OPTICS

P. Ewart

1. Geometrical Optics

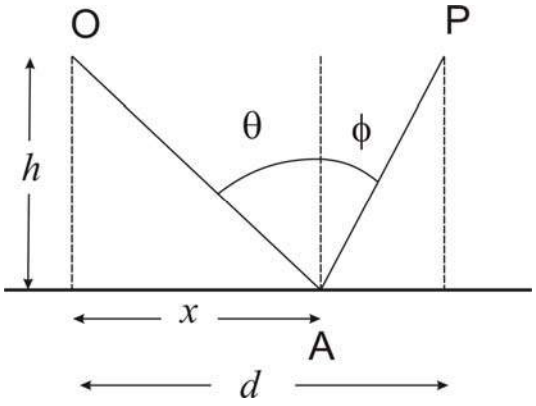
1.1Fermat's Principle

Light has been studied for a long time. Archimedes and other ancient Greek thinkers made original contributions but we mention here Heron of Alexandria (c. 10 - 75 AD) as he was the first to articulate what has come to be known as Fermat's Principle. Fermat, stated his principle as "Light travelling between two points follows a path taking the least time." The modern, and more correct version, is as follows:

"Light propagating between two points follows a path, or paths, for which the time taken is an extremum."

The principle has a theoretical basis in the quantum theory of light that avoids the question of how the light "knows" what direction to go in so that it will follow the maximal path! [Basically the wave function for the light consists of all possible paths but all, except the one corresponding to the classical path, destructively interfere owing to variations in the phase over the different paths.]

Fermat's principle is the basis of Geometrical optics which ignores the wave nature of light. The principle may be used to derive Snell's Laws of reflection and refraction.



Optical path length OAP = L, given by:

$$L = (x^2 + h^2)^{1/2} + [(d - x)^2 + h^2]^{1/2}$$

For a maximum or minimum

$$\frac{dL}{dx} = 0$$

from which we find $x = d/2$

Hence the incidence angle θ = reflection angle ϕ: Snell's law of reflection.

Using a similar procedure we can derive Snell's law of refraction:

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

where θ_1 and θ_2 are the angles between the light ray and the normal to the surface between media of refractive index n_1 and n_2 respectively.

Geometrical optics uses the effective rule of thumb that light travels in straight lines in a homogeneous medium of uniform refractive index. Deviations occur at boundaries between media of different refractive index or if the index varies in space. The path of light indicated by a ray can be plotted using Fermat's Principle or its more useful form as Snell's laws. This allows us to locate images of objects formed when light travels through complicated lens systems or, in the case of mirages, through a medium of spatially varying refractive index.

1.2 Lenses and Principal Planes

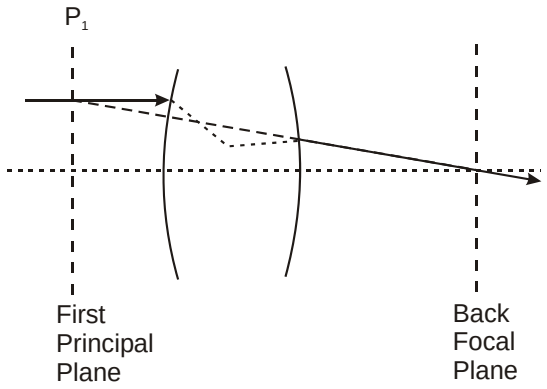


Figure 1.1

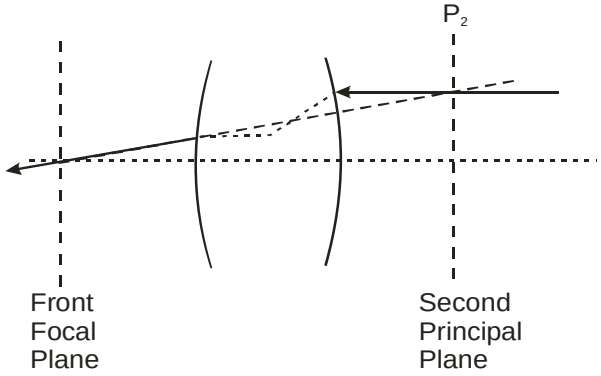


Figure 1.2

Thin lens formula:

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$$

u, v and f are measured to centre of lens of ``zero" thickness. For an object at infinity $u = \infty$ parallel rays are focussed in the image plane where $v = f$. This defines the focal plane of a thin lens.
For a thick or compound lens (composed of several individual lenses) the principal planes locate the position of an equivalent thin lens. (See Figures 1.1 and 1.2) The effective focal length is the distance from the principal plane to the associated focal plane.

1.3 Compound lens systems

1.3.1 Telephoto lens

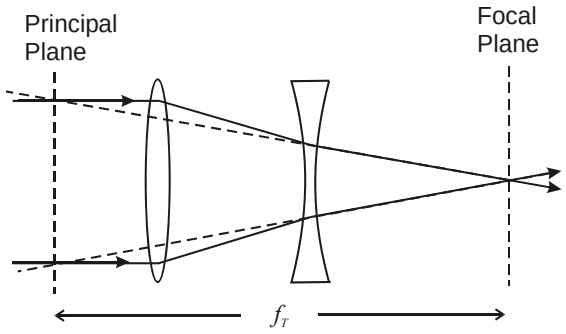


Figure 1.3

1.3.2 Wide angle lens

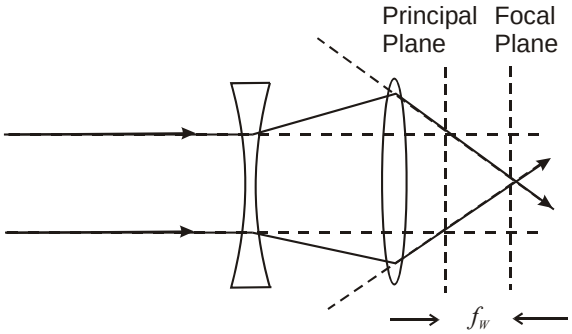


Figure 1.4

1.3.3 Telescope (Astronomical)

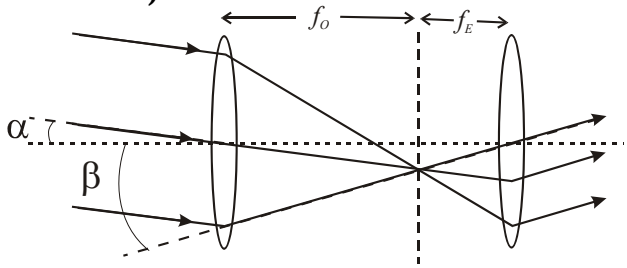


Figure 1.5

Angular magnification:

$$M = \beta / \alpha = f_o / f_E$$

1.3.4 Telescope (Galilean)

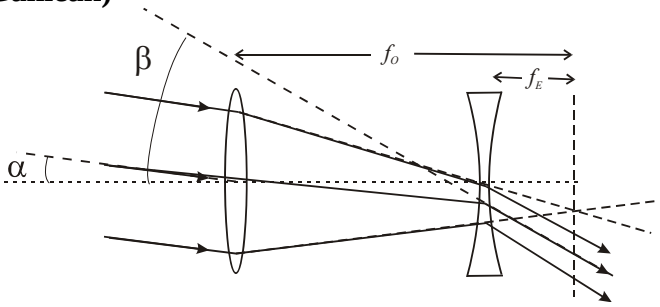


Figure 1.6

Angular magnification: $M = \beta / \alpha = f_o / f_E$

1.3.5 Telescope (Newtonian)

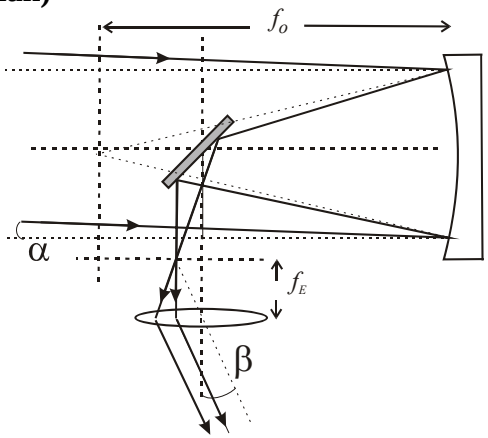


Figure 1.7

Angular magnification: $M = \beta / \alpha = f_o / f_E$

f_o is the focal length of the objective mirror and (for a spherical mirror surface) equals half the radius of curvature.

1.3.6 Compound Microscope

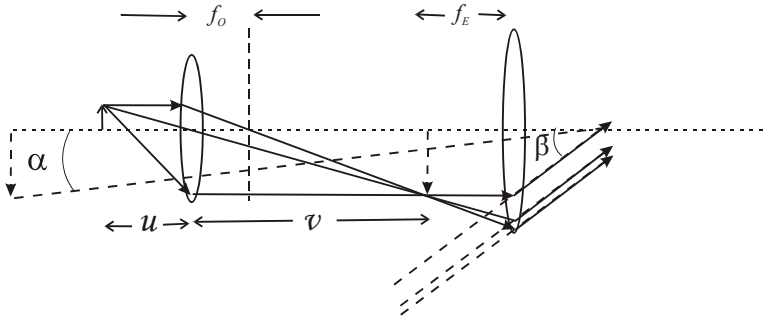


Figure 1.8

The object at distance u from objective with focal length f_o is imaged at distance v. Real image is at focal length f_E from eyepiece giving angular magnification β/α where α is the angle subtended by the real image if it was at the near point of the eye.

1.4 Illumination of optical systems

The brightness of an image is determined by the f-number (*f/no.*):

$$f/no. = \frac{focal\ length}{Diameter\ of\ Aperture}$$

Aperture stops determine the amount of light reaching the image of an image forming optical system.

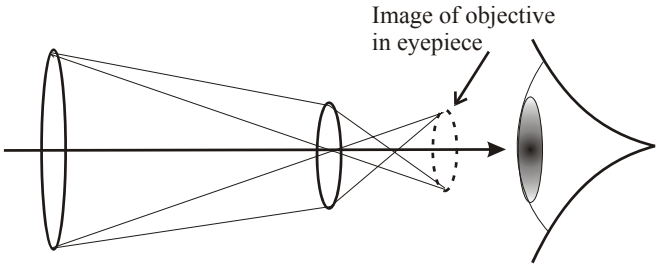


Figure 1.8 Size of objective is the effective aperture stop in a telescope.

For lenses of given focal length, the size (diameter) of the lenses are chosen to match the aperture of the image recording system e.g. pupil of eye. Aperture stops for a telescope (or binocular) system will be chosen as follows:

Aperture stop (eye piece) ≈ pupil of eye (~ 10 mm)
Aperture stop (objective) such that the size of the image of the objective formed by the eyepiece ≈ pupil of eye

Aperture stops for a camera lens are set by an iris diaphragm over a range of *f/no* from 1.2 t0 22. A small *f/no.* is a large aperture and allows in more light. This then allows a shorter shutter time to expose the film (or CCD) - useful for moving objects. The disadvantage is a reduced depth of field (range of distances in focus in the image).

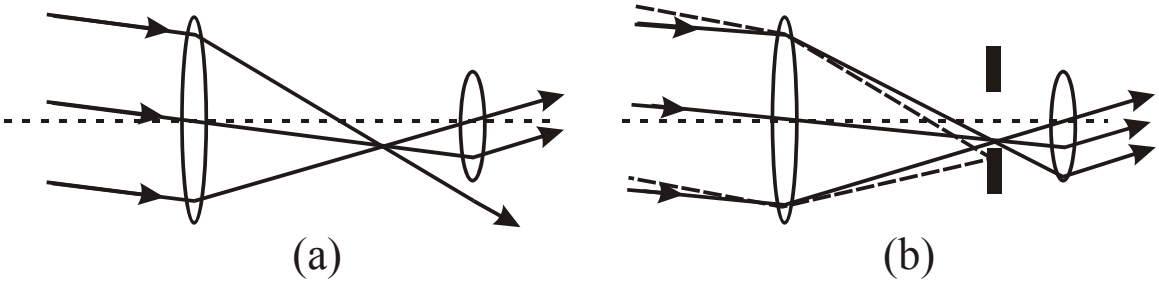


Figure 1.9 Field stop provides even illumination to image formed by eyepiece.

Off axis light in the telescope in Figure 1.9(a) spills past the eyepiece - the light to the image at this angle is reduced in intensity, the image darkens towards the edge. In Figure 1.9(b) all light at a given angle through the objective reaches the image. There is equal brightness across the image at the expense of the field of view.
Field stops: these ensure uniform illumination across the image by eliminating light at large off-axis angles. Light from the edges of the field of view reaches the image with the same intensity as light from the centre.

Physical Optics

2. Waves and Diffraction

2.1 Mathematical description of a wave

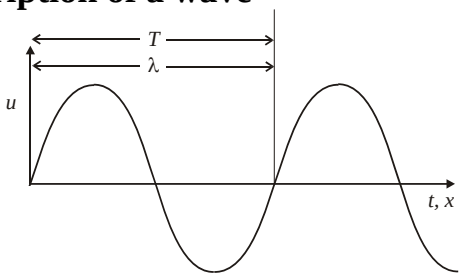


Figure 2.1

$$u = u_o \cos(\omega t - kz - \alpha) \quad \text{or} \quad u = u_o e^{-i\alpha} e^{-i(\omega t - kz)}$$

ωt : phase change with time, $\omega = \frac{2\pi}{T}$

kz : phase change with distance, $k = \frac{2\pi}{\lambda}$

α : arbitrary initial phase

2.2 Interference

Addition of amplitudes from two sources gives interference e.g. Young's slits:

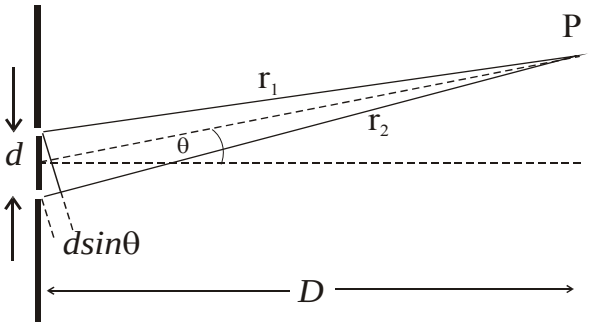


Figure 2.2 Young's slits

Two slits separated by d illuminated by monochromatic plane waves
Amplitude u_p at a point P a large distance, D, from the slits

$$u_p = \frac{u_o}{r_1} e^{-i(\omega t - kr_1)} + \frac{u_o}{r_2} e^{-i(\omega t - kr_2)}$$

Putting $(r_1 - r_2) = d\sin\theta$, $r_1 \approx r_2 = r$, intensity is:

$$I_p = 4\left(\frac{u_o}{r}\right)^2 \cos^2\left(\frac{1}{2} kd\sin\theta\right)$$

2.3 Phasors

Amplitude of wave is represented by length of a ``vector'' on an Argand diagram.
Phase of wave represented by angle of vector relative to Real axis of the Argand diagram.
The **phasor** is then: $ue^{i\delta}$

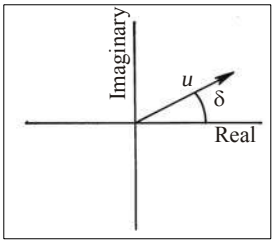


Figure 2.3 Phasor diagram

Example: Young's slits.

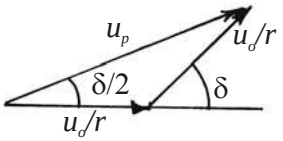


Figure 2.4 Phasor diagram for two slit problem.

Amplitude from each slit on screen: $\frac{u_o}{r}$
Phase difference δ , owing to path difference $d\sin\theta$: $\delta = kd\sin\theta$
Resultant amplitude is then

$$u_p = 2\frac{u_o}{r}\cos(\delta/2)$$

The intensity is therefore:

$$I_p = 4\left(\frac{u_o}{r}\right)^2\cos^2(\frac{1}{2}kd\sin\theta)$$

2.4 Diffraction from a finite slit

Monochromatic plane wave incident on aperture of width a
Observation plane at large distance D from aperture.
Amplitude in plane of aperture: u_o per unit length.

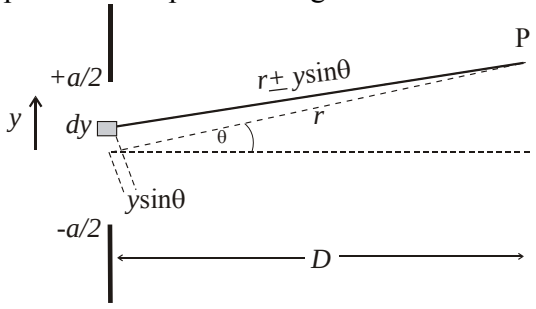


Figure 2.5 Contributions to amplitude at P from elements dy ***in slit.***

An infinitesimal element of length d y at position y contributes at P an amplitude:

$$\frac{u_o \mathrm{d}y}{r} e^{i\delta(y)}$$

The phase factor $\delta(y) = k(r \pm y\sin\theta)$.
The total amplitude at P arising from all contributions across the aperture:

$$u_p = \frac{u_o}{r} e^{ikr} \int_{-a/2}^{a/2} e^{-ik\sin\theta . y} \mathrm{d}y$$

The intensity is then: $I_p = I(0)\,\mathrm{sinc}^2\beta$

where $\beta = \frac{1}{2} ka \sin\theta$

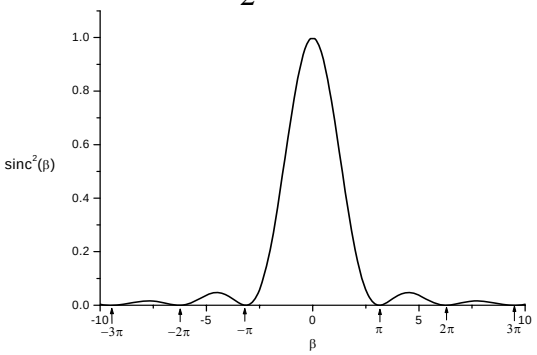


Figure 2.6 Intensity pattern from single slit, $I_p = I(0)$ ***sinc*** $^2\beta$ **.**

The first minimum is at $\beta = \pi$, $2\pi = \frac{2\pi}{\lambda} a\sin\theta$

Hence angular width θ of diffraction peak is:

$$\theta = \frac{\lambda}{a}$$

2.5 Diffraction from a finite slit: *phasor treatment*

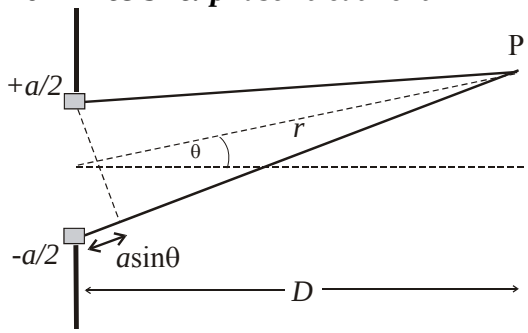


Figure 2.7 Construction showing elements at extreme edges of aperture contributing first and last phasors

On axis, $\theta = 0$ the phasor elements sum to R_o Off axis, $\theta \neq 0$ successive phase shifts between adjacent phasors bend the phasor sum to form a section of a regular polygon.

