OPTICS: the science of light

2nd year Physics FHS A2

P. Ewart

Introduction and structure of the course.

The study of light has been an important part of science from its beginning. The ancient Greeks and, prior to the Middle Ages, Islamic scholars provided important insights. With the coming of the Scientific Revolution in the 16th and 17th centuries, optics, in the shape of telescopes and microscopes, provided the means to study the universe from the very distant to the very small. Newton introduced a scientific study also of the nature of light itself. Today Optics remains a key element of modern science, not only as an enabling technology, but in Quantum Optics, as a means of testing our fundamental understanding of Quantum Theory and the nature of reality itself.

Geometrical optics, studied in the first year, ignored the wave nature of light and so, in this course, we focus particularly on Physical Optics where the primary characteristic of waves viz. *interference*, is the dominant theme. It is interference that causes diffraction – the bending of light around obstacles. So we begin with a brief résumé of elementary diffraction effects before presenting, in chapter 2, the basics of scalar diffraction theory. By using scalar theory we ignore the vector nature of the electric field of the wave, but we return to this aspect at the end of the course when we think about polarization of light. Scalar diffraction theory allows us to treat mathematically the propagation of light and the effects of obstructions or restrictive apertures in its path. We then introduce a very powerful mathematical tool, the Fourier transform and show how this can help in solving difficult diffraction problems. Fourier methods are used very widely in physics and recognise the inter-relation of variables in different dimensions such as "time and frequency" or "space and spatial frequency". The latter concept will be useful to us in understanding the formation of images in optical systems. Having established the mathematical basis for describing light we turn to methods of analysing the spectral content of light. The spectrum of light is the primary link between optics and atomic physics and other sciences such as astrophysics. The basis for almost all instruments for spectral analysis is, again, interference. The familiar Young's slit, two-beam, interference effect in which the interference arises from division of the wave-front is generalised to multiple slits in the diffraction grating spectrometer. The alternative method of producing interference, by division of amplitude, is then considered. Again we begin with the case of two beams: the Michelson interferometer and move on to multiple-beam interference in the Fabry-Perot interferometer. These devices are important tools and play a key role in modern laser physics and quantum optics. The reflection and transmission of light at boundaries between dielectric media is an important feature of almost all optical instruments and so we then consider how the physics of wave reflection at boundaries can be engineered to produce surfaces with high or partial reflectivity or even no reflectivity at all. Finally we return to the vector nature of the electric field in the light wave. The direction in which the E-field points defines the polarization and this can be random in un-polarized light, fixed in space in linearly polarized light or rotating in space in the case of elliptically or circularly polarized light. We will study how to produce, manipulate and analyse the state of polarization of light.

1. Waves and Diffraction

1.1 Mathematical description of a wave

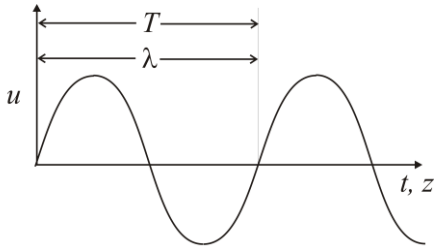


Figure 1.1

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

ωt : phase change with time, $\omega = 2\pi/T$
kz : phase change with distance, $k = 2\pi/\lambda$
α : arbitrary initial phase

Note: this convention for a wave travelling from left to right i.e. in the positive *z* direction follows that used in Quantum Mechanics to describe wave functions.

1.2 Interference

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

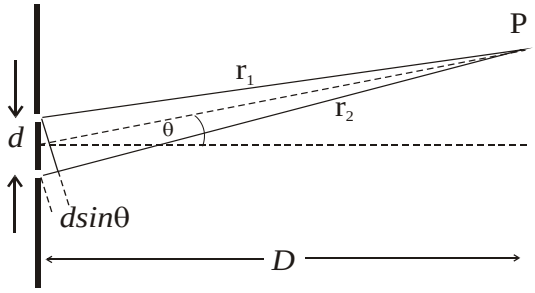


Figure 1.2 Young's slits

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

$$u_p = \frac{u_o}{r_1} e^{i(kr_1 - \omega t)} + \frac{u_o}{r_2} e^{i(kr_2 - \omega t)} \qquad (1.1)$$

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

$$I_p = 4\left(\frac{u_o}{r}\right)^2 \cos^2\left(\frac{1}{2} kd\sin\theta\right) \qquad (1.2)$$

1.3 Phasors

The amplitude of a wave is represented by the length of a "vector" on an Argand diagram. The phase of the wave is represented by the angle of the vector relative to the

Real axis of the Argand diagram.
The **phasor** is then: $ue^{i\delta}$

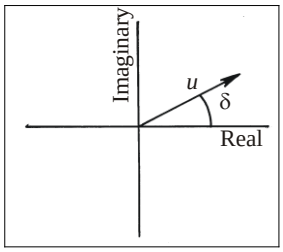


Figure 1.3 Phasor diagram

Example: Young's slits.

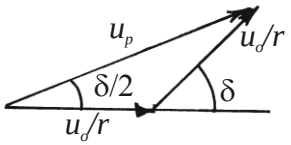


Figure 1.4 Phasor diagram for two slit problem.

Amplitude from each slit on screen: 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:

$$I_p = 4\left(\frac{u_o}{r}\right)^2 \cos^2(\frac{1}{2}kd\sin\theta) \qquad (1.3)$$

1.4 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.

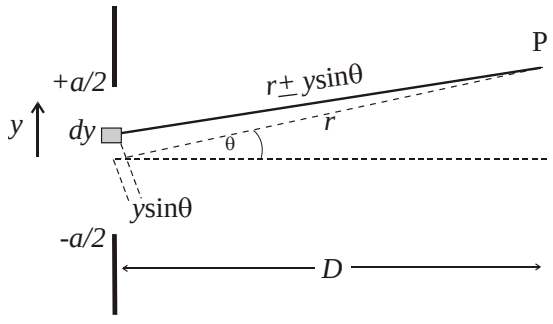


Figure 1.5 Contributions to amplitude at P from elements** dy **in slit.

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

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

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

$$u_p = \frac{u_o}{r} e^{ikr} \int_{-a/2}^{a/2} e^{ik\sin\theta.y} dy \qquad (1.4)$$

The intensity is then: $I_p = I(0)\,\mathrm{sinc}^2\beta$ (1.5)

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

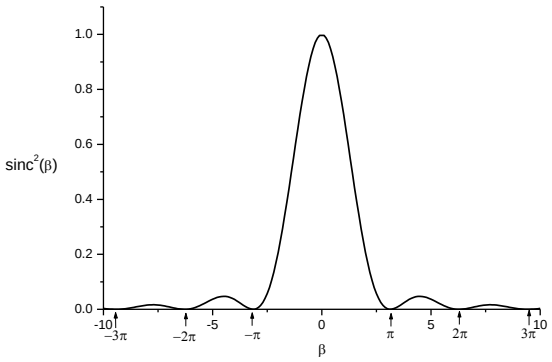


Figure 1.6 Intensity pattern from single slit, $I_p = I(0)\mathbf{sinc}^2\beta$.

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

Hence angular width θ of diffraction peak is:

$$\theta = \frac{\lambda}{a} \qquad (1.6)$$

1.5 Diffraction from a finite slit: *phasor treatment*

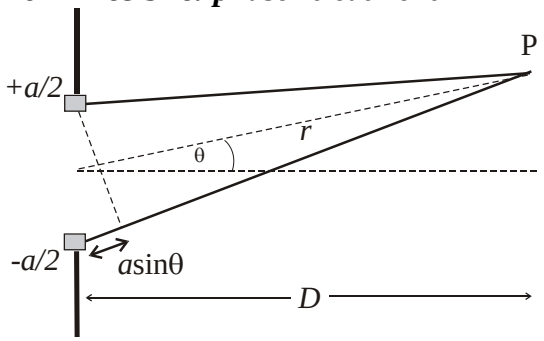


Figure 1.7 Construction showing elements at extreme edges of aperture contributing first and last phasors

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

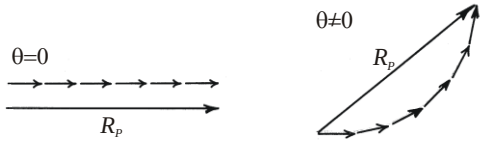


Figure 1.8 (a) Phasor diagram for finite slit showing resultant R_p for $\theta = 0$ and $\theta \neq 0$

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

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

In the limit as the phasor elements $\rightarrow 0$ 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 .

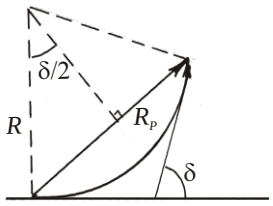


Figure 1.8 (b) Phasor diagram in the limit as phasor elements $\rightarrow 0$.

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

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

Then the intensity at θ:

$$I(\theta) = I(0)\ \mathrm{sinc}^2(\delta/2) = I(0)\ \mathrm{sinc}^2\beta \qquad (1.7)$$

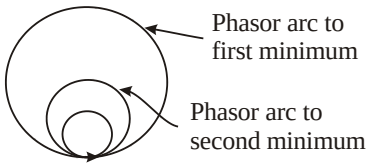


Figure 1.9 Phasor diagram showing mimina for increasing phase shift δ between extremes of slit as θ increases.

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 $\delta = 2\pi$, the angular width is:

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

1.6 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 (eqn 1.4):

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

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 :

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

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

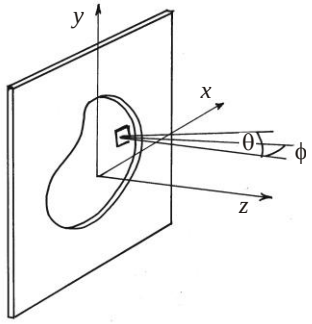


Figure 1.10 General 2-D aperture in** x,y **plane.

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:

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

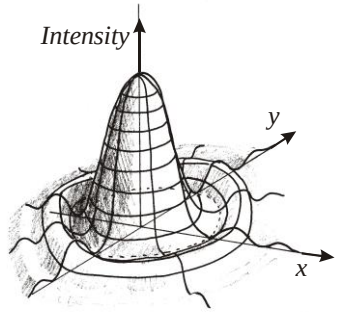


Figure 1.11. Point Spread Function for circular aperture.

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

2. 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.

2.1 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.

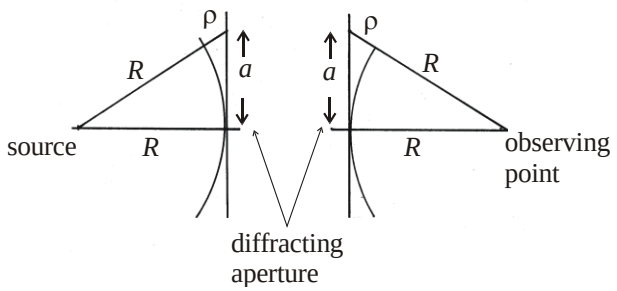


Figure 2.1 Wavefronts incident on and exiting from a plane aperture.

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

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

Alternatively,

"*Fraunhofer diffraction is the diffraction observed in the image plane of an optical system*."

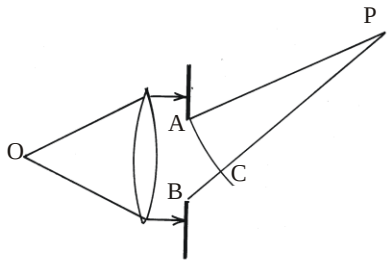


Figure 2.2 Fraunhofer condition for plane waves: image is at infinity as source is at focal length from lens.

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 $BC \leq \lambda/20$

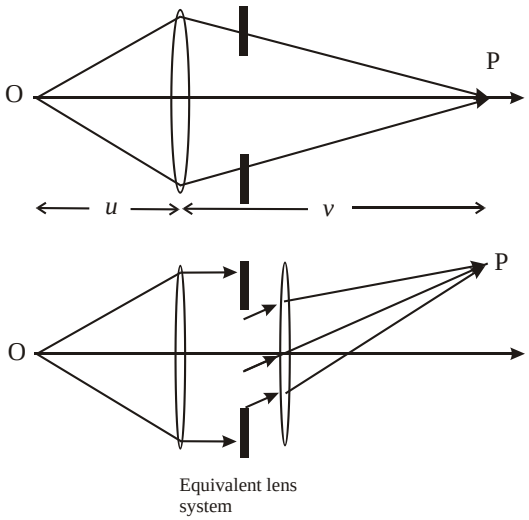


Figure 2.3 Fraunhofer diffraction observed in the image plane of a lens.

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.

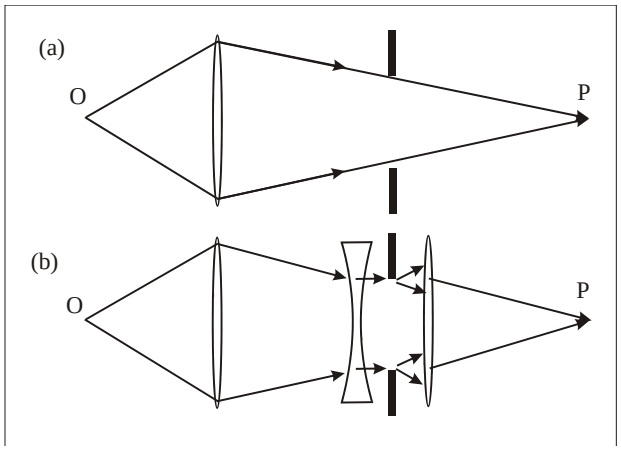


Figure 2.4 Equivalent lens system showing that Fraunhofer diffraction is independent of position of aperture

2.2 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.

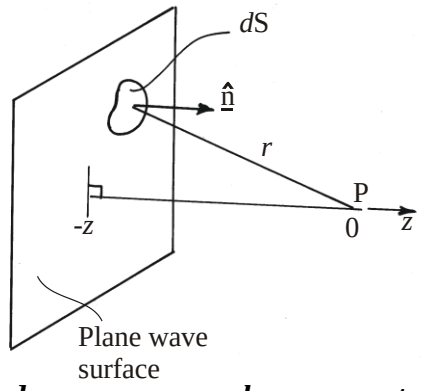


Figure 2.5 Huygens secondary sources on plane wave at -z contribute to wave at P.

