INTERFERENCE

A wave travelling in the ***z*** direction (for example) satisfies the differential wave equation.
For a light wave:

$$\frac{\partial^2 E}{\partial z^2} = \frac{1}{v^2}\frac{\partial^2 E}{\partial t^2}$$

One property of this equation is that for any set of waves:

$E_1(z,t)$, $E_2(z,t)$, $E_2(z,t)$...... $E_n(z,t)$,

Any **linear combination** (sum) of these waves will satisfy the wave equation so that if:

$$\boldsymbol{E} = \sum_{i=1}^{n} c_i \boldsymbol{E}_i(z,t)$$

then, **E** is a solution to the equation if all the $\boldsymbol{E_i}$s are solutions:

$$\frac{\partial^2}{\partial z^2}\left(\sum_{i=1}^{n} c_i \mathrm{E}_\mathrm{i}(z,t)\right) = \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\left(\sum_{i=1}^{n} c_i \mathrm{E}_\mathrm{i}(z,t)\right)$$

This suggests that if a number of waves overlap in time and space, then the resultant disturbance at any point in the medium is just the sum of the different waves.

This is known as **interference.**

For light waves this just means **"add the E fields".**
However we can never measure the resultant E field as it **oscillates too rapidly**.
What we see is the **resultant Intensity**.

Therefore in order to calculate interference effects we add the E fields and then use this to calculate the resultant intensity.

Consider two plane waves:

$$\boldsymbol{E}_1(r,t) = \boldsymbol{E}_{01} Cos(\boldsymbol{k}_1 \bullet \boldsymbol{r} - \omega t + \varepsilon_1)$$
$$\boldsymbol{E}_2(r,t) = \boldsymbol{E}_{02} Cos(\boldsymbol{k}_2 \bullet \boldsymbol{r} - \omega t + \varepsilon_2)$$

Remember $\mathbf{k_1}$ and $\mathbf{k_2}$ are vectors.

If these waves combine, the resultant field is:

$$\boldsymbol{E} = \boldsymbol{E_1} + \boldsymbol{E_2}$$

The intensity is defined as:

$$I = \varepsilon v \langle \boldsymbol{E}^2 \rangle$$

(Chapter 3 Hecht) where <$\mathbf{E}^2$> means the value of **E averaged over time.**

ε: permittivity
v: wave velocity

$$I = \varepsilon v \langle \boldsymbol{E} \bullet \boldsymbol{E} \rangle$$

$$\boldsymbol{E}^2 = (\boldsymbol{E}_1 + \boldsymbol{E}_2) \bullet (\boldsymbol{E}_1 + \boldsymbol{E}_2)$$

$$\boldsymbol{E}^2 = \boldsymbol{E}_1^2 + \boldsymbol{E}_2^2 + 2\boldsymbol{E}_1 \bullet \boldsymbol{E}_2$$

Hence the Intensity is given by:

$$I = I_1 + I_2 + I_{12}$$

$$I_1 = \varepsilon v \left\langle \mathrm{E}_1^2 \right\rangle$$

$$I_2 = \varepsilon v \left\langle \mathrm{E}_2^2 \right\rangle$$

$$I_{12} = 2\varepsilon v \left\langle \mathrm{E}_1 \bullet \mathrm{E}_2 \right\rangle$$

Note that the intensity is not just the sum of the intensities.
There is another term, *I_{12}*, the **interference term**.

$$\boldsymbol{E_1}(r,t) = \boldsymbol{E_{01}}Cos(\boldsymbol{k_1} \bullet \boldsymbol{r} - \omega t + \varepsilon_1)$$
$$\boldsymbol{E_2}(r,t) = \boldsymbol{E_{02}}Cos(\boldsymbol{k_2} \bullet \boldsymbol{r} - \omega t + \varepsilon_2)$$

Hence

$$I_{12} = 2\varepsilon v \boldsymbol{E_{01}} \bullet \boldsymbol{E_{02}} \times$$
$$\left\langle Cos\left(\boldsymbol{k_1} \bullet \boldsymbol{r} - \omega t + \varepsilon_1\right) Cos\left(\boldsymbol{k_2} \bullet \boldsymbol{r} - \omega t + \varepsilon_2\right)\right\rangle$$

We are averaging over time so want to *separate time dependent parts from the rest*. So, we can rewrite the average as:

$$\left\langle Cos(k_1 \bullet r - \omega t + \varepsilon_1) Cos(k_2 \bullet r - \omega t + \varepsilon_2)\right\rangle =$$
$$\left\langle \left(Cos(A_1) Cos\omega t + Sin(A_1) Sin\omega t\right) \times\right.$$
$$\left.\left(Cos(A_2) Cos\omega t + Sin(A_2) Sin\omega t\right)\right\rangle$$

Writing $\boldsymbol{k_1} \bullet \boldsymbol{r} + \varepsilon_1 = A_1$

etc

Multiplying out:

$$Cos(A_1)Cos(A_2)\langle Cos^2\omega t\rangle +$$
$$(Cos(A_1)Sin(A_2)+Cos(A_2)Sin(A_1))\langle Sin\omega t Cos\omega t\rangle +$$
$$Sin(A_1)Sin(A_2)\langle Sin^2\omega t\rangle$$

(remember the average is taken over **time.**)
Using trig identities we can show:

$$\langle Sin\omega t Cos\omega t\rangle = \frac{1}{2}\langle Sin2\omega t\rangle$$

But the time average of the sine of any angle is zero so

$$\langle Sin\omega t Cos\omega t\rangle = \frac{1}{2}\langle Sin2\omega t\rangle = 0$$

and

$$\langle Sin^2\omega t\rangle = \langle Cos^2\omega t\rangle = \frac{1}{2}$$

Putting these values in and using one last trig identity (Cos(A-B)=CosACosB+SinASinB) gives:

$$I_{12} = \varepsilon v \boldsymbol{E}_{01} \bullet \boldsymbol{E}_{02} Cos(k_1 \bullet r + \varepsilon_1 - k_2 \bullet r - \varepsilon_2)$$

Here the argument of the Cosine is just the difference between the phases of the two waves.

$$I_{12} = \varepsilon v \mathbf{E}_{01} \bullet \mathbf{E}_{02} Cos(\varphi_1 - \varphi_2)$$

$$I_{12} = \varepsilon v \mathbf{E}_{01} \bullet \mathbf{E}_{02} Cos\delta$$

Where $\delta = \varphi_1 - \varphi_2$
$= k_1 \bullet r + \varepsilon_1 - k_2 \bullet r - \varepsilon_2$

Here, δ, is the **phase difference** between the 2 waves.If the polarisation of the two waves are the same then $\mathbf{E_{01}}$ and $\mathbf{E_{02}}$ point in the same direction.Then:

$$I_{12} = \varepsilon v E_{01} E_{02} Cos\delta$$

$$I_1 = \varepsilon v \left\langle E_1^2 \right\rangle = \varepsilon v E_{01}^2 / 2$$

$$I_2 = \varepsilon v \left\langle E_2^2 \right\rangle = \varepsilon v E_{02}^2 / 2$$

becomes

$$I_{12} = 2\sqrt{I_1 I_2} Cos\delta$$

giving, the total intensity as:

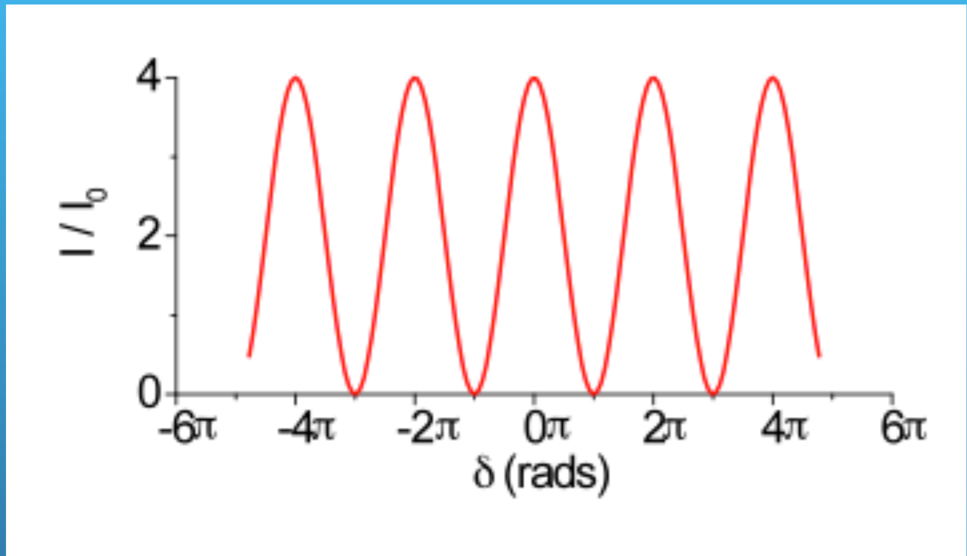
$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} Cos\delta$$

In addition, if, $I_1 = I_2 = I_0$, then

$$I = 2I_0(1 + Cos\delta) = 4I_0 Cos^2 \frac{\delta}{2}$$

The intensity of the resultant wave combination is plotted below as a function of the phase difference between the waves, δ.

$$I = 4I_0 Cos^2\delta / 2$$



It is clear that the resultant intensity depends strongly on δ.
The intensity is a **maximum** when

$\delta = 0, \pm 2\pi, \pm 4\pi, \pm 6\pi$………

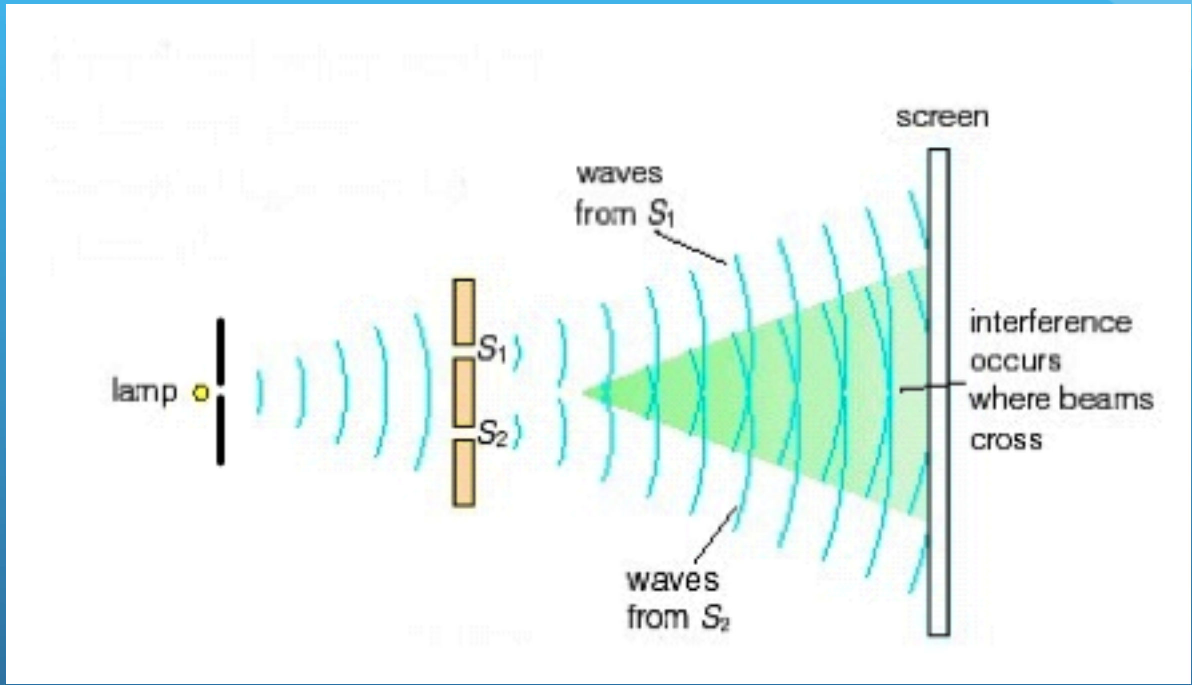
This is **total constructive interference** and occurs when the two waves are **perfectly in phase.** The intensity is a **minimum** when

$\delta = \pm\pi, \pm 3\pi, \pm 5\pi, \pm 7\pi$………

This is **total destructive interference** and occurs when the two waves are **perfectly out of phase**.

