5. INTERFERENCE.

Introduction.

Previously, we have spent time considering the amplitude of an electromagnetic wave and have focused on its vector quality when discussing polarization. This is an important property because it is frequently the case that we have several electromagnetic waves from different sources or more frequently different parts of the same source, and need to find the resultant field in a particular region of space by invoking the superposition of fields. In that superposition we obviously need to take account of the vector nature of the field by formally superposing those fields using vector addition.

In the previous analysis of polarisation we noted the change in direction of the electric field vector as a result of superposition and that this lead to different polarization states, left and right handed elliptically and circularly polarised light and plane polarised light. In this section our interest is to be the change in amplitude of the electric field vector (or equivalently light intensity) as a result of superposition or addition of waves from multiple sources or from one source first split and then later recombined at some new location. This brings about a variety of effects grouped collectively under the title of interference; effects including, two slit interference (Young's Slits), thin film interference and various interferometers.

Diffraction phenomena are closely related to interference phenomena and are dealt with using the tools that are developed for interference in what follows. We will however leave diffraction as a topic to be considered separately.

When considering interference it is convenient to continue using plane waves to describe the electromagnetic waves as these are more easily manipulated mathematically. It will later be important to consider how useful this plane wave approach may be in describing reality and this will bring us on to the concept of coherence.

Interference

To establish the mathematical background required to describe interference we continue to consider the transverse plane wave written as;

$$\vec{E}(z,t) = \vec{E}_0 \cos\left[2\pi\left(\frac{z}{\lambda} - \nu t\right)\right] = \vec{E}_0 \cos[kz - \omega t] \qquad (5.1a)$$

or equivalently

$$\vec{E}(z,t) = \vec{E}_0 \exp\ j[kz - \omega t] \qquad (5.1b)$$

as the most simple solution to the electromagnetic wave equation yet containing all essential features.

As before we can establish the electric field in a region of space resulting from two separate sources of field by adding the two fields

$$\vec{E} = \vec{E}_1 + \vec{E}_2 \qquad (5.2)$$

We note that when discussing optics these fields oscillate at typically 5×10^{14}Hz and any measurement we make will not be of the instantaneous field but of the light intensity I which is related to the time averaged square of the electric field as previously demonstrated

$$I = \frac{\langle E^2 \rangle}{\eta} \qquad (5.3)$$

where η is the impedance of the medium (see previous notes) and the triangular brackets represent a time average over several cycles.

$$\vec{E}^2 = \vec{E} \bullet \vec{E} = (\vec{E}_1 + \vec{E}_2) \bullet (\vec{E}_1 + \vec{E}_2) = \vec{E}_1^2 + \vec{E}_2^2 + 2\vec{E}_1 \bullet \vec{E}_2 \qquad (5.4)$$

We may take the time average of each side and assume that everything takes place in the same medium and that therefore η is common to all fields, this will lead to the cancellation of η in much of what follows.

NB. It is important to note here that when taking the dot product in 5.4 we are taking the value of the projection of one field on the other. This automatically takes care of the possibility that the two fields are of different polarization when adding them to discover interference effects.

$$I = I_1 + I_2 + I_{12} \qquad (5.5)$$

$$I_1 = \frac{\left\langle \vec{E}_1^2 \right\rangle}{\eta}, \qquad I_2 = \frac{\left\langle \vec{E}_2^2 \right\rangle}{\eta} \qquad (5.6)$$

The third term is the interference term and is of interest to us in this section

$$I_{12} = \frac{2\left\langle \vec{E}_1 \bullet \vec{E}_2 \right\rangle}{\eta} \qquad (5.7)$$

NB Hecht chooses

(i) ***to omit the impedance in his derivations, a perfectly reasonable course to take as the medium doesn't change and the impedance will cancel when converting from fields to intensity. I choose to keep the impedance in as this is formally correct.***

(ii) ***To use the symbol ε to represent the additional phase of the plane wave whereas I will use the symbol δ to represent the additional phase and ϕ to represent the total phase.***

(iii) ***To use δ to represent the phase difference between two waves whereas I will use $\Delta\phi$ to represent the phase difference between two waves.***

In order to evaluate the interference term we write our two electric fields as two plane waves.

We previously called attention to one of the most important properties of the electromagnetic wave, its phase, and in the case of plane waves travelling in the z direction the phase is simply the argument of the cosinusoid or exponential,

$$\phi = kz - \omega t$$

and this phase does not depend on *x* or *y* once *z* has been fixed, ie. ***the phase is the same at any value of x and y for a given value of z and t***, or ***otherwise stated the phase is constant over a plane perpendicular to the direction of travel*** hence the name plane wave.

When we have two or more plane waves we need to specify the intrinsic phase of each wave as their peaks and troughs do not generally coincide in time and space. To malke things more general to account for this the phase of a wave traveling in the *z* direction becomes $\phi = kz - \omega t + \delta$.

For a plane wave traveling in an arbitrary direction the information concerning this direction is in the fact that $\vec{k}$ is a vector (a fact up until now ignored when the light was assumed propagating in the *z* direction) and the phase is more precisely written as

$$\phi = \vec{k} \bullet \vec{r} - \omega t + \delta$$

The two electric fields are then;

$$\vec{E}_1 = \vec{E}_{01} \cos\left(\vec{k}_1 \bullet \vec{r} - \omega t + \delta_1\right) \quad (5.8a)$$

$$\vec{E}_2 = \vec{E}_{02} \cos\left(\vec{k}_2 \bullet \vec{r} - \omega t + \delta_2\right) \quad (5.8b)$$

NB Whilst the wavevectors k are identified by a subscript 1 or 2 we have no subscript on ω. This is because we are using fields oscillating at the same frequency, the magnitude of the k vectors are the same for both fields and it is the direction of propagation of the wave (or equivalently wavevector) and the electric field vectors (polarisation and amplitude) that are different between the two fields.

We can then find the interference term as follows;

$$\vec{E}_1 \bullet \vec{E}_2 = \vec{E}_{01} \bullet \vec{E}_{02} \cos\left(\vec{k}_1 \bullet \vec{r} - \omega t + \delta_1\right) \times \cos\left(\vec{k}_2 \bullet \vec{r} - \omega t + \delta_2\right) \quad (5.9)$$

We are going to want the time average of $\vec{E}_1 \bullet \vec{E}_2$ and recognizing this we rewrite the equation with the time dependence separated out using the trigonometric identity

$$\cos(A - B) = \cos A \cos B + \sin A \sin B$$

$$\vec{E}_1 \bullet \vec{E}_2 = \vec{E}_{01} \bullet \vec{E}_{02} \left\{ \cos\left(\vec{k}_1 \bullet \vec{r} + \delta_1\right)\cos(\omega t) + \sin\left(\vec{k}_1 \bullet \vec{r} + \delta_1\right)\sin(\omega t) \right\}$$

$$\times$$

$$\left\{ \cos\left(\vec{k}_2 \bullet r + \delta_2\right)\cos(\omega t) + \sin\left(\vec{k}_2 \bullet r + \delta_2\right)\sin(\omega t) \right\} \tag{5.10}$$

We have now separated the time dependent variable (represented in the cosinusoids) making the time average $\left\langle \vec{E}_1 \bullet \vec{E}_2 \right\rangle$ easier to evaluate. The time average only applies to the $\sin^2(\omega t)$ terms and the $\cos^2(\omega t)$ terms and averaging over a time period, T, much greater that the period of the cosinusoids, $\tau = \frac{2\pi}{\omega}$, they both average to:

$$\left\langle \cos^2 \omega t \right\rangle = \left\langle \sin^2 \omega t \right\rangle = \frac{1}{2}.$$

Also the time average of the product of the sin and cosine is zero;

$$\left\langle \cos(\omega t)\sin(\omega t) \right\rangle = 0.$$

This leads to a tremendous simplification in the time average of 5.10 to

$$\left\langle \vec{E}_1 \bullet \vec{E}_2 \right\rangle = \frac{1}{2}\vec{E}_{01} \bullet \vec{E}_{02}\left[\cos\left(\vec{k}_1 \bullet \vec{r} + \delta_1\right)\cos\left(\vec{k}_2 \bullet \vec{r} + \delta_2\right) + \sin\left(\vec{k}_1 \bullet \vec{r} + \delta_1\right)\sin\left(\vec{k}_2 \bullet \vec{r} + \delta_2\right)\right] \tag{5.11}$$

Using further trigonometric identities;

$$\cos A \cos B = \frac{1}{2}\left[\cos(A-B) + \cos(A+B)\right] \qquad \sin \mathrm{A} \sin \mathrm{B} = \frac{1}{2}\left[\cos(\mathrm{A}-\mathrm{B}) - \cos(\mathrm{A}+\mathrm{B})\right]$$

$$\cos A \cos B + \sin A \sin B = \cos(A-B)$$

we can simplify further and 5.11 becomes;

$$\left\langle \vec{E}_1 \bullet \vec{E}_2 \right\rangle = \frac{1}{2}\vec{E}_{01} \bullet \vec{E}_{02}\cos\left(\vec{k}_1 \bullet \vec{r} + \delta_1 - \vec{k}_2 \bullet \vec{r} - \delta_2\right) \tag{5.12}$$

$$I_{12} = \frac{2\langle \vec{E}_1 \bullet \vec{E}_2 \rangle}{\eta} = \frac{E_{01} \bullet E_{02}}{\eta} \cos(\Delta\phi) \tag{5.13}$$

$$\Delta\phi = \vec{k}_1 \bullet \vec{r} - \vec{k}_2 \bullet \vec{r} + \delta_1 - \delta_2 \tag{5.14}$$

is the total phase difference between the two plane waves at the position, $\vec{r}$, where the interference is to be determined. This phase difference is the sum of the phase difference due to different optical path lengths traversed by each wave and due to the initial phases of the two waves, δ_1 and δ_2.

In a common situation, where the polarization of the two interfering waves are identical, ie. the electric fields are parallel;

$$I_{12} = \frac{E_{01}E_{02}}{\eta} \cos\Delta\phi \tag{5.15}$$

We can write I_{12} in terms of intensities by using the earlier expressions relating fields and intensities

$$I_1 = \frac{\langle \vec{E}_{01}^2 \rangle}{\eta} = \frac{E_{01}^2}{2\eta} \qquad\qquad I_2 = \frac{\langle \vec{E}_{02}^2 \rangle}{\eta} = \frac{E_{02}^2}{2\eta} \tag{5.16}$$

$$I_{12} = 2\sqrt{I_1 I_2} \cos\Delta\phi \tag{5.17}$$

And the total irradiance (power per unit area or intensity) is then

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos\Delta\phi \tag{5.18}$$

The total irradiance then varies from point to point in space as $\cos\Delta\phi$ varies between +1 and -1

The maximum irradiance is

$$I_{Max} = I_1 + I_2 + 2\sqrt{I_1 I_2} \tag{5.19a}$$

when the phase difference is $\Delta\phi = 0, \pm 2\pi, \pm 4\pi, \pm 6\pi$ …….

In this case of total constructive interference the two waves are in phase.
And the minimum irradiance

$$I_{Min} = I_1 + I_2 - 2\sqrt{I_1 I_2} \qquad (5.19b)$$

when $\Delta\phi = \pm\pi, \pm 3\pi, \pm 5\pi, \ldots\ldots$
and in this case of total destructive interference the two waves are 180^0 out of phase.

Another commonly encountered situation is where the electric fields have not only the same polarization but also the same intensity, $I_1 = I_2 = I_0$. In this case the total irradiance may be written

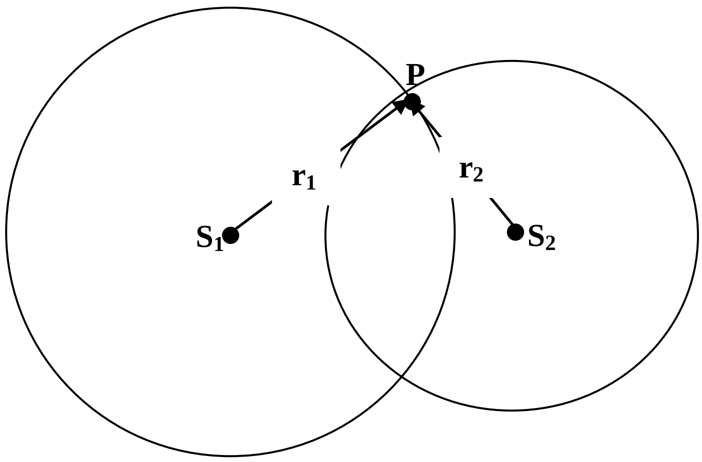
$$I = 2I_0(1 + \cos\Delta\phi) = 4I_0 \cos^2\frac{\Delta\phi}{2} \qquad (5.20)$$

Where the trigonometric identity $1 + \cos\theta = 2\cos^2\frac{\theta}{2}$ has been used.

From this we have;

$I_{\mathrm{Min}} = 0,$ $\qquad I_{\mathrm{Max}} = 4I_0$

Identical arguments and results apply to the interference of two spherical waves emanating from two point sources S_1 and S_2 that overlap at a point P.



We have seen previously that the spherical waves propagating in free space may be written as

$$\vec{E}_1(r_1,t) = \frac{\vec{E}_{01}}{r_1} exp[i(k_0 r_1 - \omega t + \delta_1)] \tag{5.21a}$$

$$\vec{E}_2(r_2,t) = \frac{\vec{E}_{02}}{r_2} exp[i(k_0 r_2 - \omega t + \delta_2)] \tag{5.21b}$$

r_1 and r_2 are the distances of point sources 1 and 2 from the point of overlap, P or equivalently the radii of curvature of the two spherical waves.

The phase in this case depends only on the distance r from the point sources and the surface of any sphere centered on a point source is a surface of constant phase.

