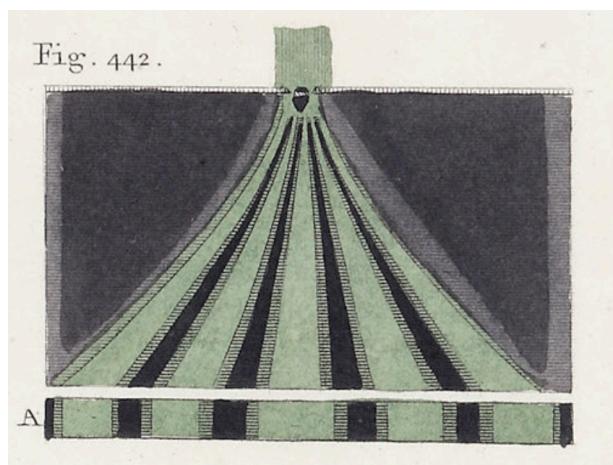


Figure 15.1: Young's double-slit experiment

**1818: Augustin-Jean Fresnel** - Developed comprehensive mathematical theory of diffraction. - Explained polarization through transverse waves. - Created Fresnel equations for reflection and refraction.

Wave optics extends our understanding beyond the limitations of geometric optics by treating light as a wave phenomenon. This approach explains effects that cannot be accounted for by ray tracing alone, such as:

- Interference (the combination of waves)

- Diffraction (the bending of waves around obstacles or through apertures)
- Color (the wavelength-dependent nature of light)

Light is part of the electromagnetic spectrum, which spans an enormous range of frequencies. The visible region, extending approximately from 400 nm (violet) to 700 nm (red), represents only a small fraction of this spectrum. This wave description is essential for understanding many optical phenomena that geometric optics cannot explain, particularly when dealing with structures comparable in size to the wavelength of light.

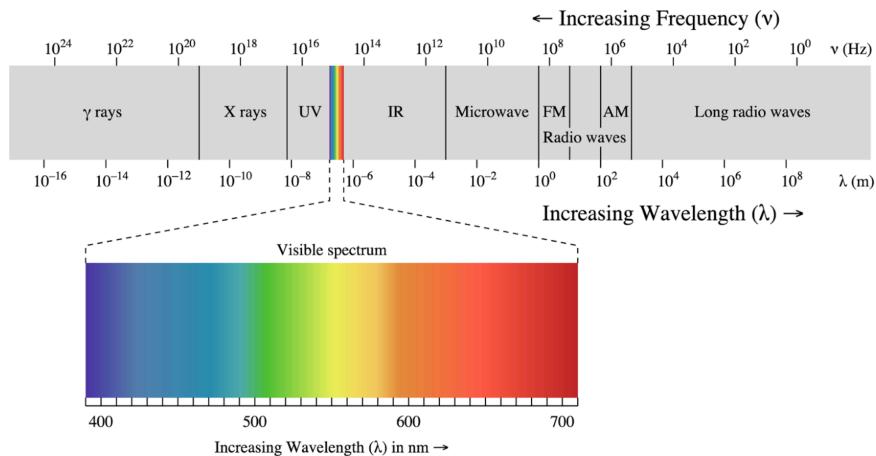


Figure 15.2: Electromagnetic Spectrum with its different regions

In the following, we would like to introduce wave by discarding the fact, that light is related to electric and magnetic fields. This is useful as the vectorial nature of the electric and magnetic field further complicates the calculations, but we do not need those yet. Accordingly we also do not understand how light really interacts with matter and we therefore have to introduce some postulates as well.

## 15.1 Postulates of Wave Optics

### **i** Wave

A wave corresponds to a physical quantity which oscillates in space and time. Its energy current density is related to the square magnitude of the amplitude.

### Wave equation

$$\nabla^2 u - \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} = 0$$

where the Laplace operator  $\nabla^2$  is defined as:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

The wave equation is a linear differential equation, which implies that the superposition principle holds. Specifically, if  $u_1(\mathbf{r}, t)$  and  $u_2(\mathbf{r}, t)$  are solutions of the wave equation, then any linear combination:

$$u(\mathbf{r}, t) = a_1 u_1(\mathbf{r}, t) + a_2 u_2(\mathbf{r}, t)$$

is also a solution, where  $a_1$  and  $a_2$  are arbitrary constants.

## Monochromatic Wave

A monochromatic wave consists of a single frequency  $\omega$ . By definition, such a wave must be infinite in time and free from phase disturbances (such as sudden jumps). The mathematical expression for a monochromatic wave is:

$$u(\mathbf{r}, t) = a(\mathbf{r}) \cos(\omega t + \phi(\mathbf{r}))$$

where:

- $a(\mathbf{r})$  represents the amplitude
- $\phi(\mathbf{r})$  represents the spatial phase
- $\omega$  represents the angular frequency

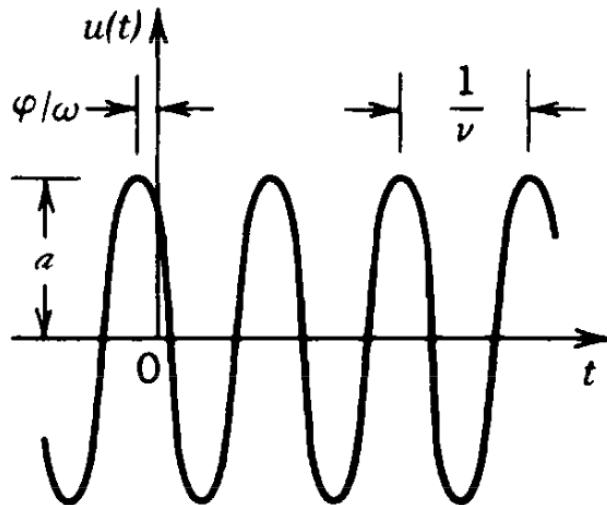


Figure 15.3: Representation of a wavefunction over time (constant position) denoting the phase  $\phi$  and the period  $T = 1/\nu$

## Complex Amplitude

The wave can be represented in complex form as:

$$U(\mathbf{r}, t) = a(\mathbf{r})e^{i\phi(\mathbf{r})}e^{i\omega t}$$

This is known as the complex wavefunction.

**Note**

A phasor displays the complex amplitude with magnitude and phase as a vector in the complex plane.

The relationship between the complex and real wavefunctions is:

$$u(\mathbf{r}, t) = \operatorname{Re}\{U(\mathbf{r}, t)\} = \frac{1}{2}[U(\mathbf{r}, t) + U^*(\mathbf{r}, t)]$$

The complex wavefunction satisfies the same wave equation:

$$\nabla^2 U - \frac{1}{c^2} \frac{\partial^2 U}{\partial t^2} = 0$$

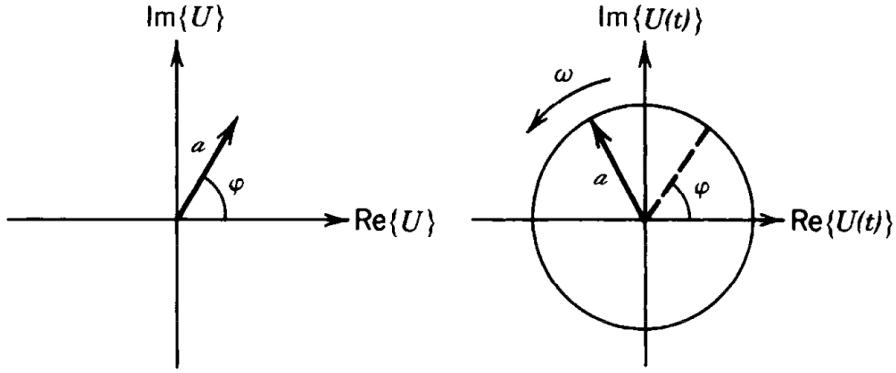


Figure 15.4: Phasor diagram of the complex amplitude  $U(\mathbf{r})$  (left) and  $U(t)$  (right)

We can separate the complex wavefunction into spatial and temporal components:

$$U(\mathbf{r}, t) = U(\mathbf{r})e^{i\omega t}$$

where

$$U(\mathbf{r}) = a(\mathbf{r})e^{i\phi(\mathbf{r})}$$

Here,  $\phi$  represents the spatial phase of the wavefunction. Substituting this into the wave equation and noting that the time derivatives bring down factors of  $i\omega$ :

$$\begin{aligned} \nabla^2[U(\mathbf{r})e^{i\omega t}] - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}[U(\mathbf{r})e^{i\omega t}] &= 0 \\ \nabla^2U(\mathbf{r})e^{i\omega t} + \frac{\omega^2}{c^2}U(\mathbf{r})e^{i\omega t} &= 0 \end{aligned}$$

The time dependence  $e^{i\omega t}$  factors out, leaving us with the **Helmholtz equation**:

$$\nabla^2U(\mathbf{r}) + k^2U(\mathbf{r}) = 0$$

where  $k = \omega/c$  is the wave number. This equation describes the spatial behavior of monochromatic waves.

### Intensity of Waves

The intensity of a wave at position  $\mathbf{r}$  and time  $t$  is defined as:

$$I(\mathbf{r}, t) = 2\langle u^2(\mathbf{r}, t) \rangle$$

where  $I$  is measured in units of  $[\frac{W}{m^2}]$ . The angle brackets  $\langle \dots \rangle$  represent a time average over one oscillation cycle of  $u$ . For visible light, this averaging occurs over an extremely brief period - for example, light with a wavelength of 600 nm has a cycle duration of just 2 femtoseconds.

The optical power  $P$  of a wave can be calculated by integrating the intensity over a surface area  $A$ :

$$P = \int_A I(\mathbf{r}, t) dA$$

Inserting the separation of the complex wavefunction into spatial and temporal components leads to the following expression for the intensity:

$$I(\mathbf{r}) = |U(\mathbf{r})|^2$$

Thus the physical quantity forming the spatial and temporal oscillation of the wavefunction is also providing the intensity of the wave when its magnitude is squared. This is a fundamental property of wavefunctions and for example not the case when temperature oscillates in space and time in a medium.

### Wavefronts

Wavefronts are surfaces in space where the phase is constant:

$$\phi(\mathbf{r}) = \text{const}$$

Typically, this constant is chosen to represent points of maximum spatial amplitude, such that:

$$\phi(\mathbf{r}) = 2\pi q$$

where  $q$  is an integer.

The direction normal to these wavefronts can be described by the gradient vector:

$$\mathbf{n} = \nabla\phi = \left( \frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y}, \frac{\partial\phi}{\partial z} \right)$$

This vector  $\mathbf{n}$  is always perpendicular to the wavefront surface and points in the direction of wave propagation. The evolution of these wavefronts in time provides important information about the wave's propagation characteristics.

## 15.2 Plane Waves

A plane wave represents a fundamental solution of the homogeneous wave equation. In its complex form, it is expressed as:

$$U(\mathbf{r}, t) = A e^{-i\mathbf{k}\cdot\mathbf{r}} e^{i\omega t} \quad (15.1)$$

where:

- The first exponential term contains the spatial phase
- The second exponential term contains the temporal phase
- $A$  is the (potentially complex) amplitude of the plane wave

The wavefront of a plane wave is defined by:

$$\mathbf{k} \cdot \mathbf{r} = 2\pi q + \arg(A)$$

where  $q$  is an integer. It just means that the projection of the position vector  $\mathbf{r}$  onto the wavevector  $\mathbf{k}$  is a multiple of  $2\pi$ . This equation describes a plane perpendicular to the wavevector  $\mathbf{k}$ . Adjacent wavefronts are separated by the wavelength  $\lambda = 2\pi/k$ , where  $k$  represents the spatial frequency of the wave oscillation.

The spatial component of the plane wave is given by:

$$U(\mathbf{r}) = A e^{-i\mathbf{k}\cdot\mathbf{r}} \quad (15.2)$$

In vacuum, the wavevector  $\mathbf{k} = \mathbf{k}_0$  is real-valued and can be written as:

$$\mathbf{k}_0 = \begin{pmatrix} k_{0x} \\ k_{0y} \\ k_{0z} \end{pmatrix} \quad (15.3)$$

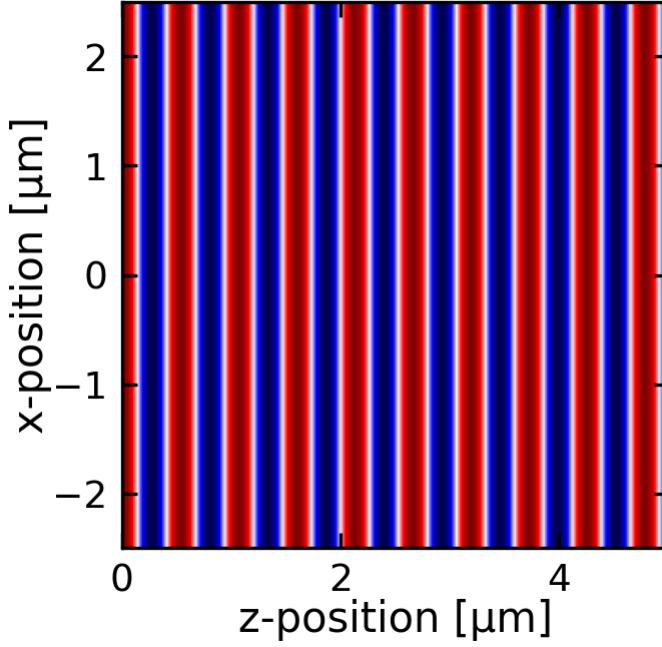


Figure 15.5: Plane wave

### 15.3 Dispersion Relation

Using the plane wave solution

$$U(\mathbf{r}, t) = A e^{-i\mathbf{k}\cdot\mathbf{r}} e^{i\omega t} \quad (15.4)$$

we can write down the sum of the spatial and temporal phase as

$$\phi(r, t) = \omega t - \mathbf{k} \cdot \mathbf{r}$$

If we select a point on the wavefront  $\mathbf{r}_m$ , and follow that over time, the phase  $\phi(t) = \text{const.}$  Taking the time derivative results in

$$\mathbf{k} \cdot \frac{d\mathbf{r}_m}{dt} = \omega$$

If we choose the direction of the wavevector for measuring the propagation speed, i.e.  $\mathbf{r}_m = r_m \mathbf{e}_k$  then we find for the propagation speed

$$\frac{dr_m}{dt} = \frac{\omega}{k}$$

or in vacuum

$$c_0 = \frac{\omega}{k_0} \quad (15.5)$$

This fundamental relationship connects:

- The momentum ( $k$ ),
- The energy ( $\omega$ )

and is called a dispersion relation despite the fact, that we do not really understand why those quantities are related to energy and momentum.

**i** Note

Light in free space exhibits a linear dispersion relation, i.e. the frequency of light changes linearly with the wavevector magnitude.

Note that if we choose a different propagation direction  $\mathbf{e}$  than the one along the wavevector  $\mathbf{e}_k$ , we can write the phase velocity as

$$\mathbf{k} \cdot \mathbf{e} \frac{dr}{dt} = k \cos(\angle \mathbf{k}, \mathbf{e}) \frac{dr}{dt} = \omega$$

or

$$\frac{dr}{dt} = \frac{\omega}{k \cos(\angle \mathbf{k}, \mathbf{e})}$$

which means that if you observe the wavepropagation not in the direction of the wavevector, the phase velocity is actually bigger than the speed of light and even tends to infinity if the angle between the wavevector and the observation direction tends to 90°.

## 15.4 Propagation in a Medium

When a wave propagates through a medium:

1. The frequency  $\omega$  remains constant (determined by the source)
2. The wave speed changes according to:

$$c = \frac{c_0}{n}$$

where  $n$  is the refractive index of the medium

This leads to changes in:

- the wavelength, which becomes shorter in the medium

$$\lambda = \frac{\lambda_0}{n}$$

- the length of the wavevector, which increases in the medium

$$k = nk_0$$

## 15.5 Snells Law

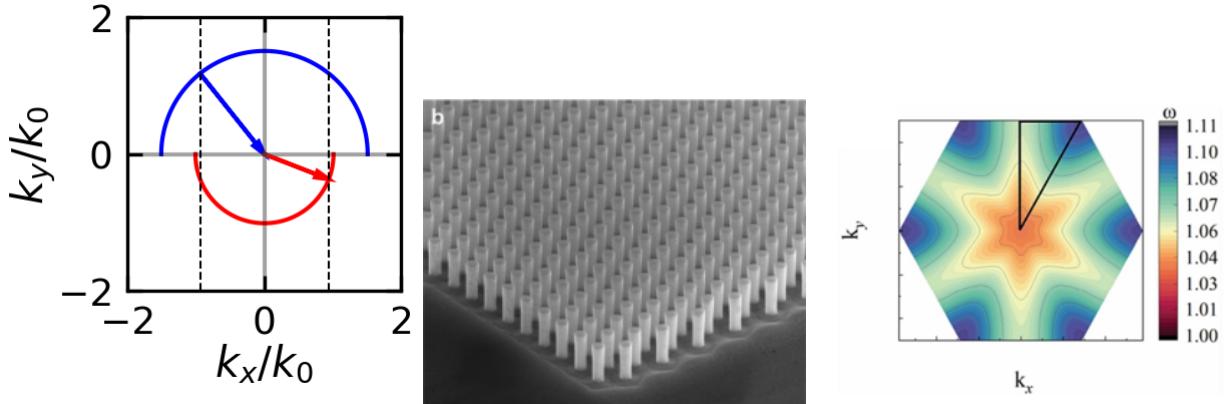
The change in the length of the wavevector has some simple consequence for Snells law. We can write Snells law as

$$n_1 k_0 \sin(\theta_1) = n_2 k_0 \sin(\theta_2)$$

where  $k_0$  is the wavevector length in vacuum. As the  $n_1 k_0$  is the magnitude of the wavevector in medium 1, and  $n_2 k_0$  is the magnitude of the wavevector in medium 2, we can rewrite Snells law as

$$k_1 \sin(\theta_1) = k_2 \sin(\theta_2)$$

which means that the component of the wavevector parallel to the interface is conserved. If the wavevector has constant length then the wavevector incident at different angles is between a point on a circle and the origin in the diagram below. The circle corresponds to an isofrequency surface.



(a) Snells law construction using the conservation of the wavevector component parallel to the interface. The vertical dashed lines indicate the parallel component of the wavevector in the two media.

(a) Electron microscopy image of a 2D photonic crystal

(a) Isofrequency surfaces of a photonic crystal

Isofrequency surfaces can have non-spherical shape. In anisotropic media, they can be ellipsoids. In photonic crystals, i.e. crystals with a periodic structure on the scale of the wavelength, they can have a more complex shape.

## 15.6 Spherical Waves

A spherical wave, like a plane wave, consists of spatial and temporal components, but with wavefronts forming spherical surfaces. For spherical waves,  $|\mathbf{k}||\mathbf{r}| = kr = \text{const}$ . Given a source at position  $\mathbf{r}_0$ , the spherical wave can be expressed as:

$$U = \frac{A}{|\mathbf{r} - \mathbf{r}_0|} e^{-ik|\mathbf{r} - \mathbf{r}_0|} e^{i\omega t} \quad (15.6)$$

! Important

The  $1/|\mathbf{r} - \mathbf{r}_0|$  factor in the amplitude is necessary for energy conservation - ensuring that the total energy flux through any spherical surface centered on the source remains constant.

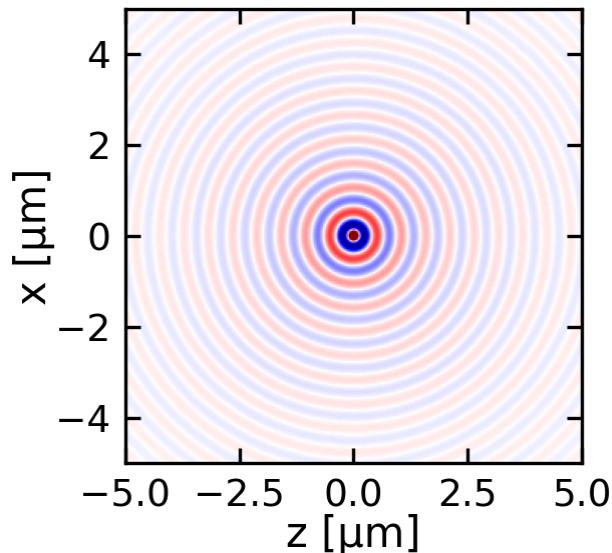


Figure 15.9: Spherical wave propagation. The wave is emitted from the origin and propagates in the positive z-direction. The wavefronts are spherical surfaces. The wave is visualized in the xz-plane.

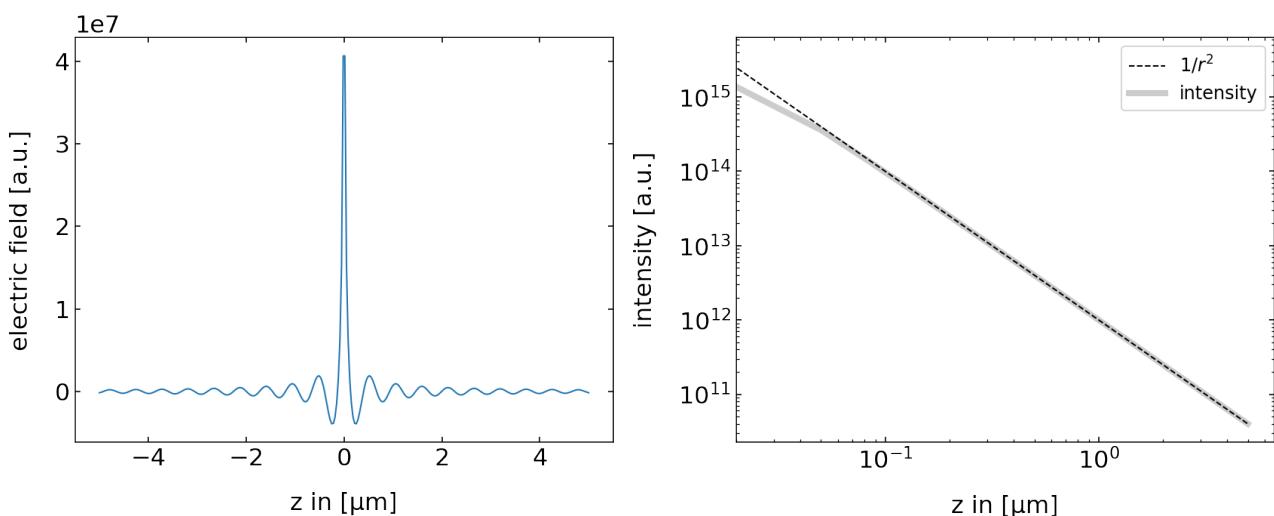


Figure 15.10: Spherical wave amplitude and intensity of the spherical wave as a function of distance from the source

The plots demonstrate that:

- The field amplitude decays rapidly with distance
- The intensity follows a  $1/r^2$  law (with slight deviations at small distances due to discretization artifacts)

Note: The direction of wave propagation can be reversed by changing the sign of the wavenumber  $k$ .

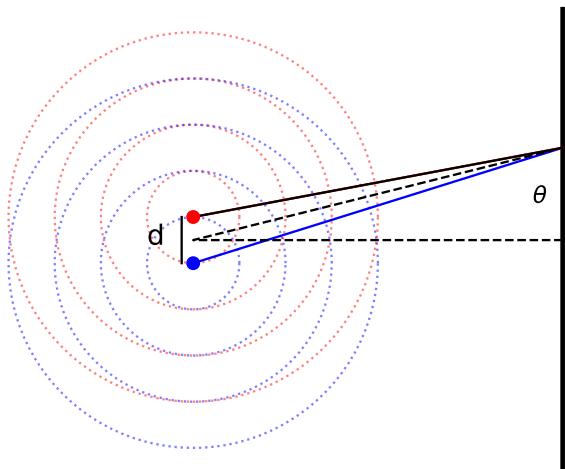


# Chapter 16

## Double Slit Interference

### Two Point Sources - Double Slit Interference

The interference of two point sources is a classic example of wave interference. It is often referred to as double slit interference. The interference pattern is created by two point sources that emit waves with the same wavelength and amplitude. The interference of the two waves depends then on the path length difference between the two waves



(a) Double slit interference as the interference from two point sources on the left and the wave amplitudes on the right. The interference pattern is created by two point sources that emit waves with the same wavelength and amplitude. The interference of the two waves depends then on the path length difference between the two waves.

The interference pattern depends on the relative phase of the two waves. The phase difference can be calculated from the path length difference between the two waves and the path length difference can be calculated considering the angle  $\theta$  between the line connecting the two sources and the line connecting the sources to the point on the screen. The path length difference is then given by

$$\Delta s = s_2 - s_1 = d \sin(\theta)$$

and consequently the phase difference is given by

$$\Delta\phi = \frac{2\pi}{\lambda} \Delta s = \frac{2\pi}{\lambda} d \sin(\theta)$$

### **i** Correct path length difference

The path length difference given above is only approximately correct. The exact calculation would involve the geometry of the problem and the path length difference would be calculated as  $\Delta s = \sqrt{d^2 + L^2 - 2dL \cos(\theta)} - \sqrt{d^2 + L^2 - 2dL \cos(\theta)}$ . Note that when observing the pattern on the screen with the help of a lens that is placed at the focal distance from the screen, the two path would correspond to parallel rays and the path length difference assumed above would be correct.

As constructive interference occurs when the phase difference is a multiple of, i.e.  $m2\pi$ , the constructive interference will be observed when

$$\sin(\theta) = m \frac{\lambda}{d}$$

### **i** Constructive interference from two sources

Constructive interference from two sources separated by a distance  $d$  will be observed at an angle  $\theta$  when  $\sin(\theta) = m \frac{\lambda}{d}$ , where  $m$  is an integer. The orders of the constructive interference are labeled as  $m = 0, 1, 2, 3, \dots$  and the  $m = 0$  constructive interference is the central maximum. The first order constructive interference angle with scale with the wavelength as  $\lambda$  and the inverse distance between the sources as  $1/d$ , i.e. larger wavelength will lead to larger angles and larger source separation will lead to smaller angles. This scaling is a common feature in many interference applications and the foundation of spectroscopy!

If the screen is at a distance  $L$  from the sources, the angle  $\theta$  can be calculated as  $\theta = \arctan(y/L)$ , where  $y$  is the distance from the center of the screen.

Inserting the phase difference into the intensity formula, we get

$$I = 2I_1 + 2I_2 + 2\sqrt{I_1 I_2} \cos\left(\frac{2\pi d}{\lambda} \sin(\theta)\right)$$

or when assuming the same intensity from the two sources

$$I = 4I_0 \cos^2\left(\frac{d\pi}{\lambda} \sin(\theta)\right)$$

The plot below shows this intensity pattern for two sources separated by a distance  $d = 2 \mu\text{m}$  and a wavelength of  $\lambda = 0.532$  micrometers.

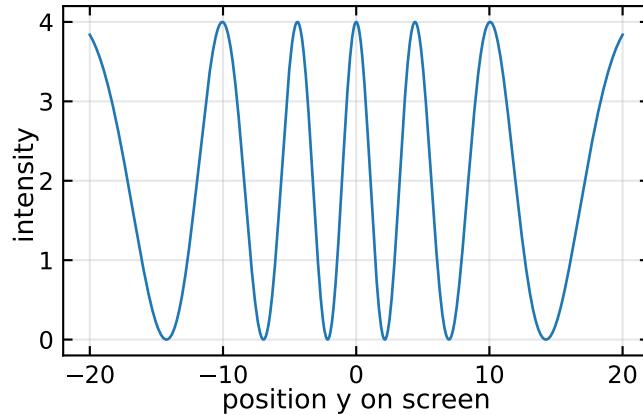


Figure 16.2: Intensity pattern of two sources at a screen at a distance  $L$ . The sources are separated by a distance  $d$  and the wavelength of the waves is  $\lambda$ .

The interference from two point sources has immediate consequences for the resolution of optical instruments. The resolution of an optical instrument is the ability to distinguish between two closely spaced objects. The Abbe criterion states the the minimum resolvable distance  $d$  between two objects is given by

$$d = \frac{\lambda}{2 \sin(\theta)}$$

where  $\lambda$  is the wavelength of the light and  $\theta$  is the angle subtended by the two objects at the lens. The Abbe criterion is derived from the condition that the microscopy lens has to collect at least the first minimum of the interference pattern of the two objects. This first destructive interference is the information that is needed to separate the two objects from one object.

### **i** Fresnel double mirror and biprism experiment

One of the first experiments that demonstrated the wave nature of light was the Fresnel double mirror experiment. In this experiment, a light source is placed in front of two tilted mirrors. The light is reflected from this mirror to a screen. The interference pattern that is observed on the screen is due to the interference of the light that is reflected from the two mirrors. The interference pattern is similar to the one that is observed in the Young's double slit experiment as the two mirrors "imitate" two virtual light sources behind the tilted mirror.

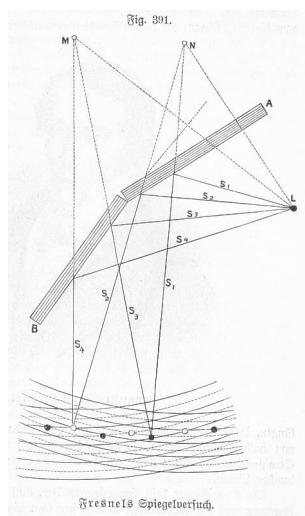


Figure 16.3: Fresnel double mirror experiment

A similar experiment is done with the so-called Fresnel biprism. The Fresnel biprism is a prism that is cut in half and the two halves are separated by a small distance. The light that is incident on the biprism is split into two beams that are then recombined on a screen. The interference pattern that is observed on the screen is due to the interference of the two beams.

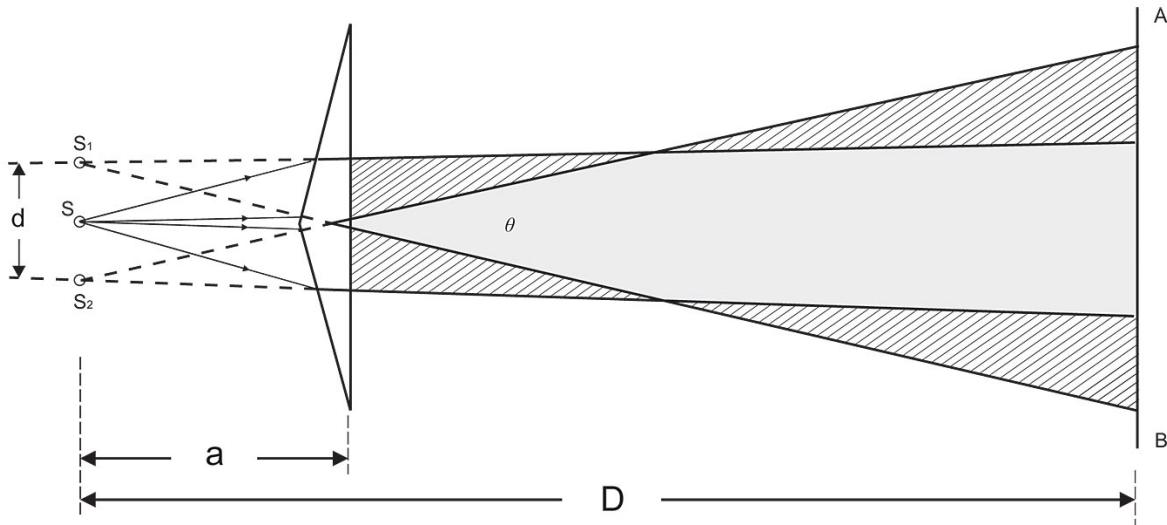


Figure 16.4: Fresnel double mirror experiment

# Chapter 17

## Thin Film Interference

The reflection and transmission of waves on a thin film can also be regarded as an interference of two waves. A light wave is incident on a thin film as depicted below. A part of the wave is reflected on the first boundary (1). Another part is transmitted through the first boundary and reflected at the second boundary to be transmitted in the same direction (2) as the first reflected part. Note that the lines and arrows denote the direction of the wavevector  $\vec{k}$  of the partial waves.

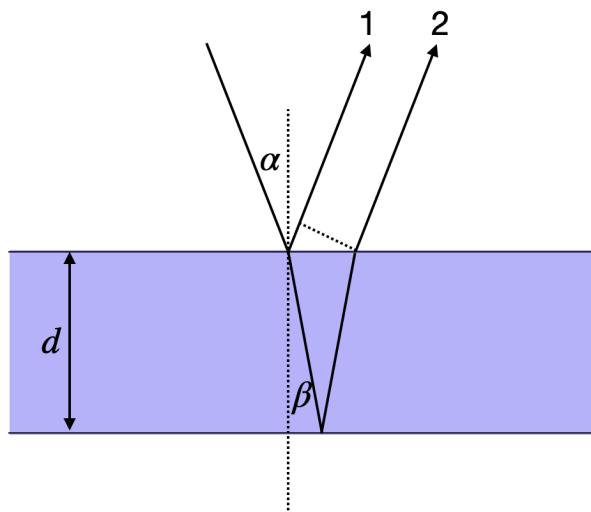


Figure 17.1: Interference on a thin film considering two partial waves.

This picture of a single reflection at each interface is a simplification. In reality, we would have multiple reflections occurring at both interfaces, leading to an infinite number of partial waves. However, for interfaces with weak reflection coefficients (like the air/glass interface where  $r = 4\%$ ), the contribution of higher-order reflections becomes negligible. After two reflections, the amplitude is already reduced to 4% of 4% = 0.16% of the incident wave. Therefore, considering just the first two partial waves provides a good approximation for weak reflections.

For the geometry shown in the figure above, we consider a medium with refractive index  $n_1$  surrounding a film with  $n_2$ . The path difference  $\Delta s$  between waves 1 and 2 consists of two contributions:

$$\Delta s = \frac{2n_2 d}{\cos(\beta)} - 2d \tan(\beta) \sin(\alpha)$$

The first term represents the optical path inside the film (wave 2), while the second term accounts for the additional path of wave 1 after reflection (shown by the dotted line).

Using Snell's law,  $n_1 \sin(\alpha) = n_2 \sin(\beta)$ , and setting  $n_1 = 1$  and  $n_2 = n$ , we can simplify the path difference:

$$\Delta s = \frac{2nd}{\cos(\beta)} - \frac{2nd \sin^2(\beta)}{\cos(\beta)} = 2nd \cos(\beta) = 2d \sqrt{n^2 - \sin^2(\alpha)}$$

The total phase difference  $\Delta$  between the waves includes both the path difference and interface effects:

$$\Delta\phi = \frac{2\pi}{\lambda} \Delta s + \pi$$

The additional term arises from the reflection at the first interface where  $n_1 < n_2$ . This phase jump occurs whenever light reflects from an optically denser medium. No such phase jump occurs at the second interface where  $n_2 > n_1$ .

### Phase Jump at Boundaries

Wave may experience phase jumps when being reflected.

A light wave will experience a phase jump of  $\pi$  when being reflected by a medium of higher refractive index.

A light wave will experience no phase jump when being reflected by a medium of lower refractive index. The physical reasons will be covered when we deal with the Fresnel formulas in electromagnetic optics.

To get to know the properties of thin film interference a bit better we consider the normal incidence  $\alpha = 0$ , which leaves us with

$$\Delta\phi = \frac{2\pi}{\lambda} 2dn + \pi$$

In case we are searching for constructive interference, this phase shift should correspond to an integer multiple of  $2\pi$ , e.g.  $\Delta\phi = m2\pi$ . From the last equation we see already, that for  $d = 0$ , we have in principle a residual phase shift of  $\pi$ , meaning that there is only destructive interference. Yet a film thickness of zero does not really make sense.

We would like to discuss two different situations in the following in an example. For that we either look at the thickness under which a constructive interference at a wavelength of  $\lambda$  occurs, or we ask what kind of wavelength do show constructive interference for a fixed thickness.

## Fixed Wavelength

For a fixed wavelength of  $\lambda$  we obtain a corresponding thickness for the constructive interference of

$$d = \frac{(2m-1)\lambda}{4n}$$

Thus we see constructive interference at from regions of the same thickness  $d$ . This means that the interference fringes correspond to iso-thickness lines that we can use to determine the thickness of the film.

## Fixed Thickness

If we use now a film of a fixed thickness of  $d$  and send in a mixture of different wavelengths, we obtain constructive interference for the wavelengths

$$\lambda_{max} = \frac{4nd}{2m-1}$$

This means, thin films appear colored under some circumstances, which we still have to define. For this, we can now have a look at some examples.

### Example - d=100 nm and below

If we look at a film thickness of  $d = 100$  nm with  $n = 1.33$ , which corresponds to water we obtain constructive interference for

$$\lambda_{max} = \frac{4 \cdot 100 \text{ nm} \cdot 1.33}{2m - 1}$$

or

$$\lambda_{max} = \frac{532 \text{ nm}}{2m - 1}$$

which yields for different values of  $m$

- $m = 1$ : 532 nm
- $m = 2$ : 177 nm
- $m = 3$ : 106 nm

and so on. We see therefore that the longest wavelength to create constructive interference is 532 nm, which is green light. There is no longer wavelength causing constructive interference. The next longest wavelength is 177 nm, which is not visible anymore, so the reflection of a  $d = 100$  nm film would look green. The left plot in the figure below shows the intensity distribution over wavelength where you recognize that the maximum is very broad. If you therefore see green light being reflected from a thin soap film, you can be sure that the film is about 100 nm thick!

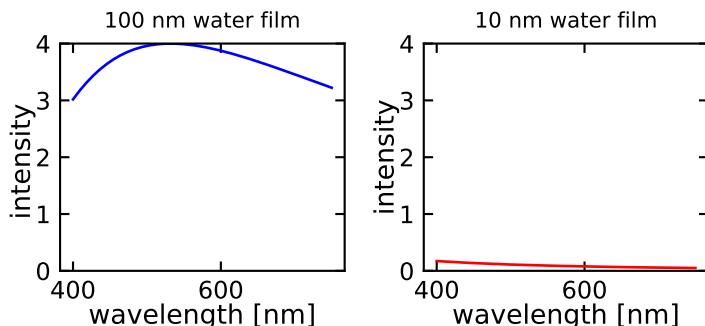


Figure 17.2: Intensity of reflection for a 100 nm and a 10 nm film of water.

An interesting effect is appearing, when the thickness of the water film gets very thin. We may ask, when is no constructive interference in the visible range observed. We therefore set the wavelength of the constructive interference to  $\lambda_{max} = 400$  nm and calculate the film thickness for which this occurs.

$$d = \frac{(2m - 1)\lambda_{max}}{2n} \approx 75 \text{ nm}$$

So for film thickness of water thinner than 75 nm, there is no constructive interference of the reflected light from the two boundaries in the visible region anymore. There will be still a reflection but no specific color. If the film gets even thinner, the intensity of the reflected light is further diminished by destructive interference and we obtain no reflection as shown on the right side of the above figure for a  $d = 10$  nm film. Such thin films, which do not show any reflection are called **Newton black films**. You might have seen them, if you look closer at soap bubbles. They will show regions, which look like holes, but of course there are no holes in soap bubbles.

**Example -  $d=1 \mu\text{m}$  and  $d=100 \mu\text{m}$** 

If the film gets thicker, e.g.  $d = 1 \mu\text{m}$  or even  $d = 100 \mu\text{m}$ , more than one constructive interference fits into the visible wavelength range. Due to that, the film may appear to have mixed colors or even look white. Below are the diagrams for those film thicknesses.

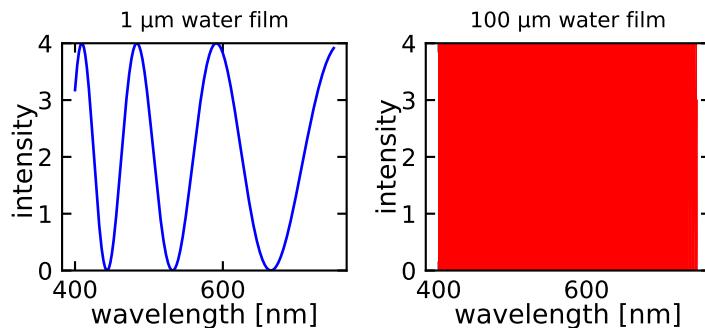


Figure 17.3: Reflection from a  $1\mu\text{m}$  (left) and a  $100 \mu\text{m}$  (right) thin water film. Experimental demonstration of the reflection of white light by a thin soap film.

**i Haidinger fringes**

This type of interference fringes observed on a plane-parallel film are called **Haidinger fringes**. Haidinger fringes are circular interference patterns observed when collimated light passes through a transparent plate at near-normal incidence. Named after Wilhelm von Haidinger who first described them in 1849, these fringes are localized at infinity (or in the focal plane of a lens) and arise from multiple reflections between the parallel surfaces of the plate. Unlike Newton's rings, which are localized near the surfaces that create them and require curved surfaces, Haidinger fringes occur with parallel surfaces and depend on the plate's thickness, its refractive index, the angle of incidence, and the wavelength of light.

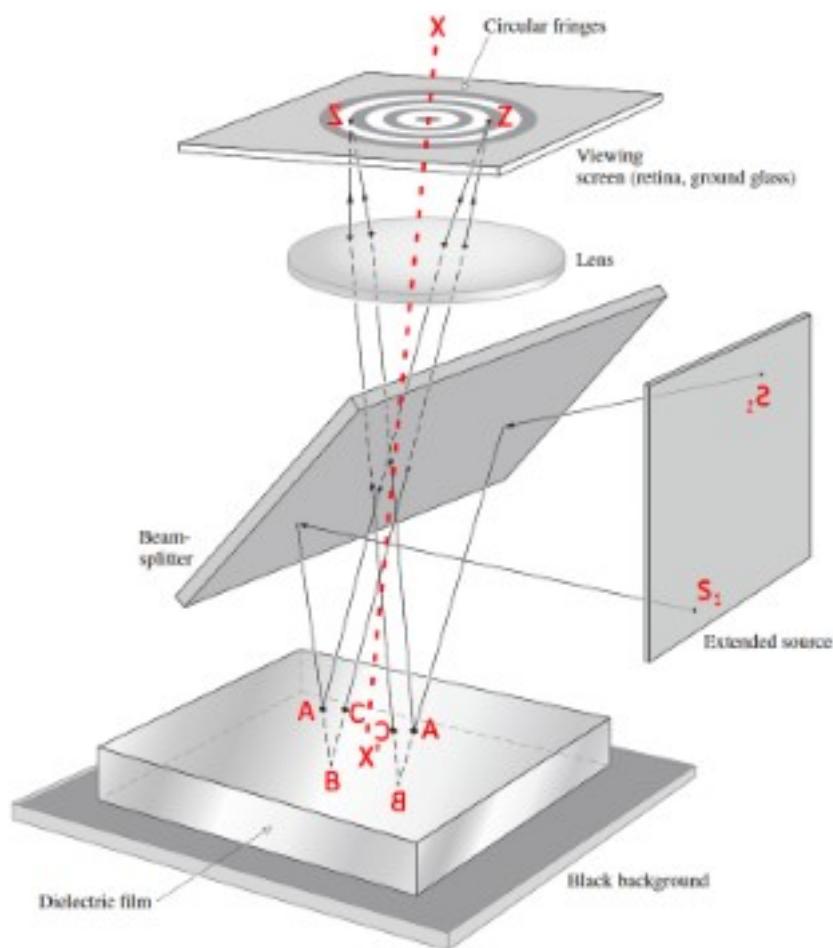


Figure 17.4: Haidinger Fringes (c) Eugene Hecht, Optics

The most beautiful example of thin film interference are of course soap bubbles or soap films. The colors of soap bubbles are due to the interference of light waves reflecting off the front and back surfaces of the thin soap film. The colors are due to the fact that the thickness of the film is not uniform. The film is thinnest at the top and thickest at the bottom. The colors are due to the interference of light waves reflecting off the front and back surfaces of the thin soap film.



Figure 17.5: Experimental demonstration of the reflection of white light by a thin soap film.

## 17.1 Newton Rings

A similar interference pattern is also observed in the case of a hemi-spherical surface touching a planar surface as sketched in the image below.

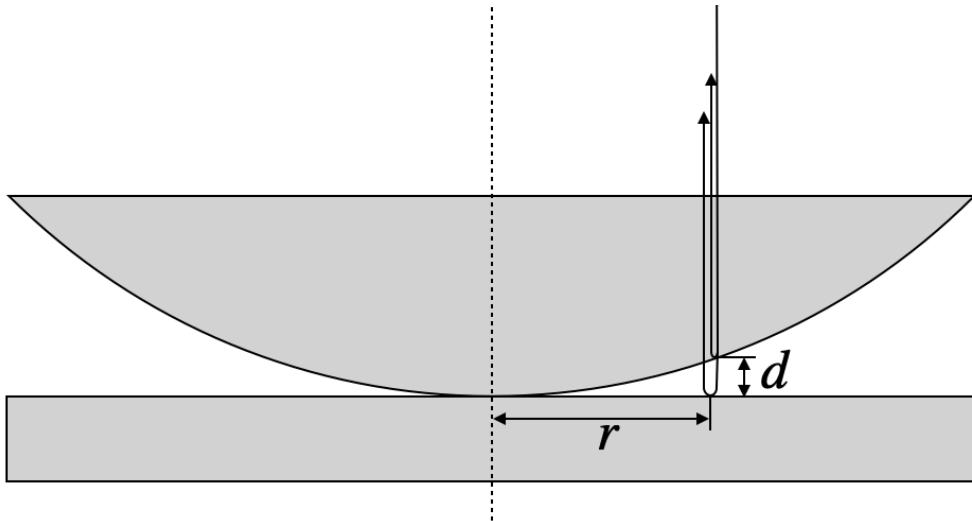


Figure 17.6: Newton Rings. Interference of waves from a spherical and a planar surface in close contact.

If light is incident normal to the top surface, reflections occur at several interfaces. The important reflections occur at the spherical surface and the planar surface below. The vertical distance between these surfaces is  $d$ , though refraction will deflect the beam slightly, making the actual path longer. If we stay close to the axis of the spherical surface ( $r \ll R$ ), where  $R$  is the radius of the spherical surface, we can neglect this refraction effect.

Under these conditions, the path length difference between a wave reflected at the curved and the planar surface is

$$\Delta s = 2d + \frac{\lambda}{2}$$

The additional term  $\lambda/2$  arises from the phase jump when reflecting at the planar boundary, as this reflection occurs at an optically denser material.

Having the path length difference, we can now calculate the condition for destructive interference:

$$\Delta s = \frac{2m+1}{2}\lambda = 2d + \frac{\lambda}{2}$$

where  $m$  is an integer. The distance  $d$  can be expressed as a function of the radial distance  $r$  from the contact point between the spherical surface and the plane surface. From the geometry of a circle, we have:

$$r^2 = d(2R - d)$$

with  $R$  being the radius of the spherical surface. Since  $d \ll R$ , the term  $d^2$  becomes negligible compared to  $2Rd$ , allowing us to simplify to:

$$r^2 = 2dR$$

from which we obtain:

$$d = \frac{r^2}{2R}$$

Inserting this distance into the interference condition yields the radius  $r_m$  where destructive interference is observed:

$$r_m = \sqrt{m\lambda R}$$

This equation shows that the radius of the interference rings increases with the square root of the integer  $m$ . Each wavelength creates its own ring pattern, with the radius depending on both the wavelength and the sphere's radius. This relationship makes Newton rings a useful tool for measuring either the wavelength of light (if  $R$  is known) or the radius of curvature of the spherical surface (if  $\lambda$  is known).

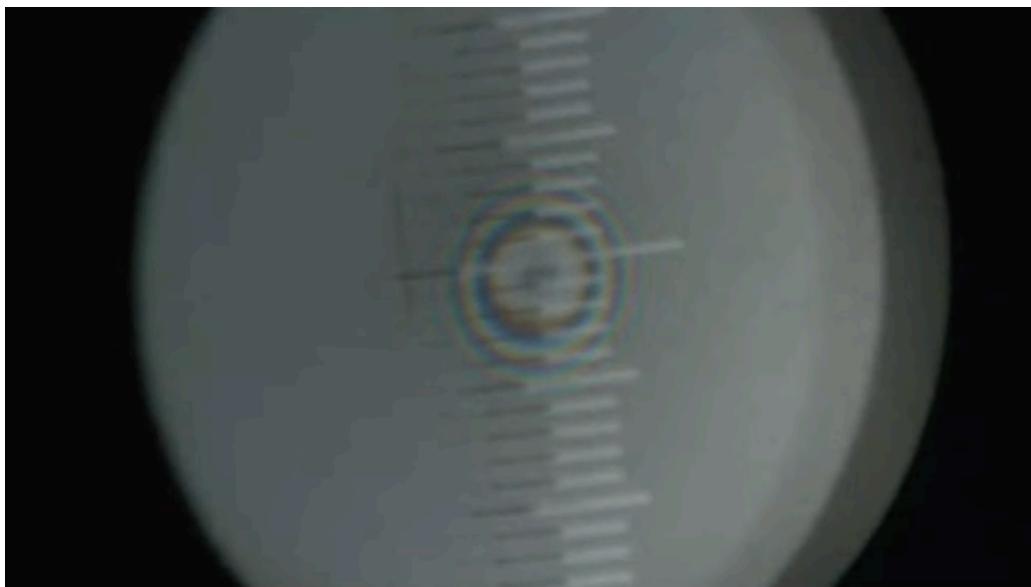


Figure 17.7: Observation of Newton Rings using white light in the lecture.

When using white light, as shown above, each wavelength creates its own set of rings, leading to the colored pattern observed. The spacing and size of these rings provide a precise method for optical measurements and quality control of optical surfaces.



# Chapter 18

## Multiple Wave Interference

So far we looked at the interference of two waves, which was a simplification as I mentioned already earlier. Commonly there will be a multitude of partial waves contribute to the observed interference. This is what we would like to have a look at now. We will do that in a quite general fashion, as the resulting formulas will appear several times again for different problems.

Nevertheless we will make a difference between

- multiwave interference of waves with the constant amplitude
- multiwave interference of waves with decreasing amplitude

Especially the latter is often occurring, if we have multiple reflections and each reflection is only a fraction of the incident amplitude.

### Multiple Wave Interference with Constant Amplitude

In the case of constant amplitude (for example realized by a grating, which we talk about later), the total wave amplitude is given according to the picture below by

$$U = U_1 + U_2 + U_3 + \dots + U_M$$

where we sum the amplitude over  $M$  partial waves. Between the neighboring waves (e.g.  $U_1$  and  $U_2$ ), we will assume a phase difference (because of a path length difference for example), which we denote as  $\Delta\phi$ .

The amplitude of the  $p$ -th wave is then given by

$$U_p = \sqrt{I_0} e^{i(p-1)\Delta\phi}$$

with the index  $p$  being an integer  $p = 1, 2, \dots, M$ ,  $h = e^{i\Delta\phi}$  and  $\sqrt{I_0}$  as the amplitude of each individual wave. The total amplitude  $U$  can be then expressed as

$$U = \sqrt{I_0} (1 + h + h^2 + \dots + h^{M-1})$$

which is a geometric sum. We can apply the sum formula for geometric sums to obtain

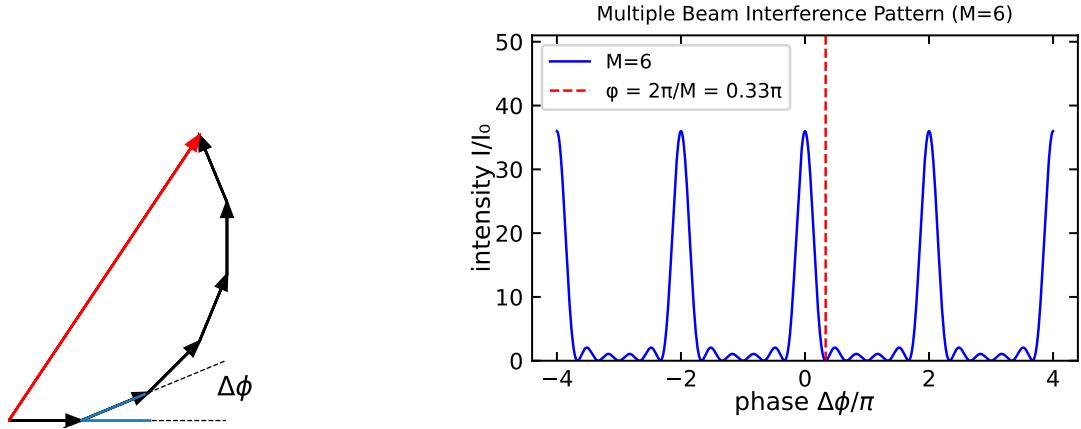
$$U = \sqrt{I_0} \frac{1 - h^M}{1 - h} = \sqrt{I_0} \frac{1 - e^{iM\Delta\phi}}{1 - e^{i\Delta\phi}}$$

We now have to calculate the intensity of the total amplitude

$$I = |U|^2 = I_0 \left| \frac{e^{-iM\Delta\phi/2} - e^{iM\Delta\phi/2}}{e^{-i\Delta\phi/2} - e^{i\Delta\phi/2}} \right|^2$$

which we can further simplify to give

$$I = I_0 \frac{\sin^2(M\Delta\phi/2)}{\sin^2(\Delta\phi/2)}$$



(a) Multiple wave interference of  $M = 6$  waves with a phase difference of  $\phi = \pi/8$ . The black arrows represent the individual waves, the red arrow the sum of all waves.

Figure 18.2: Multiple beam interference pattern for  $M=6$  beams. The intensity distribution is shown as a function of the phase shift  $\phi$ . The first minimum is at  $\phi = 2\pi/M$ . The intensity distribution is symmetric around  $\phi = 0$ .

The result is therefore an oscillating function. The numerator  $\sin^2(M\Delta\phi/2)$  shows an oscillation frequency, which is by a factor of  $M$  higher than the one in the denominator  $\sin^2(\Delta\phi/2)$ . Therefore the intensity pattern is oscillating rapidly and creating a first minimum at

$$\Delta\phi = \frac{2\pi}{M}$$

This is an important result, since it shows that the number of sources  $M$  determines the position of the first minimum and the interference peak gets narrower with increasing  $M$ . Since the phase difference  $\Delta\phi$  between neighboring sources is the same as for the double slit experiment, i.e.  $\Delta\phi = 2\pi d/\lambda \sin(\theta)$ , we can also determine the angular position of the first minimum. This is given by

$$\sin(\theta_{\min}) = \frac{1}{M} \frac{\lambda}{d}$$

This again has the common feature that it scales as  $\lambda/d$ . A special situation occurs, whenever the numerator and the denominator become zero. This will happen whenever

$$\Delta\phi = m2\pi$$

where  $m$  is an integer and denotes the interference order, i.e. the number of wavelength that neighboring partial waves have as path length difference. In this case, the intensity distribution will give us

$$I = I_0 \frac{0}{0}$$

and we have to determine the limit with the help of l'Hospitals rule. The outcome of this calculation is, that

$$I(\Delta\phi = m2\Delta\pi) = M^2 I_0$$

which can be also realized when using the small angle approximation for the sine functions.

### Wavevector Representation

We would like to introduce a different representation of the multiple wave interference of the grating, which is quite insightful. The first order ( $m = 1$ ) constructive interference condition is given by

$$\frac{1}{\lambda} \sin \theta = \frac{1}{d}$$

which also means that

$$\frac{2\pi}{\lambda} \sin \theta = \frac{2\pi}{d}$$

This can be written as

$$k \sin \theta = K$$

where  $k$  is the magnitude of the wavevector of the light and  $K$  is the wavevector magnitude that corresponds to the grating period  $d$ . As the magnitude of the wavevector of the light is conserved, the wavevectors of the incident light and the light traveling along the direction of the first interence peak form the sides of an equilateral triangle. This is shown in the following figure.

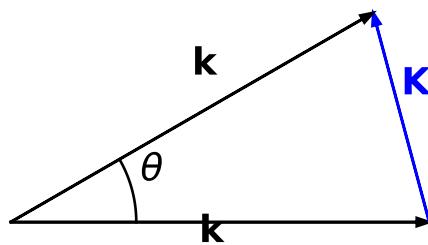


Figure 18.3: Wavevector summation for the diffraction grating. The wavevector of the incident light  $k$  and the wavevector of the light traveling along the direction of the first interence peak  $K$  form an equilateral triangle.

This means that the diffraction grating is providing a wavevector  $K$  to alter the direction of the incident light. This is again a common feature reappearing in many situations as for example in the X-ray diffraction of crystals.

## Multiple Wave Interference with Decreasing Amplitude

We will turn our attention now to a slight modification of the previous multiwave interference. We will introduce a decreasing amplitude of the individual waves. The first wave shall have an amplitude  $U_1 = \sqrt{I_0}$ . The next wave, however, will not only be phase shifted but also have a smaller amplitude.

$$U_2 = hU_1$$

where  $h = re^{i\phi}$  with  $|h| = r < 1$ .  $r$  can be regarded as a reflection coefficient, which diminishes the amplitude of the incident wave. According to that the intensity is reduced by

$$I_2 = |U_2|^2 = |hU_1|^2 = r^2 I_1$$

The intensity of the incident wave is multiplied by a factor  $r^2$ , while the amplitude is multiplied by  $r$ . Note that the phase factor  $e^{i\Delta\phi}$  is removed when taking the square of this complex number.

### i Intensity at Boundaries

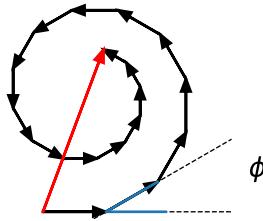
The amplitude of the reflected wave is diminished by a factor  $r \leq 1$ , which is called the reflection coefficient. The intensity is diminished by a factor  $R = |r|^2 \leq 1$ , which is the **reflectance**.

In the absence of absorption, reflectance  $R$  and **transmittance**  $T$  add to one due to energy conservation.

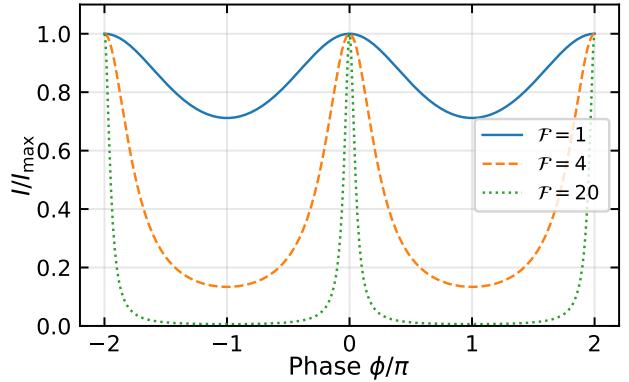
$$R + T = 1$$

Consequently, the third wave would be now  $U_3 = hU_2 = h^2U_1$ . The total amplitude is thus

$$U = U_1 + U_2 + U_3 + \dots + U_M = \sqrt{I_0}(1 + h + h^2 + \dots)$$



(a) Phase construction of a multiwave interference with  $M$  waves with decreasing amplitude due to a reflection coefficient  $r = 0.95$ .



(a) Multiple wave interference with decreasing amplitude. The graph shows the intensity distribution over the phase angle  $\phi$  for different values of the Finesse  $\mathcal{F}$ .

This yields again

$$U = \sqrt{I_0} \frac{(1 - h^M)}{1 - h} = \frac{\sqrt{I_0}}{1 - re^{i\Delta\phi}}$$

Calculating the intensity of the waves is giving

$$I = |U|^2 = \frac{I_0}{|1 - re^{i\Delta\phi}|^2} = \frac{I_0}{(1 - r)^2 + 4r \sin^2(\Delta\phi/2)}$$

which is also known as the **Airy function**. This function can be further simplified by the following abbreviations

$$I_{\max} = \frac{I_0}{(1-r)^2}$$

and

$$\mathcal{F} = \frac{\pi\sqrt{r}}{1-r}$$

where the latter is called the *Finesse*. With those abbreviations, we obtain

$$I = \frac{I_{\max}}{1 + 4 \left(\frac{\mathcal{F}}{\pi}\right)^2 \sin^2(\Delta\phi/2)}$$

for the interference of multiple waves with decreasing amplitude.

This intensity distribution has a different shape than the one we obtained for multiple waves with the same amplitude.

We clearly observe that with increasing Finesse the intensity maxima, which occur at multiples of  $\pi$  get much narrower. In addition the regions between the maxima show better contrast and for higher Finesse we get complete destructive interference.



# Chapter 19

## Fabry Perot Interferometer

The Fabry-Perot interferometer demonstrates multiple-wave interference with decreasing amplitude. It consists of two parallel mirrors separated by a distance  $d$ , creating multiple reflections of incident light.

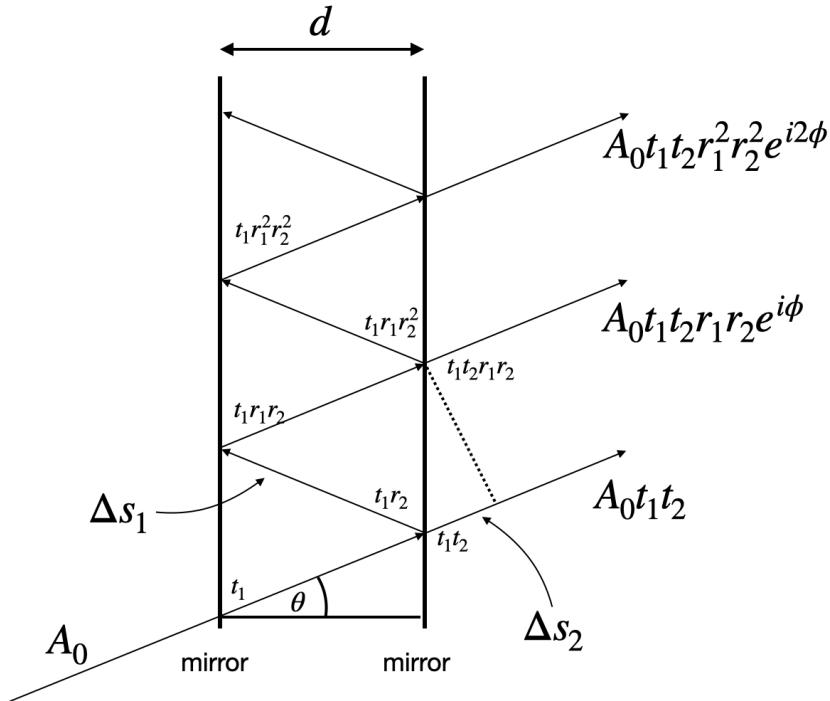


Figure 19.1: Simplified Sketch of a Fabry-Perot Interferometer.

