PHAS1224 Waves, Optics and Acoustics
Part II: Optics

R.S. Thorne

Preface

This is the second part of the PHAS1224 lecture course, and is of roughly equal length to the first part. The topics studied will be:

- Introduction. A brief description of what light is in terms of electric **E** and magnetic **B** fields, i.e. light is an electromagnetic wave. The propagation of light. Wavefronts and light rays. Fermat's theorem and Huygens's principle. An outline of the reflection, refraction and transmission of waves. Polarisation, both linear and circular.

- Basic interference of light waves via amplitude division, i.e. interference between waves reflected from different boundaries. Simple examples, e.g. films, Newton's rings.

- Devices using interference – interferometers. The Michelson interferometer and Fabry-Perot etalon. Use of these for resolution of light of different wavelengths.

- Interference using wavefront division – diffraction. Examples of double slits, multiple equally spaced slits, i.e. a diffraction grating, and a slit of finite width. Again, the use for resolution, in this case both of different wavelengths and of objects close together. The eye as an example in this context.

- Mirrors and lenses. Ray diagrams. Magnification and real and virtual images. The mirror and lens equations. Combination of lenses and mirrors. Optical instruments consisting of lenses and mirrors. The eye, microscopes and telescopes.

Books

The most suitable books for this part of the course are

Chapters 35-38 of "Physics for Scientists and Engineers with modern physics", by Jewett and Serway, Brooks Cole.

"Optics", by Hecht, Pearson Education publishers.

"An Introduction to Optics", by Pedrotti, Pedrotti and Pedrotti, Pearson Education publishers.

1 Introduction

Optics is the study of the behaviour of light. However, in order to understand this properly we need to underpin this with a basic understanding of what light is. In the classical way of thinking light is the part of the spectrum of electromagnetic radiation waves which our eyes have become attuned to detecting, i.e. "light" is not the same for all observers, like sound, although the variation is mainly between different types of organism. Hence, the whole of this course will essentially be about electromagnetic waves, but the only parts of it our eyes will "see" is light.

1.1 Electromagnetic Waves - *(Subsection non-examinable.)*

Electricity and magnetism are described by Maxwell's equations (1873). Both electric and magnetic fields are vector quantities generated usually by charges and currents respectively,

$$\mathbf{E} = \frac{Q\hat{\mathbf{r}}}{4\pi\epsilon_0 r^2}, \qquad \int \mathbf{B}\cdot d\mathbf{l} = \mu_0 I, \tag{1}$$

where the permittivity of free space $\epsilon_0 = 8.85\times 10^{-12}\mathrm{Fm}^{-1}$ and the permeability of free space $\mu_0 = 4\pi \times 10^{-7}\mathrm{Hm}^{-1}$. In a medium we have to make the replacements $\epsilon_0 \rightarrow \epsilon_r\epsilon_0$ and $\mu_0 \rightarrow \mu_r\mu_0$, where ϵ_r and μ_r are known as the relative permittivity and relative permeability respectively. For all but highly magnetic materials, which are not transparent to light, $\mu_r = 1$ to a good approximation. However, in many materials ϵ_r deviates from unity quite significantly because the electric dipoles set up in materials due to electrons and nuclei being pushed in opposite directions tend to oppose the external electric field. Hence, in the vast majority of cases $\epsilon_r > 1$, though for gases the departure is very small.

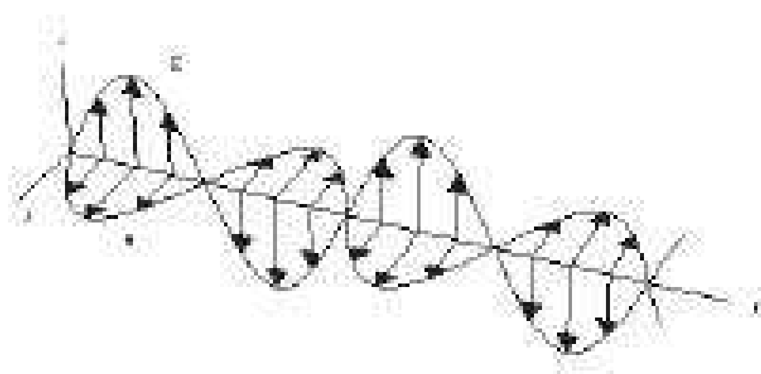

Figure 1: The electric and magnetic field and the direction of travel for an electromagnetic wave.

It is also the case that time-varying electric and magnetic fields generate fields, as in the electromagnetic induction discovered by Faraday. In particular, if we consider an electric field described by

$$\mathbf{E} = E_x\mathbf{i} = E_0 \exp(ikz - i\omega t)\mathbf{i}, \tag{2}$$

where $k = 2\pi/\lambda$ and $\omega = 2\pi f$, then Maxwell's equations say

$$\frac{\partial E_x}{\partial z} = -\frac{\partial B_y}{\partial t} \qquad \text{and} -\frac{\partial B_y}{\partial z} = \epsilon_0\mu_0\frac{\partial E_x}{\partial t}. \tag{3}$$

So a time-varying **E** field generates a z-dependent **B** field and also a time-varying **B** field generates a z-dependent **E** field. By taking the partial derivative of the first equation again with respect to z and the second with respect to t, and using commutativity of partial derivatives we find that

$$\frac{\partial^2 E_x}{\partial z^2} = \epsilon_0\mu_0\frac{\partial^2 E_x}{\partial t^2} \tag{4}$$

or alternatively taking the partial derivative of the first equation with respect to t and the second with respect to z and using the same argument we obtain

$$\frac{\partial^2 B_y}{\partial z^2} = \epsilon_0\mu_0\frac{\partial^2 B_y}{\partial t^2}. \tag{5}$$

But these are exactly the wave equation, i.e.

$$\frac{\partial^2 E_x}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 E_x}{\partial t^2}; \qquad\qquad \frac{\partial^2 B_y}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 B_y}{\partial t^2} \tag{6}$$

with

$$c = \frac{1}{\sqrt{\epsilon_0\mu_0}}. \tag{7}$$

So Maxwell's equations, which predate relativity by over 30 years, predict electromagnetic waves which travel with velocity $c = 3 \times 10^8 \text{ms}^{-1}$. In a medium c is replaced by the wave velocity

$$v = \frac{1}{\sqrt{\epsilon_r\epsilon_0\mu_r\mu_0}}. \tag{8}$$

A feature of the solutions is that the **E** and **B** fields are perpendicular to the direction of motion, i.e. the waves are transverse, and also perpendicular to each other. In fact the direction of motion of the wave is in the same direction as **E** × **B**. Slightly more precisely, if vector **S** represents the energy flux of the wave then

$$\mathbf{S} = \mathbf{E} \times \mathbf{B}/\mu_0. \tag{9}$$

.

From the equations which relate the variation of **E** and **B** we also see that

$$k|\mathbf{E}| = \omega|\mathbf{B}| \tag{10}$$

and hence that

$$|\mathbf{E}| = \frac{\omega}{k}|\mathbf{B}| = c|\mathbf{B}|. \tag{11}$$

The energy density in the wave, and therefore the intensity of the light, is proportional to the square of the **E** and **B** fields, so the vectors **E** and **B** may be though of as the amplitude of the waves. Again more precisely, if $U =$ energy density,

$$U = \tfrac{1}{2}\epsilon_0|\mathbf{E}|^2 + \tfrac{1}{2}|\mathbf{B}|^2/\mu_0. \tag{12}$$

using the relationship between $|\mathbf{E}|$ and $|\mathbf{B}|$, we find that the energy contained in these two contributions is identical, i.e.

$$\tfrac{1}{2}\frac{|\mathbf{B}|^2}{\mu_0} = \tfrac{1}{2}\frac{|\mathbf{E}|^2}{c^2\mu_0} = \tfrac{1}{2}\frac{\epsilon_0\mu_0|\mathbf{E}|^2}{\mu_0} \equiv \tfrac{1}{2}\epsilon_0|\mathbf{E}|^2. \tag{13}$$

More recently (from 1905 onwards) we have understood that electromagnetic waves may also be thought of in terms of particles (photons). However, these particles are also understood to display very wave-like behaviour. In this case we have the relationships that if E and $\mathbf{p}$ are the energy and momentum of the photon

$$E = \frac{h}{2\pi}\omega \qquad\qquad \mathbf{p} = \frac{h}{2\pi}\mathbf{k}, \tag{14}$$

so higher frequency electromagnetic waves have higher energy and momentum. For the purposes of this course we will think of light as a classical wave.

This concludes the introduction to electromagnetic waves, and none of the above in this section is strictly examinable. However, it will hopefully help you understand the behaviour we will discuss in the rest of the course.

1.2 Light in the Electromagnetic Spectrum

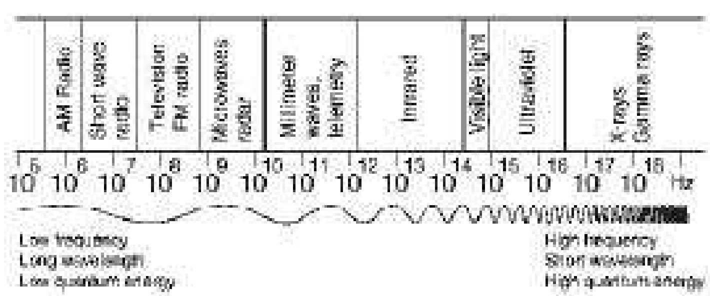

Figure 2: The electromagnetic spectrum as a function of frequency.

Returning to the place of light in the electromagnetic spectrum this is illustrated in figure 2, where we see it lies in a very narrow band of frequencies above those used for TV and radio, microwaves and infrared radiation, but below ultraviolet, x-rays and gamma rays. This band does, however, correspond to the maximum range of emittance for typical stars, such as our sun, and also to the frequencies of many common atomic transitions. Hence, our eye is attuned to these frequencies because they are the most useful for us to "see".

2 Light as a Wave - Propagation.

From now on we will think of light as a wave with amplitude described by

$$\mathbf{A} = \mathbf{A_0} \exp(i\mathbf{k} \cdot \mathbf{r} - iwt) \tag{15}$$

where we implicitly take the real part, and where $\mathbf{A}$ effectively means $\mathbf{E}$ or $\mathbf{B}$. The transverse nature of the wave means $\mathbf{A} \cdot \mathbf{k} = 0$, i.e. $\mathbf{A}$ is in the plane perpendicular to the direction of travel. The motion of the wave can be described using two different features.

The **wavefront** is the plane of constant amplitude or phase for the wave. In the same way that waves in the sea have well-defined crests and troughs of finite extent, the electric and magnetic fields in light waves have planes where amplitude is maximum or minimum. In particular if the wave behaves like

$$\mathbf{A} = \mathbf{A_0} \exp(ikz - iwt) \tag{16}$$

then we define $\phi = kz - wt$ as the phase. The planes where $kz - wt = 0, 2\pi, 4\pi$ etc. correspond to amplitude vector $\mathbf{A} = \mathbf{A_0}$, i.e. maximum amplitude. Similarly the planes where $kz - wt = \pi, 3\pi, 5\pi$ etc. correspond to amplitude vector $\mathbf{A} = -\mathbf{A_0}$, i.e. maximum negative amplitude. For $kz - wt = \pi/2, 3\pi/2, 5\pi/2$ etc. the real part of the amplitude $= 0$. All points with a particular value of ϕ define the wavefront. Equivalent points on the wave are separated by a whole wavelength or period, i.e. $\Delta\phi = 2\pi$,

A **ray** is best thought of the the normal to the local wavefront, and hence, for transverse waves is the direction of travel of the wave. We can describe light waves in terms of wavefronts or rays, but the latter are usually the most convenient.

2.1 Index of Refraction and Optical Path Length

The index of refraction (refractive index) of a medium is defined as the speed of light in a vacuum divided by the velocity in the particular medium,

$$n = c/v. \tag{17}$$

Since we saw earlier that

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} \qquad v = \frac{1}{\sqrt{\epsilon_r \epsilon_0 \mu_r \mu_0}}, \tag{18}$$

and we will always consider materials where $\mu_r = 1$ to a good approximation then we have the straightforward definition that

$$n = \sqrt{\epsilon_r}. \tag{19}$$

For gases ϵ_r is only very slightly greater than 1 and in most common transparent liquids and solids ϵ_r is in the range $1.5 - 5$. For example, $n_{\text{water}} = 1.33$,

$n_{\text{glass}} \approx 1.5$ and $n_{\text{diamond}} = 2.42$. In general denser, more tightly-packed materials have higher refractive index. Usually ϵ_r is weakly dependent on the angular frequency ω of the wave leading to dispersion, as discussed in the first part of the course. We will ignore this effect here.

Note that in general ϵ_r can have an imaginary component,

$$\epsilon_r = \epsilon_{\text{re}} + i\epsilon_{\text{im}}. \tag{20}$$

Using the relationship $k/\omega = \sqrt{\epsilon_r}/c$ this leads to an imaginary component to k

$$k = k_{\text{re}} + ik_{\text{im}}. \tag{21}$$

Substituting this into the expression for the wave we get

$$\mathbf{A} = \mathbf{A_0}\exp(ik_{\text{re}}z - k_{\text{im}}z - iwt) \equiv \mathbf{A_0}\exp(-k_{\text{im}}z)\exp(ik_{\text{re}}z - iwt). \tag{22}$$

Hence the imaginary contribution to the permittivity results in a wave which travels like a wave as usual, but with an exponential decay of the amplitude. Hence, a significant imaginary component leads to a quick decay and is typical of materials where the energy of the wave is absorbed by damping of vibration of atoms. In transparent materials which allow light to be transmitted this effect is very small indeed (though it does exist - little light penetrates very deep underwater), and it is such materials we will deal with in the course.

The concept of refractive index leads to optical path length. For fixed time the phase difference between two points along the ray of a wave is

$$\Delta\phi = k\Delta x = \frac{\omega}{v}\Delta x = \frac{\omega}{c}n\Delta x = k_0 n\Delta x, \tag{23}$$

where $k_0 = (2\pi)/\lambda_0$ and λ_0 is the wavelength in free space, i.e. $k = nk_0$ and hence $\lambda = \lambda_0/n$.

The number of wavelengths between the two points on the ray is

$$\frac{\Delta\phi}{2\pi} = \frac{n\Delta x}{\lambda_0}. \tag{24}$$

Hence, in order to find how far apart in phase two points are for given λ_0 the important unit of measurement is

$$\text{optical path length} \quad = \quad n\Delta x, \tag{25}$$

i.e. optical path length = refractive index × distance travelled. It is also a measure of the time taken to travel the distance Δx in a given medium, i.e. $\Delta t = n\Delta x/c$. The concept of optical path length will be useful in many parts of the course.

2.2 Laws of Light Propagation

There are two laws governing the propagation of light which predate the understanding in terms of electromagnetic waves by centuries, but which will prove to be very useful in deriving the behaviour of light.

1. Fermat's Theorem (1662).

"The actual path between two points taken by a beam of light is the one which is transversed in the least time."
or

"Light, in going between two points, transverses the route having the smallest optical path length."

So we see that light can be thought of as being in a real hurry. Not only does it travel at a speed faster than anything else in the universe, but it always takes the most direct route, refusing to make any detours. This principle has subsequently been generalised to describe the motion of all particles (the Principle of Least Action), but has the simplest interpretation for light.

2. Huygens' Principle (1629-1695).

"Every point on a wavefront acts as a source of a new wavefront propagating radially outwards."

This is sufficient for a quotation of Huygens' principle if it is asked for, but it is useful to be a bit more descriptive. Every point on a primary wavefront acts as the source for secondary wavelets, such that the primary wavefront at a later time is the envelope of these secondary wavelets. The wavelets advance at a speed and frequency which are the same as the primary wave at every point in space.

The two principles provide the same results, one sometimes more easily than the other.

2.2.1 Example 1 - Propagation in a Single Medium

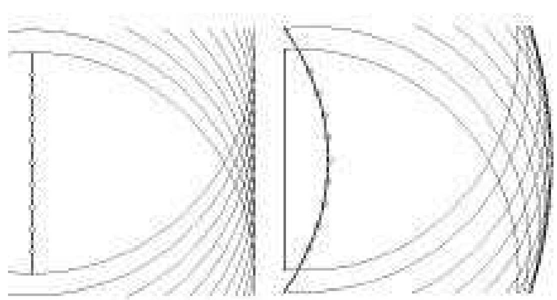

Figure 3: Huygens' principle applied to the propagation of a plane wave.

Fermat's theorem simply says that light gets from A to B in the shortest time and hence using the shortest distance, i.e. it travels in a straight line since this is defined to be the shortest distance between two points. However, the principle can be applied without modification to the more complicated space obtained when considering general relativity. In this case massive stars or black holes can bend space significantly and the bending of light is still described by light travelling past the star or black hole adopting the shortest path in the curved spacetime.

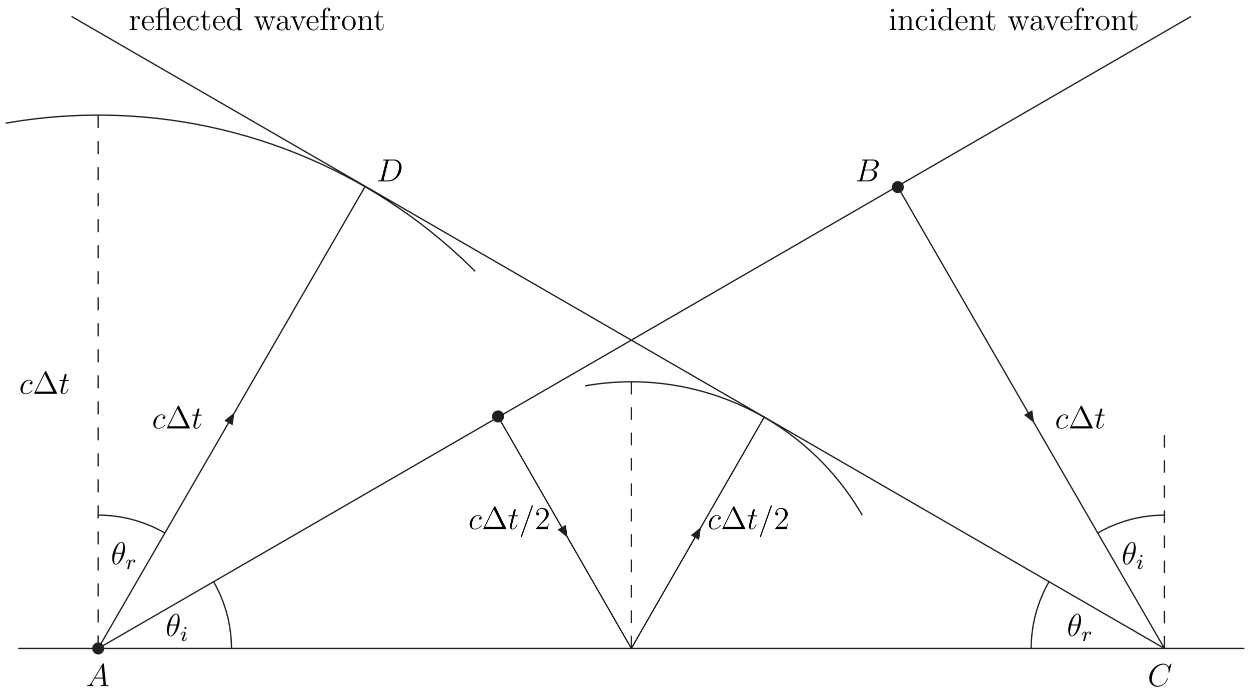

Figure 4: The wavefronts for the incident and reflected light for light incident on a plane reflecting surface.