Youngs Slits

Lets apply this to Youngs slits. This is the experiment that allowed Thomas Young to demonstrate the wave nature of light.

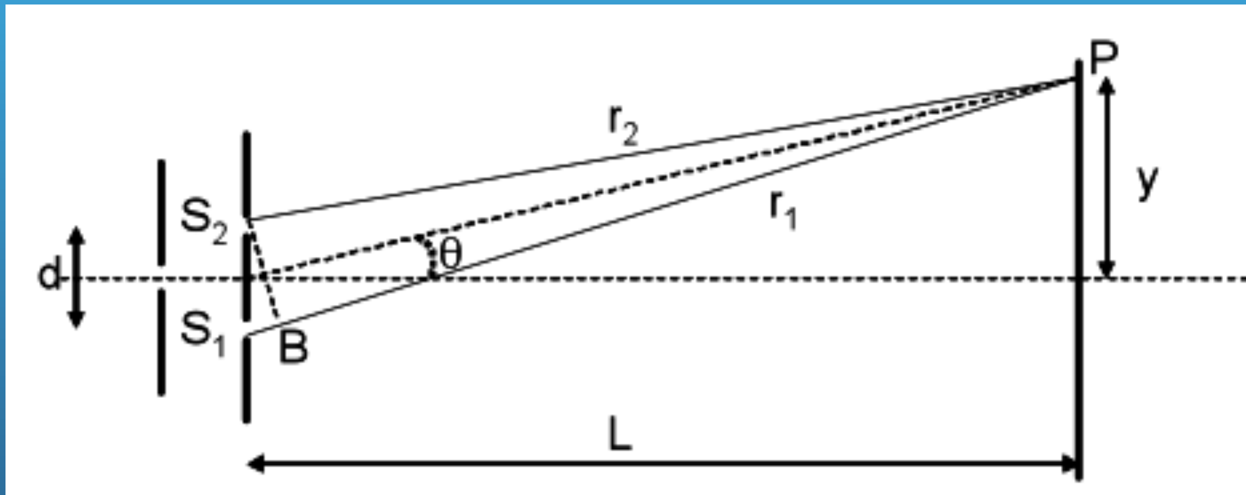


This setup consists of a single slit, illuminated from behind, that acts as a line source of light.The light from this slit then falls on two parallel slits which act as coherent light sources. This just means the light waves emanating from each slit are always in phase as they leave the slit.The light from both slits then falls on a screen some distance away.

What effect does interference have on the intensity of the light falling on the screen?

$$I = 4I_0 Cos^2 \frac{\delta}{2}$$

Remember δ is the phase difference between the waves.
Where does this phase difference come from?
We said the two waves were in phase when they left the slits.



However consider the portions of the wave that recombine on the screen at point P. These have travelled different distances and so now have a relative phase difference. However if one wave travels **exactly one wavelength** further than the other, they are in phase again at that point. This is also true if one wave has travelled **any whole number of wavelengths** further than the other.

The difference in distance travelled is known as the **path difference,** $(r_1 - r_2)$. In the figure the path difference is the distance S_1B. If the screen is very far from the slits, ie, L>>d, then:

$$S_1B = r_1 - r_2 = dSin\theta$$

But if the screen is very far away then θ is small and so

$$\theta \approx Sin\theta \approx Tan\theta = y / L$$

At point P, *if the waves are in phase*, we get **constructive interference**. This then occurs when $(r_1 - r_2) = m\lambda$

where m is an integer.

In this special situation, we call: $\theta = \theta_m$ *and* $y = y_m$

Hence $r_1 - r_2 = dSin\theta_m = y_m d / L = m\lambda$

This means we will get constructive interference at positions

$$y_m = m\lambda \frac{L}{d}$$

where m is an integer, 0, ±1, ±2…

We will see a maximum in the light intensity at these points.

What about the rest of the points on the screen?

We can get a more complete description by looking at the resultant intensity when the waves combine on the screen.

After time t, if a wave has travelled a distance r_1, its phase is

$\varphi_1 = kr_1 - \omega t$

But if a similar wave has travelled a distance r_2 in time t, its phase is

$\varphi_2 = kr_2 - \omega t$

Then the phase difference is given by $\delta = k(r_1 - r_2)$

Here, $r_1 - r_2$ is the **path difference.**

(We can ignore ε_1 and ε_2 as they were the same at the slits so $\varepsilon_2 - \varepsilon_1 = 0$)

If we can work out the **path difference for any point on the screen** we can work out δ, and so the **light intensity at any point on the screen.**

As before, the path difference is

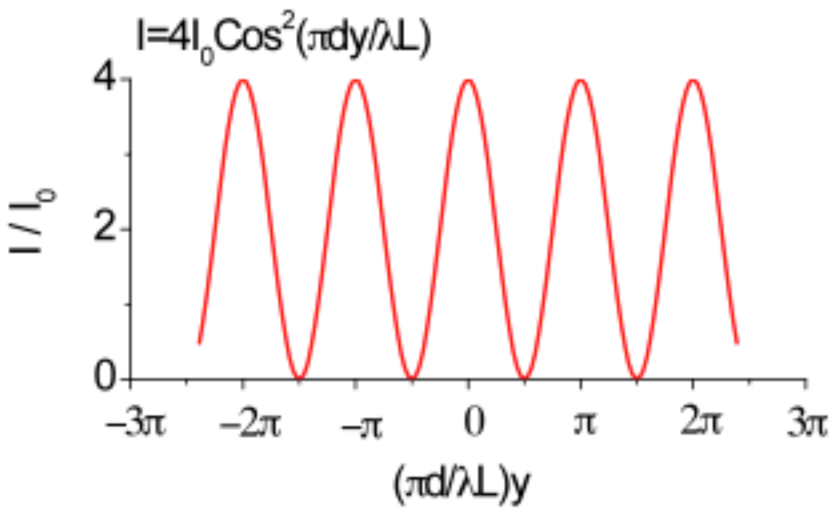
$$S_1B = r_1 - r_2 = dSin\theta$$

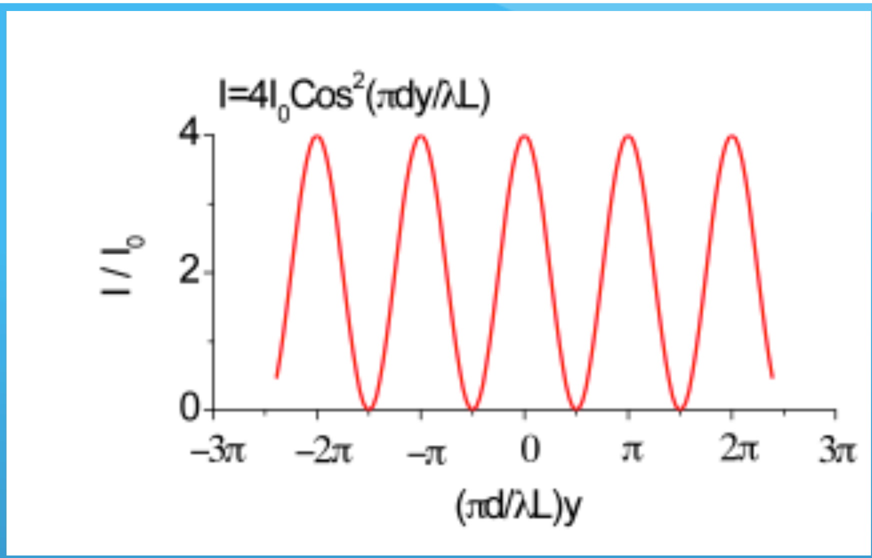
But if the screen is very far away then θ is small and so $r_1 - r_2 = yd/L$ as before. **Hence**

$$\delta = k(r_1 - r_2) = k\frac{yd}{L} = \frac{2\pi\, yd}{\lambda L}$$

$$I = 4I_0 Cos^2\frac{\delta}{2} = 4I_0 Cos^2\frac{\pi\, yd}{\lambda L}$$

Remember, ***y*** is the position of the point P on the screen where we worked out the intensity. This means that the light intensity varies with position, **y**.





The light intensity is a **maximum (constructive interference)** when $\delta = 2m\pi$

or $\dfrac{\pi dy}{\lambda L} = m\pi$ which we can write as $y_m = m\lambda \dfrac{L}{d}$

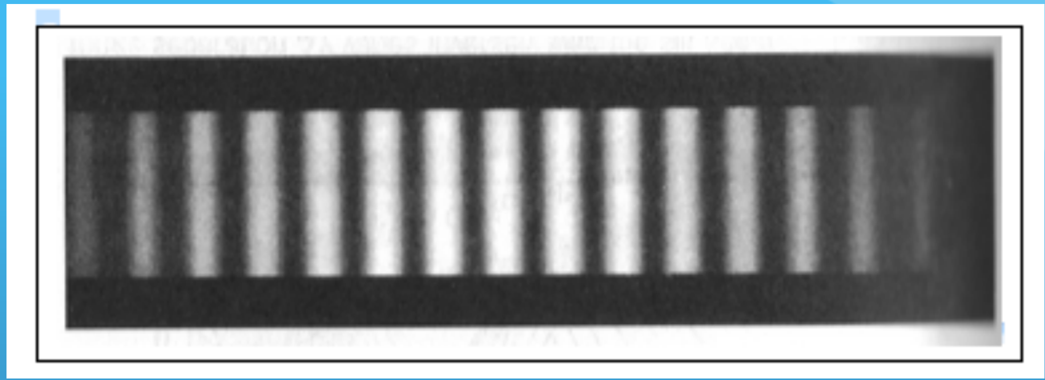
which is what we got before.

The light intensity is a **minimum (destructive interference)** when

$\dfrac{2\pi dy}{\lambda L} = (2m+1)\pi$ or $y_{\min} = \dfrac{(2m+1)}{2}\dfrac{\lambda L}{d}$

This **interference pattern** just looks like a set of bright and dark strips on the screen, with a bright stripe at the centre of the pattern.

These stripes are known as **fringes**.



We can get the spacing between the bright fringes, Δy_m, by noting

$$\Delta y_m = y_{m+1} - y_m = \frac{\lambda L}{d}$$

This does not depend on m, so the fringes are equally spaced.
For red light, λ~600nm. If the slits are 1mm apart and the screen is 5m away, then Δy_m = 3mm. This is easy to see!

Conditions for Interference

What conditions must be satisfied in order to see an interference pattern?

First the waves must be of **very nearly the same frequency**.
Otherwise the **phase difference** $\delta = \varphi_1 - \varphi_2$ will oscillate rapidly. This will result in the **interference term** averaging to zero.
In this case all we will see is the **sum of the individual intensities and no interference.**

If white light interferes with white light we get the red interfering with the red, blue with blue etc and we get a sum of the interference patterns for all frequencies.

However the most important condition for interference is that the **initial phase difference** between the waves, ie $\varepsilon_1 - \varepsilon_2$ is **constant in time.**
This is known as **coherence.**
This is why for example we use an **initial slit** in Young's experiment.