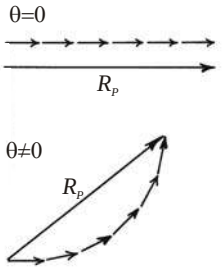


Figure 2.8 (a) Phasor diagram for finite slit showing resultant R_o for $\theta = 0$ and $\theta \neq 0$

The phase difference between first and last phasors for $\theta \neq 0$ is

$$\delta = ka\sin\theta$$

In the limit as the phasor elements $\to 0$ the phasors form an arc of a circle of radius R. The length of the arc is R_o and the length of the chord representing the resultant is R_p .

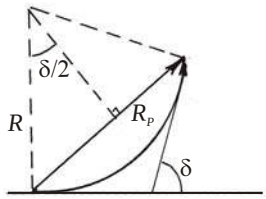


Figure 2.8 (b) Phasor diagram in the limit as phasor elements $\to 0$.

The amplitude at θ relative to the amplitude at $\theta = 0$:

$$\frac{length\ of\ chord}{length\ of\ arc} = \frac{2.R\sin(\delta/2)}{R.\delta} = \text{sinc}(\delta/2)$$

Then the intensity at θ:

$$I(\theta) = I(0)\ \text{sinc}^2(\delta/2) = I(0)\ \text{sinc}^2\beta$$

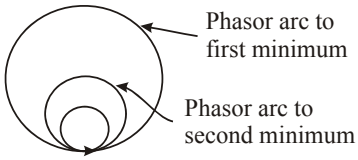


Figure 2.9 Phasor diagram showing mimina for increasing phase shift δ between extremes of slit as θ increases.

The first minimum occurs when the phasor arc bends to become a full circle i.e. the phase difference between first and last phasor elements $\delta = 2\pi$ angular radius is:

$$\theta = \frac{\lambda}{a}$$

2.6 Diffraction in 2 dimensions

Recall that the amplitude resulting from a plane wave illumination of an aperture of the form of a slit of width a in the y - direction :

$$u_p = \frac{u_o}{r} e^{ikr} \int_{-a/2}^{a/2} e^{-ik \sin\theta . y} \mathrm{d}y$$

Consider the aperture to have a width b in the x -direction, then the angular variation of the diffracted amplitude in the x -direction is :

$$u_p = \frac{u_o}{r} e^{ikr} \int_{-b/2}^{b/2} e^{-ik \sin\phi . x} \mathrm{d}x$$

$$\text{In 2-D we have : } u_p \propto e^{ikr} \int_{-b/2}^{b/2} \int_{-a/2}^{a/2} u(x,y) . e^{-ik(\sin\phi . x + \sin\theta . y)} \mathrm{d}x\mathrm{d}y$$

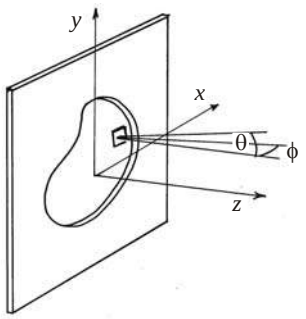


Figure 2.10 General 2-D aperture in** x,y **plane.

u(x,y) is the amplitude distribution function for the aperture. For a circular aperture of diameter a the diffraction pattern is a circular Bessel Function. The angular width to the first minimum is:

$$\theta = 1.22 \frac{\lambda}{a}$$

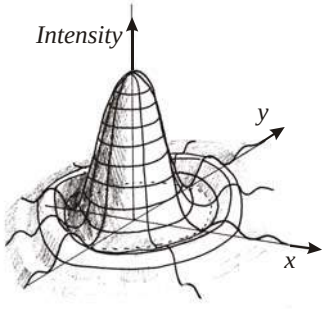


Figure 2.11. Point Spread Function for circular aperture.

A point source imaged by a lens of focal length f and diameter a gives a pattern with a minimum of radius f. This is the *Point Spread Function* or *instument function* analogous to the *impulse function* of an electrical circuit giving its response to a δ-function impulse.

3. Fraunhofer Diffraction

So far we have considered diffraction by
(a) Apertures or slits illuminated by plane waves
(b) Observation at a large distance where the phase difference between contributions from secondary sources in the diffracting plane separated by y is given to a good approximation by:

$$\delta = k \sin\theta . y$$

These are *special cases* where the phase difference δ is a *linear function* of the position y in the diffracting aperture.

3.1 Fraunhofer diffraction

Definition: ``A diffraction pattern for which the phase of the light at the observation point is a linear function of position for all points in the diffracting aperture is *Fraunhofer Diffraction*."
By *linear* we mean that the wave front deviates from a plane wave by less than $\lambda/20$ across the diffracting aperture.

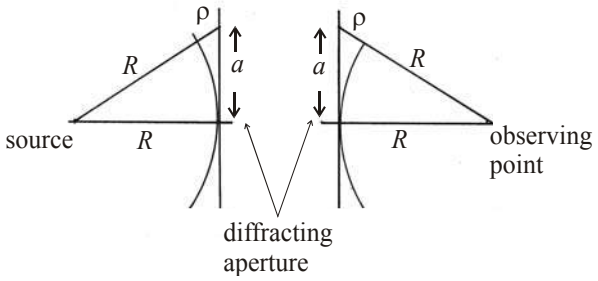


Figure 3.1 Wavefronts incident on and exiting from a plane aperture.

$$(R + \rho)^2 = R^2 + a^2$$

for $\rho \leq \lambda/20$ $\qquad R \approx 10a^2/\lambda$

Alternatively,
"*Fraunhofer diffraction is the diffraction observed in the plane of an image formed by an optical system*."

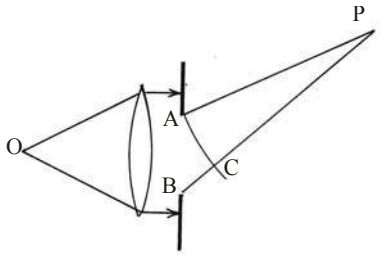


Figure 3.2 Fraunhofer condition for plane waves: image is at infinity as source is at focal length from lens.

Consider a point source at the focal point of a lens so that collimated light (plane waves) are incident on an aperture behind the lens. The image of the source is at ∞.
Fraunhofer Diffraction however will be observed at P if $BC \leq \lambda/20$

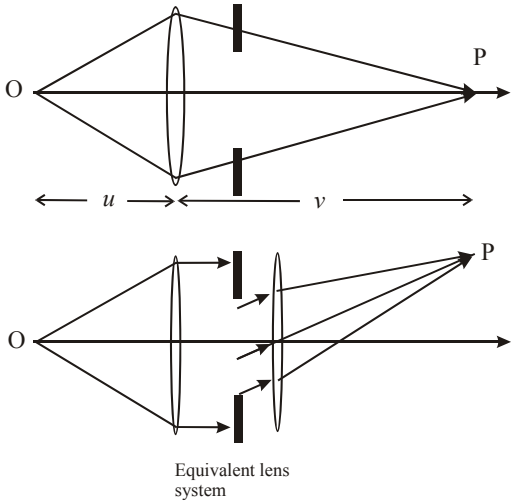


Figure 3.3 Fraunhofer diffraction observed in the image plane of a lens.

If the observation point P lies in the image plane of the lens so that curved wavefronts converge from the lens to P then no plane waves are involved. The lens and diffracting aperture however can be replaced by an equivalent system where diffraction of plane waves occurs. Note however that this means plane waves are *not* necessary to observe Fraunhofer diffraction. The key criterion is that ...

the phase varies linearly with position in the diffracting aperture.

A further consequence of noting that Fraunhofer diffraction is observed in the image plane is that the position of the aperture is not important.

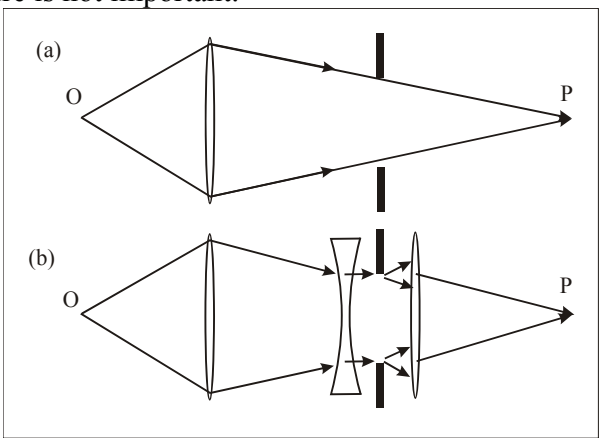


Figure 3.4 Equivalent lens system showing that Frauhofer diffraction is independent of position of aperture

3.2 Diffraction and wave propagation

Consider a plane wave surface at - z . This reproduces itself at a second plane $z = 0$. Huygens secondary sources in the wave front radiate to a point P in the second plane.

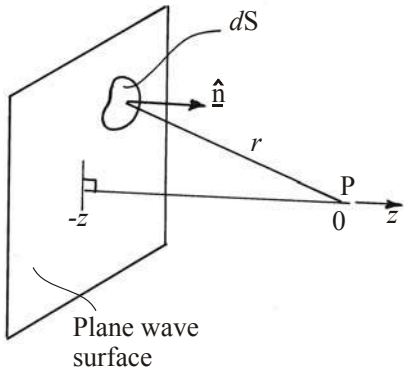


Figure 3.5 Huygens secondary sources on plane wave at $-z$ contribute to wave at P.

The amplitude at P is the resultant of all contributions from the plane at $-z$.

$$u_p = \alpha \int \frac{u_o \mathrm{d}S}{r} \eta(\underline{\mathrm{n}}, \underline{\mathrm{r}})\, e^{ikr}$$

u_o is the amplitude from element of area d S .

$\eta(\underline{\mathrm{n}},\underline{\mathrm{r}})$ is the obliquity factor - this accounts for the fact that the wave propagates only in the forward direction. $\underline{\mathrm{n}}$ is a unit vector normal to the wave front

α is a proportionality constant - to be determined.

We determine α by a self-consistency argument i.e. the plane wave at -z must reproduce itself at z = 0

We consider the amplitude at a point P a distance q from the wave such that $q = m\lambda$ where m is an integer and $m>>1$. i.e. P is a large distance from O a point on the wave front lying on a normal through P.

We construct elements of the wavefront of equal area δA centred on O.

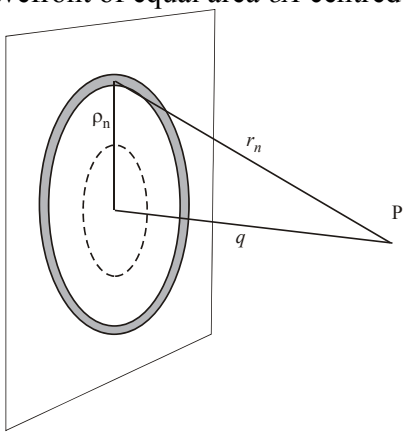


Figure 3.6 Construction of elements of equal area on plane wavefront.

The first element is a circle, the n th is an annulus of outer radius ρ_n

$$\pi(\rho_{n+1}^2 - \rho_n^2) = \delta A$$

Consequently the difference in distance δr from successive elements to P is constant

$$\delta r = r_{n+1} - r_n \simeq \frac{\delta A}{2\pi q}$$

Therefore the phase difference between waves from successive elements is also constant:

$$\delta\phi = \frac{\delta A}{\lambda q}$$

Hence we may treat contributions from each element of the wavefront as a Phasor.
[Note: we ignore, for the moment, the small difference in amplitude at P between successive elements arising from the small increase in distance r_n as ρ_n increases. We also ignore the small change in $\eta(\underline{n},\underline{r})$
Add contributions of elements (phasors) until the last phasor added is π out of phase with the first. The area of the wavefront covered by these elements is the First Half Period Zone, 1st HPZ.

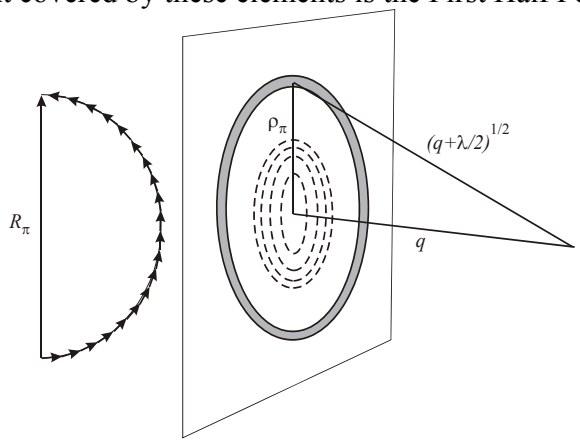


Figure 3.7 Construction of the First Half Period Zone

The difference in path-length from the outer element of the 1st HPZ to P and from O to P is $\lambda/2$.

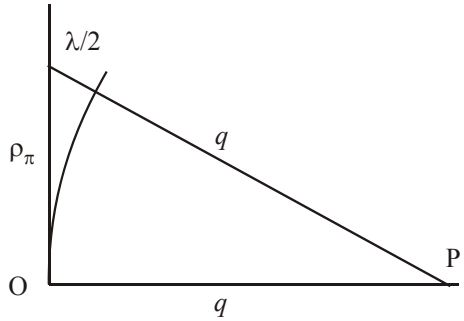


Figure 3.8 Phase shift of $\lambda/2$ arises at edge of 1st HPZ

The radius of the 1 st HPZ is ρ_π is given by:

$$\rho_\pi^2 = \lambda q$$

Recalling our diffraction integral above we may write the contribution to the amplitude at P from the 1st HPZ:

$$\alpha \frac{u_o \pi \rho_\pi^2}{q} = \alpha u_o \pi \lambda$$

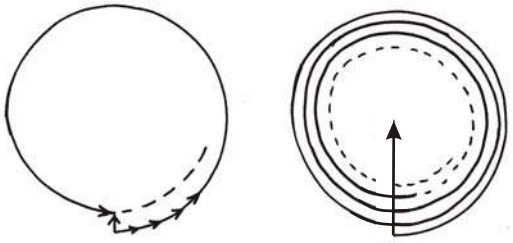
From the phasor diagram, the amplitude from the 1st HPZ is the length of the phasor arc, $\alpha u_o \pi \lambda$.
The resultant R_π is then the diameter of the circle of which the phasor arc defines half the circumference:

$$\frac{\pi R_\pi}{2} = \alpha u_o \pi \lambda$$

The resultant phasor lies along the imaginary axis so:

$$R_\pi = 2i\alpha u_o \lambda$$

Add further elements until the final phasor is in phase with the first i.e. a phase difference of 2 π.
The area of the wavefront now defines the first Full Period Zone 1st FPZ.
The resultant from the 1st FPZ is not exactly zero owing to the term $1/r$ (inverse square law for intensity) and the obliquity factor $\eta(\underline{n},\underline{r})$.
Adding further elements gives a slow spiral.



First Full-Period Zone Resultant of *n* FPZs

Figure 3.9 As n → ∞ resultant of zones tends to half the resultant of the 1st HPZ

Adding contributions from the whole wave (integrating over infinite surface) gives resultant = $\frac{1}{2}$ 1st HPZ. Therefore

$$R_\infty = i\alpha u_o \lambda$$

Self-consistency demands that this wave at P matches the original wave at O:

$$u_o = i\alpha u_o \lambda$$

$$\therefore \quad \alpha = -\frac{i}{\lambda}$$

Hence

$$u_p = -\frac{i}{\lambda} \int \frac{u_o \mathrm{d}S}{r} \eta(\underline{\mathrm{n}}, \underline{\mathrm{r}}) \, e^{ikr}$$

This is the Fresnel-Kirchoff diffraction integral.

4. Fourier methods in Optics

4.1 The Fresnel-Kirchoff integral as a Fourier Transform

The Fresnel-Kirchoff diffraction integral tells us how to calculate the field U_p in an observation plane using the amplitude distribution u_o in some initial plane

$$u_p = -\frac{i}{\lambda}\int \frac{u_o \mathrm{d}S}{r}\eta(\underline{n},\underline{r})e^{ikr}$$

We simplify as follows:

$$\eta(\underline{n},\underline{r}) = 1$$

Restrict to one dimension: $\mathrm{d}S \rightarrow \mathrm{d}x$

Ignore $\frac{1}{r}$ term by considering only a small range of r.

Use the Fraunhofer condition:

$$e^{ikr} = e^{ikr'}\, e^{ik\sin\theta\, x}$$

Absorb $e^{ikr'}$ into the constant of proportionality:

The amplitude u_p as a function of angle θ is then:

$$u_p \Rightarrow A(\beta) = \alpha \int_{-\infty}^{\infty} u(x)e^{i\beta x}\mathrm{d}x$$

where $\beta = k\sin\theta$.

We note that $A(\beta)$ is the Fourier transform of $u(x)$**.**

Important result:

The Fraunhofer diffraction pattern is the Fourier transform of the amplitude function of the diffracting aperture.

More precisely: the Fraunhofer diffraction pattern expressed as the amplitude as a function of angle is the Fourier transform of the function representing the amplitude of the incident wave as a function of position in the diffracting aperture. The Fraunhofer diffraction is expressed as a function of $\beta = k\sin\theta$ where θ is the angle of the diffracted wave relative to the wave vector $\underline{\mathbf{k}}$ of the wave incident on the aperture.

The inverse transform relation is:

$$u(x) = \frac{1}{\alpha}\int_{-\infty}^{\infty} A(\beta)e^{-i\beta x}\mathrm{d}\beta$$

4.2 The Convolution Theorem

The convolution of two functions $f(x)$ and $g(x)$ is a new function, $h(x)$, defined by:

$$h(x) = f(x) \circledast g(x) = \int_{-\infty}^{\infty} f(x')g(x - x')dx'$$

The Convolution Theorem states that the Fourier transform of a convolution of two functions is the product of the Fourier transforms of each of the two functions.

