FIELDS AND WAVES

2004

8.0 PHYSICAL OPTICS

There were really no losers in the dispute about the nature of light: was it made of "corpuscles" (particles) as had been supposed by Isaac Newton, or was it wave-like as had been advocated by Christian Huygens. Both points of view are reconciled in the way that quantum mechanics describes light: it has both particle- and wave-like properties. This particle- wave duality is characteristic of the quantum world. The particle-like properties of light are emphasised when we speak of *photons*, a mode of description utilised by Albert Einstein in 1905 in his paper explaining the photo-electric effect (it was for this work that he was awarded the Nobel Prize). Photons are the quanta of light. But we will not pursue these ideas in this course, and will in this section be concerned with the other way of talking about light, namely as a wave phenomenon.

The compelling evidence that light has wave-like properties came from the experiments of Thomas Young in 1801. His "double-slit" experiment is one of the classic experiments in physics and has implications and applications far beyond what we will consider here. For example, analogous experiments show that electrons have wave-like properties too! And the theoretical analysis of this phenomenon is at the heart of quantum mechanics.

8.1 YOUNG'S DOUBLE SLIT EXPERIMENT

The idea behind the experiment was to demonstrate the interference effect which are so characteristic of waves. As we have already seen when thinking about the Huygens construction, the waves from two sources can interfere with one another where they overlap, either constructively by reinforcing the displacement when they are in phase, or destructively when they are out of phase. But only if the sources are *coherent* will the pattern of reinforcement here and cancellation there remain steady over sufficiently many cycles of oscillation of the waves for it to be observable. So the first step is to devise a way to establish coherence for two sources. Light coming from a lamp, or from the Sun, is incoherent. This is because what we see is the effect of very many individual sources of light within the lamp or on the surface of the sun, all of them contributing independently and so not "in step" with one another. In fact since the time taken for a radiating atom to emit a wave-train is about 1 ns (10^{-9} s), the wave-train emitted by an individual atom has a length which is given by multiplying this by the speed of light ($c = 3 \times 10^8$ m $\cdot$ s^{-1}) so around 30 cm. This is half a million or so wavelengths long. But if the atom collides with another atom during the time it is radiating, the phase of the wave changes abruptly, so that the *coherence length* is much shorter than this. Typically, the wave changes phase abruptly every millimetre or so along its length; the average interval between such abrupt changes of phase is called the coherence length.

The light passing through a narrow slit in a screen has been emitted by only a small part of the source, and so when it in turn is allowed to fall on a second screen, the illumination of a small enough portion of that that screen will be coherent. This is because the light will have travelled essentially the same distance from the slit in the first screen to all points of the small region on the second screen. And then if there are two narrow slits in that second screen the light emerging from them (the *double slits* of the experiment) will be coherent; the waves will be in step with one another. And so where the waves emerging from the double slits cross one another they can establish a steady pattern of reinforcement and cancellation, of constructive and destructive interference, which can be seen when a third screen is placed where it is illuminated by light which has passed from the source, through the single slit in the first screen and then through the double slits in the second screen. A colour filter between the lamp and the first screen can be inserted so as to select light of a limited range of frequencies. [You can find a diagram of the set-up at the foot of page 341 in the text-book. There is an amusing web site at http://surendranath.tripod.com/DblSlt/DblSltApp.html]

When the apparatus is correctly adjusted, what is observed on the final screen is an alternation of light and dark *fringes*. The spacing y between adjacent bright fringes (or adjacent dark fringes) is related to the wavelength λ of the light, and to the distance d (between the centres) of the double slits as well as the distance X of the final screen from the screen with the double slits by the formula

$$\lambda = \frac{yd}{X}.$$

To understand how this formula arises, consider a point P on the final screen. The light reaching it from the slit S_1 will have travelled a different path than that from slit S_2 , and therefore there will be a different number of wavelengths between the two. If the difference between the path-lengths S_1P and S_2P is an integer number of wavelengths,

$$S_1PS_2P = m\lambda, \qquad \text{where} \qquad m = \ldots, -3, -2, -1, 0, 1, 2, 3, \ldots,$$

then the waves from the two slits S_1 and S_2, which set out in phase with one another will still be in phase when they arrive at P. So the point P will be a place of constructive interference, and so will be on a bright fringe. Conversely, if the path difference is half a wavelength more than an integer, the waves arriving at P will be exactly out of phase with one another, and so will interfere destructively, leading to darkness. So whether P is at a bright or a dark fringe can be determined simply from the geometry of the triangle S_1S_2P. You should refer to figure 18.1G on page 343 of the text-book and the explanation which follows for the derivation.

If the light source is white (unfiltered), the spacing between the fringes will be different for each of the wavelengths which go to make up the white light, and so the fringes are coloured. With monochromatic light (light of a single colour, or more precisely, of a single wavelength), the fringes are sharper and all of that colour. The spacing y of the fringes is proportional to the wavelength λ of the light, and so a double-slit *interferometer* can be used to measure the wavelength. Note also that the separation of the fringes is proportional to X/d, which is why the double slits have to be close together, typically a fraction of a millimetre, with X something like a thousand times greater. Then since the wavelength of light is in the range 250 - 650 nm, the fringe spacing, which is X/d times greater, is around a half a millimetre, which can be observed quite easily.