When the light from the source is emitted, it is **in phase**. When it arrives at the two slits both parts of the wave still have a **constant phase relationship**.
Then when it arrives at the screen the only phase differences should be due to **path length differences.**

The problem is that real sources do not emit light as **infinitely long Sine waves**.
A nice sine wave is emitted for approximately 1 ns before the **phase of the wave shifts abruptly**. Thus the phase is constant for only a short period of time.
The light emitted during this time is called a **wavetrain, wavepacket or wavegroup**.
This time can be thought of as a **coherence time**, Δt_c.

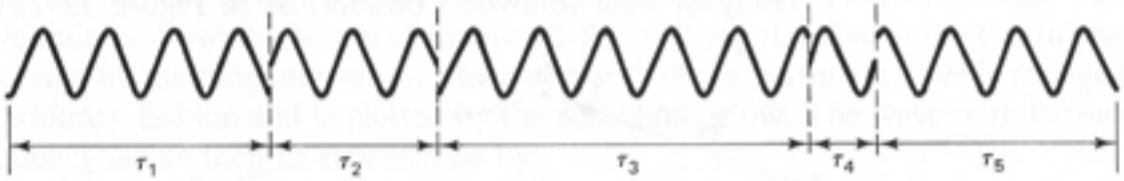


Figure 12-6 Sequence of harmonic wavetrains of varying finite lengths or lifetimes τ. The wavetrain may be characterized by an average lifetime, the coherence time τ_0.

The distance the wave travels in this time is the **coherence length**,
Δl_c and **$\Delta l_c = c\Delta t_c$.**
Both these properties are measures of the temporal coherence of the source.
For a source with $\Delta t_c \sim 1ns$, $\Delta l_c \sim 0.3m$.

Thus we can expect the two waves in the Youngs experiment to be coherent because they come from the same source.

However if we tried to use **two separate sources** there would be **no phase relationship** between them, as the phases of the two waves would change from time to time in an uncorrelated fashion.

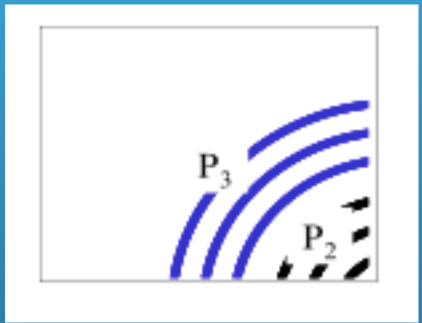
However we have to be careful **even** when using a setup similar to Youngs slits. This is because, if the path length between waves is too long we might have *one wavetrain arriving from one slit at the same time as the next wavetrain arrives from the other slit*.

This will occur if the path length is greater than the coherence length:

$$r_1 - r_2 > \Delta l_c.$$

This may happen at the edge of the screen in Youngs experiment.

In this event, the waves will not have any phase relationship (the phase of each successive wavetrain is completely unrelated). This means the interference term averages to zero and **no interference is observed.**



Consider light emitted from a **point source**. This is a light source with no appreciable size.

Above, one wavetrain is represented by the blue wavefronts and the next wavetrain by the black wave fronts.

Then the phase at P_1 is **correlated perfectly** with the phase at P_2 but **not correlated at all** with the phase at P_3 as this is part of a different wavetrain.

However every point on a given wavefront has the same phase. Thus the phase is the same at points P_3 and P_4.

Thus waves emitted from a point source are **spatially coherent** all along a wavefront but only spatially coherent over **short distances in the propagation direction**.

If the source has appreciable size, however, we can lose spatial coherence. This is because different parts of the source can emit their own wavepackets. These have no phase relationship and so lack spatial coherence. This can result in the loss of interference if the source is too big.

Thus, in order to observe interference we need a source with reasonable temporal and spatial coherence.

Δl_c and Δt_c for typical sources are given below.

This is why **Lasers** are generally used for interference work.

Source	Coherence time	Coherence length
Natural light	~3×10^{-15} s	~900 nm
Mercury Arc lamp	~0.1 ns	~0.03 m
Kr discharge lamp	~1 ns	~0.3 m
Stabilised He-Ne Laser	~1 μs	~ 300m

Interferometers

Youngs apparatus is a type of interferometer. This just a device that produces **interference between two light waves.**
In order for the two waves to be **coherent** they are generally generated from the **same source**.

There are two classes of interferometer that do this, **wavefront splitting and amplitude splitting interferometers**.

Youngs apparatus is a wavefront splitting interferometer as used the two slits to isolate two different parts of the wavefront.

Other wavefront splitting interferometers
Another type of wavefront splitting interferometer is the **Fresnel double mirror**.

Here a coherent source is provided by a slit.
Different parts of the wavefront coming from the slit are then **reflected from two mirrors** at a small angle to each other.
The reflected waves then combine at a screen.

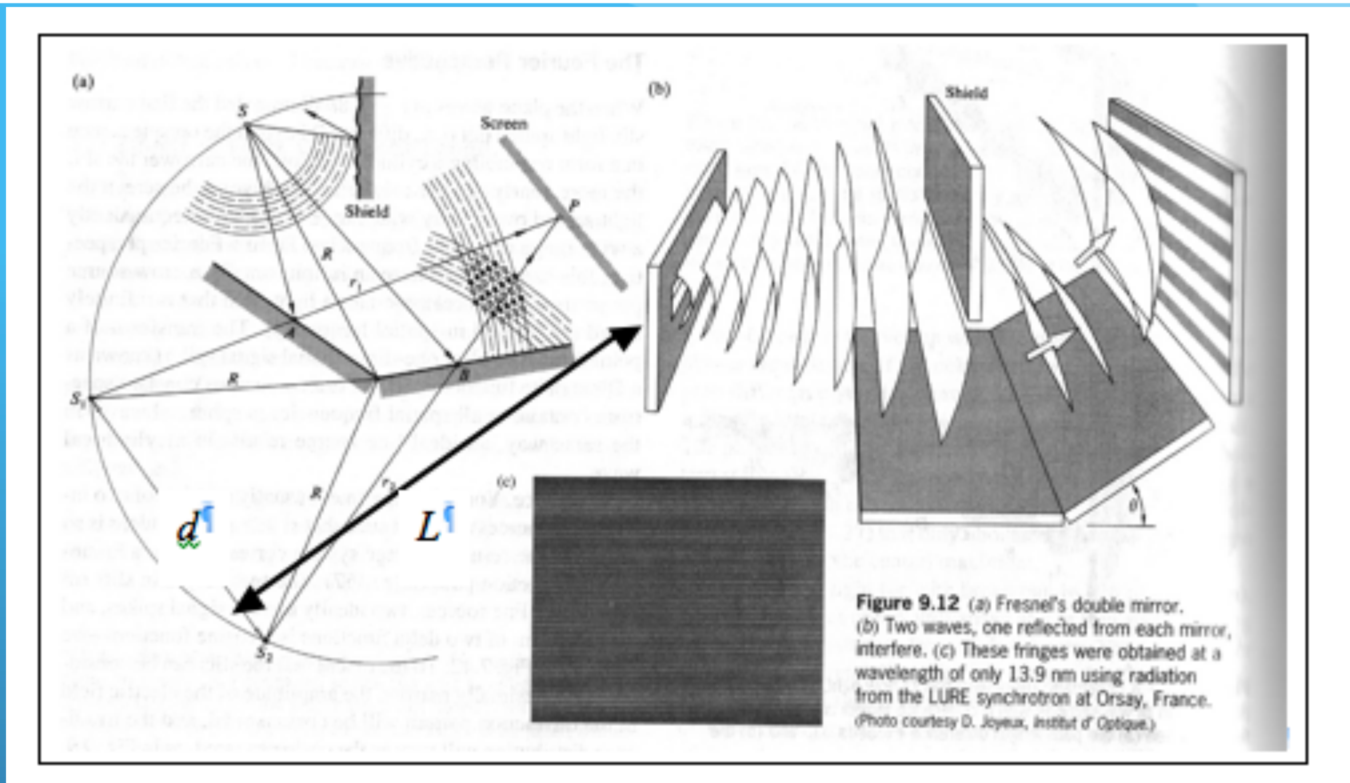


Figure 9.12 (a) Fresnel's double mirror. (b) Two waves, one reflected from each mirror, interfere. (c) These fringes were obtained at a wavelength of only 13.9 nm using radiation from the LURE synchrotron at Orsay, France. (Photo courtesy D. Joyeux, Institut d' Optique.)

The reflected waves appear to be coming from two virtual sources behind the mirrors. These are located at the positions of the image of the slit in each mirror.

Thus this is effectively a combination of **two coherent sources and a screen**, just like Youngs apparatus.

Another type of wavefront splitting interferometer is **Lloyds mirror**.

Here, again, a coherent source is provided by a slit.

Part of the wavefront coming from the slit is then **reflected off a mirror** before travelling to a screen while part travels **directly to the screen**.

The two waves then combine at a screen.

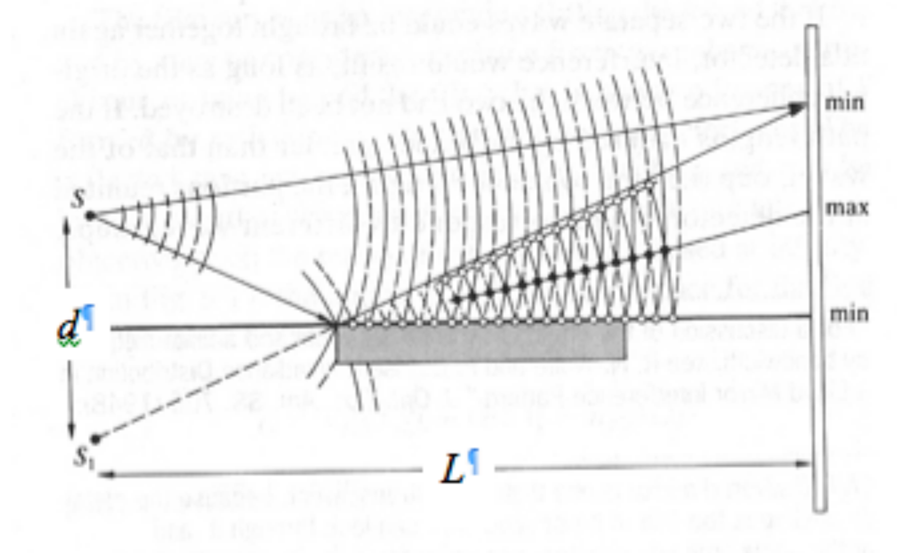


Figure 9.14 Lloyd's mirror.

The waves appear to be coming two sources, a real source (the slit) and a virtual source (the image of the slit in the mirror).

In both cases the interferometer is just a combination of **two coherent sources and a screen.** This means the **path difference** can be calculated in exactly the same way as for Youngs slits.
In all cases the fringes are equally spaced, with a fringe separation of

$$\Delta y_m = y_{m+1} - y_m = \frac{\lambda L}{d}$$

Exactly the same as for Youngs slits.

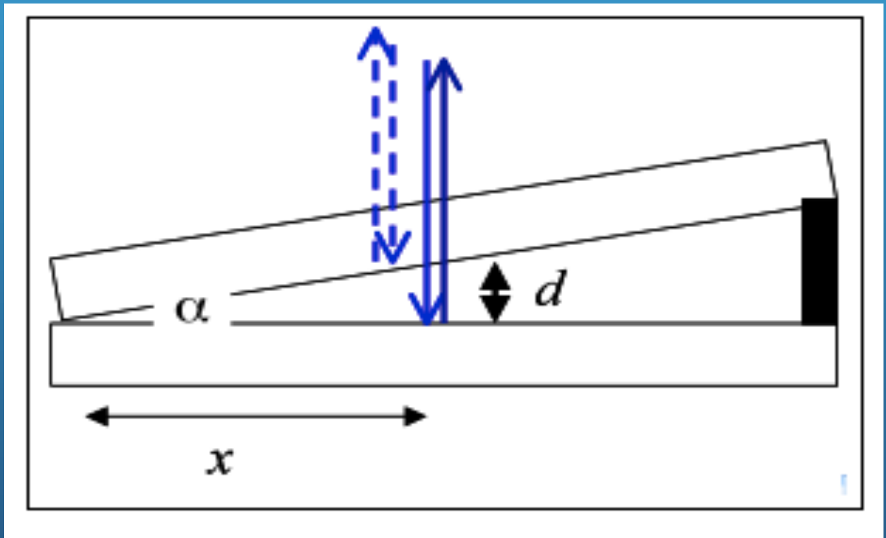
Amplitude splitting interferometers

These are Interferometers based on **reflection and refraction** from dielectrics such as glass.