For Huygens' principle the propagation from a straight, or curved wavefront is shown in figure 3. Since the ray is perpendicular to the wavefront we see that again the propagation in straight lines is obtained.

2.2.2 Example 2 - Reflection from a Plane Surface

In this case we consider the situation in figure 4 where the light ray making an angle θ_i with the normal to the surface which is the same as the wavefront making angle θ_i to the plane of the reflecting surface. θ_i is called the angle of incidence and we want to know the angle of reflection. The diagram showing the application of Huygens' principle with the incident wavefront at point A touching the surface, and that at point B being perpendicular distance $d = c\Delta t$ from the surface is shown, as is the reflected wavefront at time Δt later. The fact that the length of AD and BC are both the same, i.e. the distance travelled by the secondary wavelet $d = c\Delta t$, and that ADC and ABC are both right-angled

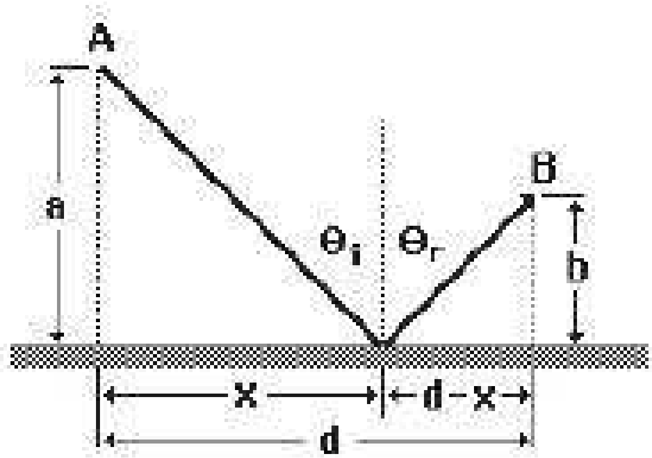

Figure 5: The law of reflection using Fermat's theorem.

triangles means that

$$\sin\theta_i = \frac{BC}{AC} = \frac{AD}{AC} = \sin\theta_r. \tag{26}$$

So we obtain the standard result for the law of reflection, angle of incidence is equal to angle of reflection, i.e. $\theta_i = \theta_r$.

We can consider instead Fermat's principle. The situation is shown in figure 5. Here we want to minimise the time, and hence the path length, for travelling from A to B while striking the mirror at some intermediate time. In the diagram the distance x is the free parameter, the variation of which changes both the incident and reflected angles. The length of the path from point A to the mirror is $\sqrt{a^2+x^2}$ and from the mirror to point B is $\sqrt{b^2+(d-x)^2}$. So the total path length is

$$L = \sqrt{a^2+x^2} + \sqrt{b^2+(d-x)^2}. \tag{27}$$

In this case the light is always travelling at the same velocity so the minimum time is the same as the minimum path length. This is obtained by finding $dL/dx = 0$, i.e. we need,

$$\frac{dL}{dx} = \frac{1}{2}\frac{2x}{\sqrt{a^2+x^2}} - \frac{1}{2}\frac{2(d-x)}{\sqrt{b^2+(d-x)^2}} = 0. \tag{28}$$

Hence,

$$\frac{x}{\sqrt{a^2+x^2}} = \frac{(d-x)}{\sqrt{b^2+(d-x)^2}}. \tag{29}$$

But this is the same as $\sin\theta_i = \sin\theta_r$ so again we have the result that the angle of incidence is equal to the angle of reflection $\theta_i = \theta_r$.

2.2.3 Example 3 - refraction and Snell's law

As light passes from one transparent medium to another it changes speed and bends. The details depend on the angle of incidence and the refractive indices

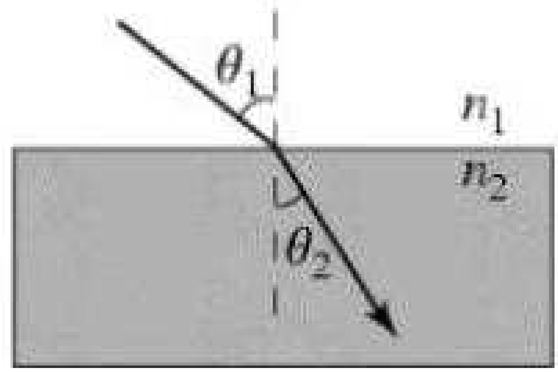

Figure 6: The Law of Refraction – Snell's Law.

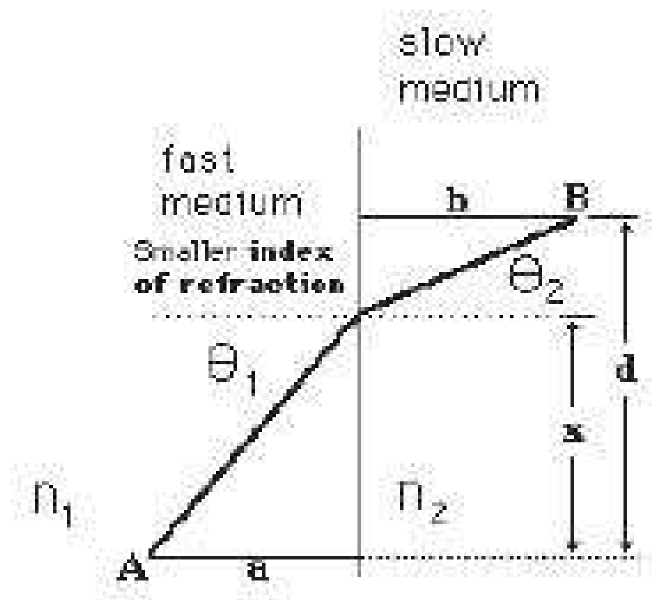

Figure 7: The law of refraction using Fermat's theorem.

of the two media. The general situation is show in figure 6. In 1621 a Dutch physicist Willebrod Snell (1591-1626) defined the relationship between θ_i and θ_t know as Snell's law

$$n_i \sin\theta_i = n_t \sin\theta_t. \tag{30}$$

It is possible to derive this by constructing the diagram of wavefronts using Huygens' principle, but this is left as an exercise. We will instead prove Snell's law using Fermat's theorem in a similar manner to that for the proof of the law of reflection. As shown in figure 7 the time taken to get from A to B is

$$t = \frac{\sqrt{a^2 + x^2}}{v_i} + \frac{\sqrt{b^2 + (d-x)^2}}{v_t}. \tag{31}$$

We need the minimum time, i.e. $dt/dx = 0$, i.e. we need,

$$\frac{dt}{dx} = \frac{x}{v_i\sqrt{a^2+x^2}} - \frac{(d-x)}{v_t\sqrt{b^2+(d-x)^2}} = 0. \tag{32}$$

Hence,

$$\frac{x}{v_i\sqrt{a^2+x^2}} = \frac{(d-x)}{v_t\sqrt{b^2+(d-x)^2}}. \tag{33}$$

But from the definition of $\sin\theta_i$ and $\sin\theta_t$ this is

$$\frac{\sin\theta_i}{v_i} = \frac{\sin\theta_t}{v_t}. \tag{34}$$

But $v_i = c/n_i$ and $v_t = c/n_t$. Therefore,

$$\frac{n_i\sin\theta_i}{c} = \frac{n_t\sin\theta_t}{c}, \tag{35}$$

and therefore

$$n_i\sin\theta_i = n_t\sin\theta_t. \tag{36}$$

Note that in the case that light propagates from a region of high refractive index to one of low refractive index the angle of refraction is larger than that of incidence

$$\sin\theta_t = \frac{n_i}{n_t}\sin\theta_i \ \text{ where } \ (n_i/n_t) > 1. \tag{37}$$

At some particular angle θ_i then θ_t becomes equal to 90^0, i.e. when $\sin\theta_i = (n_t/n_i)$. For incident angles greater than this $\sin\theta_t > 1$, so there is no real angle which satisfies this. Hence, above incident angles of

$$\theta_c = \sin^{-1}(n_t/n_i) \tag{38}$$

we have total internal reflection. No energy propagates away from the boundary on the transmitted side. As an example, for glass with $n_i = 1.4$ to air with $n_t = 1$ (to a good approximation), $\theta_c = 45.6^0$. This phenomenon is made use of, for example, in transmission along optic fibres.

Note that in fact for $\theta_i > \theta_c$ there is some transmitted light, but it decays exponentially rather than travelling as a wave. It is possible to set up a second boundary so that the light can propagate into a new medium. This phenomenon is known as tunnelling. It is often presented as a quantum phenomenon, but is really a feature of wave dynamics.

In general a light wave incident on the boundary between two media of different refractive index will be both reflected and refracted. Before looking at the details of this we need to consider polarisation.

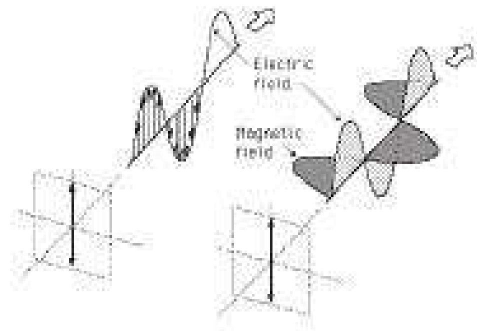

Figure 8: Linearly polarised light.

2.3 Polarisation

2.3.1 Linear Polarisation

So far we have implicitly always considered a light wave described by

$$\mathbf{E} = \mathbf{E_0} \exp(i\mathbf{k} \cdot \mathbf{r} - i\omega t) \tag{39}$$

i.e. the $\mathbf{E}$ field, and consequently also the $\mathbf{B}$ field, is in a particular direction. This is known as linearly polarised light. It appears as in figure 8, where we remember that the $\mathbf{B}$ field is always perpendicular to the $\mathbf{E}$ field and direction of motion $\hat{\mathbf{k}}$.

For given direction of travel there are two independent planes of polarisation, e.g. if the wave travels in the z direction i.e. in the direction of the unit vector $\mathbf{k}$ then $\mathbf{E}$ could be in the x or y directions, i.e. along the unit vectors $\mathbf{i}$ or $\mathbf{j}$, which would be two independent polarisations. However, any linear combination

$$\mathbf{E_{0,1}} = a\mathbf{i} + b\mathbf{j} \tag{40}$$

would be an example of linear polarisation. Then

$$\mathbf{E_{0,2}} = b\mathbf{i} - a\mathbf{j} \tag{41}$$

would be an independent orthogonal polarisation since

$$\mathbf{E_{0,1}} \cdot \mathbf{E_{0,2}} = 0. \tag{42}$$

In quantum mechanics the photon also has two independent degrees of freedom which can be interpreted as two different polarisations. In general, though, light is unpolarised, i.e. it is made up of a random combination of all possible polarisations.)

Law of Malus

The law of Malus states that if we have plane-polarised light incident at angle θ to the plane of polarisation of a perfect polariser, then only the component of the field along the direction of the polariser is transmitted, i.e. the amplitude E_0 is reduced to $E_0 \cos\theta$. However, the intensity $I \propto E_0^2$ so I_0 is reduced to $I_0 \cos^2\theta$.

If unpolarised light is incident on a polariser then each component at angle θ to the plane of polarisation will have intensity fraction $\cos^2\theta$ transmitted. Hence, the total intensity fraction transmitted will be $I_0 < \cos^2\theta >$ (where $< A >$ denotes the average of A). This means the intensity I_1 transmitted by one linear polariser is $\frac{1}{2}I_0$.

As a slightly more complicated example consider light which is unpolarised with intensity I_0 incident on one polariser with polarisation axis along the x axis, and the a second polariser with its axis at 30^0 to the x axis. In this case

$$I_2 = \cos^2 30^0 I_1 = 3/4 I_1. \tag{43}$$

But $I_1 = \frac{1}{2}I_0$, therefore

$$I_2 = 3/8 I_0. \tag{44}$$

There are also other types of polarisation, not just linear polarisation.

2.3.2 Circular Polarisation

Circularly polarised light consists of two perpendicular plane waves of equal amplitude and 90^0 difference in phase, e.g.

$$\mathbf{E} = E_0 \cos(kz - \omega t)\mathbf{i} + E_0 \sin(kz - \omega t)\mathbf{j}. \tag{45}$$

In this case

$$|\mathbf{E}| = \sqrt{E_0^2 \cos^2(kz - \omega t) + E_0^2 \sin^2(kz - \omega t)} = E_0, \tag{46}$$

independent of position or time, but the orientation of the wave changes with time and/or position. Letting $\phi = kz - \omega t$, then for $\phi = 0$ $\mathbf{E}$ is in the x direction but for $\phi = \pi/2$ $\mathbf{E}$ is in the y direction. At the intermediate point $\phi = \pi/4$

$$\mathbf{E} = \frac{E_0}{\sqrt{2}}\mathbf{i} + \frac{E_0}{\sqrt{2}}\mathbf{j}. \tag{47}$$

Circularly polarised light appears as in figure 9. For this particular choice of circularly polarised light the plane of polarisation viewed by an observer towards whom the light is travelling rotates in a clockwise direction. This is known as right-circularly polarised light. (Remember that the direction of rotation is opposite as a function of time to that of a function of position because of the relative $-$ sign in $\phi = kz - \omega t$.) For left-circularly polarised light the polarisation viewed by an observer towards whom the light is travelling rotates in an anti-clockwise direction

Circularly polarised light is obtained by first producing linear polarised light, for example with equal component along the x and y directions, and then delaying (or advancing) the component along the y direction by $\pi/2$, i.e. a quarter

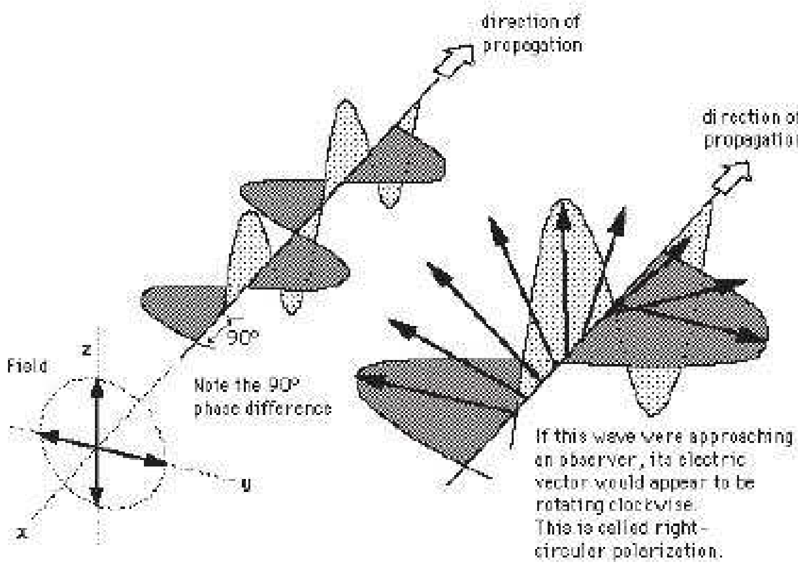

Figure 9: Circularly polarised light.

of a wavelength. A receiver for circularly polarised light must do things in the opposite order, i.e. put the x and y components back in phase and then pick out the resulting linearly polarised light. Circular polarisers are used in 3D cinemas where the film is shown in the two polarisations and glasses receive one type of circular polarisation in one eye and the other type in the other eye. This means the glasses have a definite back and front. The way one wears them they receive circularly polarised light, but turned about so the side normally facing the screen becomes the side away from the screen they instead produce circularly polarised light. If put directly through a plane polariser, circularly polarised light will result in plane polarised light of half the intensity of the circularly polarised light since on average half is aligned along one direction of polarisation and half along the perpendicular direction.

2.3.3 Elliptically Polarised Light

Elliptically polarised light can be made from two unequal amplitude waves plane polarised in perpendicular directions, e.g.

$$\mathbf{E} = 2E_0 \cos(kz - \omega t)\mathbf{i} + E_0 \sin(kz - \omega t)\mathbf{j}. \tag{48}$$

An equivalent wave can be made by two equal amplitude waves out of phase by an angle different to 90^0. It appears as in figure 10, i.e. like circularly polarised light the field vector rotates (in either the clockwise or anti-clockwise direction), but the amplitude changes periodically with time.

In fact both circular and plane polarisations are special cases of elliptical polarisation, where the phase difference between two perpendicular polarisations are $\pi/2$ and 0 respectively.

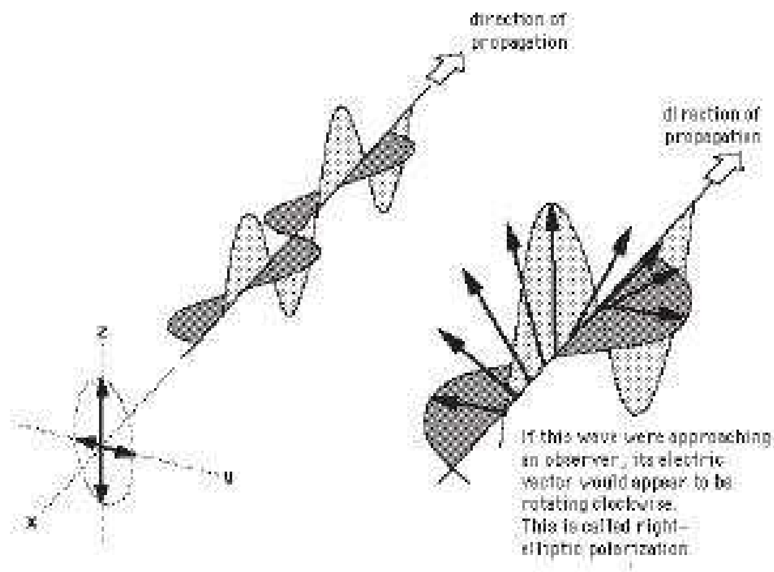

Figure 10: Elliptically polarised light.

2.4 Reflection and Transmission

To derive the general expressions for the reflection and transmission of light at the interface of two media one needs to consider a full treatment of electric and magnetic fields and their boundary conditions as they cross media. This will come in later courses. Here we will just quote the results and examine some of their consequences.