We can easily identify the phase difference between the two waves as

$$\Delta\phi = k_0(r_1 - r_2) + (\delta_1 - \delta_2) \tag{5.22}$$

Using the expression for total irradiance as found previously for plane waves, 5.20, and the new phase difference

$$I = 4I_0 \cos^2\left[\frac{1}{2}\{k_0(r_1 - r_2) + (\delta_1 - \delta_2)\}\right] \tag{5.23}$$

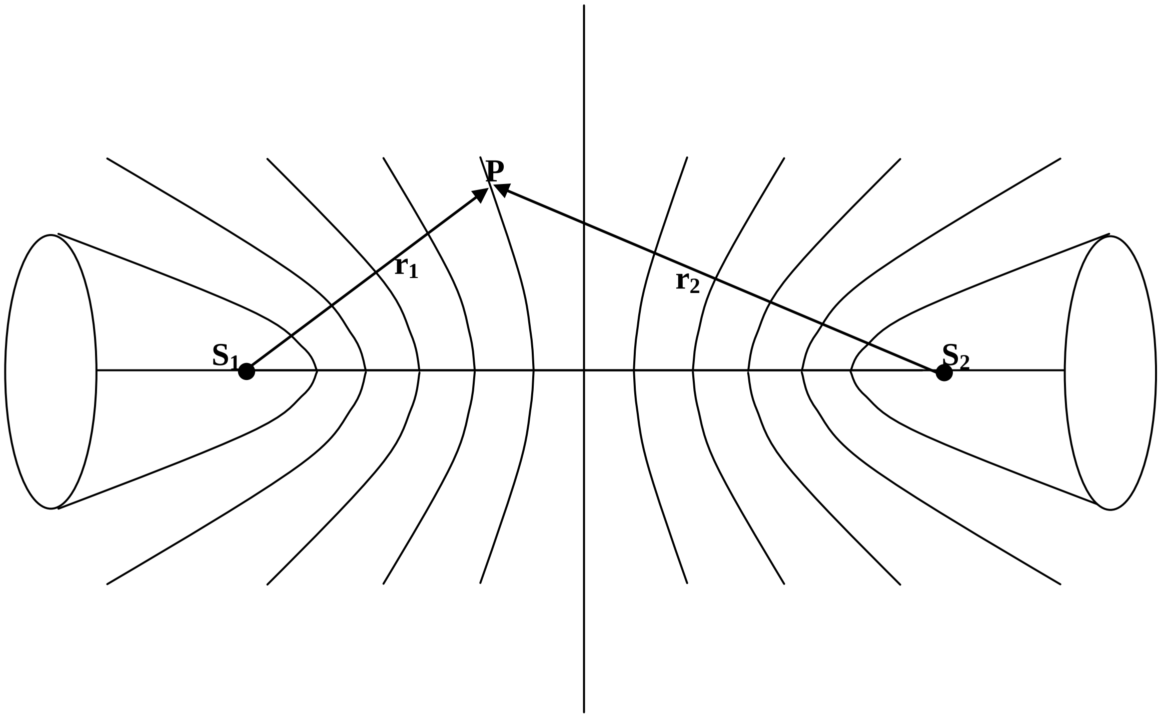
The condition for maxima and minima are as before ie. $\Delta\phi = \pm 2m\pi$ for maxima and $\Delta\phi = \pm(2m + 1)\pi$ where $m = 0, 1, 2, 3, \ldots\ldots$

Using the expression for $\Delta\phi$ we obtain constructive interference when

$$(r_1 - r_2) = \frac{\pm 2m\pi + (\delta_2 - \delta_1)}{k_0} \tag{5.24a}$$

And destructive interference when

$$(r_1 - r_2) = \frac{\pm(2m+1)\pi + (\delta_2 - \delta_1)}{k_0} \tag{5.24b}$$



Both equations 5.24a and b describe a family of hyperboloids as depicted in the above figure where the hyperboloid surfaces represent the points P where constructive/destructive interference occur. If we hold the phases from the two sources, S_1 and S_2 , to be equal, ie, $\delta_1 = \delta_2$ and choose them to be zero then 5.24 may be rewritten as

$$(r_1 - r_2) = \frac{\pm 2m\pi}{k_0} = \pm m\lambda_0 \tag{5.25a}$$

$$(r_1 - r_2) = \frac{\pm(2m+1)\pi}{k_0} = \pm\left(m + \frac{1}{2}\right)\lambda_0 \tag{5.25b}$$

Each value of *m* is represented by a hyperboloid with the positive values on the right hand side of the mid-line with $r_1 > r_2$ and the midline (or in 3D a plane perpendicular to the page) where $r_1 = r_2$ and the left hand side hyperboloids for $r_1 < r_2$. We can imagine a screen placed as a plane intersecting this set of hyperboloids where at the points of intersection there is constructive interference. Any point, P, on this midpoint plane has $r_1 = r_2$ and therefore represents 5.25a with $m = 0$, the zeroth order interference fringe.

These equations can be used to describe the appearance of interference fringes between two line or point sources such as Young's slits. We now need to explore some examples of the ways in which interference effects are manifested.

Interference Effects in Practice

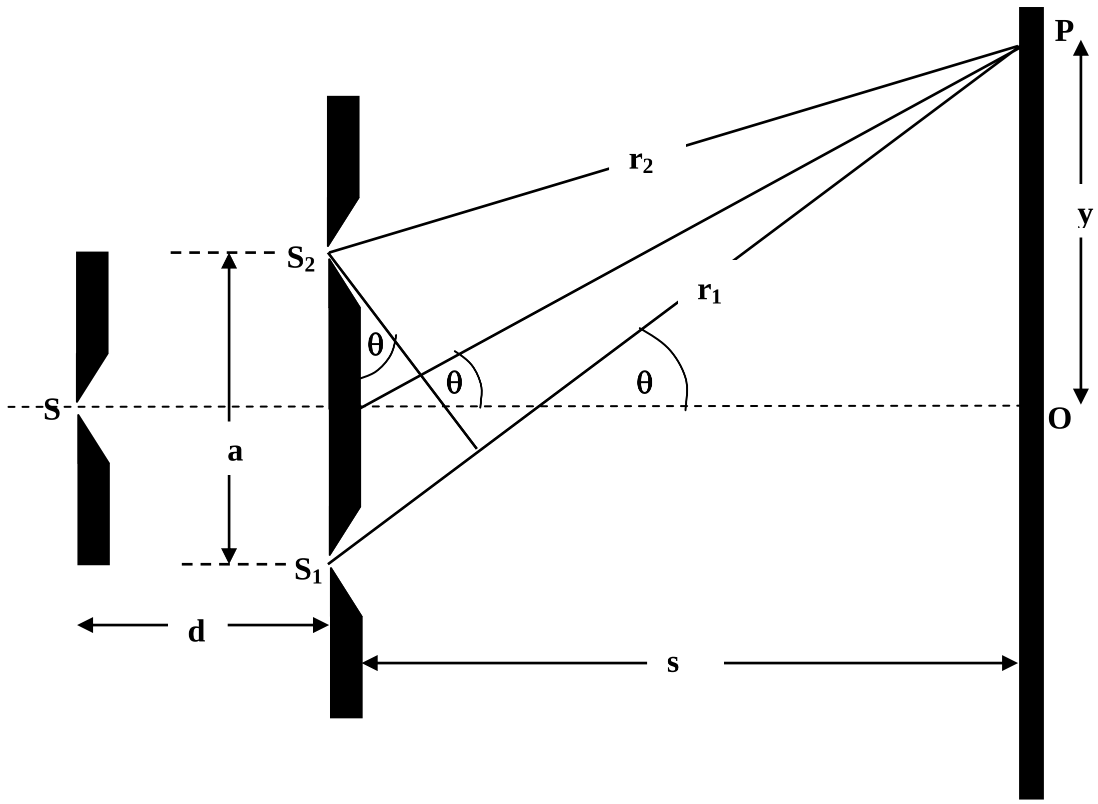
(i) Interference by Division of Wavefront.

Young's Slits

Thomas Young, in 1801, was one of the first people to demonstrate the wave nature of light carrying out what has since become a classic experiment that has been applied to particles to demonstrate matter waves as well as light. In Young's original experiment he used a pinhole with a monochromatic light source behind it to define a point source with two further pinholes in a screen at a distance from the source pinhole much greater than the wavelength of light. The two pinholes in the screen act as two separate but related light sources that have been derived from an original light source by division of the wavefront of that original source. That the two secondary sources are related allow interference effects to be observed by placing a second screen at a similar distance from the two secondary light sources.
To find the light intensity at the second screen we need to add the electric fields of the two light sources taking into account the difference in phase between those two fields as usual. If the two holes providing the interfering sources are of the same size then we can make the approximation that the intensities of the two sources are equal and use equation 5.20 to find the intensity at any point P on the screen;

$$I = 4I_0 \cos^2 \frac{\Delta\phi}{2} \qquad (5.20)$$

Before continuing further, it is necessary to keep the observation point P close to the centre of the screen and the diagram below is exaggerated for demonstration purposes. In fact we are interested in small values of θ.



It remains to establish the phase difference, $\Delta\phi$, at the observation point, P.

The two point sources act as sources of spherical waves and we have the phase difference earlier in 5.22

$$\Delta\phi = k_0(r_1 - r_2) + (\delta_1 - \delta_2) = k_0(r_1 - r_2) \qquad (5.22)$$

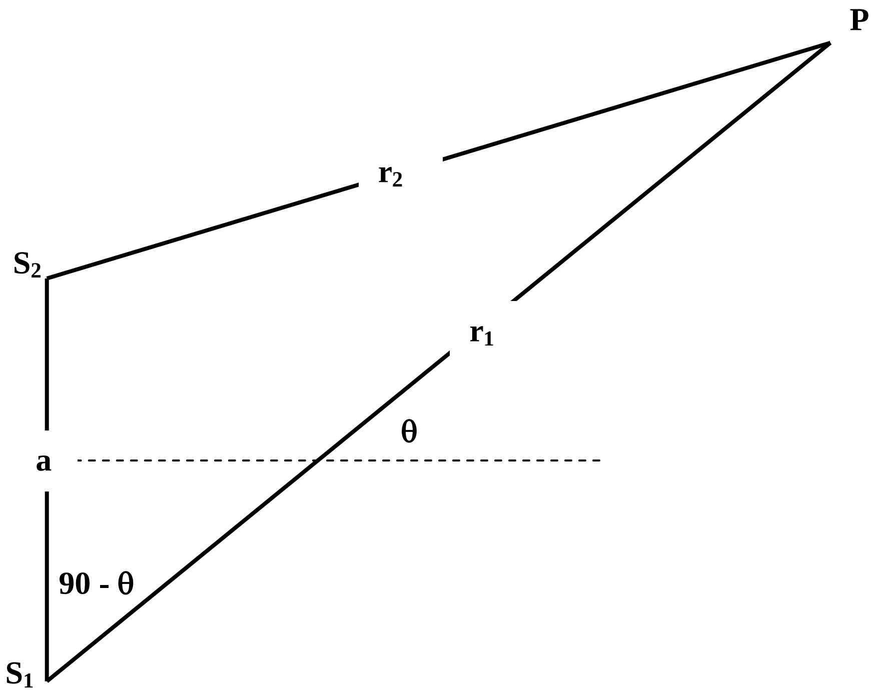
If the screen with the two point sources is far enough from the original point source, S, (orders of magnitude larger than a wavelength) then the spherical wave at S_1 and S_2 will be approximately plane and there will be no other source of phase difference apart from the path difference and the term (δ_1 - δ_2) is zero. With the aid of the figure above we may find the path difference which is the same as the optical path difference, Λ, (as the waves propagate in air with n = 1) and hence the phase difference $\Delta\phi = k_0\Lambda$.

$$\Lambda = S_1P - S_2P = r_1 - r_2 \qquad (5.26)$$

We may try to get a simplified expression for $r_1 - r_2$ beginning with the law of cosines,

$$c^2 = a^2 + b^2 - 2ab\cos C$$

as a applied to the triangle S_1S_2P



$$r_1^2 = a^2 + r_2^2 - 2ar_2\cos(90-\theta) = a^2 + r_2^2 - 2ar_2\sin\theta \tag{5.27}$$

$$r_1^2 - r_2^2 = a^2 - 2ar_2\sin\theta \tag{5.28}$$

$$(r_1 - r_2)(r_1 + r_2) \approx (r_1 - r_2)2r_2 = a^2 - 2ar_2\sin\theta \tag{5.29}$$

$$\Lambda = r_1 - r_2 \approx a\sin\theta \approx a\theta \tag{5.30}$$

Where we used the fact that a << r_1 and $r_1 \approx r_2$ to simplify 5.29

And a further simplification

$$\theta \approx \frac{y}{s} \tag{5.31}$$

may be used to obtain

$$\Delta\phi = k_0 \Lambda = k_0 a \frac{y}{s} = \frac{2\pi}{\lambda_0} a \frac{y}{s} \tag{5.32}$$

Equation 5.25a and b gave us conditions for maxima and minima respectively and using this result and 5.25 we find the condition for a bright fringe

$$(r_1 - r_2) = \frac{ay}{s} = \pm m\lambda_0 \tag{5.33}$$

Or in terms of position on the screen, y

$$y_m = m\frac{s}{a}\lambda_0 \tag{5.34}$$

Gives the position of the m^{th} bright fringe on the screen. The fringes are equally spaced with a separation

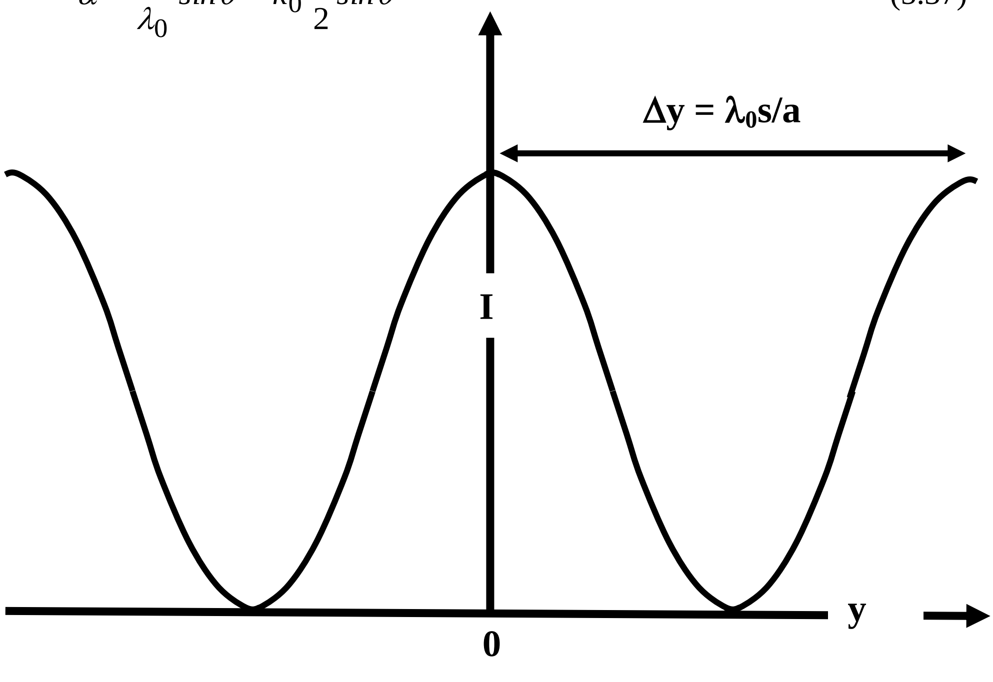
$$\Delta y = \frac{s}{a}\lambda_0 \tag{5.35}$$

Now that we have $\Delta\phi$ we can write the intensity as a function of *y* by using this in 5.20

$$I = 4I_0 \cos^2\left(\frac{\pi a y}{\lambda_0 s}\right) = 4I_0 \cos^2\left(\frac{\pi a s \sin\theta}{\lambda_0 s}\right) = 4I_0 \cos^2 \alpha \tag{5.36}$$

