Module P6.1 Light — a wave phenomenon

- 1 Opening items
 - 1.1 Module introduction
 - 1.2 Fast track questions
 - 1.3 <u>Ready to study?</u>
- 2 The wave model of light
 - 2.1 Light in a vacuum
 - 2.2 <u>Polarization</u>
 - 2.3 Light in materials
- 3 Propagation and Huygens' principle
 - 3.1 The propagation of an unrestricted wavefront
 - 3.2 <u>The propagation of a restricted wavefront</u> <u>diffraction</u>
 - 3.3 Propagating wavefronts and rays of light
 - 3.4 <u>The reflection of light</u>
 - 3.5 The refraction of light

- 4 <u>Illumination and the inverse square law</u>
 - 4.1 Intensity
 - 4.2 <u>The inverse square law of illumination</u>
- 5 Interference and the superposition principle
 - 5.1 <u>The superposition principle</u>
 - 5.2 Young's two-slit experiment
 - 5.3 <u>Coherence a necessary condition for</u> <u>interference</u>
 - 5.4 Interference of light reflected from thin films
 - 5.5 <u>The diffraction grating</u>
 - 5.6 Fraunhofer diffraction at a single slit
- 6 Closing items
 - 6.1 Module summary
 - 6.2 <u>Achievements</u>
 - 6.3 Exit test

Exit module

1 Opening items

1.1 Module introduction

Light is vital to life. There would be no life on Earth without the energy from the Sun and much of that energy is transferred by visible light. How is the energy transmitted through space and how it is delivered on arrival? Are each of these processes analogous to the passage of particles (like bullets from a gun) or waves (like waves on a beach)? Experiments must decide which is the more appropriate mechanism, waves or particles, but let us look at the alternatives for a moment.

Particles might be expected to travel in straight lines and to form sharp edged shadows, but waves are able to travel around obstacles and might therefore blur the edges of shadows. Particles would be expected to travel large distances carrying their energy with them, but a wave usually results from the coordinated local motion of many neighbouring parts that pass energy from one to another — the parts themselves don't usually travel very far, and what movements they do make may take place in any direction relative to the direction of energy flow. Despite these gross differences between waves and particles, the battle over the true nature of light raged for a very long time. Isaac Newton (1642–1727), for example, favoured a sophisticated form of particle, while Christiaan Huygens (1629–1695) favoured the wave mechanism.

The matter finally appeared settled in 1801 when a British physician, Thomas Young (1773–1829), detected clear signs of wave-like behaviour. It only remained, it seemed, for those who came after to elaborate the wave model and to determine the nature of the waves.

This was achieved in 1864, by James Clerk Maxwell (1831–1879), when he showed that oscillating electric and magnetic fields can propagate (i.e. travel) as an *electromagnetic wave* through space or various materials. But even these discoveries were later to be seen in a new way as the development of *quantum theory* allowed the debate to continue into the 20th century.

It is not only physicists who have been fascinated for hundreds of years by the nature of light. Almost everyone has marvelled at the way the colours of a peacock's tail, or of an oil film on water, change as they are viewed from different angles. In this module you will see how both these observations provide powerful clues to the nature of light.

This module describes the *wave model* of light. Section 2 will introduce you to Maxwell's *electromagnetic waves* and show how they account for various properties of light such as colour and *polarization*. Section 3 sets out *Huygens' principle*, an idea that predates Maxwell's work but which is nonetheless able to provide insight into the transmission or propagation of light, including such important phenomena as *diffraction*, *reflection* and *refraction*. In Section 4 the *intensity* of light waves is defined. This leads to a discussion of the transfer of energy by light and of the *inverse square law of illumination*. Finally, in Section 5, the *superposition principle* is introduced and used to explain various aspects of *interference*, including *Young's experiment*, reflection from *thin films*, the operation of *diffraction gratings* and the phenomenon of *Fraunhofer diffraction* at a single slit. It is this final section that deals with the oil film and the peacock's tail.

Study comment Having read the introduction you may feel that you are already familiar with the material covered by this module and that you do not need to study it. If so, try the *Fast track questions* given in Subsection 1.2. If not, proceed directly to *Ready to study?* in Subsection 1.3.

1.2 Fast track questions

Study comment Can you answer the following *Fast track questions*?. If you answer the questions successfully you need only glance through the module before looking at the *Module summary* (Subsection 6.1) and the *Achievements* listed in Subsection 6.2. If you are sure that you can meet each of these achievements, try the *Exit test* in Subsection 6.3. If you have difficulty with only one or two of the questions you should follow the guidance given in the answers and read the relevant parts of the module. However, *if you have difficulty with more than two of the Exit questions you are strongly advised to study the whole module*.

Question F1

Draw diagrams to show how Huygens' principle may be used to demonstrate the laws of reflection and refraction of light.

Question F2

Light of wavelength 550 nm in a vacuum, passes into a block of glass where its speed is $2.10 \times 10^8 \text{ m s}^{-1}$. What is the frequency and wavelength of the light in the glass?



Question F3

What is meant by polarized light? Give your answer in terms of the directions of the electric and magnetic fields. What is unpolarized light?

Question F4

In a particular Young's two-slit experiment, light of wavelength 600 nm is used. Find the separation of the slits if the first order interference maximum appears at an angle of 0.15°.

Question F5

What is meant by the term diffraction? Describe the essential characteristics of a diffraction grating.





Study comment Having seen the *Fast track questions* you may feel that it would be wiser to follow the normal route through the module and to proceed directly to <u>*Ready to study?*</u> in Subsection 1.3.

Alternatively, you may still be sufficiently comfortable with the material covered by the module to proceed directly to the *Closing items*.

1.3 Ready to study?

Study comment In order to study this module you will need to be familiar with the following terms: <u>Cartesian</u> coordinates, displacement, direction (of a vector), electric field, energy, magnetic field, magnitude (of a vector), oscillation, power, <u>SI units</u>, speed and vector. You will also need an everyday notion of what constitutes wave motion. The mathematical requirements include the use of <u>radians</u> to measure angles, the basic geometry of triangles, and basic trigonometry. If you are uncertain about any of these terms, or how to use them, refer to the Glossary which will indicate where in FLAP they are explained. The following Ready to Study questions will help you to check that you have the required level of skills and knowledge.

Question R1

Describe in one sentence what is meant by an *oscillation*.

Question R2

How would you detect the presence of an <u>electric field</u> in some region of space?

Question R3

If the village of Smallplace is 20 miles north of the city of Bigplace, what is the magnitude and direction of the *displacement* vector from Smallplace to Bigplace?





2 The wave model of light

2.1 Light in a vacuum

In his pioneering investigation of *electric* and *magnetic fields* James Clerk Maxwell was able to construct a theory of electromagnetism which made a crucial prediction. This was that oscillating electric and magnetic fields could travel through space as an <u>electromagnetic wave</u>— a pattern of fluctuating electric and magnetic fields that could tumble together through a vacuum at a speed equal to the known speed of light.

The German physicist Heinrich R. Hertz (1857–1894) succeeded in generating electromagnetic waves (of the type we now call *radiowaves*) in 1888. He showed that they behaved like light in many respects and that their speed was indeed the same as the speed of light. This confirmed Maxwell's prediction and helped to establish the idea that light itself is an electromagnetic wave phenomenon.

The nature of an electromagnetic wave needs careful consideration. Various kinds of waves are important in physics but few are as subtle as electromagnetic waves. For many sorts of waves it is easy to visualize what is actually 'waving'. In the case of *sound waves* in air, for example, the *density* or *pressure* of the air varies as individual molecules oscillate back and forth along the path of the sound wave. In the case of the waves that travel along a horizontal rope being shaken at one end, it is the rope itself that moves as its various parts oscillate at right angles to the direction of the travelling wave.

Electromagnetic waves are rather different; nothing *moves* as such, but at every point along the path of the wave there will be an oscillating electric field \boldsymbol{E} and an oscillating magnetic field \boldsymbol{B} . Both of these fields are *vector* quantities, so each has its own <u>magnitude</u> and direction that may change from place to place and from one moment to the next. It is these magnitudes and directions that vary during the passage of an electromagnetic wave, just as the air pressure at a point varies during the passage of a sound wave.



An instantaneous snapshot showing the varying magnitude and direction of the electric field E and magnetic field B at points along the path of the simplest kind of electromagnetic wave.

The oscillating fields in an electromagnetic wave always have to be related in a very particular way. This is indicated in Figure 1 which shows a snapshot of the simplest kind of electromagnetic wave at a particular instant of time. There are several points to notice about this wave:



An instantaneous snapshot showing the varying magnitude and direction of the electric field E and magnetic field B at points along the path of the simplest kind of electromagnetic wave.

• The wave is travelling along a fixed direction known as the <u>direction of propagation</u>. In this case we have chosen to make it the *z*-axis of a system of Cartesian coordinates.

• The wave is travelling at the <u>speed of light</u>. In a vacuum this is defined to be 299792458 m s⁻¹. \leq It is customary to represent this speed by *c* and it is worth noting that to three significant figures

 $c = 3.00 \times 10^8 \,\mathrm{m\,s^{-1}}$ (in a vacuum)



An instantaneous snapshot showing the varying magnitude and direction of the electric field E and magnetic field B at points along the path of the simplest kind of electromagnetic wave.

• At every point on the z-axis the electric and magnetic fields are at right angles to the direction of propagation—electromagnetic waves are therefore said to be <u>transverse waves</u>.



An instantaneous snapshot showing the varying magnitude and direction of the electric field E and magnetic field B at points along the path of the simplest kind of electromagnetic wave.

• At every point on the z-axis the electric and magnetic fields are at right angles to one another. In this case we have chosen to orientate the axes so that the electric and magnetic fields oscillate along the x and y-axes, respectively.



An instantaneous snapshot showing the varying magnitude and direction of the electric field E and magnetic field B at points along the path of the simplest kind of electromagnetic wave.

If you look along the direction of propagation (i.e. the *z*-axis in Figure 1), a clockwise rotation of 90° is needed to go from the electric field direction to the magnetic field direction at any instant of time.



An instantaneous snapshot showing the varying magnitude and direction of the electric field E and magnetic field B at points along the path of the simplest kind of electromagnetic wave.