For normal incidence on a plane boundary there is no dependence on polarisation, since all fields are perpendicular to the boundary, and we obtain

$$\text{reflection coefficient function} = r = \frac{n_i - n_t}{n_i + n_t}, \tag{49}$$

and

$$\text{transmission coefficient function} = t = \frac{2n_i}{n_i + n_t}. \tag{50}$$

So if $n_i > n_t$ then r is positive and there is no phase change in the reflected wave and if $n_i < n_t$, i.e. reflection from a more optically dense medium, then r is negative and there is a phase change of π.

In general the incident wave can have field components both perpendicular and parallel to the plane of incidence (or to the boundary), as seen in figure 11. The rules governing the reflection and transmission of electromagnetic waves at an interface are called Fresnel's equations. They are

$$\begin{aligned} r_\parallel &= \frac{n_t \cos\theta_i - n_i \cos\theta_t}{n_t \cos\theta_i + n_i \cos\theta_t} = \frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)} \\ r_\perp &= \frac{n_i \cos\theta_i - n_t \cos\theta_t}{n_i \cos\theta_i + n_t \cos\theta_t} = -\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} \end{aligned}$$

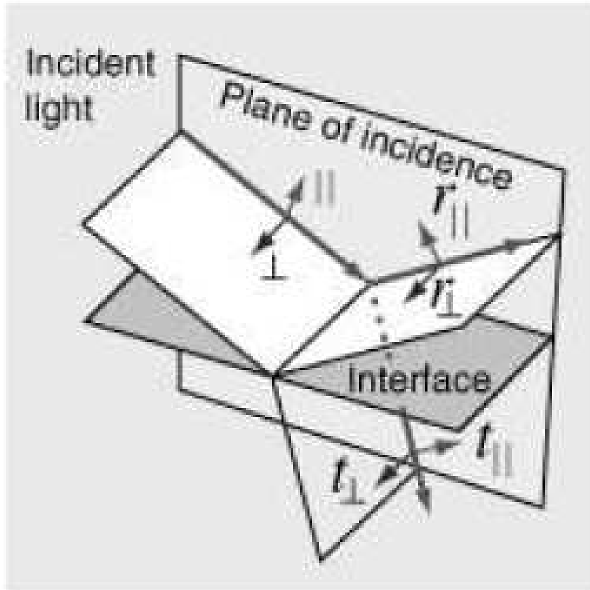

Figure 11: The geometry for Fresnel's equations.

$$
\begin{aligned}
t_{\parallel} &= \frac{2n_i\cos\theta_i}{n_t\cos\theta_i + n_i\cos\theta_t} = \frac{2\sin(\theta_t)\cos(\theta_i)}{\sin(\theta_i+\theta_t)\cos(\theta_i+\theta_t)} \\
t_{\perp} &= \frac{2n_i\cos\theta_i}{n_i\cos\theta_i + n_t\cos\theta_t} = \frac{2\sin(\theta_t)\cos(\theta_i)}{\sin(\theta_i+\theta_t)},
\end{aligned} \tag{51}
$$

where in each case we have used Snell's law to obtain the second expression. (These do not need to be learn. If you have to use any of these expressions they will be given.) The features of the reflection and transmission depend on whether n_t is greater than or less than n_i. However, in either case it is possible that the reflection coefficient for light polarised parallel to the plane of incidence, $r_{\parallel}$, can be zero. From the equation above we can see that this happens if $\theta_i + \theta_t = \pi/2$ because if this condition is satisfied then $\tan(\theta_i + \theta_t) \to \infty$, and

$$r_{\parallel} = \frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)} \to 0. \tag{52}$$

But from Snell's law,

$$n_i \sin\theta_i = n_t \sin\theta_t, \tag{53}$$

and if $\theta_t = \pi/2 - \theta_i$ then $\sin\theta_t = \cos\theta_i$. Therefore

$$n_t \cos\theta_i = n_i \sin\theta_i, \tag{54}$$

and

$$r_{\parallel} = 0 \;\text{ if }\; \theta_i \equiv \theta_p = \tan^{-1}(n_t/n_i). \tag{55}$$

This angle θ_p is known as Brewster's angle, where p denotes that the reflected light is polarised perpendicular to the plane of incidence, the component parallel

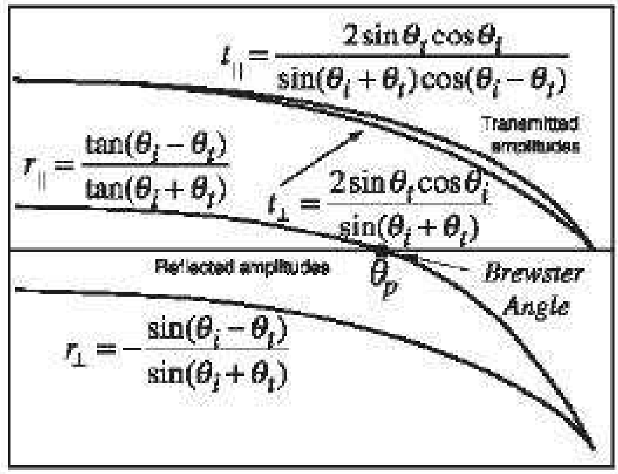

Figure 12: The amplitude coefficients of reflection and transmission for an interface with $n_i < n_t$.

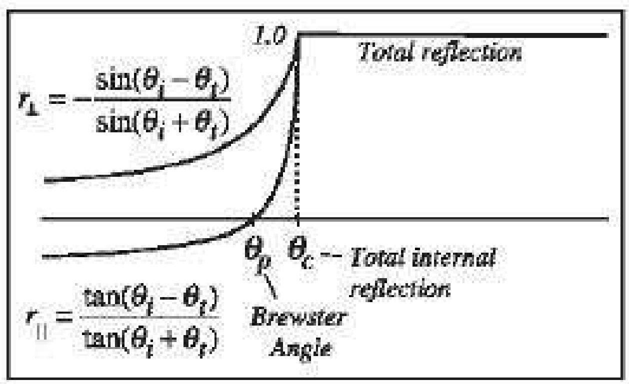

Figure 13: The amplitude coefficients of reflection for an interface with $n_i > n_t$.

to the plane of incidence being zero. This can happen for $n_i > n_t$ and $n_i < n_t$ since $\tan\theta$ can take all values between 0 and ∞. Note also that $\tan\theta > \sin\theta$ so it happens for a smaller angle than that for total internal reflection $\theta_c = \sin^{-1} n_t/n_i$ if $n_i > n_t$.

In brief, the form of r and t, both parallel and perpendicular, can be summarised as in figures 12 and 13. These are for the amplitudes. In terms of the reflected intensity $R = |r|^2$ for $n_i > n_t$ is shown in figure 14. We see that at $\theta_1 = \theta_p = \tan^{-1}(n_t/n_i)$ the reflected light is polarised perpendicular to the plane of incidence, but at all angles there is more light polarised in this direction than parallel to the plane of incidence. θ_2 corresponds to the angle at which total internal reflection sets in, θ_c.

2.5 Methods of Polarisation

There are four common methods for producing polarised light.

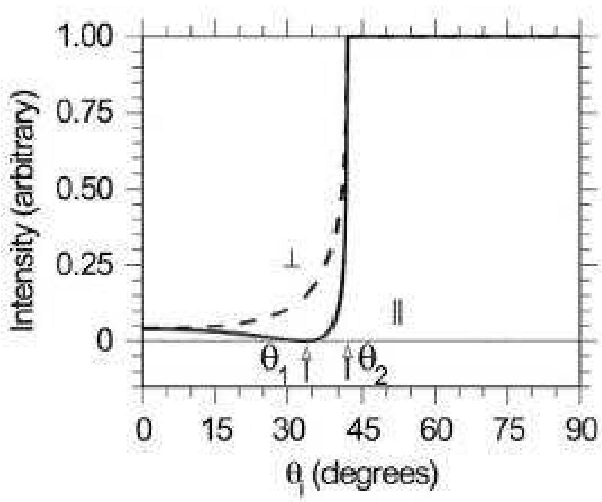

Figure 14: The intensity of reflection for the two polarisations of light for an interface with $n_i > n_t$.

2.5.1 Polarisation by Reflection

As just seen there is preferential reflection of light polarised perpendicular to the plane of incidence, and entirely polarised in this direction for Brewster's angle $\tan\theta_p = n_t/n_i$.

2.5.2 Polarisation by Absorption

Some crystalline materials absorb more light in one direction than another, so light passing through them becomes polarised. This anisotropy is called dichroism. Polaroid film and original sunglasses work like this. Long molecules absorb light in the direction of the molecules by setting up oscillations. The light perpendicular to the long molecules is transmitted. This is illustrated in figure 15.

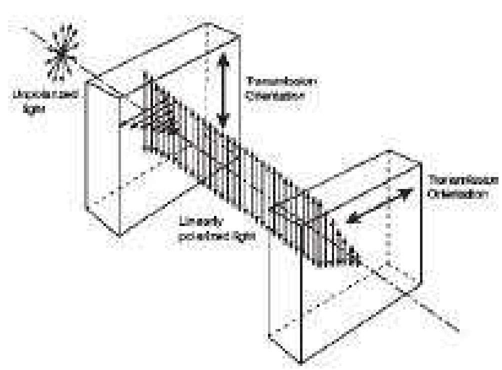

Figure 15: Polarisation by absorption.

2.5.3 Polarisation by Birefringence

Crystalline materials may have different indices of refraction associated with different directions. This property is called birefringence. Mineral crystals commonly have two distinct indices. If the y and z directions are equivalent in terms of crystalline forces the x axis is unique and is called the optic axis. Propagation along the optic axis where **E** is perpendicular to the optic axis is called the ordinary or o wave, whereas the wave with **E** parallel to the optic axis is called the extraordinary or e wave. Prisms can be arranged to give total internal reflection and eliminate one plane of polarisation, as seen in figure 16.

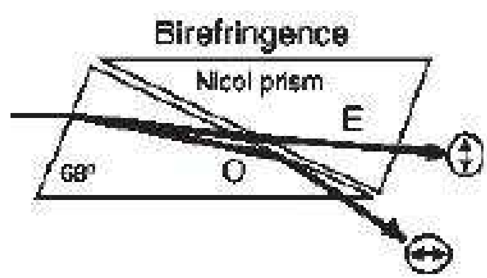

Figure 16: Polarisation by birefringence.

2.5.4 Polarisation by Scattering

When light scatters off air molecules it causes the electrons and nuclei to oscillate in the direction of the field which is perpendicular to the direction of the light. An oscillating charge dipole does not emit in the direction of oscillation. If the wave travels in the z direction then oscillation in the x direction will give light polarised along the x-axis but not emitted in the x direction and oscillation in the y direction will give light polarised along the y-axis but not emitted in the y direction. So the scattered light will be polarised, that in the x-direction along the y axis and vice-versa, as in figure 17. This means that the light under going Rayleigh scattering forming the blue nature of the sky is partially polarised.

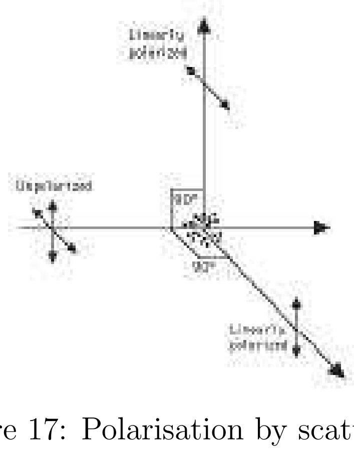

Figure 17: Polarisation by scattering.

3 Interference

Interference is the study of what happens when two or more light waves exist at the same time within the same spatial region. Light waves obey the law of linear superposition, i.e. when one has two waves the resultant is obtained simply by adding the amplitudes of the two. Hence, if the waves are in phase with one another for a significant region of space one gets constructive interference, as illustrated in figure 18. If waves are exactly out of phase with each other then the sum reduces the amplitude and we get destructive interference. The example of perfect destructive interference, i.e. two waves with equal amplitudes with a phase difference of exactly π is shown in figure 19. However, to have appreciable interference we must also have a high level of coherence.

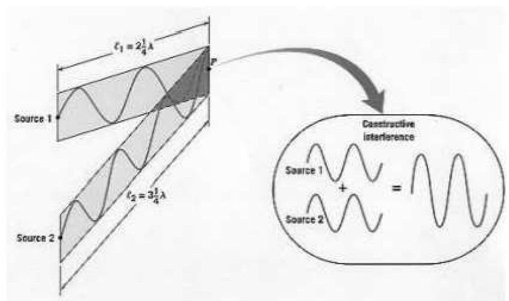

Figure 18: Example of constructive interference.

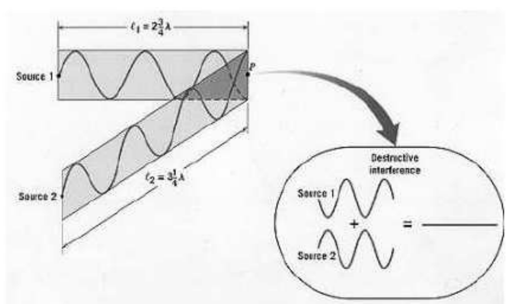

Figure 19: Example of destructive interference.

3.1 Coherence

Coherence is a measure of how well a real wave approximates a perfect infinite plane wave, i.e. we have described a wave amplitude by

$$\mathbf{A} = \mathbf{A_0} \exp(ikz - \omega t), \tag{56}$$

which means that the wave exists in exactly the same sinusoidal form for all values of z at all times. Real waves are not like this. They will only behave like a perfect wave of the form above for a limited time and space. This introduces the two concepts:

Spatial coherence is the distance one goes along a wave with constant separation of peaks and troughs, i.e. over what distance the phase change is given by $\Delta\phi = (2\pi/\lambda)\Delta x$.

Temporal coherence is a measure of how long a wave at a particular point maintains a clear phase relationship between equivalent parts of the wave, i.e. for how long $\Delta\phi = \omega\Delta t$.

Ordinary light is not actually very coherent because it is generated by electronic transitions in atoms which emit on time scales of 10^{-8} seconds and are spatially separated, and the total light comes from lots of different atoms. Rather than one continuous wave it is more like a series of uncorrelated bursts (except in a laser where the method of light generation from simulated emission does lead to coherence). Hence, in order to get stable interference effects one cannot just take two independent standard light sources. The standard method is to obtain interference by dividing a single wave in some manner and then recombining. However, in general the path difference introduced must be some limited number of wavelengths (depending on the source) otherwise coherence starts to fail and the interference pattern breaks down.

3.2 Interference by Amplitude Division - Double Reflection

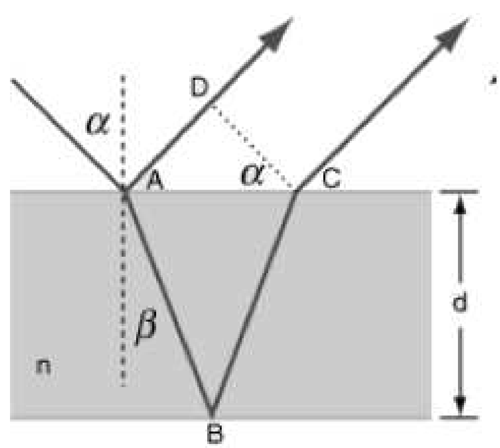

Figure 20: Typical thin film layer and amplitude division of a light ray.

We frequently obtain interference patterns when light is reflected from a surface and then from a second surface some small number of wavelengths later. The general situation is shown in figure 20. Note that the two waves leaving C and D are parallel, but we assume that they are viewed at a very large distance so even almost exactly parallel rays will converge, or that they are focused together by some instrument, e.g. an eye. The optical path difference Γ introduced between the two resultant waves if n_i is the refractive index of the medium from which the light is approaching and n is the refractive index of the layer of thickness d is

$$\Gamma = n(AB + BC) - n_i AD \equiv 2nAB - n_i AD. \tag{57}$$

If the light is incident at angle α to the normal and refracted at angle β then

$$AB = BC = d/\cos\beta, \qquad AC = 2d\tan\beta. \tag{58}$$

But $AD = AC\sin\alpha$ and so

$$\Gamma = \frac{2nd}{\cos\beta} - 2n_i d\tan\beta\sin\alpha. \tag{59}$$

But from Snell's law $n_i \sin\alpha = n\sin\beta$ and

$$\begin{aligned}\Gamma &= 2nd\left(\frac{1}{\cos\beta} - \frac{\sin^2\beta}{\cos\beta}\right) \\ &= 2nd\cos\beta.\end{aligned} \tag{60}$$

All dependence on the refractive index in the incident material has disappeared and ignoring any phase change at either of the boundaries where reflection has taken place, constructive interference requires the optical path length introduced between the waves to be a whole number of wavelengths. Hence, we get constructive interference if $2nd\cos\beta = m\lambda$, where m is an integer and for destructive interference we need $2nd\cos\beta = (m + \frac{1}{2})\lambda$.

However, it is possible that there is a relative phase change of π between the two reflections from the two different surfaces which is equivalent to an optical path difference of half a wavelength. In this case we get constructive interference if $2nd\cos\beta = (m + \frac{1}{2})\lambda$, and destructive interference if $2nd\cos\beta = m\lambda$.

In general we would need to find β from Snell's law, but we will usually consider the cases where $\alpha \approx \beta \approx 0$, i.e. normal incidence where $\cos\beta \approx 1$. We will now consider several standard examples.

3.2.1 Soap Films

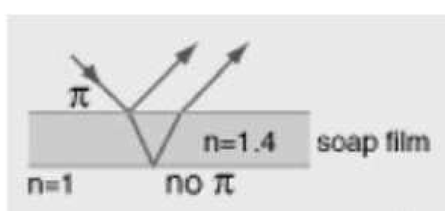

Figure 21: Reflection from a soap film.

We consider a single layer of a soap film in air, as in figure 21. There is reflection from the front and back surfaces. If $n_i > n_t$ there is no phase change on reflection, whereas if $n_i < n_t$ there is a phase change of π. In this case this means there is a phase change of π from the reflection at the front boundary. Therefore, there is constructive interference if $2nd\cos\beta = (m + \frac{1}{2})\lambda$.

3.2.2 Oil Film

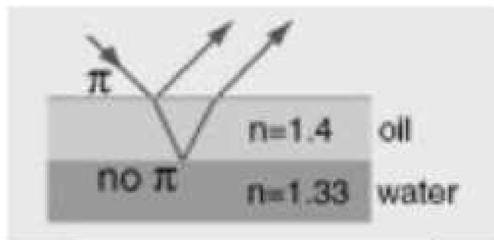

Figure 22: Reflection from a oil film on water.

We consider a thin layer of a oil on water, as in figure 22. The refractive index of oil is slightly greater than that of water so again there is a phase change of π from the reflection at the front boundary. Hence, again there is constructive interference if $2nd\cos\beta = (m + \frac{1}{2})\lambda$.