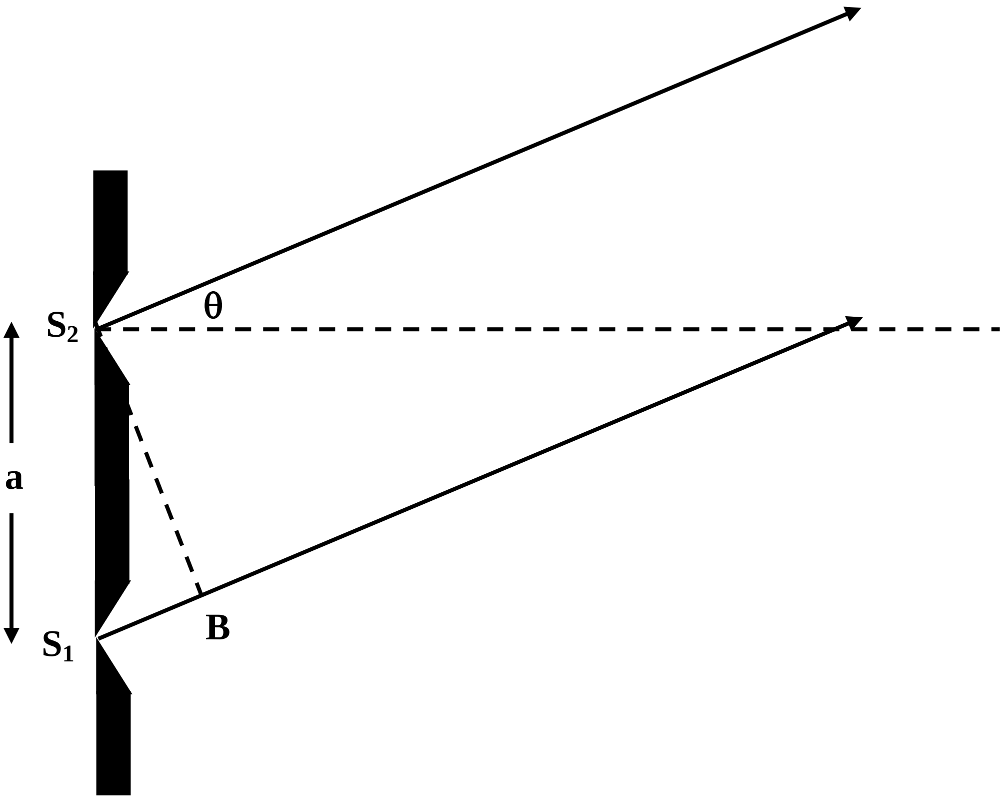
With

$$\alpha = \frac{\pi a}{\lambda_0}\sin\theta = k_0 \frac{a}{2}\sin\theta \tag{5.37}$$



α is the phase difference between rays emanating from each slit when the screen is at infinity (ie the two rays are parallel) as shown in the diagram below.

$$\alpha = k_0 S_1 B \tag{5.38}$$



We have found from this analysis that;

(i) to have well separated fringes we need a small separation, *a*, between pinholes or slits and that

(ii) longer wavelengths will give rise to broader fringes.

We have seen that Young had some stringent requirements on the geometry of his experiment in order to see the interference fringes;

(i) The slits cannot be too far apart and generally $a << s$

(ii) The original source must be at a distance, *d*, from the screen such that the spherical wave approximates a plane wave in order that $\delta_1 \approx \delta_2$

(iii) Fringes will only be observed near the centre of the screen where $r_1 - r_2$ is not too large.

(iv) The value of s/a must be large if fringe separation is to be large enough to observe due to the small size of the wavelength of visible light.

All of the above apart from (iv) result from the lack of coherence in the sources available to Young. We return to the question of coherence after we examine the second type of interference.

Interference by Division of Amplitude.

An important way of achieving the conditions for exhibiting interference effects, ie. obtaining two or more oscillating electric fields, sending them on different paths before recombining them to observe interference is to split a primary wave into two separate secondary waves by partial reflection of the primary and using the reflected and transmitted waves as the secondary waves to be recombined. This may be achieved in many ways, sometimes coming about in a quite natural manner while often achieved with a particular optical arrangement.

Dielectric Thin Film Interference.

One of the most commonly observed examples of interference through division of amplitude occurs where light incident on a thin layer of dielectric undergoes reflections from the top and bottom surface of the layer and under the right conditions constructive or destructive interference occurs. Examples of this are the colour effects seen when a thin layer of oil is floating on water and the colours seen in a soap bubble. The interference effects caused by multiple reflections in the thin scales of a butterflies wings give rise to their spectacular iridescence. The effect is also of use in many technological applications where thin layers are designed and constructed with the intention of creating thin dielectric film interference.

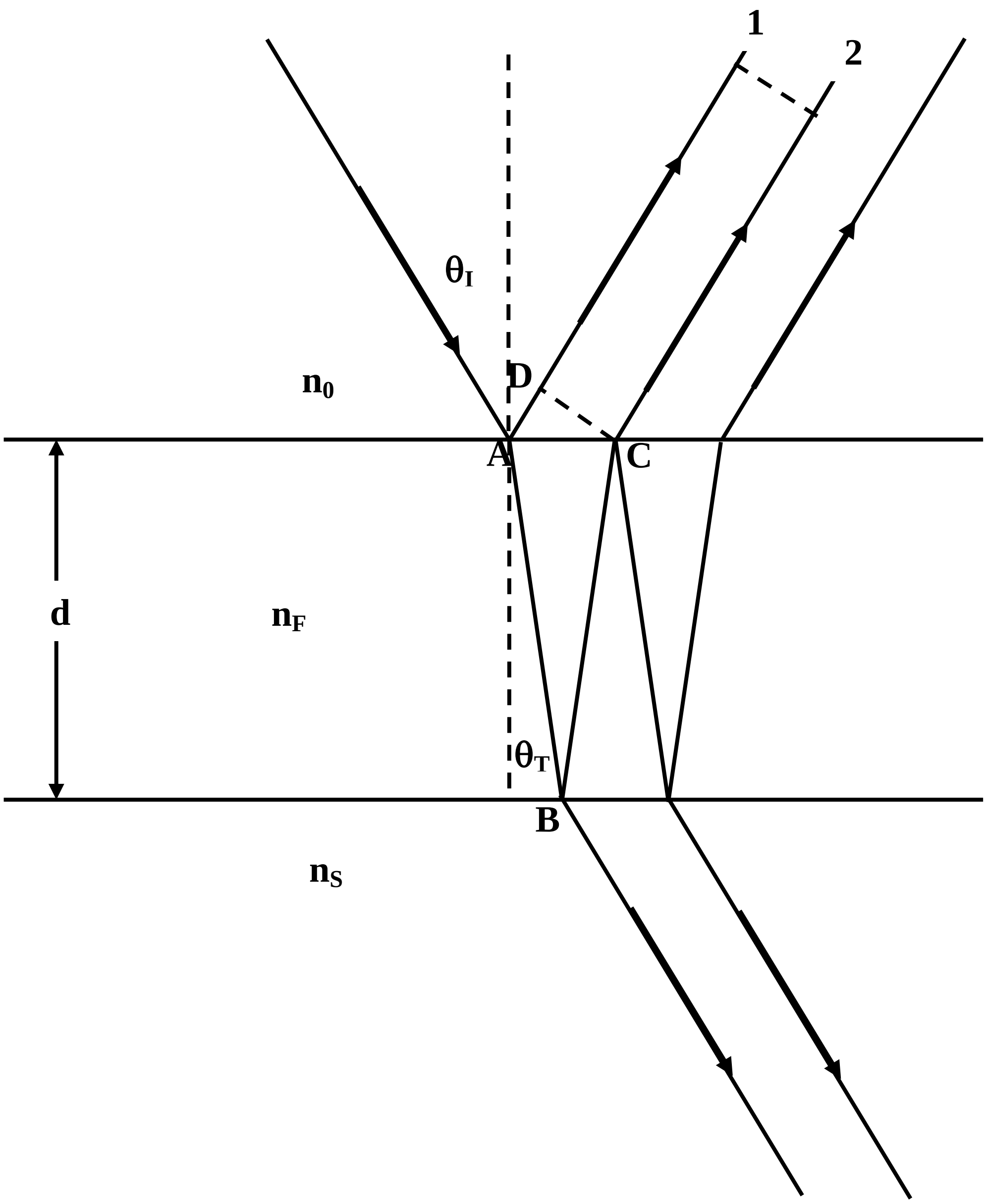
The diagram below indicates schematically how interference effects are produced as a result of partial reflection/transmission at a thin dielectric film. For purposes of this analysis the dielectric film of refractive index, n_F , is standing on a substrate of refractive index, n_S, and light is incident from a medium of refractive index, n_0 (typically air). There will in general be multiple reflections at the air/film interface and at the film/substrate interface. Transmission will also occur into the substrate. We can use equations 5.19 to discover the conditions for constructive or destructive interference by first calculating the phase difference between the two rays, 1 and 2, shown propagating upwards from the top of the film in the lower figure.

The optical path is the real space distance traveled multiplied by refractive index of the medium and therefore the optical path difference, Λ, between rays 1 and 2 after their separation on arrival at A is

$$\Lambda = n_F\left(AB + BC\right) - n_0 AD \qquad (5.39)$$

And the phase difference due to the optical path difference is the optical path difference multiplied by the ***magnitude*** of the wavevector in free space (vacuum).

$$\Delta\phi_{OPD} = k_0\Lambda = k_0\left[n_F\left(AB + BC\right) - n_0 AD\right] = \frac{2\pi}{\lambda_0}\left[n_F\left(AB + BC\right) - n_0 AD\right] \quad (5.40)$$



NB. For the total phase change we need to include the phase change that occurs on reflection which may be different for each ray as we recall from the Fresnel equations. Thus,

$$\Delta\phi = k_0\left[n_F\left(AB+BC\right)-n_0AD\right]+\left(\delta_2-\delta_1\right)=\frac{2\pi}{\lambda_0}\left[n_F\left(AB+BC\right)-n_0AD\right]+\left(\delta_2-\delta_1\right) \quad (5.41)$$

From geometry and Snell's law we obtain everything in terms of θ_T,

$$AB = BC = \frac{d}{\cos\theta_T} \qquad AC = 2d\tan\theta_T \qquad AD = AC\sin\theta_I = 2d\tan\theta_T\frac{n_F}{n_0}\sin\theta_T$$

$$AB + BC = \frac{2d}{\cos\theta_T} \qquad AD = 2d\frac{n_F}{n_0}\frac{\sin^2\theta_T}{\cos\theta_T}$$

whence

$$\Delta\phi = \frac{2\pi}{\lambda_0}\left[\frac{2n_F d}{\cos\theta_T}\left(1-\sin^2\theta_T\right)\right]+\left(\delta_2-\delta_1\right)=\frac{4\pi n_F d}{\lambda_0}\cos\theta_T+\left(\delta_2-\delta_1\right) \quad (5.42)$$

It is often convenient to have the phase difference in terms of the angle of incidence and we can use Snell's law to achieve this

$$\Delta\phi = \frac{4\pi n_F d}{\lambda_0}\sqrt{1-\sin^2\theta_T}+\left(\delta_2-\delta_1\right)=\frac{4\pi d}{\lambda_0}\sqrt{n_F^2-n_0^2\sin^2\theta_I}+\left(\delta_2-\delta_1\right) \quad (5.43)$$

Before we address ourselves to the outstanding question of the reflection phase shifts δ_1 and δ_2 we note that n_F may be greater than or less than n_0 and n_S, eg an air gap between two parallel separated glass slides or a freestanding soap film in air respectively. We also recall that there are two types of reflection namely ***internal reflection*** where the refractive index of the sourced region is greater than that of the unsourced region and ***external reflection*** where the refractive index of the unsourced region is greater than that of the sourced region. With these possible types of reflection in mind we can identify several possibilities that ***apply at near normal incidence, $\theta \le 30^0$***;

(i) External reflection plus internal reflection

An example is the soap film/bubble free standing in air, $n_F > n_0 = n_S = n_{Air}$. In this case for near normal incidence, $\theta \leq 30^0$, we will have $(\delta_2 - \delta_1) \approx \pm\pi$ and the total phase difference is

$$\Delta\phi = \frac{4\pi n_F d}{\lambda_0} \cos\theta_T \pm \pi \tag{5.44}$$

(ii) Internal reflection plus external reflection.

An example of this is an air gap between two glass slides. Again for near normal incidence $(\delta_2 - \delta_1) \approx \pm\pi$ and the total phase difference is

$$\Delta\phi = \frac{4\pi n_F d}{\lambda_0} \cos\theta_T \pm \pi \tag{5.45}$$

(iii) Both reflections are internal (or external)

If the reflection at the top and bottom interfaces are both internal reflections (or external reflections) the factor $(\delta_2 - \delta_1) = 0$ and

$$\Delta\phi = \frac{4\pi n_F d}{\lambda_0} \cos\theta_T \tag{5.46}$$

(iv) Reflections from partially metallised surfaces

Examples are a thin metallic film or two partially metallised surfaces separated by a dielectric (eg. air gap between two partial mirrors. In this circumstance then $\delta_1 = \delta_2 = 0$ as there is no phase shift upon reflection at a metal where boundary conditions require that the light electric field is zero and again

$$\Delta\phi = \frac{4\pi n_F d}{\lambda_0} \cos\theta_T \tag{5.47}$$

(i) Maximum Reflection

As we saw from equation 5.18, giving the intensity of the recombined light in terms of the intensities of the original waves and their phase difference, there is from 5.19a a maximum irradiance where $\cos\Delta\phi = 1$ ie. when $\Delta\phi = \pm 2m\pi$, $m = 0, 1, 2, 3$…….

We can apply this to any of our four situations above and taking scenarios (iii) or (iv) we find that maximum light is reflected when

$$\Delta\phi = \frac{4\pi n_F d}{\lambda_0}\cos\theta_T = \pm 2m\pi \tag{5.48}$$

Or re-written the condition becomes

$$\frac{2n_F d}{\lambda_0}\cos\theta_T = \pm m \tag{5.49}$$

The LHS depends on d, λ_0 or θ and if two of these are held constant we can obtain maximum reflected irradiance by varying the other, eg. Keeping d and λ_0 fixed we can obtain maximum irradiance by varying θ.

$$\cos\theta_T = \pm m\frac{\lambda_0}{2n_F d} = \pm m\frac{\lambda_F}{2d} \tag{5.50}$$

for maximum irradiance.

We can note here that for the special situation of normal incidence ($\cos\theta_T = 1$) the condition for maximum reflection is

$$1 = \pm m\frac{\lambda_F}{2d} \qquad \text{or} \qquad \pm m\frac{\lambda_F}{2} = d$$

Physically this represents half integer wavelengths, λ_F fitting between the two surfaces of the film. This is easily recognized as the requirement that standing waves exist between the two surfaces for maximum reflection.

Alternatively, with a white light source, at a given angle for a fixed thickness, only wavelengths satisfying 5.50 will be strongly reflected.