The amplitude at P is the resultant of all contributions from the plane at –*z*.

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

u_o is the amplitude from element of area *dS*.

$\eta(\underline{\mathrm{n}},\underline{\mathrm{r}})$ 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 and 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 now construct elements of the wavefront of equal area δA centred on O.

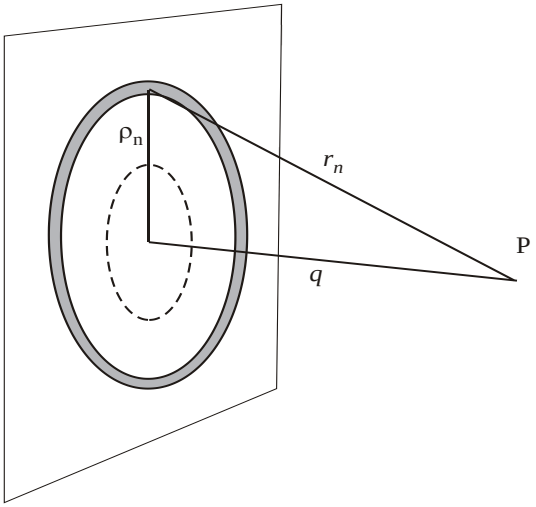


Figure 2.6 Construction of elements of equal area on plane wavefront.

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

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

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

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

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

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

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.

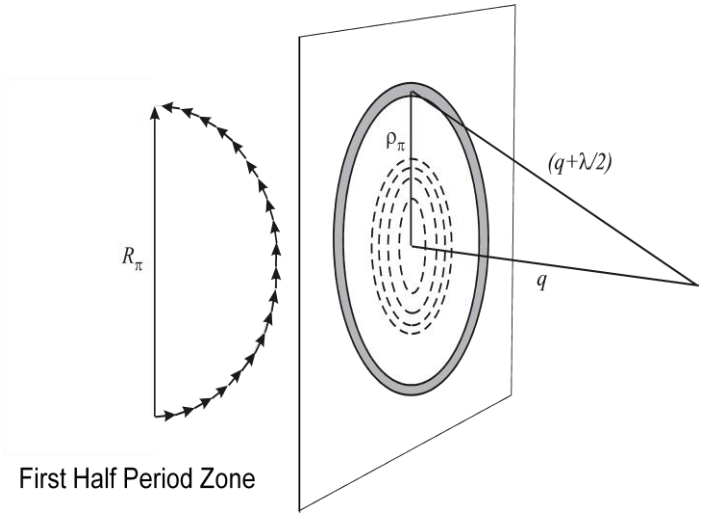


Figure 2.7 Construction of the First Half Period Zone

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

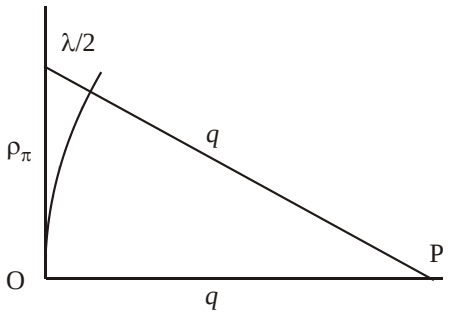


Figure 2.8 Phase shift of $\lambda/2$ arises at edge of 1st HPZ

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

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

Recalling our diffraction integral we 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 2π. 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.

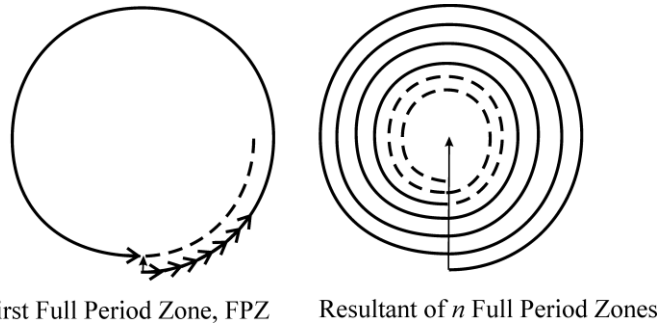


Figure 2.9 As $n \to \infty$ resultant of zones tends to half the resultant of the 1st HPZ

Adding contributions from the whole wave (integrating over infinite surface) gives resultant equal to ½ the 1st HPZ. Therefore

$$\mathrm{R}_{\infty} = i\alpha u_o \lambda$$

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

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

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

Hence

$$u_p = -\frac{i}{\lambda}\int \frac{u_o \mathrm{dS}}{r}\,\eta(\mathrm{n},\mathrm{r})\,e^{ikr} \qquad (2.1)$$

This is the Fresnel-Kirchoff diffraction integral.

3. Fourier methods in Optics

3.1 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

$$u_p = -\frac{i}{\lambda}\int \frac{u_o \mathrm{d}S}{r}\eta(\underline{n},\underline{r})e^{ikr}, \tag{3.1}$$

where the limits of integration will be defined by the boundary of the aperture. We simplify by:

Ignoring the obliquity factor i.e. put $\eta(\underline{n},\underline{r}) = 1$,
Restricting to one dimension: $\mathrm{d}S \rightarrow \mathrm{d}x$,
Ignoring the $\frac{1}{r}$ term by considering only a small range of r,
Using the Fraunhofer condition:

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

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

Since the integral will be zero wherever the amplitude function *u(x)* is zero the limits of integration can be safely extended to infinity. The amplitude u_p as a function of angle θ is then:

$$u_p \Rightarrow A(\beta) = \alpha \int_{-\infty}^{\infty} u(x)e^{i\beta x}\mathrm{d}x, \tag{3.2}$$

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

We note that *A(β)* is the Fourier transform of *u(x)***.**

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 **k** of the wave incident on the aperture.

The inverse transform relation is:

$$u(x) = \frac{1}{\alpha}\int_{-\infty}^{\infty} A(\beta)e^{-i\beta x}\mathrm{d}\beta \tag{3.3}$$

3.2 The Convolution Theorem

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

$$h(x) = f(x) \otimes g(x) = \int_{-\infty}^{\infty} f(x')g(x-x')dx' \qquad (3.4)$$

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 $H(\beta)$

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:

$$H(\beta) = F(\beta).G(\beta) \qquad (3.5)$$

3.3 Some useful Fourier transforms and convolutions

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

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

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

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

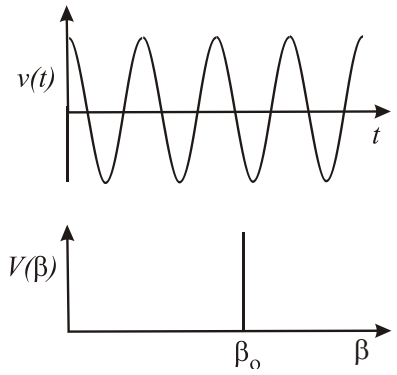


Figure 3.1 A wave of constant frequency (monochromatic) and its Fourier transform

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

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

(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:

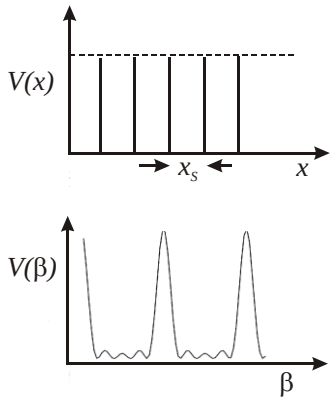


Figure 3.2 A comb of δ -functions and their transform

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

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 $e^{i\alpha}$ 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:

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

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

[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:

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

(f) The grating function:

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

(g) The triangle function:

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

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

3.4 Fourier Analysis

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

$$V(t) = c_o + \sum_{p=1}^{\infty} c_p \cos(p\omega_o t) + \sum_{p=1}^{\infty} s_p \sin(p\omega_o t) \qquad (3.6)$$

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.

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

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

$$c_o = \frac{1}{\tau}\int_0^\tau V(t)\,\mathrm{d}t$$

In general:

$$V(t) = \sum_{p=1}^{\infty} A_p e^{-ip\omega_o t} \qquad (3.7)$$

$$A_p = \frac{1}{\tau}\int_0^\tau V(t)\, e^{ip\omega_o t}\mathrm{d}t \qquad (3.8)$$

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.

3.5 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 at $\theta = \pm\lambda\omega_n / 2\pi$.

[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.]

3.6 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.*

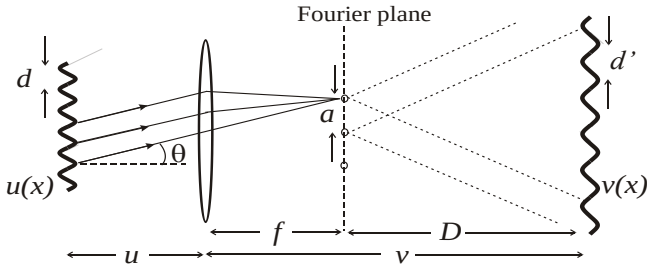


***Figure 3.3 Object** u(x) **imaged by lens to v(x)**.*

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 a of the lens and object distance u determine the maximum angle $\theta_{\max}$ from which light 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.]

3.7 The Compound Microscope

Figure 3.4 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.

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

Spatial frequencies having dimensions smaller than $d_{\min}$ will diffract to larger angles, miss the objective, and thus not appear in the image. The minimum spatial dimension d_{min}

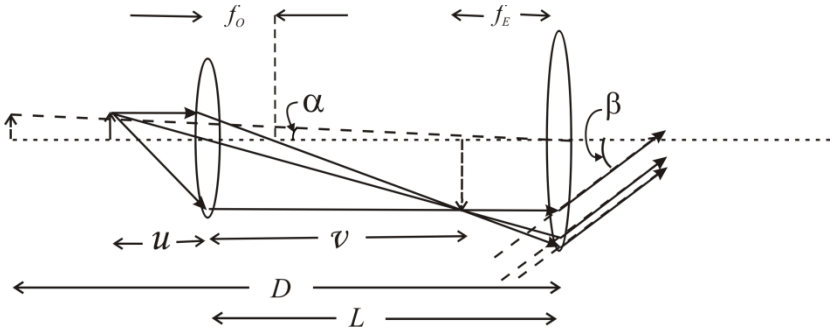


Figure 3.4 The Compound Microscope. The object at distance u from objective with focal length f_O ***is imaged at distance v. Real image is at focal length*** f_E ***from eyepiece giving angular magnification*** β/α ***where*** α ***is the angle subtended by the real image if it was at the near point of the eye, distance D. Approximately,*** $u = f_O$ ***and*** $v = L$***, the length of the tube. In this approximation the magnification is*** $M = DL / f_O f_E$

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.

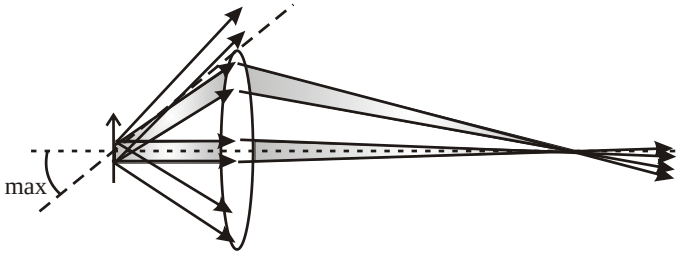


Figure 3.5. First order diffracted waves from spatial structures < d_{min} are collected by the lens and interfere in the image plane with zero order waves to form sinusoidal structure in the image. Light from smaller spatial structures (higher spatial frequencies) are diffracted to angles > $\theta_{\max}$, miss the objective and do not interfere with zero order in the image.

3.7 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* ~ λ/d_A light is lost from the image by diffraction. This is diffraction limited imaging.

For non-diffraction limited imaging: *Image brightness* $\propto d_A^2$

For diffraction limited imaging: *Image brightness* $\propto d_A^4$

4 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.

4.1 Division of wavefront

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

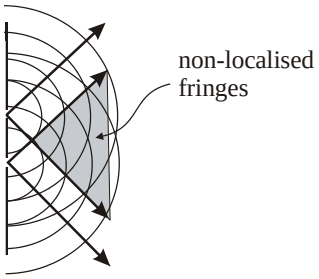


Figure 4.1 Young's slit fringes are observed throughout the region beyond the screen containing the two slits.

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

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

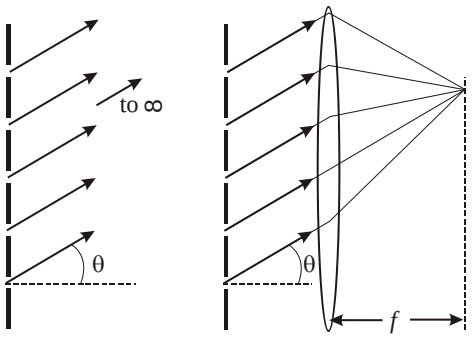


Figure 4.2 Diffraction grating fringes.

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.

4.2 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.

4.2.1 Point source

(a) Wedge.

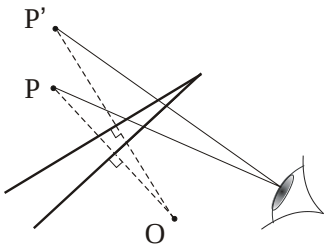


Figure 4.3 A point source O provides images P, P' in reflecting surfaces forming a wedge.

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

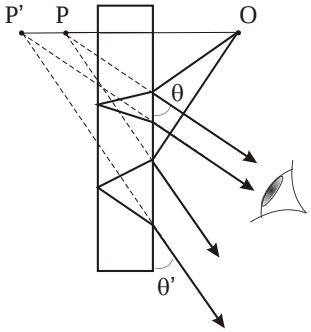


Figure 4.4 A point source reflected in two parallel surfaces again provides two images P, P'

This is similar to the wedge situation with 2-point sources. The fringes are *non-localized* fringes of *equal inclination.*

4.2.2 Extended source

(a) Wedge

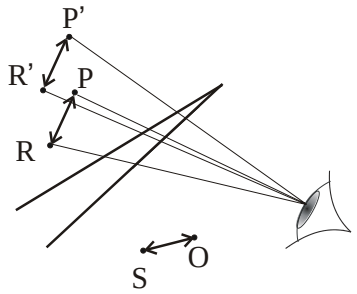


Figure 4.5. Extended source OS provides two images PR and P'R' by reflection at wedged reflecting surfaces.