3.2.3 Anti-reflection Coating

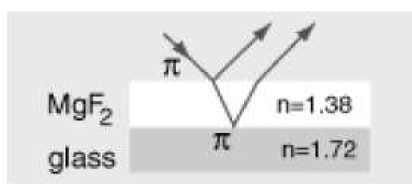

Figure 23: Anti-reflection coating.

Lenses of cameras are often coated with a film of some material with refractive index lower that that of the glass in the lens (which is usually rather high, e.g. ~ 1.7). The example of magnesium fluoride with $n = 1.38$ is shown in figure 23. In this case there is a phase change of π at both boundaries so it cancels out and there is constructive interference if $2nd\cos\beta = m\lambda$ and destructive interference if $2nd\cos\beta = (m + \frac{1}{2})\lambda$.

If we consider normally incident light and a coating exactly 1/4 of a wavelength thick then

$$nd = \lambda/4 \quad \rightarrow \quad 2nd = \lambda/2 \quad \rightarrow \quad \text{destructive interference.} \qquad (61)$$

A quarter-wavelength coating therefore eliminates the reflected wave completely since the two contributions cancel out. This leads to perfect transmission of light into the lens, as desired. However, it only works perfectly for one particular wavelength, which is usually chosen to be somewhere near the middle of the visible range, i.e. $\lambda \approx 475$nm. This means there is some reflection at the blue and red ends of the spectrum, and the reflected light is a mixture of these.

3.2.4 Wedges

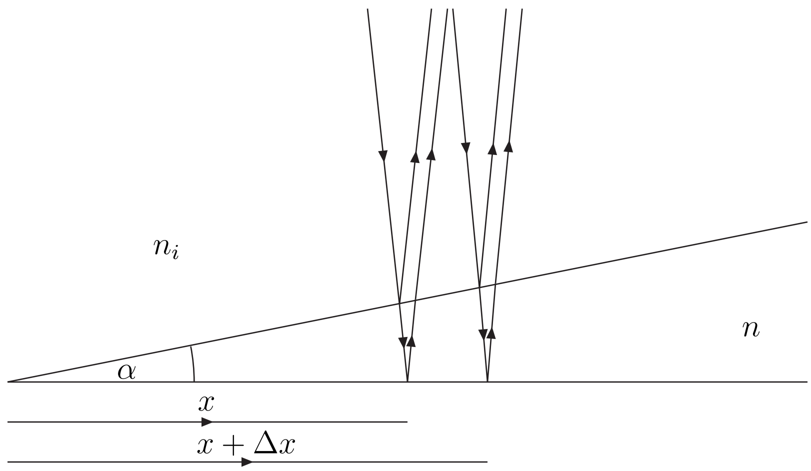

Figure 24: The reflected light at two points separated by Δx for a wedge.

By a wedge we mean a thin sliver of material of different refractive index to its surroundings which tapers to zero at one end. We could have $n > n_i$, e.g a thin strip of glass in air, or $n < n_i$, e.g. a gap between two glass plates touching at one end and separated by a small gap at the other, filled with air, or possibly some other medium. We assume the two planes of reflection make a small angle α with each other. The general situation is shown in figure 24. We consider light incident at an angle to the normal small enough that we can assume normal incidence and reflection at both boundaries. In this case the optical path difference between the two reflected waves at distance x is $2n$ times the thickness of the wedge at this point, which is $d = x \tan\alpha \approx \alpha x$. Hence the optical path difference is given by

$$\Gamma = 2n\alpha x. \tag{62}$$

If the material below the sliver has refractive index n_i (which is usually, but not necessarily the case) there must be a π phase change induced at one of the two boundaries, though which one depends on whether $n_i > n_t$ or vice versa. Hence, there is constructive interference if $2n\alpha x = (p + \frac{1}{2})\lambda$, where p is an integer, i.e.

for

$$x = \frac{(p + \frac{1}{2})\lambda}{2\alpha n}. \tag{63}$$

However, now the result of the interference depends on the distance from the apex. if there is indeed constructive interference, and hence a bright fringe at $x_1 = ((p_1 + \frac{1}{2})\lambda)/(2n\alpha)$ then the next constructive interference and bright fringe will be at $x_1 + \Delta x = ((p_1 + 1 + \frac{1}{2})\lambda)/(2n\alpha)$, so the fringe separation is

$$\Delta x = \frac{\lambda}{2n\alpha}. \tag{64}$$

3.2.5 Newtons Rings

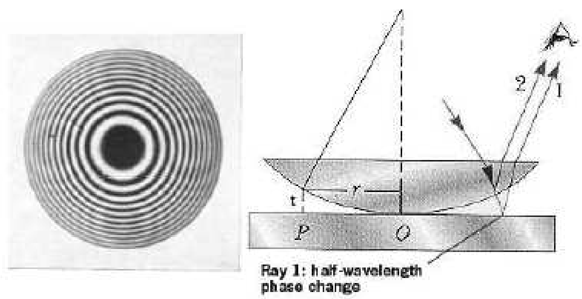

Figure 25: Pattern (left) and geometry (right) for Newton's rings.

One of the first detailed observations of the interference of light is an elaboration of the wedge set-up and is called Newton's rings, since the phenomenon was first studied by Newton. The experimental set-up consists of a section of a large lens with radius of curvature R lying curved side down on a very flat reflecting surface. The gap between the lens and the surface then forms a gap of thickness t which depends on the radial distance from the centre of the lens r. This produces a pattern of light circular fringes if light is incident at near-normal angles. The required situation and the resulting pattern are shown in figure 25.

We can derive the details of the interference pattern by generalising the argument for the wedge in the previous sub-section. From Pythagoras' theorem we see that R, r and t are related by

$$r^2 = R^2 - (R - t)^2 = 2Rt - t^2 \approx 2Rt, \tag{65}$$

where in the last stage we have assumed $t \ll R$, which will be true for fringes near the centre. So the effective wedge thickness is

$$t = \frac{r^2}{2R}. \tag{66}$$

As in the previous discussion there is one phase change of π so we obtain constructive interference, i.e. bright fringes, if

$$2t = (p + \tfrac{1}{2})\lambda \quad \rightarrow \quad r^2 = (p + \tfrac{1}{2})R\lambda, \tag{67}$$

which is the expression for the radial position of the rings. If the gap were filled with a liquid of refractive index n then the optical path length would change from $2t$ to $2nt$ and we would experience the bright fringes at $nr^2 = (p + \frac{1}{2})\lambda$.

If we consider the explicit example of a lens with radius of curvature $R = 1$m, and light of wavelength $\lambda = 500$nm, then we get the first bright fringe when $p = 0$, i.e

$$r_1 = \sqrt{\tfrac{1}{2} \times 5 \times 10^{-7}} = 0.5\text{mm}, \tag{68}$$

and the thirteenth bright fringe is when $p = 12$, i.e. at

$$r_{10} = \sqrt{12\tfrac{1}{2} \times 5 \times 10^{-7}} = 2.5\text{mm}. \tag{69}$$

4 Interferometers

We have looked at the details of interference patterns in the previous section, but in these cases the patterns produced have been fixed, so they can be explained, but there is not much scope for actually using them. An interferometer is a device which produces a variable interference pattern which can be controlled by varying the optical path length between the interfering beams in some manner.

The two examples in this section both use amplitude division of a common light source and the resulting fringe patterns can be used to investigate a variety of topics. For example they can be used to measure the refractive index n of materials, or the thickness of materials if very thin. It is also the case that many spectral lines from light emitted by atoms are in fact a series of very closely separated lines, and interferometers can be used to find and measure these small separations.

4.1 The Michelson Interferometer

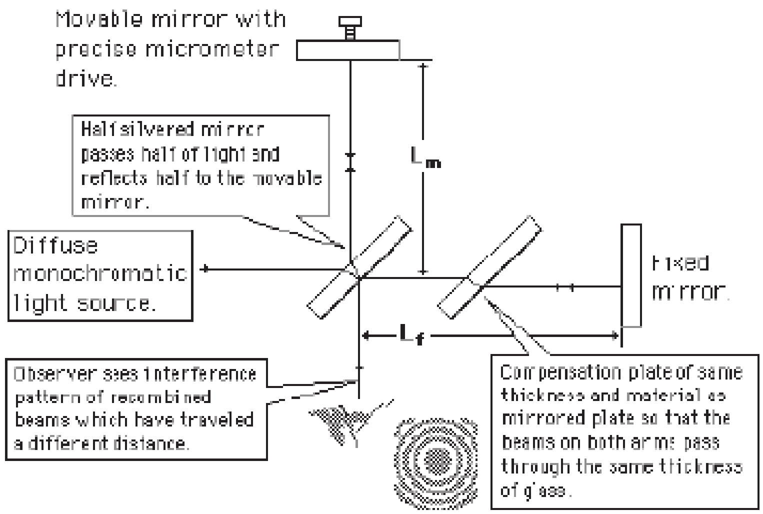

Figure 26: The experimental set-up for the Michelson interferometer.

This is an apparatus which splits a single beam into two equal amplitude parts, which travel down two arms of the interferometer at right angles to each other. One of the beams hits a fixed mirror and is reflected and one a movable mirror. By altering the position of the movable mirror one controls the path difference.The apparatus is shown in detail in figure 26. The amplitude division is made by the beam splitter. One slight complication is that one beam travels through the splitter three times and the other only once. In order to make the path of each beam through glass the same a compensator plate of the same thickness as the beam splitter is added to the second path. This is mainly important if we use light of different wavelengths where n_{glass} may be slightly wavelength dependent, and add a λ-dependent phase difference if the path difference in glass is not the same for each beam.

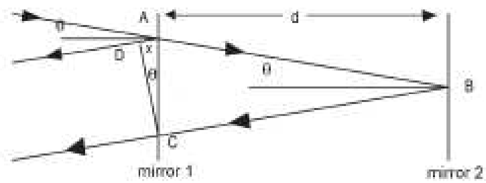

Figure 27: The details of the path difference between the two beams for the Michelson interferometer.

The path difference between the two beams is illustrated in figure 27. It is exactly the same situation as for reflection from two successive boundaries in the previous section, except that it has been achieved without the need for two materials of different refractive index and with a path difference which is variable. As can be seen from the figure, and using exactly the same argument as in eqs. (57-60) in Section 3.2, if the distance between the length of the two arms of the interferometer is $L_f - L_m = d$, then the path difference introduced for light which strikes the mirrors at angle θ to the normal is

$$\Delta = 2dn\cos\theta, \tag{70}$$

where the arms of the interferometer are usually filled with air, so we assume $n = 1$ to a good approximation. There is also one extra phase shift of π in the beam which strikes the fixed mirror, coming from the reflection off the back of the beam splitter. This means that there is constructive interference and bright fringes when

$$2d\cos\theta = (m + \tfrac{1}{2})\lambda \quad \rightarrow \quad \cos\theta = \frac{(m + \frac{1}{2})\lambda}{2d}, \tag{71}$$

where m is an integer. There is destructive interference, i.e. dark fringes, when

$$2d\cos\theta = m\lambda \quad \rightarrow \quad \cos\theta = \frac{m\lambda}{2d}. \tag{72}$$

We can be more quantitative. Assuming that the beam is indeed split exactly into two equal amplitude parts then the total amplitude detected is

$$\begin{aligned} A &= E\exp(i2kL_m) + E\exp(i2kL_m + ik\Delta + i\pi) \\ &= E\exp(i2kL_m) - E\exp(i2kL_m + ik\Delta) \\ &= -e^{(i2kL_m + ik\Delta/2)}(E\exp(ik\Delta/2) - E\exp(-ik\Delta/2)) \\ &= -2iEe^{(i2kL_m + ik\Delta/2)}\sin(k\Delta/2)), \end{aligned} \tag{73}$$

where we have used $\exp(i\pi) = -1$. The intensity is proportional to the amplitude squared so

$$I(\theta) \;\propto\; 4|E|^2\sin^2(k\Delta/2)$$

$$\rightarrow I(\theta) = 4I_0 \sin^2((2\pi/\lambda)d\cos\theta), \quad (74)$$

where I_0 is the intensity from one wave, $I_0 \propto |E|^2$. The pattern is hence such that the intensity falls to exactly half when halfway between the brightest and darkest part of the pattern. As d increases the value of θ for a given fringe increases so they expand outwards and new fringes appear at the centre. Hence, rings appear from the centre and expand, a new one appearing every time $(2\pi/\lambda)d\cos\theta$ increases by π, i.e. a new one appears at the centre ($\cos\theta = 1$) for each increase in d of $\lambda/2$. A slightly idealised version of this is shown in figure 28. In practice a less than perfect interferometer will result in less clear fringe.

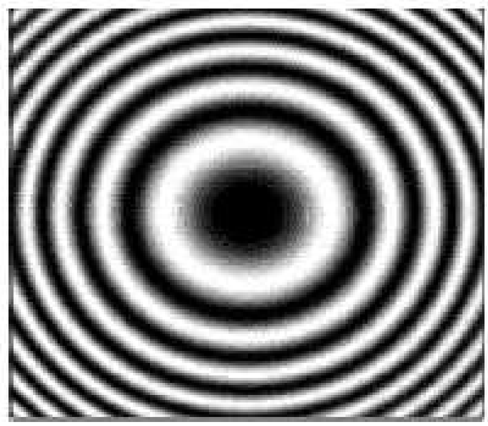

Figure 28: The (slightly idealised) pattern of fringes from a Michelson interferometer using monochromatic light.

4.1.1 Measurement of Refractive Index

If we have air in the arms of the interferometer, and then we introduce an object of refractive index n and thickness t, the optical path of that arm will increase by an amount $(n - n_{\text{air}})t$. The light goes through the object twice, so the optical path difference between the two arms changes by $2(n - n_{\text{air}})t$. Therefore the pattern changes by

$$\frac{2(n - n_{\text{air}})t}{\lambda} \quad (75)$$

fringes. Hence, we can use this result to measure the refractive index of the object, or if this is known to high accuracy already, measure the thickness of objects accurate to less than the wavelength of light.

4.1.2 Resolution of Spectral Lines

Many well-known spectral lines arising from atomic transitions are in fact a series of very closely separated lines. If a source has two lines λ_1 and λ_2 where $\Delta\lambda = \lambda_2 - \lambda_1 \ll \lambda_1$ then we obtain a very slightly different diffraction pattern for the two lines. We can choose $2d\cos\theta$ so that it gives a bright fringe for each, e.g. if this is at the centre ($\theta = 0$)

$$2d = (p_1 + \tfrac{1}{2})\lambda_1 = (p_2 + \tfrac{1}{2})\lambda_2, \tag{76}$$

for two different integers p_1 and p_2. Consequently, p_1, p_2 satisfy

$$p_1 - p_2 = 2d\Big(\frac{1}{\lambda_1} - \frac{1}{\lambda_2}\Big). \tag{77}$$

As d varies the patterns diverge, but we will next get coincident bright fringes at the centre when the difference between the integers increases by exactly one unit, i.e. p_1 increases to $p_1 + m + 1$ and p_2 increases to $p_2 + m$ where m is an integer. When this happens

$$p_1 - p_2 + 1 = 2(d + \Delta d)\Big(\frac{1}{\lambda_1} - \frac{1}{\lambda_2}\Big) = 2(d + \Delta d)\frac{\Delta\lambda}{\lambda_1\lambda_2}. \tag{78}$$

Therefore, this happens if

$$1 = \frac{2\Delta\lambda\Delta d}{\lambda_1\lambda_2}, \tag{79}$$

which leads to

$$\Delta\lambda = \frac{\lambda_1\lambda_2}{2\Delta d} \approx \frac{\lambda_1^2}{2\Delta d}. \tag{80}$$

So in principle we are able to resolve any two wavelengths if we vary the separation between the two mirrors sufficiently to return to the original comparison between patterns. In fact we wish to do this over a distance such that the pattern repeats a number of times and then to divide the total Δd by the number of pattern repeats for maximum accuracy. However, in practice this is limited by $\Delta\lambda$. If this is sufficiently small one pattern repeat may not be possible due to finite coherence of the light source.

4.2 Fabry-Perot Interferometer (Etalon)

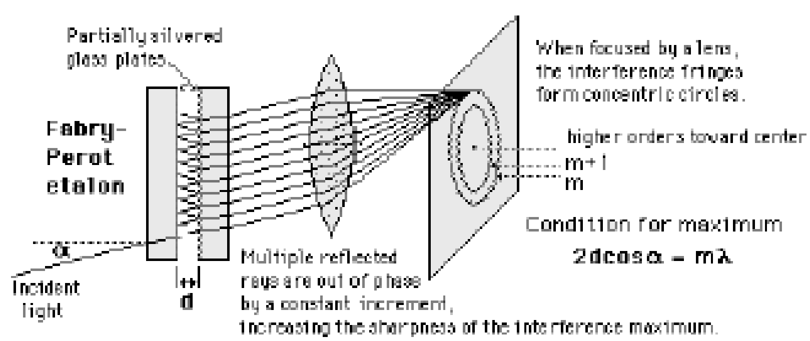

Figure 29: The experimental set up for the Fabry-Perot etalon.

The Fabry-Perot interferometer, or etalon, makes use of multiple reflections between two closely separated mirrored surfaces, as shown in figure 29. At each stage part of the light is transmitted but the majority is reflected back into the etalon. This leads to a larger number of interfering waves than the two we have considered in each of our previous examples. In principle it can be an infinite number, but the amplitude of each successive wave falls, so at some point the amplitude will become negligible. This depends on the size of the reflection coefficients r for the amplitude of the light reflected.

The general condition for bright fringes is as before that there is a constant path difference between successive beams, i.e. if the light enters at angle θ to the normal to the etalon plates, separated by a distance t, then we require $2t\cos\theta = m\lambda$, where m is an integer (this time there is a phase change of π from both reflections, so overall each beam has a change of 2π, which is the same as no phase change). Each successive transmitted beam has amplitude $|r|^2 \exp(i2kt\cos\theta) \equiv R\exp(i2\gamma)$, compared to the previous one, where $\gamma = kt\cos\theta \equiv (2\pi/\lambda)t\cos\theta$, and R is the reflectance. One can show that this results in an transmitted intensity of the form

$$I = I_{\max} \times \frac{1}{1 + F\sin^2\gamma}, \tag{81}$$

where $F = 4R/(1-R)^2$ is defined to be the finesse of the etalon. It is a large number, e.g. for $R = 0.9$ we obtain $F = 360$.

The maximum transmitted amplitude $I = I_{\max}$ when $\sin\gamma = 0$, i.e. when

$$\gamma = \frac{2\pi}{\lambda}t\cos\theta = m\pi \quad \rightarrow \quad 2t\cos\theta = m\lambda. \tag{82}$$

Small values of m means small $\cos\theta$, i.e. wide angles. In fact there is formally a fringe at $m = 0$ but this will never be seen in practice since it is at right angles to the etalon where no light is transmitted. The lowest order fringe is for $m = 1$. As t increases we can obtain more fringes.