(ii) Maximum Transmission

From 5.18 and 5.19b we have minimum reflection when $\Delta\phi = \pm\pi, \pm 3\pi, \pm 5\pi.....$ This is also the condition for maximum transmission.

$$\Delta\phi = \frac{4\pi n_F d}{\lambda_0}\cos\theta_T = \pm(2m+1)\pi \tag{5.51}$$

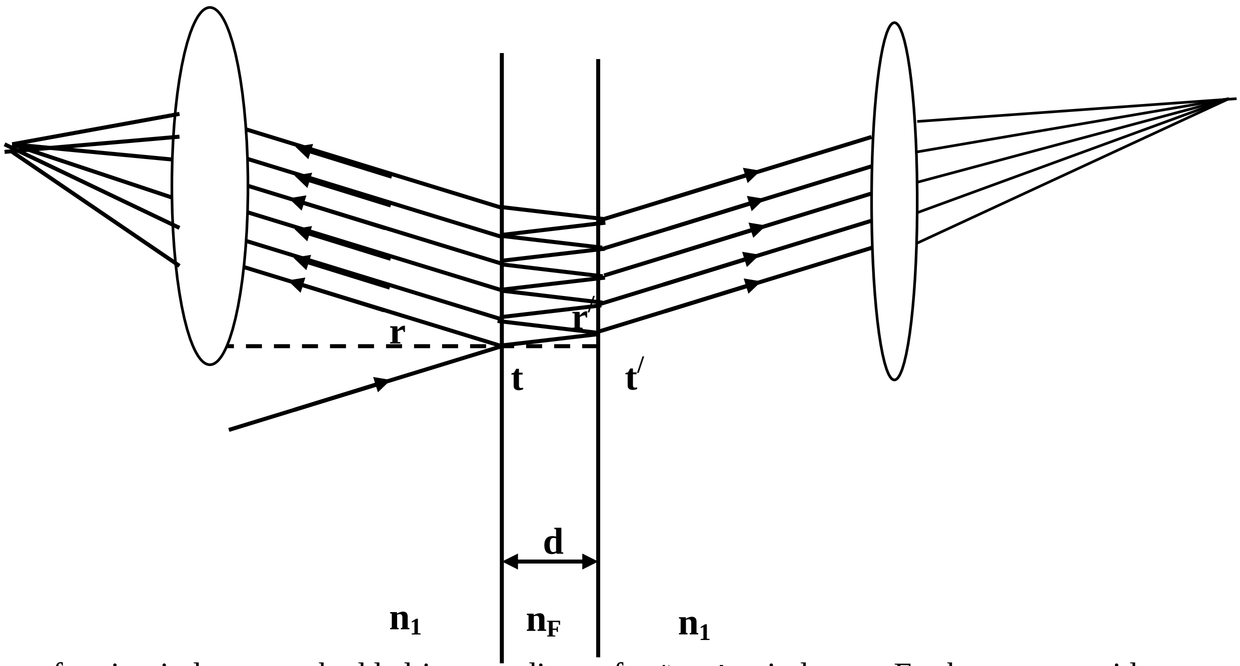
Or re-expressed;

$$\cos\theta_T = \pm(2m+1)\frac{\lambda_0}{4n_F d} = \pm(2m+1)\frac{\lambda_F}{4d} \tag{5.52}$$

So far nothing has been said about the intensities of the two combining beams, I_1 and I_2 and so nothing is known about the actual irradiance, I_R , achieved in the reflected beam. Neither has the possibility of multiple reflections been considered nor the intensity of the final transmitted beam, I_T. We return to this when multiple beam interference and the Fabry Perot interferometer are discussed.

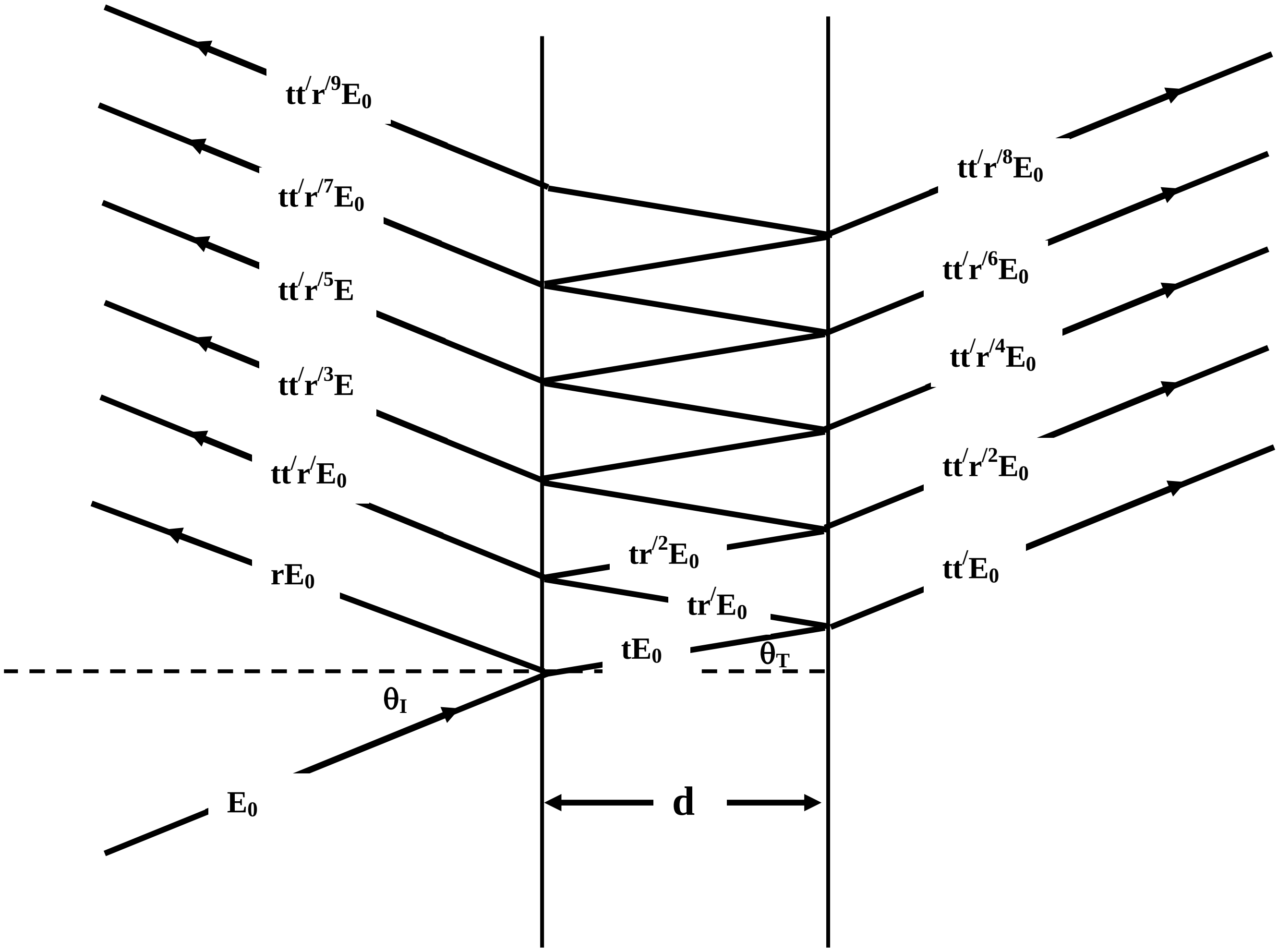
Multiple Beam Interference.

When considering the thin dielectric film interference it was recognized that multiple reflections could occur and this possibility is now explored.



refractive index n_F embedded in a medium of refractive index n_1. Further, we consider th

Consider the above system of a thin dielectric film of thickness t and e film to be non absorbing at wavelengths of interest. An electromagnetic wave impinges from the left hand side and is partially transmitted and partially reflected undergoing a series of subsequent reflections and transmissions at the left hand and right hand interface. To find the total reflected or transmitted fields we need to add all of the reflected or transmitted waves taking into account phase shifts between the waves as usual. We denote the fraction of the electric field amplitude transmitted on entering the film as t and on leaving the film as $t^{/}$ and the fraction reflected at the external interface as r and at the internal interface as $r^{/}$. These quantities were discussed earlier when the Fresnel equations were established and they depend, in general on the angle of incidence and the difference in refractive indices at the interfaces. In particular we established earlier, using an argument due to Stokes, that $r(\theta_I)= -r^{/}(\theta_T)$, where θ_I and θ_T come as a pair of angles related by Snell's law.



The above diagram establishes the sequence of reflected partial ***amplitudes*** and of transmitted partial ***amplitudes***. The reflected partial amplitudes are as follows

rE_0

$tt^{/}r^{/}E_0$

$tt^{/}r^{/3}E_0$

$tt^{/}r^{/5}E_0$

$tt^{/}r^{/7}E_0$

$tt^{/}r^{/9}E_0$

etc.

$tt^{/}r^{/(2n+1)}E_0$

The total reflected amplitude is simply the sum of the partial amplitudes ***with account taken of phase differences.*** We can assume that the polarization is unchanged upon reflection/transmission as the rays remain parallel and we can for this reason safely treat the electric field as a scalar.

Reflected rays.

When summing the reflected rays we need to note

(i) All but the first reflection only undergo reflections inside the film. There is therefore a phase difference between the first and the rest manifest in the minus sign in the relation between r and $r^{/}$.

(ii) Each other reflected ray undergoes an odd integer multiple of internal reflections , 1, 3, 5 etc.

(iii) The geometric path difference between adjacent reflected rays in the above diagram is $2d\,cos\,\theta_T$ and the phase difference $\Delta\phi = \dfrac{4\pi}{\lambda_0} n_F\, d\, cos\,\theta_T$.

(iv) The internal incidence angle, $\theta_T < \theta_C$
ie. we have light escaping the film and so no total internal reflection and thus the internal angle is less than the critical angle. In this case for light polarised perpendicular to the plane of incidence (the page) the phase change on reflection is zero. For light polarised parallel to the plane of incidence the phase change is either zero or π. There are an even number of additional reflections for each adjacent ray and therefore a phase shift of 2π, ie. the same relative phase and thus no phase change.

We may now choose to examine some simple situations as follows;

1. When the optical path difference (geometric path difference multiplied by refractive index of the medium) between two adjacent rays is equal to an integer multiple of wavelengths

$$\Lambda = 2n_F d \cos\theta_T = m\lambda_0 \qquad (5.53a)$$

Otherwise stated

$$\Delta\phi = \frac{4\pi}{\lambda_0} n_F\, d\, cos\,\theta_T = \pm 2m\pi \qquad (5.53b)$$

In this circumstance there is no effective phase difference between the reflected rays other than the first with a phase difference of π taken care of by a minus sign. The sum of electric fields is then

$$E_{0r} = rE_0 - \left(tt^{/}rE_0 + tt^{/}r^3E_0 + tt^{/}r^5E_0 + tt^{/}r^7E_0 + tt^{/}r^9E_0 + ...\right) \quad (5.54)$$

$$E_{0r} = rE_0 - tt^{/}rE_0\left(1 + r^2 + r^4 + r^6 + r^8 + ...\right) \quad (5.55)$$

Recalling that a geometric progression is

$$1 + a + a^2 + a^3 + a^4 + ... = \frac{1}{1-a} \quad (5.56)$$

The term in brackets on the RHS of 5.55 is a geometric progression with a = r^2 thus

$$E_{0r} = rE_0 - \frac{tt^{/}rE_0}{1-r^2} \quad (5.57)$$

Recalling from the Stokes relation that

$$tt^{/} = 1 - r^2 \quad (5.58)$$

We get the reflected field as $E_{0r} = 0$

The condition that the optical path difference between adjacent rays is an integer multiple of λ means that the first reflected ray is exactly cancelled by the sum of second, third and subsequent reflections. In the absence of any absorption this is precisely the condition for all of the incident power to be transmitted.

2. The second special case we can examine is where $\Lambda = n_F\left(m + \frac{1}{2}\right)\lambda_0$

$$\Delta\phi = \frac{4\pi}{\lambda_0} n_F d \cos\theta_T = \pm(2m+1)\pi$$

This leaves the first and second reflected waves in phase and all other partial waves $\lambda/2$ out of phase ***with their adjacent partial waves***. The resultant scalar amplitude is now

$$E_{0r} = rE_0 - \left(-tt^/ rE_0 + tt^/ r^3 E_0 - tt^/ r^5 E_0 + tt^/ r^7 E_0 - tt^/ r^9 E_0 + ...\right) \tag{5.59}$$

Rewritten,

$$E_{0r} = rE_0 + tt^/ rE_0 \left(1 - r^2 + r^4 - r^6 + r^8 - .....\right) \tag{5.60}$$

The series in brackets is $1 - r^2 + r^4 - r^6 + r^8 - ... = \dfrac{1}{1 + r^2}$ ie. another geometric progression where we have used $a = -r^2$ for our geometric progression.

And now

$$E_{0r} = rE_0 \left(1 + \frac{tt^/}{1 + r^2}\right) \tag{5.61}$$

Again using the Stokes relation

$$E_{0r} = \frac{2r}{1 + r^2} E_0 \tag{5.62}$$

This situation, $2d \cos\theta_T = n_F \left(m + \dfrac{1}{2}\right)\lambda_0$, gives the maximum reflected and minimum transmitted wave. We find the intensity by the usual means of taking the time average of the square of the electric field divided by the impedance of the medium in which the wave is propagating, η_1.

$$I_r = \frac{4r^2}{\left(1 + r^2\right)^2} \frac{E_0^2}{2\eta_1} = \frac{4r^2}{\left(1 + r^2\right)^2} I_I = \frac{4R}{(1 + R)^2} I_I \tag{5.63}$$

Where we have used the fact that the intensity reflection coefficient, R, and the amplitude reflection coefficient, r , are related by $r^2 = R$

3. We now consider the general case and include the amplitude ***and the phase*** explicitly where the reflected partial amplitudes including the phases are now

$rE_0\exp(j\omega t)$

$tt^{/}r^{/}E_0 \exp j(\omega t-\Delta\phi)$

$tt^{/}r^{/3}E_0 \exp j(\omega t-2\Delta\phi)$

$tt^{/}r^{/5}E_0 \exp j(\omega t-3\Delta\phi)$

.

.

.