When light with amplitude  $A_0$  enters the interferometer, it undergoes a series of transmissions and reflections. The first transmitted wave has amplitude:

$$U_1 = A_0 t_1 t_2$$

where  $t_1$  and  $t_2$  are the transmission coefficients of the first and second mirrors. The second transmitted wave includes reflections from both mirrors ( $r_1$  and  $r_2$ ) and a phase shift  $\Delta\phi$ :

$$U_2 = A_0 t_1 t_2 r_1 r_2 e^{i\phi} = U_1 r_1 r_2 e^{i\Delta\phi}$$

This follows our earlier treatment of multiple-wave interference with decreasing amplitude, where  $\sqrt{I_0} = A_0 t_1 t_2$  and  $r = r_1 r_2$ . The phase shift between successive reflections is:

$$\phi = \frac{2\pi}{\lambda} \Delta s = \frac{2\pi}{\lambda} 2d \cos(\theta)$$

where  $\Delta s = 2d \cos(\theta)$  represents the path difference between adjacent rays.

The resulting intensity distribution is:

$$I = |U|^2 = \frac{I_0}{|1 - r e^{i\phi}|^2} = \frac{I_0}{(1 - r)^2 + 4r \sin^2\left(\frac{2\pi}{\lambda} d \cos(\theta)\right)}$$

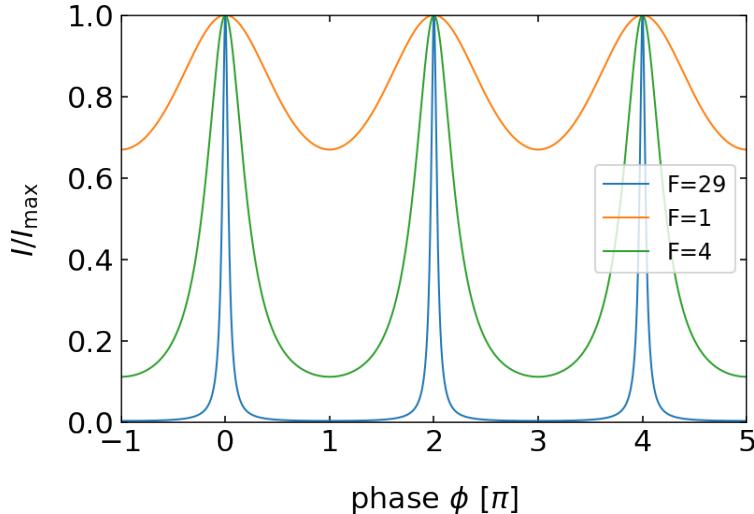


Figure 19.2: Fabry Perot Interferometer.

## Finesse and Spectral Properties

The quality of interference in a Fabry-Perot interferometer is characterized by the Finesse  $\mathcal{F}$ :

$$\mathcal{F} = \frac{\pi\sqrt{r}}{1 - r}$$

where  $r = r_1 r_2$  is the product of the mirrors' reflection coefficients. As  $r$  approaches 1 (higher reflectivity), the Finesse increases, resulting in sharper interference peaks.

For normal incidence ( $\theta = 0$ ), the phase difference simplifies to:

$$\Delta\phi = \frac{4\pi d}{\lambda}$$

Constructive interference occurs when  $\Delta\phi = m2\pi$  (where  $m$  is an integer), giving us the wavelengths of transmission maxima:

$$\lambda_m = \frac{2d}{m}$$

## Free Spectral Range

The spacing between adjacent transmission peaks in an optical system can be expressed in terms of either wavelength or frequency. This spacing is known as the **free spectral range (FSR)**.

### In Wavelength:

The difference in wavelength between adjacent transmission peaks is given by:

$$\delta\lambda = \lambda_m - \lambda_{m+1} = \frac{\lambda_m}{m+1}$$

Here,  $\lambda_m$  is the wavelength corresponding to the  $m$ -th transmission peak.

### In Frequency:

The difference in frequency between adjacent transmission peaks is given by:

$$\delta\nu = \nu_{m+1} - \nu_m = \frac{c}{2d}$$

Here,  $c$  is the speed of light, and  $d$  is the distance between the reflecting surfaces in the optical system.

The free spectral range (FSR) represents the interval between successive transmission peaks and is a crucial parameter in the design and analysis of optical systems, such as Fabry-Pérot interferometers and optical resonators.

## Spectral Resolution

The spectral resolution of an interferometer is determined by the width of its interference peaks, which indicates the instrument's ability to distinguish between closely spaced wavelengths. This width is often characterized by the full width at half maximum (FWHM) of the peaks.

To find the FWHM, we start with the intensity ratio at half maximum:

$$\frac{I}{I_{\max}} = \frac{1}{2} = \frac{1}{1 + \left(\frac{\mathcal{F}}{\pi}\right)^2 \Delta\phi_{1/2}^2}$$

Solving for the phase difference  $\Delta\phi_{1/2}$  at half maximum, we can determine the corresponding frequency width:

$$\Delta\nu = \frac{c}{2d\mathcal{F}} = \frac{\delta\nu}{\mathcal{F}}$$

Here,  $\delta\nu$  is the free spectral range,  $c$  is the speed of light,  $d$  is the distance between the reflecting surfaces, and  $\mathcal{F}$  is the finesse of the interferometer.

The finesse  $\mathcal{F}$  is defined as the ratio of the free spectral range to the FWHM:

$$\mathcal{F} = \frac{\delta\nu}{\Delta\nu} = \frac{\lambda}{\Delta\lambda}$$

This ratio provides a measure of the interferometer's spectral resolution, indicating how well it can separate two closely spaced spectral lines.

The overall resolving power  $\mathcal{R}$  of the interferometer is given by:

$$\mathcal{R} = m\mathcal{F}$$

where  $m$  is the order of the interference. The resolving power  $\mathcal{R}$  quantifies the ability of the interferometer to resolve spectral features, with higher values indicating better resolution.

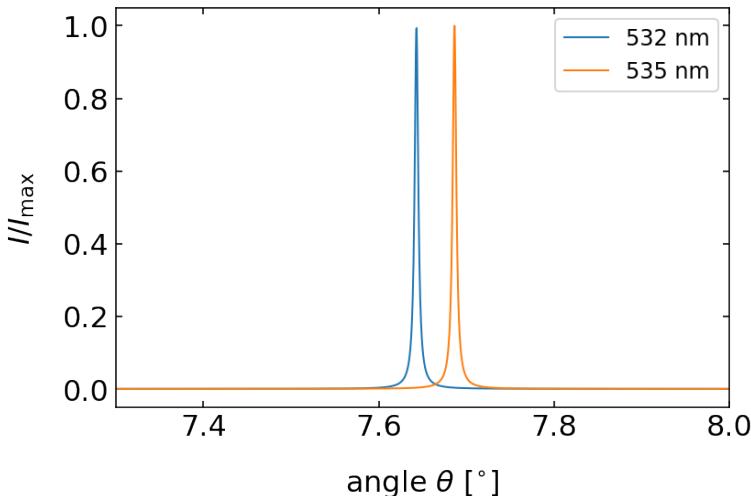


Figure 19.3: Two different wavelengths interfering constructively in a Fabry Perot interferometer.

### **i** Free Spectral Range and Spectral Resolution

#### **Free Spectral Range (FSR):**

The free spectral range is the spacing between adjacent transmission maxima in a Fabry-Perot interferometer. It can be expressed in terms of wavelength or frequency:

- In wavelength:

$$\delta\lambda = \lambda_m - \lambda_{m+1} = \frac{\lambda_m}{m+1}$$

- In frequency:

$$\delta\nu = \nu_{m+1} - \nu_m = \frac{c}{2d}$$

The FSR indicates the range over which the interferometer can distinguish between different wavelengths or frequencies before the next order of interference occurs.

#### **Spectral Resolution:**

The spectral resolution of a Fabry-Perot interferometer is determined by the width of the interference peaks. It is often quantified by the Finesse ( $\mathcal{F}$ ), which is the ratio of the free spectral range to the full width at half maximum (FWHM) of the peaks:

$$\mathcal{F} = \frac{\delta\nu}{\Delta\nu} = \frac{\lambda}{\Delta\lambda}$$

The resolving power ( $\mathcal{R}$ ) of the interferometer is given by:

$$\mathcal{R} = m\mathcal{F}$$

where  $m$  is the interference order. The resolving power indicates the ability of the interferometer to distinguish between closely spaced spectral lines.

## Ring Pattern Formation

When a Fabry-Perot interferometer is used with an extended monochromatic light source and appropriate optics, it produces a characteristic ring pattern:

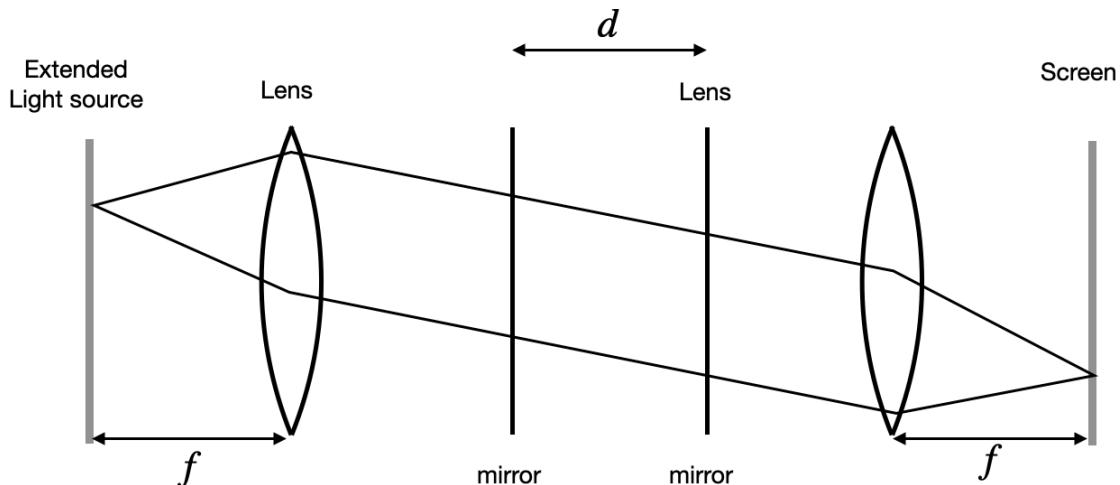


Figure 19.4: Fabry Perot Interferometer and interference pattern observed in the lecture.

The rings represent contours of constant phase difference, becoming more closely spaced with increasing radius as demonstrated in these experimental observations:



Figure 19.5: Fabry Perot Interferometer and interference pattern observed in the lecture.

This ring pattern is a powerful tool for spectroscopic analysis, as different wavelengths produce distinct ring patterns, allowing precise wavelength measurements and spectral analysis.

### Applications in Modern Research and Technology

#### Spectroscopy:

- **High-Resolution Spectroscopy:** Fabry-Perot interferometers are used to achieve high spectral resolution, allowing precise measurements of spectral lines. This is crucial in fields like astrophysics, where detailed analysis of stellar spectra can reveal information about the composition, temperature, and motion of celestial objects.

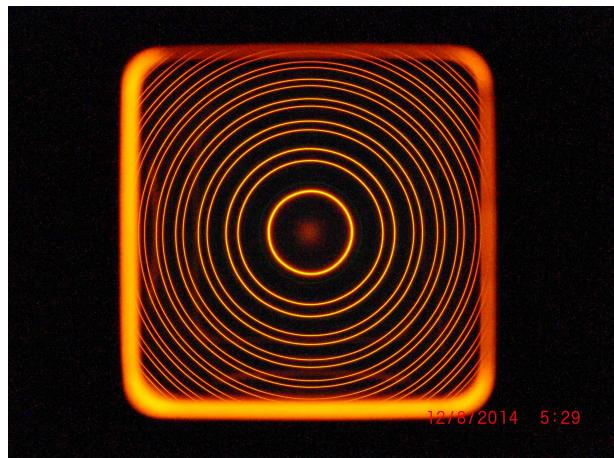


Figure 19.6: Sodium D Lines measured as  $593.0 \pm 4.5$  nm,  $591.1 \pm 5.0$  nm with a difference of  $0.604 \pm 0.025$  (matching the accepted values of 589.5924 nm, 588.9951 with a difference of 0.5973 nm). Also courtesy of Lake Forest College

- **Laser Spectroscopy:** They are used to analyze the spectral properties of lasers, including linewidth, mode structure, and stability.

#### Optical Communications:

- **Wavelength Division Multiplexing (WDM):** Fabry-Perot filters are used in WDM systems to separate and combine different wavelength channels, increasing the data-carrying capacity of optical fibers.
- **Laser Stabilization:** They help stabilize the wavelength of lasers used in optical communication systems, ensuring consistent performance and reducing signal degradation.

#### Metrology:

- **Precision Measurement:** Fabry-Perot interferometers are used for precise distance and displacement measurements. They can measure changes in length with sub-nanometer accuracy, making them valuable in applications like semiconductor manufacturing and materials science.
- **Refractive Index Measurement:** They are used to measure the refractive index of gases, liquids, and solids with high precision.

#### Laser Technology:

- **Mode-Locking:** Fabry-Perot cavities are used in mode-locked lasers to produce ultra-short pulses of light, which are essential for applications in time-resolved spectroscopy, medical imaging, and telecommunications.
- **Laser Tuning:** They are used to tune the wavelength of lasers, enabling precise control over the output wavelength for various applications.

#### Environmental Monitoring:

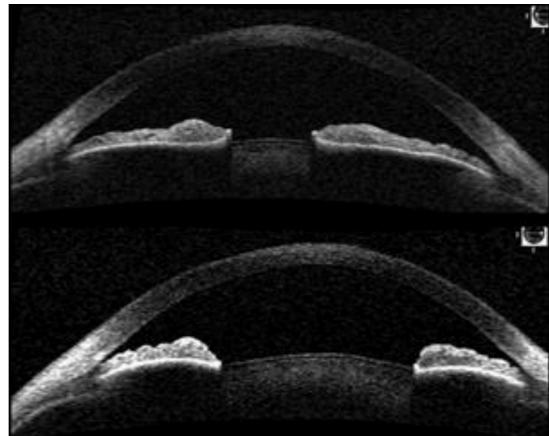
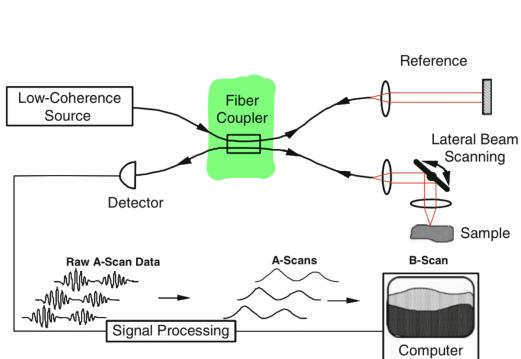
- **Gas Analysis:** Fabry-Perot interferometers are used in gas analyzers to detect and quantify trace gases in the atmosphere. This is important for monitoring air quality, greenhouse gas emissions, and industrial processes.
- **Remote Sensing:** They are used in remote sensing instruments to analyze the spectral properties of reflected or emitted light from the Earth's surface and atmosphere, providing valuable data for climate studies and environmental monitoring.

#### Astronomy:

- **Interferometric Imaging:** Fabry-Perot interferometers are used in telescopes to enhance the resolution of astronomical images. They can be used to study fine details of celestial objects, such as the structure of galaxies and the dynamics of star-forming regions.
- **Doppler Spectroscopy:** They are used to measure the Doppler shift of spectral lines, allowing astronomers to determine the radial velocity of stars and planets, which is crucial for the detection of exoplanets.

#### Biomedical Applications:

- **Optical Coherence Tomography (OCT):** Fabry-Perot interferometers are used in OCT systems to achieve high-resolution cross-sectional imaging of biological tissues. This is valuable for medical diagnostics, particularly in ophthalmology and dermatology.



- **Fluorescence Microscopy:** They are used to enhance the spectral resolution of fluorescence microscopes, enabling detailed analysis of biological samples.

#### Quantum Optics:

- **Cavity Quantum Electrodynamics (CQED):** Fabry-Perot cavities are used to study the interaction between light and matter at the quantum level. This research is fundamental for the development of quantum information technologies and quantum computing.
- **Single-Photon Sources:** They are used to create and manipulate single-photon sources, which are essential for quantum communication and cryptography.

These applications highlight the versatility and importance of Fabry-Perot interferometers in advancing scientific research and technological innovation across various fields.



# Chapter 20

## Diffraction

### 20.1 Huygens' Principle

Formulated by Christiaan Huygens in 1678, **Huygens' principle** states that every point on a wavefront acts as a source of secondary spherical wavelets that spread out in the forward direction. The new position of the wavefront at any later time is found by constructing a surface tangent to these secondary wavelets. This principle provides a powerful method for analyzing wave propagation and explains various wave phenomena such as reflection, refraction, and diffraction.

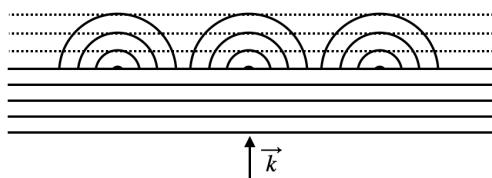


Figure 20.1: Illustration of Huygens' principle for a plane wave incident with a wave vector  $\vec{k}$ . Each point on the wavefront acts as a source of secondary wavelets, and the new wavefront is the envelope of these wavelets.

Huygens' principle can be demonstrated numerically and visually. By placing a large number of spherical wave sources closely along a line and allowing their waves to interfere, we can reconstruct a plane wavefront propagating in the forward direction, illustrating how a plane wave advances according to Huygens' concept.

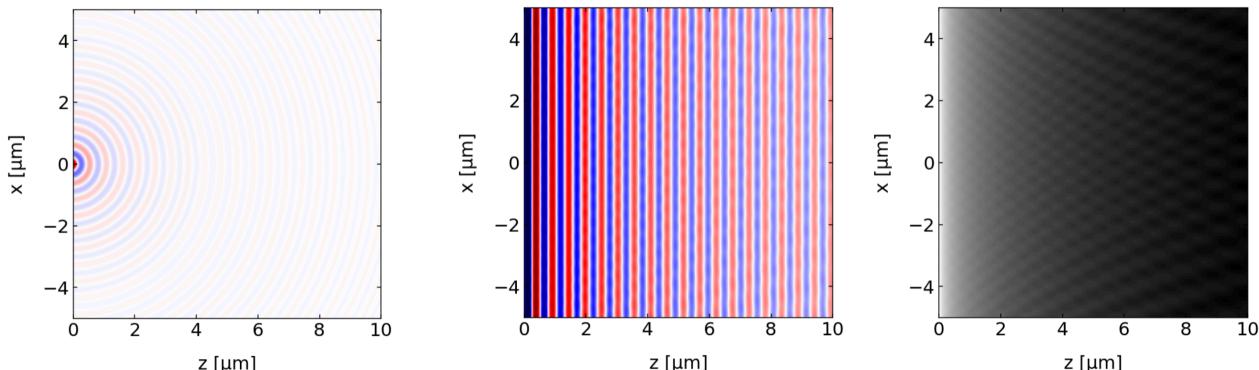


Figure 20.2: Numerical demonstration of Huygens' principle used to recreate a plane wave from a set of spherical waves. The graph on the left shows the amplitude of a single spherical wave of wavelength  $\lambda = 532$  nm. By arranging 500 spherical wave sources densely along the x-axis at  $z = 0$ , all in phase (representing the phase of the incident plane wave at  $z = 0$ ), we can recreate the plane wavefronts for  $z > 0$  (middle) and the constant intensity distribution (right).

Mathematically, this phenomenon can be described using our earlier treatment of multi-wave interference. Consider  $M$  spherical wave sources arranged along the x-axis at  $z = 0$ , each separated by a small distance  $d$  from its neighbor. At a point far away from the sources (in the far-field approximation), the path difference between waves from adjacent sources leads to a phase difference given by:

$$\Delta\phi = \frac{2\pi d \sin \theta}{\lambda}$$

where  $\theta$  is the angle relative to the z-axis (the forward direction), and  $\lambda$  is the wavelength of the waves. The superposition of these waves results in an intensity pattern described by:

$$I(\theta) = I_0 \frac{\sin^2(M \frac{\pi d \sin \theta}{\lambda})}{\sin^2(\frac{\pi d \sin \theta}{\lambda})}$$

This expression arises from the interference of  $M$  waves with a constant phase difference  $\Delta\phi$  between neighboring waves. The numerator represents the constructive and destructive interference due to the finite number of sources, and the denominator accounts for the spacing between them.

This mathematical framework serves as the foundation for understanding diffraction phenomena, particularly in two important cases:

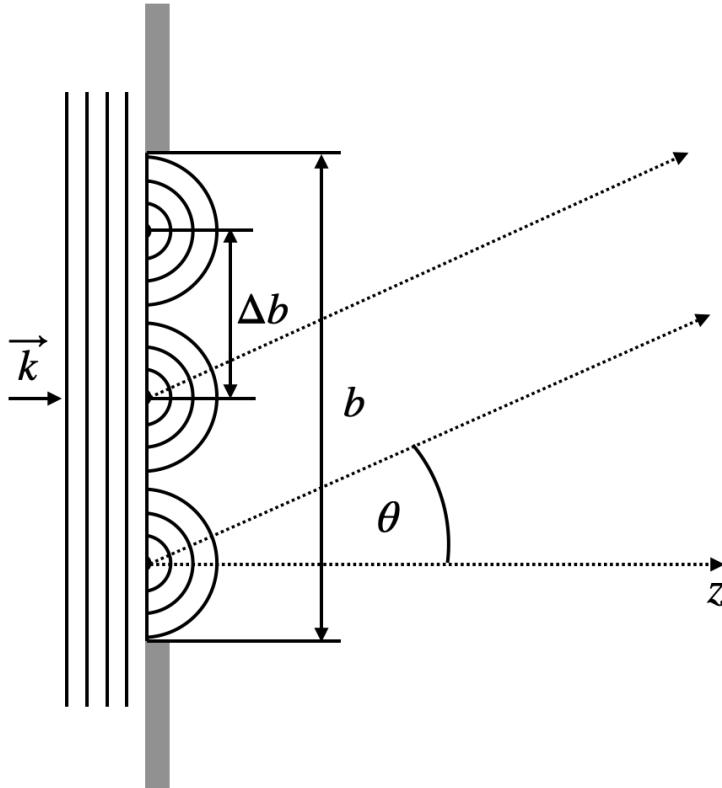
1. **Single-Slit Diffraction:** When Huygens' wavelets are confined to a finite width (the slit width), the interference between these wavelets produces a characteristic diffraction pattern with a central maximum and diminishing side lobes.
2. **Diffraction Gratings:** When multiple slits are arranged periodically, the interference of the transmitted waves leads to sharp diffraction maxima at specific angles, making diffraction gratings powerful tools for spectroscopic analysis.

While we commonly use the term “diffraction” to describe these phenomena, they are fundamentally due to the interference of waves, as explained by Huygens' principle. By considering every point on a wavefront as a source of secondary wavelets, we can understand and predict the complex patterns that arise when waves encounter obstacles or apertures.

## 20.2 Single Slit Diffraction

We now apply our interference formula to study the diffraction of an incident plane wave (wavevector  $\vec{k}$ ) on a single slit of width  $b$ . We can model this by placing a series of Huygens sources along the slit opening. While

the sketch below shows just 3 sources for clarity, we'll generalize this to  $M$  sources.



We divide the slit into segments of width  $\Delta b$  such that we have  $M = b/\Delta b$  Huygens sources, each with amplitude  $A_0 = \sqrt{I_0}$ . Applying our previous multi-wave interference formula with spacing  $d = \Delta b$ , we obtain:

$$I = I_0 \frac{\sin^2(M\pi \frac{\Delta b}{\lambda} \sin(\theta))}{\sin^2(\pi \frac{\Delta b}{\lambda} \sin(\theta))}$$

Since  $M\Delta b = b$ , we can rewrite this as:

$$I = I_0 \frac{\sin^2(\pi \frac{b}{\lambda} \sin(\theta))}{\sin^2(\pi \frac{b}{M\lambda} \sin(\theta))}$$

For convenience, let's substitute  $x = \pi \frac{b}{\lambda} \sin(\theta)$ , giving:

$$I = I_0 \frac{\sin^2(x)}{\sin^2(x/M)}$$

In reality, we have a continuous distribution of sources across the slit width, corresponding to  $M \rightarrow \infty$ . In this limit, for the denominator,  $x/M$  becomes very small, and we can use the small-angle approximation:

$$\sin^2\left(\frac{x}{M}\right) \approx \left(\frac{x}{M}\right)^2$$

Therefore, our final expression becomes:

$$I(\theta) = I_s \frac{\sin^2 (\pi \frac{b}{\lambda} \sin(\theta))}{(\pi \frac{b}{\lambda} \sin(\theta))^2}$$

where  $I_s = M^2 I_0$  represents the total intensity from all sources. This expression is often written using the sinc function (sinus cardinalis):

$$I(\theta) = I_s \operatorname{sinc}^2 \left( \pi \frac{b}{\lambda} \sin(\theta) \right)$$

This formula describes the characteristic diffraction pattern of a single slit, with a central maximum and symmetric side lobes of decreasing intensity.

### **i Single Slit Diffraction**

The intensity distribution generated by the diffraction of monochromatic light on a single slit and observed in the far field is given by

$$I(\theta) = I_s \frac{\sin^2 (\pi \frac{b}{\lambda} \sin(\theta))}{(\pi \frac{b}{\lambda} \sin(\theta))^2}$$

where  $\lambda$  is the wavelength of the light and  $b$  the width of the slit. The angle of observation is given by  $\theta$ . Note that the diffraction pattern on any aperture is resulting from the fact that you remove Huygens sources that would be normally needed to form a plane wavefront for example.

#### **Fourier Transform and Diffraction**

The intensity distribution generated by the diffraction of monochromatic light on a single slit and observed in the far field is given by

$$I(\theta) = I_s \frac{\sin^2 (\pi \frac{b}{\lambda} \sin(\theta))}{(\pi \frac{b}{\lambda} \sin(\theta))^2}$$

where  $\lambda$  is the wavelength of the light and  $b$  the width of the slit. The angle of observation is given by  $\theta$ . Note that the diffraction pattern on any aperture is resulting from the fact that you remove Huygens sources that would be normally needed to form a plane wavefront for example.

This diffraction pattern can be understood as the Fourier transform of the aperture function. In the case of a single slit, the aperture function is a rectangular function, and its Fourier transform is a sinc function. This relationship between the aperture and its diffraction pattern is a fundamental concept in wave optics and is widely used in various applications, including imaging and signal processing.

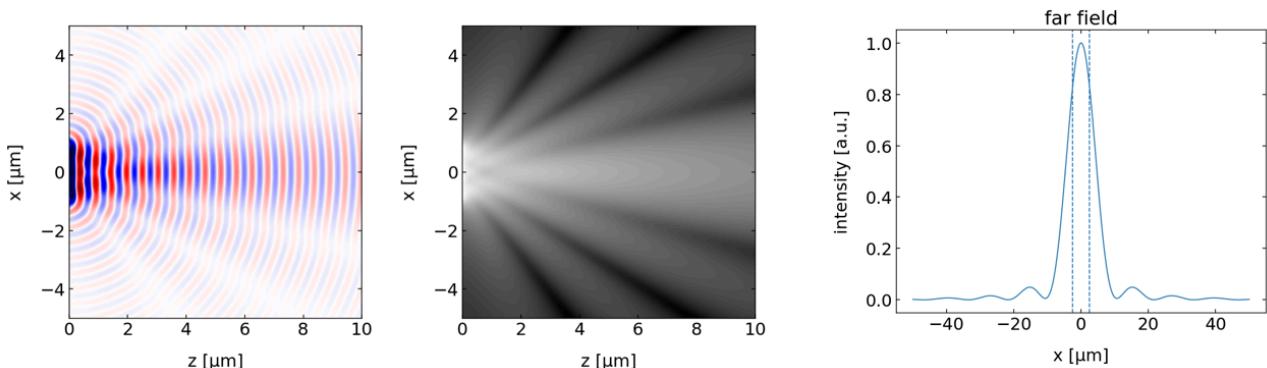


Figure 20.3: Total wave amplitude behind a slit ( $b=2\mu\text{m}$ ) for an incident wave of 532 nm wavelength. The plot in the middle shows the intensity in the space behind the slit. The graph on the right displays the diffraction pattern at a screen at 100  $\mu\text{m}$  distance from the slit.

Let's have a look at some of the properties of the intensity distribution.

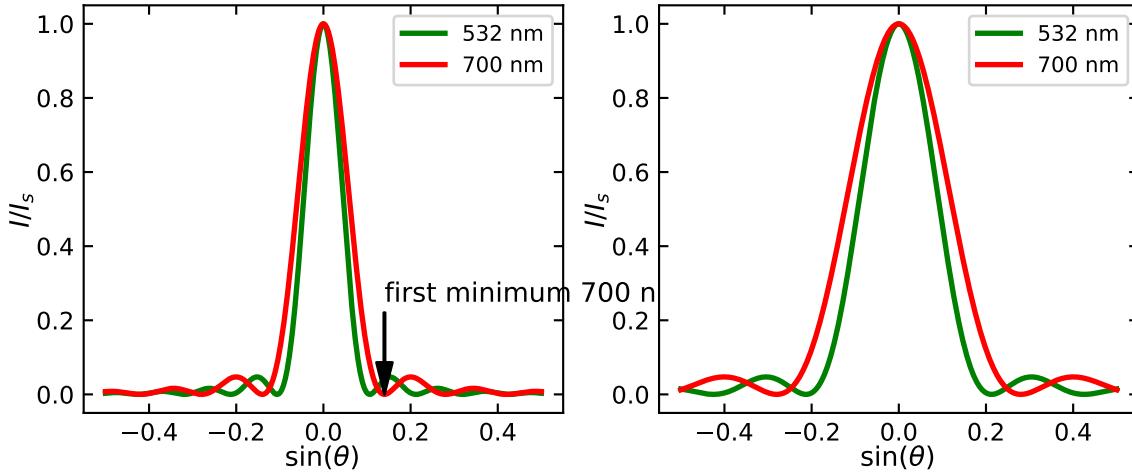


Figure 20.4: Caption

The single-slit diffraction pattern shows characteristic features that we can observe in both theoretical calculations and experimental measurements. The intensity distribution is described by an oscillating function with decreasing amplitude. The oscillations arise from the  $\sin^2$  term in the numerator, while the decay comes from the square term in the denominator.

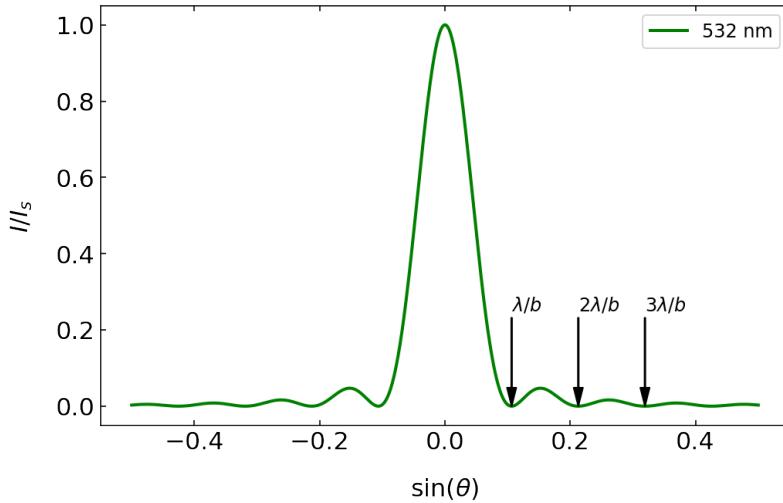


Figure 20.5: Diffraction patterns as a function of the sine of the diffraction angle. The minima of the diffraction pattern in this plot are at integer multiples of  $\lambda/b$ .

The graphs above illustrate two key relationships in single-slit diffraction:

1. The effect of wavelength: When comparing patterns for different wavelengths (with fixed slit width  $b = 5$  m), longer wavelengths produce broader diffraction patterns
2. The effect of slit width: For the same wavelength, reducing the slit width to  $b = 2.5$  m results in a broader diffraction pattern

These observations can be quantified by analyzing the positions of intensity minima. The intensity goes to zero when the argument of the sine function in the numerator equals multiples of :

$$\pi \frac{b}{\lambda} \sin(\theta) = m\pi$$

where  $m$  is an integer. This simplifies to:

$$\sin(\theta) = m \frac{\lambda}{b}$$

This relationship reveals a fundamental principle in diffraction: the angular spread of the pattern is proportional to the ratio of wavelength to the size of the diffracting object ( $\lambda/b$ ). While the exact mathematical form may vary for different geometries, this basic scaling remains valid.

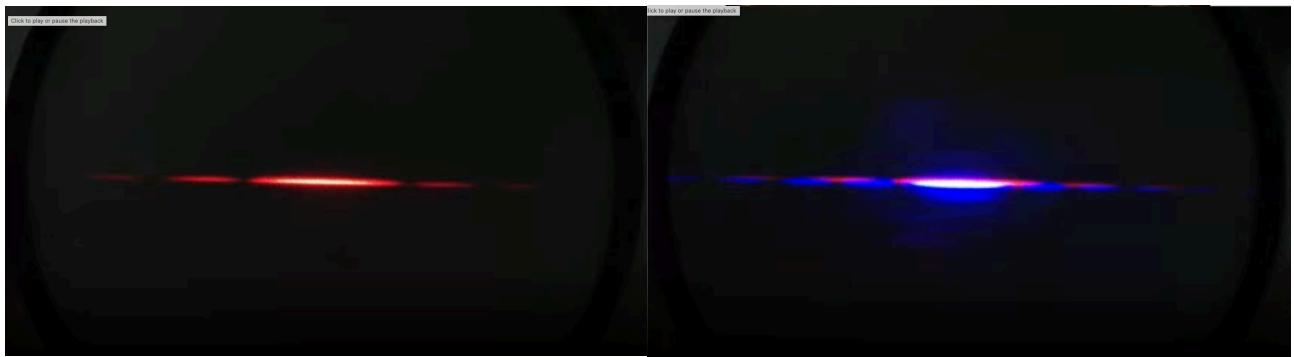


Figure 20.6: Diffraction patterns on a single slit as observed in the lecture. The left image shows the diffraction pattern for red light, while the right image combines two different wavelengths (red, blue), where one clearly recognizes the wider diffraction peaks for the longer red wavelength.

The experimental observations above clearly demonstrate these principles, particularly showing how red light (longer wavelength) produces a broader diffraction pattern than blue light (shorter wavelength).

### 20.3 Circular Aperture

For a circular aperture, the diffraction pattern follows a more complex mathematical form involving Bessel functions. The intensity distribution is given by:

$$I(\theta) = I_0 \left( \frac{2J_1(x)}{x} \right)^2$$

where  $J_1$  is the [Bessel function](#) of the first kind, and

$$x = \frac{2\pi R}{\lambda} \sin(\theta)$$

Here,  $R$  is the radius of the aperture. While similar to the sine function, the Bessel function has zeros at different positions:  $x_1 = 1.22\pi$ ,  $x_2 = 2.23\pi$ , and so on.

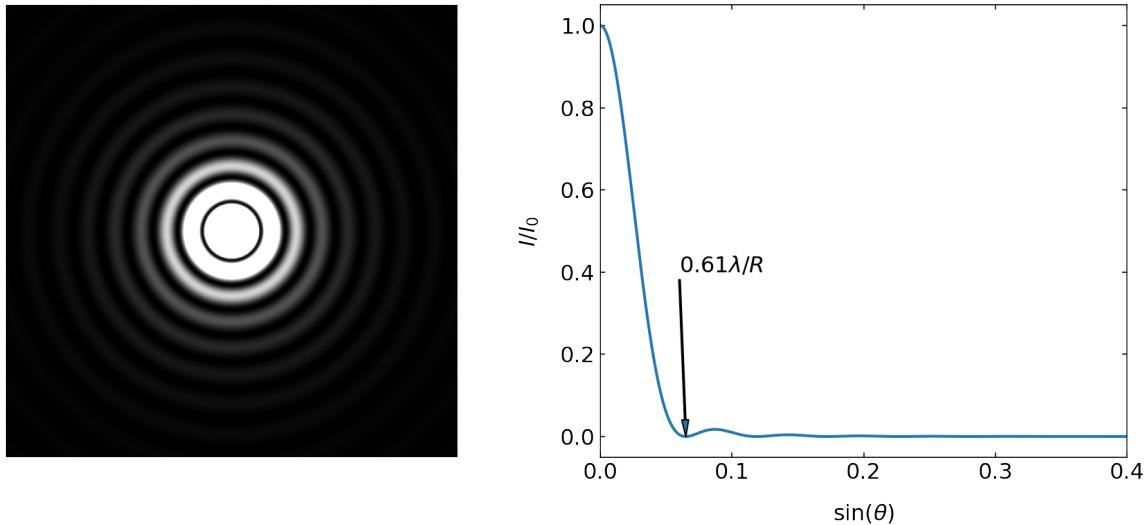


Figure 20.7: Diffraction pattern of a circular aperture of radius 5  $\mu\text{m}$ . Note that the intensity scale is saturated. The diffraction rings would otherwise not be visible. The minima of the diffraction pattern in this plot are at integer multiples of  $0.61\lambda/R$ .

The first minimum of the diffraction pattern occurs when:

$$x_1 = 1.22\pi = \frac{2\pi R}{\lambda} \sin(\theta_1)$$

Solving for  $\sin(\theta_1)$ :

$$\sin(\theta_1) = 0.61 \frac{\lambda}{R}$$

This follows the same general principle we've seen before: the angular spread is proportional to wavelength divided by aperture size. The central bright region up to this first minimum is known as the **Airy disc**, and in microscopy, this defines a resolution element or **resel**.

## 20.4 Application: Diffraction Grating

We now combine the concepts of single-slit diffraction and multiple-wave interference to understand the behavior of diffraction gratings. Diffraction gratings are important in spectroscopy and the compression of short laser pulses.

Consider a diffraction grating with  $N$  slits, each of width  $b$ , and separated by a distance  $d$ . Each slit acts as a source of diffraction, producing an intensity pattern that oscillates with decreasing amplitude. The width of this diffraction pattern is determined by  $\lambda/b$ , and the pattern is given by:

$$I(\theta) = I_s \frac{\sin^2 \left( \pi \frac{b}{\lambda} \sin(\theta) \right)}{\left( \pi \frac{b}{\lambda} \sin(\theta) \right)^2}$$

Now, we have multiple slits, each separated by a distance  $d$ . The interference of waves from these slits is described by:

$$I = I_0 \frac{\sin^2(N\phi/2)}{\sin^2(\phi/2)}$$

where the phase difference  $\phi$  between waves from neighboring slits is given by:

$$\phi = k\Delta s = \frac{2\pi}{\lambda} d \sin(\theta)$$

Combining these expressions, the intensity distribution for a diffraction grating is:

$$I(\theta) = I_0 \frac{\sin^2\left(\pi \frac{b}{\lambda} \sin(\theta)\right)}{\left(\pi \frac{b}{\lambda} \sin(\theta)\right)^2} \frac{\sin^2\left(N\pi \frac{d}{\lambda} \sin(\theta)\right)}{\sin^2\left(\pi \frac{d}{\lambda} \sin(\theta)\right)}$$

This formula describes the intensity pattern produced by a diffraction grating, which is the product of the single-slit diffraction pattern and the multiple-slit interference pattern.

### **i** Diffraction Grating

The intensity distribution generated by a diffraction grating from monochromatic light and observed in the far field is given by

$$I(\theta) = I_0 \frac{\sin^2\left(\pi \frac{b}{\lambda} \sin(\theta)\right)}{\left(\pi \frac{b}{\lambda} \sin(\theta)\right)^2} \frac{\sin^2\left(N\pi \frac{d}{\lambda} \sin(\theta)\right)}{\sin^2\left(\pi \frac{d}{\lambda} \sin(\theta)\right)}$$

where  $\lambda$  is the wavelength of the light,  $b$  is the width of the slit,  $d$  is the distance between the slits, and  $N$  is the number of slits illuminated. The angle of observation is given by  $\theta$ .

## Properties of the Diffraction Pattern

Let's examine the properties of this intensity distribution. The graph below shows the intensity distribution for a diffraction grating with  $N = 8$  slits, a slit distance of  $d = 4 \mu\text{m}$ , a slit width of  $b = 2 \mu\text{m}$ , and a wavelength of 532 nm. We observe the following general properties:

- The intensity pattern consists of main maxima, called **diffraction orders**, characterized by integer numbers. The central peak is the 0th order peak, the first main peak to the right is the 1st diffraction order, and so on.
- The main peaks are separated by  $N - 2$  secondary peaks and  $N - 1$  minima.
- The intensity distribution is characterized by an envelope, which is the diffraction pattern of a single slit (dashed line). In the example below, the 2nd order peak is suppressed. The envelope becomes wider if the slits become narrower.

## Position of the Main Peaks

The position of the main peaks is determined by the condition that the denominator of the multiple-slit interference term is zero. This occurs when the argument is an integer multiple of  $\pi$ , i.e.,  $\pi \frac{d}{\lambda} \sin(\theta) = m\pi$ , or:

$$\sin(\theta) = m \frac{\lambda}{d}$$

where  $m$  is an integer. The first-order diffraction maximum is found at  $\sin(\theta) = \frac{\lambda}{d}$ , independent of the number of slits  $N$ . This means that the position of the main peaks increases linearly with the wavelength  $\lambda$  and decreases with increasing slit distance  $d$ .

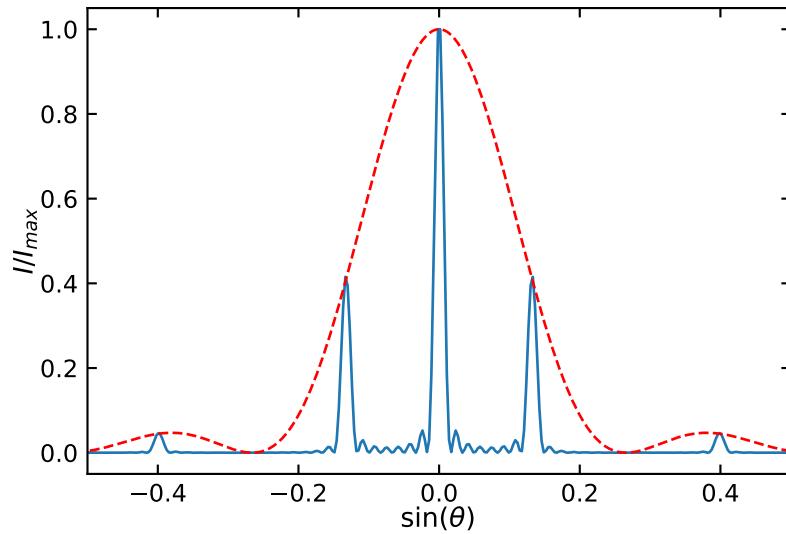


Figure 20.8: Diffraction pattern of a grating where 8 slits with a width of 2  $\mu\text{m}$  and a distance of 4 micrometers are illuminated by a wavelength of 532 nm.

### Influence of the Slit Width

The two plots below show the influence of the slit width while keeping the slit distance the same. We have  $N = 8$  slits with  $d = 4 \mu\text{m}$ , while the slit width is  $b = 2 \mu\text{m}$  on the left side and  $b = 1 \mu\text{m}$  on the right side. The result is an increased width of the envelope. The first minimum of the slit diffraction pattern occurs at  $\sin(\theta) = \frac{\lambda}{b}$ .

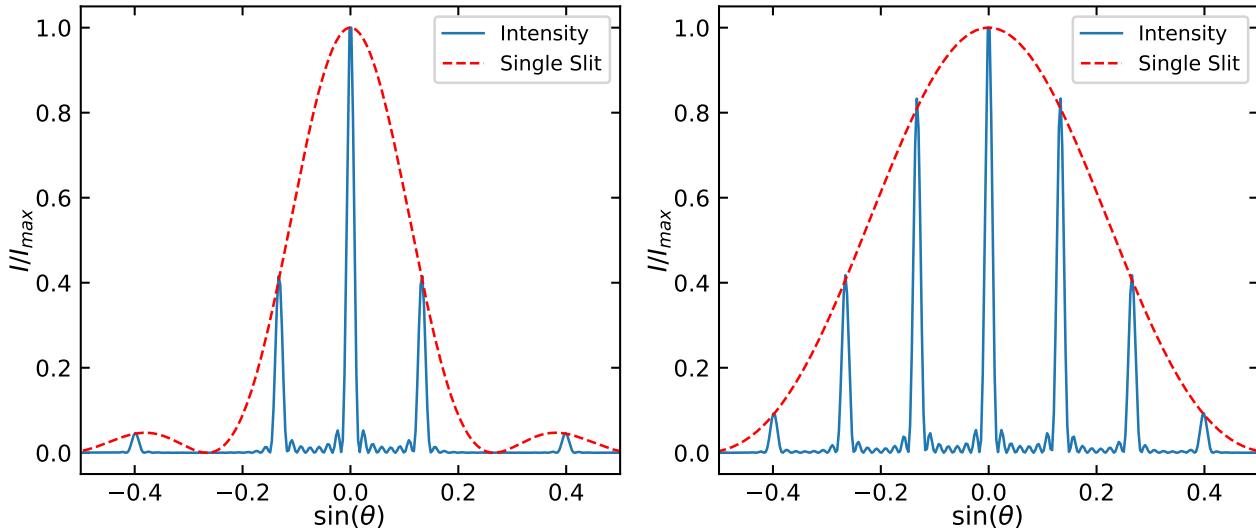


Figure 20.9: Diffraction pattern of a grating with  $N = 8$  slits ( $d = 4 \mu\text{m}$ ,  $b = 1 \mu\text{m}$ ) with  $\lambda = 532 \text{ nm}$ .

### Influence of the Slit Number

When increasing the number of slits, the main diffraction peaks become sharper. The location of the main peaks for a given wavelength remains unchanged, but there are now  $N - 2$  secondary maxima in between. This decreased width of the main peaks is important for the spectral resolution of the grating.

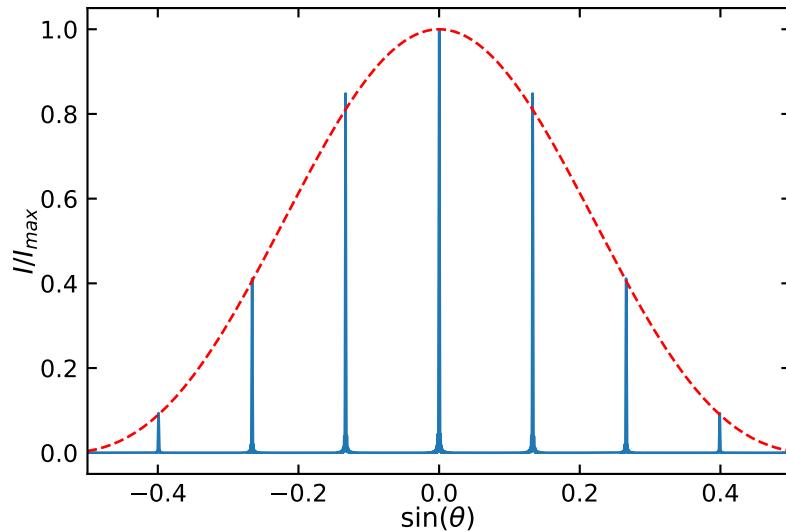


Figure 20.10: Diffraction pattern of a grating with  $N = 100$  slits ( $d = 4 \mu\text{m}$ ,  $b = 1 \mu\text{m}$ ) with  $\lambda = 532 \text{ nm}$ .

## Spectral Resolution

To quantify the spectral resolution, we use a criterion similar to the optical resolution of a microscope: two peaks are separable if the second peak is located at the minimum of the first diffraction pattern. Here, the diffraction patterns refer to different wavelengths  $\lambda_1$  and  $\lambda_2$ .

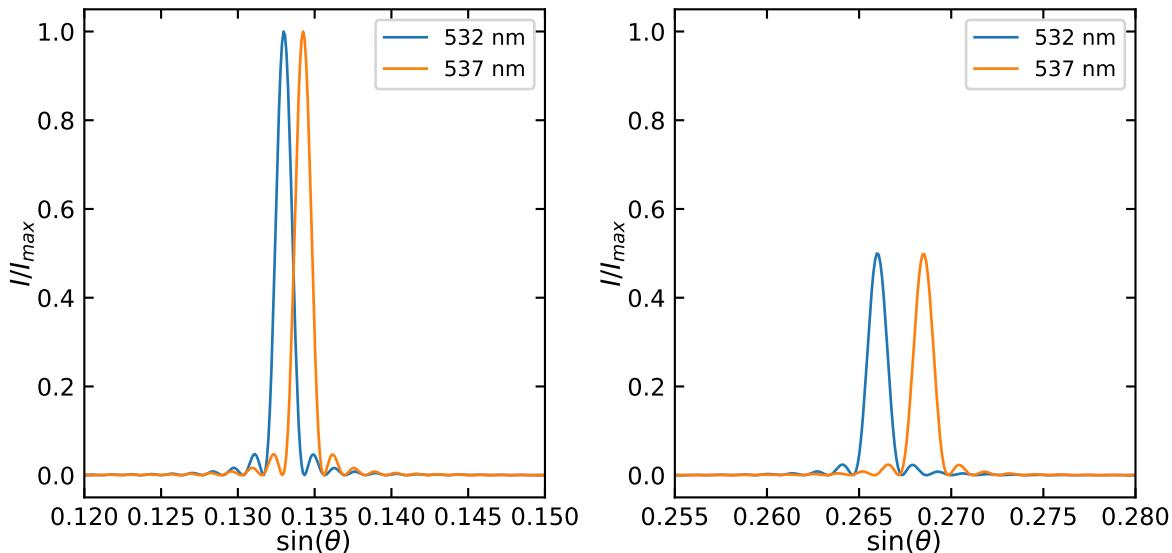


Figure 20.11: Rayleigh resolution limit of a grating with  $N = 100$  slits ( $d = 4 \mu\text{m}$ ,  $b = 1 \mu\text{m}$ ) with  $\lambda_1 = 532 \text{ nm}$  and  $\lambda_2 = 537 \text{ nm}$  in the first order diffraction peak (left) and the second order peak (right).

Consider the  $m$ th order diffraction peak for the wavelength  $\lambda_1$ . This occurs at:

$$\sin(\theta) = m \frac{\lambda_1}{d}$$

The next secondary minimum to larger angles of the diffraction pattern is located where the numerator of the multiple-wave interference term:

$$\sin^2 \left( N\pi \frac{d}{\lambda} \sin(\theta) \right)$$

becomes zero, or the argument:

$$N\pi \frac{d}{\lambda} \sin(\theta) = l\pi$$

becomes a multiple  $l$  of  $\pi$ . For the first-order main peak, we have  $N - 2$  intermediate peaks as well as the 0th and now the first-order peak. Therefore,  $m = l/N$ , and the next minimum after the 1st order peak is at:

$$\sin(\theta_1) = \frac{l+1}{N} \frac{\lambda_1}{d}$$

This angle must correspond to the position of the main peak of the first-order diffraction of the wavelength  $\lambda_2$ , so:

$$\sin(\theta_1) = m \frac{\lambda_2}{d}$$

Combining both equations for the two wavelengths yields:

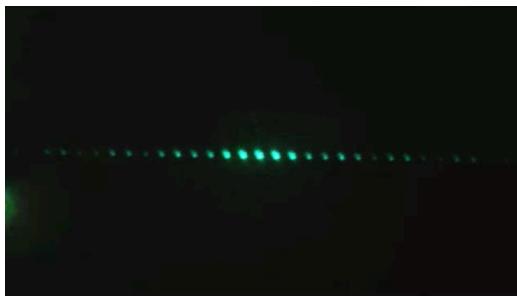
$$\left( m + \frac{1}{N} \right) \frac{\lambda_1}{d} = m \frac{\lambda_2}{d}$$

and after some rearrangements (setting  $\lambda_1 = \lambda$ ):

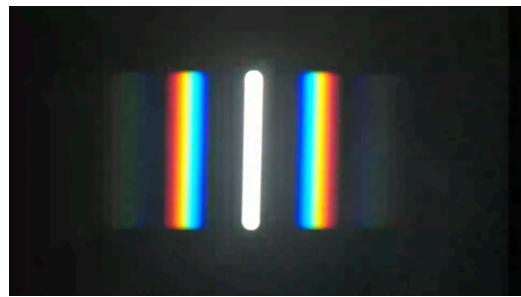
$$R = \frac{\lambda}{\Delta\lambda} = mN$$

This is the resolving power  $R$  of a grating. The ability to resolve two wavelengths increases with the diffraction order  $m$  and the number of slits used for the diffraction. However, the intensity of higher diffraction orders rapidly decreases due to the grating envelope. Therefore, the main parameter to change is the number of illuminated slits.

Our finding is illustrated in the figure above, where we achieve a resolution of about 5 nm when using  $N = 100$  slits at a distance of  $d = 4 \mu\text{m}$ .



(a) Diffraction pattern observed for a grating in the lecture with red light (left) and white light (right).



(b) Diffraction pattern observed for a grating in the lecture with red light (left) and white light (right).

Figure 20.12: Diffraction pattern observed for a grating in the lecture with red light (left) and white light (right).

**i** Diffraction Grating Playground

App to play with the parameters of a grating with a slit distance of 4  $\mu\text{m}$  and a slit width of 2  $\mu\text{m}$ .

# Chapter 21

## Diffraction in Applications

### 21.1 Application: Diffraction in the Human Eye

The human eye provides an excellent real-world example of circular aperture diffraction through the pupil (the opening in the iris). By examining how light diffracts as it enters the eye, we can understand fundamental limits on visual resolution and how the eye is naturally optimized to these constraints.

#### Calculating the Diffraction Limit of the Eye

When light passes through the circular aperture of the pupil, it undergoes diffraction, producing an Airy disk pattern on the retina. The angle to the first minimum (dark ring) of the diffraction pattern is given by:

$$\sin(\theta_1) = 0.61 \frac{\lambda}{R}$$

where: -  $\theta_1$  is the angle to the first minimum, -  $\lambda$  is the wavelength of the light, -  $R$  is the radius of the aperture (pupil).

For an average pupil radius of  $R = 2.5\text{ mm}$  and green light with a wavelength of  $\lambda = 532\text{ nm}$  (to which the human eye is most sensitive), we have:

$$\sin(\theta_1) = 0.61 \times \frac{532 \times 10^{-9} \text{ m}}{2.5 \times 10^{-3} \text{ m}} = 0.61 \times 2.128 \times 10^{-4} \approx 1.298 \times 10^{-4}$$

Thus, the angle to the first minimum is approximately:

$$\theta_1 \approx \sin^{-1}(1.298 \times 10^{-4}) \approx 0.00744^\circ$$

#### Determining the Size of the Airy Disk on the Retina

The distance from the pupil to the retina (the image plane) is approximately  $L = 20\text{ mm}$  (or 2 cm). The linear radius  $r$  of the Airy disk on the retina is calculated by:

$$r = L \sin(\theta_1) = 20 \text{ mm} \times 1.298 \times 10^{-4} = 2.596 \times 10^{-3} \text{ mm} = 2.596 \mu\text{m}$$

So, the diameter  $D$  of the Airy disk (central bright spot) is:

$$D = 2r = 2 \times 2.596 \mu\text{m} = 5.192 \mu\text{m}$$

This means that the smallest spot of light that can be formed on the retina due to diffraction is about  $5.19 \mu\text{m}$  in diameter.

### Comparing with Photoreceptor Spacing in the Fovea

The fovea is a small region in the retina responsible for sharp central vision. It contains a high density of cone photoreceptor cells. The average density of cones in the fovea is approximately 150,000 cells per square millimeter. To find the average spacing  $d$  between these cells, we proceed as follows:

#### Area per Cell:

$$\text{Area per cell} = \frac{1 \text{ mm}^2}{150,000} = 6.667 \times 10^{-6} \text{ mm}^2 = 6.667 \mu\text{m}^2$$

#### Linear Spacing Between Cells:

Assuming a square packing (for simplicity), the linear spacing  $d$  is:

$$d = \sqrt{\text{Area per cell}} = \sqrt{6.667 \mu\text{m}^2} \approx 2.58 \mu\text{m}$$

In reality, the photoreceptors are more closely packed in a hexagonal arrangement, but this calculation gives a good approximation.

#### Analysis of the Results

The analysis reveals that the diameter of the Airy disk is approximately  $5.192 \mu\text{m}$ , while the center-to-center spacing of photoreceptors in the human eye is about  $2.58 \mu\text{m}$ . This observation is significant because the diameter of the Airy disk is roughly twice the photoreceptor spacing, indicating that the central maximum of the diffraction pattern spans about two photoreceptors.

The close correspondence between the diffraction limit of the eye and the spacing of photoreceptor cells is noteworthy for several reasons. Firstly, the diffraction limit establishes the fundamental constraint on the resolving power of the eye, determining the smallest angular separation between two points of light that can be distinguished. Secondly, the density of photoreceptors is sufficiently high to sample the details provided by the optical system up to this diffraction limit.

Increasing the density of photoreceptors within the area of the Airy disk would not enhance visual resolution due to two primary factors. The first factor is the physical limitation imposed by the diffraction limit, which is a fundamental constraint arising from the wave nature of light and the size of the pupil. Consequently, resolution cannot be improved beyond this limit merely by increasing photoreceptor density. The second factor is related to signal intensity. Adding more photoreceptors in the same area would result in each cell receiving less light, given that the total light intensity is fixed. This reduction in light per photoreceptor could potentially decrease the signal-to-noise ratio, making it more challenging to detect light.

### Biological Optimization

The design of the human eye exemplifies a natural optimization process. The density of photoreceptors is matched to the optical resolving power of the eye, ensuring that the visual system extracts the maximum amount of information without unnecessary redundancy. This efficient use of resources reflects an evolutionary adaptation, where biological systems have evolved to align anatomical structures with physical laws, optimizing functions such as vision to confer survival advantages. Over time, this alignment has resulted in a visual system that is finely tuned to the constraints and capabilities imposed by the physics of light and the anatomy of the eye.

- Land, M. F., & Nilsson, D.-E. (2012). *Animal Eyes* (2nd ed.). Oxford University Press.
- Williams, D. R. (1988). Topography of the foveal cone mosaic in the living human eye. *Vision Research*, 28(3), 433–454.

## 21.2 Application: Resolution of an Optical Microscope

The resolution of an optical microscope is fundamentally limited by the diffraction of light as it passes through the optical components, particularly the objective lens. Diffraction causes point sources of light to produce blurred images rather than perfect points, affecting the microscope's ability to distinguish between two closely spaced objects.

### Rayleigh's Criterion for Resolution

**Key Question:** *How close can two point sources be while still being perceived as distinct entities by an optical system?*

To answer this, we need to consider two essential aspects of how a lens modifies light:

1. **Wavefront Transformation:** A lens alters the curvature of incoming wavefronts, focusing parallel rays (plane waves) to a point in the focal plane.
2. **Finite Aperture Effects:** The lens has a finite size and acts as a circular aperture, introducing diffraction effects that spread the image of a point source into a diffraction pattern known as the Airy disk.

#### Visual Representation:

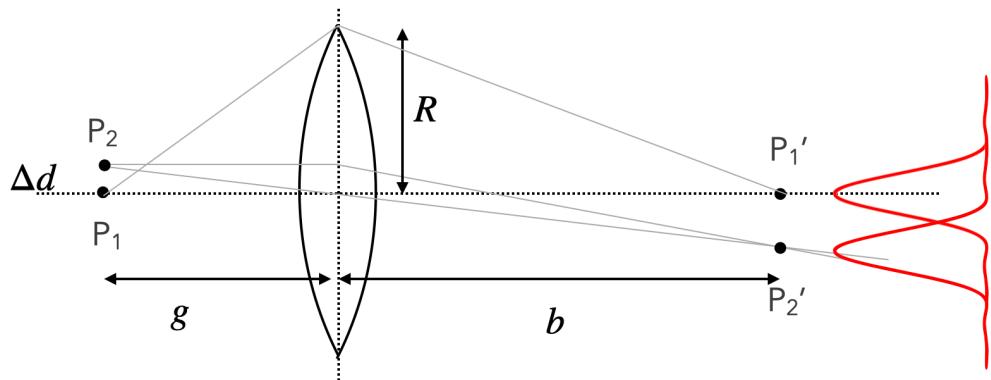


Figure 21.1: Illustration showing two point sources and their overlapping diffraction patterns as they approach each other.

Each point source produces its own diffraction pattern. As the sources move closer, their patterns begin to overlap, making it harder to distinguish between them.

#### Rayleigh's Resolution Criterion:

- Two point sources are considered **just resolvable** when the principal maximum (center) of one Airy pattern coincides with the first minimum (dark ring) of the other.

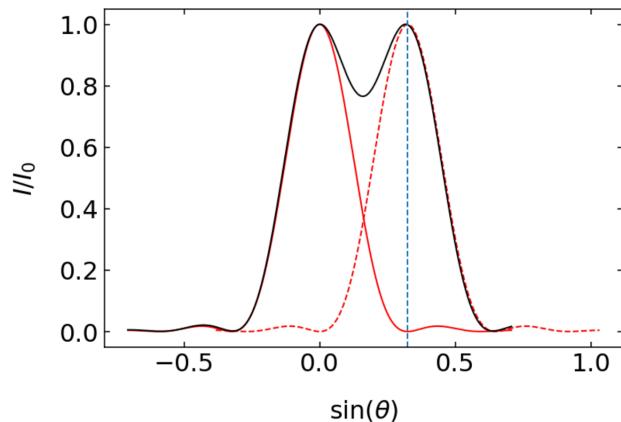


Figure 21.2: Graph depicting Rayleigh's criterion, showing the intensity profiles of two overlapping Airy patterns.