• The electric and magnetic fields vary <u>sinusoidally</u> and they do so together (i.e. they are <u>in phase</u>). A characteristic property of each field is its maximum magnitude. This is called its <u>amplitude</u> and may be denoted by E_0 and B_0 for the electric and magnetic fields, respectively.



An instantaneous snapshot showing the varying magnitude and direction of the electric field E and magnetic field B at points along the path of the simplest kind of electromagnetic wave.

Although it is not obvious from the figure, the ratio of the field magnitudes, E/B at any point, is fixed and in a vacuum is equal to the speed of light, so

 $E/B = E_0/B_0 = c$ (in a vacuum)



An instantaneous snapshot showing the varying magnitude and direction of the electric field E and magnetic field B at points along the path of the simplest kind of electromagnetic wave.

• Both the electric and the magnetic fields are characterized by a common wavelength λ which is equal to the distance between corresponding parts of the wave (e.g. from one peak to the next).



Since the wave has a particular wavelength and moves with a known speed the number of wavelengths that 0 will pass a fixed point in one second is easy to calculate. That number per second is called the frequency fof the wave and is given by $f = c/\lambda$. Thus, in a vacuum

 $c = f\lambda$ (in a vacuum)

Figure 1



• Since *f* determines the number of wavelengths that pass a fixed point in one second, the time required for *one* wavelength to pass a fixed point will be 1/f. This time is called the **period** *T* of the wave and is given by

T = 1/f

• (a) The SI unit of frequency is the hertz (Hz), where $1 \text{ Hz} = 1 \text{ s}^{-1}$. What is the SI unit of period *T*?

(b) The SI unit of electric field strength (i.e. magnitude) is the volt per metre $(V m^{-1})$ where $1 V m^{-1} = 1 \text{ kg m s}^{-2} C^{-1}$, and the SI unit of magnetic field strength is the tesla (T), where $1 T = 1 \text{ kg s}^{-1} C^{-1}$. Confirm that the ratio of electric and magnetic field amplitudes, E_0/B_0 , can be measured in the same units as the speed of light in a vacuum, *c*, as implied above.



Electromagnetic waves of different wavelength differ markedly in their properties. Collectively they make up the <u>electromagnetic</u>

spectrum shown in Figure 2a, of which visible light is just one small segment.



Figure 2a The electromagnetic spectrum.

Those parts of the spectrum with shorter wavelengths than visible light are notorious for their powers of penetration and their potential for causing biological damage. They include the gamma rays associated with nuclear phenomena, the X-rays used in hospitals and the ultraviolet radiation associated with a good sun-tan or a bad skin cancer.





On the other side of visible radiation, at the long wavelength end of the spectrum, are the more benign regions of *infrared* and *microwave* radiation that assist us in cooking, and the *radiowaves* that provide us with entertainment and worldwide communications.



Figure 2a The electromagnetic spectrum.

Notice that the wavelength and frequency scales in <u>Figure 2a</u> are <u>logarithmic</u> (i.e. they involve <u>powers</u> of ten) so the range covered by visible light (shown expanded in Figure 2b) really is very narrow indeed.

The human eye senses different frequencies of light as different colours. We usually define the colours by the corresponding wavelengths in vacuum and the sensible unit to use is the *nanometre* $(1 \text{ nm} = 10^{-9} \text{ m})$. Light in the range 400–500 nm is usually perceived as 'blue', 500–560 nm as 'green', 560–600 nm as 'yellow' and 600–750 nm as 'red'. You may be aware that the eye is most sensitive to small wavelength changes as colour differences in the centre of the visible range, at around 550 nm; the trained eye can see a colour difference corresponding to a 10 nm change at around 550 nm but few can do this at around 650 nm. Evolution has equipped us with an eye which is perfectly tuned to sunlight!

Question T1

Hertz measured the frequency of his electromagnetic waves to be 100 MHz. Calculate the wavelength of the radiation he generated. \Box



Figure 2b The part of the electromagnetic spectrum corresponding to visible light.



2.2 Polarization

For any electromagnetic wave the E and B vectors at any point must always be mutually perpendicular and must always lie in a plane which is perpendicular to the direction of propagation. Figure 1 illustrated a simple example in which the electric field vector E was confined to oscillate only in the x-direction and the magnetic field vector B only in the y-direction, but more complicated arrangements are also possible.



For instance, it is easy to imagine a wave travelling in the z-direction in which the E and B vectors are mutually perpendicular and confined to the (x, y) plane, but in which their orientation within that plane changes continuously along the z-axis. In such a wave the electric and magnetic fields would wind around the direction of propagation and, in contrast to Figure 1, there would be no unique direction associated with either field.

An electromagnetic wave that does have a unique direction associated with its electric field vector is said to be **linearly polarized** or **plane polarized**. For such a wave:

The **plane of polarization** is the plane that contains the direction of the electric field vector and the direction of propagation.



An instantaneous snapshot showing the varying magnitude and direction of the electric field E and magnetic field B at points along the path of the simplest kind of electromagnetic wave.

• What is the plane of polarization of the wave in Figure 1?



Linearly polarized light is very special and rather artificial. Light from ordinary sources is more likely to be **unpolarized**; this means that the direction of the electric field within the plane perpendicular to the direction of propagation varies rapidly and unpredictably (usually over a time interval of the order of 10^{-9} s), and in such a way that on average there is no preferred direction of oscillation for the electric field. Processes such as reflection can cause unpolarized light to become polarized, so much of the light we see in everyday life is actually **partially polarized** — there is a preferred direction of oscillation for *E* but it is not as pronounced as it would be in the case of linearly polarized light. However, in *any* light wave, whatever its state of polarization, the instantaneous electric and magnetic fields are always perpendicular to each other and to the direction of propagation.

It is possible to produce linearly polarized light by passing a beam of unpolarized light through a **polarizing filter** or **polaroid**. These filters are often made by aligning certain long-chain molecules in one direction within a plastic sheet. This arrangement will preferentially absorb those light waves in which the electric field is oscillating parallel to the axis of alignment. Light emerging from a polarizing filter will then be polarized perpendicular to the aligned molecules. In polaroid sunglasses such filters are mounted so as to absorb light which is horizontally polarized; the light reflected from horizontal surfaces is partially polarized in this direction so polaroid sunglasses can reduce the glare from roads and swimming pools by preferentially removing this component of the light.

Question T2

A certain beam of partially polarized light is composed of equal amounts of vertically polarized light (with electric field amplitude E_0) and unpolarized light. In what direction is the beam polarized and what will be the average electric field measured over a period of a few seconds? (Think carefully!)



2.3 Light in materials

Now we will consider the nature of light in different (transparent) materials.

When an electromagnetic wave travels through a transparent material its speed $\leq \\$ is always less than c (the speed in a vacuum). For light travelling through air the difference is only 0.03% but the speed of light in glass or water is about two-thirds or three-quarters of c, respectively. We define the ratio of the speed of light in a vacuum to the speed of light in a material to be the <u>refractive index</u> of the material. The refractive index is usually represented by μ (the Greek letter 'mu'), so

 $\mu = \frac{\text{speed of light in a vacuum}}{\text{speed of light in the material}}$

(1)

The speed of visible light and the corresponding refractive index for various common materials is given in Table 1.

For any given material, μ is found to vary with the wavelength of the light; different wavelengths travel at different speeds. Thus, the refractive indices in Table 1 are really averages taken over a range of wavelengths, though the variations are rather small. Other optical properties of materials vary even more strikingly with wavelength. Some materials are transparent at one wavelength and opaque at another.

Table 1The speed of light in variousmaterials.

Material	$\frac{Speed}{/10^8ms^{-1}}$	Refractive index
air	3.00	1.00*
water	2.26	1.33
glass†	1.6 to 2.0	1.9 to 1.5
diamond	1.24	2.42

* 1.0003 is a more accurate value for air.

† For a range of glass types.

When light passes from one material into another, its speed usually changes. Since the speed is always given by the product of the wavelength and the frequency, it is interesting to ask whether it is the wavelength or the frequency which changes, or do both change?

To answer this, consider what happens at the boundary of the two materials. The fields on both sides of this boundary must oscillate in the same way, so they must have a common *frequency*. The wavelength and speed of light may vary across the boundary, but the frequency must remain the same.

$$\lambda_{\text{material}} = \frac{\lambda_{\text{vacuum}}}{\mu} \tag{2}$$

Question T3

The refractive index μ of a glass block is 1.50 for light of frequency 6.00×10^{14} Hz. Find the wavelength of the light in a vacuum and inside the block.


3 Propagation and Huygens' principle

3.1 The propagation of an unrestricted wavefront

The electromagnetic wave model introduced in <u>Section 2</u> can provide a complete account of the propagation of light that includes details of its behaviour at surfaces which reflect or transmit the light that falls upon them. However, the full theory is mathematically complicated and rather cumbersome. Fortunately it is possible to obtain a good deal of insight into the propagation of light by means of a much simpler approach introduced in the 17th century by the Dutch physicist Christiaan Huygens.

In its modern version, Huygens' approach makes use of *wavefronts*. In a region occupied by waves, a <u>wavefront</u> is a line (in two dimensions) or a surface (in three dimensions) that only passes through points at which the wave is at the same stage (or <u>phase</u>) in its oscillatory cycle. For instance, in the case of the waves spreading out from a point of disturbance on the surface of a pond, the wavefronts would be concentric circles that might, for example, run along the crests (or troughs) of the waves at some instant. These circular wavefronts would expand outwards, along with the waves until some obstruction was encountered. In the absence of such restrictions the wavefront through any point will be perpendicular to the direction of propagation of the wave at that point.

Note In this module we will often use lines or curves to represent wavefront surfaces in three dimensions. We can safely do this when there is an obvious symmetry in the situation. For example, in the case of waves expanding outwards uniformly from a point in three-dimensional space, the wavefronts will be spherical surfaces centred on the source point but, thanks to the spherical symmetry, they can safely be indicated diagramatically by circles drawn around the source point. Also note that when using wavefronts it is usual to draw them at intervals of one wavelength. Thus, if one wavefront passes only through points at which the wave is reaching its maximum, the next wavefront will normally have the same property.