When light falls on glass, some is reflected and some is transmitted (refracted). Thus the **amplitude (or intensity) is split**. The two portions can then be made to recombine and hence interfere.

Here the solid rays travel further than the dashed rays by a distance 2d. This introduces a phase difference

The simplest type can be made from two glass slides on top of each other with a spacer between them at one end forming a thin air wedge.



Here the solid rays travel further than the dashed rays by a distance **2d**. This introduces a phase difference

The phase difference due to the path difference is given by $\Delta\varphi = kz_1 - kz_2 = 2kd$ where z_1 and z_2 are the distances travelled by the two rays.

However there is another source of phase difference:
phase shift on reflection.
When light reflects of the glass-air interface (dashed ray), it just reflects normally.

However when it reflects from an air-glass interface (solid ray, $n_{glass} > n_{air}$) something different happens:

One can show for a ray, travelling from material 1 and then reflecting of the interface with material 2, the reflected field is related to the incident field by:

$$E_r = \frac{n_1 - n_2}{n_1 + n_2} E_i$$

If $n_1 > n_2$ (ie a glass-air interface), E_r has the same sign as E_i.

However, if $n_1 < n_2$ (ie an air-glass interface), E_r has the opposite sign to E_i.

This is equivalent to the instantaneous reversal of the E field at the instant of reflection, which is the same as *the reflection* ***inducing a phase shift of*** *π rads (180^0).*

We note the above equation also gives the ratio of reflected to incident electric fields. Thus it can be used to work out the reflected intensity.

However, this expression is only strictly true for normal incidence.

This phase shift on reflection means the total phase shift in our wedge is given by

$$\Delta\varphi = 2kd \pm \pi$$

Constructive interference occurs when the two reflected waves are in phase:

$$2m\pi = 2kd \pm \pi$$

Rearranging and noting that $k=2\pi/\lambda$ gives

$$2d = \lambda(m \pm 1/2)$$

The position of the fringe can be obtained by noting that

$$d = x\alpha$$ giving $$x = \frac{\lambda(m \pm 1/2)}{2\alpha}$$

The fringe spacing is $$\Delta x = x_{m+1} - x_m = \frac{\lambda}{2\alpha}$$

The Michelson Interferometer

This is, historically, one of the most important types of interferometer.
It employs a **beam splitter**, M_0 (half silvered mirror) to divide the incident wave into two parts travelling in different directions.
The two parts then reflect of **two mirrors, M_1 and M_2**, before being recombined on a **screen** (after beam 2 has been redirected by the beam splitter).

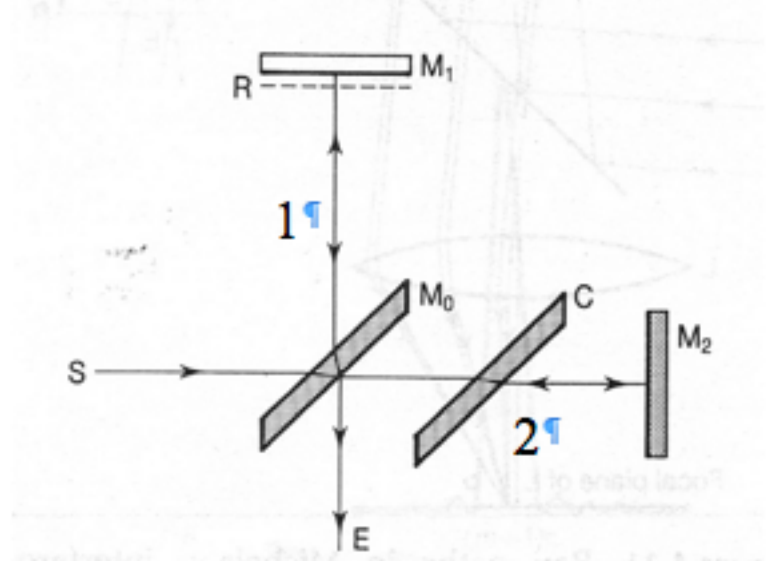


Figure 4.10 Michelson interferometer

Interference will occur if the two beams have travelled **different distances** (path difference).

Note that beam 1 passes through the glass in the beam splitter (M_0) three times, while beam 2 only passes through it once. Thus a **compensator** is inserted in the path of beam 2 to ensure both beams have traversed the same thickness of glass.

The nature of the path distance can be seen by noting the position of plane **R**, which is the reflection of M_2 in M_0.
The path difference is controlled by the distance between R and M_1.

If light from a **point source** falls on the beam splitter, the light rays have a **range of angles of incidence**, depending on where exactly they are incident.

We will consider one angle of incidence and work out the path difference between real rays **reflecting from M_1 and virtual rays reflecting from R.**

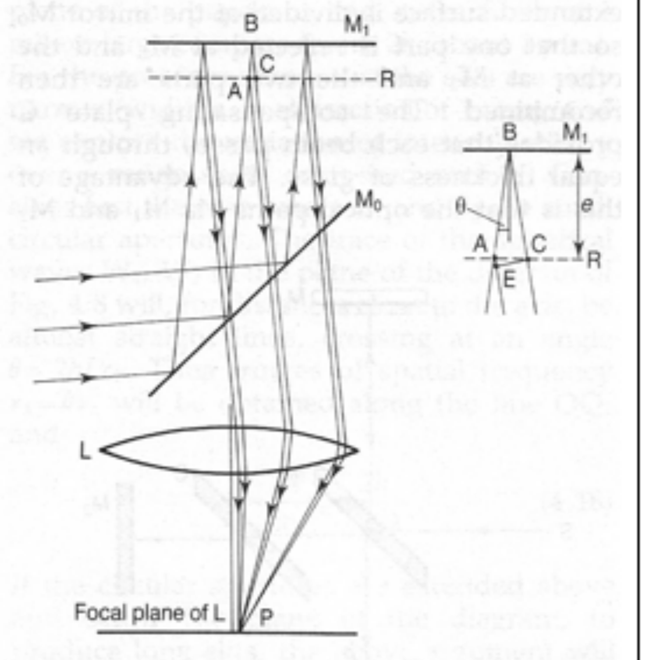


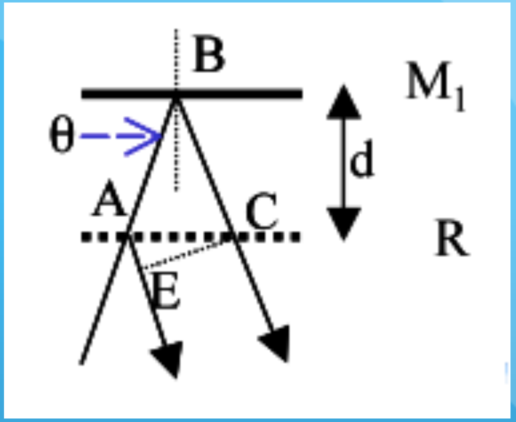
Figure 4.11 Ray paths in Michelson interferometer

(Virtual rays reflecting from R travel equal path lengths to real rays reflecting from M_2 as R is the reflection of M_2 in M_0)

Path difference = AB+BC-AE

$$AB + BC - AE = \frac{2d}{Cos\theta} - 2d\frac{Sin\theta}{Cos\theta}Sin\theta$$

$$AB + BC - AE = 2dCos\theta$$



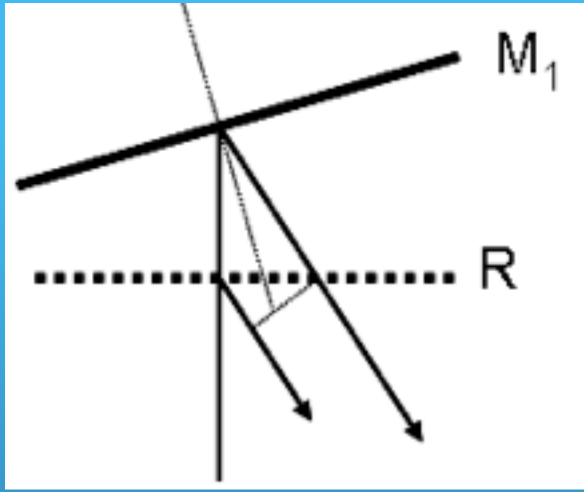
This means that there will be a path difference unless the mirrors are **equidistant from the beam splitter** (d=0), or the rays are **normally incident** on the mirrors (θ=0).

For constructive interference, the path difference must satisfy

$$2dCos\theta = m\lambda$$

Thus for a point source, interference maxima occur for constant angle of incidence, θ. This is fulfilled for a **cone** of light coming from the source. This means the fringes will be **circular**.

We can also use **plane waves** as a source. Then if mirror 1 is **tilted at a small angle** so that it is not parallel to **R**, the path difference will be due to a thin wedge as in the "thin air wedge" interferometer.



The fringes will now be straight lines because the path difference depends on where exactly the ray reflects of M_1.

In addition if M_1 is moved further from or nearer to R, the fringes will appear to move parallel to **R**. This allows the measurement of very small displacements of M_1.

It is the basis for Gravitational wave detection.

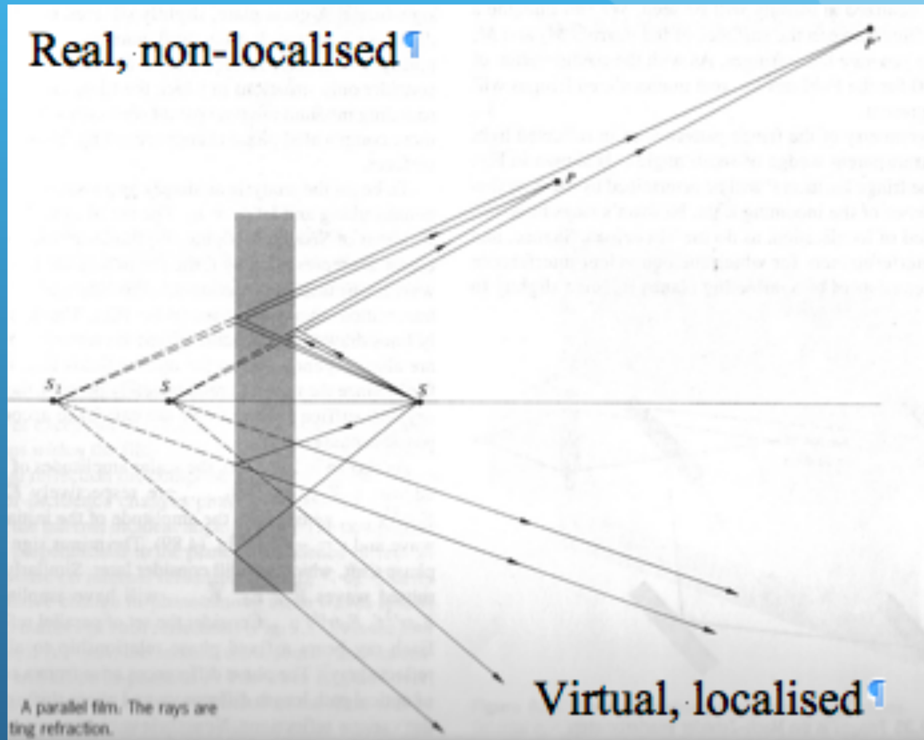
Types and Localisation of Fringes

Fringes can be **real or virtual.**
Real fringes are those that **can** be seen on a **screen without** use of a lens.
Virtual fringes **cannot** be projected onto a **screen without using a lens**.
Fringes can also be **localised or non-localised.**

Non-Localised fringes are real and can be found everywhere in a volume of space. (The fringes in Youngs Ext are non-localised as it doesn't matter how far the screen is from the slits)
Localised fringes can only be seen at a particular position in space.