The Fourier transform, F.T., of $f(x)$ is $F(\beta)$

The Fourier transform, F.T., of $g(x)$ is $G(\beta)$

The Fourier transform, F.T., of $h(x)$ is $H(\beta)$

$$H(\beta) = F(\beta).G(\beta)$$

4.3 Some useful Fourier transforms and convolutions

(a) We can represent a wave of constant frequency β_o as a function of time t

$$v(t) = V_o e^{-i\beta_o t}$$

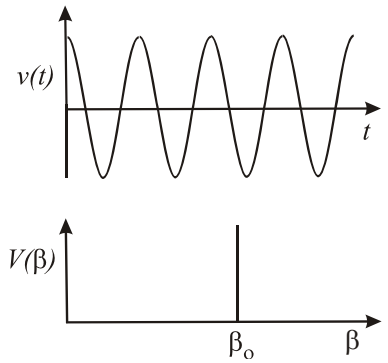


Figure 4.1 A wave of constant frequency (monochromatic) and its Fourier transform

$$F.T.\{v(t)\} = V(\beta) = V_o\ \delta(\beta - \beta_o)$$

i.e. $V(\beta)$ represents the spectrum of a monochromatic wave of frequency β_o and is a delta function in frequency space.

Alternatively the inverse transform relations allow us to represent the F.T. of a delta function:

$$v(t) = V_o\ \delta(t - t_o) \text{ as inverse } F.T.\{v(t)\} = V(\beta) = V_o\ e^{-i\beta t_o}$$

(b) The double slit function, i.e. two delta-functions separated by d :

$$v_b(x) = \delta(x \pm \frac{d}{2})$$

$$V_b(\beta) = 2\cos(\frac{1}{2}\beta d)$$

(c) A comb of delta functions:

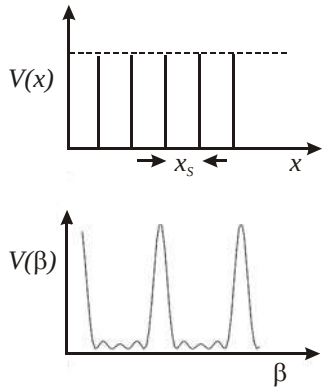


Figure 4.2 A comb of δ -functions and their transform

$$v_c(x) = \sum_{m=0}^{N-1} \delta(x - mx_s)$$

The F.T. of $v_c(x)$ is:

$$V_c(\beta) = e^{i\alpha} \frac{\sin(\frac{1}{2}N\beta x_s)}{\sin(\frac{1}{2}\beta x_s)}$$

where

$$\alpha = \frac{1}{2}(N-1)\beta x_s$$

The factor α is simply the consequence of starting our comb at $x = 0$. This factor can be eliminated by shifting our comb to sit symmetrically about the origin. This result illustrates the "Shift Theorem".

(d) The top-hat function:

$$v_d(x) = 1 \text{ for } |x| \leqslant \frac{a}{2} \qquad v_d(x) = 0 \text{ for } |x| > \frac{a}{2}$$

$$V_d(\beta) = a\,\mathrm{sinc}(\frac{1}{2}\beta a)$$

[What would be the result if the top-hat was shifted to sit between $x = 0$ and $x = a$?]

Now some useful convolutions:

(e) The double slit:

$$v_s(x) = v_b(x) \circledast v_d(x)$$

(f) The grating function:

$$v_g(x) = v_c(x) \circledast v_d(x)$$

(g) The triangle function:

$$v_\Delta(x) = v_d(x) \circledast v_d(x)$$

This is a self-convolution. The self-convolution is known also as the *autocorrelation function.*

4.4 Fourier Analysis

A periodic function *V(t)* may be represented by a Fourier series.

$$V(t) = c_o + \sum_{p=1}^{\infty} c_p \cos(p\omega_o t) + \sum_{p=1}^{\infty} s_p \sin(p\omega_o t)$$

V(t) is the result of *synthesis* of the set of Fourier components.
Fourier *analysis* is the reverse process - finding the components (amplitude and phase) that make up *V(t)*. The coefficients are found by integrating the function over a period τ of the oscillation.

$$s_p = \frac{2}{\tau}\int_0^\tau V(t)\sin(p\omega_o t)\,\mathrm{d}t$$
$$c_p = \frac{2}{\tau}\int_0^\tau V(t)\cos(p\omega_o t)\,\mathrm{d}t$$
$$c_o = \frac{2}{\tau}\int_0^\tau V(t)\,\mathrm{d}t$$

In general:

$$V(t) = \sum_{p=1}^{\infty} A_p e^{-ip\omega_o t}$$
$$A_p = \frac{1}{\tau}\int_0^\tau V(t)\, e^{ip\omega_o t}\mathrm{d}t$$

This last expression represents a Fourier transform - suggesting that this operation analyses the function *V(t)* to find the amplitudes of the Fourier components A_p.

4.5 Spatial frequencies

Consider a plane wave falling normally on an infinite screen with amplitude transmission function: $u(x) = 1 + \sin(\omega_s x)$ i.e. a grating with periodic pattern of width

$$d = \frac{2\pi}{\omega_s}$$

This defines the spatial frequency:

$$\omega_s = \frac{2\pi}{d}$$

The Fraunhofer diffraction pattern is then:

$$A(\beta) = \alpha \int_{-\infty}^{\infty} u(x) e^{i\beta x} \mathrm{d}x$$

where $\beta = k \sin\theta$. We find:

$$A(\beta) = 0 \text{ except for } \beta = 0, \pm\omega_s$$

$$\text{i.e. } \sin\theta \approx \theta = 0 \text{ or } \pm\frac{\lambda}{d}$$

The sinusoidal grating has a Fraunhofer diffraction pattern consisting of zero order and ± first orders $\theta = \pm\lambda / d = \pm\lambda\omega_s / 2\pi$.

An additional spatial frequency ω_n will lead to additional first orders a $\theta = \pm\lambda\omega_n / 2\pi$.

[Note: a finite screen will result in each order being spread by the diffraction pattern of the finite aperture, i.e. the "spread function" of the aperture.]

4.6 Abbé theory of imaging

We consider an object consisting of an infinite screen having a sinusoidal transmission described by a function *u(x)* so that the amplitude transmission repeats with a spacing *d*. This acts as an object at a distance *u* from a lens of focal length *f*.

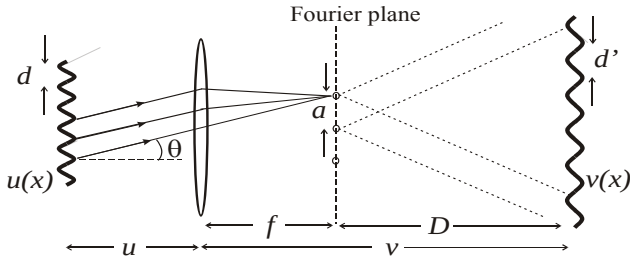


***Figure 4.3 Object** u(x) **imaged by lens to v(x)**.*

Diffraction orders are waves with parallel wave vectors at angles $\theta = 0$ and ±λ/*d*.
A lens brings these parallel waves to a focus as "points" in the focal plane separated by $a = f\lambda/d$.

Apart from a phase factor, the amplitude in the focal plane is the F.T. of $u(x)$. This plane is the *Fourier plane*.
Zero and first order "points" act as coherent sources giving two-beam interference at positions beyond the focal plane. In the image plane, distance *v* from the lens, the interference pattern is maximally sharp, $v = f + D$.
The interference pattern is a sinusoidal fringe system with spacing:

$$d' = \frac{D\lambda}{a}$$

From geometry

$$\frac{d}{u} = \frac{d'}{v}$$

Hence:

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$$

For a finite grating the "points" will be spread by diffraction at the effective aperture of the grating. [Note that we can describe such a grating as a convolution of an infinite sine wave with a top-hat function.]
Any object amplitude distribution may be synthesised by a set of sinusoidal functions. Each Fourier component with a specific spatial frequency contributes $\pm$ orders to the diffraction pattern at specific angles θ to the axis. The aperture α of the lens and object distance *u* determine the maximum angle $\theta_{\max}$ that may be collected. Diffraction orders at angles greater than $\theta_{\max}$ do not contribute to the final image. The corresponding spatial frequencies will be missing from the image. Higher spatial frequencies contribute to sharp edges in the object distribution. The lack of high spatial frequencies in the image leads to blurring and loss of resolution.
[Note: the discussion so far is valid only for *coherent* light i.e. light waves having a fixed phase relationship across the aperture in the object plane. In practice for microscopic objects this condition is partially fulfilled even for white light illumination.]

4.7 The Compound Microscope

Figure 4.4 shows the arrangement of the compound microscope. Basically a very short focal length lens, the objective, forms a real, inverted, image of the specimen in the image plane, giving a linear magnification of *v/u*. The eye-piece is basically a simple magnifier used to view the real image which is located at the focal length of the eyepiece giving a virtual image at infinity. This allows viewing with minimum eyestrain. The minimum dimension of spatial structure in the object d_{min} that can be resolved is such that the associated diffraction order will be at the maximum angle $\theta_{\max}$ that can be collected by the objective lens.

$$\sin\theta_{\max} = \lambda / d_{\min}$$

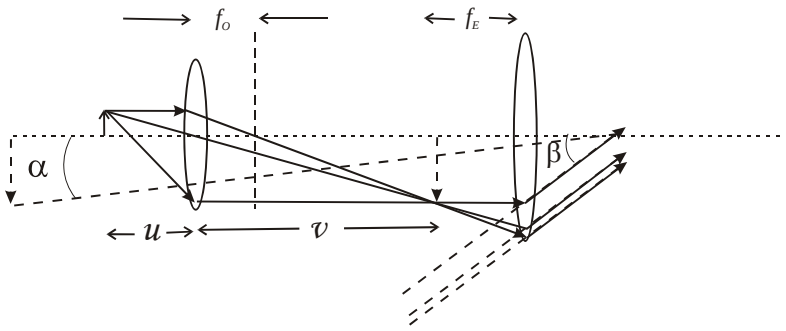


Figure 4.4 The Compound Microscope. The object at distance u from objective with focal length f_o is imaged at distance v. Real image is at focal length f_E from eyepiece giving angular magnification β/α where α is the angle subtended by the real image if it was at the near point of the eye.

Spatial frequencies having dimensions smaller than $d_{\min}$ will lead to diffraction to larger angles, miss the objective and thus not appear in the image. The minimum spatial dimension d_{min} that can be resolved may be increased by immersing the objective and object in oil of refractive index n_o; the oil immersion objective:

$$n_o \sin\theta_{\max} = \frac{\lambda}{d_{\min}}$$

$n_o \sin\theta_{\max}$ is the Numerical Aperture and defines the ultimate resolution of the device.

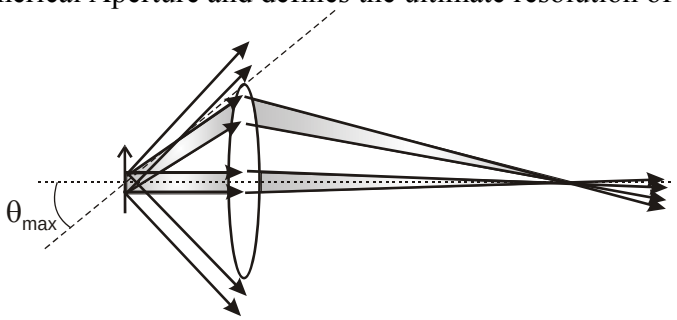


Figure 4.5. First order diffracted waves from spatial structures < d_{min} are collected by the lens and interfere in the image plane with zero order waves to form sinusoidal structure in the image. Light from smaller spatial structures (higher spatial frequencies) are diffracted to angles > $\theta_{\max}$, miss the objective and do not interfere with zero order in the image.

4.7 Diffraction effects on image brightness

Normal image brightness is determined by the *f/no*. of the optical system i.e. f/d_A where d_A is the limiting aperture. When the image size approaches the order of the $PSF \sim \lambda/d_A$ light is lost from the image by diffraction. This is diffraction limited imaging.

For non-diffraction limited imaging: $\textit{Image brightness} \propto d_A^2$

For diffraction limited imaging: $\textit{Image brightness} \propto d_A^4$

5 Optical instruments and fringe localization

Optical instruments for spectroscopy use interference to produce a wavelength-dependent pattern. The interfering beams are produced either by *division of wavefront* or by *division of amplitude*. The diffraction grating divides the wavefront into multiple beams. The Michelson divides the amplitude into two beams and the Fabry-Perot interferometer divides the amplitude into multiple beams. It is important to know where to look for the fringes. Before looking at specific instruments we consider the general question of fringe localization.

5.1 Division of wavefront

(a) Two-slit interference, Young's Slits

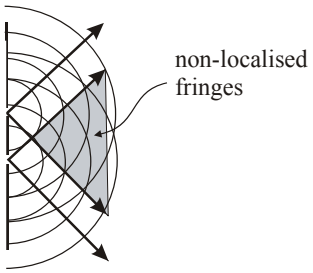


Figure 5.1 Young's slit fringes are observed throughout the region beyond the screen containing the two slits.

The fringes are *non-localized* and usually observed under the Fraunhofer condition.

(b) N-slit diffraction, the diffraction grating.

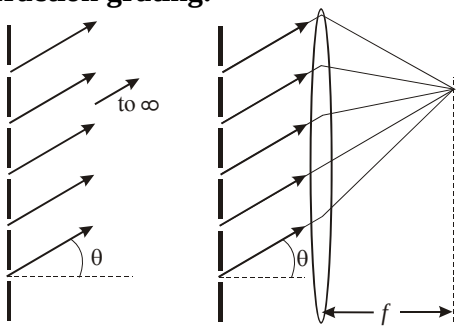


Figure 5.2 Diffraction grating fringes.

Again we usually observe the Fraunhofer condition. A monochromatic plane wave is diffracted i.e. suffers constructive interference at angle θ. Parallel light interferes at infinity or in the focal plane of a lens. The fringes are localized at infinity or in the image plane of the instrument.

5.2 Division of amplitude

The interference may involve two beams (Michelson) or multiple beams (Fabry-Perot). The situations are modelled by reflection of light from a source at two surfaces. The source may be a point or extended and the surfaces may be at an angle (wedged) or parallel. The images of the source in the reflecting surfaces act as two effective sources.

5.2.1 Point source
(a) Wedge.

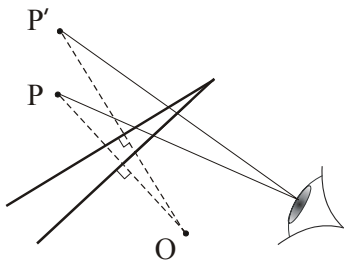


Figure 5.3 A point source O provides images P, P' in reflecting surfaces forming a wedge.
This system is equivalent to 2-point sources or Young's slit situation. Therefore the fringes are *non-localized* fringes of *equal thickness.* .

(b) Parallel

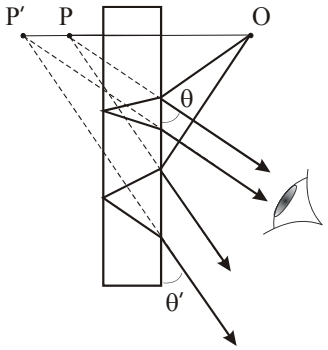


Figure 5.4 A point source reflected in two parallel surfaces again provides two images P, P'
This is similar to the wedge situation with 2-point sources. The fringes are *non-localized* fringes of *equal inclination.*

5.2.2 Extended source
(a) Wedge

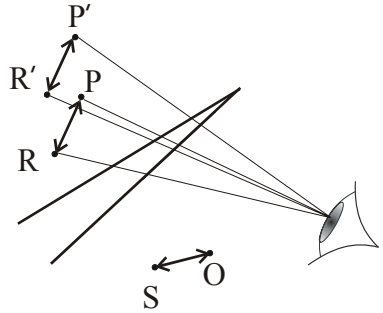


Figure 5.5. Extended source OS provides two images PR and P'R' by reflection at wedged reflecting surfaces.
Each point on the extended source produces non-localized fringes. Overlap of all these patterns gives no visible fringes. However at the apex of the wedge the path difference is zero and is the same for all points on the effective sources so fringes are visible in this region. The zero order fringe is a straight line fringe in the plane of the wedge. Other low-order fringes may be seen if the source is not too large and the wedge angle not too big. The fringes are of *equal thickness* and *localized* in the plane of the wedge e.g. Newton's Rings.

(b) Parallel

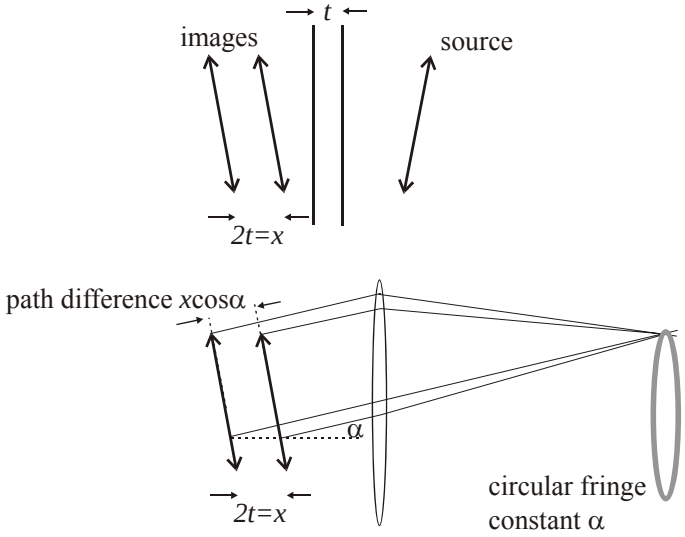


Figure 5.6 Upper figure shows two images of extended source by reflection in parallel slab of thickness *t.* ***Lower figure shows fringes of equal inclination formed in focal plane of a lens by light from the two images of the source.***

Close to plate overlapping patterns lead to no visible fringes. At large distance the fringes become wider and exceed the displacement of the overlap. Fringes become visible and are fringes of *equal inclination* and *localized at infinity.* These fringes are more conveniently observed in the focal plane of a lens. e.g. the eye.
Reflecting surfaces separated by *t* lead to two images separated by *2t* or *x = 2t*. Parallel light at an angle of inclination α to the axis from equivalent points on the effective sources are brought together in the focal plane. The path difference is $x\cos\alpha$ and the phase difference δ:

$$\delta = \frac{2\pi}{\lambda} x\cos\alpha$$

Bright fringes (constructive interference) occurs when the phase difference $\delta = p2\pi$ (*p = integer*) or

$$x\cos\alpha = p\lambda$$

For small angles the angular size of the fringes is given by

$$\alpha_p^2 - \alpha_{p+1}^2 = \frac{2\lambda}{x}$$

Hence radii of fringes in focal plane of lens with focal length f:

$$r_p^2 - r_{p+1}^2 = \frac{2f^2\lambda}{x}$$

As *x* increases, fringes get closer together
As *x* decreases → 0 fringes get larger and fill the field of view.
The behaviour of the fringes formed by parallel surfaces will be important for the Michelson and Fabry-Perot interferometers.