For **incoherent light sources** (where the light waves are not in phase), this criterion corresponds to a 26% dip in intensity between the two peaks, which is generally sufficient for the human eye or detectors to distinguish the two sources as separate.

#### Angle to the First Minimum:

The angle  $\theta_1$  to the first minimum of the diffraction pattern from a circular aperture of radius  $R$  is given by:

$$\sin(\theta_1) = 1.22 \frac{\lambda}{2R}$$

Since the diameter of the aperture  $D = 2R$ , this can also be written as:

$$\sin(\theta_1) = 1.22 \frac{\lambda}{D}$$

The factor 1.22 arises from the first zero of the Bessel function  $J_1$  that describes the diffraction pattern of a circular aperture.

#### Small Angle Approximation:

For small angles (common in optical systems),  $\sin(\theta_1) \approx \theta_1$  in radians.

#### Relating Angular to Linear Separation in the Image Plane:

The angular resolution  $\theta_1$  corresponds to a linear separation  $\Delta x$  in the image plane (at image distance  $b$ ):

$$\theta_1 = \frac{\Delta x}{b}$$

#### Combining the Equations:

Substituting  $\theta_1$ :

$$\frac{\Delta x}{b} = 1.22 \frac{\lambda}{D}$$

Solving for  $\Delta x$ :

$$\Delta x = 1.22 \frac{\lambda b}{D}$$

### Considering the Object Plane:

The magnification  $M$  of the optical system is:

$$M = \frac{b}{g}$$

where  $g$  is the object distance. The corresponding separation in the object plane ( $\Delta d$ ) is:

$$\Delta d = \frac{\Delta x}{M} = \frac{\Delta x g}{b}$$

Substituting  $\Delta x$ :

$$\Delta d = 1.22 \frac{\lambda b}{D} \times \frac{g}{b} = 1.22 \frac{\lambda g}{D}$$

### Introducing Numerical Aperture (NA):

The **numerical aperture** (NA) of the lens is defined as:

$$\text{NA} = n \sin(\alpha)$$

where:

- $n$  is the refractive index of the medium between the object and the lens.
- $\alpha$  is the half-angle of the maximum cone of light that can enter or exit the lens.

Since  $\sin(\alpha) = \frac{R}{g}$ , we have:

$$D = 2R = 2g \sin(\alpha)$$

Substituting  $D$  into  $\Delta d$ :

$$\Delta d = 1.22 \frac{\lambda g}{2g \sin(\alpha)} = \frac{1.22 \lambda}{2 \sin(\alpha)} = \frac{0.61 \lambda}{\sin(\alpha)}$$

Therefore, incorporating the refractive index  $n$ :

$$\Delta d = \frac{0.61 \lambda}{n \sin(\alpha)} = \frac{0.61 \lambda}{\text{NA}}$$

Under Rayleigh's resolution criterion, several key factors influence the resolving power of an optical system.

Firstly, a **higher numerical aperture (NA)** improves resolution. This increase can be achieved by enhancing either the refractive index  $n$  of the medium between the object and the lens or the sine of the collection angle  $\sin(\alpha)$ . Since the NA is defined as  $\text{NA} = n \sin(\alpha)$ , a larger NA allows the lens to gather more diffracted light, thereby resolving finer details in the image. In air, where the refractive index is approximately  $n \approx 1$ , the **maximum achievable NA is less than 1**, which limits the resolution. This limitation arises because the maximum value of  $\sin(\alpha)$  is 1 (when  $\alpha = 90^\circ$ ), so the NA in air cannot exceed 1. In practical systems, the collection angle  $\alpha$  is much less than  $90^\circ$ , further reducing the NA and thus the resolution. Immersion lenses, which use a medium with a higher refractive index (such as water or oil), can achieve higher NAs, overcoming this limitation and improving resolution.

Secondly, using **shorter wavelengths** ( $\lambda$ ) of light leads to better resolution. According to the formula  $\Delta d = \frac{0.61 \lambda}{\text{NA}}$ , the minimum resolvable distance  $\Delta d$  is directly proportional to the wavelength. Therefore, decreasing the wavelength reduces  $\Delta d$ , allowing the optical system to distinguish smaller features of the object.

**i Note****Rayleigh's Resolution Criterion**

Two incoherent point sources can be resolved when their minimum separation  $\Delta d$  satisfies:

$$\Delta d \geq \frac{0.61\lambda}{NA}$$

**Where:**

- $\Delta d$  is the minimum resolvable distance between the two point sources.
- $\lambda$  is the wavelength of the light used for imaging.
- $NA = n \sin(\alpha)$  is the numerical aperture of the optical system.
  - $n$  is the refractive index of the medium.
  - $\alpha$  is the half-angle of the maximum cone of light entering the lens.

**Abbe's Criterion for Resolution****Limitations of Rayleigh's Criterion:**

- Rayleigh's criterion applies to **incoherent** light sources, where the intensities of the diffraction patterns add directly.
- It does not fully account for the effects of coherence and interference in the imaging process.

**Ernst Abbe** developed a theory that considers the imaging of **coherent** light sources, where the phases of the waves are correlated. It emphasizes the importance of diffracted orders and spatial frequencies in the formation of images.

The key concepts in Abbe's theory are

**Diffraction Grating Analogy:**

- An object with fine details can be thought of as a diffraction grating that scatters light into multiple diffraction orders.
- The ability to resolve these fine details depends on the optical system's capacity to capture these diffracted orders.

**Spatial Frequencies:**

- The finer the details in the object, the higher the spatial frequency.
- High spatial frequencies correspond to larger angles in the diffraction pattern.

**Optical Transfer Function:**

- Describes how different spatial frequencies are transmitted through the optical system.
- An optical system with a larger NA can transmit higher spatial frequencies, improving resolution.

Following that analogy, Abbe's criterion states that the minimum resolvable distance  $\Delta d$  between two points in the object plane is given by:

$$\Delta d = \frac{\lambda}{2NA}$$

**i Rayleigh's and Abbe's criteria**

- **Abbe's Limit:**
  - $\Delta d = \frac{\lambda}{2NA}$
  - Emphasizes coherent imaging and the transmission of at least two diffracted orders (zeroth and first) for resolution.
- **Rayleigh's Limit:**
  - $\Delta d = \frac{0.61\lambda}{NA}$

- Based on the visibility of intensity dips in the overlapping Airy patterns of incoherent sources.

#### Implications in Microscopy:

##### Coherent Illumination:

- Techniques like phase-contrast or interference microscopy rely on coherent light.
- Abbe's criterion is more appropriate for these methods.

##### Incoherent Illumination:

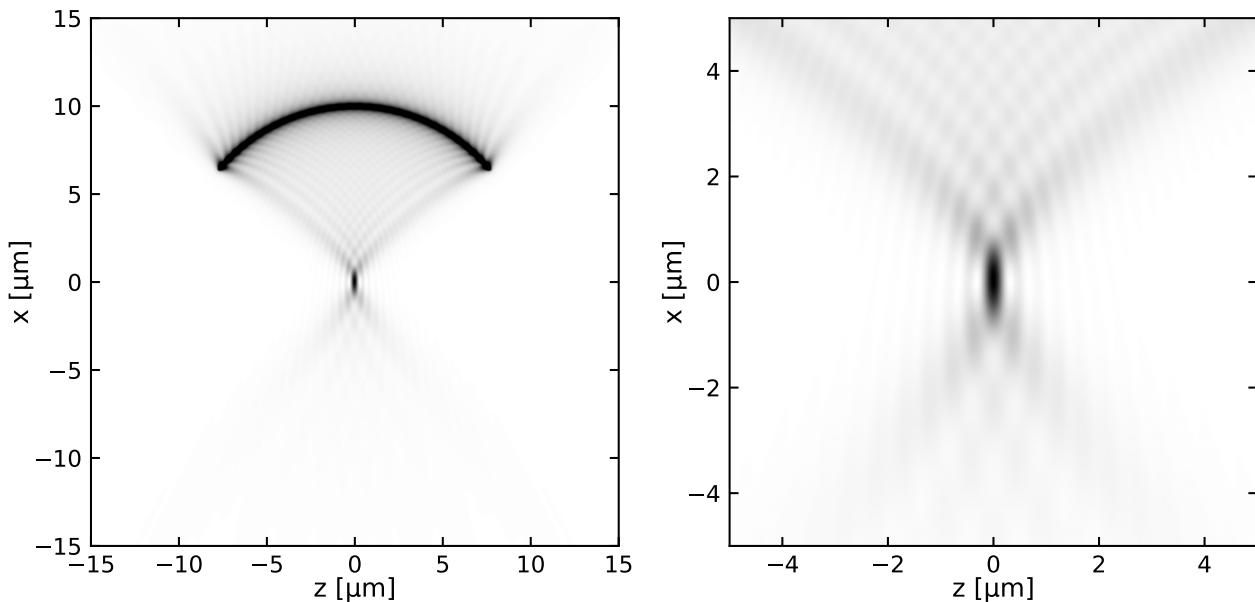
- Common in conventional bright-field microscopy.
- Rayleigh's criterion provides a practical resolution limit.

##### Importance of Abbe's Criterion:

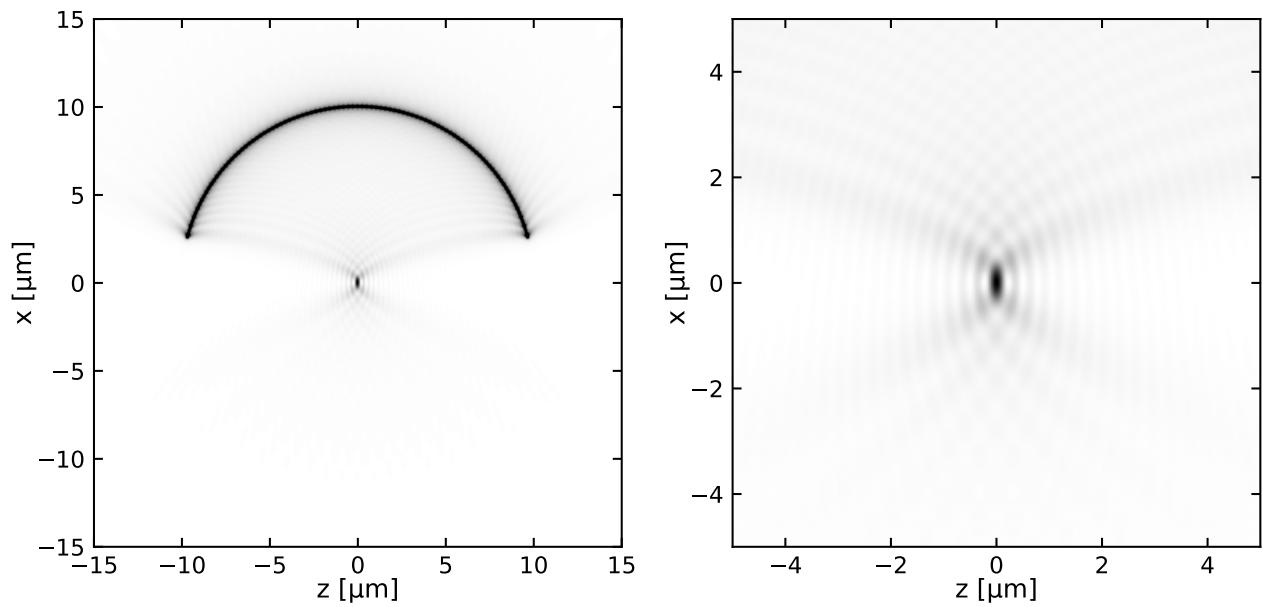
- Highlights the role of interference between diffracted waves in image formation.
- Demonstrates that resolution is fundamentally limited by the wavelength of light and the NA of the system.
- Suggests that capturing higher spatial frequencies (larger NA) leads to better resolution.

## 21.3 Using Huygens Sources for more

The Huygens principle is a powerful tool to calculate the diffraction pattern of an aperture. The idea is to consider the aperture as a collection of point sources, which emit spherical waves. The superposition of all these waves will then give the total wave field. Below is a Python code which demonstrates the calculation of the diffraction pattern of a spherical mirror. It uses Huygens sources placed on an arc (see left). The right image shows the resulting diffraction pattern in the focal plane of the mirror.



The pattern changes with increasing angle of the arc, which is consistent with our knowledge of the numerical aperture defining the resolution of an optical system. The changes are especially visible along the vertical axis.



# Chapter 22

## Fresnel Zones

We want to take a more general look at diffraction by exploring a concept known as Fresnel zones. Consider spherical waves of wavelength  $\lambda$  emitted from a source, as indicated by the solid line in the sketch below.

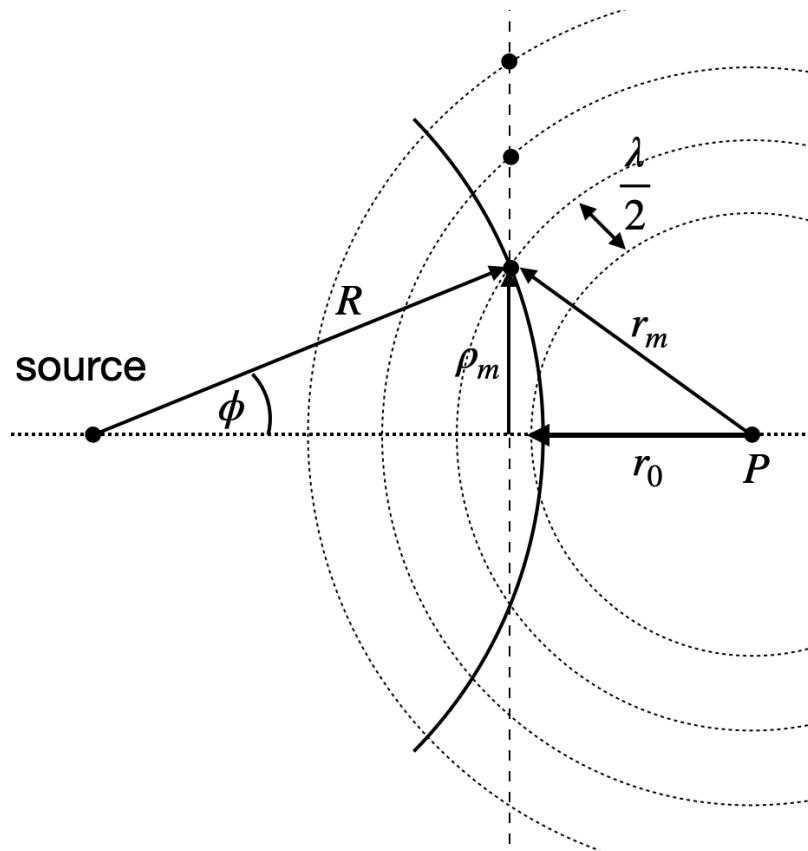


Figure 22.1: Construction of the Fresnel Zones.

We will examine the intensity of the wave at a point  $P$ . To do this, we consider the amplitude contributions from all points on the wavefront, as each point on the wavefront acts as a Huygens source contributing to the intensity at point  $P$ .

Instead of calculating the intensity explicitly, we will analyze the distances of individual points on the wavefront from point  $P$ . Specifically, we look at concentric circles around point  $P$ , where the radius of each circle increases by  $\lambda/2$ , i.e.,

$$r_m = r_0 + m \frac{\lambda}{2}$$

where  $m$  is an integer. The regions between  $r_m$  and  $r_{m+1}$  are called **Fresnel zones**. If we consider two neighboring zones, each zone contains pairs of points that are exactly  $\lambda/2$  out of phase. This means that these pairs of points would lead to destructive interference. If we remove these points, we are left with constructive interference along the optical axis only. We can construct such an aperture by calculating the ring radius

$$\rho_m^2 = \left( r_0 + m \frac{\lambda}{2} \right)^2 - r_0^2$$

according to the sketch above. This yields

$$\rho_m^2 = r_0 m \lambda + m^2 \frac{\lambda^2}{4}$$

For  $r_0 \gg \lambda$ , we can simplify the above formula to

$$\rho_m = \sqrt{mr_0\lambda}$$

which gives the radius of the individual zones. The width of the zones is given by

$$\Delta\rho_m = \rho_{m+1} - \rho_m = \sqrt{r_0\lambda}(\sqrt{m+1} - \sqrt{m})$$

## 22.1 Fresnel Zone Plate

If we now fill the ring from  $\rho_m$  to  $\rho_{m+1}$  on a glass slide but leave the ring from  $\rho_{m+1}$  to  $\rho_{m+2}$  transparent, we create a so-called **Fresnel zone plate**. Here, the radius in the first zone  $r$  ranges from  $r_0$  to  $r_0 + \lambda/2$ . The next zone will range from  $r_0 + \lambda/2$  to  $r_0 + \lambda$  but is removed from its contribution to the point.

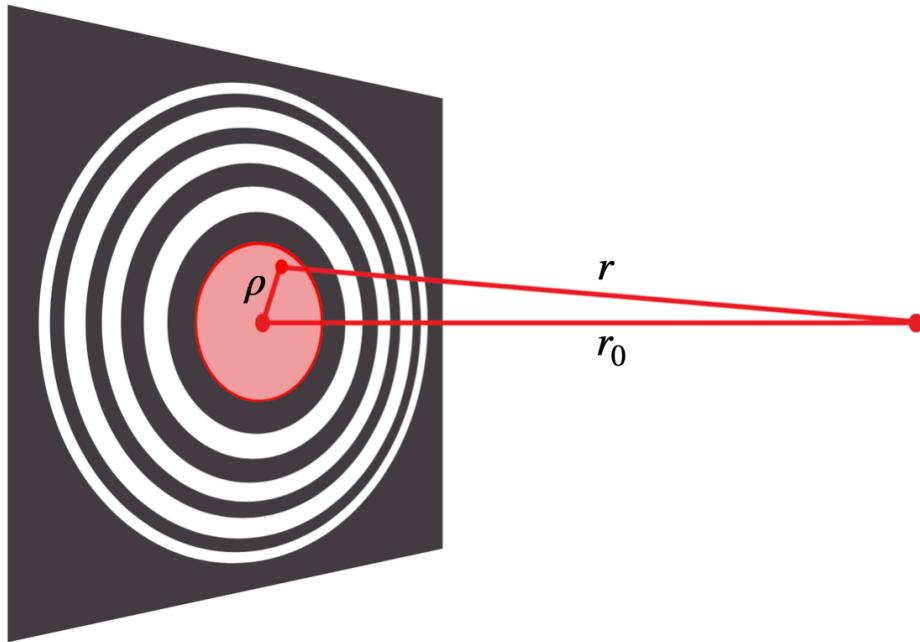


Figure 22.2: Fresnel zone plate removing destructive interference to the point on the optical axis.

The Fresnel zone plate can be constructed by defining the inner reference zone in an arbitrary way. One may either block or transmit the direct path from the light source along the optical axis, resulting in either the odd or even zones being transparent.

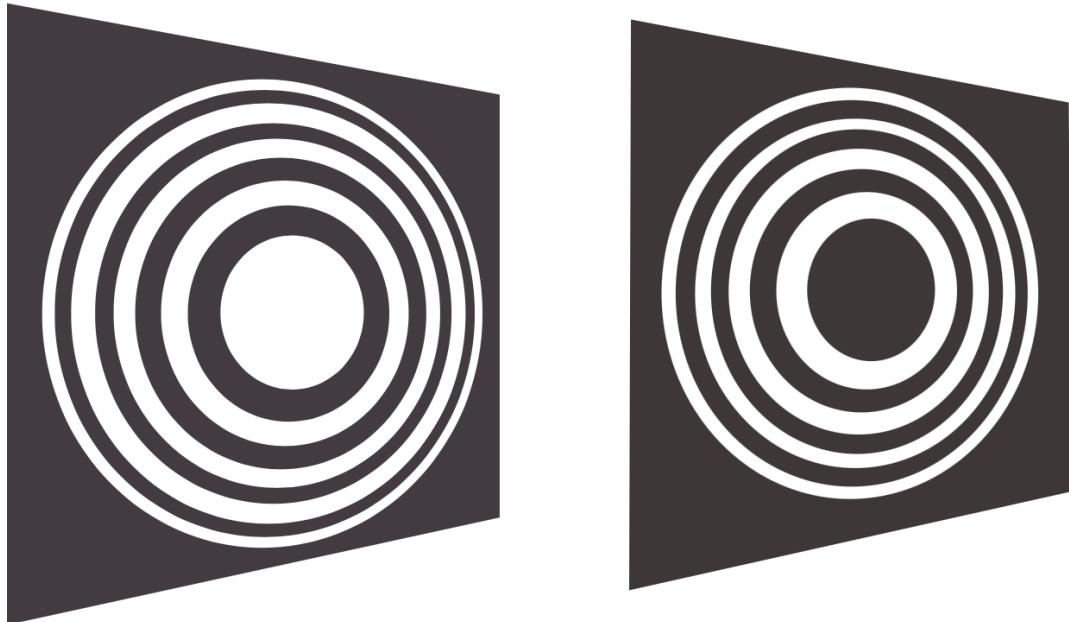


Figure 22.3: Fresnel zone plates with odd (left) or even (right) zones transparent delivering the same result.



Such zone plates are important for applications where focusing of radiation is required but the refractive indices are not large enough to create strong enough refraction. This is especially true for X-ray radiation.

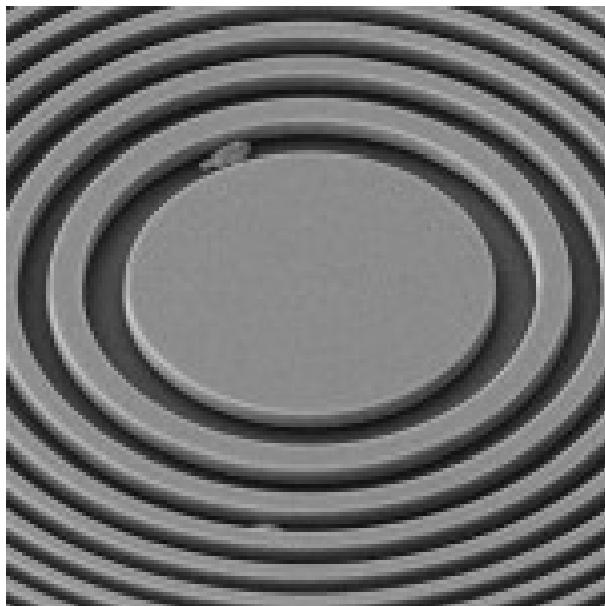


Figure 22.4: Fresnel zone plates for X-ray radiation. Image taken from Ion beam lithography for Fresnel zone plates in X-ray microscopy - Optics Express, Vol. 21 Issue 10, pp.11747-11756 (2013).

Below is an calculation of the intensity pattern at the focal distance of a zone plate from many spherical wave sources if the destructively interfering waves are not removed (left) and if they are removed.

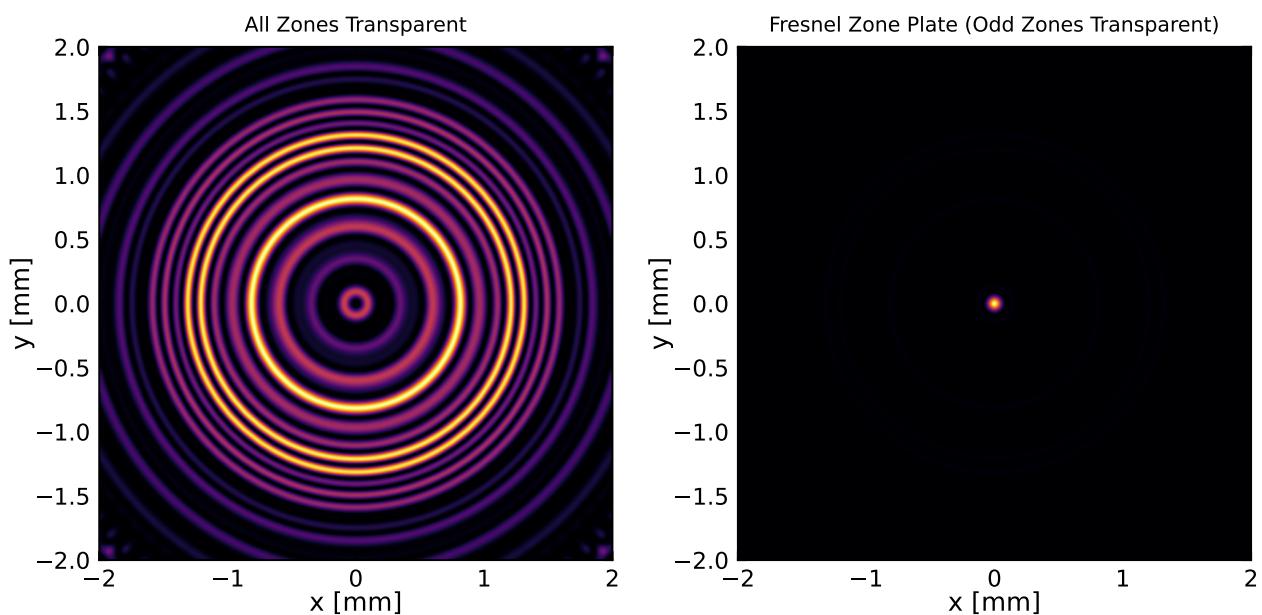


Figure 22.5: Consider with care. Need to check the result again.

## 22.2 Applications and Importance of Fresnel Zone Plates

Fresnel zone plates are used in various applications, particularly where traditional lenses are ineffective. Some key applications include:

**X-ray Microscopy:** Fresnel zone plates are used to focus X-rays, which have very short wavelengths and

require special techniques for focusing. Traditional lenses are not effective for X-rays due to their low refractive indices.

**Optical Systems:** In optical systems, Fresnel zone plates can be used to create focal points without the need for bulky lenses. This is particularly useful in compact optical devices.

**Holography:** Fresnel zone plates are used in holography to create and reconstruct holograms. They help in manipulating the wavefronts to produce the desired holographic images.

**Astronomy:** In astronomy, Fresnel zone plates can be used in telescopes to focus light from distant stars and galaxies. They offer an alternative to traditional lenses and mirrors.



## Chapter 23

# Diffraction Integral

In the last section about Fresnel zones and the zone plate, we considered how different paths contribute to the intensity at a point on the optical axis. We would like to generalize this idea to an integral formulation that allows us to calculate any kind of diffraction pattern.

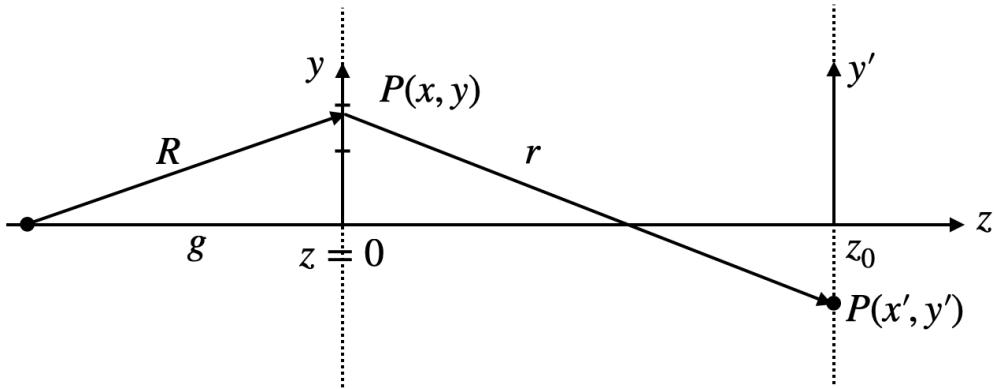


Figure 23.1: Diffraction integral.

Assume we have a light source  $S$  as shown in the image above, which emits a spherical wave (though it does not necessarily have to be a spherical wave). The spatial amplitude of this wave at the point  $P(x, y)$  at a tiny aperture element  $d\sigma$  is given by:

$$U_s(x, y) = U_0(x, y)e^{i\phi(x, y)}$$

where

$$U_0 = \frac{A}{R} = \frac{A}{\sqrt{g^2 + x^2 + y^2}}$$

and

$$\phi(x, y) = -kR$$

This represents the amplitude of the Huygens wave, which emanates from the point  $P(x, y)$  and propagates towards the screen at  $P(x', y')$ . This Huygens wave contributes a fraction of an amplitude  $dU_p$  to the total amplitude at point  $P(x', y')$ , which is given by:

$$dU_p = C \frac{U_s d\sigma}{r} e^{-ikr}$$

with  $C = \frac{i \cos(\theta)}{\lambda}$ , known as the obliquity factor, found through a more detailed calculation.

The total amplitude at the point  $P(x', y')$  is then given by the integral over all contributions:

$$U_p = \iint C U_s \frac{e^{-ikr}}{r} dx dy$$

where  $dx dy = d\sigma$ . The integral runs over all positions in the aperture plane  $(x, y)$  where there is an opening. This integral is called the **Fresnel-Kirchhoff diffraction integral** and allows us to calculate complex scalar diffraction patterns.

This formulation generalizes the concept of Fresnel zones and provides a powerful tool for analyzing and predicting diffraction patterns for various aperture shapes and configurations.

## 23.1 Fresnel Approximation

The diffraction integral does not always need to be calculated in full; we can use approximations to obtain diffraction patterns in different regimes. One such approximation is the **Fresnel approximation**, which yields the diffraction pattern in the **near field**.

The distance  $r$  from the point  $P(x, y)$  to the point  $P(x', y')$  can be written as:

$$r = \sqrt{z_0^2 + (x - x')^2 + (y - y')^2}$$

Using a binomial expansion for small angles, we can approximate this as:

$$r \approx z_0 \left( 1 + \frac{(x - x')^2}{2z_0^2} + \frac{(y - y')^2}{2z_0^2} + \dots \right)$$

In this approximation, we assume that  $\cos(\theta) = z_0/r \approx 1$  and  $C = i/\lambda$ , considering small diffraction angles. Using this approximation, we find the amplitude of the wave at a point  $P(x', y')$ :

$$U(x', y', z_0) = i \frac{e^{-ikz_0}}{\lambda z_0} \iint U_s(x, y) \exp \left[ -\frac{ik}{2z_0} ((x - x')^2 + (y - y')^2) \right] dx dy$$

As the integration is over  $x$  and  $y$ , we can factor out all screen coordinate elements, yielding:

$$U(x', y', z_0) = i \frac{e^{-ikz_0}}{\lambda z_0} e^{-\frac{ik}{2z_0}(x'^2 + y'^2)} \iint U_s(x, y) e^{-\frac{ik}{2z_0}(x^2 + y^2)} e^{\frac{ik}{2z_0}(xx' + yy')} dx dy$$

This is the **Fresnel approximation**. It simplifies the calculation of the diffraction pattern in the near field by making reasonable assumptions about the geometry and angles involved.

## 23.2 Fraunhofer Approximation

If we further assume that the aperture is small as compared to the distance at which we observe the diffraction pattern, we can further simplify the Fresnel approximation to yield the Fraunhofer approximation giving the diffraction pattern in the far field. The condition is

$$z_0 \gg \frac{1}{\lambda} (x^2 + y^2)$$

In this case we can neglect the term

$$e^{-\frac{ik}{2z_0}(x^2+y^2)} \approx 1$$

which results in

$$U(x', y', z_0) = i \frac{e^{-ikz_0}}{\lambda z_0} e^{-\frac{ik}{2z_0}(x'^2+y'^2)} \iint U_s(x, y) e^{\frac{ik}{2z_0}(xx'+yy')} dx dy$$

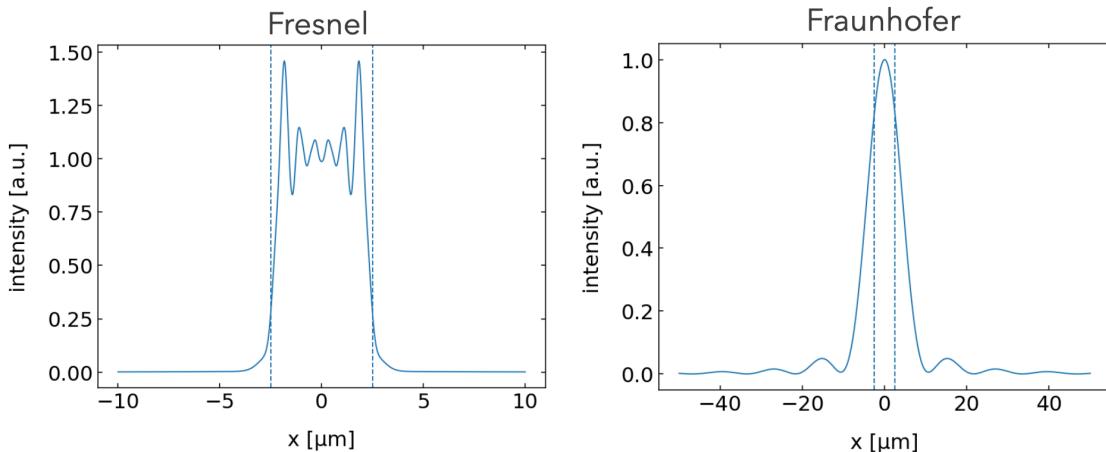


Figure 23.2: Diffraction pattern of a slit in the near field (Fresnel diffraction, left) and the far field (Fraunhofer diffraction, right).

While these formulas provide the mathematical tools, we may obtain a more intuitive idea about the different approximation in the following way. Consider the image below, where we would like to know about the diffraction intensity of a slit of width  $b$  at the optical axis at a distance  $D$ .

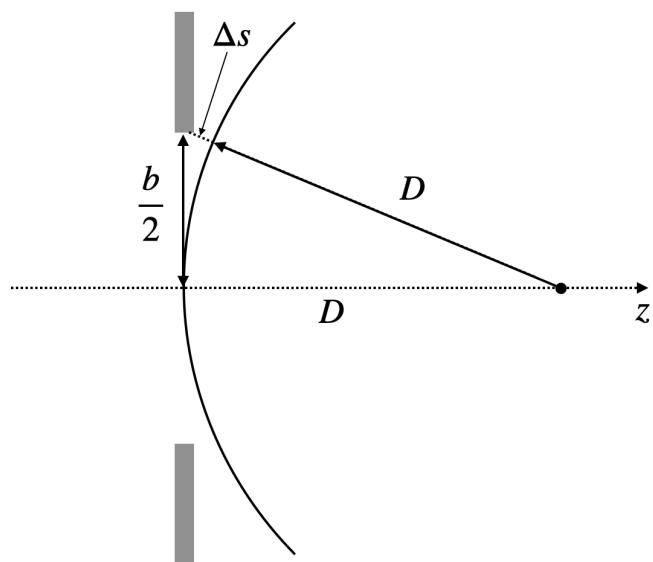


Figure 23.3: Illustration of the importance of additional geometrical path length difference for the discrimination of Fresnel (near-field) and Fraunhofer (far-field) diffraction.

The waves from the center of the slit and the edge have to travel towards that point a different pathlength, which we may calculate to

$$\Delta s = \sqrt{\frac{b^2}{4} + D^2} - D \quad (23.1)$$

$$= D\sqrt{\frac{b^2}{4D^2} + 1} - D \quad (23.2)$$

We may develop the square root into a Taylor series and obtain

$$\Delta s = \frac{b^2}{8D} - \frac{b^4}{128D^3} + O(4) \quad (23.3)$$

$$\approx \frac{b^2}{8D} \quad (23.4)$$

The second order correction term  $\frac{b^2}{8D}$  decreases quadratic with the distance  $D$  of the point, which means that at large distances, we can safely assume  $\Delta s = 0$  on the axis, i.e. all waves arriving at that point have to travel the same distance. This corresponds to the far-field approximation. To be more specific we require

$$\frac{b^2}{8D^2} < \frac{\lambda}{8}$$

or

$$\frac{b^2}{\lambda D} < 1$$

to be fulfilled to be in the far field.

$$F = \frac{b^2}{\lambda D} \begin{cases} \ll 1, & \text{Fraunhofer} \\ \approx 1, & \text{Fresnel} \\ \gg 1, & \text{Full vector} \end{cases}$$

This number  $F$  is called the Frensel number and gives us an idea by how far the dimensions of the opening contribute to the diffraction pattern rather than the direction of the wave propagation only.

### 23.3 Babinet's Principle

The above considerations of diffraction have some intriguing consequence. Consider the two apertures in the image below.

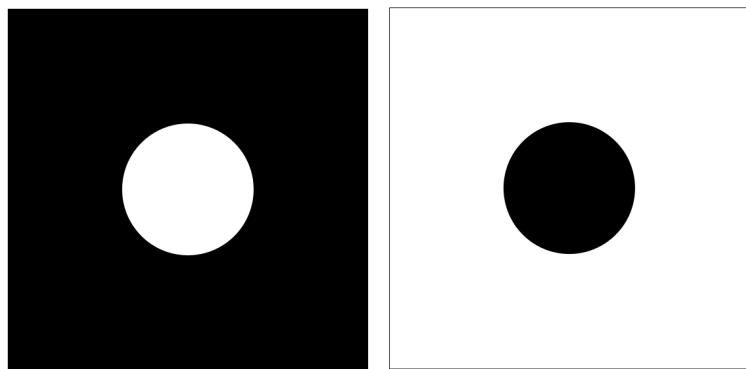


Figure 23.4: Two complementary apertures, which have the same diffraction pattern in the far field.

The left aperture will create in the far field an amplitude distribution  $U_h$ , while the inverse aperture on the right will cause an amplitude  $U_d$ . If we combine both amplitudes in the far field, we obtain a total amplitude distribution

$$U = U_h + U_d$$

In the case when we have two complementary apertures, that total amplitude has to be zero, when hole and dot are placed at the same position. We therefore obtain

$$U_h = -U_d$$

and therefore

$$I_h = I_d$$

This is the Principle of Babinet which states:

### **i** Babinet's Principle

Babinet's principle states that the far-field diffraction intensity distribution of complementary apertures is identical. This means that an opaque object and its complementary aperture (where the object is replaced by a transparent region and vice versa) produce the same diffraction pattern in the far field.

The images below show an experimental demonstration of Babinet's principle on a slit and a wire.

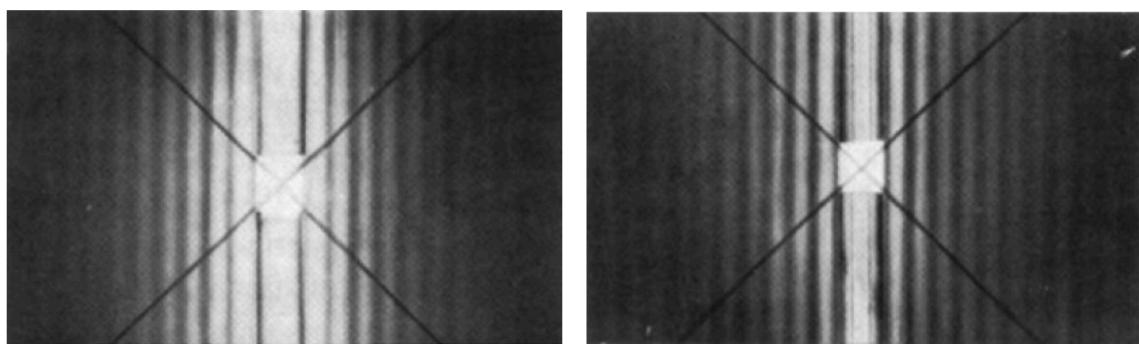


Figure 23.5: Babinet's principle demonstrated experimentally on a slit (left) and a wire (right).



## Chapter 24

# Interferometers and other Coherence Applications

### Michelson Interferometer

The Michelson interferometer is an essential optical instrument used to measure the interference of light waves. It consists of a coherent light source, such as a laser, which emits a beam directed towards a beam splitter. The beam splitter divides the light into two beams: one reflected towards a fixed mirror and the other transmitted towards a movable mirror. After reflecting off their respective mirrors, the beams are recombined at the beam splitter, where they interfere.

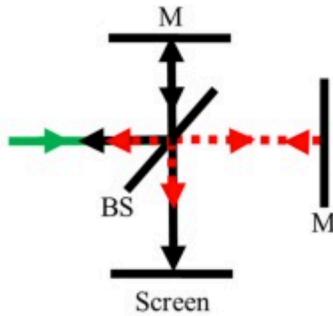


Figure 24.1: Michelson Interferometer

The interference pattern depends on the difference in the optical path lengths of the two beams. Constructive interference occurs when the path lengths are equal or differ by an integer multiple of the wavelength  $\lambda$ , given by the condition  $2d = m\lambda$ , where  $d$  is the path length difference and  $m$  is an integer. Destructive interference occurs when the path lengths differ by an odd multiple of half the wavelength, given by  $2d = (m + \frac{1}{2})\lambda$ .

By adjusting the position of the movable mirror, the path length difference changes, altering the interference pattern. This sensitivity to path length variations makes the Michelson interferometer useful for precise measurements, such as determining the wavelength of light, measuring small distances, and detecting changes in refractive index. It also played a crucial role in the Michelson-Morley experiment, which provided evidence against the existence of the luminiferous ether and supported the theory of special relativity.

### LIGO Interferometer Overview

The Laser Interferometer Gravitational-Wave Observatory (LIGO) is designed to detect gravitational waves—ripples in spacetime caused by massive accelerating objects, such as merging black holes or neutron stars. LIGO uses a Michelson interferometer configuration with two perpendicular arms, each several kilometers long. A laser

beam is split into two beams that travel down these arms, reflect off mirrors at the ends, and then recombine at the beam splitter. Under normal conditions, the lengths of the arms are such that the beams interfere destructively, resulting in no light reaching the detector.



Figure 24.2: Virgo detector of LIGO

When a gravitational wave passes through the interferometer, it causes a tiny but measurable change in the lengths of the arms. This change alters the interference pattern of the recombined beams, allowing the detection of the gravitational wave. The sensitivity of LIGO is such that it can detect changes in arm length smaller than a thousandth of the diameter of a proton.

The phase shift  $\Delta\phi$  caused by a gravitational wave can be expressed as:

$$\Delta\phi = \frac{4\pi\Delta L}{\lambda}$$

where  $\Delta L$  is the change in the length of the interferometer arms due to the gravitational wave, and  $\lambda$  is the wavelength of the laser light used in the interferometer.

### Derivation of the Phase Shift

To understand the phase shift in LIGO, consider the effect of a gravitational wave passing through the interferometer. The wave causes a differential change in the lengths of the two arms, denoted as  $\Delta L$ . This change in length affects the travel time of the laser beams in each arm.

The time difference  $\Delta T$  between the beams traveling in the two arms can be expressed as:

$$\Delta T = \frac{\Delta L}{c}$$

The phase shift  $\Delta\phi$  is then related to this time difference by the equation:

$$\Delta\phi = \frac{2\pi\Delta T}{T} = \frac{4\pi\Delta L}{\lambda}$$

This phase shift alters the interference pattern observed at the detector, allowing LIGO to measure the presence and properties of gravitational waves. The extraordinary precision of LIGO's measurements enables it to detect incredibly small disturbances in spacetime, providing valuable insights into some of the most energetic events in the universe.

### LIGO Technical Details

LIGO consists of two large interferometers located in the United States: one in Hanford, Washington, and the other in Livingston, Louisiana. These facilities are operated by the LIGO Scientific Collaboration (LSC), which includes scientists from various institutions around the world. Here is a link to the [LIGO website](#) and a direct link to an [overview document](#).

## Interferometer Design

### 1. Michelson Interferometer Configuration:

- LIGO uses a Michelson interferometer configuration with 4-kilometer-long arms.
- Each interferometer has two perpendicular arms, forming an “L” shape.
- A laser beam is split into two beams that travel down the arms, reflect off mirrors, and recombine at the beam splitter.

### 2. Fabry-Pérot Cavities:

- Each arm of the interferometer contains a Fabry-Pérot cavity to increase the effective path length of the laser beams.
- The cavities are formed by highly reflective mirrors placed at the ends of the arms.
- The laser beams bounce back and forth multiple times within the cavities, effectively increasing the arm length to several hundred kilometers.

### 3. Laser System:

- LIGO uses a high-power, stabilized laser operating at a wavelength of 1064 nm (infrared).
- The laser power is typically around 200 watts, but the effective power in the interferometer arms is increased to several kilowatts using power recycling techniques.

### 4. Suspension and Isolation:

- The mirrors and other optical components are suspended by a system of pendulums to isolate them from ground vibrations and other noise sources.
- The suspension system includes multiple stages of isolation, including active and passive damping mechanisms.

### 5. Vacuum System:

- The interferometer arms are housed in ultra-high vacuum tubes to eliminate air molecules that could scatter the laser beams and introduce noise.
- The vacuum system maintains a pressure of around  $10^{-9}$  torr.

## Detection Principle

### 1. Gravitational Waves:

- Gravitational waves are ripples in spacetime caused by accelerating massive objects, such as merging black holes or neutron stars.
- As a gravitational wave passes through the interferometer, it stretches and compresses the spacetime along the arms, causing tiny changes in the arm lengths.

### 2. Interference Pattern:

- The changes in arm lengths cause a phase shift in the laser beams when they recombine at the beam splitter.
- This phase shift results in a change in the interference pattern, which is detected by photodetectors.

### 3. Sensitivity:

- LIGO is designed to detect changes in arm lengths as small as  $10^{-19}$  meters, which is less than one-thousandth the diameter of a proton.
- The sensitivity is achieved through advanced noise reduction techniques, including seismic isolation, thermal noise reduction, and quantum noise reduction.

## Data Analysis

### 1. Signal Processing:

- The data from the photodetectors are processed to identify potential gravitational wave signals.
- Advanced algorithms and computational techniques are used to filter out noise and extract the signals.

### 2. Event Detection:

- When a potential gravitational wave event is detected, the data are analyzed to determine the properties of the source, such as the masses and spins of merging black holes or neutron stars.
- The detection is confirmed by comparing data from both LIGO detectors and, if available, data from other gravitational wave observatories like Virgo.

## Mach Zehnder Interferometer

The Mach-Zehnder interferometer is an optical device used to measure phase shifts between two light beams. It consists of a coherent light source, such as a laser, which emits a beam that is split into two paths by a beam splitter. Each beam travels along a different path, encountering mirrors that redirect them towards a second beam splitter where they are recombined. The recombined beams then interfere, producing an interference pattern that can be observed at the output ports.

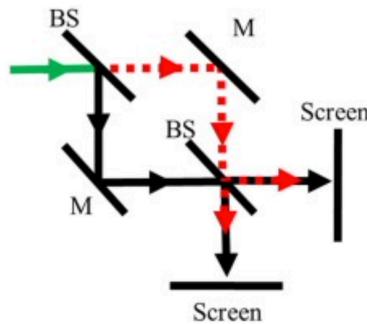


Figure 24.3: Mach Zehnder Interferometer

The interference pattern depends on the phase difference between the two beams, which is influenced by the optical path lengths they travel. If the path lengths are equal, constructive interference occurs, resulting in maximum intensity at one output port and minimum intensity at the other. If the path lengths differ, the phase difference causes a shift in the interference pattern, leading to varying intensities at the output ports.

The phase difference  $\Delta\phi$  between the two beams is given by  $\Delta\phi = \frac{2\pi\Delta L}{\lambda}$ , where  $\Delta L$  is the difference in path lengths and  $\lambda$  is the wavelength of the light. By introducing a phase shift in one of the paths, such as by changing the length of the path or altering the refractive index of the medium through which the beam travels, the interference pattern can be controlled and measured.

The Mach-Zehnder interferometer is widely used in applications requiring precise measurements of phase shifts, such as in optical communication systems, quantum mechanics experiments, and the study of optical properties



(a) Mach Zehnder Interferometer with a Gaussian

(a) Mach Zehnder Interferometer with a Bessel Beam

of materials.

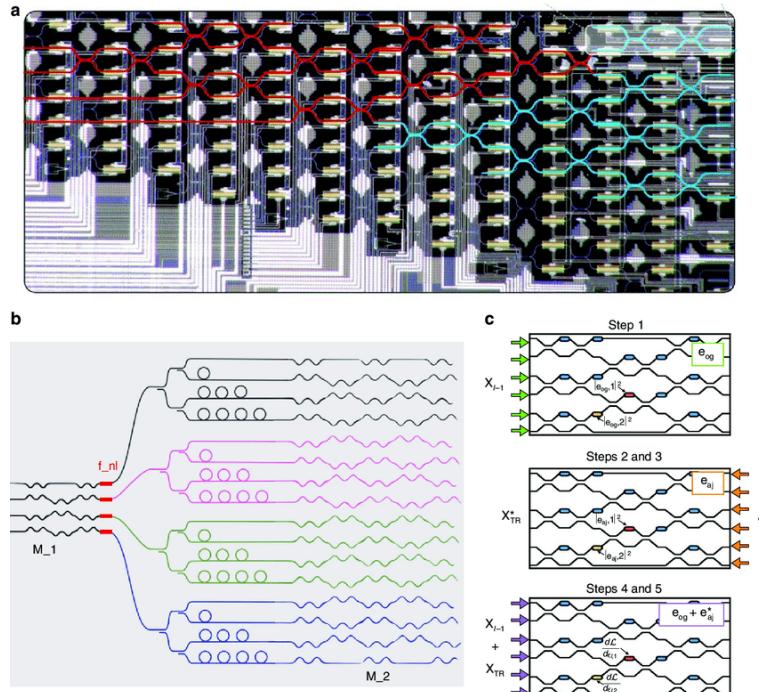


Figure 24.6: Network of Mach Zehnder Interferometers forming an all optical neural network. Zhou, H. et al. Photonic matrix multiplication lights up photonic accelerator and beyond. Light: Sci. Appl. 11, 30 (2022).

## Sagnac Interferometer Overview

A Sagnac interferometer operates by splitting a beam of light into two separate beams that travel in opposite directions around a closed loop. These beams are then recombined at the end of the loop, resulting in an interference pattern. If the interferometer is rotating, the path lengths of the two beams differ, leading to a phase shift.

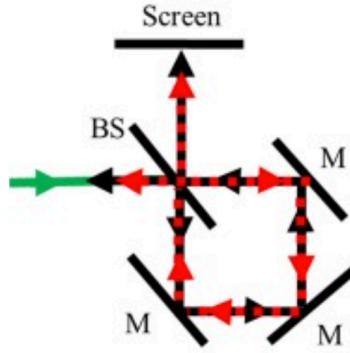
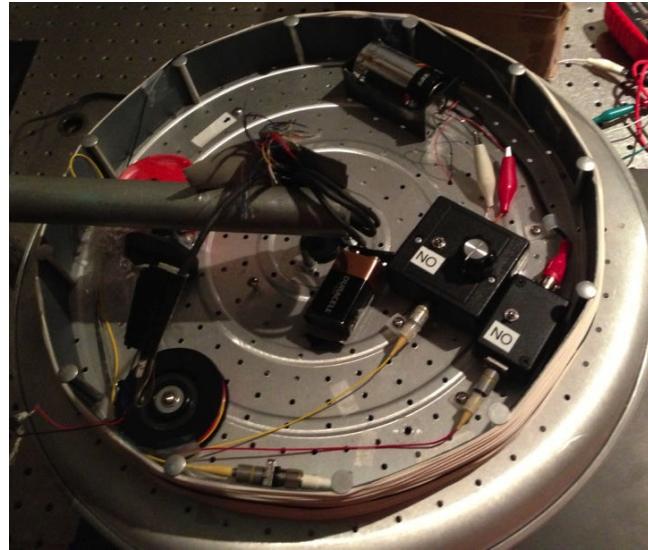


Figure 24.7: Sagnac Interferometer

The phase shift, denoted as  $\Delta\phi$ , can be calculated using the formula:

$$\Delta\phi = \frac{8\pi A\Omega}{\lambda c}$$

where  $A$  represents the area enclosed by the light path,  $\Omega$  is the angular velocity of the rotation,  $\lambda$  is the wavelength of the light, and  $c$  is the speed of light.

Figure 24.8: Sagnac interferometer made of a fibre rolled up on a cylinder taken from [here](#)

### Derivation of the Phase Shift

To derive the phase shift, we start by considering the path length difference. Assume a loop with a perimeter  $L$  and an area  $A$ . In the absence of rotation, the time taken for light to travel around the loop is given by  $T = \frac{L}{c}$ .

When the interferometer rotates with an angular velocity  $\Omega$ , the effective path length changes. For the beam traveling in the direction of rotation, the path length increases, while for the beam traveling opposite to the direction of rotation, the path length decreases.

The time difference  $\Delta T$  between the two beams can be expressed as:

$$\Delta T = \frac{4A\Omega}{c^2}$$

The phase shift  $\Delta\phi$  is related to this time difference by the equation:

$$\Delta\phi = \frac{2\pi\Delta T}{T} = \frac{8\pi A\Omega}{\lambda c}$$

### Derivation Details

To derive the formula for the time difference  $\Delta T$  between two counter-propagating beams in a Sagnac interferometer, we start by considering a loop of perimeter  $L$  and area  $A$ . The interferometer is rotating with an angular velocity  $\Omega$ . Light travels in opposite directions around the loop, creating two counter-propagating beams.

In a non-rotating frame, the time taken for light to travel around the loop is  $T = \frac{L}{c}$ . When the interferometer rotates with angular velocity  $\Omega$ , the effective path lengths for the two beams differ due to the rotation. For the beam traveling in the direction of rotation, the effective path length increases, while for the beam traveling opposite to the direction of rotation, the effective path length decreases.

The relative velocity of light with respect to the rotating frame is  $c \pm v$ , where  $v = \Omega R$  is the tangential velocity at the perimeter of the loop. For small angular velocities, we can approximate the effect using the area  $A$  and the angular velocity  $\Omega$ .

The time taken for the beam traveling in the direction of rotation is:

$$T_+ = \frac{L}{c + v}$$

and the time taken for the beam traveling opposite to the direction of rotation is:

$$T_- = \frac{L}{c - v}$$

For small  $v$ , we can use the binomial expansion to approximate the times:

$$T_+ \approx \frac{L}{c} \left(1 - \frac{v}{c}\right)$$

$$T_- \approx \frac{L}{c} \left(1 + \frac{v}{c}\right)$$

The time difference between the two beams is:

$$\begin{aligned} \Delta T &= T_+ - T_- = \frac{L}{c} \left(1 - \frac{v}{c}\right) - \frac{L}{c} \left(1 + \frac{v}{c}\right) \\ \Delta T &= \frac{L}{c} \left(-\frac{v}{c} - \frac{v}{c}\right) \\ \Delta T &= -\frac{2Lv}{c^2} \end{aligned}$$

The tangential velocity  $v$  is related to the angular velocity  $\Omega$  and the radius  $R$  of the loop:

$$v = \Omega R$$

The area  $A$  of the loop is related to the radius  $R$  and the perimeter  $L$ :

$$A = \pi R^2$$

Combining these, we get:

$$R = \frac{L}{2\pi}$$

$$v = \Omega \frac{L}{2\pi}$$

Substituting  $v = \Omega \frac{L}{2\pi}$  into the expression for  $\Delta T$ :

$$\Delta T = -\frac{2L(\Omega \frac{L}{2\pi})}{c^2}$$

$$\Delta T = -\frac{2L^2\Omega}{2\pi c^2}$$

$$\Delta T = -\frac{L^2\Omega}{\pi c^2}$$

Using the relationship  $L^2 = 4A$ , we get:

$$L^2 = 4A$$

Substituting  $L^2 = 4A$  into the expression for  $\Delta T$ :

$$\Delta T = -\frac{4A\Omega}{c^2}$$

Since the time difference  $\Delta T$  is typically considered in magnitude, we take the absolute value:

$$\Delta T = \frac{4A\Omega}{c^2}$$

## **Part III**

# **Electromagnetic Optics**

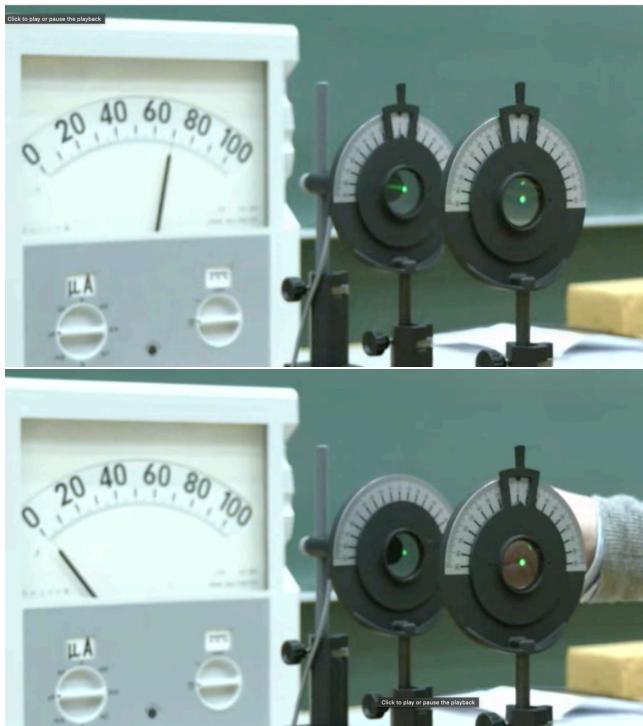


# Chapter 25

## Electromagnetic Optics

So far we have looked at the propagation of light in form of ray's and its description in Geometrical Optics. We made a number of assumptions that we formulated as postulates. We then extended this description by a scalar Wave Optics description to allow for a description of interference and diffraction, which can not be explained by Geometrical Optics. Yet concepts of refractive index and light matter interaction and the intensities are not covered by Wave Optics and also just postulates.

Electromagnetic Optics allows us to define these missing things. We discover light as electromagnetic waves consisting of electric and magnetic fields, which allow us to describe the interaction with charges in atoms, which is the foundation for the refractive index for example. The new thing is therefore the fact that we now need vectors for the description of light.

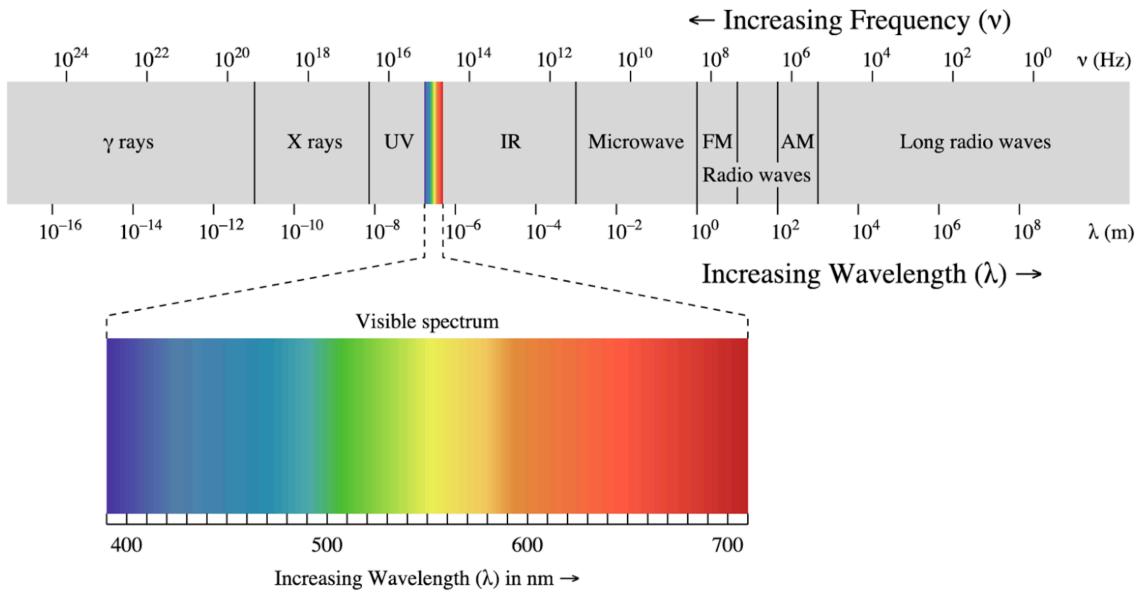


Demonstration of the vectorial nature of the electric field of electromagnetic waves with the help of a set of polarizers. (Left) The first polarizer on the right side of the image is transmitting only the vertical direction of the electric field. This vertical direction of the electric field is probed with the second polarizer. If pointing in the same direction, the intensity as measured with the meter behind is maximum. When probed with a second polarizer along the horizontal direction (Right), the detector shows zero intensity.

### 25.1 Electromagnetic Spectrum

While most of our considerations are focused on electromagnetic waves in the visible region, all of them can be generalized to other regions of the very broad electromagnetic spectrum. The electromagnetic theory is scale

free, meaning that the same effect on specific structures occur also at smaller scales if you scale the wavelength of the wave.



(a) Electromagnetic Spectrum

Electromagnetic wave spectrum with its specific regions.

## Maxwell Equations

To obtain a wave equation in terms of electric  $\vec{E}$  and magnetic fields  $\vec{B}$  we need Maxwell's equations. We will consider them in vacuum, i.e. for zero charge  $\rho = 0$  and current density. The electric and magnetic permeabilities of vacuum are given by  $\epsilon_0$  and  $\mu_0$ .

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad (\text{ME.1})$$

$$\nabla \cdot \vec{E} = 0 \quad (\text{ME.2})$$

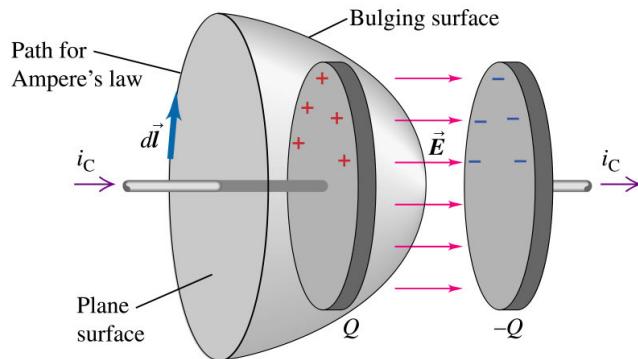
$$\nabla \times \vec{B} = \epsilon_0 \mu_0 \frac{\partial \vec{E}}{\partial t} \quad (\text{ME.3})$$

$$\nabla \cdot \vec{B} = 0 \quad (\text{ME.4})$$

Maxwell's third equation (ME.3) is significant because it expands upon Ampère's law, which states that magnetic field lines form closed loops around electric currents. This principle can be applied to a capacitor circuit, where we can calculate the magnetic field by integrating along a circular path around the current-carrying wire.