This can be illustrated using the **Pohl Interferometer** and a point source.

This is just a sheet of dielectric such as glass.
Fringes are formed in two ways:



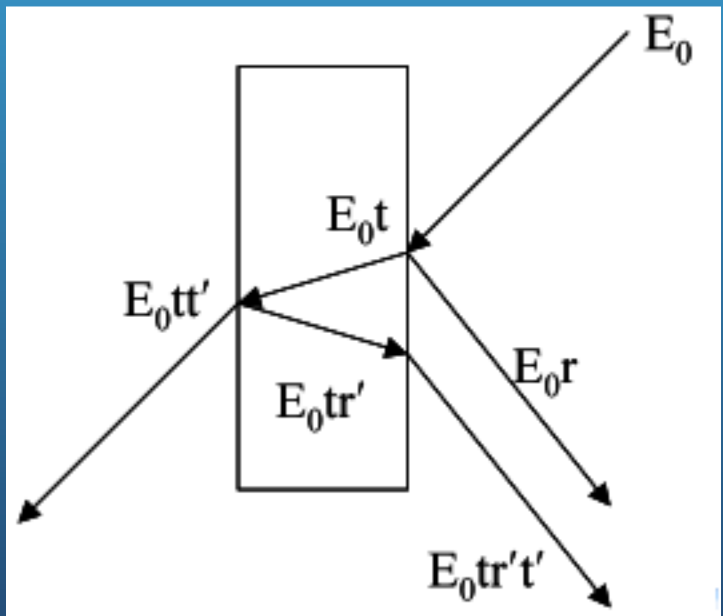
Light reflected from the front surface interfering with light reflected from the back surface (different θ_i). Above top.
By two rays reflected from opposite surfaces interfering with each other (same θ_i). Above bottom.

In the top half of the figure the fringes are **real and non-localised**. This is because rays converge (**real**) onto a screen anywhere (**non-localised**) to the right of the dielectric.

In the bottom half of the figure the fringes are **virtual and localised**. This is because the rays are parallel and need a lens to bring the rays together (**virtual**). The fringes will then be found at the focal plane of the lens (**localised**).

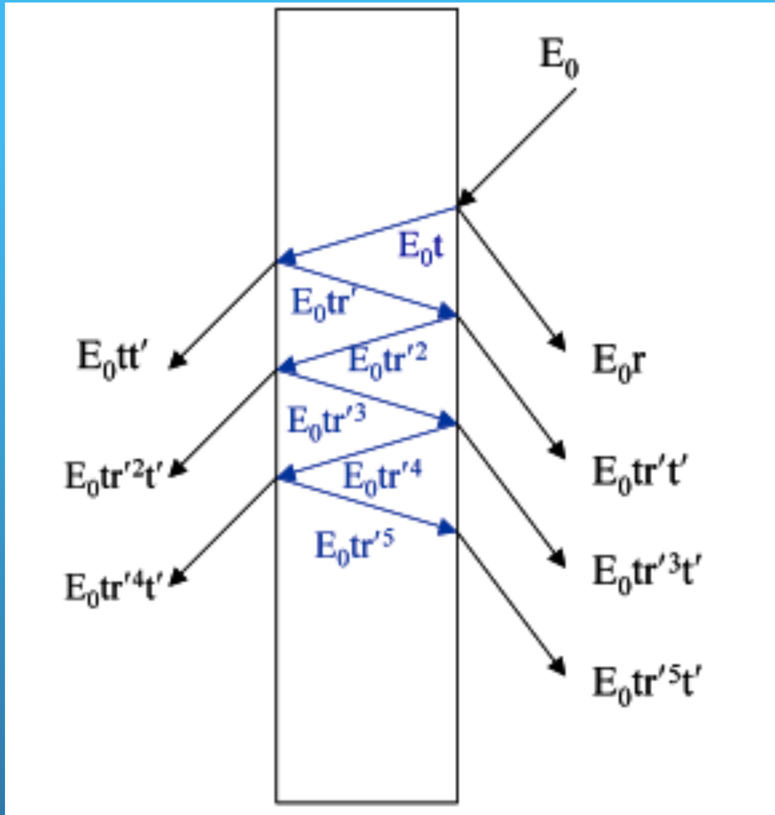
Multiple beam interference

Imagine a light ray with amplitude $\mathbf{E_0}$, incident on a dielectric plate. The fraction of the amplitude transmitted is ***t***. The fraction of the amplitude reflected is ***r***.



We need to keep in mind phase changes on reflection so we write ***r*** for an external reflection and ***r′*** for an internal reflection. Similarly we write ***t*** for a wave entering the material and ***t′*** for a ray leaving the material.

Above we have only shown a fraction of the rays formed by multiple reflections.



$$t + r = 1$$
$$and$$
$$t' + r' = 1$$

Here are some more rays resulting from multiple reflections. Discounting the initially reflected ray, (E_0r), all rays leaving the material have amplitude proportional to E_0tt'. This is because all had to both enter and leave the dielectric. Again, discounting the initially reflected ray, (E_0r), all rays leaving on the right have amplitudes containing **odd** powers of r′.

Also all transmitted rays have amplitudes containing **even** powers of r′.

What is the difference between ***r* and *r′***?

The difference is the phase change! Remember:

When light reflects of an glass-air interface, it just reflects normally.

However when it reflects from a air-glass interface ($n_{glass} > n_{air}$), the reflection **induces a phase shift of π** rads (180^0). This is equivalent to the instantaneous reversal of the E field at the instant of reflection.

This means that the **amplitude changes sign** when a light wave reflects from a **high n/low n interface.**

This means that ***r= -r′***.

$$t + r = 1$$
$$and$$
$$t' + r' = 1$$

therefore $tt' = (1-r)(1-r')$

but $r = -r'$ so $tt' = (1-r)(1+r) = 1 - r^2$

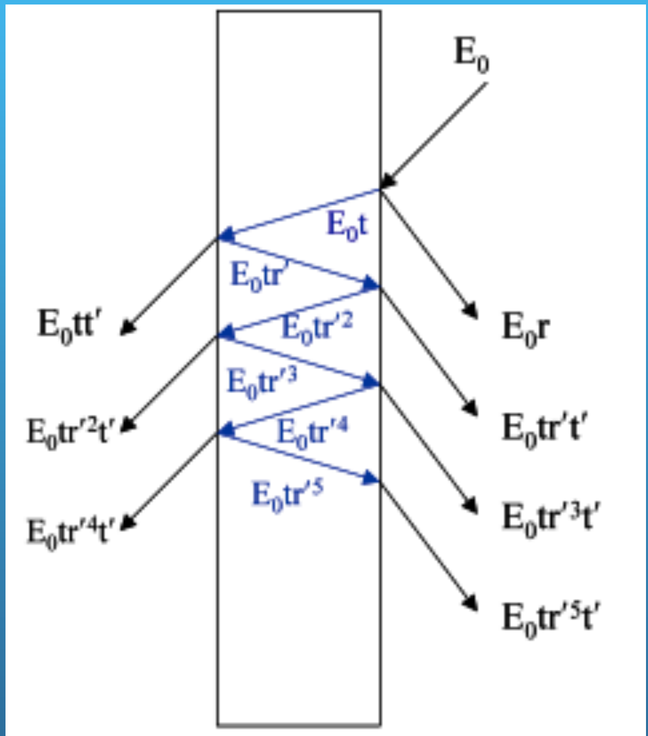
$$tt' + r^2 = 1$$

$r^2 = R$ the reflection coefficient and $tt' = T$ the transmission coefficient

So that in terms of intensity $I_t = TI_0$ and $I_r = RI_0$

How do we calculate the interference for either the reflected or transmitted light?

As an example we will look at the reflected light.
In order to calculate the inference we need to add the electric fields for all the (multiply) reflected rays.



The amplitudes of these rays are: E_0r

$E_0tr't'$

$E_0tr'^3t'$

$E_0tr'^5t'$

etc

However we need to also consider the **phase** of each wave **relative to the phase of the first wave** entering the material. If we consider complex notation, the first wave, just an instant before reflection (position z= 0) is described by:

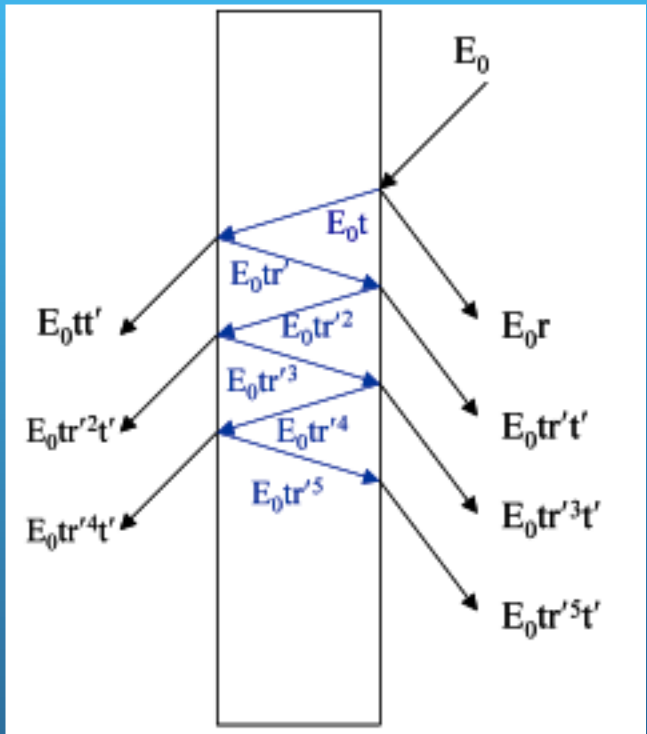
$$E = E_0 e^{i(\omega t - \varepsilon)}$$

where $\boldsymbol{\varepsilon}$ is the phase of the wave when it enters the material (t=0)
For convenience we will write $\boldsymbol{\varepsilon=0}$

How do we calculate the interference for either the reflected or transmitted light?

As an example we will look at the reflected light.

In order to calculate the inference we need to add the electric fields for all the (multiply) reflected rays.