6 The diffraction grating spectrograph

6.1 Interference pattern from a diffraction grating

Consider a plane wave of wavelength λ incident normally on a reflecting or transmitting grating of N slits separated by d. The amplitude contributed by each slit is u and the intensity of the interference pattern is found by adding amplitudes and taking the squared modulus of the resultant.

(1) $N = 2$

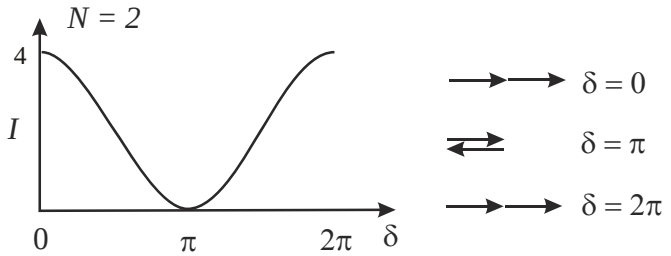


Figure 6.1 Intensity pattern and associated phasor diagram for 2-slit interference

$$I(\theta) = 4u^2 \cos^2(\frac{\delta}{2})$$

where

$$\delta = \frac{2\pi}{\lambda} d \sin\theta$$

Principal maxima at $\delta = 0, n2\pi$, of intensity $4u^2$. 1 minimum beween principal maxima.

(2) $N = 3$

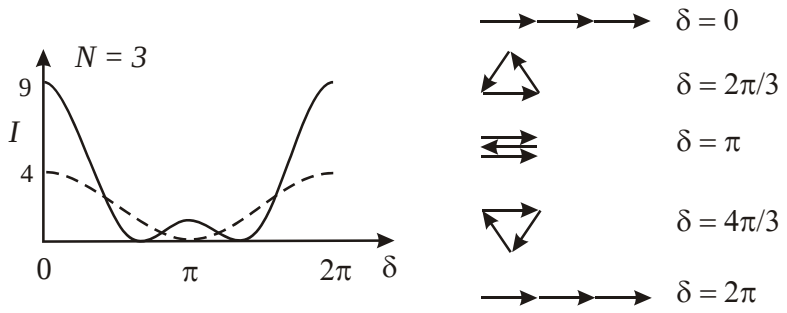


Figure 6.2 Phasor diagrams for 3-slit interference and intensity pattern

Using phasors to find resultant amplitude

(a) $\delta = 0, n2\pi$ Principal maxima of intensity $9u^2$. 2 minima between principal maxima.

(b) $\delta = 2\pi/3$ Minimum / zero intensity

(c) $\delta = \pi$ Subsidiary maxima of intensity u^2

(d) $\delta = 4\pi/3$ Minimum / zero intensity

(3) $N = 4$

Principal maxima at $\delta = 0, n2\pi$ of intensity $16u^2$. 3 minima between principa maxima.

In general:

Principal maxima at $\delta = 0, n2\pi$, intensity $\propto N^2$. $(N-1)$ minima at $\frac{n2\pi}{N}$

and width of principal maxima $\propto 1/N$.

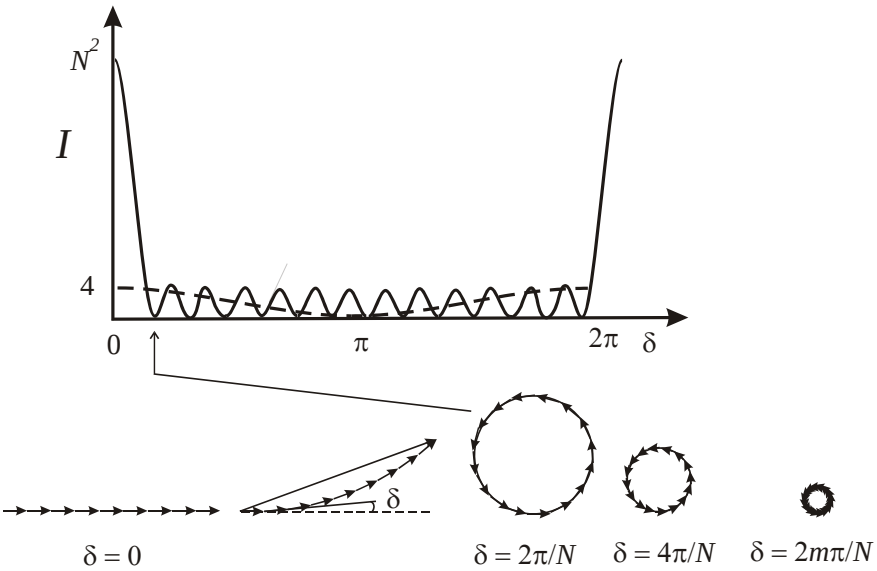


Figure 6.3 Phasor diagrams for N-slit interference and intensity pattern

Amplitude of N phasors:

$$A = u + ue^{-i\delta} + ue^{-i2\delta} + \ \ldots\ + ue^{-i(N-1)\delta}$$

Hence intensity:

$$I(\theta) = I(0)\frac{\sin^2(\frac{N\delta}{2})}{\sin^2(\frac{\delta}{2})}$$

6.2 Effect of finite slit width

Grating of N slits of width a separated by d is a convolution of a comb of N δ–functions with a single slit (top-hat function):

$$f(x) = \sum_{p=0}^{N-1} \delta(x - pd) \quad ; \quad g(x) = 1, \ \text{for} \ \{-\frac{a}{2} \leqslant x \leqslant \frac{a}{2}\}; \ g(x) = 0, \ \text{for} \ \{|x| > \frac{a}{2}\}$$

$$h(x) = f(x) \circledast g(x)$$

Using the Convolution Theorem with,

$$F(\beta) = F.T.\{f(x)\}, \quad G(\beta) = F.T.\{g(x)\} \ \text{and} \ \ H(\beta) = F.T.\{h(x)\}$$

$$H(\beta) = F(\beta).G(\beta)$$

Hence

$$|H(\beta)|^2 = I(\theta) = I(0)\frac{\sin^2(\frac{N\delta)}{2})}{\sin^2(\frac{\delta}{2})}.\frac{\sin^2(\frac{\gamma}{2})}{(\frac{\gamma}{2})^2}$$

where $\delta = kd\sin\theta$ and $\gamma = ka\sin\theta$

6.3 Diffraction grating performance

6.3.1 The diffraction grating equation

The equation for $I(\theta)$ gives the positions of principal maxima, $\delta = 0, n2\pi$, n is an integer: the order of diffraction (this is also the number of wavelengths in the path difference).
For simplicity we consider normal incidence on the grating. Then principal maxima occur for

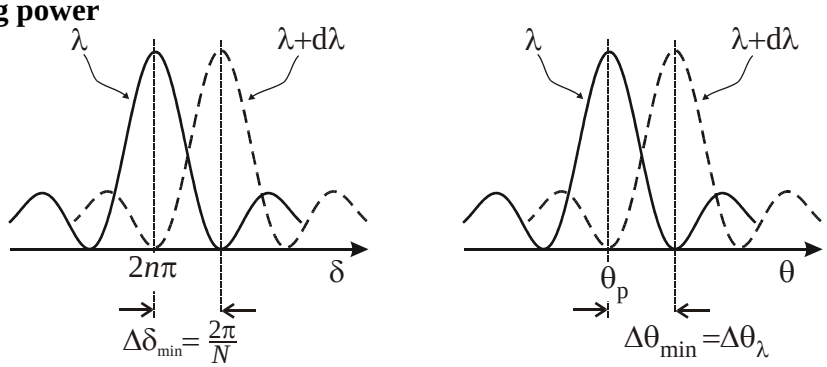
$$d\sin\theta = n\lambda$$

6.3.2 Angular dispersion

The angular separation $d\theta$ between spectral components differing in wavelength by $d\theta$:

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = \frac{n}{d\cos\theta}$$

6.3.3 Resolving power



(a) (b)

Figure 6.4 (a) Principal maxima for wavelength λ and for $\lambda + d\lambda$ such that the phase shift δ for the two differs by the change in δ between the maxima and first minima. (b) The same situation plotted as a function of diffraction angle θ. The angular width to the first minimum $\Delta\theta_{min}$ equals the angular separation $\Delta\theta_\lambda$ between the two wavelengths.

Principal maxima for wavelength λ occur for a phase difference of $\delta = n2\pi$. The change in phase difference δ between the maximum and the first minimum is $\Delta\delta_{min}$.

$$\Delta\delta_{\min} = \pm\frac{2\pi}{N}$$

and

$$\delta = \frac{2\pi}{\lambda} d\sin\theta$$

Angular width to first minimum $\Delta\theta_{min}$ is found from

$$\frac{\mathrm{d}\delta}{\mathrm{d}\theta} = \frac{2\pi}{\lambda} d\cos\theta$$

So

$$\Delta\delta_{\min} = \frac{2\pi}{\lambda} d\cos\theta.\Delta\theta_{\min}$$

The angular separation $\Delta\theta_\lambda$ of principal maxima for λ and $\lambda + \Delta\lambda$ is found from:

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = \frac{n}{d\cos\theta}$$

$$\Delta\theta_\lambda = \frac{n}{d\cos\theta}\Delta\lambda$$

The resolution criterion is:

$$\Delta\theta_\lambda = \Delta\theta_{\min}$$

Hence the Resolving Power is:

$$\frac{\lambda}{\Delta\lambda} = nN$$

6.3.4 Free Spectral Range

n^{th} order of λ and $(n + 1)^{th}$ order of $(\lambda + \Delta\lambda_{\text{FSR}})$ lie at same angle θ.
$\{\ n\lambda = d\sin\theta = (n+1)(\lambda - \Delta\lambda)\ \}$ Hence overlap occurs for these wavelengths at this angle.
The Free Spectral Range is thus:

$$\Delta\lambda_{FSR} = \frac{\lambda}{(n+1)}$$

Note: the Resolving Power $\propto n$ and the FSR $\propto 1/n$.

6.4 Blazed (reflection) gratings

The Blaze angle ξ is set to reflect light into the same direction as the diffracted order of choice for a given wavelength. For incident angle ϕ and diffracted angle θ the blaze angle will be：

$$\xi = \frac{1}{2}(\phi + \theta)$$

where ϕ and θ satisfy the grating equation

$$d(\sin\theta \pm \sin\phi) = n\lambda$$

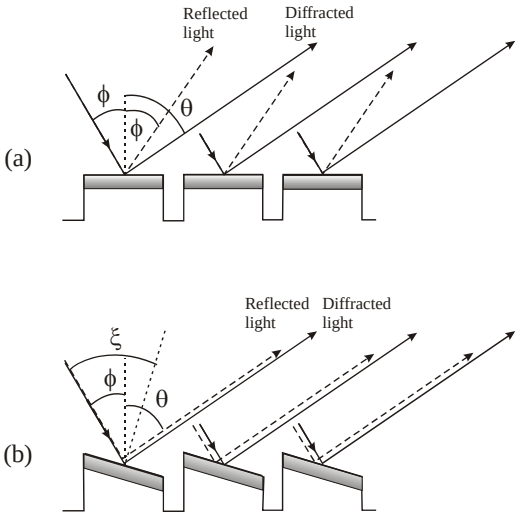


Figure 6.5 (a) Diffraction angle $\theta \neq$ Reflection angle ϕ for ordinary grating. (b) Blazed grating reflects light at same angle as diffracted order

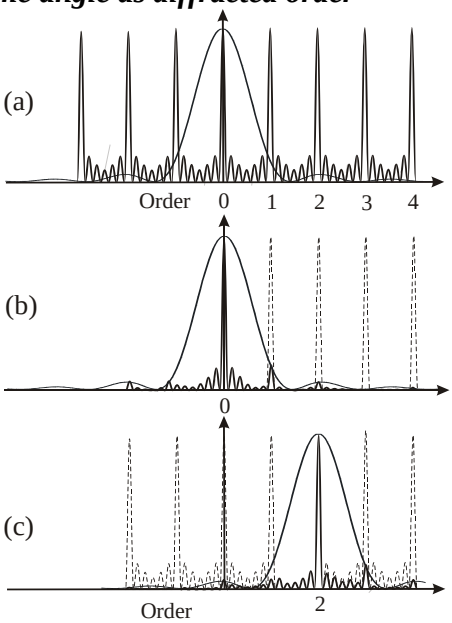


Figure 6.6 (a) Grating intensity pattern and single slit diffraction pattern.
(b) Effect of single slit diffraction envelope on grating diffraction intensity for unblazed grating. (c) Grating intensity pattern for blaze set to reflect light into 2nd order.

6.5 Effect of slit width on resolution and illumination

Consider the imaging forming system consisting of two lenses of focal length f_1 and f_2 . The image of a slit of width Δx_s has a width:

$$\Delta x_i = \left(\frac{f_2}{f_1}\right)\Delta x_s$$

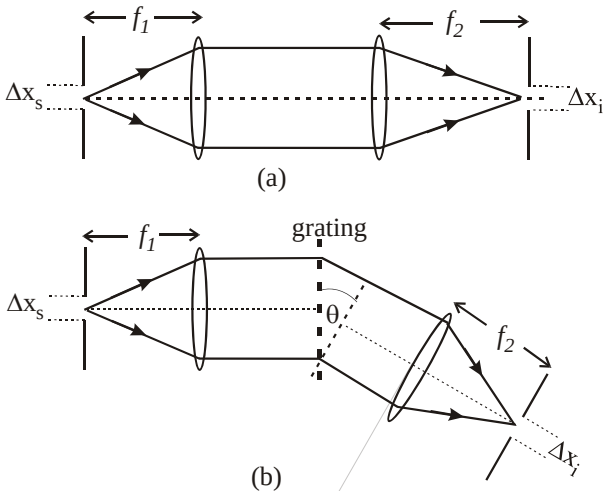


Figure 6.7 (a) Image forming system to image slit of width Δx_s to image Δx_i . (b) Images of slit are spectrally dispersed by diffraction at grating. Slit is imaged at angle θ from diffraction grating leading to foreshortening by $\cos\theta$.

In a diffraction grating spectrograph the image is viewed at the diffraction angle θ and so is foreshortened by $\cos\theta$.

$$\Delta x_i = \left(\frac{f_2}{f_1 \cos\theta}\right)\Delta x_s$$

The minimum resolvable wavelength difference, $\Delta\lambda_R$, has an angular width $\Delta\theta_R$:

$$\Delta\theta_R = \left(\frac{n}{d\cos\theta}\right)\Delta\lambda_R$$

Wavelengths having difference $\Delta\lambda_R$ are separated in the image plane of lens f_2 by Δx_R :

$$\Delta x_R = \frac{f_2\lambda}{Nd\cos\theta}$$

where we used $\Delta\lambda_R = \dfrac{\lambda}{nN}$

Resolution is achieved provided: $\Delta x_i \leqq \Delta x_R$

The limiting slit width Δx_s is then:

$$\Delta x_s \leqslant \frac{f_1 \lambda}{Nd} \quad \text{or} \quad \Delta x_s \leqslant \frac{f_1 \lambda}{W}$$

Note: the optimum slit width is such that the diffraction pattern of the slit just fills the grating aperture, $W = Nd$.

$\Delta x_s > \Delta x_R$: resolution reduced by overlap of images at different wavelengths

$\Delta x_s < \Delta x_R$: resolution not improved beyond diffraction limit but image brightness is reduced.

7 The Michelson (Fourier Transform) Interferometer

A two-beam interference device in which the interfering beams are produced by division of amplitude at a 50:50 beam splitter.

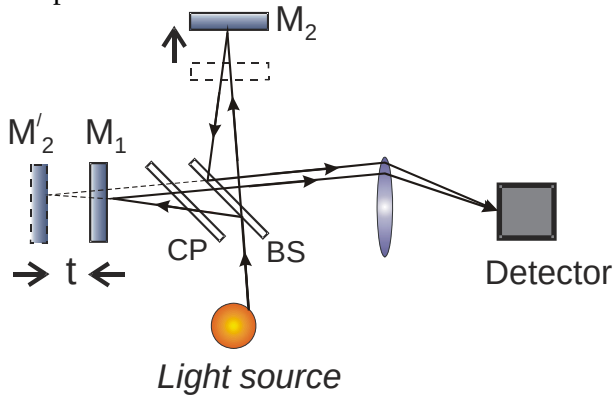


Figure 7.1 The Michleson interferometer. The beam splitter BS sends light to mirrors M_1 and M_2 in two arms differing in length by t. M_2' is image of M_2 in M_1 resulting effectively in a pair of parallel reflecting surfaces illuminated by an extended source as in figure 5.6. CP is a compensating plate to ensure beams traverse equal thickness of glass in both arms.

7.1 Michelson Interferometer

Distance from beam splitter to mirrors differs by t in the two paths, and α is the angle of interfering beams to the axis. Resulting phase difference between beams:

$$\delta = \frac{2\pi}{\lambda} 2t\cos\alpha = \frac{2\pi}{\lambda} x\cos\alpha \tag{7.1}$$

Constructive interference at $\delta = 2p\pi$, where p is an integer.

$$x\cos\alpha = p\lambda$$

On axis the order of interference is $p = x/\lambda$.
Symmetry gives circular fringes about axis. The fringes are of equal inclination and localized at infinity. They are viewed therefore in the focal plane of a lens. Fringe of order p has radius r_p in the focal plane of a lens (focal length, f, see section 5.2.2(b).

$$r_p^2 - r_{p+1}^2 = \frac{2f^2\lambda}{x} \tag{7.2}$$

Two-beam interference pattern:

$$I(x) = I(0)\cos^2\left(\frac{\delta}{2}\right)$$
$$I(x) = \frac{1}{2} I(0)\left[1 + \cos 2\pi\bar{\nu}x\right] \tag{7.3}$$

where $\bar{\nu} = \frac{1}{\lambda}$, the wavenumber.

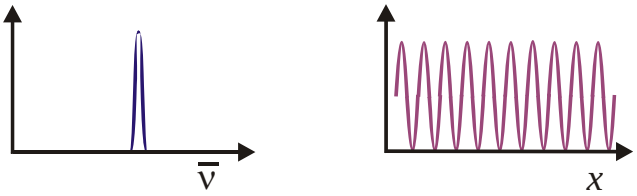


Figure 7.2 Input spectrum of monochromatic source and resulting interferogram obtained from scanning Michelson interferometer.