The fringes produced by the Fabry-Perot interferometer are far sharper than the Michelson interferometer. The intensity never becomes zero, the minimum relative value being $1/(1+F)$, but this is very small for large F. However, the intensity falls to half it's maximum when

$$\frac{1}{1 + F\sin^2\gamma} = \tfrac{1}{2} \quad \rightarrow \quad \sin\gamma = \pm 1/\sqrt{F}, \tag{83}$$

where $1/\sqrt{F} \ll 1$. Let us say this happens at $\gamma = \gamma_{\max} + \Delta\gamma$, where $\gamma_{\max} = m\pi$. This results in

$$\sin(\gamma_{\max} + \Delta\gamma) = \sin\gamma_{\max}\cos\Delta\gamma + \cos\gamma_{\max}\sin\Delta\gamma = \pm 1/\sqrt{F}. \tag{84}$$

But $\sin\gamma_{\max} = 0$ and $\cos\gamma_{\max} = \pm 1$ and so

$$\sin\Delta\gamma \approx \Delta\gamma = \pm 1/\sqrt{F}. \tag{85}$$

This results in very narrow bands if F is large and is illustrated in figure 30.

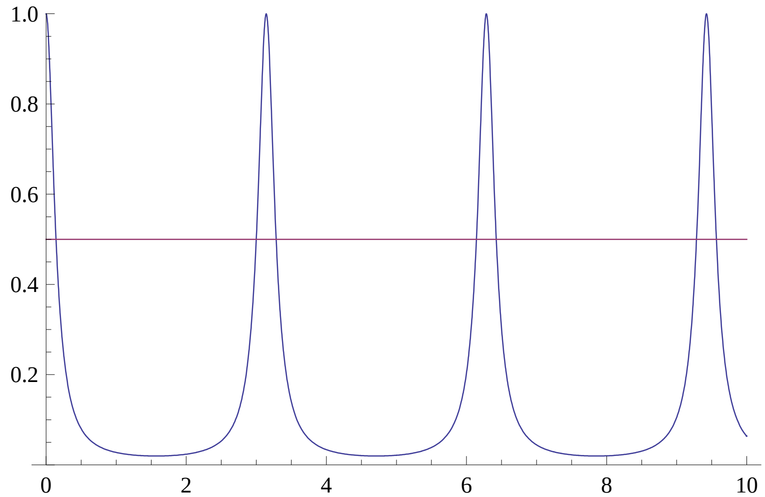

Figure 30: The pattern of intensity from the Fabry-Perot etalon. The x-axis is in units of $\gamma = (\pi/\lambda)2t\cos\theta$. The line showing half-maximum intensity is included.

4.2.1 Chromatic Resolution of a Fabry-Perot Interferometer

If we have two lines of similar wavelength, i.e. λ_1 and λ_2 with $\lambda_2 - \lambda_1 = \Delta\lambda$, the interferometer is deemed to be just able to resolve the fringes at a given order m if the maximum of one falls at the same place as the half-maximum of the other. If we consider the distance from maximum to half-maximum for one wavelength λ_1 then we saw in the previous section that this is given by $\Delta\gamma = \pm 1/\sqrt{F}$, where $\gamma = (2\pi/\lambda_1)t\cos\theta$. For fixed λ_1 and t this means

$$\Delta(\cos\theta) = \pm\frac{\lambda_1}{\sqrt{F}2\pi t}. \tag{86}$$

But maxima are defined to be at $2t\cos\theta = m\lambda$, so the change in the angle at which a maximum is seen as a function of wavelength is

$$2t\Delta(\cos\theta) = m\Delta\lambda \quad \rightarrow \quad \Delta\cos\theta = \frac{m\Delta\lambda}{2t}. \tag{87}$$

Equating the two expressions for $\Delta\cos\theta$ we obtain

$$|\lambda_1/\Delta\lambda| = m\pi\sqrt{F}, \tag{88}$$

which defines the chromatic resolving power of the interferometer. It improves with increasing order of the fringe and increasing reflectance. The former can be increased by making t larger, so allowing fringes of higher order.

5 Diffraction

Diffraction is a study of the light produced when an infinite plane wavefront is interrupted and light from only a finite region on the wavefront is allowed to propagate further. Using Huygens' principle we can treat each point where the light is not interrupted as the source of a new wavelet and investigate the net result of adding the light from all these sources.

We will first consider the simplest example of this, the case of two very narrow slits. We can generalise this by either having a larger number of discrete sources, or we can consider the effect of the finite width of the slits. The general example of the pattern formed in the latter case is shown in figure 31. In all cases we will consider the limit where the distance at which we view the diffraction pattern is much greater than either the size of the image or the the limited region that the diffracted light originates from, and we will first show that in this limit we can assume all the interfering light rays are parallel to each other to a good approximation.

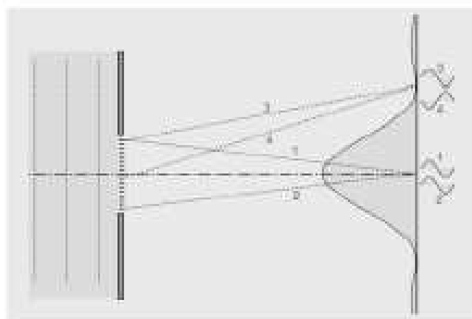

Figure 31: An example of diffraction, i.e. the pattern formed from transmission through a single slit.

5.1 Interference by Wavefront Division – Young's Double Slits

We can allow light from a given source to only pass through two narrow apertures, or slits, separated by distance d and then view the result on a screen at large distance L. This is known as Young's double slit experiment, and is illustrated in figure 32. In the limit that the two light rays are converging sufficiently far from the screen that they are effectively parallel then the path difference between them is simply $d \sin\theta$, where θ is the angle to the normal of the interfering light rays. In this case there is constructive interference if $d \sin\theta = m\lambda$ for integer m. If this approximation is good then we are working in the so-called Fraunhofer limit. If the approximation is not good then we are in the near-field or Fresnel limit, and it is rather more difficult to investigate the details of interference. In practice we will always be working in the Fraunhofer limit, but we should first consider when this limit is indeed justified.

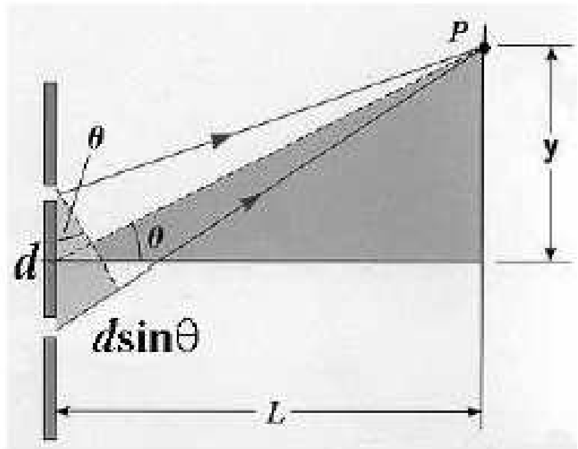

Figure 32: Young's double slit experiment.

Being precise the distance travelled by the upper light ray in figure 32 to the screen is $r_1 = \sqrt{L^2 + (y - d/2)^2}$ and by the lower ray is $r_2 = \sqrt{L^2 + (y + d/2)^2}$. So the exact optical path difference is

$$\Gamma = r_2 - r_1 = \sqrt{L^2 + (y + d/2)^2} - \sqrt{L^2 + (y - d/2)^2}. \tag{89}$$

We can pull L out of the two distances as a common factor and write

$$\Gamma = r_2 - r_1 = L[(1 + (y + d/2)^2/L^2)^{\frac{1}{2}} - (1 + (y - d/2)^2/L^2)^{\frac{1}{2}}]. \tag{90}$$

Assuming that $L \gg (y \pm d/2)$, and using the binomial expansion $(1 + x)^{\frac{1}{2}} \approx 1 + \frac{1}{2}x$, we obtain

$$r_2 - r_1 \approx L + \tfrac{1}{2}\frac{y^2 + yd + d^2/4}{L} - L - \tfrac{1}{2}\frac{y^2 - yd + d^2/4}{L} = \frac{yd}{L}. \tag{91}$$

But $y/L = \tan\theta \approx \sin\theta$ if $\theta \ll 1$ i.e. $y \ll L$. So in the limit that $y, d \ll L$ we obtain

$$r_2 - r_1 = d\sin\theta, \tag{92}$$

and the Fraunhofer limit is satisfied.

Having established the validity of the expression for the path difference we now know we get bright fringes at $d\sin\theta = m\lambda$. However, we can easily be more quantitative about the interference pattern. If the wave from the top slit at a given time is described by $E\exp(ikr_1)$ then that from the bottom slit is $E\exp(ikr_1 + i\delta)$, where $\delta = kd\sin\theta$. Therefore, at angle θ the total amplitude of light on the screen is

$$A = E\exp(ikr_1) + E\exp(ikr_1 + i\delta). \tag{93}$$

It is useful to pull out $E\exp(ikr_1 + i\delta/2)$ as a common factor, obtaining

$$A = E\exp(ikr_1 + i\delta/2)(\exp(i\delta/2) + \exp(-i\delta/2))$$

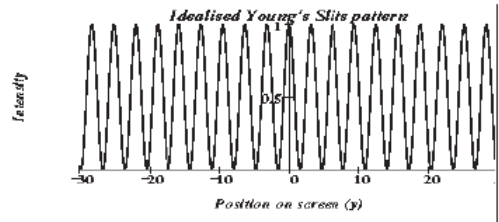

Figure 33: The pattern of intensity in Young's double slit experiment.

$$= \quad E\exp(ikr_1 + i\delta/2) \times 2\cos(\delta/2). \tag{94}$$

The light intensity

$$I \propto |A|^2 = 4|E|^2\cos^2(\delta/2) \equiv 4|E|^2\cos^2(\pi d\sin\theta/\lambda), \tag{95}$$

whereas the intensity from one slit goes like $I \propto |E|^2$.

Hence, the pattern of light intensity from Young's double slits looks as in figure 33. It reaches it maximum $|E|^2$ when $\delta/2 = m\pi$, i.e. when $d\sin\theta = m\lambda$. There is zero intensity for $\delta/2 = (m + \frac{1}{2})\pi$, i.e. $d\sin\theta = (m + \frac{1}{2})\lambda$. For the intermediate case of $d\sin\theta = (m + 1/4)\lambda$ or $d\sin\theta = (m + 3/4)\lambda$ we have $\cos\delta/2 = \pm 1/\sqrt{2}$ and the intensity is half its maximum. If the light is travelling through a medium of refractive index n we simply replace λ by λ/n, or equivalently d by nd in all the above.

Note that if the path difference becomes large, i.e. $m \gg 1$ the coherence may start to be lost, in which case the fringe pattern will start to degrade. However, it is also in this limit that y is becoming larger, so the Fraunhofer approximation will also start to break down, also affecting the pattern.

5.1.1 Modification of the Simplest Case

We will consider two modifications to the simplest case of Young's double slit experiment. First we change the path length from one slit. If a strip of material of thickness t and refractive index n is placed over one slit then it adds a path difference $(n-1)t$. This results in the fringes being shifted, e.g., if $(n-1)t = \lambda/2$ then this is equivalent to a phase shift of π between the two beams and will turn points of constructive interference to destructive interference and vice versa.

Alternatively we could consider letting different amplitudes of light through the two slits. If we let that through the upper slit be amplitude E_1 and that from the lower slit be E_2 then at a point on the screen the total amplitude is

$$\begin{aligned} A &= E_1\exp(ikr) + E_2\exp(ikr + i\delta) \\ &= \exp(ikr_1 + i\delta/2)(E_1\exp(i\delta/2) + E_2\exp(-i\delta/2) \\ &= \exp(ikr_1 + i\delta/2)((E_1 + E_2)\cos(\delta/2) + i(E_1 - E_2)\sin(\delta/2)). \end{aligned} \tag{96}$$

So in this case the intensity is proportional to

$$|A|^2 \quad = \quad (E_1 + E_2)^2\cos^2(\delta/2) + (E_1 - E_2)^2\sin^2(\delta/2)$$

$$= E_1^2 + E_2^2 + 2E_1E_2(\cos^2(\delta/2) - \sin^2(\delta/2))$$
$$= E_1^2 + E_2^2 + 2E_1E_2\cos\delta. \quad (97)$$

For this situation we get maximum intensity $I_{\text{max}} \propto (E_1+E_2)^2$ when $\cos\delta = 1$, i.e. when $\delta = 2m\pi$, and we get minimum intensity $I_{\text{min}} \propto (E_1 - E_2)^2$ when $\cos\delta = -1$, i.e. when $\delta = 2(m+\frac{1}{2})\pi$. Therefore, we have maximum intensity if

$$\frac{2\pi}{\lambda}d\sin\theta = 2m\pi \quad \rightarrow \quad d\sin\theta = m\lambda, \quad (98)$$

and minimum intensity if

$$\frac{2\pi}{\lambda}d\sin\theta = 2(m+\tfrac{1}{2})\pi \quad \rightarrow \quad d\sin\theta = (m+\tfrac{1}{2})\lambda, \quad (99)$$

which are the same conditions as for the case of equal amplitude waves. However, in this case the maximum and minimum intensities will depend on the two amplitudes, and if these are sufficiently different the fringes may be difficult to see. This is quantified in terms of fringe visibility V defined by

$$V = \frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{max}} + I_{\text{min}}}, \quad (100)$$

which needs to be greater than some value, perhaps $V = \frac{1}{2}$, for fringes to be seen with reasonable clarity.

5.2 Multiple Slits – Diffraction Grating

The resolution of the fringes for the double slit experiment is not that good, and for the purposes of resolving spectral lines we would like to improve this. We can try the same way that the Fabry-Perot etalon improved on the Michelson interferometer and increase the number of interfering light sources. Hence, we consider the example of an array of N equally spaced slits, with separation d forming a straight line. This is the generalisation of Young's double slits to cases where $N > 2$. The light rays from each slit, which will ultimately interfere with each other at a distant screen, is shown in figure 34. If we view the light at an angle θ to the normal of the line of slits, the path difference between successive rays is $d\sin\theta$. We can then calculate the total amplitude of light using the same approach as in Section 5.1. If the light received at the screen from the top slit is $E\exp(ikr_1)$, where E is the magnitude of the electric field and $k = 2\pi/\lambda$, then if each slit transmits the same amplitude of light the total amplitude is

$$A_{\text{Tot}} = E\exp(ikr_1) + E\exp(ikr_1 + i\delta) + E\exp(ikr_1 + 2i\delta) + \cdots + E\exp(ikr_1 + i(N-1)\delta), \quad (101)$$

where $\delta = (2\pi/\lambda)d\sin\theta$. We can pull out $E\exp(ikr_1)$ as a common factor and re-express as

$$A_{\text{Tot}} = E\exp(ikr_1)(1 + \exp(i\delta) + \exp(2i\delta) + \cdots \exp((N-1)i\delta))$$

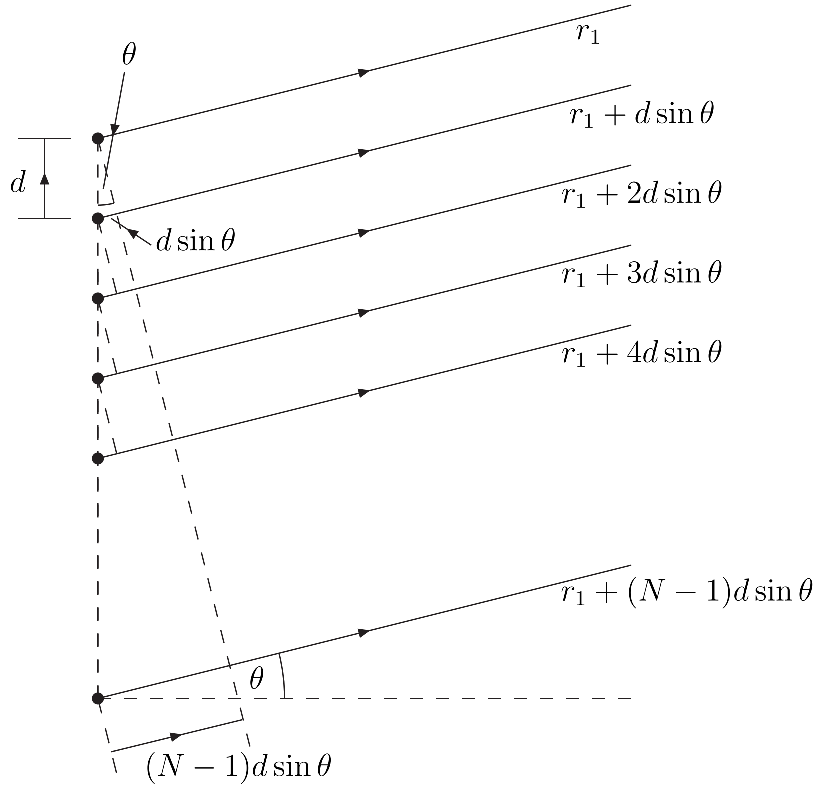

Figure 34: The path difference introduced from transmission through N very narrow slits.

$$= \quad E\exp(ikr_1)(1 + \exp(i\delta) + (\exp(i\delta))^2 + \cdots (\exp(i\delta))^{N-1}) \quad (102)$$

The term within brackets is a geometric series, and can be summed using

$$1 + x + x^2 + \cdots + x^{N-1} = \frac{1 - x^N}{1 - x} \equiv \frac{x^N - 1}{x - 1}. \quad (103)$$

This results in

$$A_{\text{Tot}} = E\exp(ikr_1)\frac{\exp(iN\delta) - 1}{\exp(i\delta) - 1}. \quad (104)$$

Using the same trick as for the double slits and Michelson interferometer we pull out common factors giving

$$\begin{aligned} A_{\text{Tot}} &= E\exp(ikr_1)\frac{\exp(iN\delta/2)}{\exp(i\delta/2)}\left(\frac{\exp(iN\delta/2) - \exp(-iN\delta/2)}{\exp(i\delta/2) - \exp(-i\delta/2)}\right) \\ &= E\exp(ikr_1 + (N-1)\delta/2)\frac{\sin(N\delta/2)}{\sin(\delta/2)}. \quad (105) \end{aligned}$$

The intensity is proportional to the total amplitude squared so

$$I \propto |A_{\text{Tot}}|^2 = |E|^2\frac{\sin^2(N\delta/2)}{\sin^2(\delta/2)}. \quad (106)$$

Introducing a normalisation constant I_0, the intensity of the light from a diffraction grating is equal to

$$I(\theta) = I_0 \frac{\sin^2(N(\pi/\lambda)d\sin\theta)}{\sin^2((\pi/\lambda)d\sin\theta)}. \tag{107}$$

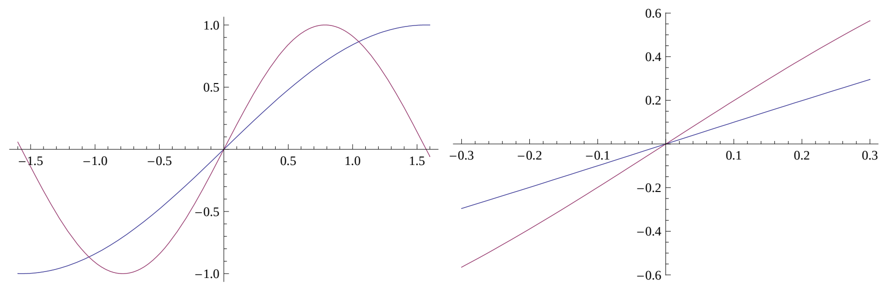

Figure 35: The relative x-dependence of $\sin x$ and $\sin 2x$ (left) and the same focusing on very small x (right).