The amplitudes of these rays are: E_0r

$E_0tr't'$

$E_0tr'^3t'$

$E_0tr'^5t'$

etc

However we need to also consider the **phase** of each wave **relative to the phase of the first wave** entering the material. If we consider complex notation, the first wave, just an instant before reflection (position z= 0) is described by:

$$E = E_0 e^{i(\omega t - \varepsilon)}$$

where ε is the phase of the wave when it enters the material (t=0)

For convenience we will write $\varepsilon = 0$

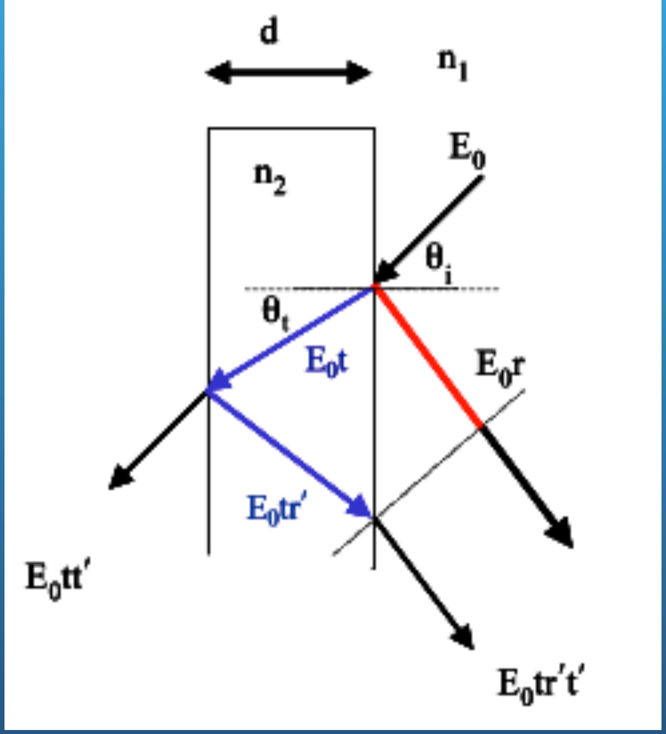
By the time the first and second reflected waves are adjacent to each other, the second has travelled further so has a different phase:

The phase difference between them is just

$$\Delta\varphi = \varphi_2 - \varphi_1 = k_2 z_2 - k_1 z_1$$

where z_1 is the **distance travelled by the first reflected ray** and and z_2 is the **distance travelled by the second ray.** k_2 refers to the **dielectric material.**

We can easily work out z_1 and z_2:



z_2 is the distance travelled by the blue ray.

and z_1 is the distance in red travelled by the first ray.

Remembering that: $k_1 = 2\pi n_1/\lambda_0$ and $k_2 = 2\pi n_2/\lambda_0$ *and* $n_1 Sin\theta_i = n_2 Sin\theta$

We can work out that $$\Delta\varphi = \frac{4\pi d}{\lambda} n_2 Cos\theta_t$$

This means that if the first reflected wave is described after reflection by

$$E_{1r} = E_0 r e^{i\omega t}$$

Then, taking into account the **amplitude reductions on (multiple) reflection** and the **phase changes**, the second reflected wave is

$$E_{2r} = E_0 t r' t' e^{i(\omega t + \Delta\varphi)}$$

But the third reflected wave must just have twice the phase difference with respect to the first

$$E_{3r} = E_0 t r'^3 t' e^{i(\omega t + 2\Delta\varphi)}$$

In general the Nth reflected wave is described by

$$E_{Nr} = E_0 t r'^{(2N-3)} t' e^{i(\omega t + (N-1)\Delta\varphi)}$$

The interference is found by adding all these waves:

$$E_r = E_0 r e^{i(\omega t)} + E_0 t r' t' e^{i(\omega t + \Delta\varphi)} + ..$$
$$+ E_0 t r'^{(2N-3)} t' e^{i(\omega t + (N-1)\Delta\varphi)}$$

This can be rewritten as a series

$$E_r = E_0 e^{i\omega t}\left\{ r + tr't' e^{i\Delta\varphi} \left[\begin{array}{l} 1 + \left(r'^2 e^{i\Delta\varphi}\right) + \left(r'^2 e^{i\Delta\varphi}\right)^2 + \\ + \left(r'^2 e^{i\Delta\varphi}\right)^{N-2} \end{array} \right] \right\}$$

If

$$\left| r'^2 e^{i\Delta\varphi} \right| < 1$$

(as it must be from the definition of r' and the properties of complex exponentials), then

$$E_r = E_0 e^{i\omega t}\left\{ r + \frac{tr't' e^{i\Delta\varphi}}{1 - r'^2 e^{i\Delta\varphi}} \right\}$$

Remember

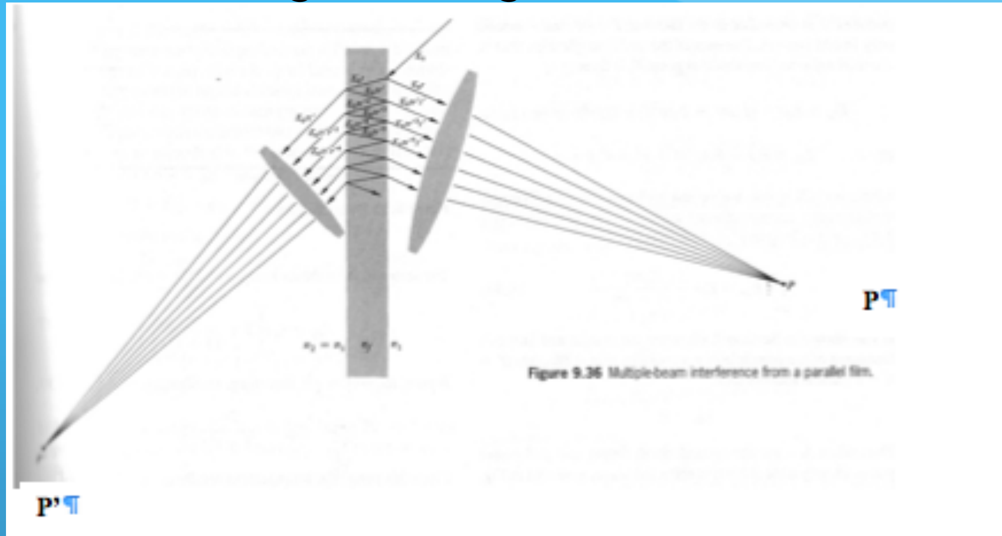
$$\begin{array}{l} tt' = 1 - r^2 \\ and \\ r = -r' \end{array}$$

so

$$E_r = E_0 e^{i\omega t}\left\{ \frac{r\left(1 - e^{i\Delta\varphi}\right)}{1 - r^2 e^{i\Delta\varphi}} \right\}$$

Where does the interference occur?
All the reflected rays are parallel. This means any interference fringes are **virtual**

We need a **lens** to bring them together.



The reflected intensity at P is then $I_r = \frac{\varepsilon c}{2} E_r E_r^*$ where E_r^* is the complex conjugate of E_r.

This means that the reflected intensity at P is

$$I_r = \frac{\varepsilon c}{2} \frac{E_0^2 r^2 \left(1 - e^{-i\Delta\varphi}\right)\left(1 - e^{i\Delta\varphi}\right)}{\left(1 - r^2 e^{-i\Delta\varphi}\right)\left(1 - r^2 e^{i\Delta\varphi}\right)}$$

which can be written as

$$I_r = I_i \frac{2r^2(1 - Cos\Delta\varphi)}{(1 + r^4) - 2r^2 Cos\Delta\varphi}$$

where $I_i = \varepsilon v E_0^2/2$ is just the incident intensity

Exactly the same analysis can be done for the transmitted waves with

$$E_{1t} = E_0 t t' e^{i(\omega t)}$$

$$E_{2t} = E_0 t r'^2 t' e^{i(\omega t + \Delta\varphi)}$$

$$E_{3t} = E_0 t r'^4 t' e^{i(\omega t + 2\Delta\varphi)}$$

In general the Nth reflected wave is described by

$$E_{Nt} = E_0 t r'^{2(N-1)} t' e^{i(\omega t + (N-1)\Delta\varphi)}$$

This gives a transmitted intensity at P' of

$$I_t = I_i \frac{(1-r^2)}{(1+r^4) - 2r^2 Cos\Delta\varphi}$$

Finally using the trig identity

$$Cos\Delta\varphi = 1 - 2Sin^2(\Delta\varphi / 2)$$

This means that

$$I_r = I_i \frac{\left[2r/(1-r^2)\right]^2 Sin^2(\Delta\varphi/2)}{1+\left[2r/(1-r^2)\right]^2 Sin^2(\Delta\varphi/2)}$$

and

$$I_t = I_i \frac{1}{1+\left[2r/(1-r^2)\right]^2 Sin^2(\Delta\varphi/2)}$$

How do we work out the fringe pattern from this equation?

Remember
$$\Delta\varphi = \frac{4\pi d}{\lambda_0} n_2 Cos\theta_t$$

This means that the **phase difference** between adjacent waves, $\Delta\varphi$, depends on $\boldsymbol{Cos\theta_t}$, the **angle of refraction**. This in turn depends on the angle of incidence, θ_i.

Thus we get a different phase difference and so different resultant intensity for different angles of incidence. For each angle of incidence, the light comes out at a given angle of reflection, θ_r, ($\theta_i = \theta_r$).

The lens transforms these different angles into corresponding points on the screen (where the screen is at the focal plane of the lens).

It can easily be shown from these equations that $I_i = I_r + I_t$ and so energy is conserved.

When is the transmitted light a maximum?

$$I_t = I_i \frac{1}{1+\left[2r/(1-r^2)\right]^2 Sin^2(\Delta\varphi/2)}$$

This is when the denominator is a minimum, ie when $Sin^2(\Delta\varphi/2)=0$
This occurs when $\Delta\varphi/2 = m\pi$ or

$$\frac{4\pi d}{\lambda} n_2 Cos\theta_t = 2m\pi$$

this defines a range of angles of refraction (and so angles of incidence) where

$$(I_t)_{max} = I_i$$
$$(I_r)_{min} = 0$$

similarly

$$(I_t)_{min} = I_i \frac{(1-r^2)^2}{(1+r^2)^2}$$

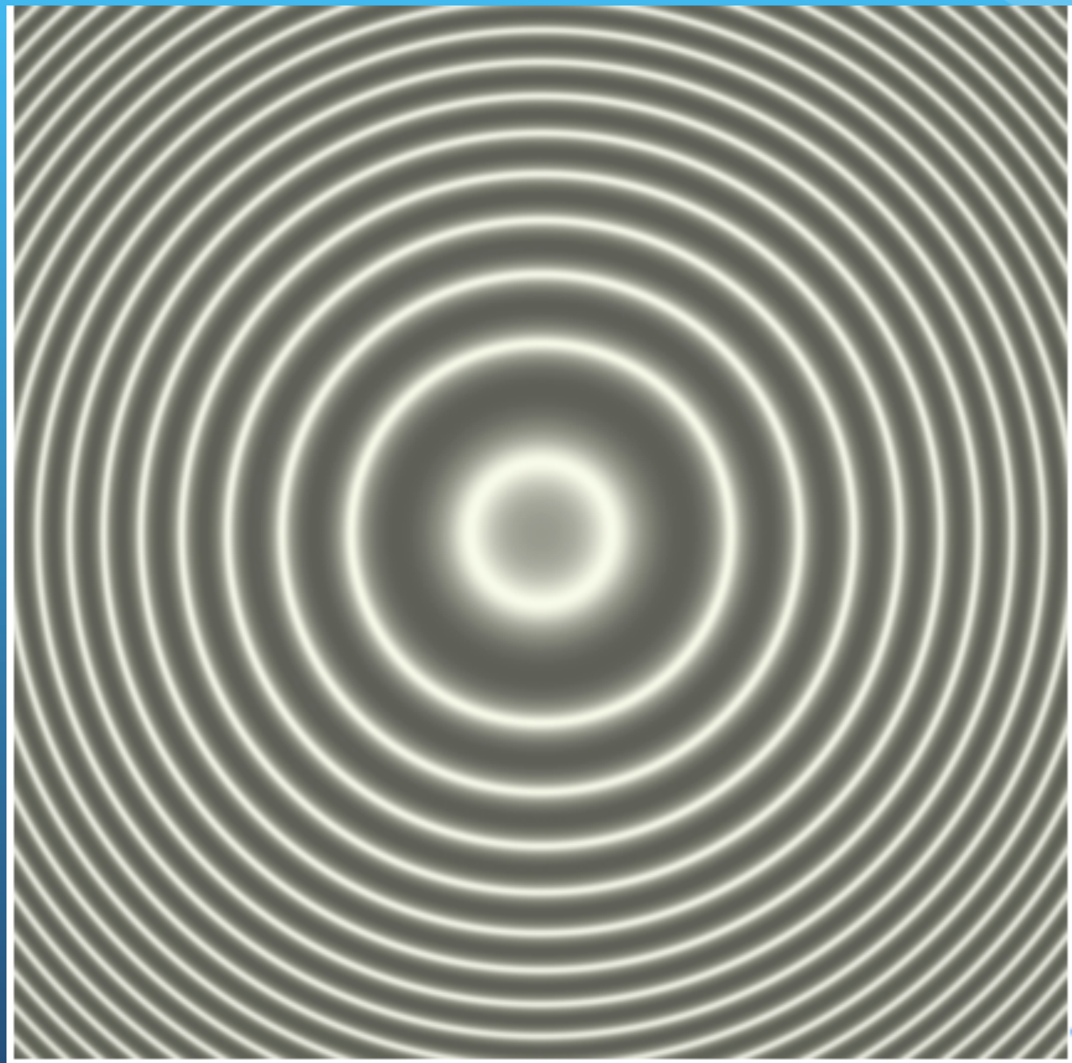
and

$$(I_r)_{max} = I_i \frac{4r^2}{(1+r^2)^2}$$

As $\boldsymbol{r^2 \to 1}$ $\boldsymbol{(I_t)_{min} \to 0}$ $\boldsymbol{(I_r)_{max} \to I_i}$