7.2 Resolving Power of the Michelson Spectrometer.

Consider that we wish to resolve two wavelengths λ_1 and λ_2 that differ by $\Delta\lambda$. The corresponding wavenumbers are $\bar{\nu}_1$ and $\bar{\nu}_2$ and they provide two independent interferograms so the resultant is the sum of the two:

$$I(x) = \frac{1}{2}I_0(\bar{\nu}_1)[1 + \cos 2\pi\bar{\nu}_1 x] + \frac{1}{2}I_0(\bar{\nu}_2)[1 + \cos 2\pi\bar{\nu}_2 x]$$

Let the two components have equal intensity: so $I_0(\bar{\nu}_1) = I_0(\bar{\nu}_2) = I_0(\bar{\nu})$ is the intensity of each interferogram at $x = 0$. Then

$$I(x) = I_0(\bar{\nu})\left[1 + \cos 2\pi\left(\frac{\bar{\nu}_1 + \bar{\nu}_2}{2}\right)x \cos 2\pi\left(\frac{\bar{\nu}_1 - \bar{\nu}_2}{2}\right)x\right] \tag{7.4}$$

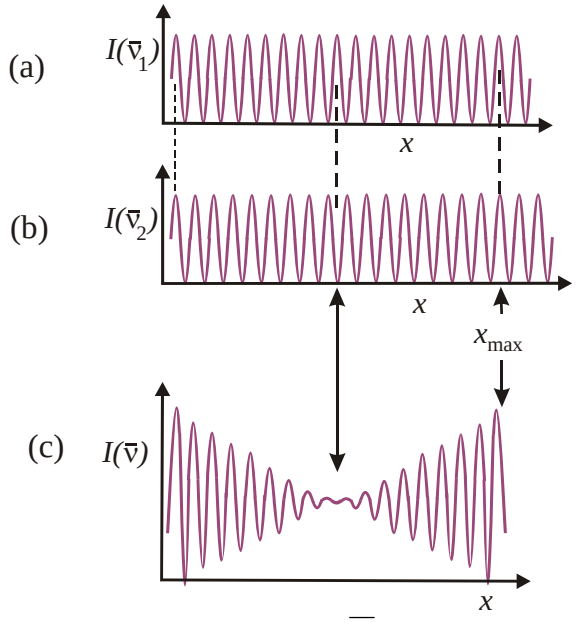


Figure 7.3 (a) Interferogram of source component $\bar{\nu}_1$ (b) interferogram of source component $\bar{\nu}_2$. (c) Interferogram of combined light showing added intensities (a) and (b). Note visibility of fringes cycles to zero and back to unity for equal intesity components. To resolve the complete cycle requires a path difference x_{max}

This looks like an interferogram of a light source with mean wavenumber $\left(\frac{\overline{\nu}_1+\overline{\nu}_2}{2}\right)$ multiplied by an envelope function $\cos 2\pi\left(\frac{\overline{\nu}_1-\overline{\nu}_2}{2}\right)x$. This envelope function goes first to a zero when a "peak" of interferogram for $\overline{\nu}_1$ first coincides with a zero in the interferogram for $\overline{\nu}_2$. The visibility (or contrast) of the fringes cycles to zero and back to unity; the tell-tale sign of the presence of the two wavelength components. The number of fringes in the range covering the cycle is determined by the wavenumber difference $\Delta\overline{\nu} = \overline{\nu}_1 - \overline{\nu}_2$. The instrument will have the power to resolve these two wavenumbers (wavelengths) if the maximum path difference available, x_{max}, is just sufficient to record this cycle in the envelope of the interferogram. The minimum wavenumber difference $\Delta\overline{\nu}_{min}$ that can be resolved is found from the value of x_{max} giving the cycle in the cosine envelope function:

$$2\pi\left(\frac{\Delta\overline{\nu}_{min}}{2}\right)x_{max} = \pi$$

$$\Delta\overline{\nu}_{min} = \frac{1}{x_{max}} \tag{7.5}$$

This minimum resolvable wavenumber difference is the *instrument function* as it represents the width of the spectrum produced by the instrument for a monochromatic wave.

$$\Delta\overline{\nu}_{Inst} = \frac{1}{x_{max}} \tag{7.6}$$

Hence the Resolving Power RP is:

$$\text{RP.} = \frac{\lambda}{\Delta\lambda_{Inst}} = \frac{\overline{\nu}}{\Delta\overline{\nu}_{Inst}} = \frac{x_{max}}{\lambda} \tag{7.7}$$

7.3 The Fourier Transform spectrometer

Comparing Figure 7.2 with figure 4.1 we see that the interferogram looks like the Fourier transform of the intensity spectrum. The interferogram produced using light of two wavenumbers $\overline{\nu}_1$ and $\overline{\nu}_2$ is

$$I(x) = \frac{1}{2}I_0(\overline{\nu}_1)[1 + \cos 2\pi\overline{\nu}_1 x] + \frac{1}{2}I_0(\overline{\nu}_2)[1 + \cos 2\pi\overline{\nu}x]$$

In the case of multiple discrete wavelengths:

$$\begin{aligned} I(x) &= \sum_i \frac{1}{2}I_0(\overline{\nu}_i)[1 + \cos 2\pi\overline{\nu}_i x] \\ &= \sum_i \frac{1}{2}I_0(\overline{\nu}_i) + \sum_i \frac{1}{2}I_0(\overline{\nu}_i)\cos 2\pi\overline{\nu}_i x \end{aligned}$$

First term on r.h.s. is ½ I_o where I_o is the total intensity at $x = 0$
Second term on r.h.s. is sum of individual interferograms.
Replacing components with discrete wavenumbers by a continuous spectral distribution:

$$I(x) = \frac{1}{2}I_0 + \int_0^\infty S(\overline{\nu})\cos 2\pi \overline{\nu}_i x.d\overline{\nu}$$

where $S(\overline{\nu})$ is the power spectrum of the source.
Now $S(\overline{\nu}) = 0$ for $\overline{\nu} < 0$, so second term may be written:

$$F(x) = \int_{-\infty}^\infty S(\overline{\nu})\cos 2\pi \overline{\nu}_i x.d\overline{\nu} \tag{7.8}$$

$F(x)$ is the cosine Fourier Transform of $S(\overline{\nu})$

$$F.T.\{S(\overline{\nu})\} = I(x) - \frac{1}{2}I_0$$

Or

$$S(\overline{\nu}) \propto F.T.\left\{I(x) - \frac{1}{2}I_0\right\} \tag{7.9}$$

Apart from a constant of proportionality the Fourier transform of the interferogram yields the Intensity or Power Spectrum of the source. See Figure 7.2.

The Michelson interferometer effectively compares a wavetrain with a delayed replica of itself. The maximum path difference that the device can introduce, x_{max}, is therefore the limit on the length of the wavetrain that can be sampled. The longer the length measured the lower the uncertainty in the value of the wavenumber obtained from the Fourier transform. Distance x and wavenumber $\overline{\nu}$ are Fourier pairs or conjugate variables.[see equation (7.9)] This explains why the limit on the uncertainty of wavenumber (or wavelength) measurement $\Delta\overline{\nu}_{Inst}$ is just the inverse of x_{max}. In essence this explains the general rule for all interferometers including diffraction grating instruments that:

$$\Delta\overline{\nu}_{Inst} = \frac{1}{\text{Maximum path difference between interfering beams}} \tag{7.10}$$

7.4 The Wiener-Khinchine Theorem

Note: this topic is NOT on the syllabus but is included here as an interesting theoretical digression.
The recorded intensity $I(x)$ is the product of two fields, $E(t)$ and its delayed replica $E(t + \tau)$ integrated over many cycles. (The delay $\tau = x/c$.) The interferogram as a function of the delay may be written:

$$\gamma(\tau) = \int E(t)E(t+\tau)\mathrm{d}t \tag{7.11}$$

Taking the integral from $-\infty$ to $+\infty$ we define the Autocorrelation Function of the field to be

$$\Gamma(\tau) = \int_{-\infty}^{\infty} E(t)E(t+\tau)\mathrm{d}t \tag{7.12}$$

The Autocorrelation Theorem states that if a function *E(t)* has a Fourier Transform *F(ω)* then

$$F.T.\{\Gamma(\tau)\} = |F(\omega)|^2 = F^*(\omega).F(\omega) \tag{7.13}$$

Note the similarity between the Autocorrelation theorem and the Convolution Theorem.
The physical analogue of the Autocorrelation theorem is the *Wiener-Kinchine Theorem.*

"The Fourier Transform of the autocorrelation of a signal is the spectral power density of the signal"

The Michelson interferogram is just the autocorrelation of the light wave (signal).
Note that ω and $\bar{\nu}$ are related by a factor $2\pi c$ where *c* is the speed of light.

7.5 Fringe visibility.

7.5.1 Fringe visibility and relative intensities

Figure 7.3 shows an interferogram made up of two independent sources of different wavelengths. The contrast in individual fringes of the pattern varies and we define the "visibility" of the fringes by

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \tag{7.14}$$

The fringe visibility "comes and goes" periodically as the two patterns get into and out of step. The example shown consisted of two sources of equal intensity. The visibility varies between 1 and 0. If however the two components had different intensity $I_1(\overline{\nu}_1)$ and $I_2(\overline{\nu}_2)$ then the envelope function of the interferogram does not go to zero. The contrast of the fringes varies from $I_1(\overline{\nu}_1) + I_2(\overline{\nu}_2)$ at zero path difference (or time delay) to a minimum value of $I_1(\overline{\nu}_1) - I_2(\overline{\nu}_2)$. Denoting the intensities simply by I_1 and I_2.

$$V_{\max} = \frac{(I_1 + I_2) - 0}{(I_1 + I_2) + 0} \quad ; \quad V_{\min} = \frac{I_1 - I_2}{I_1 + I_2}$$

Hence $\dfrac{V_{\min}}{V_{\max}} = \dfrac{I_1 - I_2}{I_1 + I_2}$ which leads to: $\dfrac{I_1}{I_2} = \dfrac{1 + V_{\min}/V_{\max}}{1 - V_{\min}/V_{\max}}$ (7.15)

Measuring the ratio of the minimum to maximum fringe visibility $V_{\min} / V_{\max}$ allows the ratio of the two intensities to be determined.

7.5.1 Fringe visibility, coherence and correlation

When the source contains a continuous distribution of wavelengths/wavenumbers the visibility decreases to zero with increasing path difference *x* and never recovers. The two parts of each of the Fourier components (individual frequencies) in each arm of the interferometer are in phase at zero path difference (zero time delay). At large path differences there will be a continuous distribution of interferograms with a range of phase differences that "average" to zero and no steady state fringes are visible. The path difference x_o introduced that brings the visibility to zero is a measure of the wavenumber difference $\Delta\overline{\nu}_L$ across the width of the spectrum of the source.

$$\Delta\overline{\nu}_L \simeq \frac{1}{x_o} \tag{7.16}$$

$\Delta\overline{\nu}_L$ is the spectral linewidth of the source.

A source having a finite spectral linewidth i.e. every light source (!) may be thought of as emitting wavetrains of a finite average length. When these wavetrains are split in the Michelson, and recombined after a delay, interference will occur only if some parts of the wavetrains overlap. Once the path difference *x* exceeds the average length of wavetrains no further interference is possible. The two parts of the divided wavetrain are no longer "coherent". The Michelson interferogram thus gives us a measure of the degree of coherence in the source. A perfectly monochromatic source (if it existed!) would give an infinitely long wavetrain and the visibility would be unity for all values of *x*. The two parts of the divided wavetrain in this case remain perfectly correlated after any delay is introduced. If the wavetrain has random jumps in phase separated in time on average by say τ_c then when the two parts are recombined after a delay $\tau_d < \tau_c$ only part of the wavetrains will still be correlated. The wavetrains from the source stay correlated with a delayed replica only for the time τ_c which is known as the *coherence time*. Thus we see that the Michelson interferogram provides us with the *autocorrelation function* or self-correlation along the length of the electromagnetic wave emitted by the source. (see section 4.3) In other words the Michelson provides a measure of the "*longitudinal coherence"* of the source.

[Note. Light sources may also be characterised by their "*transverse coherence"*. This is a measure of the degree of phase correlation the waves exhibit in a plane transverse to the direction of propagation. Monochromatic light emanating from a "point" source will give spherical wavefronts i.e. every point on a sphere centred on the source will have the same phase. Similarly a plane wave is defined as a wave originating effectively from a point source at infinity. Such a source will provide Young's Slit interference no matter how large the separation of the slits. (The fringes, of course, will get very tiny for large separations.) If the slits are illuminated by two separate point sources, with the same monochromatic wavelength but with a small displacement, then two sets of independent fringes are produced. The displacement of the sources gives a displacement of the two patterns. For small slit separation this may be insignificant and fringes will be visible. When, however, the slit separation is increased the "peaks" of one pattern overlap the "troughs" of the other pattern and uniform illumination results. The separation of the slits in this case therefore indicates the extent of the spatial correlation in the phase of the two monochromatic sources i.e. this measures the extent of the *transverse coherence* in the light from the extended source.]

8. The Fabry-Perot interferometer

This instrument uses multiple beam interference by division of amplitude. Figure 8.1 shows a beam from a point on an extended source incident on two reflecting surfaces separated by a distance *d*. Note that this distance is the optical distance i.e. the product of refractive index *n* and physical length. For convenience we will omit *n* from the equations that follow but it needs to be included when the space between the reflectors is not a vacuum. An instrument with a fixed *d* is called an *etalon*. Multiple beams are generated by partial reflection at each surface resulting in a set of parallel beams having a relative phase shift δ introduced by the extra path $2d\cos\theta$ between successive reflections which depends on the angle θ of the beams relative to the axis. (See section 5.2.2 (b)). Interference therefore occurs at infinity - the fringes are of equal inclination and localized at infinity. In practice a lens is used and the fringes observed in the focal plane where they appear as a pattern of concentric circular rings.

8.1 The Fabry-Perot interference pattern

This is done in all the text books (consult for details). The basic idea is as follows:

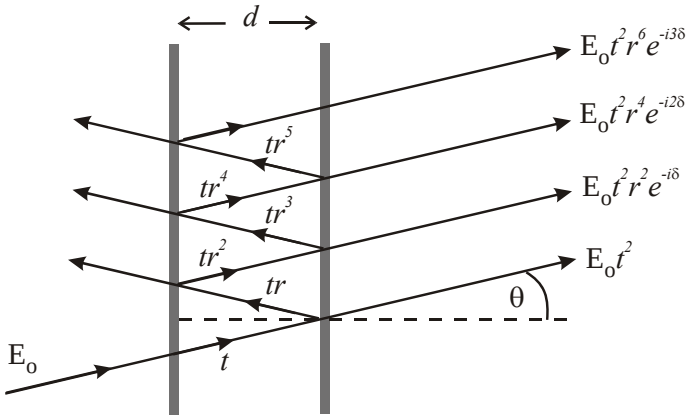


Figure 8.1 Multiple beam interference of beams reflected and transmitted by parallel surfaces with amplitude reflection and transmission coefficients r_i, t_i respectively.

Amplitude reflection and transmission coefficients for the surfaces are r_1, t_1 and r_2, t_2, respectively.
The phase difference between successive beams is:

$$\delta = \frac{2\pi}{\lambda} 2d\cos\theta \qquad (8.1)$$

An incident wave $E_o e^{-i\omega t}$ is transmitted as a sum of waves with amplitude and phase given by:

$$E_t = E_0 t_1 t_2 e^{-i\omega t} + E_0 t_1 t_2 r_1 r_2 e^{-i(\omega t-\delta)} + E_0 t_1 t_2 r_1^2 r_2^2 e^{-i(\omega t-2\delta)} + \ldots etc.$$

Taking the sum of this Geometric Progression in $r_1 r_2 e^{i\delta}$

$$E_t = E_0 t_1 t_2 e^{i\omega t}\left[\frac{1}{1 - r_1 r_2 e^{i\delta}}\right]$$

and multiplying by the complex conjugate to find the transmitted Intensity:

$$I_t = E_t E_t^* = E_0^2 t_1^2 t_2^2\left[\frac{1}{1 + r_1^2 r_2^2 - 2 r_1 r_2 \cos\delta}\right]$$

writing $E_0^2 = I_0$, $r_1 r_2 = R$ and $t_1 t_2 = T$, and $\cos\delta = (1 - 2\sin^2\delta/2)$:

$$I_t = I_0 \frac{T^2}{(1-R)^2}\left[\frac{1}{1+\frac{4R}{(1-R)^2}\sin^2\delta/2}\right] \tag{8.2}$$

If there is no absorption in the reflecting surfaces $T = (1 - R)$ then defining

$$\frac{4R}{(1-R)^2} = \Phi \tag{8.3}$$

$$I_t = I_0\left[\frac{1}{1+\Phi\sin^2\delta/2}\right] \tag{8.4}$$

This is known as the Airy Function. See figure 8.2

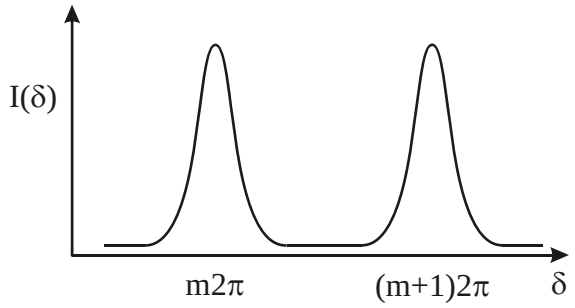


Figure 8.2 The Airy function showing fringes of order m, m+1 as function of δ .

8.2 Observing Fabry-Perot fringes

The Airy function describes the shape of the interference fringes. Figure 8.2 shows the intensity as a function of phase shift δ. The fringes occur each time δ is a multiple of 2π .

$$\delta = \frac{2\pi}{\lambda} 2d\cos\theta = m2\pi \tag{8.5}$$

m is an integer, the order of the fringe. The fringes of the Airy pattern may be observed by a system to vary *d,* λ, or θ. A system for viewing many whole fringes is shown in figure 8.3. An extended source of monochromatic light is used with a lens to form the fringes on a screen. Light from any point on the source passes through the F.P. at a range of angles illuminating a number of fringes. The fringe pattern is formed in the focal plane of the lens.

