PHYS231 Waves and Optics

Michaelmas Term 2009

Volker Perlick, v.perlick@lancaster.ac.uk

Office Hour: Wed 1-2 pm (C35)

Recommended reading:
F. Pedrotti and L. Pedrotti: ***Introduction to Optics***

I. Revision of basic geometric optics

(cf. Pedrotti and Pedrotti, Chapter 3)

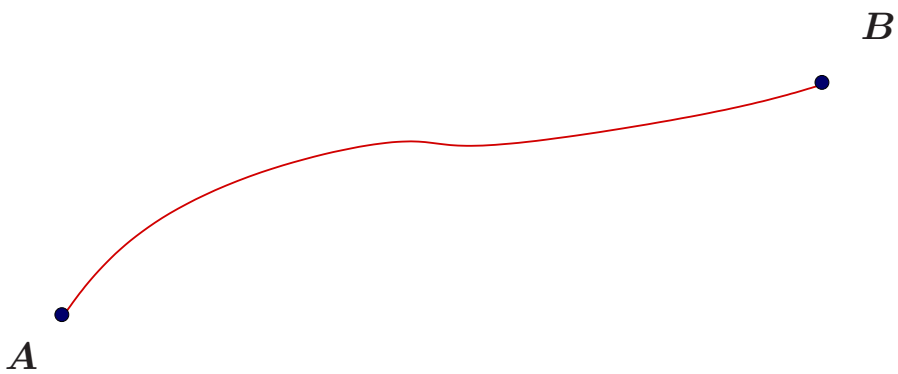
The simplest way to describe light is in terms of rays.
This is called ray optics or geometric optics.
All of geometric optics can be based on Fermat's Principle.

a) Fermat's Principle

Pierre de Fermat (1601–1665)
French lawyer, politician, mathematician

Original version of Fermat's Principle:

> **The actual path between two points taken by a light ray is the one which is traversed in the least time.**
>
> **Pierre de Fermat (1657)**



Applying Fermat's Principle to light rays in some medium requires to know along each trial path the velocity of light

$$\boxed{v = \frac{c}{n}}$$

in this medium. Here

$$c = \text{velocity of light in vacuo } (\approx 300\,000 \text{ km/sec})$$

$$n = \text{index of refraction}$$

Caveats:

- **Modifications of Fermat's Principle are necessary for moving media (e.g., light traveling in moving fluid).**
- **Frequency of light has to be fixed if velocity of light depends on frequency ("dispersive medium").**
- **Actually, the travel time need not be minimal, only stationary, i.e., a minimum, a maximum or a saddle, see examples below.**

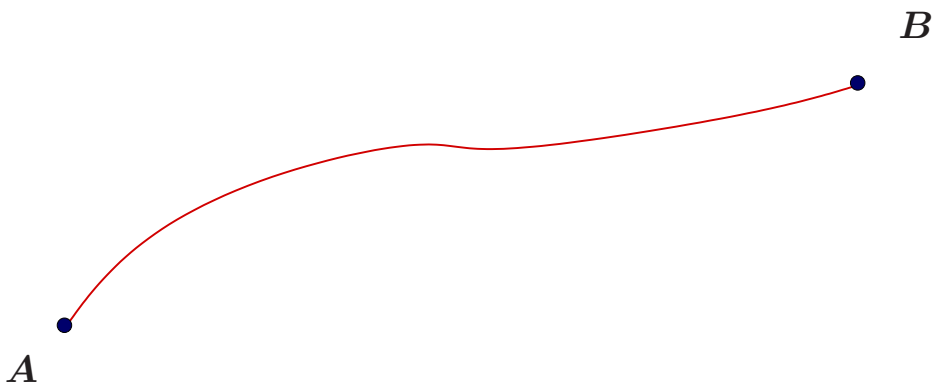
Fermat's Principle has analogue in mechanics ("Optical-mechanical analogy"):

ray optics	$\longleftrightarrow$	**classical mechanics**
ray = path of photon	$\longleftrightarrow$	**path of particle**
Fermat's Principle	$\longleftrightarrow$	**Maupertuis' Principle**
wave optics	$\longleftrightarrow$	**quantum mechanics**
electromagnetic wave	$\longleftrightarrow$	**probability wave**

In the following we discuss several applications of Fermat's Principle to basic geometric optics.

b) Medium with n = constant

Assume that the index of refraction n is constant.
(Recall: The velocity of light is $v = \frac{c}{n}$.)



$$\text{travel time } t = \text{geometric length} \cdot \frac{n}{c}$$

minimal travel time = minimal geometric length

Fermat's Principle $\Rightarrow$ light rays are straight lines

[maxima or saddles cannot occur in this case]

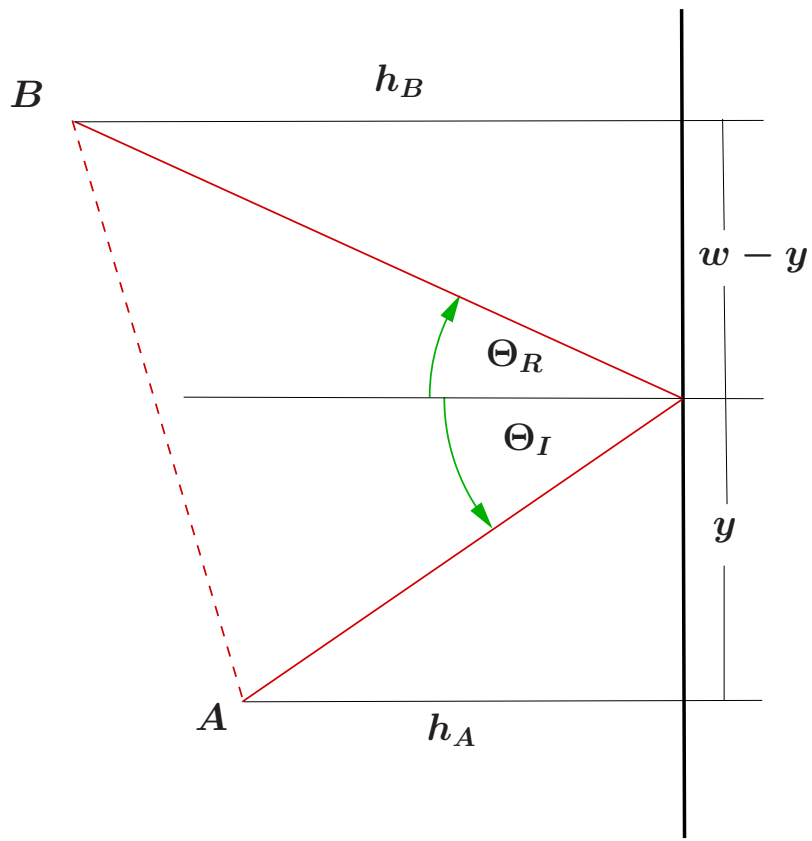
c) Reflection law

(was known already to ancient Greeks)

Consider medium of constant index of refraction n, bounded by reflecting plane.

Then Fermat's Principle yields two solutions for rays from A to B: The connecting straight line, and a reflected ray.

Derivation of the reflection law:



By symmetry, rays must lie in plane perpendicular to reflecting plane.

$$\text{travel time } t = \text{geometric length} \cdot \frac{n}{c} =$$

$$= \left(\sqrt{h_A^2 + y^2} + \sqrt{h_B^2 + (w-y)^2}\right) \frac{n}{c}$$

$$0 \stackrel{!}{=} \frac{dt}{dy} = \left(\frac{\cancel{2}y}{\cancel{2}\sqrt{h_A^2+y^2}} + \frac{\cancel{2}(w-y)(-1)}{\cancel{2}\sqrt{h_B^2+(w-y)^2}}\right) \frac{n}{c}$$

$$\frac{y}{\sqrt{h_A^2+y^2}} = \frac{w-y}{\sqrt{h_B^2+(w-y)^2}}$$

$$\sin \Theta_I = \sin \Theta_R$$

As $-\frac{\pi}{2} \leq \Theta_I \leq \frac{\pi}{2}$ and $-\frac{\pi}{2} \leq \Theta_R \leq \frac{\pi}{2}$:

$$\boxed{\Theta_I = \Theta_R}$$ This is the reflection law

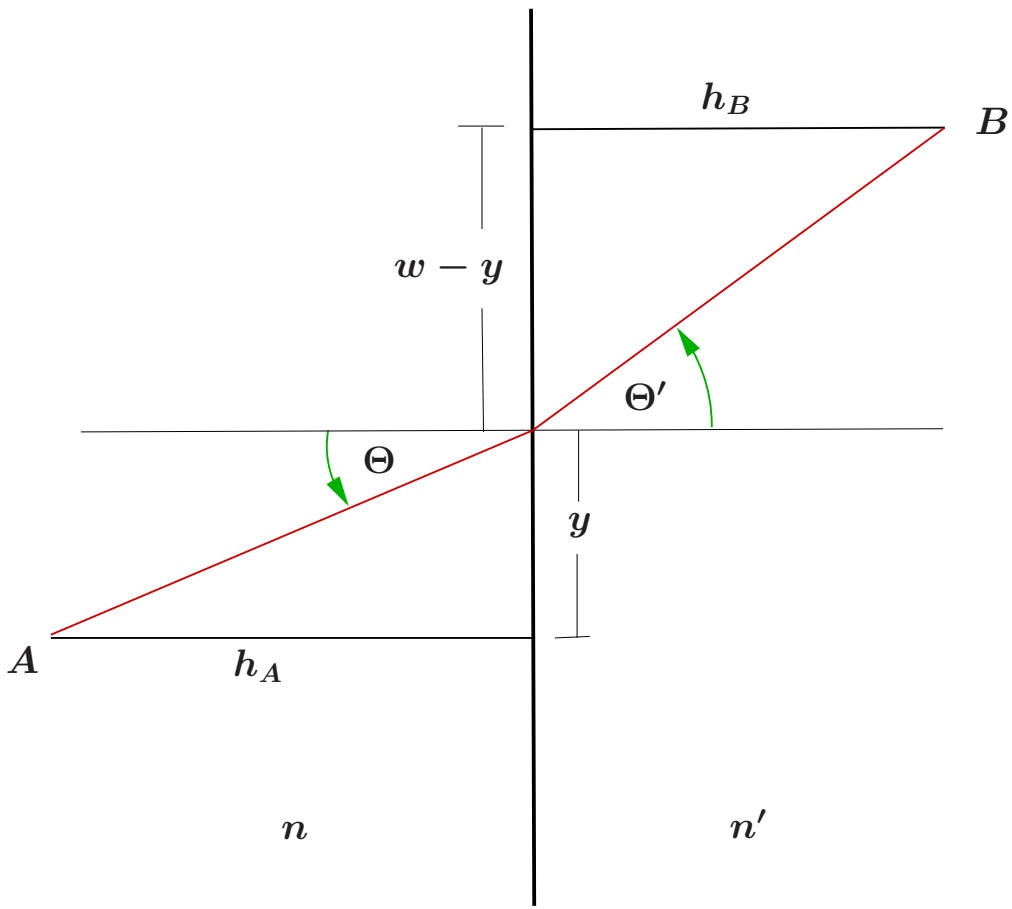
- The solution is always a minimum with respect to all broken straight lines. (Proof: Consider second derivative.) However, it is a saddle if arbitrarily curved trial paths are allowed.
- The reflection law holds for curved reflecting surfaces as well. (Proof: Apply above argument to the tangent plane.)

d) Refraction law

(experimentally discovered by Snellius in 1621)

Consider medium with constant index of refraction n in one half-space, medium with constant index of refraction n' in the other half-space.

Derivation of the refraction law:



By symmetry, rays must lie in plane perpendicular to boundary plane.

$$\text{travel time } t =$$

$$= \frac{n}{c}\sqrt{h_A^2 + y^2} + \frac{n'}{c}\sqrt{h_B^2 + (w-y)^2}$$

$$0 \stackrel{!}{=} \frac{dt}{dy} = \frac{\cancel{2}ny}{\cancel{2}c\sqrt{h_A^2+y^2}} + \frac{\cancel{2}n'(w-y)(-1)}{\cancel{2}c\sqrt{h_B^2+(w-y)^2}}$$

$$\frac{ny}{\sqrt{h_A^2+y^2}} = \frac{n'(w-y)}{\sqrt{h_B^2+(w-y)^2}}$$

$$\boxed{n \sin\Theta = n' \sin\Theta'}$$

This is Snell's law.

- The solution is always a minimum. (Proof: Consider second derivative.)
- The refraction law holds for curved surfaces as well, e.g. for lenses. (Proof: Apply above argument to the tangent plane.)

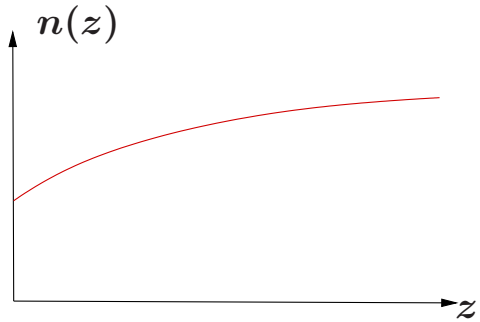
e) Medium with spatially variable n

If n is not a constant, it is not the geometric length $\int d\ell$ but the

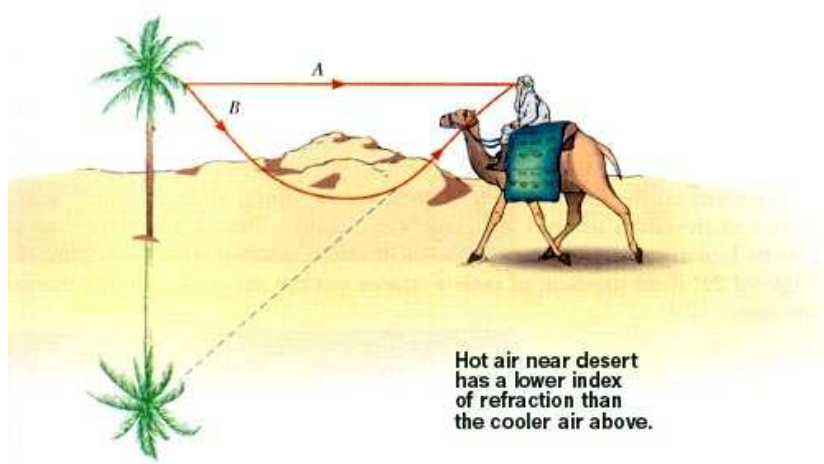
$$\boxed{\text{optical path length} = \int n\, d\ell}$$

that is minimal (or stationary). As a consequence, the rays are curved.

Example:
Air above a hot surface
(z is the height)



(inferior) mirage:

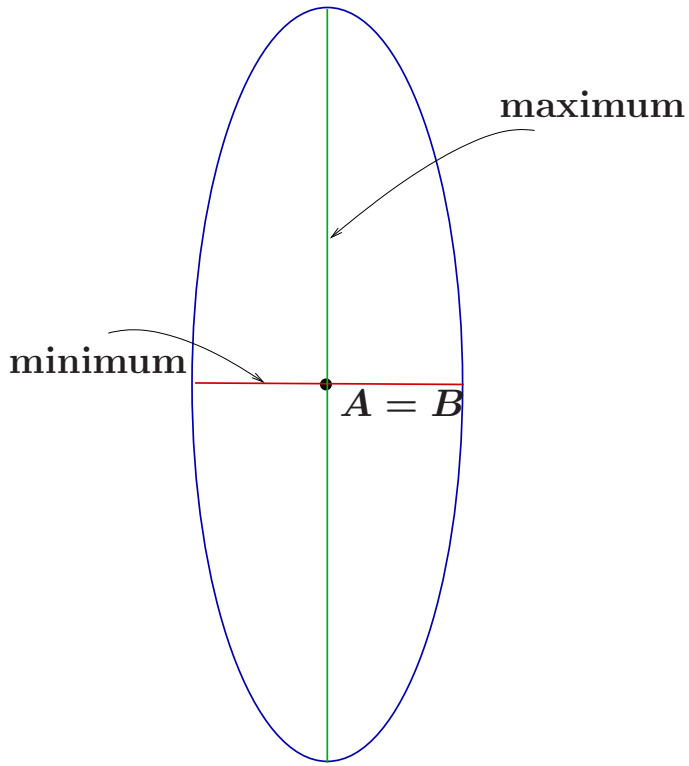


Picture from http://sol.sci.uop.edu

Requires temperature gradients of at least 2 degrees per meter.

f) Example where solution is not a minimum

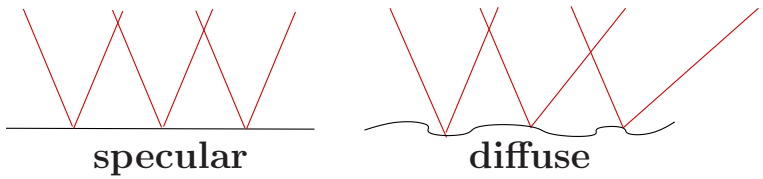
Consider the reflection at the inner side of an ellipsoid. Then, with respect to broken straight lines, the solution may be a minimum or a maximum.



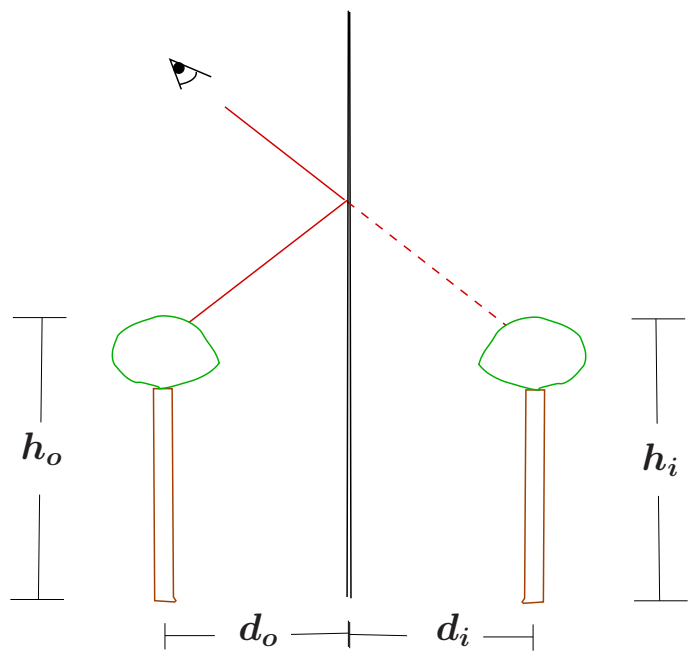
Remark: With respect to arbitrary trial curves (i.e., not only broken straight lines), Fermat's Principle never yields a maximum (either minimum or saddle).

g) Imaging by mirrors of various shapes

Always consider specular reflection



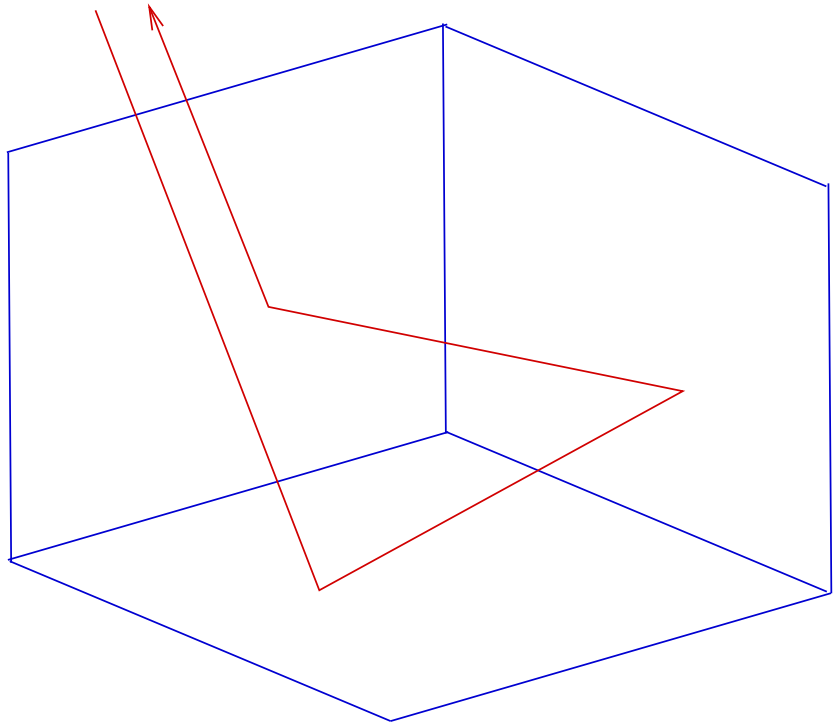
α) Plane mirror



Virtual image, upright and side-inverted,

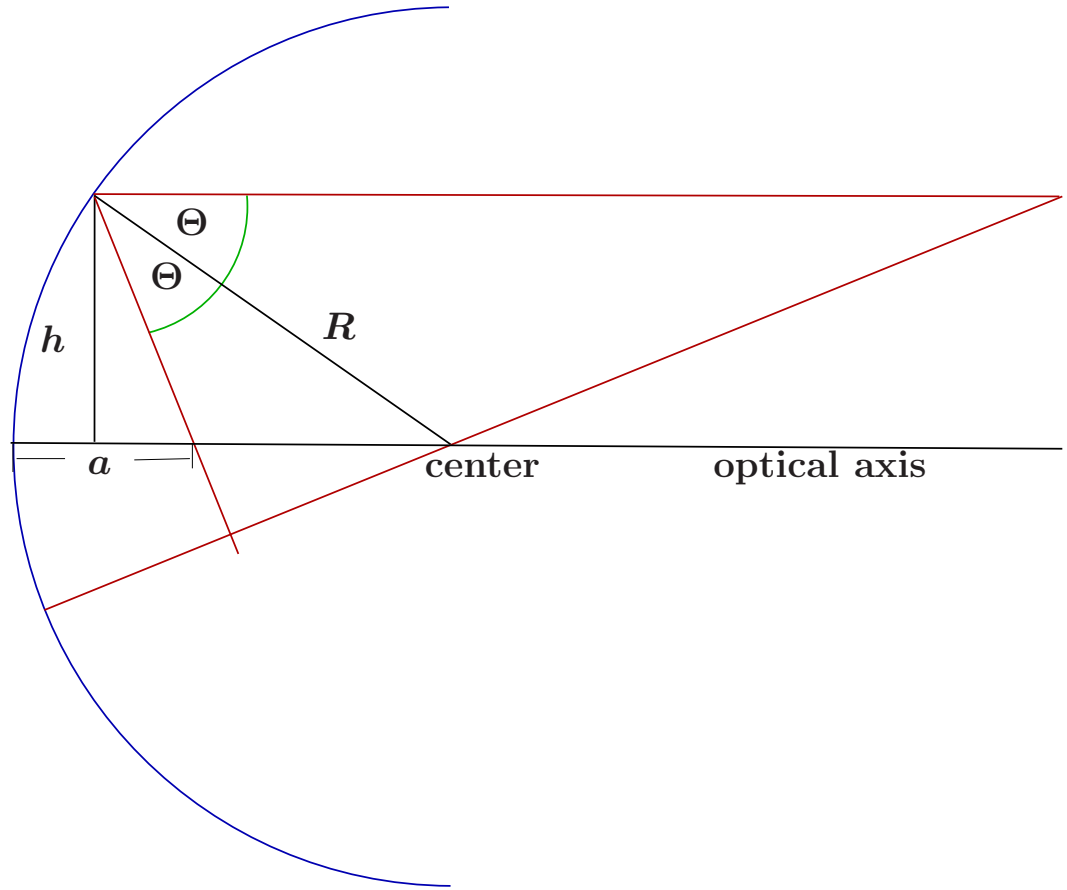
$$d_o = d_i \text{ and } h_o = h_i.$$

β) Corner mirror



outgoing ray always parallel to incoming ray

γ) Concave spherical mirror



- **Ray through center is always reflected in itself.**
- **What about ray parallel to optical axis?**
  h given, determine a:

$$\text{sine theorem:} \qquad \frac{R}{\sin(\pi - 2\Theta)} = \frac{R-a}{\sin\Theta}$$

$$R\sin\Theta = (R-a)\sin 2\Theta \;=\; (R-a)\,2\sin\Theta\cos\Theta$$

$$R \;=\; 2\,(R-a)\cos\Theta \qquad (*)$$

$$\frac{h}{R} = \sin\Theta\,, \qquad \cos\Theta = \sqrt{1-\frac{h^2}{R^2}} \qquad (**)$$

$$(*) \text{ and } (**) \;\Longrightarrow R = 2(R-a)\sqrt{1-\frac{h^2}{R^2}}$$

$$a = R - \frac{R^2}{2\sqrt{R^2-h^2}}$$

a depends on h, i.e., no perfect focusing

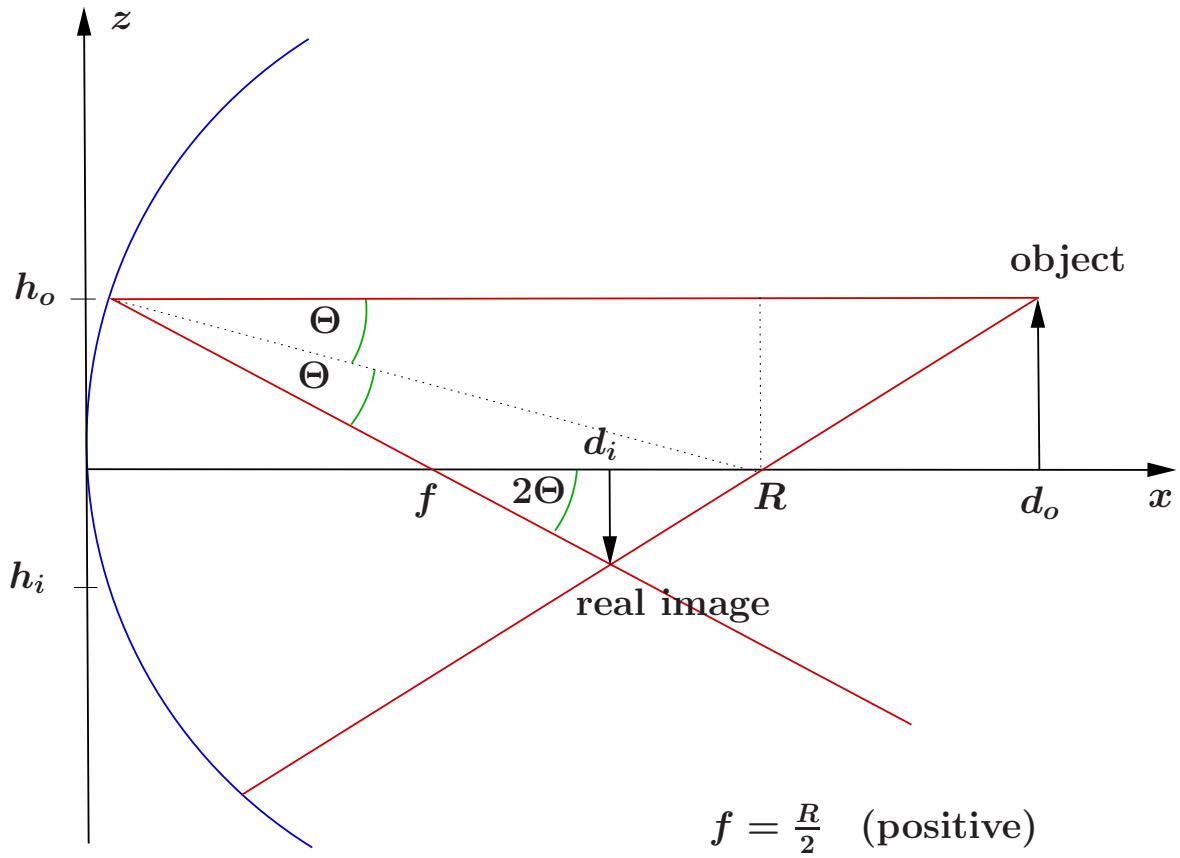
First order approximation (Gaussian optics, paraxial rays):

$\sin\Theta \approx \Theta\,,\ \cos\Theta \approx 1\,,\ R^2 - h^2 \approx R^2\,,\ a \approx \frac{R}{2}$

First order approximation for a gives focal length of spherical concave mirror:

$$\boxed{f = \frac{R}{2}}$$

Imaging properties of concave spherical mirror in first order approximation:



d_o distance of object (positive by choice of axes)

h_o height of object (positive by choice of axes)

d_i distance of image (positive if on the same side of the mirror as object)

h_i height of image (negative if inverted)

Goal: Determine d_i and h_i if f, d_o and h_o are given.

$$\frac{h_o}{d_o} = -\frac{h_i}{d_i} \qquad (*)$$

$$\left.\begin{aligned} \tan(2\Theta) &= \frac{-h_i}{d_i - f} \approx 2\Theta \\ \sin\Theta &= \frac{h_o}{2f} \approx \Theta \end{aligned}\right\} \Longrightarrow \frac{-h_i}{d_i - f} = \frac{h_o}{f} \qquad (**)$$

$$(*) \quad \text{and} \quad (**) \quad \Longrightarrow \quad \frac{d_o}{di} = \frac{f}{d_i - f}$$

$$d_o\, d_i \;-\; d_o\, f \;=\; d_i\, f \qquad \mid\; : \; d_i\, d_o\, f$$

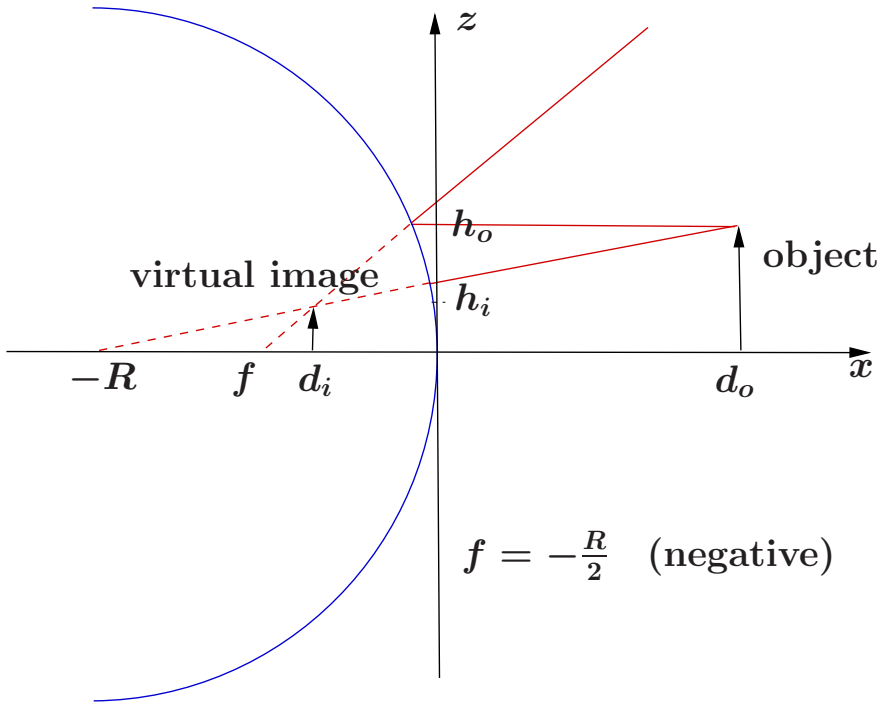
$$\boxed{\frac{1}{f} = \frac{1}{d_i} + \frac{1}{d_o}}$$

This is the mirror equation. With d_i known, $(*)$ now gives the magnification

$$\boxed{m = \frac{h_i}{h_o} = -\frac{d_i}{d_o}}$$

For $d_o > f$, the image is real and inverted, thus $m < 0$.
For $d_o < f$, the image is virtual and upright, thus $m > 0$.

δ) Convex spherical mirror



d_o distance of object (positive by choice of axes)

h_o height of object (positive by choice of axes)

d_i distance of image (negative if on other side of the mirror than object)

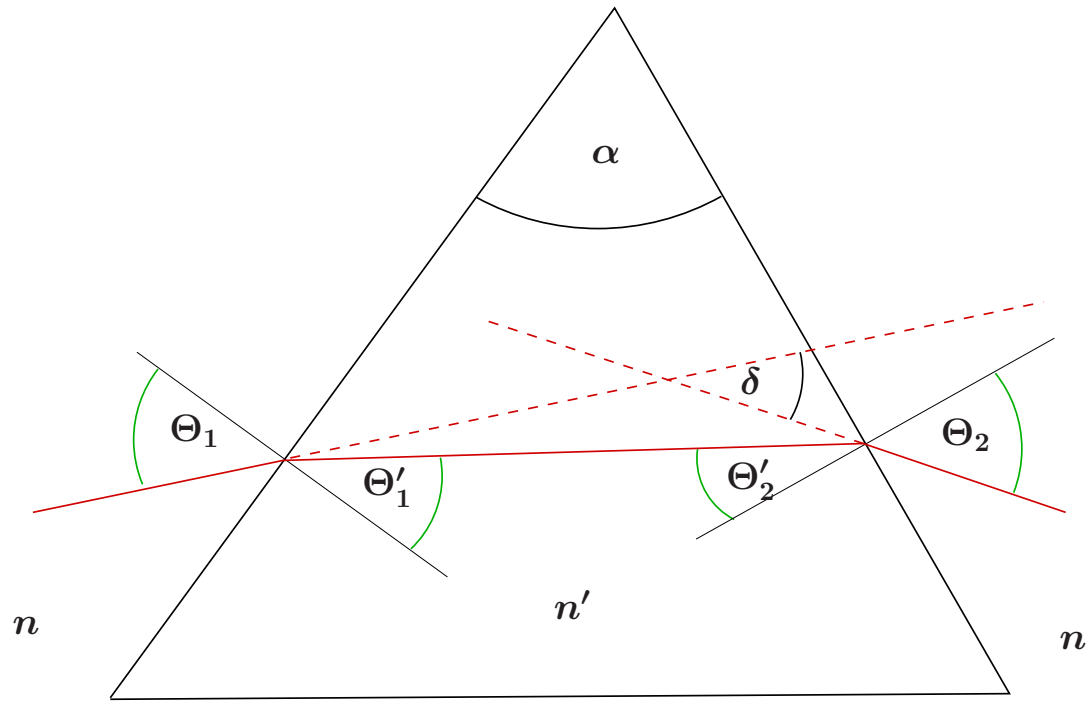
h_i height of image (positive if upright)

Calculation analogous to concave case: $\boxed{\frac{1}{f} = \frac{1}{d_i} + \frac{1}{d_o}}$ $\boxed{m = \frac{h_i}{h_o} = -\frac{d_i}{d_o}}$

Image is always virtual and upright, thus $m > 0$.

h) Refraction by bodies of various shapes

α) Prisms



Prism: Index of refraction n'

Ambient medium: Index of refraction n

$n' > n \implies \delta > 0$, i.e., bending away from apex

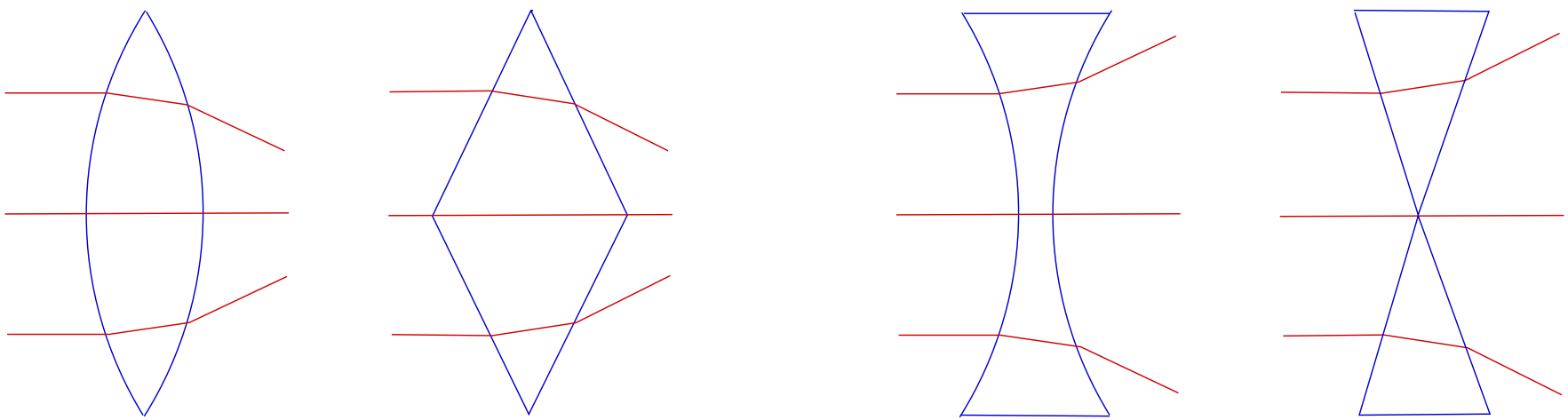
β) Spherical lenses

Lens: Index of refraction n'

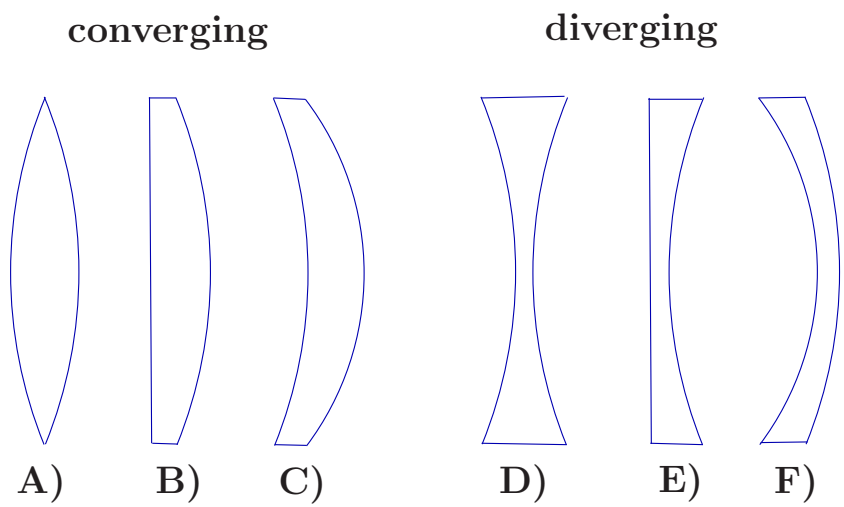
Ambient medium: Index of refraction n

Assume $n' > n$

Spherical lens can be approximated by prisms:



converging effect **diverging effect**



A) double-convex
B) plano-convex
C) concave-convex

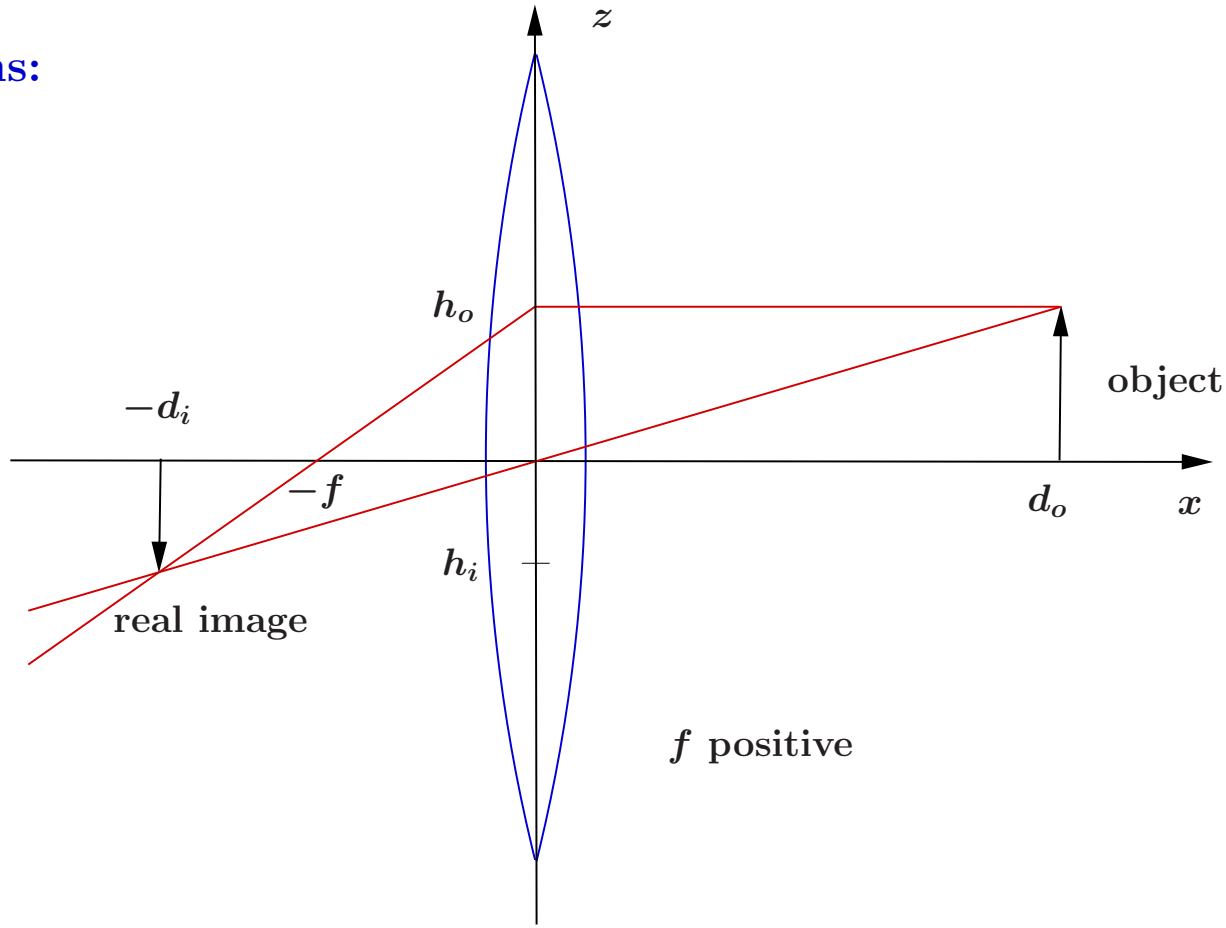
D) double-concave
E) plano-concave
F) convex-concave

C) and F) are also called meniscus-lenses

Like spherical mirrors, spherical lenses focus parallel rays only in first order approximation (paraxial rays, Gaussian optics). Imaging properties are simple under the following two assumptions.

- First order approximation: Parallel rays go through focus.
- Thin lens approximation: Rays through center of lens are not being deflected.

Converging lens:



d_o distance of object (positive by choice of axes)

h_o height of object (positive by choice of axes)

d_i distance of image (positive if on the other side of the lens than object)

h_i height of image (negative if inverted)

Goal: Determine d_i and h_i if f, d_o and h_o are given.

$$\frac{h_o}{d_o} = -\frac{h_i}{d_i} \qquad (*)$$

$$\frac{h_o}{f} = \frac{-h_i}{d_i - f} \qquad (**)$$

$$(*) \quad \text{and} \quad (**) \quad \Longrightarrow \quad \frac{d_o}{di} = \frac{f}{d_i - f}$$

$$d_o\, d_i \; - \; d_o\, f \; = \; d_i\, f \qquad | \; : \; d_i\, d_o\, f$$

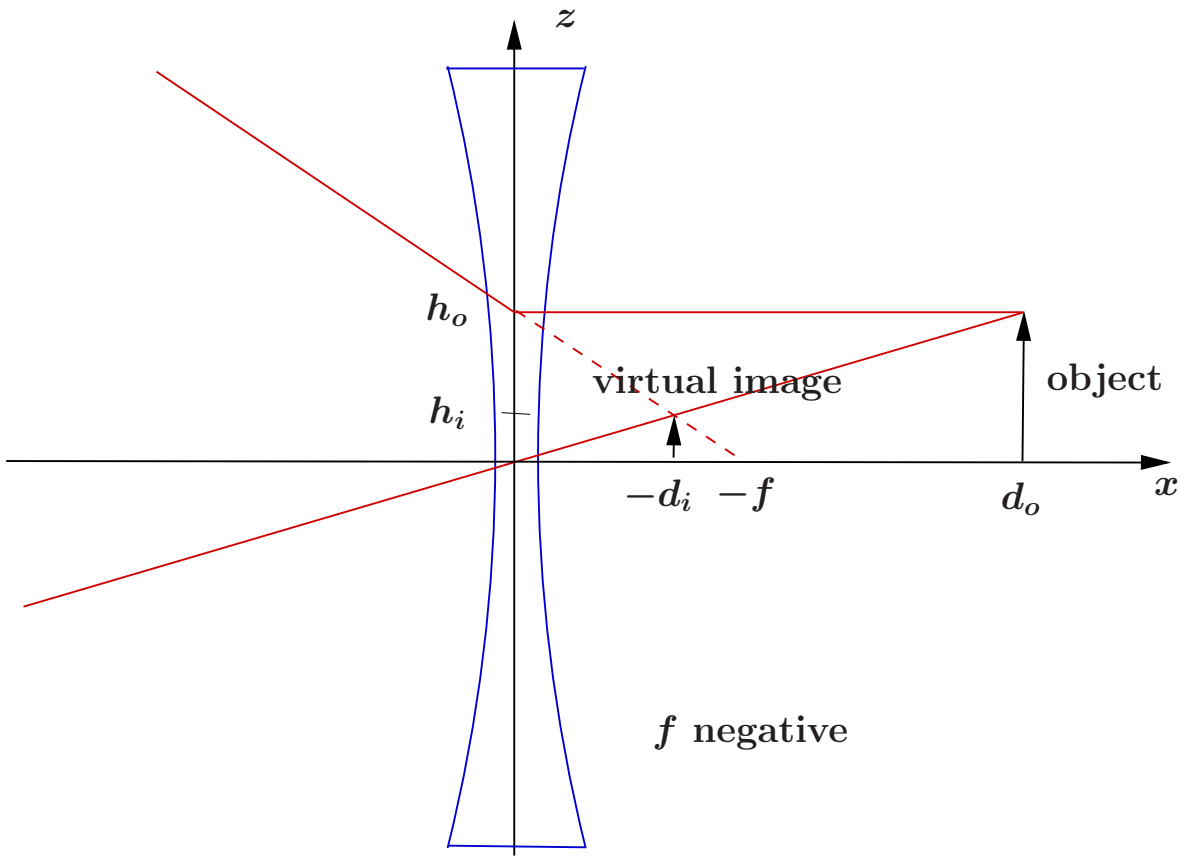
$$\boxed{\frac{1}{f} = \frac{1}{d_i} + \frac{1}{d_o}}$$

This is the (thin) lens equation. With d_i known, $(*)$ now gives the magnification

$$\boxed{m = \frac{h_i}{h_o} = -\frac{d_i}{d_o}}$$

For $d_o > f$, the image is real and inverted, thus $m < 0$.
For $d_o < f$, the image is virtual and upright, thus $m > 0$.

Diverging lens:



d_o distance of object (positive by choice of axes)

h_o height of object (positive by choice of axes)

d_i distance of image (negative if on the same side of the lens as object)

h_i height of image (positive if upright)

Calculation analogous to converging lens: $\boxed{\frac{1}{f} = \frac{1}{d_i} + \frac{1}{d_o}}$ $\boxed{m = \frac{h_i}{h_o} = -\frac{d_i}{d_o}}$

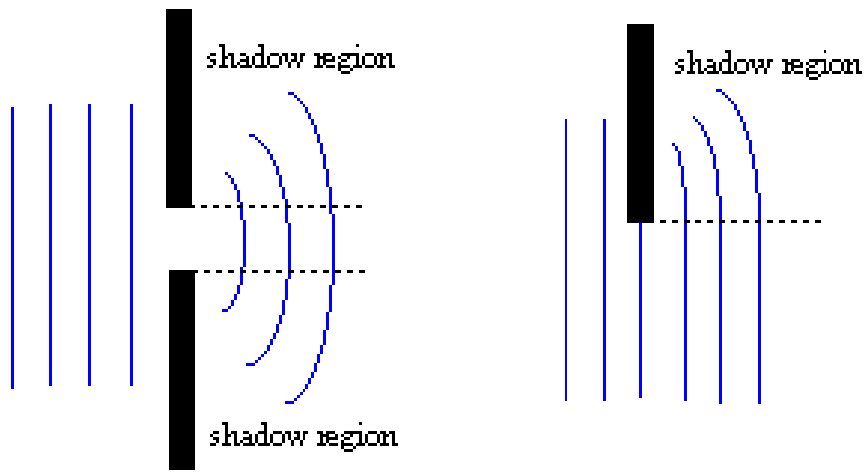
Image is always virtual and upright, thus $m > 0$.

II. Waves

(cf. Pedrotti and Pedrotti, Chapters 8 and 9)

There are two observational facts that demonstrate that light is actually a wave phenomenon. Ray optics is only an approximation.

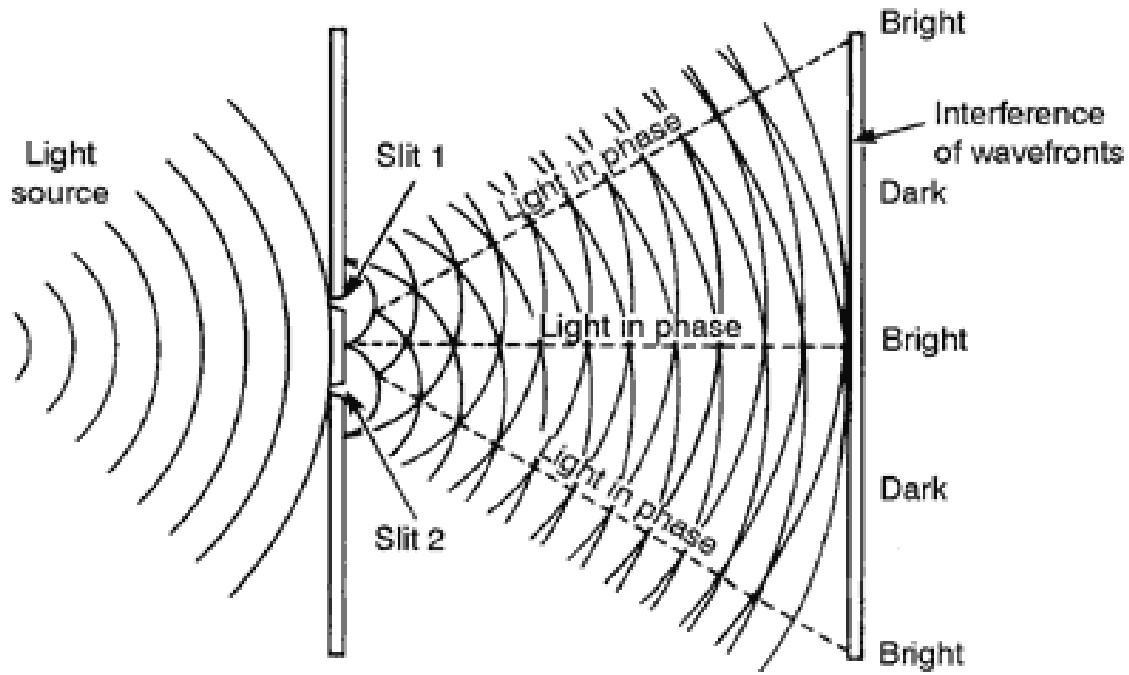
First, light can travel into the geometric shadow. This is called "diffraction".



picture from http://www.astrophys-assist.com

Diffraction of light was first observed by Francesco Grimaldi (1618–1663) whose observations were published in 1665.

Second, light coming from two sources shows "interference".



picture from http://www.astrophys-assist.com

This was demonstrated in the famous double-slit experiment by Thomas Young (1773-1829) in 1801.

In every-day life the wave character of light is not obvious because its wave length ($< 1\,\mu$m) is very small compared to the dimension of obstacles.

Later we discuss diffraction and interference in great detail. In this section we deal with general properties of waves.

a) General features of waves

Waves are special fields. A field is a function of space of time.

Distinguish:

- **one-dimensional scalar field**

$$u(x,t)$$

  e.g. elongation of a string, temperature along a (thin) rod, ...