This means that as $\boldsymbol{r^2 \to 1}$ intense light is only transmitted / reflected for certain values of $\Delta\varphi$. All other things being equal, this means for certain θ_t and hence certain θ_r and hence positions on the screen → ***Interference fringes***

As a given fringe occurs for a given angle of incidence and so a given angle of reflection, a point source gives circular fringes

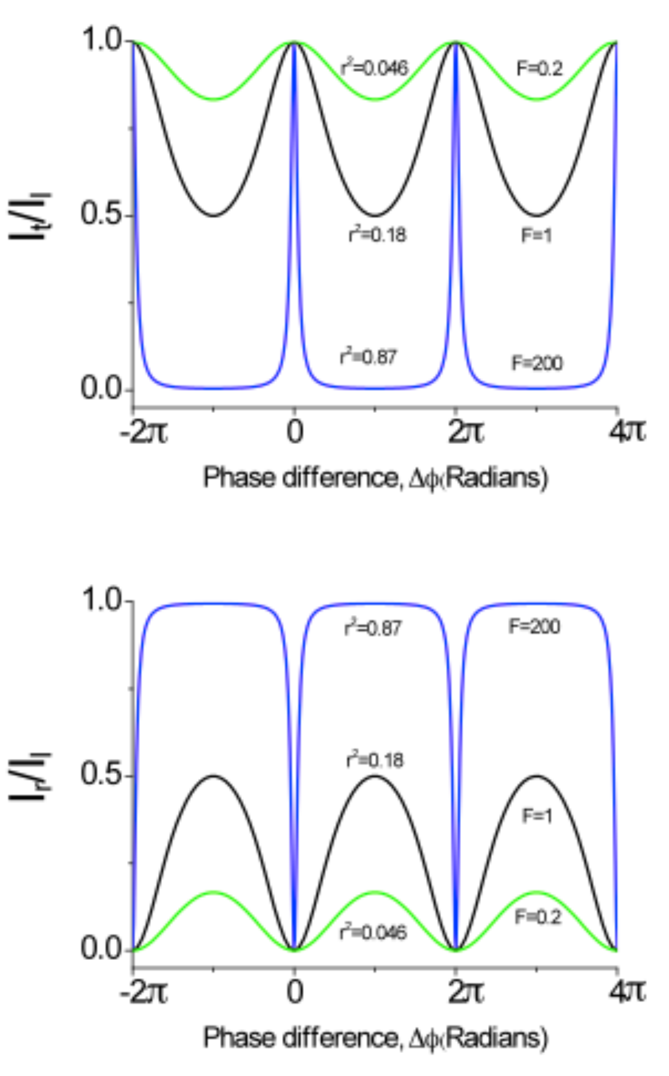


Finally we can simplify the above equations by writing the combination of constants:

$$F = \left[2r/(1-r^2)\right]^2$$

this is called the **Finesse**. Then we can write I_t/I_i as the Airy function:

$$\frac{I_t}{I_i} = \frac{1}{1+F\,Sin^2(\Delta\varphi/2)} \qquad \frac{I_r}{I_i} = \frac{F\,Sin^2(\Delta\varphi/2)}{1+F\,Sin^2(\Delta\varphi/2)}$$



Transmitted

$$\Delta\varphi = \frac{4\pi d}{\lambda} n_2 Cos\theta_t$$

Reflected

This is useful because when $\boldsymbol{r^2 \rightarrow 1}$ or $F \rightarrow \infty$ only light with $\Delta\varphi$ very close to $2m\pi$ is Transmitted.

Remember $(I_t)_{\max} = I_i$ when

$$\frac{4\pi d}{\lambda} n_2 Cos\theta_t = 2m\pi$$

For light normally incident (θ_t=0) the transmitted intensity is maximised for wavelengths

$$\lambda_m = \frac{2dn_2}{m}$$

where m is an integer

If $\boldsymbol{r^2 \rightarrow 1}$ only wavelengths very close to λ_m will be transmitted.

This means we have a wavelength specific filter or ***interference filter***.
We can make interference filters with values of d and n_2 tailored to transmit pretty much any wavelength we choose.

The Fabry Perot Interferometer

This interferometer or **etalon** (if the thickness *d* is fixed) just consists of a dielectric with thin metal films on either side to increase *r*. It is thus very similar to the scenario just discussed. For an etalon, the dielectric is a slab of material such as glass with fixed thickness. For a F-P interferometer we have 2 parallel sheets of glass separated by distance d, coated (on the inside) with a thin metal layer (to increase r). The distance d is then variable.

When light from an extended source is incident on the etalon, circular fringes are formed (all points on any circle of incidence have the same θ_i and hence θ_t ***and*** θ_r).

Remember, the fringes are when the transmitted / reflected intensity is a maximum ie when

$$\frac{4\pi d}{\lambda} n_2 Cos\theta_t = 2m\pi$$

For a given λ, their position on the screen is controlled by the values of θ_t defined by the equation.

If d is varied, θ_t will change and the fringes will appear to move.
This can be used to measure the wavelength of unknown light.
If you move the plates of the interferometer (and so change d) by an amount L and count then number of fringes, N, that pass a given point on the screen.
Then the wavelength of the light is given by

$$\lambda = \frac{2L}{N}$$

This apparatus can also be used to **resolve** light consisting of two or more **wavelengths very close together**.
As the transmitted light is a maximum (constructive interference) when

$$\frac{4\pi d}{\lambda} n_2 Cos\theta_t = 2m\pi$$

then two different wavelengths, λ_1 and λ_2 will have their maximum intensity for a different θ_t and hence different θ_i ***and*** θ_r. This means their respective fringes will be separated on the screen.
This allows resolution of closely spaced **spectral lines** that would be otherwise impossible.

Visibility of fringes

While calculating interference we have generally discussed light that is **totally coherent.**
However we also noted that no real source is **totally spatially or temporally coherent.**

This is because:
Light is emitted in **wavetrains** with **finite coherence length.**
Different wavetrains with **randomly different phases** are emitted from **different points on the source.**

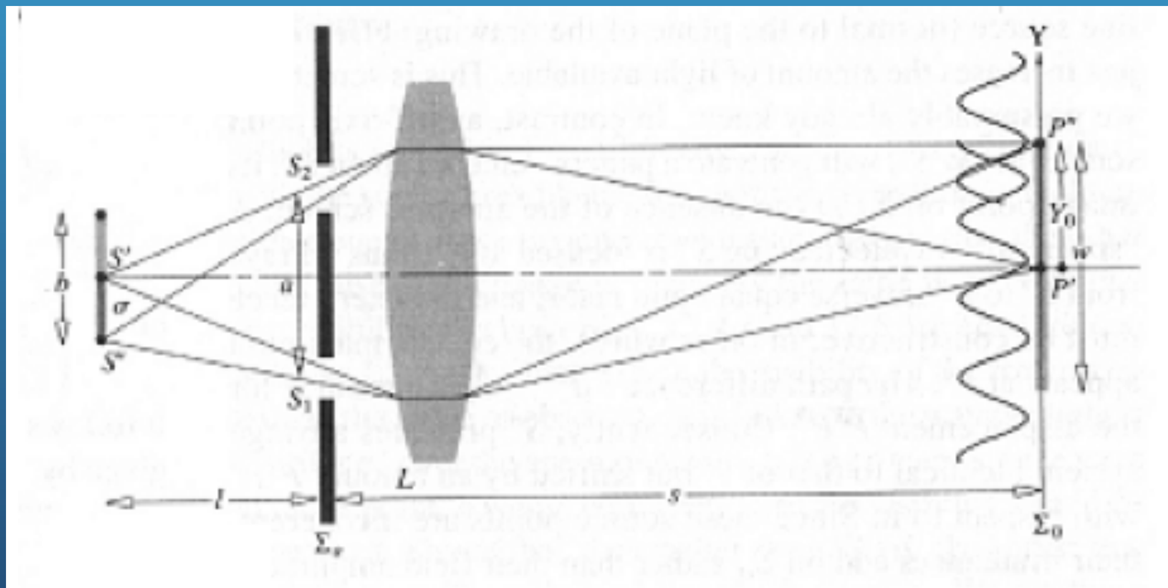
We said the first factor resulted in the loss of the interference pattern at the edge of the screen. We ignored the effect of the second factor because we always assumed **point sources.**

What about a **non-point source**?

If a source has size we have to deal with light coming from different regions of the source. *Each point on the source emits wavetrains that have no relative phase Relationship*. This must disrupt the quality of the interference pattern.

If the source is **small** we would expect **mimimal** disruption.
However if it is **large** we would expect **significant** disruption.

To see the effect of this we will look at **Youngs Experiment** using **two pinholes** and a **finite sized source.**



Consider light coming from a point *S'* on the centre of the source.
This will produce a normal interference pattern centred at the middle of the screen, *P'*.

Any point above or below *S'* on a line perpendicular to the page will produce the same pattern just shifted *below or above P'* (on line perpendicular to the page).

However, consider a point *S''*, away from *S'* along a line **parallel to the pinhole Spacing**. This point will generate an interference pattern at *P''*, ie **shifted from *P'***.

This brings us to 3 important points relating to this extended source.

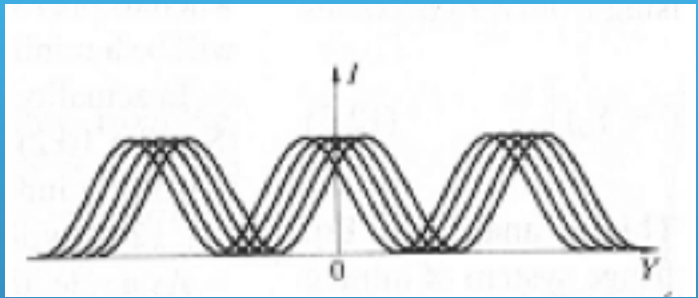
Thus we can think of the source as a **set of line sources**, each pointing Perpendicular to the inter-pinhole direction (out of the page).

Each of these line sources **forms its own pattern**, shifted a given distance from the centre of the screen.

Each of these line sources is **incoherent** with respect to the rest (no fixed phase relationship).

The fact that these sources are **incoherent** is important. It means that light waves from **different sources cannot interfere** with each other (the interference term averages to zero).

This means we can add their intensities.



Summing a number of interference patterns.