Huygens realized that it is possible to predict how a given wavefront will advance by means of geometrical constructions similar to those of Figure 3. These constructions are based on the following principle:

Huygens' principle

Each point on a wavefront may be treated as a source of **secondary** wavelets that expand radially from their source with the same speed as the original wave.

In practice it is only necessary to select as sources a set of points that are separated by about one wavelength. After a short period of time (typically about the period of the wave) the new wavefront can be constructed by drawing a smooth line that just touches each wavelet.

Figure 3 shows how this works for a wave expanding from a **point source** in three dimensions when the wavefront is unrestricted. In Figure 3a we are looking near the source, and the time interval for the construction is the wave period, so a circular (i.e. spherical) wavefront creates a new circular (i.e. spherical) wavefront that has expanded by one wavelength.



Figure 3 The Huygens' constructions for an unrestricted wavefront (shown in two dimensions) in the case of (a) a spherical wavefront and (b) a plane wavefront.

In Figure 3b we are looking at the situation far away from the source; now the radii of the wavefronts are so large that any small part of the wavefront will appear to be a flat plane and will give rise to other planes as it advances.

Note A point source is ideally one of zero size. A real source acts as a point source when its size is less than a wavelength of light. In practice a source can often be treated as a point source if the distance to the source is very large compared to the size of the source.

Thus, Figure 3a depicts an expanding <u>spherical wave</u> with <u>spherical</u> <u>wavefronts</u> and Figure 3b a <u>plane wave</u> with <u>plane wavefronts</u>. At sufficiently large distances from a point source any expanding spherical wave approximates a plane wave, as the radius of curvature of the wavefronts become very large.



Figure 3 The Huygens' constructions for an unrestricted wavefront (shown in two dimensions) in the case of (a) a spherical wavefront and (b) a plane wavefront.

3.2 The propagation of a restricted wavefront — diffraction 塗

Figure 4 shows a plane wavefront propagating to the right, through an aperture. In this case the aperture is a slit of width *w*, where *w* happens to be five times the wavelength of the light. This aperture represents a significant restriction to the passage of the wavefront since its width is comparable to the wavelength. In using Huygens' construction to determine the form of the wavefront to the right of the slit we must include only those secondary wavelets originating from points within the aperture since all the others are blocked off by the slit walls. The striking feature of the resulting wavefront, shown in Figure 4, is that *near the edges of the slit the wavefronts are curved, corresponding to waves moving away from the original direction of propagation.* The curvature arises as a result of the missing contributions which are now blocked off by the walls of the slit. The inescapable conclusion is that the beam, which initially was travelling parallel to the axis, has been spread out by the restriction imposed by the slit. This spreading effect is known as <u>diffraction</u>.

Note You may be wondering why, in the Huygens' construction, we only consider the expanding secondary wavelets travelling in the direction of the original wave, rather than also in the opposite direction. If so, you deserve congratulations! This is one of the flaws in the Huygens' approach and it needs Maxwell's theory to resolve it.



Figure 4 The Huygens' construction for diffraction at an aperture of width equal to five wavelengths.

The extent to which light is diffracted by a slit depends on the size of the slit. If the slit width is very large compared to the wavelength then the diffraction effect is rather insignificant, unless we look very close to the shadow region near the edge of the slit. On the other hand, if the slit width is made narrower than five wavelengths, the diffraction becomes more significant. If the slit width were less than one wavelength then it would act essentially as a line of point sources with circular wavefronts beyond the slit, corresponding to the beam having been spread out uniformly over all angles. The diffraction of light by narrow apertures is a striking phenomenon and provides a strong indication that light is a wave phenomenon.

Home experiment If you have access to a pair of binoculars, mask one of the large front lenses using card to produce a narrow slit about 1 mm wide. Look through the eyepiece and focus on a distant street lamp at night. You should see a wider image of the slit and several subsidiary images — the former effect is clear evidence of diffraction, the latter is explained in Subsection 5.6.

Diffraction has been described as a consequence of *restricting* an infinite wavefront by interposing a narrow *aperture*. Alternatively we could have demonstrated diffraction by placing a narrow *obstacle* in front of the infinite wavefront. Either a hair or a narrow slit is an equally suitable object with which to produce diffraction from an infinite wavefront.

Diffraction is the spreading of a propagating wavefront away from its original direction of propagation that occurs whenever the wavefront encounters an aperture or an obstacle. The effects are especially significant when they are caused by an object whose size is comparable with the wavelength of light.

• In terms of Huygens' principle the diffraction of plane waves by a narrow slit can be attributed to the absence of the secondary wavelets from points outside the slit. How can you explain the diffraction of light by a hair?

3.3 Propagating wavefronts and rays of light

Before going on to discuss *reflection* and *refraction* we will introduce the convenient concept of a **ray** of light. Light rays are directed lines (i.e. they have arrowheads on them) used to show the direction in which the light is propagating. They are drawn at right angles to propagating wavefronts. In our example of the spherical wave the light rays are straight lines spreading radially from the source, and in the case of the plane wave they are straight parallel lines, perpendicular to the wavefronts.

In any uniform medium, rays travel in straight lines and it is very convenient to use rays to track the passage of light through a system — but we must remember that we are dealing with a wave phenomenon, that the rays represent propagating wavefronts and that diffraction will spread the beam when restrictions (apertures or obstacles) are involved. Apertures are *always* involved to some extent since we can never work with an infinite wavefront! Therefore, we should really draw our rays not as sharp lines but as 'bands' of increasing width, as the wavefronts spread out. In fact this is never done on ray diagrams and the convention is rather that ray diagrams are not used where diffraction effects are significant. Fortunately there are many cases where the apertures involved are much greater than the wavelength, so the diffraction effects are insignificant and we can safely represent the passage of light by rays. The range of topics that can be adequately treated by methods based on the use of rays constitutes the field of geometrical optics, and the territory where rays are inadequate and a fuller version of the wave model must be used constitutes physical optics.

3.4 The reflection of light

Huygens' principle can easily explain the behaviour of light when it is reflected from a plane mirror. A plane mirror is a flat reflecting surface, on which any irregularities are much less than a wavelength of light. Such mirrors are often made by depositing a silver film on to a flat glass surface. In this situation we can represent light arriving at the mirror from some particular direction by means of an incoming ray, called the **incident ray**. The direction of this ray can be specified by defining a line called the **normal** that is perpendicular to the mirror's surface at the point of incidence, and assigning a value to the angle of **incidence** θ_i between this normal and the incident ray. Similarly, the direction of the light reflected from the point of incidence can be indicated by a <u>reflected ray</u> and specified by the <u>angle of reflection</u> $\theta_{\rm R}$ between the normal and the reflected ray. These angles are shown in Figure 6.



Figure 6 Experimental observation of the reflection and refraction of a light ray at the boundary between two transparent materials.

Experiments support the following:

Law of reflection

- The *reflected ray*, the *incident ray* and the *normal* all lie in the same plane.
- The angle of reflection is always equal to the angle of incidence: $\theta_{R} = \theta_{i}$



Figure 6 Experimental observation of the reflection and refraction of a light ray at the boundary between two transparent materials.

The Huygens' construction that accounts for this is shown in Figure 5. The line ABC is part of an extensive plane wavefront (where AC $\gg \lambda$), moving in the direction shown by the arrows and beginning to arrive at the plane mirror at an angle of incidence θ_i . Point A on the wavefront is just arriving at the mirror. Points on the wavefront corresponding to B and C will arrive later, at points N and C' on the mirror. Secondary wavelets produced at the same time from points A, B and C will expand outwards from these points, all travelling at the same speed. In the time taken for the wavelet produced at C to reach the mirror at point C', the wavelet from A will have covered an equal distance and will be arriving at A' while that from B will be arriving at B' having also covered the same distance (BN + NB').



Figure 5 The reflection of a wavefront. Note that the angles marked θ_i and θ_R are not themselves the angles of incidence and reflection, but they are equal to those angles.

The resulting wavefront, after reflection at the mirror, will be the line A'B'C' that is tangential to each of the wavelets. Now, the right-angled triangles ACC' and AA'C' are identical, so the lines ABC and A'B'C' make the same angle with the mirror surface. Since the light rays are perpendicular to the wavefronts it follows that these angles are also equal to the angles of incidence θ_i and reflection θ_R , respectively, so these two angles are equal — in accord with observation.

Question T4

Using a ruler, compasses and protractor construct a diagram similar to Figure 5 but with the angle $\hat{CAC'} = 20^{\circ}$.

What is the angle of reflection on your figure? \Box



Figure 5 The reflection of a wavefront. Note that the angles marked θ_i and θ_R are not themselves the angles of incidence and reflection, but they are equal to those angles.

3.5 The refraction of light

If you place a pencil into a glass of water, you will observe that the pencil appears to be bent at the surface of the water. This effect is due to the <u>refraction</u> of light rays at the air-water surface. In general, when a light ray passes from one transparent material into another there will be both a reflected ray and a <u>refracted ray</u> which passes into the second material, having been bent or refracted at the surface. This is illustrated in Figure 6. The direction of the refracted ray can be specified by the <u>angle of refraction</u> θ_r between the refracted ray and the normal drawn in the second material.



Figure 6 Experimental observation of the reflection and refraction of a light ray at the boundary between two transparent materials.

Experiments support the following:

Law of refraction

- The *incident ray*, the *refracted ray* and the *normal* all lie in the same plane.
- The *angle of incidence* and the *angle of refraction* are related by

$$\frac{\sin \theta_{\rm i}}{\sin \theta_{\rm r}} = \text{constant} \qquad (3)$$

where the constant depends on the two materials involved and on the wavelength of the light.



Figure 6 Experimental observation of the reflection and refraction of a light ray at the boundary between two transparent materials.

We can use the Huygens' construction to explain these observations. The argument follows closely along the lines used in <u>Subsection 3.4</u>.

In Figure 7 ABC is part of an extensive plane wavefront (where AC >> λ), moving in the direction shown by the arrows and beginning to arrive at the boundary between the two materials. It is now possible for light to travel in either material, though if the materials have different *refractive indices*, μ_1 and μ_2 say, the light must change speed as it moves from one material to the other. Let us suppose that in material 1 the speed of light is $v_1 = c/\mu_1$ and that in material 2 it is $v_2 = c/\mu_2$. Now, consider what happens to secondary wavelets emitted simultaneously from points A, B and C on the incident wavefront.