Let us consider the form of the pattern this equation describes. It is easiest to start with the limit $\theta \to 0$. In this case all the light has travelled the same distance to the screen and there is no path difference between successive waves. Hence, this is the situation where we have perfect constructive interference. In order to obtain the expression for the intensity we need to examine the limit of the expression

$$I(\theta) = I_0 \frac{\sin^2 Nx}{\sin^2 x} \tag{108}$$

as $x \to 0$, where $x = (\pi/\lambda)d\sin\theta$. For small argument x we have the limit $\sin x \to x$ and $\sin Nx \to Nx$. Hence, both expressions $\to 0$, but $\sin x$ is smaller in this region than $\sin Nx$. This is shown in figure 35, where we consider $\sin 2x$ compared to $\sin x$. The former rises twice as quickly and reaches it maximum at $x = \pi/4$ whereas the latter takes until $\pi/2$. Regions nearer to 0 are shown in the right-hand figure and one can see that $\sin x$ is always smaller than $\sin 2x$ in this region. Using the expressions valid in the limit $x \to 0$ we obtain

$$I(0) = I_0 \frac{\sin^2 Nx}{\sin^2 x}\bigg|_{x\to 0} \to I_0 \frac{N^2x^2}{x^2} = I_0 N^2. \tag{109}$$

Since this is due to perfect constructive interference it is the maximum intensity and is known as a primary maximum. It will happen for any other value of $(\pi/\lambda)d\sin\theta$ where $\sin((\pi/\lambda)d\sin\theta)) \to 0$, i.e. $(\pi/\lambda)d\sin\theta = m\pi$, since in all these cases both $\sin^2((\pi/\lambda)d\sin\theta))$ and $\sin^2(N(\pi/\lambda)d\sin\theta))$ will behave in the same way as for $\theta \to 0$. The condition for a primary maximum is therefore

$$(\pi/\lambda)d\sin\theta = m\pi \quad \to \quad d\sin\theta = m\lambda, \tag{110}$$

i.e. the expected result that the path difference between each successive ray is an integer number of wavelengths.

Since this is due to perfect constructive interference it is the maximum intensity and is known as a primary maximum. It will happen for any other value of $(\pi/\lambda)d\sin\theta$ where $\sin((\pi/\lambda)d\sin\theta)) \to 0$, i.e. $(\pi/\lambda)d\sin\theta = m\pi$, since in all these cases both $\sin^2((\pi/\lambda)d\sin\theta))$ and $\sin^2(N(\pi/\lambda)d\sin\theta))$ will behave in the same way as for $\theta \to 0$. The condition for a primary maximum is therefore

$$(\pi/\lambda)d\sin\theta = m\pi \quad \to \quad d\sin\theta = m\lambda, \tag{110}$$

i.e. the expected result that the path difference between each successive ray is an integer number of wavelengths.

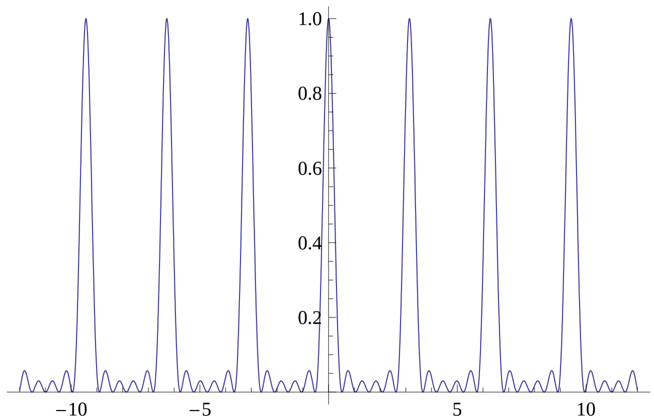

Figure 36: The diffraction pattern from a grating with 6 slits with intensity in units of I_0N^2.

The intensity pattern will have zeros for all θ where $\sin(N(\pi/\lambda)d\sin\theta) = 0$, i.e. for $(\pi/\lambda)d\sin\theta = (k/N)\pi$ (k is an integer), except for the special case where $\sin((\pi/\lambda)d\sin\theta))$ is also zero, i.e. $(\pi/\lambda)d\sin\theta = m\pi$, where we have the primary maxima. This corresponds to the cases where $k = mN$. For given N there will be $N-1$ values of k between two such conditions, so there are $N-1$ zeros in the intensity between two primary maxima. The pattern is shown in figure 36. Between two successive zeros there must be another maximum. These will be approximately when $\sin^2(N(\pi/\lambda)d\sin\theta)$ is at its maximum value of 1. It is straightforward to check that the height of these maxima is much smaller than the height of the principal maxima. These maxima are called subsidiary maxima. There are $N-2$ of them between the $N-1$ zeros which exists between two principal maxima.

We note that the condition for a principal maximum, $d\sin\theta = m\lambda$, means that a maximum of given order m occurs at smaller θ for smaller wavelength light. This means it may be possible to have more maxima (which must be for $\theta < 90^0$) for some lower wavelength (e.g. blue) than a longer wavelength (e.g. red). It also means that, as for the Fabry-Perot etalon, we can use a diffraction grating to distinguish between two closely separated spectral lines of slightly different wavelength.

5.2.1 Chromatic Resolution of a Diffraction Grating

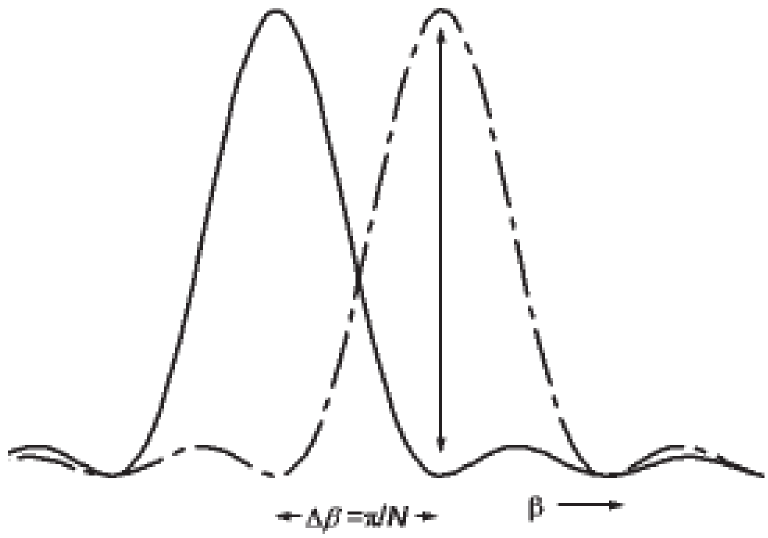

Figure 37: The criterion for telling apart two patterns from slightly different wavelengths.

We consider two wavelengths of light λ_1 and $\lambda_1 + \Delta\lambda$, where $\Delta\lambda \ll \lambda_1$. We wish to define the chromatic resolving power of a grating $|\lambda_1/(\Delta\lambda_{\text{min}})|$. Two patterns are said to be just resolvable if the primary maximum of the light from one wavelength lies exactly at the position of the first zero of the other wavelength, as shown in figure 37.

Let us consider the pattern for the light with wavelength λ_1. If the primary maximum is at $(\pi/\lambda_1)d\sin\theta = m\pi$, the first zero is at $m\pi + \pi/N$. Hence for fixed λ_1 the change in angle from the maximum to 0 may be expressed using

$$(\pi/\lambda_1)d\Delta\sin\theta = \pi/N \quad \rightarrow \quad \Delta\sin\theta = \frac{\lambda_1}{Nd}. \tag{111}$$

Considering instead how the position of the maxima move as λ changes, we have the maximum at $d\sin\theta = m\lambda$, which means

$$d\Delta\sin\theta = m\Delta\lambda \quad \rightarrow \quad \Delta\sin\theta = \frac{m\Delta\lambda}{d}. \tag{112}$$

Equating the two angular changes to obtain the minimum $\Delta\lambda$ we obtain

$$\frac{\lambda_1}{Nd} = \frac{m\Delta\lambda_{\text{min}}}{d} \quad \rightarrow \quad \frac{\lambda_1}{\Delta\lambda_{\text{min}}} = mN. \tag{113}$$

This defines the chromatic resolving power of the grating. As with the Fabry-Perot etalon the resolution increases with order of the fringe, since at wider angles for the grating there is more dispersion of wavelengths. For the grating the resolution also increases with the number of slits.

5.3 Diffraction from a Finite Width Slit

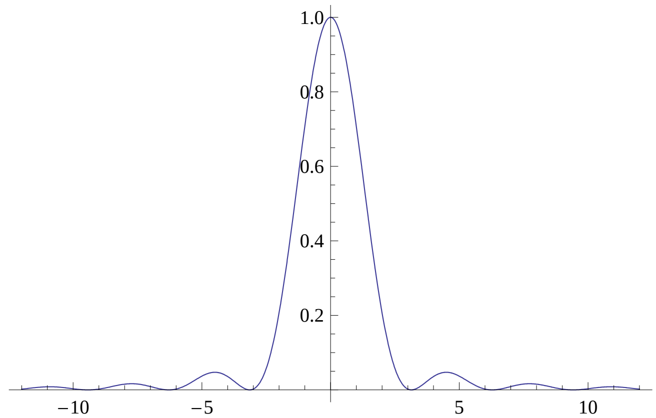

Figure 38: The diffraction pattern from a single slit of finite width.

In order to fully derive the diffraction pattern from a single slit we have to use the technique of Fourier transforms, which you will cover in maths next year. Hence, here we will obtain the pattern using an argument based on the diffraction pattern we have already derived. We can think of a single slit with width a as a very large number of equally spaced narrow slits, each at the centre of a line of length d (so the separation between slits is also d). If there are N narrow slits, each occupying distance d the width of the single slit is $Nd = a$. We can then let $N \to \infty$ and $d \to 0$ such that $Nd = a$ is a finite constant. Doing this we know that the intensity of the array of narrow slits is

$$\begin{aligned} I(\theta) &= I_0 \frac{\sin^2(N(\pi/\lambda)d\sin\theta)}{\sin^2((\pi/\lambda)d\sin\theta)} \\ &= I_0 \frac{\sin^2((\pi/\lambda)a\sin\theta)}{\sin^2((\pi/\lambda)(a/N)\sin\theta)}. \end{aligned} \quad (114)$$

But as $N \to \infty$, $a/N \to 0$ and

$$\sin((\pi/\lambda)(a/N)\sin\theta) \to (\pi/\lambda)(a/N)\sin\theta. \quad (115)$$

This means that

$$I = I_0 \frac{N^2 \sin^2((\pi/\lambda)a\sin\theta)}{((\pi/\lambda)a\sin\theta)^2}. \quad (116)$$

However, we have implicitly always been talking about energy density of light which is proportional to $|E|^2$. As we let $N \to \infty$ the amount of light transmitted from each region of the slit shrinks like $1/N$. Hence, the effective amplitude of the light undergoing interference should be divided by N to reflect this, and therefore the intensity should be divided by N^2. Doing this we obtain the correct answer for a slit of finite width a,

$$I = I_0 \frac{\sin^2((\pi/\lambda)a\sin\theta)}{((\pi/\lambda)a\sin\theta)^2}. \quad (117)$$

This is of the form $\frac{\sin^2 x}{x^2}$ where $x = (\pi/\lambda)a\sin\theta$. This is maximum as $x \to 0$ and we get

$$\frac{\sin^2 x}{x^2} \to \frac{x^2}{x^2} = 1. \tag{118}$$

It is a minimum for $x \equiv (\pi/\lambda)a\sin\theta = m\pi$ for integer m, i.e. for $a\sin\theta = m\lambda$. There are further maximum very near to where $\sin^2((\pi/\lambda)a\sin\theta)$ is a maximum, i.e. at $(\pi/\lambda)a\sin\theta = (m+\frac{1}{2})\pi$, i.e. at $a\sin\theta = (m+\frac{1}{2})\lambda$. The diffraction pattern from a single slit is shown in figure 38. One can see that the central maximum is much higher than the other maxima.

5.3.1 Resolution of Images – Rayleigh Criterion.

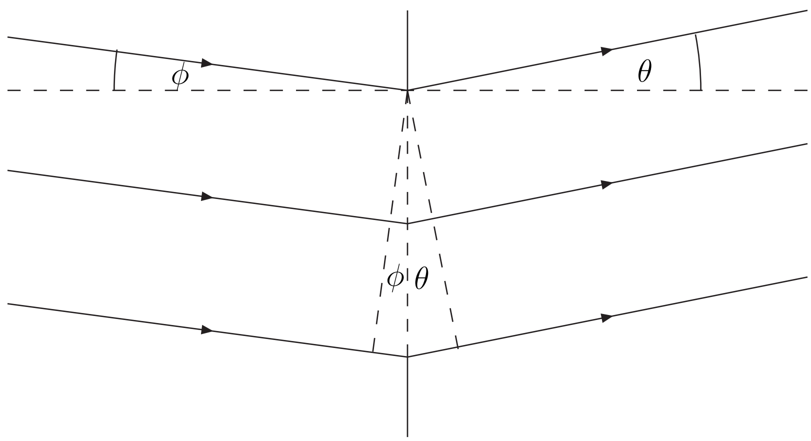

Figure 39: Light incident at a small angle ϕ to a slit.

For the single slit it is interesting to consider the resolution of images of sources at slightly different points. Let us consider that we have one object which emits light which arrives exactly perpendicular to the slit, and another where the light arrives at a small angle ϕ, i.e. the two sources are separated by the angle ϕ. The light arriving at the slit from the first source produces the diffraction pattern we have just derived. That arriving from the second appears as in figure 39. In this case the path difference between waves from difference parts of the slit is proportional to $a\sin\theta$ from the path difference after travelling through the slit, and there is an additional difference of $a\sin\phi$ in the light arriving at the slit. This second path difference could be $-\sin\phi$ is the small angle ϕ is the other side of the normal to to slit. Hence, for the second source the path difference is

$$a(\sin\theta \pm \sin\phi) \approx a\sin(\theta \pm \phi), \tag{119}$$

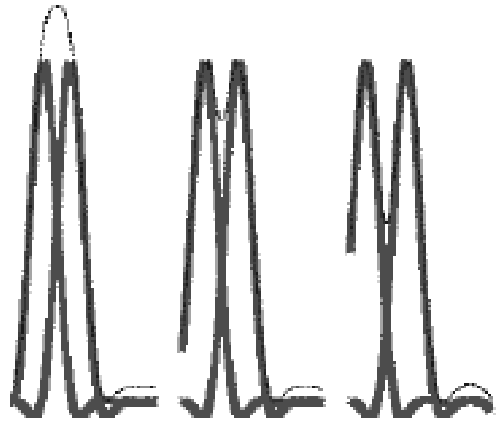

Figure 40: The Rayleigh criterion for telling apart the images from two closely separated sources.

where we have used the approximation that both angles are small, which will be true if we are looking at the limits of resolution.

This means that the diffraction pattern from the second source is

$$I = I_0 \frac{\sin^2((\pi/\lambda)a\sin(\theta \pm \phi))}{((\pi/\lambda)a\sin(\theta \pm \phi))^2}, \tag{120}$$

which is the same as before but shifted by an angle of $\pm\phi$. We can distinguish the two sources if we can see that there are two separate patterns. The criterion for this is called Rayleigh's criterion, and is that the two patterns are said to be distinguishable if the maximum of one is at the minimum of the other. The minimum of the first pattern is at $(\pi/\lambda)a\sin\theta = \pi$. This is the maximum of the second pattern if $\theta = \pm\phi$. This means the two objects are distinguishable if

$$|(\pi/\lambda)a\sin\phi| = \pi \quad \rightarrow \quad |\sin\phi| = \lambda/a = \sin\theta_R \approx \theta_R. \tag{121}$$

So the minimum resolvable angle using Rayleigh's criterion, θ_R is (λ/a). The patterns from the two objects are shown in figure 40, where we move from separations less than Rayleigh's criterion on the left, to exactly at minimum separation, to greater than minimum separation. When the two patterns are separated by the criterion amount the minimum intensity in between the two peaks can be found easily by noting that in the middle both patterns have argument $\pi/2$. Hence the intensity is

$$I = 2 \times I_0 \frac{\sin^2(\pi/2)}{(\pi/2)^2} = 2 \times \frac{1}{\pi^2/4} = \frac{8}{\pi^2} \approx 0.8. \tag{122}$$

If we have a circular aperture rather than a 1-dimensional slit, the pattern formed and argument for the limit of resolutions is very similar, but the diffraction pattern is a Bessel function rather than depending on the sin function. The features are much the same, but a detailed study shows that in this case the limit of resolution is $\theta_R = 1.22\lambda/D$ where D is the diameter of the circle.

5.3.2 The Limit of Resolution of the Eye

Typically the human eye has a pupil diameter (for normal light intensity) of about 2mm. Therefore the limit of resolution is

$$\Delta\theta = \frac{1.22 \times 500 \times 10^{-9}}{2 \times 10^{-3}} = 3 \times 10^{-4} \,\text{rads} = 0.02^0. \qquad (123)$$

At 100m, using $\tan\theta \approx \theta$, which is an excellent approximation for these small angles, this corresponds to a separation of two objects of about 3cm. This is about the best an eye can possibly do. The image of any object is formed on the retina at the back of the eye, and will subtend the angle formed by the object. Since one cannot resolve angles smaller than 0.02^0 this means there is no point having light receptors in the retina closer together than this angle. Indeed, this turns out to be true. The density of receptors is no greater than required by this limit.