Earlier, we derived an equation describing the interference pattern for light from two narrow slits (two line sources).
This applies equally well pinholes as slits.
Where ***y*** is the distance on the screen from the centre of the pattern,
d is the distance between sources and ***L*** is the distance from sources to screen.

$$I = 4I_0 Cos^2\left(\frac{y\pi d}{\lambda L}\right)$$

Where y is the distance on the screen from the centre of the pattern, d is the distance between sources and L is the distance from sources to screen.

How do we describe a pattern centred on a point other than the centre of the screen?

$$I = 4I_0 Cos^2\left[\frac{\pi d}{\lambda L}(y - y_0)\right]$$

Where y_0 is the distance from the **centre of the pattern** to the **centre of the screen**.

Thus, each individual line source contributes an intensity distribution given by the above equation. Each intensity distribution has a different y_0.

Thus to calculate the total intensity distribution (interference pattern) we need to **all up all the contributions** from all the different line sources.
We need to sum over all the different y_0s.

We do this by integration:

$$dI(y) = I(y_0)\frac{dy_0}{w} = \frac{4I_0}{w} Cos^2\left[\frac{\pi d}{\lambda L}(y - y_0)\right] dy_0$$

Here dI is the fraction of the total sum of interference patterns coming from individual interference patterns whose centres fall in $dy_{0.}$

Integrating this is just adding the contributions from all the **different interference patterns** due to all the **different incoherent line sources.**
We integrate over the spatial extent of the **centres of the different interference patterns**, *w*, and so our limits of integration are *w/2* and -*w/2*.
This is equivalent to **integrating over the size of the source. Then**

$$I(y) = \frac{4I_0}{w} \int_{-w/2}^{w/2} Cos^2 \left[\frac{\pi d}{\lambda L} (y - y_0) \right] dy_0$$

This can be integrated quite simply to give

$$I(y) = 2I_0 \left[1 + \frac{\lambda L}{w\pi d} Sin\left(\frac{\pi d}{\lambda L} w \right) Cos\left(2\frac{\pi d}{\lambda L} y \right) \right]$$

Which is of course the same as

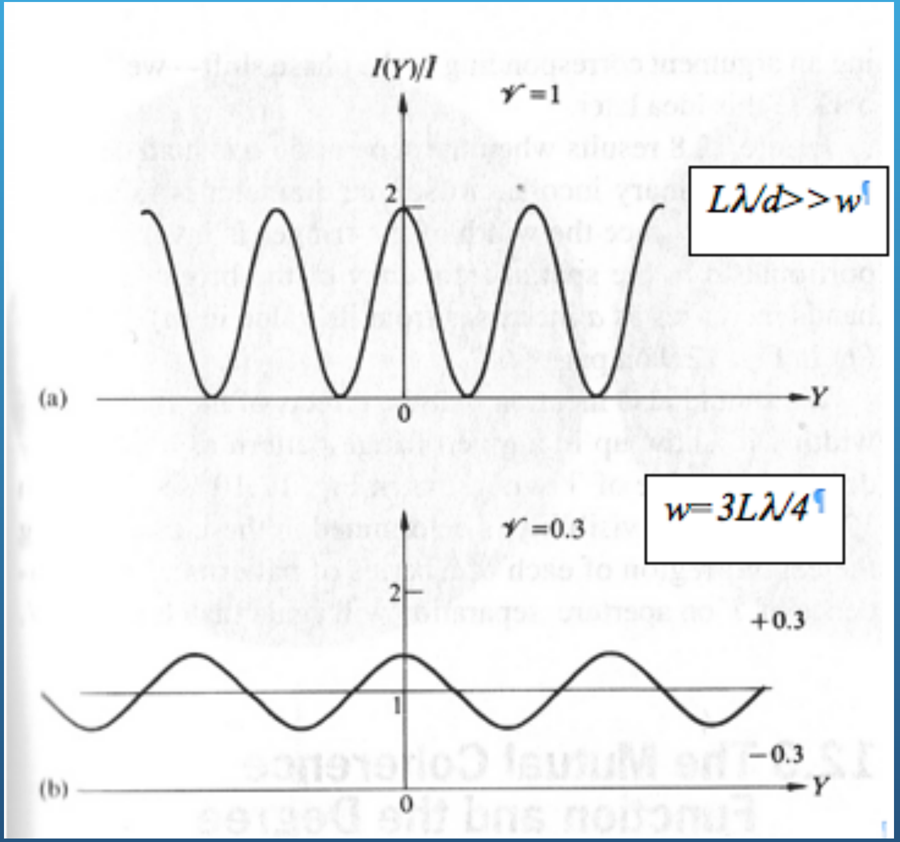
$$\frac{I(y)}{2I_0} = \left[1 + Sinc\left(\frac{\pi d}{\lambda L} w \right) Cos\left(2\frac{\pi d}{\lambda L} y \right) \right]$$

where $[\mathrm{Sinc(x)} = \mathrm{Sin(x)/x}]$

The intensity distribution for an **extended source**, then can be compared with our previous result for a **point source.**

Earlier $$\frac{I(y)}{2I_0} = 1 + Cos\delta = 1 + Cos\frac{2\pi y d}{\lambda L}$$

The difference is the *Sinc* part.
This causes the light intensity in the interference pattern never to reach zero **for an extended source.**
This means we never get **total destructive interference.**



In short we can rate the **coherence of the source** by how pronounced the **difference in light intensity** is between **constructive and destructive interference.**

Or more simply
How big is the difference between maxima and mimima?
This can be described by the visibility of the fringes, *V*:

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}$$

The intensity is at a maximum when

$$Cos\left(2\frac{\pi d}{\lambda L}y\right) = 1$$

and a minimum when

$$Cos\left(2\frac{\pi d}{\lambda L}y\right) = -1$$

and $\mathrm{sinc}(\pi dw/L\lambda) > 0$

This means

$$\frac{I_{\max}}{2I_0} = 1 + \left|Sinc\left(\frac{\pi d}{\lambda L}w\right)\right|$$

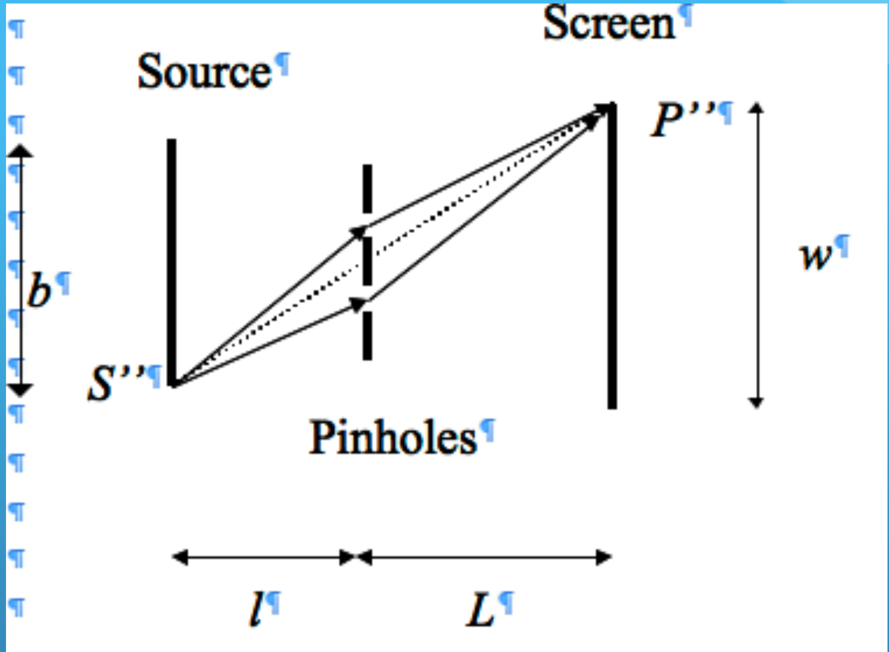
$$\frac{I_{\min}}{2I_0} = 1 - \left|Sinc\left(\frac{\pi d}{\lambda L}w\right)\right|$$

and

$$V = \left|Sinc\left(\frac{\pi d}{\lambda L}w\right)\right|$$

This can be simplified slightly by noting that for a given point near the edge of the source *S''*, the centre of the interference pattern on the screen is at *P''*(near the edge of the pattern).

A line from S'' to P'' goes through the point midway between the pinholes.



We can use this to relate the size of the interference pattern, ***w***, to the size of the source ***b***: $\frac{b}{l} = \frac{w}{L}$ This allows us to write V as

$$V = \left| Sinc\left(\frac{\pi d}{\lambda l} b\right) \right|$$

As the source size, ***b***, is reduced, ***V*** approaches 1 and the fringes become perfectly visible.

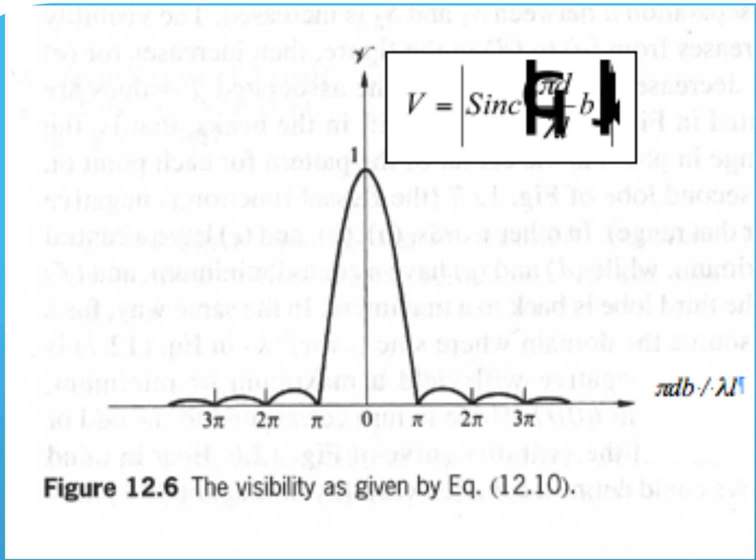
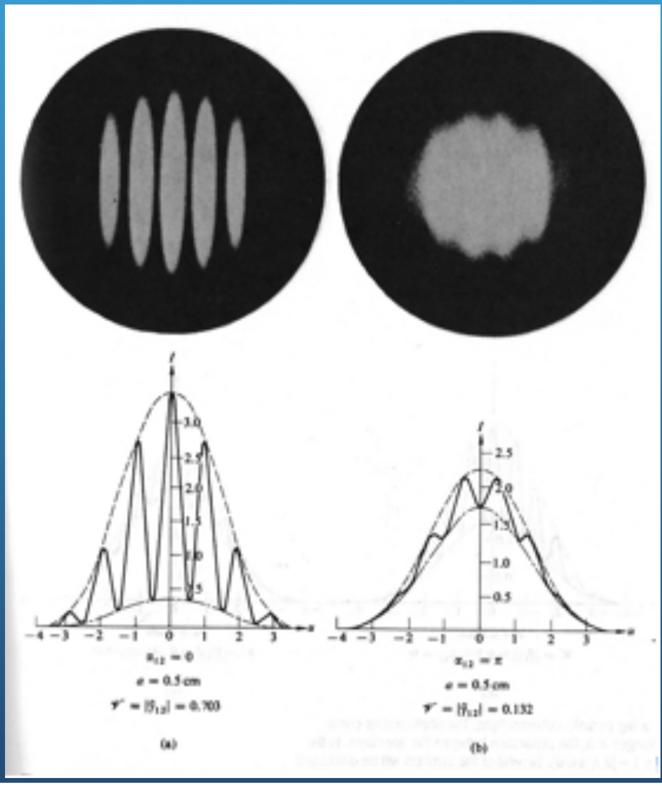


Figure 12.6 The visibility as given by Eq. (12.10).



In reality, the interference pattern from two pinholes will be localised in a circular region of the screen. This is due to diffraction of light at the pinholes.

The interference pattern formed by two pinholes illuminated by an extended source is shown on left along with the interference distribution calculated including diffraction.