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

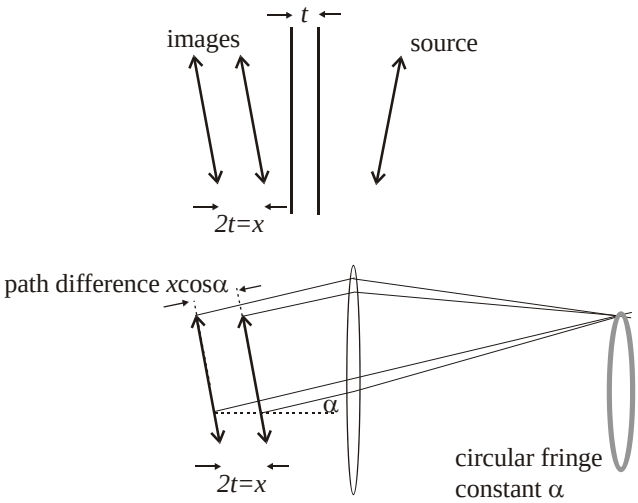


Figure 4.6 Upper figure shows two images of extended source by reflection in parallel slab of thickness *t.* ***Lower figure shows fringes of equal inclination formed in focal plane of a lens by light from the two images of the source.***

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:

$$\delta = \frac{2\pi}{\lambda} x \cos\alpha \qquad (4.1)$$

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

$$x\cos\alpha = p\lambda \qquad (4.2)$$

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 :

$$r_p^2 - r_{p+1}^2 = \frac{2f^2\lambda}{x} \qquad (4.3)$$

As x increases, fringes get closer together. As x decreases $\rightarrow 0$ 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.

5 The diffraction grating spectrograph

5.1 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) $N = 2$

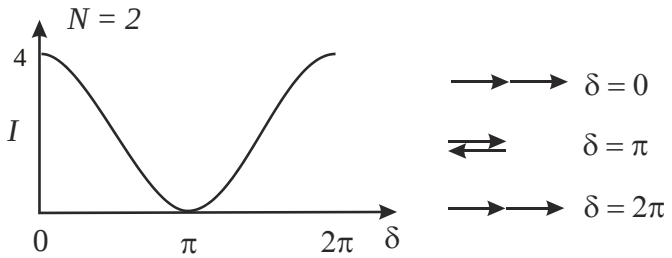


Figure 5.1 Intensity pattern and associated phasor diagram for 2-slit interference

$$I(\theta) = 4u^2 \cos^2(\frac{\delta}{2}) \qquad (5.1)$$

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

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

(2) $N = 3$

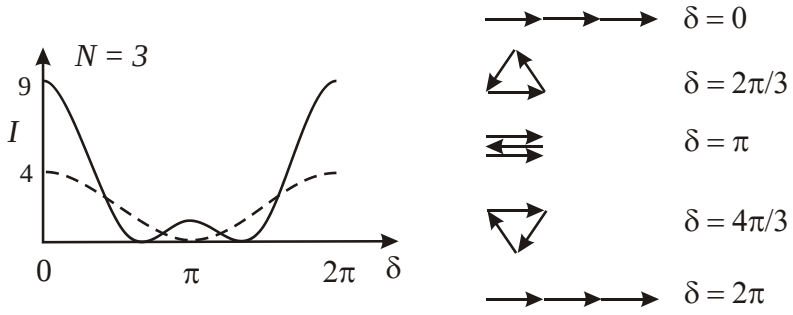


Figure 5.2 Phasor diagrams for 3-slit interference and intensity pattern

Using phasors to find resultant amplitude

(a) $\delta = 0, n2\pi$ Principal maxima of intensity $9u^2$. Two 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 $16u^2$. Three minima between principal maxima.

In general we have principal maxima at $\delta = 0, n2\pi$, intensity $\propto N^2$. $(N-1)$ minima at $n2\pi/N$ and width of principal maxima $\propto 1/N$.

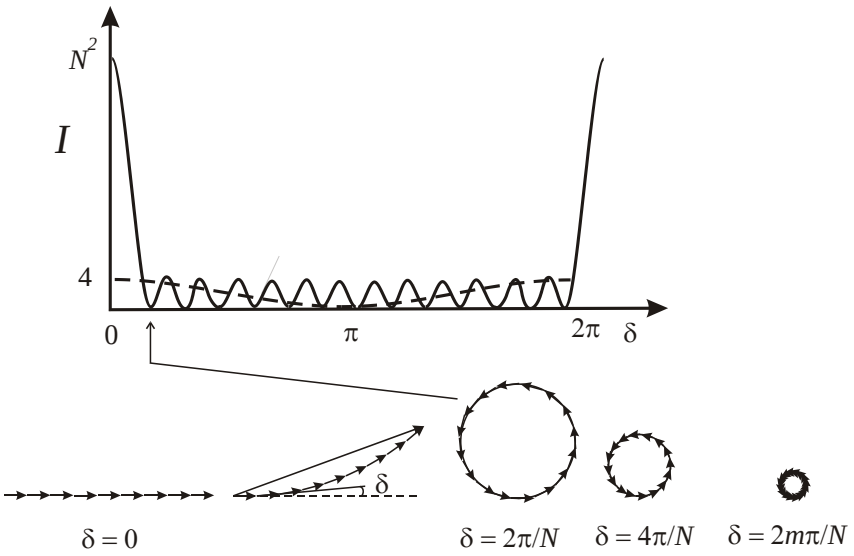


Figure 5.3 Phasor diagrams for N-slit interference and intensity pattern

Amplitude of N phasors:

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

Hence intensity:

$$I(\theta) = I(0)\frac{\sin^2(\frac{N\delta}{2})}{\sin^2(\frac{\delta}{2})} \qquad (5.2)$$

5.2 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 $f(x)$ with a single slit (top-hat function) $g(x)$:

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

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

Using the Convolution Theorem with,

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

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

Hence $$|H(\beta)|^2 = I(\theta) = I(0)\frac{\sin^2(\frac{N\delta}{2})}{\sin^2(\frac{\delta}{2})}.\frac{\sin^2(\frac{\gamma}{2})}{(\frac{\gamma}{2})^2} \quad (5.3)$$

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

5.3 Diffraction grating performance

5.3.1 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. Principal maxima occur for

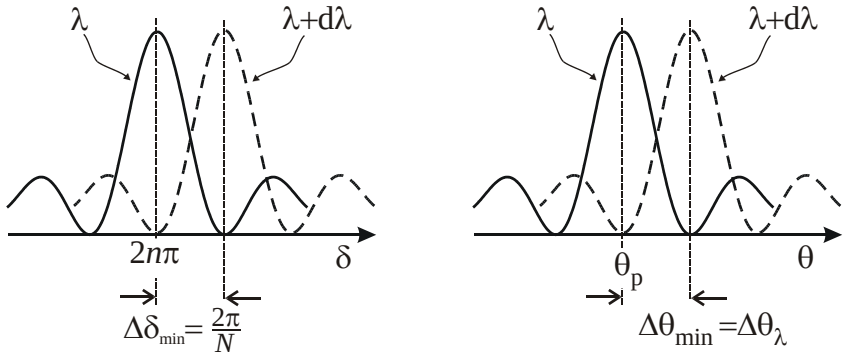
$$d\sin\theta = n\lambda \quad (5.4)$$

5.3.2 Angular dispersion

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

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = \frac{n}{d\cos\theta} \quad (5.5)$$

5.3.3 Resolving power



(a) (b)

Figure 5.4 (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 $\Delta\theta_\lambda$ of 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 $\Delta\delta_{min}$

$$\Delta\delta_{\min} = \pm\frac{2\pi}{N} \quad (5.6)$$

and

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

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

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

Thus the phase difference between the maximum and first minimum is:

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

$$\therefore \Delta\theta_{\min} = \frac{\lambda}{Nd\cos\theta} \qquad (5.7)$$

The angular separation $\Delta\theta_\lambda$ of principal maxima for λ and $\lambda + \Delta\lambda$ is found from (5.5):

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

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

The resolution criterion is:

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

Hence the Resolving Power is:

$$\frac{\lambda}{\Delta\lambda} = nN \qquad (5.8)$$

5.3.4 Free Spectral Range

The 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:

$$\Delta\lambda_{FSR} = \frac{\lambda}{(n+1)} \qquad (5.9)$$

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

5.4 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

$$d(\sin\theta \pm \sin\phi) = n\lambda \qquad (5.10)$$

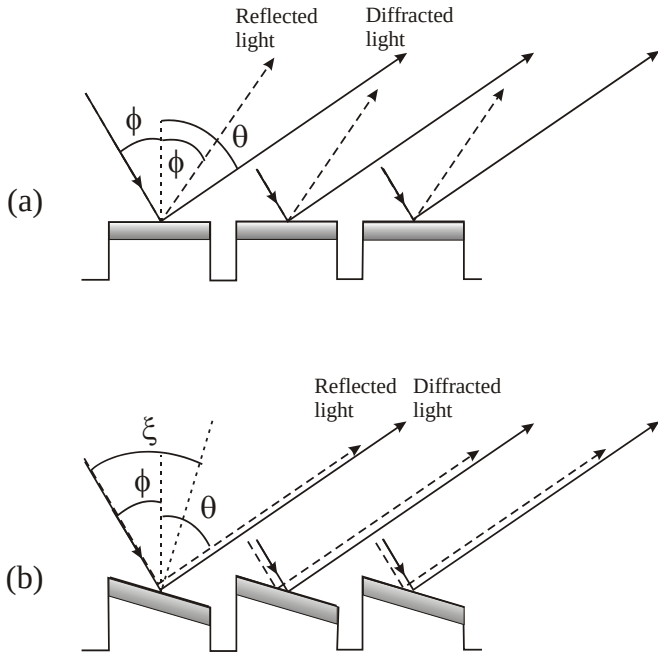


Figure 5.5 (a) Diffraction angle $\theta \neq$ Reflection angle ϕ for ordinary grating. (b) Blazed grating reflects light at same angle as diffracted order

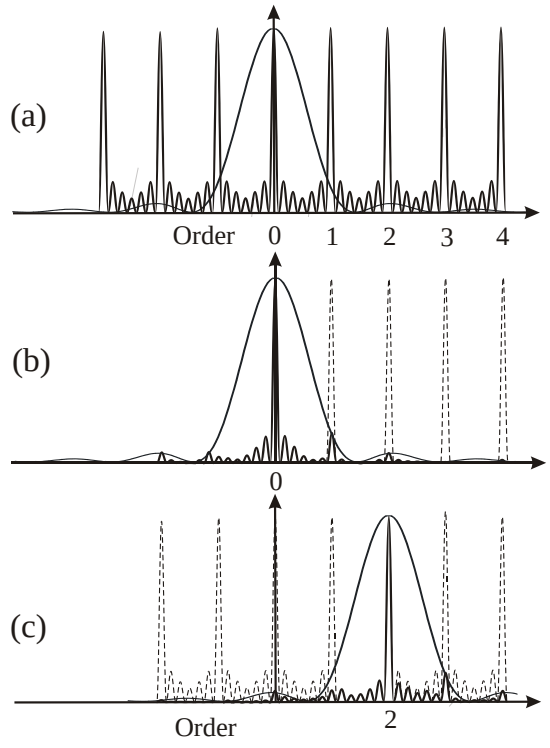


Figure 5.6 (a) Grating intensity pattern and single slit diffraction pattern. (b) Effect of single slit diffraction envelope on grating diffraction intensity for unblazed grating. (c) Grating intensity pattern for blaze set to reflect light into 2nd order.

5.5 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:

$$\Delta x_i = \left(\frac{f_2}{f_1} \right) \Delta x_s \tag{5.11}$$

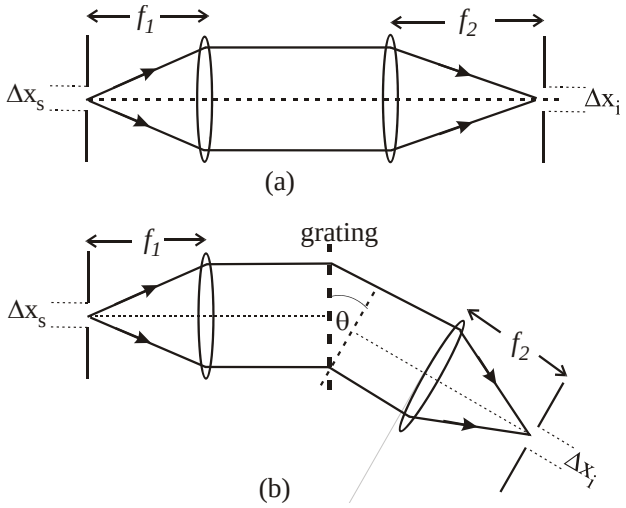


Figure 5.7 (a) Image forming system to image slit of width Δx_s to image Δx_i . (b) Images of slit are spectrally dispersed by diffraction at grating. Slit is imaged at angle θ from diffraction grating leading to foreshortening by $\cos\theta$.

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 :

$$\Delta x_R = \frac{f_2 \lambda}{Nd \cos\theta} \tag{5.12}$$

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

Resolution is achieved provided: $\Delta x_i \le \Delta x_R$ and the limiting slit width Δx_s is then:

$$\Delta x_s \le \frac{f_1 \lambda}{Nd} \quad \text{or} \quad \Delta x_s \le \frac{f_1 \lambda}{W} \tag{5.13}$$

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 brightness is reduced.

The point of this exercise is to compare the size of the slit image in the "real" spectrograph with the size predicted using the theoretical resolving power. The key difference between the two is that the image in the *real* spectrograph, because it is observed at the diffraction angle θ, is foreshortened by a factor $\cos\theta$. So this means that we have to make the observed image size Δx_i larger by a factor $1/\cos\theta$ in order to compare like with like. The theoretical resolving power leads us to a value for the angular separation of wavelengths that differ by $\Delta\lambda_R$. Using our expression for the angular dispersion we find this angular separation to be:

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

This is equation (5.13). We now need to find the spatial extent of the image, Δx_R, corresponding to this theoretical resolution and this is given simply by the relation $f\Delta\theta$ where the focal length in this case is f_2

$$\Delta x_R = f_2\left(\frac{n}{d\cos\theta}\right)\Delta\lambda_R$$

then using, $\Delta\lambda_R = \frac{\lambda}{nN}$: $$\Delta x_R = \frac{f_2\lambda}{Nd\cos\theta}$$

This is equation (5.14).