$tt^{/}r^{/(2N-3)}E_0 \exp j(\omega t-(N-1)\Delta\phi)$

Writing, as usual,

$$\Delta\phi = k_0\Lambda + \delta = \frac{2\pi}{\lambda_0}2n_F d\cos\theta_T + \delta \qquad (5.64)$$

as the total phase difference due to the path difference ***between adjacent reflected waves*** , $2n_Fd\cos\theta_T$,plus the phase difference, δ, due to the extra reflection suffered by each successive reflected wave.

$$E_{0r} = rE_0e^{j\omega t} + tt^{/}r^{/}E_0e^{j(\omega t-\Delta\phi)} + tt^{/}r^{/3}E_0e^{j(\omega t-2\Delta\phi)} + tt^{/}r^{/5}E_0e^{j(\omega t-3\Delta\phi)} + ... \qquad (5.65)$$

Rewritten as

$$E_{0r} = E_0e^{j\omega t}\left[r + tt^{/}r^{/}e^{-j\Delta\phi}\left(1 + (r^{/2}e^{-j\Delta\phi}) + (r^{/2}e^{-j\Delta\phi})^2 + (r^{/2}e^{-j\Delta\phi})^3 + ... + (r^{/2}e^{-j\Delta\phi})^{N-2}\right)\right] \qquad (5.66)$$

The series on the RHS converges if $\left|r^{/2}e^{-j\Delta\phi}\right| < 1$

The geometric progression in the curved brackets on the RHS of 5.66 has $a = r^{/2}e^{-j\Delta\phi}$ and the total reflected field may be compactly rewritten as

$$E_{0r} = E_0e^{j\omega t}\left[r + \frac{tt^{/}r^{/}e^{-j\Delta\phi}}{1 - r^{/2}e^{-j\Delta\phi}}\right] \qquad (5.67)$$

We can again use the Stokes relations, $r=-r^{/}$ and $tt^{/}=1-r^2$ to obtain the reflected field in a more compact form as follows

$$E_{0r}=E_0e^{j\omega t}\left[r-\frac{r\left(1-r^2\right)e^{-j\Delta\phi}}{1-r^2e^{-j\Delta\phi}}\right]$$

Getting everything over a common denominator

$$E_{0r}=E_0e^{j\omega t}\left[\frac{r\left(1-r^2e^{-j\Delta\phi}\right)-r\left(1-r^2\right)e^{-j\Delta\phi}}{1-r^2e^{-j\Delta\phi}}\right]$$

Finally

$$E_{0r}=E_0e^{j\omega t}\left[\frac{r\left(1-e^{-j\Delta\phi}\right)}{1-r^2e^{-j\Delta\phi}}\right] \tag{5.68}$$

The reflected intensity is as usual given by the time average of the square of the electric field divided by the impedance. There is only the one time dependent exponential and the time average will as usual provide a factor $1/2$ thus

$$I_{0r}=\frac{E_{0r}E_{0r}^*}{2\eta_1}=\frac{r^2E_0^2}{2\eta_1}\frac{\left(1-e^{-j\Delta\phi}\right)\left(1-e^{+j\Delta\phi}\right)}{\left(1-r^2e^{-j\Delta\phi}\right)\left(1-r^2e^{+j\Delta\phi}\right)} \tag{5.69}$$

Noting that the incident light intensity $I_i=\dfrac{E_0^2}{2\eta_1}$ and also using de Moivre's theorem this can be written more compactly as

$$I_{0r}=I_i\frac{2r^2\left(1-\cos\Delta\phi\right)}{\left(1+r^4\right)-2r^2\cos\Delta\phi}=I_i\frac{2R(1-\cos\Delta\phi)}{\left(1+R^2\right)-2R\cos\Delta\phi} \tag{5.70}$$

We can do the same for the transmitted waves by adding all the partial transmitted waves

$$E_{0t} = tt^{/}E_0 e^{j\omega t} + tt^{/}r^{/2}E_0 e^{j(\omega t - \Delta\phi)} + tt^{/}r^{/4}E_0 e^{j(\omega t - 2\Delta\phi)} + ... + tt^{/}r^{/(2N-1)}E_0 e^{j(\omega t - (N-1)\Delta\phi)} \tag{5.71}$$

$$E_{0t} = tt^{/}E_0 e^{j\omega t}\left[1 + r^{/2}e^{-j\Delta\phi} + r^{/4}e^{-j2\Delta\phi} + ... + r^{/(2N-1)}e^{-j(N-1)\Delta\phi}\right] \tag{5.72}$$

The geometric progression in square brackets on the RHS has $a = r^{/2}e^{-j\Delta\phi} = r^2 e^{-j\Delta\phi}$ again and 5.72 may be compactly rewritten as

$$E_{0t} = E_0 e^{j\omega t}\left[\frac{1 - r^2}{1 - r^2 e^{-j\Delta\phi}}\right] \tag{5.73}$$

Multiplying E by it's complex conjugate and dividing by $2\eta_1$ gives the intensity as usual and for the transmitted intensity we obtain

$$I_{0t} = \frac{I_i(1 - r^2)^2}{\left(1 + r^4\right) - 2r^2\cos\Delta\phi} = I_i \frac{\left(tt^{/}\right)^2}{1 + R^2 - 2R\cos\Delta\phi} \tag{5.74}$$

Looking at 5.70 and 5.74 it is difficult to interpret what is occurring physically. To make each of these expressions more transparent we finally use the trigonometric identity, $\cos\theta = 1 - 2\sin^2\frac{\theta}{2}$ to get expressions for the reflected and transmitted intensities from 5.70 and 5.74

$$I_r = I_i \frac{\left[\frac{2r}{1 - r^2}\right]^2 \sin^2\frac{\Delta\phi}{2}}{1 + \left[\frac{2r}{1 - r^2}\right]^2 \sin^2\frac{\Delta\phi}{2}} \tag{5.75}$$

And

$$I_t = I_i \frac{1}{1 + \left[\frac{2r}{1 - r^2}\right]^2 \sin^2\frac{\Delta\phi}{2}} \tag{5.76}$$

It is clear now that to maximize the transmitted intensity the denominator in 5.76 needs to be minimized, ie. the sinusoid in the denominator needs to be zero and therefore

$$\frac{\Delta\phi}{2} = m\pi \qquad (5.77)$$

for maximum transmission. Bearing in mind the relation, 5.64, between phase difference and wavelength, $\Delta\phi = k_0\Lambda + \delta = \frac{2\pi}{\lambda_0}2n_F d\cos\theta_T + \delta$, this is also a condition on wavelength;

By rearrangement of 5.64 with the condition 5.77 we obtain this condition

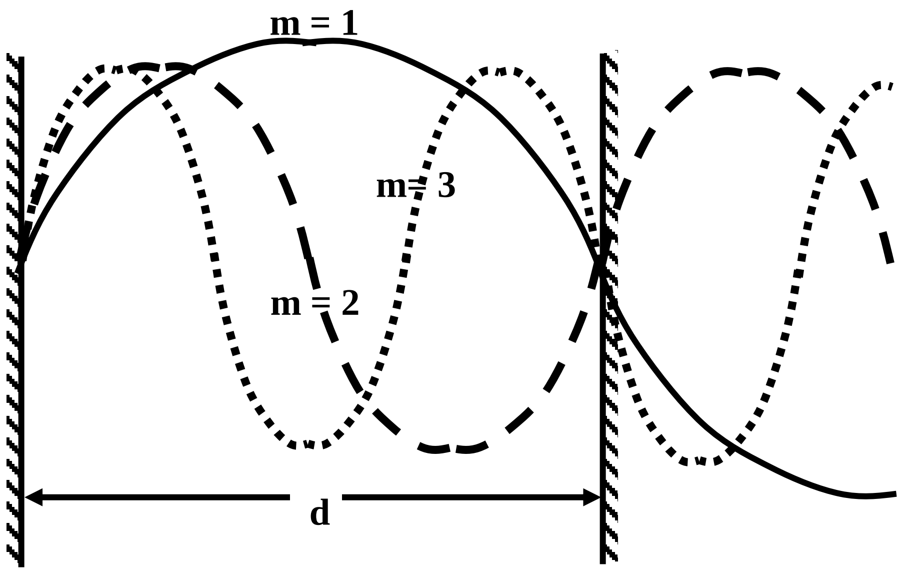
$$m\frac{\lambda_0}{2n_F} = m\frac{\lambda_F}{2} = 2d\cos\theta_T \qquad (5.78)$$

where we take δ ≈ 0.

The physical meaning of 5.78 becomes clear when we consider normal incidence on the film and 5.78 becomes

$$m\frac{\lambda_F}{2} = 2d \qquad (5.78a).$$

This is the requirement that half integer wavelengths of the wave fit between the two inner surfaces of the film, ie. that when standing waves are supported there is maximum transmission.



What is required to maximize the reflected intensity is not so clear but its minimum is zero where the sinusoid is zero in the numerator of 5.75, and therefore $\frac{\Delta\phi}{2} = m\pi$. Unsurprisingly we obtain minimum reflection where transmission is maximum. We will examine this more closely.

First we introduce a quantity that appears frequently from now on called the **coefficient of finesse,** F which we will see is a commonly used figure of merit for a Fabry Perot device

$$F = \left(\frac{2r}{1-r^2}\right)^2 = \frac{4R}{(1-R)^2} \tag{5.79}$$

And obtain a compact form for the reflected and transmitted intensities

$$I_r = I_i \frac{F\sin^2\frac{\Delta\phi}{2}}{1+F\sin^2\frac{\Delta\phi}{2}} \tag{5.80}$$

$$I_t = I_i \frac{1}{1+F\sin^2\frac{\Delta\phi}{2}} \tag{5.81}$$

In the absence of absorption as is the case here, conservation of energy requires that the transmitted and reflected intensities should add to the input intensity

$$I_i = I_t + I_r \tag{5.82}$$

Checking

$$I_t + I_r = I_i\left[\frac{1}{1+F\sin^2\frac{\Delta\phi}{2}} + \frac{F\sin^2\frac{\Delta\phi}{2}}{1+F\sin^2\frac{\Delta\phi}{2}}\right] = I_i \tag{5.83}$$

As noted earlier maximum transmission occurs where the denominator of 5.81 (or 5.76) is as small as possible which is when $\sin^2 \frac{\Delta\phi}{2} = 0$.

When this condition holds

(i) **the maximum transmitted intensity** is

$$(I_t)_{Max} = I_i \tag{5.84}$$

And

(ii) ***the minimum reflected intensity*** as we have seen before is

$$(I_r)_{Min} = 0$$

(iii) ***the minimum transmitted intensity*** occurs when the denominator of 5.79 (or 5.76) is as large as possible ie. for $\sin^2 \frac{\Delta\phi}{2} = 1$ or for

$$\frac{\Delta\phi}{2} = \left(m + \frac{1}{2}\right)\pi \qquad\qquad \Delta\phi = (2m+1)\pi \tag{5.85}$$

And the value of the minimum transmitted intensity using $\sin^2 \frac{\Delta\phi}{2} = 1$ in 5.79 is;

$$(I_t)_{Min} = I_i \frac{1}{1+F} = I_i \left[\frac{1}{1 + \frac{4R}{(1-R)^2}}\right] = I_i \left[\frac{(1-R)^2}{(1-R)^2 + 4R}\right] = I_i \left[\frac{(1-R)^2}{1 - 2R + R^2 - 4R}\right] \tag{5.86}$$

$$(I_t)_{Min} = I_i \frac{(1-R)^2}{(1+R)^2} \tag{5.87}$$

And finally

(iv) ***the maximum reflected intensity*** occurs when $\sin^2 \frac{\Delta\phi}{2} = 1$ and is

$$\left(I_r\right)_{Max} = I_i \frac{F}{1+F} = I_i \frac{\left(\frac{4R}{(1-R)^2}\right)}{1+\left(\frac{4R}{(1-R)^2}\right)^2} = I_i \frac{4R}{(1+R)^2} \tag{5.88}$$

The fractional reflected and transmitted intensities (compared to input) are given by 5.78 as

$$\frac{I_r}{I_i} = \frac{F\sin^2\frac{\Delta\phi}{2}}{1+F\sin^2\frac{\Delta\phi}{2}} \tag{5.89}$$

and

$$\frac{I_t}{I_i} = \frac{1}{1+F\sin^2\frac{\Delta\phi}{2}} = \mathscr{A}(\Delta\phi) \tag{5.90}$$

The function

$$\frac{1}{1+F\sin^2\theta} = \mathscr{A}(\theta) \tag{5.91}$$

appears elsewhere in physical problems and is a tabulated function called the Airy function. The fractional transmission and reflection are plotted below.

Summarising.

1. Reflection

First draw the diagram with the multiple reflections and then list and add together all of the partial reflected waves.

$$E_{0r} = E_0 e^{j\omega t}\left[r + tt^/ r^/ e^{-j\Delta\phi}\left(1 + (r^{/2} e^{-j\Delta\phi}) + (r^{/2} e^{-j\Delta\phi})^2 + (r^{/2} e^{-j\Delta\phi})^3 + ... + (r^{/2} e^{-j\Delta\phi})^{N-2}\right)\right]$$

Tidy up the geometric progression in curved brackets using $1 + a + a^2 + ... = \dfrac{1}{1-a}$

$$E_{0r} = E_0 e^{i\omega t}\left[r + \frac{tt^/ r^/ e^{-j\Delta\phi}}{1 - r^{/2} e^{-j\Delta\phi}}\right]$$

Use the Stokes relations, $tt^/ = 1 - r^2$ and $r = -r^/$ to tidy up

$$E_{0r} = E_0 e^{i\omega t}\left[\frac{r\left(1 - e^{-j\Delta\phi}\right)}{1 - r^2 e^{-j\Delta\phi}}\right]$$

Now find the intensity by the standard route;

$$I_{0r} = \frac{E_{0r}E_{0r}^*}{2\eta_1} = \frac{r^2 E_0^2}{2\eta_1}\frac{\left(1 - e^{-j\Delta\phi}\right)\left(1 - e^{+j\Delta\phi}\right)}{\left(1 - r^2 e^{-j\Delta\phi}\right)\left(1 - r^2 e^{+j\Delta\phi}\right)}$$

Tidy up using De Moivre

$$I_{0r} = I_i \frac{2r^2\left(1 - \cos\Delta\phi\right)}{\left(1 + r^4\right) - 2r^2\cos\Delta\phi} = I_i \frac{2R(1 - \cos\Delta\phi)}{\left(1 + R^2\right) - 2R\cos\Delta\phi}$$

Introduce the coefficient of Finesse; $F = \left(\dfrac{2r}{1 - r^2}\right)^2 = \dfrac{4R}{(1 - R)^2}$ and finish.

$$I_r = I_i \frac{F\sin^2\frac{\Delta\phi}{2}}{1+F\sin^2\frac{\Delta\phi}{2}}$$

2. Transmission

First draw the diagram with the multiple reflections and then list and add together all of the partial reflected waves.

$$E_{0t} = tt^/ E_0 e^{j\omega t} + tt^/ r^{/2} E_0 e^{j(\omega t - \Delta\phi)} + tt^/ r^{/4} E_0 e^{j(\omega t - 2\Delta\phi)} + ... + tt^/ r^{/(2N-1)} E_0 e^{j(\omega t-(N-1)\Delta\phi)}$$

$$E_{0t} = tt^/ E_0 e^{j\omega t}\left[1 + r^{/2} e^{-j\Delta\phi} + r^{/4} e^{-j2\Delta\phi} + ... + r^{/(2N-1)} e^{-j(N-1)\Delta\phi}\right]$$

Tidy up the geometric progression in curved brackets using $1 + a + a^2 + ... = \frac{1}{1-a}$