- **three-dimensional scalar field**

$$u(\underline{r}\,,\,t\,)$$

  e.g. temperature in a room, mass density in a fluid or in a gas, ...

- **three-dimensional vector field**

$$\underline{v}(\underline{r}\,,\,t\,)$$

  e.g. velocity of a fluid, electric field, magnetic field, ...

There are other types of fields. E.g., the gravitational field in general relativity is a "tensor field".

A wave is a field that describes how a physically measurable quantity travels through space in the course of time.

Notes:

- **The time-dependence of a wave need not be periodic. (However, we can always do a series expansion, called Fourier analysis, in terms of sinusoidal waves.)**
- **A wave does not necessarily need a material medium to travel in. E.g., light waves can travel in vacuo. This was hard to grasp for 19th and early 20th century physicists who believed that light waves needed a material medium, called the ether, to travel in. A. Einstein postulated, within the framework of his Special Relativity Theory of 1905, that the ether does not exist.**

Distinguish the following properties of waves:

$$\text{linear} \longleftrightarrow \text{non-linear}$$

For linear waves the superposition principle holds, for non-linear waves it does not. Small-amplitude waves are typically linear.

$$\text{damped} \longleftrightarrow \text{undamped}$$

For a damped wave, part of the wave energy is converted into other forms of energy, typically heat, for an undamped wave the wave energy is preserved.

traveling $\longleftrightarrow$ standing

In a standing wave the energy does not travel if averaged over appropriate time intervals, in a traveling wave it does. (A linear standing wave can be written as a superposition of two waves that travel into opposite directions.)

free $\longleftrightarrow$ driven

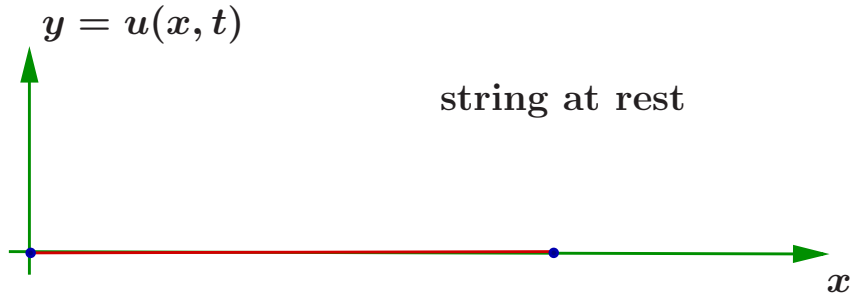
A driven wave has a permanent driving force as its source, in a free wave a perturbation travels without driving force; e.g.:

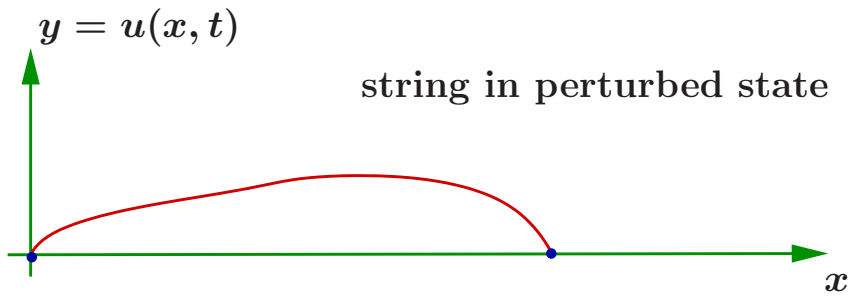
- stone thrown into water: free wave
  (after stone has sunk)
- rod rhythmically moved in water: driven wave

We will now discuss one-dimensional scalar waves in detail Then we turn to three-dimensional scalar and vector waves. The latter case includes light.

b) One-dimensional scalar waves

α) Derivation of the one-dimensional wave equation for a string



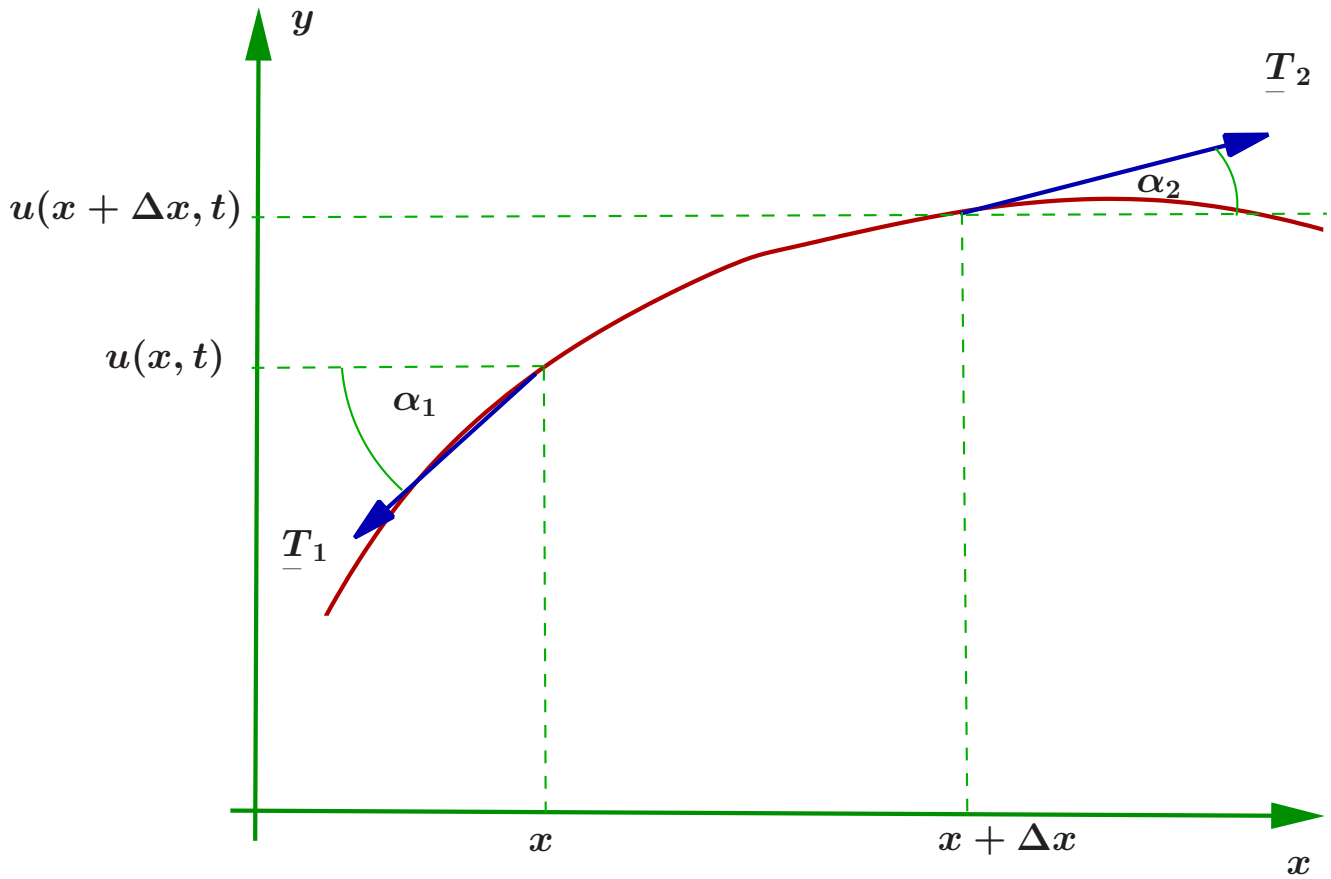


Assumptions:

- motion in $x - y-$plane
- μ = mass/length in rest state = constant
- motion purely transverse, i.e., each line element moves vertically (justified if elongation sufficiently small)
- no forces other than tension (i.e., ignore gravity, friction, ...)

Goal: Derive differential equation for $u(x,t)$.

Consider short line element between x and $x + \Delta x$



$$\mu \cdot \Delta x = \text{mass of line element}$$

$$\underline{T}_1 = \text{tension at } x$$

$$\underline{T}_2 = \text{tension at } x + \Delta x$$

Total force on line element: $\underline{F} = \underline{T}_1 + \underline{T}_2$

Newton's second law: $\underline{T}_1 + \underline{T}_2 = \mu \cdot \Delta x \cdot \underline{a}$

- $x-$component of Newton's second law (motion is transverse):

$$-|\underline{T}_1| \cos\alpha_1 \; + \; |\underline{T}_2| \cos\alpha_2 \; = \; 0$$

$$|\underline{T}_1| \cos\alpha_1 \; = \; |\underline{T}_2| \cos\alpha_2 \; =: \; T \qquad (*)$$

- $y-$component of Newton's second law:

$$-|\underline{T}_1| \sin\alpha_1 \; + \; |\underline{T}_2| \sin\alpha_2 \; = \; \mu \cdot \Delta x \cdot \frac{\partial^2 u(x + \frac{1}{2}\Delta x, t)}{\partial t^2} \qquad (**)$$

where $x + \frac{1}{2}\Delta x =$ center of inertia of line element.

From geometry:

$$\tan\alpha_1 \; = \; \frac{\partial u(x,t)}{\partial x} \qquad (***)$$

$$\tan\alpha_2 \; = \; \frac{\partial u(x + \Delta x, t)}{\partial x}$$

Insert $(***)$ into $(**)$:

$$-|\underline{T}_1|\cos\alpha_1\,\frac{\partial u(x,t)}{\partial x} + |\underline{T}_2|\cos\alpha_2\,\frac{\partial u(x+\Delta x,t)}{\partial x} = \mu\cdot\Delta x\cdot\frac{\partial^2 u(x+\frac{1}{2}\Delta x,t)}{\partial t^2}$$

Insert $(*)$ into the last equation:

$$T\,\frac{1}{\Delta x}\left(\frac{\partial u(x+\Delta x,t)}{\partial x} - \frac{\partial u(x,t)}{\partial x}\right) = \mu\cdot\frac{\partial^2 u(x+\frac{1}{2}\Delta x,t)}{\partial t^2}$$

Send $\Delta x \longrightarrow 0$:

$$T\,\frac{\partial^2 u(x,t)}{\partial x^2} = \mu\cdot\frac{\partial^2 u(x,t)}{\partial t^2}$$

With $v = \sqrt{T/\mu}$:

$$\boxed{\frac{\partial^2 u(x,t)}{\partial x^2} - \frac{1}{v^2}\,\frac{\partial^2 u(x,t)}{\partial t^2} = 0}$$

one-dimensional wave equation

- The differential operator

$$\frac{\partial^2}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}$$

is called the one-dimensional wave operator with velocity v.

- v has, indeed, dimension of velocity:

$$\frac{1}{\text{length}^2} = \frac{1}{\dim[v]^2} \cdot \frac{1}{\text{time}^2}$$

$$\dim[v] = \frac{\text{length}}{\text{time}}$$

- For elastic string, $v = \sqrt{T/\mu}$ increases with increasing tension T and decreases with increasing mass density μ.

- For all mechanical waves:

$$v^2 = \frac{\text{measure of tension}}{\text{measure of inertia}}$$

β) Solutions to the one-dimensional wave equation

Claim 1: Let $u(x,t) = \varphi(x - vt)$, where φ is *any* two times differentiable function. Then $u(x,t)$ satisfies the wave equation

$$\frac{\partial^2 u(x,t)}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 u(x,t)}{\partial t^2} = 0\,.$$

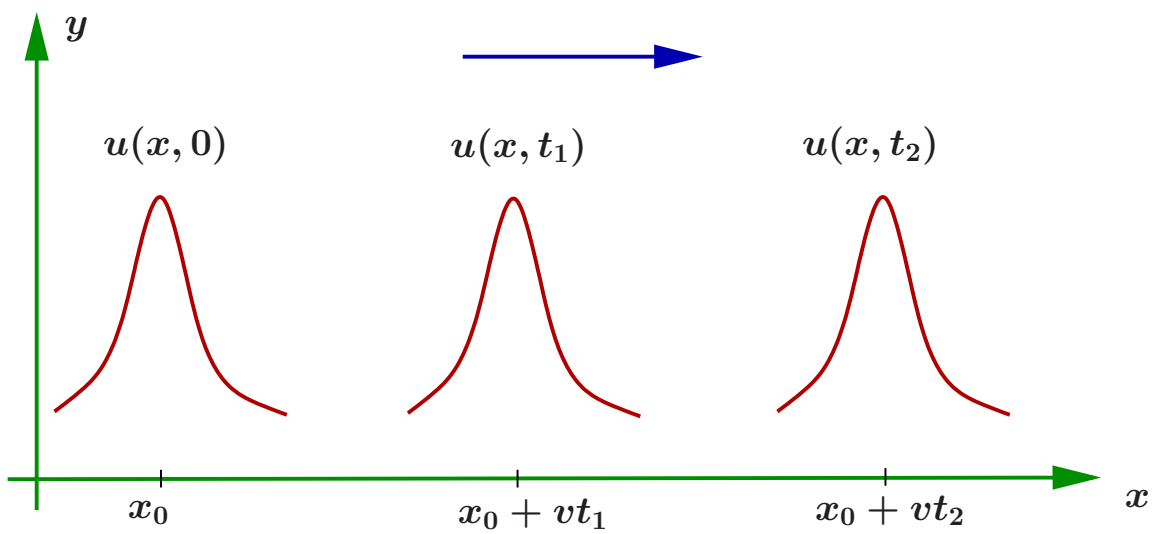
Proof:

$$\frac{\partial u(x,t)}{\partial x} = \varphi'(x - vt) \qquad \frac{\partial u(x,t)}{\partial t} = -v\,\varphi'(x - vt)$$

$$\frac{\partial^2 u(x,t)}{\partial x^2} = \varphi''(x - vt) \qquad \frac{\partial^2 u(x,t)}{\partial t^2} = v^2\,\varphi''(x - vt)$$

$\square$

Such solutions describe waves of arbitrary shape moving in positive $x-$direction:

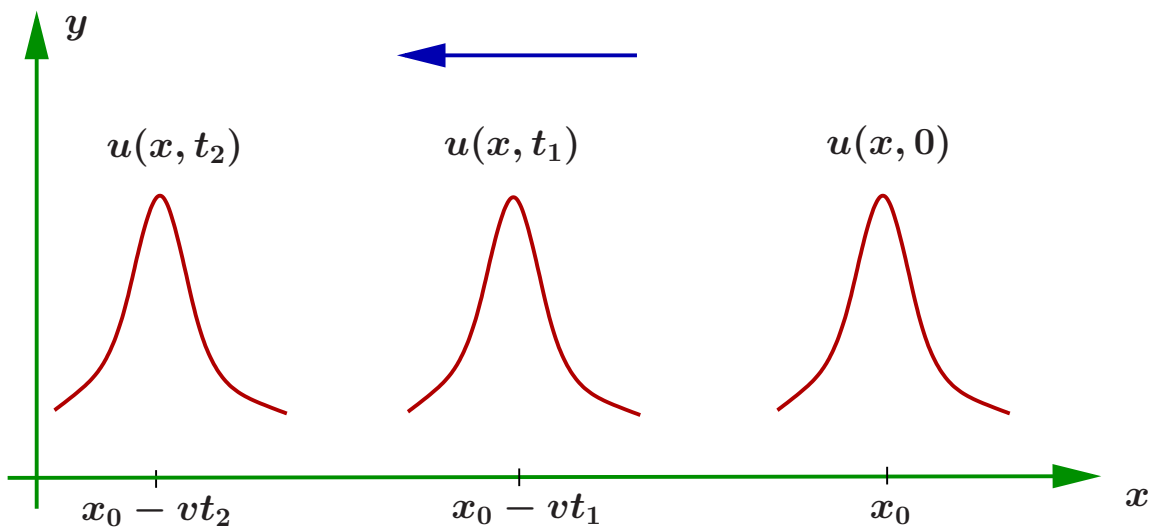


Claim 2: Let $u(x,t) = \psi(x+vt)$, where ψ is *any* two times differentiable function. Then $u(x,t)$ satisfies the wave equation

$$\frac{\partial^2 u(x,t)}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 u(x,t)}{\partial t^2} = 0\,.$$

The proof is analogous to that of Claim 1.

Such solutions describe waves of arbitrary shape moving in negative $x-$direction:



Superposition principle (wave equation is linear): Every function of the form $u(x,t) = \varphi(x-vt) + \psi(x+vt)$ solves the wave equation.

One can show that $u(x,t) = \varphi(x-vt) + \psi(x+vt)$ is actually the general solution of the wave equation. Indeed, φ and ψ can be adapted to arbitrary initial conditions:

Want to find a solution $u(x,t)$ of the wave equation such that

$$u(x,0) = f(x) \quad \text{and} \quad \frac{\partial u}{\partial t}(x,0) = h(x) \qquad (*)$$

where f and h are two given functions. Choose x_0 and define functions φ and ψ by

$$\varphi(x) = \frac{1}{2} f(x) - \frac{1}{2v} \int_{x_0}^{x} h(\xi) d\xi \,,$$

$$\psi(x) = \frac{1}{2} f(x) + \frac{1}{2v} \int_{x_0}^{x} h(\xi) d\xi \,.$$

Then

$$u(x,t) = \varphi(x-vt) + \psi(x+vt)$$

satisfies the wave equation and the initial conditions $(*)$. This represents the general solution to the wave equation in terms of its initial conditions:

$$\boxed{\begin{aligned} u(x,t) &= \frac{1}{2}\Big(u(x+vt,0) + u(x-vt,0) \Big) \\ &\quad + \frac{1}{2v} \int_{x-vt}^{x+vt} \frac{\partial u}{\partial t}(\xi,0)\, d\xi \end{aligned}}$$

d'Alembert's formula

Example: Determine the solution $u(x,t)$ to the one-dimensional wave equation that satisfies the initial conditions

$$u(x,0) = 0 \quad \text{and} \quad \frac{\partial u}{\partial t}(x,0) = C\cos(kx)$$

where C and k are real constants.

D'Alembert's formula:

$$u(x,t) \;=\; \frac{1}{2}\Big(\,u(x+vt,0) \;+\; u(x-vt,0)\,\Big) \;+\; \frac{1}{2v}\int_{x-vt}^{x+vt} \frac{\partial u}{\partial t}\,(\xi,0)\,d\xi$$

$$= \;0 \;+\; \frac{1}{2v}\int_{x-vt}^{x+vt} C\cos(k\xi)\,d\xi \;=\; \frac{C}{2v}\,\frac{1}{k}\sin(k\xi)\,\Big|_{x-vt}^{x+vt}$$

$$= \frac{C}{2vk}\Big(\,\sin\big(k(x+vt)\big) - \sin\big(k(x-vt)\big)\,\Big)$$

$$= \frac{C}{vk}\cos(kx)\sin(kvt)$$

Check that this $u(x,t)$ satisfies indeed the wave equation and the initial conditions!

γ) Harmonic waves in one space dimension

A one-dimensional harmonic wave is a function that depends on x and t in the form

$$u(x,t) \;=\; A\sin(kx - \omega t + \delta) \qquad (*)$$

A : amplitude
δ : phase shift
k : wave number
ω : angular frequency

Note that $(*)$ is the same as

$$u(x,t) \;=\; A\cos(kx - \omega t + \tilde{\delta})$$

where $\tilde{\delta} = \delta - \frac{\pi}{2}$.

Also, with the identity

$$\sin(\alpha + \beta) \;=\; \sin\alpha\cos\beta \;+\; \cos\alpha\sin\beta$$

$(*)$ can be equivalently rewritten as

$$u(x,t) \;=\; B\sin(kx - \omega t) \;+\; C\cos(kx - \omega t)$$

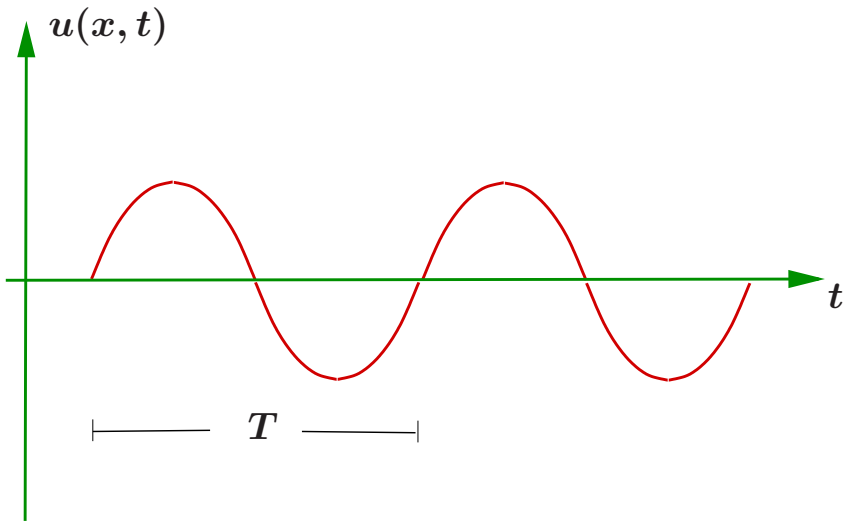
where $B \;=\; A\cos\delta$ and $C \;=\; A\sin\delta$.

At fixed x:

Oscillation with

period $T = \frac{2\pi}{\omega}$ [in seconds]

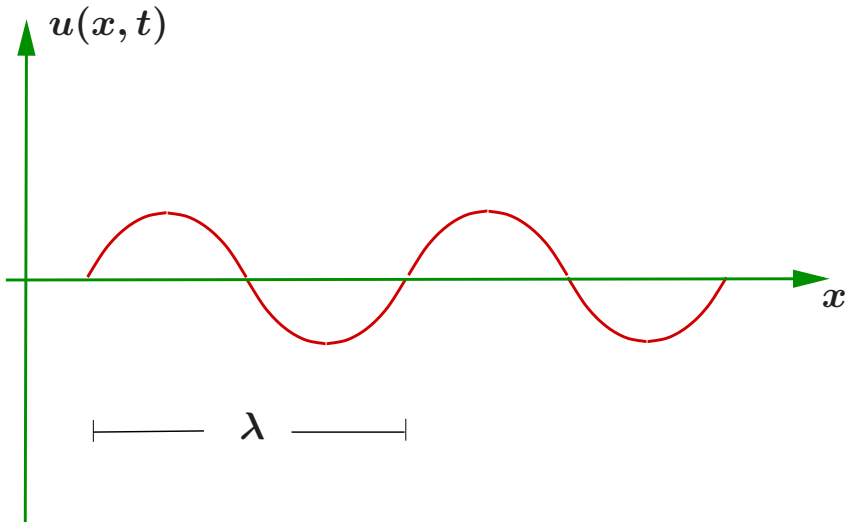
frequency $\nu = \frac{\omega}{2\pi}$ [in Hz = c.p.s.]



At fixed t:

Snapshot of a wave with

wave length $\lambda = \frac{2\pi}{k}$ [in meters]



The differential equation that determines the physical system (or the "medium") gives a relation between ω and k, called the dispersion relation, which is usually written in the form

$$\boxed{\omega = \omega(k)}$$

From the dispersion relation one derives the

$$\boxed{\text{phase velocity} \qquad v_p = \frac{\omega}{k}}$$

and the

$$\boxed{\text{group velocity} \qquad v_g = \frac{d\omega}{dk}}$$

Explanation of the names "phase velocity" and "group velocity":

Phase velocity:

If someone moves with the phase velocity $v_p = \omega/k$, his x-coordinate depends on time according to $x = x_0 + \omega t/k$, so for him the phase $kx - \omega t + \delta = kx_0 + \delta$ is a constant (independent of t). This explains the name "phase velocity".

Group velocity:

Consider two harmonic waves with the same amplitude and the same phase shift:

$$u_1(x,t) = A\cos(k_1 x - \omega_1 t + \delta) \qquad \text{and} \qquad u_2(x,t) = A\cos(k_2 x - \omega_2 t + \delta)\,.$$

The identity

$$\cos\alpha \;+\; \cos\beta \;=\; 2\cos\frac{\alpha-\beta}{2}\cos\frac{\alpha+\beta}{2}$$

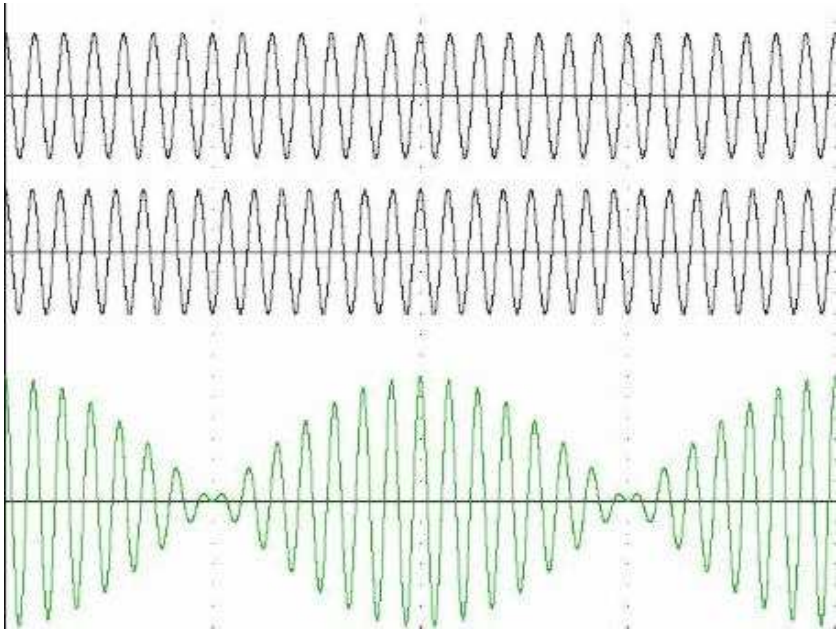
yields

$$u(x,t) = u_1(x,t) + u_2(x,t) =$$

$$2\,A\cos\Big(\frac{k_1-k_2}{2}\,x - \frac{\omega_1-\omega_2}{2}\,t\Big)\cos\Big(\frac{k_1+k_2}{2}\,x - \frac{\omega_1+\omega_2}{2}\,t + \delta\Big)$$

This is a cosine function whose amplitude is modulated by another cosine function, with modulation wave number and modulation frequency

$$k_{\text{mod}} = \frac{k_1-k_2}{2}\,, \qquad \omega_{\text{mod}} = \frac{\omega_1-\omega_2}{2}$$



picture from http://www.tinpan.fortunecity.com

The maximum of the modulation function moves at velocity

$$v_{\text{mod}} = \frac{\omega_{\text{mod}}}{k_{\text{mod}}} = \frac{\omega_1 - \omega_2}{k_1 - k_2}\,.$$

If $k_1 - k_2$ is small,

$$v_{\text{mod}} \approx d\omega/dk = v_g\,.$$

This explains the name "group velocity". (Here we considered a "group" consisting of only two waves. The argument can be generalised to larger groups of waves.)

Example: Derive the dispersion relation for the wave equation

$$\frac{\partial^2 u(x,t)}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 u(x,t)}{\partial t^2} = 0\,.$$

Ansatz $u(x,t) = A\sin(kx - \omega t + \delta)$ yields

$$\frac{\partial^2 u(x,t)}{dx^2} = -A\,k^2 \sin(kx - \omega t + \delta)$$

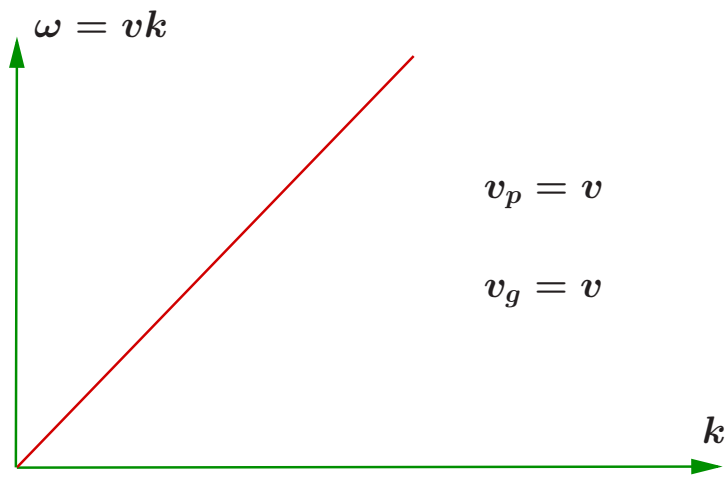
$$\frac{\partial^2 u(x,t)}{dt^2} = -A\,\omega^2 \sin(kx - \omega t + \delta)\,.$$

Wave equation requires

$$\left(-k^2 + \frac{\omega^2}{v^2}\right) A\sin(kx - \omega t + \delta) = 0\,.$$

So the dispersion relation for the wave equation reads

$$\boxed{\omega = v\,k}$$



Quite generally, if v_p is independent of k, the medium is said to be "non-dispersive".

δ) Complex notation and phasors

Some quantities in physics are complex, e.g. the wave function in quantum mechanics. However, even when dealing with real quantities complex notation is often useful.

Basic idea: Write the real quantity you are interested in as the real part of an appropriately defined complex quantity; use the more convenient calculational rules for complex numbers.

Recap complex numbers:

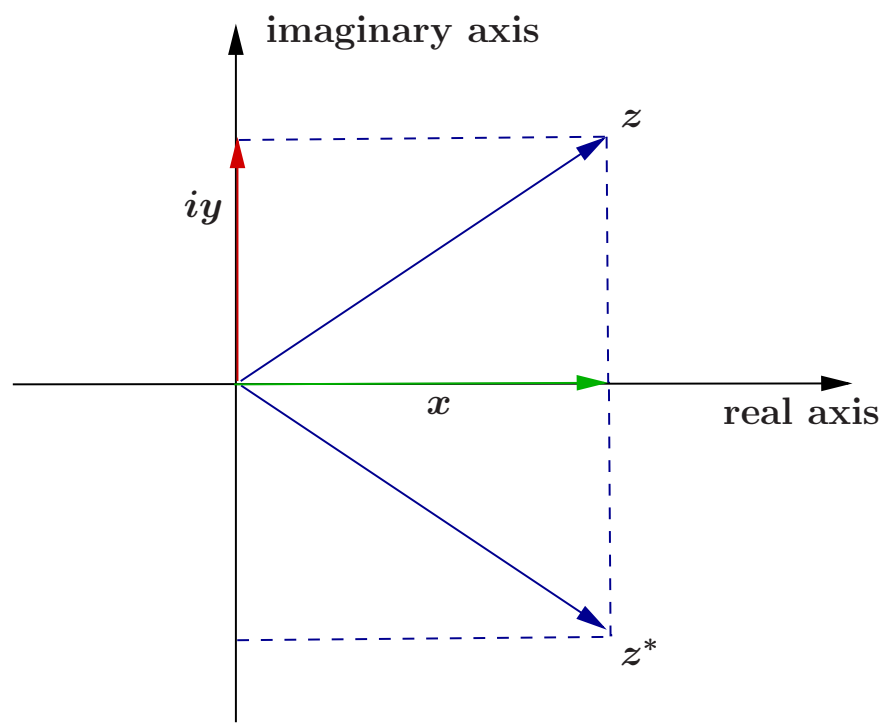
$$z = x + iy \quad \text{where} \quad i^2 = -1$$

$$x = \mathrm{Re}(z)\,, \qquad y = \mathrm{Im}(z)$$

$$\text{Complex conjugate}: \; z^* = x - iy$$

$$\text{Modulus}: \; |z| = \sqrt{x^2 + y^2}$$

$$|z|^2 = z^*\, z = z\, z^*$$

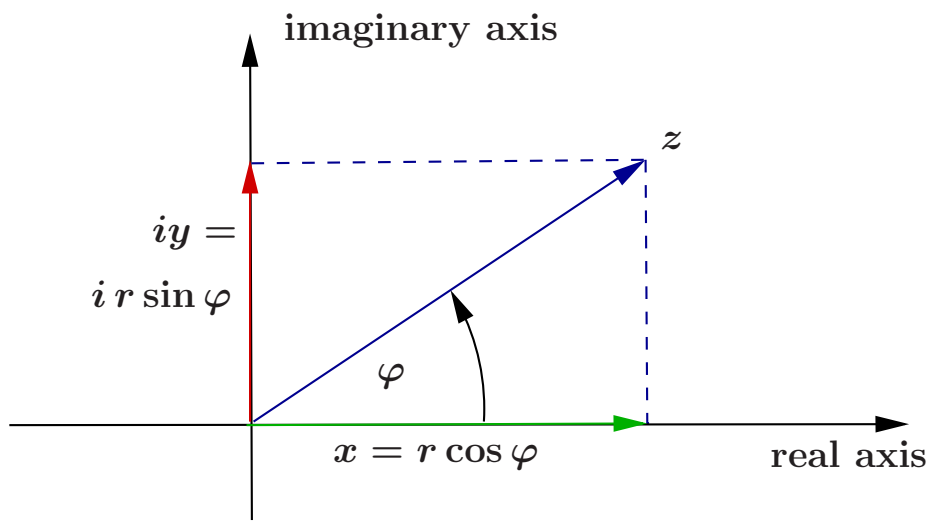


$\{\text{complex numbers}\} = \{\text{pairs of real numbers with multiplication such that } i^2 = -1\}$.

Polar cordinates for complex numbers:

$$x = r\cos\varphi\,, \qquad y = r\sin\varphi$$

$$r = \sqrt{x^2+y^2}\,, \qquad \tan\varphi = \frac{y}{x}$$



$$z = x + i\,y = r\left(\cos\varphi + i\sin\varphi\right)$$

Euler's formula: $\boxed{\cos\varphi + i\sin\varphi = e^{i\varphi}}$

Proof of Euler's formula: Compare Taylor series

$$e^x = \sum_{n=0}^{\infty}\frac{x^n}{n!}\,, \qquad \sin x = \sum_{m=0}^{\infty}(-1)^m\frac{x^{2m+1}}{(2m+1)!}\,, \qquad \cos x = \sum_{m=0}^{\infty}(-1)^m\frac{x^{2m}}{(2m)!}\,.$$

$\square$

Representation of complex numbers in polar coordinates with Euler's formula:

$\boxed{z = r\,e^{i\varphi}}$ $\quad r$ and φ real, $r = |z| \geq 0$.

Harmonic wave at fixed position in space is given by a function

$$u(t) \;=\; A\cos\big(\omega t + \beta\big)$$

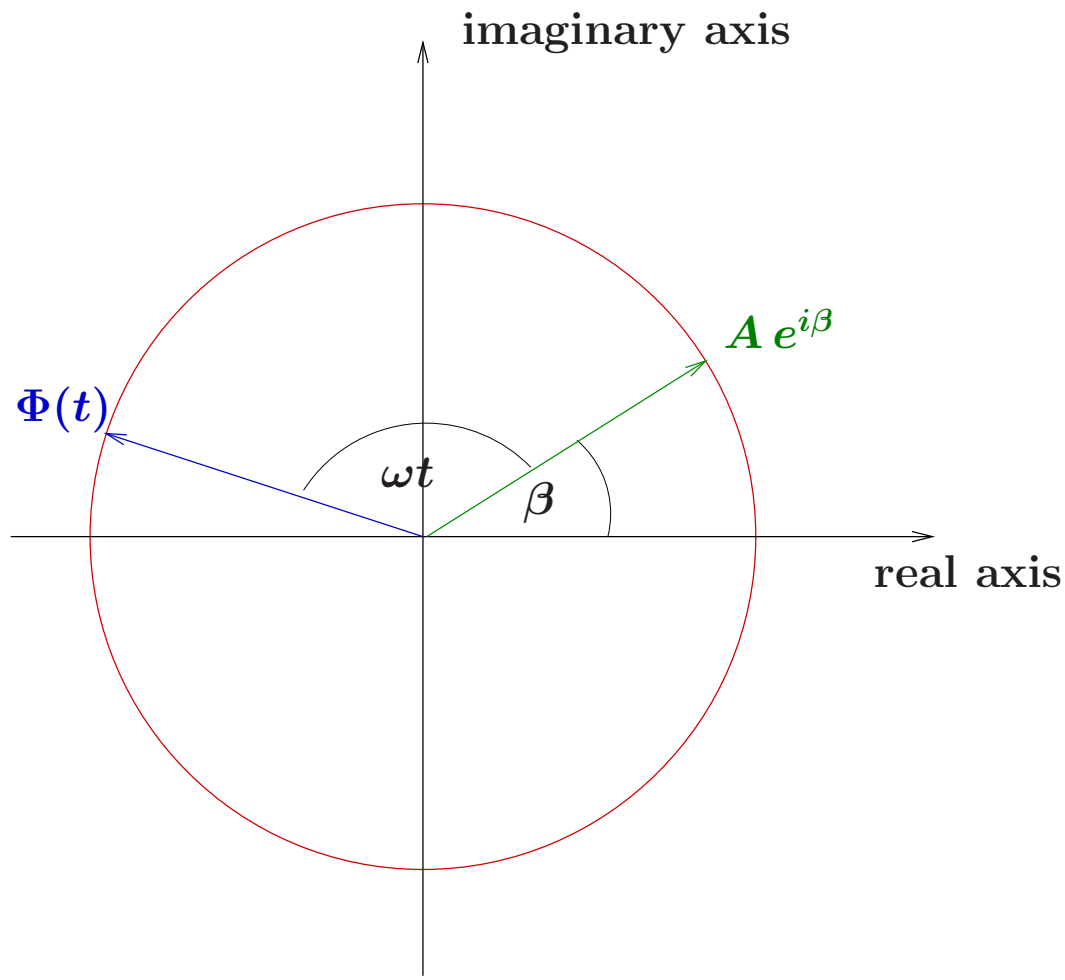
where ω, A and β are real constants.

$u(t)$ can be written as the real part of an appropriately defined complex function $\Phi(t)$:

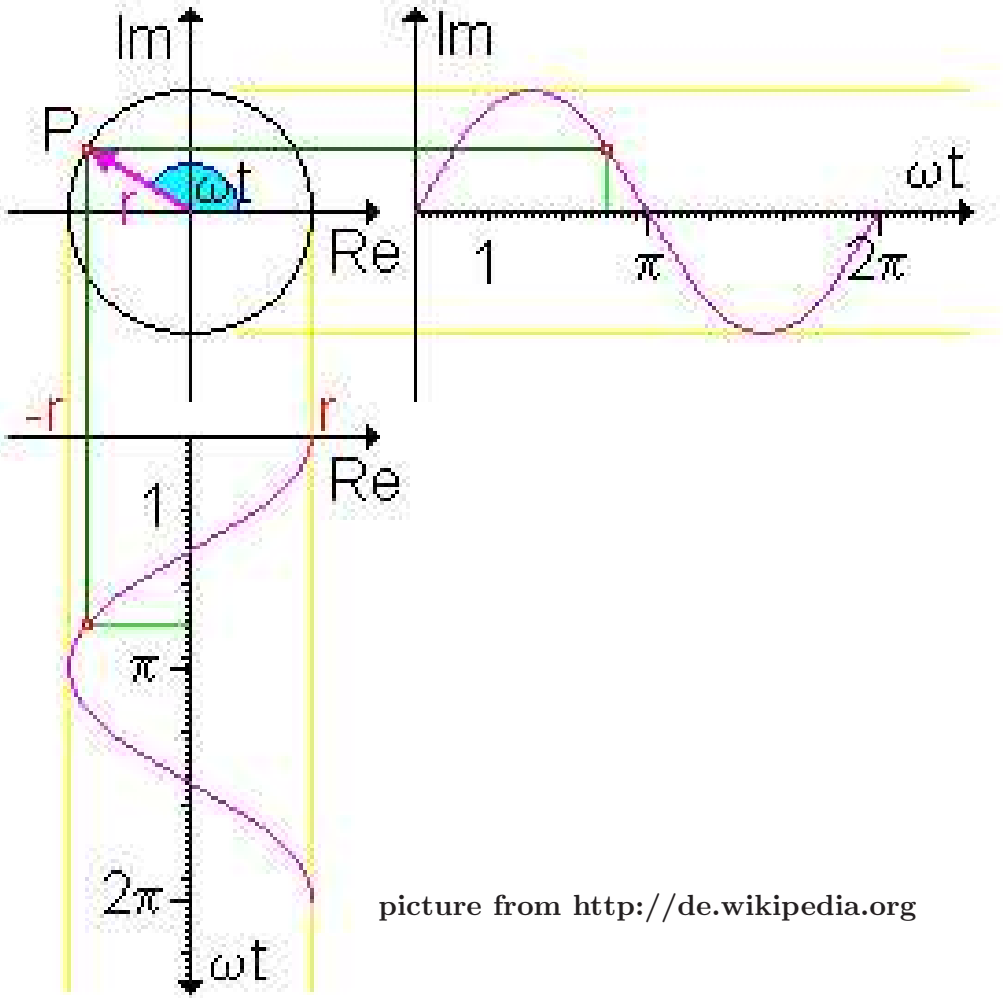
$$u(t) = \mathrm{Re}\big(\Phi(t)\big)$$

$$\Phi(t) = A\, e^{i(\omega t+\beta)}$$

$$= Ae^{i\beta}e^{i\omega t}$$



Time-dependence $\widehat{=}$ circular motion in complex plane : "Phasor diagram"



picture from http://de.wikipedia.org

Example 1: Two waves are given, at a fixed point x, as

$$u_1(t) \;=\; \sqrt{3}\,A\cos(\omega t) \quad \text{and} \quad u_2(t) \;=\; A\sin(\omega t)$$

where A and ω are positive real constants. Write $u(t) \;=\; u_1(t) \;+\; u_2(t)$ in the form $u(t) \;=\; r\cos(\omega t + \varphi)$. Show u_1, u_2 and u in a phasor diagram.

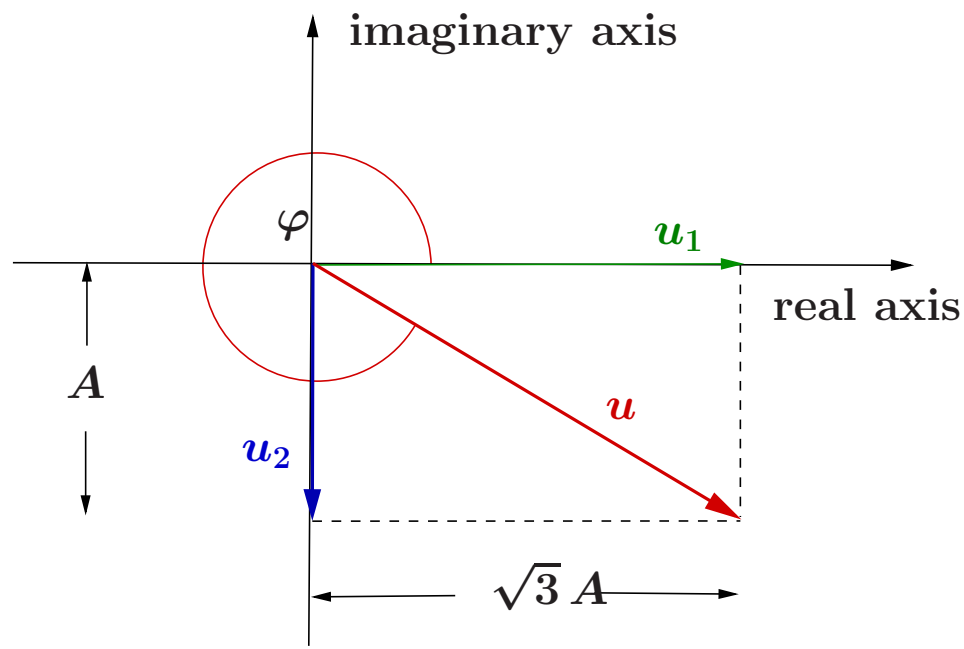
$$u_1(t) \;=\; \mathrm{Re}\{\,\sqrt{3}\,A\,e^{i\omega t}\,\} \quad \text{and} \quad u_2(t) \;=\; \mathrm{Re}\{\,A\,e^{i(\omega t-\pi/2)}\,\} \;=\; \mathrm{Re}\{\,-\,i\,A\,e^{i\omega t}\,\}$$

$$u(t) \;=\; \mathrm{Re}\{\,(\sqrt{3}-i)\,A\,e^{i\omega t}\,\}$$

Set $(\sqrt{3}-i)A = re^{i\varphi}$. Then $r = \sqrt{3+1}\,A = 2A$ and $\tan\varphi = -1/\sqrt{3}$, so $\varphi = -\pi/3(+2\pi)$.

$$u(t) \;=\; \mathrm{Re}\{\,2\,A\,e^{-i\pi/3}\,e^{i\omega t}\,\} \;=\; \mathrm{Re}\{\,2\,A\,e^{i(\omega t-\pi/3)}\,\} \;=\; 2\,A\cos\big(\omega t - \pi/3\big)\,.$$

Phasor diagram for $t = 0$:



Example 2: Rederive the dispersion relation for the wave equation

$$\frac{\partial^2 u(x,t)}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 u(x,t)}{\partial t^2} = 0$$

with the ansatz $u(x,t) = W e^{i(kx-\omega t)}$ where W is complex and k and ω are real.

$$\frac{\partial u(x,t)}{\partial x} = W\, i\, k\, e^{i(kx-\omega t)}\,, \qquad \frac{\partial^2 u(x,t)}{\partial x^2} = -W\, k^2\, e^{i(kx-\omega t)}$$

$$\frac{\partial u(x,t)}{\partial t} = -W\, i\, \omega\, e^{i(kx-\omega t)}\,, \qquad \frac{\partial^2 u(x,t)}{\partial t^2} = -W\, \omega^2\, e^{i(kx-\omega t)}$$

So the wave eqation requires

$$\left(-k^2 + \frac{\omega^2}{v^2} \right) W\, e^{i(kx-\omega t)} = 0\,, \quad \text{hence} \quad \omega = v\, k\,.$$

[Note: If we write $W = Ae^{i\delta}$, we have

$$\mathrm{Re}\{u(x,t)\} = A\cos(kx - \omega t + \delta) \quad \text{and} \quad \mathrm{Im}\{u(x,t)\} = A\sin(kx - \omega t + \delta)\,.$$

The complex function $u(x,t)$ satisfies the wave equation if and only if both its real and its imaginary part satisfies the wave equation. This relates the use of the exponential function to our earlier use of sine or cosine functions.]

ϵ) Fourier analysis

Why are harmonic waves important?

Fourier Theorem: Every continuous function $F(t)$ that is periodic, $F(t+T) = F(t)$, can be written as a Fourier series

$$F(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\frac{2n\pi t}{T} + \sum_{n=1}^{\infty} b_n \sin\frac{2n\pi t}{T},$$

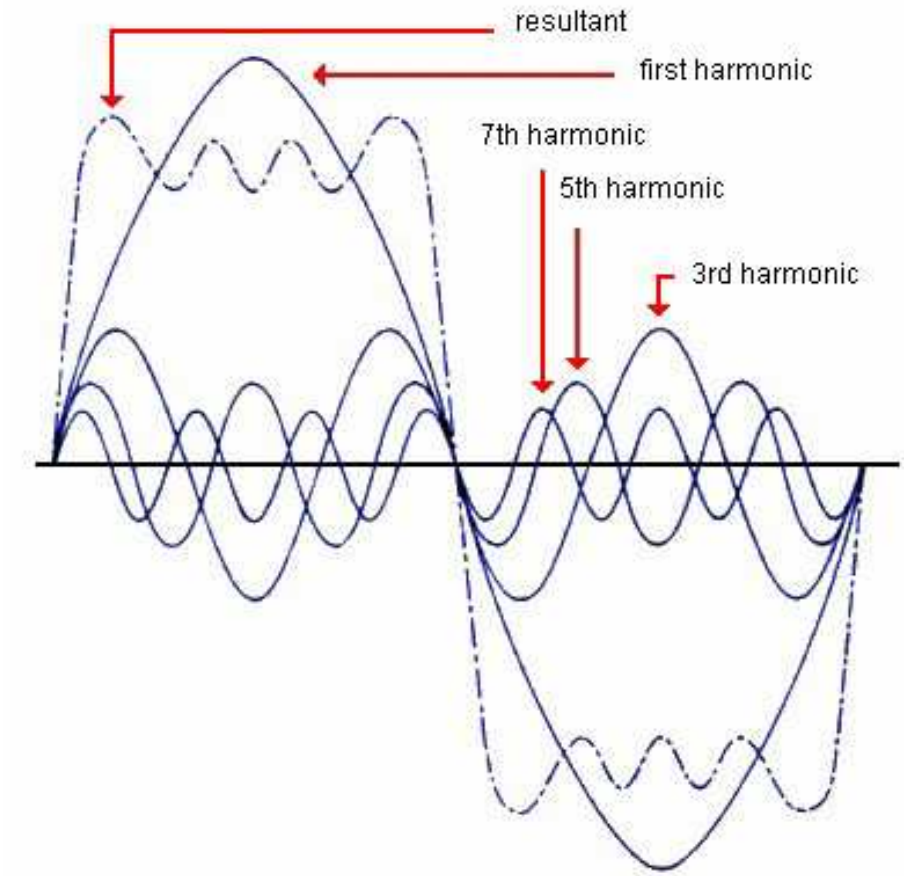
where

$$a_n = \frac{2}{T}\int_0^T F(t)\cos\frac{2n\pi t}{T}\,dt\,, \qquad b_n = \frac{2}{T}\int_0^T F(t)\sin\frac{2n\pi t}{T}\,dt\,.$$

- Determining the coefficients a_n and b_n for a given function $F(t)$ is called "Fourier analysis" or "harmonic analysis". For practical purposes, the infinite series can often be approximated by a finite sum.
- Fourier analysis can be done with respect to the time coordinate t and with respect to the spatial coordinate x.
- The Fourier theorem even holds for some discontinuous functions. (F must satisfy the "Dirichlet condition": There are only finitely many discontinuities in a finite interval and at each discontinuity the limit from above and from below exists.)
- If a function is defined on a finite interval $[0, T]$, one can extend it periodically to the whole real line and apply the Fourier theorem to the resulting periodic function.

Example: Fourier analysis of a rectangle function.

The picture shows the first four sine functions that contribute to a rectangle function. (Cosine functions do not contribute because the rectangle function is odd.)



picture from http://www.nemesi.net/mp3

Switch to complex notation:

Introduce c_n for $n \in \mathbb{Z}$ by

$$c_0 = \frac{a_0}{2}$$

$$c_n = \frac{1}{2}\left(a_n - i\, b_n\right) \quad \text{if } n > 0\,,$$

$$c_n = \frac{1}{2}\left(a_{-n} + i\, b_{-n}\right) \quad \text{if } n < 0\,.$$

Then

$$F(t) = \sum_{n=-\infty}^{\infty} c_n\, e^{i\omega_n t}\,, \quad \omega_n = \frac{2\pi n}{T}$$

$$c_n = \frac{1}{T}\int_0^T F(t)\, e^{-i\omega_n t}\, dt$$

Works for real-valued and for complex-valued periodic functions $F(t)$.

What if $F(t)$ is not periodic?

Replace Fourier series by Fourier integral:

$$F(t) = \int_{-\infty}^{\infty} G(\omega)\, e^{i\omega t}\, d\omega\,,$$

where

$$G(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(t)\, e^{-i\omega t}\, dt\,,$$

is the Fourier transform of $F(t)$.

$G(\omega)$ tells with which weight each frequency ω occurs in the Fourier analysis of $F(t)$.

Keep in mind:

Periodic functions have discrete spectrum c_n, $\omega_n = 2\pi n/T$, non-periodic functions have continuous spectrum $G(\omega)$.

c) Three-dimensional scalar waves

α) The three-dimensional wave equation

From the one-dimensional wave equation

$$\frac{\partial^2 u(x,t)}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 u(x,t)}{\partial t^2} = 0$$

for a function $u(x,t)$ one generalises to the three-dimensional wave equation

$$\boxed{\frac{\partial^2 u(\underline{r},t)}{\partial x^2} + \frac{\partial^2 u(\underline{r},t)}{\partial y^2} + \frac{\partial^2 u(\underline{r},t)}{\partial z^2} - \frac{1}{v^2}\frac{\partial^2 u(\underline{r},t)}{\partial t^2} = 0}$$

for a function $u(\underline{r},t) = u(x,y,z,t)$.