Figure 7 The refraction of a wavefront (reflection not shown). Note that the angles marked θ_i and θ_r are not themselves the angles of incidence and refraction, but they are equal to those angles.

In the time taken for the wavelet from C to travel at speed v_1 through material 1 to the point C', the wavelet from A will have travelled at speed v_2 through material 2 to A' and the wavelet from B will have travelled first with speed v_1 to B" and then at speed v_2 to B'. (Figure 7 shows this for the situation where $v_2 < v_1$, (i.e. $\mu_2 > \mu_1$) but we could equally well have dealt with the case where $v_2 > v_1$.) As usual the propagated wavefront is tangential to the secondary wavelets, so in this case it is the line A'B'C'. As you can see this new wavefront is not parallel to the original wavefront ABC—the waves have been refracted. We can find the relationship between the angles θ_i and θ_r by equating the time required for the wavelet in material 1 to cover the distance CC' with the time required for the wavelet in material 2 to cover the distance AA'.



Figure 7 The refraction of a wavefront (reflection not shown). Note that the angles marked θ_i and θ_r are not themselves the angles of incidence and refraction, but they are equal to those angles.

recalling that the wavelets travel at different speeds we obtain

The triangles ACC' and AA'C' are both rightangle triangles, where the angle $C\hat{A}C' = \theta_i$ and $A'\hat{C}'A = \theta_r$, so we can use the trigonometric relations

$$\sin \theta_{i} = CC'/AC' \quad \sin \theta_{r} = AA'/AC' \quad \text{and} \\ \sin \theta_{i} / \sin \theta_{r} = CC'/AA' \quad \text{to give}$$



Figure 7 The refraction of a wavefront (reflection not shown). Note that the angles marked θ_i and θ_r are not themselves the angles of incidence and refraction, but they are equal to those angles.

$\sin \theta_{i}$	$\underline{\mu}_2$
$\sin \theta_{\rm r}$	$-\mu_1$

(4)

$\frac{\sin \theta_{\rm i}}{\sin \theta_{\rm r}} = \frac{\mu_2}{\mu_1}$	(Eqn 4)
Equation 4 is now known as Snell's law . It shows that the constant in Equation 3	
$\frac{\sin \theta_{\rm i}}{\sin \theta_{\rm r}} = {\rm constant}$	(Eqn 3)

is just the ratio of the refractive indices of the two materials and thus explains why it depends on the materials themselves and on the wavelength of the light. This association of μ with refraction also explains why μ is called the refractive index.

Question T5

Using a ruler, compasses and protractor, construct a diagram similar to Figure 7 but with $\theta_i = 20^\circ$ (take $\mu_1 = 1.0$ and $\mu_2 = 1.5$). Find θ_r and compare it with the theoretical prediction from Snell's law.



Figure 7 The refraction of a wavefront (reflection not shown). Note that the angles marked θ_i and θ_r are not themselves the angles of incidence and refraction, but they are equal to those angles.

It is interesting to contrast the account of refraction given by Huygens' wave model of light with that of Newton's particle model. The particle model had to claim that particles of light were 'attracted' by the higher refractive index material so as to bend them towards the normal. The mechanism of this attraction was unknown but the consequence of the attraction would be an *increase* in the speed of light in the higher refractive index material. The wave model made exactly the opposite prediction, that the speed of light is *reduced* in the higher refractive index material. It is difficult to measure the speed of light and this issue was not settled until the mid-19th century, when the French physicist Jean Foucault (1819–1868) made a precise measurement of the speed of light in a vacuum and also in various transparent materials (see <u>Table 1</u>). This confirmed the prediction of the wave theory.

The variation of refractive index with wavelength leads to the phenomenon of **dispersion**, that is the variation of the speed of light in a transparent material with wavelength. Its occurrence implies that the angle through which a given incident ray is refracted by a particular material will generally depend on the wavelength. This effect was first used systematically by Isaac Newton when he employed a triangular glass *prism* to study the spectrum of colours which make up white light from the Sun. Sometimes the term *dispersion* is specifically used to describe the separation of a beam of light into its constituent wavelengths when it is refracted by a material whose refractive index depends on wavelength.

4 Illumination and the inverse square law

4.1 Intensity

As the warmth of sunlight proves, light can carry energy through space. The general study of the transfer of energy by light is very complicated; in this section we describe some simple cases and discuss the way in which the wave model accounts for the phenomenon of energy transfer by light.

Figure 8 shows a plane drawn perpendicular to an idealized uniform parallel beam of light. If the *rate* at which the beam transfers energy across the area A is denoted by P (for *power*) \leq then the <u>intensity</u> of the beam passing through the plane is given by



Figure 8 An idealized uniform parallel beam of light passing through area A of a plane perpendicular to the direction of propagation.

I = P/A

(5)

According to the electromagnetic wave model of light, a uniform light beam of the kind shown in Figure 8 may be represented by a <u>uniform plane wave</u>, that is a <u>plane wave</u> in which the electric field has the same instantaneous magnitude E at all points on any given wavefront.

Remember, at any given point on the path of an electromagnetic wave the magnitude (and direction) of the electric field changes with time, varying between zero and E_0 — the <u>amplitude</u> of the electric field. Now, according to the theory of electromagnetic waves, a uniform plane wave transfers energy in the direction of propagation at a rate that is *proportional to the square of its electric field amplitude*. Thus we may define the intensity of such a wave by

 $I = kE_0^2$

where k is an appropriate constant of proportionality. The form of this relationship should not come as too much of a surprise to you since the energy of oscillating systems is often proportional to the square of the amplitude.

• If the above definition of the intensity of a uniform plane wave is to be consistent with the definition of the intensity of a uniform beam of light, what are suitable SI units for k?



While it is convenient to define intensity in terms of the idealized situation shown in Figure 8, in all real cases a beam is never perfectly parallel nor perfectly uniform; it always diverges to some extent, and its intensity varies over its cross section. An important practical light source that at least approximates the idealized case is the laser. The beam from a laser is very nearly parallel, though its intensity is not uniform over the cross section. The finite width of the beam implies that it must spread a little due to diffraction since it behaves as though diffracted by an aperture of width equal to the diameter of the beam the narrower the beam the more it spreads. A small helium-neon laser (which has a red beam) might typically radiate 1 mW of power in a 1 mm diameter beam, the edges of which diverge from one another at an angle of only one milliradian.



Figure 8 An idealized uniform parallel beam of light passing through area *A* of a plane perpendicular to the direction of propagation.

Question T6

Use the figures given above to calculate the average intensity of this laser (a) at the laser itself and (b) at a distance of 100 m from the laser. In (b) you may assume that the laser acts as a point source, but with the full divergence given above. \Box



4.2 The inverse square law of illumination

Real light sources produce a diverging beam for which the intensity (i.e. the rate at which energy is transferred across a unit area perpendicular to the direction of propagation) must decrease with distance from the source. It is often important to know the intensity of a beam at its point of arrival but to predict this we need to know *how* the intensity decreases with distance from the source. For the idealized case of a perfectly parallel beam the answer is that the intensity is *independent* of the distance, but such a case never occurs in practice — even the laser has an intensity which decreases with distance, as you saw in Question T6.

Fortunately, we can predict how the intensity will decrease with distance for the idealized case of a point source radiating uniformly in all directions, producing spherical waves and spherical wavefronts. This is often a very good approximation to real sources, particularly at distances much greater than the size of the source.

Consider a point source (located in a vacuum) that consumes power P and radiates all of that power, uniformly in all directions, at visible wavelengths. Suppose that the source is at the centre of an imaginary sphere of radius r. The rate at which energy flows across the whole imaginary spherical surface must also be P and since this power is spread uniformly across the surface each part of the sphere will receive the same power per unit area.

This power per unit area on the surface is the *intensity* at the surface and, since the surface area of the sphere is $4\pi r^2$, it must be given by

$$I(r) = \frac{P}{4\pi r^2} \tag{6}$$

Equation 6 shows that the intensity due to a uniformly radiating point source decreases as the inverse square of the distance from the source. This is known as the **inverse square law of illumination**. In deriving this we have relied only on the principle of *energy conservation* and on the geometry of a sphere. The inverse square law also holds in the case of a *spherical* source of any size, providing it radiates uniformly in all directions and provided the distance r (measured from the centre of the sphere) is much larger than the radius of the source.

Question T7

On a clear dark night an average person can *just* see the light from a 40 W tungsten bulb at a distance of 30 km. Estimate the minimum intensity that the human eye can detect (remember the inefficiency of the tungsten bulb). Assuming the aperture (*pupil*) of the eye has radius 2 mm calculate the minimum power that the eye can detect. In terms of power suggest why a *telescope* increases the range of visibility. \Box



Question T8

Taking the intensity of sunlight as 1.4 kW m^{-2} at the Earth's orbital distance $(1.5 \times 10^{11} \text{ m})$, estimate the total power transferred into space from the Sun by sunlight. \Box

As a consequence of geometry and conservation, inverse square laws arise in many areas other than illumination. For example, <u>gamma radiation</u> from small <u>radioactive</u> sources is emitted uniformly in all directions so the intensity of the radiation decreases as the inverse square of the distance from the source. A person standing at a distance of 10 m from a source will receive a radiation dose 1% of that received by a person 1 m away, over the same time interval. (This calculation neglects the small absorption of gamma radiation by the air.) The inverse square law is the best and most reliable radiation shield we have!

5 Interference and the superposition principle

In <u>Subsection 3.2</u> we described *diffraction* and explained that this strongly indicates that light travels as a wave. In this section we describe <u>interference</u>, which provides even clearer evidence of wave-like behaviour.

If two beams of light of equal intensity come together at a point in space, what will be the intensity at the intersection point? If the beams are thought of as streams of particles, each carrying its own small share of the energy, the expected answer is obvious. The intensity must be twice the intensity of each beam alone, since the total number of particles will be equal to the sum of the numbers in each beam — it's as easy as '1 + 1 = 2'. However, experimentally it is found that under the right circumstances the resulting light intensity can be anything from zero to four times the intensity of a single beam — the outcome depends critically on the source of the two beams and is not determined by their intensities alone. This extraordinary phenomenon, first demonstrated by Thomas Young in 1801, is easily explained by a wave model of light but is almost impossible to account for in terms of particles. Its discovery provided convincing evidence that light was a wave phenomenon even though many decades were to pass before Maxwell uncovered the electromagnetic nature of those waves.