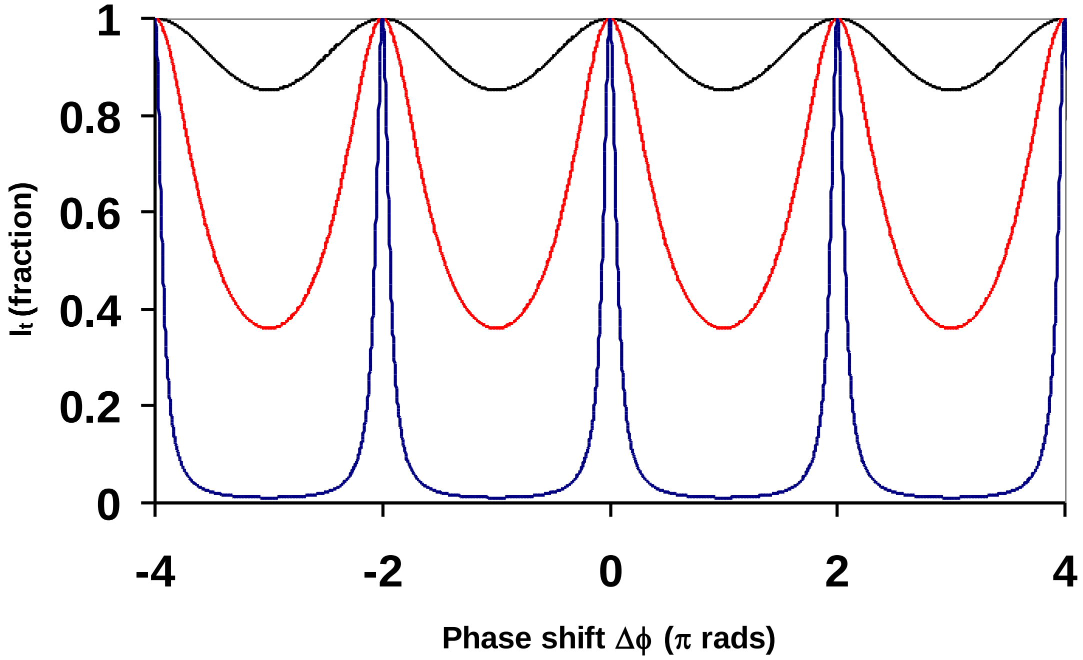
$$E_{0t} = E_0 e^{j\omega t}\left[\frac{1-r^2}{1-r^2 e^{-j\Delta\phi}}\right]$$

$$I_{0t} = \frac{I_i(1-r^2)^2}{\left(1+r^4\right) - 2r^2\cos\Delta\phi} = I_i \frac{\left(tt^/\right)^2}{1+R^2-2R\cos\Delta\phi}$$

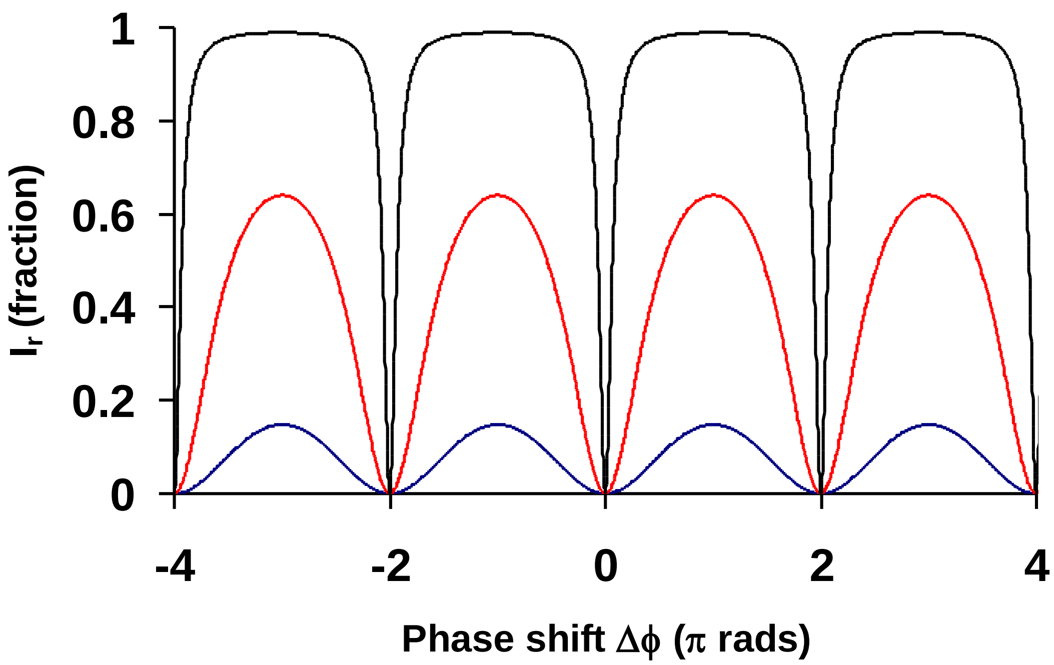
$$I_t = I_i \frac{1}{1+\left[\frac{2r}{1-r^2}\right]^2 \sin^2\frac{\Delta\phi}{2}}$$

$$I_t = I_i \frac{1}{1+F\sin^2\frac{\Delta\phi}{2}}$$

Transmission through thin film vs phase shift, r = 0.2, 0.5, 0.9



Reflection from thin film vs phase difference, r = 0.2, 0.5, 0.9



The reflection/transmission intensities from a thin dielectric film as described by equations 5.89 and 5.90 are shown in the two graphs above where three values of *r* are used, the sharpest minima/maxima occur for the largest values of *r*.

There is a maximum in the transmission where $\Delta\phi$ is an integer multiple of 2π and as *r* is increased this maximum becomes increasingly sharp around these values. The inverse is true of the reflection. Such behavior suggests potential use as a tuned bandpass filter.

It is important to pause and to recall once again what it is that $\Delta\phi$ represents physically. It represents the difference in phase between adjacent reflected (or transmitted) waves. and that phase difference is determined by the extra distance traveled by one wave with respect to the other and so varies with angle of incidence and film thickness. It also depends inversely on the wavelength of the light within the film, $\lambda = \lambda_0 / n_F$. This means that the horizontal axis, $\Delta\phi$, in the above graphs could as easily be a plot of either inverse wavelength or of angle of incidence. We see this when writing the phase difference explicitly as $\Delta\phi = 2dn_F \, 2\pi/\lambda_0 \cos\theta_T$.

In other words there are two variables that could be plotted on the x axis in place of $\Delta\phi$, namely by holding θ_T constant the inverse wavelength, λ_0^{-1} (or frequency $\nu = c\dfrac{1}{\lambda_0}$) could be plotted on the x axis with a series of wavelengths/frequencies at which sharp resonant transmission occurs given by

$$\Delta\phi = 2m\pi = 2dn_F \frac{2\pi}{\lambda_0}\cos\theta_T \,, \qquad \lambda_0 = \frac{2dn_F}{m}\cos\theta_T \,, \qquad \theta_T = cos^{-1}\left(\frac{m\lambda_0}{2dn_F}\right)$$

NB the resonant frequencies are equally spaced the wavelengths are not!

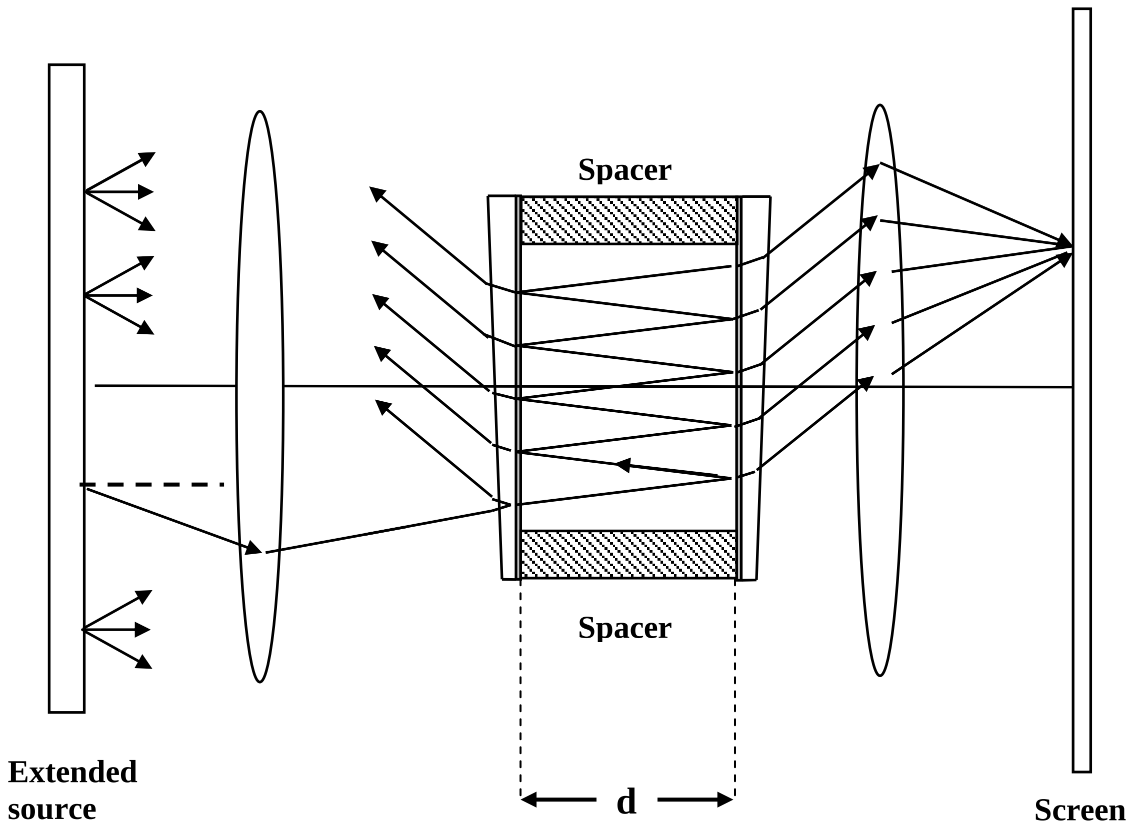
In the above equations the integer *m* is also known as the order of the transmission (or reflection). An extended white light source would then be split into transmitted spectral components where the wavelength criteria is satisfied. Depending on the order, m, this would be a wavelength band around some peak that is transmitted whilst the other wavelengths are reflected and lost until a second order (third order etc) transmission peak allows further transmission of another band of wavelengths around a second (third etc.) wavelength. The peak transmission (or reflection) will also depend on the viewing direction and it is this that gives rise to the colours observed in a thin oil film floating on water or the colours observed in a soap bubble.

Looking at the above graph for transmission through the thin film we see an increasingly narrow range of $\Delta\phi$ (or θ_T or λ_0) over which transmission occurs as r is increased. Ie. the transmission of the thin film may be highly tuned. This highly defined directionality and wavelength range, comes about due to the large number of coherent sources that contribute to the overall beam.

The thin film dielectric with multiple interference clearly demonstrates useful properties and the potential for constructing a useful device. **The Fabry Perot interferometer** is an engineered structure using the basic principles that have just been discussed and finds many uses in optics from spectroscopy through high resolution optical filters to laser resonators. For this reason it is worth examining in some detail.

b)Fabry Perot Interferometer.

One of the simplest realizations of an optical structure, using the principles of the multiple reflection interference of the thin film previously discussed, is to take a pair of parallel partially reflecting surfaces separated by a careful engineered spacer.



The above diagram shows such a structure, known as a Fabry Perot interferometer or Fabry Perot etalon, with a pair of metallised partially reflecting surfaces, in this case evaporated onto two transparent substrates. The reflecting surfaces are held precisely parallel to one another and usually the substrates will be slightly wedged on the non-metallised surface in order to suppress the formation of secondary parallel reflecting systems that would interfere with the operation of the primary system. The two reflecting surfaces of the primary system are held at a precise separation, d. There is in the example shown an extended light source to the left and a screen to the right. Choosing a ray from a given portion of the extended source traveling at an arbitrary angle, its progress is followed as it undergoes multiple reflection/transmission events before the transmitted rays from this one coherent point source ar collected by a second lens and brought to a focus at some point P on a viewing screen (photographic plate, retina etc.). Of course, any ray traveling at the same angle from an equivalent point on the extended source (at the same distance from the system axis) would have been brought to a focus on the screen at the same distance from the system axis resulting in the appearance of a series of concentric rings centered on that axis.

This system has the essential characteristics of the previous situation examined with the exception that;

(i) The metallised reflecting surfaces will be a source of dissipation/absorption and this will mean that the Stokes relations that were used frequently in the previous discussion no longer hold ie.

$I_i \neq I_r + I_t \qquad T + R \neq 1 \qquad$ and $tt^{/} + r^2 \neq 1$

Rather these are modified to account for the absorbance, A, in the following way;

$$I_i = I_r + I_t + I_A \qquad T + R + A = 1 \tag{5.92}$$

Where 5.92 again represents the conservation of energy but with the dissipation (absorption) term included.

(ii) The gap, *d*, is much larger than in our thin film scenario and can be from microns up to centimeters as it is now a factor under control of engineering.

and

(iii) Metallic films will introduce a phase shift, $\delta(\theta)$, upon reflection which may not be zero or π and that may depend on the angle of incidence, θ_T.

Now, the phase difference between two successive waves is as usual

$$\Delta\phi = \frac{4\pi n_F d}{\lambda_0}\cos\theta_T + 2\delta \tag{5.93}$$

If we consider the action to take place at angles close to normal incidence, ie. $\theta \approx 0$ then δ is approximately constant and because we are discussing a system where $d >> \lambda_0$ the first term on the RHS dominates and the second term, 2δ, may be neglected. This allows us to re-express the transmission as given for a thin film in 5.74 as

$$\frac{I_t}{I_i} = \frac{T^2}{1 + R^2 - 2R\cos\Delta\phi} \tag{5.94}$$

Or, using the trigonometric identity, $\cos\theta = 1 - 2\sin^2\frac{\theta}{2}$ to re-express 5.94 as

$$\frac{I_t}{I_i} = \left(\frac{T}{1-R}\right)^2 \frac{1}{1 + \left(4R\Big/(1-R)^2\right)\sin^2\frac{\Delta\phi}{2}} \tag{5.95}$$

This re-expression of 5.94 makes it much easier to understand what is happening with the transmission as it did with the thin film analysis as we simply get maxima when the sinusoid in the denominator goes to zero

We can use 5.80, $T = 1 - R - A$, to re-express this as

$$\frac{I_t}{I_i} = \left(1 - \frac{A}{1-R}\right)^2 \frac{1}{1 + \left(4R\Big/(1-R)^2\right)\sin^2\frac{\Delta\phi}{2}} = \left(1 - \frac{A}{1-R}\right)^2 \mathscr{A}(\theta) \tag{5.96}$$

Where $\mathscr{A}$ is the Airy function. For the case of zero absorption examined previously we had

$$\frac{I_t}{I_i} = \frac{1}{1 + \left(\frac{4R}{(1-R)^2} \right) \sin^2 \frac{\Delta\phi}{2}} \qquad = \mathscr{A}(\theta) \tag{5.90}$$

which is what 5.96 reverts to if we make $A = 0$.