Short-hand notation: With the Laplace operator

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

the wave equation reads

$$\boxed{\Delta u - \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} = 0\,.}$$

Even shorter: With the wave operator (or d'Alembert operator)

$$\Box = \Delta - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}, \qquad \left(\Box = \Box_v\right)$$

the wave equation reads

$$\boxed{\Box u = 0\,.}$$

With the help of the vectorial operator del (or nabla)

$$\nabla(\,\cdot\,) = \frac{\partial(\,\cdot\,)}{\partial x}\,\underline{\hat{x}} + \frac{\partial(\,\cdot\,)}{\partial y}\,\underline{\hat{y}} + \frac{\partial(\,\cdot\,)}{\partial z}\,\underline{\hat{z}}$$

where $\underline{\hat{x}}$, $\underline{\hat{y}}$, $\underline{\hat{z}}$ are the unit vectors in x, y, z directions:

$$\Delta = \nabla^2 = \nabla\cdot\nabla\,,$$

$$\Box = \nabla^2 - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\,.$$

β) Sound waves in air

What we notice as sound is a pressure wave in air. Approximately, the pressure $p(\underline{r}, t)$ satisfies the wave equation

$$\Delta p - \frac{1}{v^2}\frac{\partial^2 p}{\partial t^2} = 0$$

where

$$v = \sqrt{\gamma p_0/\rho_0}\,,$$

$$\gamma = \text{adiabatic index}\,,$$
$$p_0 = \text{pressure at rest}\,,$$
$$\rho_0 = \text{mass density at rest}\,.$$

For air at 0^oC :

$$\gamma \approx 1.4\,, \quad p_0 \approx 10^6\,\text{dyne/cm}^2\,, \quad \rho_0 \approx 1.3 \cdot 10^{-3}\,\text{g/cm}^3\,,$$

$$v \approx 330\text{m/s}$$

v increases with temperature ($v \sim \sqrt{T}$).

The frequency of sound waves is perceived as pitch.

A440: Standard for musical pitch (440 Hz).

The same pitch played by different instruments sounds different because of overtones (higher terms in Fourier series).

Musical intervals are frequency ratios. E.g., an octave is the frequency ratio 1:2.

Audio range: 20 Hz to 20 kHz (varies significantly with age).

Maximum of sensitivity: 1 kHz to 3.5 kHz.

Some animals (elephants, alligators, ...) can produce and hear frequencies lower than 20 Hz (infrasound).

Some animals (bats, mice, ...) can produce and hear frequencies higher than 20 kHz (ultrasound).

The intensity of a sound wave is measured as the time-average

$$\overline{p(\underline{r}\,,t)^2} = \frac{1}{(t_2 - t_1)} \int_{t_1}^{t_2} p(\underline{r}\,,t)^2 dt\,.$$

Human perception of the intensity is measured in decibel (dB):

$$L(\underline{r}) = 10 \log_{10} \left(\frac{\overline{p(\underline{r}\,,t)^2}}{{p_0}^2} \right) \text{ dB}\,.$$

Note: For all human senses, the relation between perception and stimulus is logarithmic (Weber-Fechner law).

immediate soft tissue damage	185 dB
threshold of pain	134 dB
jet engine, 100 m distant	110 ... 140 dB
moving passenger car, 10 m distant	60 ... 80 dB
normal talking, 1 m distant	40 ... 60 dB
quiet rustling leaves	10 dB

γ) Harmonic plane waves in three space dimensions

From one-dimensional harmonic waves

$$u(x,t) \;=\; A\sin(kx - \omega t + \delta)$$

we generalise to three-dimensional harmonic plane waves

$$u(\underline{r}\,,t) \;=\; A\sin(\underline{k}\cdot\underline{r} - \omega t + \delta)\,.$$

A : amplitude

δ : phase shift

$\underline{k}$: wave vector

ω : angular frequency

Angular frequency determines period: $T = 2\pi/\omega$.

Modulus of wave vector determines wave length: $\lambda = 2\pi/|\underline{k}|$.

Direction of wave vector determines the surfaces of constant phase:

The surfaces $\underline{k}\cdot\underline{r} - \omega t + \delta \;=\; \text{constant}$ are planes perpendicular to $\underline{k}$.

The intensity of a plane harmonic wave

$$u(\underline{r}, t) = A \sin(\underline{k} \cdot \underline{r} - \omega t + \delta) .$$

is the time-average of $u(\underline{r}, t)^2$:

$$I(\underline{r}) = \overline{u(\underline{r}, t)^2} = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} u(\underline{r}, t)^2 \, dt =$$

$$= A^2 \frac{\omega}{2\pi} \int_0^{2\pi/\omega} \sin^2(\underline{k} \cdot \underline{r} - \omega t + \delta) \, dt .$$

Substituting $\xi = \omega t$ and using $\frac{1}{2\pi} \int_0^{2\pi} \sin^2(\xi + \alpha) d\xi = \frac{1}{2}$,

$$I(\underline{r}) = \overline{u(\underline{r}, t)^2} = \frac{A^2}{2} .$$

The dispersion relation for plane harmonic waves is of the form $\boxed{\omega = \omega(\underline{k})}$

In the special case that ω depends only on $k = |\underline{k}|$, the medium is called **isotropic**.

Phase velocity vector:

$$\underline{v}_p(\underline{k}) = \frac{\omega(\underline{k})}{|\underline{k}|^2}\,\underline{k}$$

has modulus

$$v_p(\underline{k}) = \frac{\omega(\underline{k})}{|\underline{k}|}\,.$$

Group velocity vector:

$$\underline{v}_g(\underline{k}) = \nabla_{\underline{k}}\,\omega(\underline{k}) = \frac{\partial\omega(\underline{k})}{\partial k_x}\,\underline{\hat{x}} + \frac{\partial\omega(\underline{k})}{\partial k_y}\,\underline{\hat{y}} + \frac{\partial\omega(\underline{k})}{\partial k_z}\,\underline{\hat{z}}$$

has modulus

$$v_g(\underline{k}) = \sqrt{\left(\frac{\partial\omega(\underline{k})}{\partial k_x}\right)^2 + \left(\frac{\partial\omega(\underline{k})}{\partial k_y}\right)^2 + \left(\frac{\partial\omega(\underline{k})}{\partial k_z}\right)^2}\,.$$

For an isotropic medium, $\omega = \omega(k)$, we recover the familiar formulas

$$v_p(k) = \frac{\omega(k)}{k}\,, \qquad v_g(k) = \frac{d\omega(k)}{dk}\,.$$

Derive the dispersion relation for the wave equation $\Delta u - \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} = 0$.

$$u(\underline{r}, t) = A \sin(\underline{k} \cdot \underline{r} - \omega t + \delta) ,$$

$$\nabla u(\underline{r}, t) = A\, \underline{k} \cos(\underline{k} \cdot \underline{r} - \omega t + \delta) ,$$

$$\nabla \cdot \nabla u(\underline{r}, t) = - A\, |\underline{k}|^2 \sin(\underline{k} \cdot \underline{r} - \omega t + \delta) ,$$

$$\frac{\partial^2}{\partial t^2} u(\underline{r}, t) = - A\, \omega^2 \sin(\underline{k} \cdot \underline{r} - \omega t + \delta) .$$

As $\nabla \cdot \nabla = \Delta$, the wave equation requires

$$\left(k^2 - \omega^2/v^2\right) A \sin(\underline{k} \cdot \underline{r} - \omega t + \delta) = 0 ,$$

where $k = |\underline{k}|$. This yields the dispersion relation for the wave equation

$$\boxed{\omega = v\, k .}$$

$$\Longrightarrow v_p = v_g = v .$$

As in the one-dimensional case, we say that the medium is non-dispersive if v_p is independent of k.

Keep in mind:

isotropic medium:

v_p is independent of the direction of $\underline{k}$, but may depend on k.

isotropic and non-dispersive medium:

v_p is a constant.

For the wave equation

$$v_p = v = \text{constant} \; ;$$

so the wave equation describes isotropic and non-dispersive media.

δ) Spherical harmonic waves

A spherical harmonic wave is a wave of the form

$$u(\underline{r}, t) = A(r) \sin(kr - \omega t + \delta)$$

where $r = |\underline{r}| = \sqrt{x^2 + y^2 + z^2}$.

The surfaces of constant phase

$$kr - \omega t + \delta = \text{constant}$$

are spheres.

Goal: Determine $A(r)$ such that the wave equation

$$\Delta u - \frac{1}{v^2} \frac{\partial^2 u}{\partial t^2} = 0$$

is satisfied.

We first calculate

$$\nabla r = \frac{\partial\sqrt{x^2+y^2+z^2}}{\partial x}\,\underline{\hat{x}} + \frac{\partial\sqrt{x^2+y^2+z^2}}{\partial y}\,\underline{\hat{y}} + \frac{\partial\sqrt{x^2+y^2+z^2}}{\partial z}\,\underline{\hat{z}} =$$

$$\frac{\cancel{2}x\,\underline{\hat{x}} + \cancel{2}y\,\underline{\hat{y}} + \cancel{2}z\,\underline{\hat{z}}}{\cancel{2}\sqrt{x^2+y^2+z^2}} = \frac{\underline{r}}{r}$$

and

$$\Delta r = \nabla\cdot\nabla r = \nabla\cdot\frac{\underline{r}}{r} = \frac{1}{r}\nabla\cdot\underline{r} - \frac{\underline{r}}{r^2}\cdot\nabla r =$$

$$\frac{1}{r}\left(\frac{\partial x}{\partial x} + \frac{\partial y}{\partial y} + \frac{\partial z}{\partial z}\right) - \frac{\underline{r}}{r^2}\cdot\frac{\underline{r}}{r} = \frac{3}{r} - \frac{1}{r} = \frac{2}{r}\,.$$

From $u(\underline{r}, t) = A(r) \sin(kr - \omega t + \delta)$ we find

$$\nabla u(\underline{r}, t) = \Big(A'(r) \sin(kr - \omega t + \delta) + A(r)\, k \cos(kr - \omega t + \delta) \Big) \nabla r \,.$$

Hence, with our results $\nabla r = \frac{\underline{r}}{r}$ and $\Delta r = \frac{2}{r}$ we get

$$\Delta u(\underline{r}, t) = \nabla \cdot \nabla u(\underline{r}, t) =$$

$$\Big(A'(r) \sin(kr - \omega t + \delta) + A(r)\, k \cos(kr - \omega t + \delta) \Big) \nabla \cdot \nabla r +$$

$$\Big(\{ A''(r) - A(r)\, k^2 \} \sin(kr - \omega t + \delta) + 2\, A'(r)\, k \cos(kr - \omega t + \delta) \Big) |\nabla r|^2 =$$

$$\Big(\frac{2\, A'(r)}{r} + A''(r) - A(r)\, k^2 \Big) \sin(kr - \omega t + \delta)$$

$$+ \Big(\frac{2\, k\, A(r)}{r} + 2\, A'(r)\, k \Big) \cos(kr - \omega t + \delta) \,.$$

On the other hand,

$$\frac{\partial^2 u}{\partial t^2} = - A(r)\, \omega^2 \sin(kr - \omega t + \delta) \,.$$

Thus, the wave equation requires

$$\left(\frac{2A'(r)}{r} + A''(r) - A(r)\,k^2 + \frac{A(r)\,\omega^2}{v^2}\right)\sin\left(kr - \omega t + \delta\right) +$$

$$\left(\frac{2\,k\,A(r)}{r} + 2\,A'(r)\,k\right)\cos\left(kr - \omega t + \delta\right) = 0\,.$$

This holds, for all r and t, if and only if the coefficients are zero,

$$\frac{2A'(r)}{r} + A''(r) - A(r)\,k^2 + A(r)\,\frac{\omega^2}{v^2} = 0 \qquad (*)$$

and

$$\frac{A(r)}{r} + A'(r) = 0\,. \qquad (**)$$

$(**)$ yields

$$\frac{dA}{dr} = -\frac{A}{r}\,, \qquad \int\frac{dA}{A} = -\int\frac{dr}{r}\,, \qquad \ln A = -\ln r + \ln A_0\,, \qquad A(r) = \frac{A_0}{r}\,.$$

With $A(r)$ known, $(*)$ reads

$$-\cancel{\frac{2\,A_0}{r^3}} + \cancel{\frac{2\,A_0}{r^3}} + \frac{A_0}{r}\left(\frac{\omega^2}{v^2} - k^2\right) = 0$$

i.e., it gives the dispersion relation $\omega = v\,k\,.$

Summing up, a spherical harmonic wave that solves the wave equation

$$\Delta u - \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} = 0$$

reads

$$u(\underline{r}, t) = \frac{A_0}{r}\sin(kr - \omega t + \delta)$$

with $\omega = vk$.

If it is not centered at the origin, but rather at a point with position vector $\underline{r}_0$, we find by a coordinate transformation $\underline{r} \mapsto \underline{r} - \underline{r}_0$:

$$u(\underline{r}, t) = \frac{A_0}{|\underline{r} - \underline{r}_0|}\sin\Big(k|\underline{r} - \underline{r}_0| - \omega t + \delta\Big) .$$

The intensity of a spherical harmonic wave

$$u(\underline{r}, t) \;=\; \frac{A_0}{r}\sin(kr - \omega t + \delta)$$

is the time-average of $u(\underline{r}\,,\,t)^2$:

$$I(\underline{r}) = \overline{u(\underline{r}\,,t)^2} \;=\; \frac{\omega}{2\pi}\int_0^{2\pi/\omega} u(\underline{r}\,,\,t)^2\,dt \;=\; \frac{A_0^2}{r^2}\,\frac{\omega}{2\pi}\int_0^{2\pi/\omega}\sin^2(kr - \omega t + \delta)\,dt\;.$$

Substituting $\xi = \omega t$ and using $\frac{1}{2\pi}\int_0^{2\pi}\sin^2(\xi + \alpha)d\xi = \frac{1}{2}$,

$$I(\underline{r}) = \overline{u(\underline{r}\,,t)^2} \;=\; \frac{A_0^2}{2\,r^2}\;.$$

So the intensity integrated over a sphere of radius r is independent of r:

$$\int_0^{2\pi}\int_0^{\pi}\overline{u(\underline{r}\,,t)^2}\,\underbrace{r^2\sin\vartheta\,d\vartheta\,d\varphi}_{\text{area element}} \;=\; \int_0^{2\pi}\int_0^{\pi}\frac{A_0^2}{2\,\cancel{r^2}}\,\cancel{r^2}\sin\vartheta\,d\vartheta\,d\varphi \;=\; \frac{A_0^2}{\cancel{2}}\,\cancel{2}\,2\pi\;.$$

d) Three-dimensional vector waves

The most important example for a vector wave in three dimensions is an electromagnetic wave.

An electromagnetic wave consists of two vector fields $\underline{E}(\underline{r}, t)$ and $\underline{B}(\underline{r}, t)$ which are dynamically coupled by Maxwell's equations.

Radio waves, visible light, X-rays and Gamma-rays are all electromagnetic waves (of different frequency ranges, see picture next page).

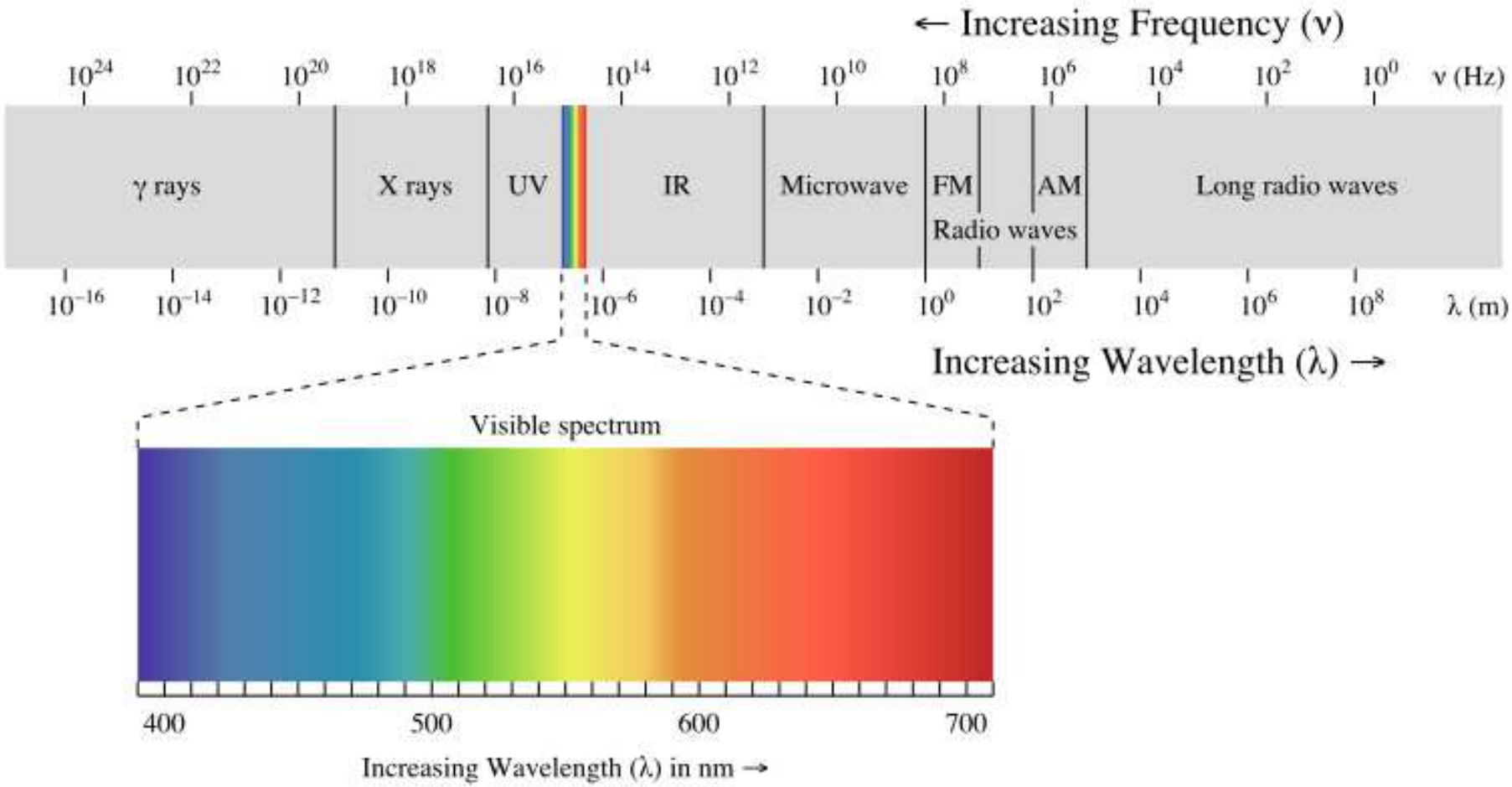
In principle, all properties of light can be derived from Maxwell's equations.

Below we derive the wave equation for $\underline{E}$ and $\underline{B}$ from Maxwell's equations.

James Clerk Maxwell (1831–1879)

picture from http://en.wikipedia.org/

Spectrum of electromagnetic waves



picture from http://en.wikipedia.org/

Maxwell's theory of electromagnetic waves was verified by Heinrich Hertz's experiments.

In 1888, Hertz produced radio waves in the laboratory in agreement with Maxwell's theory.

Heinrich Hertz (1857 – 1894)

picture from http://en.wikipedia.org/

"It's of no use whatsoever [...] this is just an experiment that proves Maestro Maxwell was right – we just have these mysterious electromagnetic waves that we cannot see with the naked eye. But they are there."

Heinrich Hertz

α) Derivation of the wave equation from Maxwell's equations

In SI units, the source-free Maxwell equations read

$$\nabla \cdot \underline{B} = 0 \,,$$

$$\nabla \times \underline{E} + \frac{\partial}{\partial t} \underline{B} = \underline{0} \,,$$

$$\nabla \cdot \underline{D} = 0 \,,$$

$$\nabla \times \underline{H} - \frac{\partial}{\partial t} \underline{D} = \underline{0} \,.$$

The vector fields $\underline{E}(\underline{r}, t)$, $\underline{B}(\underline{r}, t)$, $\underline{D}(\underline{r}, t)$ and $\underline{H}(\underline{r}, t)$ are related by constitutive equations that characterise the medium. We consider the simplest case of constitutive equations,

$$\underline{D}(\underline{r}, t) = \varepsilon_r \, \varepsilon_0 \, \underline{E}(\underline{r}, t) \,, \qquad \underline{B}(\underline{r}, t) = \mu_r \, \mu_0 \, \underline{H}(\underline{r}, t) \,,$$

where ε_r and μ_r are constants that characterise the medium. For vacuum, $\varepsilon_r = \mu_r = 1$.

ε_0 and μ_0 are constants of nature. In SI units:

$\varepsilon_0 =$ absolute permittivity $= 8.85 \cdot 10^{-12} \, \frac{C^2}{N \, m^2}$

$\mu_0 =$ absolute permeability $= 1.26 \cdot 10^{-6} \, \frac{N}{A^2}$

Then Maxwell's equations read

$$\nabla \cdot \underline{B} = 0\,,$$

$$\nabla \times \underline{E} + \frac{\partial}{\partial t}\underline{B} = \underline{0}\,,$$

$$\nabla \cdot \underline{E} = 0\,,$$

$$\nabla \times \underline{B} - \frac{1}{v^2}\frac{\partial}{\partial t}\underline{E} = \underline{0}\,,$$

where $v = \dfrac{1}{\sqrt{\varepsilon_r\,\mu_r\,\varepsilon_0\,\mu_0}}$. From the second Maxwell equation:

$$\nabla \times (\nabla \times \underline{E}) = -\nabla \times \frac{\partial \underline{B}}{dt}$$

$$\nabla(\nabla \cdot \underline{E}) - (\nabla \cdot \nabla)\underline{E} = -\frac{\partial}{\partial t}\left(\nabla \times \underline{B}\right)$$

$$\underline{0} - \Delta \underline{E} = -\frac{\partial}{\partial t}\left(\frac{1}{v^2}\frac{\partial \underline{E}}{\partial t}\right)$$

$$\boxed{\Delta \underline{E} - \frac{1}{v^2}\frac{\partial^2 \underline{E}}{\partial t^2} = 0\,.}$$

An analogous calculation yields

$$\boxed{\Delta \underline{B} - \frac{1}{v^2}\frac{\partial^2 \underline{B}}{\partial t^2} = 0\,.}$$

Every component of $\boldsymbol{E}$ and every component of $\boldsymbol{B}$ satisfies the three-dimensional wave equation. Thus, Maxwell's equations predict the existence of electromagnetic waves that propagate with velocity

$$\boxed{v = \frac{1}{\sqrt{\varepsilon_r\,\mu_r \varepsilon_0\,\mu_0}} = \frac{c}{n}}$$

where

index of refraction: $\boxed{n = \sqrt{\varepsilon_r\,\mu_r}}$

vacuum velocity of light: $\boxed{c = \frac{1}{\sqrt{\varepsilon_0\,\mu_0}} \approx 300,000 \text{ km/sec}}$

Propagation of electromagnetic waves (light) in a medium with constant ε_r and constant μ_r is isotropic and non-dispersive. Propagation of electromagnetic waves in gases or fluids is, in general, isotropic and dispersive, in crystals it is anisotropic and dispersive. (Glass is an extremely viscous fluid, not a crystal.)

The interstellar medium is dispersive, i.e., the velocity of electromagnetic waves depends on the wave length.

Arrival of radiation from supernova 1987A in the Large Magellanic Cloud:

type	date	time
[neutrino	February 23	07:35]
X rays	February 23	09:00
light	February 24	07:00
radio	February 25	10:00

(Distance $\approx 160,000$ light years)

picture from http://www.dsd.lbl.gov/

β) Plane harmonic electromagnetic waves

Ansatz for harmonic plane electromagnetic waves:

$$\underline{E}(\underline{r},t) = \underline{E}_0 \cos(\underline{k}\cdot\underline{r} - \omega t + \alpha)\,, \qquad \underline{B}(\underline{r},t) = \underline{B}_0 \cos(\underline{k}\cdot\underline{r} - \omega t + \beta)\,.$$

Feed this into Maxwell's equations for medium with constant ε_r and constant μ_r:

$$\nabla\cdot\underline{B} = 0\,,$$

$$\nabla\times\underline{E} + \frac{\partial}{\partial t}\underline{B} = \underline{0}\,,$$

$$\nabla\cdot\underline{E} = 0\,,$$

$$\nabla\times\underline{B} - \frac{1}{v^2}\frac{\partial}{\partial t}\underline{E} = \underline{0}\,,$$

where

$$v = \frac{c}{n} = \frac{1}{\sqrt{\varepsilon_r\,\mu_r\,\varepsilon_0\,\mu_0}}$$

is the velocity of light.

Consider the first and the third of the Maxwell equations:

$$\nabla \cdot \big(\underline{E}_0 \cos(\underline{k} \cdot \underline{r} - \omega t + \alpha)\big) = -\underline{E}_0 \cdot \underline{k} \sin(\underline{k} \cdot \underline{r} - \omega t + \alpha) = 0 \ ,$$

$$\nabla \cdot \big(\underline{B}_0 \cos(\underline{k} \cdot \underline{r} - \omega t + \beta)\big) = -\underline{B}_0 \cdot \underline{k} \sin(\underline{k} \cdot \underline{r} - \omega t + \beta) = 0 \ .$$

These imply

$$\underline{E}_0 \cdot \underline{k} \ = \ 0 \, , \qquad \underline{B}_0 \cdot \underline{k} \ = \ 0 \, .$$

i.e., electromagnetic waves are transverse.

Now consider the remaining two Maxwell equations:

$$\frac{\partial\big(\underline{E}_0 \cos(\underline{k} \cdot \underline{r} - \omega t + \alpha)\big)}{\partial t} - v^2 \nabla \times \big(\underline{B}_0 \cos(\underline{k} \cdot \underline{r} - \omega t + \beta)\big) =$$

$$= \underline{E}_0 \, \omega \sin(\underline{k} \cdot \underline{r} - \omega t + \alpha) + v^2 \underline{k} \times \underline{B}_0 \sin(\underline{k} \cdot \underline{r} - \omega t + \beta)\big) = \underline{0} \ ,$$

$$\frac{\partial\big(\underline{B}_0 \cos(\underline{k} \cdot \underline{r} - \omega t + \beta)\big)}{\partial t} + \nabla \times \big(\underline{E}_0 \cos(\underline{k} \cdot \underline{r} - \omega t + \alpha)\big) =$$

$$= \underline{B}_0 \, \omega \sin(\underline{k} \cdot \underline{r} - \omega t + \beta) - \underline{k} \times \underline{E}_0 \sin(\underline{k} \cdot \underline{r} - \omega t + \alpha)\big) = \underline{0} \ .$$

These imply

$$\alpha = \beta$$

i.e., electric and magnetic field are in phase, and

$$\underline{E}_0 = -\frac{v^2}{\omega}\,\underline{k} \times \underline{B}_0\,, \qquad \underline{B}_0 = \frac{1}{\omega}\,\underline{k} \times \underline{E}_0\,.$$

Thus,

$$\boxed{\underline{k}\,,\ \underline{E}_0 \text{ and } \underline{B}_0 \text{ are pairwise orthogonal}}$$

and their magnitudes satisfy

$$\left|\underline{E}_0\right| = v\left|\underline{B}_0\right|, \qquad \omega = v\left|\underline{k}\right|.$$

Summing up: For a harmonic plane electromagnetic wave in a medium with constant ε_r and constant μ_r, the electric field

$$\underline{E}(\underline{r}\,,t) = \underline{E}_0 \cos(\underline{k}\cdot\underline{r} - \omega t + \alpha)\,,$$

determines the corresponding magnetic field to be

$$\underline{B}(\underline{r}\,,t) = \frac{1}{v}\,\frac{\underline{k}}{|\underline{k}|} \times \underline{E}(\underline{r}\,,t)\,.$$

Thus, it suffices to know $\underline{E}$.

The "intensity" ("irradiance") of a plane harmonic electromagnetic wave with electric field $\underline{E}(\underline{r}, t) = \underline{E}_0 \cos(\underline{k} \cdot \underline{r} - \omega t + \alpha)$ is the time-average of $|\underline{E}(\underline{r}, t)|^2$:

$$I(\underline{r}) = \overline{|\underline{E}(\underline{r}, t)|^2} = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} |\underline{E}(\underline{r}, t)|^2 \, dt =$$

$$= |\underline{E}_0|^2 \frac{\omega}{2\pi} \int_0^{2\pi/\omega} \cos^2(\underline{k} \cdot \underline{r} - \omega t + \delta) \, dt = \frac{1}{2} |\underline{E}_0|^2 .$$

In light from an ordinary light source (Sun, light bulb, ...), the direction of the vector $\underline{E}_0$ varies randomly over all directions perpendicular to $\underline{k}$. This is called unpolarised light.

If a certain direction of $\underline{E}_0$ is filtered out, one gets polarised light. We will discuss polarisation of light later in detail.

The vector character of light is crucial for the understanding of polarisation experiments, but it is irrelevant for diffraction and interference experiments. For the latter case, one often models light in terms of a fictitious scalar field $u(\underline{r}, t)$, in analogy to sound waves, with the irradiance given as the time-average of $u(\underline{r}, t)^2$. This is called scalar optics.

III. Diffraction

(cf. Pedrotti and Pedrotti, Sec. 3-1, Chapters 16, 17 and 18)

a) Huygens' Principle

Huygens' Principle

- was formulated by C. Huygens in his "Traitée de la lumière" (submitted to the Académie Française in 1678);
- describes propagation of waves in terms of wave fronts;
- applies to waves described by a large class of differential equations, including the classical wave equation;
- allows to rederive the reflection law and the refraction law in terms of wave fronts;
- gives a basic understanding of diffraction.

Christiaan Huygens (1629 - 1695)

picture from http://de.wikipedia.org/

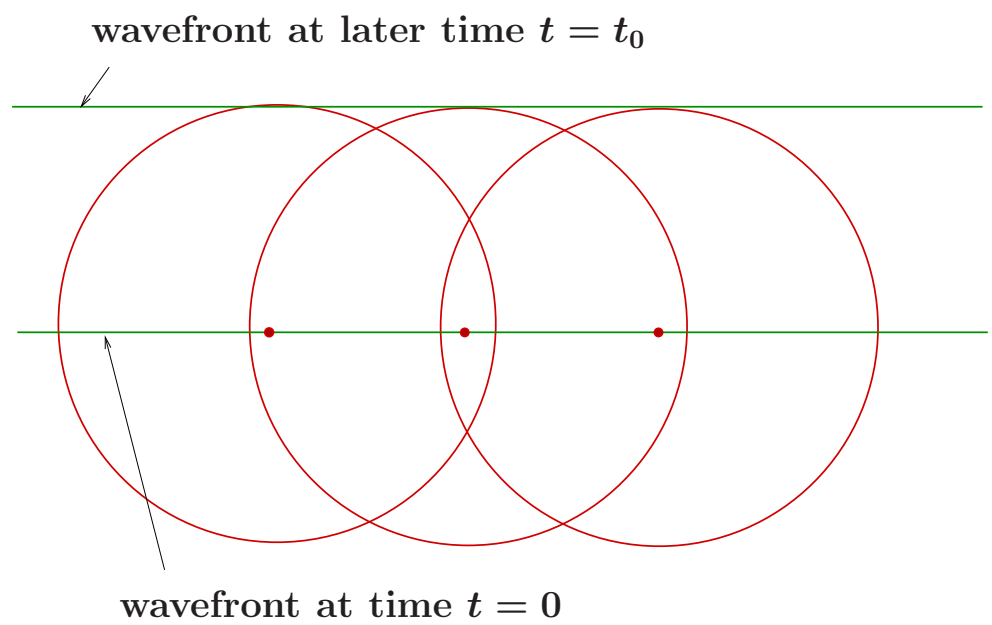
Huygens' Principle: Every point in a wavefront is the source of an elementary wave (called "wavelet") which propagates in all directions with the speed of the wave. Later wavefronts are the envelopes of these wavelets at fixed time.

We now give several applications of Huygens' Principle to optics.

α) Medium with constant velocity of light

In a medium with constant velocity of light v (i.e., constant index of refraction $n = c/v$), the wavelets are spheres that expand with velocity v.

A plane wavefront remains a plane wavefront parallel to itself.



β) Reflection law in terms of wavefronts

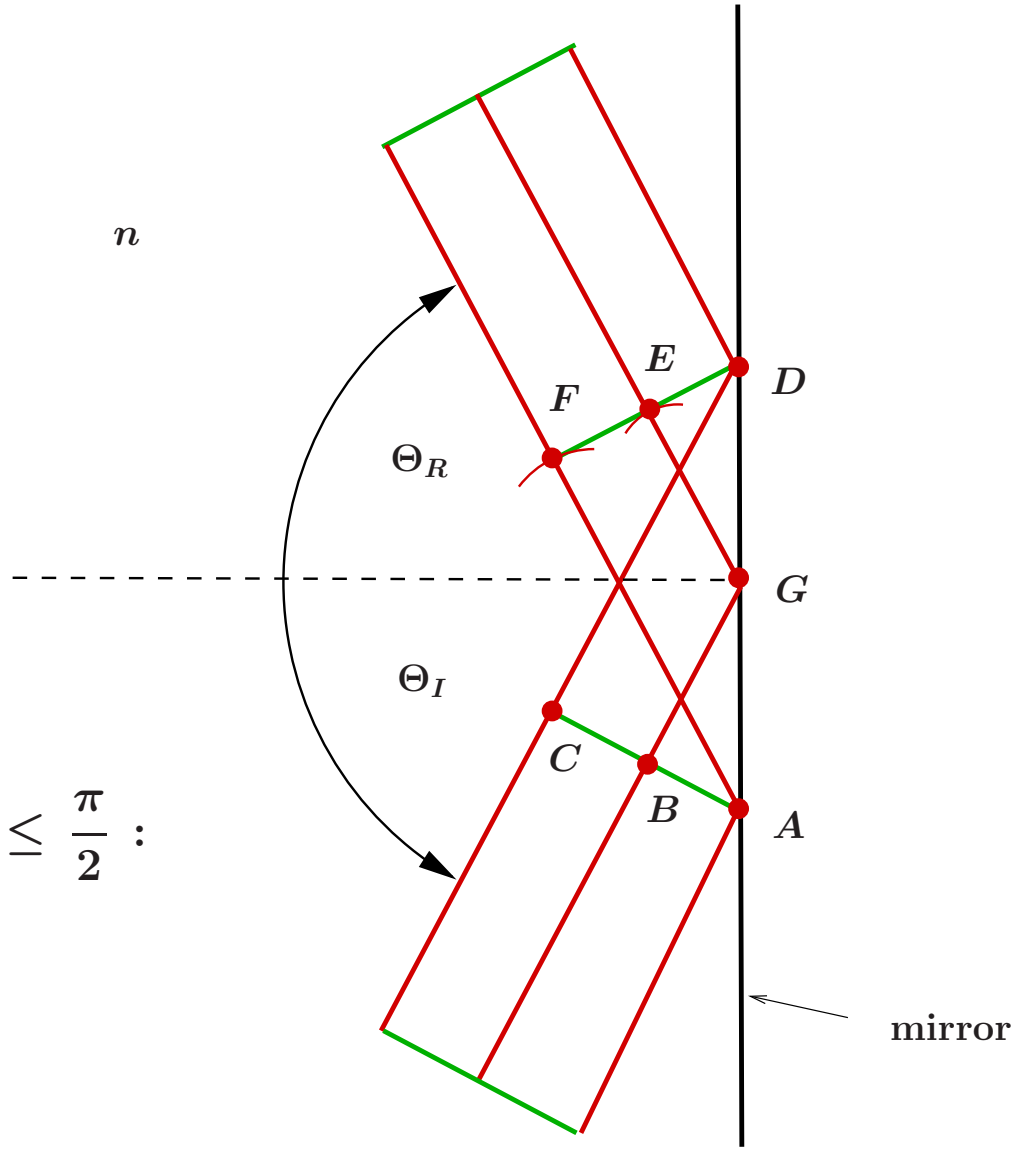
$$\frac{n}{c}\,\overline{\mathrm{CD}} = \frac{n}{c}\left(\overline{\mathrm{BG}} + \overline{\mathrm{GE}}\right) = \frac{n}{c}\,\overline{\mathrm{AF}}$$

$$\sin\Theta_I = \overline{\mathrm{CD}} : \overline{\mathrm{AD}}\,, \qquad \sin\Theta_R = \overline{\mathrm{AF}} : \overline{\mathrm{AD}}\,,$$

$$\sin\Theta_I = \sin\Theta_R$$

As $\quad -\dfrac{\pi}{2} \le \Theta_I \le \dfrac{\pi}{2} \quad$ and $\quad -\dfrac{\pi}{2} \le \Theta_R \le \dfrac{\pi}{2}$:

$$\boxed{\Theta_I = \Theta_R}$$



γ) Refraction law in terms of wavefronts

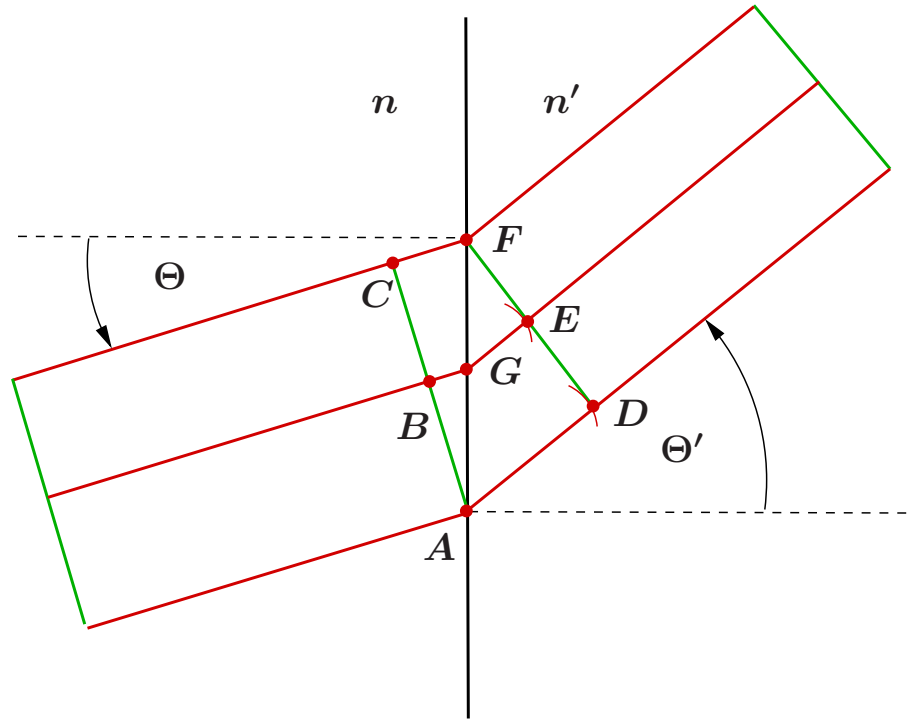
$$\frac{n'}{c}\,\overline{\mathrm{AD}} = \frac{n}{c}\,\overline{\mathrm{BG}} + \frac{n'}{c}\,\overline{\mathrm{GE}} = \frac{n}{c}\,\overline{\mathrm{CF}} \qquad (*)$$

$$\sin\Theta = \overline{\mathrm{CF}} : \overline{\mathrm{AF}}\,, \qquad \sin\Theta' = \overline{\mathrm{AD}} : \overline{\mathrm{AF}}\,,$$

$$\overline{\mathrm{AF}} = \overline{\mathrm{CF}} : \sin\Theta = \overline{\mathrm{AD}} : \sin\Theta' \qquad (**)$$

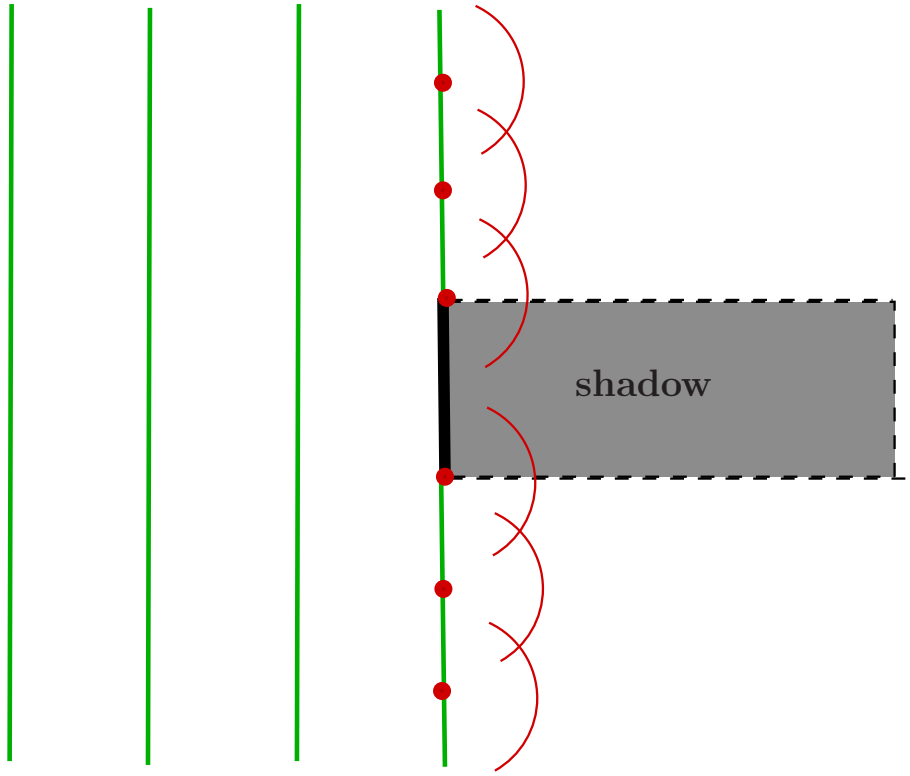
$(*)$ and $(**)$ $\Longrightarrow$

$$\boxed{n\sin\Theta = n'\sin\Theta'}$$



δ) Huygens' Principle and diffraction

Huygens' Principle indicates that waves propagate into the geometric shadow.



This is called diffraction.

With light, diffraction was first observed by Francesco Grimaldi (1618–1663) whose observations were published in 1665.

Huygens did not know about Grimaldi's observation and tried to find arguments why, in spite of the principle named after him, light should show no diffraction.

In its original version Huygens' principle cannot explain diffraction quantitatively; it does not allow to calculate the wave field $u(\underline{r}, t)$.

Augustin Fresnel added to the Huygens principle the assumption that the wave field can be calculated as a superposition of all wavelets originating from one wave front, taking their amplitudes and phases into account. This is called the Huygens-Fresnel Principle.

In the following we work out some mathematical details of the Huygens-Fresnel Principle.

Augustin Fresnel (1788–1827)

picture from http://en.wikipedia.org/

Recall (p. 69) :

A spherical harmonic wave that solves the classical wave equation

$$\Delta u - \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} = 0$$

is of the form

$$u(\underline{r}, t) = \frac{A_0}{|\underline{r} - \underline{r}_0|}\sin\big(k|\underline{r} - \underline{r}_0| - \omega t + \delta\big)$$

where $\omega = vk$.

In complex notation, using $\sin\varphi = \mathrm{Re}\big\{ -ie^{i\varphi}\big\}$,

$$u(\underline{r}, t) = \mathrm{Re}\Big\{\frac{-i\,A_0}{|\underline{r} - \underline{r}_0|}\,e^{i\left(\frac{\omega}{v}|\underline{r} - \underline{r}_0| - \omega t + \delta\right)}\Big\} = \mathrm{Re}\Big\{ -i\,A_0\,e^{i\delta}\,e^{-i\omega t}\,\frac{e^{i\frac{\omega}{v}|\underline{r} - \underline{r}_0|}}{|\underline{r} - \underline{r}_0|}\Big\}$$

The Huygens-Fresnel Principle says that superposition of such wavelets, with A_0, δ, ω fixed and $\underline{r}_0$ ranging over one wavefront, gives the wave field.

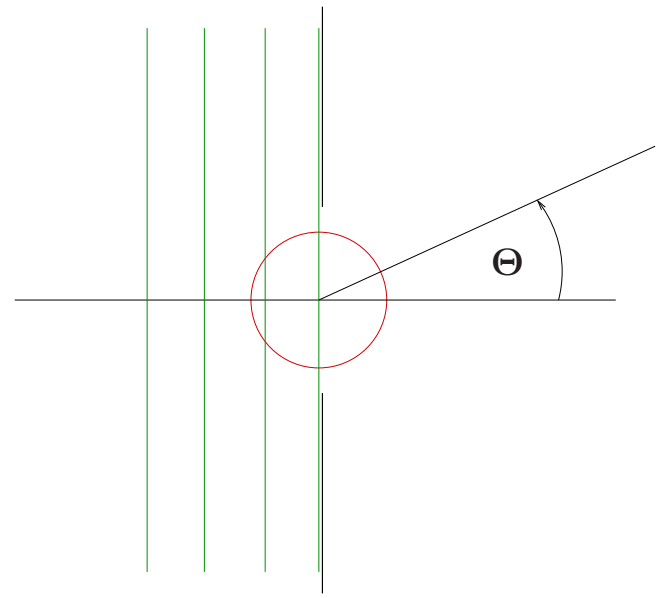
$-ie^{i\delta} = e^{i(\delta - \pi/2)}$ can be made equal to 1 by shifting the zero on the time axis, $t \mapsto t - t_0$ with $t_0 = \frac{1}{\omega}\big(\frac{\pi}{2} - \delta\big)$.

The Huygens-Fresnel Principle is the main tool for calculating diffraction patterns.

However, something must be wrong with the Huygens-Fresnel Principle in this form:

Actually, Huygens wavelets should not radiate in the backward direction because no light goes back from the aperture in the direction from which the light came in.

Therefore, Fresnel introduced the socalled "obliquity factor" $\frac{1}{2}\left(1 + \cos\Theta\right)$ with which the amplitude of each wavelet must be multiplied.



Forward direction, $\Theta = 0$: $\frac{1}{2}(1 + \cos\Theta) = 1$

Backward direction, $\Theta = \pi$: $\frac{1}{2}(1 + \cos\Theta) = 0$

[A mathematical justification for the obliquity factor was later given by Gustav Kirchhoff (1824–1887).]

So the Huygens-Fresnel wavelets finally read

$$u(\underline{r}, t) = \mathrm{Re}\left\{ \frac{A_0}{2}\left(1 + \cos\Theta\right) e^{-i\omega t} \frac{e^{i\frac{\omega}{v}|\underline{r} - \underline{r}_0|}}{|\underline{r} - \underline{r}_0|} \right\} .$$

This is the basis of the diffraction theory as it was worked out by Kirchhoff. In the following we discuss the diffraction patterns for apertures of various shapes.

b) Fraunhofer diffraction

We will discuss the diffraction patterns for planar apertures of simple geometry, with the help of the Huygens-Fresnel principle. One distinguishes:

Fraunhofer diffraction:

- **Light source so far away from aperture that incoming wavefront can be considered as plane.**
- **Observation screen so far away from aperture that the fronts of the wavelets, originating from the points of the aperture, can be considered as plane when they arrive at the screen.**

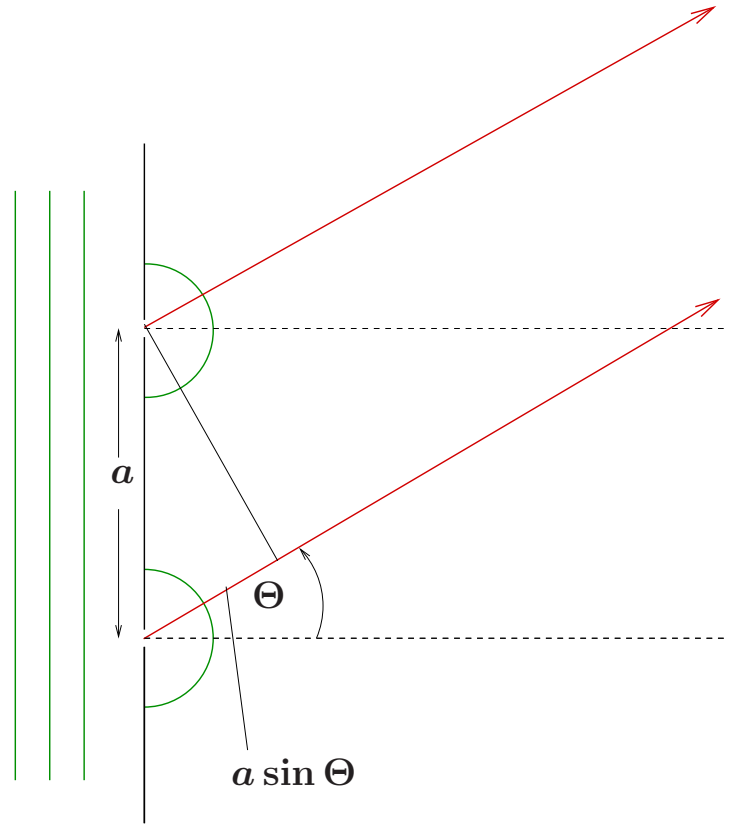
Fresnel diffraction: Otherwise

The conditions of Fraunhofer diffraction can be achieved with the help of collimator lenses (see 3rd worksheet).

For all examples in this section we

- assume that the index of refraction is constant;
- use scalar optics (i.e., we describe light in terms of a scalar function $u(\underline{r}, t)$);
- assume that Fraunhofer diffraction is applicable and that Θ is so small that the obliquity factor $\frac{1}{2}(1 + \cos\Theta)$ can be approximated by 1.

α) Two narrow slits



Consider two narrow slits with separation a.

Restrict to plane perpendicular to the slits.

If the slits are very narrow, the resulting wave is the superposition of just two wavelets, centered in the slits.

As we assume Fraunhofer diffraction, they can be approximated as plane harmonic waves near a screen point.

Wavelet from first slit:

$$u_0(\underline{r}, t) = \mathrm{Re}\left\{ A\, e^{i(\underline{k} \cdot \underline{r} - \omega t)} \right\}$$

For the second wavelet, the distance to the screen is longer by $a \sin \Theta$ (see picture), so with respect to the first it has a phase shift $\Delta = k\, a \sin \Theta$.