5.4 Gratings with Finite Width Slits

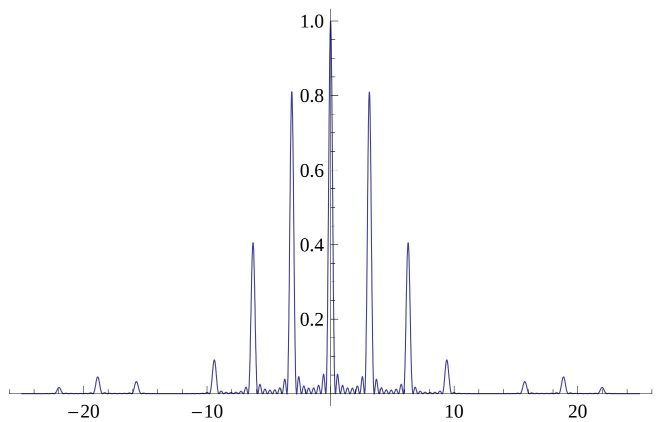

Figure 41: The diffraction pattern from a diffraction grating with 8 slits of finite width $a = d/4$. The x axis is in units of $(\pi/\lambda)d\sin\theta$.

We have seen that the diffraction pattern from a grating with N very narrow slits of separation d is

$$I(\theta) = I_0 \frac{\sin^2(N(\pi/\lambda)d\sin\theta)}{\sin^2((\pi/\lambda)d\sin\theta)}, \qquad (124)$$

and that from a single slit with width a is

$$I(\theta) = I_0 \frac{\sin^2((\pi/\lambda)a\sin\theta)}{((\pi/\lambda)a\sin\theta)^2}. \qquad (125)$$

When we consider a grating to have finite width slits the resulting pattern is just the product of that for the grating and the shape of that for the single slit, i.e.

$$I(\theta) = I_0 \frac{\sin^2(N(\pi/\lambda)d\sin\theta)}{\sin^2((\pi/\lambda)d\sin\theta)} \times \frac{\sin^2((\pi/\lambda)a\sin\theta)}{((\pi/\lambda)a\sin\theta)^2}. \qquad (126)$$

It must be the case that $d > a$, so the first minimum of the single slit pattern must be at a greater angle than the first primary maximum away from the centre. The general form of the pattern is shown in figure 41. One sees that the primary maxima fall until the first single slit minimum, then rise and fall again, with much smaller size. This example has $d = 4a$, but good gratings will have d/a greater and so will have slower fall of the primary maxima intensity. Note that the primary maxima are at $d \sin\theta = m\lambda$ and the diffraction minima for the single slit are at $a \sin\theta = k\lambda$, where m and k are integers. If the ratio d/a is an integer then primary maxima and diffraction minima will coincide, and when $d/a = m/k$ the order m grating maxima will be missing. Indeed, in figure 41 there are no maxima for $m = 4$ or $m = 8$, or indeed if we plotted further for any multiple of 4.

6 Geometrical Optics

In this section we will consider mirrors, lenses, and the instruments which can be made using them. In order to study this we only ever really need to deal with one equation. If we define p as the object distance, q as the image distance and f as a focal length then both mirrors and lenses obey the same common equation

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{f}. \tag{127}$$

However, in various situations we have to consider the definitions of what p, q and f are, and also look at the general consequences.

6.1 Mirrors

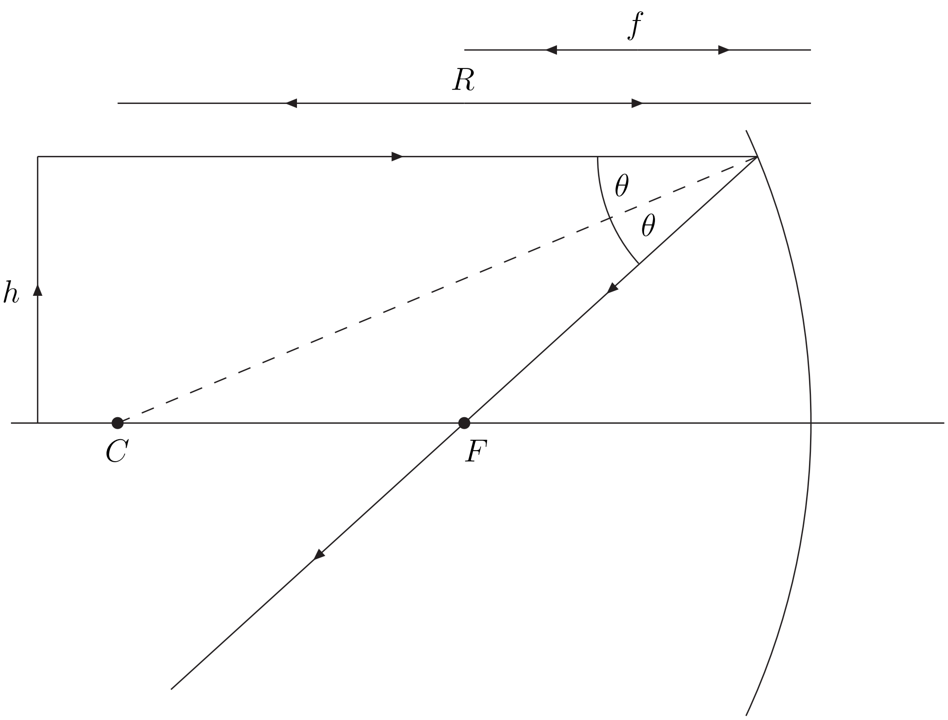

Figure 42: The relationship of centre of curvature to focal point.

We will always consider a mirror to be a section of a spherical shell with a given centre of curvature C, and a radius of curvature R. The reflecting surface may be on the inner surface of the spherical shell, in which case it is a concave mirror, or on the outer surface of the spherical shell, in which case it is convex. A concave mirror will focus light rays towards a point and a convex one will cause them to diverge from a point. More precisely, we define the principal axis of the mirror as the line which is the normal to the centre of the mirror, and then the focal point is that point at which all light rays parallel to the to the principal axis will be focused. For a convex mirror this point is implicitly behind the mirror.

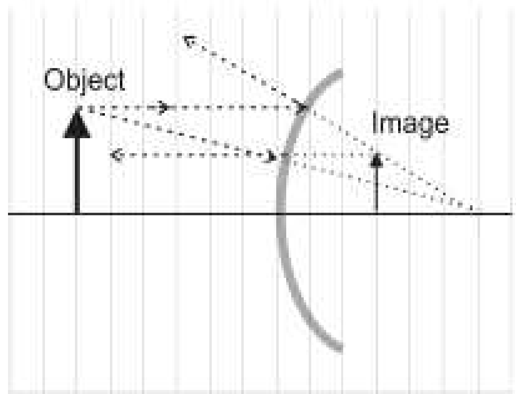

Figure 43: The focus of light rays for a convex mirror.

We can determine the position of the focal point by considering the geometry shown in figure 42. Using the fact that when a light ray strikes the surface at any point the angle of incidence relative to the line from C, which is normal to the mirror, is equal to the angle of reflection, we see we have the results

$$\frac{h}{R} = \tan\theta, \qquad \frac{h}{f} = \tan 2\theta, \tag{128}$$

where f is the distance from the focal point to the mirror. If θ is small this can be approximated by

$$\frac{h}{R} = \theta, \qquad \frac{h}{f} = 2\theta, \tag{129}$$

and clearly in this limit the focal length f is defined by $f = R/2$. This is taken as a general definition, but it should be remembered that it assumes that all rays striking the mirror which are parallel to the principal axis are making a small angle to the normal of the mirror. This assumes the mirror is only a very small part of the full spherical shell.

The situation for the convex mirror is shown in figure 43. It is the same as for the concave mirror, but parallel rays are reflected outwards and it is only the continuation of these reflected rays behind the mirror which meet at the focal point. This means that f is negative for a convex mirror.

6.1.1 Images from Concave Mirrors

We begin the study of the image formed by a concave mirror by making an explicit derivation of the mirror equation. This can be done in a similar manner for any situation, but this is perhaps the most straightforward to present. We consider the case in figure 44, where we have an object O of height h at a

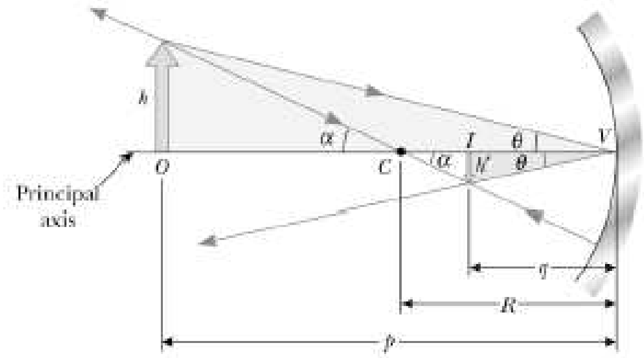

Figure 44: The image formed by a spherical concave mirror when the object O is placed outside the centre of curvature C.

distance $p > R$ from the vertex of the principal axis with the mirror at point V. We look at two separate rays from the top of the object. One goes through the centre of curvature C, and striking the mirror at right angles, is reflected back along the same line. There is another ray which strikes the mirror at V, and hence has angle of incidence θ equal to angle of reflection. The two rays meet at one point, and this forms the "top" of the image I at distance q from the mirror. This has height h', but it is inverted, so h' is negative.

To find the relationship between p and q we can consider the two triangles which have their vertex at C. O and I form the opposite sides of these triangles. The angle at the vertex is the same, let us call it α, it each case. This means we can define $\tan\alpha$ for each triangle and equate the results, i.e.

$$\tan\alpha = \frac{h}{p-R} = \frac{-h'}{R-q}. \tag{130}$$

But, from consideration of the two triangles with angles θ at the vertex, and I and O as the opposite sides we see that

$$\frac{-h'}{h} = \frac{q}{p}. \tag{131}$$

This is a general result which is true for all mirrors and lenses. Putting this in the expression for $\tan\alpha$ we obtain

$$\frac{-h'}{h} = \frac{q}{p} = \frac{R-q}{p-R}. \tag{132}$$

A little rearrangement gives

$$\frac{R-q}{q} = \frac{p-R}{p} \quad\rightarrow\quad \frac{R}{q} - 1 = 1 - \frac{R}{p} \quad\rightarrow\quad R\left(\frac{1}{p} + \frac{1}{q}\right) = 2. \tag{133}$$

This immediately becomes

$$\left(\frac{1}{p} + \frac{1}{q}\right) = \frac{2}{R} = \frac{1}{f}, \tag{134}$$

i.e. we obtain the mirror equation.

We now consider the type of image formed by a concave mirror. For all image distances $p > f$ the general situation is as illustrated in figure 44. The light rays reflected from a mirror do indeed really focus at a point, so the image formed is real. The image is in front of the mirror, which corresponds to the image distance q being positive. The image is also inverted, and in this case reduced in size.

Let us consider an explicit example, $f = 5$cm and $p = 20$cm. Using the lens equation

$$\frac{1}{q} = \frac{1}{f} - \frac{1}{p} = \frac{p-f}{fp} \quad \rightarrow \quad q = \frac{fp}{p-f}. \tag{135}$$

In this case $q = 100/15 = 20/3$cm. We now introduce one more quantity, the lateral magnification M. This is defined by

$$M = \frac{h'}{h} = -\frac{q}{p}. \tag{136}$$

M will be positive if the image is upright and negative if it is inverted. In this example

$$M = -(20/3)/20 = -1/3, \tag{137}$$

and the image is inverted and reduced.

All these features in the example are common to the situation $p > f$ for a concave mirror, except the image is not always reduced (equivalently q is not always less than p. It is easy to see that if $2f > p > f$ then $q > p$, and the image will be enlarged. Indeed as p approaches f from above the image distance $q \rightarrow \infty$.

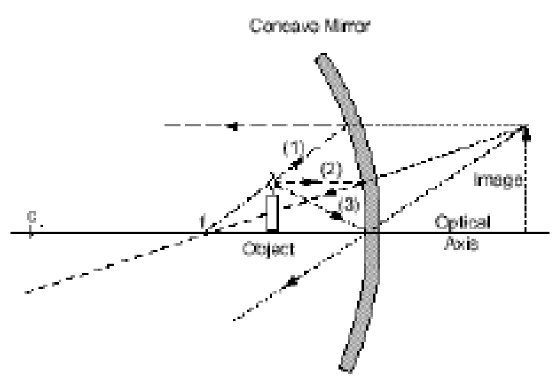

Figure 45: The image formed by a concave mirror when $f > p$.

The features change qualitatively if $f > p$. The general situation is shown in the diagram in figure 45, where this time a light ray incident parallel to the principal axis and focused at F, and one incident from F and reflected parallel to the principal axis are drawn, as well as one which is reflected at equal angle at the vertex. We now see that the general feature is that the light rays from the mirror do not meet, but will appear to do so by an observer looking at the mirror, because they appear to originate from a point behind the mirror. Hence, a virtual image is obtained which is behind the mirror. This image is upright and enlarged.

As a specific example we take $f = 10$cm and $p = 6$cm. This time

$$q = \frac{fp}{p-f} = 60/(-4)\text{cm} = -15\text{cm}. \tag{138}$$

The lateral magnification $h'/h = -q/p = 15/6 = 2.5$, and the image is indeed enlarged.

6.1.2 Images from Convex Mirrors

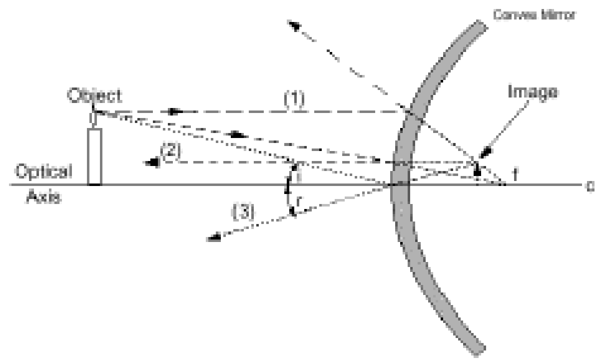

Figure 46: The image formed by a convex mirror.

In this case the equation is the same, but light is focused (in a virtual sense) behind the mirror, so $f < 0$. The situation is shown in figure 46, where again the lines either incident or reflected parallel to the principal axis are shown. The rays diverge away from the mirror and to an observer appear to meet at a point between the mirror and the focal point. The image is upright, virtual and reduced. As a specific example we take $p = 15$cm and $f = -5$cm. This gives

$$q = \frac{fp}{p-f} = -75/20\text{cm} = -15/4\text{cm}. \tag{139}$$

The magnification is $M = -(-15/4)/15 = 1/4$. The magnitude of the image distance q is always less than either p or f, and the magnification is always < 1.

6.1.3 Image in a Plane Mirror

The focal length of a plane mirror is $f \to \infty$, so this is the limit of infinite focal length for either the concave or convex mirror. The mirror equation gives

$$1/p + 1/q = 0, \tag{140}$$

i.e. $p = -q$, and the image is upright, virtual, and has magnification $M = 1$. The ray diagram consistent with this is shown in figure 47.

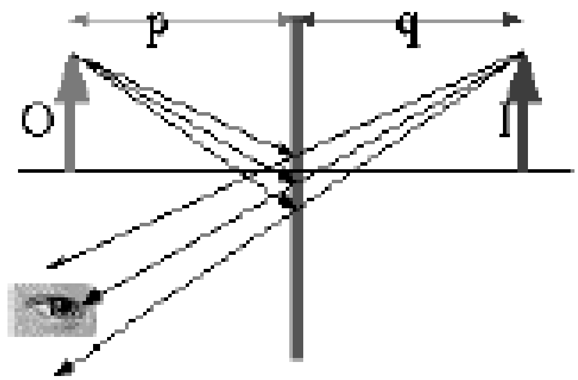

Figure 47: The image in a plane mirror.

6.2 Lenses

6.2.1 Definitions

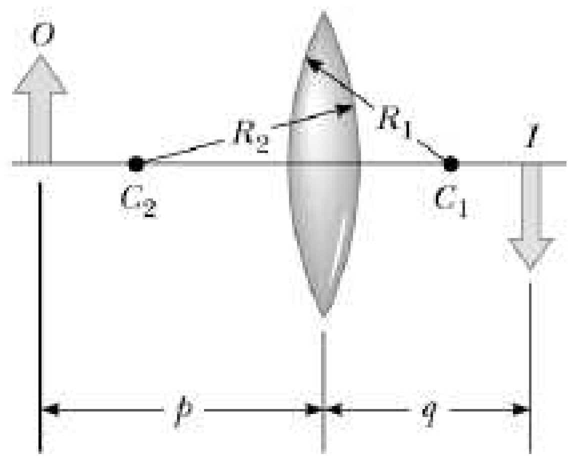

Figure 48: The geometry for a thin lens.

The general situation for a lens is shown in figure 48, where a biconvex lens is shown, but either, or both surfaces could be convex. As with the mirror the surface of the lens is assumed to be a section of the surface of a sphere, but the lens has two distinct surfaces, each with a centre of curvature, C_1 and C_2. Hence, the situation for the lens is similar to the mirror, but there are more degrees of freedom. We make the following definitions.

The object is conventionally in front (in diagrams usually left) of the lens, so this is classified as $p > 0$. If the object for a lens was behind the lens we would have $p < 0$. This can be the case in practice if we consider a system of lenses, as discussed later.

If the image is behind (right of) the lens then $q > 0$. This corresponds to a real image. An image in front of the lens has $q < 0$, and is virtual

The focal length f is defined as $f > 0$ for a converging lens and $f < 0$ for a diverging lens.

The radii of curvature R_1, R_2 are each > 0 if C_1, C_2 are behind the lens and are < 0 if C_1, C_2 are in front of the lens.

As with the mirror the lateral magnification $M = h'/h = -q/p$, and $M > 0$ is an upright image and $M < 0$ is an inverted image.

Figure 48 shows a biconvex lens with $R_1 > 0$ and $R_2 < 0$. A biconcave lens would have $R_1 < 0$ and $R_2 > 0$. It is possible to have one convex and one concave surface, but we will not consider an example of this. Unlike a mirror the lens has a focal point either side of the lens, as illustrated in figure 49.

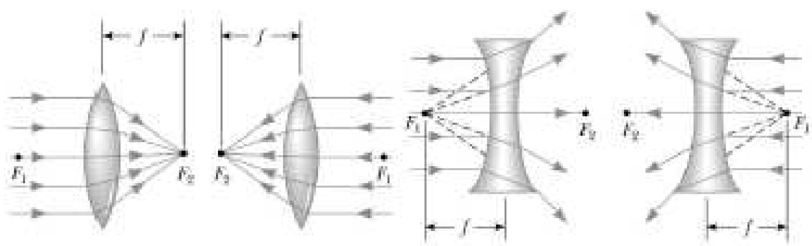

Figure 49: The focal points of a converging (left) and diverging (right) lens.