NB. This calculated "theoretical" image size does not include any foreshortening effect.
Now consider the size of the image in the "real" spectrograph that *will* exhibit foreshortening. The size of the image formed in the lens imaging system of the spectrograph is determined by the linear magnification of the lens system: $M = v/u = f_2/f_1$. So the size of any image is related to the entrance slit width Δx_s by the relation:

$$\Delta x_i = \frac{f_2}{f_1}\Delta x_s$$

When an image is viewed in the diffracted beam at angle θ there is a foreshortening by a factor of $\cos\theta$, i.e. $\Delta x_i' = \Delta x_i \cos\theta$. Conversely any image of size $\Delta x_i'$ viewed in the diffracted beam corresponds to an un-foreshortened image that will be larger by the same factor i.e. $1/\cos\theta$.

$$\Delta x_i = \frac{f_2\Delta x_s}{f_1\cos\theta}$$

This is equation (5.12). It is this image size – without the effect of foreshortening – that needs to be compared to the theoretical image size. In order to achieve the theoretical resolving power we need to have: $\Delta x_i = \Delta x_R$
In which case:

$$\frac{f_2\Delta x_s}{f_1\cos\theta} = \frac{f_2\lambda}{Nd\cos\theta}$$

Or the limiting slit width is

$$\Delta x_s \le \frac{f_1\lambda}{Nd}$$

We note that if the slit is less than this limiting value the resolution is not improved – the image simply becomes less bright – we have diffraction limited imaging.

6 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.

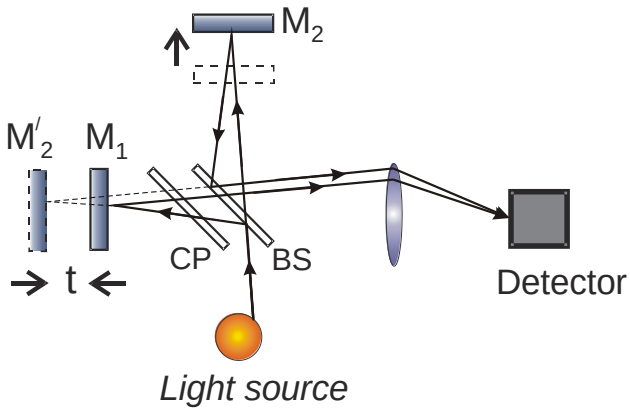


Figure 6.1 The Michleson interferometer. The beam splitter BS sends light to mirrors M_1 and M_2 in two arms differing in length by t. M'_2 is image of M_2 in M_1 resulting effectively in a pair of parallel reflecting surfaces illuminated by an extended source as in figure 5.6. CP is a compensating plate to ensure beams traverse equal thickness of glass in both arms.

6.1 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:

$$\delta = \frac{2\pi}{\lambda} 2t\cos\alpha = \frac{2\pi}{\lambda} x\cos\alpha \qquad (6.1)$$

Constructive interference occurs at $\delta = 2p\pi$, where p is an integer, $x\cos\alpha = p\lambda$. Thus 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).

$$r_p^2 - r_{p+1}^2 = \frac{2f^2\lambda}{x} \qquad (6.2)$$

Two-beam interference pattern:

$$I(x) = I(0)\cos^2\left(\frac{\delta}{2}\right)$$
$$I(x) = \frac{1}{2} I(0)\left[1+\cos 2\pi\bar{\nu}x\right] \qquad (6.3)$$

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

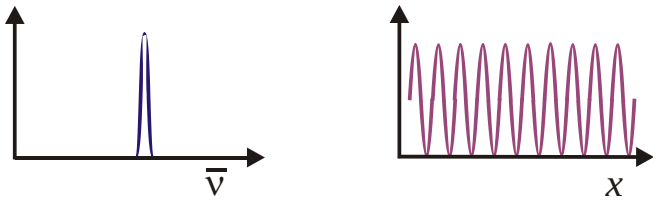


Figure 6.2 Input spectrum of monochromatic source and resulting interferogram obtained from scanning Michelson interferometer.

6.2 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

$$I(x) = I_0(\bar{\nu})\left[1 + \cos 2\pi\left(\frac{\bar{\nu}_1 + \bar{\nu}_2}{2}\right)x \cos 2\pi\left(\frac{\bar{\nu}_1 - \bar{\nu}_2}{2}\right)x\right] \qquad (6.4)$$

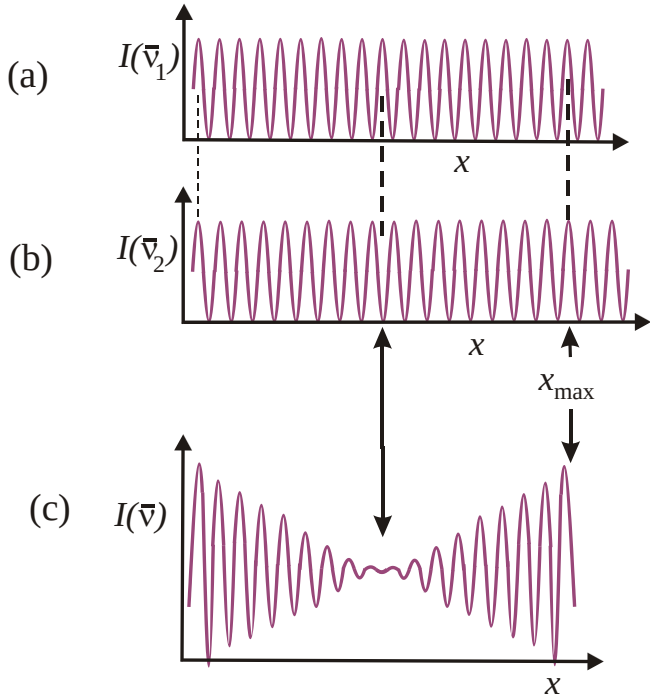


Figure 6.3 (a) Interferogram of source component $\bar{\nu}_1$ (b) interferogram of source component $\bar{\nu}_2$. (c) Interferogram of combined light showing added intensities (a) and (b). Note visibility of fringes cycles to zero and back to unity for equal intesity components. To resolve the complete cycle requires a path difference x_{max}

This looks like an interferogram of a light source with mean wavenumber $\left(\frac{\bar{\nu}_1+\bar{\nu}_2}{2}\right)$ multiplied by an envelope function $\cos 2\pi\left(\frac{\bar{\nu}_1-\bar{\nu}_2}{2}\right)x$. This envelope function goes first to a zero when a "peak" of interferogram for $\bar{\nu}_1$ first coincides with a zero in the interferogram for $\bar{\nu}_2$. 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 $\Delta\bar{\nu} = \bar{\nu}_1 - \bar{\nu}_2$. 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 $\Delta\bar{\nu}_{\min}$ that can be resolved is found from the value of $x_{\max}$ giving the cycle in the cosine envelope function:

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

$$\Delta\bar{\nu}_{\min} = \frac{1}{x_{\max}} \tag{6.5}$$

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.

$$\Delta\bar{\nu}_{Inst} = \frac{1}{x_{\max}} \tag{6.6}$$

Hence the Resolving Power RP is:

$$\text{RP.} = \frac{\lambda}{\Delta\lambda_{Inst}} = \frac{\bar{\nu}}{\Delta\bar{\nu}_{Inst}} = \frac{x_{\max}}{\lambda} \tag{6.7}$$

6.3 The Fourier Transform spectrometer

In Figure 6.1 we see that the interferogram looks like the Fourier transform of the intensity spectrum. The interferogram produced using light of two wavenumbers $\bar{\nu}_1$ and $\bar{\nu}_2$ is

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

In the case of multiple discrete wavelengths:

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

First term on r.h.s. is ½ I_o where I_o is the total intensity at x = 0 and the Second term is a sum of individual interferograms.

Replacing components with discrete wavenumbers by a continuous spectral distribution, *I(x)* becomes:

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

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

$$F(x) = \int_{-\infty}^{\infty} S(\bar{\nu}) \cos 2\pi \bar{\nu} x . d\bar{\nu} \tag{6.8}$$

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

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

Or

$$S(\bar{\nu}) \propto F.T.\left\{ I(x) - \frac{1}{2} I_0 \right\} \tag{6.9}$$

Apart from a constant of proportionality the Fourier transform of the interferogram yields the Intensity or Power Spectrum of the source. See Figure 6.2.

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 $\bar{\nu}$ are Fourier pairs or conjugate variables.[see equation (7.9)] This explains why the limit on the uncertainty of wavenumber (or wavelength) measurement $\Delta\bar{\nu}_{Inst}$ is just the inverse of $x_{\max}$. In essence this explains the general rule for all interferometers including diffraction grating instruments that:

$$\Delta\bar{\nu}_{Inst} = \frac{1}{\text{Maximum path difference between interfering beams}} \tag{6.10}$$

6.4 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 + τ)* integrated over many cycles. (The delay $\tau = x/c.$) The interferogram as a function of the delay may be written:

$$\gamma(\tau) = \int E(t) E(t+\tau) \mathrm{d}t \tag{6.11}$$

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

$$\Gamma(\tau) = \int_{-\infty}^{\infty} E(t) E(t+\tau) \mathrm{d}t \tag{6.12}$$

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

$$F.T.\{\Gamma(\tau)\} = \left|F(\omega)\right|^2 = F^*(\omega).F(\omega) \qquad (6.13)$$

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.

6.5 Fringe visibility.

6.5.1 Fringe visibility and relative intensities

Figure 6.3 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

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \qquad (6.14)$$

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 $I_1(\bar{\nu}_1)$ and $I_2(\bar{\nu}_2)$ then the envelope function of the interferogram does not go to zero. The contrast of the fringes varies from $I_1(\bar{\nu}_1) + I_2(\bar{\nu}_2)$ at zero path difference (or time delay) to a minimum value of $I_1(\bar{\nu}_1) - I_2(\bar{\nu}_2)$. 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 $$\frac{V_{\min}}{V_{\max}} = \frac{I_1 - I_2}{I_1 + I_2}$$

which leads to: $$\frac{I_1}{I_2} = \frac{1 + V_{\min}/V_{\max}}{1 - V_{\min}/V_{\max}} \qquad (6.15)$$

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

6.5.1 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 $\Delta\bar{\nu}_L$ across the width of the spectrum of the source.

$$\Delta\bar{\nu}_L \approx \frac{1}{x_o} \tag{6.16}$$

$\Delta\bar{\nu}_L$ 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.]

7. The Fabry-Perot interferometer

This instrument uses multiple beam interference by division of amplitude. Figure 7.1 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 4.2.2 (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.

7.1 The Fabry-Perot interference pattern

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

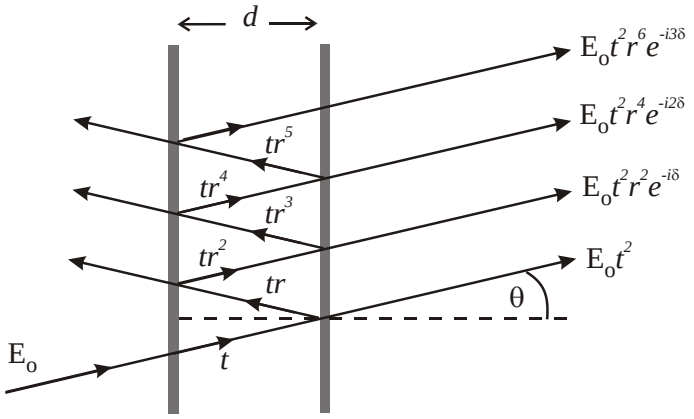


Figure 7.1 Multiple beam interference of beams reflected and transmitted by parallel surfaces with amplitude reflection and transmission coefficients r_i, t_i respectively.

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:

$$\delta = \frac{2\pi}{\lambda} 2d\cos\theta \qquad (7.1)$$

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

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

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:

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

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

$$I_t = I_0\frac{T^2}{(1-R)^2}\left[\frac{1}{1+\frac{4R}{(1-R)^2}\sin^2\delta/2}\right] \tag{7.2}$$

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

$$\frac{4R}{(1-R)^2} = \Phi \tag{7.3}$$

$$I_t = I_0\left[\frac{1}{1+\Phi\sin^2\delta/2}\right] \tag{7.4}$$

This is known as the Airy Function. See figure 7.2

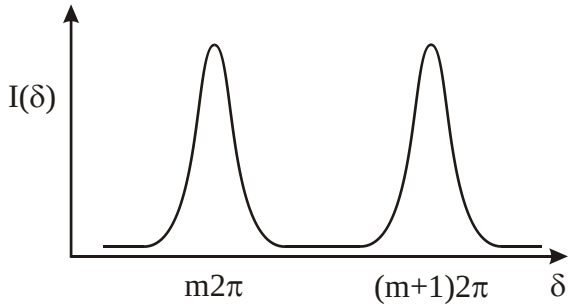


Figure 7.2 The Airy function showing fringes of order m, m+1 as function of δ .

7.2 Observing Fabry-Perot fringes

The Airy function describes the shape of the interference fringes. Figure 7.2 shows the intensity as a function of phase shift δ. The fringes occur each time δ is a multiple of 2π.

$$\delta = \frac{2\pi}{\lambda}2d\cos\theta = m2\pi \tag{7.5}$$

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 7.3. 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.