Wavelet from second slit:

$$u_1(\underline{r}, t) = \mathrm{Re}\left\{ A\, e^{i(\underline{k} \cdot \underline{r} - \omega t + \Delta)} \right\}$$

The superposition of them is

$$u(\underline{r},t) \;=\; u_0(\underline{r},t) \;+\; u_1(\underline{r},t) \;=\; \mathrm{Re}\{\, A\, e^{i(\underline{k}\cdot\underline{r}-\omega t)}\,(1+e^{i\Delta})\,\}$$

$$=\; \mathrm{Re}\{\, A\, e^{i(\underline{k}\cdot\underline{r}-\omega t)}\, e^{i\frac{\Delta}{2}}\,(e^{-i\frac{\Delta}{2}}+e^{i\frac{\Delta}{2}})\,\} \;=\; \mathrm{Re}\{\, A\, e^{i(\underline{k}\cdot\underline{r}-\omega t+\frac{\Delta}{2})}\, 2\cos(\frac{\Delta}{2})\,\}$$

$$=\; 2\,A\cos(\underline{k}\cdot\underline{r}-\omega t+\frac{\Delta}{2})\cos(\frac{\Delta}{2})$$

The intensity (irradiance) is the time-average of the square of $u(\underline{r},t)$:

$$I(\Theta) \;=\; \frac{\omega}{2\,\pi}\int_{t_0}^{t_0+\frac{2\,\pi}{\omega}} u(\underline{r},t)^2\,dt \;=\; 4\,A^2\cos^2(\frac{\Delta}{2})\,\frac{\omega}{2\,\pi}\int_{t_0}^{t_0+\frac{2\,\pi}{\omega}} \cos^2(\underline{k}\cdot\underline{r}-\omega t+\frac{\Delta}{2})\,dt$$

With

$$\frac{\omega}{2\,\pi}\int_{t_0}^{t_0+\frac{2\,\pi}{\omega}} \cos^2(\alpha-\omega t)\,dt \;=\; \frac{1}{2}$$

$$I(\Theta) \;=\; 2\,A^2\cos^2(\frac{\Delta}{2})\,.$$

With $\Delta = k\,a \sin\Theta = \frac{2\pi}{\lambda}\,a \sin\Theta$:

$$I(\Theta) = 2\,A^2 \cos^2\left(\frac{\pi\,a}{\lambda}\sin\Theta\right)$$

As $I(0) = 2A^2$:

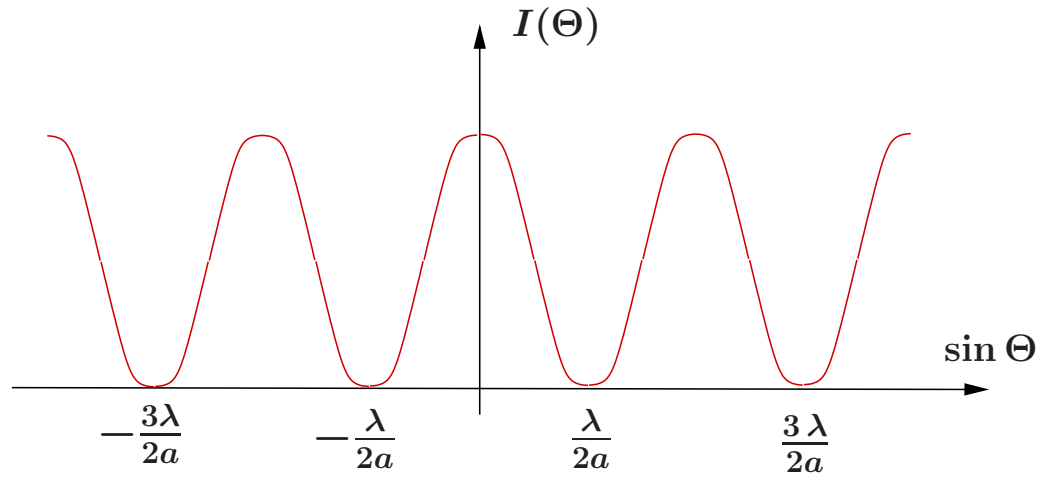
$$I(\Theta) = I(0)\cos^2\left(\frac{\pi\,a}{\lambda}\sin\Theta\right) .$$

m^{th} minimum $(m = 0, \pm 1, \pm 2, \ldots)$:

$$\boxed{\sin\Theta_m = \frac{(2m-1)}{2}\frac{\lambda}{a}}$$

m^{th} maximum $(m = 0, \pm 1, \pm 2, \ldots)$:

$$\boxed{\sin\Theta_m = m\frac{\lambda}{a}}$$



So we get a diffraction pattern of equally spaced bright stripes.

The spacing depends on wave length (i.e. colour).

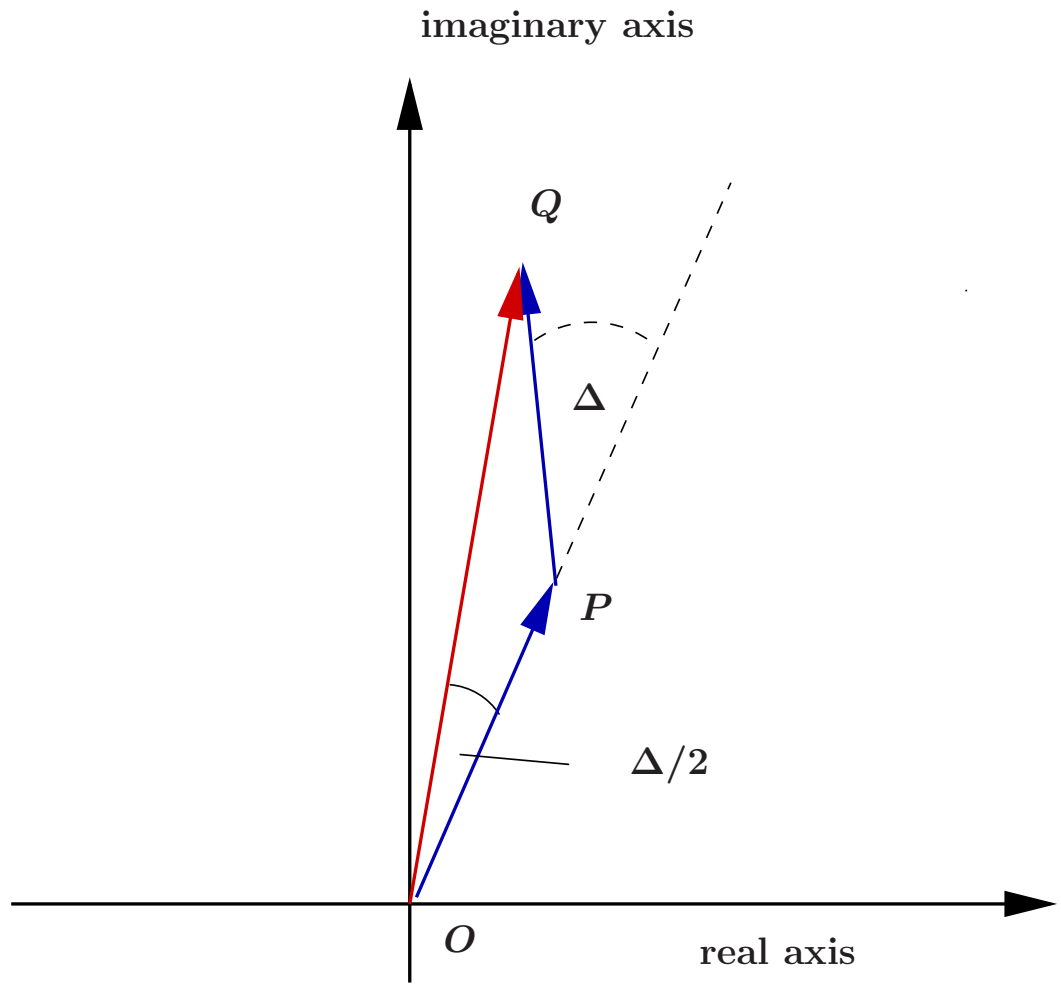
picture from www.itp.uni-hannover.de

The same result can be found by means of a phasor diagram:

Notation: $OP =$ complex number represented by arrow from O to P in complex plane

$$u_0(\underline{r}, t) = \mathrm{Re}\{OP\}$$
$$u_1(\underline{r}, t) = \mathrm{Re}\{PQ\} = \mathrm{Re}\{OP\, e^{i\Delta}\}$$
$$u(\underline{r}, t) = \mathrm{Re}\{OQ\}$$

Phasor diagram rotates rigidly with angular frequency ω.



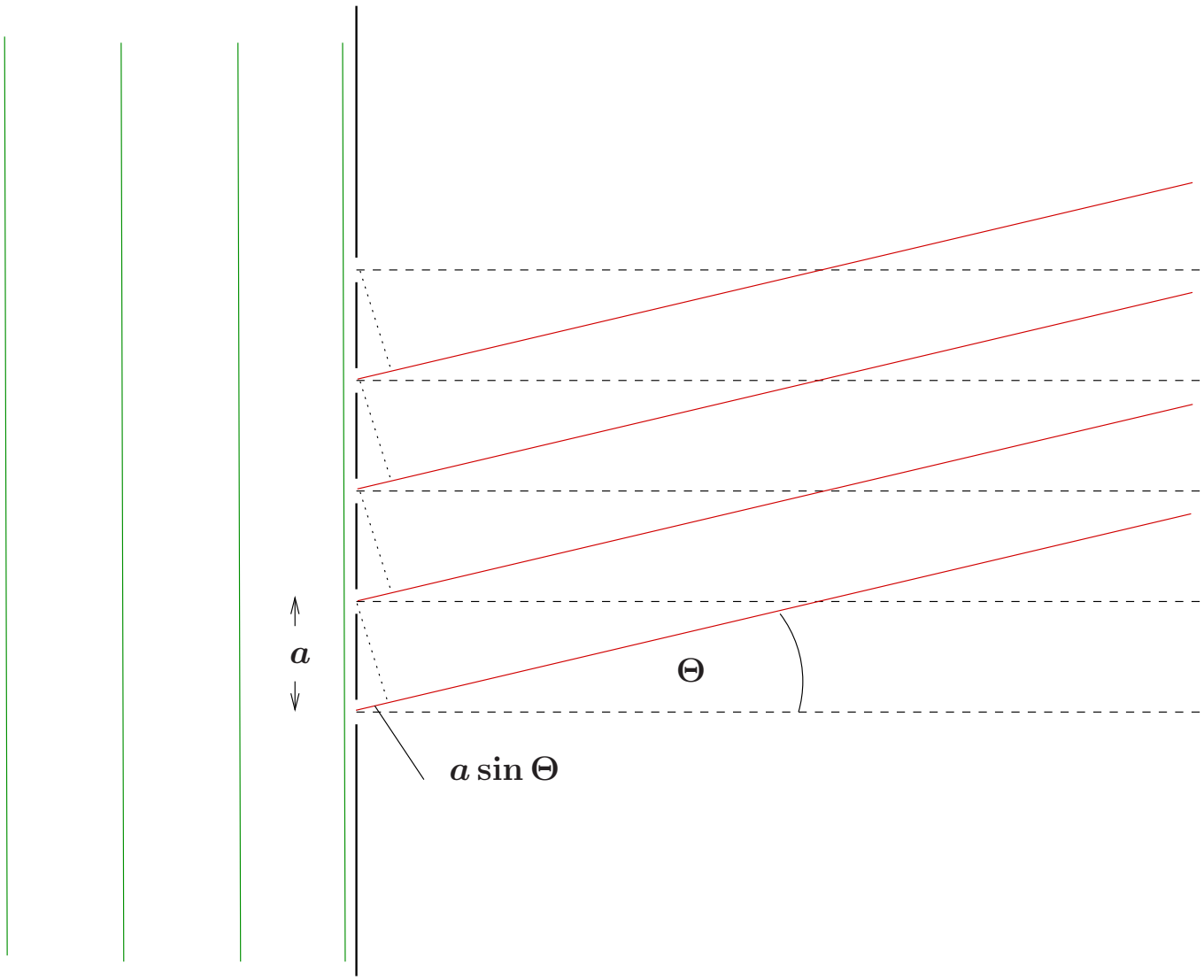
From the picture one reads:

$$\frac{1}{2}|OQ| = |OP| \cos\left(\frac{\Delta}{2}\right)$$

$$|OQ|^2 = 4\,|OP|^2 \cos^2\left(\frac{\Delta}{2}\right)$$

β) Grating

From two narrow slits we generalize to N narrow slits with equal separation a. This is called a "grating".



Now we have to add up N wavelets.

The wavelet from each subsequent slit gains a phase shift

$$\Delta = \frac{2\pi}{\lambda} a \sin\Theta$$

relative to the previous one.

Summing up all wavelets:

$$u(\underline{r},t) = \mathrm{Re}\left\{A\left(e^{i\left(\underline{k}\cdot\underline{r}-\omega t\right)} + e^{i\left(\underline{k}\cdot\underline{r}-\omega t+\Delta\right)} + \right.\right.$$

$$\left.\left. + e^{i\left(\underline{k}\cdot\underline{r}-\omega t+2\Delta\right)} + \ldots + e^{i\left(\underline{k}\cdot\underline{r}-\omega t+[N-1]\Delta\right)}\right)\right\} =$$

$$= \mathrm{Re}\left\{A\, e^{i\left(\underline{k}\cdot\underline{r}-\omega t\right)}\left(1 + e^{i\Delta} + e^{i2\Delta} + \ldots + e^{i[N-1]\Delta}\right)\right\}$$

Geometric progression: $\quad 1 + q + q^2 + \ldots + q^{N-1} = \dfrac{1-q^N}{1-q}$

$$u(\underline{r},t) = \mathrm{Re}\left\{A\, e^{i\left(\underline{k}\cdot\underline{r}-\omega t\right)}\,\frac{\left(1-e^{i\Delta N}\right)}{\left(1-e^{i\Delta}\right)}\right\} =$$

$$= \mathrm{Re}\left\{A\, e^{i\left(\underline{k}\cdot\underline{r}-\omega t\right)}\,\frac{e^{iN\Delta/2}\left(e^{-iN\Delta/2}-e^{iN\Delta/2}\right)}{e^{i\Delta/2}\left(e^{-i\Delta/2}-e^{i\Delta/2}\right)}\right\} =$$

$$= \mathrm{Re}\left\{A\, e^{i\left(\underline{k}\cdot\underline{r}-\omega t+\frac{N\Delta}{2}-\frac{\Delta}{2}\right)}\,\frac{\cancel{(-2i)}\sin(N\Delta/2)}{\cancel{(-2i)}\sin(\Delta/2)}\right\} =$$

$$= A\cos\left(\underline{k}\cdot\underline{r}-\omega t+[N-1]\Delta/2\right)\frac{\sin(N\Delta/2)}{\sin(\Delta/2)}$$

Gives intensity
$$I(\Theta) = \frac{\omega}{2\pi}\int_{t_0}^{t_0+\frac{2\pi}{\omega}} u(\underline{r},t)^2\,dt$$

$$= A^2\,\frac{\sin^2(N\Delta/2)}{\sin^2(\Delta/2)}\,\underbrace{\frac{\omega}{2\pi}\int_{t_0}^{t_0+\frac{2\pi}{\omega}} \cos^2\big(\underline{k}\cdot\underline{r} - \omega t + [N-1]\Delta/2\big)\,dt}_{1/2}$$

$$= \frac{A^2}{2}\,\frac{\sin^2(N\Delta/2)}{\sin^2(\Delta/2)}\,.$$

With $\Delta = \frac{2\pi}{\lambda}\,a\sin\Theta$:

$$I(\Theta) = \frac{A^2}{2}\,\frac{\sin^2\big(\frac{N\pi a}{\lambda}\sin\Theta\big)}{\sin^2\big(\frac{\pi a}{\lambda}\sin\Theta\big)}\,.$$

To calculate $I(0)$, we use the Taylor series of the sine function:

$$\frac{\sin(N\zeta)}{\sin\zeta} = \frac{N\zeta - \frac{1}{3!}N^3\zeta^3 + \ldots}{\zeta - \frac{1}{3!}\zeta^3 + \ldots} = \frac{N - \frac{1}{3!}N^3\zeta^2 + \ldots}{1 - \frac{1}{3!}\zeta^2 + \ldots} \longrightarrow N \quad \text{for } \zeta \to 0\,,$$

so

$$I(0) = A^2N^2/2\,.$$

Thus, the intensity is

$$I(\Theta) = I(0)\,\frac{\sin^2\left(\frac{N\pi a}{\lambda}\sin\Theta\right)}{N^2\sin^2\left(\frac{\pi a}{\lambda}\sin\Theta\right)}\,.$$

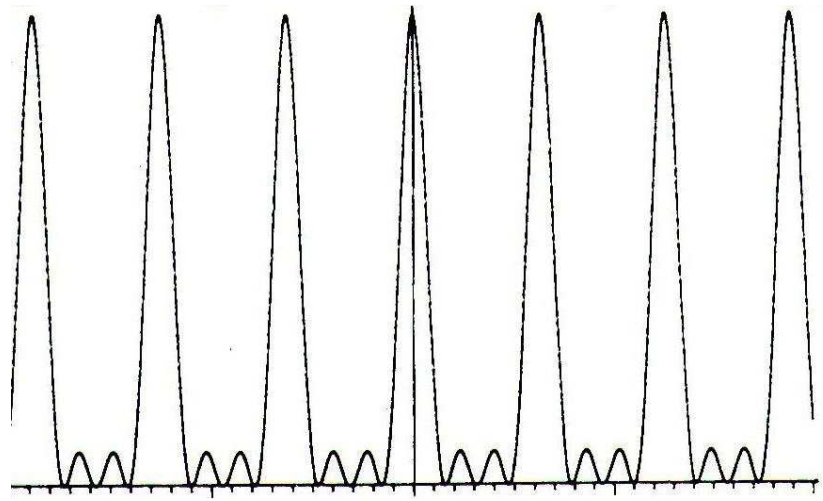
The pictures show $I(\Theta)$ over $\sin\Theta$ for two values of N.

With increasing N, the primary maxima become sharper and the secondary maxima are more and more suppressed.

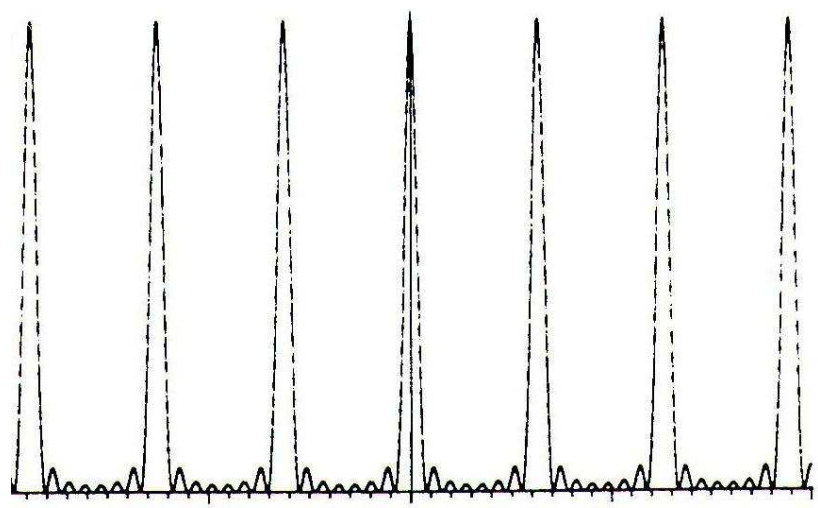
Primary maxima occur at

$$\boxed{\sin\Theta_m = m\,\frac{\lambda}{a}}$$

$$m = 0, \pm 1, \pm 2, \ldots$$



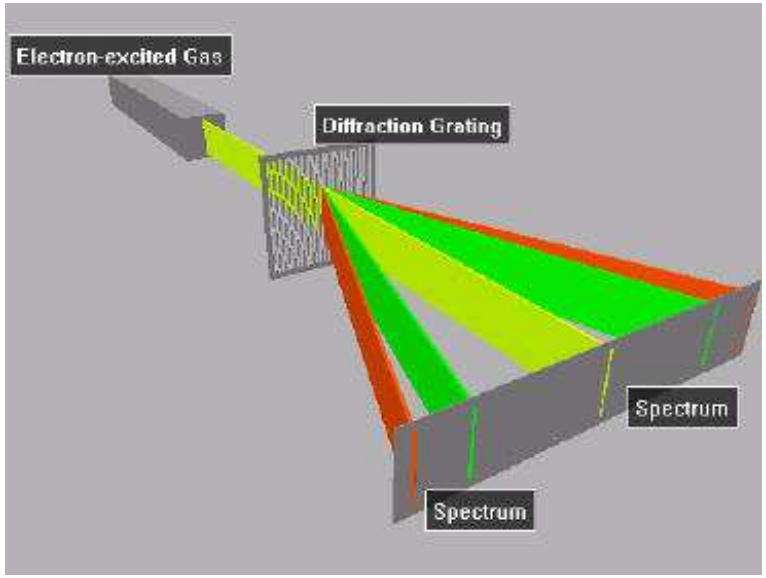
$N = 4$

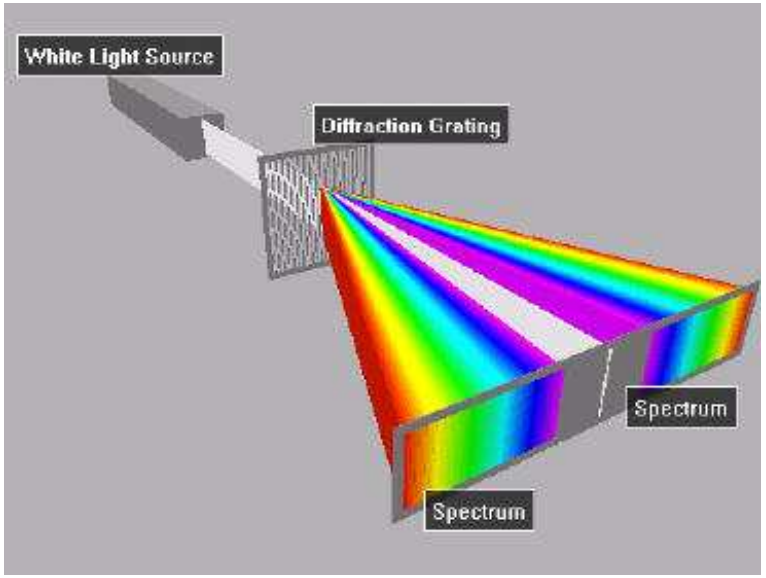


$N = 8$

picture from http://www.physik.fu-berlin.de/ brewer/

Dependence on λ implies that a grating can be used for spectral decomposition:

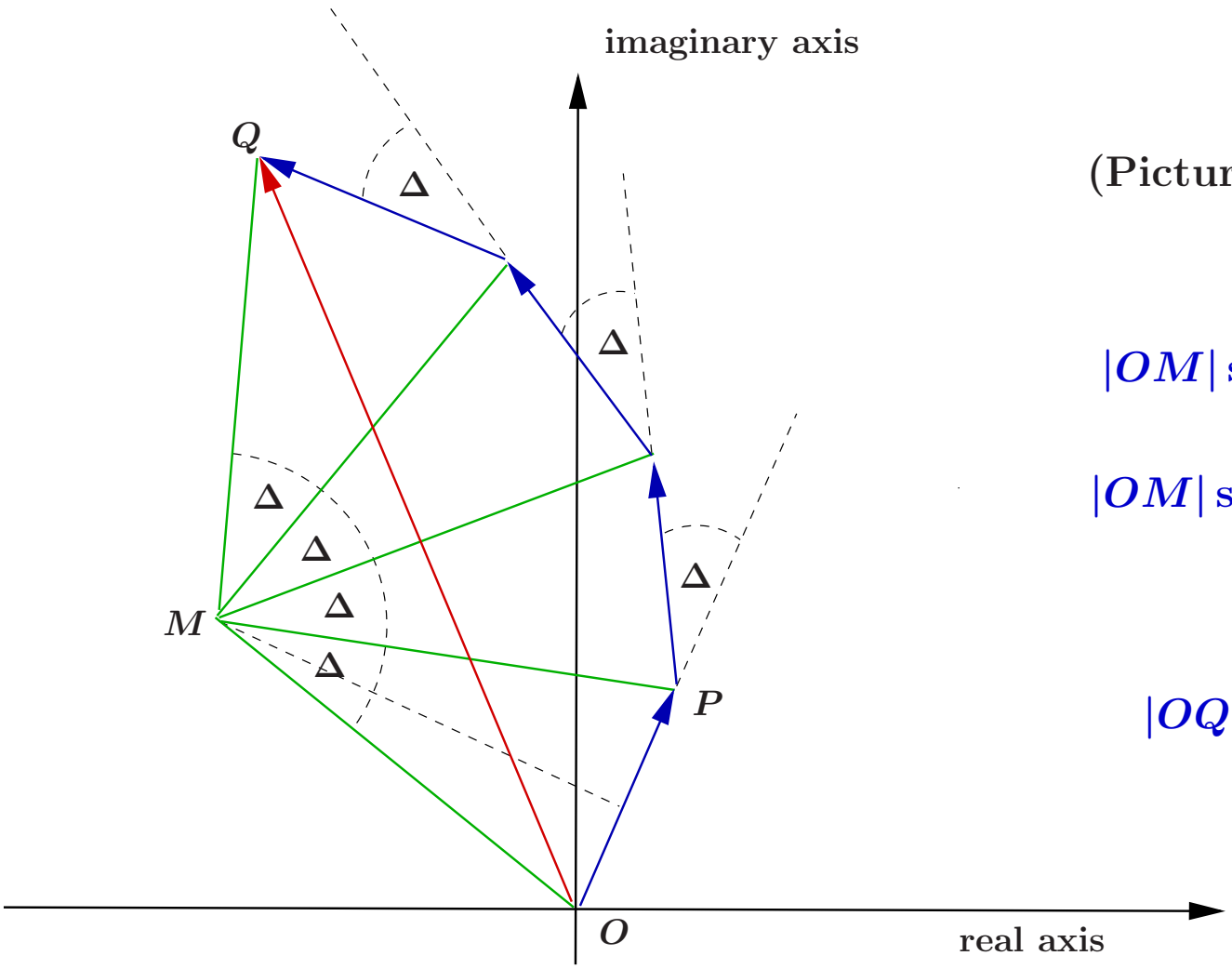




pictures from http://des.memphis.edu

For instance, in astronomy gratings are used for spectral analysis of celestial bodies.

Geometric construction of the N slits diffraction pattern with the help of phasors:



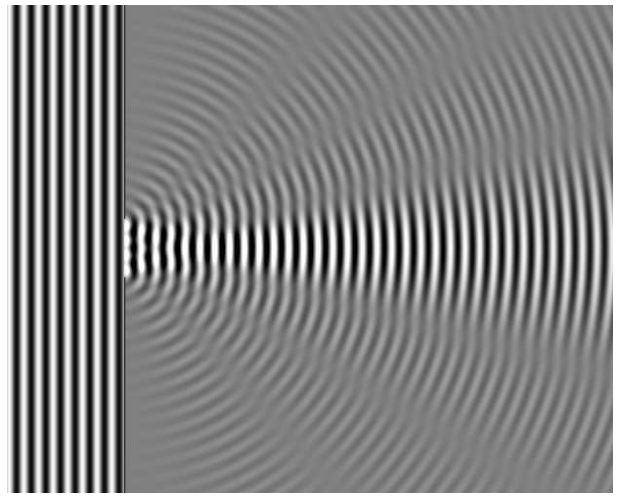
(Picture for $N = 4$ slits)

$$\left.\begin{aligned} |OM| \sin\frac{\Delta}{2} &= \frac{1}{2}|OP| \\ |OM| \sin\frac{N\Delta}{2} &= \frac{1}{2}|OQ| \end{aligned}\right\} \Longrightarrow$$

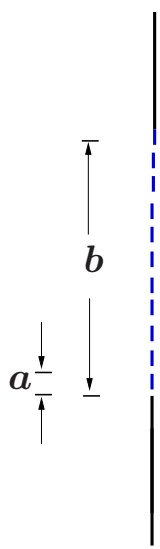
$$|OQ|^2 = |OP|^2 \frac{\sin^2\frac{N\Delta}{2}}{\sin^2\frac{\Delta}{2}}$$

γ) Single slit

Diffraction pattern produced by a single slit with water waves.



picture from http://content.answers.com



Consider single slit of finite width b as the limit of a grating:

Let $b = Na$, keep b fixed, send $N \to \infty$ (thus $a \to 0$).

For finite N, we know (p.98):

$$I(\Theta) = I(0)\,\frac{\sin^2\left(\frac{N\,\pi\,a\sin\Theta}{\lambda}\right)}{N^2\sin^2\left(\frac{\pi\,a\sin\Theta}{\lambda}\right)} = I(0)\,\frac{\sin^2\left(\frac{\pi\,b\sin\Theta}{\lambda}\right)}{N^2\sin^2\left(\frac{\pi\,b\sin\Theta}{N\lambda}\right)}$$

To calculate the limit $N \to \infty$, we use the Taylor expansion of the sine function:

$$N \sin\left(\frac{\xi}{N}\right) = N\left(\frac{\xi}{N} - \frac{1}{3!}\frac{\xi^3}{N^3} + \dots\right) = \xi - \frac{1}{3!}\frac{\xi^3}{N^2} + \dots \longrightarrow \xi \quad \text{for } N \to \infty$$

Thus, in the limit $N \to \infty$ we get

$$I(\Theta) = I(0)\,\frac{\sin^2\left(\frac{\pi\, b \sin\Theta}{\lambda}\right)}{\left(\frac{\pi\, b \sin\Theta}{\lambda}\right)^2}\,.$$

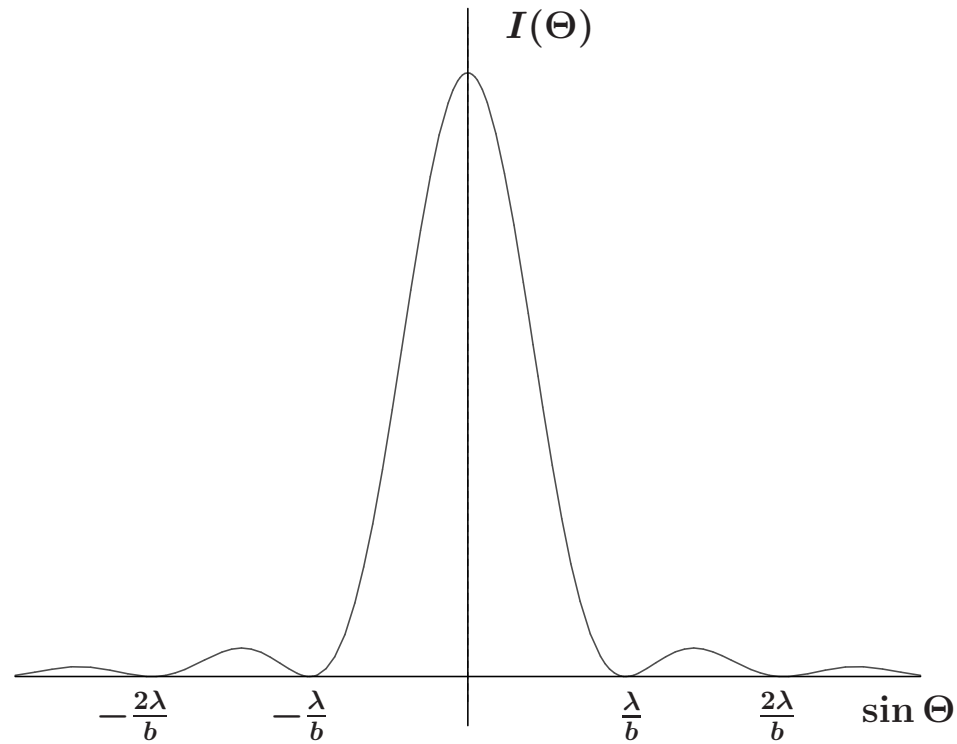
This is sometimes written as

$$I(\Theta) = I(0)\,\mathrm{sinc}^2\left(\frac{\pi\, b \sin\Theta}{\lambda}\right)$$

with the "sinc-function" $\quad \mathrm{sinc}(\xi) = \dfrac{\sin\xi}{\xi}\,.$

Minima occur at

$$\boxed{\sin\Theta_m = \frac{m\,\lambda}{b}} \qquad (m = \pm 1, \pm 2, \dots)$$

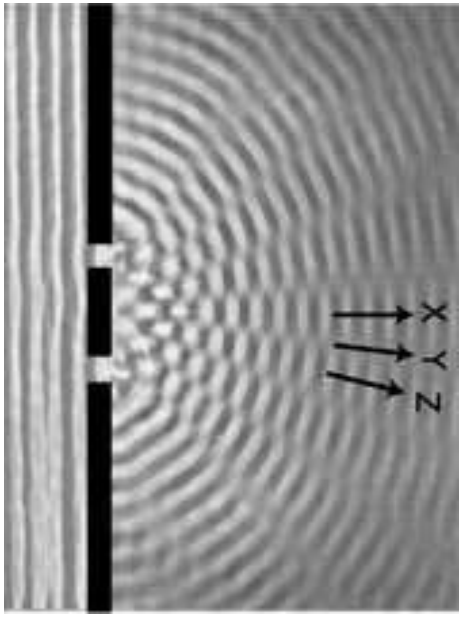


picture from nebula.deanza.fhda.edu

δ) Double slit

Thomas Young
(1773-1829)

picture from http://en.wikipedia.org



Diffraction pattern produced by a double slit with water waves

picture from http://www.lightandmatter.com/

With light, the double-slit experiment was first described by Thomas Young in 1801. It was the "experimentum crucis" which decided in favour of the wave theory of light.

The double-slit experiment with matter waves (electrons) played an important role for the understanding of quantum mechanics.

Consider two slits of finite width b and separation a.

Proceed as for single slit (p.101).

Treat each of the two slits as the limit of a grating (N narrow slits with separation g).

Keep $b = Ng$ fixed, send $N \to \infty$ (thus $g \to 0$).

For finite N, wave from first slit is (p.96)

$$u_1(\underline{r}, t) =$$

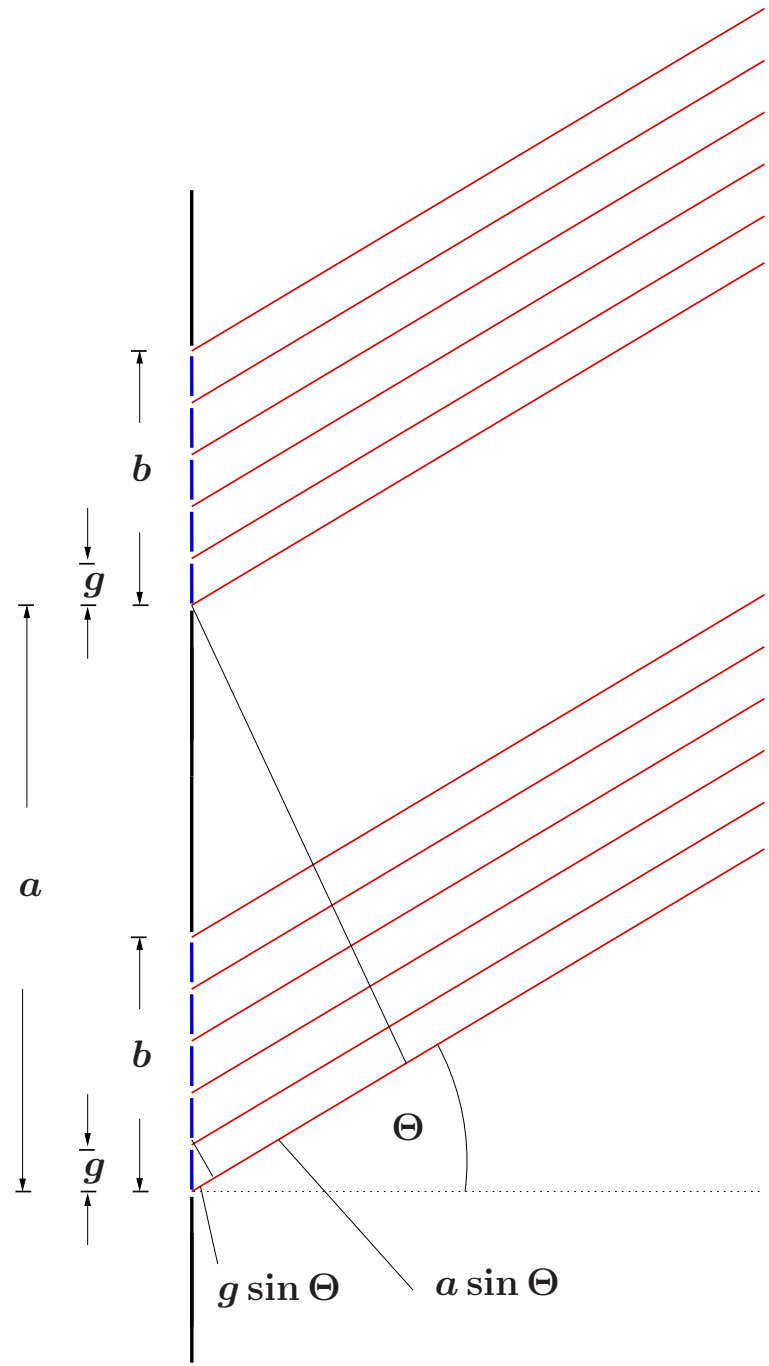
$$\mathrm{Re}\left\{ A\, e^{i\left(\underline{k}\cdot\underline{r} - \omega t + \frac{(N-1)\Delta}{2}\right)} \frac{\sin(N\Delta/2)}{\sin(\Delta/2)} \right\}$$

where $\Delta = k\, g \sin\Theta = \frac{k\, b}{N} \sin\Theta$.

Wave from second slit has additional phase shift $k\, a \sin\Theta$:

$$u_2(\underline{r}, t) =$$

$$\mathrm{Re}\left\{ A\, e^{i\left(\underline{k}\cdot\underline{r} - \omega t + \frac{(N-1)\Delta}{2}\right)} \frac{\sin(N\Delta/2)}{\sin(\Delta/2)}\, e^{i\, k\, a \sin\Theta} \right\}$$



Adding up:

$$u(\underline{r},t) \;=\; u_1(\underline{r},t) \;+\; u_2(\underline{r},t) \;=$$

$$\mathrm{Re}\left\{ A\, e^{i\left(\underline{k}\cdot\underline{r}\;-\omega t+\frac{(N-1)\Delta}{2}\right)}\;\frac{\sin(N\Delta/2)}{\sin(\Delta/2)}\left(1\;+\;e^{i\,k\,a\sin\Theta}\right)\right\} \;=$$

$$\mathrm{Re}\left\{ A\, e^{i\left(\underline{k}\cdot\underline{r}\;-\omega t+\frac{(N-1)\Delta}{2}\right)}\;\frac{\sin(N\Delta/2)}{\sin(\Delta/2)}\,e^{\frac{i}{2}\,k\,a\sin\Theta}\underbrace{\left(e^{-\frac{i}{2}\,k\,a\sin\Theta}\;+\;e^{\frac{i}{2}\,k\,a\sin\Theta}\right)}_{2\cos\left(\frac{k}{2}\,a\sin\Theta\right)}\right\} \;=$$

$$2\,A\cos\left(\underline{k}\cdot\underline{r}\;-\;\omega t\;+\;\frac{(N-1)\Delta}{2}\;+\;\frac{k}{2}\,a\sin\Theta\right)\,\frac{\sin(N\Delta/2)}{\sin(\Delta/2)}\cos\left(\frac{k}{2}\,a\sin\Theta\right).$$

Intensity:

$$I(\Theta)\;=\;\frac{2\pi}{\omega}\int_0^{\frac{2\pi}{\omega}} u(\underline{r},t)^2 dt \;=$$

$$4\,A^2\,\frac{\sin^2(N\Delta/2)}{\sin^2(\Delta/2)}\cos^2\left(\frac{k}{2}\,a\sin\Theta\right)\underbrace{\frac{2\pi}{\omega}\int_0^{\frac{2\pi}{\omega}}\cos^2\left(\underline{k}\cdot\underline{r}\;-\;\omega t\;+\;\frac{(N-1)\Delta}{2}\;+\;\frac{k}{2}\,a\sin\Theta\right)dt}_{1/2}\,.$$

With $\Delta = \frac{k\,b}{N}\sin\Theta$: $\quad I(\Theta) = 2\,A^2\,\dfrac{\sin^2\left(\frac{k\,b}{2}\sin\Theta\right)}{\sin^2\left(\frac{k\,b}{2\,N}\sin\Theta\right)}\cos^2\left(\dfrac{k}{2}\,a\sin\Theta\right).$

By Taylor expansion of the sine function,

$$\frac{\sin\left(\frac{k\,b}{2}\sin\Theta\right)}{\sin\left(\frac{k\,b}{2\,N}\sin\Theta\right)} = \frac{\frac{k\,b}{2}\cancel{\sin\Theta} - \frac{1}{3!}\left(\frac{k\,b}{2}\right)^3\sin^{\cancel{3}2}\Theta + \ldots}{\frac{k\,b}{2\,N}\cancel{\sin\Theta} - \frac{1}{3!}\left(\frac{k\,b}{2\,N}\right)^3\sin^{\cancel{3}2}\Theta + \ldots} \to N \quad \text{for} \quad \Theta \to 0\,,$$

we find $I(0) = 2\,A^2\,N^2$, hence

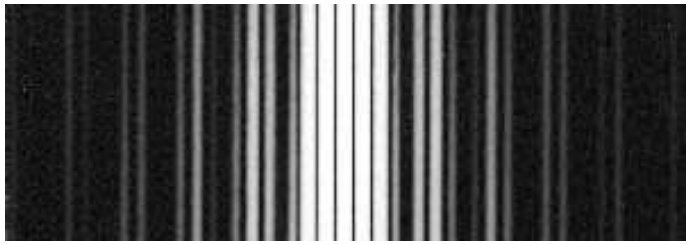
$$I(\Theta) = I(0)\,\frac{\sin^2\left(\frac{k\,b}{2}\sin\Theta\right)}{N^2\sin^2\left(\frac{k\,b}{2\,N}\sin\Theta\right)}\cos^2\left(\frac{k}{2}\,a\sin\Theta\right) =$$

$$= I(0)\left(\frac{\sin\left(\frac{k\,b}{2}\sin\Theta\right)}{\cancel{N}\left(\frac{k\,b}{2\,\cancel{N}}\sin\Theta - \frac{1}{3!}\left(\frac{k\,b}{2}\right)^3\frac{1}{N^{\cancel{3}2}}\sin^3\Theta + \ldots\right)}\right)^2\cos^2\left(\frac{k}{2}\,a\sin\Theta\right).$$

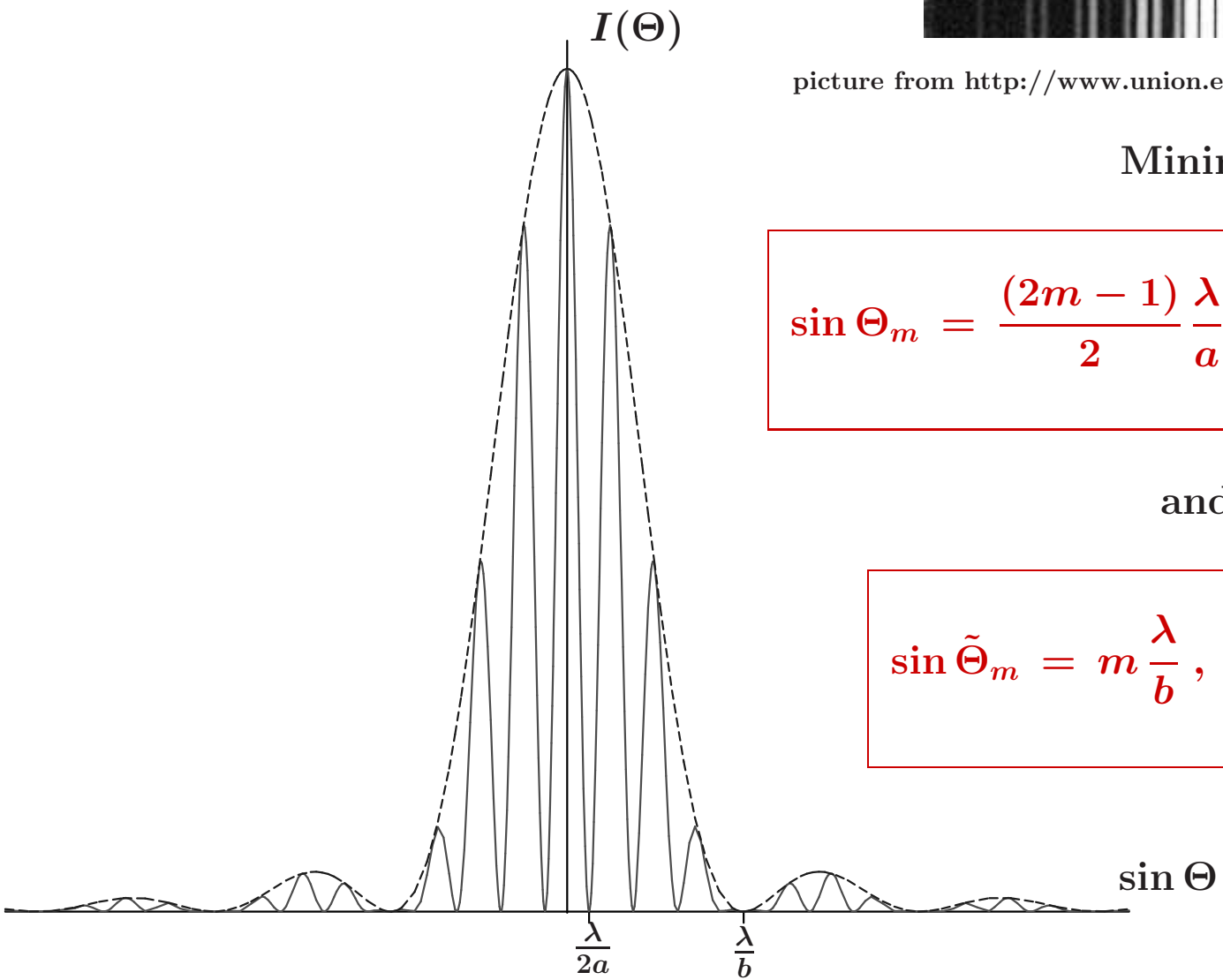
Limit $N \to \infty$ gives the double-slit diffraction pattern. With $k = 2\,\pi/\lambda$:

$$I(\Theta) = I(0)\underbrace{\left(\frac{\lambda}{\pi\,b\sin\Theta}\sin\frac{\pi\,b\sin\Theta}{\lambda}\right)^2}_{=\,\mathrm{sinc}^2\left(\frac{\pi\,b\sin\Theta}{\lambda}\right)}\left(\cos\frac{\pi\,a\sin\Theta}{\lambda}\right)^2$$

Note: $I(\Theta)$ is the intensity function for two narrow slits with separation a, modulated with the intensity function for a single slit with width b.



picture from http://www.union.edu/PUBLIC/PHYDEPT/jonesc/



Minima at

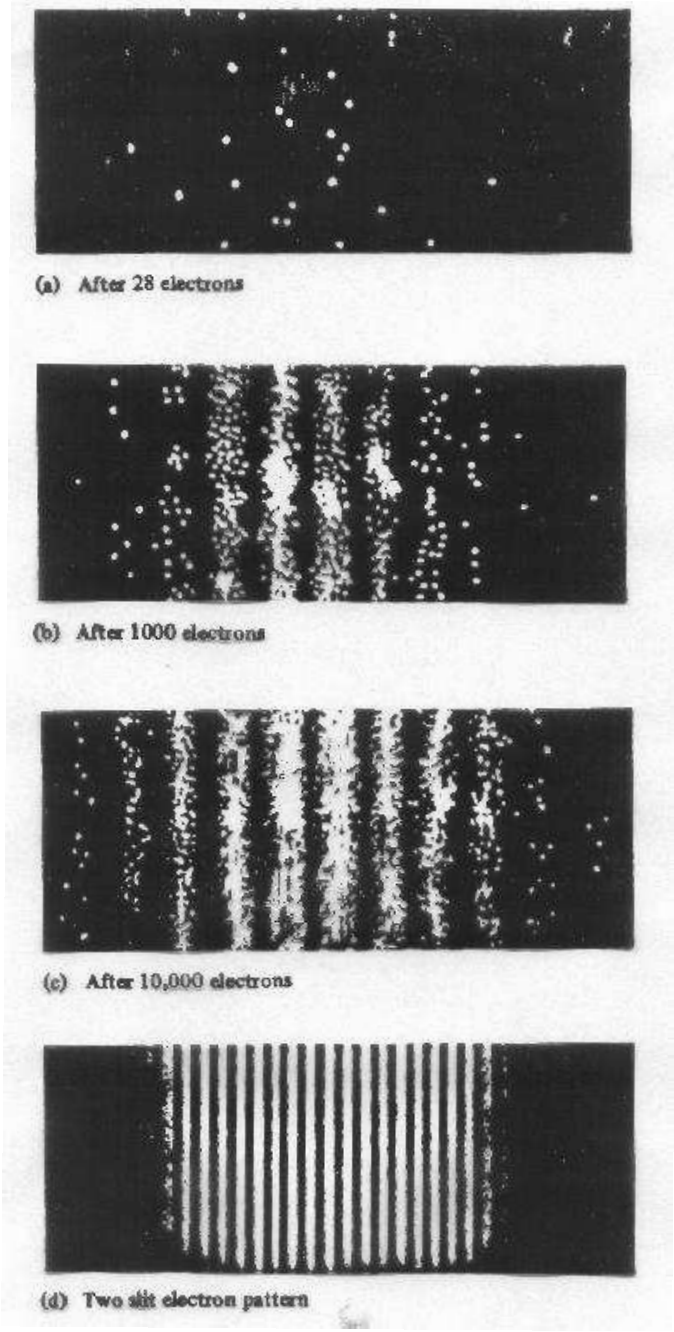
$$\sin\Theta_m = \frac{(2m-1)}{2}\frac{\lambda}{a}, \quad m = 0, \pm 1, \pm 2, \ldots$$

and at

$$\sin\tilde{\Theta}_m = m\frac{\lambda}{b}, \quad m = \pm 1, \pm 2, \ldots$$

Double-slit experiment with electrons:

picture from http://www.physics.utoledo.edu/ ljc



(a) After 28 electrons

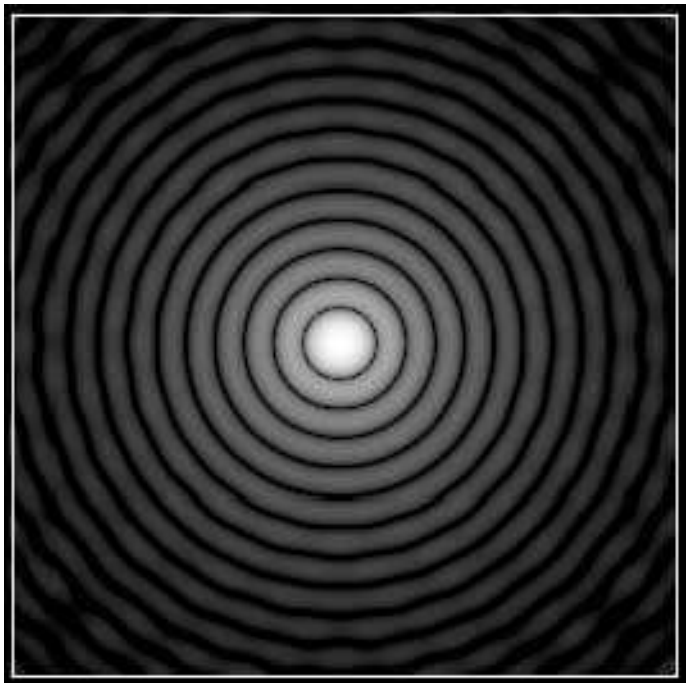
(b) After 1000 electrons

(c) After 10,000 electrons

(d) Two slit electron pattern

ε) Circular aperture

The diffraction pattern of a circular aperture was investigated by the British astronomer George B. Airy (1801–1892) and is named after him.



picture from

http://www.astrotelescope.com/optique

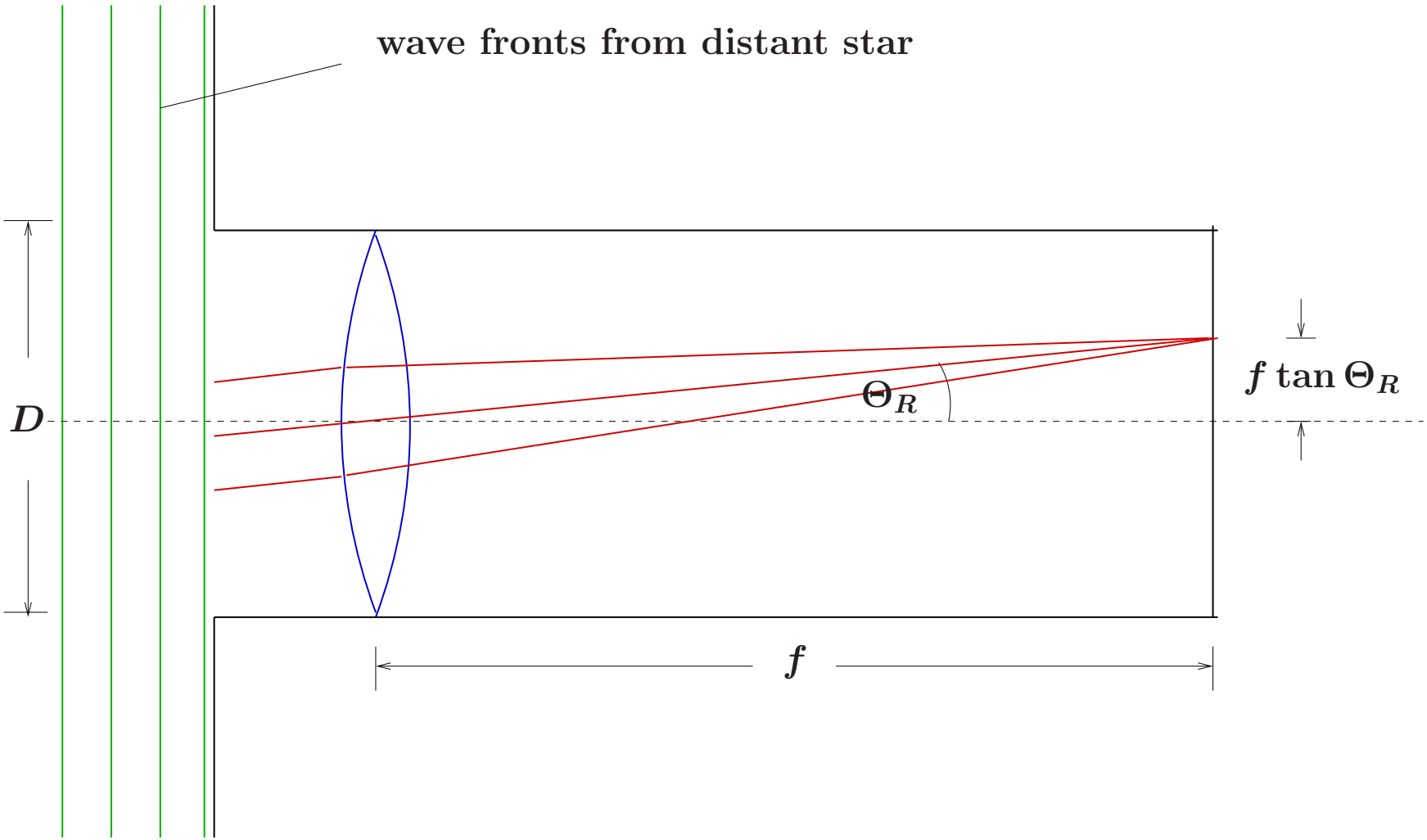
Fraunhofer diffraction pattern produced by a a circular aperture of diameter D:

$$I(\Theta) = I(0) \left(\frac{2\,\lambda}{\pi\, D \sin\Theta} J_1\!\left(\frac{\pi}{\lambda} D \sin\Theta \right) \right)^2$$

where J_1 denotes the first Bessel function of the first kind.