In this setup, the current density flows through a surface bounded by the circular path. Importantly, we can choose any surface that shares the same boundary circle (mathematically known as path-independent surface integration). However, this creates an apparent paradox: Ampère's law must work both for: 1. A simple surface through which the conduction current flows 2. A surface that passes through the capacitor gap where no actual charges flow

To resolve this inconsistency, Maxwell introduced the concept of displacement current. This additional current exists even in regions without flowing charges and is proportional to the time rate of change of the electric field, multiplied by the vacuum permittivity ( $\epsilon_0$ ).



Copyright © 2004 Pearson Education, Inc., publishing as Addison Wesley.

The displacement current makes physical sense because current only flows in the capacitor's wires when the electric field between the plates is changing. This addition to Ampère's law was crucial, as it completed the set of equations that describe electromagnetic waves.

This term is essential for deriving the electromagnetic wave equation and understanding how electromagnetic waves propagate through space.

## Deriving the Wave Equation

We will take the first (ME.1) of the four equations and apply another rotation  $\nabla \times$  to both sides

$$\nabla \times \nabla \times = -\nabla \times \frac{\partial \vec{B}}{\partial t}$$

We can exchange the time and spatial derivate on the right side as  $\nabla$  is not depending on time to get

$$\nabla \times \nabla \times \vec{E} = -\frac{\partial \nabla \times \vec{B}}{\partial t} \quad (25.1)$$

$$= -\epsilon_0 \mu_0 \frac{\partial^2 \vec{E}}{\partial t^2} \quad (25.2)$$

where we used the third equation (ME.3) to replace the rotation of the magnetic field. We now have to expand the left side with the identity

$$\nabla \times \nabla \times \vec{E} = \nabla(\nabla \cdot \vec{E}) - \nabla(\nabla \vec{E})$$

Note the the first term on the right side is the gradient of the divergence of  $\vec{E}$ , while the second term is the divergence of the gradient of  $\vec{E}$ . We know that in vacuum the divergence of the elecric field is zero (no sources of the electric field) and therefore  $\nabla \times \nabla \times \vec{E} = -\nabla(\nabla \vec{E})$  and we have our wave equation

$$\frac{\partial^2 \vec{E}}{\partial r^2} - \epsilon_0 \mu_0 \frac{\partial^2 \vec{E}}{\partial t^2} = 0 \quad (\text{Wave Equation})$$

### **i** Wave Equation

The wave equation for the propagation of electric fields in vaccum is given by

$$\frac{\partial^2 \vec{E}}{\partial r^2} - \epsilon_0 \mu_0 \frac{\partial^2 \vec{E}}{\partial t^2} = 0$$

The phase velocity of the wave is

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}} = 299792458 \frac{\text{m}}{\text{s}}$$

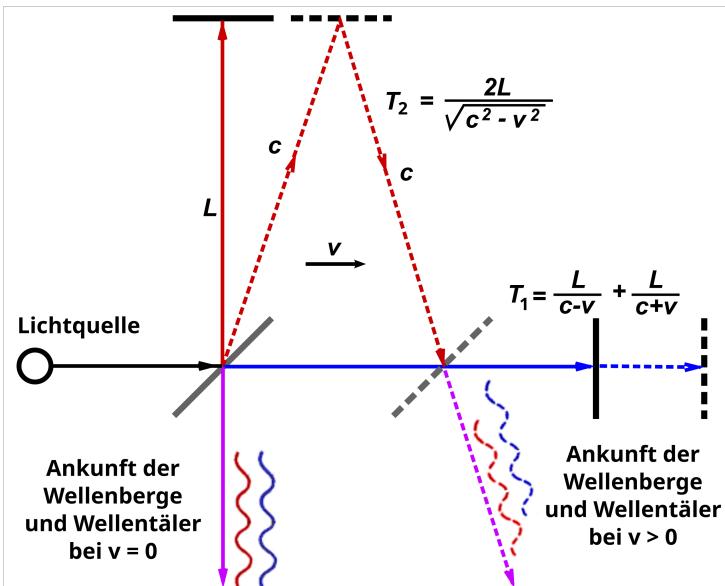
One of the interesting relations to electrostatics is now, that the static permeabilities  $\epsilon_0 \cdot \mu_0$  determine the speed of light. Note that the above wave equation is a vectorial equation., That means there is a wave equation for each component of the electric field, e.g.

$$\frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2 E_x}{\partial t^2} = 0$$

for the x-component of the electric field. Equivalent equations exist for the other field components.

The same mathematical treatment can be done for the magnetic field  $\vec{B}$  and the same wave equation will follow from that.

### Michelson Morley Experiment



The Michelson-Morley experiment of 1887 was designed to detect the hypothetical luminiferous ether through which light was thought to propagate. Using an interferometer, they split a light beam into two perpendicular paths and recombined them to create an interference pattern. The theoretical derivation considered the time for light to travel in both directions: along the direction of Earth's motion through the ether, the outward and return journey times are given by:

$$t_1 = \frac{L}{c-v}$$

$$t_2 = \frac{L}{c+v}$$

where L is the arm length, c is the speed of light, and v is Earth's velocity through the ether. The total time for this path is therefore:

$$T_1 = t_1 + t_2 = \frac{L}{c-v} + \frac{L}{c+v} = \frac{2Lc}{c^2 - v^2}$$

For the perpendicular arm, the time calculation involved the Pythagorean theorem, as light would travel diagonally relative to the ether, giving:

$$T_2 = \frac{2L}{\sqrt{c^2 - v^2}}$$

The time difference  $\Delta T = T_2 - T_1$ , when expanded using the binomial theorem and keeping terms to second order in  $v/c$ , yields:

$$\Delta T = \frac{L}{c} \times \frac{v^2}{c^2}$$

This time difference corresponds to a path difference of:

$$\Delta d = 2L \frac{v^2}{c^2}$$

However, Michelson and Morley observed no significant fringe shift, contradicting the ether theory and paving the way for special relativity, which established the constancy of the speed of light in all inertial reference frames.

## 25.2 Plane Waves, Spherical Waves

### Plane Waves

We will first have a look at elementary solutions of the wave equation again as we have done that in the wave optics sections. First of all we can write the solution of the wave equation as a product of a spatial and a temporal amplitude, i.e.

$$\vec{\mathcal{E}}(\vec{r}, t) = \mathcal{R}e\{\vec{E}(\vec{r})e^{-i\omega t}\} \quad (25.3)$$

$$\vec{\mathcal{B}}(\vec{r}, t) = \mathcal{R}e\{\vec{B}(\vec{r})e^{-i\omega t}\} \quad (25.4)$$

We use again the complex notation and remember that the measurable physical quantity  $\vec{\mathcal{E}}$  or  $\vec{\mathcal{B}}$  has to be a real valued quantity. Therefore we may calculate with the complex quantities, but finally need to calculate the real part ( $\mathcal{R}e$ ) if required. In the following we will use the complex notation throughout the calculation and only refer to the real value if this is really useful. When inserting the complex ansatz above into the wave equation, we can take the time derivative which yields for the electric field

$$\Delta \vec{E}(\vec{r}) + \frac{\omega^2}{c^2} \vec{E}(\vec{r}) = 0 \quad (\text{Helmholtz Equation})$$

The latter equation is known as the **Helmholtz equation**. It is the differential equation for the spatial amplitude of the wave. We may also insert the solutions into the first and the third Maxwell equation which results in

$$\nabla \times \vec{E} = i\omega \vec{B} \quad (25.5)$$

$$\nabla \times \vec{B} = -\epsilon_0 \mu_0 i\omega \vec{E} \quad (25.6)$$

We obtain finally a plane wave with our knowledge from the wave optics section.

$$\vec{E}(\vec{r}) = \vec{E}_0 e^{i\vec{k} \cdot \vec{r}} \quad (25.7)$$

$$\vec{B}(\vec{r}) = \vec{B}_0 e^{i\vec{k} \cdot \vec{r}} \quad (25.8)$$

$$(25.9)$$

Taking the rotation of those two equations yields

$$\vec{k} \times \vec{E}_0 = \omega \vec{B}_0 \quad (25.10)$$

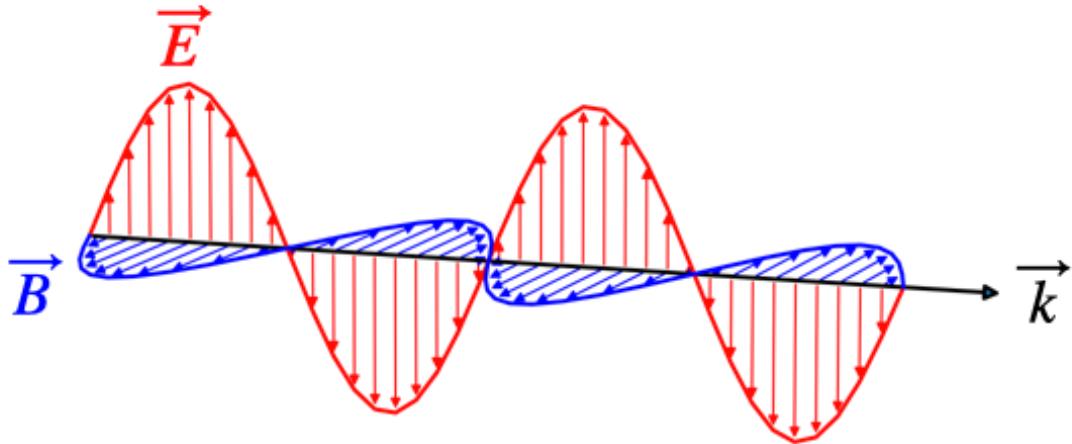
$$\vec{k} \times \vec{B}_0 = -\frac{\omega}{c^2} \vec{E}_0 \quad (25.11)$$

The latter two equations tell essentially two things. First of all they state that the vectors  $\vec{k}$ ,  $\vec{E}_0$  and  $\vec{B}_0$  stand perpendicular to each other. This is why electromagnetic waves are termed transverse waves. The physical quantity of a transverse wave change perpendicular to its propagation direction.

The second thing is that the amplitudes of the two waves are not independent of each other but rather

$$B_0 = \frac{1}{c} E_0$$

This is quite helpful, as we may just do calculations for the electric field and transfer them with the help of this conversion to the magnetic field.



(a) Plane Wave

Plane wave propagating along the y-direction, with the electric field oscillating along the z-direction.

## Spherical Waves

Spherical waves are more complex than plane waves and require a different mathematical approach. We can describe them using an auxiliary function called the **Vector potential**, defined as:

$$\vec{A}(\vec{r}) = A_0 U(\vec{r}) \hat{x}$$

where  $\hat{x}$  is the unit vector in the x-direction, and  $U(\vec{r})$  is the scalar spherical wave function from wave optics:

$$U(\vec{r}) = \frac{1}{r} e^{-ikr}$$

This vector potential satisfies the Helmholtz equation:

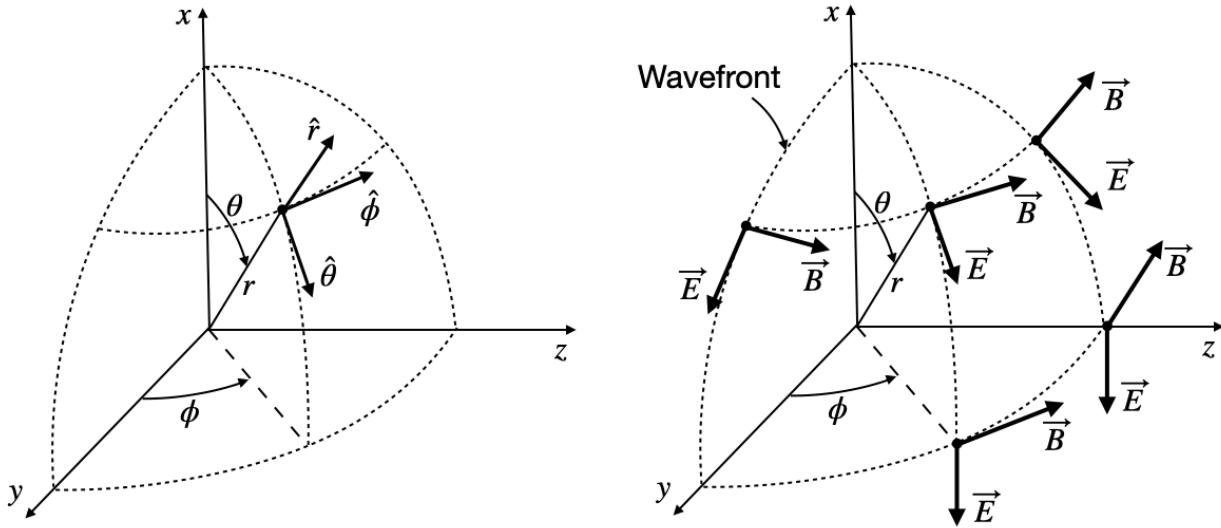
$$\Delta \vec{A} + k^2 \vec{A} = 0$$

When we solve for the electric and magnetic fields in spherical coordinates, and consider large distances where  $r \gg \lambda$  or  $kr \gg 2\pi$ , we get:

$$\vec{E}(\vec{r}) = E_0 \sin(\theta) U(\vec{r}) \hat{\theta} \quad (25.12)$$

$$\vec{B}(\vec{r}) = B_0 \sin(\theta) U(\vec{r}) \hat{\phi} \quad (25.13)$$

These equations reveal that both the electric and magnetic fields lie tangent to the spherical wavefront. Since the wave propagates radially while the fields are perpendicular to this direction, we can classify spherical waves as transverse electromagnetic waves, just like plane waves.



(Left) Definition of the unit vectors in a spherical coordinate system. (Right) Vectors of the electric and magnetic field for a spherical wave.



# Chapter 26

## Polarization of EM Waves

The vectorial nature of electric and magnetic fields are the new property that we inserted into our description of wave propagation. Electric and magnetic fields, while propagating along a certain direction, have a specific direction in the lab frame, which can change over time in specific ways. This stationary state of the direction of the electric field vector is termed **polarization**. You may also use the magnetic field to define polarization, but commonly this is done using the electric field direction.

### Polarization of Electromagnetic Waves

The polarization of an electromagnetic wave is defined by the direction of its electric field vector in our laboratory frame.

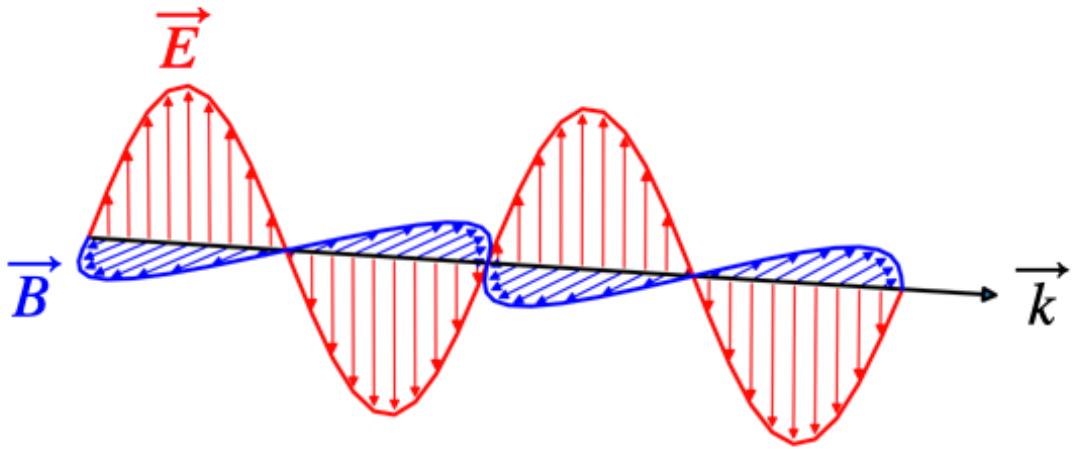
We differentiate between different states of polarization, e.g. linear, circular, elliptical polarization. But we may, depending on the properties of the light source also have unpolarized light.

The polarization of light is important for many applications including for example 3D cinema or TV. The polarization state is commonly changed when light interacts with matter, therefore polarization is also a very important tool to characterize materials. For example, the technique of **ellipsometry** is studying the polarization state of light reflected from a material thereby gaining important information about the electronic properties of the material. This is an important tool in solid-state physics. The polarization of light is also important for interference as two waves of orthogonal directions of the electric field cannot interfere (please check this idea yourself). It can be also used to encode information for quantum cryptography.

In the following sections we will shortly define these different polarization states.

### 26.1 Linearly Polarized Waves

Light is called linearly polarized if the electric field vector oscillates in a single plane during light propagation. The picture below, which we introduced earlier for a plane wave shows such a linearly polarized wave.



We can generalize this description a bit more. Our plane wave shall be given by

$$\vec{E} = \vec{E}_0 e^{i(\omega t - kz)}$$

The wave propagates along the z-direction. The polarization is given by the vector  $\vec{E}_0$ . This vector can be split into its components along the x- and the y-direction

$$\vec{E}_0 = E_{0x} \hat{e}_x + E_{0y} \hat{e}_y$$

For a linearly polarized electromagnetic wave, the magnitude of  $E_{0x}$  and  $E_{0y}$  are fixed over time, such that the angle of the electric field vector with the x-axis, for example, is fixed in time. This also requires that the phase of the total electric field components is the same, i.e.

$$E_x = E_{0x} e^{i(\omega t - kz)} \quad (26.1)$$

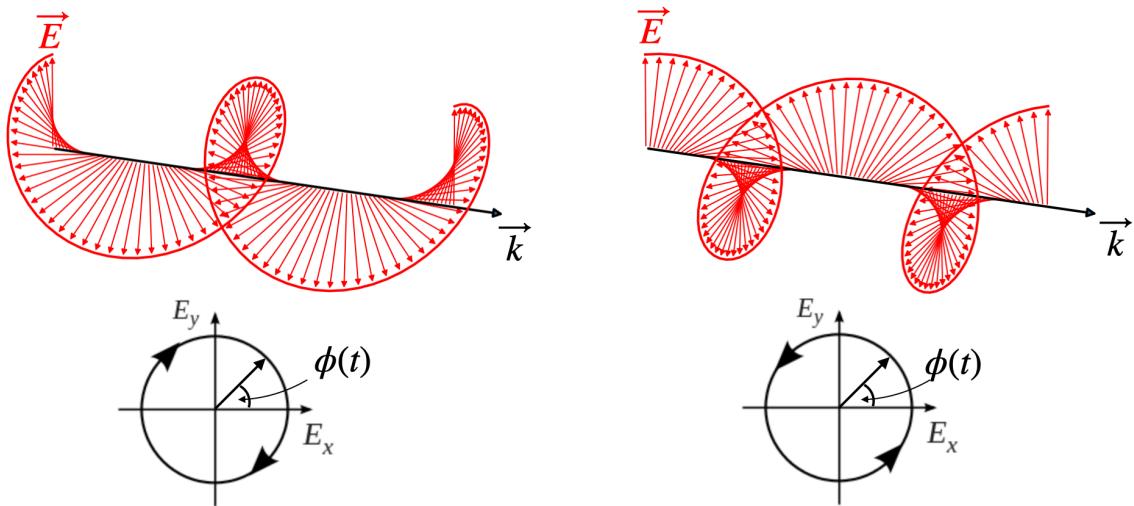
$$E_y = E_{0y} e^{i(\omega t - kz)} \quad (26.2)$$

$$(26.3)$$

The linear polarization state, independent of the polarization direction is also called  **$\pi$ -polarization**. We will use this term later in the description of light-atom interaction.

## 26.2 Circularly Polarized Waves

In circular polarized light, the end of the electric field vector describe a circle around the propagation direction. This means the electric field vector is rotating around the propagation direction.



To obtain circularly polarized light we can consider our two polarization components again but with an additional phase shift of  $\pm\pi/2$  of one of the components (here the y-component) together with constraining the amplitude of the components to the same value.

$$E_x = E_{0x} e^{i(\omega t - kz)} \quad (26.4)$$

$$E_y = E_{0y} e^{i(\omega t - kz \pm \frac{\pi}{2})} \quad (26.5)$$

$$E_{0y} = E_{0x} \quad (26.6)$$

The consequence of that additional phase shift of one component is, that  $E_y$  reaches its maximum amplitude at a different position  $z$ , when keeping the time fixed, i.e. making a snapshot. If  $z$  is fixed, the above formulas describe a rotation of the electric field around the direction of  $\vec{k}$  (here z-direction). The third equation ensures that during rotation the magnitude of the electric field vector is unchanged and thus the end of the electric field vectors describes a circle (bottom row of figure) around the direction of propagation.

Right circularly polarized light is also known as  **$\sigma^+$  polarization**, while left circularly polarized light is  **$\sigma^-$  polarization**.

## 26.3 Elliptically Polarized

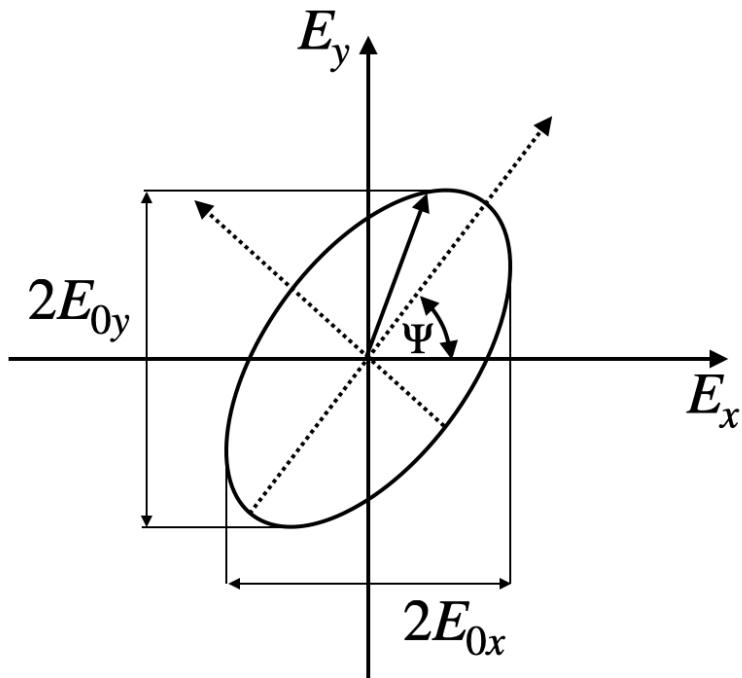
If both components of the electric field carry an additional phase delay  $\delta_1$  and  $\delta_2$  and both component amplitudes are different  $E_{0x} \neq E_{0y}$ , then we find elliptically polarized light

$$E_x = E_{0x} e^{i(\omega t - kz + \delta_1)} \quad (26.7)$$

$$E_y = E_{0y} e^{i(\omega t - kz + \delta_2)} \quad (26.8)$$

$$E_{0y} \neq E_{0x} \quad (26.9)$$

which means that the end point of the electric field vector describes an ellipse around the propagation direction.



The ellipse is then described by

$$\frac{E_x(z, t)^2}{E_{0x}^2} + \frac{E_y(z, t)^2}{E_{0y}^2} - \frac{2E_x(z, t)E_y(z, t)}{E_{0x}E_{0y}} \cos(\delta) = \sin^2(\delta)$$

where  $\delta = \delta_2 - \delta_1$ . The ellipse is in general not oriented along a specific axis but has an angle  $\Psi$  with the x-axis, which is fixed in time. This angle can be calculated by

$$\tan(2\Psi) = \frac{2E_{0x}E_{0y}}{E_{0x}^2 - E_{0y}^2} \cos(\delta)$$

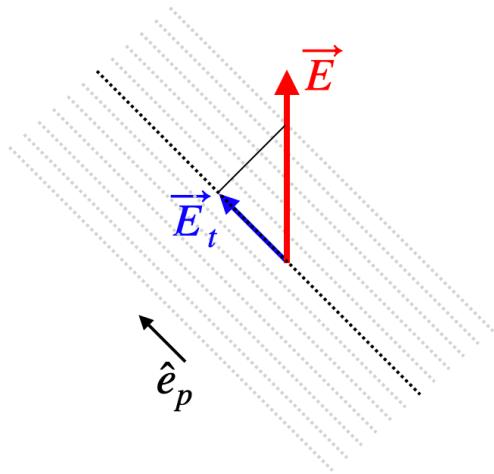
The circular polarized state is a special case of the elliptically polarized state of light.

## 26.4 Unpolarized Light

Unpolarized light is obtained when the electric field vector fluctuates in polarization statistically.

## 26.5 Analyzing Polarization

The polarization state of light is analyzed with the help of polarizers. These can be thin polymeric films, where the order of the polymer chain in the film provides a preferential axis or polarizers can also be made of other birefringent materials like Mica, for example. Those materials are called anisotropic as they have a refractive index, which depends on the direction of the electric field in the material. Depending on the design of the polarizer, you may generate linearly polarized light, you may change the polarization direction or even create circularly polarized light. We will cover anisotropic materials in a later lecture.



In mathematical terms, a *linear polarizer* projects the electric field vector to a specific direction. If  $\vec{E}(z, t)$  is our electric field vector of the electromagnetic wave and  $\hat{e}_p$  is the unit vector in the polarization direction the resulting electric field after the light has passed through the polarizer is given by

$$\vec{E}_t(z, t) = (\vec{E}(z, t) \cdot \hat{e}_p)\hat{e}_p = E(z, t) \cos(\theta)\hat{e}_p$$

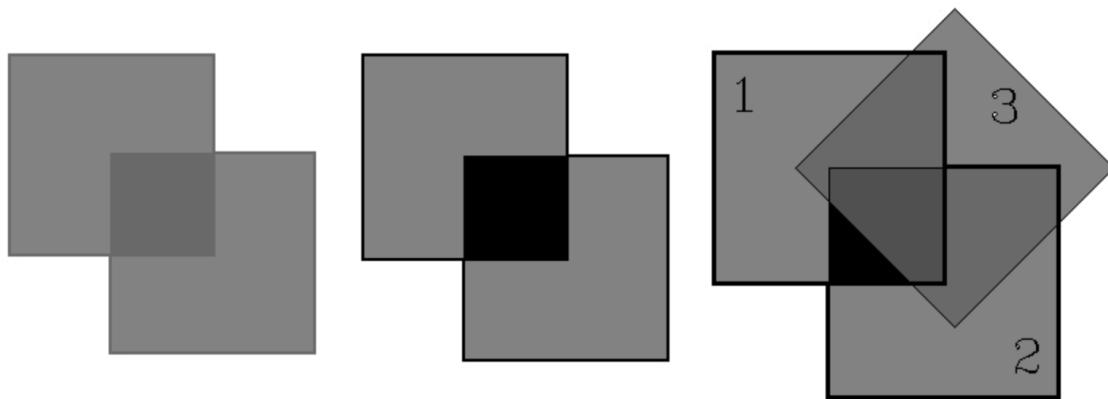
where  $\theta$  is the angle between the electric field vector and the polarizers orientation. This means the intensity of a linearly polarized wave transmitted through a linear polarizer only disappears if polarizer and electric field are orthogonal ( $\theta = 90^\circ$ )

### Law of Malus

The above description is also the essence of the **Law of Malus**, which concerns the transmission intensity. Taking the magnitude square of the left and the right side of the above equation yields

$$I_t = I_0 \cos^2(\theta)$$

It follows that the intensity drops to 50 %, when we adjust an angle of  $45^\circ$ . As soon as we have a lecture in the lecture hall again, we will have a look at some nice demonstrations that illustrate Malus's law similar to the picture below.



Two polarizers  
Parallel orientation

Two polarizers  
Orthogonal orientation

Three polarizers  
1 and 2 orthogonal  
3 between 1 and 2



## Chapter 27

# Energy Transport and Momentum Transport of EM Waves

### 27.1 Energy Transport

A characteristic property of waves is that they transport energy rather than mass. It is this energy which we typically detect with the help of a photodetector and not the electric field. We therefore need a quantity that describes this energy transport. From electrostatics, we know that the energy density of the electric field is given by  $w_e = \frac{1}{2}\epsilon_0\vec{E} \cdot \vec{E}$ . This was the energy needed to assemble a collection of charges at a place. The energy density of the magnetic field is given by  $w_m = \frac{1}{\mu_0}\vec{B} \cdot \vec{B}$ .

To obtain such a quantity we recall the energy density of the electric and magnetic field which in sum give

$$w = w_e + w_m = \frac{1}{2}\epsilon_0\vec{E} \cdot \vec{E} + \frac{1}{\mu_0}\vec{B} \cdot \vec{B} \quad (27.1)$$

Using the relation between the electric and the magnetic field amplitude we can further simplify the above expression for electromagnetic waves to  $w = \epsilon_0 E^2$ , but we will keep the full expression in the following.

Let's assume that we look at a volume  $V$  in which we have an electromagnetic wave, then the change in the energy inside that volume with time is given by

$$\frac{dW}{dt} = \int \frac{dw}{dt} dV$$

the integral over the time derivative of the energy density  $w$ . Applying energy conservation, the loss of energy in our volume must be due to some energy flow out of the volume that is described by a quantity  $\vec{S}$  that is describing that energy current density. If we consider a spherical surface without loss of generality, then the integrated energy flux through the surface should amount for the loss of energy in the volume, i.e.

$$\int \frac{dw}{dt} dV = - \oint_A \vec{S} \cdot \vec{n} da$$

where  $\vec{n}$  is a unit vector normal to the outside of the surface. The integral on the right side is running over the whole surface of our volume. Note that the minus sign on the right side occurs due to the fact that the volume is loosing energy. We may now apply Gauss' theorem and convert the surface integral on the right side by a volume integral over the divergence of the energy current density  $\vec{S}$ , i.e.

$$\oint_A \vec{S} \cdot \vec{n} da = \int \nabla \cdot \vec{S} dV = - \int \frac{dw}{dt} dV$$

We thus obtain the continuity equation

$$\frac{dw}{dt} = -\nabla \cdot \vec{S}$$

under vacuum conditions, meaning that we have no free charges and no charge current density.

We may use this equation to calculate the energy current density  $\vec{S}$  by the time derivative of our energy density of the electromagnetic wave.

$$\frac{\partial w}{\partial t} = \epsilon_0 \vec{E} \frac{\partial \vec{E}}{\partial t} + \frac{1}{\mu_0} \vec{B} \frac{\partial \vec{B}}{\partial t} \quad (\text{ED})$$

We may simplify that expression with the help of Maxwell's equations

$$\frac{\partial \vec{B}}{\partial t} = -\nabla \times \vec{E} \quad (27.2)$$

$$\frac{\partial \vec{E}}{\partial t} = \frac{1}{\mu_0 \epsilon_0} \nabla \times \vec{B} \quad (27.3)$$

and the  $\vec{H} = \vec{B}/\mu_0$  relating the magnetic field  $\vec{H}$  to the magnetic flux density  $\vec{B}$ . Inserting this in eq. ??, we find

$$\frac{dw}{dt} = \vec{E}(\nabla \times \vec{H}) - \vec{H}(\nabla \times \vec{E})$$

Using the vector identity

$$\nabla \cdot (\vec{H} \times \vec{E}) = \vec{E}(\nabla \times \vec{H}) - \vec{H}(\nabla \times \vec{E})$$

we find

$$\frac{dw}{dt} = \nabla(\vec{H} \times \vec{E}) = -\nabla(\vec{E} \times \vec{H}) = -\nabla \cdot \vec{S}$$

which is valid for vacuum and our energy current density can be identified as

$$\vec{S} = \vec{E} \times \vec{H} \quad (\text{Poynting Vector})$$

The vector  $\vec{S}$  is called the **Poynting vector** and describes the energy transport of an electromagnetic wave. The equation  $\frac{dw}{dt} + \nabla \cdot \vec{S} = 0$  can be also generalized to a situation where charge current densities  $\vec{j}$  are present. In this case we have to include in our consideration the work done by the electric field on the charge current density, which is  $\vec{E} \cdot \vec{j}$

$$\frac{dw}{dt} + \nabla \cdot \vec{S} = \vec{E} \cdot \vec{j} \quad (\text{Poynting theorem})$$

This equation is known as Poynting theorem and just a way of writing energy conservation.

Let's have a closer look at the Poynting vector, which we can write with the help of the magnetic flux density  $\vec{B}$  as

$$\vec{S} = \epsilon_0 c^2 (\vec{E} \times \vec{B})$$

The magnitude of the Poynting vector is then given by

$$S = |\vec{S}| = \epsilon_0 c^2 |\vec{E}| |\vec{B}| \quad (27.4)$$

$$= \epsilon_0 c E^2 = I \quad (27.5)$$

which is the same as the intensity. The magnitude of the Poynting vector describes the intensity of an electromagnetic wave or the energy flow through an area. It therefore has the unit of an intensity, which is  $W/m^2$ .

If we now have a plane wave

$$\vec{E} = \vec{E}_0 \cos(\omega t - \vec{k} \cdot \vec{r})$$

in the real value description, then its intensity is

$$I = I_0 \cos^2(\omega t)$$

at a position  $\vec{r} = 0$  with  $I_0 = c\epsilon_0 E_0^2$ . As we commonly calculate the intensity as the time average over one cycle of oscillation, we find for this wave the intensity

$$\langle I \rangle = I_0 \langle \cos^2(\omega t) \rangle = \frac{1}{2} I_0$$

This is the intensity we would record with the help of a detector and the flow of energy is set by the direction of the Poynting vector. This is not to be confused with the flow of the wavefronts, which go in the direction of the wavevector.

An example where this happens are birefringent materials, where the wavefronts and the Poynting vector are not parallel to each other. In this case the energy flow is not in the direction of the wavefronts. This is a very important concept in optics, as it is the energy flow that is important for the heating of materials and not the direction of the wavefronts.

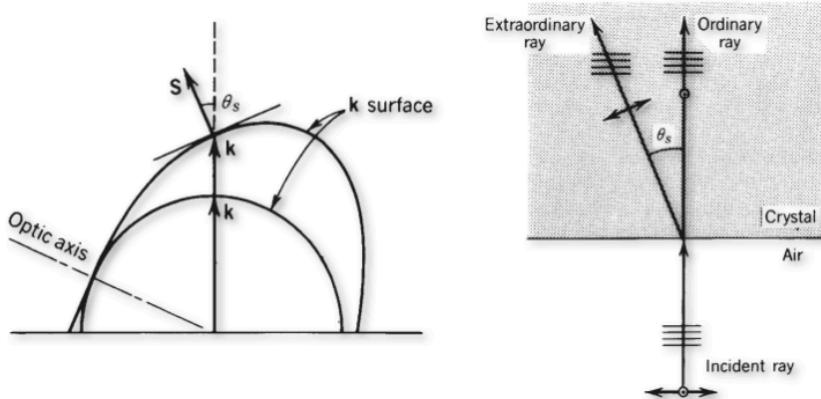


Figure 27.1: Birefringence teaser. The wavefronts of the two waves are not parallel to the Poynting vector, which is the direction of the energy flow.

## 27.2 Momentum Transport and Radiation Pressure

Like the energy of a wave that can be transferred to objects like photodetectors, electromagnetic waves also transport momentum, which can be turned into a motion of objects, when waves collide with massive objects.

The momentum is a property of the electromagnetic wave. We would like to describe that so-called **radiation pressure** (the flow of momentum through an area) in a very simple way.

For this purpose we need the relativistic energy

$$W^2 = p^2 c^2 + m^2 c^4$$

which can be calculated from the momentum  $p$ , the speed of light  $c$  and the mass  $m$ . As lightwaves propagate with the speed of light, they cannot have a mass and the second term is zero. The momentum of an electromagnetic wave is thus given by its energy divided by the speed of light

$$p = \frac{W}{c}$$

Therefore the momentum density in a volume must be also equal to the energy density divided by the speed of light.

$$\frac{dp}{dV} = \frac{1}{c} w = \frac{\epsilon_0 E^2}{c} = \frac{S}{c^2}$$

Therefore the momentum density  $dp/dV$  is directly related to the magnitude of the Poynting vector and the intensity.

The momentum that is therefore transported through an area  $A$  in a time  $dt$  by electromagnetic radiation is given by

$$\frac{dp}{Acdt} = \frac{S}{c^2}$$

since the volume from which the momentum comes is  $dv = Acdt$ . Consequently, the momentum current density (momentum per time and area) is after a slight transformation given by

$$\frac{1}{A} \frac{dp}{dt} = \frac{S}{c}$$

As  $dp/dt$  can be identified as a force the left side corresponds to a force divided by an area and thus the radiation pressure we are looking for.

$$p_{\text{rad}} = \frac{S}{c} \quad (\text{radiation pressure})$$

So far, this is a hypothetical radiation pressure, which we relate to the flow of momentum. It becomes a real pressure, if the radiation interacts with some surface.

If we consider a perfectly absorbing surface of area  $A$ , then the momentum of the electromagnetic wave is completely transferred to the surface and

$$p_{\text{rad}} = \frac{S}{c} \quad (\text{perfect absorption})$$

is the radiation pressure for perfect absorption.

If we have, however, a perfectly reflecting surface, we transfer due to the reflection, twice the momentum to the wall and therefore the radiation pressure for perfect reflection is

$$p_{\text{rad}} = 2 \frac{S}{c} \quad (\text{perfect reflection})$$

Thus if you want to measure radiation pressure, its best to use reflecting surfaces. This has been done for the first time in an experiment by Nichols and Hull in the years from 1900-1903.

### Radiation Pressure measurements



(a) Original apparatus

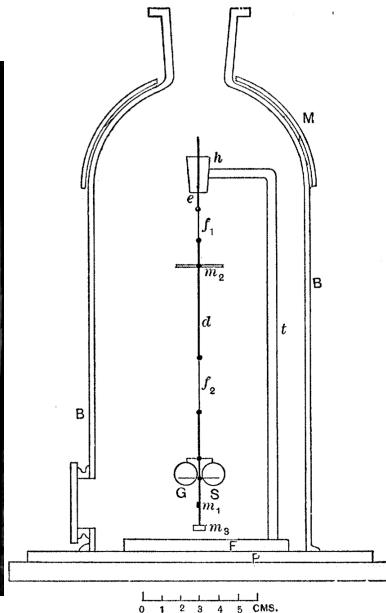


Fig. 2.

(b) Schematic setup

Figure 27.2: Experiment by Nichols and Hull at Dartmouth college to measure radiation pressure of light being reflected by two mirrors.

Nichols and Hull mounted therefore the two mirrors on a torsion spring in a geometry much like the Cavendish experiment. The light falling on both mirrors creates tiny forces which elongate the torsionspring which is mounted in a vessel where air is removed. From the elongation one may determine the forces and thus the pressure. From a separate measurement of the absorption and conversion of light into heat (with a Boulometer), one can determine the intensity or energy contained in the radiation.

### Comet Tails

Radiation pressure is also of importance in astronomy and in particular visible in the tail of comets. Comets show tails composed of light gas ions and dust particles which separate under the influence of radiation pressure. While orbiting around a star, the radiation pressure pushes the light gas ions radially away from the star, while the dust particles follow a curved shape bent towards the orbit due to their larger mass. This is also visible in the photograph we took in July 2020 for the comet Neowise which passed earth in a very spectacular way.

### Optical Tweezers and Magneto-optical Traps

Radiation pressure has emerged as a crucial tool for manipulating both microscopic particles and atomic species. Figure ?? (left) illustrates optical tweezers, where a tightly focused laser beam traps colloidal particles. While radiation pressure tends to push particles along the beam direction, additional gradient forces arising from the intense light field's spatial variation maintain the particle's position in the focal region. This technique has become invaluable in biophysics for:

- Measuring piconewton forces generated by molecular motors
- Studying protein folding mechanisms
- Investigating enzymatic processes such as CRISPR/Cas gene editing



Figure 27.3: Ion and dust tail of the comet Neowise photographed in 07/2020 by B.Cichos.

but also for fundamental physics

- like the measurement of the Casimir force
- understanding the interaction of light with matter
- and the measurement of the radiation pressure itself
- or even measuring the Maxwell Boltzmann distribution of single particles

At the atomic scale, radiation pressure combined with magnetic fields enables the trapping and cooling of atoms to temperatures of a few millikelvin in Magneto-Optical Traps (MOTs). Through additional cooling mechanisms, these atomic gases can reach even lower temperatures where they transition into a quantum state known as a Bose-Einstein condensate. This process relies on precise control of atomic hyperfine transitions.

These atomic trapping techniques form the foundation of modern atomic clocks, essential for GPS navigation and gravitational wave detection. Figure ?? shows a fountain atomic clock design, where laser-cooled atoms are propelled upward through a microwave cavity by radiation pressure. Here, the atoms' hyperfine energy levels interact with the microwave field.

In their fountain-like trajectory, atoms pass through the microwave cavity twice, enabling precise measurements of atomic transition frequencies. These measurements establish a fundamental time reference with unprecedented accuracy, used to synchronize timekeeping systems worldwide.

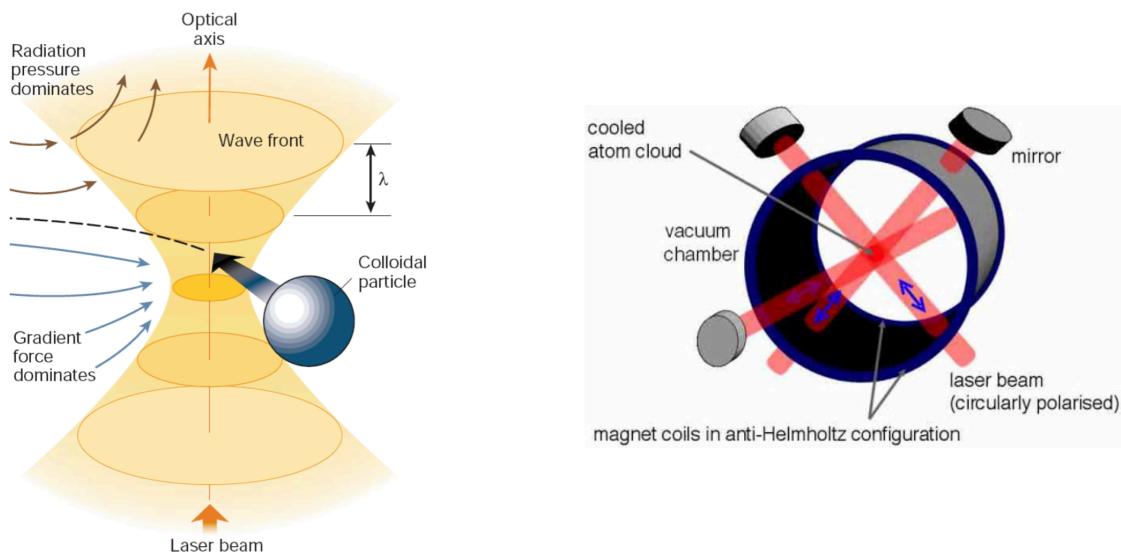


Figure 27.4: Principle sketch for optical tweezers (left, taken from a publication by D. Grier) and a sketch of a magneto-optical trap (MOT, right) as used for cooling and trapping of atomic gases for atomic clocks or Bose Einstein condensation.

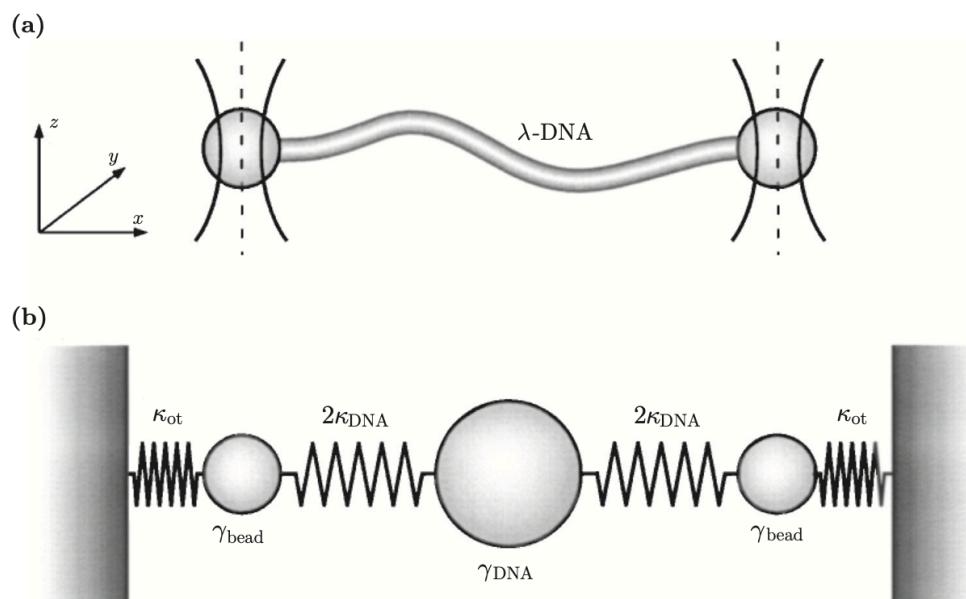


Figure 27.5

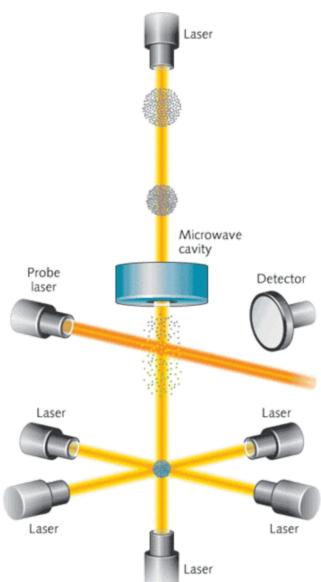


Figure 27.6: Schematic diagram of an atomic fountain clock, showing the path of laser-cooled atoms through a microwave cavity for precise frequency measurements.

# Chapter 28

# Electromagnetic Waves in Matter

The behavior of electromagnetic waves changes dramatically when they propagate through materials rather than vacuum. This interaction between light and matter underlies numerous phenomena in our daily lives, from the colors we see to modern optical technologies. Understanding these interactions requires bridging two perspectives: the microscopic view of how individual atoms respond to electromagnetic fields, and the macroscopic description of wave propagation through bulk materials.

## 28.1 Microscopic Picture of Dielectric Response

### Single Atom Response

To understand how materials respond to electromagnetic waves, let's first examine how individual atoms become polarized in an electric field. Our model atom consists of:

1. An electron charge  $-q$  uniformly distributed in a spherical cloud of radius  $a$
2. A positive nucleus at the center of this negative charge distribution

The charge density of the electron cloud is:

$$\rho = \frac{-q}{\frac{4}{3}\pi a^3} = -\frac{3q}{4\pi a^3}$$

This configuration creates a radial electric field inside the cloud:

$$\vec{E}(r) = -\frac{1}{4\pi\epsilon_0} \frac{q}{a^3} \vec{r}$$

where  $\vec{r}$  is the position vector from the center. The linear dependence on  $\vec{r}$  creates a restoring force similar to a spring.

### Atomic Polarizability

When we apply an external electric field  $\vec{E}_{\text{ex}}$ , the nucleus is displaced by a distance  $\vec{d}$  until forces balance:

$$q\vec{E}_{\text{ex}} - \frac{1}{4\pi\epsilon_0} \frac{q^2\vec{d}}{a^3} = \vec{0}$$

This force balance between the external field (pushing the nucleus) and the internal field (trying to restore the nucleus to the center) determines the atomic polarization, leading to the displacement:

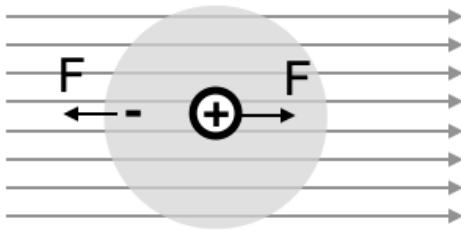


Figure 28.1: Polarization of an electric cloud of an atom: An external electric field displaces the positive nucleus relative to the negative electron cloud, creating an induced dipole moment.

$$\vec{d} = 4\pi\epsilon_0 a^3 \frac{\vec{E}_{\text{ex}}}{q}$$

This displacement creates a dipole moment:

$$\vec{p} = q\vec{d} = 4\pi\epsilon_0 a^3 \vec{E}_{\text{ex}}$$

The dipole moment increases linearly with the external field, with the proportionality constant:

$$\alpha = 4\pi\epsilon_0 a^3$$

known as the electronic polarizability. Note that this polarizability scales with atomic volume.

### From Single Atoms to Collective Response

When many atoms in a material are exposed to an electric field, their individual dipole moments combine to create a macroscopic polarization. The collective behavior can be characterized by the polarization density:

$$\vec{P} = N\vec{p}$$

where  $N$  is the number of dipoles per unit volume and  $\vec{p}$  is the individual atomic dipole moment. However, this simple picture of independent atoms needs modification for dense materials where atoms are closely packed and interact with each other.

In real materials, each atom experiences not only the external field but also the fields from neighboring dipoles. This leads us to consider the local field effects and the transition from microscopic to macroscopic descriptions, which we'll explore in the next section.

## 28.2 Macroscopic Description of Dielectric Materials

### Polarization Density and Surface Charges

The transition from microscopic to macroscopic behavior becomes apparent when we examine a sample of polarized material. For a cylindrical sample with cross-sectional area  $A$  and height  $s$ , the microscopic dipoles create observable macroscopic effects:

This alignment leads to:

- Total dipole moment:  $\vec{p}_{\text{cyl}} = As\vec{P}$
- End charges:  $q_{\text{end}} = \frac{p_{\text{cyl}}}{s} = AP$
- Surface charge density:  $\sigma = \frac{q_{\text{end}}}{A} = P$

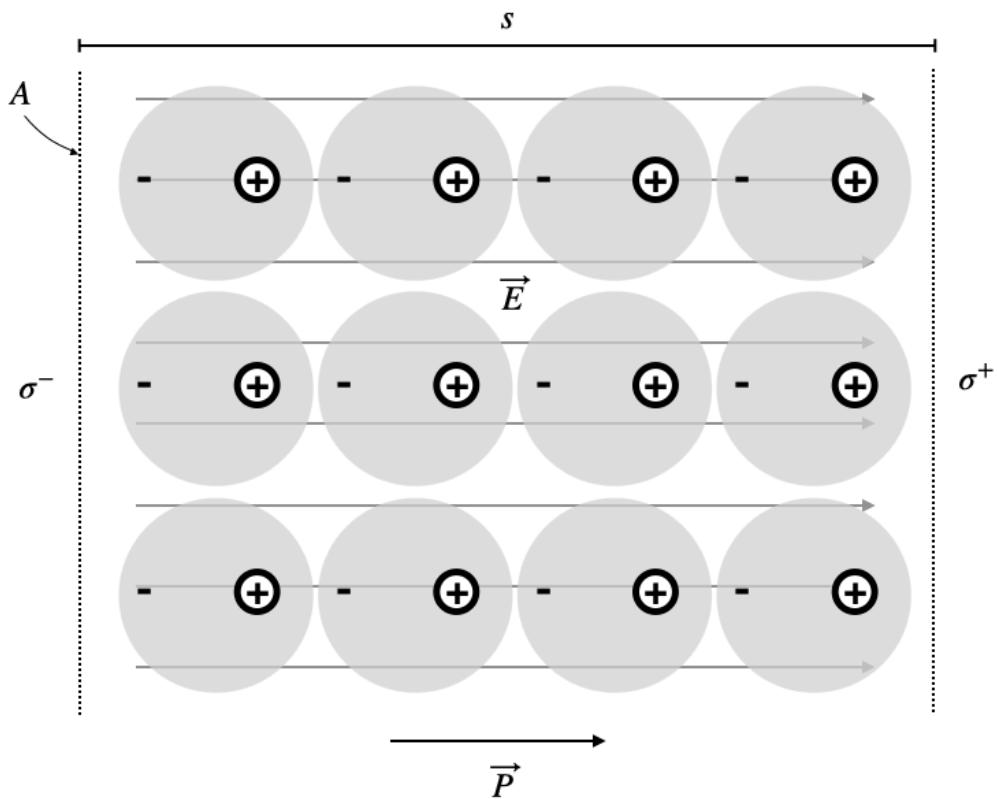


Figure 28.2: Polarization in a dielectric material showing aligned atomic dipoles and resulting surface charges.

More generally, the surface charge density at any boundary is:

$$\sigma_b = \vec{P} \cdot \hat{n}$$

where  $\hat{n}$  is the surface normal. These surface charges are “bound” to atoms, distinct from free charges.

### Volume Charge Density

When polarization isn’t uniform throughout a material, we need to consider both surface and volume charges. All bound charges in the material must sum to zero since the material was initially neutral:

$$\int \sigma_b dA + \int \rho_b dv = 0$$

Using the surface charge relation  $\sigma_b = \vec{P} \cdot \hat{n}$ , we can write:

$$\int \sigma_b dA = \int \vec{P} \cdot \hat{n} dA = \int \vec{P} \cdot d\vec{A}$$

Applying Gauss’s theorem:

$$\int \vec{P} \cdot d\vec{A} = \int (\nabla \cdot \vec{P}) dv$$

This leads to the local relationship:

$$\rho_b = -\nabla \cdot \vec{P}$$

This fundamental equation reveals that bound volume charges appear wherever the polarization has a divergence - that is, wherever it changes spatially in a way that creates an imbalance of positive and negative charges.

## Electric Displacement Field

In materials, the electric field's divergence must account for both free charges and bound charges:

$$\nabla \cdot \vec{E} = \frac{\rho_f + \rho_b}{\epsilon_0}$$

Using  $\rho_b = -\nabla \cdot \vec{P}$ :

$$\nabla \cdot \vec{E} = \frac{\rho_f - \nabla \cdot \vec{P}}{\epsilon_0}$$

This suggests defining the electric displacement field:

$$\vec{D} = \epsilon_0 \vec{E} + \vec{P}$$

which simplifies Gauss's law to:

$$\nabla \cdot \vec{D} = \rho_f$$

This is beautiful, since it brings back the same dependence on free charges as in vacuum, but now with the displacement field  $\vec{D}$  instead of the electric field  $\vec{E}$ .

## Linear Dielectric Response

For many materials, the polarization density is proportional to the applied electric field:

$$\vec{P} = \epsilon_0 \chi \vec{E}$$

where  $\chi$  is the electric susceptibility. This linear relationship defines linear dielectric materials, though real materials can exhibit nonlinear responses under strong fields.

The displacement field becomes:

$$\vec{D} = \epsilon_0 \vec{E} + \vec{P} = \epsilon_0 (1 + \chi) \vec{E} = \epsilon \vec{E}$$

with electric permittivity:

$$\epsilon = \epsilon_0 (1 + \chi) \quad (\text{electric permittivity})$$

The dimensionless ratio:

$$\epsilon_r = \frac{\epsilon}{\epsilon_0} = 1 + \chi$$

is the relative permittivity or dielectric constant, though it typically depends on frequency.

### Clausius-Mossotti Relation

In dense materials, the local electric field  $\vec{E}_{loc}$  experienced by an atom differs from the average field  $\vec{E}$ :

$$\vec{E}_{loc} = \vec{E} + \vec{E}_{dep}$$

where  $\vec{E}_{dep} = -\frac{\vec{P}}{3\epsilon_0}$  is the depolarization field from the cavity surface charges.

The polarization density relates to the local field through:

$$\vec{P} = N\alpha\vec{E}_{loc} = N\alpha(\vec{E} - \frac{\vec{P}}{3\epsilon_0})$$

Solving this equation and comparing with  $\vec{P} = \epsilon_0(\epsilon_r - 1)\vec{E}$  yields the Clausius-Mossotti relation:

$$\frac{\epsilon_r - 1}{\epsilon_r + 2} = \frac{N\alpha}{3\epsilon_0}$$

This fundamental relation connects microscopic polarizability to macroscopic permittivity, accounting for local field effects in dense media.

## 28.3 Maxwell's Equations in Matter

The presence of bound charges and currents in materials requires a modified form of Maxwell's equations. The four fundamental equations become:

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad (\text{ME.1})$$

Faraday's law remains unchanged, describing electromagnetic induction regardless of medium.

$$\nabla \cdot \vec{D} = \rho_f \quad (\text{ME.2})$$

Gauss's law now involves the displacement field  $\vec{D}$  and only free charges  $\rho_f$ .

$$\nabla \times \vec{H} = \frac{\partial \vec{D}}{\partial t} + \vec{j}_f \quad (\text{ME.3})$$

Ampère's law includes both the displacement current  $\frac{\partial \vec{D}}{\partial t}$  and free current density  $\vec{j}_f$ .

$$\nabla \cdot \vec{B} = 0 \quad (\text{ME.4})$$

The absence of magnetic monopoles remains fundamental. The material properties enter through the constitutive relations:

$$\vec{D} = \epsilon \vec{E} = \epsilon_0 \epsilon_r \vec{E} \quad (28.1)$$

$$\vec{B} = \mu \vec{H} = \mu_0 \mu_r \vec{H} \quad (28.2)$$

Similar to electric polarization, materials respond to magnetic fields through magnetization  $\vec{M}$ :

$$\vec{B} = \mu_0(\vec{H} + \vec{M})$$

For linear magnetic materials:

$$\vec{B} = \mu_0(1 + \chi_m)\vec{H}$$

where  $\chi_m$  is the magnetic susceptibility.

## 28.4 Wave Propagation in Non-conducting Matter

For isotropic media without free charges ( $\rho_f = 0$ ) and currents ( $\vec{j}_f = 0$ ), Maxwell's equations lead to the wave equation:

$$\nabla^2 \vec{E} - \frac{1}{v^2} \frac{\partial^2 \vec{E}}{\partial t^2} = 0$$

where the phase velocity is:

$$v = \frac{c}{\sqrt{\epsilon_r \mu_r}}$$

For non-magnetic materials ( $\mu_r = 1$ ), we define the refractive index:

$$n = \frac{c}{v} = \sqrt{\epsilon_r} \quad (\text{refractive index})$$

This fundamental relation connects our microscopic understanding of atomic polarization to the macroscopic phenomenon of light propagation. In a later section we will provide a more detailed discussion of the microscopic origins of the refractive index. Monochromatic waves in matter take the form:

$$\vec{E} = \vec{E}_0 e^{i(\omega t - \vec{k} \cdot \vec{r})}$$

The wavevector  $\vec{k}$  relates to the vacuum wavevector  $\vec{k}_0$  through:

$$\vec{k} = n \vec{k}_0$$

When electromagnetic waves enter a material, their wavelength changes while the frequency remains constant. This is because:

$$\lambda = \frac{\lambda_0}{n}$$

where  $\lambda_0$  is the vacuum wavelength. The wave frequency  $\omega$  remains unchanged:

$$\omega = \frac{2\pi c}{\lambda_0} = \frac{2\pi v}{\lambda}$$

## 28.5 Special Cases

### Negative Refraction

While most materials have  $n > 1$ , a remarkable class of engineered materials called metamaterials can exhibit negative refraction when both  $\epsilon_r < 0$  and  $\mu_r < 0$ . In these materials:

$$v = -\frac{c}{\sqrt{\epsilon_r \mu_r}} \quad (\text{negative refraction})$$

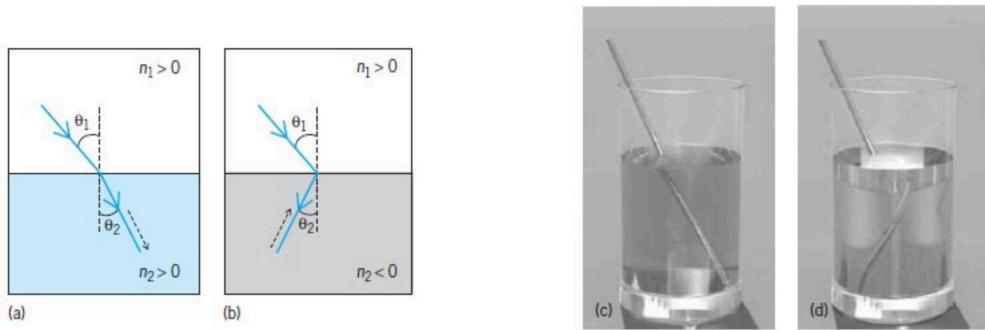


Figure 28.3: Refraction: Diagrams of (a) positive refraction and (b) negative refraction; and calculated images of a metal rod (c) in a glass filled with regular water ( $n = 1.3$ ), and (d) in a glass filled with “negative-index water” ( $n = -1.3$ ). In parts a and b, solid lines with arrows indicate the direction of the energy flows, broken lines with arrows show the direction of the wave vectors. (Parts c and d from G. Dolling et al., Opt. Express, 14:1842–1849, 2006).

### Energy Flow in Negative Index Materials

The Poynting vector describes energy flow in electromagnetic waves:

$$\vec{S} = \vec{E} \times \vec{H}$$

where  $\vec{H}$  relates to  $\vec{B}$  through  $\vec{B} = \mu \vec{H}$ . In negative index materials, both permittivity and permeability are negative, leading to unusual behavior.

For a plane wave,  $\vec{B} = \frac{1}{\omega} \vec{k} \times \vec{E}$ , giving:

$$\vec{H} = \frac{1}{\mu\omega} \vec{k} \times \vec{E}$$

Using the vector triple product identity:

$$\vec{S} = \frac{1}{\mu\omega} [\vec{k}(\vec{E} \cdot \vec{E}) - \vec{E}(\vec{E} \cdot \vec{k})]$$

Since  $\vec{E} \cdot \vec{k} = 0$  in a plane wave:

$$\vec{S} = \frac{1}{\mu\omega} \vec{k} E_0^2$$

In negative index materials ( $\vec{k} = -\hat{\vec{k}}$ ):

$$\vec{S} = -\frac{k E_0^2 \hat{\vec{k}}}{\mu\omega}$$

This remarkable result shows that energy flows opposite to wave propagation, a unique characteristic of negative index materials.