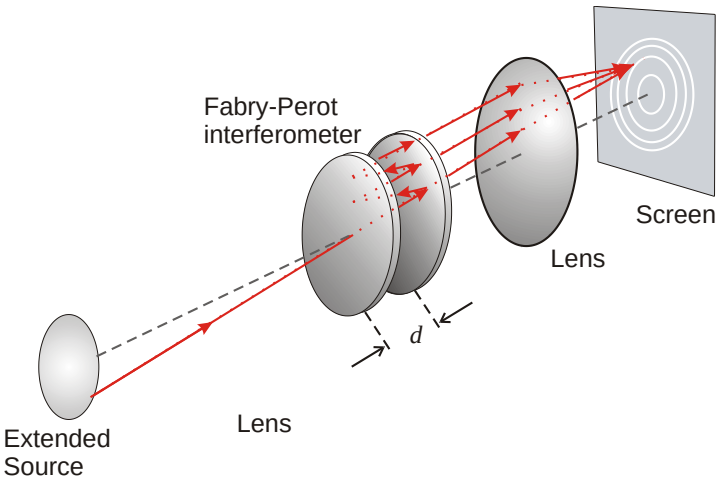


Figure 7.3. Schematic diagram of arrangement to view Fabry-Perot fringes. Parallel light from the Fabry-Perot is focussed on the screen.

From equation (7.5) the m^{th} fringe is at an angle θ_m

$$\cos\theta_m = \frac{m\lambda}{2d} \tag{7.6}$$

The angular separation of the m^{th} and $(m+1)^{th}$ fringe is $\Delta\theta_m$ is small so $\theta_m \approx \theta_{m+1} = \theta$
For small angles $\cos\theta_{m+1} - \cos\theta_m$ leads to:

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

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

$$\rho(\lambda) = f\theta_m = f\cos^{-1}\left(\frac{m\lambda}{2d}\right) \tag{7.7}$$

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 $\bar{\nu}$, $(1/\lambda)$. This also has the advantage that all the available light is put into the detected fringe on axis.

7.3 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*

$$F=\frac{2\pi}{\Delta\delta}=\frac{\pi\sqrt{\Phi}}{2} \qquad (7.8)$$

or

$$F=\frac{\pi\sqrt{R}}{(1-R)} \qquad (7.9)$$

So the sharpness of the fringes is determined by the reflectivity of the mirror surfaces. [Note: $F\sim\frac{3}{(1-R)}$. checks that the quadratic equation for *R* has been solved correctly!]

7.4 The Instrument width

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

$$\Delta\delta_{Inst}=\frac{2\pi}{F} \qquad (7.10)$$

$\Delta\bar{\nu}_{Inst}$ 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 :

$$\Delta\bar{\nu}_{Inst}=\frac{1}{2dF} \qquad (7.11)$$

7.5 Free Spectral Range, FSR

Figure 7.4 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 $\bar{\nu}$.

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

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

$$\Delta\bar{\nu}_{FSR} = \frac{1}{2d} \tag{7.12}$$

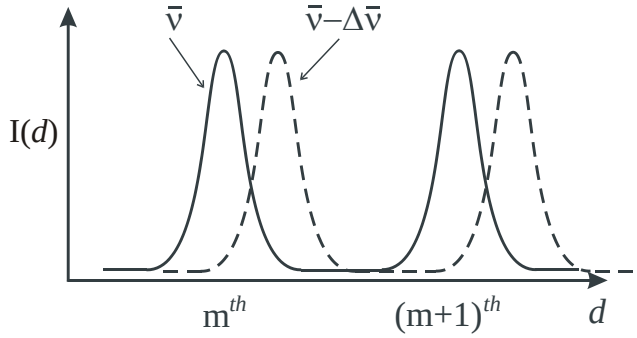


Figure 7.4 Fabry-Perot fringes for wavenumber $\bar{\nu}$ and $(\bar{\nu}-\Delta\bar{\nu})$ observed in centre-spot scanning mode. The mth -order fringe of $\bar{\nu}$ and $(\bar{\nu}-\Delta\bar{\nu})$ appear at a slightly different values of the interferometer spacing d. When the wavenumber difference $\Delta\bar{\nu}$ increases so that the m^{th} 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 7.4 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 (7.5). In figure 7.3 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.

7.6 Resolving Power

The instrumental width may now be expressed as:

$$\Delta\bar{\nu}_{Inst} = \frac{\Delta\bar{\nu}_{FSR}}{F} = \frac{1}{2dF} \tag{7.13}$$

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

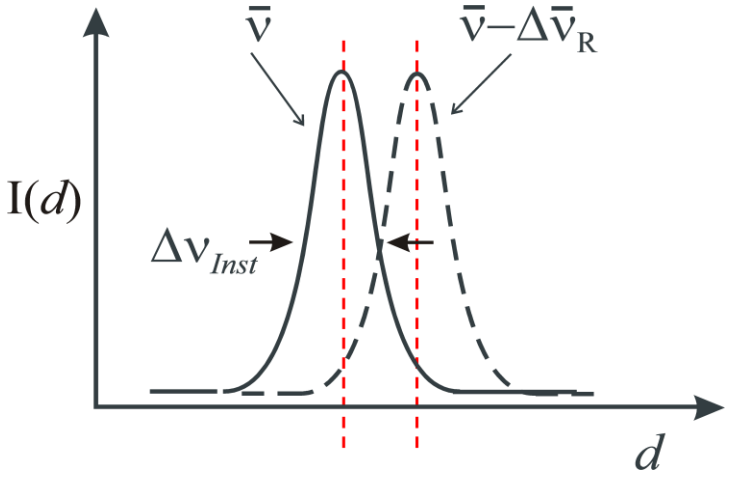


Figure 7.5 Resolution criterion: light of two wavenumbers $\bar{\nu}$, $\bar{\nu} - \Delta\bar{\nu}_R$ ***is resolved when the separation of fringes for*** $\bar{\nu}$ ***and*** $\bar{\nu} - \Delta\bar{\nu}_R$ ***is equal to the instrument width*** $\Delta\bar{\nu}_{Inst}$.

As in Figure 7.4 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

$$R.P. = \frac{\bar{\nu}}{\Delta\bar{\nu}_R} = mF \qquad (7.14)$$

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.

7.7 Practical matters

7.7.1 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

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

Hence

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

The required reflectivity *R* is then found from

$$F = \frac{\pi\sqrt{R}}{(1-R)} \tag{7.15}$$

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

This sets the maximum radius of the pinhole to be used.

7.7.3 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 not worthwhile making the reflectivity finesse > $x/2$.

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

$$R+T+A=1$$

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.

Instrument function and instrument width

The **instrument function** is the "mathematical" function that describes the shape of the spectrum produced in response to a delta-function or monochromatic input. For example, in the case of a diffraction grating spectrometer this is a function made up from a ratio of sine functions: $\sin^2(N\delta/2)/\sin^2(\delta/2)$. This function describes the "shape" of the interference pattern i.e. a spectral line.

The **instrument width** is the "width" of this function defined in some (arbitrary!) way e.g. in the case of the diffraction grating we choose the separation of the minima on either side of the peak of each order. The pattern produced by a diffraction grating spectrograph consists of lines, sometimes known as *spectral lines*, that are simply images of the entrance slit formed by light of different wavelengths.

In the case of the Michelson interferometer the instrument does not produce a "spectral line" directly. Rather the spectral line has to be calculated by finding the Fourier Transform (F.T.) of the interferogram. For a monochromatic input wave the interferogram would be a sine wave going on to infinity! The F.T. and hence the instrument function would then be a delta function. In practice however there is a limit to how long the interferogram can be - set by the maximum displacement of the mirror. So the interferogram is a sine wave of finite length. The instrument function would then be the F.T. of this finite sine wave, - broader than a delta function and having some finite width. This function however is rarely used. Instead we usually focus on the effective spectral width that the instrument produces for a monochromatic wave - "the instrument width" - and this is found from the inverse of the maximum path difference - in units of reciprocal metres.

So because different instruments produce different shapes of "spectral lines" from their interference patterns their instrument functions are different - so it's hard to compare them. Instead we usually refer to the instrument width of each type of device - Grating, Michleson or Fabry-Perot etc... as this allows a more practical way of comparing their usefulness for resolving spectral lines.

8. Reflection at dielectric surfaces and boundaries

8.1 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:

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

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.

The constants μ_r and ε_r characterise the response of the medium to the incident electric field. In general they are represented by complex tensors. This is because they affect the amplitude and phase of the light wave in the medium and the tensor nature reflects their dependence on the vector nature of the fields and the symmetry of the medium in which they propagate. For the time being we will consider only *linear isotropic homogeneous* (LIH) media and use scalar quantities for the relative permeability μ_r and permittivity ε_r. In these cases the refractive index is then also a scalar quantity.

8.1.1 Reflection and transmission at normal incidence

When a wave is incident on a boundary between two different media the change in electrical and magnetic response amounts to a change in impedance to the travelling wave. At such impedance boundaries some of the wave energy is reflected and the rest transmitted. We consider first *normal* incidence i.e. the angle between the propagation direction $\underline{k}$ and the normal to the surface separating the two media is 0^o. In this case there is symmetry with respect to the axis of propagation: the response will be the same no matter what the direction of the electric field $\underline{E}$ (the magnetic field $\underline{H}$ is orthogonal to both $\underline{E}$ and $\underline{k}$.) Thus the vector nature of the fields may be ignored in this case. When, however, the wave is incident obliquely at a boundary, the symmetry is broken and the direction of the $\underline{E}$ and $\underline{H}$ fields must be considered i.e. we must take account of the vector nature of the fields and we will deal with this case later.

At any boundary the fields must satisfy *boundary conditions* that are determined by conservation laws for the electro-magnetic fields. The case of normal incidence at the interface of two media of different refractive index n_1 and n_2 is illustrated in figure 8.1.

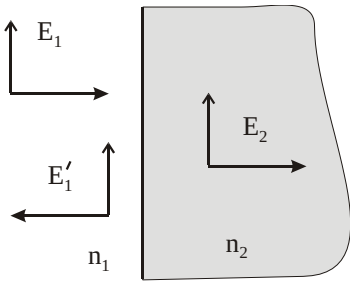


Figure 8.1 Reflection of an electromagnetic wave incident normally from medium of refractive index n_1 on a medium of index n_2

Boundary conditions demand that the perpendicular component of *D* is continuous and the tangential components of *E* and *H* are continuous. Incident and reflected electric field amplitudes are E_1 and $E_1^{'}$ respectively. Considering the boundary conditions for both *E* and *H* allows us to calculate the ratio of the incident and reflected amplitudes:

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

The intensity reflection coefficient is therefore:

$$R = \left(\frac{n_2 - n_1}{n_2 + n_1} \right)^2 \qquad (8.1)$$

For an air/glass interface ($n_2 = 1.5$), $R \sim 4\%$.

8.2 Reflection properties of dielectric layers.

The reflection at a dielectric boundary may be thought of as arising from an impedance *mis-match*. The larger the mis-match the larger the reflected proportion of the incident energy. The possibility arises of using two boundaries with intermediate mis-matches to generate two reflected waves. If the mis-matches can be arranged to give approximately equal amplitude reflected waves it may be possible to arrange their relative phases to produce destructive interference for the backward travelling waves i.e. an anti-reflection device.

8.2.1 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 8.2. 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. The wave vectors are k_i.

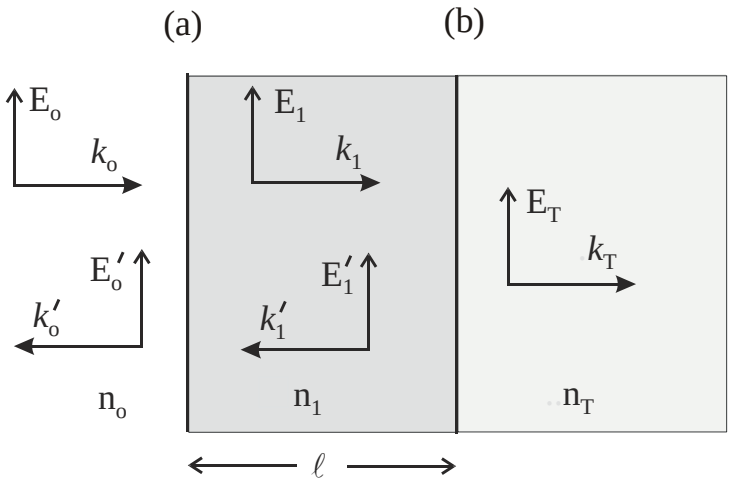


Figure 8.2 Reflected and transmitted waves for a wave incident normally from medium of index n_o on a dielectric layer of thickness ℓ, index n_1 on a substrate of index n_T.

At boundary (a)

$$E_o + E_o^{'} = E_1 + E_1^{'} \tag{8.2}$$

$$H_o - H_o^{'} = H_1 - H_1^{'} \tag{8.3}$$

using $H = n^2\sqrt{\frac{\varepsilon_o}{\mu_o}}E$:

$$n_o(E_o - E_o^{'}) = n_1(E_1 - E_1^{'}) \tag{8.4}$$

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

$$E_1 e^{ik_1\ell} + E_1^{'} e^{-ik_1\ell} = E_T \tag{8.5}$$

$$n_1(E_1 e^{ik_1\ell} - E_1^{'} e^{-ik_1\ell}) = n_T E_T \tag{8.6}$$

Eliminating E_1 and $E_1^{'}$ from (8.2), (8.4), (8.5) and (8.6):

$$E_o + E_o^{'} = \left[\cos k_1\ell - i\left(\frac{n_T}{n_1}\right)\sin k_1\ell\right]E_T \tag{8.7}$$

$$n_o(E_o - E_o^{'}) = \left[-in_1 \sin k_1\ell + n_T \cos k_1\ell\right]E_T \tag{8.8}$$

Writing:

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

$$C = -in_1 \sin k_1\ell \qquad D = \cos k_1\ell \tag{8.9}$$

We find:

$$\frac{E_o^{'}}{E_o} = r = \frac{An_o + Bn_o n_T - C - Dn_T}{An_o + Bn_o n_T + C + Dn_T} \tag{8.10}$$

and

$$\frac{E_T}{E_o} = t = \frac{2n_o}{An_o + Bn_o n_T + C + Dn_T} \tag{8.11}$$

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

$$R = |r|^2 = \left|\frac{n_o n_T - n_1^2}{n_o n_T + n_1^2}\right|^2 \tag{8.12}$$

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

$$R = |r|^2 = \left|\frac{n_o - n_T}{n_o + n_T}\right|^2 \tag{8.13}$$

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 (8.12). 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.