The angular radius Θ_R of the central bright disk ("Airy disk") can be found from tabulated values of J_1.

$$\sin\Theta_R \approx 1.22\, \frac{\lambda}{D}$$



Because of diffraction, even an optically perfect telescope does not produce a point image of a point source (distant star).

Image in the focal plane has radius $f \tan \Theta_R \approx 1.22\, f\, \lambda / D$ (with approximation $\tan \Theta_R \approx \sin \Theta_R$ for small angles).

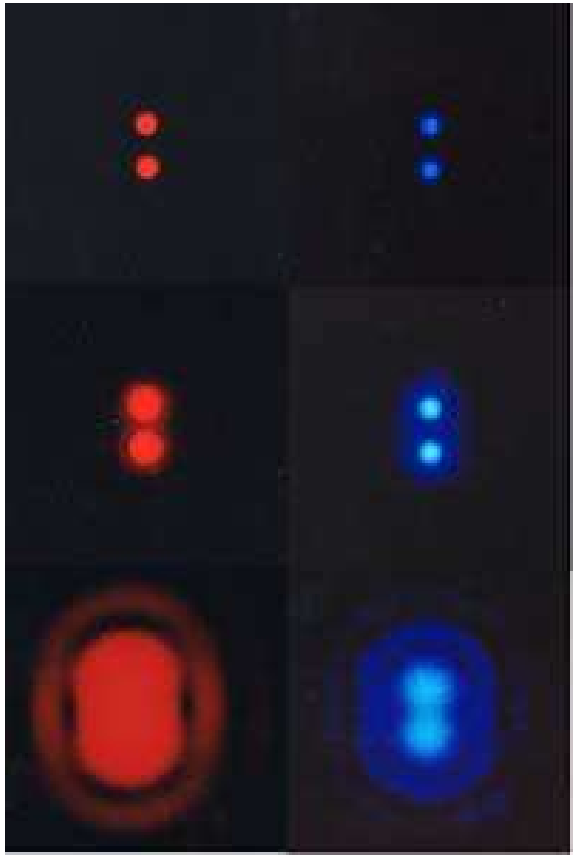
Rayleigh criterion for resolution of two point sources:

Angular separation α of the two sources should be larger than the angular radius of their Airy disks,

$$\boxed{\sin\alpha \;>\; 1.22\,\frac{\lambda}{D}}$$

where D is the diameter of the aperture of the optical instrument (e.g. telescope).

- Bigger aperture gives higher resolving power.
- Smaller wave length gives higher resolving power.
- A radio telescope must have a much bigger aperture than an optical telescope to give the same resolving power.
- In spite of the last observation, today radio telescopes have higher resolving power than optical telescopes because they can be combined by interferometric methods (see below).



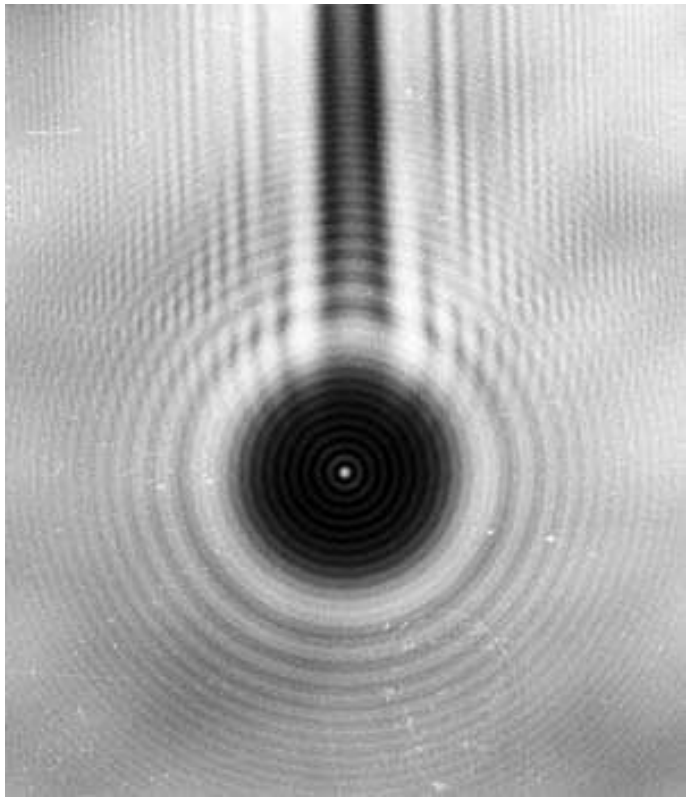
Top to bottom: Decreasing D.

Left to right: Decreasing λ.

picture from

http://www.union.edu/PUBLIC/PHYDEPT/jonesc

α) Opaque circular disk



picture from

http://www.union.edu/PUBLIC/PHYDEPT/jonesc

With Fraunhofer diffraction, wavelets originating in the aperture plane show no phase difference in the forward direction ($\Theta = 0$), so there is always brightness at the center of the screen.

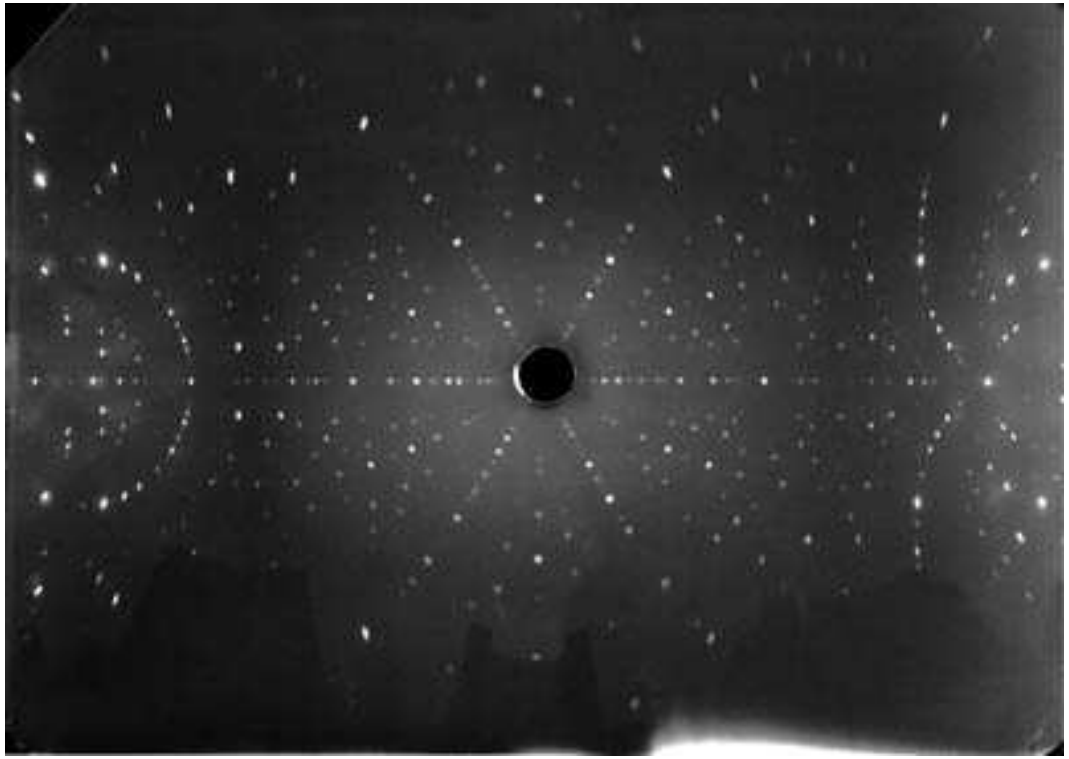
In particular, the diffraction pattern of an opaque circular disk should show a bright spot in the center. Denis Poisson considered this to be absurd and therefore rejected Fresnel's diffraction theory in 1818. The spot was observed by François Arago shortly thereafter (and had actually already been observed by Giacomo Maraldi in 1723). It is now called "Poisson's spot" or "Arago's spot".

Thus, by diffraction, an opaque circular disk has a focusing effect on light.

η) Note on (Fraunhofer) diffraction by three-dimensional structures

The above techniques apply to planar apertures only. For three-dimensional obstacles, one needs other techniques, even if light source and observation screen are very far from the obstacle.

An important example is the diffraction of X rays by a three-dimensional crystal. This was first observed by Max von Laue in 1912 (with his collaborators Friedrich and Knipping). For this discovery he won the Nobel Prize in 1914.



picture from www.union.edu/PUBLIC/PHYDEPT/jonesc

Laue diagram of an Si crystal

c) Fresnel diffraction

Recall:

Fraunhofer diffraction:

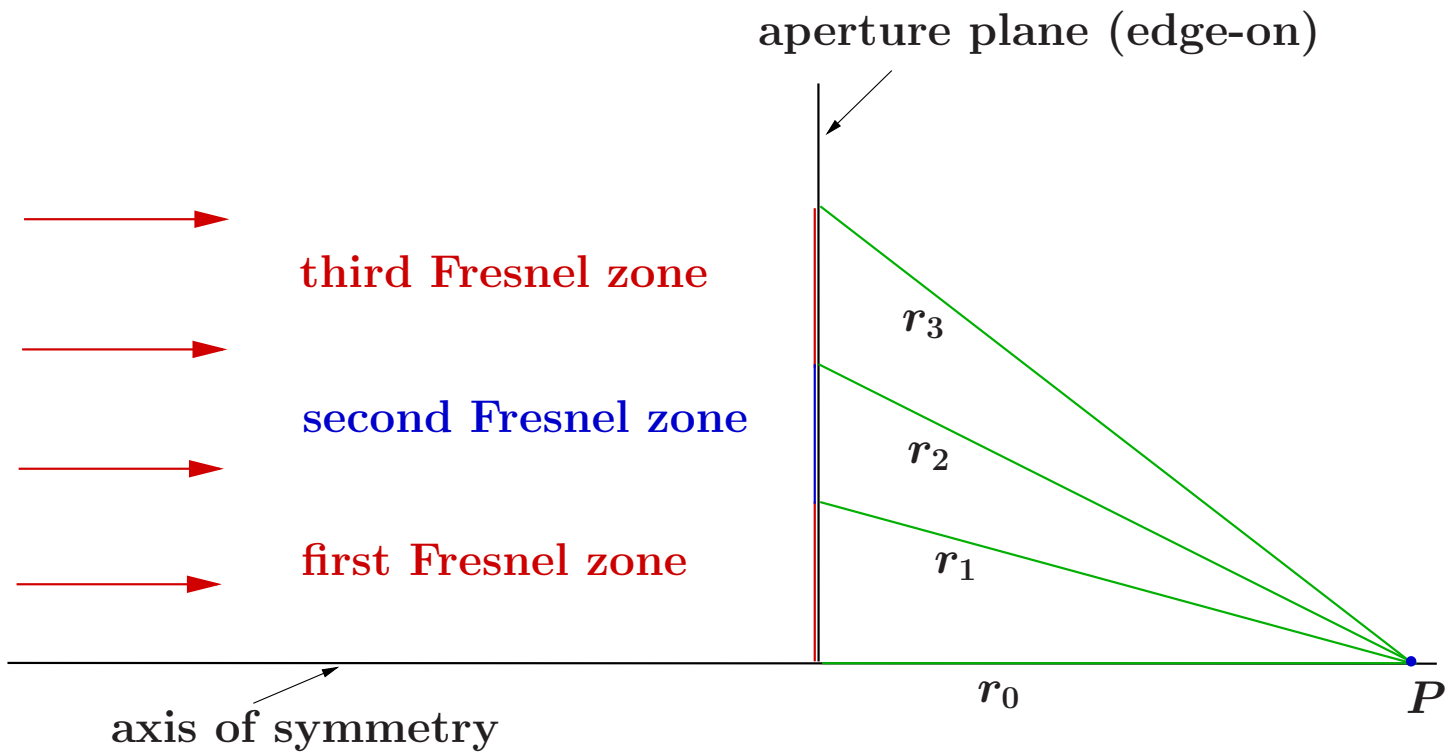
- **Light source so far away from aperture that incoming wavefront can be considered as plane.**
- **Observation screen so far away from aperture that the fronts of the wavelets, originating from the points of the aperture, can be considered as plane when they arrive at the screen.**

Fresnel diffraction: Otherwise

Fresnel diffraction is much more difficult to calculate than Fraunhofer diffraction. We will do only one example:

Consider light source so far away from aperture plane that incoming wavefront can be considered as plane, but observation screen at finite distance. Calculate intensity on the axis of symmetry for a circular aperture.

Divide aperture into Fresnel zones, defined by $r_N = r_0 + \frac{N\lambda}{2}$



Phase difference at P of wavelets from subsequent Fresnel zones is π.

Construction of Fresnel zones depends on r_0 and λ, i.e., it is made for a particular point P and a particular wave length.

Now subdvide each Fresnel zone into K concentric rings, with constant phase difference π/K between subsequent rings.

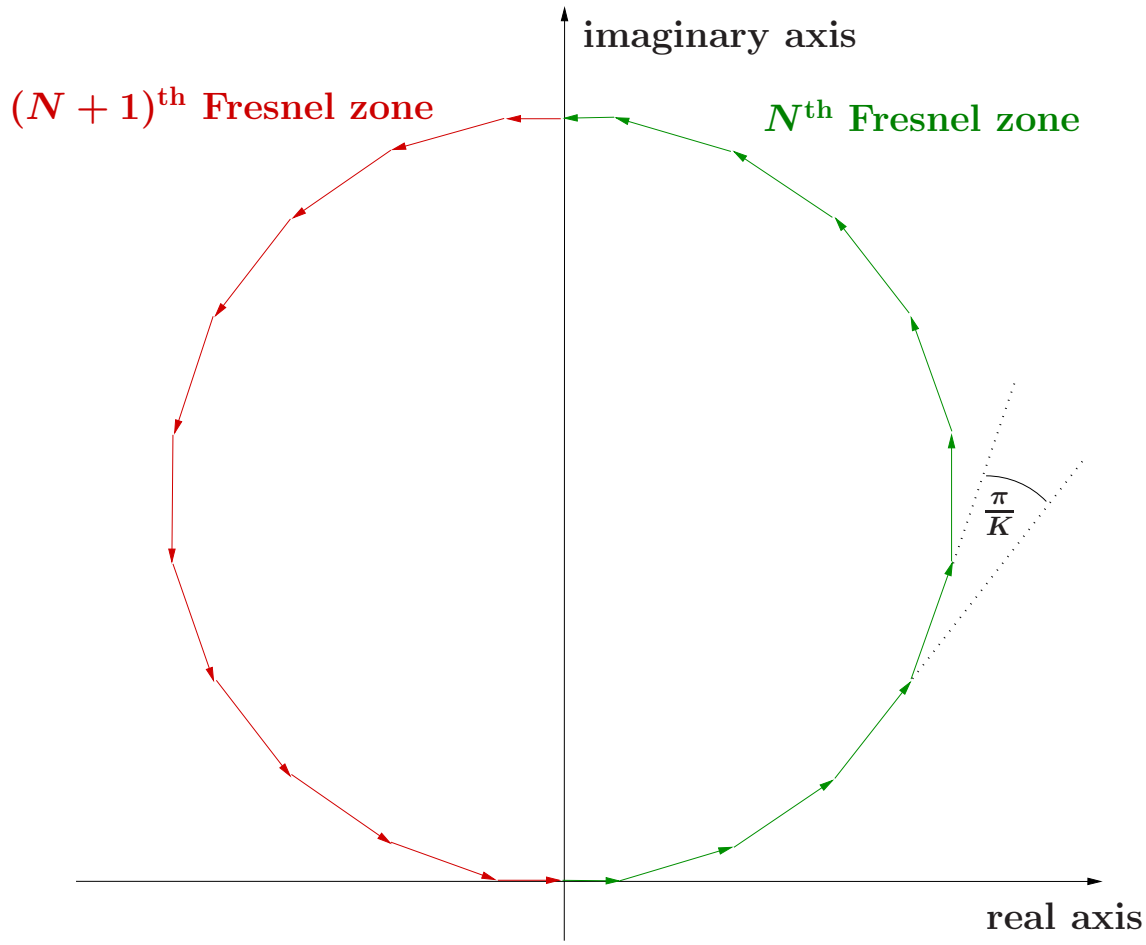
Consider one phasor for all the light coming from one ring zone, i.e., K phasors for each Fresnel zone.

(Picture for $K = 10$; actually, we divide each Fresnel zone into 9 full-size and 2 half-size sections.)

For $K \to \infty$ we get two half-circles that close up.

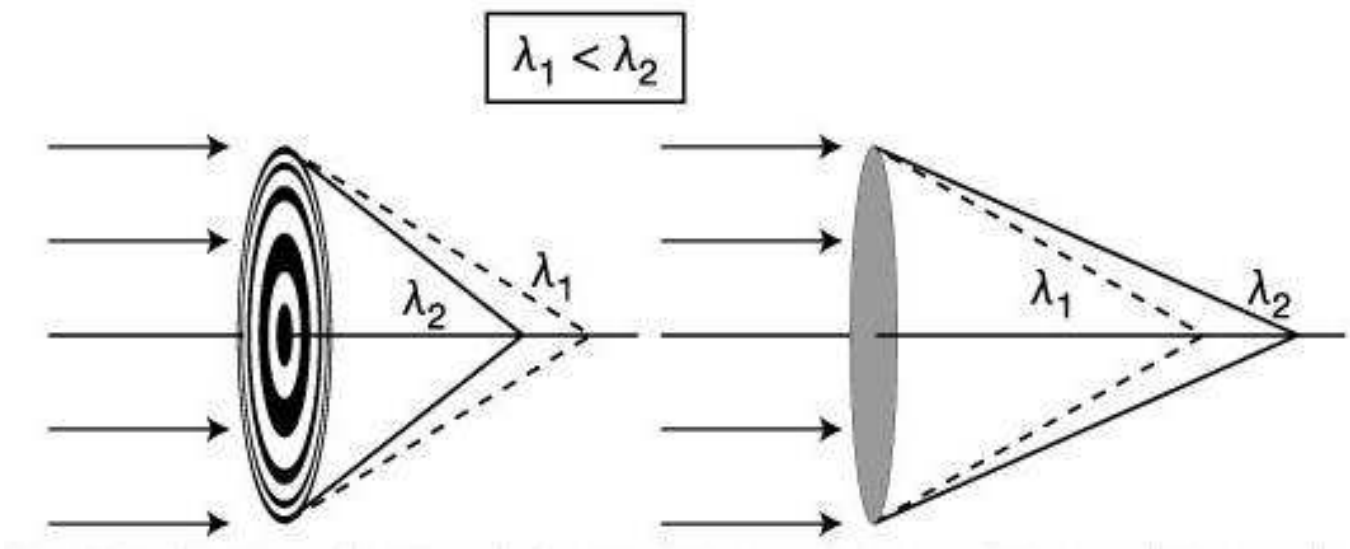
Phasors from N^{th} and $(N+1)^{\text{th}}$ Fresnel zones add up to zero!

Any two subsequent Fresnel zones give waves that cancel each other.



- If circular aperture contains even number of Fresnel zones: Darkness at P.
- If circular aperture contains odd number of Fresnel zones: Same intensity at P as from first Fresnel zone alone.

Fresnel zone plate:
Make every other Fresnel zone opaque, get maximal intensity at P, i.e., effect similar to converging lens.



picture from http://www.nature.com

- **A Fresnel zone plate produces an image similar to a lens:**

picture from http://www.alternativephotography.com

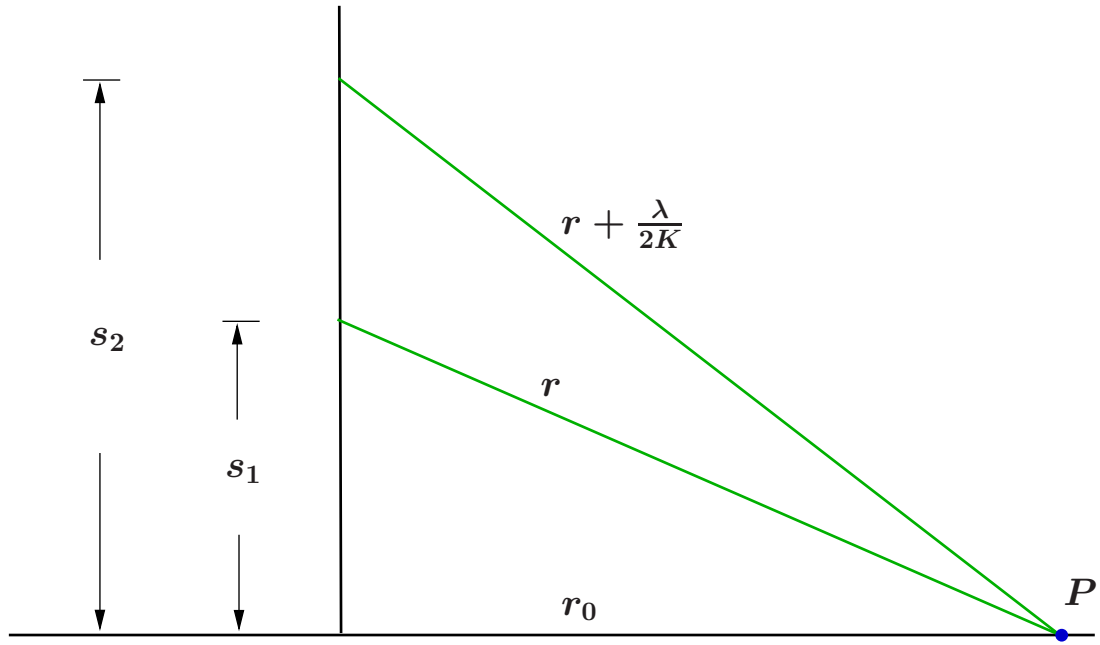
- **Fresnel zone plates can be easily produced for wave lengths outside of the visible spectrum.**
- **A space telescope using a Fresnel zone plate is in the planning stage ('Fresnel imager').**

We have assumed that the intensity from all ring zones is the same (phasors of equal length).

This is true as long as the obliquity factor can be approximated by 1, i.e., Θ so small that

$$\frac{1}{2}\left(1 + \cos\Theta\right) = 1 \,.$$

Then the amplitude of each wavelet falls off equally in all spatial directions; the total amplitude from the ring zone at P is



$\sim \mathcal{A}$ = area of ring zone

$\sim 1/r$ = inverse distance from P to ring zone.

$$\frac{\mathcal{A}}{r} = \frac{\pi\, s_2^2 - \pi\, s_1^2}{r} = \frac{\pi}{r}\left[\left(\left(r + \frac{\lambda}{2K}\right)^2 - r_0^2\right) - \left(r^2 - r_0^2\right)\right] =$$

$$= \frac{\pi}{r}\left(r^2 + \frac{2r\lambda}{2K} + \frac{\lambda^2}{4K^2} - r_0^2 - r^2 + r_0^2\right) = \frac{\pi\,\lambda}{K}\left(1 + \frac{\lambda}{4Kr}\right) \approx \frac{\pi\,\lambda}{K}$$

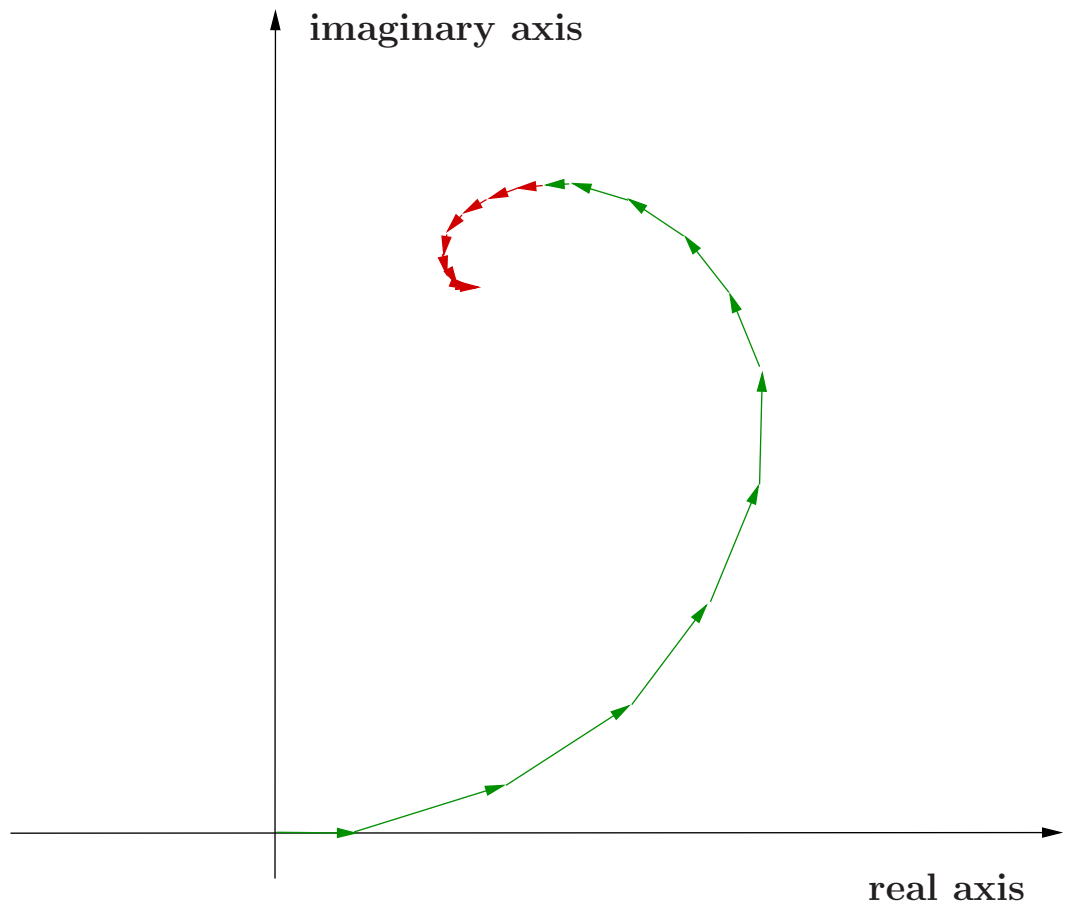
is, indeed, independent of r. (We used that $\lambda \ll r$ for all practical purposes.)

Taking obliquity factor into account, the intensity decreases from ring zone to ring zone if we move outwards (i.e., phasors become shorter and shorter).

For $K \to \infty$, we get a spiral instead of half-circles that close up.

Thus, exact cancellation of waves from subsequent Fresnel zones holds true only as long as Θ is so small that the obliquity factor can be approximated by 1.

Quantitative criterion for distinction between Fraunhofer and Fresnel diffraction:



$$\boxed{\text{Fresnel number}: \quad F = \frac{a^2}{r_0\,\lambda}}$$

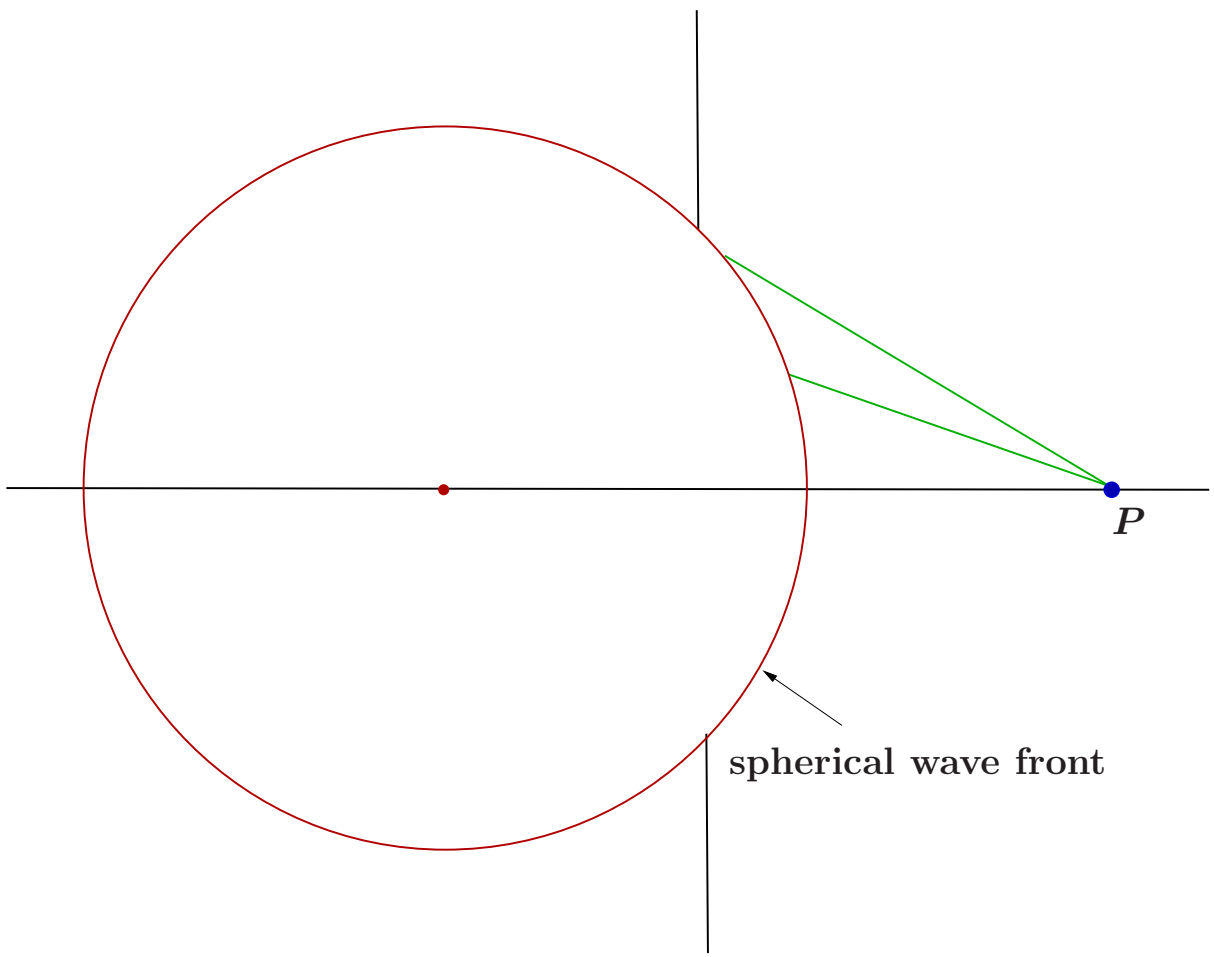
r_0 = distance of observation point from aperture
a = radius of aperture
λ = wave length

F counts how many Fresnel zones are in the aperture.

Fraunhofer diffraction is applicable if $F \ll 1$.

Construction of Fresnel zones was given here for parallel incoming light (i.e., light source at infinity).

For light source at finite distance, construction must be modified. Then Huygens wavelets originate from a spherical wave front, not from a plane.



IV. Interference

(cf. Pedrotti and Pedrotti, Chapters 10, 11 and 12)

a) Interference and coherence

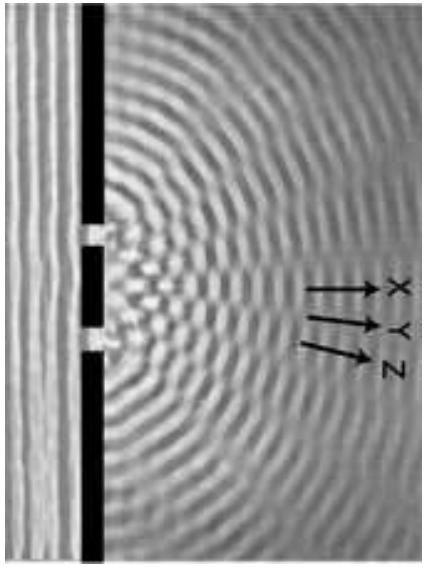
"Interference" means superposition of two or more waves which

- enhance each other in some regions ("constructive interference");
- cancel each other, exactly or approximately, in other regions ("destructive interference").

This results in characteristic "interference patterns".

Diffraction is a special interference phenomenon: According to the Huygens-Fresnel principle, diffraction can be explained in terms of interference of wavelets.

Interference patterns can be observed only if the superposed waves are "coherent", i.e., if their phases are synchronised.



picture from http://www.lightandmatter.com/

How to achieve coherence?

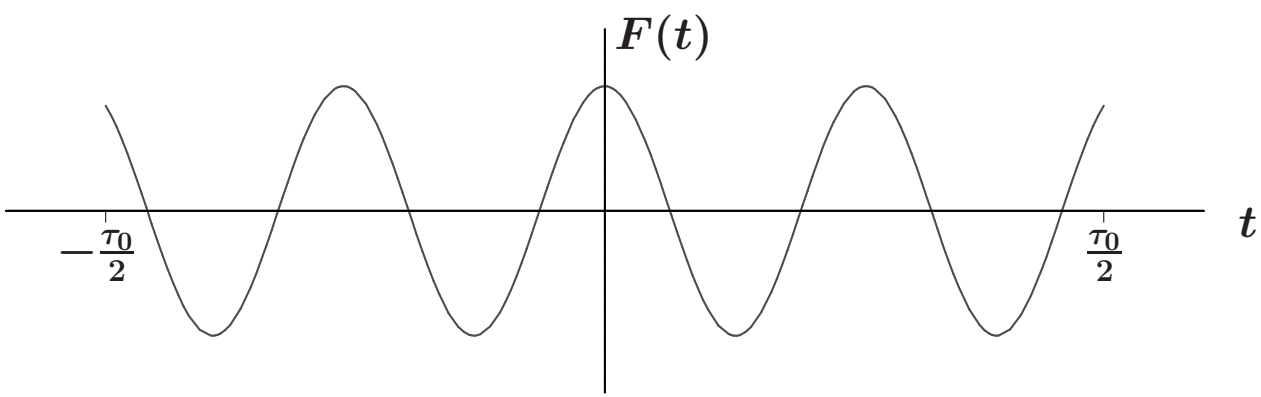
- water waves: wave machine with two rods on the same axle
- sound waves: two loudspeakers fed from the same generator
- radio waves: two antennas fed from the same emitter
- light: see examples in this section

Light emitted from two points of an ordinary light source (e.g., the Sun or a light bulb) is not coherent; it consists of short "wave trains" which are not synchronised.

As quantitative measures for coherence one uses the notions of "coherence time", "coherence length" and "band width":

If we idealise, for simplicity, a wave train as strictly harmonic over its lifetime, it is given at a fixed point in space by a function of the form

$$F(t) = \begin{cases} A\cos\omega_0 t & \text{if} \quad \frac{-\tau_0}{2} < t < \frac{\tau_0}{2} \\ 0 & \text{otherwise} \end{cases}$$



The average duration τ_0 is called the coherence time of the light source.

What is the Fourier transform $G(\omega)$ (i.e., the frequency spectrum) of such a wave train?

Recall:
$$F(t) = \int_{-\infty}^{\infty} G(\omega)\, e^{i\omega t}\, d\omega$$

where
$$G(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(t)\, e^{-i\omega t}\, dt = \frac{1}{2\pi} \int_{-\tau_0/2}^{\tau_0/2} A\cos(\omega_0 t)\, e^{-i\omega t}\, dt =$$

$$= \frac{A}{2\pi} \int_{-\tau_0/2}^{\tau_0/2} \frac{1}{2} \left(e^{i\omega_0 t} + e^{-i\omega_0 t} \right) e^{-i\omega t}\, dt = \frac{A}{4\pi} \int_{-\tau_0/2}^{\tau_0/2} \left(e^{-i(\omega-\omega_0)t} + e^{-i(\omega+\omega_0)t} \right) dt =$$

$$= \frac{A}{4\pi} \left[\frac{e^{-i(\omega-\omega_0)t}}{-i(\omega-\omega_0)} + \frac{e^{-i(\omega+\omega_0)t}}{-i(\omega+\omega_0)} \right]_{-\tau_0/2}^{\tau_0/2} =$$

$$= \frac{A}{2\pi} \left(\frac{\sin\left((\omega-\omega_0)\tau_0/2\right)}{\omega-\omega_0} + \frac{\sin\left((\omega+\omega_0)\tau_0/2\right)}{\omega+\omega_0} \right) =$$

$$= \frac{A\tau_0}{4\pi} \left(\frac{\sin\left((\omega-\omega_0)\tau_0/2\right)}{(\omega-\omega_0)\tau_0/2} + \frac{\sin\left((\omega+\omega_0)\tau_0/2\right)}{(\omega+\omega_0)\tau_0/2} \right)$$

If the wave train contains many periods, $\omega_0\tau_0 \gg 2\pi$, the second term is small near $\omega = \omega_0$. We can then approximate $G(\omega)$ near ω_0 by

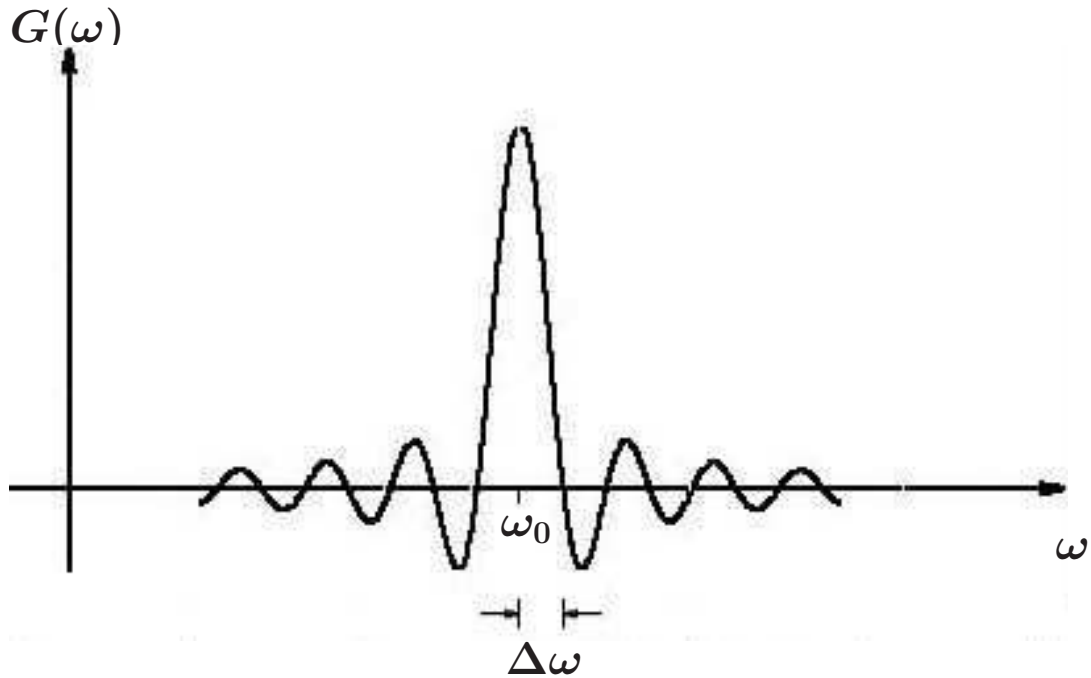
$$G(\omega) \approx \frac{A\tau_0}{4\pi} \frac{\sin\left((\omega-\omega_0)\tau_0/2\right)}{(\omega-\omega_0)\tau_0/2}.$$

Recall: We already encountered the (square of the) function $\mathrm{sinc}(x) = \frac{1}{x}\sin x$ for the single-slit diffraction pattern.

The "band width"

$$\boxed{\Delta\omega = \frac{2\pi}{\tau_0}}$$

(see picture) is a measure for the frequency interval that contributes essentially to the wave train.



The shorter the wave train, the broader the frequency distribution:

$$\tau_0 \to 0 : \quad \Delta\omega \to \infty$$

The longer the wave train, the narrower the frequency distribution:

$$\tau_0 \to \infty : \quad \Delta\omega \to 0$$

To express the band width in terms of wave length, rather than frequency, introduce

coherence length ℓ_0 = length traveled by light in coherence time τ_0.

In medium with constant velocity of light $v = c/n$: $\ell_0 = v\tau_0 = \dfrac{2\pi v}{\Delta\omega}$

light bulb: $\ell_0 \approx$ a few μm

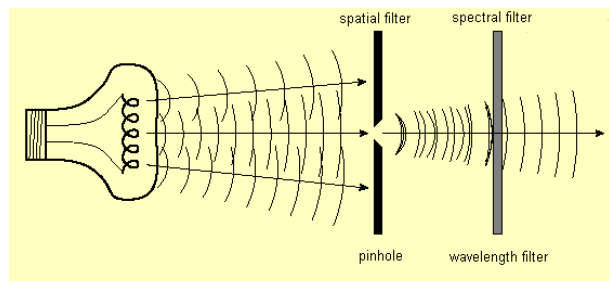
laser: ℓ_0 from a few cm up to more than 100 km

Band width in terms of wave length:

$$\omega = \frac{2\pi v}{\lambda} \quad \Rightarrow \quad \Delta\omega \approx \frac{2\pi v}{\lambda^2}\Delta\lambda \quad \Rightarrow \quad \boxed{\Delta\lambda \approx \frac{\lambda^2}{\ell_0}}$$

Because of higher coherence, interference experiments are easier to carry through with laser light than with light from a light bulb or from a gas discharge lamp.

Light from a light bulb can be made more coherent by sending it through a small aperture and through a filter. (The aperture singles out wave trains from a smaller spatial region, the filter reduces the band width $\Delta\omega$.)



picture from http://electron9.phys.utk.edu

b) Two-beam interference in complex notation

Superposition of two harmonic waves with the same frequency at a fixed point in space:

$$
\begin{aligned}
u_1(t) &= A_1 \cos(\omega t + \delta_1) \\
u_2(t) &= A_2 \cos(\omega t + \delta_2) \\
u(t) &= u_1(t) + u_2(t)
\end{aligned}
$$

Intensities:

$$
\begin{aligned}
I_1 &= \frac{\omega}{2\,\pi} \int_0^{2\pi/\omega} u_1(t)^2\, dt = A_1^2/2 \\
I_2 &= \frac{\omega}{2\,\pi} \int_0^{2\pi/\omega} u_2(t)^2\, dt = A_2^2/2 \\
I &= \frac{\omega}{2\,\pi} \int_0^{2\pi/\omega} u(t)^2\, dt = \ ?
\end{aligned}
$$

Here we have used that

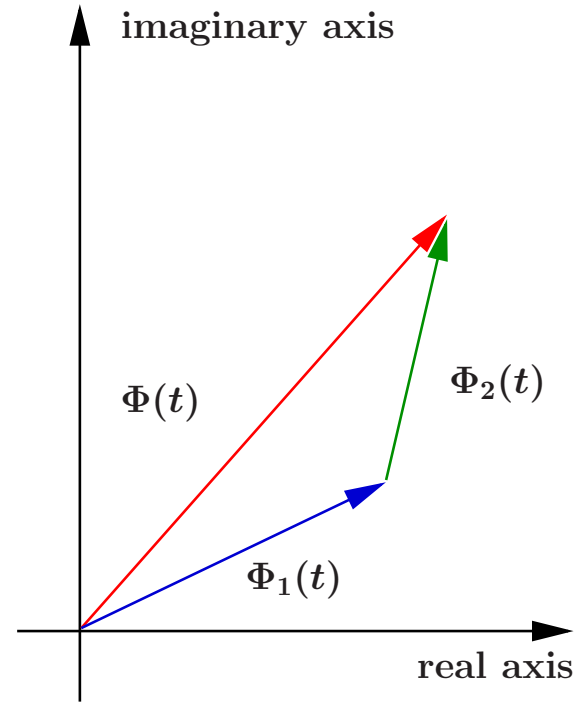
$$
\frac{1}{2\pi} \int_0^{2\pi} \cos^2\xi\, d\xi = \frac{1}{2}
$$

Introduce complex notation:

$$\Phi_1(t) \;=\; A_1\, e^{i(\omega t+\delta_1)}\,, \qquad u_1(t) \;=\; \mathrm{Re}\big(\Phi_1(t)\big)$$

$$\Phi_2(t) \;=\; A_2\, e^{i(\omega t+\delta_2)}\,, \qquad u_2(t) \;=\; \mathrm{Re}\big(\Phi_2(t)\big)$$

$$\Phi(t) \;=\; \Phi_1(t) \;+\; \Phi_2(t)\,, \qquad u(t) \;=\; \mathrm{Re}\big(\Phi(t)\big)$$



$$\Phi(t) \;=\; A_1\, e^{i(\omega t+\delta_1)} \;+\; A_2\, e^{i(\omega t+\delta_2)} \;=\; Z\, e^{i\omega t}$$

where $Z \;:=\; A_1\, e^{i\delta_1} \;+\; A_2\, e^{i\delta_2}$ is independent of t.

Phasor diagram rotates rigidly with angular frequency ω.

Write Z in polar coordinates:

$$Z \;=\; x \;+\; i\,y \;=\; r\, e^{i\varphi}\,, \qquad Z^* \;=\; x \;-\; i\,y \;=\; r\, e^{-i\varphi}$$

$$r^2 = Z^* Z = \big(\, A_1\, e^{-i\delta_1} \;+\; A_2\, e^{-i\delta_2}\,\big)\,\big(\, A_1\, e^{i\delta_1} \;+\; A_2\, e^{i\delta_2}\,\big) = A_1^2 \;+\; A_2^2 \;+\; 2\, A_1\, A_2 \cos(\delta_2 - \delta_1)$$

$$u(t) \;=\; \mathrm{Re}\big(\Phi(t)\big) \;=\; \mathrm{Re}\big(r e^{i(\omega t+\varphi)}\big) \;=\; r\cos(\omega t + \varphi)$$

$$I \;=\; \frac{\omega}{2\,\pi}\int_0^{2\pi/\omega} u(t)^2\, dt \;=\; r^2/2 \;=\; \big(\, A_1^2 \;+\; A_2^2 \;+\; 2\, A_1\, A_2 \cos(\delta_2 - \delta_1)\,\big)/2$$

$$\boxed{I \;=\; I_1 \;+\; I_2 \;+\; 2\sqrt{I_1 I_2}\cos(\delta_2 - \delta_1)}$$

- If $\delta_2 - \delta_1$ varies randomly over observation time (incoherent beams), cosine averages to zero, $I = I_1 + I_2$.
- If $\delta_2 - \delta_1$ is constant over observation time (coherent beams):

  Interference maxima if $\cos(\delta_2 - \delta_1) = 1$

$$\delta_2 - \delta_1 = 2m\pi \quad \text{for integer } m$$

$$I \;=\; (\sqrt{I_1} + \sqrt{I_2}\,)^2$$

  Interference minima if $\cos(\delta_2 - \delta_1) = -1$

$$\delta_2 - \delta_1 = (2m+1)\pi \quad \text{for integer } m$$

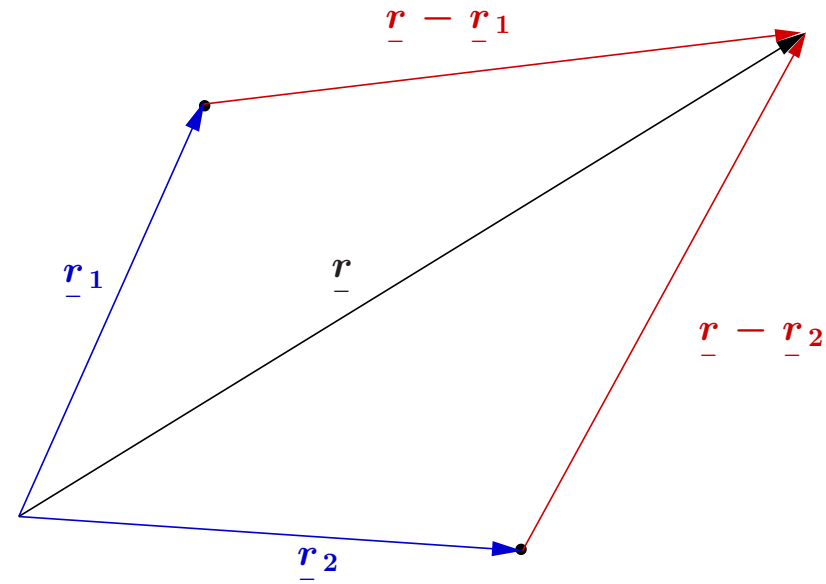
$$I \;=\; (\sqrt{I_1} - \sqrt{I_2}\,)^2$$

- If intensities are equal, $I_1 = I_2 =: I_0$,

$$I \;=\; 2\,I_0\,\big(\,1 \;+\; \cos(\delta_2 - \delta_1)\,\big) \;=\; 4\,I_0\cos^2\!\frac{(\delta_2 - \delta_1)}{2}$$

Relation between phase difference $\delta_2 - \delta_1$ and difference in optical path length:

In a medium with constant index of refraction n, consider two coherent sources of frequency ω at points with position vector $\underline{r}_1$ and $\underline{r}_2$.