5.1 The superposition principle

A property which is common to all waves, and which can be used to find their combined effect, is enshrined in the following principle.

The superposition principle

If two or more waves meet in a region of space, then at each instant of time the net disturbance they cause at

any point is given by the sum of the disturbances caused by each of the waves individually. $\$

By 'disturbance' we mean the change in whatever physical quantity is marking the passage of the wave. For water waves it would be the vertical displacement of the wave from the normal water level; for sound waves it would be the change in air pressure from its undisturbed value; for light waves it would be the change in the transverse electric and magnetic fields at each point. In general, if at time t two waves individually create disturbances $y_1(t)$ and $y_2(t)$ are a particular point then their superposition will create a disturbance $y(t) = y_1(t) + y_2(t)$ at that point. This is illustrated in Figure 9a where two waves of equal amplitude and wavelength (shown by differently dashed curves), both propagating in the x-direction, are combined to form a resultant (shown by a solid line). In Figure 9a the waves being combined are exactly in step (or in phase) and, in accordance with the superposition principle, their resultant causes twice the disturbance of the individual waves. Note in particular that the *amplitude* y_{max} of the resultant is twice the amplitude of each of the original waves.



Figure 9a Superposition of two waves with both waves in step (zero phase difference or *in phase*)

Figure 9b shows what happens when the waves are somewhat out of step, the superposition principle still applies but the resulting wave has an amplitude that is less than twice the original amplitude.

Figure 9c shows the extreme case in which the two superposed waves are causing exactly opposite disturbances at each point; their effects now cancel one another completely so the resultant has zero amplitude.

Figure 9b/c Superposition of two waves (b) with the waves somewhat out of step (an intermediate phase difference), and (c) with the waves totally out of step (a phase difference of π radians or 180° or in *anti-phase*).



The key point to note in this case is that the resultant in each part of Figure 9 has an amplitude that depends on the extent to which the waves being combined are out of *step*. This latter concept can be expressed quantitatively in terms of the **phase difference** ϕ between the waves. Phase difference is usually quoted as an *angle*, either in degrees (0 to 360) or radians (0 to 2π), and indicates the fraction of a wavelength (or period) by which the waves are out of step. For example, if at some particular instant two waves of identical wavelength that occupy the same region of space both cause their maximum disturbance at the same place (as in Figure 9a) then the phase difference between them is zero and we say they are **in phase**.



Figure 9a Superposition of two waves (a) with both waves in step (zero phase difference or *in phase*).

If on the other hand the peaks of one wave are separated from those of the other wave by half a wavelength (as in Figure 9c) the phase difference is π radians or 180° and the waves are said to be totally <u>out of phase</u> or <u>in antiphase</u>. Any other fraction of a wavelength separating the peaks would be described by the corresponding fraction of 2π radians or 360° .



Figure 9c Superposition of two waves (c) with the waves totally out of step (a phase difference of π radians or 180° or in *anti-phase*).

• Estimate the phase difference ϕ between the two waves being combined in Figure 9b.

When dealing with waves it is important to remember that the disturbances they cause vary from place to place and from one moment to the next. Thus, although we have just been discussing the phase difference between two waves at some particular *instant*, we might equally well have chosen to discuss the difference in phase between the oscillations that the waves caused at some particular *point* (or even at a pair of different points). If so, the phase difference between the waves would correspond to a fraction of a period rather than a fraction of a wavelength, but the essential idea would remain the same. It is hard to overemphasize the importance of realizing that waves involve changes in time and space and that along the path of a wave there is an oscillation taking place at every point.



Figure 9b Superposition of two waves with the waves somewhat out of step (an intermediate phase difference).

Mathematically, if at some particular point $\leq t$ the oscillations caused by two different waves with the same wavelength can be described by expressions of the form $y_1(t) = A \sin(\omega t)$ and $y_2(t) = B \sin(\omega t + \phi)$ where *A*, *B*, ω , and ϕ are constants and *t* represents time, then the phase difference between those waves is ϕ .

The term <u>constructive interference</u> is used to describe the condition in which two waves combine to produce a resultant with an amplitude which exceeds that of either of the original waves; fully constructive interference is when the waves are in phase and the amplitude of the resultant is the sum of the amplitudes of the combined waves. <u>Destructive interference</u> describes the condition in which two waves combine to produce a resultant with an amplitude which is less than that of either of the original waves; fully destructive interference is when the waves are totally out of phase and the amplitude of the resultant wave is the <u>modulus</u> of the difference in the amplitudes of the individual waves.

In all cases of interference, the *intensity* of the resultant is given by the square of its amplitude (see <u>Subsection 4.1</u>). It follows that this intensity is *not* generally equal to the sum of the intensities of the two individual waves.

An example illustrates this important point. If two waves, each of amplitude A interfere fully constructively (Figure 9a) the combined wave will have amplitude 2A and intensity $(2A)^2 = 4A^2$. Separately the two waves each have intensity A^2 , and adding the intensities would give just $2A^2$; this is almost a case of '1 + 1 = 4', or more informatively '(1 + 1)² = 2² = 4'.



Figure 9a Superposition of two waves with both waves in step (zero phase difference or *in phase*)

More striking still is the case of two waves, each of amplitude A, interfering fully destructively (Figure 9c). Here the superposition has both amplitude and intensity equal to zero. Thus, when dealing with energy transported by waves we have the possibility of two energy flows combining to give zero energy flow at one place and maximum energy flow at another!



Figure 9c Superposition of two waves with the waves totally out of step (a phase difference of π radians or 180° or in *anti-phase*).

Question T9

For the more general case, where the two waves of equal amplitude and wavelength interfere with a phase difference ϕ , the amplitude of the resultant is $2A \cos(\phi/2)$.

Prove this result by taking $y_1(t) = A \sin(\omega t)$ and $y_2(t) = A \sin(\omega t + \phi)$ and then showing that $y_1(t) + y_2(t) = 2A \cos(\phi/2) \sin(\omega t + \phi/2)$.


5.2 Young's two-slit experiment

We are now in a position to consider **Young's experiment**, performed first in 1801 by Thomas Young; it remains an excellent demonstration of interference as it highlights all the conditions necessary for interference effects to be observed **Construct**. The schematic layout is shown in Figure 10.



A narrow slit is placed in front of a source of light such as a sodium lamp (you will have seen such lamps as the yellow **monochromatic** street lamps) or an ordinary (white) incandescent lamp with a filter to select a single colour (i.e. wavelength) of light. Light from this slit illuminates a double slit which is placed a centimetre or so from the single slit. Each of the three slits is sufficiently narrow to act as a line of point sources or as a line source, with diffraction producing expanding cylindrical wavefronts (shown as circular in the two-dimensional plot) beyond the slits.



The two expanding sets of wavefronts from the double slit overlap on a screen, placed about a metre away, where they interfere. The next subsection discusses why this particular arrangement is adopted, but first let us consider the observations.

The **interference pattern** seen on the screen includes regions where the light intensity is high, signifying constructive interference, separated by regions where the intensity is low, where destructive interference *consection* cocurs. These light and dark regions take the form of linear bands, called **interference fringes** (see Figure 10).



We can use Huygens' construction and the principle of superposition to follow the light through the apparatus and to explain these observations. Each wavefront from the source slit arrives at S_1 and S_2 simultaneously so that these act as independent line sources emitting secondary wavelets that are in phase. The secondary wavelets expand as semicircles and the net wave disturbance at any point on the screen is calculated by adding the contributions from S_1 and S_2 . The intensity at that point is then calculated from the square of the amplitude of the resultant disturbance.



Consider Figure 10 and look first at what happens at position B_0 on the screen. Wavefronts leaving S_1 and S_2 simultaneously *always* arrive at B_0 at the same time so they are still in phase and there is fully constructive interference with an amplitude which is double that from a single slit; this produces a bright fringe with an intensity four times that from either slit alone.

Similar situations exist at points B_1 , B_{-1} , B_2 and B_{-2} where the two wavelets still arrive in phase despite the fact that the wave from one slit has travelled an additional distance of one or two complete wavelengths relative to that from the other slit.



The difference in the distances from each slit to any point on the screen is called the **path difference**.

An additional path difference of one or more complete wavelengths between the two wavelets arriving at B_1 , B_{-1} , B_2 and B_{-2} corresponds to an extra phase difference equal to some multiple of 2π , and this will not harm the fully constructive interference that occurs at each of those points. (This can also be seen in Figure 9a where it is impossible to tell whether one wave leads the other by zero, one, two or several whole wavelengths.)



If we ignore the tiny difference in amplitude arising from the slightly different distances from the slits, the intensity of the light at B_1 , B_{-1} , B_2 and B_{-2} will also be four times the intensity due to each slit acting alone.

Now consider what happens at positions D_1 , D_{-1} , D_2 and D_{-2} on the screen. D_1 and D_{-1} correspond to points on the screen for which there is a path difference from the slits of one-half wavelength and so the wavelets from each slit arrive totally out of phase at the screen, interfering destructively to give zero intensity and a dark fringe.



Dark fringes are also produced at D_2 and D_{-2} , where the path difference is one and a half wavelengths. At position C the situation is intermediate, where there is a partially constructive interference of the two waves.



It is straightforward to calculate the positions on the screen where fully constructive and fully destructive interference take place. In Figure 11 the path lengths from the two slits to a point P on the screen are shown. The distance *D* is much greater than the slit separation *d* so that the angle θ from the 'straight-through' direction to point P is very small. Under these conditions the two paths S₁P and S₂P are essentially parallel and are at the same small angle, θ , to the straight-through direction.

For fully constructive interference the path difference must be an integral number of wavelengths, $n\lambda$, where *n* is an integer, so

$$S_2P - S_1P = NS_2 = n\lambda$$
 with $n = 0, \pm 1, \pm 2, \pm 3, ...$

For fully destructive interference the path difference is

 $S_2P - S_1P = NS_2 = (n + 1/2)\lambda$ with $n = 0, \pm 1, \pm 2, \pm 3, \dots$

Look at the right-angled triangle S_1NS_2 . The angle $S_2\hat{S}_1N$ also equals θ so that $NS_2 = d \sin \theta$.