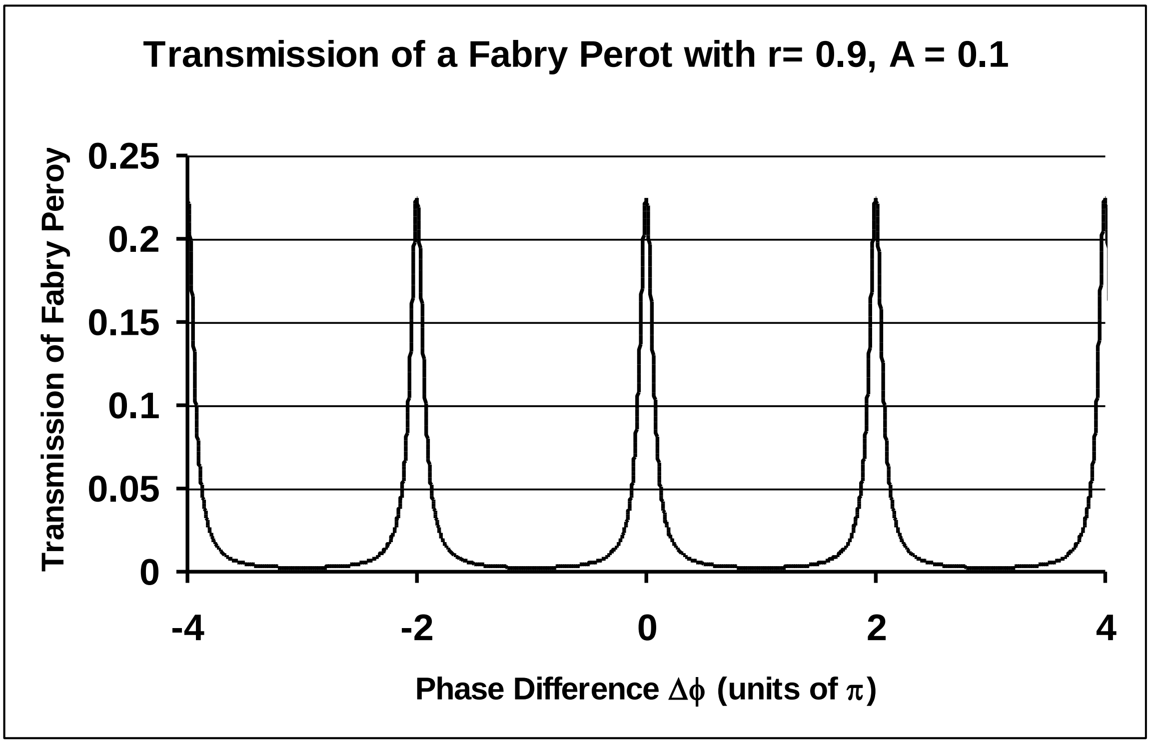
In the current case with absorption according to 5.96 we have maximum transmission when $\sin^2 \frac{\Delta\phi}{2} = 0$ or when $\Delta\phi = 2m\pi$ as previously only now the maximum transmission is reduced by the absorption to

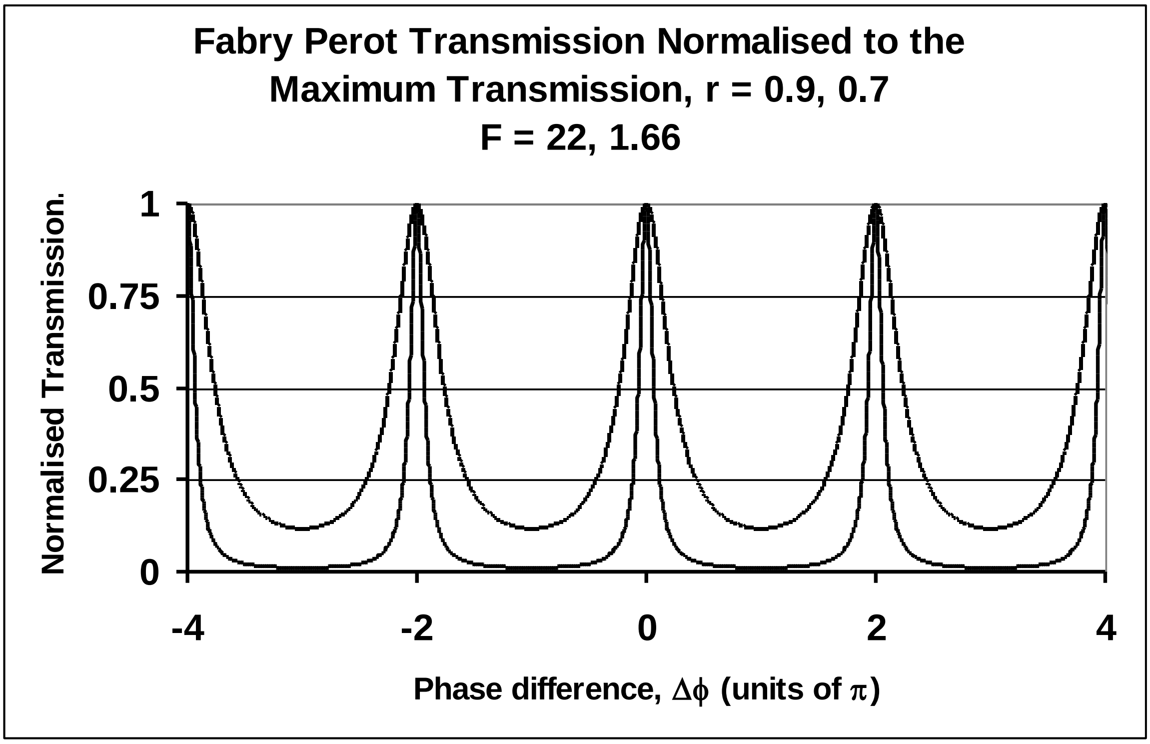
$$\left(\frac{I_t}{I_i} \right)_{Max} = \left(1 - \frac{A}{1-R} \right)^2 \tag{5.97}$$

We can use 5.97 to normalize the transmission to the maximum transmission as given by 5.97

$$\frac{I_t}{(I_t)_{Max}} \mathscr{A} \tag{5.98}$$

The normalized intensity is given by the Airy function as in the first of the two graphs shown below. Here, the effect of the absorbance has been included. The effect of A is to reduce the maximum value of the transmission.The graphs show the transmission of the Fabry Perot with an absorbance, $A = 0.1$ as a function of phase difference (equivalently, θ or λ_0). Also shown is the normalised transmission. Note in the first that the maximum is well below 1 as a result of the absorbance. In those two graphs the amplitude reflectance is changed from 0.9 to 0.7 and this has big effects on the transmission with the width of the transmission getting much larger and the transmission never dropping to zero in the case of a reflectance of 0.7.


Transmission of a Fabry Perot with r= 0.9, A = 0.1
Transmission of Fabry Peroy
0.25
0.2
0.15
0.1
0.05
0
-4
-2
0
2
4
Phase Difference Δϕ (units of π)


Fabry Perot Transmission Normalised to the Maximum Transmission, r = 0.9, 0.7
F = 22, 1.66
Normalised Transmission.
1
0.75
0.5
0.25
0
-4
-2
0
2
4
Phase difference, Δϕ (units of π)

When using the Fabry Perot as a filter or monochromator (wavelength selective device cf prisms and gratings) several important questions arise concerning **the resolution of the Fabry Perot**, ie. how well defined is the transmitted wavelength/frequency?
To answer this we need to look at a number of things.

i) **<u>The free spectral range.</u>** This is the separation of the transmission peaks of adjacent orders which occur at $\Delta\phi = 2m\pi$ so peaks are separated in phase by 2π. The separation of the peaks of adjacent modes in terms of wavelength is called the ***free spectral range*** of the Fabry Perot and indicates the wavelength range over which the interferometer may be operated without adjacent modes mixing. We see how this matters by considering operating the interferometer as a variable plate separation device and as d is altered $\Delta\phi$ will change with it until at some point $\Delta\phi$ satisfies the condition for maximum transmission and the light is transmitted. We know d and $\Delta\phi$ and should therefore know λ. But there is a problem as we don't know m. There is a second transmission peak in the neighbourhood of this one where the order number (or mode number) is $m + 1$ and we may have a wavelength corresponding to λ_{m+1} and not λ_m. To avoid any such confusion we need to know how far apart the two transmission windows are in terms of wavelength (or frequency). The peak separation or free spectral range is simply found as

$$\lambda_m - \lambda_{m+1} = \frac{2n_F d\cos\theta_T}{m} - \frac{2n_F d\cos\theta_T}{m+1} = \frac{2n_F d\cos\theta_T}{m(m+1)} = \frac{\lambda_m}{m+1} = \Delta\lambda_{FSR} \tag{5.100}$$

The free spectral range in terms of frequency is

$$\nu_{m+1} - \nu_m = \frac{c}{\lambda_{m+1}} - \frac{c}{\lambda_m} = \frac{c(m+1)}{2n_F d\cos\theta_T} - \frac{cm}{2n_F d\cos\theta_T} = \frac{c}{2n_F d\cos\theta_T} = \frac{1}{T} \tag{5.101}$$

$$\Delta\nu_{FSR} = \frac{c}{2n_F d\cos\theta_T} = \frac{c}{m(m+1)\Delta\lambda_{FSR}} \tag{5.101a}$$

Generally the device is used with an air gap and at normal incidence in which case 5.100 simplifies to

$$\Delta\lambda_{FSR} = \frac{2d}{m(m+1)} = \frac{\lambda_m}{m+1}$$

And 5.101 simplifies to

$$\Delta\nu_{FSR} = \frac{c}{2d} = \frac{c}{m(m+1)\Delta\lambda_{FSR}} = \frac{1}{T}$$

Where T is the time taken for the wave to travel from one reflecting surface to the other and back again, also known as the round trip time within the interferometer.
If we make the free spectral range larger than the interval over which we are looking for any emission the problem of confusing orders can be avoided.

ii) How wide is the window of transmission

As a measure of the width of the transmission window we look for the value of $\Delta\phi$ where the maximum transmission has dropped by half, $\Delta\phi_{1/2}$ and the width of the transmission window is defined as twice this ie Full Width at Half Maximum, $\Delta\phi_{FWHM} = 2\times\Delta\phi_{1/2}$, a commonly employed criterion.

The curve is simply the Airy function $\mathscr{A}(\theta)$ and we need to know where this has dropped by 50%.

$$\mathscr{A}\left(\frac{\Delta\phi_{1/2}}{2}\right) = \frac{1}{1+\left(4R/(1-R)^2\right)\sin^2\frac{\Delta\phi_{1/2}}{2}} = \frac{1}{1+F\sin^2\frac{\Delta\phi_{1/2}}{2}} = 0.5 \tag{5.102}$$

Re arranging to make $\Delta\phi_{1/2}$ the subject of the equation

$$F\sin^2\frac{\Delta\phi_{1/2}}{2} = 1 \tag{5.103}$$

$$\sin\frac{\Delta\phi_{1/2}}{2} = \sqrt{\frac{(1-R)^2}{4R}} \tag{5.104}$$

$$\Delta\phi_{1/2} = 2\sin^{-1}\sqrt{\frac{1-R^2}{4R}} = 2\sin^{-1}\sqrt{\frac{1}{F}} \tag{5.105}$$

For large coefficient of finesse, F, the approximation

$$\sin^{-1}\sqrt{\frac{1}{F}} \approx \sqrt{\frac{1}{F}}$$

may be used

$$\Delta\phi_{1/2} = 2\sqrt{\frac{1-R^2}{4R}} = 2\sqrt{\frac{1}{F}} \tag{5.106}$$

And the width of the peak, at half the maximum is

$$\Delta\phi_{FWHM} = 2\Delta\phi_{1/2} = \frac{4}{\sqrt{F}}$$

It is the case that the phase cannot be measured and only relative phases have any practical effect. We are more interested in asking questions about wavelength or frequency when discussing light waves. These are, of course, related to the phase as seen many times, $\Delta\phi = 2dn_F\,\frac{2\pi}{\lambda_m}\cos\theta_T = m2\pi$ for maximum transmission of the m^th^ order wavelength λ_m. If we know about the phase but wish to know about the wavelength we proceed as follows;

$$\frac{d(\Delta\phi)}{d\lambda_0} = -4\pi n_F\, d\,\frac{1}{\lambda_m^2} \tag{5.107}$$

$$\Delta\lambda_{1/2} = -\lambda_m^2 \frac{\Delta\phi_{1/2}}{4\pi n_F d}\cos\theta_T = -\lambda_m^2 \frac{1}{2\pi n_F d}\sqrt{\frac{1}{F}}\cos\theta_T \qquad (5.108)$$

Using the previously established condition for maximum transmission of m^{th} order

$$\lambda_m = \frac{2n_F d\cos\theta_T}{m}$$

We may simplify 5.108

$$\Delta\lambda_{1/2} = -\lambda_m \frac{1}{m\pi}\sqrt{\frac{1}{F}}\cos\theta_T \qquad (5.109)$$

And used in normal incidence

$$\Delta\lambda_{1/2} = -\lambda_m \frac{1}{m\pi}\sqrt{\frac{1}{F}} \qquad (5.109a)$$

And the resolution of the Fabry Perot used in m$^{\text{th}}$ order is defined as

$$\frac{\lambda_m}{\Delta\lambda_{FWHM}} = \frac{m\pi}{2}\sqrt{F}\sec\theta_T \qquad (5.110)$$

And used in normal incidence

$$\frac{\lambda_m}{\Delta\lambda_{FWHM}} = \frac{m\pi}{2}\sqrt{F} \qquad (5.110a)$$

We may want this in terms of frequency where we use the relationship between frequency and wavelength, $\lambda_m \nu_m = c$ for the m^{th} order. Using a procedure identical to that used to translate from phase to wavelength we translate from wavelength to frequency as follows;

$$\frac{d\nu_m}{d\lambda_m} = -\frac{c}{\lambda_m^2}$$

$$\Delta\nu_{FWHM} = -\frac{c}{\lambda_m^2}\Delta\lambda_{FWHM} = -\frac{c}{\lambda_m}\frac{\Delta\lambda_{FWHM}}{\lambda_m} = -\nu_m\frac{2}{m\pi}\frac{1}{\sqrt{F}}\cos\theta_T \qquad (5.111)$$

Therefore the resolution in m^{th} order is as previously found for this quantity in terms of wavelength

$$\frac{\nu_m}{\Delta\nu_{FWHM}} = \frac{m\pi}{2}\sqrt{F}\sec\theta_T \qquad (5.112)$$

NB the minus signs that occur in these equations when differentiating has been subsequently ignored as the meaning is simply that for example on going from λ to $\lambda + \Delta\lambda$ the frequency will go from ν to $\nu - \Delta\nu$ and in this context it carries no significance.

iii) What is a good figure of merit when defining the performance of a Fabry Perot? Comparing the separation of peaks or free spectral range with the width of a single peak we may find a figure of merit for the resolution of the Fabry Perot.