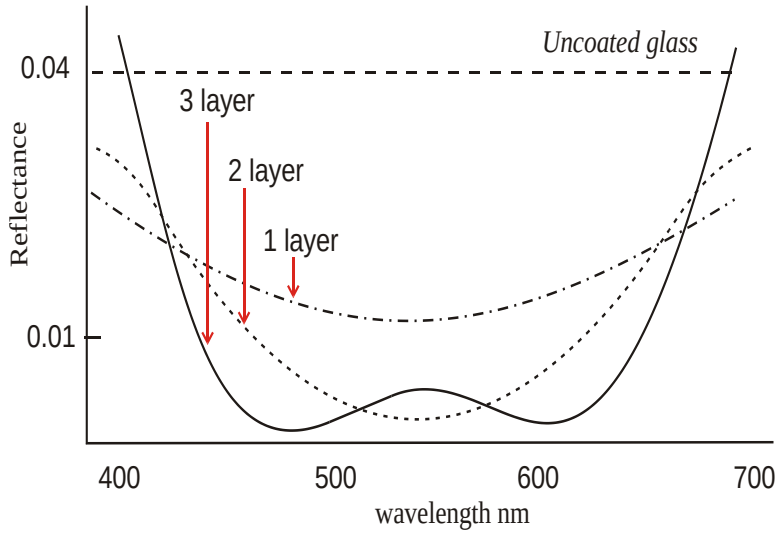


Figure 8.3 Anti-reflection dielectric coatings. A single $\lambda/4$ layer can reduce reflection from 4% to ~1%. Further reduction at specific wavelength regions is achieved by additional layers at the expense of increased reflectivity elsewhere. This enhanced reflection at the blue and red end of the visible is responsible for the purple-ish hue or blooming on camera or spectacle lenses.

8.2.2 Multiple dielectric layers: matrix method.

Write equations (8.7) and (8.8) 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:

$$\begin{pmatrix}1\\ n_o\end{pmatrix}+\begin{pmatrix}1\\ -n_o\end{pmatrix}r=\begin{pmatrix}A & B\\ C & D\end{pmatrix}\begin{pmatrix}1\\ n_T\end{pmatrix}t \tag{8.14}$$

The characteristic matrix is

$$M=\begin{pmatrix}A & B\\ C & D\end{pmatrix} \tag{8.15}$$

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

$$M_m=\begin{pmatrix}0 & -i/n_m\\ -in_m & 0\end{pmatrix} \tag{8.16}$$

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

$$M_{Stack}=M_1M_2M_3.....M_N \tag{8.17}$$

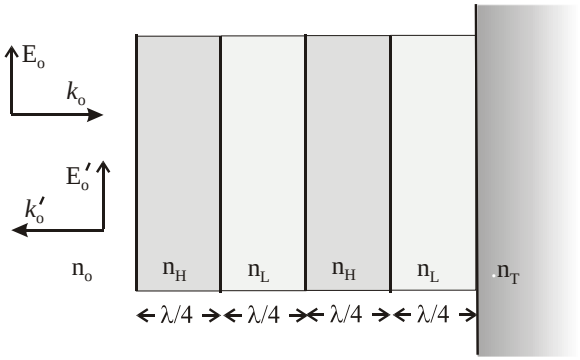


Figure 8.4 Multiple quarter-wave stack

8.2.3 High reflectance mirrors

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

$$M_{HL}=\begin{pmatrix}-n_L/n_H & 0\\ 0 & -n_H/n_L\end{pmatrix} \tag{8.18}$$

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 (8.10) we find the reflectivity of the composite stack:

$$R_{Stack} = \left\{ \frac{1 - \frac{n_T}{n_o}\left(\frac{n_H}{n_L}\right)^{2N}}{1 + \frac{n_T}{n_o}\left(\frac{n_H}{n_L}\right)^{2N}} \right\}^2 \qquad (8.19)$$

8.2.4 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 Free Spectral Range, 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.

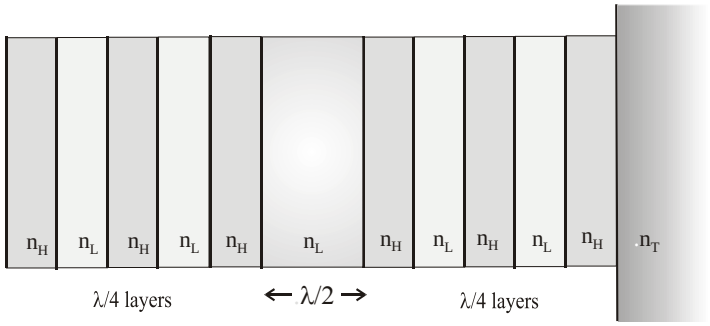


Figure 8.5 Interference filter constructed using multiple dielectric layers consisting of two high-reflectance stacks separated by a $\lambda/2$ layer which acts as a spacer in the Fabry-Perot type interference device. The spacer may be made in integer multiples of $\lambda/2$ to alter the FSR.

8.3 Reflection and transmission at oblique incidence

We now consider the more general case of a light wave striking a dielectric boundary at an angle of incidence θ that is not 90^o. We define the plane of incidence as that plane containing the wave propagation vector $\underline{k}$ and the normal to the dielectric boundary. This is illustrated in figure 8.6. The incident wave has the form: $E_0 e^{i\omega(t-nr/c)}$, where *r* is the distance along the propagation axis *OP*. In general the electric field vector $\underline{E}$ can lie at any angle around the wave vector $\underline{k}$. However we can always resolve this vector into two components: one lying in (i.e. parallel to) the the plane of incidence, E_P, and one perpendicular to this plane, E_S. [A light wave with its E-vector parallel to the plane of incidence is known as p-polarized light. When the E-vector is perpendicular to this plane it is s-polarized, from the German *senkrecht* meaning perpendicular.] The H-vectors will be perpendicular to the respective E-vectors in each case.

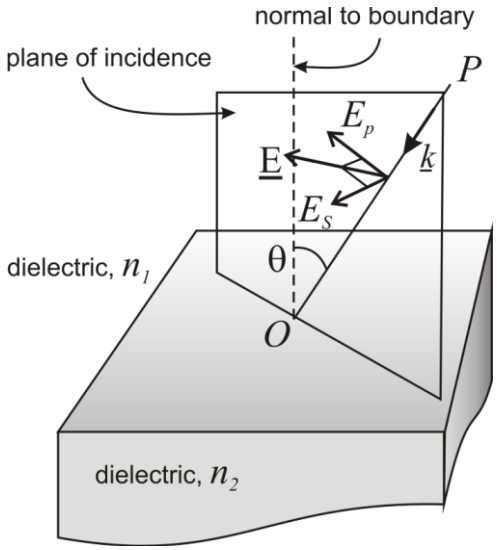


Figure 8.6 Wave with electric field E incident at oblique angle of incidence θ from dielectric with refractive index n_1 to dielectric with index n_2. The electric field $\underline{E}$ has a component E_P parallel to the plane of incidence and E_S perpendicular to the plane of incidence.

8.3.1 Reflection and transmission of p-polarized light

We consider first the case of p-polarized light. This is represented in figure 8.7 in which the plane of incidence (the *x*-*z* plane) is the plane of the paper. The incident, reflected and transmitted E-fields all lie in this plane and are denoted E_1^P, $E_1^{\vec{P}}$ and E_2^P respectively, lying respectively at angles θ_1, θ_2 and ϕ to the normal or z-axis.

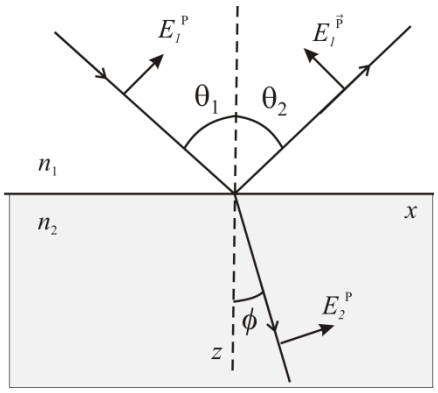


Figure 8.7

The boundary conditions require:

$$E_1^P \cos\theta_1 e^{i\omega(t-n_1 x\sin\theta_1/c)} - E_1^{\vec{P}} \cos\theta_2 e^{i\omega(t-n_1 x\sin\theta_2/c)} = E_2^P \cos\phi\, e^{i\omega(t-n_2 x\sin\phi/c)} \qquad (8.20)$$

and the H-field components out of the paper are:

$$H_1^P + H_1^{\vec{P}} = H_2^P \qquad (8.21)$$

Since these conditions are obeyed at all times the exponential terms must be identical. Hence,

$$\theta_1 = \theta_2 = \theta$$
$$n_1 \sin\theta = n_2 \sin\phi\,, \qquad (8.22)$$

i.e this is Snell's law for reflection and refraction.

From section 8.1 we can write $H = nE$ and the boundary conditions then become,

$$(E_1^P - E_1^{\vec{P}})\cos\theta = E_2^P \cos\phi$$
$$n_1(E_1^P + E_1^{\vec{P}}) = n_2 E_2^P$$

From this we find, using Snell's law (8.22),

$$\frac{E_2^P}{E_1^P} = \frac{2\cos\theta\sin\phi}{\sin\theta\cos\theta + \sin\phi\cos\phi} = \frac{2\cos\theta\sin\phi}{\sin(\theta+\phi) + \cos(\theta-\phi)} \qquad (8.23)$$

$$\frac{E_1^{\vec{P}}}{E_1^P} = \frac{\sin\theta\cos\theta - \sin\phi\cos\phi}{\sin\theta\cos\theta + \sin\phi\cos\phi} = \frac{\tan(\theta-\phi)}{\tan(\theta+\phi)} \qquad (8.24)$$

8.3.2 Reflection and transmission of s-polarized light

When the light is polarized such that the E-vector is perpendicular to the plane of incidence the H-fields will lie in this plane as shown in figure 8.8.

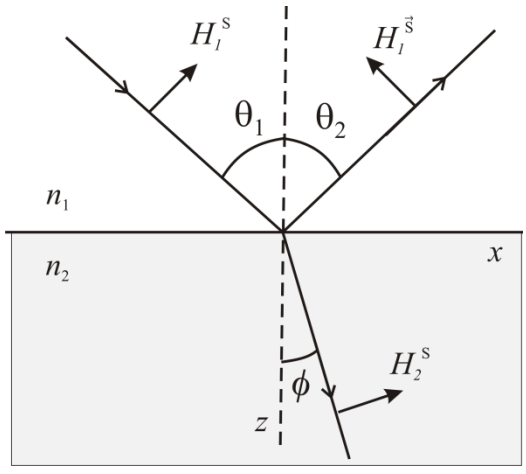


Figure 8.8. The E-vectors are out of the plane of the figure

Following the same procedure, using the boundary conditions on E and H we find the following relationships between the incident, reflected and transmitted amplitudes,

$$\frac{E_2^S}{E_1^S} = \frac{2\cos\theta\sin\phi}{\cos\theta\sin\phi + \sin\theta\cos\phi} = \frac{2\cos\theta\sin\phi}{\sin(\theta+\phi)} \qquad (8.25)$$

$$\frac{E_2^{\vec{S}}}{E_1^S} = \frac{\cos\theta\sin\phi - \sin\theta\cos\phi}{\cos\theta\sin\phi + \sin\theta\cos\phi} = -\frac{\sin(\theta-\phi)}{\sin(\theta+\phi)} \qquad (8.26)$$

Equations (8.23), (8.24), (8.25) and (8.26), are the Fresnel Equations. [*non-examinable*] They allow us to predict the reflection and transmission coefficients for light of various polarizations incident on a dielectric surface i.e. a boundary between two different dielectric media such as air and glass.

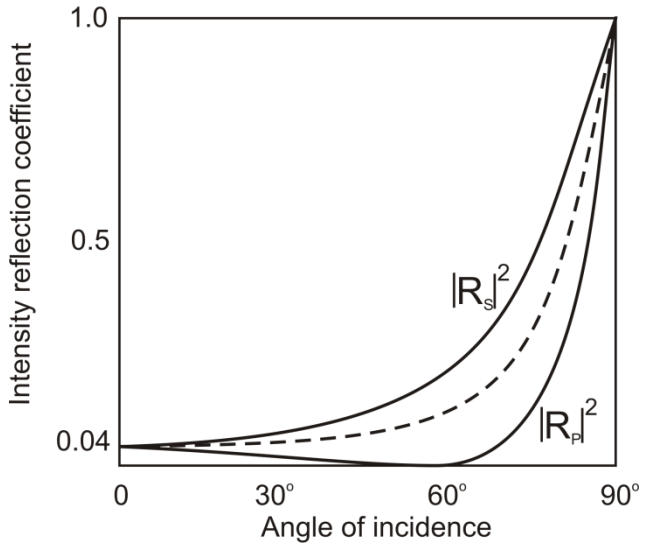


Figure 8.9 Intensity reflection coefficients for s-polarized and p-polarized light as function of incidence angle from air to glass. The dashed line is the average and represents the behaviour for unpolarized light.

The Fresnel equations show that the reflection coefficients for both s- and p-polarized light vary with angle of incidence as shown in figure 8.9

8.4 Deductions from Fresnel's equations

8.4.1 Brewsters' Angle

The amplitude reflection coefficient, *r*, is given for p-polarized light by equation (8.24),

$$\frac{E_1^{\vec{P}}}{E_1^P} = r = \frac{\tan(\theta-\phi)}{\tan(\theta+\phi)} \qquad (8.24)$$

We can see that as $(\theta+\phi)$ approaches π/2, *r* will tend to zero [$\tan(\theta+\phi)\to\infty$]. In the limit $(\theta+\phi)=\pi/2$, then $\cos\theta=\sin\phi$. The angle of incidence at which this condition is satisfied is known as Brewster's angle θ_B and, from (8.22),

$$\sin\theta_B=\frac{n_2}{n_1}\sin\phi=\frac{n_2}{n_1}\cos\theta_B$$

or

$$\tan\theta_B=\frac{n_2}{n_1} \qquad (8.27)$$

Thus for p-polarized light incident at angle θ_B there will be no reflected wave.