Having made clear our definitions it would be possible to derive the lens equation using the sort of geometrical arguments we used for the mirror in combination with Snell's law for refraction. This is rather complicated and we will not go through details. We will merely note that the result, in the limit that the lens is thin (equivalent to the statement that the surface of the lens is only a very small section of a spherical surface), is the **lens maker's formula**

$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right), \tag{141}$$

where n is the refractive index of the lens and that of the surrounding medium is implicitly equal to 1. So a bi-convex lens is converging, a bi-concave lens diverging, and a mixed lens depends on the relative radii of curvature. This result for the focal length of the lens can then be used in the lens equation,

$$\frac{1}{f} = \frac{1}{p} + \frac{1}{q}. \tag{142}$$

This results in the images for lenses being much the same as for mirrors, but where light was reflected back in the previous case, it is now refracted through.

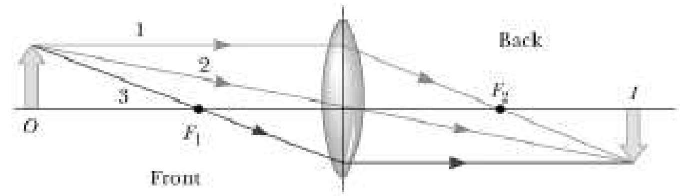

Figure 50: The image for a converging lens with $p > f$.

6.2.2 Images using a Converging Lens.

As for the concave (focusing) mirror there are two qualitatively different cases – $p > f$ and $p < f$. The situation for the former is shown in figure 50. The light rays converge at a point behind the mirror to form a real, inverted image which may be reduced or enlarged, depending on the position of the object compared to the focal length. As an explicit example we consider $p = 11$cm and $f = 5$cm. This leads to

$$q = \frac{pf}{p-f} = 55/6\,\text{cm}, \tag{143}$$

and q is indeed positive. The lateral magnification $M = -q/p = -(55/6)/11 = -5/6$, so the image is reduced and inverted. The image will be enlarged if $f < p < 2f$, e.g for the same lens if $p = 8$cm, $q = 40/3$cm and $M = -5/3$.

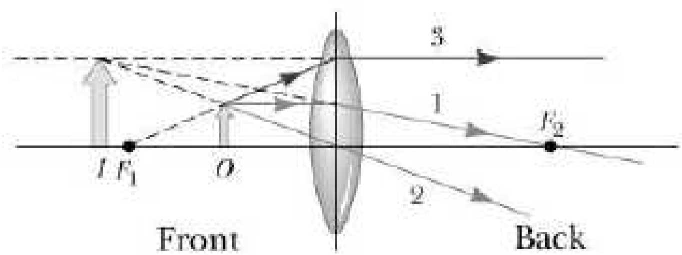

Figure 51: The image for a converging lens with $p < f$.

For $p < f$ the image formation is illustrated in figure 51. The light rays emergent from the lens are divergent and an observer sees them appear to converge in front of the lens. The image is virtual, upright and enlarged. For example, if $f = 10$cm and $p = 6$cm then

$$q = \frac{pf}{p-f} = 60/(-4)\,\text{cm} = -15\,\text{cm}. \tag{144}$$

The magnification is $M = -(-15)/6 = 2.5$, and the image is indeed enlarged.

6.2.3 Images using a Diverging Lens.

As for the convex (diverging) mirror the qualitative situation is the same whatever the value of p, and is shown in figure 52 for a bi-concave lens. The emergent

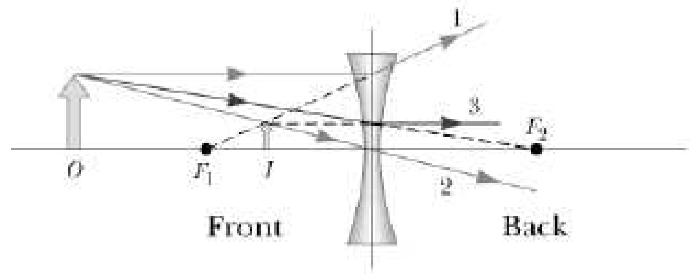

Figure 52: The image for a diverging lens.

rays diverge, so again an observer sees a virtual image appearing to originate from in front of the lens. It is upright, reduced and between the lens and the focal point. For example if $f = -5$cm and $p = 9$cm then

$$q = \frac{pf}{p-f} = -45/14\,\text{cm}. \tag{145}$$

The lateral magnification $M = -(45/14)/9 = 5/14$, and the image is reduced.

6.2.4 Systems of Lenses

It is rather easier to set up a system of more than one lens than it is a system of mirrors. This can be used to obtain enhanced magnification in optical instruments such as microscopes and telescopes, and we will soon consider these specific examples. However, in principle any system is possible. We will only explicitly consider the case of two lenses.

The generalisation from one lens to a system of two with a common principal axis is straightforward. If we have the first lens with focal length f_1 and an object at distance p_1, and the second lens has focal length f_2, the only other parameter to be specified is the separation of the lenses L. The image for the first lens is given by

$$\frac{1}{p_1} + \frac{1}{q_1} = \frac{1}{f_1}, \tag{146}$$

and the image from the first lens then becomes the object for the second. The object distance p_2, i.e. the distance of the first image from the second lens is just $p_2 = L - q_1$, where it is possible that $p_2 < 0$. The resultant image from the second lens is then at position q_2, given by

$$\frac{1}{p_2} + \frac{1}{q_2} = \frac{1}{f_2}. \tag{147}$$

The overall lateral magnification is the ratio of the final image height to initial object height, which is the product of the individual magnifications, since the first image height is identical to the second object height. So overall

$$M = M_1 \times M_2. \tag{148}$$

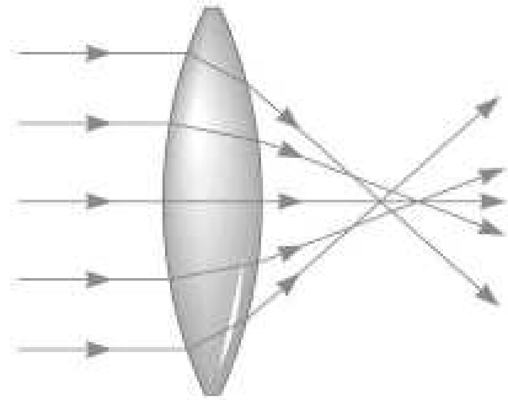

Figure 53: Spherical aberration.

6.2.5 Lens Aberrations

All the above has assumed lenses behave perfectly. This is not always the case, so we quickly review the reasons for this. The departures from ideal behaviour are called aberrations. One type is called **spherical aberration**, and is shown in figure 53. If a lens surface really is part of a section of a perfect sphere then for large sections the focal point is not constant (remember that for the mirror $f = R/2$ was derived in the small angle limit). Hence, light refracted near the edges does not go through the normal focal point. For large lenses this is countered by altering the shape at distances far from the lens centre in the appropriate manner.

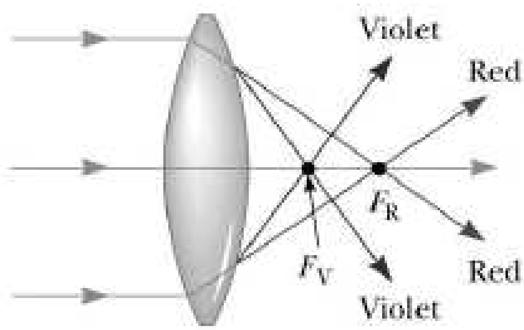

Figure 54: Chromatic aberration.

There is also a feature called **chromatic aberration**. For many materials the refractive index n is slightly higher for lower wavelength light, causing lower wavelengths to be focused slightly nearer to the lens. This is shown in figure 54. It can be minimised by making the complete lens from a combination of materials of different refractive index.

6.3 Optical Instruments

6.3.1 The Eye and Vision Correction

Very simply speaking the eye is a system where a single converging lens focuses light and forms an image on the retina at the back of the eye. The muscle surrounding the eye is able to change its shape to some extent, altering the focal point depending on what is being viewed. However, it does not have unlimited flexibility so the eye has two limiting points. The **near point** is the smallest object distance at which the eye can effectively focus properly. For a normal eye it is about 25cm. Hence, at about 25cm one can see fine detail most precisely, but if the object is closer than this its image will start to blur. The **far point** is the furthest distance at which the lens can focus onto the retina. For a normal eye it is effectively at infinity. However, both the near and far point can change from these values due to imperfections of the eye. There are two standard problems.

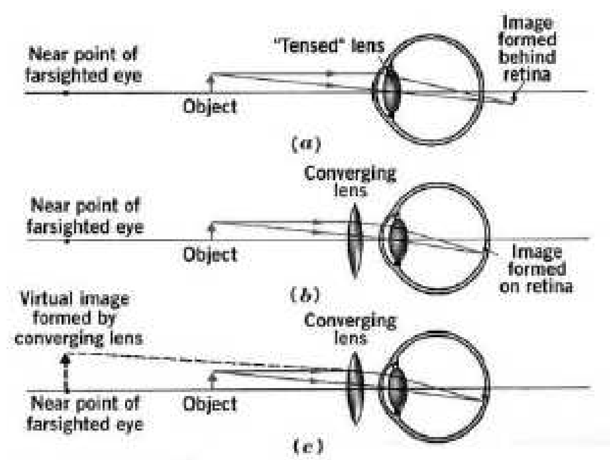

Figure 55: An illustration of farsightedness.

Farsightedness or hyperopia occurs when the light from relatively nearby objects focuses behind the retina. It can be due to the distance between the lens and retina being small, or the curvature of the lens not being able to become great enough. It can be corrected by placing an appropriate converging lens in front of the eye. The object, whose image is now focused onto the retina, then appears to be at the real near point of the eye. This is illustrated in figure 55.

Nearsightedness or myopia occurs when the light from distant objects focuses in front of the retina. It can be due to the distance between the lens and retina being too large, or the cornea – the outer surface of the eye – giving too much of a focusing effect. It can be corrected by placing an appropriate diverging lens in front of the eye. The object, whose image is now focused onto the retina, then appears to be at the real far point of the eye. This is illustrated in figure 56.

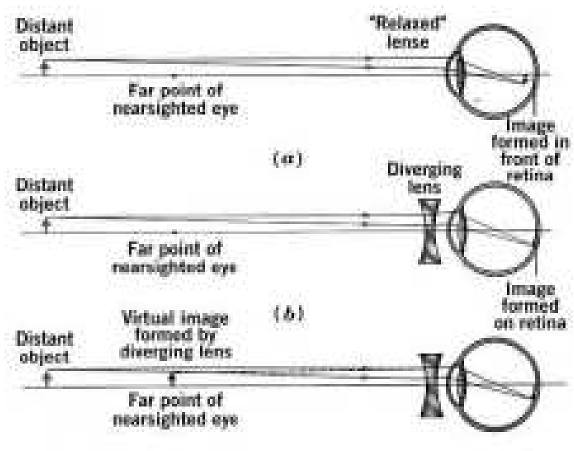

Figure 56: An illustration of nearsightedness.

6.3.2 The Compound Microscope

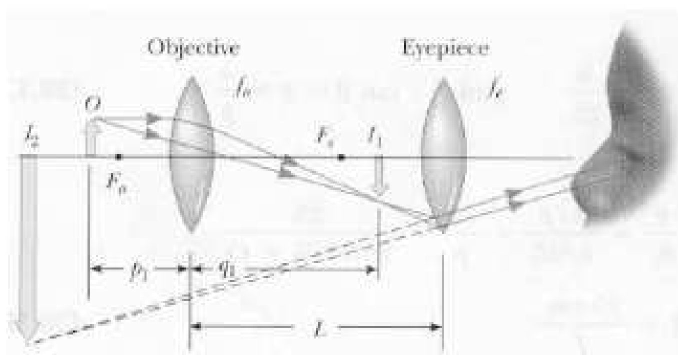

Figure 57: The arrangement for a compound microscope.

A single lens can only provide so much magnification. This can be improved by combining two lenses in a device known as a compound microscope. The arrangement is shown in figure 57. It consists of two focusing lenses. The *objective*, which is nearer to the object, has a very short focal length $f_o < 1$cm, and the *eyepiece*, which is used to view the image, has a focal length f_e of a few cm. The two lenses are separated by a distance $L \gg f_o$ or f_e, and which in practice will be a few tens of cm.

The object is placed slightly outside the focal point of the objective. The first lens forms a real, enlarged, inverted image I_1 just inside the focal point of the eyepiece. The eyepiece then forms the image I_2, which is viewed by the observer, and is virtual, enlarged, and the same orientation as I_1, i.e. inverted compared to the object.

The lateral magnification M_1 of the first image is $-q_1/p_1$. But from the positions described above, $p_1 \approx f_o$ and $q_1 \approx L$. Thus, the lateral magnification by the objective is

$$M_1 \approx -\frac{L}{f_o}. \tag{149}$$

The overall magnification of the microscope is sometimes defined using the angular magnification of the eyepiece, but we will only introduce this concept for the telescope. For the microscope it makes little difference to the answer. Instead we will consider the lateral magnification of the eyepiece. It is natural to take the image distance q_2 to be at the near point of the eye, i.e. $q_2 = -25$cm. The object distance for the lens is the distance of I_1 from the lens, which is given by $p_2 \approx f_e$. Hence, the lateral magnification of the eyepiece is

$$M_2 = -\frac{-25\text{cm}}{f_e} = \frac{25\text{cm}}{f_e}. \tag{150}$$

The overall lateral magnification of the compound microscope is just the product of the magnifications of the objective and of the eyepiece, i.e.

$$M = M_1 \times M_2 \approx -\frac{25L}{f_o f_e}\,\text{cm}, \tag{151}$$

Where the negative sign indicates that the image is inverted. Using, for example, $f_o = 1$cm, $f_e = 5$cm and $L = 50$cm, $M = -250$, which is a typical value.

6.3.3 The Telescope

For the telescope we are viewing objects at very large distances, e.g. planets in the solar system, so the objects are very large, but extremely far away, i.e. the image distance is tending to infinity compared to the focal length. In this case it is not the lateral magnification which is important, but the angular magnification, which is defined by the ratio of the angle subtended by the image, e.g. some region of the moon, compared to the angle subtended by the object viewed without a telescope. There are two different types of telescope, refracting, which uses two lenses, and reflecting, where one mirror is used along with a lens. We will consider both, but the former in more detail.

6.3.4 Refracting Telescope

The basic arrangement is shown in figure 58. As with the compound microscope there is an objective and eyepiece. The separation between the two focusing lenses is very close to the sum of the focal lengths, i.e. $L \approx f_o + f_e$. The focal length of the objective is much larger than that of the eyepiece, i.e. $f_o/f_e \gg 1$. The object is at a very large distance, i.e. $p_1 \to \infty$, so the image distance is given by $q_1 \approx f_o$. The image is real and inverted. The distance between the top and bottom of the image is taken to subtend an angle θ_o from the objective. The image I_1 is formed very near to the focal point of both lenses. The image I_1 subtends an angle θ_0 from the objective, so we have the identity

$$\tan\theta_0 \approx \theta_0 = \frac{-h'}{f_o}. \tag{152}$$

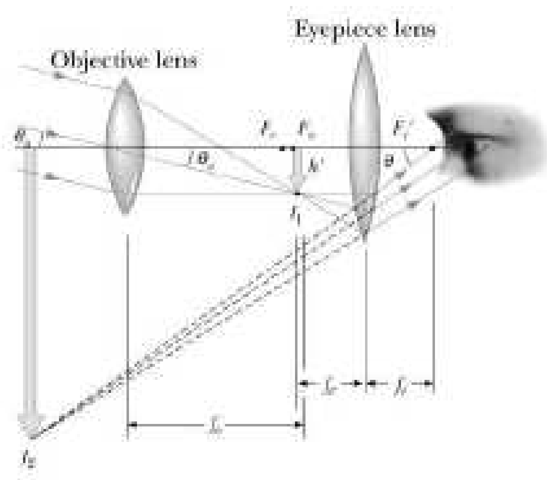

Figure 58: The arrangement for a refracting telescope.

The eyepiece forms an enlarged, virtual image I_2 of the initial object, which is inverted because its object, I_1 is inverted. Since the object distance $p_2 \approx f_e$ the image I_2 has image distance $q_2 \to \infty$. However, the angle θ subtended by the final image at the eye is the same as the angle a ray coming from the tip of I_2 and travelling parallel to the principal axis makes with the axis after it passes through the lens. This ray is focused through the focal point of the eyepiece, so

$$\tan\theta \approx \theta = \frac{h'}{f_e}. \tag{153}$$

The total angular magnification is then expressed as

$$m = \frac{\theta}{\theta_o} = \frac{h'/f_e}{-h'/f_o} = -\frac{f_o}{f_e}, \tag{154}$$

where the negative sign shows that the image is inverted.

In a terrestrial rather than astronomical telescope one generally wants an upright image. This can be achieved by using a diverging lens as eyepiece. This can be understood by noticing in figure 57 that a ray travelling along the line parallel to the principal axis would then diverge out from the focal point left of the eyepiece, rather than towards the focal point right of the eyepiece. It can be shown that again the angular magnification is $m = -f_0/f_e$, which is now a positive number because f_e is negative. Similarly $L = f_o + f_e = f_o - |f_e|$.

Finally we note that whatever the angular magnification the resolution of the telescope is limited by $\theta_R = 1.22\lambda/D$, where D is the objective diameter due to image spreading from diffraction. This leads to reflecting telescopes.

6.3.5 Reflecting Telescope

In order to have the greatest resolution, as well as greatest intensity, the objective of a telescope needs to be as large as possible. It is difficult to make very

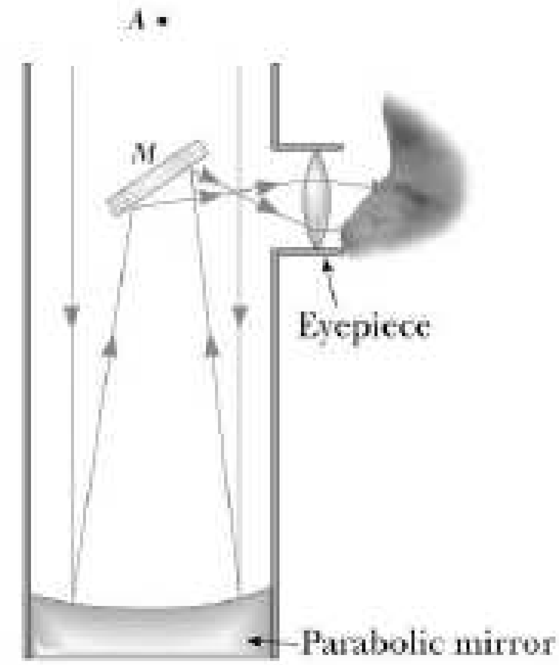

Figure 59: The arrangement for a reflecting telescope.

large lenses since they are very heavy, so can sag causing spatial aberrations, and increased diameter also means thickness which leads to more chromatic aberration due to increased path lengths in the refracting material. Large mirrors can be supported since they do not require light to enter one side and leave the other, and chromatic aberration is not an issue. An example of a reflecting telescope is shown in figure 59, where there is a converging mirror and lens. Reflecting telescopes can be 10m, whereas refracting are limited at about 1m.