At point with position vector $\underline{r}$:

$$u_1(\underline{r}, t) = \frac{B_1}{|\underline{r} - \underline{r}_1|} \cos\big(\omega t - k|\underline{r} - \underline{r}_1|\big) ,$$

$$u_2(\underline{r}, t) = \frac{B_2}{|\underline{r} - \underline{r}_2|} \cos\big(\omega t - k|\underline{r} - \underline{r}_2|\big) .$$

Phase difference:

$$\delta_2 - \delta_1 = k \left(|\underline{r} - \underline{r}_1| - |\underline{r} - \underline{r}_2| \right) .$$

With $k = \omega/v = \omega\, n/c$:

$$\delta_2 - \delta_1 = \frac{\omega}{c} \cdot \left(n\,|\underline{r} - \underline{r}_1| - n\,|\underline{r} - \underline{r}_2| \right)$$

$$\text{Phase difference} = \frac{2\,\pi}{\text{vacuum wave length}} \cdot \text{difference in optical path length}$$

Interference maxima:

$$\frac{\text{difference in optical path length}}{\text{vacuum wave length}} = m\,, \quad m = 0, \pm 1, \pm 2, \ldots$$

Interference minima:

$$\frac{\text{difference in optical path length}}{\text{vacuum wave length}} = \frac{2m+1}{2}\,, \quad m = 0, \pm 1, \pm 2, \ldots$$

If the index of refraction n varies from point to point, the optical path length must be written as an integral,

$$n\,|\underline{r} - \underline{r}_1| \mapsto \int_{P_1} n\,ds\,, \qquad n\,|\underline{r} - \underline{r}_2| \mapsto \int_{P_2} n\,ds\,,$$

where P_1 and P_2 are the rays from the point with position vector $\underline{r}_1$ and $\underline{r}_2$, respectively, to the point with the position vector $\underline{r}$.

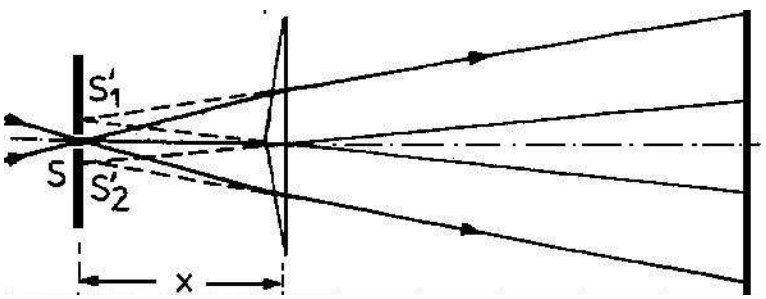
c) Interference experiments

We now give an overview of interference experiments.

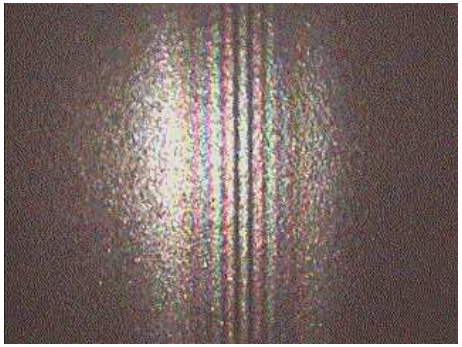
For all examples in this section we

- use scalar optics (i.e., we describe light in terms of a scalar function $u(\underline{r}, t)$);
- assume that the experiment is done in a medium with constant index of refraction unless otherwise stated.

α) Fresnel's biprism



S_1' and S_2' act like a double-slit with built-in coherence. Thus, we get the same interference pattern as for the double-slit.



picture from http://www.physics.umd.edu

β) Lloyd's mirror

Humphrey Lloyd (1800–1881)

Same geometry as for double slit, so one would expect the same interference pattern:

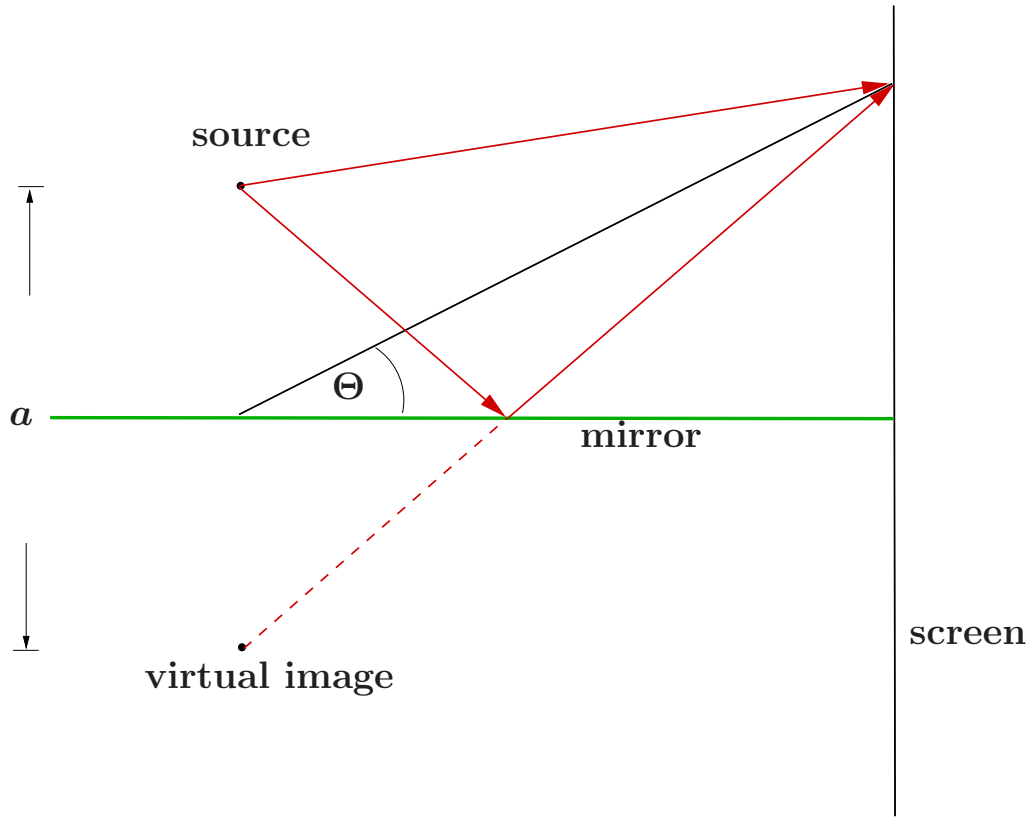
center bright

minima at $\sin\Theta_m = \frac{(2m-1)}{2}\frac{\lambda}{a}$.

That's wrong! Experiment shows:

center dark

maxima at $\sin\Theta_m = \frac{(2m-1)}{2}\frac{\lambda}{a}$.



Explanation:

Wave undergoes phase shift of π when reflected at mirror.

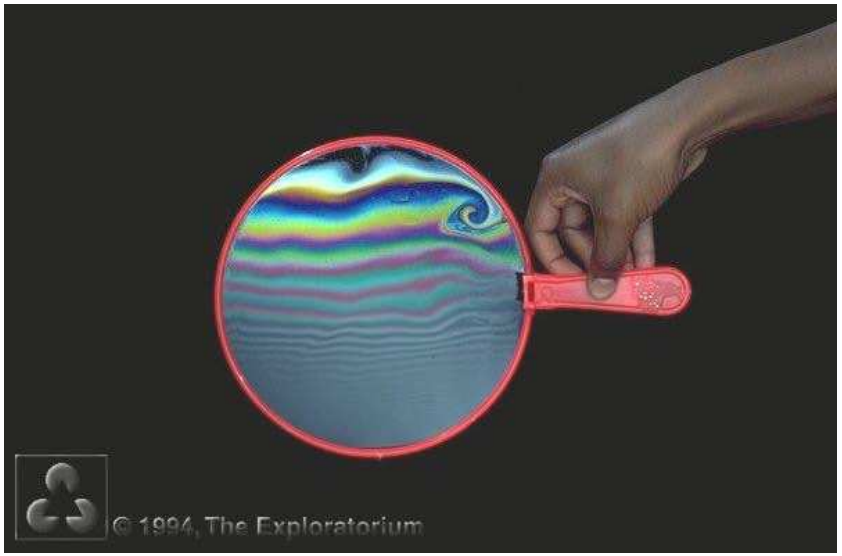
(We will show later how to derive this from Maxwell's equations.)

γ) Thin films

Observe thin transparent material, e.g.

- soap bubble
- oil layer
- plastic layer of CD
- wing of a butterfly

in white light.



picture from http://www.exploratorium.edu/imagery/

picture from http://homepages.compuserve.de/kunz2andy/

See colored interference pattern, resulting from light reflected at front side interfering with light reflected at back side.

Consider a layer of thickness d with index of refraction n', sandwiched between medium with index of refraction n and medium with index of refraction n''.

Difference in optical path length for the two beams:

$$\Delta = n'\left(\overline{AB} + \overline{BC}\right) - n\,\overline{AD} \qquad (*)$$

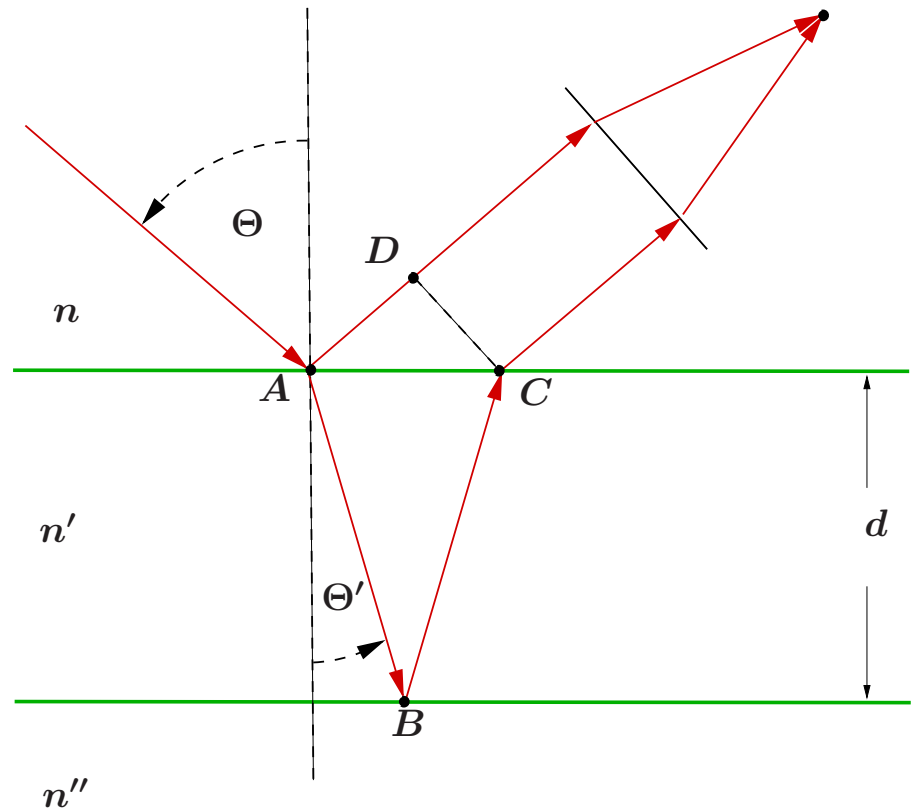
From geometry:

$$\overline{AB} = \overline{BC} = \frac{d}{\cos\Theta'} \qquad (**)$$

$$\overline{AD} = \overline{AC}\sin\Theta = 2\,d\tan\Theta'\sin\Theta$$

The last equation can be rewritten with Snell's law:

$$\overline{AD} = 2\,d\tan\Theta'\,\frac{n'}{n}\sin\Theta' \qquad (***)$$



Insert $(**)$ and $(***)$ into $(*)$:

$$\Delta = \frac{2\,n'\,d}{\cos\Theta'}\left(1 - \sin^2\Theta'\right) = 2\,n'\,d\cos\Theta'$$

Additional phase jump of π occurs at reflection from optically thinner to optically thicker medium. (We will see later how to derive this from Maxwell's equations.) Thus:

Difference in optical path length $= 2\,n'\,d\cos\Theta' + \Delta_r$

$$\Delta_r = \begin{cases} 0 \quad \text{if} \quad n < n' < n'' \quad \text{or} \quad n > n' > n'' \\ \lambda/2 \qquad\qquad \text{otherwise} \end{cases}$$

where $\lambda =$ vacuum wave length.

Minima: $\quad 2\,n'\,d\cos\Theta' + \Delta_r = \frac{2m+1}{2}\,\lambda$

Maxima: $\quad 2\,n'\,d\cos\Theta' + \Delta_r = m\,\lambda$

$m = 0, \pm 1, \pm 2, \ldots$

Θ' can be expressed in terms of Θ by means of Snell's law.

Thin-film interference is an example for "amplitude division" (amplitude of incoming wave is split into two halves) as opposed to "wave front division" (wave front is split into two halves, as by Fresnel's biprism or Lloyd's mirror).

- **If only first and second beam are considered, as in the picture, we get a two-beam interference pattern which is similar to the double-slit diffraction pattern.**
- **If in addition beams are considered that undergo several reflections inside the layer, a multi-beam interference pattern results. Note that the intensity of these beams decreases with increasing number of reflections. This decrease is strongest for small Θ.**
- **Thin-film interference can be used for measuring d or n'.**
- **An inportant application of thin film interference is anti-reflective coating which will be discussed later.**

Thin film of variable thickness leads to fringes of equal thickness ("Newton's rings").

picture from http://www.physics.ucsd.edu/

Newton's rings allow to determine the radius of curvature R of a spherical lens surface:

mth minimum:
(Assume $n > n'$ and $n'' > n'$)

$$2\,n'\,d\cos\Theta' + \frac{\lambda}{2} = \frac{2m+1}{2}\,\lambda$$

$$n' \approx 1 \text{ (air)}$$

$$\cos\Theta' \approx 1 \text{ (vertical incidence)}$$

$$2\,d = m\,\lambda$$

Radius of mth dark ring:

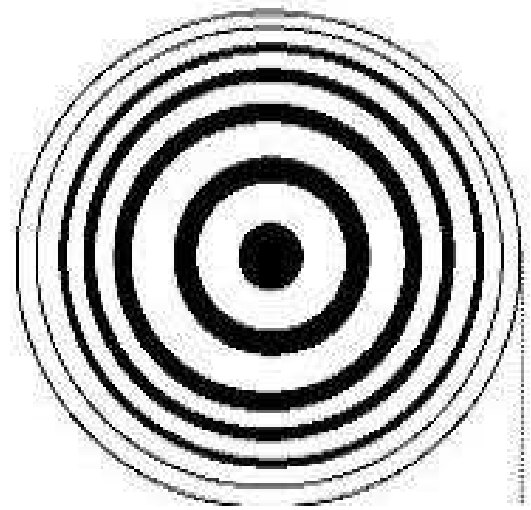
$$r_m = \sqrt{R^2 - (R-d)^2}$$

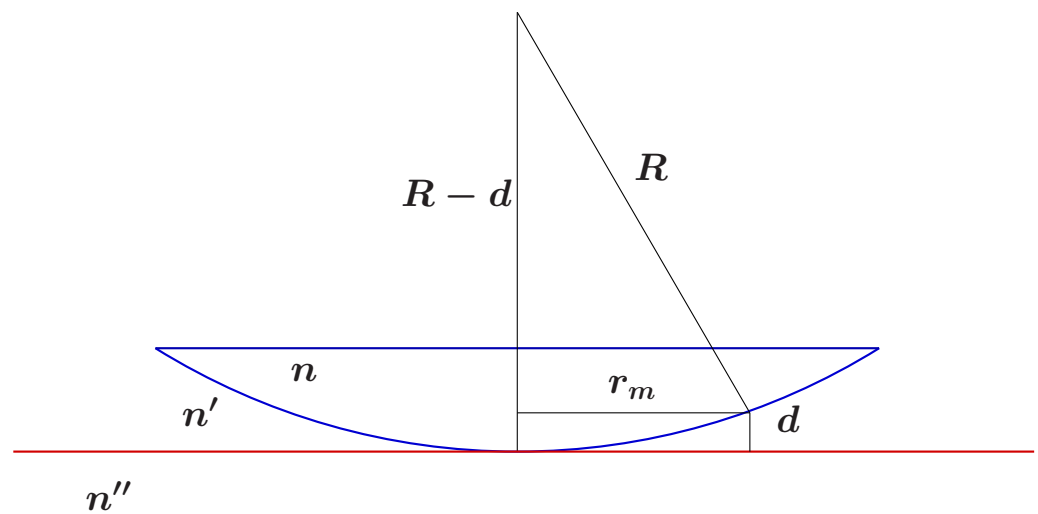
$$= \sqrt{\cancel{R^2} - \cancel{R^2} + 2Rd - d^2}$$

$$= \sqrt{2Rd\left(1 - \frac{d^{\cancel{2}}}{2R\cancel{d}}\right)}$$

$$= \sqrt{2Rd}\,\sqrt{1 - \frac{d}{2R}} \approx \sqrt{2Rd}$$

$$R \approx \frac{r_m^2}{m\,\lambda}$$

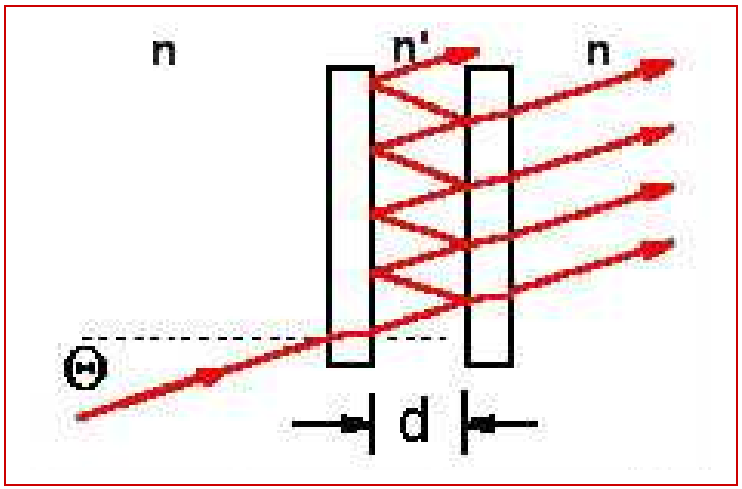




picture from http://www.practicalphysics.org/imageLibrary

δ) Fabry-Perot interferometer

Charles Fabry and Alfred Perot (1899)



picture from http://www.chemicool.com

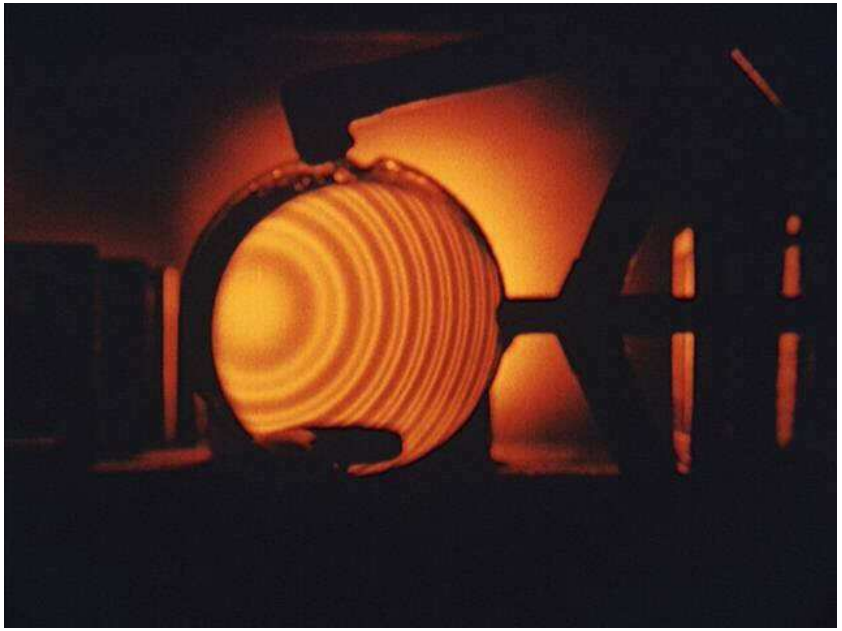
A Fabry-Perot interferometer with fixed d is often called an "etalon" (french for standard or gauge).

- point source on axis: circular pattern
- line source (slit): parallel stripes

Difference in optical path length for two successive beams can be calculated in analogy to thin films, cf. p.136. (If the refraction index of the glass plates is bigger than n', there is a phase jump of π at each reflection. As there is always an *even* number of reflections, this is irrelevant because the phase is defined only up to integer multiples of 2π.)

$$\boxed{\begin{array}{ll} \text{minima:} & 2\,n'\,d\cos\Theta' \;=\; \frac{2m+1}{2}\;\lambda \\ \text{maxima:} & 2\,n'\,d\cos\Theta' \;=\; m\,\lambda \end{array}} \qquad m = 0, \pm 1, \pm 2, \ldots$$

where $n\sin\Theta \;=\; n'\sin\Theta'$ and $\lambda =$ vacuum wave length .



picture from http://www.engineering.sdstate.edu

Interference pattern produced by a Fabry-Perot interferometer, with a sodium gas discharge lamp. The light source emits two yellow spectral lines at $\lambda_1 = 588.9950\,\mathrm{nm}$ and $\lambda_2 = 589.5924\,\mathrm{nm}$ ("sodium doublet").

Condition that, for a wave length interval $[\lambda, \lambda + \Delta\lambda]$, the mth order maximum and the $(m+1)$th order maximum do not overlap:

$$(m+1)\,\lambda > m\,(\lambda + \Delta\lambda)$$

$$\Delta\lambda < \lambda/m =: (\Delta\lambda)_{fsr}$$

$(\Delta\lambda)_{fsr}$ is called the free spectral range of the Fabry-Perot interferometer.

For $\Theta = \Theta' = 0$, the order number m satisfies $2\,d\,n' = m\,\lambda$, thus

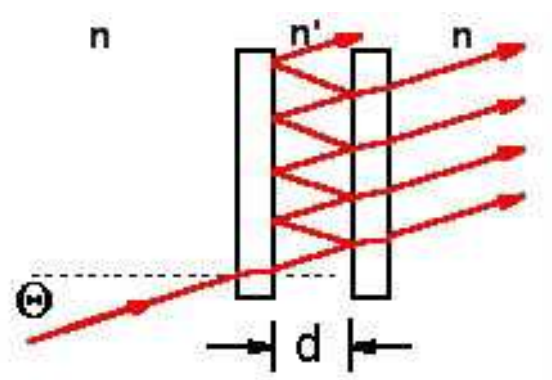
$$(\Delta\lambda)_{fsr} = \frac{\lambda^2}{2\,d\,n'}$$

Now restrict to the case $n' = n$, thus $\Theta = \Theta'$.

Goal: Calculate the total intensity as a function of Θ.

This requires calculating the superposition of infinitely many beams.

The calculation is based on two observations:



- Whenever beam meets glass plate, a fraction r of the amplitude is reflected $(0 < r < 1)$, the rest is transmitted.
- Two subsequent beams have difference in optical path length of $\Delta = 2\,d\,n\cos\Theta$, thus a phase difference of

$$\delta = \frac{2\,\pi}{\lambda}\,\Delta = \frac{4\,\pi}{\lambda}\,d\,n\cos\Theta$$

Incoming beam: $\mathrm{Re}\left\{ A\,e^{i(\underline{k}\cdot\underline{r}-\omega t)} \right\}$

First transmitted beam: $\mathrm{Re}\left\{ (1-r^2)\,A\,e^{i(\underline{k}\cdot\underline{r}-\omega t)} \right\}$

Second transmitted beam: $\mathrm{Re}\left\{ (1-r^2)\,r^2 A\,e^{i(\underline{k}\cdot\underline{r}-\omega t+\delta)} \right\}$

...

K^{th} transmitted beam: $\mathrm{Re}\left\{ (1-r^2)\,r^{2(K-1)} A\,e^{i(\underline{k}\cdot\underline{r}-\omega t+[K-1]\delta)} \right\}$

Summing up all transmitted beams:

$$u(\underline{r},t) \;=\; \mathrm{Re}\left\{ (1-r^2)\,A\,e^{i(\underline{k}\cdot\underline{r}-\omega t)} \left(1 \;+\; r^2 e^{i\delta} \;+\; (r^2 e^{i\delta})^2 \;+\; \ldots \;+\; (r^2 e^{i\delta})^{K-1} \;+\; \ldots \right) \right\}$$

Infinite geometric progression ($q = r^2 e^{i\delta}$, $|q\,| = r^2 < 1$):

$$1 \;+\; q \;+\; q^2 \;+\; \ldots \;+\; q^{K-1} \;+\; \ldots \;=\; \frac{1}{1-q}$$

$$u(\underline{r},t) \;=\; \mathrm{Re}\left\{ Z\,e^{i(\underline{k}\cdot\underline{r}-\omega t)} \right\}, \qquad Z \;:=\; \frac{(1-r^2)\,A}{1-r^2 e^{i\delta}}$$

Decompose Z into modulus and phase, $Z \;=\; |Z|\,e^{i\varphi}$:

$$u(\underline{r},t) \;=\; \mathrm{Re}\left\{ |Z|\,e^{i(\underline{k}\cdot\underline{r}-\omega t+\varphi)} \right\} \;=\; |Z|\cos(\underline{k}\cdot\underline{r} - \omega t + \varphi)$$

Intensity of all transmitted beams:

$$I_{\mathrm{trans}} \;=\; \overline{u^2} \;=\; \frac{|Z|^2}{2} \;=\; \frac{A^2}{2}\,\frac{\left(1-r^2\right)^2}{\left|1-r^2 e^{i\delta}\right|^2} \;=\; \frac{A^2}{2}\,\frac{\left(1-r^2\right)^2}{\left(1-r^2 e^{i\delta}\right)\left(1-r^2 e^{i\delta}\right)^*}$$

$$=\; \frac{A^2}{2}\,\frac{\left(1-r^2\right)^2}{\left(1-r^2 e^{i\delta}\right)\left(1-r^2 e^{-i\delta}\right)} \;=\; \frac{A^2}{2}\,\frac{\left(1-r^2\right)^2}{\left(1+r^4-2r^2\cos\delta\right)}$$

With the identity $\cos\delta = 1 - 2\sin^2\frac{\delta}{2}$ and $A^2/2 = I_{\text{inc}}$ = incoming intensity:

$$I_{\text{trans}} = \frac{I_{\text{inc}}\,(1-r^2)^2}{(1-r^2)^2 + 4\,r^2\sin^2\frac{\delta}{2}} = \frac{I_{\text{inc}}}{1 + \dfrac{4\,r^2}{(1-r^2)^2}\sin^2\frac{\delta}{2}}\,.$$

The resulting expression for the ratio $I_{\text{trans}}/I_{\text{inc}}$ is known as the "Airy function":

$$\boxed{T = \frac{I_{\text{trans}}}{I_{\text{inc}}} = \frac{1}{1 + f\sin^2\frac{\delta}{2}}\,, \qquad \delta = \frac{4\,\pi}{\lambda}\,d\,n\cos\Theta}$$

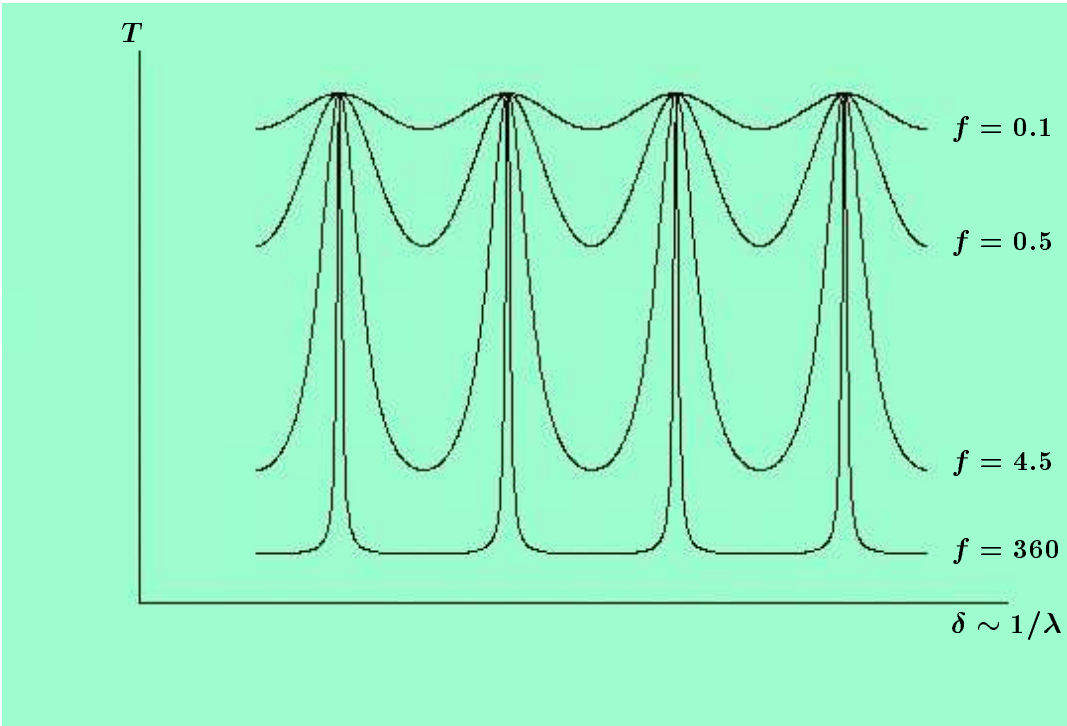
where

$$f = \frac{4\,r^2}{(\,1 - r^2\,)^2}$$

is the "coefficient of finesse". The "finesse" F is defined as $F = \frac{\pi}{2}\sqrt{f}$.

If the reflectance r and, thus, the coefficient of finesse f is known, the Airy function gives the ratio $T = I_{\text{trans}}/I_{\text{inc}}$ as a function of $\delta \sim 1/\lambda$ (for fixed Θ).

Plot of the Airy function $T = \dfrac{1}{1 + f \sin^2 \frac{\delta}{2}}$ for different values of f:

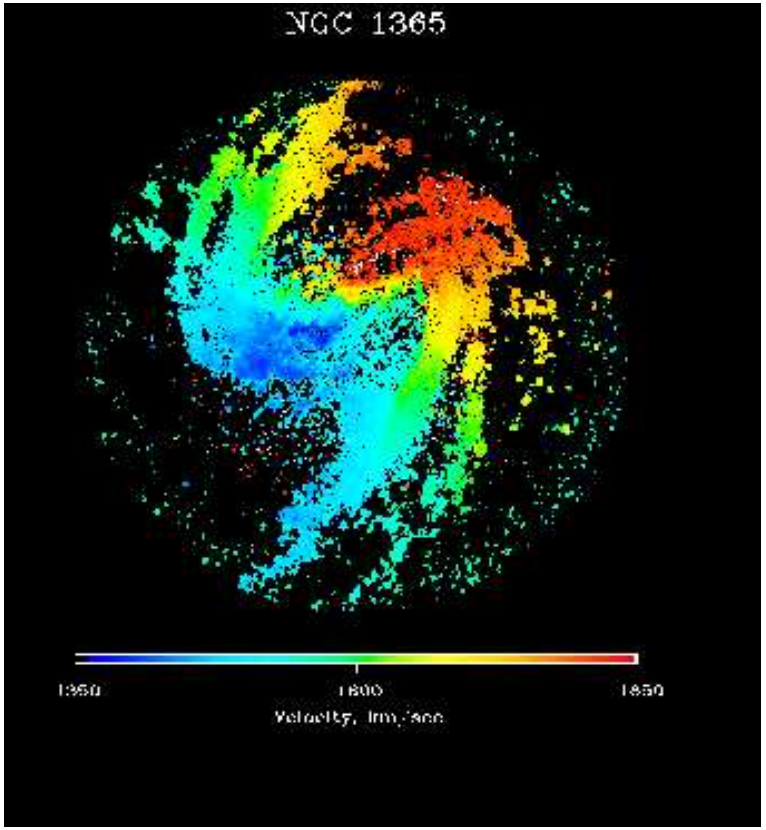


picture from http://lqcc.ustc.edu.cn/cui/

Resolving power with respect to wave length increases with increasing f.

Applications of Fabry-Perot:

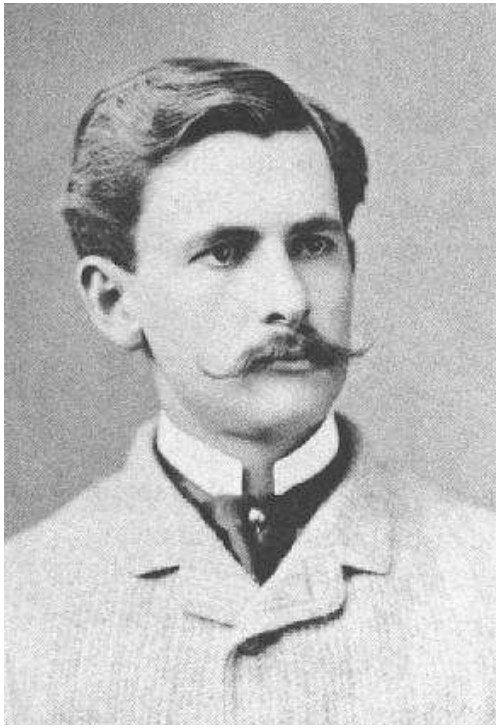
- high precision spectroscopy
- optical filters
- length standard
- laser resonator
- (astrophysical) photography in a small spectral range
- ...



picture from http://mingus.as.arizona.edu/ bjw/

Velocity field of the galaxy NGC 1365, obtained with the Rutgers Fabry-Perot, CTIO 1.5-m telescope. Radial motion (red: away from us, blue: towards us) is measured in terms of Doppler shift of spectral lines.

ε) Michelson interferometer



picture from http://www.lightandmatter.com/

Albert A. Michelson (1852–1931)

Nobelprize 1907

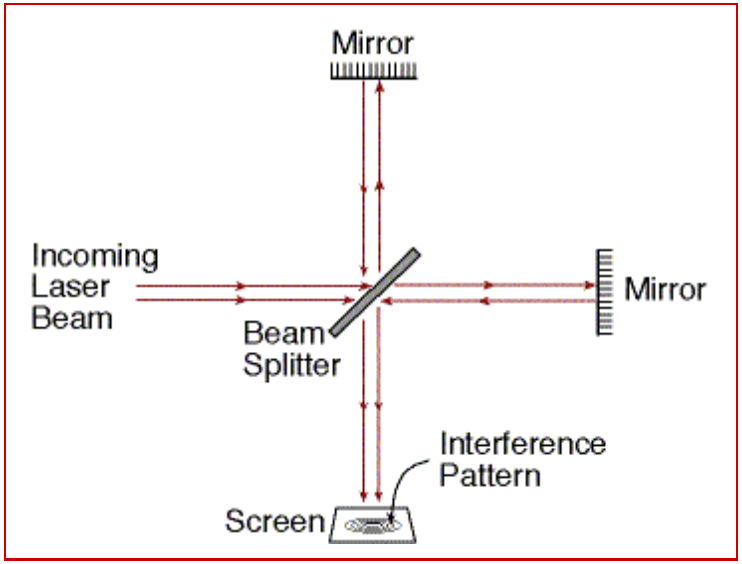
The Michelson interferometer was designed for measuring the velocity of the Earth with respect to the (hypothetical) ether.

The experiment was carried through

- by Michelson in Berlin and Potsdam, Germany, 1881
- by Michelson and Morley in Cleveland, Ohio, 1888

The negative outcome of this experiment was crucial for the advent of Special Relativity.

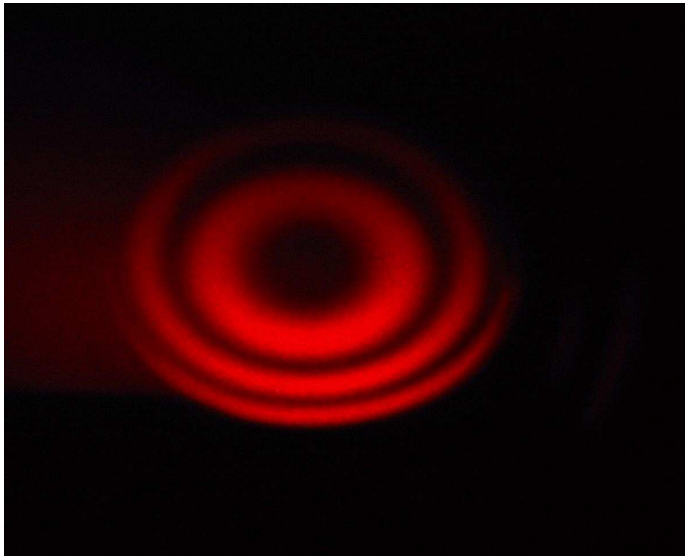
Sketch of a Michelson interferometer



picture from http://web.phys.ksu.edu/vqm/laserweb

If the mirrors are perfectly aligned, the fringes are circular.

The fringes move whenever the difference in optical path length of the two beams is changed (e.g. by moving one mirror, or by placing a sample into one arm).



picture from http://www1.union.edu

If one of the mirrors is moved by $\lambda/4$, the center changes from dark to bright.

More generally:

If difference in optical path lengths changes by

$$\Delta - \tilde{\Delta} = m\,\lambda\,,$$

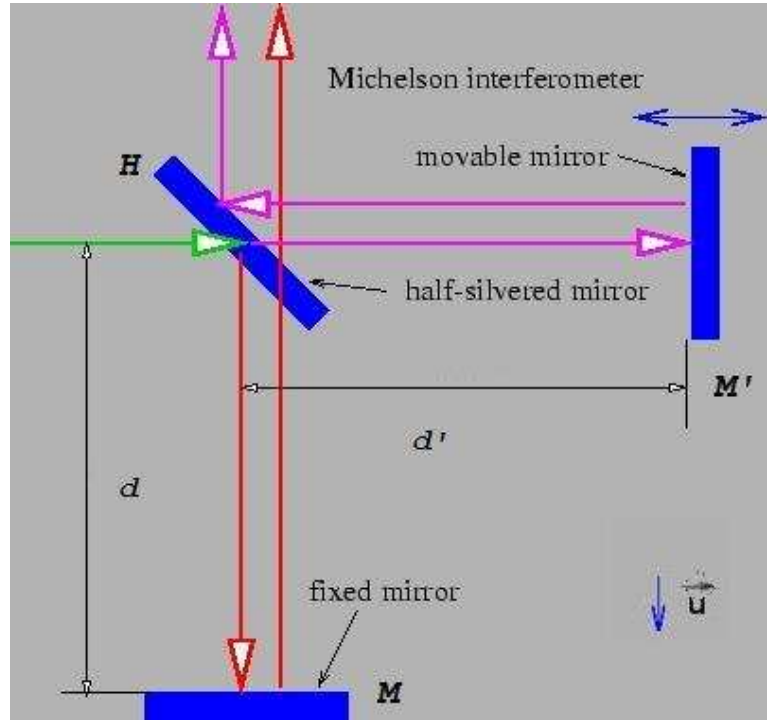
m fringes move by.

Michelson's attempt to measure the velocity of the Earth relative to the ether:

$c =$ velocity of light relative to the ether

$u =$ velocity of the Earth relative to the ether

Assume first that, in the ether system, the Earth moves in the direction from H to M, as indicated in the picture.



picture from http://www2.selu.edu/Academics

First beam goes

- from H to M in time t_1 : $\quad ct_1 = d + ut_1\,, \qquad t_1 = \frac{d}{c-u}$
- back from M to H in time t_2 : $\quad ct_2 = d - ut_2\,, \qquad t_2 = \frac{d}{c+u}$

Second beam goes

- from H to M' in time t_1' : $\quad (ct_1')^2 = d'^2 + (ut_1')^2\,, \qquad t_1' = \frac{d'}{\sqrt{c^2-u^2}}$
- back from M' to H in time t_2' : $\quad (ct_2')^2 = d'^2 + (ut_2')^2\,, \qquad t_2' = \frac{d'}{\sqrt{c^2-u^2}}$

Difference in optical path length:

$$\Delta = c(t_1 + t_2 - t_1' - t_2') = \frac{c\,d}{c-u} + \frac{c\,d}{c+u} - \frac{2\,c\,d'}{\sqrt{c^2-u^2}}$$

$$= \frac{2\,d}{1-\frac{u^2}{c^2}} - \frac{2d'}{\sqrt{1-\frac{u^2}{c^2}}} \approx 2d\left(1+\frac{u^2}{c^2}\right) - 2d'\left(1+\frac{u^2}{2c^2}\right)$$

$$= 2d - 2d' + (2d-d')\frac{u^2}{c^2}$$

Now rotate apparatus by 90°, such that, in the ether system, the Earth moves in the direction from H to M'.

First beam goes

- from H to M in time $\tilde{t}_1$: $(c\tilde{t}_1)^2 = d^2 + (u\tilde{t}_1)^2\,,\qquad \tilde{t}_1 = \frac{d}{\sqrt{c^2-u^2}}$
- back from M to H in time $\tilde{t}_2$: $(c\tilde{t}_2)^2 = d^2 + (u\tilde{t}_2)^2\,,\qquad \tilde{t}_2 = \frac{d}{\sqrt{c^2-u^2}}$

Second beam goes

- from H to M' in time $\tilde{t}_1'$: $c\tilde{t}_1' = d' + u\tilde{t}_1'\,,\qquad \tilde{t}_1' = \frac{d'}{c-u}$
- back from M' to H in time $\tilde{t}_2'$: $c\tilde{t}_2' = d' - u\tilde{t}_2'\,,\qquad \tilde{t}_2' = \frac{d'}{c+u}$

Difference in optical path length:

$$\tilde{\Delta} = c\,(\,\tilde{t}_1 + \tilde{t}_2 - \tilde{t}'_1 - \tilde{t}'_2\,) = \frac{2\,c\,d}{\sqrt{c^2 - u^2}} - \frac{cd'}{c-u} - \frac{cd'}{c+u} = \frac{2\,d}{\sqrt{1-\frac{u^2}{c^2}}} - \frac{2\,d'}{1-\frac{u^2}{c^2}}$$

$$\approx 2\,d\left(1 + \frac{u^2}{2\,c^2}\right) - 2\,d'\left(1 + \frac{u^2}{c^2}\right) = 2\,d - 2\,d' + (d - 2\,d')\frac{u^2}{c^2}$$

The quantity $\Delta - \tilde{\Delta}$ gives the number m of fringes that move by during the rotation process:

$$\Delta - \tilde{\Delta} = (\,d + d'\,)\,\frac{u^2}{c^2} = m\,\lambda\,.$$

One expected $u \approx$ velocity of the Earth relative to the Sun $\approx 30\,\mathrm{km/sec} \approx 10^{-4}\,c$.

In the Michelson-Morley 1888 experiment, it was $d + d' \approx 22\,\mathrm{m}$ (effective armlength enlarged with the help of mirrors) and $\lambda \approx 500\,\mathrm{nm}$, so one expected

$$m = \frac{(\,d + d'\,)}{\lambda}\,\frac{u^2}{c^2} \approx 0.4.$$

This would have been easily observable for Michelson. However, no change of the fringe pattern was actually observed.

Explanation given by Albert Einstein (1905): The ether does not exist. Light in vacuum propagates in all inertial systems with velocity c.

Some applications of Michelson interferometer:

- Measuring distance with accuracy in the order of λ
- Measuring thickness or index of refraction of samples placed into one arm
- Gravitational wave detectors

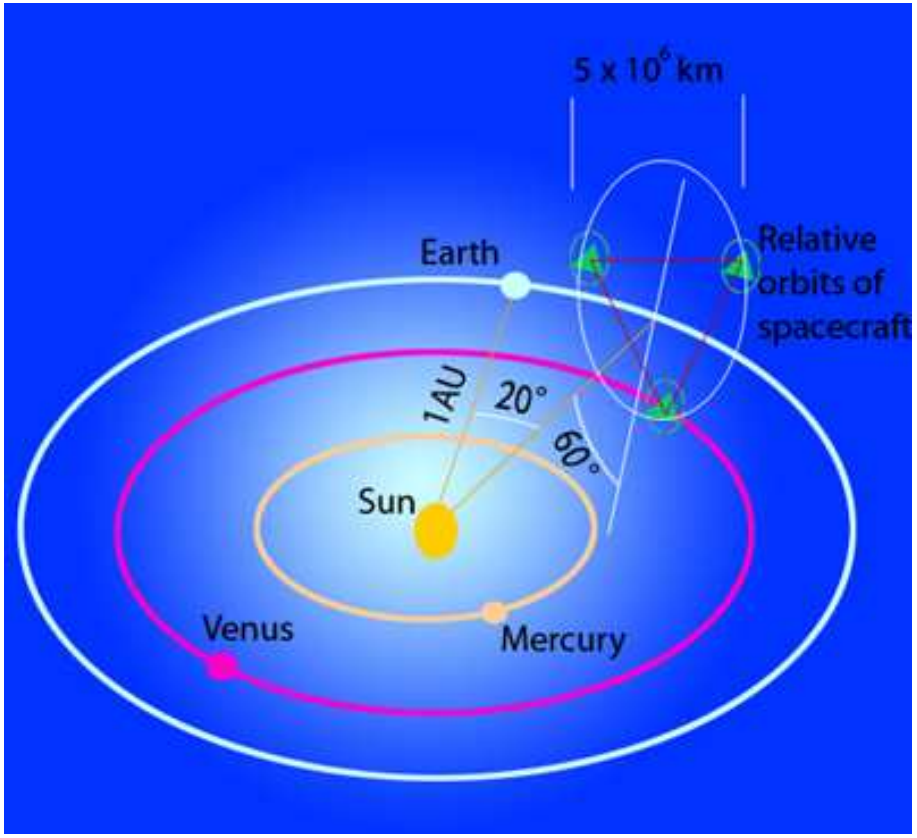
Gravitational waves are predicted by Einstein's general theory of relativity. Until now they have been detected only indirectly. (Hulse-Taylor pulsar loses energy which is interpreted as being radiated away in the form of gravitational waves, Nobelprize 1993). Direct detection of gravitational waves is attempted with Michelson interferometers, e.g.

Geo600, German-British project, Hannover, Germany.

picture from http://www.sr.bham.ac.uk/ adf

Gravitational wave detector in space:

Laser Interferometer Space Antenna (LISA)
ESA-NASA project, launch scheduled 2015



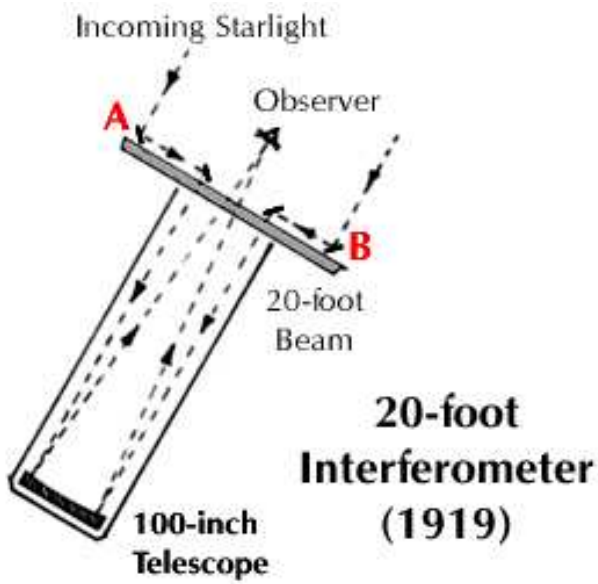
picture from http://www.esa.int/techresources

ζ) Other interferometers

There is a large variety of other interferometers. Here are two further examples. Both are two-beam interferometers, like the Michelson interferometer.

Michelson's stellar interferometer

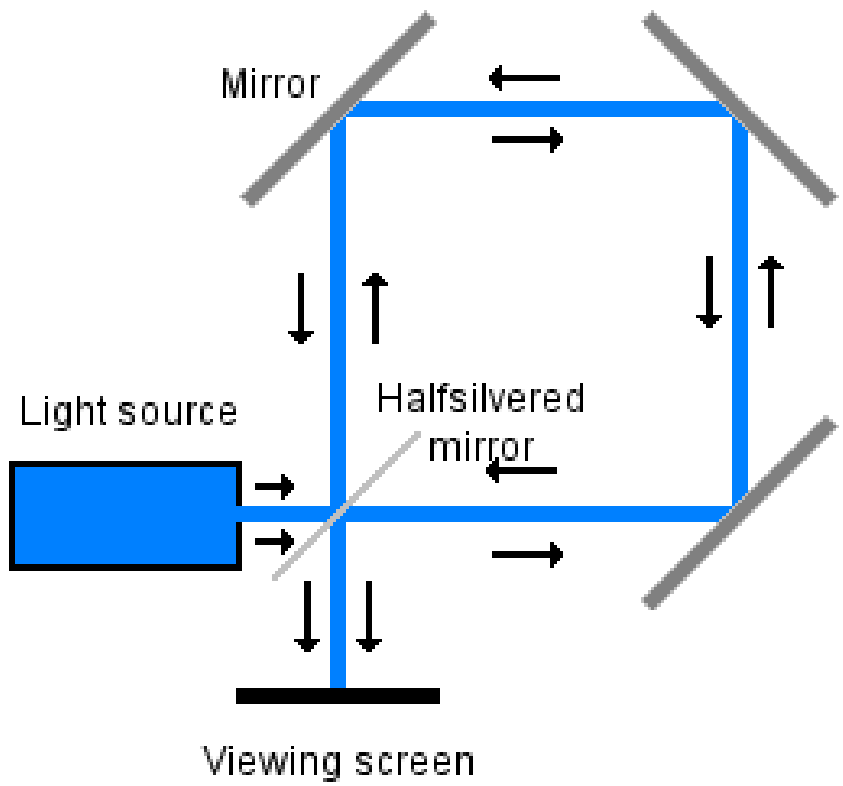
- was used by Michelson in 1919 to measure the angular diameter of the star Beteigeuze (red shoulder star in Orion)
- result was 0.044 arcseconds



picture from http://www.mtwilson.edu/vir/

Sagnac interferometer (Georges Sagnac, 1913)

- **ring interferometer, mounted on rotating platform**
- **was used by Michelson to measure the rotation of the Earth**
- **is now used e.g. in inertial guidance systems.**



picture from http://en.wikipedia.org

η) Radio interferometry

A radio telescope has a much lower resolving power than an optical telescope of the same aperture (recall Rayleigh criterion).

However, the resolving power of radio telescopes can be greatly enhanced by using a radio interferometric method, called "aperture synthesis", invented by Martin Ryle in the 1950s.

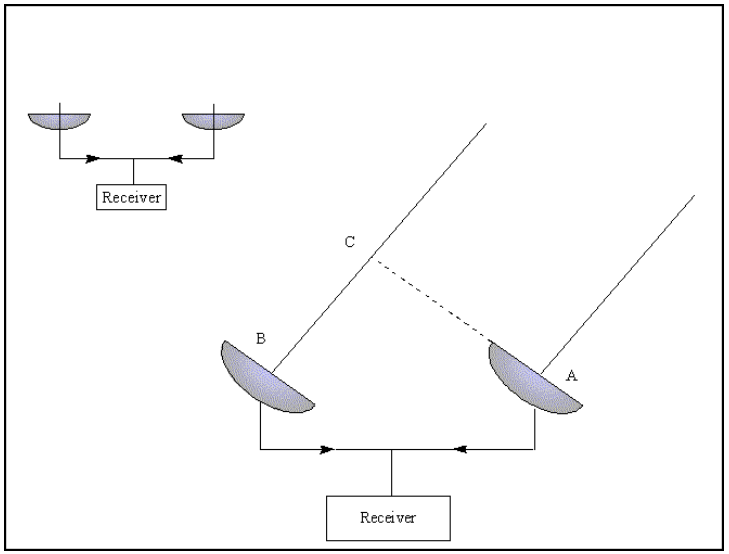
Basic idea: Connect a collection of telescopes together, such that the resulting resolving power is the same as for an instrument the size of the entire collection.