Another consequence of this is that unpolarized light, which consists, as we will see later, of waves with E-vectors varying randomly in all possible orientations, will become plane polarized with its E-vector perpendicular to the plane of incidence. This is because any E-vector may be composed of s- and p-polarizations and only the s-polarization will be reflected.
An important application of Brewster's angle is in enabling p-polarized light to suffer no reflection losses when passing through a glass window. This is very useful for transmitting laser light where it is important to minimize reflection losses at window surfaces.

8.4.2 Phase changes on reflection

We first consider the predictions of the Fresnel equations for normal incidence. For $\theta = 0$, using (8.22) we find,

$$\frac{E_2^P}{E_1^P}=\frac{2\cos\theta\sin\phi}{\sin\theta\cos\theta+\sin\phi\cos\phi}$$

$$=\frac{2\cos\theta(n_1/n_2)\sin\theta}{\sin\theta\cos\theta+\sin\phi\cos\phi}$$

$$=\frac{2(n_1/n_2)}{1+\dfrac{\sin\phi\cos\phi}{\sin\theta\cos\theta}} \qquad = \qquad \frac{2(n_1/n_2)}{1+\dfrac{n_1}{n_2}\dfrac{\cos\phi}{\cos\theta}}$$

Hence,

$$\frac{E_2^P}{E_1^P}=\frac{2n_1}{n_2+n_1} \qquad (8.23a)$$

Similarly we find,

$$\frac{E_1^{\vec{P}}}{E_1^P}=\frac{n_2-n_1}{n_2+n_1} \qquad (8.24a)$$

$$\frac{E_2^S}{E_1^S}=\frac{2n_1}{n_2+n_1} \qquad (8.25a)$$

$$\frac{E_2^{\vec{S}}}{E_1^S}=-\frac{n_2-n_1}{n_2+n_1} \qquad (8.26a)$$

There appears to be a discrepancy between the equations for the reflected light polarized in orthogonal planes, (8.24a) and (8.26a) since, at normal incidence, the two planes are indistinguishable, yet they have opposite signs! The sign discrepancy arises because we have taken the incident and reflected fields, E_1^P and $E_1^{\vec{P}}$, to be in the opposite directions (see figure 8.7). Assuming $n_2 > n_1$ and since the ratio $E_1^P/E_1^{\vec{P}}$ is positive, our theory predicts that there is a phase shift of π in the reflected E-vector relative to the incident wave. Note that as the incident and reflected H-vectors are in the same direction there is no phase shift of the H-field. In the case of the s-polarized E-vectors in figure 8.8, they are both perpendicular to the plane of the paper but in opposite directions i.e. corresponding to a π-phase shift.
In the case $n_2 < n_1$ there is no phase shift in E but a π-phase shift occurs in the H-vector. In both cases there is no phase shift in the transmitted wave.

8.4.3 Total (internal) reflection and evanescent waves

When a wave is incident from a medium of index n_2 obliquely on a less dense medium, index n_1, as shown in figure 8.10, we know that if the angle of incidence exceeds a critical value θ_{crit}, given by $\theta_{crit} = \sin^{-1}(n_1/n_2)$, then the wave is totally reflected. In this case the angle ϕ in the less dense medium is an imaginary quantity, leading to phase shifts that lie between 0 and π and are different for E and its corresponding H. The analysis is tedious and off-syllabus (thankfully!) but leads to the following predictions,

1. *The amplitude of the reflected beam and incident beams are equal – total reflection.*

2. *There is a transmitted beam with the following characteristics:*
   (a) its wave velocity is $v_1/\sin\phi$
   (b) it travels parallel to the interface
   (c) its amplitude decays exponentially with distance perpendicular to the surface
   (d) it is neither a plane nor a transverse wave – it has a component along the surface
   (e) its Poynting vector is zero.

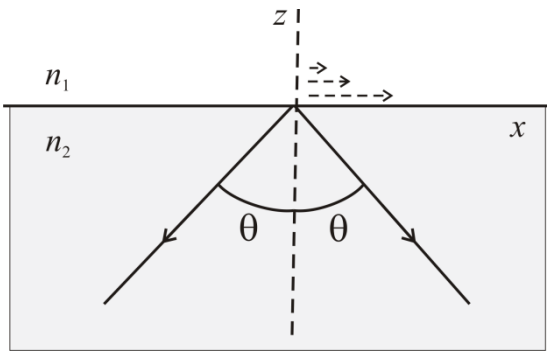


Figure 8.10 Internal reflection, the evanescent wave is shown as dotted lines

This, slightly mysterious, transmitted wave is known as an ***evanescent wave***. Its presence can be detected by bringing another dielectric, say of index also n_2, close to the surface as shown in figure 8.11.

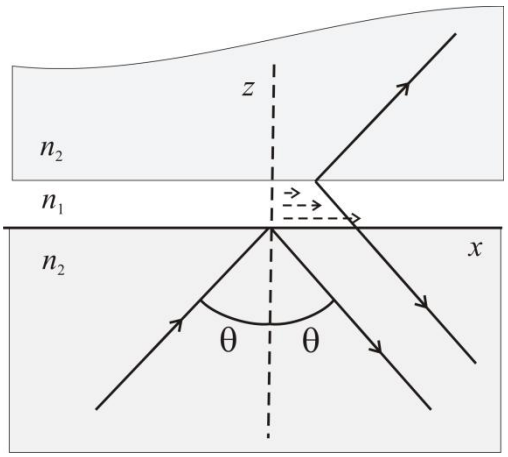


Figure 8.10 Frustrated internal reflection

As the gap between the two dielectrics of index n_2 gets smaller (of the order of a few wavelengths) a transmitted wave appears simultaneously with a weakening of the reflected wave. This effect arises because there is a reflection from the second n_1 /n_2 boundary which destructively interferes with the reflected wave from the first boundary. This effect – optical tunnelling – is the wave equivalent of quantum mechanical tunnelling of particles through a potential barrier. The effect, known as frustrated total internal reflection, finds application in a type of microscopy in which the surface layers of biological samples are excited by the evanescent waves. Since these penetrate only a short distance, of the order of 100 nm, fluorescence is generated by only the molecules in this thin layer. The fluorescence image is then no longer swamped by fluorescence from the bulk of the sample that would otherwise overwhelm the signal from the surface layer.

9. 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** in the case of polarized light:

(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*.

9.1 Polarization states

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

$$\begin{aligned} E_y &= E_{oy}\cos(kx-\omega t)\underline{\mathbf{j}} \\ E_z &= E_{oz}\cos(kx-\omega t+\delta)\underline{\mathbf{k}} \end{aligned} \qquad (9.1)$$

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

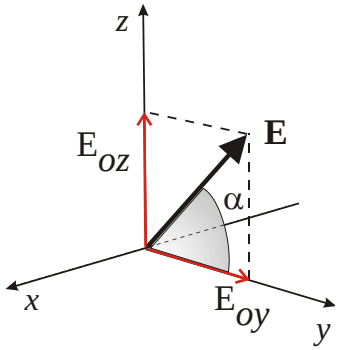


Figure 9.1 Electric field vector in light wave has components E_{oz} and E_{oy} in plane orthogonal to propagation direction along x-axis

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

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

$$\mathbf{E}_P = \{E_{oy}\underline{\mathbf{j}} + E_{oz}\underline{\mathbf{k}}\}\cos(kx-\omega t) \qquad (9.2)$$

at a fixed angle α to the *y*-axis

$$\tan\alpha = \frac{E_{oz}}{E_{oy}} \qquad (9.3)$$

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

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

$$E_y = E_o \cos(kx - \omega t)\underline{\mathbf{j}}$$
$$E_z = E_o \sin(kx - \omega t)\underline{\mathbf{k}} \tag{9.4}$$

$$\tan\alpha = \frac{\sin(kx-\omega t)}{\cos(kx-\omega t)} = \tan(kx-\omega t) \tag{9.5}$$

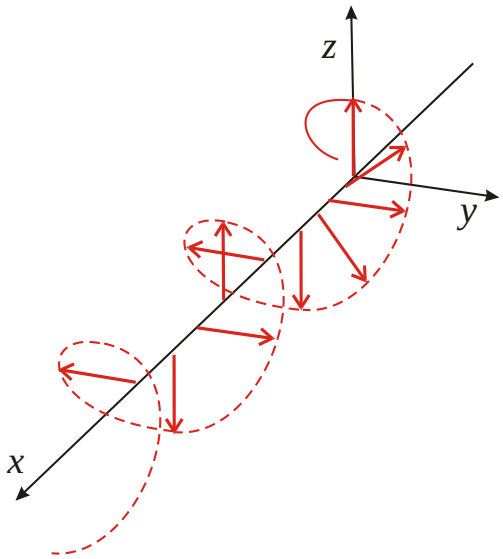


Figure 9.2 Right circularly polarized light propagating in the positive x-direction.

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 $t = kx_o/\omega$:

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

As viewed back towards the source the **E**-vector has rotated clockwise. See figure 9.3. This is Right Circularly Polarized light ($\delta = -\pi/2$). *Right* circularly polarized light advances like a *Left*-handed screw!

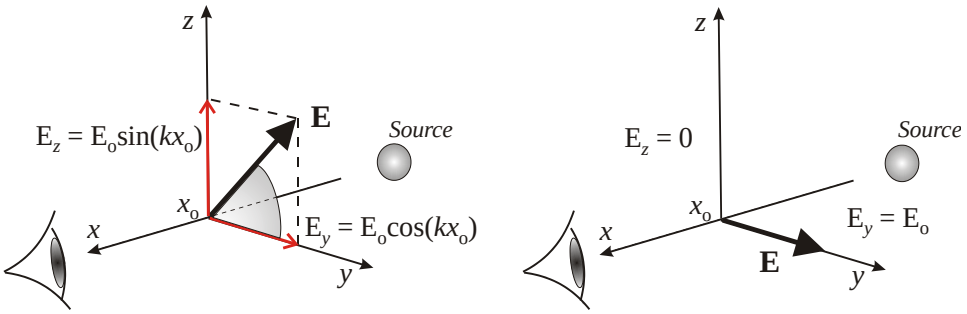


Figure 9.3 Direction of circular polarization is determined by looking back towards the source. (a) and (b) show E-vector at a point $x = x_o$ ***at time t = 0, and a later time*** $t = kx_o / \omega$. ***In this case the E-vector has rotated clockwise and is denoted Right Circularly Polarized.***

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:

$$\begin{aligned} \mathbf{E}_R &= E_o\left[\cos(kx-\omega t)\underline{\mathbf{j}} + \sin(kx-\omega t)\underline{\mathbf{k}}\right] \\ \mathbf{E}_L &= E_o\left[\cos(kx-\omega t)\underline{\mathbf{j}} - \sin(kx-\omega t)\underline{\mathbf{k}}\right] \end{aligned} \tag{9.6}$$

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

$$\mathbf{E}_P = \mathbf{E}_R + \mathbf{E}_L = 2E_o \cos(kx-\omega t)\underline{\mathbf{j}} \tag{9.7}$$

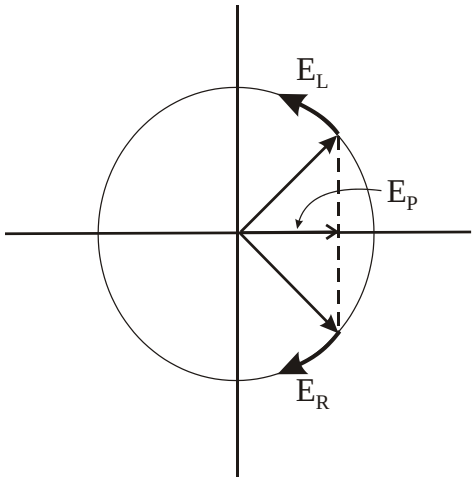


Figure 9.4 Plane polarized light is a superposition of a right- and a left-circularly polarized component.

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

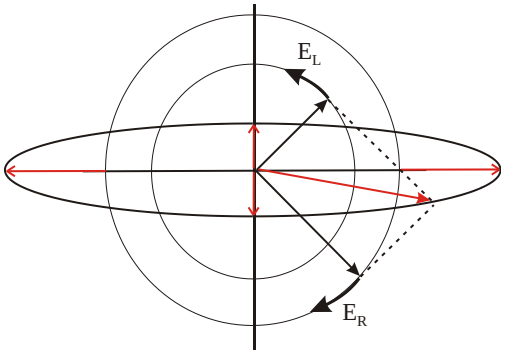


Figure 9.5 A superposition of right- and left-circularly polarized components of unequal magnitude gives elliptically polarized light.

Case 3: Elliptically polarized light.

In general there is a relative phase δ between *y* and *z* components. From (9.1):

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

Writing

$$E_z = E_{oz}[\cos(kx-\omega t)\cos\delta - \sin(kx-\omega t)\sin\delta] \tag{9.8}$$

Substitute in (9.8) 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:

$$\frac{E_y^2}{E_{oy}^2} + \frac{E_z^2}{E_{oz}^2} - 2\frac{E_y}{E_{oy}}\frac{E_z}{E_{oz}}\cos\delta = \sin^2\delta \tag{9.9}$$

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

$$\frac{E_y^2}{E_{oy}^2} + \frac{E_z^2}{E_{oz}^2} = 1 \tag{9.10}$$

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.

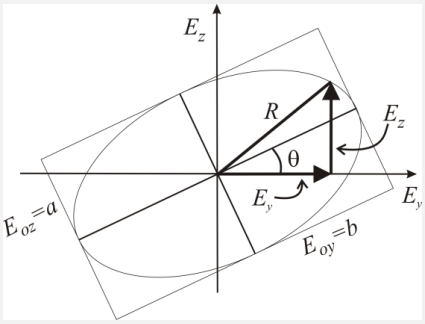
For $\delta \neq \pm\pi/2$ the axes of symmetry of the ellipse are rotated relative to the *y / z* axes by an angle θ [see box below for derivation]:

$$\tan 2\theta = 2\frac{E_{oy}E_{oz}}{E_{oy}^2 - E_{oz}^2}\cos\delta \tag{9.11}$$

Equation of ellipse describing E-vector for polarized light: $$\frac{E_z^2}{E_{oz}^2}+\frac{E_y^2}{E_{oy}^2}-\frac{2E_zE_y}{E_{oz}E_{oy}}\cos\delta-\sin^2\delta=0$$

Where δ is the phase shift of E_{oz} relative to E_{oy}.