Figure 11 The path lengths from the slits to a distant screen.

The condition for fully constructive interference then becomes:

 $d\sin\theta = n\lambda \quad \text{with } n = 0, \pm 1, \pm 2, \pm 3, \dots \tag{7}$

and the condition for fully destructive interference becomes:

 $d\sin\theta = (n + 1/2)\lambda$ with $n = 0, \pm 1, \pm 2, \pm 3, ...$ (8)

From Figure 11 the distance OP is given by $OP = D \tan \theta$.

Since θ is small, sin $\theta \approx \tan \theta \approx \theta = OP/D$ and the condition for bright fringes is therefore

 $OP = D\theta = nD\lambda/d \quad \text{with } n = 0, \pm 1, \pm 2, \pm 3, \dots$

The condition for dark fringes is

OP = $(n + 1/2)D\lambda/d$ with $n = 0, \pm 1, \pm 2, \pm 3, ...$

The whole number *n* determines the **<u>order of interference</u>** for the bright fringes.



Figure 11 The path lengths from the slits to a distant screen.

Equation 7

 $d\sin\theta = n\lambda$ with $n = 0, \pm 1, \pm 2, \pm 3, \dots$

(Eqn 7)

has n = 0 for $\theta = 0$, the zero order or 'straight-through' intensity maximum. Similarly $n = \pm 1$ defines the first order maxima, $n = \pm 2$ the second order maxima, and so on.

In some situations it is necessary to take into account the refractive index, μ , of the material through which light passes. The conditions for constructive/destructive interference arises from the path difference, expressed in terms of the *wavelength in the material* and this will generally differ from the wavelength in a vacuum by a factor of the refractive index, as we saw in <u>Subsection 2.3</u>:

$$\lambda_{\text{material}} = \lambda_{\text{vacuum}} / \mu$$
 (Eqn 2)

Interference conditions in a material can be expressed in terms of the <u>optical path length</u> in the material, and the wavelength in a vacuum. The optical path length is defined as

optical path length = $\mu \times$ geometrical path length (9)

The next question should help to make clear the value of this concept.

Question T10

The space between the slits and the screen in a Young's two-slit experiment is filled with glass of refractive index μ . Show that the angles of the interference maxima are given by $\sin \theta = n\lambda/\mu d$.

5.3 Coherence — a condition necessary to observe interference

Our discussion of interference so far has ignored an important condition which is necessary to *observe* interference. This condition is known as <u>coherence</u> and it determines how *predictable* the phase difference is between two waves. The production of an observable interference fringe requires that the conditions for constructive or destructive interference (set by the phase relationships) should persist long enough in time to permit an observation, and extend far enough in space for the fringe to have some observable width. If the two interfering waves were both ideal plane waves with exactly the same wavelength or frequency then perfect coherence would be possible and knowledge of their phase difference at one point and time would ensure that the phase difference at any other point and time could be predicted — the waves would have perfect temporal coherence and spatial coherence. Unfortunately the reality for actual light sources are quite different.

- No source really produces light of a single frequency (i.e. <u>monochromatic light</u>) but rather a range of different frequencies so when we superimpose these the phase differences of the waves are no longer perfectly defined. The frequency of light is so high (about 5×10^{14} Hz) that even a tiny spread of frequencies produces a range of phase differences. In practical terms each frequency in the source produces its own independent set of fringes and when these are superimposed the visibility of all the fringes will be impaired since each set of intensity maxima will be at a different position. (There is an exception to this when *all* wavelengths give a *minimum* at the point of observation then a 'black fringe' will be produced.)
- No source is a true point source so when we collect light from different points to make a beam there will inevitably be different path lengths involved and consequently a range of phase differences. In practical terms each *point* on the source produces its own independent set of fringes and when these are superimposed the visibility of the fringes will again be impaired. Even for a monochromatic source the visibility of the fringes will be compromised by the size of the source.

Both of these problems contribute to the phase uncertainties in any interference experiment and the coherence is limited by the source available. A source with high coherence is said to be <u>coherent</u> and one with very low coherence is said to be <u>incoherent</u>. When two coherent waves add together they do so by the superposition principle, their displacements are added and the resultant amplitude must be squared to give the intensity. If the waves are incoherent then the two waves act independently and the resultant intensity is just the sum of the two separate intensities. The reason for this is that any interference condition holds only instantaneously and the average over time at any point becomes the sum of the separate intensities.

A point monochromatic source would give perfect coherence but such a situation is unattainable. Despite this, we can obtain reasonable fringes, as in Young's experiment, by using a coloured filter to isolate a narrow frequency range (providing adequate temporal coherence) and narrow slits to restrict the source size (providing adequate spatial coherence). In this process we gain coherence at the expense of light intensity and we must seek a compromise between these two to produce the most visible fringes.

5.4 Interference of light reflected from thin films

When light of a given wavelength from a point on a source falls on to a thin film the wavefronts reflected from the top and bottom surfaces of the film are coherent, so if these two sets of wavefronts are then brought together there will be interference. It is convenient to discuss this problem in terms of the ray representation, since we are not intending to restrict the wavefronts significantly and the ray paths give the optical path lengths, path differences and hence the phase differences. Figure 12 shows a ray of light of a given wavelength leaving a point P on a source and having an angle of incidence θ_i to a thin film of transparent material of thickness τ and refractive index μ . (If a white light source is used the discussion that follows can be applied to each separate wavelength in the light from P.)



Figure 12 Rays reflected from a thin film.

The ray arriving at the film is partially reflected at the top and bottom surfaces of the film to produce two parallel rays which can be made to converge on a screen by means of a <u>lens</u>. The lens of a human eye is shown to emphasize that the effects can be viewed directly. We will assume that the film has refractive index $\mu > 1.0$ and that elsewhere $\mu = 1.0$. A soap film in air would be an example.



Figure 12 Rays reflected from a thin film.

Figure 13 The details of the reflection and refraction at the thin film.

The details of the reflection and refraction are shown in Figure 13.

From Equation 9

optical path length = $\mu \times$ geometrical path length (Eqn 9)

the optical path difference x between the two rays reflected from the top and bottom of the film is

 $x = \mu(AB + BC) - AD$

If θ_r is the angle of refraction and τ is the film thickness, trigonometry gives

 $x = 2\mu\tau\cos\theta_{\rm r} \tag{10}$

You might expect that the condition for constructive interference, when the two waves are recombined by the lens, is $x = n\lambda$; but this is wrong! It turns out that the reflection of one of the rays from the upper surface (such as at A in Figure 13) where the refractive index increases along the direction of propagation introduces an additional phase difference of π . (The reflection from the lower surface, such as at B, where the refractive index is decreasing, doesn't produce any compensating phase difference.)



Figure 13 The details of the reflection and refraction at the thin film.

The condition for fully constructive interference when one reflection introduces the π phase difference is then

 $x = (n + 1/2)\lambda$ with n = 1, 2, 3, ...

and for fully destructive interference

 $x = n\lambda$ with n = 1, 2, 3, ...

Substitution in Equation 10

 $x = 2\mu\tau\cos\theta_{\rm r}$

(Eqn 10)

then gives:

The interference conditions when one reflection introduces the π phase difference:				
	$2\mu\tau\cos\theta_{\rm r} = (n+1/2)\lambda$	for constructive interference	(11)	
and	$2\mu\tau\cos\theta_{\rm r}=n\lambda$	for destructive interference	(12)	

If the film is illuminated by an extended source of monochromatic light then an interference pattern will be seen in the film with bright fringes wherever Equation 11 is satisfied.

In practice changes in film thickness τ are usually more significant than changes in the refraction angle θ_r , so the fringes follow contours of equal film thickness. If the monochromatic light source is replaced by an extended source of white light that contains many different wavelengths then what is seen will be the result of adding together many interference patterns and will consist of a pattern of coloured bands. This is why a soap film or an oil film on water appears coloured when viewed in daylight $\underline{\mathscr{D}}$.

Alternatively, if a *parallel beam of white light* is incident then, *at a fixed angle of view*, the reflected light appears white to the eye but closer examination reveals that it is deficient in those particular wavelengths which have suffered destructive interference from the film.

Question T11

Green light is reflected at normal incidence ($\theta_i = \theta_r = 90^\circ$) from a film of soap solution in air. Calculate the smallest thickness at which constructive interference takes place. (Assume $\lambda = 550 \text{ nm}$ and $\mu = 1.30$)



There are many important applications of these principles in optical technology. First, the process described above can be used to remove a particular wavelength by passing light through an <u>interference filter</u>, made by applying a thin transparent coating to a glass base. Second, <u>anti-reflection coatings</u> applied to the many air-glass surfaces in a multi-component camera lens can reduce the amount of light reflected at each of those surfaces to well below the usual 4% of the incident intensity. This coating material has a refractive index intermediate between glass and air; magnesium fluoride ($\mu = 1.38$) is a common choice because of its durability. The film thickness is set to give destructive interference for light reflected from the front and back surfaces of the coating and there is then minimum reflected light and maximum transmitted light. In this case the reflections at the upper and lower surfaces of each coating occur where the refractive index is increasing and so *each* reflection involves the π phase difference. Thus the relevant conditions are:

The interference conditions when both reflections (or neither) introduce the π phase difference:					
	$2\mu\tau\cos\theta_{\rm r}=n\lambda$	for constructive interference	(13)		
and	$2\mu\tau\cos\theta_{\rm r} = (n+1/2)\lambda$	for destructive interference	(14)		

Destructive interference at normal incidence (cos $\theta_r = 1$) then requires a minimum film thickness (n = 0) given by $\tau = \lambda/4\mu$.

Question T12

Calculate the minimum thickness of a magnesium fluoride anti-reflection coating required for a camera lens. (Assume λ is about 550 nm.)

Another application of thin film technology is the production of a *high reflectivity* mirror for a particular wavelength. This is done by coating glass with several layers and adjusting the thickness of each layer to give reinforcement of the front and back reflections. Mirrors of this type may have very many transparent layers, with alternating high and low refractive index, and can reflect more than 99% of the incident intensity at the selected wavelength — far more than can be achieved by a single coating of silver on glass (which gives about 80%).