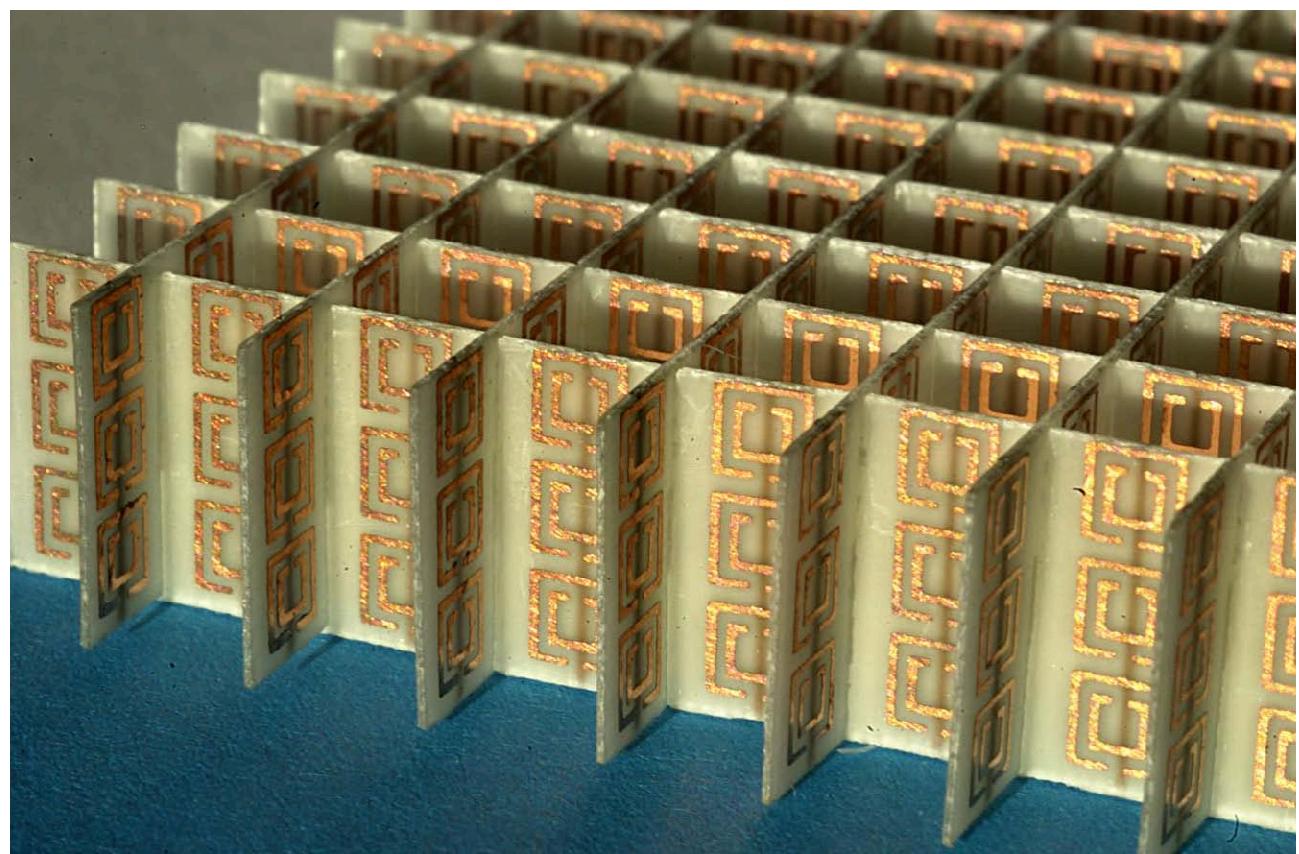


Figure 28.4: Metamaterial with negative refraction. Split-ring resonators provide negative permeability while metallic wires provide negative permittivity. Smith, D. R., et al. (2004). Science, 305(5685), 788-792.

## Metamaterial Realization

### Split-Ring Resonators (SRRs)

A split-ring resonator typically consists of a pair of concentric metallic rings, each with a gap. These rings can be thought of as forming an LC circuit, where the rings themselves provide the inductance ( $L$ ) and the gaps provide the capacitance ( $C$ ).

#### Resonant Frequency of the SRR

The SRR behaves as an LC resonator with a specific resonant frequency ( $\omega_0$ ). The resonant frequency can be expressed as:

$$\omega_0 = \frac{1}{\sqrt{LC}}$$

where: -  $L$  is the inductance of the rings. -  $C$  is the capacitance of the gaps.

#### Magnetic Response of the SRR

When an external alternating magnetic field is applied perpendicular to the plane of the SRR, it induces a circulating current around the rings. This induced current creates a magnetic dipole moment that opposes the change in the external magnetic field (Lenz's Law).

#### Magnetic Susceptibility

The magnetic susceptibility ( $\chi_m$ ) of the SRR can be related to the magnetic moment ( $m$ ) induced in response to the external magnetic field ( $H$ ):

$$m = \alpha H$$

where  $\alpha$  is the polarizability of the SRR. The susceptibility is given by:

$$\chi_m = \frac{m}{H} = \alpha$$

#### Polarizability of the SRR

The polarizability  $\alpha$  can be modeled using the Lorentz oscillator model for the resonant behavior of the SRR. This gives:

$$\alpha(\omega) = \frac{F\omega^2}{\omega_0^2 - \omega^2 - i\gamma\omega}$$

where: -  $F$  is a geometric factor related to the SRR. -  $\omega_0$  is the resonant frequency. -  $\gamma$  is the damping factor (related to losses). -  $\omega$  is the angular frequency of the applied magnetic field.

#### Effective Permeability

The effective permeability  $\mu(\omega)$  of the metamaterial containing SRRs can be expressed in terms of the magnetic susceptibility  $\chi_m(\omega)$ :

$$\mu(\omega) = 1 + \chi_m(\omega)$$

Substituting  $\chi_m(\omega)$  with  $\alpha(\omega)$ :

$$\mu(\omega) = 1 + \frac{F\omega^2}{\omega_0^2 - \omega^2 - i\gamma\omega}$$

This equation describes the frequency-dependent effective permeability of a metamaterial containing split-ring resonators.

### Negative Permeability

For negative permeability to occur, the term in the denominator ( $\omega_0^2 - \omega^2 - i\gamma\omega$ ) must be such that the real part of the permeability becomes negative. This generally happens in the frequency range around the resonant frequency  $\omega_0$ . Specifically, when  $\omega$  is slightly below  $\omega_0$ , the real part of  $\mu(\omega)$  can become negative due to the resonance.

The final equation for the effective permeability of a metamaterial with split-ring resonators is:

$$\mu(\omega) = 1 + \frac{F\omega^2}{\omega_0^2 - \omega^2 - i\gamma\omega}$$

### Interpretation

- When  $\omega$  is near  $\omega_0$ , the permeability  $\mu(\omega)$  can exhibit negative values.
  - The parameters  $F$ ,  $\omega_0$ , and  $\gamma$  are determined by the geometry and material properties of the SRRs.
- This equation highlights how the resonant properties of the SRRs lead to a negative permeability in the metamaterial, enabling unique electromagnetic properties such as a negative refractive index.

The concept of negative refractive index, first theorized by Victor Veselago in 1968, predicted materials with simultaneous negative permittivity and permeability would exhibit:

- Reversed Snell's law
- Reversed Doppler effect
- Reversed Cherenkov radiation

Modern realizations use carefully designed structures with:

- Split-ring resonators for negative
- Wire arrays for negative
- Precise geometric arrangements to maintain wave propagation

These materials enable novel applications including:

- Superlenses breaking the diffraction limit
- Electromagnetic cloaking
- Novel waveguiding devices

The concept of negative refractive index was first theorized by Veselago [[@veselago1968electrodynamics](#)]. Later, Pendry showed how these materials could be used to create perfect lenses [[@pendry2000negative](#)].

## References

## Chapter 29

# Dispersion & Absorption

When light interacts with matter, its electric field  $\vec{E}(\vec{r}, t)$  influences the charged particles within atoms. For visible light ( $\lambda \approx 500 \text{ nm}$ ) interacting with atoms (size  $\approx 0.1 \text{ nm}$ ), we can make two important approximations:

1. **Dipole Approximation:** Since  $\lambda \gg a$  (atomic size), the field appears uniform across each atom:

$$\vec{E}(\vec{r}, t) \approx \vec{E}(t)$$

2. **Local Field Approximation:** The field acting on each atom is approximately the macroscopic field:

$$\vec{E}_{\text{local}} \approx \vec{E}$$

For a linearly polarized wave along the x-direction:

$$\vec{E}(t) = E_0 \hat{x} e^{-i\omega t}$$

For isotropic media, the dipole moment  $\vec{p}$  of an atom is proportional to the local electric field:

$$\vec{p} = \alpha \vec{E}$$

The equation of motion for the electron displacement vector follows:

$$\ddot{\vec{r}} + \sigma \dot{\vec{r}} + \omega_0 \vec{r} = \frac{q}{m} \vec{E}(t)$$

For the x-component, we can write:

$$\ddot{x} + \sigma \dot{x} + \omega_0 x = \frac{q}{m} E_x(t)$$

The solution has the form:

$$x(t) = x_0 e^{i\omega t}$$

This leads to:

$$\vec{r}(t) = \frac{1}{\omega_0^2 + i\omega\sigma - \omega^2} \frac{q}{m} \vec{E}(t)$$

The oscillating dipole moment becomes  $\vec{p} = q\vec{r}(t) = \alpha \vec{E}(t)$ , and the polarization density:

$$\vec{P} = Nq \frac{1}{\omega_0^2 + i\omega\sigma - \omega^2} \frac{q}{m} \vec{E}(t) = \epsilon_0 \chi \vec{E}(t)$$

From this, we obtain the electronic susceptibility:

$$\chi = \chi_0 \frac{1}{\omega_0^2 + i\omega\sigma - \omega^2}$$

with

$$\chi_0 = \frac{q^2 N}{m\epsilon_0}$$

The susceptibility is complex, written as  $\chi = \chi' + i\chi''$ , making both the dielectric function  $\epsilon_r = 1 + \chi$  and refractive index complex quantities.

The complex refractive index takes the form:

$$n = n_r - i\kappa = \sqrt{\epsilon_r} = \sqrt{1 + \chi}$$

where the negative sign convention for the imaginary part is standard but not universal. Explicitly:

$$n = 1 + \frac{Nq^2}{2\epsilon_0 m} \frac{(\omega_0^2 - \omega^2) - i\sigma\omega}{(\omega_0^2 - \omega^2)^2 + \omega^2\sigma^2} = n_r - i\kappa$$

These real and imaginary components,  $n_r$  and  $\kappa$ , significantly affect light propagation. The two components can be written as

$$n_r = 1 + A \frac{(\omega_0^2 - \omega^2)}{(\omega_0^2 - \omega^2)^2 + \omega^2\sigma^2}$$

and

$$\kappa = A \frac{\sigma\omega}{(\omega_0^2 - \omega^2)^2 + \omega^2\sigma^2}$$

with  $A = \frac{Nq^2}{2\epsilon_0 m}$ .

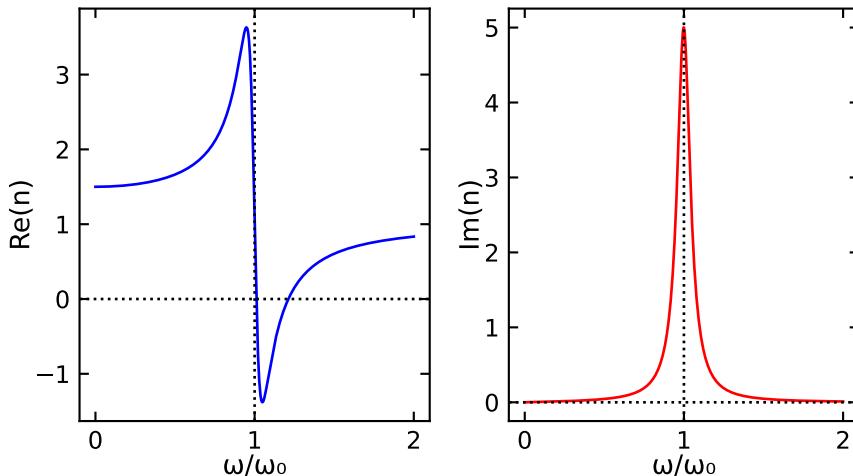


Figure 29.1: Real and imaginary part of the refractive index.

The plots show how both components vary with frequency due to the atomic resonance. Near resonance, the real part exhibits strong dispersion, transitioning from values above 1 to below 1. The imaginary part shows a Lorentzian peak with width determined by the damping coefficient  $\sigma$ .

### Measuring Refractive Index

A simple way to measure the refractive index is through the critical angle principle. By measuring the critical angle at which total internal reflection occurs, one can determine the refractive index of a material. This technique is widely used in refractometers and optical sensors.

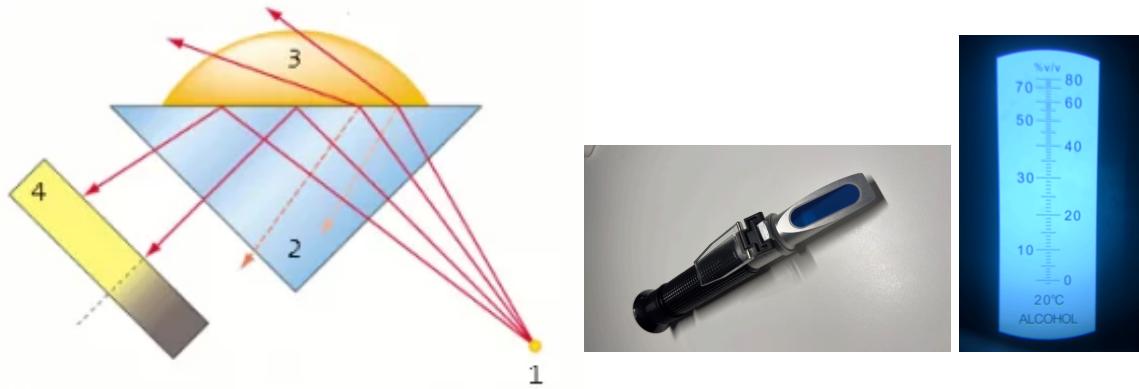


Figure 29.2: Principle of the Abbe refractometer (left). Abbe refractometer for alcohol content measurement (right).

## 29.1 Absorption

### Important

The imaginary component of the refractive index leads to the Lambert-Beer Law, a fundamental principle in optics and spectroscopy.

The imaginary component  $\kappa$  determines wave attenuation. For a plane wave propagating in the z-direction, the electric field is:

$$\vec{E}(z, t) = E_0 \hat{x} e^{i(kz - \omega t)}$$

The spatial part follows:

$$\vec{E}(z) = E_0 \hat{x} e^{-ikz} = E_0 \hat{x} e^{-ink_0 z}$$

Including the complex refractive index:

$$\vec{E}(z) = E_0 \hat{x} e^{-i(n-i\kappa)k_0 z} = E_0 \hat{x} e^{-in_r k_0 z} e^{-\kappa k_0 z}$$

The exponential decay factor modifies the wave amplitude with distance.

The intensity, which is proportional to the time-averaged Poynting vector magnitude, follows:

$$I = \frac{1}{2} \epsilon_0 c |\vec{E}|^2 \quad (29.1)$$

$$= \frac{1}{2} \epsilon_0 c |E_0|^2 |e^{-in_r k_0 z} e^{-\kappa k_0 z}|^2 \quad (29.2)$$

$$= I_0 e^{-2\kappa k_0 z} \quad (29.3)$$

$$= I_0 e^{-\alpha z} \quad (29.4)$$

where  $\alpha = 2k_0\kappa = \frac{4\pi\kappa}{\lambda}$  represents the absorption coefficient. This exponential decay of intensity with distance is known as the **Lambert-Beer Law**:

$$\frac{I}{I_0} = e^{-\alpha z} \quad (\text{Lambert Beer Law})$$

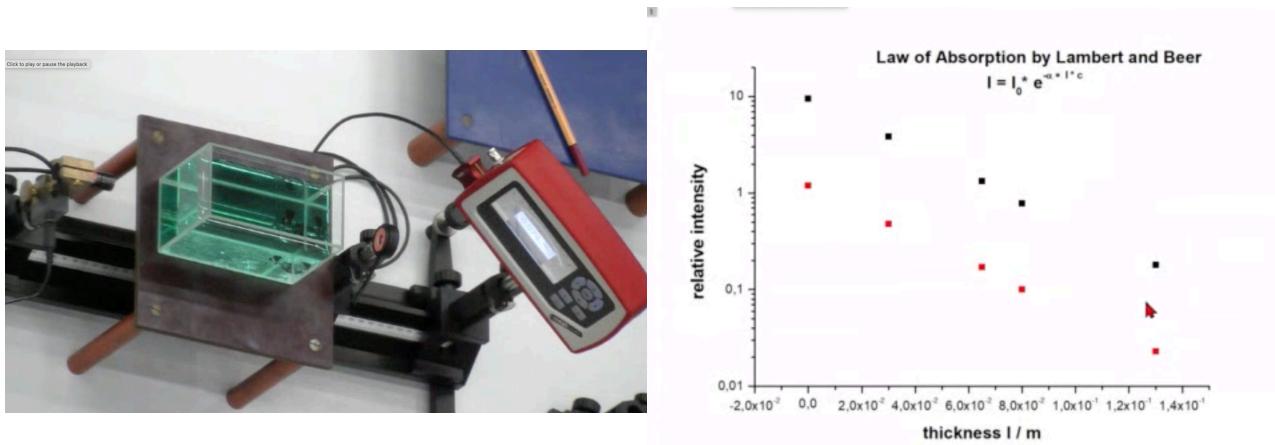


Figure 29.3: Measurement of the absorption coefficient.

The frequency-dependent absorption reflects resonant behavior, with strong absorption near resonance and weak absorption elsewhere. Real materials typically have multiple resonances, creating complex absorption spectra that serve as unique chemical fingerprints.

### **i** Absorption Measurements

The measurement of absorption is fundamental to many fields including chemistry, materials science, and biological studies. The basic principle relies on the Lambert-Beer Law:

$$\ln \left( \frac{I}{I_0} \right) = -\alpha(\lambda)L = -\epsilon(\lambda)cL$$

where:

- $I/I_0$  is the transmittance
- $\alpha(\lambda)$  is the absorption coefficient
- $\epsilon(\lambda)$  is the molar extinction coefficient
- $c$  is the molar concentration
- $L$  is the path length

The above logarithm can be used to define the absorbance  $A$  as

$$A = -\ln \left( \frac{I}{I_0} \right) / \ln(10)$$

which is the typical quantity that is plotted against concentration in absorption measurements and linearly

depends on the concentration. According to Lambert Beers law, the absorbance is calculated from

$$A = \epsilon cL$$

and provides the molar extinction coefficient  $\epsilon$  for the material at a given wavelength.

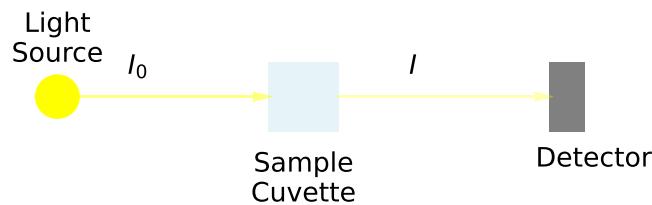


Figure 29.4: Basic setup for absorption measurements

## Concentration Measurements

The linear relationship between absorbance and concentration enables quantitative analysis:

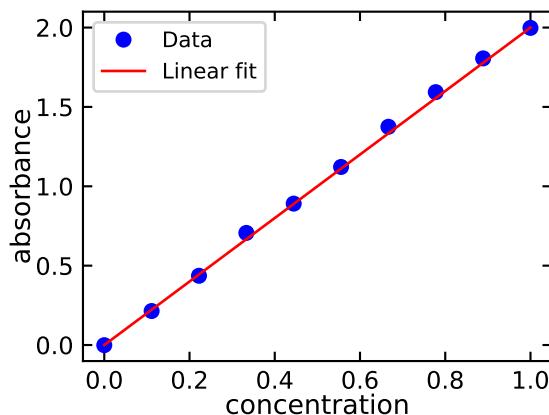


Figure 29.5: Beer-Lambert law verification plot

### Note

The linear range typically extends up to absorbance values of ~1.0. Beyond this, deviation from linearity may occur.

## Applications

### 1. Chemical Analysis

Chemical analysis provides quantitative information about composition and concentration in solutions. Absorption measurements enable precise determination of concentration levels across many compounds. Reaction kinetics can be monitored in real-time by tracking absorption changes. Quality control in industrial processes relies on rapid and accurate absorption measurements to ensure product consistency.

## 2. Materials Characterization

Optical properties of materials reveal crucial information about their electronic structure and physical properties. Band gap determination helps classify semiconductors and predict their behavior in devices. Film thickness measurements using absorption techniques provide non-destructive ways to characterize thin films and coatings. Understanding these properties is essential for developing new materials and optimizing their performance in applications.

## 3. Biological Studies

Protein quantification through absorption measurements forms a cornerstone of biochemical analysis. DNA and RNA analysis relies on characteristic absorption peaks to determine concentration and purity. Enzyme assays monitor reaction progress through changes in absorption, providing insight into biological processes. These techniques are fundamental to modern biological research and medical diagnostics.

## 29.2 Dispersion

### Note

Dispersion manifests through two key velocities:

- Phase velocity: speed of wave fronts
- Group velocity: speed of wave packets

The frequency dependence of the real refractive index affects wave propagation speeds in two important ways: through the phase velocity and group velocity.

### Phase Velocity

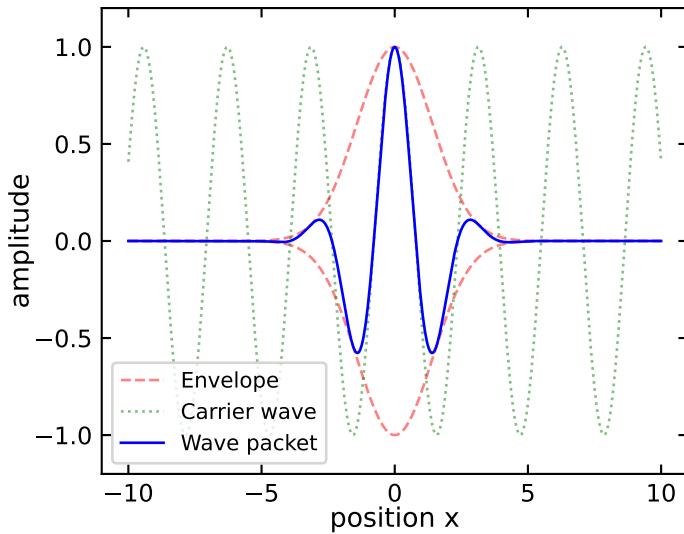
The phase velocity represents the speed of wave phase fronts:

$$v = \frac{\omega}{k} = \frac{c}{n_r} \quad (\text{phase velocity})$$

### Group Velocity

The group velocity represents the speed of wave packets

$$v_g = \frac{d\omega}{dk} \quad (\text{group velocity})$$



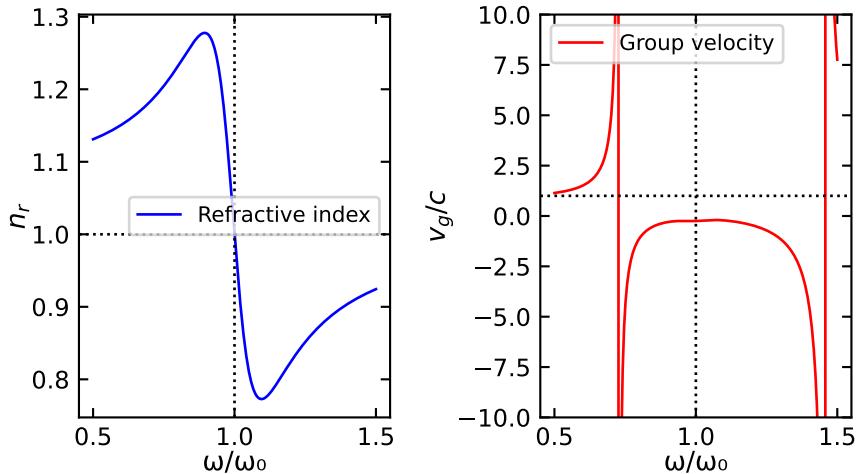
To relate group velocity to the refractive index, we can derive:

$$v_g = \frac{d}{dk}(vk) = v + k \frac{dv}{dk}$$

Using  $k = k_0 n_r$  and  $v = c/n_r$ , and after some algebra, we obtain:

$$v_g = \frac{c}{n_r + \omega \frac{dn_r}{d\omega}} \quad (\text{group velocity dispersion})$$

The group velocity is distinct from the phase velocity and represents how fast the envelope of a wave packet propagates through space. In a dispersive medium, where the refractive index depends on frequency, the group velocity can differ significantly from the phase velocity. This is particularly important in optical communications where information is carried by wave packets rather than single-frequency waves.



For optical pulses in fibers or other waveguides, group velocity dispersion leads to pulse spreading because different frequency components travel at different speeds. This effect becomes critical in long-distance optical communication systems where pulse broadening can lead to signal distortion and intersymbol interference.

In vacuum or non-dispersive media where the refractive index is constant ( $\frac{dn_r}{d\omega} = 0$ ), the group velocity equals the phase velocity. However, in most practical situations involving light propagation through materials, dispersion causes these velocities to differ, necessitating careful consideration in optical system design.

## Types of Dispersion

The frequency dependence of the refractive index leads to two distinct regimes of dispersion:

### 1. Normal Dispersion:

$$\frac{dn_r}{d\omega} > 0$$

In the normal dispersion regime, the refractive index increases with frequency, meaning that higher frequency (blue) light travels more slowly through the medium than lower frequency (red) light. This behavior is commonly observed in transparent materials at frequencies far from their resonances. The effects of normal dispersion are ubiquitous in optics: when white light passes through a prism, it separates into its spectral components creating a rainbow pattern; in optical systems, it causes chromatic aberration where different colors focus at different points; and in optical fibers, it leads to temporal spreading of pulses as different frequency components travel at different speeds through the medium.

**i** Note

Normal dispersion dominates in transparent materials at frequencies well below their resonances. This is why prisms separate white light into its spectral components with blue light bending more than red light.

### 2. Anomalous Dispersion:

$$\frac{dn_r}{d\omega} < 0$$

Anomalous dispersion occurs in the vicinity of absorption resonances, where the conventional relationship between frequency and refractive index is reversed. In this regime, higher frequencies propagate faster than lower frequencies, leading to unique optical phenomena. The behavior is characterized by strong frequency-dependent absorption accompanied by rapid variations in the refractive index. This unusual dispersion can even result in negative group velocities under certain conditions. The practical applications of anomalous dispersion are particularly important in modern optics, where it is used for pulse compression in ultrafast laser systems, dispersion compensation in optical communications, and various applications in ultrafast optics.

**!** Important

The transition between normal and anomalous dispersion typically occurs near absorption resonances, where the refractive index varies rapidly with frequency. This behavior is described by the Kramers-Kronig relations, which connect the real and imaginary parts of the refractive index.

Real materials often exhibit both types of dispersion across different frequency ranges, resulting in complex optical behavior. Understanding and controlling these dispersion effects is crucial for the development of optical communication systems, the generation and manipulation of ultrashort pulses, the design of optical components, and the implementation of nonlinear optical processes. The interplay between normal and anomalous dispersion enables sophisticated control over light propagation, essential for modern optical technologies.

## Superluminal Group Velocity

The concept of superluminal group velocity emerges in regions of anomalous dispersion, where the group velocity can mathematically exceed the speed of light  $c$ . This occurs when:

$$n_r + \omega \frac{dn_r}{d\omega} < 1$$

In our classical oscillator model, this condition is met near the resonance frequency when:

$$|\omega_0 - \omega| < \frac{\sigma}{2}$$

This seemingly paradoxical result warrants careful interpretation. The group velocity, while traditionally associated with the speed of energy or information transport, becomes problematic near resonances. Several key points help clarify this phenomenon:

- Physical Interpretation:** The superluminal effect is associated with pulse reshaping rather than actual faster-than-light signal propagation. The pulse peak appears to emerge from the medium before it would in vacuum due to preferential absorption and dispersion of different frequency components.
- Pulse Distortion:** Near resonances, pulses undergo severe distortion, making the group velocity less meaningful as a measure of signal propagation. The pulse envelope no longer maintains its shape, and different parts of the pulse travel at different velocities.
- Causality and Energy Transport:** While  $v_g$  may exceed c, the energy transport velocity:

$$v_E = \frac{\vec{S}}{w_{em}} = \frac{I}{w_{em}} \leq c$$

where  $\vec{S}$  is the Poynting vector and  $w_{em}$  is the electromagnetic energy density, always remains subluminal, preserving causality.

#### i Note

The apparent superluminal propagation has been observed experimentally in various systems, including:

- Gain media
- Photonic crystals
- Atomic gases near absorption lines (see @superluminal)
- Optical fibers with specially engineered dispersion

#### ! Important

Despite superluminal group velocity, no information can be transmitted faster than c. This is ensured by:

1. The severe pulse distortion in regions of anomalous dispersion
2. The Kramers-Kronig relations linking absorption and dispersion
3. The principle of causality in electromagnetic theory

The study of superluminal group velocity has practical applications in: - Understanding fundamental limits of signal propagation - Designing fast-light and slow-light devices - Developing optical delay lines and buffers - Quantum information processing

This phenomenon illustrates the subtle interplay between classical electromagnetic theory, special relativity, and the practical limitations of optical pulse propagation.

#### i Kramers-Kronig Relations

The Kramers-Kronig relations express a fundamental connection between the real and imaginary parts of the complex refractive index (or susceptibility). These relations arise from causality - the principle that effects cannot precede their causes - and the analytical properties of response functions in the complex frequency plane.

#### Physical Origin

Consider a general response function  $\chi(\omega)$  describing how a system responds to electromagnetic perturbations. For this response to be causal,  $\chi(\omega)$  must be analytic in the upper half of the complex frequency plane. This analyticity requirement, combined with the reality condition  $\chi(-\omega) = \chi^*(\omega)$ , leads to the Kramers-Kronig relations.

### Principal Relations

For the complex susceptibility  $\chi(\omega) = \chi'(\omega) + i\chi''(\omega)$ , the Kramers-Kronig relations take the form:

$$\chi'(\omega) = \frac{1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} \frac{\chi''(\omega')}{\omega' - \omega} d\omega'$$

$$\chi''(\omega) = -\frac{1}{\pi} \mathcal{P} \int_{-\infty}^{\infty} \frac{\chi'(\omega')}{\omega' - \omega} d\omega'$$

where  $\mathcal{P}$  denotes the Cauchy principal value of the integral.

#### Note

Similar relations hold for the complex refractive index  $n(\omega) = n_r(\omega) + i\kappa(\omega)$ :

$$n_r(\omega) - 1 = \frac{2}{\pi} \mathcal{P} \int_0^{\infty} \frac{\omega' \kappa(\omega')}{\omega'^2 - \omega^2} d\omega'$$

$$\kappa(\omega) = -\frac{2\omega}{\pi} \mathcal{P} \int_0^{\infty} \frac{n_r(\omega') - 1}{\omega'^2 - \omega^2} d\omega'$$

### Implications

The Kramers-Kronig relations reveal profound connections in optical physics. Most significantly, they demonstrate that absorption and dispersion are intrinsically linked phenomena. It is impossible to have a perfectly transparent medium that exhibits dispersion, as the presence of dispersion necessarily implies some absorption at certain frequencies. This connection is mathematically expressed through sum rules, such as the integral relation  $\int_0^{\infty} \omega' \chi''(\omega') d\omega' = 0$ , which places constraints on physically possible optical responses.

The relations enforce causality in light-matter interactions, ensuring that optical responses cannot precede their stimuli. This causality principle manifests in the observed behavior of optical materials, particularly near resonances where anomalous dispersion invariably accompanies absorption peaks.

#### Important

The Kramers-Kronig relations demonstrate that dispersion and absorption are not independent phenomena but are fundamentally linked through causality. This connection explains why anomalous dispersion always occurs near absorption resonances.

### Applications

The practical significance of the Kramers-Kronig relations extends throughout optical physics and engineering. Scientists and engineers use these relations to determine complete optical responses from partial measurements, as measuring either the real or imaginary part allows reconstruction of the other. This capability proves invaluable in materials characterization and optical device design. The relations also serve as a theoretical framework for validating experimental optical data and understanding the fundamental limits of optical materials.

In optical device design, these relations guide the development of components with specific dispersion properties, though they also highlight the inherent trade-offs between dispersion and absorption that must be considered. The understanding provided by the Kramers-Kronig relations has become essential in fields ranging from spectroscopy to telecommunications, where precise control of optical properties is crucial.

The Kramers-Kronig relations stand as one of the most elegant examples of how fundamental physical principles - in this case, causality - manifest in measurable optical properties, providing both practical tools for optical engineering and deep insights into the nature of light-matter interactions.

# Chapter 30

## Reflection and Refraction of Electromagnetic Waves

### 30.1 Effect of the Refractive Index

The effect of the material on the wave propagation can be understood when considering a plane wave  $E_s$  incident on a thin material slab of thickness  $\Delta z$ .

$$\begin{aligned} E &= E_0 \cdot e^{-i(\omega t - kz - (n-1)k\Delta z)} \\ &= E_0 e^{-i(n-1)k\Delta z} e^{i(\omega t - kz)} \\ &= e^{-i\theta} E_s \quad \text{with } \theta = k(n-1)\Delta z \end{aligned} \tag{30.1}$$

For small values of  $\theta$  the exponential function can be approximated by

$$e^x \sim 1 + x + \frac{x^2}{2} \tag{30.2}$$

such that we obtain

$$e^{-ik(n-1)\Delta z} \approx 1 - ik(n-1)\Delta z - \frac{k^2(n-1)^2 \Delta z^2}{2} \tag{30.3}$$

The total field behind the thin slab therefore is

$$E(z) = \underbrace{E_0 e^{i(\omega t - kz)}}_{E_e} - \underbrace{ik(n-1)\Delta z E_0 e^{i(\omega t - kz)}}_{E_{\text{medium}}} \tag{30.4}$$

As shown in Figure ??, the resulting wave is delayed by a phase factor  $k(n-1)\Delta z$  (see Equation ??) and the amplitude is reduced by a factor  $k(n-1)\Delta z$ .

### 30.2 General Description of Reflection and Refraction

Before we examine the behavior of electromagnetic waves at boundaries, we need to define the geometry and examine the electric and magnetic fields present. Figure ?? shows the wavevectors of the incident  $\vec{k}_I$ , reflected  $\vec{k}_R$  and transmitted  $\vec{k}_T$  waves in the two materials with the refractive indices  $n_1$  and  $n_2$ .

These wavevectors are connected to the following plane waves:

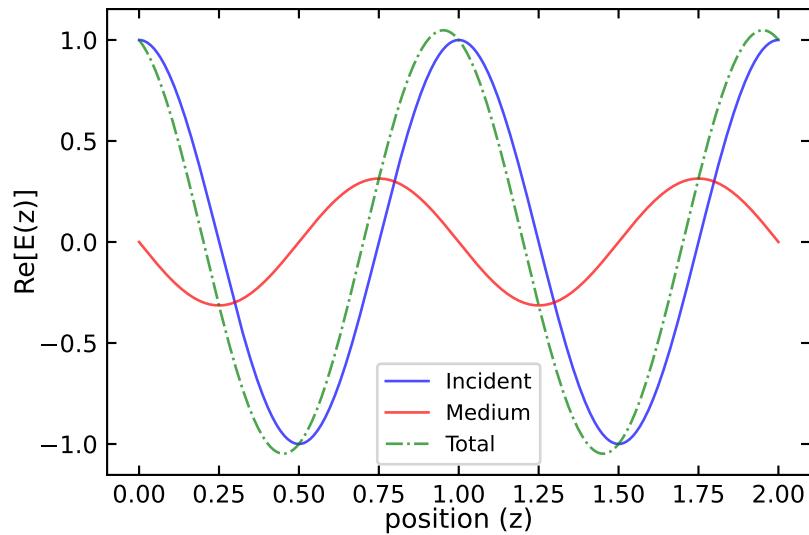


Figure 30.1: Components of the electric field in medium

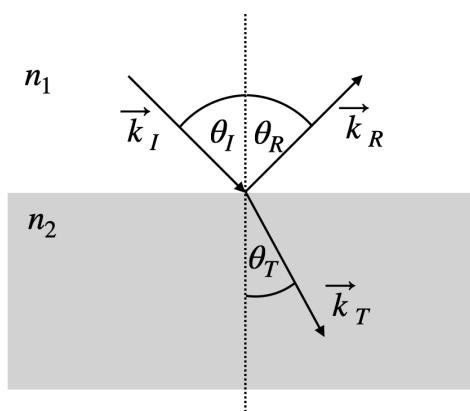


Figure 30.2: Reflection and refraction of an electromagnetic wave at a boundary.

$$\vec{E}_{inc} = \vec{E}_I e^{i(\omega_I t - \vec{k}_I \cdot \vec{r})} \quad (30.5)$$

$$\vec{E}_{ref} = \vec{E}_R e^{i(\omega_R t - \vec{k}_R \cdot \vec{r})} \quad (30.6)$$

$$\vec{E}_{tra} = \vec{E}_T e^{i(\omega_T t - \vec{k}_T \cdot \vec{r})} \quad (30.7)$$

As discussed in Equation ?? through Equation ??, we must consider the direction of polarization of their electric or magnetic fields. We differentiate between:

- **p-polarized light:** electric field is in the plane of incidence (given by the  $k$ -vector and the surface normal), also called transverse magnetic (TM)
- **s-polarized light:** electric field is perpendicular to the plane of incidence, also called transverse electric (TE)

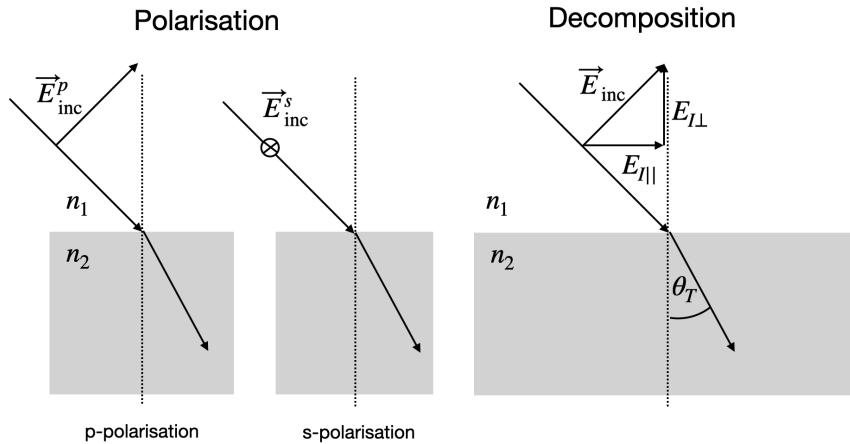


Figure 30.3: Decomposition of the incident electric field into components normal and tangential to the dielectric boundary.

As shown in Figure ??, we need to split these polarisation vectors into components that are **parallel** ( $\parallel$ ) or **perpendicular** ( $\perp$ ) to the interface. This is required for applying boundary conditions for the electric and magnetic fields.

## Boundary Conditions

The boundary conditions for the electric or magnetic field passing an interface are derived from the Maxwell equations.

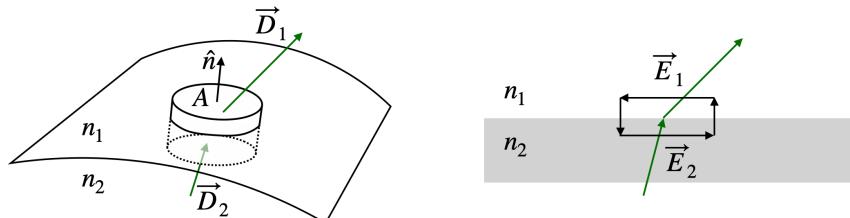


Figure 30.4: Integration over a closed surface (left) and close path (right) to obtain the boundary conditions for the electric field components.

Let us take the divergence of the displacement field  $\nabla \cdot \vec{D} = \rho_f$  which is equal to the density of free charges. When integrating both sides over the volume:

$$\int \nabla \vec{D} dV = \int \rho_f dV \quad (30.8)$$

we can apply Gauss' theorem and replace the volume integral over the divergence with an integral over closed surface of that volume:

$$\oint \vec{D} d\vec{A} = q_f \quad (30.9)$$

where  $d\vec{A}$  is a vector standing perpendicular on the surface element  $dA$  and  $q_f$  are the free charges in the volume. When we consider a pillbox-shaped volume straddling the interface between two materials, the surface integral can be broken down into three parts:

1. Top surface (in material 1)
2. Bottom surface (in material 2)
3. Side surface (cylindrical part)

The total surface integral becomes:

$$\oint \vec{D} \cdot d\vec{A} = \int_{\text{top}} \vec{D}_1 \cdot d\vec{A} + \int_{\text{bottom}} \vec{D}_2 \cdot d\vec{A} + \int_{\text{side}} \vec{D} \cdot d\vec{A}$$

For a pillbox of height  $h$  and radius  $r$ :

- The top surface contributes:  $\vec{D}_1 \cdot \hat{n} A$  (negative because  $d\vec{A}$  points along  $\hat{n}$ )
- The bottom surface contributes:  $-\vec{D}_2 \cdot \hat{n} A$
- The side surface contribution goes to zero as  $h \rightarrow 0$  (as area  $\sim 2\pi r h$ )

Therefore:

$$\oint \vec{D} \cdot d\vec{A} = A(\vec{D}_1 - \vec{D}_2) \cdot \hat{n} = q_f$$

### Key Points

1. The side surface contribution vanishes as  $h \rightarrow 0$  because:
  - Its area scales with  $h$  ( $2\pi r h$ )
  - The field components remain finite
2. Only the normal components of  $\vec{D}$  contribute because:
  - $d\vec{A}$  is parallel to  $\hat{n}$  for top and bottom surfaces
  - The tangential components don't contribute to the dot product
3. The ratio  $q_f/A$  becomes the surface charge density  $\sigma_f$  as  $h \rightarrow 0$

Following that, we obtain the boundary condition for the normal component of the displacement field:

$$D_{1\perp} = D_{2\perp} \quad (30.10)$$

This implies a jump in the normal electric field component:

$$\frac{E_{1\perp}}{E_{2\perp}} = \frac{\epsilon_2}{\epsilon_1} \quad (30.11)$$

as  $D = \epsilon E$ . Another boundary condition arises from the curl of the electric field. Using the Maxwell equation:

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \quad (30.12)$$

Integrating both sides over an area and applying Stokes theorem:

$$\oint \vec{E} d\vec{l} = -\frac{\partial}{\partial t} \int \vec{B} d\vec{A} \quad (30.13)$$

Consider a rectangular loop straddling the interface between two media, with a height  $h$  a width  $w$  and a normal to the interface  $\hat{n}$  the line integral can be broken down into four parts:

1. Top segment (in medium 1)
2. Bottom segment (in medium 2)
3. Two vertical segments connecting them

The line integral becomes:

$$\oint \vec{E} \cdot d\vec{l} = w\vec{E}_1 \cdot \hat{t} - w\vec{E}_2 \cdot \hat{t} + (\text{vertical segments})$$

where  $\hat{t}$  is the unit vector tangent to the interface.

As  $h \rightarrow 0$ :

1. The contribution from vertical segments vanishes (as  $h \rightarrow 0$ )
2. The area of the loop approaches zero, making the right-hand side zero:

$$-\frac{\partial}{\partial t} \int \vec{B} \cdot d\vec{A} \rightarrow 0$$

Therefore:

$$w(\vec{E}_1 - \vec{E}_2) \cdot \hat{t} = 0$$

Since  $w \neq 0$ , this implies:

$$(\vec{E}_1 - \vec{E}_2) \cdot \hat{t} = 0$$

This can be rewritten in terms of the cross product with the normal vector:

$$\hat{n} \times (\vec{E}_2 - \vec{E}_1) = 0 \quad (30.14)$$

Therefore:

$$E_{2||} = E_{1||} \quad (30.15)$$

indicating that the tangential component of the electric field is conserved.

### Boundary Conditions for the Magnetic Field

For the magnetic field, we can derive boundary conditions using similar approaches. Starting with Ampère's law:

$$\nabla \times \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t}$$

Integrating over an area and applying Stokes' theorem:

$$\oint \vec{H} \cdot d\vec{l} = \int \vec{J} \cdot d\vec{A} + \frac{\partial}{\partial t} \int \vec{D} \cdot d\vec{A}$$

Using the same rectangular loop approach as with the electric field, but now in the limit as  $h \rightarrow 0$ :

1. The area integral of the displacement current ( $\partial \vec{D}/\partial t$ ) vanishes
2. The surface current density  $\vec{K} = \vec{J}dA$  remains finite

This yields:

$$\hat{n} \times (\vec{H}_2 - \vec{H}_1) = \vec{K}$$

In the absence of surface currents ( $\vec{K} = 0$ ), we have:

$$H_{2||} = H_{1||} \quad (30.16)$$

For the normal component, starting from  $\nabla \cdot \vec{B} = 0$  and using the pillbox approach as with the electric field:

$$\oint \vec{B} \cdot d\vec{A} = 0$$

This leads directly to:

$$B_{1\perp} = B_{2\perp} \quad (30.17)$$

Since  $\vec{B} = \mu \vec{H}$ , this implies a jump in the normal component of the H-field:

$$\frac{H_{1\perp}}{H_{2\perp}} = \frac{\mu_2}{\mu_1} \quad (30.18)$$

These boundary conditions for the magnetic field complement those for the electric field and are essential for determining the reflection and transmission coefficients at interfaces.

## **i** Summary Boundary Conditions at an Interface

For a boundary between two media (1 and 2) with surface normal  $\hat{n}$ :

### Electric Field

- Normal component (with surface charge density  $\sigma_f$ ):

$$\epsilon_1 E_{1\perp} - \epsilon_2 E_{2\perp} = \sigma_f$$

- Tangential component:

$$E_{1||} = E_{2||}$$

### Magnetic Field

- Normal component:

$$B_{1\perp} = B_{2\perp}$$

- Tangential component (with surface current density  $\vec{K}$ ):

$$\hat{n} \times (\vec{H}_2 - \vec{H}_1) = \vec{K}$$

In the absence of free charges ( $\sigma_f = 0$ ) and currents ( $\vec{K} = 0$ ), these reduce to:

- $\epsilon_1 E_{1\perp} = \epsilon_2 E_{2\perp}$
- $E_{1||} = E_{2||}$
- $B_{1\perp} = B_{2\perp}$
- $H_{1||} = H_{2||}$

## Reflection/Refraction

### Frequency and Wavevector Matching

Referring to Figure ??, we can explicitly write down the components of the wavevectors of the three waves in our coordinate system:

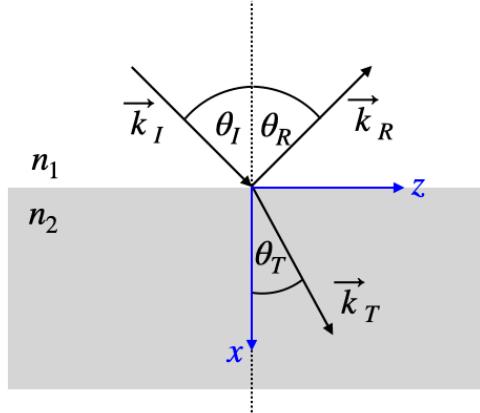


Figure 30.5: Coordinate system for wave at interfaces.

According to the coordinate system shown in Figure ??, we have:

$$\begin{aligned}\vec{k}_I &= k_I \cos(\theta_I) \hat{e}_x + k_I \sin(\theta_I) \hat{e}_z \\ \vec{k}_R &= -k_R \cos(\theta_R) \hat{e}_x + k_R \sin(\theta_R) \hat{e}_z \\ \vec{k}_T &= k_T \cos(\theta_T) \hat{e}_x + k_T \sin(\theta_T) \hat{e}_z\end{aligned}\quad (30.19)$$

Note that the component  $\hat{e}_x$  always provides the wavevector component perpendicular to the interface, while  $\hat{e}_z$  is the tangential (parallel) component. The total field on both sides is given by:

$$\vec{E} = \vec{E}_{inc} + \vec{E}_{ref} \quad (\text{for } x < 0) \quad (30.20)$$

and

$$\vec{E} = \vec{E}_{tra} \quad (\text{for } x > 0) \quad (30.21)$$

For these fields to match according to our previously described boundary conditions, we require:

$$\omega_I = \omega_R = \omega_T \quad (30.22)$$

This frequency matching confirms our initial intuition. Along the interface, we also need phase matching of the waves as discussed in the wave optics chapter:

$$\vec{k}_I \cdot \vec{r} = \vec{k}_R \cdot \vec{r} = \vec{k}_T \cdot \vec{r} \quad (30.23)$$

at all positions  $\vec{r}$  that belong to the interface (i.e.,  $x = 0$ ). Therefore  $\vec{r} = \{0, y, z\}$  and the equalities yield:

$$k_I \sin(\theta_I) = k_R \sin(\theta_R) = k_T \sin(\theta_T) \quad (30.24)$$

Since the magnitude of the wavevector of the incident and the reflected light is the same (both waves travel in the same material), we find:

$$\theta_I = \theta_R \quad (30.25)$$

For the incident and transmitted waves, we must account for the change in wavenumber:

$$\begin{aligned} n_1 k_0 \sin(\theta_I) &= n_2 k_0 \sin(\theta_T) \\ n_1 \sin(\theta_I) &= n_2 \sin(\theta_T) \end{aligned} \quad (30.26)$$

Equation ?? represents Snell's law, which results from the conservation of the parallel component of the wavevector across an interface, while the normal component must have a jump according to the refractive indices.

# Chapter 31

## Fresnel Equations

In the last lecture, we have discussed the matching of the frequency and the wavevector at an interface between two materials. We now also need to have a look at the matching of the electric field amplitudes at the boundaries.

In the last lecture, we have discussed the matching of the frequency and the wavevector at an interface between the two materials. We now also need to have a look at the matching of the electric field amplitudes at the boundaries.

### 31.1 Reflection

#### s-polarized light

We would first like to have a look at light that is polarized perpendicular to the incident plane - so-called s-polarized light. According to our previous definition, s-polarized light is directed along the y-axis of our coordinate system. The incident light amplitude is therefore written as

$$\vec{E}_I = E_I \hat{e}_y$$

This electric field vector is perpendicular to the plane of incidence. With this polarization the reflected and the refracted electric field also have to have this polarization and thus

$$\vec{E}_R = E_R \hat{e}_y$$

and

$$\vec{E}_T = E_T \hat{e}_y$$

From this it also follows that at the boundary, the incident and the reflected field must be the same as the transmitted field in our matching condition. This is due to the fact that s-polarized light is always parallel to the interface.

$$E_I + E_R = E_T \quad (\text{matching})$$

The magnetic field lies then in the plane of incidence as it is perpendicular to  $\vec{k}_I$  and  $\vec{E}_I$ . It has components parallel  $\parallel$  and perpendicular  $\perp$  to the interface. The tangential components as well as the perpendicular ones are conserved so we can write for the tangential ones

$$B_I \cos(\theta_I) - B_R \cos(\theta_R) = B_T \cos(\theta_T)$$

Given the fact that

$$\vec{B} = \frac{1}{v}(\hat{v} \times \vec{E})$$

and that all three vectors are orthogonal we can use the relation  $B = E/v$  for the magnetic and electric field amplitudes and  $\theta_I = \theta_R$  to obtain

$$\frac{E_I - E_R}{v_1} \cos(\theta_I) = \frac{E_T}{v_2} \cos(\theta_T)$$

With the help of the condition *matching* above we can replace the transmitted electric field  $E_T$  and using  $v_1 = c/n_1$  and  $v_2 = c/n_2$  we finally find

$$\frac{E_R}{E_I} = \frac{n_1 \cos(\theta_I) - n_2 \cos(\theta_T)}{n_1 \cos(\theta_I) + n_2 \cos(\theta_T)} = r_s$$

This is the **Fresnel coefficient for the reflection of s-polarized light**  $r_s$ . If we replace not the transmitted but the reflected electric field in the formula above, we may obtain the **Fresnel coefficient for the transmission of s-polarized light**  $t_s$

$$\frac{E_T}{E_I} = \frac{2n_1 \cos(\theta_I)}{n_1 \cos(\theta_I) + n_2 \cos(\theta_T)} = t_s$$

### p-polarized light

If the electric field is parallel to the plane of incidence, the we need to split it into parallel  $\parallel$  and perpendicular  $\perp$  components. The incident field may be written as

$$\vec{E}_I = E_I \sin(\theta_I) \hat{e}_x + E_I \cos(\theta_I) \hat{e}_z$$

where  $\hat{e}_z$  and  $\hat{e}_x$  are the unit vectors in the z- and x-direction, respectively. The first term is the normal component to the boundary and the second the parallel component. We may pick our the parallel part, for which we know that the electric field is just continuous across the interface.

$$E_I \cos(\theta_I) + E_R \cos(\theta_R) = E_T \cos(\theta_T)$$

Now the magnetic field is perpendicular to the plane of incidence and we may write

$$B_I - B_R = B_T$$

for the matching condition of the magnetic field. From this follows that

$$\frac{E_I - E_R}{v_1} = \frac{E_T}{v_2}$$

We may now replace again the transmitted field, which results finally in the **Fresnel coefficient for the reflection of p-polarized light**  $r_p$

$$\frac{E_R}{E_I} = \frac{n_2 \cos(\theta_I) - n_1 \cos(\theta_T)}{n_1 \cos(\theta_T) + n_2 \cos(\theta_I)} = r_p$$

Finally, we may also do the same for the **Fresnel coefficient of the transmission of p-polarized light**  $t_p$

$$\frac{E_T}{E_I} = \frac{2n_1 \cos(\theta_I)}{n_1 \cos(\theta_T) + n_2 \cos(\theta_I)} = t_p$$

**i** Note

**Fresnel Equations**

The Fresnel equations give the relations for the amplitude of the transmitted and reflected electric fields to the incident electric field amplitude as a function of the angle of incident and the light polarization.

**s-polarization**

$$\frac{E_R}{E_I} = \frac{n_1 \cos(\theta_I) - n_2 \cos(\theta_T)}{n_1 \cos(\theta_I) + n_2 \cos(\theta_T)} = r_s \quad (\text{Fresnel Coefficient } r_s)$$

$$\frac{E_T}{E_I} = \frac{2n_1 \cos(\theta_I)}{n_1 \cos(\theta_I) + n_2 \cos(\theta_T)} = t_s \quad (\text{Fresnel Coefficient } t_s)$$

**p-polarization**

$$\frac{E_R}{E_I} = \frac{n_2 \cos(\theta_I) - n_1 \cos(\theta_T)}{n_1 \cos(\theta_T) + n_2 \cos(\theta_I)} = r_p \quad (\text{Fresnel Coefficient } r_p)$$

$$\frac{E_T}{E_I} = \frac{2n_2 \cos(\theta_I)}{n_1 \cos(\theta_T) + n_2 \cos(\theta_I)} = t_p \quad (\text{Fresnel Coefficient } t_p)$$

## Air to Glass

Lets discuss the results we obtained with the help of specific examples. We will consider the interface between air ( $n_1 = 1$ ) and glass ( $n_2 = 1.5$ ) and vary the angle of incidence. The transmission angle can be obtained from Snell's law  $n_1 \sin(\theta_I) = n_2 \sin(\theta_T)$ . Besides the Fresnel coefficients, we plot also the phase. This phase gives us an idea about possible phase changes upon reflection and refraction. We have previously assumed for example, that under normal incidence we obtain a phase jump of  $\pi$  upon reflection when coming from air to glass. This is something we may check now.

To do so, we just represent the complex Fresnel coefficients as  $r_s = |r_s|e^{i\phi}$

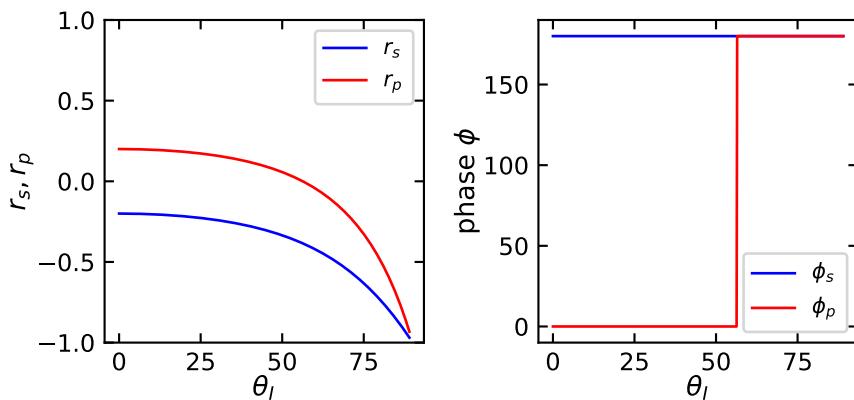


Figure 31.1: Fresnel coefficients (left) and phase (right) of the reflected light observed for the reflection at an air( $n_1 = 1$ )/glass( $n_2 = 1.5$ ) interface.

The above graph displays the reflection coefficients  $r_s, r_p$  as a function of the angle of incidence. The value of  $r_s$  is negative for the whole range indicating there is a phase jump by an angle of  $\pi$  as we assumed already in the thin film interference section. This phase jump is also confirmed in the second plot on the right. Note that the reflection coefficient for the parallel polarization  $r_p$  is positive up to an angle of about  $56^\circ$ . This also

means that there is no phase jump for this component up to this angle. Beyond this angle we find also a phase jump of  $\phi$ . Note that normal incidence  $\theta_I = 0$  is a special point, where all incident electric field are tangential to the interface. So also  $r_p$  will be negative at that point and for this incident angle there is no plane of incidence definition.

The special angle where the  $r_p = 0$  is called the **Brewster Angle**. It is special since at this angle of incidence unpolarized light will be turned into completely s-polarized light in reflection. This also means that you can get rid of a reflection from an air/glass interface, when observing this interface with a linear polarizer.

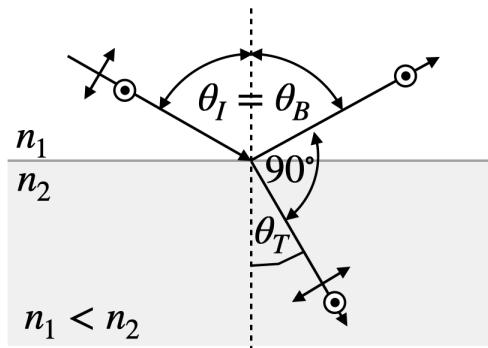


Figure 31.2: Reflections observed with linear polarizer.

Following the Fresnel formula for  $r_p$ , the reflection coefficient becomes zero when  $n_2 \cos(\theta_I) = n_1 \cos(\theta_T)$ . Using Snells law as well, we find that

$$\theta_I + \theta_T = \frac{\pi}{2}$$

and thus the following definition for the Brewster angle ( $\theta_B$ )

$$\tan(\theta_B) = \frac{n_2}{n_1}$$

The Brewster angle arises from the fact that the dipoles which are induced by the incident light in the material oscillate along the direction of the reflected light. We will see later that dipoles do not emit light along its oscillation direction and hence, there is no reflection for the in plane polarization.

### Note

#### Brewster Angle

The Brewster angle is the angle under which the reflection of light with a polarization in the plane of incidence vanishes.

$$\tan(\theta_B) = \frac{n_2}{n_1}$$

#### Brewsters Pyramid

#### Brewster Angle Microscopy

Brewster angle microscopy (BAM) is a powerful surface-sensitive technique that takes advantage of the Brewster angle phenomenon. At the Brewster angle, p-polarized light is completely transmitted through an interface, resulting in zero reflection. However, if there are molecules or thin films present at the interface, they will disturb this condition and create a reflection that can be detected.



Figure 31.3: Reflections from a pyramid observed with linear polarizer. The pyramid sides are arranged at an angle, which corresponds to the Brewster angle. If polarized light in a certain direction is falling on the pyramid faces, the reflection disappears.

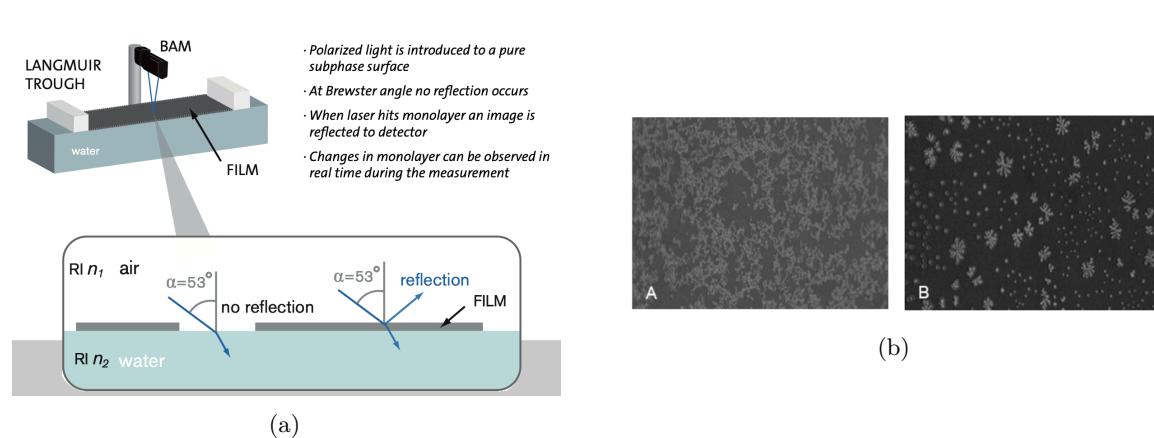


Figure 31.4: MODEL SKIN LIPIDS a) without and b) with 40 % addition of oleic acid at 10 mN/m imaged with KSV NIMA BAM. The size of the images is ~720 m (W)  $\times$  400 m (H). With permission from Langmuir 2013, 29 (15), pp 4857–4865. Copyright 2013 American Chemical Society.