Writing $E_{oz}=a,\ E_{oy}=b$ $$\frac{E_z^2}{a^2}+\frac{E_y^2}{b^2}-\frac{2E_zE_y}{ab}\cos\delta-\sin^2\delta=0 \qquad (1)$$



we can represent this ellipse graphically in the figure:

The ellipse is at an angle θ to the *z*-axis and $\tan\theta=\left|\frac{E_z}{E_y}\right|_{\text{maximum}}$ when the point *R* on the ellipse is at the major axis.

We find the the major axis by finding the maximum value of $E_y^2+E_z^2=R^2$. Differentiating this equation and setting the result equal to zero for maximum:

$$2E_ydE_y+2E_zdE_z=0 \qquad (2)$$

Differentiating (1) :

$$2\frac{E_z}{a^2}dE_z+2\frac{E_y}{b^2}dE_y-\left(\frac{2E_ydE_z+2E_zdE_y}{ab}\right)\cos\delta=0$$

$$\left(\frac{E_z}{a^2}-\frac{E_y\cos\delta}{ab}\right)dE_z+\left(\frac{E_y}{b^2}-\frac{E_z\cos\delta}{ab}\right)dE_y=0 \qquad (3)$$

Comparing coefficients in equations (2) and (3):

$$E_y=\frac{E_y}{b^2}-\frac{E_z\cos\delta}{ab};\quad \text{and} \qquad E_z=\frac{E_z}{a^2}-\frac{E_y\cos\delta}{ab}$$

$$1=\frac{1}{b^2}-\frac{\cos\delta}{ab}\frac{E_z}{E_y};\quad \text{and} \qquad 1=\frac{1}{a^2}-\frac{\cos\delta}{ab}\frac{E_y}{E_z}$$

Combining these last two equations:

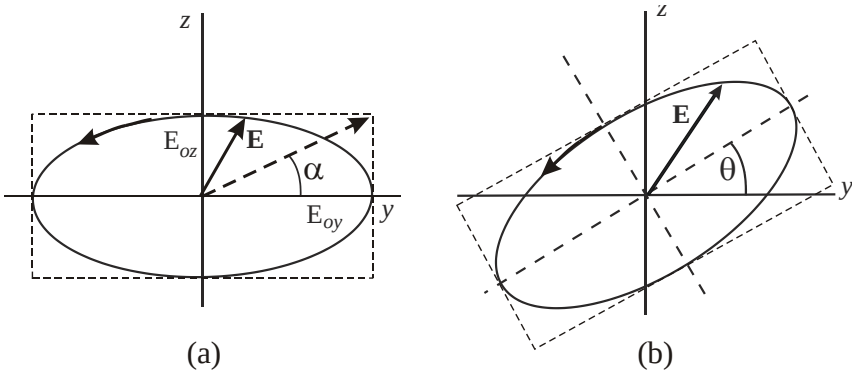
$$\frac{1}{a^2}-\frac{1}{b^2}=\left(\frac{E_y}{E_z}-\frac{E_z}{E_y}\right)\frac{\cos\delta}{ab}$$

Using $\tan\theta=\frac{E_z}{E_y}$: $$\frac{b^2-a^2}{a^2b^2}=\left(\tan\theta-\cot\theta\right)\frac{\cos\delta}{ab}$$

$$\frac{b^2-a^2}{ab}=\left(2\cot 2\theta\right)\cos\delta \quad \text{gives,} \qquad \tan 2\theta=\frac{2ab\cos\delta}{b^2-a^2}$$

Hence,

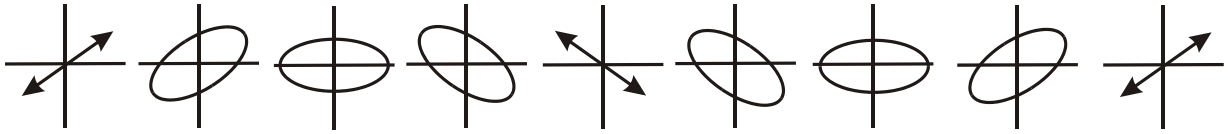
$$\tan 2\theta=\frac{2E_{oy}E_{oz}}{E_{oy}{}^2-E_{oz}{}^2}\cos\delta$$



(a) (b)

Figure 9.6 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.



$\delta = 0,$ $\pi/4$ $\pi/2$ $3\pi/4$ π $5\pi/4$ $3\pi/2$ $7\pi/4$ 2π

Figure 9.7 General elliptical state of polarization for different values of relative phase δ ***between the components.***

9.2 Optics of anisotropic media; birefringence.

Firstly; some background information that is not specifically on 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:

$$\begin{pmatrix} \mathbf{D}_x \\ \mathbf{D}_y \\ \mathbf{D}_z \end{pmatrix} = \varepsilon_o \begin{pmatrix} \varepsilon_X & 0 & 0 \\ 0 & \varepsilon_Y & 0 \\ 0 & 0 & \varepsilon_Z \end{pmatrix} \begin{pmatrix} \mathbf{E}_x \\ \mathbf{E}_y \\ \mathbf{E}_z \end{pmatrix} \qquad (9.12)$$

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.*

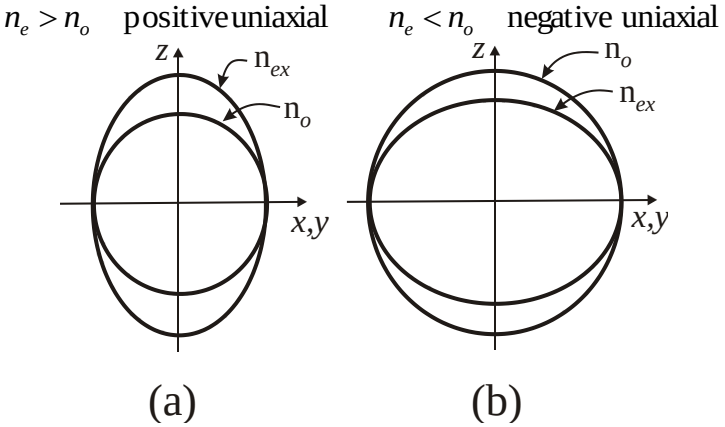


Figure 9.8 (a) positive and (b) negative uniaxial birefringent crystals.

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.

Note: Hecht section 8.4.2 Birefringent Crystals (3rd Ed. p 342, 2nd Ed. p 288), defines what he means by positive and negative uniaxial crystals.

$$\Delta n = (n_e - n_o)$$

This will be negative if $n_e < n_o$ and positive if $n_e > n_o$. This is exactly the same as the definition used here

Note that figure 9.8 of the notes is not showing the same thing as Hecht's figure 8.24 (8.28 in 2nd Ed). These notes show the "refractive index surfaces," Hecht is showing the *wavefront* of the ordinary and extra-ordinary waves.

In the paragraph above figure 9.8 we are talking about the direction of the E-vector. When the E-vector lies in the *x*,*y* plane it is an ordinary ray. The refractive index is given by the distance from the origin to the surface. For the *o*-ray this will be n_o and will be the same irrespective of the direction of the wave propagation.
If the E-vector lies along the z-axis the refractive index for this ray (the *e*-ray) is given by the distance from the origin to the intercept of the z-axis with the elliptical surface and will be n_e.
Note that if the E-vector lies at an angle to the z-axis the distance to the surface giving the refractive index is n_{ex} as shown in the figure 9.8(a). You can see that n_{ex} will take a value lying between n_e and n_o.

9.3 Production and manipulation of polarized light

At the end of section 9.1 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:

$$\delta = \frac{2\pi}{\lambda}|n_o - n_e|\ell \qquad (9.14)$$

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:	phase shift, δ:	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)

Angle of linearly polarized input	**Phase shift introduced by birefringent plate**	**Output polarization**
$\theta = 45^o$ ($E_y = E_z$)	$\delta = +/-\frac{\pi}{2}$ (Quarter-wave, $\lambda/4$ plate)	Left/Right Circular
$\theta \neq 45^o$ ($E_y \neq E_z$)	$\delta = +/-\frac{\pi}{2}$ (Quarter-wave, $\lambda/4$ plate)	Left/Right Elliptical
$\theta \neq 45^o$ ($E_y \neq E_z$)	$\delta = +/-\pi$ (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.

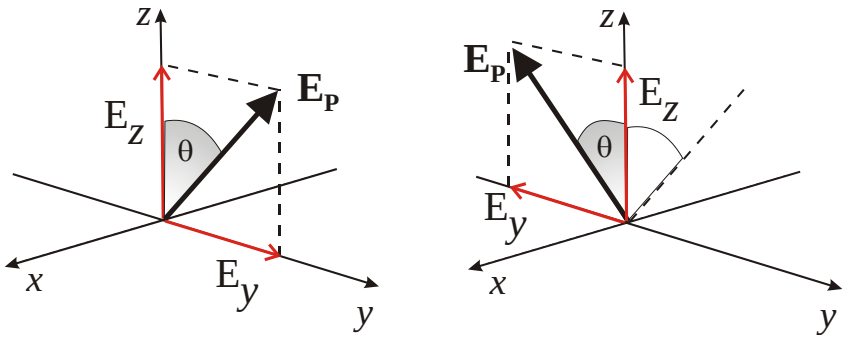


Figure 9.9 Action of a $\lambda/2$-plate with axis at θ to E-vector of plane polarized light shifts phase of one component; E_y by π relative to original phase resulting in a rotation by 2θ of the resultant E-vector.

Polarized light may be produced from unpolarized light using:

(a) Fresnel reflection at Brewster's angle.
(b) "Polaroid-type" material: crystals (tourmaline) in a plastic matrix absorb 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.

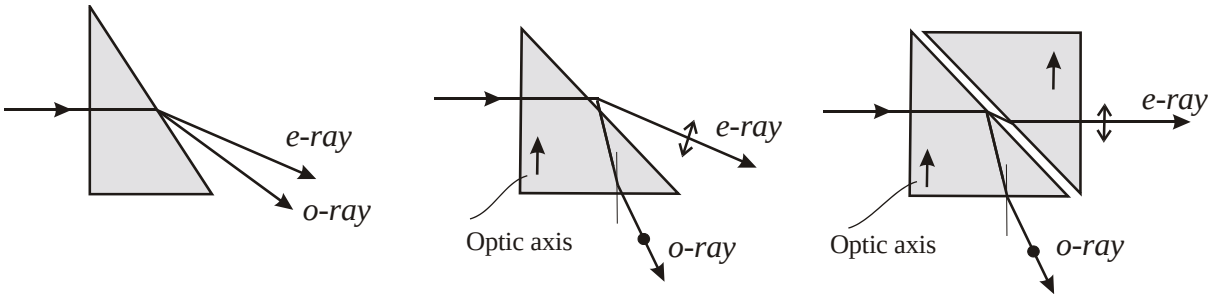


Figure 9.10 Prism polarizers

9.4 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 9.6) The following method may be used to specify the state by determination of these parameters.

(i) Pass the light through a linear polarizer. Rotate linear polarizer to determine approximately the orientation of the major/minor axes of ellipse – this will be the angle at which a maximum and minimum transmission is obtained.

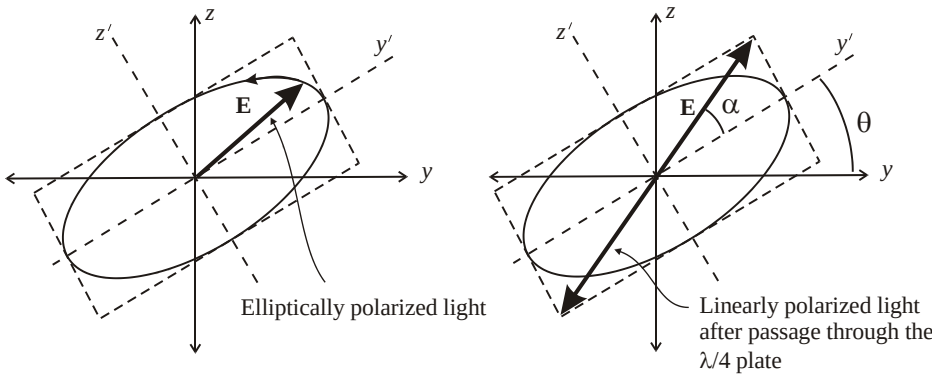
(ii) Set linear polarizer for maximum transmission. Remember that if the coordinate axes are chosen to coincide with the principal axes of the ellipse there is then a phase difference of $\delta = \pm\pi/2$ between the components.

(iii) Insert a $\lambda/4$-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. The $\lambda/4$-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, the components of the *E*-vector relative to the axes *y', z'* at this orientation have a relative phase, $\delta = \pm\pi/2$.



(a) (b)

Figure 9.11 (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 elliptical polarization.

9.5 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 uncorrelated randomly polarized waves from uncorrelated atoms still produce an interference pattern.

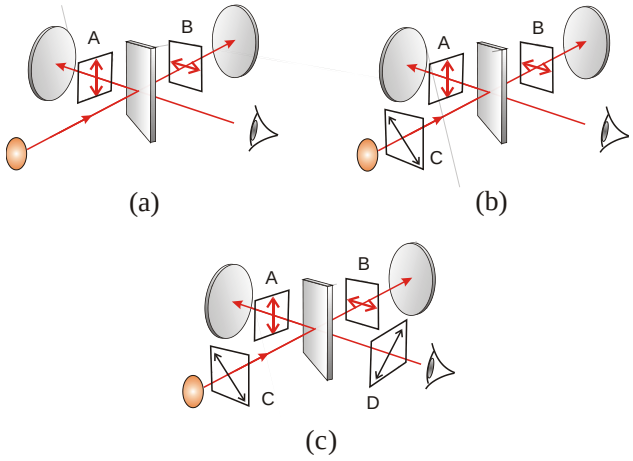


Figure 9.12 Interference of polarized light in a two-beam (Michelson) interferometer A and B are linear polarizers i.e. pass only light polarized in directions shown. Unpolarized light from the source is split at the beam splitter.
(a) Light in paths A and B are orthogonally polarized; no interference.
(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} transmits phase correlated components from polarizer C that are parallel and so interference is produced.

Note that unpolarized light cannot 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.