5.5 The diffraction grating

The <u>diffraction grating</u> is a very important device which is used in the analysis of the wavelength components of a beam of light. It is found in all student physics laboratories and research laboratories and its performance is superior to that of a glass prism. Rather than 'diffraction grating' a better name would be 'interference grating' since it functions like Young's slits except that it is has thousands of slits rather than two. Instead of two interfering sources the diffraction grating is a multiple beam interference device, which produces much sharper interference maxima than the double slit. To understand the diffraction grating we will return to the two-slit experiment and ask what happens to the interference pattern as the number of slits is increased.

Figure 14 shows the interference patterns produced by monochromatic (i.e. single wavelength) light passing through sets of 2, 3, 4 and 5 equally spaced narrow (line source) slits. You should note particularly that:

- The main maxima or **principal maxima** occur at the *same positions* for any number of slits.
- The principal maxima become *more intense* and *sharper* as the number of slits increases.
- Smaller maxima or **subsidiary maxima** appear between the principal maxima when there are more than two slits.
- The subsidiary maxima become more numerous but *relatively* smaller and *less intense* as the number of slits increases.



3 slits



Figure 14 The interference patterns produced by sets of 2, 3, 4 and 5 equally spaced narrow slits.

A more detailed examination of Figure 14 or a theoretical analysis of the pattern reveals two additional points, both of which play a vital part in making the diffraction grating so important:

- \circ The peak intensity of the principal maxima rises with the *square* of the number of slits.
- \circ The width of the principal maxima decreases *linearly* with the number of slits.





5 slits

Figure 14 The interference patterns produced by sets of 2, 3, 4 and 5 equally spaced narrow slits.

These two additional points can be justified, without too much mathematics, in the following way. From what was said in <u>Subsections 5.2</u> and <u>5.4</u> you will know that at a point of fully constructive interference the intensity due to the coherent super-position of waves from *two* slits is *four* times that from a single slit. It is reasonable therefore that the coherent superposition of waves from *N* slits provides N^2 times the intensity from one slit. The width is a little more tricky but this too can be deduced from energy considerations. The *total* amount of energy reaching the screen from *N* identical slits must be *N* times that from one slit. If we neglect all light outside the principal maxima and remember that these have peak intensities rising as N^2 then they must narrow in proportion to *N* so that the *product* of their width and peak intensity (i.e. their area and the total energy they pass) can increase as *N*.

Figure 15 shows the situation for a real diffraction grating, with than 1000 more equispaced slits. In this figure the intensity is plotted on the vertical axis and $\sin\theta$ along the horizontal axis, where d, the grating spacing, is the distance between the centres of adjacent slits. The angle θ is defined as in Figure 11 for Young's slits. With so many slits, only the principal maxima are now significant and they are so sharp that they can be drawn as vertical lines.



Figure 15 The intensity distribution from a diffraction grating with a slit separation of about four times the width of the slits.

The plot of Figure 15 emphasizes that Equation 7 is just as valid for locating the principal maxima of a diffraction grating as it was in the two-slit case. Regardless of the number of slits it is still true that at each principal maximum

 $d\sin\theta = n\lambda$ with n = 0, $\pm 1, \pm 2, \pm 3, \dots$ (Eqn 7)

Though in the case of a diffraction grating the integer n is often referred to as the **order of diffraction** rather than the *order of interference*.



Figure 15 The intensity distribution from a diffraction grating with a slit separation of about four times the width of the slits.

and perhaps A new surprising feature can be seen in Figure 15. The various principal maxima no longer have the same intensity, as they did in Figure 14. This is because these slits cannot be taken as line sources. Their width is about a quarter of their separation in this case. The diffraction from these slits is less than it would be from ideal line sources, so less light reaches large angles and the principal maxima at higher values of *n*, have a reduced intensity.



Figure 15 The intensity distribution from a diffraction grating with a slit separation of about four times the width of the slits.

In practice, of course, the slits must not be too narrow if sufficient light is to be passed to see the orders! In the next subsection we will consider the diffraction pattern from a single slit when the slit is appreciably wider than a line source, and then we can return to this point. Meanwhile, it is important to remember that although the *relative intensity* of the principal maxima depends on details of the slit geometry they always appear at angles θ_n given by a rearrangement of Equation 7:

$$\sin \theta_n = \frac{n\lambda}{d} \quad \text{with } n = 0, \pm 1, \pm 2, \dots \tag{15}$$

In practice, a typical grating has several thousand slits, each just a few micrometres from its nearest neighbours. Any particular principal maximum will be very narrow and will cover a very tiny range of angles around the value of θ_n given by Equation 15. Nonetheless, that range of angles *is* finite and given its value we could use Equation 15 to work out the various different wavelengths that might have principal maxima within that narrow angular range. This is an interesting quantity because it corresponds to the range of wavelengths that cannot be clearly distinguished from one another at that particular principal maximum. In practice this range of wavelengths is very small; it might well be less than 0.1 nm at the first order maximum for light of wavelength 550 nm. It is the narrowness of these maxima that gives the diffraction grating its great ability to separate light waves with only slightly different wavelengths.

So far we have described a diffraction grating as an array of slits which transmit light; this type is properly described as a **transmission diffraction grating**. The essential feature of a diffraction grating is that it is a device which imposes a periodic variation onto an incident wavefront. Producing a periodic spatial variation in transmitted intensity is just one way of achieving this. There are several other possibilities, including reflecting the wavefront from a surface of periodic *reflectivity*. For example, a regular array of narrow lines etched onto a mirror surface produces a **reflection diffraction grating**. This will produce several reflected orders of diffraction for any incident beam. The zero order maximum obeys the usual law of reflection, having an angle of reflection equal to the angle of incidence, but higher order maxima may be seen on either side of this zero order maximum.

Some diffraction gratings occur naturally. For example, the regular spacing of the atoms in a crystal is about 0.1 nm and the separate atomic layers in a crystal act as a very good reflection diffraction grating for X-rays, which have wavelengths in this region. There are also naturally-occurring diffraction gratings for visible light, which brings us to the peacock's tail!

Question T13

1

The feathers of a peacock's tail contain layers of cells in a regular pattern, repeating at intervals of about 500 nm. This forms a natural reflection diffraction grating for light. If white light is shining at normal incidence on to the tail, show that only the green to violet part of the spectrum can appear with first order interference maxima. Explain briefly why the reflected colours change slightly as the viewing angle changes.

(*Hint*: Equation 15 is valid at normal incidence for a reflection grating.)

$$\sin \theta_n = \frac{n\lambda}{d} \quad \text{with } n = 0, \pm 1, \pm 2, \dots \tag{Eqn 15}$$



5.6 Fraunhofer diffraction at a single slit

Finally in this module, we consider in more detail the diffraction pattern from a single slit when the slit is wide compared with a line source. In <u>Subsection 3.2</u> we showed that when light is diffracted by a single slit the diffraction increases as the slit width is reduced. We can now investigate the intensity pattern on the screen as the slit width changes. If the slit were less than one wavelength wide then the light would be spread uniformly over the screen, but what is the diffraction pattern from a slit of width *w* when *w* is considerably greater than λ ? The following discussion shows that this topic is correctly placed here, in the section on interference.

We can demonstrate what is called **Fraunhofer diffraction** at a wide single slit by using an arrangement similar to that for Young's two-slit experiment (Figure 10), except that the slits S_1 and S_2 are replaced by the single slit. This wide slit must be sufficiently far from the source that it is illuminated by a plane wavefront and the screen needs to be about a metre from the wide slit. P Figure 16 shows the wide slit and the screen, although the distance to the screen has been reduced for ease of illustration; the angle θ is actually very small.



Figure 16 Fraunhofer diffraction at a single slit.

Figure 17 shows the intensity on the distant screen for the Fraunhofer diffraction from a slit of width w plotted against sin θ . Notice the following features of this plot:

- As expected there is a central maximum, the width of which increases as the wavelength increases or as the slit width decreases.
- There are several subsidiary maxima which are much less intense and only about half as wide as the central maximum.
- The first *minimum* on either side of the central maximum occurs at an angle given by $\sin \theta = \pm \lambda/w$.



Figure 17 The observed intensity distribution on the screen for Fraunhofer diffraction at a slit.

Notice that the last point is in sharp contrast with the interference pattern of a diffraction grating or a double slit, where the first *maximum* appears at $\sin \theta = \pm \lambda/d$. The reason for this difference will be apparent from the analysis that follows.

Our aim now is to calculate the positions of all the *minima* in this diffraction pattern, using secondary wavelets from all points along the slit. The task of locating the subsidiary maxima is beyond the scope of this module, except that it is clear they are about mid-way between the adjacent minima.

First, referring to Figure 16, consider wavelets from point A at the top edge of the slit and from point B in the centre. Following our treatment of Young's two-slit experiment, the path difference from these two points to point P on the screen is $(w/2)\sin\theta$. Consequently, there will be (first order) fully destructive interference between these wavelets at P if

$$\frac{w}{2}\sin\theta = \frac{\lambda}{2}$$
 i.e. when $\sin\theta = \frac{\lambda}{w}$

Now, this is exactly the location of the first minimum in the Fraunhofer diffraction pattern, but so far we have considered only the wavelets coming from two of the many points within the slit. Is there any reason to suppose that a location of fully destructive interference for these two contributions might also be a location of fully destructive interference

for *all* the contributions from within the slit? Indeed there is; the *symmetry* of the situation tells us that destructive interference must also take place at P for wavelets from *any* pair of points such as A' and B' that are separated by w/2, and we can uniformly fill the entire slit with such pairs. Thus, every wavelet from within the slit that arrives at P interferes destructively with a wavelet from another part of the slit, and we can conclude that the amplitude at P is zero and hence the intensity there is also zero. This accounts for the first minimum in the Fraunhofer diffraction pattern.





We can also divide the aperture into quarters rather than halves and perform the same analysis with pairs of points a distance w/4 apart. In this case the values of $\sin \theta$ at which fully destructive interference takes place are given by

$$\frac{w}{4}\sin\theta = \frac{\lambda}{2}$$
 i.e. when $\sin\theta = \frac{2\lambda}{w}$

dividing further into eighths, sixteenths, etc. gives the general expression for intensity *minima* in Fraunhofer diffraction:

$$\sin \theta_n = \frac{n\lambda}{w} \quad \text{with } n = 1, 2, 3, \dots \tag{16}$$



Figure 17 The observed intensity distribution on the screen for Fraunhofer diffraction at a slit.