### Reflection Removal

#### Glass to Air

If we invert the order of the materials and have light incident to a glass/air interface we observe a new effect in the **total internal reflection of light**. First of all the p-polarization obtains now a phase shift by  $\pi/2$  up to the Brewster angle, which is now at  $33^\circ$ , when it gets back in phase with the incident light. The s-polarized light has no phase jump at the interface, yet both components reach a magnitude of 1 at the critical angle of total internal reflection  $\theta_C$ . Starting from this incident angle, all light is reflected by the glass/air interface. As we know from the geometrical optics, the angle of total internal reflection is obtained when the transmission angle becomes  $90^\circ$ .

$$\sin(\theta_C) = \frac{n_2}{n_1} \quad (n_2 < n_1)$$

For the glass/air interface as shown below, this total internal reflection occurs for an angle  $\theta_C = 41.8^\circ$ . While the magnitude of the reflection coefficient is then 1, the phase of the reflected light changes continuously from  $0^\circ$  to  $90^\circ$ .

We will investigate the transmitted light in the case of total internal reflection below in the transmission section. You will see in the images below also, that on the reflection side a standing interference pattern is observed, due to the superposition of incident and reflected wave.



Figure 31.5: Reflections from surfaces which are observed under an angle are often partially polarized such that the observation with the help of a polarizer (e.g. in photography), can greatly reduce the highlights from reflections and make objects behind windows visible.

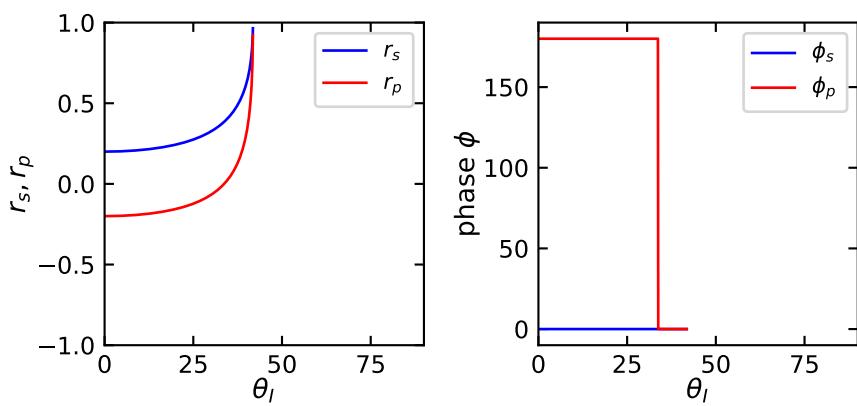


Figure 31.6: Fresnel coefficients (left) and phase (right) of the reflected light observed for the reflection at an air( $n_1 = 1$ )/glass( $n_2 = 1.5$ ) interface.

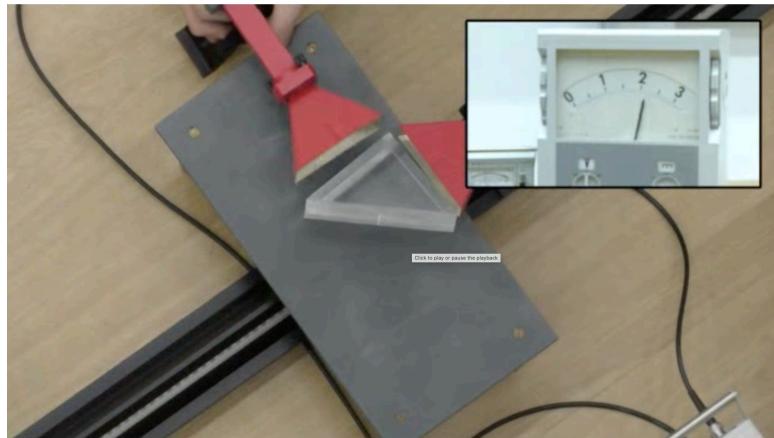


Figure 31.7: Total internal reflection of microwaves by a prism.

## 31.2 Transmission

As compared to the reflection, the transmission of light is more or less boring.

### Air to Glass

The transmission of light from air to glass only changes weakly the intensity, except for large angles of incidence. Also the phase angle is not changed at all upon transmission.

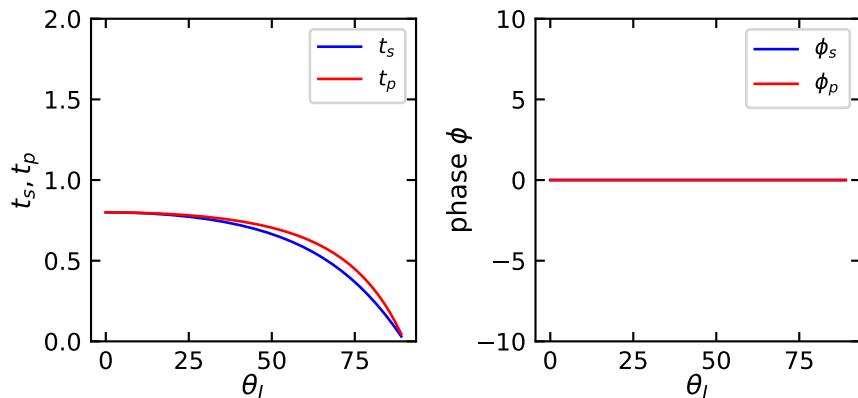


Figure 31.8: Fresnel coefficients (left) and phase (right) of the transmitted light observed for the transmission at an air( $n_1 = 1$ )/glass( $n_2 = 1.5$ ) interface.

### Glass to Air

The situation, however, becomes a bit more interesting the the case of the transmission of light from glass to air. The first thing we notice here is that the Fresnel coefficients for transmission  $t_s, t_p$  are both bigger than one and even grow to larger values at the edge of total internal reflection. This seems to say that the amplitude of the electric field that is transmitted is bigger than the one incident. This is in fact true, but does not violate energy conservation, but is rather resulting from the changed speed of light in the glass. We will have a closer look at the intensities down below.

Another thing that is occurring but not directly visible is the electric field amplitude in the region directly behind the glass interface in air. Even though all of the light is reflected, this does not mean that no electric field may exist on the transmission side. In fact, an electric field is present and will decay in amplitude exponentially. To see what is happing directly behind the interface we

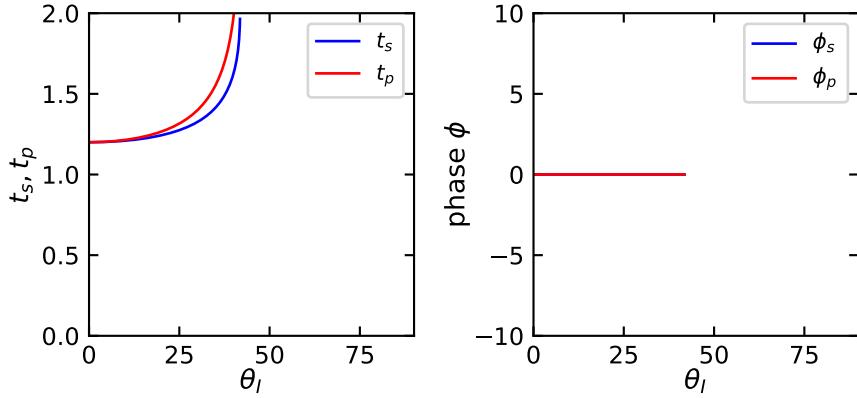


Figure 31.9: Fresnel coefficients (left) and phase (right) of the transmitted light observed for the transmission at a glass( $n_1 = 1.5$ )/air( $n_2 = 1.0$ ) interface.

What is interesting for the total internal reflection is now the electric field that exists on the side, where usually the transmission would go. Here is our coordinate system again such that we can have a look at the wavevectors in the different regions.

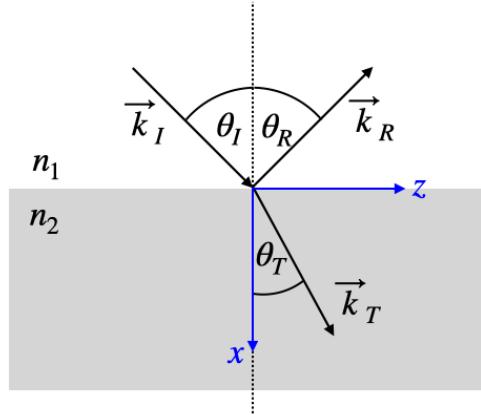


Figure 31.10: Coordinate system for wave at interfaces.

The wavevector in the transmission region is given by

$$\vec{k}_T = k_{Tx} \hat{e}_x + k_{Tz} \hat{e}_z$$

According to the condition that the tangential component of the wavevector is conserved we can write down

$$k_{Iz} = \vec{k}_I \cdot \hat{e}_z = \vec{k}_T \cdot \hat{e}_z = k_{Tz} = \frac{\omega}{v_1} \sin(\theta_I) \hat{e}_z$$

where  $\omega$  is the frequency of light and  $v_1$  the phase velocity in medium 1, which is in this case glass. Besides the tangential component of the wavevector we can also calculate the magnitude square of the wavevector that is transmitted, which is

$$|\vec{k}_T|^2 = \frac{\omega^2}{v_2^2}$$

Since the wavevector is in the x-z plane we may calculate the x-component from

$$\vec{k}_T \cdot \hat{e}_x = \pm \sqrt{|\vec{k}_T|^2 - (\vec{k}_T \cdot \hat{e}_z)^2} \quad (31.1)$$

$$= \pm \frac{\omega}{v_2} \sqrt{1 - \frac{v_2^2 \sin^2(\theta_I)}{v_1^2}} \quad (31.2)$$

$$= \pm \frac{\omega}{v_2} \sqrt{1 - \frac{n_1^2 \sin^2(\theta_I)}{n_2^2}} \quad (31.3)$$

The last equation will yield a negative value under the square root

$$\frac{n_1}{n_2} \sin(\theta_I) > 1$$

which is in fact the region, where total internal reflection occurs. If this is the case, the transmitted wavevector along the x-direction will be imaginary and we may write

$$k_{Tx} = \vec{k}_T \cdot \hat{e}_x = \pm i \frac{\omega}{v_2} \alpha$$

where  $\alpha$  is the given by

$$\alpha = \sqrt{\frac{n_1^2 \sin^2(\theta_I)}{n_2^2} - 1}$$

and is a positive number due to a little trick taking the negative value of the previous square root term. We may now use this wavevector with the x- and the z-component and insert that in our plane wave, which would be propagating in the transmission region (here air).

$$\vec{E}_{\text{trans}} = \vec{E}_T e^{i(\omega t - k_{Tx}x - k_{Tz}z)}$$

Inserting the results for the components of the wavevector obtained above yields

$$\vec{E}_{\text{trans}} = \vec{E}_T e^{i(\omega t - n_1 k_0 \sin(\theta_I) z)} e^{-n_2 k_0 \alpha x}$$

The imaginary -xcomponent of the wavevector is creating a real valued factor for the amplitude of the electric field in the transmission region. The electric field in this region is decaying exponentially with the distance  $x$  from the interface. Yet, the field is only oscillating along the interface but not in the x-direction. Such an exponentially decaying field without oscillation is called **evanescent wave**. Equivalent exponential decays do also exist in quantum mechanics, however, they concern in this case the decay of the probability to find a particle inside a region.

The image below shows the electric field (left) and the intensity (right) for the total internal reflection at a boundary between glass and air, which the exponential decay is nicely visible for the electric field. To calculate the intensity, we would have to calculate the Poynting vector. As it turns out, the Poynting vector is zero in the transmission region and hence there is no energy flow in the transmission region. Nevertheless, there is a time oscillating electric field and we may calculate the magnitude square of the electric field decay, which is shown in the right figure.

To gain a bit more insight into the lengthscale on which the electric field in Figure ?? is decaying, we may have a look at the prefactor in the above exponential term.

This prefactor denotes an inverse length scale, i.e.

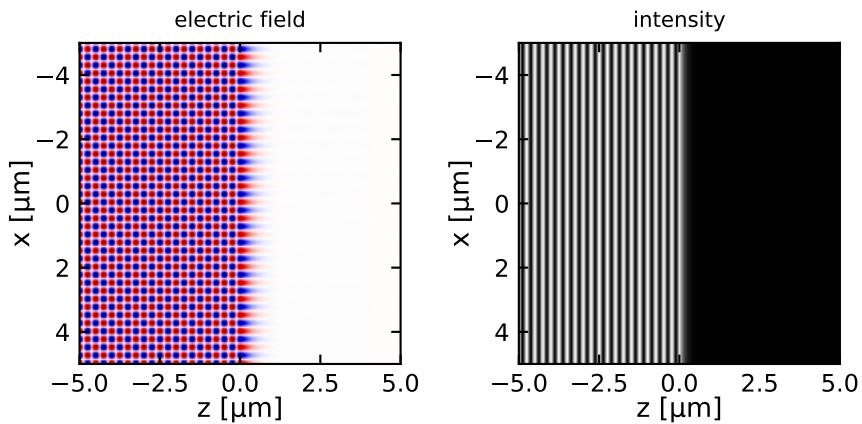


Figure 31.11: Electric field amplitude (left) and intensity (right) at a glass( $n_1 = 1.5$ )/air( $n_2 = 1.0$ ) interface.

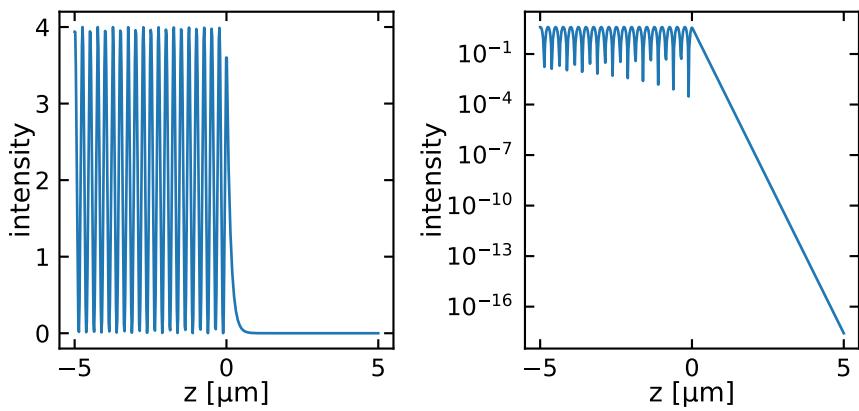


Figure 31.12: Electric field amplitude (left) and intensity (right) along the propagation direction at a glass( $n_1 = 1.5$ )/air( $n_2 = 1.0$ ) interface.

$$\delta = \frac{1}{n_2 k_0 \alpha}$$

and measures the distance on which the electric field decayed to  $1/e$ . The plot below shows that the decay length for the electric field amplitude quickly drops with increasing incident angle

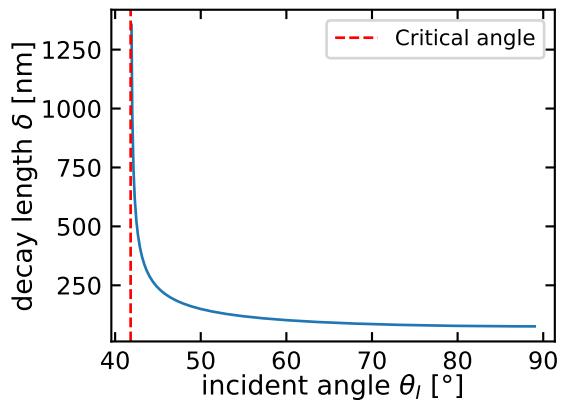


Figure 31.13: Decay length of the evanescent wave in air for total internal reflection at a glass/air interface (wavelength = 532nm).

The evanescent field in the transmission region can be converted back into a propagating wave if a second interface is brought in the range of the evanescent wave. This can be demonstrated easily with microwaves, since their wavelength is long.

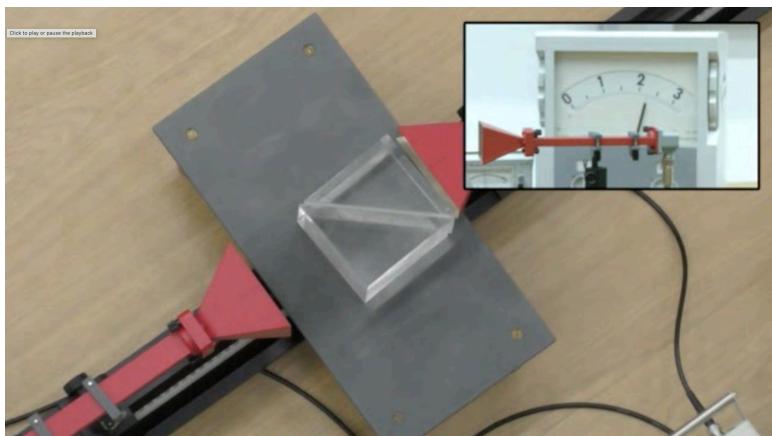


Figure 31.14: Total internal reflection of microwaves by a prism (left).

In the visible region, this decay length is on the order of a few 100 nanometers. Evanescent field play an important role for optical microscopy for example, where they are either used in total internal reflection fluorescence microscopy (TIRF), exciting only a tiny sheet above an interface and thus reducing fluorescence background. Other techniques use the evanescent field on tiny tips as sources for excitation of fluorescence or scattering (NSOM). Evanescent waves are also used for coupling light into small mechanical resonators to cool them to extremely low temperatures (opto-mechanics).

### 31.3 Intensities of Reflected and Transmitted Waves

So far we calculated the Fresnel coefficients, from which we may obtain the electric and magnetic field amplitudes of the reflected and transmitted waves. We now want to calculate the intensities. The Poynting vector of a

wave is given by

$$\vec{S} = \vec{E} \times \vec{H} = \frac{1}{\mu_r \mu_0} (\vec{E} \times \vec{B}) = \epsilon_r \epsilon_0 v^2 (\vec{E} \times \vec{B})$$

The magnitude of the Poynting vector is therefore given by

$$S = \epsilon_r \epsilon_0 v |\vec{E}|^2 = n^2 \epsilon_0 c |\vec{E}|^2$$

The intensity is then obtained from the integration over one cycle of oscillation, which finally results in

$$I = \frac{1}{2} n \epsilon_0 c |\vec{E}|^2$$

With the help of this equation it can be seen that the reflected intensity and the incident intensity contain the same refractive index, thus only the magnitude square of the Fresnel coefficients for the reflection is important giving

$$R = |r|^2$$

which is valid for either s- or p-polarization.

For the transmission we have to be a bit more careful as first of all the light propagates with different phase velocities (refractive index  $n$ ) and the cross section of a beam changes due to the refraction at the interface as shown in the image below.

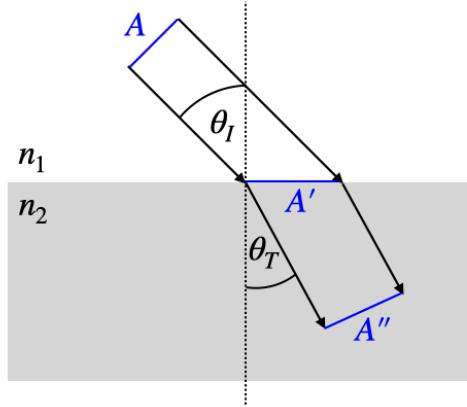


Figure 31.15: Relevant areas for the calculation of the intensities from the transmission Fresnel coefficients.

The two cross-sections behave as

$$\frac{A''}{A} = \frac{\cos(\theta_T)}{\cos(\theta_I)}$$

The intensities directly contain the refractive index and this

$$T = \frac{n_2 \cos(\theta_T)}{n_1 \cos(\theta_I)} |t|^2$$

As energy conservation is valid, the reflected energy and the transmitted energy have to sum up to the incident energy in nonabsorbing media, which states that

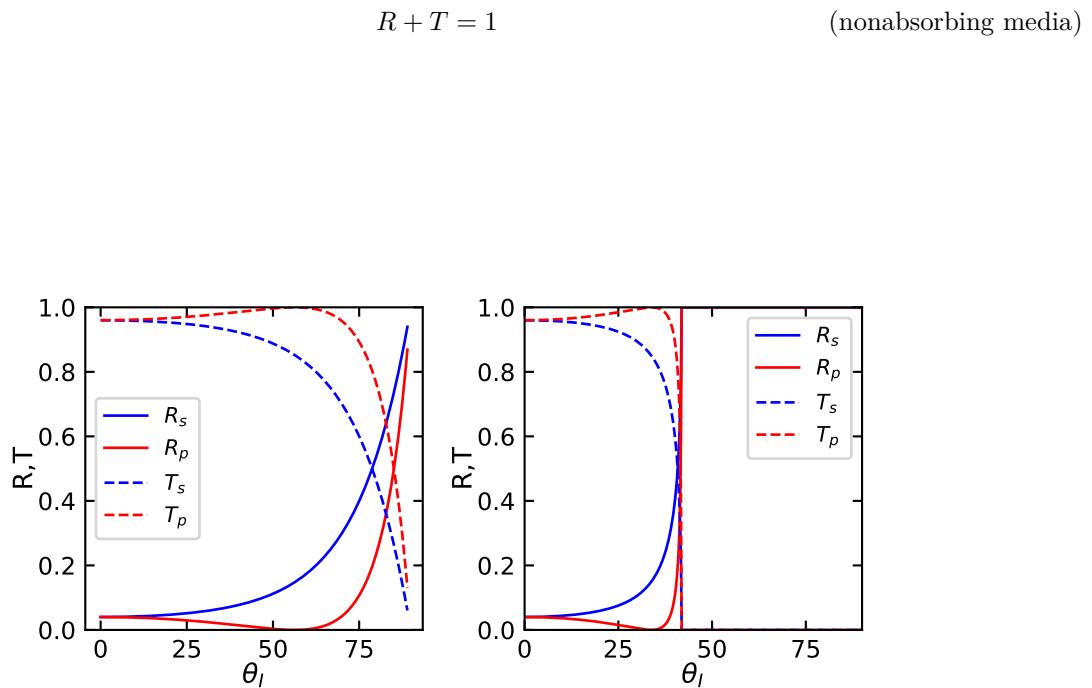


Figure 31.16: Reflected and transmitted intensities for air to glass (left) and glass to air (right) interfaces.

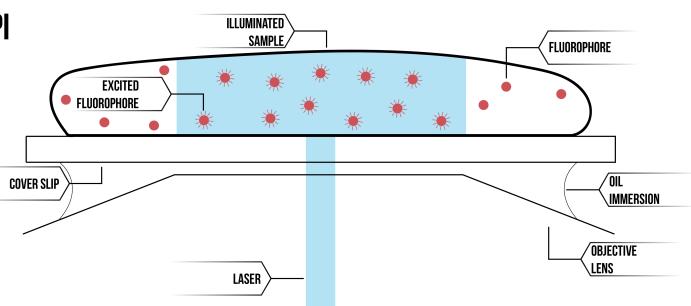
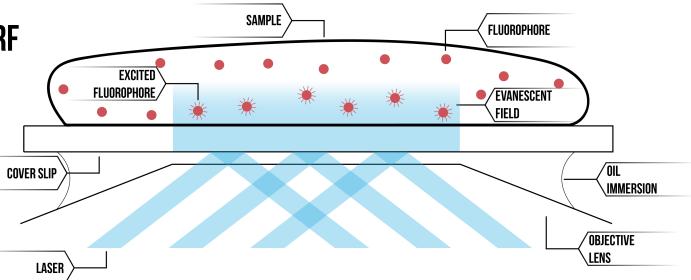
The displayed reflection and transmission coefficients will reappear in quantum mechanics for the tunneling effect on a potential energy barrier.

### **i** Total Internal Reflection Fluorescence (TIRF) Microscopy

TIRF microscopy is a powerful imaging technique that exploits the properties of evanescent waves to achieve exceptional signal-to-noise ratio in fluorescence imaging. When light undergoes total internal reflection at a glass-water interface (typically a microscope coverslip-sample interface), the resulting evanescent field only penetrates about 100-200 nm into the sample. This shallow excitation depth provides an excellent optical sectioning capability, as only fluorophores within this thin layer are excited. This makes TIRF microscopy particularly well-suited for studying processes at or near the cell membrane, such as:

- Membrane protein dynamics
- Endocytosis and exocytosis
- Cell adhesion
- Single-molecule tracking

The signal-to-background ratio in TIRF microscopy can be 100-1000 times better than conventional fluorescence microscopy because out-of-focus fluorescence is virtually eliminated. The technique has become indispensable in cell biology, especially for studying membrane-associated processes in living cells. The exponential decay of the evanescent field intensity with distance from the interface provides an additional advantage: the fluorescence intensity can be used to estimate the distance of fluorescent molecules from the surface with nanometer precision.

**EPI****TIRF**

# Chapter 32

## Electromagnetic Waves in Metals

Metals have free charges, which modify the propagation of light. Their influence is large and gives metals their characteristic reflection and color. In this case, we have to include the free charge density  $\rho_f$  and the current density  $\vec{j}$  in the Maxwell equations and derive the wave equation accordingly. Instead of modifying the wave equation, we would like to go a different way here. We would like to derive the dielectric function for metals based on a microscopic description, which is only approximate but captures some basic features.

### 32.1 Drude Model

The model we would like to put forward is the Drude model. It considers the motion of a charge in the electromagnetic field of a wave. As compared to our previous attempts on bound electrons in atoms, there is no direct restoring force for the free charges in the metal. The equation of motion therefore looks as

$$m_e \frac{d^2 \vec{r}}{dt^2} + m_e \Gamma \frac{d\vec{r}}{dt} = e \vec{E}_0 e^{i\omega t}$$

Here, the coefficient  $\Gamma$  is given by

$$\Gamma = \frac{v_f}{l} = \frac{1}{\tau}$$

where  $v_f$  is the Fermi velocity and  $l$  is the mean free path electrons travel before colliding with a lattice site inside the metal. To obtain an expression for the refractive index, we use again the ansatz

$$\vec{r} = \vec{r}_0 e^{i\omega t}$$

from which we obtain an expression for  $\vec{r}_0$ . With this solution we can again calculate the polarization density and finally also the dielectric function

$$\epsilon_r(\omega) = n^2 = 1 - \frac{\omega_p^2}{\omega^2 - i\Gamma\omega}$$

with

$$\omega_p = \sqrt{\frac{Ne^2}{\epsilon_0 m}}$$

which is called the **plasma frequency**. It is a characteristic for each metal, since the density of free charges in the metal enters the equation. It is located in the UV region of the electromagnetic spectrum.

Parameter	Symbol	Gold Value	Silver Value	Copper Value	Units
Free electron density	$N$	$5.9 \times 10^2$	$5.86 \times 10^2$	$8.47 \times 10^2$	$\text{m}^{-3}$
Electron mass	$m_e$	$9.109 \times 10^{-31}$	$9.109 \times 10^{-31}$	$9.109 \times 10^{-31}$	kg
Elementary charge	e	$1.602 \times 10^{-1}$	$1.602 \times 10^{-1}$	$1.602 \times 10^{-1}$	C
Resulting plasma frequency	$\omega_p$	$13.8 \times 10^1$	$13.7 \times 10^1$	$16.5 \times 10^1$	rad/s
Frequency	$f$	$2.18 \times 10^1$	$2.18 \times 10^1$	$2.62 \times 10^1$	Hz
Plasma wavelength	$\lambda_p$	138	138	114	nm

The complex refractive index  $n$  is composed of two parts: the real part  $n_r$ , which describes the phase velocity of light in the material, and the imaginary part  $\kappa$  (kappa) which describes the absorption. These are related by:

$$n = n_r - i\kappa$$

Since the dielectric function  $\epsilon$  is equal to the square of the refractive index  $n$ , we can expand this relationship to separate the real and imaginary components:

$$n^2 = n_r^2 - \kappa^2 - 2in_r\kappa = \epsilon' + i\epsilon''$$

Here,  $\epsilon'$  represents the real part of the dielectric function, which determines how much polarization occurs in response to an applied electric field, while  $\epsilon''$  represents the imaginary part, which determines how much energy is absorbed by the material. From the Drude model, we can express these components explicitly:

$$\epsilon' = n_r^2 - \kappa^2 = 1 - \frac{\tau^2(\omega^2 - \omega_p^2)}{1 + \omega^2\tau^2}$$

$$\epsilon'' = 2n_r\kappa = \frac{\omega_p^2\tau}{\omega(1 + \omega^2\tau^2)}$$

These equations directly connect the optical properties ( $n_r$  and  $\kappa$ ) to the physical parameters of the metal like the plasma frequency  $\omega_p$  and the relaxation time  $\tau$ . When  $\epsilon'$  is negative, which occurs below the plasma frequency, the metal reflects light strongly. When  $\epsilon''$  is large, the metal absorbs light efficiently. Together, these components fully characterize how the metal interacts with electromagnetic radiation.

The figure above shows the dielectric function for Gold as determined from the parameters  $\omega_p = 13.8 \times 10^{15} \text{ s}^{-1}$  and  $\Gamma = 1.075 \times 10^{14} \text{ s}^{-1}$ . The real value of the dielectric function is negative over the whole visible range stating that there is no propagating wave inside the metal. It just becomes positive below the plasma frequency. The imaginary part increases continuously with the wavelength.

## Conductivity

Als we calculate the position  $\vec{r}$  of the electrons in the oscillating electric field, we may also obtain its velocity, which helps us to calculate a new quantity, which is the conductivity  $\sigma$  of the metal. This conductivity will be frequency dependent as well as the charges will not follow the electric field oscillations in the same way at different frequencies. The conductivity is following from Ohm's law, which is given by

$$\vec{j} = \sigma \vec{E} = Ne\vec{v}$$

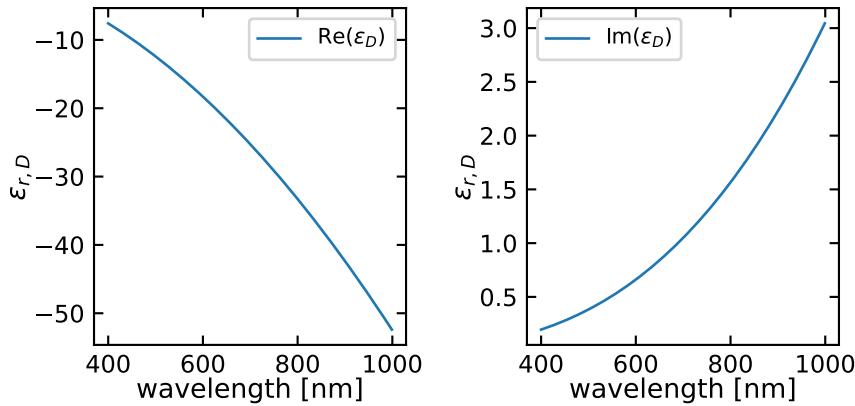


Figure 32.1: Real (left) and imaginary part (right) of the dielectric function according to the Drude model. The dashed line in the left plot corresponds to the plasma frequency. The calculations are for Gold.

where we explicitly wrote the current density on the right side. Following this relation, the conductivity is obtained as

$$\sigma = \epsilon_0 \omega_p^2 \frac{\tau(1 + i\omega\tau)}{1 + \omega^2\tau^2}$$

from the Drude model and we may express the real and imaginary parts of the dielectric function also in relation to the conductivity by

$$\epsilon' = 1 - \frac{\mathcal{R}(\sigma)\tau}{\epsilon_0}$$

$$\epsilon'' = \frac{\mathcal{I}m(\sigma)}{\epsilon_0 \omega^2 \tau}$$

which readily tells us, that we obtain information on the conductivity at high frequencies from optical measurements.

## 32.2 Optical Properties of Metals

### Reflectivity of Metals

The reflectivity of a metal is closely related to its dielectric function. The reflectivity  $R$  of a metal can be calculated from the complex dielectric function  $\epsilon$  using the Fresnel equations:

$$R = \left| \frac{n - 1}{n + 1} \right|^2$$

where  $n = \sqrt{\epsilon}$  is the complex refractive index.

The plot below shows the square of complex refractive index  $n$  for gold according to the Drude model. Interestingly the square of the real part of the refractive index is negative below the plasma frequency. This means that the wave is evanescent and cannot propagate in the metal. This is the reason why metals are opaque in the visible range.

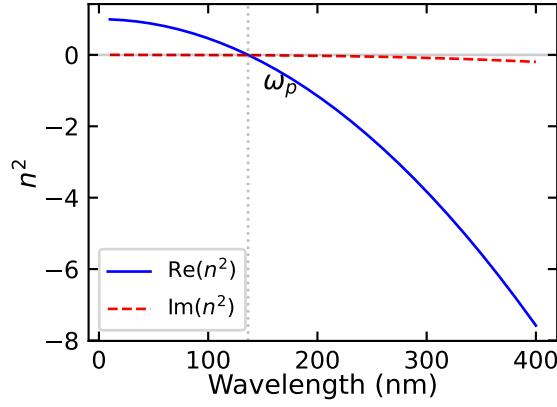


Figure 32.2: Real and imaginary parts of  $n^2$  for gold according to the Drude model. The dashed line indicates the plasma frequency.

Table 32.2: Real and imaginary components of the refractive index squared ( $n^2$ ) at different wavelengths

Wavelength (nm)	$\text{Re}(n^2)$	$\text{Im}(n^2)$
10.0	0.99	-0.00
88.1	0.58	-0.00
166.2	-0.48	-0.01
244.2	-2.20	-0.04
322.3	-4.57	-0.10

From the square of the refractive index, we can calculate the reflectivity of the metal. The plot below shows the reflectivity of gold with and without damping. Without damping, the reflectivity is exactly 1 below the plasma frequency and drops sharply above it. The damping broadens the reflectivity curve and reduces the reflectivity below the plasma frequency.

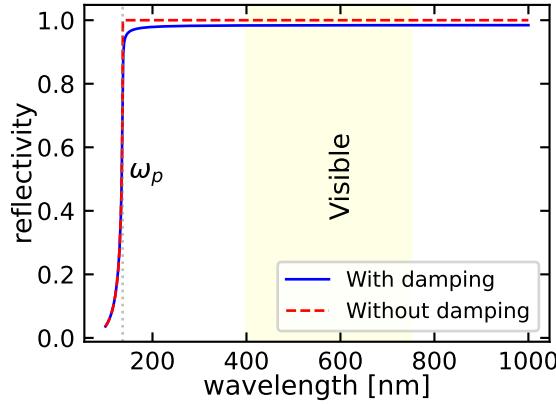


Figure 32.3: Comparison of metal reflectivity with and without damping. Without damping ( $\gamma = 0$ ), the reflectivity is exactly 1 below the plasma frequency and drops sharply above it.

Table 32.3: Comparison of reflectivity ( $R$ ) values with and without damping effects at different wavelengths

Wavelength (nm)	$R$ (with damping)	$R$ (no damping)
100.0	0.036	0.036
280.2	0.982	1.000
460.4	0.984	1.000
640.5	0.984	1.000
820.7	0.984	1.000

## Dispersion Relation in Metals

To understand how electromagnetic waves propagate in metals, we need to derive the dispersion relation. Starting from Maxwell's equations, the dielectric function can be derived from the Drude model, and it is given by:

$$-k^2 \vec{E} + \frac{\omega^2}{c^2} \vec{E} + i\omega\mu_0\sigma\vec{E} = 0$$

This leads to the dispersion relation:

$$k^2 = \frac{\omega^2}{c^2} \left( 1 + i\frac{\sigma}{\epsilon_0\omega} \right) = \frac{\omega^2}{c^2} \epsilon_r(\omega)$$

Using the Drude model expression for  $\epsilon_r(\omega)$ , we can write:

$$k^2 = \frac{\omega^2}{c^2} \left( 1 - \frac{\omega_p^2}{\omega^2 - i\Gamma\omega} \right)$$

This dispersion relation shows two important regimes:

1. For  $\omega < \omega_p$ : The wave vector  $k$  becomes imaginary, meaning the waves are evanescent and decay exponentially into the metal. This explains why metals are reflective at visible frequencies.
2. For  $\omega > \omega_p$ : The wave vector  $k$  is real, and electromagnetic waves can propagate through the metal. This typically occurs in the ultraviolet region for most metals.

The penetration depth (skin depth)  $\delta$  of the electromagnetic wave into the metal can be calculated from the imaginary part of  $k$ :

$$\delta = \frac{1}{\text{Im}(k)}$$

This skin depth is typically on the order of tens of nanometers in the visible region, explaining why metals are opaque even in thin films.

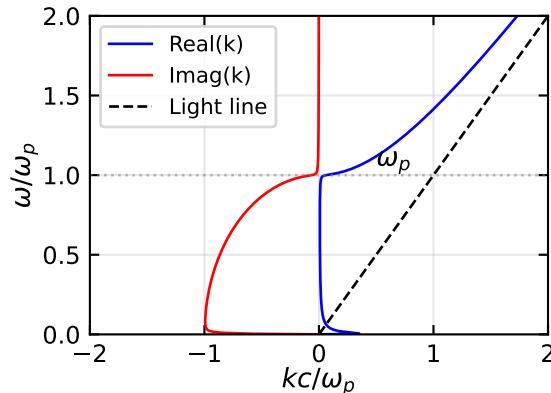


Figure 32.4: Dispersion relation for electromagnetic waves in a metal according to the Drude model. The light line (dashed) represents the dispersion in vacuum. Note how below the plasma frequency, no propagating waves exist in the metal.

The figure above shows the dispersion relation for electromagnetic waves in a metal according to the Drude model. Several important features can be observed:

1. Below the plasma frequency ( $\omega < \omega_p$ ), the real part of  $k$  is zero while the imaginary part is large, indicating evanescent waves that decay exponentially into the metal.
2. Above the plasma frequency ( $\omega > \omega_p$ ), the real part of  $k$  becomes dominant, allowing wave propagation through the metal.

3. The light line (dashed) represents the dispersion relation in vacuum ( $\omega = ck$ ). The metal's dispersion relation deviates significantly from this line, especially near and below the plasma frequency.
4. The presence of damping ( $\gamma$ ) leads to a smooth transition around  $\omega_p$  rather than an abrupt change.

This behavior explains why metals are highly reflective at frequencies below the plasma frequency (typically in the visible range) but can become transparent at higher frequencies in the ultraviolet region.

In the simple interpretation of the Drude model, we often say that below the plasma frequency, waves cannot propagate in the metal because the dielectric function is negative. However, this is only strictly true when we ignore damping ( $\Gamma = 0$ ).

When we include damping ( $\Gamma \neq 0$ ), the dielectric function becomes complex:

$$\epsilon_r(\omega) = 1 - \frac{\omega_p^2}{\omega^2 - i\Gamma\omega}$$

In the low-frequency limit ( $\omega \ll \omega_p$ ), this becomes:

$$\epsilon_r(\omega) \approx 1 - \frac{\omega_p^2}{-i\Gamma\omega} \approx -\frac{\omega_p^2}{-i\Gamma\omega} = \frac{\omega_p^2}{\Gamma\omega} i$$

The wave vector is then:

$$k = \frac{\omega}{c} \sqrt{\epsilon_r} \approx \frac{\omega}{c} \sqrt{\frac{\omega_p^2}{\Gamma\omega} i} = \frac{\omega}{c} (1+i) \sqrt{\frac{\omega_p^2}{2\Gamma\omega}}$$

This shows that both the real and imaginary parts of  $k$  are non-zero and equal in magnitude in this limit. This represents a heavily damped wave that can propagate a short distance into the metal.

The wave in the metal has the form:

$$E(x) = E_0 e^{ikx} = E_0 e^{i\text{Re}(k)x} e^{-\text{Im}(k)x}$$

The skin depth is then:

$$\delta = \frac{1}{\text{Im}(k)} = \frac{c}{\omega} \sqrt{\frac{2\Gamma\omega}{\omega_p^2}}$$

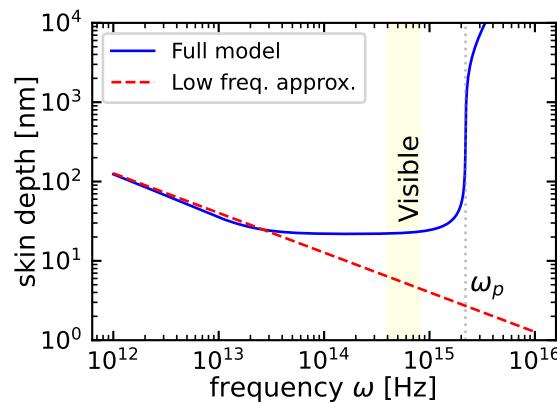


Figure 32.5: Skin depth as a function of frequency for a metal (Gold parameters). The dashed line shows the typical one over square root of frequency behavior at low frequencies.

Table 32.4: Penetration depth (delta) comparison between full model and low frequency approximation at different frequencies

freq (Hz)	delta_full (nm)	delta_low (nm)
1.00e+12	123.63	127.08
6.32e+12	44.97	50.54
4.00e+13	23.13	20.10
2.53e+14	21.91	8.00
1.60e+15	31.64	3.18

## Skin Depth

At low frequencies, the skin depth can be approximated as:

$$\delta \approx \sqrt{\frac{2}{\omega_p^2 \Gamma}} \frac{c}{\sqrt{\omega}} = \sqrt{\frac{2}{\sigma_0 \mu_0 \omega}}$$

where  $\sigma_0 = \frac{\omega_p^2 \epsilon_0}{\Gamma}$  is the DC conductivity. This shows the characteristic  $\delta \propto 1/\sqrt{\omega}$  behavior.

Some important observations from the plot:

1. At low frequencies (below the plasma frequency), the skin depth follows the  $1/\sqrt{\omega}$  dependence
2. For visible light ( $\omega \approx 10^{14} - 10^{15} \text{ Hz}$ ), the skin depth is typically tens of nanometers
3. Above the plasma frequency, the skin depth increases rapidly as the metal becomes transparent
4. For Gold at room temperature and visible frequencies, the skin depth is approximately 20-30 nm

This small skin depth explains why:

- Thin metal films ( $>100 \text{ nm}$ ) are opaque
- Surface effects dominate in metals at optical frequencies
- Metal nanoparticles can support localized surface plasmons
- Why RF shielding requires only thin metal layers

The skin effect is crucial in many applications:

- RF and microwave engineering
- Electromagnetic shielding
- Design of electrical conductors for AC applications
- Plasmonic devices

## 32.3 Plasmonics

One of the exciting fields of research in modern optics is plasmonics. Plasmons are collective oscillations of free electrons in a metal that can interact strongly with light. These interactions lead to fascinating phenomena, such as enhanced light-matter interactions, subwavelength imaging, and sensing applications.

### Localized Surface Plasmons

In metal nanoparticles, the confined electron oscillations create localized surface plasmons (LSPs). These are non-propagating excitations of the conduction electrons coupled to the electromagnetic field. The resonance condition depends strongly on the particle size, shape, and dielectric environment. A striking demonstration of LSPs can be observed in solutions of gold nanoparticles, where the interaction between light and the collective electron oscillations creates vivid colors.



Figure 32.6: Backscattering and transmission through a 65 nm gold nanoparticle solution. The backscattering is green due to the surface plasmon resonance of the conduction band electrons, while the transmission is red, since the green light is removed by scattering.

## Surface Plasmon Polaritons

Surface plasmon polaritons (SPPs) are electromagnetic excitations propagating along a metal-dielectric interface. These waves result from the coupling of the electromagnetic fields to oscillations of the conductor's electron plasma. The dispersion relation for SPPs is given by:

$$k_{SPP} = \frac{\omega}{c} \sqrt{\frac{\epsilon_m \epsilon_d}{\epsilon_m + \epsilon_d}}$$

where  $\epsilon_m$  and  $\epsilon_d$  are the dielectric functions of the metal and dielectric, respectively. This unique dispersion relation allows SPPs to concentrate light into subwavelength volumes, enabling applications in sensing, waveguiding, and imaging beyond the diffraction limit.

### Surface Plasmon Polariton Sensor

A surface plasmon polariton sensor is a device that exploits the sensitivity of SPPs to changes in the local refractive index. By monitoring the resonance condition of the SPPs, one can detect minute changes in the dielectric environment near the metal surface. This principle forms the basis of label-free biosensing, where the binding of biomolecules to the metal surface can be detected in real-time.

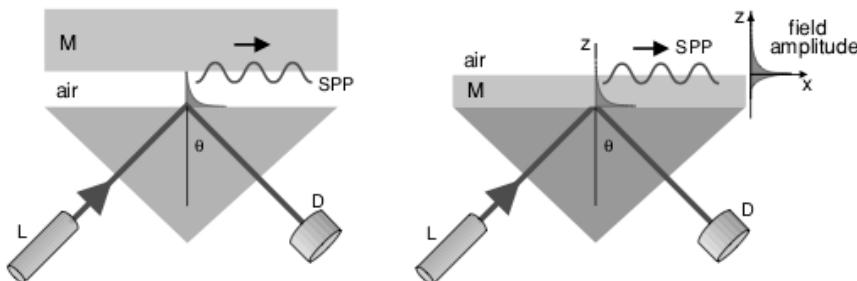


Figure 32.7: Two different excitation schemes of surface plasmon polaritons for sensing applications. The prism coupling method (left) uses total internal reflection to excite SPPs, while the Kretschmann configuration (right) involves a thin metal film on a dielectric substrate.

To observe SPPs, one can use a prism coupling method or the Kretschmann configuration. In both cases, the

resonance condition is sensitive to changes in the refractive index near the metal surface, making SPP sensors highly versatile and sensitive.

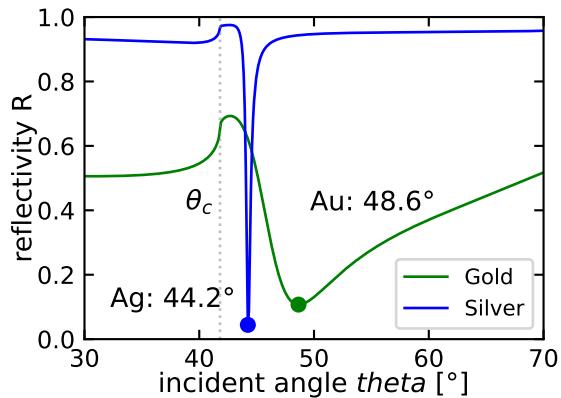


Figure 32.8: Comparison of reflectivity versus incident angle for 50 nm gold and silver films on glass at 532 nm. Silver shows a sharper and deeper resonance due to lower losses.

Table 32.5: Comparison of surface plasmon resonance (SPR) characteristics for gold and silver films at 532 nm wavelength

Metal	SPR Angle (°)	Minimum Reflectivity	Dielectric Constant ( ) at 532 nm
Gold	48.6	0.107	-4.8 + 2.4i
Silver	44.2	0.044	-11.7 + 0.4i

## Historical and Modern Applications

The use of plasmonic effects in materials dates back to ancient times, though the underlying physics was not understood. Perhaps the most famous example is the Lycurgus Cup from the 4th century AD. This remarkable artifact appears green in reflected light but red in transmitted light due to the presence of gold and silver nanoparticles embedded in the glass matrix.

Modern applications of plasmonics have expanded far beyond decorative uses. Surface plasmon resonance sensors enable highly sensitive biochemical detection by monitoring changes in the local refractive index near metal surfaces. Plasmonic waveguides can guide light in dimensions far below the diffraction limit, promising applications in next-generation optical computing and telecommunications. In medicine, plasmonic nanoparticles are being developed for targeted therapy and diagnostic imaging.

The field of metamaterials leverages plasmonic effects to create artificial materials with properties not found in nature, such as negative refractive indices and perfect lensing. These developments continue to push the boundaries of what's possible in optics and photonics, leading to innovations in solar energy harvesting, quantum information processing, and ultrasensitive chemical detection.



Figure 32.9: The Lycurgus Cup demonstrates the dramatic optical effects of metallic nanoparticles. When viewed in reflected light (left), the cup appears green, while in transmitted light (right), it glows ruby red due to the surface plasmon resonance of embedded gold and silver nanoparticles.

# Chapter 33

## Anisotropic Materials

### 33.1 Light propagation

So far we have only discussed isotropic materials, meaning that the speed of light was independent by the direction of light propagation in the material. Often, the light propagation is, however, not isotropic as the underlying materials have an anisotropic structure.

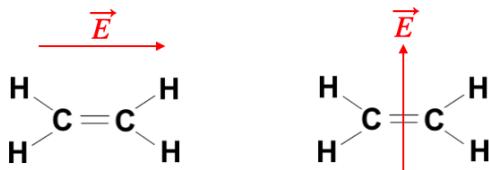


Figure 33.1: Ethene molecules as an example for an anisotropically polarizable object. The red arrows denote the electric field directions.

Just consider the above molecule  $C_2H_4$  (Ethen), where the two carbon atoms are bound by a double bond. In an external electric field aligned along that bond, we certainly expect the electrons to be more easily displaced over a larger distance, when the field is aligned parallel to the double bond. The dipole induced will therefore depend on the orientation of molecule and electric field. We therefore need to express the electronic polarizability  $\alpha$  for the calculation of the dipole moment  $\vec{p} = \alpha \vec{E}$  as a tensor  $\vec{\alpha}$ .

$$\vec{\alpha} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix}$$

Consequently also the susceptibility  $\chi$ , the dielectric function  $\epsilon_r$  and refractive index  $n$  will be tensors in anisotropic materials. The entries of these tensors now depend on the choice of the coordinate system, since the polarisabilities are connected to the structures. Accordingly, there is also coordinate system in which the tensor becomes completely diagonal.

$$\vec{\epsilon}_r = \begin{pmatrix} \epsilon_{11} & 0 & 0 \\ 0 & \epsilon_{22} & 0 \\ 0 & 0 & \epsilon_{33} \end{pmatrix} \quad (\text{dielectric tensor})$$

This coordinate systems is the principle coordinate system of the tensor. The electric displacement field  $\vec{D}$  is related to the electric field  $\vec{E}$  by

$$\vec{D} = \epsilon_0 \vec{\epsilon}_r \vec{E}$$

and in the principle coordinate system of the tensor we obtain

$$\begin{pmatrix} D_1 \\ D_2 \\ D_3 \end{pmatrix} = \begin{pmatrix} \epsilon_0 \epsilon_{11} E_1 \\ \epsilon_0 \epsilon_{22} E_2 \\ \epsilon_0 \epsilon_{33} E_3 \end{pmatrix}$$

From this it becomes evident, that the displacement field  $\vec{D}$  is not anymore parallel to the electric field  $\vec{E}$ . The dielectric tensor and refractive index tensor are directly related through  $n_{ij} = \sqrt{\epsilon_{ij}}$ . In the principal coordinate system, where both tensors are diagonal, the principal refractive indices are simply the square roots of the principal dielectric constants. This relationship allows us to visualize the optical properties through the index ellipsoid. This also means now, that the wavevector  $\vec{k}$  is now perpendicular to the displacement field and not to the electric field. The magnetic field is still directed along the same direction and thus the Poynting vector is not parallel to the wavevector. Energy and phasefronts flow in different directions. The sketch below indicates the relations of the vectors.

### ! Energy Flow in Anisotropic Materials

In anisotropic materials, the direction of energy flow (given by the Poynting vector  $\vec{S} = \vec{E} \times \vec{H}$ ) generally differs from the direction of phase propagation (given by the wavevector  $\vec{k}$ ). This is because  $\vec{E}$  and  $\vec{D}$  are not parallel in anisotropic media. While the wavevector  $\vec{k}$  determines the direction of phase velocity and is perpendicular to the wavefronts, the energy actually flows in the direction of the Poynting vector  $\vec{S}$ . This direction is always perpendicular to the k-surface at the point where  $\vec{k}$  intersects it. This explains why a single incident beam can split into two beams (ordinary and extraordinary) traveling in different directions through the material, even though their wavefronts remain parallel.

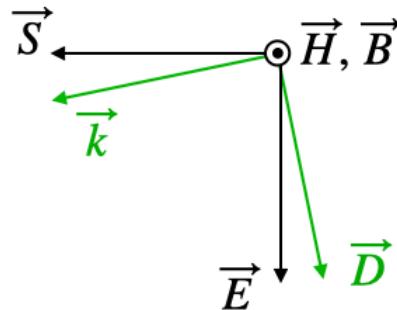


Figure 33.2: Sketch of the relations between the electric field  $\vec{E}$ , the displacement field  $\vec{D}$ , the wavevector  $\vec{k}$  and the magnetic field  $\vec{H}$ .

The tensorial nature of the dielectric function has now consequences for the propagation of light. In general different symmetries are considered. If we have for example two of the diagonal elements be equal to each other

$$\epsilon_{11} = \epsilon_{22} \neq \epsilon_{33}$$

then we call the material **uniaxial**. The system has one optical axis. In case

$$\epsilon_{11} \neq \epsilon_{22} \neq \epsilon_{33} \neq \epsilon_{11}$$

we call the material **biaxial** and the material has two optical axes. We will have a look at the meaning of an optical axis a bit later.

A tensor as the dielectric tensor (similar to the tensor of inertia) can be geometrically represented by an ellipsoid with three different half axes as depicted below. This can be done for the refractive index or the dielectric tensor.

When doing so with the refractive index, we obtain an index ellipsoid, where the half-axes correspond to the three refractive indices  $n_1 = \sqrt{\epsilon_{11}}$ ,  $n_2 = \sqrt{\epsilon_{22}}$  and  $n_3 = \sqrt{\epsilon_{33}}$ .

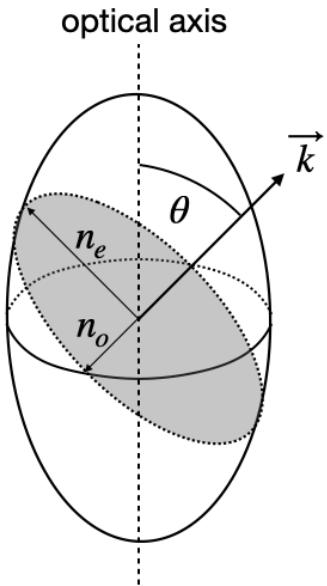


Figure 33.3: Index ellipsoid for a uniaxial material and a construction of the refractive indices. The  $k$ -vector cuts the ellipsoid in one point. The plane perpendicular to the  $k$ -vector intersects the ellipsoid in an ellipse. The long and short axis of the ellipse correspond to the normal modes of the material.

To find out the relevant refractive indices for a certain propagation direction one undertakes the following steps:

- chose a direction of light propagation with the direction of the wavevector  $\vec{k}$
- construct a plane perpendicular to the wavevector
- take the intersection of this plane with the ellipsoid, which is in general an ellipse
- the ellipse has a long and a short axis, which correspond to the direction of the so-called normal modes
- a  $\vec{D}$  field along those half-axes propagates with the corresponding refractive index  $n_o$  (ordinary),  $n_e$  (extra\_ordinary)

### Normal Modes in Anisotropic Materials

Normal modes are specific directions of the displacement field  $\vec{D}$  for which the light wave can propagate through the anisotropic medium without changing its polarization state. These modes correspond to the major and minor axes of the ellipse formed by intersecting the index ellipsoid with a plane perpendicular to the propagation direction. For a given propagation direction, there are always two orthogonal normal modes, each associated with a specific refractive index (ordinary or extraordinary). Any other polarization state can be expressed as a superposition of these two normal modes, but will generally change as the wave propagates through the medium due to the different phase velocities of the two modes.

An **optical axis** is now a propagation direction, for which the refractive index does not depend on the direction of the electric field. For a biaxial material, there are two distinct directions for a propagation, while there is only one for uniaxial.

Considering now in more detail the propagation of light as a function of the direction of light propagation one obtains a dispersion relation (how the wavenumber depends on the direction for a given frequency of light). This surface is the so-called **k-surface**, which in general consists of 2 sheets. This k-surface is obtained from Maxwell's equations

$$\vec{k} \times \vec{H} = -\omega \vec{D}$$

$$\vec{k} \times \vec{E} = \omega \mu_0 \vec{H}$$

which leads to

$$\vec{k} \times (\vec{k} \times \vec{E}) + \omega^2 \mu_0 \epsilon_0 \vec{\epsilon}_r \vec{E} = 0$$

This equation leads to a system of equations for the components of the wavevector ( $\vec{k} = \{k_1, k_2, k_3\}$ ) that define the k-surface for a given frequency.

The k-surface represents all possible wavevectors at a given frequency in the material. For a uniaxial crystal, this surface consists of two sheets: a sphere representing the ordinary waves and an ellipsoid representing the extraordinary waves. Where these surfaces intersect defines the optical axis, along which the ordinary and extraordinary waves travel with the same phase velocity. This geometric representation helps visualize how light propagates in different directions through the crystal.

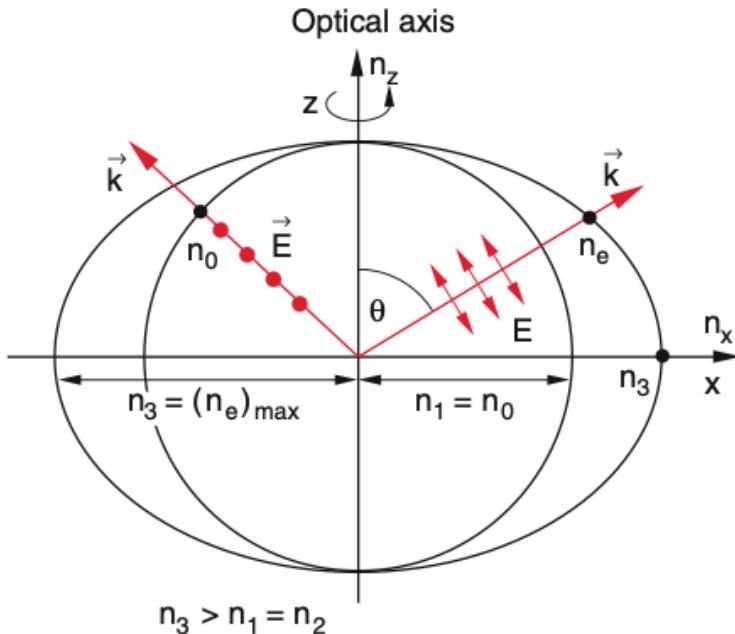


Figure 33.4: K-surface for a uniaxial material. The circular surface corresponds to the ordinary refractive index and the elliptical surface to the extra-ordinary refractive index. The polarization of the electric field is indicated.

If we select now a specific direction of propagation for the light with the direction of  $\vec{k}$ , the intersections of the k-surface with the k-vector deliver the solutions for the refractive indices. From the figure above, it becomes obvious, that the electric field vector must be perpendicular to the plane in the case of the ordinary refractive index. Only in this case, the wavevector can be rotated in the above sketch such that the direction of the electric field does not change. If the electric field is within the plane of the above sketch, the field direction with respect to the crystal structure changes and therefore the refractive index changes.

The ordinary ray corresponds to light where the electric field is perpendicular to the plane containing the optical axis and the direction of propagation, while the extraordinary ray corresponds to light where the electric field has a component parallel to this plane. The association with s- or p-polarization depends on the orientation of the crystal's optical axis relative to the plane of incidence.

Let's discuss the sketches above, which show the general case of **birefringence**. The image on the left side shows an interface between air on the bottom and the anisotropic material on the top. The wavevector is incident from below and is normal to the surface. Its direction will intersect both k-surfaces and thereby select the ordinary refractive index and the extra-ordinary refractive index for the two different polarization directions.

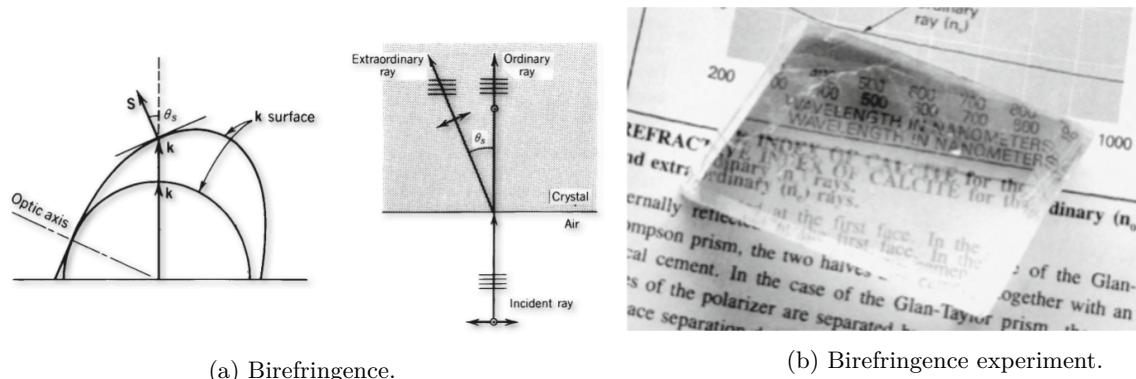


Figure 33.5: Birefringence. The left side shows the general case of birefringence. The right side shows the birefringence in a calcite crystal.

Both waves propagate in the same direction, so their wavefronts are perpendicular to the k-vectors as shown in the middle sketch. Yet, the Poynting vectors go in different directions. The Poynting vector, which points in the direction of energy propagation, is always perpendicular to the k-surface, since the k-surface is an isoenergy surface (*think about that for a moment*). This means that the beam is split into two beams with parallel wavefronts as indicated. This phenomenon is called birefringence and separates two orthogonal polarizations into two beams as indicated.

This effect is probably best visible, when you put a birefringent material like the common calcite crystal on a newspaper or book. You will immediately see the birefringence in the double images as in the figure above to the right.

The different refractive index for both polarizations make those materials very suitable for creating polarizing optical elements like beam splitters. A selection of different polarizing beam splitters is shown in the image below.

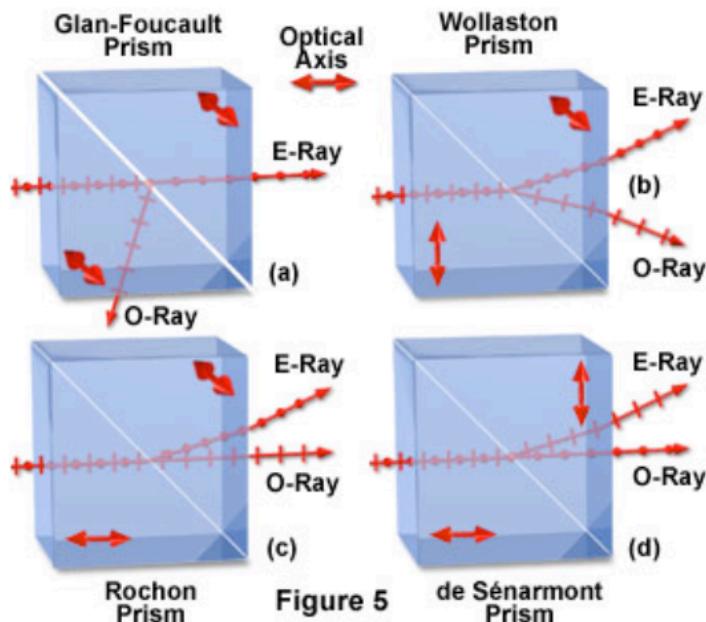


Figure 33.6: Polarizing beam splitters. Glan-Foucault prism, Wollaston prism, Rochon prism and de Séarmont prism. All of them are based on birefringence and consist of two prisms with different refractive indices for the two polarizations.