Aperture synthesis with radio telescopes on different continents can resolve angular distances of less than 0.0001 arcseconds ("Very Long Baseline Interferometry" = VLBI).

picture from http://www.eb.com/

Martin Ryle (1918-1984)

Nobelprize 1974



picture from

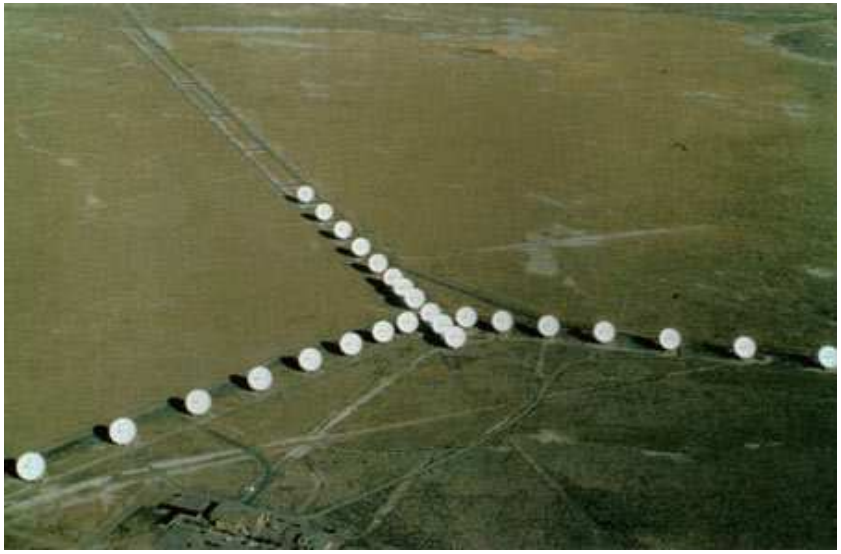
http://astrosun2.astro.cornell.edu/academics/courses//astro201

Resolving power is determined by length of baseline, not by diameter of the individual radio telescopes.

The larger the number of different baselines, the higher the quality of the pictures.

Very Large Array (VLA) near Socorro, New Mexico

- 27 radio telescopes in Y shaped array
- 25 meter diameter each
- 36 km longest baseline
- angular resolution $\approx$ 0.05 arcseconds

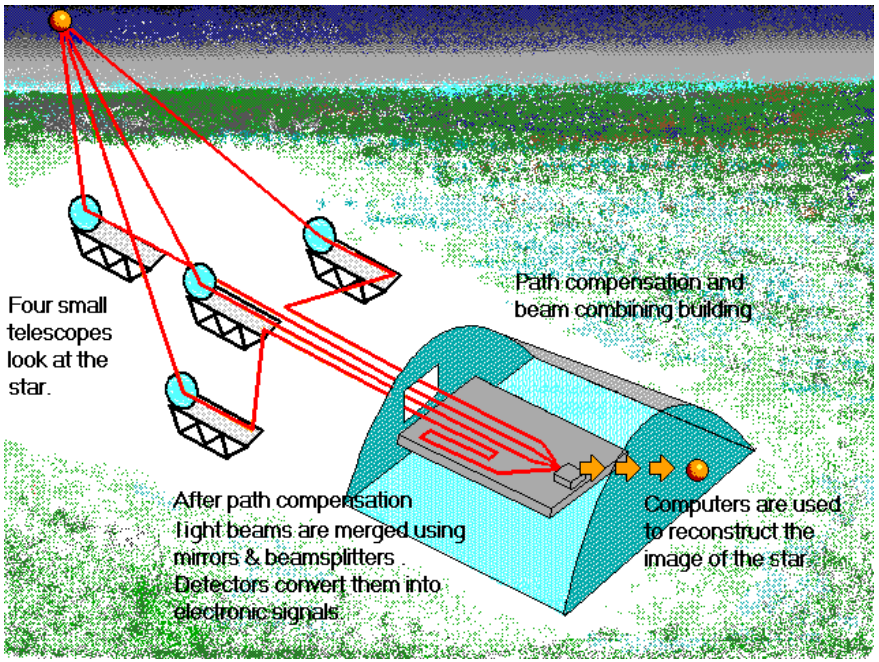


picture from http://deepspace.jpl.nasa.gov

Aperture synthesis with infrared or optical telescopes is much more difficult and more recent.

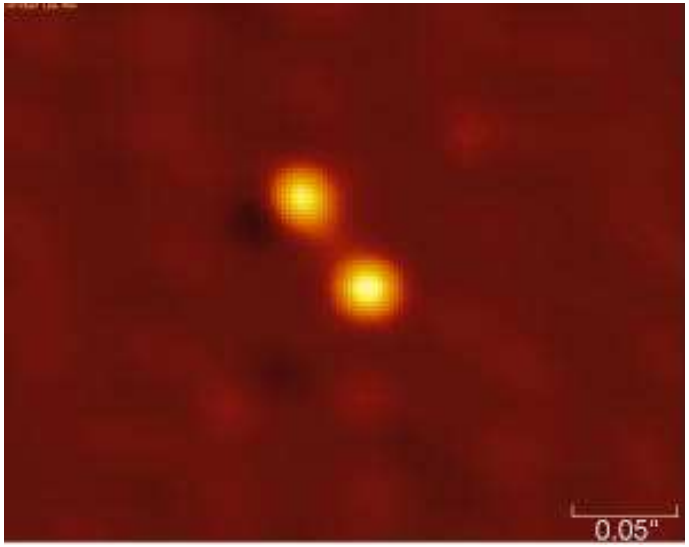
Cambridge Optical Aperture Synthesis Telescope (COAST)

- 4 telescopes, 40 cm diameter each
- 100 meter longest baseline
- observation in the red and near infrared
- aims at angular resolution of 0.001 arcseconds



picture from http://www.mrao.cam.ac.uk/telescopes/coast

Infrared picture of binary star Capella, taken by COAST



picture from http://138.238.143.191/astronomy/Chaisson

V. Polarisation

(cf. Pedrotti and Pedrotti, Chapter 15 [Chapters 14, 19, 20 treat more advanced aspects])

a) Linearly and circularly polarised light

Recall (p. 80):

Light is not a scalar wave, but rather described by two vector fields $\underline{E}$ and $\underline{B}$ that satisfy Maxwell's equations. The vector character is largely irrelevant for diffraction and interference, but it is crucial for understanding polarisation.

In a medium with constant permittivity ε_r and constant permeability μ_r,

$$\underline{D} = \varepsilon_r \varepsilon_0 \underline{E}\,, \qquad \underline{B} = \mu_r \mu_0 \underline{H}\,,$$

a plane harmonic wave is completely determined by the electric field

$$\underline{E}(\underline{r}\,,t) \;=\; \underline{E}_0 \cos(\underline{k}\cdot\underline{r} - \omega t + \alpha)\;;$$

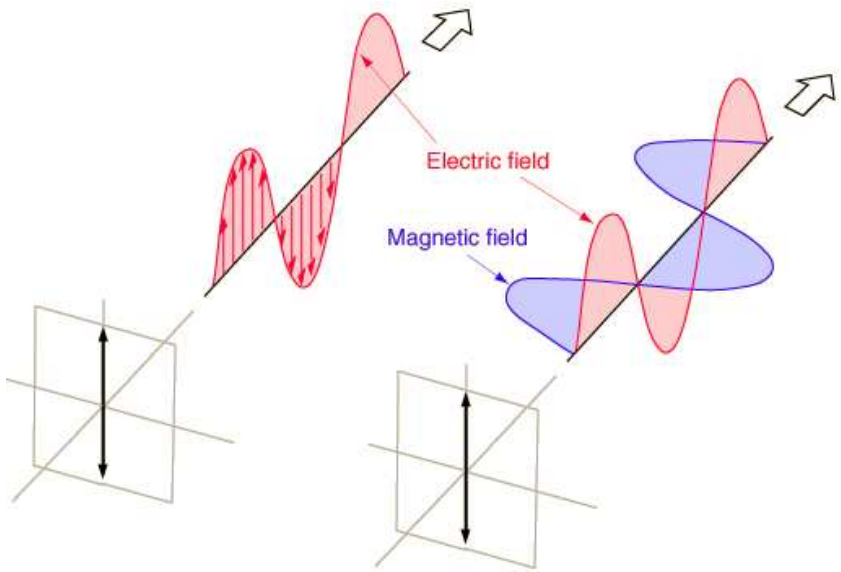
The corresponding magnetic field is given by $\underline{B}(\underline{r}\,,t) \;=\; \dfrac{1}{v}\dfrac{\underline{k}}{|\underline{k}|}\times\underline{E}(\underline{r}\,,t)$ where

$$v \;=\; \frac{c}{n} \;=\; \frac{c}{\sqrt{\varepsilon_r\,\mu_r}} \;=\; \frac{1}{\sqrt{\varepsilon_0\,\mu_0\,\varepsilon_r\,\mu_r}}$$ is the velocity of light in the medium.

An electromagnetic wave of the form

$$\underline{E}(\underline{r}, t) = \underline{E}_0 \cos(\underline{k} \cdot \underline{r} - \omega t + \alpha),$$

with constant $\underline{E}_0$, is called linearly polarised.



picture from

http://hyperphysics.phy-astr.gsu.edu/hbase/phyopt/

For an observer watching such a wave coming towards him, the $\underline{E}$-vector seems to oscillate in a line.

Superposition of two linearly polarised waves with the same wave vectors ($\underline{k}_1 = \underline{k}_2$), hence the same frequencies ($\omega_1 = \omega_2$), and the same phase shifts (i.e., $\alpha_1 = \alpha_2$) gives again a linearly polarised wave.

If the phase shifts are different, this is no longer true. Some special cases:

Two linearly polarised waves

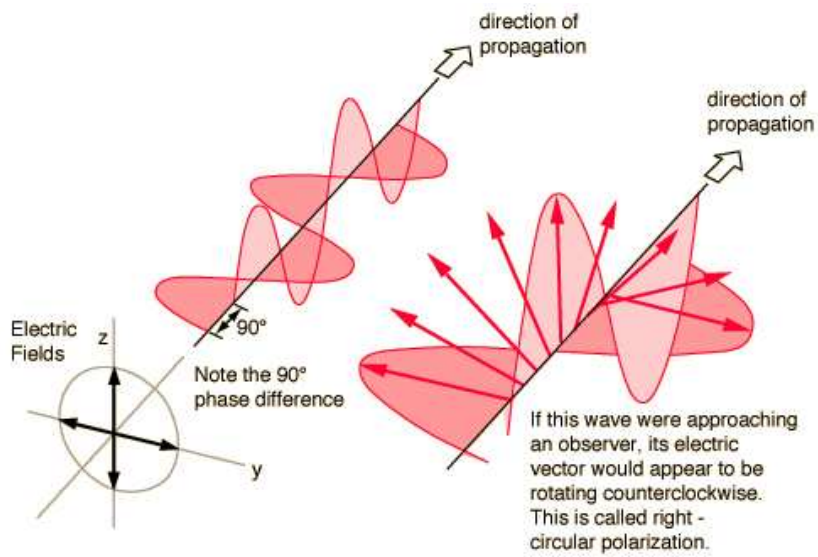
$$\underline{E}_1(\underline{r}\,,t) \;=\; \underline{E}_{01}\cos(\underline{k}\cdot\underline{r}\;-\omega t+\alpha_1)\,,$$

$$\underline{E}_2(\underline{r}\,,t) \;=\; \underline{E}_{02}\cos(\underline{k}\cdot\underline{r}\;-\omega t+\alpha_2)\,,$$

with

$$\underline{E}_{01}\cdot\underline{E}_{02} \;=\; 0\,, \qquad |\underline{E}_{01}| \;=\; |\underline{E}_{02}|\,, \qquad |\alpha_2-\alpha_1| \;=\; \pi/2$$

give a circularly polarised wave.



picture from

http://hyperphysics.phy-astr.gsu.edu/hbase/phyopt/

corkscrew: right-handed (as in picture)

anti-corkscrew: left-handed

For an observer watching such a wave coming towards him, the tip of the $\underline{E}$-vector seems to move on a circle.

Two linearly polarised waves

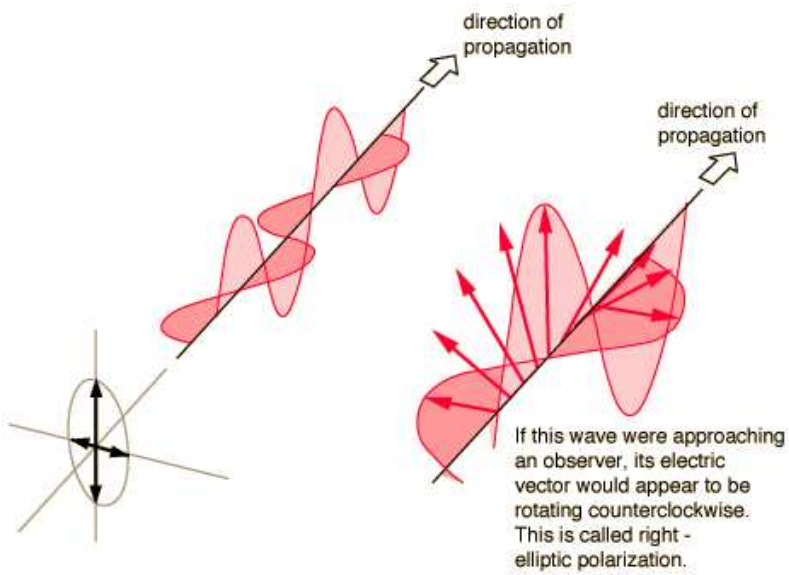
$$\underline{E}_1(\underline{r}, t) = \underline{E}_{01} \cos(\underline{k} \cdot \underline{r} - \omega t + \alpha_1) ,$$

$$\underline{E}_2(\underline{r}, t) = \underline{E}_{02} \cos(\underline{k} \cdot \underline{r} - \omega t + \alpha_2) ,$$

with

$$\underline{E}_{01} \cdot \underline{E}_{02} = 0 , \qquad |\underline{E}_{01}| \neq |\underline{E}_{02}| , \qquad |\alpha_2 - \alpha_1| = \pi/2$$

give an elliptically polarised wave.



picture from

http://hyperphysics.phy-astr.gsu.edu/hbase/phyopt/

corkscrew: right-handed (as in picture)

anti-corkscrew: left-handed

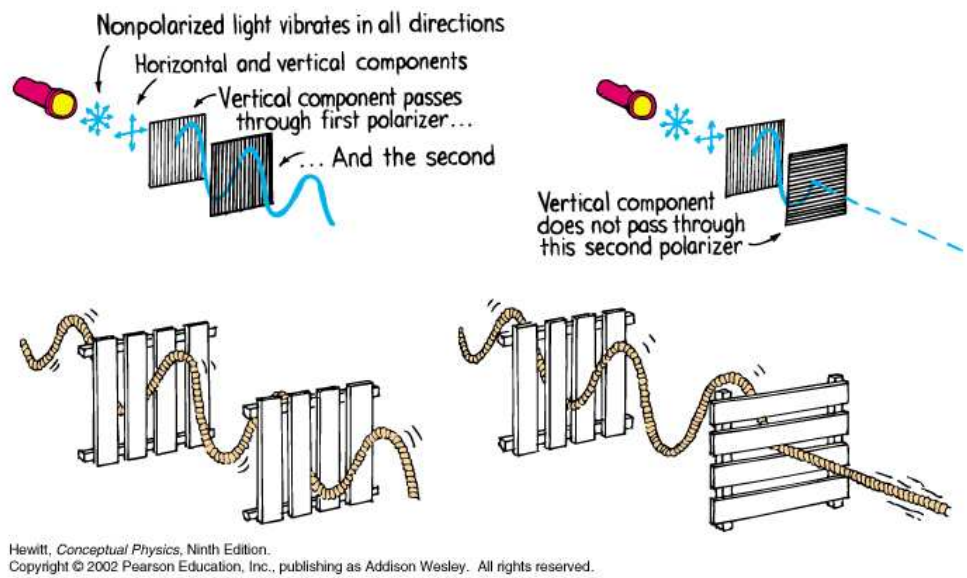
For an observer watching such a wave coming towards him, the tip of the $\underline{E}$-vector seems to move on an ellipse.

Waves from ordinary light sources (our Sun, light bulb, etc.) consists of wave trains with the direction of $\underline{E}_0$ varying randomly. This is called unpolarised light.

Unpolarised light can be made (partly) polarised by various methods to be discussed below.

Note: The fact that light can be polarised proves that light is a transverse vector wave. For longitudinal vector waves, or for scalar waves, polarisation does not exist.

Polarisation is often visualised by the "picket-fence analogy":



picture from http://physics.usc.edu/ bars/

b) Optical effects and their relation to polarisation

α) Reflection and refraction

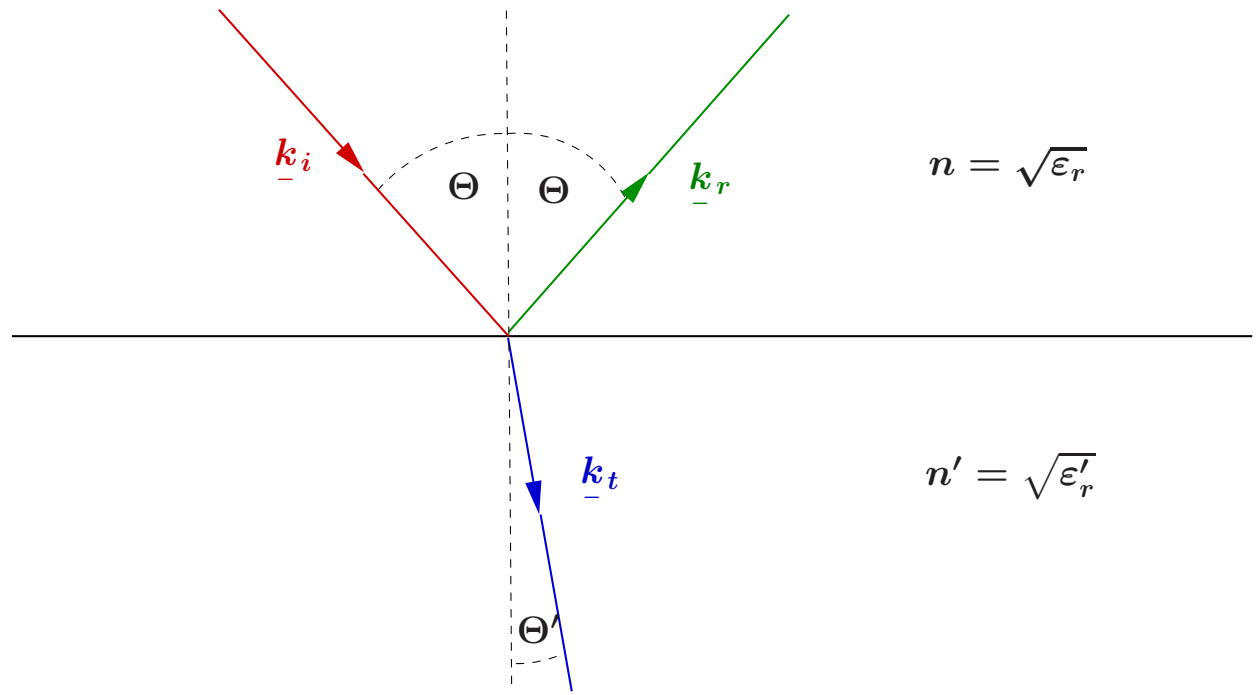
We know the basic laws of reflection and refraction. With the help of Maxwell's equations, taking the vector character of light waves into account, we will get additional information on the amplitudes of reflected and refracted waves.

Boundary between two dielectric media ($\mu_r = \mu'_r = 1$):

$\underline{E}_i$ = incident wave with wave vector $\underline{k}_i$

$\underline{E}_r$ = reflected wave with wave vector $\underline{k}_r$

$\underline{E}_t$ = transmitted wave with wave vector $\underline{k}_t$



The source-free Maxwell equations (in SI units)

$$\nabla \cdot \underline{D} = 0$$

$$\nabla \cdot \underline{B} = 0$$

$$\frac{\partial \underline{D}}{\partial t} - \nabla \times \underline{H} = \underline{0}$$

$$\frac{\partial \underline{B}}{\partial t} + \nabla \times \underline{E} = \underline{0}$$

imply:

At a boundary surface,

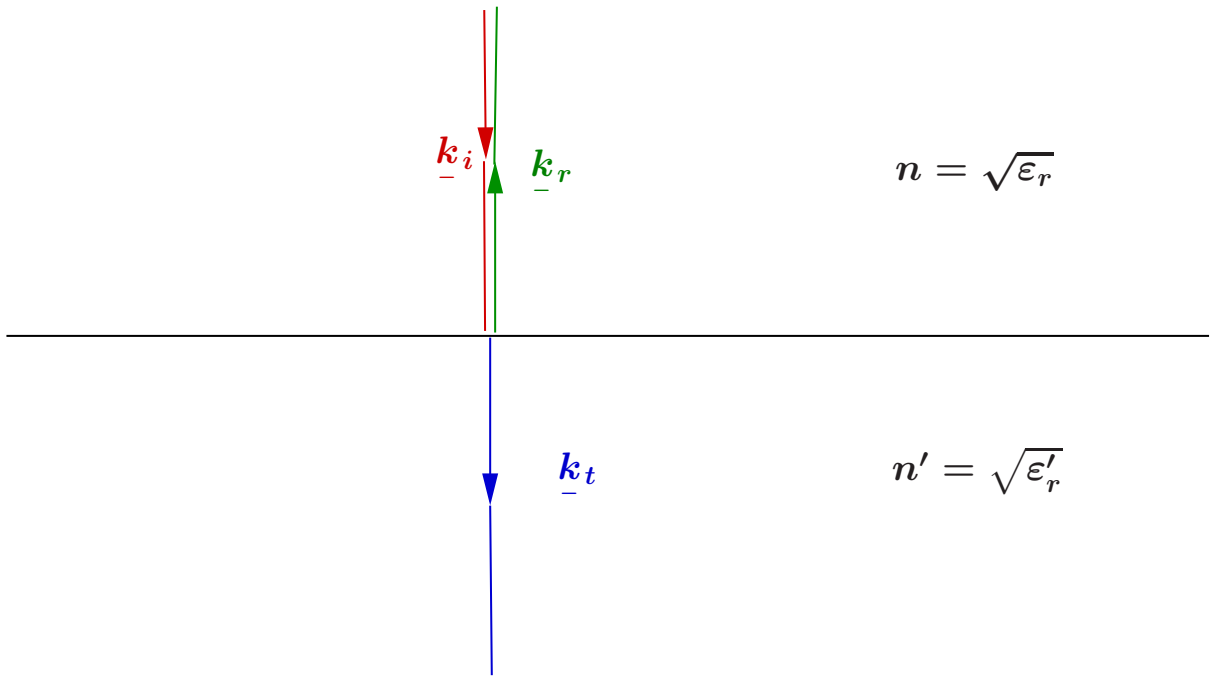
- tangential components of $\underline{E}$ and $\underline{H}$
- normal components of $\underline{D}$ and $\underline{B}$

are continuous.

For a proof see, e.g., J. D. Jackson: "Classical Electrodynamics" Second Edition, Wiley (1975), pp. 19–20.

With this information one can calcaluate $\underline{E}_r$ and $\underline{E}_t$ if $\underline{E}_i$ is given.

Consider simplest case first:
Normal incidence



- z-axis can be chosen perpendicular to boundary:

$$\underline{k}_i = -k_i\,\hat{\underline{z}}\ , \qquad \underline{k}_r = k_r\,\hat{\underline{z}}\ , \qquad \underline{k}_t = -k_t\,\hat{\underline{z}}\ .$$

- Waves are transverse and the situation is symmetric with respect to the $\underline{k}_i - \underline{E}_i$−plane, so we can choose the $x-$ and $y-$axes such that

$$\underline{E}_{0i} = E_i\,\hat{\underline{x}}\ , \quad \underline{E}_{0r} = E_r\,\hat{\underline{x}}\ , \quad \underline{E}_{0t} = E_t\,\hat{\underline{x}}\ .$$

(E_i, E_r and E_t can be positive or negative.)

E_t and E_r are determined by E_i in the following way:

- Continuity at boundary of tangential component of $\underline{E}$ requires

$$E_i + E_r = E_t \,. \qquad (*)$$

- Energy flux of EM wave is $\sim n|\underline{E}|^2$. (This was calculated in the third worksheet, Problem 1.) So conservation of energy requires

$$n\,E_i^2 = n\,E_r^2 + n'\,E_t^2 \,. \qquad (**)$$

If we eliminate E_r from $(*)$ and $(**)$ we get $(n' + n)\,E_t = 2\,n\,E_i$.
This equation implies that E_t has the same sign as E_i, i.e. $\underline{E}_t$ and $\underline{E}_i$ are in phase.

If we eliminate E_t from $(*)$ and $(**)$ we get $(n' + n)\,E_r = (n - n')\,E_i$.
Thus

if $n > n'$: $\underline{E}_r$ and $\underline{E}_i$ are in phase.

if $n < n'$: $\underline{E}_r$ gets a phase shift of π relative to $\underline{E}_i$.

This was already mentioned, as an observational fact, on p.133 and p.136. Nowe we have *derived* it from Maxwell's equations.

Ratio of reflected energy flux to incident energy flux gives the *reflectance* (cf. p.141):

$$r^2 = \frac{n\,E_r^2}{n\,E_i^2} = \frac{(n - n')^2}{(n + n')^2} \,.$$

Air-glass: $n = 1, n' = 1.5 \Rightarrow r^2 = 0.04$

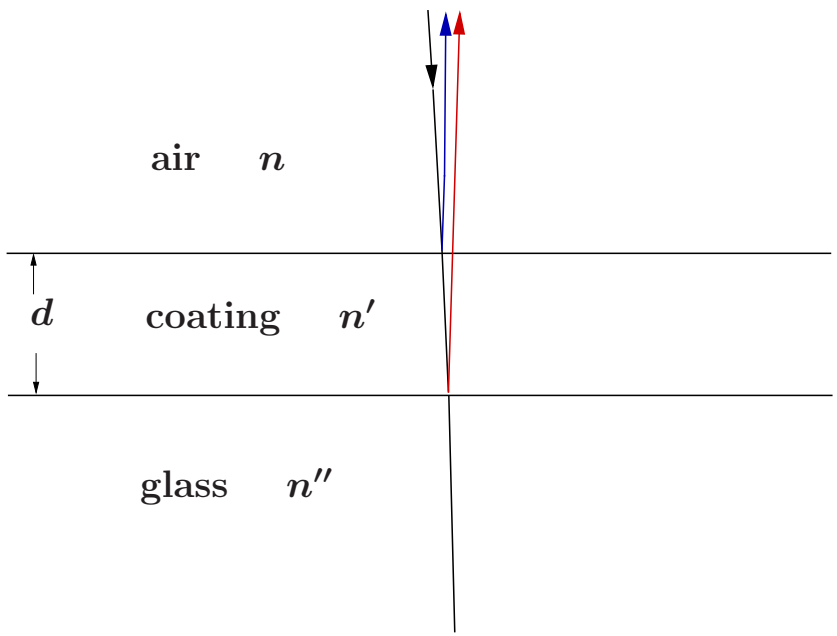
Air-diamond: $n = 1, n' = 2.5 \Rightarrow r^2 = 0.18$

Not easy to fake!

Note: For non-normal incidence, R increases, $r \to 1$ for $\Theta \to \pi/2$.

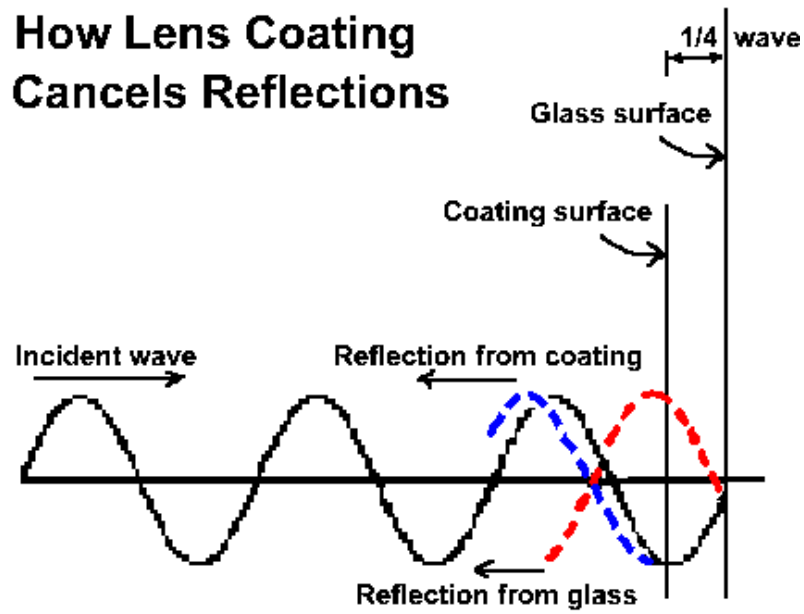
Application: Antireflective coating

Assume $n > n' > n''$, (approximately) normal incidence.



Destructive interference occurs for the two reflected waves if $n'd = \lambda/4$.

The picture on the right is actually not correct because it ignores the phase jump of π for the two reflected waves. (However, as both reflected waves get the same phase jump, the relative phase and, thus, the cancellation effect comes out correctly.)



picture from http://rick_oleson.tripod.com/coating

For $n'd = \lambda/4$, the two reflected waves cancel completely if they have the same amplitude. This turns out to be the case if

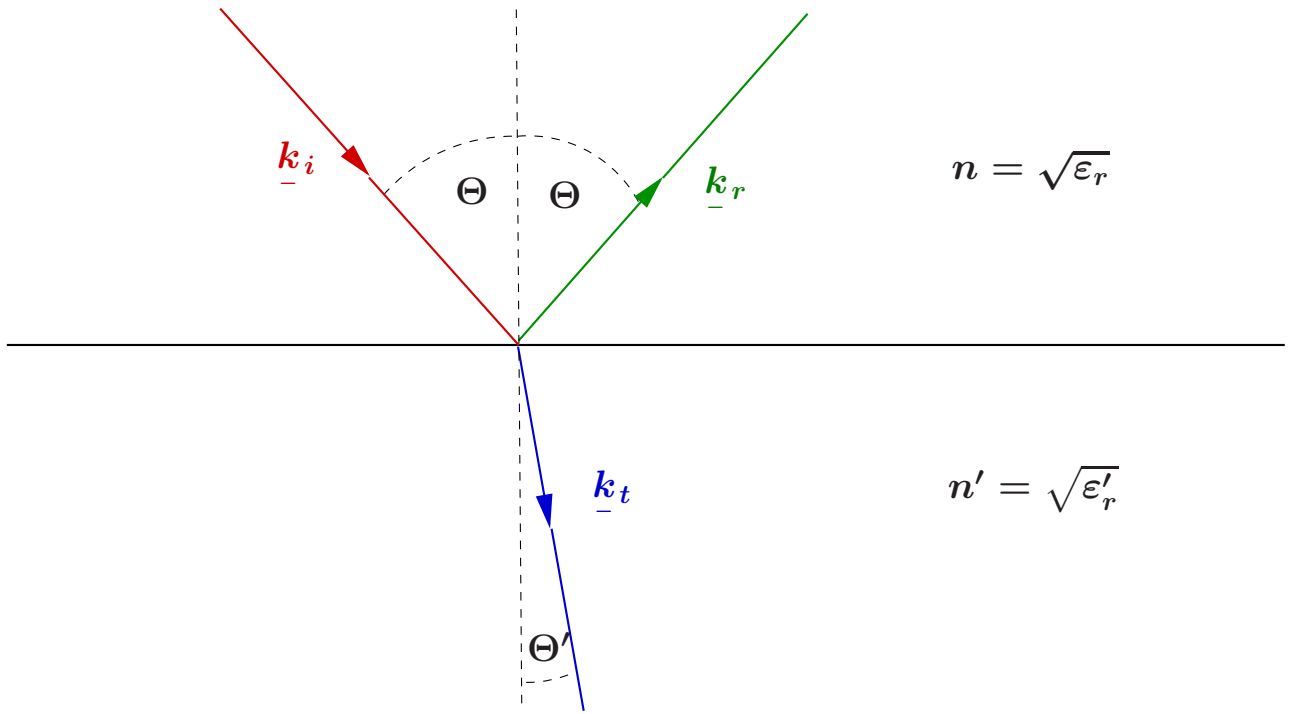
$$n' = \sqrt{n\,n''}\,.$$

For $n = 1$ (air) and $n'' = 1.52$ (glass), one wants to have $n' = \sqrt{nn''} = 1.23$. The closest one gets is cryolite ($Na_3Al\,F_6$) with $n' = 1.31$.

With multi-layer films one can achieve anti-reflective coatings that reflect less than 0.1 % (and high-reflective coatings that reflect more than 99.9 %).

Example: Computer screens

Generalisation to non-normal incidence:



$\underline{k}_i$, $\underline{k}_r$ and $\underline{k}_t$ define "plane of incidence".

Decompose $\underline{E}$ into component in the plane of incidence ($||$) and component perpendicular to the plane of incidence ($\perp$). Then an argument analogous to the above yields the Fresnel equations, shown on the next page. (Nobody expects you to know them by heart!)

$\underline{E} = \underline{E}_\perp$:

$$\frac{E_{t\perp}}{E_{i\perp}} = \frac{2\cos\Theta}{\cos\Theta + \sqrt{\left(\frac{n'}{n}\right)^2 - \sin^2\Theta}}$$

$$\frac{E_{r\perp}}{E_{i\perp}} = \frac{\cos\Theta - \sqrt{\left(\frac{n'}{n}\right)^2 - \sin^2\Theta}}{\cos\Theta + \sqrt{\left(\frac{n'}{n}\right)^2 - \sin^2\Theta}}$$

$\underline{E} = \underline{E}_{||}$:

$$\frac{E_{t||}}{E_{i||}} = \frac{2\,\frac{n'}{n}\cos\Theta}{\left(\frac{n'}{n}\right)^2\cos\Theta + \sqrt{\left(\frac{n'}{n}\right)^2 - \sin^2\Theta}}$$

$$\frac{E_{r||}}{E_{i||}} = \frac{\left(\frac{n'}{n}\right)^2\cos\Theta - \sqrt{\left(\frac{n'}{n}\right)^2 - \sin^2\Theta}}{\left(\frac{n'}{n}\right)^2\cos\Theta + \sqrt{\left(\frac{n'}{n}\right)^2 - \sin^2\Theta}}$$

$E_{t\perp} = 0$ only if $\Theta = \pi/2$.

$E_{r\perp} = 0$ only if $n' = n$.

$E_{t||} = 0$ only if $\Theta = \pi/2$.

$E_{r||} \stackrel{?}{=} 0$ leads to "Brewster angle":

Define

$$\boxed{\tan\Theta_B = \frac{n'}{n}, \qquad \text{Brewster angle } \Theta_B}$$

With Snells' law,

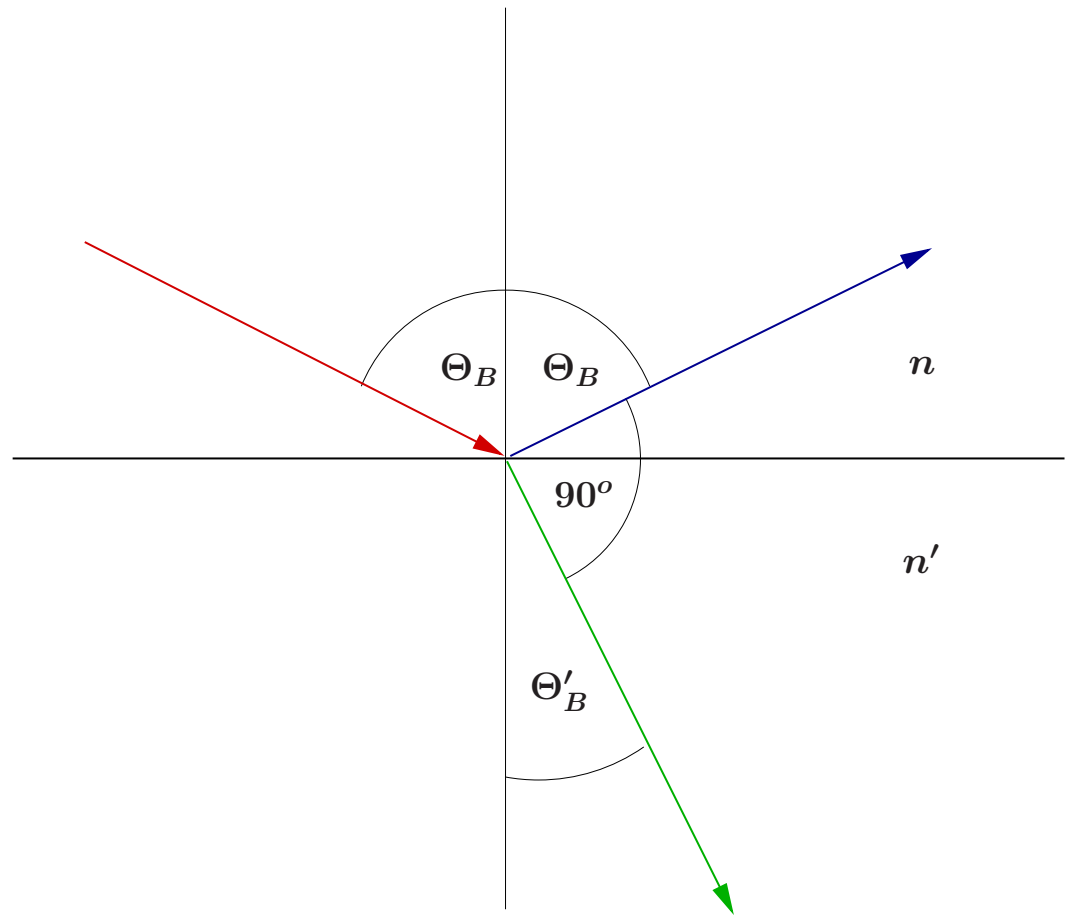
$$\sin\Theta' = \frac{n}{n'}\sin\Theta$$

we find for $\Theta = \Theta_B$,

$$\sin\Theta'_B = \frac{\sin\Theta_B}{\tan\Theta_B} = \cos\Theta_B$$

$$\Longrightarrow \qquad \Theta'_B = \frac{\pi}{2} - \Theta_B\,.$$

Hence, for $\Theta = \Theta_B$ the transmitted ray is perpendicular to the reflected ray.

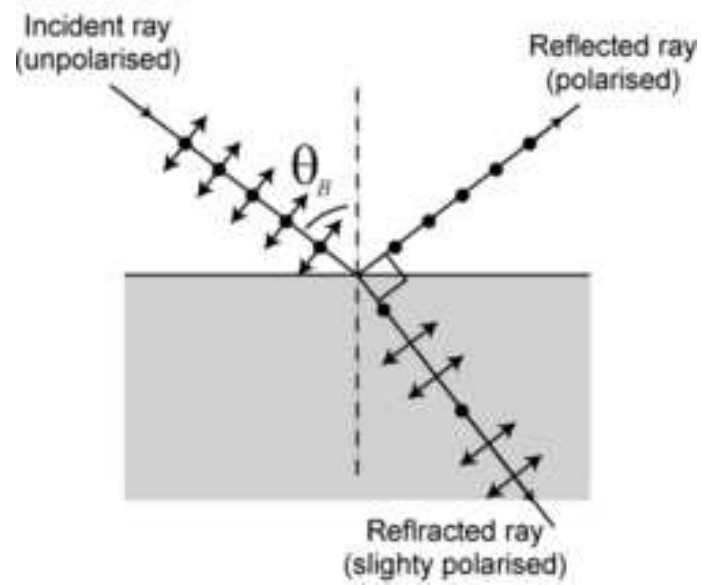


As the definition of the Brewster angle implies $\left(\frac{n'}{n}\right)^2 \cos\Theta_B = \sqrt{\left(\frac{n'}{n}\right)^2 - \sin^2\Theta_B}$, the Fresnel equations yield for $\Theta = \Theta_B$:

$$E_{t\perp} = \frac{2\,n^2}{n^2 + n'^2} E_{i\perp}\,, \quad E_{t||} = \frac{n}{n'} E_{i||}\,, \quad E_{r\perp} = \frac{n^2 - n'^2}{n^2 + n'^2} E_{i\perp}\,, \quad E_{r||} = 0.$$

Thus:

If unpolarised light is incident under the Brewster angle, the reflected light is linearly polarised perpendicular to the plane of incidence. (The transmitted ray is partially polarised.)



picture from

http://content.answers.com/main

Air-glass: $n = 1, \quad n' = 1.5 \quad \Rightarrow \quad \Theta_B \approx 57^o$.

For $\Theta \neq \Theta_B$, the reflected ray is partially polarised.

The following picture was taken with a polarisation filter. Two effects are observable:

- **The intensity over the glass surface varies stronger than when seen with the naked eye (because the degree of polarisation depends on the angle of incidence).**
- **The mirror image in the lower right-hand corner is clearer than when seen with the naked eye (because the polarisation filter was rotated such that the polarised wave reflected in this area was maximally transmitted).**

picture from http://en.wikipedia.org

β) Scattering

Scattering of light can be microscopically explained in the following way:

- An incident electromagnetic wave causes electrons to oscillate around the nucleus to which they are bound.
- These oscillating particles radiate the acquired energy away (dipole radiation).

We speak of Rayleigh scattering if the wave length is much larger than the elongation of the radiating particles, $\lambda >> r_0$,

$$r(t) \;=\; r_0 \cos(\omega t + \alpha)\,, \qquad \omega \;=\; \frac{2\pi c}{\lambda}$$

The radiated power is proportional to the square of the particle's acceleration ("Larmor formula"), $P(t) \;\sim\; |\,\ddot{r}(t)\,|^2 \;=\; \omega^4\, r_0^2 \cos^2(\omega t + \alpha)$

For Rayleigh scattering, the radiation from various sources adds up incoherently, so the averaged radiated power is $\overline{P} \;\sim\; \omega^4 r_0^2$. Hence:

> For Rayleigh scattering, the averaged radiated power is proportional to $\omega^4 \sim \lambda^{-4}$.

E.g., 400-nm light (violet) is scattered approximately 10 times as much as 700-nm light (red).

That is why the sky is blue and the sun is yellow (or even red, if viewed through a thick layer of atmosphere). If viewed from space, the sun is perfectly white.

For scattering by bigger particles, the situation is different. If the elongation of the particles is bigger than λ, the radiation adds up coherently within an area, transverse to the ray, of order λ^2. It turns out that then $\overline{P} \sim \omega^4\lambda^4 = (2\pi c)^4$, i.e., $\overline{P}$ is independent of the frequency.
This is why

- milk is white (droplets of fat suspended in water);
- clouds are white (droplets of water suspended in air).

What has all this to do with polarisation?

- An oscillating dipole radiates in all directions except the direction of oscillation.
- Electromagnetic waves are transverse.

Both observations together imply the following.

By scattering, unpolarised light becomes partly polarised:

Along the incident direction, the scattered light is unpolarised.

In the transverse directions, the longitudinal components are suppressed.

At intermediate angles, there is a partial polarisation effect.

The following picture of a clear sky was taken with a polarisation filter. The polarisation of skylight is maximal if we observe in directions that make a right angle with the direction to the sun. Bees use this effect for orientation.

picture from www.weather-photography.com

γ) Absorption

For microwaves, polarisation by absorption can be realised in the following way:

Consider a grid with grid stepping $g << \lambda$. (This can be realised with wires only for microwaves, not for light.)

Then $\underline{E}$ along the wire is absorbed by the grid, so incoming unpolarised waves become polarised across the grid.

This is contrary to intuition. The reason is that the electrons in the wire can freely move along the wire, so $\underline{E}-$fields pointing in the direction along the wire are absorbed.

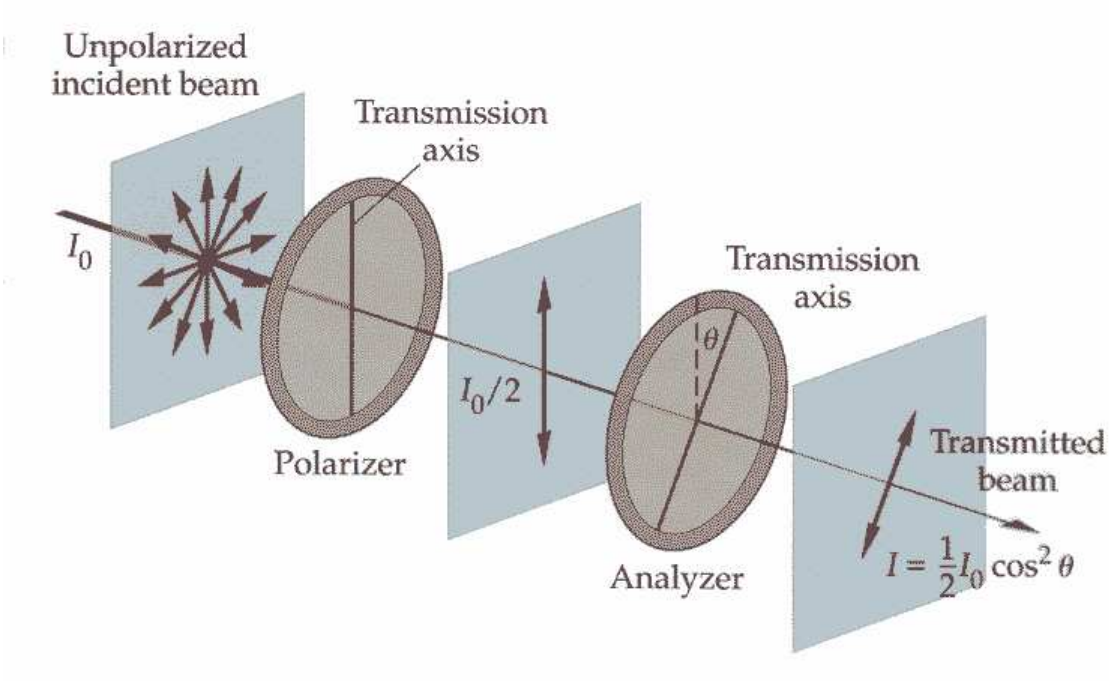
For optical wave lengths, a somewhat similar effect is produced by the molecular structure of certain materials. Such materials are called dichroic.

"Dichroism" means polarisation by selective absorption.

Polarisation filters (e.g. for photography) and also some types of sunglassses are made from dichroic materials.

The first polarisation filters, consisting of polymer films with embedded crystals, were produced in 1933 by E. H. Land under the brand name "Polaroid".

Two polarisation filters ("polariser" and "analyser"):



picture from http://courses.dce.harvard.edu/~physe1b

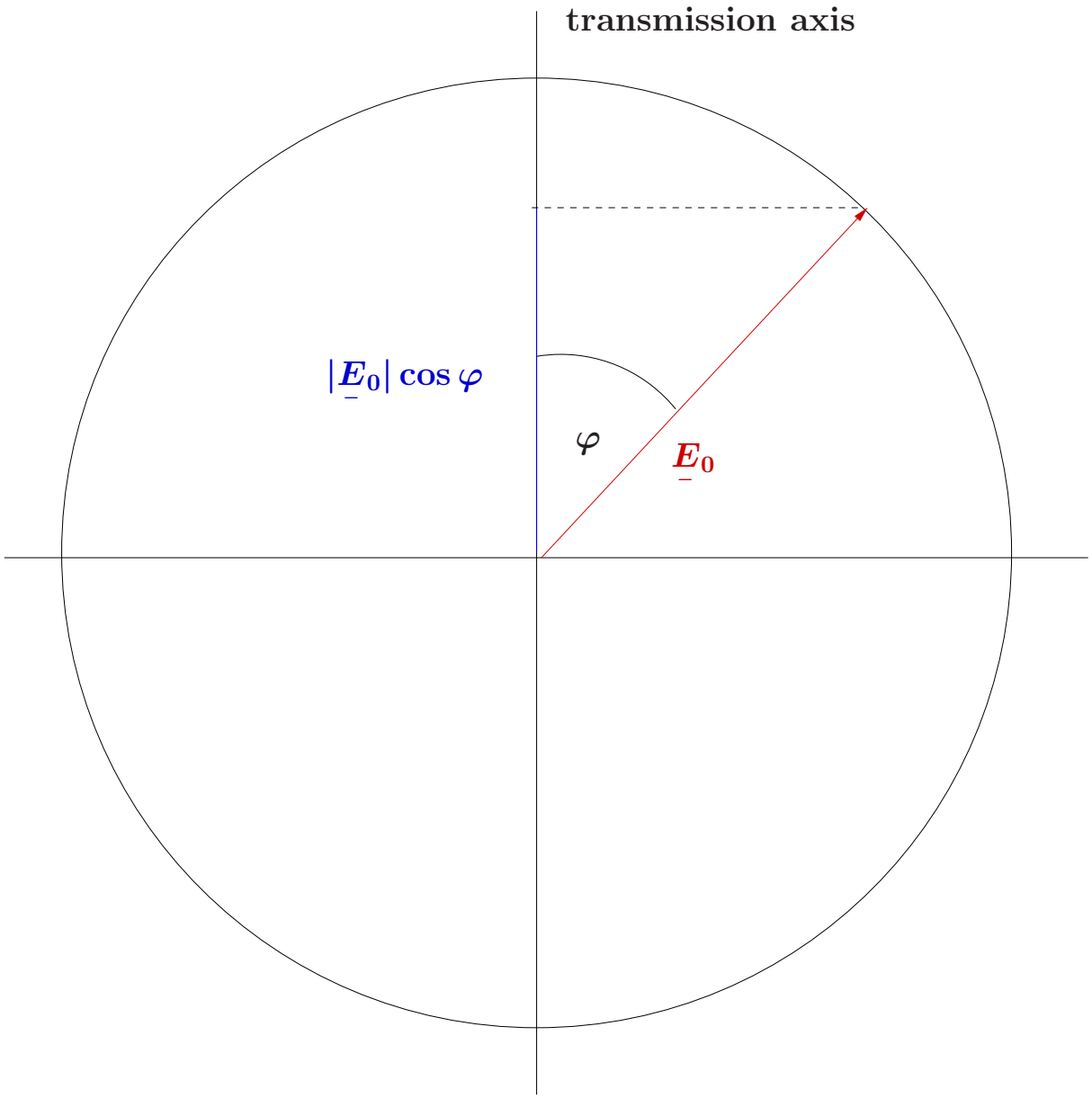
Intensity of transmitted light depends on Θ:

$$\text{Malus' law}: \quad I = \frac{1}{2} I_0 \cos^2\Theta$$

polariser and analyser parallel: maximal intensity

polariser and analyser across: darkness

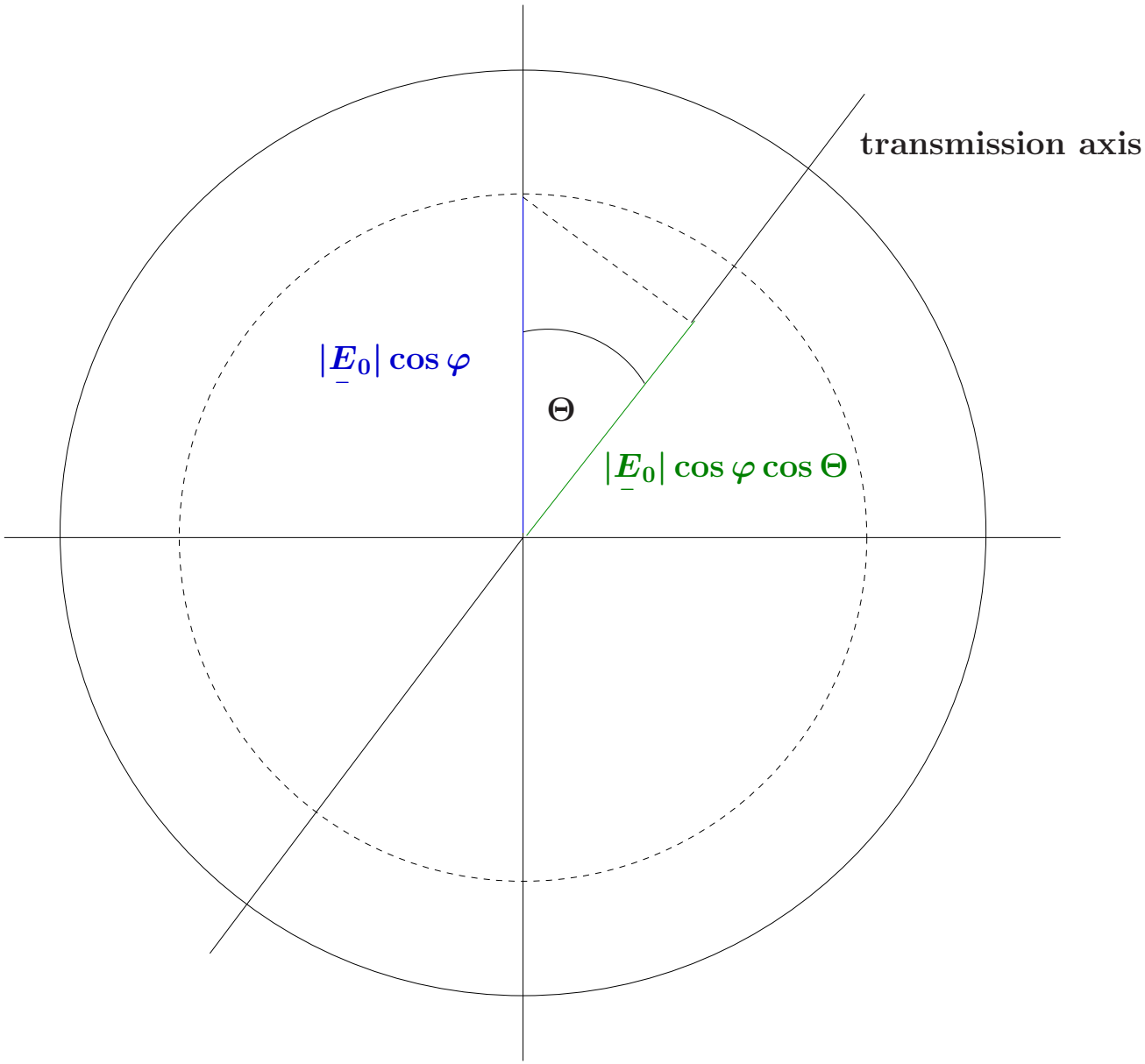
Derivation of Malus' law:



Assume incoming light is polarised at angle φ with respect to first polariser, with amplitude vector $\underline{E}_0$.