The straight-through position, $\theta = 0$, is always a maximum in intensity because constructive interference takes place between wavelets from pairs of points placed symmetrically about the centre of the slit (e.g. A' and B'). Equation 16 tells us that the minima in the diffraction pattern are equally spaced with sin θ in units of λ/w from the central maxima, as is shown in Figure 17. A more complicated calculation confirms that the subsidiary maxima are much less intense than the central maximum.

Question T14

If the angle θ is small, show that the positions of the Fraunhofer diffraction minima, on a screen placed a distance *D* from the aperture, are given by

$$OP = \frac{n\lambda D}{w} \quad \Box$$

Finally we can return to the diffraction grating and to Figure 15 to explain why the intensity of the principal maxima decrease with order as they do.



Figure 15 The intensity distribution from a diffraction grating with a slit separation of about four times the width of the slits. $\underline{\textcircled{}}$
Diffraction effects from finite slit widths are important in a diffraction grating because the slits have width, w, although this must be smaller than the slit separation d. Each slit produces an identical diffraction pattern, of the form shown in Figure 17, which must be wider than the separation of the interference peaks because w < d. The net effect is that the common single slit diffraction pattern forms an overall envelope of intensity for the interference pattern.



Figure 17 The observed intensity distribution on the screen for Fraunhofer diffraction at a slit.

This is shown as the dashed curve in Figure 15. The first diffraction minimum shown in Figure 15 lies just beyond the fourth order interference maximum, confirming that $d \approx 4w$ in our example. The dashed curve in Figure 15 begins to rise again after $n = \pm 4$, since a singleslit diffraction pattern has smaller subsidiary maxima beyond the first minima on each side.



Figure 15 The intensity distribution from a diffraction grating with a slit separation of about four times the width of the slits. $\underline{\textcircled{m}}$

6 Closing items

6.1 Module summary

- 1 Transverse <u>electromagnetic waves</u>, in which mutually perpendicular electric and magnetic fields oscillate at right angles to the <u>direction of propagation</u>, are able to account for many of the observed features of light. In a vacuum such waves travel at <u>the speed of light</u> ($c = 3.00 \times 10^8 \text{ ms}^{-1}$) and are characterized by their <u>amplitude</u> and their <u>wavelength</u> λ or <u>frequency</u> f, where $c = f\lambda$. The wavelengths corresponding to light are just a small part of the full <u>electromagnetic spectrum</u>.
- 2 An electromagnetic wave that has a unique direction associated with its electric field is said to be *linearly polarized*. For such a wave the *plane of polarization* is the plane that contains the direction of the electric field vector and the direction of propagation.
- 3 In a transparent material the speed and wavelength of light are smaller than they are in a vacuum by a factor called the *refractive index*, μ :

 $\mu = \frac{\text{speed of light in a vacuum}}{\mu}$

(Eqn 1)

 $r = \frac{1}{\text{speed of light in the material}}$

4 <u>*Huygens' construction*</u> predicts the behaviour of propagating <u>wavefronts</u> by treating each point on the wavefront as a source of <u>secondary wavelets</u> in accordance with <u>*Huygens' principle*</u>.

- 5 *Diffraction* is the spreading of a propagating wavefront away from its original direction of propagation that occurs whenever the wavefront encounters an aperture or an obstacle. It is an observed property of light and its effects are especially significant when they are caused by an object whose size is comparable with the wavelength of light. When diffraction effects are insignificant the advance of the wavefront can be represented by a <u>ray</u>.
- 6 The wave model of light can also explain the <u>law of reflection</u> ($\theta_i = \theta_R$) and

the <u>law of refraction (Snell's law)</u> $\frac{\sin \theta_i}{\sin \theta_r} = \frac{\mu_2}{\mu_1}$ (Eqn 4)

- 7 The *intensity* of a uniform beam of light is the power transferred per unit area perpendicular to the beam. The intensity of a uniform plane electromagnetic wave is proportional to the square of its electric field amplitude. The intensity of light from a uniformly radiating *point source* decreases as the inverse square of the distance from the source.
- 8 According to the *superposition principle* if two or more waves meet in a region of space, then at each instant of time the net disturbance they cause at any point is given by the sum of the disturbances caused by each of the waves individually.

9 Light from a <u>coherent</u> source can <u>interfere constructively</u> or <u>destructively</u>. This is demonstrated by the pattern of <u>interference fringes</u> seen in Young's two-slit experiment. With slits separated by distance d and <u>monochromatic</u> light of wavelength λ , the directions of maximum intensity at <u>order</u> n are given by

$$\sin \theta_n = \frac{n\lambda}{d} \tag{Eqn 15}$$

- 10 Interference is responsible for the colours seen in reflection from thin transparent films such as soap or oil films. It also has many practical applications, such as its use in low-reflectivity coatings for lenses and high-reflectivity coatings for mirrors.
- 11 The interference of monochromatic light from a *diffraction grating* results in sharply defined intensity maxima at angles that are again given by Equation 15. Diffraction gratings are important in the measurement of wavelength.
- 12 The *Fraunhofer diffraction* pattern from a single slit of width *w* consists of a central intensity maximum, with several *minima* on either side at angles given by

$$\sin \theta_n = \frac{n\lambda}{w} \tag{Eqn 16}$$

6.2 Achievements

When you have completed this module, you should be able to:

- A1 Define the terms that are emboldened and flagged in the margins of the module.
- A2 Draw a sketch indicating the instantaneous electric and magnetic fields in a linearly polarized plane wave and discuss the way in which the properties of such waves relate to properties of light such as speed of propagation, colour and intensity.
- A3 State Huygens' principle (in your own words), relate it to the electromagnetic wave model of light, and use it to explain the propagation, diffraction and interference of light, including the laws of reflection and refraction.
- A4 Recall and use the laws of reflection and refraction (Snell's law).
- A5 Relate the intensity of a uniform parallel light beam to the power of its source and, for a point-like source, derive, interpret and use the inverse square law of illumination.
- A6 Describe Young's two-slit experiment and the pattern of interference fringes that it produces. Use the superposition principle and the wave model of light to account for the origin of the fringes and to derive formulae relating the angles of interference maxima and minima to the slit separation and the wavelength.

- A7 Explain the meaning of coherence and its significance in the observation of interference.
- A8 Describe and explain how interference can be observed in reflection from thin transparent films. Derive and use the formulae which give the conditions for constructive or destructive interference in terms of the angle of refraction, the film thickness and the wavelength (given the expression for the optical path difference).
- A9 Explain the physical principles of anti-reflection coatings and high-reflectivity coatings, and be able to calculate the thickness required for these, given the wavelength and the relevant refractive indices.
- A10 Describe and explain the principles of the diffraction grating, as used in transmission and in reflection.
- A11 Recall and use the diffraction grating relation between the angles of the interference maxima, the wavelength and the slit separation.
- A12 Describe Fraunhofer diffraction at a single slit and recall how the angles of the minima depend on the wavelength and on the slit width.

Study comment You may now wish to take the *Exit test* for this module which tests these Achievements. If you prefer to study the module further before taking this test then return to the *Module contents* to review some of the topics.

6.3 Exit test

Study comment Having completed this module, you should be able to answer the following questions each of which tests one or more of the Achievements.

Question E1

(A2) An electromagnetic wave of wavelength 550 nm is travelling along the z-axis with its electric field of amplitude 0.10 V m^{-1} oscillating along the x-axis. What is the frequency of the wave? What is the direction of polarization? What is the amplitude and direction of the magnetic field?

Question E2

(A3 and A4) Make sketches showing how Huygens' secondary wavelets predict (a) diffraction at a slit, and (b) Snell's law of refraction.

(A5) Estimate the distance from Earth of the furthest star which is visible to the unaided eye. (Make the drastic assumptions that the power of the Sun $(4 \times 10^{26} \text{ W})$ is typical of all stars, that the minimum intensity detectable by the human eye is $7.1 \times 10^{-10} \text{ W m}^{-2}$ and that there is no absorption of energy in space!)

Question E4

(A5) The intensity of light from a small source decreases as the inverse square of the distance from the source. What does this imply about how the amplitudes of the oscillating electric and magnetic fields change with distance from the source?





(A6) Green light of wavelength 546.1 nm is used in a Young's two-slit experiment where the slits are 1.5 mm apart.

- (a) Find the angles of the first and second order interference maxima.
- (b) If the screen is placed 2 m from the slits, find the distance on the screen between the second order minimum and the zero order maximum.
- (c) What colour of light should be used so that its second order maximum appears at the same angle as the second order minimum for the green light?

Question E6

(*A7*) What is meant by 'a coherent light source'? Why does the production of interference fringes require coherence? How is coherence achieved (a) in a Young's two-slit apparatus? (b) in thin film interference?



(A8 and A9) A film of oil of refractive index $\mu = 1.40$ covers the surface of a pool of water ($\mu = 1.33$) to a depth of 300 nm. Find the angles (θ_r) at which blue ($\lambda = 450$ nm), green ($\lambda = 550$ nm) and red ($\lambda = 650$ nm) light are reflected with maximum intensity in first order (i.e. n = 1).

Question E8

(A10 and A11) A diffraction grating has 12 000 slits ruled regularly on to a glass plate 2 cm wide. Find the angles of the first and second order interference maxima for the two yellow wavelengths $\lambda = 589.0$ nm and $\lambda = 589.6$ nm (sodium light) — remember to use four significant figures throughout your calculation.





(A12) An aperture of width w is illuminated by plane waves of wavelength 589.0 nm. Find the three values of w such that the first diffraction minima appear at 20° , 40° and 60° , respectively. At each of the three calculated values of w find the angles of the first diffraction minimum for green light of wavelength 546.1 nm.



Study comment This is the final *Exit test* question. When you have completed the *Exit test* go back to Subsection 1.2 and try the *Fast track questions* if you have not already done so.

If you have completed **both** the *Fast track questions* and the *Exit test*, then you have finished the module and may leave it here.