The ratio of peak separation to peak width is known as the finesse, $\mathscr{F}$,

$$\mathscr{F} = \frac{\Delta\lambda_{FSR}}{2\Delta\lambda_{1/2}} = \frac{\lambda_m/(m+1)}{2\lambda_m\frac{1}{m\pi}\sqrt{\frac{1}{F}}} = \frac{\pi\sqrt{F}}{2} \qquad (5.113)$$

Where we assume $m >> 1$

We see by comparing the frequency resolution expression 15.96 And the definition of finesse, $\mathscr{F}$ that the frequency resolution of the m^{th} order may also be written as

$$\frac{\nu_m}{\Delta\nu_{FWHM}} = \frac{m\pi}{2}\sqrt{F}\sec_T = m\mathscr{F}\sec\theta_T \qquad (5.114)$$

For normal incidence

$$\frac{\nu_m}{\Delta\nu_{FWHM}} = \frac{m\pi}{2}\sqrt{F} = m\mathscr{F} \qquad (5.114a)$$

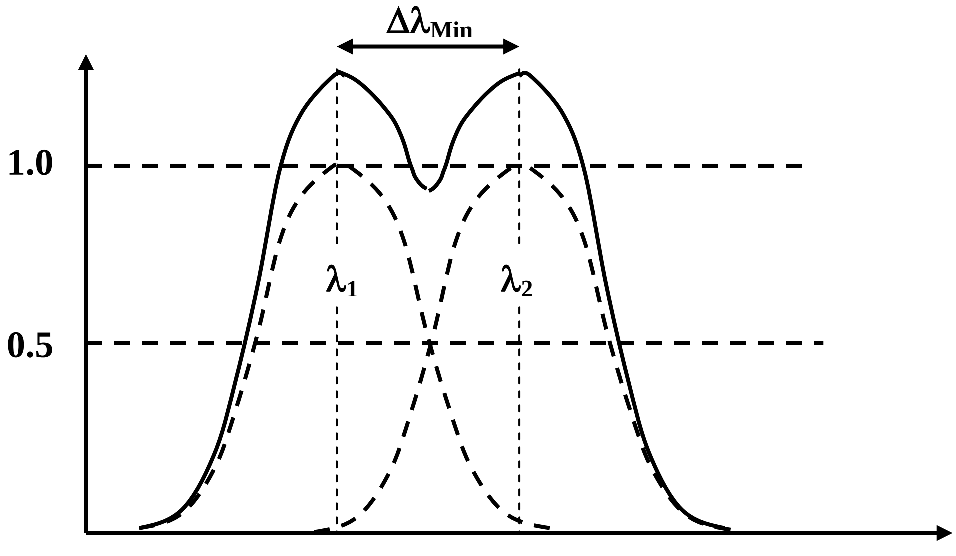
and

$$\frac{\lambda_m}{\Delta\lambda_{FWHM}} = \frac{m\pi}{2}\sqrt{F} = m\,\mathscr{F} \qquad (5.110b)$$

NB ***The finesse*** and ***the coefficient of finesse*** , while closely related are not the same thing!!

iv) What is the resolving power, $\mathcal{R}$, of the Fabry Perot?

This is a question that asks, " if we have a source of two wavelengths very close together what is the minimum wavelength separation, $\Delta\lambda_{Min}$ that a Fabry Perot would be able to "see""?



The criteria for being able to resolve two slightly different wavelengths is that their half maximum height values just cross as shown in the diagram above.

$$\lambda_1 + \Delta\lambda_{1_{1/2}} = \lambda_2 - \Delta\lambda_{2_{1/2}}$$

Therefore

$$\Delta\lambda_{Min} = 2\Delta\lambda_{1/2} = 2\lambda_m \frac{1}{m\pi}\sqrt{\frac{1}{F}}$$

The resolving power is simply defined as

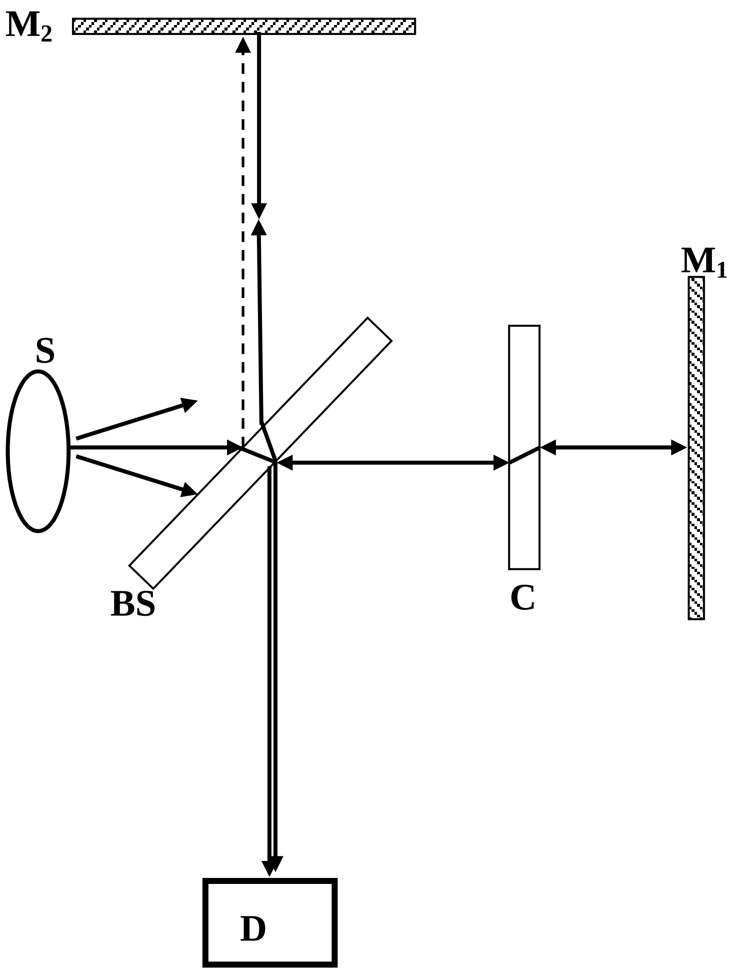
$$\mathcal{R} = \frac{\lambda}{\Delta\lambda_{Min}} = \frac{\lambda_m}{\Delta\lambda_{Min}} = \frac{m\pi\sqrt{F}}{2} = m\mathcal{F}$$

Finally, if we know F or $\mathcal{F}$ we can find the minimum resolvable wavelength difference for a given order, m

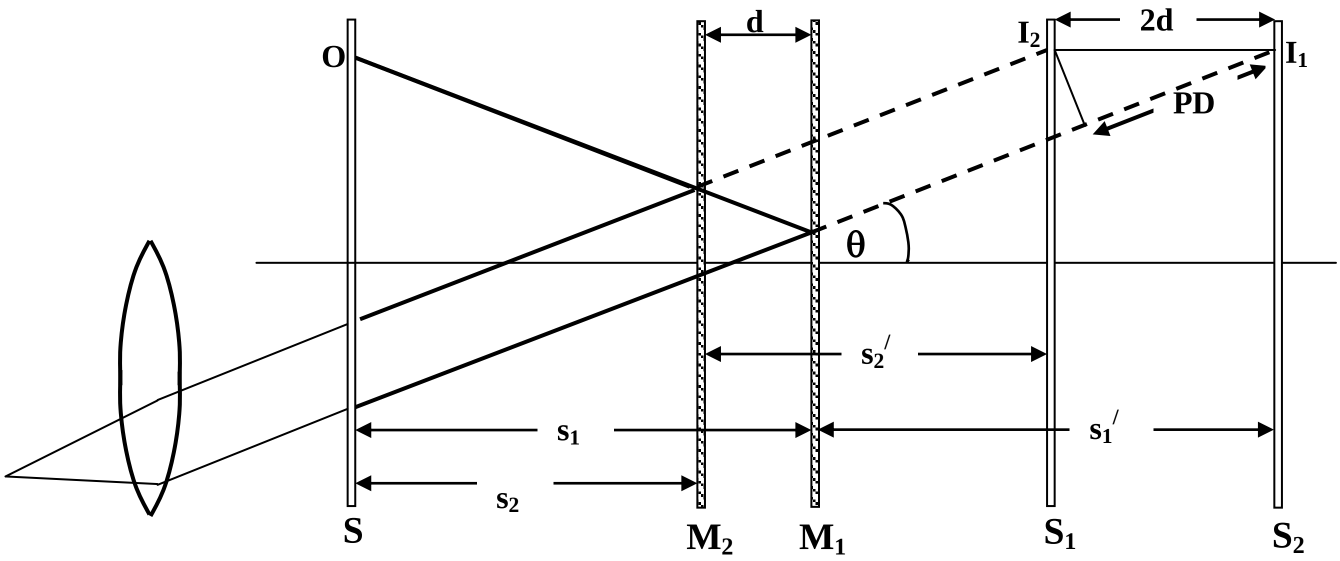
$$\Delta\lambda_{Min} = \frac{2\lambda_m}{m\pi\sqrt{F}} = \frac{\lambda_m}{m\mathcal{F}}$$

Michelson Interferometer.

Another well known amplitude splitting device for recombining two beams to produce interference is the Michelson interferometer.



The basic structure of a Michelson interferometer is laid out in the above figure. Light from a light source, S is incident upon a 50:50 beam splitter, BS, inclined at 45^0. The beam is split into two beams of equal intensity one traveling to mirror M_1 , the other traveling to mirror M_2 before both are reflected back to the beam splitter and on to a detector, D. One of the mirrors should be on a moveable track. A compensator, C, is placed in one of arm of the interferometer to allow for the fact that the beam traveling via M_2 has to traverse the beam splitter twice whereas the beam traveling to M_1 traverses the beam splitter only once, the compensator is to ensure that the optical path lengths are equal. At the detector interference fringes are formed. Considering the symmetry of the situation we can see that these will be in the form of concentric circles of light and dark. In order to understand these fringes we need to construct a virtual lay out of the optical components from the point of view of an observer looking from the detector towards the beam splitter. The observer will see a co-linear view of the source, the two mirrors and the image of the source in each mirror as shown below.



Taking an arbitrary point on the source to act as an object O that is an individual source of spherical waves and choosing an arbitrary ray from this point source to follow around the system, the two mirrors with a path difference of d = $s_1 - s_2$ will give rise to two virtual images, I_1 and I_2 that will behave as two coherent point sources to be recombined at the detector. The irradiance will be given therefore by 5.18. The path difference, PD, between the two virtual sources is

$$PD = 2d\cos\theta \qquad \Delta\phi = nk_0 2d\cos\theta = 2d\frac{2\pi}{\lambda_0}\cos\theta \qquad (5.115)$$

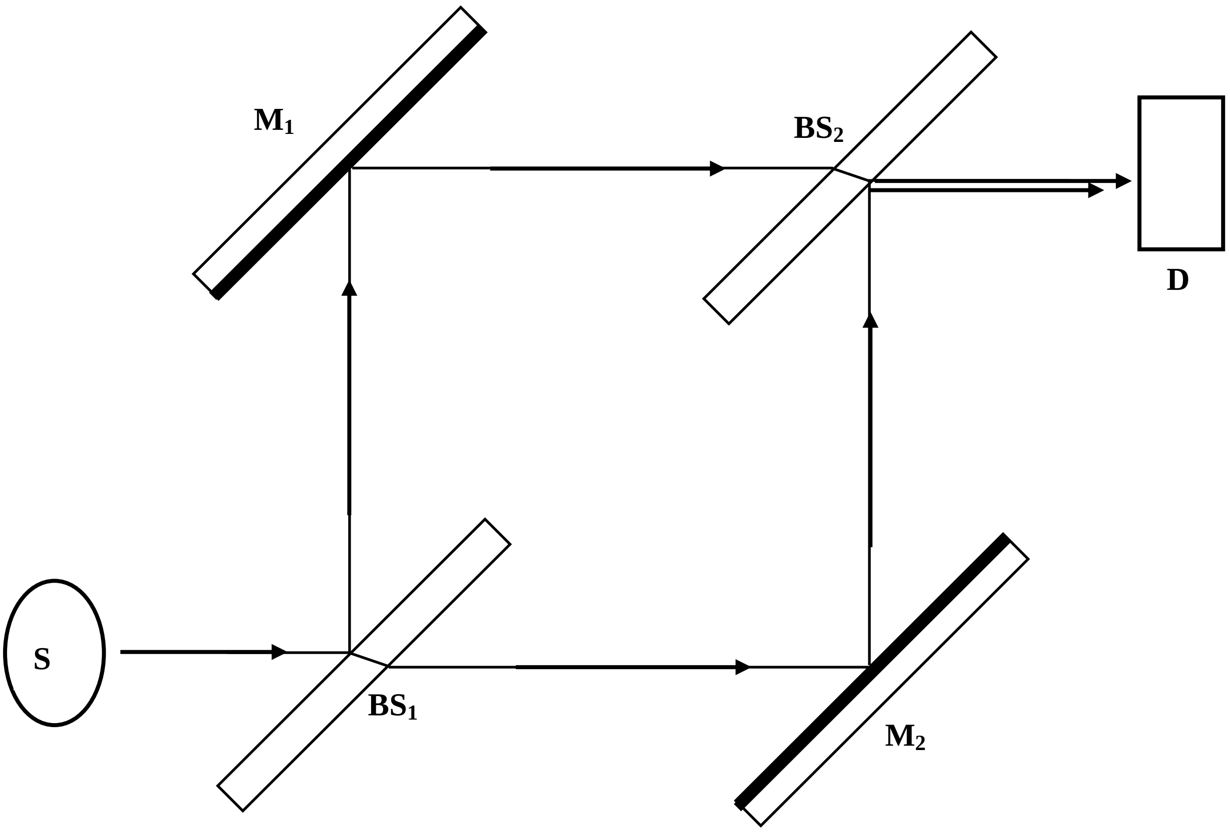
For constructive interference $\cos\Delta\phi = +1$ or $\Delta\phi = \pm 2m\pi$, thus for constructive interference the condition is

$$2d\cos\theta = \pm m\lambda_0$$

Any point source object O at the same radius from the centre of the source will satisfy the same equation and the result is a series of concentric rings at the detector of light and dark.

c) Mach Zhender Interferometer.

The Mach Zhender interferometer is another simple design that achieves interference effects by division of amplitude. It has similarities with the Michelson previously examined.



There are two arms that take light from the source to the detector, the amplitude splitting occurring at the first beam splitter, BS_1 . If the lengths of the two arms are kept equal there will be no phase difference, $\Delta\phi = 0$, and constructive interference occurs at the detector. The use of the device comes in applying it to analyse phase changes in some experimental system that can be placed in one of the arms. If the experimental system causes a change in the phase of the light traveling in that arm there will be a phase difference, $\Delta\phi \neq 0$ and interference fringes will be detected by the detector. The fringes can be used to analyse the experimental system and how changes made to that system affect the interaction of the system on light.