(We will later average over all possible values of φ.)

Amplitude after first polariser: $|\underline{E}_0| \cos\varphi$.



Angle between the transmission axes of the two polarisers: Θ

Amplitude after second polariser: $|\underline{E}_0| \cos\varphi \cos\Theta$.

Intensity after second polariser: $I_\varphi = I_0 \cos^2\varphi \cos^2\Theta$

Averaging over all angles φ (recall: average of $\cos^2$ function is $1/2$) gives Malus' law: $I = \frac{1}{2} I_0 \cos^2\Theta$

δ) Birefringence (= double refraction)

Anisotropic materials (crystals, polymers, ...) have different indices of refraction, depending on the polarisation direction of the incoming ray.

Here we consider only the simplest case, a uniaxial crystal. In such a crystal, there is a prefered axis; the laws of light propagation are rotationally symmetric around this axis.

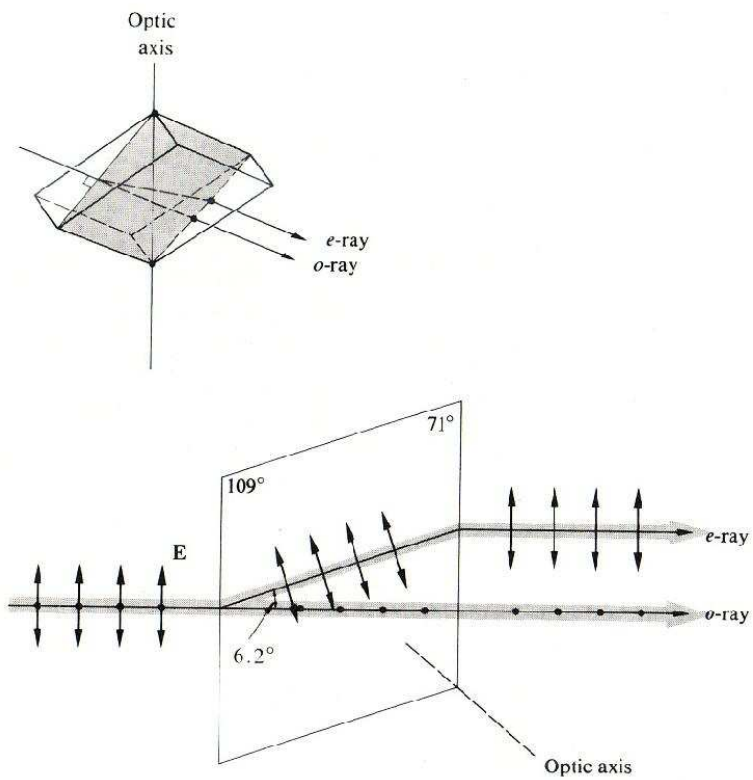
- If the wave vector $\underline{k}$ (i.e., the ray direction) is parallel to the axis, the velocity of light is independent of the polarisation direction.
- If the wave vector is perpendicular to the axis, the velocity of light is different for $\underline{E}$ parallel to the axis and $\underline{E}$ perpendicular to the axis.

Thus, for light that propagates non-parallel to the axis, there are two different indices of refraction. The difference is maximal for rays perpendicular to the axis.

As a consequence, on refraction an incoming ray splits into two.

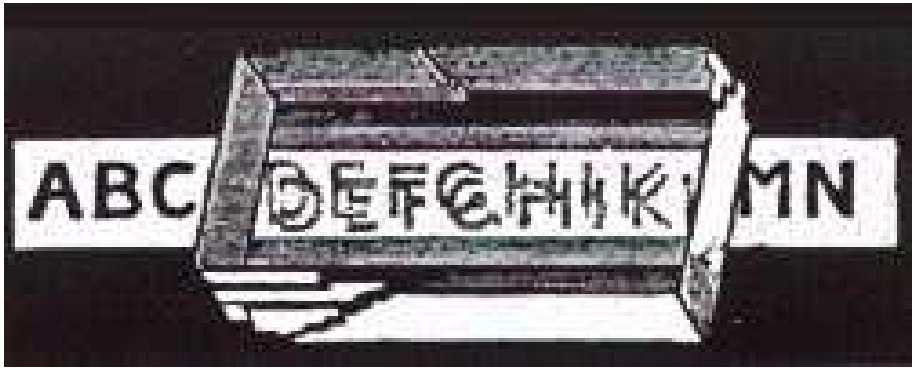
One ray is called the ordinary ray (o-ray, $E-$vector perpendicular to optical axis), the other one the extraordinary ray (e-ray).

Calcite (Iceland spar): $n_{||} = 1.486$, $n_{\perp} = 1.658$



picture from http://www.star.le.ac.uk/∼rw

Double refraction of calcite:



picture from http://www.casdn.neu.edu/~geology/department/staff/colgan

Double refraction was first described in 1669 by Danish scientist Erasmus Bartholinus.

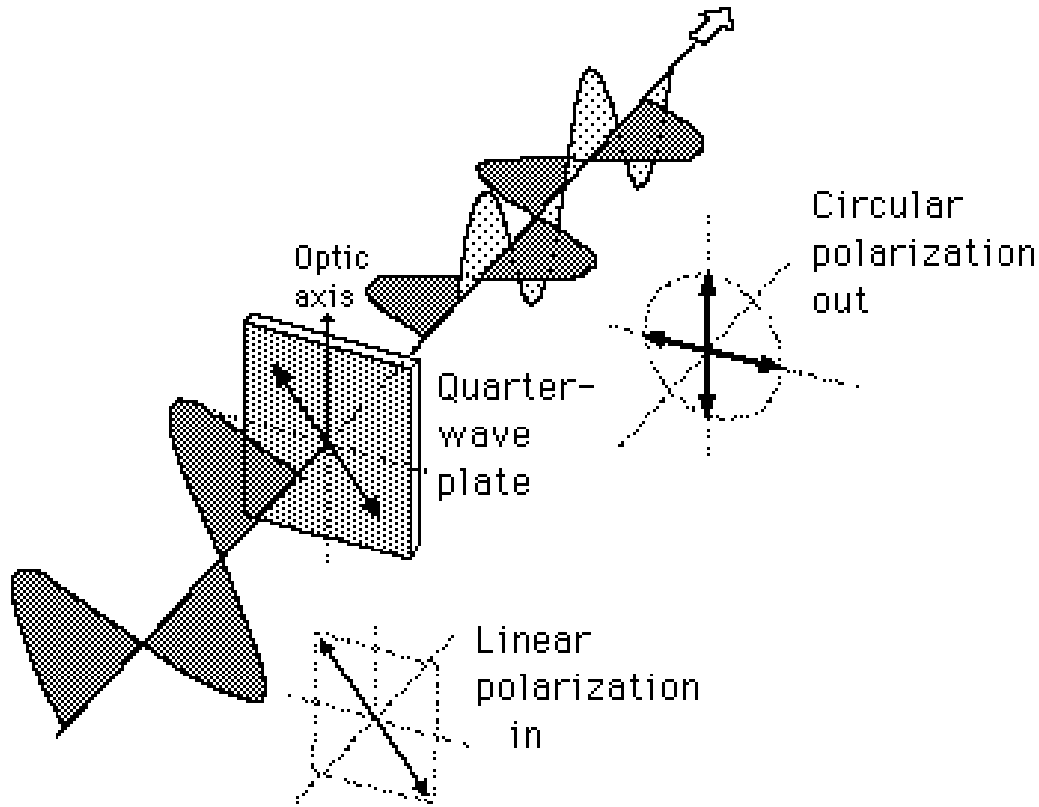
Phase retarder: Cut birefringent material (e.g. mica) parallel to optical axis. Choose thickness d such that phase difference of ordinary and extraordinary ray is a specific fraction of 2π (e.g., $\pi/2$ or π). Then the difference in optical path length is a specific fraction of the vacuum wave length (e.g., $\lambda/4$ or $\lambda/2$).

Quarter wave plate ($\lambda/4$ plate):

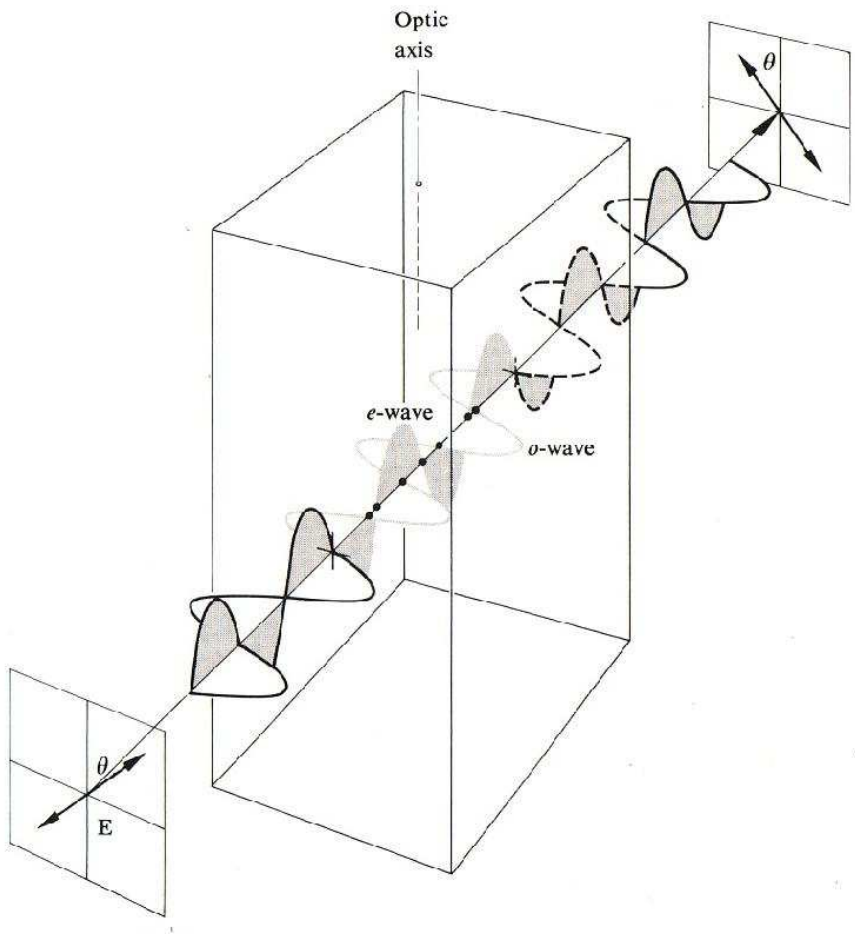
$$|n_{||} - n_{\perp}|\, d = \lambda/4$$

In: linear polarisation under 45° with respect to the optical axis

Out: circular polarisation



Half wave plate ($\lambda/2$ plate):



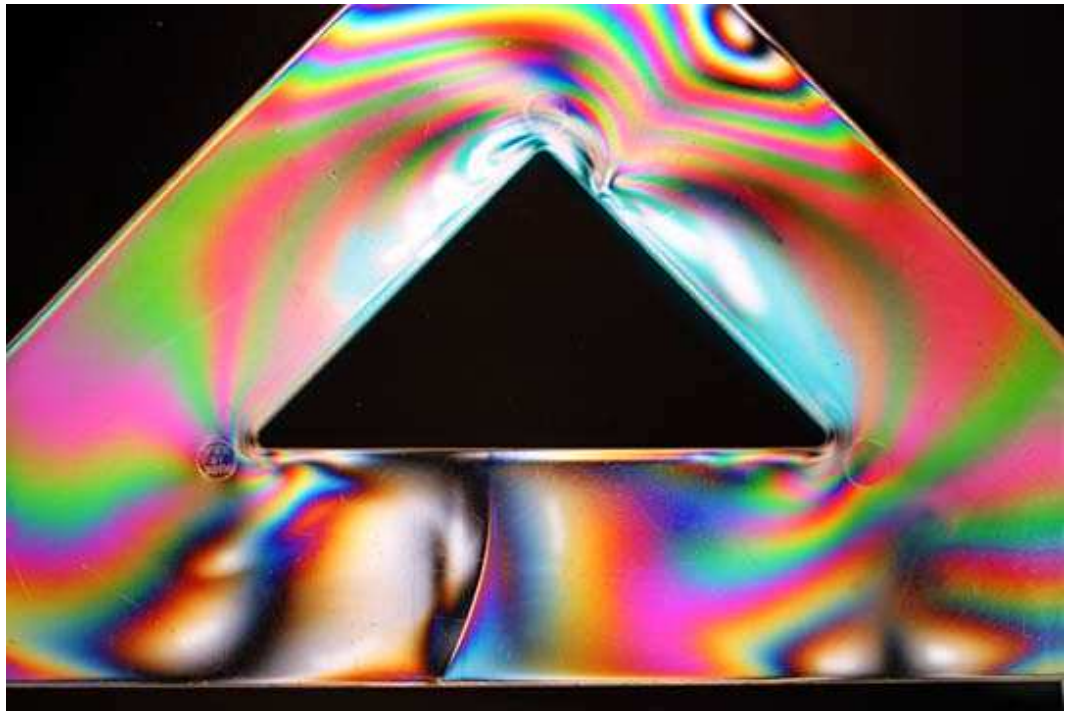
$$|n_{||} - n_{\perp}|\, d = \lambda/2$$

In: linear polarisation

Out: linear polarisation

polarisation direction rotated by $\pi/2$

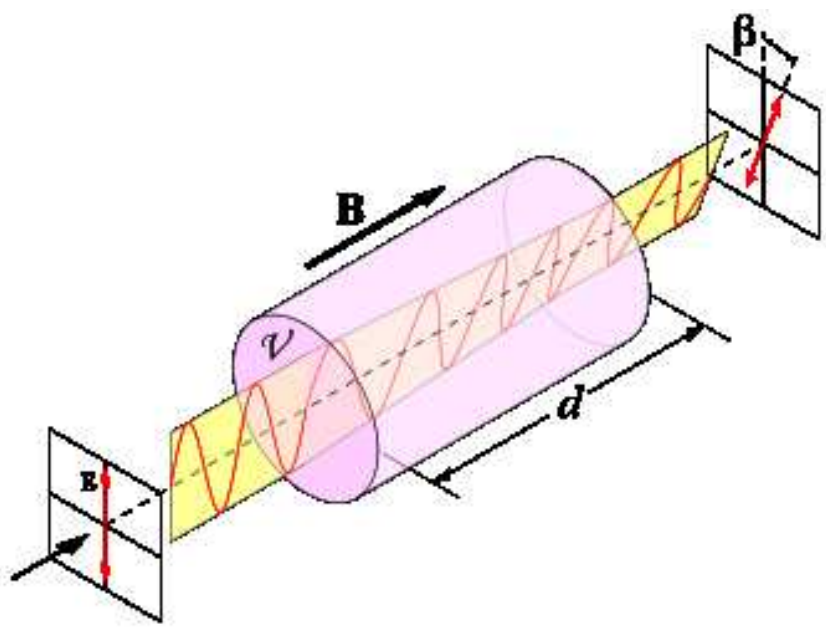
Some isotropic materials become anisotropic and thus birefringent under stress. This phenomenon is called photoelasticity. The degree of birefringence is proportional to the strain, so the strain in the material becomes visible in a polariscope, if white light is used. The colors are produced by interference of ordinary and extraordinary beams, with the optical path length proportional to the strain, so different strains give maximal intensity for different wave lengths.



picture from http://www.rit.edu/∼andpph

One speaks of circular birefringence if the propagation velocity of circularly polarised waves is different for right-handed and left-handed waves.

Some media become circular birefringent if a magnetic field is applied ("Faraday effect"). The Faraday effect is observed, e.g., in the interstellar medium. Linearly polarised light remains linearly polarised, but the polarisation direction is rotated.



picture from http://en.wikipedia.org

VI. Lasers

(cf. Pedrotti and Pedrotti, Chapters 21 [and 22, 23])

Lasers cannot be understood in terms of classical (ray or waves) optics; the quantum theory of light is required.

Light quanta were introduced by A. Einstein in 1905 (Nobel Prize 1922).

The name photon for the light quantum was introduced by G. Lewis in 1926.

$$\text{energy of photon} \quad \longleftrightarrow \quad \text{frequency of light}$$

$$E = h\nu$$

$$h = \text{Planck's constant} = 6.626 \cdot 10^{-34}\, J \cdot s \;\; (\text{has dimension energy} \times \text{time})$$

Visible light: $\nu \approx 10^{15}\,\text{Hz}$

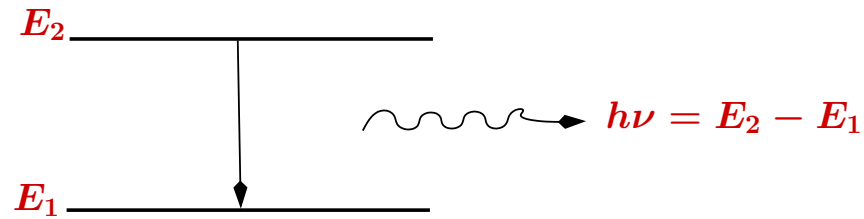
Why is a medium (e.g., a hot gas) emitting light ?

- Electrons in atoms occupy different energy levels.
- In equilibrium, higher energy levels are less populated than lower ones.
- If an electron jumps from a higher level to a lower level, the energy difference is emitted as a photon.

There are two different ways of emission, spontaneous and stimulated:

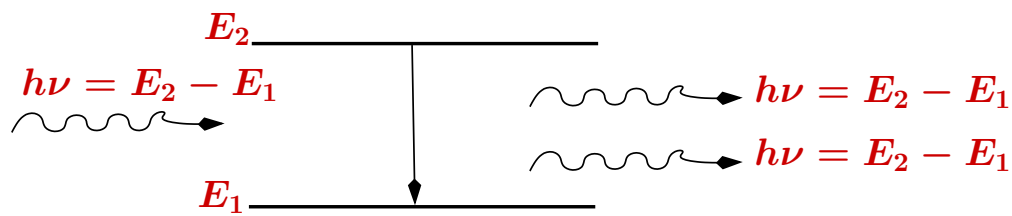
Emission may occur spontaneously, with a certain probability; the resulting phases vary randomly, so the emitted light is incoherent.

spontaneous emission



If light with frequency $\nu = \frac{1}{h}(E_2 - E_1)$ comes in, emission of a photon with the same frequency is stimulated. Phase and direction coincide with that of the incoming light. In this way one can produce coherent light.

stimulated emission



laser = light amplification by stimulated emission of radiation

- requires population inversion ("pumping"), i.e., a higher level must be more populated than a lower level;
- amplification is reached by placing the medium in a "resonator" ("optical cavity") such that the light is constantly reflected back and forth through the medium.

Historical remarks:

- **Stimulated emission was discovered by A. Einstein in 1916.**

  > **"A splendid light has dawned on me about the absorption and emission of radiation."**
  >
  > **A. Einstein in a letter to Michele Besso (1916)**

- **The first realization was with microwaves, called**
  maser = microwave amplification by stimulated emission of radiation
  by C. Townes in 1954 (Nobel Prize with N. Basow and A. Prokhorov in 1964)

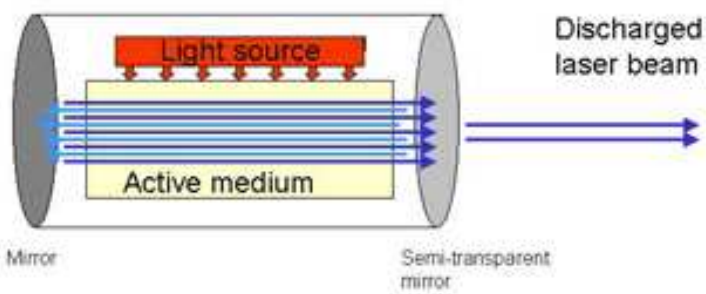
- **The basic ideas of how to transfer this to optical wave lengths are due to**
  C. Townes and A. Schawlow (1957): "optical maser"
  G. Gould (1957): "laser"

- **The first lasers that were actually built were by**
  T. Maiman (1960): ruby laser
  A. Javan (1960): HeNe laser

- **In the 1980s, G. Gould after a long legal battle got the most important patents for the invention of the laser.**

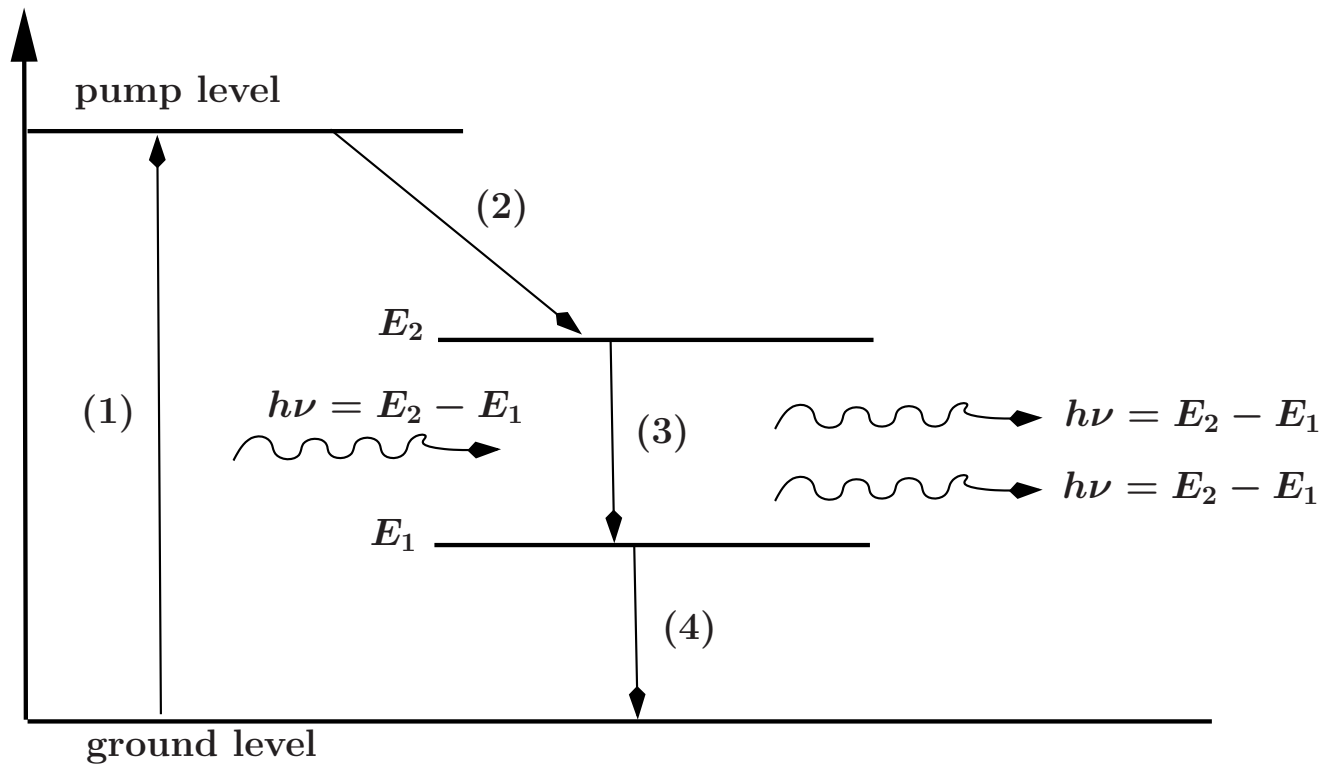
Basic elements of a laser:

- **medium**: determines wave length (ruby: 693 nm, neon: 633 nm, CO_2: 11 μm)
  - gas
  - solid (crystal doped with ions, semi-conductor, ...)
  - liquid (dye, ...)
- **pumping**: energy source that produces population inversion
  - optical
  - electrical
  - thermal
  - nuclear
- **resonator**: plane or (slightly) curved mirrors
  Resonance: $d = m\lambda/2$, m integer (recall Fabry-Perot)



picture from http://www.merck.de/servlet/PB

Laser energy cycle (for 4-level system):



(1) Electrons are pumped to the pump level.

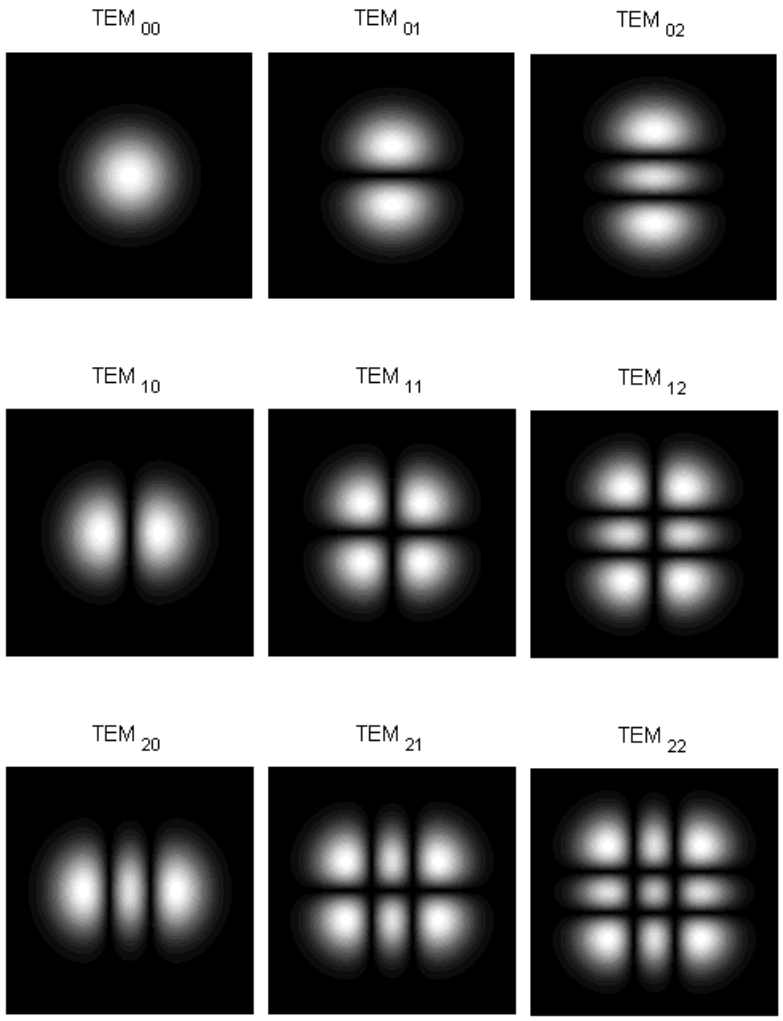
(2) The pump level quickly ($\approx 10^{-8}s$) decays.

(3) The metastable (life time $\approx 10^{-3}s$) level at E_2 decays by stimulated emission; this is the process one is interested in.

(4) E_1 quickly decays to the ground level, so that the laser process can go on.

Intensity of laser beam over beam cross section (“burning pattern”)

Ground mode and higher order modes excited in cavity



picture from http://de.wikipedia.org

Characteristics of laser light:

- **Monochromaticity:** The frequency is determined by $E_2 - E_1$. The spread of the radiated frequency in a laser is up to 10^{-7} times smaller than for the same transition occurring spontaneously.

- **Coherence:** For some lasers, the coherence length is more than 1000 km.

- **Directionality:** Laser light is highly collimated and can be forced into an area with radius of order λ.

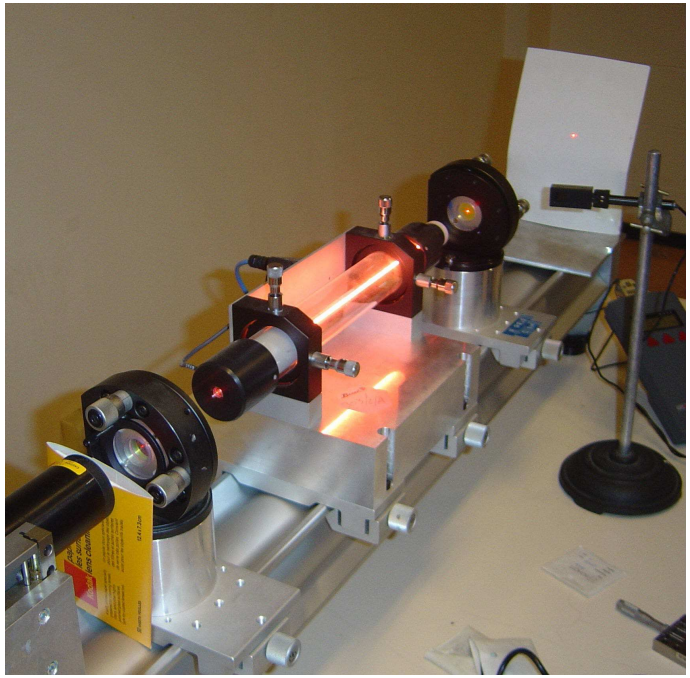
- **Intensity:** The Nova Laser at the Lawrence Livermore Laboratory (Berkeley, California) could send pulses of $\approx 10^5$J in ≈ 2.5 ns. At a star in 50 light years distance, in this time 8 photons per square meter would be received. (From our sun, in the same time only 0.00025 photons per square meter would be received.)

Laser light can be (linearly) polarised by inserting "Brewster windows".

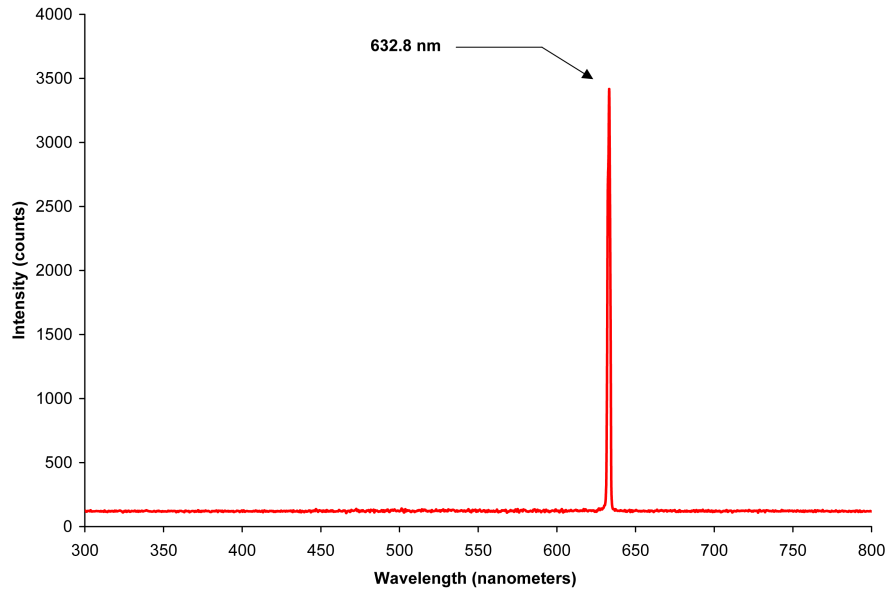
Some laser types:

Gas lasers:

- HeNe laser: $\lambda = 633$ nm (red), power up to 50 mW, beam diameter > 0.5 mm



picture from http://en.wikipedia.org



picture from http://en.wikipedia.org

- CO_2 laser

  $\lambda = 10.6\ \mu$m (infrared), power up to 100 W, beam diameter > 3 mm

  Used e.g. for industrial cutting

Solid state lasers:

- **Ruby laser**

  Ruby = aluminium oxide crystal doped with chromium

  $\lambda = 694$ nm (red), energy up to 100 J per pulse, beam diameter > 2.5 cm

- **Nd-YAG laser**

  YAG = Yttrium – Aluminium – Garnet

  $Y_3Al_5O_{12}$ crystal doped with Neodymium

  $\lambda = 1.1$ μm (infrared), power up to 600 W, beam diameter > 6 mm

Semiconductor lasers (laser diodes):

- **Gallium-arsenide laser diode**

  $\lambda = 780 - 900$ nm (red), power up to 40 mW, beam diverges rapidly

  Laser diodes are used in CD players, laser pointers (should be limited to 5 mW), laser printers, ...

Dye lasers:

Discovered accidentally by F. Schäfer in 1966, wave length tunable, used e.g. in laser spectroscopy

Free electron laser:

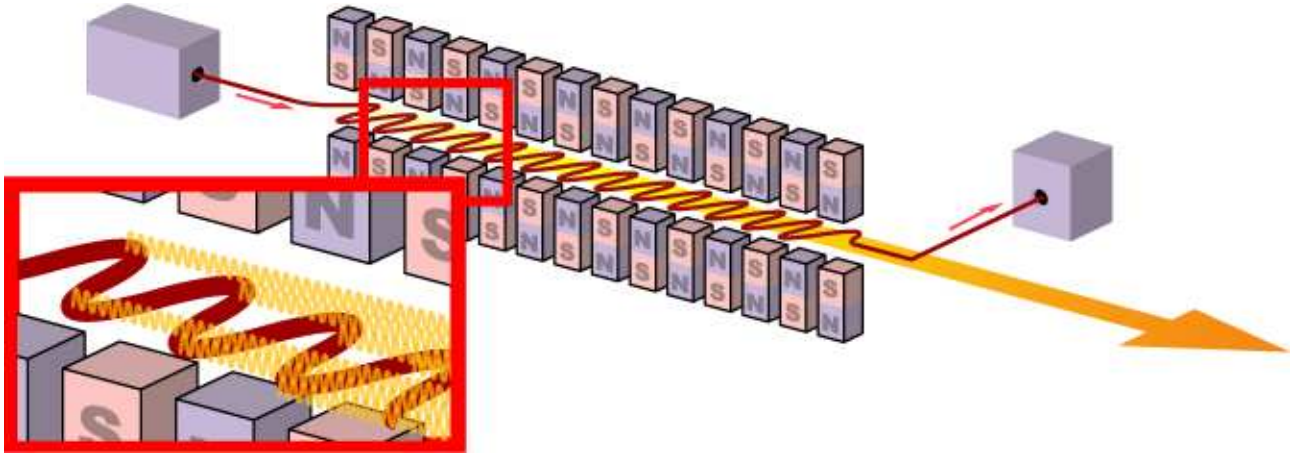
Completely different from all other laser types discussed above

Uses beam of free electrons (not bound in atoms)

Electrons radiate when accelerated very intense beam

Wave length is tunable

Used e.g. for studies in solid state physics and in medicine



picture from http://en.wikipedia.org

Some applications of lasers (very incomplete list):

- **cutting, welding, etc. of materials**
- **printing**
- **medicine (e.g. eye surgery)**
- **science, e.g.**
  - **laser spectroscopy**
  - **laser gyroscope for measuring rotations**
  - **lunar laser ranging**
  - **gravitational wave detectors**
- **CD and DVD players**
- **laserscanners**
- **military applications**
- **arts and entertainment**
- **holography**

VII. Holography

This part was not presented in the lectures, for lack of time, and will not be the subject of any exam questions

(cf. Pedrotti and Pedrotti, Chapter 13)

Ordinary photograph:

- 2-dimensional
- no information about depths
- information only about intensity coming from various directions, not about phase

Hologram:

- information on intensity and phase
- allows to reconstruct the complete wave field near the observer's eye
- gives spatial (3-dimensional) impression of object

holos (greek): the whole, everything graphein (greek): to write

holography: writing the whole (information about an object to a photographic plate)

- Principle idea of holography:
  D. Gábor (1947), Nobel Prize 1971
- Realization with laser and invention of off-axis technique:
  E. Leith and J. Upatnieks (1962)

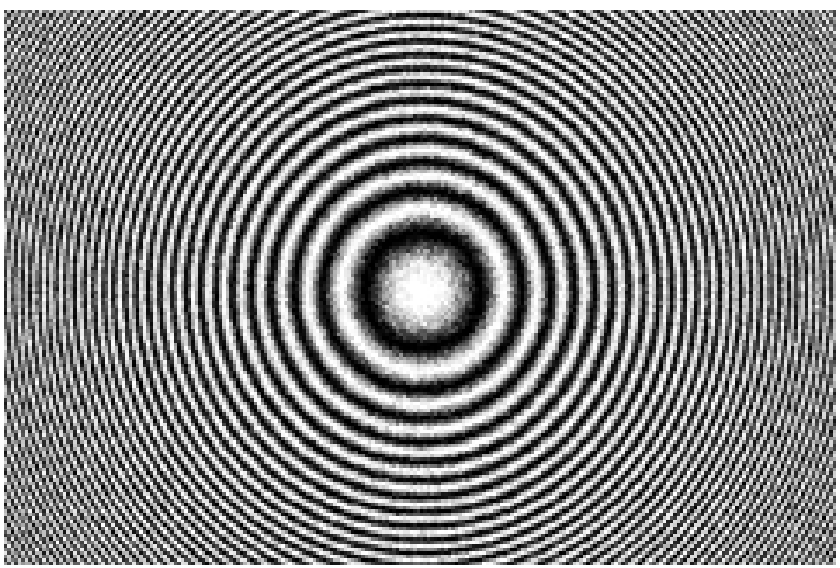
Hologram of point source:

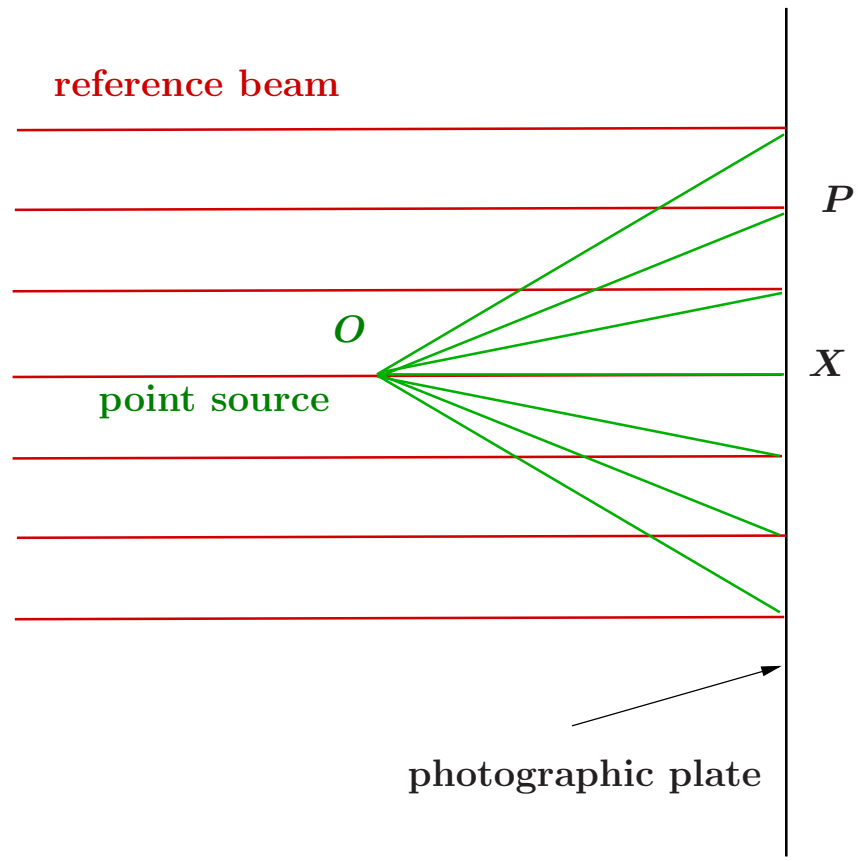
Reference beam and light from point source at O give interference pattern on photograhic plate

Difference in optical path length at point P : $\overline{OP} - \overline{OX}$

Interference maxima if
$\overline{OP} - \overline{OX} = m\lambda$
with $m = 0, \pm 1, \pm 2, \ldots$

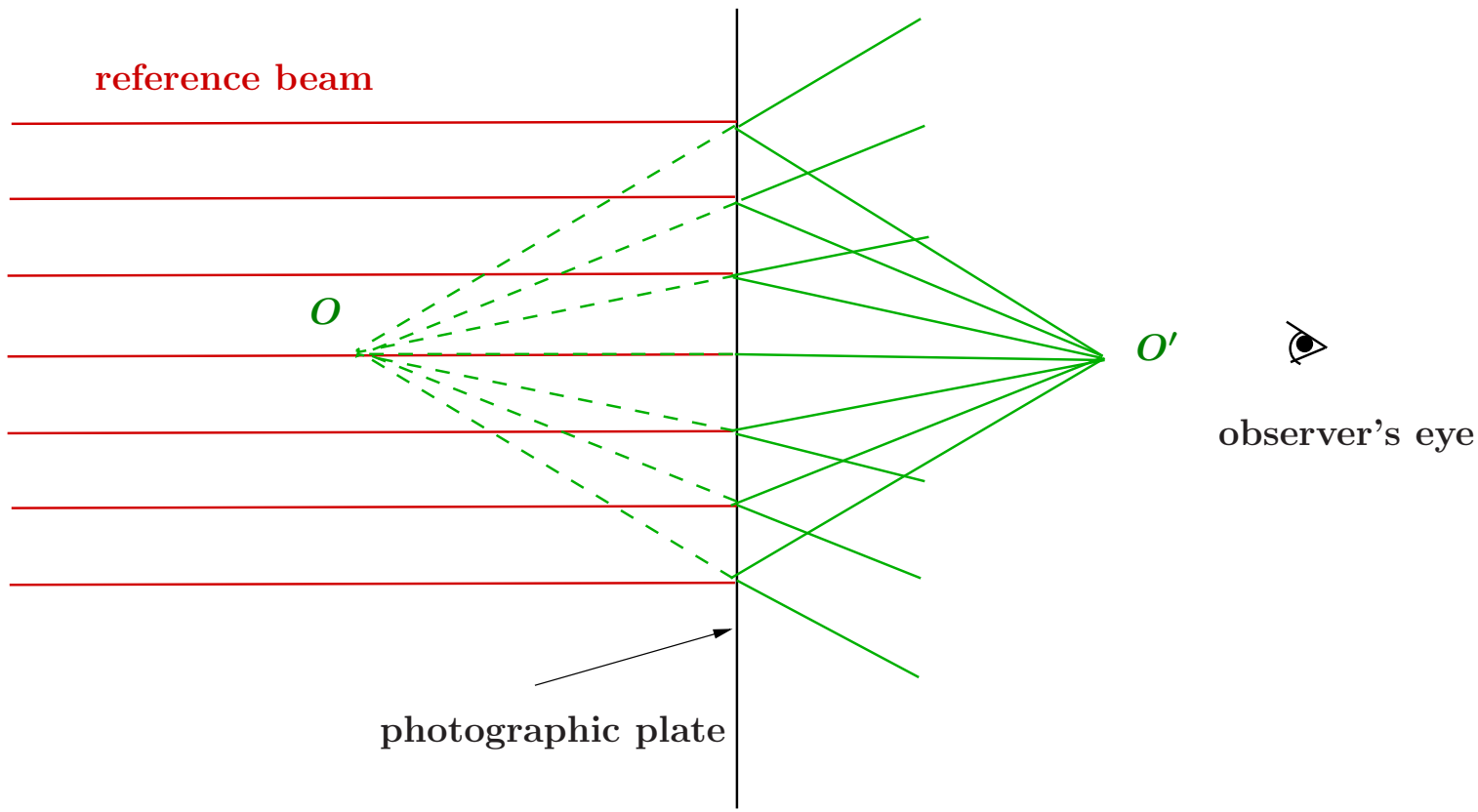
So one gets a circular interference pattern ("Gábor zone plate"):





If the photographic plate has been developed, it is again brought in the same position with respect to the reference beam, but now without the point source at O.

Diffraction pattern of the Gábor zone plate reproduces a virtual image of the point source at O and a real image at the mirror-symmetric point O'.

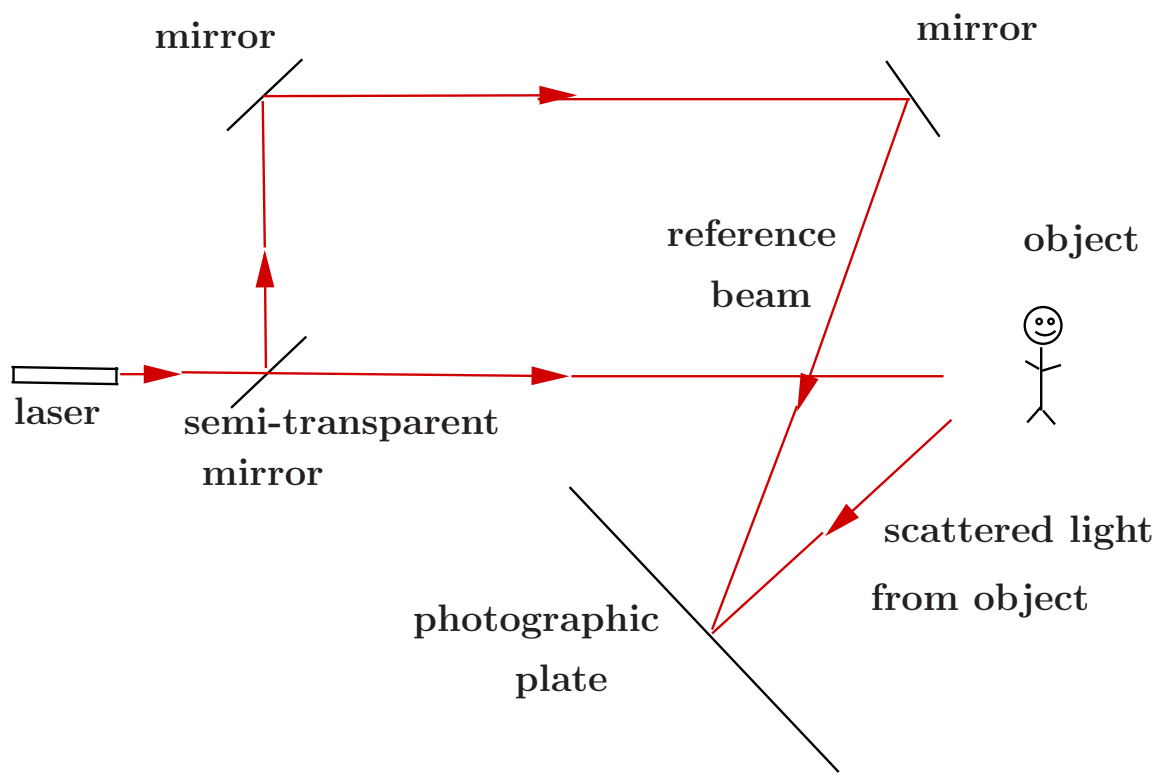


- The observer's eye sees a virtual image at O. That is what one wants.
- The real image at O' is unwanted.
- There are also unwanted higher order images, corresponding to higher order diffraction maxima.

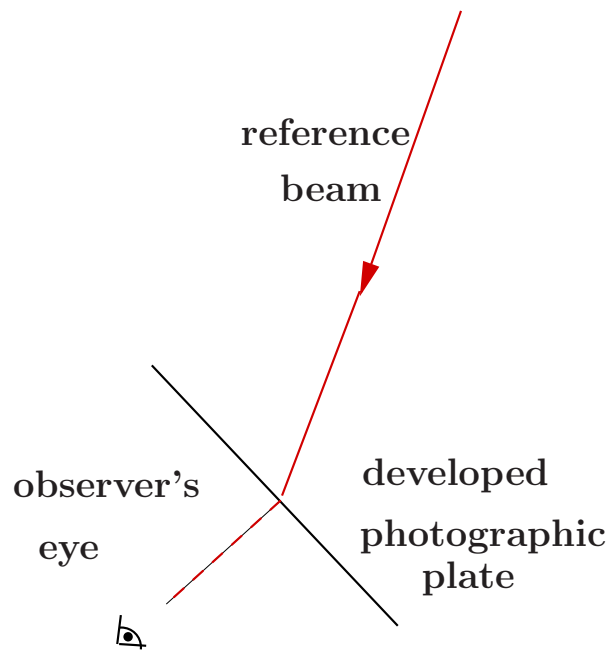
One uses the socalled off-axis technique to separate the virtual image from the unwanted images:

- reference beam comes in at an angle
- real and virtual images are separated in direction

Hologram of an extended 3-dimensional scene is the superposition of the holograms of its points.



Reconstruction:



Observer's eye sees 3-dimensional virtual image of object

Making hologram:

- **Photographic plate perpendicular to line to object**
- **Reference beam makes angle $\alpha \neq 0$ with this line**

Viewing hologram:

- **View line perpendicular to photographic plate**
- **Reference beam comes in at angle α**

picture from http://www.holography.dragonseye.com

Characteristic features of holograms:

- **The reconstructed image is 3-dimensional.**
- **If observer position is changed, objects move relative to each other, like when viewing the real scene.**
- **Every part of a hologram contains information about the whole scene.**
- **Making a hologram requires a coherent light source (laser).**
- **Viewing a hologram with an ordinary light source (light bulb) gives slightly blurred image; if white light is used, image has colored fringes.**
- **If reference beam for viewing has other λ than reference beam for making the hologram, the object is displaced.**
- **White light holograms that reproduce objects in natural colors can be produced with thick emulsions ("volume holograms").**

Modifications:

- **Instead of transmisson holograms, one can also produce reflection holograms.**
- **Other recording materials than photographic plates are used, e.g. photothermoplastics and polymers.**
- **Dynamic holograms, rather than static ones, have been produced.**

Some applications:

- **Arts and entertainment**
- **Measurements of small displacements**
- **Microscopy**
- **Data storage**
- **Security feature on credit cards etc.**

Notes on exam preparation:

The exam will be in May 2010. There will be a revision seminar two or three weeks before the exam. Here are some suggestions for exam preparation.

- **Look at past exam papers. Concentrate on the years 2007/2008/2009 (before that the course was taught by someone else who partly emphasised other things). You will see that several questions frequently reoccur, with minor modifications.**
- **Work through the lecture notes. I know that it is a lot of material. I tried to highlight particularly relevant things by red frames.**
- **Revise the worksheet questions. About 40 % of the exam questions will be similar to worksheet questions. The rest are socalled bookwork problems, like "State Fermat's Principle" or "Sketch a Fabry-Perot interferometer".**