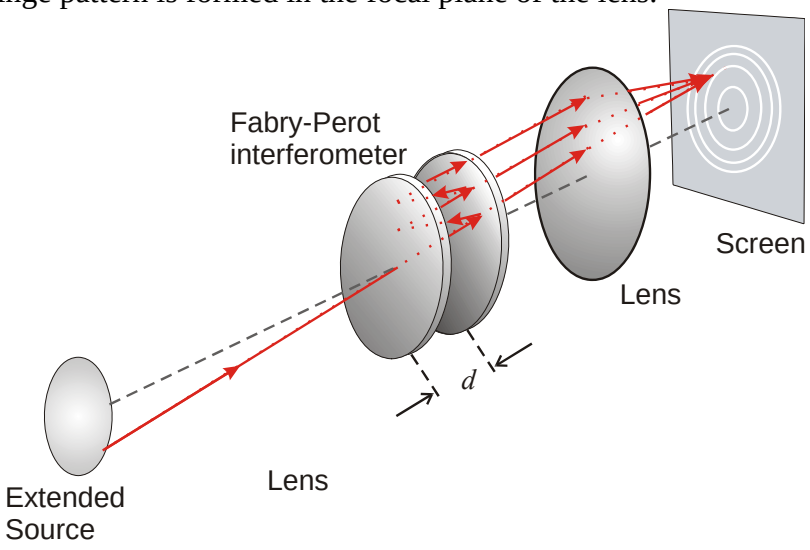


Figure 8.3. Schematic diagram of arrangement to view Fabry-Perot fringes. Parallel light from the Fabry-Perot is focussed on the screen.

From equation (8.5) the m^{th} fringe is at an angle θ_m

$$\cos\theta_m = \frac{m\lambda}{2d} \qquad (8.6)$$

The angular separation of the m^{th} and $(m+1)^{th}$ fringe is $\Delta\theta_m$ is small so $\theta_m \approx \theta_{m+1}$

$$\theta\Delta\theta_m = \frac{\lambda}{2d} = \text{constant}$$

Therefore the fringes get closer together towards the outside of the pattern. The radius of the θ_m fringe is

$$\rho(\lambda) = f\theta_m = f\cos^{-1}\left(\frac{m\lambda}{2d}\right) \qquad (8.7)$$

An alternative method to view fringes is “Centre spot scanning”. A point source or collimated beam may be used as the source and imaged on a “pinhole”. Light transmitted through the pinhole is monitored as a function of d or λ. Fringes are produced of order *m* linearly proportional to *d* or $\bar{\nu}$, (1/λ). This also has the advantage that all the available light is put into the detected fringe on axis.

8.3 Finesse

The separation of the fringes is 2π in δ-space, and the width of each fringe is defined by the half-intensity point of the Airy function i.e. $I_t/I_0 = 1/2$ when

$$\Phi \sin^2 \delta/2 = 1$$

The value of δ at this half-intensity point is $\delta_{1/2}$

$$\sin^2\left(\frac{\delta_{1/2}}{2}\right) = \frac{1}{\Phi}$$

$\delta_{1/2}$ differs from an integer multiple of 2π by a small angle so we have:

$$\delta_{1/2} = \frac{2}{\sqrt{\Phi}}$$

The full width at half maximum FWHM is then $\Delta\delta$

$$\Delta\delta = \frac{4}{\sqrt{\Phi}}$$

The sharpness of the fringes may be defined as the ratio of the separation of fringes to the halfwidth FWHM and is denoted by the Finesse F

$$F = \frac{2\pi}{\Delta\delta} = \frac{\pi\sqrt{\Phi}}{2} \tag{8.8}$$

or

$$F = \frac{\pi\sqrt{R}}{(1-R)} \tag{8.9}$$

So the sharpness of the fringes is determined by the reflectivity of the mirror surfaces.

[Note that $F \sim \frac{3}{(1-R)}$. This gives a quick check to ensure the quadratic equation for R has been solved correctly!]

8.4 The Instrumental width

The width of a fringe formed in monochromatic light is the instrumental width:

$$\Delta\delta_{Inst} = \frac{2\pi}{F} \tag{8.10}$$

$\Delta\bar{\nu}_{Inst}$ is the instrument width in terms of the apparent spread in wavenumber produced by the instrument for monochromatic light. For on-axis fringes ($\cos\theta = 1$):

$$\delta = 2\pi\bar{\nu}2d$$

$$\mathrm{d}\delta = 2\pi 2d\,\mathrm{d}\bar{\nu}$$

Hence $\mathrm{d}\delta = \Delta\delta_{Inst} = 2\pi 2d\,\Delta\bar{\nu}_{Inst} = \frac{2\pi}{F}$ and:

$$\Delta\bar{\nu}_{Inst} = \frac{1}{2dF} \tag{8.11}$$

8.5 Free Spectral Range, FSR

Figure 8.4 shows two successive orders for light having different wavenumbers, $\bar{\nu}$ and ($\bar{\nu} - \Delta\bar{\nu}$). Orders are separated by a change in δ of 2π The $(m+1)^{th}$ order of wavenumber $\bar{\nu}$ may overlap the m^{th} order of ($\bar{\nu} - \Delta\bar{\nu}$). i.e. changing the wavenumber by $\Delta\bar{\nu}$ moves a fringe to the position of the next order of the original wavenumber $\bar{\nu}$.

$$\bar{\nu}2d = (m+1) \qquad \text{and} \qquad (\bar{\nu} - \Delta\bar{\nu})2d = m$$
$$\therefore \quad \Delta\bar{\nu}2d = 1$$

This wavenumber span is called the Free Spectral Range, FSR:

$$\Delta\bar{\nu}_{FSR} = \frac{1}{2d} \tag{8.12}$$

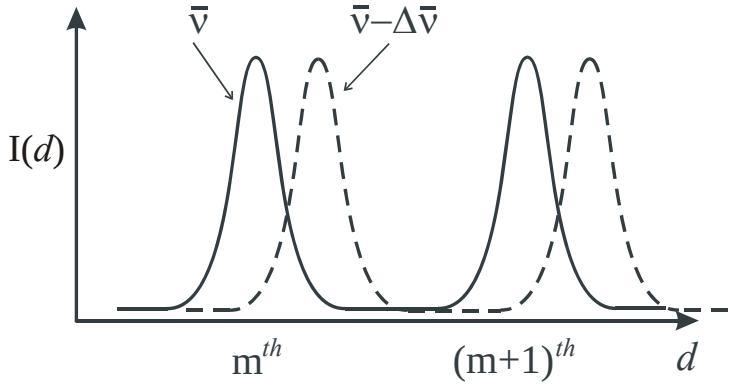


Figure 8.4 Fabry-Perot fringes for wavenumber $\bar{\nu}$ and $\bar{\nu}$ - $\Delta\bar{\nu}$ observed in centre-spot scanning mode. The mth -order fringe of $\bar{\nu}$ and $\bar{\nu}$ - $\Delta\bar{\nu}$ appear at a slightly different values of the interferometer spacing d. When the wavenumber difference $\Delta\bar{\nu}$ increases so that the mth order fringe of $\bar{\nu}$ - $\Delta\bar{\nu}$ overlaps the (m+1) th order of $\bar{\nu}$ the wavenumber difference equals the Free Spectral Range, FSR

In figure 8.4 the different orders for each wavelength (wavenumber) are made visible by changing the plate separation d. (Because changing d will change δ). The phase δ can be varied by changing d, λ or θ. See equation (8.5). In figure 8.3 the different orders for a given wavelength are made visible by the range of values of θ. If the source emits different wavelengths, fringes of the same order will appear with different radius on the screen.

8.6 Resolving Power

The instrumental width may now be expressed as:

$$\Delta\bar{\nu}_{Inst} = \frac{\Delta\bar{\nu}_{FSR}}{F} = \frac{1}{2dF} \tag{8.13}$$

Two monochromatic spectral lines differing in wavenumber by $\Delta\bar{\nu}_R$ are just resolved if their fringes are separated by the instrumental width: $\Delta\bar{\nu}_R = \Delta\bar{\nu}_{Inst}$

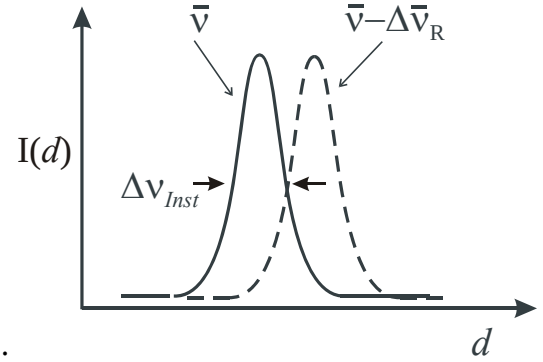


Figure 8.5 Resolution criterion: light of two wavenumbers $\bar{\nu}$, $\bar{\nu} - \Delta\bar{\nu}_R$ ***is resolved when the separation of fringes for*** $\bar{\nu}$ ***and*** $\bar{\nu} - \Delta\bar{\nu}_R$ ***is equal to the instrument width*** $\Delta\bar{\nu}_{Inst}$

As in Figure 8.4 the fringes of the same order for each spectral line separated in wavenumber by $\Delta\bar{\nu}_R$ could be recorded by varying d or θ.

The Resolving Power is then given by:

$$\frac{\bar{\nu}}{\Delta\bar{\nu}_R} = \frac{\bar{\nu}}{\Delta\bar{\nu}_{Inst}}$$

Now $\bar{\nu} = m/2d$:

$$\frac{\bar{\nu}}{\Delta\bar{\nu}_{Inst}} = \frac{m}{2d}\frac{2dF}{1}$$

Hence

$$R.P. = \frac{\bar{\nu}}{\Delta\bar{\nu}_R} = mF \tag{8.14}$$

Note, F defines the effective number of interfering beams and m is the order of interference. Alternatively, F determines the maximum effective path difference:

$$\text{Maximum path difference} \;=\; (2d\cos\theta) \times F \quad \text{and} \quad (2d\cos\theta = m\lambda)$$

So

$$\frac{\text{Maximum path difference}}{\lambda} = mF$$

i.e. the Resolving Power is the number of wavelengths in the maximum path difference.

8.7 Practical matters

8.7.1 Designing a Fabry-Perot

(a) FSR: The FSR is small so F.P.s are used mostly to determine small wavelength differences. Suppose a source emits spectral components of width $\Delta\bar{\nu}_c$ over a small range $\Delta\bar{\nu}_S$. We will require $\Delta\bar{\nu}_{FSR} > \Delta\bar{\nu}_S$. This determines the spacing d : $\frac{1}{2d} > \Delta\bar{\nu}_S$ or

$$d < \frac{1}{2\Delta\bar{\nu}_S}$$

(b) Finesse (Reflectivity of mirrors). This determines the sharpness of the fringes i.e. the instrument width.

We require

$$\Delta\bar{\nu}_{Inst} \lesssim \Delta\bar{\nu}_c \qquad \text{or} \qquad \frac{\Delta\bar{\nu}_{FSR}}{F} \lesssim \Delta\bar{\nu}_c$$

Hence

$$F \geq \frac{1}{2d\Delta\bar{\nu}_c}$$

The required reflectivity R is then found from

$$F = \frac{\pi\sqrt{R}}{(1-R)} \tag{8.15}$$

8.7.2 Centre spot scanning

The pin-hole admitting the centre spot must be chosen to optimize resolution and light throughput. Too large and we lose resolution; too small and we waste light and reduce signal-to-noise ratio. We need to calculate the radius of the first fringe away from the central fringe:

$$\cos\theta_m = \frac{m\lambda}{2d}$$

If m^{th} fringe is the central fringe, $\theta_m = 0$ and so $m = 2d/\lambda$ The next fringe has angular radius:

$$\theta_{m-1} = \cos^{-1}\left(1 - \frac{\lambda}{2d}\right)$$

The fringe radius in focal plane of lens of focal length f:

$$\rho_{m-1} = f\theta_{m-1}$$

8.7.3 Limitations on Finesse

The sharpness of the fringes is affected if the plates are not perfectly flat. A "bump" of $\lambda/10$ in height is visited effectively 10 times if the reflectivity finesse is 10 and thus the path difference is altered by λ. If the flatness is λ/x it is therefore not worthwhile to make the reflectivity finesse > $x/2$.

We assumed $T = (1-R)$ i.e. no absorption. In practice,

$$R + T + A = 1$$

where A is the absorption coefficient of the coatings. The coefficient in equation (8.2) modifies the transmitted intensity:

$$\frac{T^2}{(1-R)^2} \Rightarrow \left(\frac{1-R-A}{1-R}\right)^2 = \left(1 - \frac{A}{1-R}\right)^2$$

Increasing $R \to 100\%$ means $(1-R) \to A$ and the coefficient in the Airy function:

$$I_0 \frac{T^2}{(1-R^2)} \Rightarrow 0$$

i.e. the intensity transmitted to the fringes tends to zero.

9. Reflection at dielectric surfaces and boundaries

9.1 Electromagnetic waves at dielectric boundaries

Maxwell's equations lead to a wave equation for electric and magnetic fields E, H :

$$\nabla^2 E = \varepsilon_o \varepsilon_r \mu_o \mu_r \frac{\partial^2 E}{\partial t^2}$$

$$\nabla^2 H = \varepsilon_o \varepsilon_r \mu_o \mu_r \frac{\partial^2 H}{\partial t^2}$$

Solutions are of the form:

$$E = E_o \exp i(\omega t - \mathbf{k}.\mathbf{r})$$

From Maxwell's equations we also have:

$$\nabla \times E = -\mu_o \mu_r \frac{\partial^2 H}{\partial t^2}$$

From which we find:

$$E = \sqrt[2]{\frac{\mu_o \mu_r}{\varepsilon_o \varepsilon_r}} H \quad \text{or} \quad E = \frac{1}{n} \sqrt[2]{\frac{\mu_o}{\varepsilon_o}} H$$

where n is the refractive index of the medium and $\sqrt{\frac{\mu_o}{\varepsilon_o}}$ is the impedance of free space.

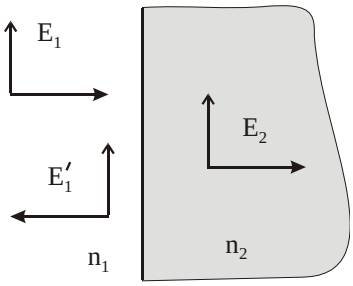


Figure 9.1 Reflection of an electromagnetic wave incident normally from medium of refractive index n_1 on a medium of index n_2

Boundary conditions at the interface of two media of different refractive index n_1 and n_2 demand that the perpendicular component of D is continuous and the tangential components of E and H are continuous. Incident and reflected field amplitudes are E_1 and $E_1^{'}$ respectively.

$$\frac{E_1}{E_1'} = r = \frac{n_2 - n_1}{n_2 + n_1}$$

The intensity reflection coefficient is therefore:

$$R = \left(\frac{n_2 - n_1}{n_2 + n_1} \right)^2 \tag{9.1}$$

For an air/glass interface $R \sim 4\%$.

9.2 Reflection properties of a single dielectric layer.

Consider a wave incident from air, refractive index n_o , on a dielectric layer of index n_1 deposited on a substrate of refractive index n_T . See figure 9.2. E_o, H_o are incident electric and magnetic wave amplitudes respectively in the air, and $E_o^{'}, H_o^{'}$ the reflected amplitudes; E_1 , H_1 and $E_1^{'}, H_1^{'}$ are incident and reflected amplitudes in the dielectric layer and E_T is the amplitude transmitted to substrate.

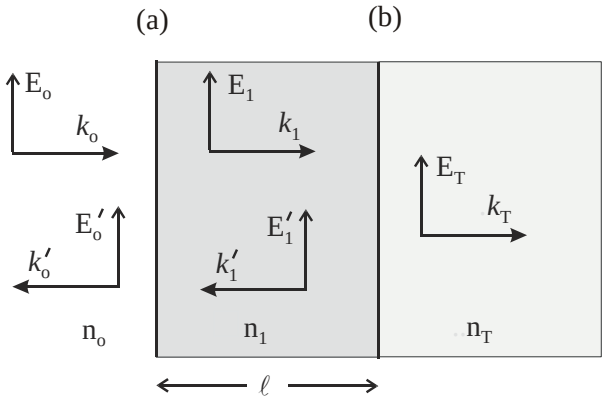


Figure 9.2 Reflected and transmitted waves for a wave incident normally from medium of index n_o on a dielectric layer on thickness ℓ and index n_1 deposited on a substrate of index n_T

At boundary (a)

$$E_o + E_o^{'} = E_1 + E_1^{'} \tag{9.2}$$

$$H_o - H_o^{'} = H_1 - H_1^{'} \tag{9.3}$$

using $H = n\sqrt{\frac{\varepsilon_o}{\mu_o}}E$:

$$n_o(E_o - E_o^{'}) = n_1(E_1 - E_1^{'}) \tag{9.4}$$

At boundary (b), E_1 has acquired a phase shift owing to propagating the thickness ℓ of the layer:

$$E_1 e^{ik_1\ell} + E_1^{'} e^{-ik_1\ell} = E_T \tag{9.5}$$

$$n_1(E_1 e^{ik_1\ell} - E_1^{'} e^{-ik_1\ell}) = n_T E_T \tag{9.6}$$

Eliminating E_1 and $E_1^{'}$ from (9.2), (9.4), (9.5) and (9.6):

$$E_o + E_o' = \left[\cos k_1 \ell - i\left(\frac{n_T}{n_1}\right) \sin k_1 \ell \right] E_T \tag{9.7}$$

$$n_o (E_o - E_o') = \left[-i n_1 \sin k_1 \ell + n_T \cos k_1 \ell \right] E_T \tag{9.8}$$

Writing:

$$A = \cos k_1 \ell \qquad B = -i\left(\frac{1}{n_1}\right) \sin k_1 \ell$$
$$C = -i n_1 \sin k_1 \ell \qquad D = \cos k_1 \ell \tag{9.9}$$

We find:

$$\frac{E_o'}{E_o} = r = \frac{A n_o + B n_o n_T - C - D n_T}{A n_o + B n_o n_T + C + D n_T} \tag{9.10}$$

and

$$\frac{E_T}{E_o} = t = \frac{2 n_o}{A n_o + B n_o n_T + C + D n_T} \tag{9.11}$$

Now consider the case when $\ell = \lambda / 4$, a quarter-wave layer; $k_1 \ell = \pi / 2$:

$$R = |r|^2 = \left| \frac{n_o n_T - n_1^2}{n_o n_T + n_1^2} \right|^2 \tag{9.12}$$

For $\ell = \lambda / 2$ a half-wave layer; $k_1 \ell = \pi$:

$$R = |r|^2 = \left| \frac{n_o - n_T}{n_o + n_T} \right|^2 \tag{9.13}$$

Note that for a half-wave layer the refractive index of the layer does not appear in the reflectivity and the result is the same as for an uncoated surface. (This effect is similar to that of a half-wave section of a transmission line.)