Birefringence is often observed in mechanically stressed materials like quickly cooled glasses or stretched polymer foils. The latter is demonstrated in the images below, where we used

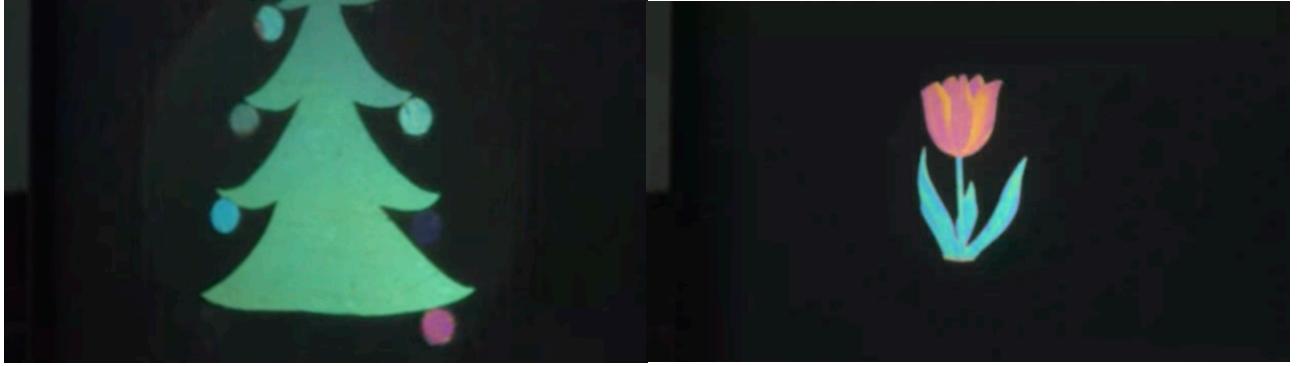


Figure 33.7: Scotch tape art. Scotch tape christmas tree and tulip between crossed polarizers showing their birefringent behavior.

## 33.2 Wave retarders

Wave retarders are essential optical components that manipulate the polarization state of light by exploiting the different phase velocities of ordinary and extraordinary waves in anisotropic materials. These components are fundamental in many optical systems, from liquid crystal displays to laser systems and optical communications. Anisotropic materials can now be cut in specific way, for example such that the optical axis is parallel to the interface as in the picture below. Under these circumstances at normal incidence both wavevectors and both Poynting vectors of the polarizations propagate along the same direction. Yet, the p-polarization has a lower phase velocity (long k-vector) than the s-polarization. If now the incident light has now electric field components parallel and perpendicular to the optical axis which are of same amplitude, then one component of the incident polarization is delayed in its phase with respect to the other component.

Let

$$\vec{E} = (E_{0x}\hat{e}_x + E_{0y}\hat{e}_y)e^{i(\omega t - kz)}$$

be the incident electric field of the plane wave travelling along the z-direction with  $E_{0x} = E_{0y}$  and the optical axis aligned along the y-axis. Then each vector component of the wave travels with a different wavevector

$$\begin{aligned} E_x &= E_{0x}e^{i(\omega t - n_e k_0 z)} \\ E_y &= E_{0y}e^{i(\omega t - n_o k_0 z)} \end{aligned}$$

After a material of thickness  $d$  both components have accumulated a phase shift of

$$\Delta\phi = (n_e - n_o)k_0 d$$

### Quarter Wave Plate

If we design now a material of a certain thickness such that the phase shift is  $\Delta\phi = \pi/2$ , then the light exiting the material will be circularly polarized as in the sketch on the right side. The minimum thickness of the crystal is then

$$d = \frac{\pi}{2} \frac{1}{k_0(n_e - n_o)} = \frac{\lambda}{4(n_e - n_o)}$$

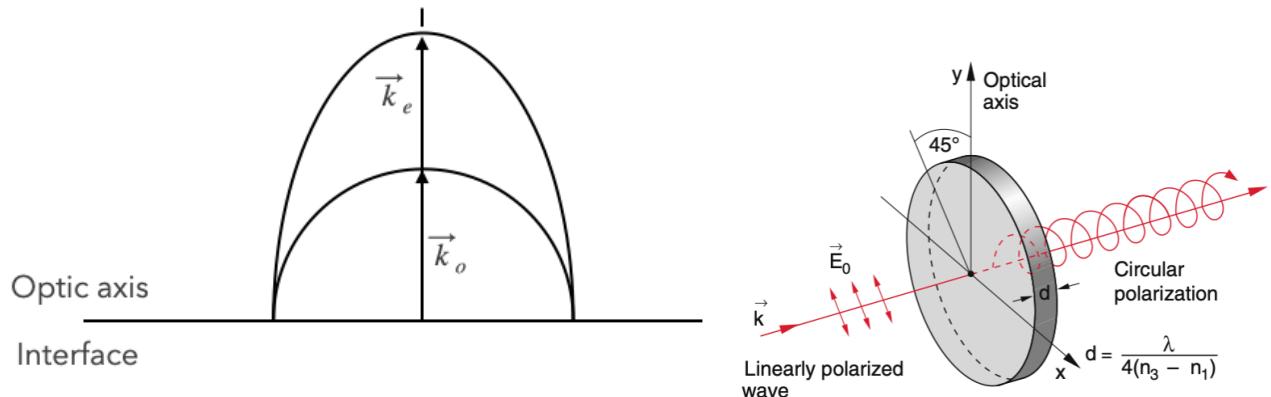


Figure 33.8: Sketch of the conditions for a wave retarder. Both ordinary and extra-ordinary wave propagate in the same direction, but with different phase velocities. The phase difference between the two components is  $\Delta\phi$ . The right side shows the resulting circular polarization.

The crystal thickness is, however, not restricted to this minimal thickness, but to any thickness which creates a phase difference of  $\pi/2$ . Such an optical element which is generating a phase difference of  $\pi/2$  between the different electric field components with respect to the optical axis is called a **Quarter Wave Plate**, since the phase difference corresponds to a quarter wavelength.

If the two electric field components are not of equal amplitude, then a quarter wave plate will generate elliptically polarized light. Note that if you send circular polarization to a QWP, linearly polarized light will be created.

### Half Wave Plate

A half wave plate creates a phase difference of  $\pi$  (half a wavelength) between the two electric field components. When linearly polarized light passes through a half wave plate, its polarization direction is rotated. The component of the electric field parallel to the optical axis experiences the extraordinary refractive index  $n_e$ , while the perpendicular component experiences the ordinary refractive index  $n_o$ . The  $\pi$  phase shift effectively flips the sign of the component perpendicular to the optical axis.

When linearly polarized light is incident at an angle  $\theta$  with respect to the optical axis, the output polarization will be rotated to an angle  $2\theta$  with respect to the optical axis but on the opposite side. This means the total rotation of the polarization direction is  $2\theta$ . For example, if the incident polarization makes a  $30^\circ$  angle with the optical axis, the output polarization will be rotated by  $60^\circ$  with respect to the original direction. The special case of  $45^\circ$  input leads to a  $90^\circ$  rotation of the polarization.

The minimum thickness of a half wave plate is given by:

$$d = \frac{\pi}{k_0(n_e - n_o)} = \frac{\lambda}{2(n_e - n_o)}$$

Similar to the quarter wave plate, any thickness which creates a phase difference of  $\pi$  can be used as a half wave plate. The actual thickness used in practice is often chosen based on manufacturing constraints and the specific wavelength range for which the plate is designed.



## Chapter 34

# Optical Activity and Rotatory Dispersion

Optical activity is a phenomenon where certain molecules rotate the plane of polarized light. While this basic definition is widely known, the underlying quantum mechanical and electromagnetic mechanisms reveal a fascinating interplay between molecular structure and light interaction.

When light interacts with chiral molecules, it induces both electric ( $\vec{p}$ ) and magnetic ( $\vec{m}$ ) dipole moments. These moments are coupled through the molecule's electronic structure:

$$\begin{aligned}\vec{p} &= \alpha \vec{E} + G \vec{B} \\ \vec{m} &= G \vec{E} + \beta \vec{B}\end{aligned}$$

where:

- $\alpha$  is the electric polarizability tensor
- $\beta$  is the magnetic polarizability tensor
- $G$  is the mixed electric-magnetic polarizability tensor
- $\vec{E}$  and  $\vec{B}$  are the electric and magnetic fields of light

The crucial parameter  $G$  exists only in chiral molecules and directly leads to circular birefringence.

### Circular Birefringence and Its Electronic Origin

The mixed polarizability  $G$  creates different responses for left and right circularly polarized light. Plane-polarized light can be decomposed into these circular components:

$$\vec{E}_{plane} = \vec{E}_{right} + \vec{E}_{left}$$

The refractive indices for right (R) and left (L) circular polarization are:

$$\begin{aligned}n_R &= n_0 + \frac{2\pi N}{n_0} (G' - iG'') \\ n_L &= n_0 - \frac{2\pi N}{n_0} (G' - iG'')\end{aligned}$$

where:

- $n_0$  is the average refractive index
- $N$  is the number density of molecules

- $G'$  and  $G''$  are the real and imaginary parts of  $G$

The rotation angle per unit length is:

$$\alpha = \frac{\pi}{\lambda} (n_L - n_R) = \frac{4\pi^2 Nl}{\lambda n_0} G'$$

### Quantum Mechanical Foundation

At a quantum level, the mixed polarizability  $G$  arises from electronic transitions:

$$G = \frac{1}{\hbar} \sum_n \frac{\omega_{n0}}{\omega_{n0}^2 - \omega^2} \Im(\langle 0 | \vec{\mu} | n \rangle \cdot \langle n | \vec{m} | 0 \rangle)$$

This expression connects molecular electronic structure to circular birefringence through:

1. Electronic transitions ( $\omega_{n0}$ )
2. Spatial arrangement of electrons (through  $\vec{\mu}$  and  $\vec{m}$ )
3. Molecular chirality (required for non-zero  $G$ )

## 34.1 Wavelength Dependence and the Cotton Effect

The wavelength dependence of optical rotation becomes more complex near electronic absorption bands, leading to the Cotton effect. This phenomenon is described mathematically by the sum of two terms:

$$[\alpha]_\lambda = \sum_i \frac{A_i \lambda}{\lambda^2 - \lambda_i^2} + \sum_j \frac{B_j \lambda^3}{(\lambda^2 - \lambda_j^2)^2}$$

In this expression,  $A_i$  and  $B_j$  represent amplitude constants that determine the strength of the optical rotation, while  $\lambda_i$  and  $\lambda_j$  correspond to the wavelengths of electronic absorption transitions in the molecule. The first term describes the normal optical rotatory dispersion away from absorption bands, while the second term becomes particularly important near absorption wavelengths, where it accounts for anomalous dispersion effects. When approaching an electronic transition, the optical rotation can change dramatically, even reversing sign, creating what is known as a Cotton curve. This behavior arises from the coupling between electronic transitions and the chiral structure of the molecule. The Cotton effect is a characteristic feature of optically active molecules and provides valuable information about their electronic structure, making it an important tool for investigating molecular conformations and electronic states in chiral systems.

## 34.2 Sugar Solutions

In sugar molecules, the optical activity arises from their asymmetric carbon centers and specific electronic structure:

1. **Coupled Chromophores:** The oxygen atoms and hydroxyl groups create a network of coupled electronic transitions.
2. **Helical Electron Displacement:** During light interaction, electrons follow a helical path:

$$\Psi_{electron}(t) = \sum_i c_i \Psi_i e^{i(\vec{k} \cdot \vec{r} - \omega t + \phi_i)}$$

The colored scattering in sugar solutions results from three combined effects:

1. **Wavelength-Dependent Rotation:** As light passes through a sugar solution, different wavelengths experience different amounts of rotation. This wavelength dependence follows a modified Drude equation:

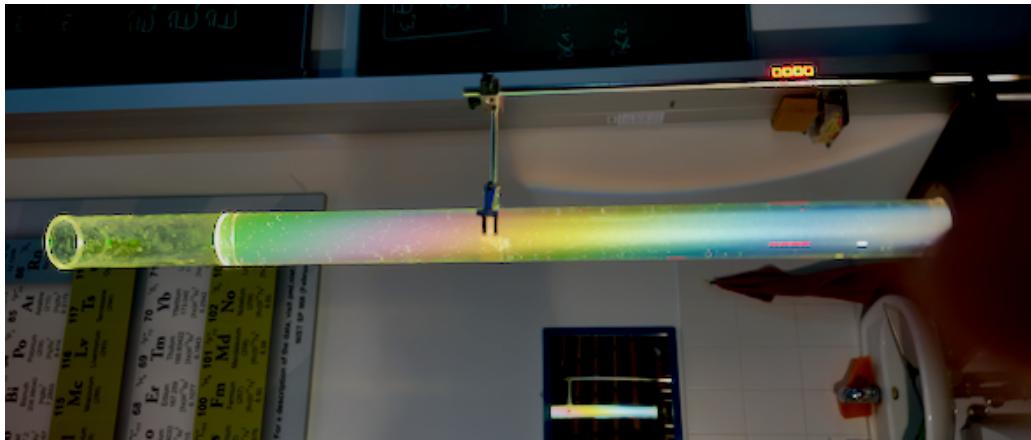


Figure 34.1: The complex interplay between molecular rotation, light scattering, and multiple scattering events leads to the observed color effects in sugar solutions.

$$\alpha(\lambda) = \frac{K}{\lambda^2} \left( 1 + \frac{a}{\lambda^2} + \frac{b}{\lambda^4} \right)$$

where  $K$  is related to the specific rotation of the sugar molecule, while  $a$  and  $b$  are correction terms accounting for electronic transitions. This relationship means that blue light (shorter wavelength) experiences significantly more rotation than red light (longer wavelength).

2. **Differential Scattering:** Different wavelengths scatter at different angles due to varying rotation angles. This creates a spatial separation of colors, as each wavelength emerges from the solution at a slightly different angle. The scattering angle  $\theta$  for each wavelength is related to the rotation angle by:

$$\theta(\lambda) \propto \alpha(\lambda)$$

This relationship leads to a rainbow-like separation of colors in the scattered light.

3. **Multiple Scattering Effects:** In real sugar solutions, light often undergoes multiple scattering events. The intensity of scattered light follows:

$$\vec{I}_{scattered}(\lambda) = \vec{I}_0(\lambda) e^{-\mu(\lambda)l} [1 - e^{-\sigma(\lambda)l}]$$

where  $\mu(\lambda)$  is the absorption coefficient,  $\sigma(\lambda)$  is the scattering coefficient, and  $l$  is the path length. Multiple scattering enhances the color separation effect and creates a more complex pattern of scattered light. The wavelength dependence of both  $\mu$  and  $\sigma$  further contributes to the observed color effects.

### 34.3 Temperature Effects

Temperature influences these processes through:

1. Molecular rotation rates:

$$\tau_c = \frac{4\pi\eta r^3}{3k_B T}$$

2. Conformational distribution:

$$N_i \propto e^{-E_i/k_B T}$$

## 34.4 Applications and Practical Implications

Understanding these mechanisms is crucial for:

1. Design of polarimetric instruments
2. Industrial crystallization processes
3. Chiral separation techniques
4. Pharmaceutical analysis methods

The complex interplay between electronic structure, circular birefringence, and light scattering explains both the fundamental nature of optical activity and its practical applications in various fields.

# Chapter 35

## Dipole Radiation

We would like to examine the origin of electromagnetic radiation in a classical picture. Especially we would like to understand why the electromagnetic fields that are radiated are transverse to the direction of propagation. We will consider for this purpose the radiation generated by an accelerated charge.

### 35.1 Electric field of an accelerated charge

We will follow for the derivation of transverse electric field the basic steps of the derivation by Larmor. We consider a charge, which is initially at rest in the point  $O$  as sketched below. The charge is then accelerated for a very short period  $\Delta t$  to the point  $O'$  and reaches a final speed  $u$ , which is small compared to the speed of light  $u \ll c$ . The charge then moves with that speed up to a point  $O''$  for a time  $t$ .

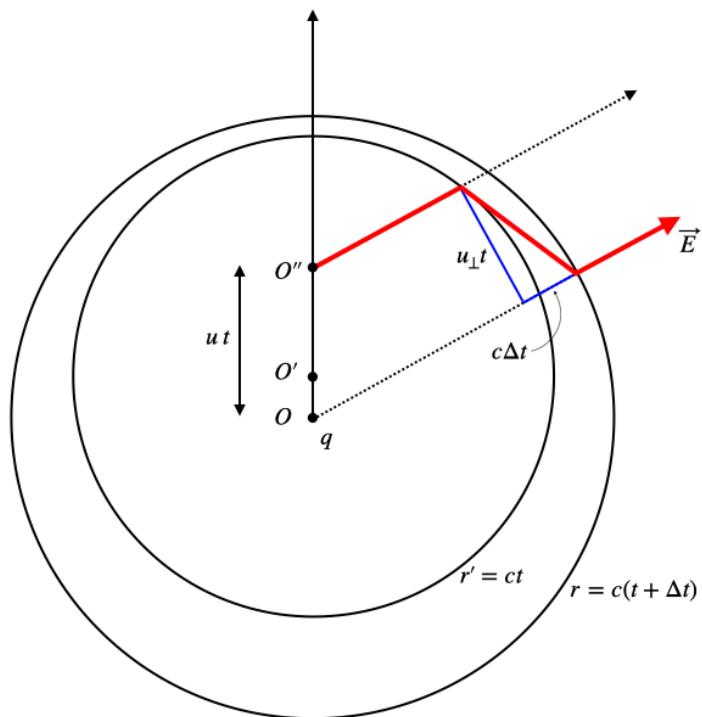


Figure 35.1: Setup of the accelerated charge.

What will be important for our consideration is the fact that any information about a system's change of state is limited in its propagation to the speed of light. Thus, the information that the charge  $q$  has been accelerated

has propagated from  $O$  a distance  $r = c(t + \Delta t)$ , when the charge has reached the point  $O''$ . The fact that the charge stopped accelerating, which is originating from point  $O'$  has traveled a distance  $r' = ct$ .

Consider now the field line of the electric field that exists outside the outer circle which would trace back to point  $O$ . This field line has now moved with the charge and is inside the inner circle corresponding to the red line. This field line needs to be connected to the remaining part outside the outer circle, as it is the same field line. Thus the field line needs to bend accordingly in the shell between the circles as indicated by the red line.

The bent part of the field vector can be decomposed into a field component perpendicular ( $E_{\perp}$ ) to the direction from  $O$  and one parallel to this radial direction ( $E_{\parallel}$ ). The perpendicular component is the new thing and we would like to calculate it.

The ratio of perpendicular and parallel component is then given by

$$\frac{E_{\perp}}{E_{\parallel}} = \frac{u_{\perp}t}{c\Delta t} = \frac{a_{\perp}\Delta t}{c\Delta t}$$

where the ratio of the electric fields is the same as of the velocity components. The velocity  $u_{\perp}$  thereby follows directly from the component of the acceleration perpendicular to the radial direction  $a_{\perp}$ . The result of this calculation is then independent of the time period  $\Delta t$  and we obtain

$$E_{\perp} = \frac{a_{\perp}r}{c^2} E_{\parallel}$$

where we replaced the time  $t$  by  $t = r/c$ . To obtain  $E_{\perp}$  we now need to know the value of  $E_{\parallel}$ , which we can obtain by Gauss's theorem considering a "pillbox" around the boundary of the sphere emanating from  $O$ . Since the field lines from  $O$  are radially outwards, this calculation yields

$$E_{\perp} = -\frac{a_{\perp}}{c^2} \frac{q}{4\pi\epsilon_0 r^2} = -\frac{a_{\perp}q}{c^2 4\pi\epsilon_0 r}$$

This electric field that is generated is now not anymore radially pointing outwards from the source, but it is tangential to a sphere around  $O$ . It further decays with the distance as  $1/r$  and not  $1/r^2$  as the common Coulomb field. The distance dependence is consistent with the one developed earlier for spherical waves. As we know that this electric field  $E_{\perp}$  is tangential to the sphere surface we may write down the radiated field of the accelerated charge as

$$\vec{E}(\vec{r}, t) = -\frac{a_{\perp}q}{c^2 4\pi\epsilon_0 r} \hat{\theta}$$

and the corresponding magnetic field as

$$\vec{B}(\vec{r}, t) = \frac{\vec{E}}{c} = -\frac{a_{\perp}q}{c^3 4\pi\epsilon_0 r} \hat{\phi}$$

where  $\hat{\theta}$  and  $\hat{\phi}$  are the unit vectors along the polar and azimuthal direction.

Note that in the above equation the electric field is observed at time  $t$  but the acceleration has happened a time  $t - \frac{r}{c}$  earlier as it propagates with finite speed. This will finally lead to our wavelike propagation.

## 35.2 Energy flow

With the help of the Poynting vector

$$\vec{S}(\vec{r}, t) = \frac{1}{\mu_0} \vec{E} \times \vec{B}$$

we may now have a look at the energy flow from the accelerated charge. Since the electric and the magnetic field are orthogonal we can readily obtain the magnitude of the Poynting vector

$$S(\vec{r}, t) = \frac{a_{\perp}^2 q^2}{\mu_0 c^5 (4\pi\epsilon_0)^2 r^2}$$

from which now follows with  $a_{\perp} = a \sin \theta$

$$S(\vec{r}, t) = \frac{a^2 q^2 \sin^2(\theta)}{\mu_0 c^5 (4\pi\epsilon_0)^2 r^2}$$

or

$$S(\vec{r}, t) = \frac{a^2 q^2 \sin^2(\theta)}{c^3 (4\pi\epsilon_0)^2 r^2}$$

The total power that is then radiated by the accelerated charge is given as the integral of the Poynting vector over a closed surface around the charge, i.e.

$$P = \int \int S dA = \frac{a^2 q^2 \sin^2(\theta)}{c^3 (4\pi\epsilon_0)^2} \int_0^{2\pi} \int_0^{\pi} \frac{\sin^2(\theta)}{r^2} r^2 \sin(\theta) d\theta d\phi$$

which upon integration finally yields **Larmor's formula**

$$P = \frac{q^2 a^2}{6\pi c^3 \epsilon_0}$$

This is the total radiated power of an accelerated charge.

### 35.3 Oscillating Dipole

With the previous section we are now ready to have a look at a situation where the charge is oscillating, for example, around a fixed positive charge. This situation can occur when an atom is polarized by the electric field of an incident light wave. Since this electric field is in the visible range of a wavelength much longer than the size of the atom, we may consider the atom as being in a homogeneous oscillating electric field as we did already earlier. This approximation is called the **Rayleigh limit** and the process is termed **Rayleigh Scattering**. Let's assume the charge is oscillating at a frequency  $\omega$  such that its displacement from the positive charge is

$$x = x_0 e^{i\omega t}$$

from which we obtain the velocity

$$\dot{x} = ix_0 \omega e^{i\omega t}$$

and finally the required acceleration

$$\ddot{x} = a = -x_0 \omega^2 e^{i\omega t}$$

The product of charge and acceleration which enters the generated electric field can then be expressed as

$$qa = -qx_0 \omega^2 e^{i\omega t} = -p\omega^2 e^{i\omega t}$$

since the dipole moment is given by  $p = qx_0$ . Consequently, the electric field radiated by an oscillating dipole is given by

$$E(\vec{r}, t) = -\frac{a_{\perp} q}{c^2 4\pi\epsilon_0 r} = \frac{p\omega^2 \sin^2(\theta)}{c^2 4\pi\epsilon_0 r} e^{i\omega t}$$

The direction of the electric and also the magnetic field can now be constructed with the appropriate unit vector in the radial direction as well as the direction of the dipole moment  $\vec{p}$ . The perpendicular component of the dipole moment including its direction is given by  $(\hat{e}_r \times \vec{p}) \times \hat{e}_r$  such that we obtain the electric and magnetic field in its vectorial beauty

$$\vec{E}(\vec{r}, t) = \frac{\omega^2}{4\pi\epsilon_0 c^2 r} (\hat{e}_r \times \vec{p}) \times \hat{e}_r e^{i(kr - \omega t)} \quad (35.1)$$

$$\vec{B}(\vec{r}, t) = \frac{\omega^2}{4\pi\epsilon_0 c r} (\hat{e}_r \times \vec{p}) e^{i(kr - \omega t)} \quad (35.2)$$

This corresponds to the radiated field of an oscillating dipole at large distances ( $r \gg \lambda$ ), which is called the far field. In the near field, there are additional components of the electric field which are not propagating and quickly decaying. The total electric field of an oscillating dipole is given by

$$\vec{E}(\vec{r}, t) = \frac{\omega^3}{4\pi\epsilon_0 c^3} \left[ ((\hat{e}_r \times \vec{p}) \times \hat{e}_r) \frac{1}{kr} + 3(\hat{e}_r (\hat{e}_r \cdot \vec{p}) - \vec{p}) \left( \frac{1}{(kr)^3} - \frac{i}{(kr)^2} \right) \right] e^{i(kr - i\omega t)}$$

With the help of the dipole field we thus obtain the intensity radiated by an oscillating dipole at

$$I(\theta) = \frac{\omega^4 |\vec{p}|^2}{32\pi^2 \epsilon_0 c^3} (1 - \cos^2(\theta))$$

and the radiated power is given by

$$P = \frac{\omega^4 |\vec{p}|^2}{12\pi\epsilon_0 c^3} = \left( \frac{2\pi}{\lambda} \right)^4 \frac{c|\vec{p}|^2}{12\pi\epsilon_0} \quad (\text{radiated power of an oscillating dipole})$$

As frequencies are not as intuitive in our daily color language, we have converted that expression to contain the wavelength of light, which tells us that the power radiated scales with the inverse of the wavelength to the power of four. This means for visible light that blue light is scattered much stronger than red light.

This  $\lambda^{-4}$  dependence, known as Rayleigh scattering, explains why the sky appears blue during the day. As sunlight travels through the atmosphere, it encounters molecules much smaller than its wavelength. These molecules scatter blue light ( $\lambda \approx 450 \text{ nm}$ ) about 10 times more strongly than red light ( $\lambda \approx 700 \text{ nm}$ ), causing the sky's characteristic blue color.

The same effect explains why sunsets appear red. When the Sun is near the horizon, sunlight travels through more atmosphere to reach our eyes, approximately through a path length  $L \propto 1/\cos\theta$ , where  $\theta$  is the angle from the zenith. The increased scattering of blue light along this longer path, following  $I \propto L\lambda^{-4}$ , leaves primarily red wavelengths to travel directly to the observer.

This fundamental process governs not only our sky's appearance but can also be observed in other natural phenomena. For instance, scattered light in colloidal suspensions follows the same wavelength dependence when the scattering particles are much smaller than the wavelength of light ( $d \ll \lambda$ ), creating similar color effects in smoke and certain liquids.

### Mie Scattering and Complex Materials

When particles become comparable to or larger than the wavelength of light ( $d \geq \lambda$ ), Rayleigh scattering no longer adequately describes the physics. In this regime, Mie scattering becomes dominant. For a dielectric sphere with relative permittivity  $\epsilon_r$ , the scattered intensity can be expressed as a series solution to Maxwell's equations:

$$I(\theta) = \frac{\lambda^2}{4\pi^2 r^2} (|S_1(\theta)|^2 + |S_2(\theta)|^2)$$

where  $S_1$  and  $S_2$  are complex scattering amplitudes containing Bessel functions and Legendre polynomials. The size parameter  $x = 2\pi r/\lambda$  and the relative refractive index  $m = \sqrt{\epsilon_r}$  determine the scattering behavior. For dielectric particles,  $m$  is real, leading to primarily directive scattering. The scattering cross-section  $\sigma$  for intermediate-sized dielectric particles can be approximated as:

$$\sigma \approx \pi r^2 \left( 2 - \frac{4}{x} \sin x + \frac{4}{x^2} (1 - \cos x) \right)$$

Modern materials engineering has extended these concepts to metamaterials, where engineered structures create effective medium properties not found in nature. For these materials, both  $\epsilon_r$  and the relative permeability  $\mu_r$  can be complex and frequency-dependent:

$$\epsilon_r(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\gamma\omega}, \quad \mu_r(\omega) = 1 - \frac{F\omega^2}{\omega^2 - \omega_0^2 + i\gamma\omega}$$

Here,  $\omega_p$  is the plasma frequency,  $\omega_0$  the resonance frequency, and  $\gamma$  the damping factor. These materials can exhibit negative refractive indices or epsilon-near-zero behavior, leading to unusual scattering patterns and applications in perfect lenses or electromagnetic cloaking. The Mie theory has been extended to these cases by allowing complex values of  $m$ , though the mathematical framework remains similar.

This generalized theory explains phenomena ranging from atmospheric optics to the design of nanophotonic devices and metamaterial-based sensors. The interaction between light and matter becomes particularly intricate when dealing with plasmonic materials or structures with engineered electromagnetic responses.



# **Part IV**

# **Quantum Mechanics**



# Chapter 36

## The Structure of Matter - Historical Experiments

Building upon our study of electromagnetic waves and classical mechanics, we now turn to the fundamental structure of matter itself. Our journey will follow three key discoveries among many others that revolutionized physics: First, we will examine evidence showing that matter is composed of discrete atoms and that atoms contain charged particles, based on groundbreaking 19th century experiments. Second, we will see how phenomena like the photoelectric effect revealed that light behaves not just as waves, but also as particles called photons. Third, we will explore how matter itself exhibits wave-like properties, leading to the Schrödinger equation and the foundations of quantum mechanics.

This exploration begins with the historical experiments that first proved the existence of atoms. We will examine Dalton's law of chemical combinations and Gay-Lussac's law of gaseous reactions, before investigating the internal structure of atoms through Thomson's and Rutherford's seminal experiments.

The concept of matter being composed of indivisible particles dates back to ancient Greece. The philosophers *Leucippus* and *Democritus* (5th century BCE) first proposed that all matter consists of infinitesimally small, space-filling particles that differ in size and shape. They argued that the properties of macroscopic objects arise from the arrangement of these fundamental particles, which they named “*ατομος*” (*atoms*) meaning “*indivisible*.” A century later, Epicurus\* advanced this atomic theory by proposing that these particles also possessed mass, adding to their spatial properties.

### 36.1 Atoms and molecules

#### Dalton's law

Based on precise quantitative analyses of chemical reactions, John Dalton published “*A New System of Chemical Philosophy*” in 1808, establishing three key postulates:

- All chemical elements are composed of indivisible atomic particles
- Atoms of the same element are identical in all properties, while atoms of different elements have different properties
- Chemical compounds form when atoms combine in simple whole-number ratios

#### Example: Formation of Water ( $H_2O$ )

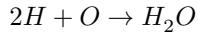
Let's look at the formation of water ( $H_2O$ ) as a classic example of Dalton's law:

- Hydrogen ( $H$ ) atoms and Oxygen ( $O$ ) atoms combine in a fixed 2 : 1 ratio
- Two hydrogen atoms always combine with one oxygen atom to form one water molecule
- This ratio never varies - you'll never find water molecules with three hydrogen atoms or two oxygen atoms

This demonstrates Dalton's three key postulates because:

1. The hydrogen and oxygen atoms are indivisible units
2. Each hydrogen atom is identical to every other hydrogen atom, and each oxygen atom is identical to every other oxygen atom
3. They combine in a simple whole-number ratio (2 : 1)

Chemical equation:



This consistent 2 : 1 ratio holds true whether you're making one molecule of water or a gallon of it - it's a fundamental principle that demonstrates Dalton's law of definite proportions.

## The law of Gay-Lussac

In 1805, building upon Dalton's atomic theory, Joseph Louis Gay-Lussac and Alexander von Humboldt discovered that hydrogen and oxygen at equal pressure and temperature combine to form water in a precise volume ratio of 2:1. This observation led Gay-Lussac to a more general law: when gases react at equal pressure and temperature, they do so in simple, whole-number volume ratios. This discovery provided further evidence for Dalton's atomic theory, as these integer ratios suggested that gases must be composed of discrete particles combining in simple proportions.

### Example: Formation of Water from Hydrogen and Oxygen Gases

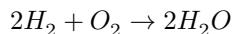
Let's examine how hydrogen and oxygen gases combine to form water vapor:

- 2 volumes of hydrogen gas ( $H_2$ ) + 1 volume of oxygen gas ( $O_2$ )  $\rightarrow$  2 volumes of water vapor ( $H_2O$ )
- At equal temperature and pressure:
  - If you start with 2L of hydrogen gas
  - And 1L of oxygen gas
  - You'll get 2L of water vapor

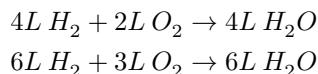
This perfectly demonstrates Gay-Lussac's law because:

1. The gases combine in simple, whole-number ratios (2 : 1)
2. The relationship holds true regardless of the actual volumes used (as long as they maintain the ratio)
3. The reaction must occur at the same temperature and pressure

Chemical equation:



If you were to scale this up or down, the ratio would remain constant. For example:



## The Avogadro number

In 1811 Amedeo Avogadro introduced the term "molecule" as the smallest particle of a gas that retains the characteristics of that gas. He hypothesized that at identical pressure and temperature, equal volumes of different gases contain the same number of molecules. This fundamental insight laid the groundwork for connecting molecular-scale properties to macroscopic measurements. The Avogadro number ( $N_A = 6.022 \cdot 10^{23} \text{ mol}^{-1}$ ) serves as a crucial bridge between these scales - it allows us to relate the mass of individual atoms and molecules to measurable quantities like grams, and connect microscopic properties like molecular energies to macroscopic properties like temperature and pressure. From these relationships emerged the concept of "molar volume" - the volume occupied by 1 mol of a gas at standard conditions ( $p = 1013 \text{ hPa}, T = 0^\circ\text{C}$ ), where the mass in grams corresponds to the molecular weight of the gas molecules. The modern, general definition of 1 mol is the number of particles that corresponds to 12 g of carbon  $^{12}\text{C}$ , which applies to both gaseous and non-gaseous substances. The molar volume is  $V_M = 22.414 \text{ L}$ .

Without the Avogadro number, we would be unable to quantitatively connect the behavior of individual particles to the bulk properties we can measure in the laboratory, making it one of the most important constants bridging

quantum and classical physics.

### **i** Oil spot experiment

One simple and practical method to determine the Avogadro number is the oil spot experiment. A small drop of oil with known volume  $V$  spreads on a water surface to form a thin film of thickness  $h$  equal to one molecule's diameter. If  $A$  is the area of the circular film, then:

$$h = \frac{V}{A}$$

Since we know the oil's density  $\rho$  and molar mass  $M$ , we can calculate the number of molecules in the drop:

$$N = \frac{\rho V}{M} N_A$$

The volume of one molecule is:

$$v = \frac{V}{N} = \frac{M}{\rho N_A}$$

Since the thickness  $h$  equals the diameter of one molecule, we can write:

$$h^3 = \frac{6M}{\pi \rho N_A}$$

Solving for  $N_A$ :

$$N_A = \frac{6M}{\pi \rho h^3}$$

Using oleic acid ( $C_{18}H_{34}O_2$ ), this method typically yields results within an order of magnitude of the accepted value.

### **i** Other Experimental Methods for Determining Avogadro's Number

#### Millikan Oil Drop Experiment

The Millikan oil drop experiment measures elementary electric charge ( $e$ ) and uses Faraday's constant ( $F$ ) to calculate Avogadro's number through the relationship:

$$N_A = F/e$$

#### X-ray Crystallography

This method determines atomic spacing in crystals and combines density measurements with molar mass. The calculation uses density ( $\rho$ ), unit cell volume ( $V_{unit}$ ), molar mass ( $M$ ), and number of atoms per unit cell ( $n_{atoms}$ ):

$$N_A = \frac{\rho V_{unit}}{M} \times n_{atoms}$$

#### Brownian Motion

Einstein's analysis of Brownian motion relates particle movement to Avogadro's number using temperature ( $T$ ), gas constant ( $R$ ), viscosity ( $\eta$ ), particle radius ( $r$ ), and diffusion coefficient ( $D$ ):

$$N_A = \frac{RT}{6\pi\eta r D}$$

#### Black Body Radiation

This approach uses the relationship between the gas constant ( $R$ ) and Boltzmann constant ( $k_B$ ):

$$N_A = \frac{R}{k_B}$$

Modern measurements typically combine multiple methods for increased accuracy, with the current accepted value being approximately  $6.022 \times 10^{23} \text{ mol}^{-1}$ .

These experimental findings and concepts paved the way for the modern understanding that matter is generally composed of atoms.

## 36.2 The Structure of atoms

### Indication of charged particles within atoms

At the end of the 19th century experimental findings accumulated indicating matter bears charged particles. The main results were:

- Experiments on electrolytic current demonstrated that molecules can dissociate, whereas the resulting ions migrate in opposite directions and transport charges and mass.
- Gas discharging phenomena are influenced through electric and magnetic fields. Thus, discharging is correlated to motion of charged particles.
- Magnetic phenomena arise from electric conduction in metals and semiconductors.
- $\alpha$  and  $\beta$  radiation are deflected through magnetic fields;  $\alpha$  and  $\beta$  particles are positively charged,

heavy particles and negatively charged, light particles, respectively. Consequently the concept of matter being composed of atoms was extended. Atoms in turn consist of positively and negatively charged particles which bear mass and charge.

Johann Wilhelm Hittorf observed in gas discharge tubes that particles emitted from a cathode propagate in straight lines. Moreover, these particles can be deflected with the aid of a magnet. As a consequence of the emission from the cathode and the direction of the magnetic poles, these particles had to be negatively charged. Later in 1897, Joseph John Thomson determined the charge-to-mass ratio  $e/m$  and demonstrated that this ratio is independent of the cathode material. In contrast to charged particles emitted from the cathode, Eugen Goldstein observed in 1886 a ray emitted from the anode which is propagating in the opposite direction than the cathode ray. In 1887 Wilhelm Wien reported from a  $10^{-4}$  reduced charge-to-mass ratio of this anode ray and concluded that these particles are charged gas ions.

In 1899 Thomson and Charles Wilson studied sinking droplets of condensed water vapor. The speed of falling was depending on the size of the droplets and the viscosity of the retarding gas. While measuring the amount of water and charges Thomson and Wilson were able to estimate the elementary charge of about  $10^{-19}$  C. In 1910 Robert Andrews Millikan refined this approach with his famous oil drop experiment. He sprayed tiny oil droplets into a chamber between two horizontal metal plates that could be charged to create an electric field. The droplets became electrically charged through friction and X-ray ionization. By carefully adjusting the electric field strength, Millikan could balance the gravitational force with the electromagnetic force, causing droplets to hover in place. By measuring the voltage needed to suspend droplets of different sizes, and accounting for air resistance, he discovered that the charges always occurred in discrete multiples of what we now know as the elementary charge:  $e = 1.602 \cdot 10^{-19}$  C.

#### Charge-to-Mass Ratio (Q/m)

The charge-to-mass ratio ( $Q/m$ ) of an object is the charge divided by the mass of the same object. This quantity is generally useful only for objects that may be treated as particles. For extended objects, total charge, charge density, total mass, and mass density are often more useful.

The charge to mass ratio can be measured in an experiment where a charged particle is accelerated by an electric field and deflected by a magnetic field. The ratio of the electric field strength to the magnetic field strength is equal to the charge-to-mass ratio of the particle.

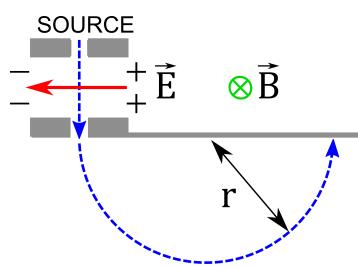


Figure 36.1: Principle Setup of the measurement of the charge to mass ratio. A charged particle is accelerated by an electric field and deflected by a magnetic field. The ratio of the electric field strength to the magnetic field strength is equal to the charge-to-mass ratio of the particle.

This method was used by Thomson to determine the charge-to-mass ratio of the electron and is still used today in modern experiments to measure the charge-to-mass ratio of particles with the help of mass spectrometers.

## Derivation

Starting from the equality of magnetic and centripetal forces:

$$qvB = m \frac{v^2}{r}$$

This can be rearranged to:

$$\frac{q}{m} = \frac{v}{Br}$$

Since electric and magnetic forces are equal:

$$F_{electric} = F_{magnetic}$$

$$Eq = Bqv$$

This gives us:

$$v = \frac{E}{B}$$

Combining these equations yields:

$$\frac{q}{m} = \frac{E}{B^2 r}$$

where:

- $q$  is charge
- $m$  is mass
- $v$  is velocity
- $B$  is magnetic field strength
- $r$  is radius of circular path
- $E$  is electric field strength

Concerning the mass of an electron, its value is still only accessible via the charge-to-mass ratio. A precise measurement of the  $e/m$  ratio is possible with a Wien filter. To do so, an electron is accelerated by means of a voltage, whereas perpendicular to the propagation direction an electric field deflects the electron beam. In addition a magnetic field is used to compensate the deflection. Thus, the  $e/m$  ratio depends only on the applied accelerating voltage and the electric and magnetic fields.

### The Thomson model

At the beginning of the 20th century only negatively charged electrons as cathode rays and positively charged  $\alpha$  particles were available for scattering experiments. However, if  $\alpha$  particles are scattered at atoms, the electrons will have only a minor effect. The scattering depends for the most part on the spatial arrangement of positive charges within the atoms.

As a first attempt, in 1904 Thomson proposed the “**Plum Pudding Model**”. According to this model every atom consists of a number of  $Z$  electrons with a cumulative charge of  $-Z \cdot e$  and a number of  $Z$  positive charges. Thus, the atoms appear neutral on the macroscopic scale. This model aligned with Paul Drude’s earlier theory of electrical conduction in metals (1900), which treated electrons as a freely moving “gas” within a positive background charge - essentially applying Thomson’s atomic model to bulk materials. Both models shared the concept of mobile electrons moving through a positively charged background.

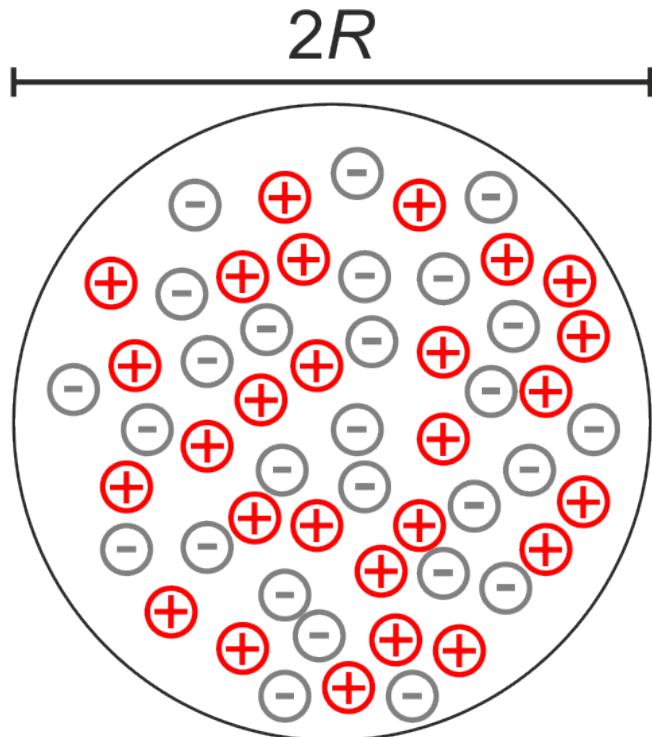


Figure 36.2: Model of an atom as proposed by Thomson. The equal amount of positive and negative charges are distributed across the atom volume.

Indication of a not-sufficient description is provided on the basis of simple calculations. If the  $Z$  electrons are equally distributed within a sphere with the radius  $R$  the electron density would measure

$$n_e = \frac{Z}{\frac{4}{3}\pi R^3}$$

Then, the plasma frequency would result in

$$\omega_p = \sqrt{\frac{n_e \cdot e^2}{\epsilon_0 \cdot m_e}} = \sqrt{\frac{3Ze^2}{4\pi\epsilon_0 R^3 m_e}}$$

This plasma frequency equation, derived from the uniform electron distribution in Thomson’s model, would predict specific absorption and emission frequencies. However, these predicted frequencies did not match the sharp spectral lines observed experimentally in atomic spectra, indicating a fundamental flaw in Thomson’s model.

## The Rutherford Model

In 1908, Rutherford, Geiger, and Marsden began their famous scattering experiments of  $\alpha$  particles (helium nuclei) on extremely thin gold foil, only a few atoms thick. Using a new, turnable apparatus, they were able to detect scattering angles up to  $180^\circ$ . The surprising observation that approximately 1 in 8000  $\alpha$  particles was scattered at large angles directly contradicted Thomson's plum pudding model, which predicted only small-angle deflections. The fact that most particles passed straight through the foil indicated that atoms are mostly empty space.

In 1911, Rutherford therefore proposed a new atomic model: all positive charge and most of the mass must be concentrated in a tiny volume (later called the nucleus) at the center of the atom, with electrons orbiting around it. Using this model, Rutherford derived the scattering formula for  $\alpha$  particles, which successfully described the experimental data. The Rutherford scattering formula is based on the Coulomb force between the  $\alpha$  particle and the nucleus, and it provides the differential cross section for scattering at a given angle.

To look into the scattering experiment we would like to define some basic quantity, the differential scattering cross section, which is not only important for the Rutherford scattering but also for other scattering experiments. The differential cross section means geometrically:

- Particles with impact parameters between  $b$  and  $b + db$  form a ring of area  $d\sigma = 2\pi b db$
- These particles are scattered into a ring on a sphere of radius  $R$  with area  $dA = 2\pi R^2 \sin(\theta) d\theta$
- The solid angle is  $d\Omega = dA/R^2 = 2\pi \sin(\theta) d\theta$

The differential cross section is defined as:

$$\frac{d\sigma}{d\Omega} = \frac{b db}{\sin(\theta) d\theta}$$

which can be rewritten as:

$$\frac{d\sigma}{d\Omega} = \frac{b}{\sin(\theta)} \frac{db}{d\theta}$$

The key physical insight is that particles hitting the target within a ring of radius  $b$  and width  $db$  are scattered into a corresponding angular range  $d\theta$ , and the cross section describes this relationship. As the scattering process involves the interaction potential between the particle and the target, the differential cross section provides a quantitative measure of the scattering process and the interaction between the particles.

### **i** Differential cross section and interaction potential

For a central potential  $V(r)$ , we can use conservation laws. The energy conservation gives:

$$E = \frac{1}{2}\mu v^2 + V(r)$$

And the angular momentum conservation yields:

$$L = \mu r^2 \dot{\phi} = \mu v b$$

as we used in Rutherford scattering.

The scattering angle  $\vartheta$  can be expressed through an integral formula:

$$\vartheta = \pi - 2 \int_{r_{min}}^{\infty} \frac{b dr}{r^2 \sqrt{1 - \frac{b^2}{r^2} - \frac{2V(r)}{E}}}$$

where  $r_{min}$  is the distance of closest approach.

This relates the impact parameter  $b$  to the scattering angle  $\vartheta$ . For the Coulomb potential  $V(r) = \frac{kqQ}{r}$ , this gives us the Rutherford result.

The general relation for the differential cross section remains:

$$\frac{d\sigma}{d\Omega} = \frac{b}{\sin \vartheta} \frac{db}{d\vartheta}$$

Different potentials  $V(r)$  will lead to different  $b(\vartheta)$  relations and thus different angular dependencies of the cross section.

### i Rutherford scattering

We consider the scattering of a charged particle in a Coulomb potential. The force between a particle A and the scattering center B is:

$$F = \frac{qQ}{4\pi\epsilon_0 r^2} = \frac{a}{r^2}$$

where  $q$  is the charge of the particle and  $Q$  the charge of the scattering center.

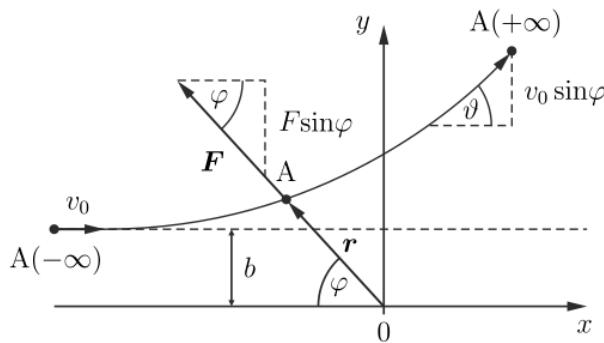


Figure 36.3: Sketch of the Rutherford scattering experiment. The  $\alpha$  particles A are scattered by the Coulomb potential of the nucleus (B at the origin). The parameter  $b$  is called the impact parameter. The angle  $\theta$  the scattering angle.

In the center of mass frame the angular momentum is:

$$L = \mu r^2 \frac{d\varphi}{dt} = \mu r v \sin \varphi \xrightarrow{(r \rightarrow \infty)} \mu v_0 b \quad (36.1)$$

with the reduced mass  $\mu = m_1 m_2 / (m_1 + m_2)$ . The particle is deflected by the force in  $y$ -direction:

$$F_y = \frac{a \sin \varphi}{r^2} = \mu \frac{dv_y}{dt}$$

With Equation ?? we get:

$$\frac{dv_y}{dt} = \frac{a \sin \varphi}{\mu v_0 b} \frac{d\varphi}{dt}$$

To get the total deflection  $\vartheta$  we need to integrate from  $A(-\infty)$  to  $A(+\infty)$ :

$$\int_0^{v_0 \sin \vartheta} dv_y = \frac{a}{\mu v_0 b} \int_0^{\pi - \vartheta} \sin \varphi d\varphi$$

$$v_0 \sin \vartheta = \frac{a}{\mu v_0 b} (1 + \cos \vartheta)$$

Using  $(1 + \cos \vartheta) / \sin \vartheta = 1 / (\tan(\vartheta/2))$  we find:

$$b = \frac{a}{\mu v_0^2} \frac{1}{\tan(\vartheta/2)} \quad (36.2)$$

From this, we can now calculate the differential cross section

$$\frac{db}{d\vartheta} = \frac{b}{\sin \vartheta} \frac{db}{d\vartheta}$$

Using  $\sin \vartheta = 2 \sin(\vartheta/2) \cos(\vartheta/2)$  we can rewrite Equation ?? as follows:

$$\frac{b}{\sin \vartheta} = \frac{1}{2} \frac{a}{\mu v_0^2} \frac{1}{\sin^2(\vartheta/2)}$$

The derivative of Equation ?? with respect to  $\vartheta$  gives:

$$\frac{db}{d\vartheta} = \frac{1}{2} \frac{a}{\mu v_0^2} \frac{1}{\sin^2(\vartheta/2)}$$

Thus, for the differential cross section we get:

$$\frac{d\sigma}{d\Omega} = \frac{1}{4} \left( \frac{a}{\mu v_0^2} \right)^2 \frac{1}{\sin^4(\vartheta/2)} = \frac{1}{4} \left( \frac{qQ}{4\pi\epsilon_0 \mu v_0^2} \right)^2 \frac{1}{\sin^4(\vartheta/2)}$$

which is the Rutherford scattering formula. This formula successfully described the experimental data except at very large angles (small impact parameters), where deviations indicated that nuclei, while much smaller than atoms ( $< 10^{-14}$  m), are not point-like. This groundbreaking experiment established the nuclear model of the atom and provided the first evidence for nuclear structure.

In an experiment, one would send  $\alpha$  particles from a source to a thin gold foil. The  $\alpha$  particles are scattered at the gold atoms and detected by a detector. The detector is placed at a distance  $R$  from the foil. The detector is often a scintillation counter or a photographic plate. In a scintillation counter, the  $\alpha$  particles hit a scintillator, which emits light when hit by a charged particle. The light is detected by a photomultiplier tube, which converts the light into an electrical signal. The signal is then processed and counted by a computer. The number of scattered particles is counted as a function of the scattering angle  $\vartheta$ . The Rutherford scattering formula describes the angular distribution of the scattered particles.

To quantitatively describe the scattering in terms of detected counts on a detector, we need to consider the following quantities:

- $\dot{N} \cdot A$ : number of particles per unit time hitting area  $A$  of scattering volume  $V$
- $\Delta\dot{N}(\vartheta, \Omega)$ : number of particles per unit time scattered into solid angle  $\Delta\Omega$  around angle  $\vartheta$

The fraction of scattered particles is given by:

$$\frac{\Delta\dot{N}}{\dot{N} \cdot A} = n_B \Delta x \frac{d\sigma}{d\Omega} \Delta\Omega$$

where  $n_B$  is the density of scattering centers,  $\Delta x$  is the scattering path length, and  $\frac{d\sigma}{d\Omega}$  is the differential scattering cross section.

The detector area at distance  $R$  is:

$$\Delta A_D = R^2 \sin(\vartheta) \Delta\Omega$$

Using the differential scattering cross section:

$$\frac{d\sigma}{d\Omega} = b \cdot \frac{db}{d\vartheta} \frac{1}{\sin(\vartheta)}$$

and calculating  $\frac{db}{d\vartheta}$ , we obtain the Rutherford scattering formula:

$$\frac{\Delta\dot{N}}{\dot{N} \cdot A} = n_{\text{Gold}} \Delta x \frac{1}{4R^2} \left( \frac{qQ}{8\pi\epsilon_0 E_{\text{kin}}} \right)^2 \frac{\Delta A_d}{\sin^4(\frac{\vartheta}{2})}$$

### Experimental verification

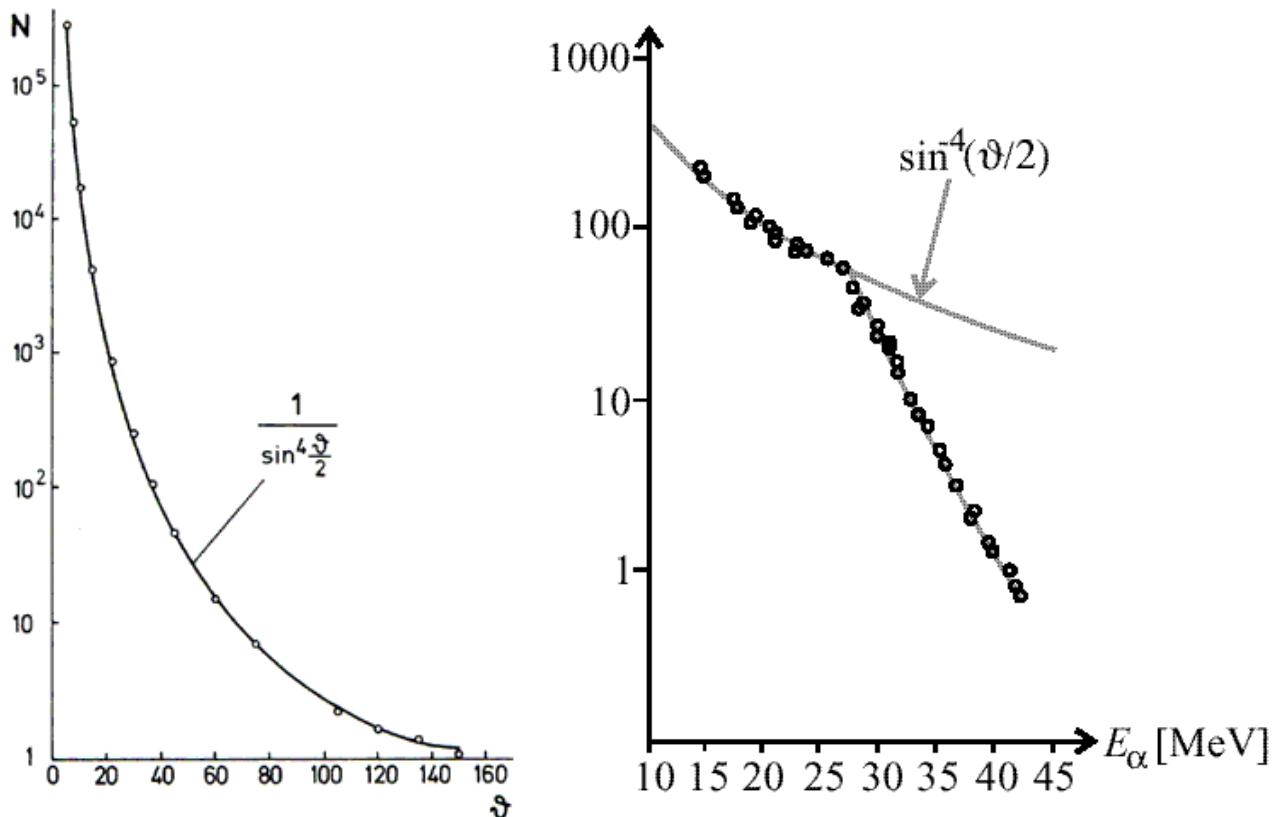


Figure 36.4: Graphical representation of Geiger and Marsden's experimental results for Rutherford scattering of  $\alpha$ -particles on a gold foil. The scattering rate  $\dot{N}$  is plotted as a function of the scattering angle  $\vartheta$  for a given energy of the  $\alpha$  particles in the left plot. The solid curve represents the theoretical prediction for Coulomb scattering.  $\alpha$  particles with a constant scattering angle of 60 degrees deviate from Rutherford scattering above a kinetic energy of approximately 25 MeV due to a change of the scattering potential as the  $\alpha$  particles penetrate the nucleus. The y-axis shows arbitrary units (scattering on gold).

### 36.3 Rutherford Scattering Trajectories

The graph below shows the calculated Rutherford scattering trajectories for  $\alpha$  particles scattered by a gold atom. The trajectories are calculated for different impact parameters  $b$  and show the deflection of the  $\alpha$  particles as they pass the gold atom using the above described forces.

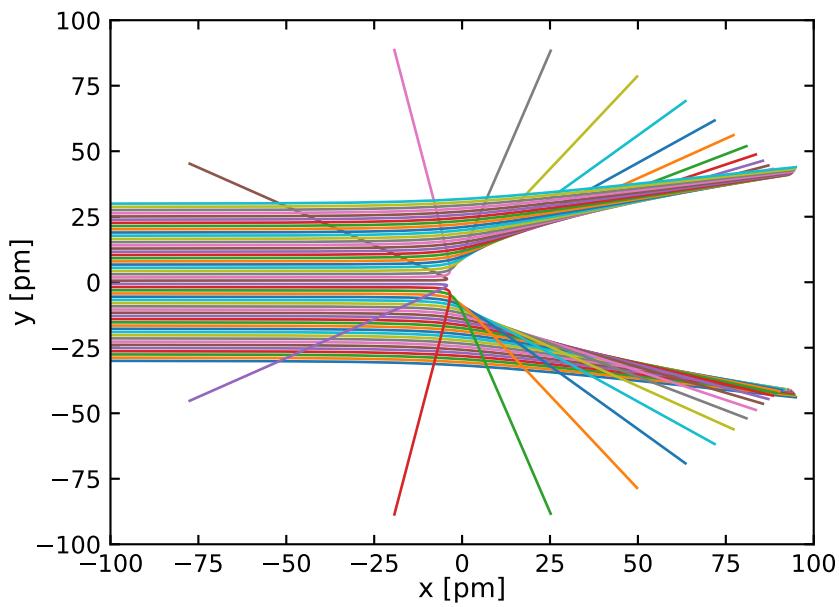


Figure 36.5: Rutherford scattering trajectory simulation

## 36.4 Spectroscopic Evidence

The development of optical tools like spectroscopy has provided further evidence for the inner structure of atoms. Of particular importance are the observation of sharp spectral lines in atomic emission and absorption spectra which seem to be a fingerprint of the atom's inner structure.

The image below shows the emission spectra of a number of basic elements revealing the characteristic spectral lines.

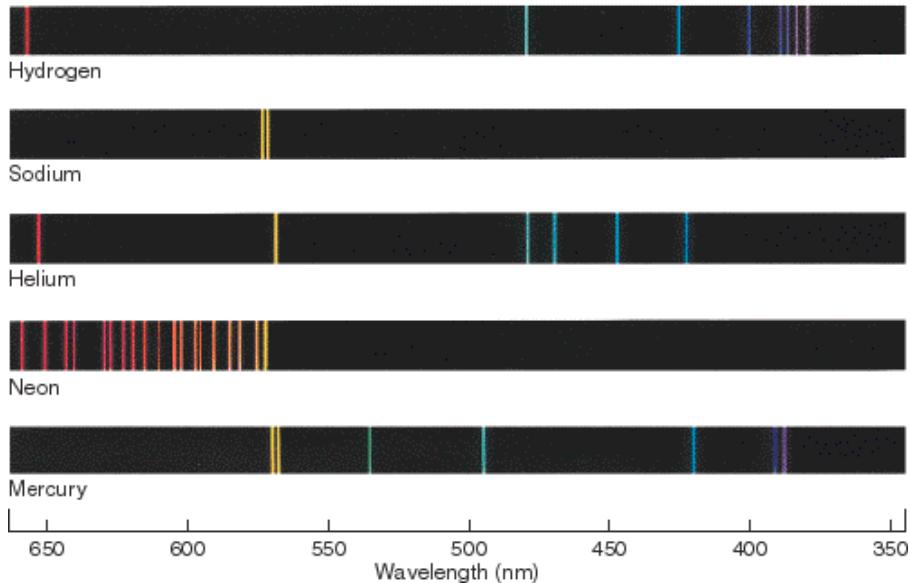


Figure 36.6: Emission spectra of Hydrogen, Sodium, Helium, Neon and Mercury

These lines have to be explained with models that also comply with the structure of the atoms that is determined by scattering experiments. The Bohr model, which will be discussed in one of the next lectures, was the first model

that could explain the spectral lines of hydrogen. It was based on the Rutherford model and the quantization of angular momentum.

# Chapter 37

## Particle Nature of Light

### 37.1 The Wave-Particle Nature of Light

The nature of light has been a subject of scientific debate since the 17th century. Newton proposed a particle theory based on light's straight propagation and refraction, while Huygens advocated a wave theory supported by interference and diffraction phenomena. The wave interpretation gained strong support when Heinrich Hertz discovered electromagnetic waves, with light being interpreted as a special spectral region governed by Maxwell's equations.

However, at the beginning of the 20th century, several experimental findings emerged that could not be explained by classical electromagnetic theory:

- The photoelectric effect, where light ejects electrons from metals
- The Compton effect, showing light scattering like particles
- The ultraviolet catastrophe in black body radiation

These observations revealed fundamental limitations in both classical mechanics and electromagnetic wave theory for describing atomic-scale phenomena. This document examines the experimental evidence that established light's particle nature while maintaining its wave characteristics.

### 37.2 The Photoelectric Effect

#### Hallwachs and Lennard's Discoveries

In 1888, Wilhelm Hallwachs discovered that ultraviolet light could cause negatively charged metal plates to lose their charge. Using charged foils connected to an irradiated metal plate, he observed that: - Negative charges decreased upon UV irradiation - Positive charges remained unchanged

This suggested that light was somehow causing electrons to leave the metal surface. #### Hallwachs and Lennard

In 1888 Wilhelm Hallwachs published an experiment with charged foils connected to a metal plate which was irradiated with ultraviolet light (we did this experiment in the last lecture before Christmas). If the foils and the plate are negatively charged and electrically isolated against the surrounding, the charge does decrease upon irradiation with ultraviolet light. In contrast, if the system is positively charged, the charge does not decrease. Hallwachs concluded that the light is responsible for negative charges leaving the metal plate.

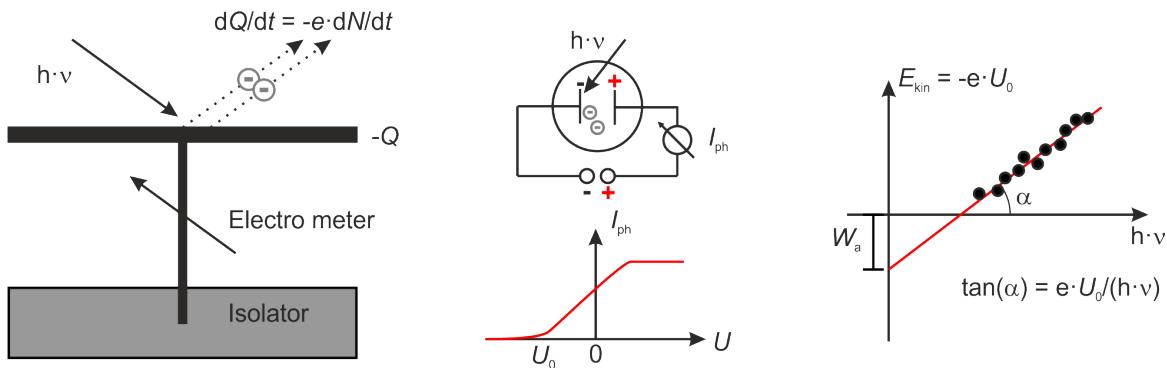


Figure 37.1: Scheme of the apparatus used by Hallwachs (left). Scheme of the apparatus used by Lennard and the corresponding photocurrent  $I_{\text{ph}}$  (center). From the onset voltage  $U_0$  one can calculate the work function  $W_a$  as intercept and Planck's constant  $h$  as part of the slope (right).

Later in 1902 Lennard measured the photocurrent between two plates in vacuum. The current set in already at a negative voltage  $U_0$  between the plates, increased with rising voltage, and reached a plateau which depended only on the light's intensity. He concluded:

- The electrons must bear a minimum energy in order to overcome the oppositely directed electric field,  $E_{\text{kin}} \leq e \cdot U_0$ .
- The kinetic energy  $mv^2/2$  of the photoelectrons depends on the frequency  $\nu$  of the light, not on the light's intensity.
- The number of photoelectrons is proportional to the light's intensity.
- There is no delay between light irradiation and electron emission.

### Classical Wave Theory Expectations

If we assume a fully wave-like behavior of light, then light with a radiation power of  $P_L$  might hit a surface with an area  $A$  and shares its energy equally between all electrons. For a penetration depth of  $\Delta z \approx \lambda$  and a density of the conducting electrons of  $N$ , then every conducting electron gets on average the energy of

$$\Delta \bar{W} = \frac{P_L}{N \cdot A \cdot \lambda} \Delta t$$

within the time interval  $\Delta t$ . Thus, the work function can be compensated in the case of

$$\Delta t > W_a \frac{N \cdot A \cdot \lambda}{P_L}.$$

Let us consider a zinc plate with a work function of  $W_a = 4 \text{ eV}$  and a light source with a spectral filter  $\lambda = 250 \text{ nm}$  emitting a power of  $P_L = 1 \text{ W}$  at a distance of  $R = 1 \text{ m}$  away from our zinc plate, there will be an intensity of

$$I_L = \frac{P_L}{4\pi R^2} \approx 8 \cdot 10^{-6} \frac{\text{W}}{\text{cm}^{-2}}$$

reaching the plate. For a penetration depth of  $\Delta z \approx \lambda$  this intensity will be distributed between

$$N = 10^{23} \cdot \text{cm}^{-3} \cdot \lambda = 2.5 \cdot 10^{18} \frac{1}{\text{cm}^{-2}}$$

electrons, whereas each electron acquires on average a power of

$$P_{\text{el}} = 3 \cdot 10^{-24} \text{ W} = 2 \cdot 10^{-5} \text{ eVs}^{-1}.$$

Thus, it will take a time of  $\Delta t = W_a/P_{\text{el}} = 2 \cdot 10^5$  s for one electron to acquire enough energy to leave the zinc plate. This result is in clear contrast to experimental findings.

### Einstein's Quantum Explanation

In 1905, Einstein proposed a revolutionary explanation based on light quanta (photons). According to this model:

- Light energy is transmitted in discrete packets of energy  $E = h\nu$
- Each photon interacts with a single electron
- The photon's entire energy transfers to the electron

This led to Einstein's photoelectric equation:

$$E_{\text{kin}}^{\max} = h \cdot \nu - W_a,$$

with  $W_a = -e(\phi_{\text{vac}} - \phi)$  being the work function of the cathode material (often the vacuum work function is set to zero,  $\phi_{\text{vac}} = 0$ ). The work function is the amount of energy one has to compensate in order to bring one electron from bulk into vacuum against the forces binding the electron in bulk.

The work function determines at which frequency or wavelength the photoelectric effect occurs. Below are typical work functions for various metals:

Table 37.1: Work functions and corresponding threshold wavelengths for various metals

Metal	Work Function (eV)	Threshold Wavelength (nm)
Cesium	1.95	636
Potassium	2.30	539
Sodium	2.75	451
Calcium	3.20	388
Zinc	4.31	288
Copper	4.70	264
Silver	4.73	262
Platinum	6.35	195

These values show why alkali metals like cesium and potassium are particularly suitable for photoelectric devices, as they respond to visible light, while metals like platinum require ultraviolet radiation.

Since one can determine the maximum kinetic energy  $E_{\text{kin}}^{\max} = -e \cdot U_0$  ( $U_0 < 0$ ) from the voltage  $U_0$  at which the photocurrent sets in and

$$-e \cdot U_0 = h \cdot \nu - W_a,$$

one is able to determine the work function on the basis of the intercept of the  $-e \cdot U_0$  vs.  $h \cdot \nu$  curve and Planck's constant from the slope of the curve.

### Experimental Verification

The definitive test of Einstein's photon theory came from Joffé and Dobronrawov in 1925. They used small, charged bismuth beads held within a Millikan capacitor and irradiated those beads with low-dose X-rays. Every change of the overall charge of the beads interferes with the equilibrium in the capacitor, and can be observed by means of a change of the bead position. Using a radiation power of  $P = 10^{-12}$  W meaning an emission rate of  $\dot{N} = 10^3$  photons per second with an energy of  $h \cdot \nu = 10^4$  eV on average every 30 minutes a change of the bead charge was detected. The number of photons arriving at one bead within a time interval  $\Delta t$  is  $Z = \dot{N} \cdot \Delta t \cdot d\Omega / (4\pi)$ , with  $d\Omega$  as the solid angle covered by the bead. The calculated time constant of  $Z$  was in well agreement with the observed rate of the charge alteration. If we again assume a wave-like explanation of the photoeffect, the emitted power within the according solid angle will be absorbed by the bead and distributed between all its electrons. As a consequence the bead as a whole will have collected enough energy in order to

release an electron within the same period of time. However it cannot be explained how all  $10^{12}$  atoms are supposed to combine their energy in one, single electron at the very same time.

### 37.3 The Compton Effect

The particle nature of light received further confirmation through X-ray scattering experiments. When X-rays of wavelength  $\lambda_0$  strike a material, the scattered radiation shows two components:

- Unshifted radiation with the original wavelength  $\lambda_0$
- Shifted radiation with increased wavelength  $\lambda_S > \lambda_0$

Remarkably, the wavelength shift depends primarily on the scattering angle rather than the target material, suggesting a fundamental interaction mechanism. Experiments concerning the Compton effect were first conducted by Arthur Compton in 1923. He used X-rays with a wavelength of  $\lambda_0 = 0.154$  nm and observed scattered radiation with a wavelength of  $\lambda_S = 0.165$  nm at a scattering angle of  $\theta = 60^\circ$ .

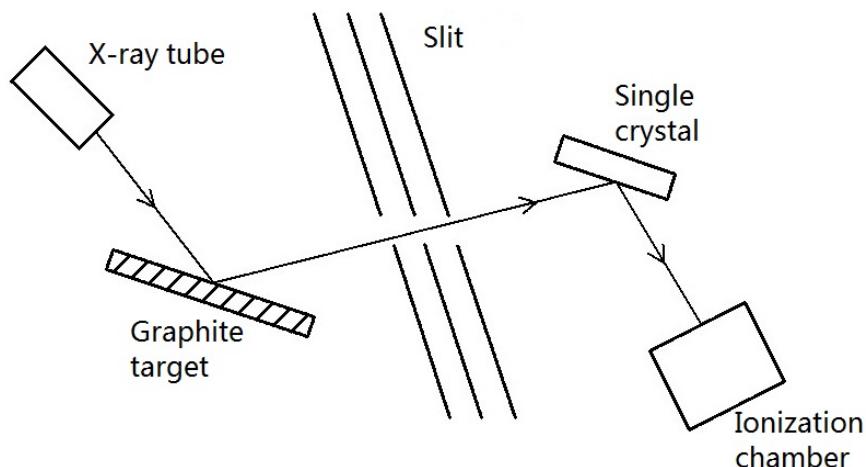


Figure 37.2: General setup for Compton scattering experiments.

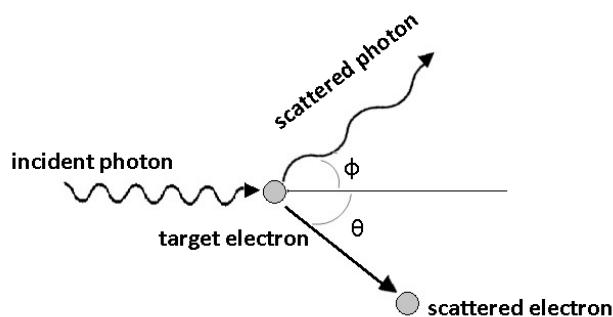
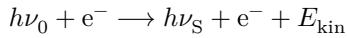


Figure 37.3: Compton Scattering Geometry

### Theoretical Analysis

This phenomenon can be explained by treating light as particles (photons) colliding with electrons. Each photon carries: - Energy  $E = h\nu = \hbar\omega$  - Momentum  $p = h/\lambda = \hbar k$

If a photon impacts into a weakly bound electron with a binding energy much smaller than the photon energy ( $E_B \leq E$ ), we can neglect the binding energy and assume the electron as free electron. In addition we simplify further and assume the electron as being at rest. During the collision event



energy and momentum are conserved. The law of conservation of energy then reads as

$$h\nu_0 = h\nu_s + E_{\text{kin}}$$

with  $E_{\text{kin}}^e$  as the relativistic kinetic energy of the electron

$$E_{\text{kin}}^e = \frac{m_0 c^2}{\sqrt{1 - \beta^2}} - m_0 c^2$$

and  $\beta = v/c$ . If we ascribe a momentum like

$$\vec{p} = \hbar \vec{k}$$

with

$$|\vec{p}| = \hbar |\vec{k}| = h \frac{1}{\lambda}$$

to the photon, we can formulate the law of momentum conservation as follows,

$$\hbar \vec{k}_0 = \hbar \vec{k}_s + \vec{p}^e$$

with

$$\vec{p}^e = \frac{m_0 \vec{v}}{\sqrt{1 - \beta^2}}.$$

If we isolate the square of the momentum of the electron, we obtain an equation depending on the squared difference between the wavevectors of the incident and scattered photon. Calculating this difference leads to a scalar product between these two vectors and necessitates the introduction of the angle between the propagation direction of the incident and scattered photon. We denote this angle as  $\phi$ ,

$$\begin{aligned} \frac{m_0^2 v^2}{1 - \beta^2} &= \hbar^2 (\vec{k}_0 - \vec{k}_s)^2 \\ &= \hbar^2 (k_0^2 + k_s^2 - 2k_0 k_s \cos(\varphi)) \\ &= \frac{h^2}{c^2} (\nu_0^2 + \nu_s^2 - 2\nu_0 \nu_s \cos(\varphi)) \end{aligned}$$

where:

- $m_0$  is the rest mass
- $v$  is velocity
- $\beta = v/c$
- $k_0, k_s$  are initial and scattered wave vectors
- $\nu_0, \nu_s$  are initial and scattered frequencies
- $\varphi$  is the scattering angle

From the law of energy conservation we get

$$\frac{m_0^2 v^2}{1 - \beta^2} = \frac{h^2}{c^2} (\nu_0 - \nu_s)^2 + 2 h m_0 (\nu_0 - \nu_s),$$

which we can compare with the law of momentum conservation and get

$$\nu_0 - \nu_s = \frac{h}{m_0 c^2} \nu_0 \nu_s (1 - \cos(\varphi)).$$

Now making use of  $1 - \cos(\varphi) = 2 \sin^2(\varphi/2)$  and  $\nu = c/\lambda$  we achieve the Compton formula

$$\begin{aligned} \lambda_s &= \lambda_0 + 2 \frac{h}{m_0 c} \sin^2(\varphi/2) \\ &= \lambda_0 + 2 \lambda_C \sin^2(\varphi/2) \end{aligned}$$

with  $\lambda_C$  denoting the Compton wavelength of the electron,

$$\lambda_C = \frac{h}{m_0 c} = 2.4262 \cdot 10^{-12} \text{ m.}$$

The Compton wavelength is a constant and represents the change of the wavelength  $\Delta\lambda = \lambda_s - \lambda_0$  at a scattering angle of  $\varphi = 90^\circ$ . Results from experiments almost perfectly coincide with the Compton formula. Furthermore the ratio between the wavelengths

$$\frac{\lambda_s}{\lambda_0} = \frac{h\nu_0}{m_0 c^2}$$

(replace with ??

$$\frac{\nu_s}{\nu_0} = \frac{1}{1 + \frac{h\nu_0}{m_0 c^2} (1 - \cos(\varphi))}$$

) represents the ratio between the energy of the incident photon and the energy of the electron at rest. Thus, if we know the mass of the electron, we can determine  $\varphi$  and  $\lambda_s$  (and therefore  $\lambda_C$ ) and calculate  $h$ .

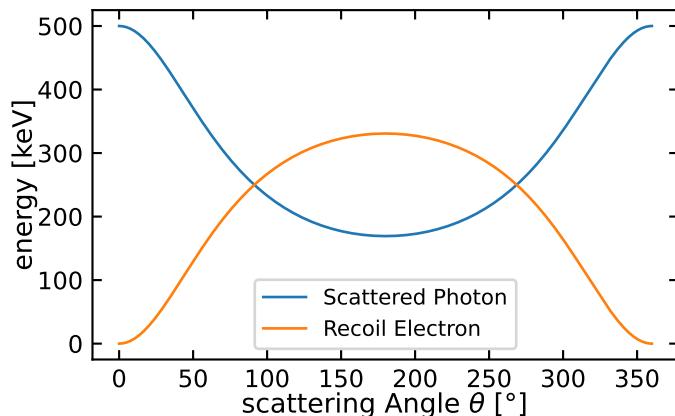


Figure 37.4: Photon and Electron Energies in Compton Scattering

## Experimental Results

The plots below show the observed relative scattering intensity of photons at different wavelength for different scattering angles from an experiment using XXX photons scattered on a YYY target. The plots comprise an elastic scattering peak at the wavelength of the incident photons and an inelastic scattering peak at a longer wavelength, in agreement with the Compton formula.

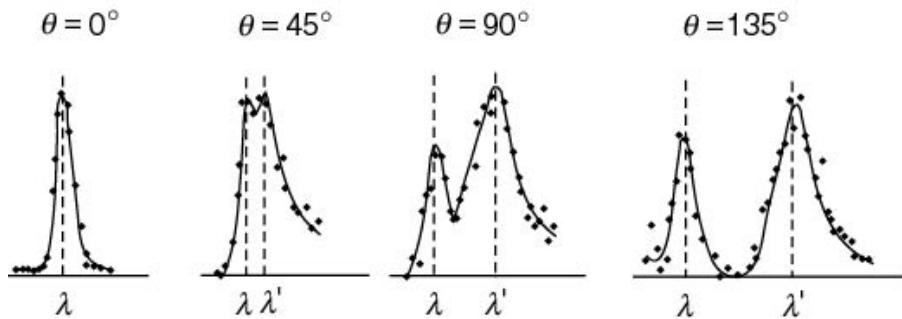


Figure 37.5: Compton Scattering of X-rays. The spectra contain an elastic peak at the incident wavelength and an inelastic peak at a longer wavelength. The elastic peak occurs since the sample contains also tightly bound electrons which do not contribute to the Compton effect.

The Klein-Nishina formula describes the differential cross-section of the Compton effect, showing how the intensity of scattered photons depends on the scattering angle and wavelength:

$$\frac{d\sigma}{d\Omega} = \frac{r_e^2}{2} \left( \frac{E'}{E_0} \right)^2 \left( \frac{E'}{E_0} + \frac{E_0}{E'} - \sin^2 \theta \right)$$

where:

- $r_e$  is the classical electron radius
- $E'$  is the scattered photon energy
- $E_0$  is the initial photon energy
- $\theta$  is the scattering angle

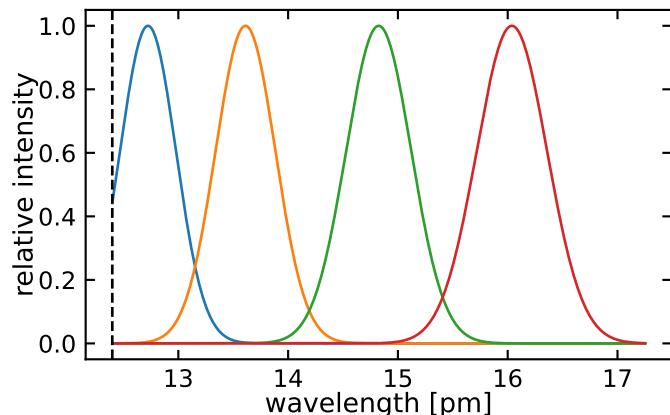


Figure 37.6: Klein-Nishina Differential Cross-Section for the Compton Effect at different scattering angles of 30°, 60°, 90°, and 120°

### **i** Detection of $\gamma$ rays

High energy photons, i.e.  $\gamma$  rays can be detected by a scintillation counter. The scintillation counter consists of a scintillator, a photomultiplier tube, and a discriminator. The scintillator is a material that emits visible light when struck by high-energy photons. The emitted light is then converted into an electrical signal by the photomultiplier tube. The photomultiplier amplifies the signal through acceleration of photoelectrons (generated by the photoelectric effect) in an electric field, creating cascades of secondary electrons upon collision with metallic dynodes. The discriminator filters the output signals to reject noise and select events within specific energy ranges of interest.

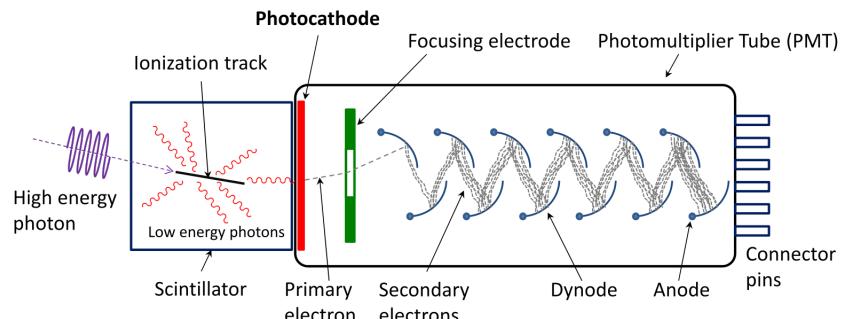


Figure 37.7: Sketch of a Scintillator

### **i** Compton Continuum and Compton Edge

#### Compton Edge

The Compton edge represents the maximum energy transfer possible in a single Compton scattering event, occurring at a scattering angle of  $180^\circ$  (backscatter). The energy of the Compton edge ( $E_{CE}$ ) can be calculated as:

$$E_{CE} = E_0 - E'_{min} = E_0 \left( 1 - \frac{1}{1 + \frac{2E_0}{m_e c^2}} \right)$$

where  $E_0$  is the initial photon energy and  $E'_{min}$  is the minimum energy of the scattered photon.

#### Compton Continuum

The Compton continuum is the energy distribution of scattered photons between:

- The photopeak (original energy  $E_0$ )
- The minimum energy after scattering ( $E'_{min}$  at  $180^\circ$ )

The intensity distribution across this continuum is described by the Klein-Nishina formula:

$$\frac{d\sigma}{d\Omega} = \frac{r_e^2}{2} \left( \frac{E'}{E_0} \right)^2 \left( \frac{E'}{E_0} + \frac{E_0}{E'} - \sin^2 \theta \right)$$

In gamma-ray spectroscopy, this appears as a broad feature in the energy spectrum, with:

1. A sharp cutoff at the Compton edge
2. Declining intensity towards lower energies
3. A backscatter peak at low energies corresponding to  $180^\circ$  scattering

## 37.4 Properties of Photons

Having established light's particle nature through the photoelectric and Compton effects, we can now summarize the fundamental properties of photons. These quantum particles unite the classical wave description of light with its particle behavior.

### Energy and Momentum

Every electromagnetic field consists of energy quanta  $h\nu$ , which we call photons. For a cavity resonator, the electromagnetic energy density is:

$$\omega_{\text{em}} = n \cdot h \cdot \nu.$$

We can further comprehend the flux  $I = \varepsilon_0 c E^2$  of an electromagnetic wave as a *particle flux*  $\dot{N}$  of photons

$$I = \dot{N} h \nu$$

with  $\dot{N} = n \cdot c$ . Therefore, if a light wave with a flux of  $I$  is shining on an area is *tantamount* with a particle flux of photons reaching this area.

As demonstrated on the basis of the Compton effect, every photon bears a momentum  $\vec{p} = \hbar \cdot \vec{k}$  with the magnitude  $|\vec{p}| = p = h\nu/c$ . As a consequence, if we are able to assign an energy density to the density of modes, we are also able to assign a *momentum density*,

$$\pi_{\text{em}} = n \cdot \hbar \cdot k.$$

Then, the relation between the energy density  $\omega_{\text{em}}$  of an electromagnetic wave and the momentum density  $\pi_{\text{em}}$  of the very same wave reads as

$$\omega_{\text{em}} = c \cdot \pi_{\text{em}}.$$

### Angular Momentum

Photons carry two distinct types of angular momentum: spin angular momentum (SAM) associated with polarization, and orbital angular momentum (OAM) associated with the spatial distribution of the wavefront.

#### Spin Angular Momentum

The spin angular momentum is related to the polarization state of light. When a circularly polarized photon is absorbed by a free atom, the atom's angular momentum changes by  $\pm \hbar$ . For left-handed circular polarized light ( $\sigma^+$ ) propagating along  $z$ , the change is  $\Delta J_z = +\hbar$ , while for right-handed polarization ( $\sigma^-$ ), it's  $\Delta J_z = -\hbar$ . The spin angular momentum vector is:

$$\vec{S}_{Ph} = \pm \hbar \frac{\vec{k}}{|\vec{k}|}$$

Linearly polarized light, being a superposition of equal amounts of  $\sigma^+$  and  $\sigma^-$  light, has zero net spin angular momentum.

### Orbital Angular Momentum

Photons can also carry orbital angular momentum, characterized by a quantum number  $l$ :

$$\vec{L}_{Ph} = l\hbar \frac{\vec{k}}{|\vec{k}|}$$

where  $l$  can be any integer. This OAM is associated with helical wavefronts and is important in applications like optical tweezers and quantum information.

The total angular momentum of a photon is the sum of both contributions:

$$\vec{J}_{Ph} = \vec{S}_{Ph} + \vec{L}_{Ph}$$

### Effect of Gravity on Photons

Despite having zero rest mass, photons interact with gravitational fields due to their energy-equivalent mass. From special relativity we know that mass is affected by the relative motion of reference systems:

$$m = \frac{1}{\sqrt{1 - (\frac{v}{c})^2}} m_0$$

Only particles with a rest mass  $m_0 = 0$  can travel at the speed of light, which is why photons must have zero rest mass. For the energy of a photon, we find:

$$\begin{aligned} E &= \sqrt{p^2 c^2 + m_0^2 c^4} \\ &= pc \\ &= \frac{h}{\lambda} c \\ &= h\nu \end{aligned}$$

in accordance with previous energy and momentum considerations. If we assign an effective mass  $m$  to the photon:

$$m = \frac{E}{c^2} = \frac{h\nu}{c^2}$$

this photon must perform work when traveling in a gravitational field. Moving from position  $\vec{r}_1$  with gravitational potential  $\Phi(\vec{r}_1)$  to position  $\vec{r}_2$  with potential  $\Phi(\vec{r}_2)$ , the work is:

$$W = m \cdot \Delta\Phi = \frac{h\nu}{c^2} (\Phi(\vec{r}_2) - \Phi(\vec{r}_1))$$

By energy conservation, the photon's energy  $h\nu$  must change by this amount, leading to a frequency shift:

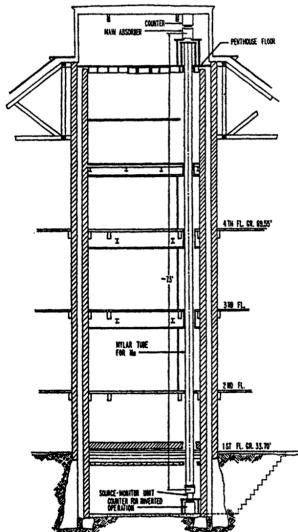
$$\nu_2 = \nu_1 \left(1 - \frac{\Delta\phi}{c^2}\right)$$

or in relative terms:

$$\frac{\Delta\nu}{\nu} = \frac{\Delta\Phi}{c^2}$$

This remarkable prediction - that photons experience a redshift (longer wavelength, smaller frequency, lower energy) when rising in a gravitational field - was first experimentally verified by Pound and Rebka in 1959. Using the Mössbauer effect, they measured the frequency shift of gamma rays from  $^{57}\text{Fe}$  (14.4 keV) traveling vertically over a height of 22.5 meters at Harvard University. The expected relative frequency shift was extremely small:

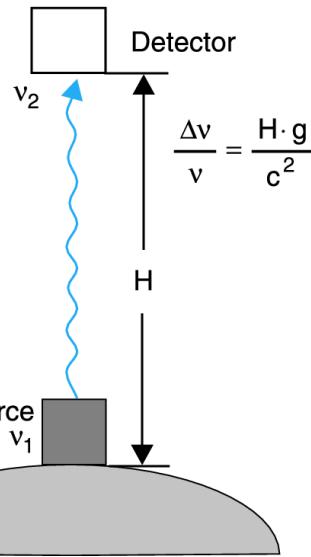
$$\frac{\Delta\nu}{\nu} \approx 2.5 \times 10^{-15}$$



(a) Architectural sketch of the Jefferson Physical Laboratory

tower showing the placement of the gravitational red-shift

experiment running from the penthouse to the basement  
Source: R. V. Pound and J. L. Snider, "Effect of Gravity" gamma ray source and detector separated vertically by 22.5  
(ref. 23), p. B 792 and Pound & Rebka, Phys. Rev. Lett. 3,  
440 (1959)



(b) Schematic of the Pound-Rebka experiment showing

Figure 37.8: Sketch of the tower at Harvard University (left) and the Pound-Rebka experiment setup (right).

### The Mössbauer Effect in the Pound-Rebka Experiment

The detection of gravitational redshift requires extremely precise frequency measurements ( $\Delta\nu/\nu \approx 10^{-15}$ ). This was made possible by the Mössbauer effect, which allows for recoil-free emission and absorption of gamma rays in crystals.

In free atoms, the emission or absorption of a gamma ray causes recoil, shifting the photon energy by about  $10^{-3}$  eV. For a gamma ray of energy  $E_\gamma$ , the recoil energy is:

$$E_R = \frac{E_\gamma^2}{2Mc^2}$$

where  $M$  is the mass of the nucleus. However, in a crystal lattice, the nucleus is not free to recoil. Instead, the recoil momentum can be transferred to the entire crystal if:

1. The nucleus is tightly bound in the crystal lattice
2. The recoil energy is less than the energy of the lowest lattice vibration (phonon)
3. The entire crystal acts as a single quantum mechanical system

In this case, the effective mass  $M$  in the recoil energy formula becomes the mass of the entire crystal, making  $E_R$  negligible. For  $^{57}\text{Fe}$  nuclei in a crystal:

- Gamma ray energy: 14.4 keV
- Natural linewidth:  $\approx 10^{-8}$  eV
- No recoil broadening
- Extremely sharp resonance

This allows for the detection of the tiny frequency shifts caused by gravity over the 22.5 m height difference in the Harvard tower.

The success of the Pound-Rebka experiment demonstrates how quantum mechanics and general relativity make consistent predictions about photon behavior, though a complete quantum theory of gravity remains one of physics' greatest challenges.

# Chapter 38

## Black Body Radiation

Black body radiation represents one of the pivotal problems that led to the birth of quantum mechanics. While seemingly a purely thermodynamic phenomenon - the electromagnetic radiation emitted by an idealized perfect absorber in thermal equilibrium - its explanation required a radical departure from classical physics.

In the late 19th century, classical physics completely failed to explain the observed spectrum of black body radiation. The classical Rayleigh-Jeans law predicted that the intensity of radiation would increase indefinitely with frequency (the ‘ultraviolet catastrophe’), which clearly contradicted experimental measurements. This crisis in physics was resolved only when Max Planck introduced the revolutionary concept that electromagnetic energy could only be emitted in discrete quantities or ‘quanta’ - an idea that would become one of the fundamental principles of quantum mechanics.

The study of black body radiation thus marks the historical transition point from classical to quantum physics. Planck’s solution not only explained the observed radiation spectrum but introduced the quantum of action  $h$  (Planck’s constant), which would become central to all of quantum mechanics. This topic demonstrates how quantum effects emerge even in seemingly classical macroscopic systems when we examine them carefully enough.

The figure below displays the emission of a light bulb with a tungsten filament. The filament is heated up to a specific temperature by different currents.



Figure 38.1: Light emission at different temperatures. The filament of the light bulb is heated up to different temperatures by varying the current. The color of the emitted light changes with the temperature of the filament.

Without dispersing the spectrum we directly notice the different color of the light emitted by the filament. While the heating mechanism is different for different materials, the emitted spectrum is always similar and solely depends on the temperature of the radiator. The emitted spectrum is called the **blackbody spectrum** and is universal for all materials.

## 38.1 Blackbody

**i** Note

A blackbody is a model of a radiation source whose emission depends only on its temperature. Its emission, however, does not depend on the material the radiator is made from, nor on its surface or any other potential characteristics.

Consider a body with a cavity as depicted below. The body is heated to a certain temperature  $T$  and the cavity is closed.

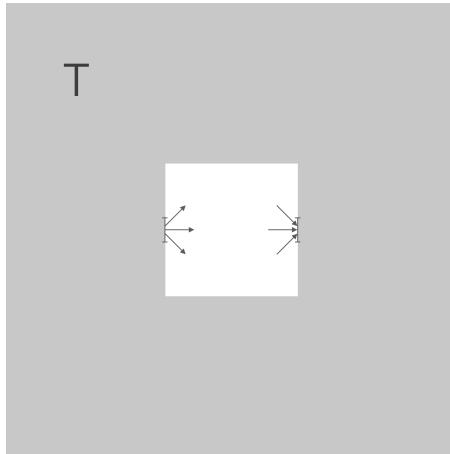


Figure 38.2: A body with a cavity. The body is heated to a certain temperature  $T$  and the cavity is closed but can be equipped with a tiny hole to probe the radiation inside the cavity without perturbation.

The system is in thermal equilibrium and thus each quadratic degree of freedom carries an energy of  $0.5k_B T$ , where  $k_B = 1.380649 \times 10^{-23}$  J/K is the Boltzmann constant.

Each surface element of the cavity emits radiation at a specific frequency  $\nu$ . The amount of power that is radiated by a surface element  $dA$  into a solid angle element  $d\Omega$  at a frequency interval  $d\nu$  relates to the property of its surface, i.e. the emissivity  $E_\nu^*$

$$\frac{dW_E}{dt} = E_\nu^* dA d\Omega d\nu$$

The emissivity is thereby the spectral radiance (energy emitted per unit area, per unit solid angle, per unit frequency), which has units of  $W \cdot m^{-2} \cdot sr^{-1} \cdot Hz^{-1}$ . In a similar way, each surface element at the cavity absorbs a certain power under a solid angle element and frequency interval. The absorbed power is given by

$$\frac{dW_A}{dt} = A_\nu S_\nu^* dA d\Omega d\nu$$

where  $S_\nu^*$  is the spectral radiance of an ideal black body, so the radiated power per unit area, per unit solid angle and per unit frequency interval. The absorption capability  $A_\nu$  is also a number between 0 and 1.

In the steady state, the body emits and absorbs the same amount of power and thus

$$\frac{dW_E}{dt} = \frac{dW_A}{dt}$$

which readily yields

$$E_{\nu}^* = A_{\nu} S_{\nu}^*$$

The spectral radiancy times the frequency dependent absorptivity is equal to the spectral emissivity. This is known as **Kirchhoff's law of thermal radiation**.

From the Kirchhoff's law we can see that a perfectly absorbing body  $A_{\nu} = 1$  is also a perfect emitter. This is the case for a *blackbody*, which absorbs all radiation incident on it and emits the maximum amount of radiation possible at a given temperature.

Note that in the case of a blackbody, the absorptivity does not depend anymore on the frequency  $\nu$ , while it will for a real body. In the case of a blackbody, the spectral radiancy  $S_{\nu}^*$  is then only a function of the temperature  $T$  and needs to be calculated from the cavity structure, i.e. the modes of the cavity, which we will calculate next.

### The Leslie Cube Experiment

The Leslie Cube, developed by John Leslie in 1804, was one of the first experimental demonstrations of how surface properties affect thermal radiation. The apparatus consists of a cubic vessel with different surface treatments on each face (e.g., polished metal, blackened surface, rough surface), filled with hot water.



Figure 38.3: Leslie Cube. The cube is filled with hot water, and each face has a different surface treatment. The radiation emitted from each face is measured using a thermopile detector. The experiment demonstrated that different surfaces emit radiation differently at the same temperature, and that good absorbers are also good emitters.

By measuring the radiation emitted from each face using a thermopile detector, Leslie showed that:

1. Different surfaces emit radiation differently at the same temperature
2. Good absorbers are also good emitters (leading to Kirchhoff's law)
3. The emissivity depends on the surface properties but not on the material inside

This simple but elegant experiment helped establish fundamental principles of thermal radiation and provided early experimental evidence for what would later be formalized as Kirchhoff's law of thermal radiation:

$$\frac{E_{\nu}^*}{A_{\nu}} = S_{\nu}^*$$

## 38.2 Spectral density of modes

To determine the spectral energy density in the cavity  $S_\nu^*$ , we need to calculate two quantities: the number of modes in the cavity and the average energy per mode. While calculating the number of modes is relatively straightforward, determining the average energy per mode is more complex and ultimately leads to Planck's law of radiation.

The number of modes in the cavity corresponds to the number of possible standing waves that can exist in the cavity. Let's first consider a simple one-dimensional case where the wave propagates along the z-direction, i.e.  $k = k_z$ . When an electric field  $\vec{E}(z, t) = E_0 \cos(\omega t - k_z z) \vec{e}_x$  is incident on a conducting surface at  $z = 0$ , the tangential components of the electric field must vanish. This leads to:

$$\vec{E}(z = 0, t) = E_{0,I} \vec{e}_x + E_{0,R} \vec{e}_x = 0$$

and

$$E_{0,I} \vec{e}_x = -E_{0,R} \vec{e}_x$$

The superposition of incident and reflected waves results in a standing wave:

$$\vec{E}(z, t) = 2E_0 \sin(k_z z) \sin(\omega t) \vec{e}_x,$$

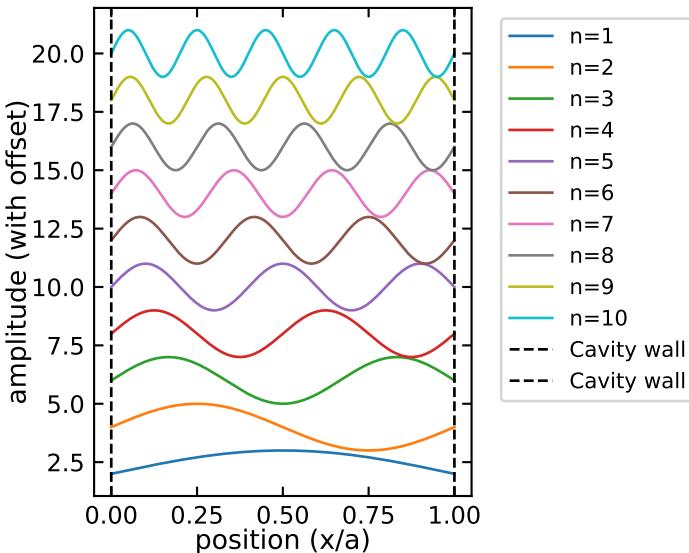
where the wavevector  $k_z$  is given by  $k_z = \omega/c_0$  with  $c_0$  being the speed of light in vacuum. For a second conducting surface at  $z = a$ , the boundary condition requires:

$$\vec{E}(z = a, t) = 2E_0 \sin(ka) \sin(\omega t) \vec{e}_x = 0$$

leading to:

$$k_z = o \frac{\pi}{a}$$

These represent the modes in a one-dimensional cavity.



For a three-dimensional cavity, we can analyze each direction independently, obtaining the following conditions for the wavevector components:

$$\begin{aligned}k_x &= n \frac{\pi}{a} \\k_y &= m \frac{\pi}{b} \\k_z &= o \frac{\pi}{c}\end{aligned}$$

where  $m, n, o$  are natural numbers. The magnitude of the wavevector is  $|\vec{k}| = \sqrt{k_x^2 + k_y^2 + k_z^2}$ , which determines the possible frequencies  $\omega$  in the cavity

$$\omega = c \cdot \pi \sqrt{\left(\frac{o}{a}\right)^2 + \left(\frac{m}{b}\right)^2 + \left(\frac{n}{c}\right)^2}$$

### i Resonator modes

The standing waves which follow from these frequencies in our resonator are given by:

$$\vec{E}_{m,n,o} = \vec{E}_0(m, n, o) \cdot \cos(\omega t)$$

with  $\vec{E}_0(m, n, o) = (E_{0,x}, E_{0,y}, E_{0,z})^T$  and

$$\begin{aligned}E_{0,x} &= A \cdot \cos\left(n \frac{\pi}{a} x\right) \sin\left(m \frac{\pi}{b} y\right) \sin\left(o \frac{\pi}{c} z\right) \\E_{0,y} &= B \cdot \sin\left(n \frac{\pi}{a} x\right) \cos\left(m \frac{\pi}{b} y\right) \sin\left(o \frac{\pi}{c} z\right) \\E_{0,z} &= C \cdot \sin\left(n \frac{\pi}{a} x\right) \sin\left(m \frac{\pi}{b} y\right) \cos\left(o \frac{\pi}{c} z\right)\end{aligned}$$

This system, comprising a box with ideally conducting walls, is known as a *cavity resonator*, and the possible standing waves are called the resonator's *principle oscillations* or **resonator modes**.

As mentioned before, we are interested in the number of modes that fit into the cavity. For that purpose we need to count the number of modes within a certain frequency range. In order to simplify the calculation a bit, we restrict our box to a cube with edge length  $a$  such that

$$\begin{aligned}\omega &= c \cdot \frac{\pi}{a} \sqrt{n^2 + m^2 + o^2} \\&\rightarrow n^2 + m^2 + o^2 = \left(\frac{a\omega}{\pi c}\right)^2 \\&\rightarrow n^2 + m^2 + o^2 = \left(\frac{a}{\pi}\right)^2 \cdot k^2\end{aligned}$$

We can visualize the possible modes in k-space (where  $k_x$ ,  $k_y$ , and  $k_z$  are our axes). The points  $(m, n, o)$  create an evenly-spaced grid with spacing  $\pi/a$  between points. Since each combination of  $(m, n, o)$  represents one mode in the resonator, counting the grid points, or better the number of unit cell cubes with a length  $\pi/a$  tells us the number of possible modes.

When considering large values where  $\sqrt{m^2 + n^2 + o^2} \gg 1$ , the sphere radius in k-space  $|\vec{k}|$  becomes much larger than  $\pi/a$ , corresponding to wavelengths  $\lambda$  much smaller than the cavity size  $a$ . In this limit, we can approximate the number of allowed modes  $N_L$  (where  $m, n, o > 0$ ) by calculating the volume occupied by unit cells within the first octant of a sphere with radius  $|\vec{k}|$ . The volume of this octant is:

$$\begin{aligned}V_S &= \frac{1}{8} \frac{4}{3} \pi |\vec{k}|^3 \text{ or} \\V_S &= \frac{1}{6} \pi \left(\frac{\omega}{c_0}\right)^3.\end{aligned}$$

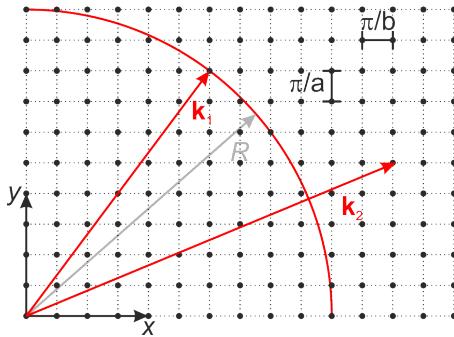


Figure 38.4: Two-dimensional k-space with a circle representing a sphere in two-dimensional space.

The number modes corresponds then to the volume of the octant to the volume of a unit cell ( $V_{UC} = (\pi/a)^3$ ) which is

$$N_L = \frac{V_S}{V_{UC}} = \frac{\pi}{6} \left( \frac{a \cdot \omega}{\pi \cdot c_0} \right)^3.$$

Since each standing wave can have two independent polarization states, the total number of modes below frequency  $\omega_S$  is twice the previous result and we find the total number of modes up to a limiting frequency  $\omega_S$

$$N(\omega \leq \omega_S) = 2 \cdot \frac{\pi}{6} \left( \frac{a \cdot \omega_S}{\pi \cdot c_0} \right)^3 = \frac{8\pi\nu_S^3}{3c_0^3}.$$

Here we made use of  $\omega_S = 2\pi\nu_S$ . The number of modes per volume of the cavity  $a^3$  is thus given by

$$\frac{N(\nu \leq \nu_S)}{a^3} = n = \frac{8\pi\nu_S^3}{3c_0^3}.$$

When we now increase the limiting frequency  $\nu_S$ , the number of modes increases. This increase is best represented by the **spectral density of modes**, that is the number of allowed modes per unit volume of the resonator within the interval  $[\nu; \nu + d\nu]$ . A straightforward calculation of the first derivative of the density of modes with respect to the frequency gives us an expression for the spectral mode density

$$dn(\nu) = \frac{8\pi\nu^2}{c_0^3} d\nu$$

So far we have calculated that only under particular conditions standing waves can be established within a cubic cavity. These eigen-oscillations are called “modes of the cavity”. Furthermore, if the wavelength is small compared to the cavity dimensions, we derived the spectral density of modes, that is the number of modes within one cubic meter of volume within the interval between  $\nu$  and  $\nu + d\nu$  is given by the equation above. This is the spectral density of optical modes per volume.

### Density of optical modes per unit volume

This equation represents the density of optical modes per unit volume and frequency interval in free space. It has important consequences for lasing:

1. The density of modes increases quadratically with frequency ( $\nu^2$ )
2. This means there are more available modes at higher frequencies
3. For lasers, this affects:
  - The threshold conditions for lasing
  - The emission probability at different frequencies

- The competition between modes

In practical terms, it helps explain why it's generally easier to achieve lasing at shorter wavelengths (higher frequencies) where there are more available modes, although other factors like gain and losses also play crucial roles.

This spectral mode density is now the basis for any further consideration. To obtain the spectral energy density we just need to multiply the mode density with the energy stored in each mode  $\bar{W}_\nu(T)$  and obtain the spectral energy density

$$\omega_\nu \, d\nu = dn(\nu) \cdot \bar{W}_\nu(T) \cdot d\nu = \frac{8\pi\nu^2}{c_0^3} \cdot \bar{W}_\nu(T) \cdot d\nu$$

This is the energy density at a frequency  $\nu$  in a tiny interval  $d\nu$  and the temperature  $T$ . This spectral energy density  $\omega_\nu$  relates to the spectral radiance  $S_\nu^*$  through:

$$S_\nu^* = \frac{c_0}{4\pi} \omega_\nu$$

### 38.3 Rayleigh–Jeans law

One way to obtain a mean energy per mode is to consider classical thermal equilibrium and equipartition. In this case, each quadratic degree of freedom contains  $0.5k_B T$  energy such that

$$\bar{W}_\nu(T) = k_B \cdot T,$$

with  $k_B$  and  $T$  being the Boltzmann constant and absolute temperature, respectively. Therefore, within the limit of the classical approach the spectral energy density,

$$\omega_\nu \, d\nu = \frac{8\pi\nu^2}{c_0^3} k_B T \, d\nu,$$

rises quadratically with respect to the frequency  $\nu$ . This quadratic relation is known as **Rayleigh–Jeans law**. As a consequence a small hole in the cavity wall will then emit radiation into the solid angle of  $d\Omega = 1$  sr with the radiance of

$$\begin{aligned} S_\nu^*(\nu) \, d\nu &= \frac{c_0}{4\pi} \omega_\nu(\nu) \, d\nu \\ &= \frac{2\nu^2}{c_0^2} k_B T \, d\nu \end{aligned}$$

If we now consider a temperature of about 5000 K we achieve a wavelength bigger than  $2 \mu\text{m}$ , being well in the infrared region. For this spectral region the measured radiance and the theoretical prediction are in agreement. However, if we reduce the wavelength, disparities between experimental findings and the prediction appear. Moreover, if the Rayleigh–Jeans law was valid, there would be the so-called **ultraviolet catastrophe**! In the case of decreasing frequencies, the spectral energy density and the integrated radiance will rise until they become infinitely big for vanishing frequencies.

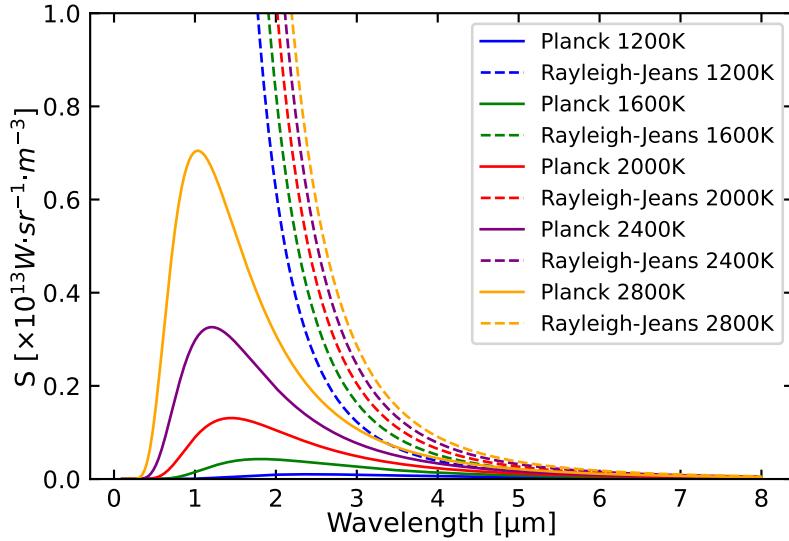


Figure 38.5: Comparison between blackbody radiation (solid lines) and radiation as described through the Rayleigh-Jeans law (dashed lines) at different temperatures.

While the Rayleigh Jeans Law is corrected with Planck's law, it demonstrates the failure of classical physics to describe the blackbody radiation. The Rayleigh-Jeans law predicts that the spectral energy density increases with frequency, leading to the ultraviolet catastrophe. This discrepancy between theory and experiment was one of the key motivations for the development of quantum mechanics.

### 38.4 Planck's law

In 1900 Max Planck tackled the problem of how to avoid the ultraviolet catastrophe and describe the blackbody radiation as a whole. He assumed that the energy of the absorbing and emitting oscillators in the wall can release or absorb small packets of energy given by  $h \cdot \nu$ , where  $h$  was used as a **helper constant**. Taking the limit  $h \rightarrow 0$  should then lead to the classical result, yet Planck found that the classical limit was not reached. Rather, the quantization of energy with  $h = 6.626 \dots \cdot 10^{-34}$  Js led to a new law of radiation, which was in perfect agreement with the experimental findings. The energy at a particular frequency  $\nu$  is then quantized in terms of the number of energy packets  $n$  as

$$W_\nu = n \cdot h \cdot \nu$$

where  $n$  is a natural number. If we now consider thermal equilibrium, the probability  $p(W_\nu)$  of finding an oscillator with energy  $W_\nu = nh\nu$  (meaning the eigenstate is occupied by  $n$  photons) is given by the Boltzmann distribution:

$$p(W_\nu) = \frac{e^{-\frac{n \cdot h \cdot \nu}{k_B T}}}{\sum_{n=0}^{\infty} e^{-\frac{n \cdot h \cdot \nu}{k_B T}}}$$

Note that a Boltzmann distribution provides the maximum entropy for a given energy, i.e. it spreads the energy as evenly as possible over the available states (modes) of the systems. The average energy per oscillator can then be calculated as the expectation value of energy, weighting each possible energy level by its probability of occupation. The averaged energy per oscillator then reads as

$$\bar{W}_\nu = \sum_{n=0}^{\infty} p(nh\nu) n h \nu$$

and further

$$\bar{W}_\nu = \frac{h \cdot \nu}{e^{\frac{h \cdot \nu}{k_B T}} - 1}.$$

The spectral energy density in the cavity is thus given by

$$\omega(\nu, T) = dn(\nu) \cdot \bar{W}_\nu(\nu, T)$$

which leads us to the famous Planck's formula

$$\omega(\nu, T) d\nu = \frac{8\pi h\nu^3}{c_0^3} \frac{d\nu}{e^{\frac{h\nu}{k_B T}} - 1}.$$

Here  $\omega(\nu, T) d\nu$  represents the spectral distribution of the energy density per frequency interval; its unit is  $[\omega(\nu, T)] = \text{Jsm}^3$ . The radiance of the area element  $dA$  emitted into the solid angle  $d\Omega$  then is

$$\begin{aligned} S_\nu^* d\nu d\Omega &= \frac{c_0}{4\pi} \omega_\nu(\nu, T) d\nu d\Omega \\ &= \frac{2h\nu^3}{c_0^2} \frac{d\nu d\Omega}{e^{\frac{h\nu}{k_B T}} - 1}. \end{aligned}$$

Planck's theory posited that oscillator energy exists in discrete units or packets denoted by  $n$ , following the relation  $W_\nu = n \cdot h \cdot \nu$ . This concept was later expanded by Einstein in his explanation of the photoelectric effect, demonstrating that light itself consists of discrete energy quanta proportional to frequency  $\nu$ . Einstein's interpretation, expressing the energy of the electromagnetic field as  $E = h\nu$ , established the foundation for understanding light as discrete particles called *photons*.

While we've previously written Planck's law in terms of frequency  $\nu$ , we can alternatively express it using wavelength  $\lambda$  through the relationship  $\lambda = c/\nu$ . Making this conversion requires noting that  $d\lambda = -(c/\nu^2) d\nu$ . This allows us to express the spectral energy density

$$\omega(\lambda, T) d\lambda = \frac{8\pi h c_0}{\lambda^5} \frac{d\lambda}{e^{\frac{hc_0}{\lambda k_B T}} - 1}$$

and the radiance

$$S_\lambda^* d\lambda d\Omega = \frac{2hc^2}{\lambda^5} \frac{d\lambda d\Omega}{e^{\frac{hc}{\lambda k_B T}} - 1}$$

in terms of wavelength  $\lambda$  and temperature  $T$ .

### Measuring Planck's Constant

Planck's constant  $h$  is a fundamental constant of nature, and its precise value is crucial for many areas of physics. The most accurate method for measuring  $h$  is through the photoelectric effect, as demonstrated by Einstein. One could also determine  $h$  by measuring the Compton edge in X-ray scattering experiments or estimating by measuring the uncertainty in the particle momentum and position from a diffraction experiment.

Another possibility arises from counting single photons from a laser beam. The particle nature of light now delivers the definition of light power by the relation

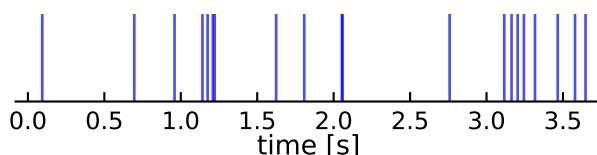
$$P = N \cdot E_{\text{photon}} = N \cdot h \cdot \nu = N \cdot h \cdot \frac{c}{\lambda}$$

where  $N$  is the number of photons per second. If we know the wavelength of the laser light, we can

determine Planck's constant by measuring the power of the laser beam and the number of photons per second. We can easily arrive at the relation

$$h = \frac{P\lambda}{Nc}$$

In reality, measuring the photon arrival rate is a challenging task. For coherent laser light, the photon arrivals follow Poisson statistics, meaning the time intervals between consecutive photons are exponentially distributed. While this makes the measurements theoretically straightforward, the uncertainty in counting photons (which follows the square root of the mean count rate according to Poisson statistics) is usually the limiting factor in the precision of the measurement. Yet, we can dim the laser light to a point where only one photon is present at a time. Note that while diming the laser light with an appropriate filter, the photon rate decreases, but the photon statistics remain Poissonian. Only the average rate of photon arrivals decreases such that most of the time no photon arrives.



By measuring the time between two photon arrivals, we can determine the photon rate and thus Planck's constant. Single photons are typically detected with the help of an Avalanche Photodiode (APD). The APD is a semiconductor device that can detect single photons by amplifying the signal from a single electron-hole pair created by the photon.



Figure 38.6: An APD is a special type of photodiode that can detect single photons by amplifying the signal from a single electron-hole pair created by the photon.

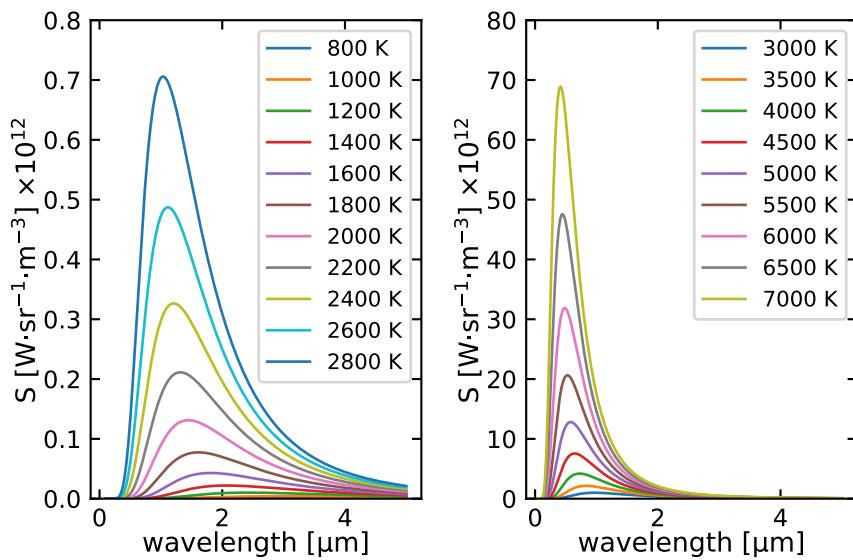


Figure 38.7: Planck's law of blackbody radiation at different temperatures.

The sun is a blackbody radiator with a temperature of about 5778 K. The spectral radiance of the sun is given by Planck's law.

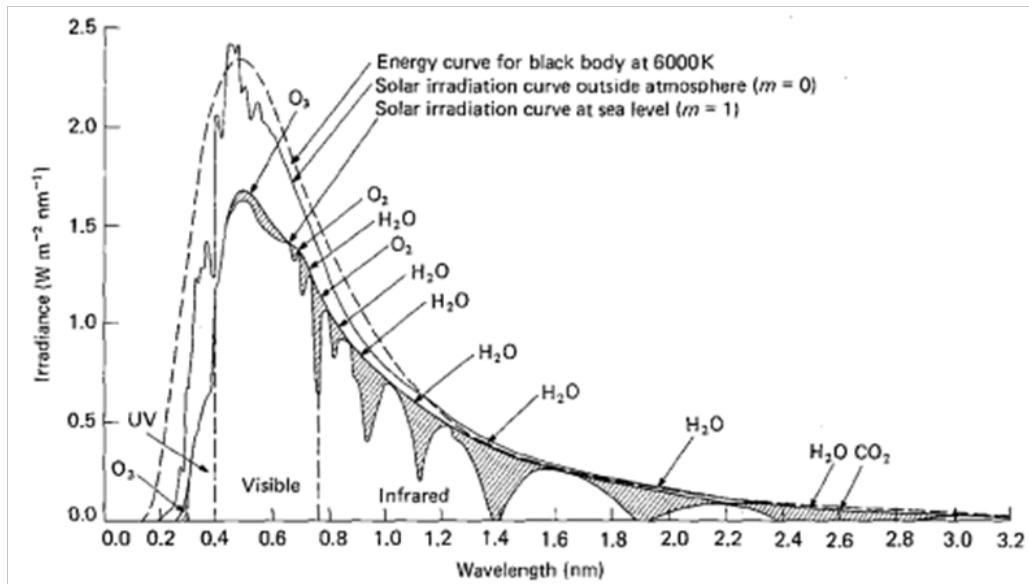


Figure 38.8: The sun spectrum at the earth as compared to Planck's law. Due to the presence of absorption lines of water and other components in the atmosphere, the spectrum is not a perfect blackbody spectrum.

Since the spectrum is completely defined by the temperature of the blackbody, it is possible to estimate the temperature of the sun or other stars by comparing the spectrum of the sun with Planck's law.

## Properties of Planck's Radiation Formula

### **i** Planck's Radiation Law

The fundamental equation describing the spectral radiance of a black body:

$$S(\lambda, T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k_B T}} - 1}$$

where:

- $S(\lambda, T)$  is the spectral radiance [ $\text{W} \cdot \text{sr}^{-1} \cdot \text{m}^{-3}$ ]
- $h$  is Planck's constant [ $\text{J} \cdot \text{s}$ ]
- $c$  is the speed of light [ $\text{m/s}$ ]
- $k_B$  is the Boltzmann constant [ $\text{J/K}$ ]
- $T$  is absolute temperature [ $\text{K}$ ]
- $\lambda$  is wavelength [ $\text{m}$ ]

Note that Planck's radiation law is only valid for propagating electromagnetic modes. Evanescent modes are not included in this formula and do not follow Planck's radiation law.

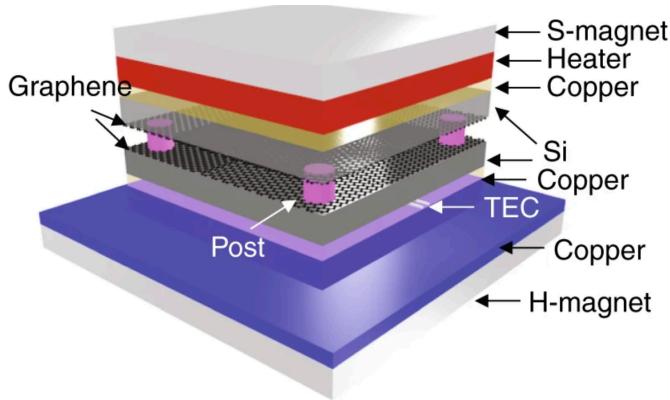


Figure 38.9: A schematic of the home-made measurement setup. From top to bottom, the emitter side consists of S-magnet, heater, copper spreader, and Si/Gr and the receiver side consisting of Gr/Si, copper spreader, TEC layer, copper spreader, and H-magnet. Four photoresist posts are used to separate the emitter and the receiver. from [Observing of the super-Planckian near-field thermal radiation between graphene sheets](#)

This opens a number of possibilities which are used for solar energy collection beyond the Shockley limit, for example in thermophotovoltaic cells, which convert the spectral absorbance into heat radiation, that is shaped spectrally by metamaterials to match the bandgap of the solar cell.

### **i** Stefan-Boltzmann Law

The total power radiated per unit area across all wavelengths:

$$E = \sigma T^4$$

This can be derived by integrating Planck's law over all wavelengths:

$$E = \int_0^\infty S(\lambda, T) d\lambda = \sigma T^4$$

where  $\sigma = \frac{2\pi^5 k^4}{15c^2 h^3} \approx 5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$