An anti-reflection (AR) coating can be made i.e. one that minimizes the reflection by selecting a dielectric material such that $n_o n_T - n_1^2 = 0$ from equation (9.12). This requires $n_1 = \sqrt[2]{n_o n_T}$. For an air/glass boundary this is not possible, the closest we can do is to have n_1 as low as possible e.g. MgF_2 has $n_1 = 1.38$ giving $R \sim 1\%$. Improved AR coatings are made using multiple layers. Coatings may also be made to enhance the reflectivity i.e high reflectance mirrors.

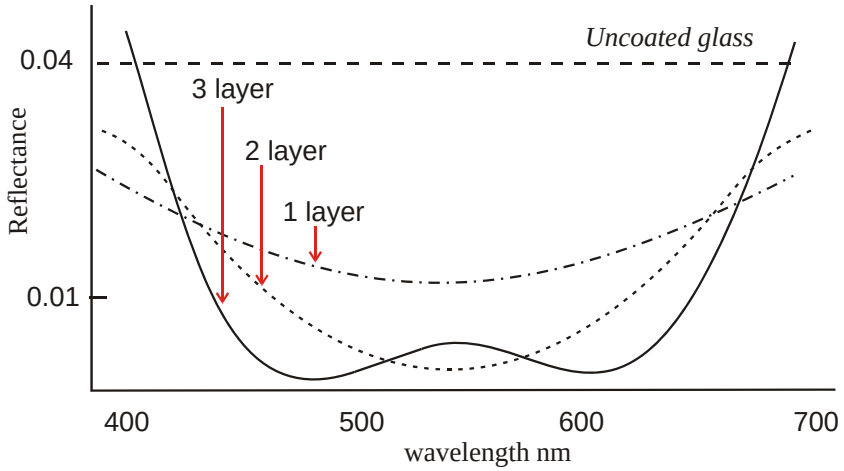


Figure 9.3 Anti-reflection dielectric coatings. A single $\lambda/4$ layer can reduce reflection from 4% to about 1%. Further reduction at specific wavelength regions is achieved by additional layers at the expense of increased reflectivity elsewhere. This enhanced reflection at the blue and red end of the visible is responsible for the purple-ish hue or blooming on camera or spectacle lenses.

9.3 Multiple dielectric layers, Matrix Method.

Write equations (9.7) and (9.8) in terms of *r* i.e. $E_o^{'}/E_o$ and *t* i.e. E_T/E_o

$$1 + r = (A + Bn_T)t$$
$$n_o(1 - r) = (C + Dn_T)t$$

or in matrix form:

$$\begin{pmatrix} 1 \\ n_o \end{pmatrix} + \begin{pmatrix} 1 \\ -n_o \end{pmatrix} r = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} 1 \\ n_T \end{pmatrix} t \qquad (9.14)$$

The characteristic matrix is

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \qquad (9.15)$$

The characteristic matrix for a $\lambda/4$ layer of index n_m is

$$M_m = \begin{pmatrix} 0 & -i/n_m \\ -in_m & 0 \end{pmatrix} \qquad (9.16)$$

A stack of *N* layers has a characteristic matrix:

$$M_{Stack} = M_1 M_2 M_3 \ldots\ldots M_N \qquad (9.17)$$

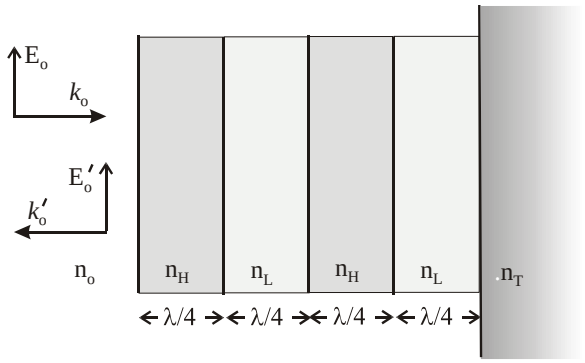


Figure 9.4 Multiple quarter-wave stack

9.4 High reflectance mirrors

A stack of 2 dielectric layers of alternate high and low index n_H, n_L, respectively has the matrix:

$$M_{HL} = \begin{pmatrix} -n_L / n_H & 0 \\ 0 & -n_H / n_L \end{pmatrix} \tag{9.18}$$

For N such pairs the matrix is M_{HL}^{N} . From this 2×2 matrix we find the values of A, B, C and D. Hence from equation (9.10) we find the reflectivity of the composite stack:

$$R_{Stack} = \left\{ \frac{1 - \frac{n_T}{n_o}\left(\frac{n_H}{n_L}\right)^{2N}}{1 + \frac{n_T}{n_o}\left(\frac{n_H}{n_L}\right)^{2N}} \right\}^2 \tag{9.19}$$

9.5 Interference Filters

A Fabry-Perot etalon structure may be constructed from two high reflectance stacks separated by a layer that is $\lambda/2$ or an integer multiple of $\lambda/2$ The half-wave layer(s) acts as a spacer to determine the FSR. The FSR will therefore be very large such that only one transmission peak may lie in the visible region of the spectrum. This is an interference filter.
Narrower range filters may be made by increasing the spacer distance and increasing the reflectance. Extra peaks may be eliminated using a broad band high or low pass filter.

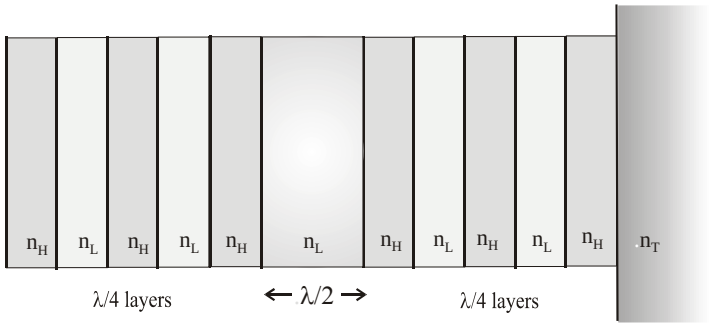


Figure 9.5 Interference filter constructed using multiple dielectric layers consisting of two high-reflectance stacks separated by a $\lambda/2$ layer which acts as a spacer in the Fabry-Perot type interference device. The spacer may be made in integer multiples of $\lambda/2$ to alter the FSR.

10. Polarized light

The polarization of light refers to the direction of the electric field vector **E** of the wave. There are three options for **E**:
(1) its direction and amplitude remains fixed in space - *linear* polarization,
(2) its direction rotates at angular frequency ω about the direction of propagation and the amplitude remains constant - *circular* polarization
(3) its direction rotates at angular frequency ω and its amplitude varies between a maximum and minimum during each complete rotation - *elliptical* polarization.
For propagation in the *x*- direction the vector **E** may be resolved into two orthogonal components E_y and E_z. Each of the three polarization states is thus characterised by a fixed phase relationship between these components. If the phase is randomly varying the light is said to be *unpolarized*.

10.1 Polarization states

An electromagnetic wave travelling in the positive *x* direction has an electric field **E** with components E_y and E_z.

$$E_y = E_{oy}\cos(kx-\omega t)\underline{\mathbf{j}}$$
$$E_z = E_{oz}\cos(kx-\omega t-\delta)\underline{\mathbf{k}} \qquad (10.1)$$

where δ is a relative phase. The light is *polarized* when δ is a constant.

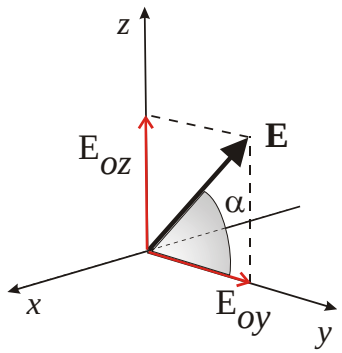


Figure 10.1 Electric field vector in light wave has components E_{oz} and E_{oy} in plane orthogonal to propagation direction along x-axis

Case 1: Linearly polarized light, $\delta = 0$

The components are in phase. The resultant is a vector $\mathbf{E}_P$:

$$\mathbf{E}_P = \{E_{oy}\underline{\mathbf{j}} + E_{oz}\underline{\mathbf{k}}\}\cos(kx-\omega t) \qquad (10.2)$$

at a fixed angle α to the *y*-axis

$$\tan\alpha = \frac{E_{oz}}{E_{oy}} \qquad (10.3)$$

Case 2: Circularly polarized light, $\delta = \pm\pi/2$

Consider $\delta = -\pi/2$ and $E_{oy} = E_{oz} = E_o$

$$E_y = E_o \cos(kx - \omega t)\underline{\mathbf{j}}$$
$$E_z = E_o \sin(kx - \omega t)\underline{\mathbf{k}} \qquad (10.4)$$

$$\tan\alpha = \frac{\sin(kx - \omega t)}{\cos(kx - \omega t)} = \tan(kx - \omega t) \qquad (10.5)$$

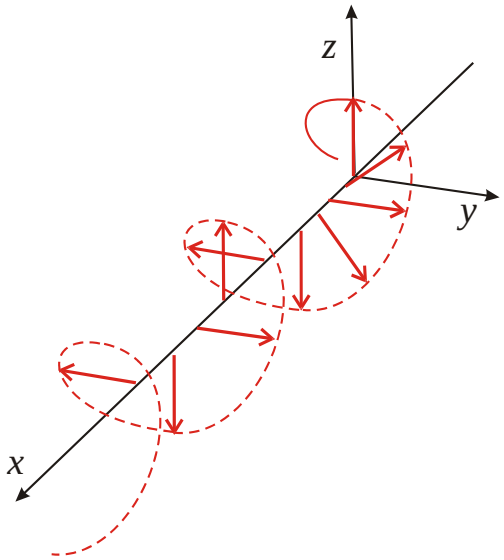


Figure 10.2 Circularly polarized light propagating in the positive x-direction.

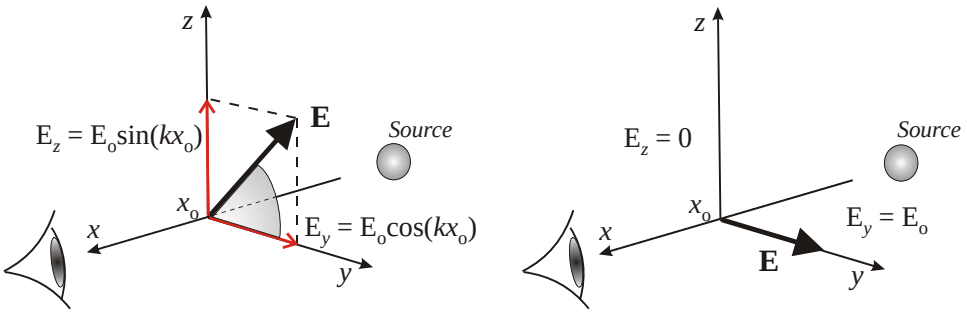
The tip of the **E**-vector rotates at angular frequency ω at any position x on the axis, and rotates by 2π for every distance λ along the x-axis. What is the direction of rotation?
Consider position $x = x_o$ and time t = 0.

$$E_y = E_o \cos(kx_0)$$
$$E_z = E_o \sin(kx_0)$$

The vector is at some angle α.
At position $x = x_o$ and time $t = kx_o/\omega$:

$$E_y = E_o$$
$$E_z = 0$$

The **E**-vector has rotated clockwise as viewed back towards the source. See figure 10.3. This is Right Circularly Polarized light ($\delta = -\pi/2$). *Right* circularly polarized light advances like a *Left*-handed screw!



(a) $x = x_o$, $t = 0$ (b) $x = x_o$, $t = kx_o/\omega$

Figure 10.3 Direction of circular polarization is determined by looking back towards the source. (a) and (b) show E-vector at a point $x = x_o$ ***at time t = 0, and a later time*** $t = kx_o / \omega$ ***. In this case the E-vector has rotated clockwise and is denoted Right Circularly Polarized.***

Conversely, $\delta = +\pi/2$ is Left Circularly Polarized light: viewed towards the source the **E**-vector rotates anti-clockwise. Thus the **E**-vector for right and left circular polarization is written:

$$\begin{aligned} \mathbf{E}_R &= E_o\left[\cos(kx-\omega t)\underline{\mathbf{j}} + \sin(kx-\omega t)\underline{\mathbf{k}}\right] \\ \mathbf{E}_L &= E_o\left[\cos(kx-\omega t)\underline{\mathbf{j}} - \sin(kx-\omega t)\underline{\mathbf{k}}\right] \end{aligned} \tag{10.6}$$

Note that a linear superposition of $\mathbf{E}_R$ and $\mathbf{E}_L$ and gives linear or plane polarized light.

$$\mathbf{E}_P = \mathbf{E}_R + \mathbf{E}_L = 2E_o\cos(kx-\omega t)\underline{\mathbf{j}} \tag{10.7}$$

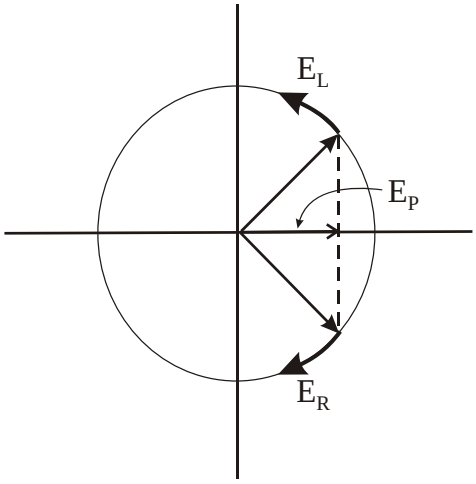


Figure 10.4 Plane polarized light is a superposition of a right- and a left-circularly polarized component.

If the components are of unequal amplitude then the resultant traces out an ellipse i.e the light is elliptically polarized.

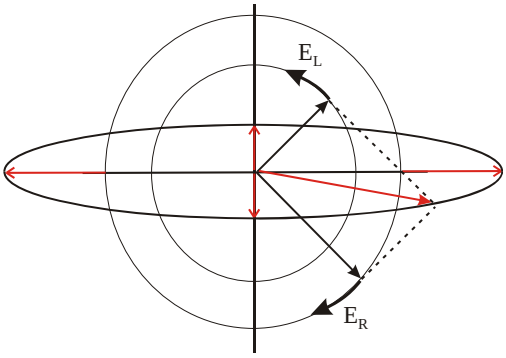


Figure 10.5 A superposition of right- and left-circularly polarized components of unequal magnitude gives elliptically polarized light.

Case 3: Elliptically polarized light.

In general there is a relative phase δ between *y* and *z* components. From (10.1):

$$E_y = E_{oy}\cos(kx - \omega t)$$
$$E_z = E_{oz}\cos(kx - \omega t - \delta)$$

Writing

$$E_z = E_{oz}[\cos(kx-\omega t)\cos\delta - \sin(kx-\omega t)\sin\delta] \qquad (10.8)$$

Substitute in (10.8) using

$$\cos(kx - \omega t) = \frac{E_y}{E_{oy}}, \quad \text{and} \quad \sin(kx - \omega t) = \left[1 - \left(\frac{E_y}{E_{oy}}\right)^2\right]^{1/2}$$

we obtain:

$$\frac{E_y^2}{E_{oy}^2} + \frac{E_z^2}{E_{oz}^2} - 2\frac{E_y}{E_{oy}}\frac{E_z}{E_{oz}}\cos\delta = \sin^2\delta \qquad (10.9)$$

So for $\delta = \pm\pi/2$

$$\frac{E_y^2}{E_{oy}^2} + \frac{E_z^2}{E_{oz}^2} = 1 \qquad (10.10)$$

This is the equation for an ellipse with E_{oy}, E_{oz} as the major/minor axes, i.e. the ellipse is disposed symmetrically about the *y / z* axes.

For $\delta = \pm\pi/2$ the axes of symmetry of the ellipse are rotated relative to the *y / z* axes by an angle θ

$$\tan 2\theta = 2\frac{E_{oy}E_{oz}}{E_{oy}^2 - E_{oz}^2}\cos\delta \qquad (10.11)$$

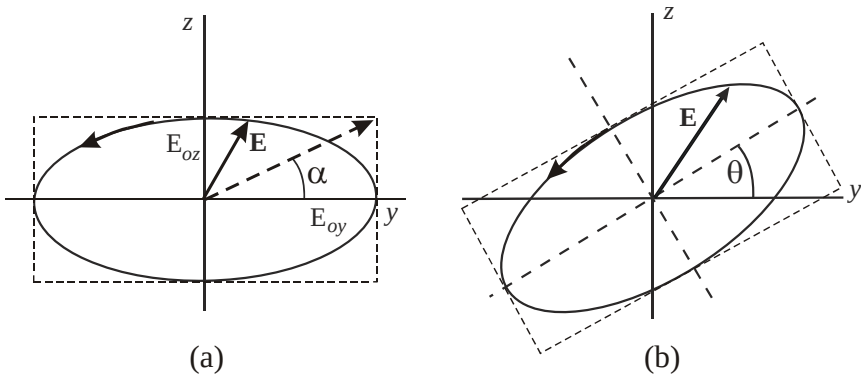


Figure 10.6 Elliptically polarized light (a) axes aligned with *y, z* ***axes***, (b) ***with axes at angle*** θ ***relative to*** *y,z* ***axes.***

As δ varies from $0 \rightarrow 2\pi$ the polarization varies from linear to elliptical and back to linear. Thus we may transform the state of polarization between linear and elliptical or vice-versa by altering the relative phase of the two components. This can be done using a material that has different refractive index for two different directions of polarization i.e. a birefringent material.

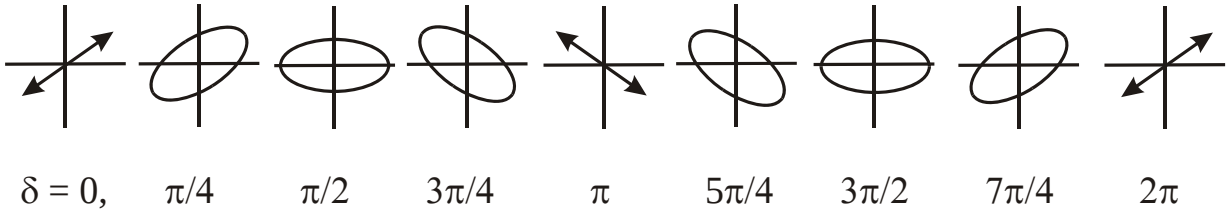


Figure 10.7 General elliptical state of polarization for different values of relative phase δ ***between the components.***

10.2 Optics of anisotropic media; birefringence.

Firstly; some background information that is not required for the syllabus, but is interesting/useful to know about. The optical properties of a material are determined by how the electric field **D** inside the medium is related to an electric field **E** incident "from outside".

$$\mathbf{D} = \varepsilon_o \varepsilon_r \mathbf{E}$$

The permittivity ε_r is a tensor so the components of **D** are related to the components of **E** by:

$$\begin{pmatrix} \mathbf{D}_x \\ \mathbf{D}_y \\ \mathbf{D}_z \end{pmatrix} = \varepsilon_o \begin{pmatrix} \varepsilon_X & 0 & 0 \\ 0 & \varepsilon_Y & 0 \\ 0 & 0 & \varepsilon_Z \end{pmatrix} \begin{pmatrix} \mathbf{E}_x \\ \mathbf{E}_y \\ \mathbf{E}_z \end{pmatrix} \qquad (10.12)$$

The permittivity ε_r of the medium is equal to the square of the refractive index, n^2 [Note that the permittivity tensor matrix has been diagonalized here for simplicity i.e. we have chosen to represent it by components $\varepsilon_X, \varepsilon_Y$ and ε_Z specifying its value along the axes of symmetry.]
An isotropic medium is represented by

$$\varepsilon_X = \varepsilon_Y = \varepsilon_Z \quad (= n^2)$$

A particular type of anisotropic medium is represented by

$$\varepsilon_X = \varepsilon_Y \neq \varepsilon_Z$$

Hence there are, in this type of material, different values of refractive index for light with its E vectors along different axes:

$$n_X^2 = n_Y^2 \neq n_Z^2$$

Now begins the stuff you need to know!

We will be concerned only with *uniaxial*, anisotropic materials i.e. crystals that have two characteristic values of refractive index i.e. *birefringent*. We identify 3 orthogonal axes in a crystal: *x, y* and *z*. If a ray of light is polarized such that the E-vector lies in the xy-plane it experiences a refractive index n_o. [i.e. $n_X = n_Y = n_o$] . Note that the ray may propagate in any direction and, provided its E-vector lies in the *x,y*-plane, it will "see" the refractive index n_o. Such a ray is called an ordinary ray or *o*-ray. n_o is the ordinary index.
If the ray is polarized with the E-vector parallel to the *z*-axis (i.e. it propagates in the *x,y*-plane) it experiences a refractive index n_e, the extra-ordinary index and is the extraordinary ray or *e*-ray. Note that if a ray propagates along the *z*-axis, its E-vector must lie in the *x,y*-plane and it will be an *o*-ray. In this case the direction of the E-vector, i.e. its polarization direction, makes no difference to the refractive index. Thus the z-axis is the axis of symmetry and is called the *Optic Axis*. This is the only axis of symmetry and the crystal is *uniaxial.*

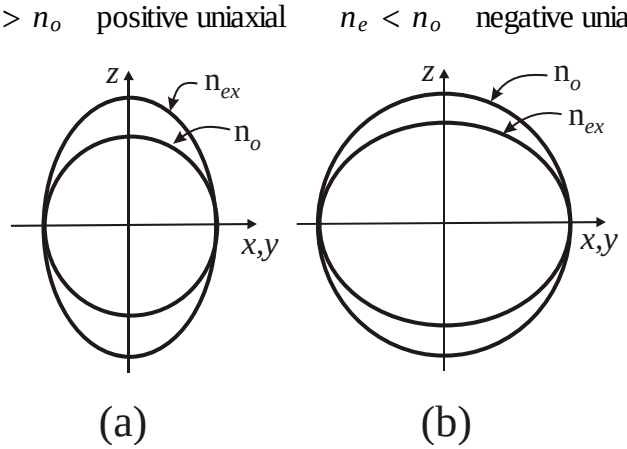


Figure 10.8 (a) positive and (b) negative uniaxial birefringent crystals.

The difference in refractive indices characterizes the degree of birefringence.

$$\Delta n = |n_o - n_e|$$

The wave front of an *o*-ray is spherical whereas the wave front of an *e*-ray is elliptical.

10.3 Production and manipulation of polarized light

At the end of section 10.1 it was noted that the polarization state of a wave may be modified by changing the phase factor δ. This can be done using a crystal cut with parallel faces normal to the x-axis i.e. such that the y,z- plane lies in the faces. A linearly polarized wave travelling in the x-direction in general will have components E_y, E_z along y, z axes which experience refractive indices n_o and n_e respectively. After traversing a length ℓ of the crystal a relative phase shift between the two components will be introduced:

$$\delta = \frac{2\pi}{\lambda}|n_o - n_e|\ell \qquad (10.14)$$

For a given birefringent material the value of δ will be determined by the length ℓ.

Input polarization:	E_y, E_z in phase:	Linear
Output polarization:	δ – phase shift:	Elliptical

The form of elliptical polarization created from a linearly polarized input depends on the value of δ and angle θ of input polarization direction relative to the optic axis (z-axis)

$\theta = 45^o$	$(E_y = E_z)$	$\delta = \pm\frac{\pi}{2}$ (Quarter-wave, $\lambda/4$ plate)	Right/Left Circular
$\theta \neq 45^o$	$(E_y \neq E_z)$	$\delta = \pm\frac{\pi}{2}$ (Quarter-wave, $\lambda/4$ plate)	Right/Left Elliptical
$\theta \neq 45^o$	$(E_y \neq E_z)$	$\delta = \pm\pi$ (Half-wave, $\lambda/2$ plate)	Linear, plane rotated by 2θ

Note: a quarter-wave plate may be used to convert linear to elliptical or vice versa.

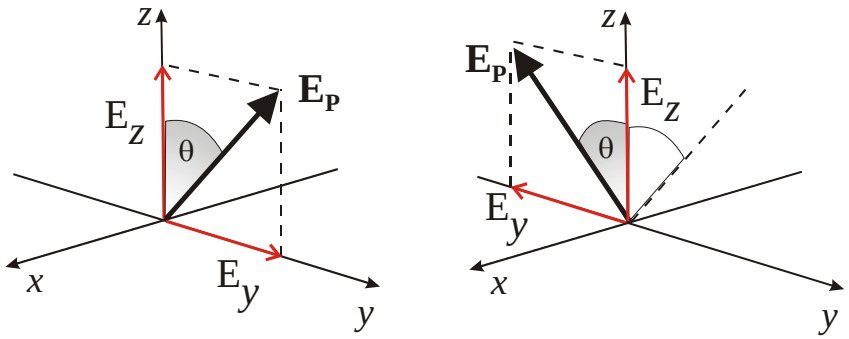


Figure 10.9 Action of a $\lambda/2$-plate with axis at θ to E-vector of plane polarized light shifts phase of one component; E_y by π relative to original phase resulting in a rotation by 2θ of the resultant E-vector.

Polarized light may be produced from unpolarized light using:
(a) Fresnel reflection at Brewster's angle.
(b) "Polaroid-type" material: absorbs one component.
(c) Birefringent prism: *o*-rays and *e*-rays have different refractive indices so different angle of refraction and different critical angles θ_c. Prism may be cut so that beam strikes angled face at incidence angle θ_i where $\theta_i > \theta_c$ for o-ray and $\theta_i < \theta_c$ for e-ray (or vice versa.) Deviation may be compensated by use of a second prism.

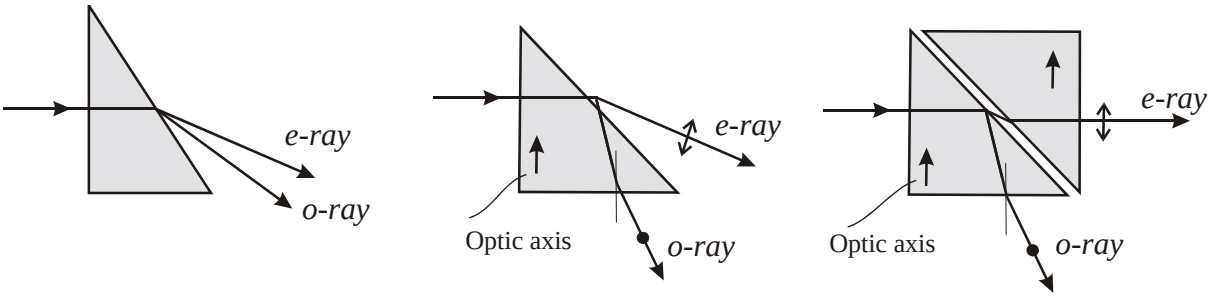


Figure 10.10 Prism polarizers

Variable phase shifts may be introduced using a *compensator*. The Babinet compensator has two wedged plates A and B with orthogonal optic axes. The phase shift in each prism depends on the position across the wedge. $\delta_A = (2\pi/\lambda)\Delta n.d_1$ and $\delta_B = (2\pi/\lambda)\Delta n..d_2$. A net phase shift δ is introduced:

$$\delta = (2\pi/\lambda)\Delta n(d_1 - d_2) \qquad (10.15)$$

The phase shift introduced varies across the aperture.
The Babinet-Soleil produces a variable phase shift that is constant across the aperture.

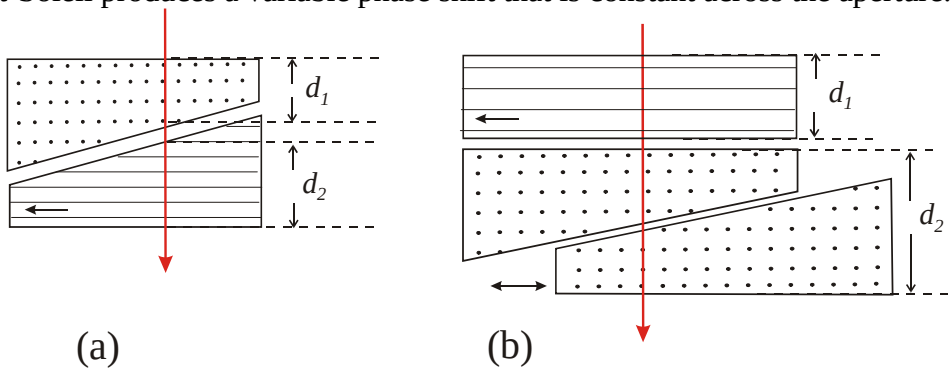


Figure 10.11 (a) Babinet compensator; phase shift introduced to transmitted beam varies across the aperture according to the value of $|d_1 - d_2|$ ***(b) Babinet-Soliel compensator; phase shift is constant across the aperture as*** d_2 ***is varied by position of wedge.***

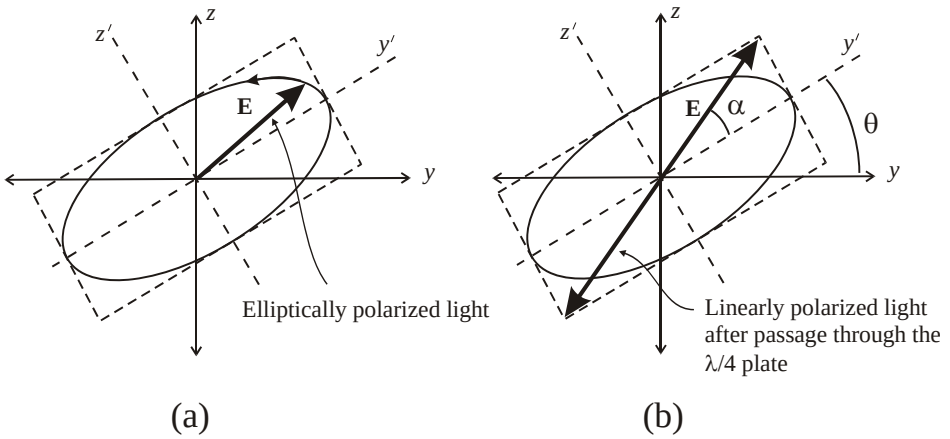
10.4 Analysis of polarized light

The general state of light polarization is elliptical. Linear and circular polarizations are special cases of elliptical; $\delta = 0$ and $\delta = \pm\pi/2$ (with $E_{oy} = E_{oz} = E_o$) respectively. Linear polarization is also a linear superposition of right and left circularly polarized components of equal amplitude. The state of polarization is specified by two parameters: the ratio of E_{oy}/E_{oz}, or $\tan\alpha$, and the phase angle δ. (see figure 10.6) The following two methods may be used to specify the state by determination of these parameters.

(a) Using a Babinet-Soleil, B-S, compensator and a linear polarizer.
If the axis of the B-S is aligned along one component the device may be adjusted to insert a phase shift δ in the other to convert the light to linear polarization. This linear polarization may be extinguished by a crossed, linear polarizer i.e. an analyzer. The B-S is rotated until a position is

found where adjustment of δ leads to light that can be totally extinguished by the analyzer. Note that rotating the linear polarizer without the B-S in position will indicate approximately the orientation of the major and minor axes of the ellipse. The B-S may then be inserted in approximately the correct orientation and both the B-S and the analyzer are adjusted to obtain extinction in a process of iteration.

(b) Using a $\lambda/4$-plate and a linear polarizer.
(i) Rotate linear polarizer to determine approximately the orientation of the major/minor axes of ellipse; the angle obtaining maximum and minimum transmission.
(ii) Set linear polarizer for maximum transmission. Remember that if the coordinate axes are chosen to coincide with the principal axes of the ellipse there is a phase difference $\delta = \pm\pi/2$ between the components.
(iii) Insert $\lambda/4$--plate before the linear polarizer. (The axes of the $\lambda/4$--plate will be known.) Align axis of $\lambda/4$--plate with approximate ellipse axis. If it is exactly along the axis then linearly polarized light will result.
(iv) Rotate linear polarizer to check for complete extinction.
(v) Iterate orientation of $\lambda/4$--plate and linear polarizer to obtain total extinction. $\lambda/4$--plate is now at angle θ to reference axes. The position of total extinction specifies the orientation of the linearly polarized E-vector. The angle between this vector and the axis of the $\lambda/4$--plate is α. The ratio of the E-vector components is $\tan\alpha$.
Thus the ratio E_{oy}/E_{oz} and the orientation of the ellipse is determined (relative to the axes y', z' at this orientation $\delta = \pm\pi/2$).



(a) (b)

Figure 10.12 (a) Elliptically polarized light with axes at arbitrary angle. (b) Linearly polarized light produced from (a) using $\lambda/4$- plate aligned with y', z' axes. Ellipticity is found from $\tan\alpha$.

Any given state of elliptically polarized light may be converted to any desired state of elliptical polarization using a sequence of $\lambda/4$--plate, $\lambda/2$--plate, $\lambda/4$--plate.
First $\lambda/4$--plate is adjusted to give linear polarization at angle set by original elliptical axes.
Axis of $\lambda/2$-plate set at θ relative to E-vector to rotate it by 2θ (θ is chosen to produce the desired orientation of linear polarized light).
Second $\lambda/4$--plate is rotated relative to E-vector of the linearly polarized light to achieve desired elliptical polarization.

10.5 Interference of polarized light.

(a) Orthogonally polarized waves do not interfere. The basic idea of wave interference is that *waves interfere with themselves* not with each other. A dipole source of electromagnetic waves, say an emitting atom, cannot emit simultaneously two orthogonally polarized waves. Thus two orthogonally polarized waves cannot have come from the same source and so cannot interfere. A source e.g. an atom may emit a linearly polarized wave that may be resolved into two orthogonal components. These components may interfere if their planes of polarization are made to be the same e.g. if one component is rotated by a $\lambda/2$ -plate to be parallel to the other. The two components are in-phase i.e. coherent.

(b) Unpolarized light has randomly varying plane of polarization. Interference occurs, for example in a Michelson, because each wave train (photon!) is split into a pair at the beam splitter. Each one of the pair has orthogonal components say E_{oy} and E_{oz}. The y-component of one of the pair interferes with the y-component of the other one of the pair. Likewise the z-components of the split wave interfere to give the composite interference pattern. Thus emission from uncorrelated atoms emitting uncorrelated randomly polarized wave trains still produces an interference pattern.

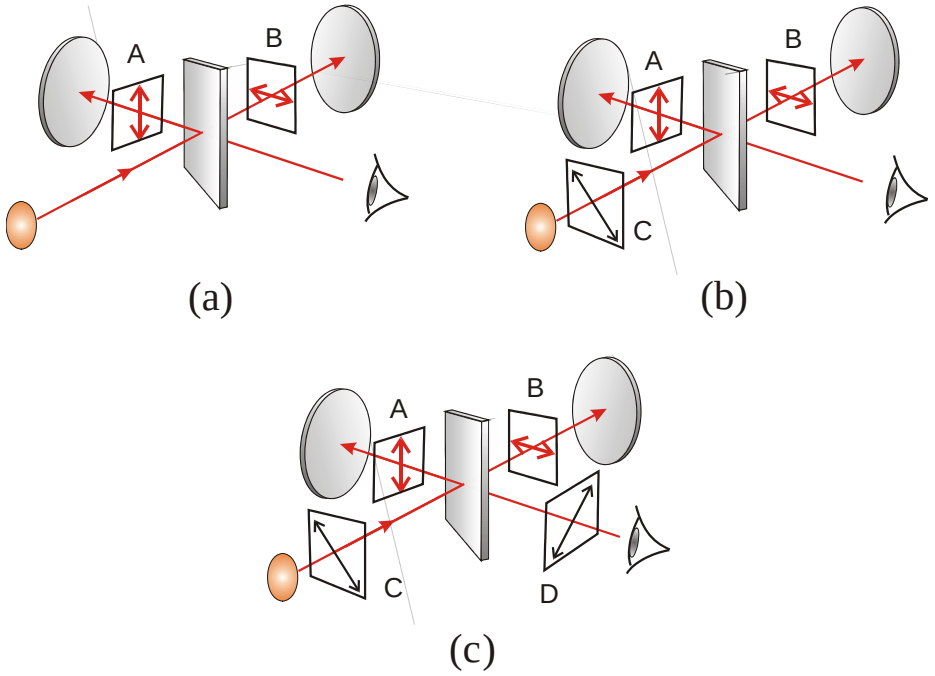


Figure 10.13 Interference of polarized light in a two-beam (Michelson) interferometer A and B are linear polarizers i.e. pass only light polarized in directions shown. Unpolarized light from the source is split at the beam splitter.
(a) Light in paths A and B are orthogonally polarized; no interference.
(b) Linear Polarizer C at 45^{o} produces phase-correlated components passed by A and B. The components are however orthogonally polarized and so no interference is produced.
(c) Linear polarizer D at 45^{o} transmitts phase correlated components from polarizer C that are parallel and so interference is produced.

It is worth noting that unpolarized light can not be fully coherent and so cannot be perfectly monochromatic. Random variation in the plane of polarization results in a random variation of the amplitude along a given axis in space e.g. the y-axis. This is essentially an amplitude modulated wave and so must contain Fourier components i.e. other frequencies. Unpolarized light is therefore not purely monochromatic or fully coherent.