The light emitted from a laser is coherent to an extent not realisable by other methods. In fact the coherence length can be several metres, and it is possible to use light from a laser in a double-slit interferometer without the need for the first screen to establish coherent illumination of the double slits. The principle exploited by lasers (Light Amplification by Stimulated Emission of Radiation) was first given in 1916 by Einstein, but it was not until the 1950s that these ideas were implemented. Today lasers have many important applications in industry and elsewhere: for example the microscopic pits etched into the surface of a CD which code the recorded sound are sensed by a laser, and the barcode on purchases is read at the check-out by a laser.

8.2 INTERFERENCE BY THIN FILMS

You will no doubt be familiar with the brilliantly coloured appearance of thin films of oil on a puddle of water. This is caused by the interference of light reflected from the upper surface of the oil with that from the oil-water interface. The light reflected from the two surfaces will travel different distances in reaching the viewer, in other words there will be a difference in path length, and so the possibility or interference. Because the maxima in the interference pattern which results will be seen in different directions for different colours, the result is to make the oil film appear to be coloured.

If we consider a thin film with parallel sides, the difference in path length at normal incidence is just twice the thickness t of the film. So there will be destructive interference when this is half a wave length. This is exploited in the *blooming* of lenses, for example in a camera or other optical device. The idea is to eliminate as far as possible the reflections from the surface of the lens. By coating the lens with a thin uniform film of thickness and refractive index designed so that the path difference between the rays reflected from its two surfaces *is* half a wave length there is destructive interference, and the reflected light is eliminated;

$$2t = \lambda/2 = \lambda_0/2n.$$

(Note that we have used the fact that the wavelength λ?of the light in the film differs from λ_0 the wavelength in air, by the factor n, the refractive index of the material used to coat the lens.) Of course, this can only be achieved at one particular wavelength, usually chosen to be in the yellow portion of the spectrum since this is near the middle of the visible range. The yellow light is then cancelled out from the reflected white light, leaving a predominance of red and blue. Reflections of white light from a bloomed lens then appear to be faint purple.

There is something else to be noted: on reflection at the interface between media with different refractive indices, light coming from the side with lower refractive index suffers a 180° phase change. If the refractive index of the film used to bloom the lens is intermediate between that of the air and the glass of the lens, each of the reflections, from the upper and the lower surface, has this phase change, so they cancel out.

But now consider what happens to light which is transmitted from outside the camera through the air-bloom interface, then reflected from the bloom-lens interface, and reflected back again at the bloom-air interface and so through the bloom-lens interface into the camera. It has had only one phase-reversal at a reflection, but has passed three times through the bloom, rather than just once if it had not been reflected at all, so a path difference again of twice the thickness of the bloom. This path difference was just such as to give a phase reversal, so the phase reversal from the reflection cancels that from the extra path length, and the transmitted rays interfere *constructively* . This result should not surprise you; if light is subtracted from the reflected component by the blooming of the lens, it is not surprising to find that it has to be added to the transmitted component!

8.3 AIR WEDGES AND NEWTON'S RINGS

If two microscope slides are placed one on top of the other, but with a thin wire holding one edge of the upper slide away from the lower, there is a wedge-shaped air-filled space between the slides with a very small opening angle. This means that light incident from above reflected from the upper surface of the lower slide can interfere with light reflected from the lower surface of the upper slide. These reflections introduce a relative phase difference of 180°, since only the first is from lower to higher refractive index. The path difference will vary as we go along the wedge from zero at one end to twice the thickness of the wire at the other. So depending on where along the wedge we look, there will be either destructive or constructive interference, and so an alternation of light and dark fringes. The mth dark fringe will occur where the twice the thickness of the wedge is m wavelengths. The spacing y between the dark fringes is therefore given by

$$y \tan \theta = \lambda_0 / 2,$$

where θ is the opening angle of the wedge.

A similar effect can be observed when a convex lens with a large radius of curvature for its surface is placed on a glass plate. This was noted by Newton, who described the alternating concentric light and dark rings which can be observed when light is reflected from such an arrangement. Again there is a film between the glass surfaces, but now the places where the separation t is an integer multiple of the wavelength are in concentric circles. If the film (the space between the lens and the plate) has a refractive index n which is between that of glass and of air, the rays reflected from the bottom of the curved surface of the lens and from the upper surface of the plate on which it rests have again a 180° phase difference coming from the reflections, and a path difference of $2t$. So there will be destructive interference whenever $2t = m\lambda_0/n$. To good approximation, Pythagoras' theorem can then be used to give an equation for the diameter d_m of the mth dark ring counting outwards from the centre (which is dark):

$$d_m^2 = 4mR\lambda_0/n,$$

where R is the radius of curvature of the lower surface of the lens.

It is perhaps surprising that Newton did not recognise that he had observed compelling evidence for the wave nature of light!

8.4 DIFFRACTION

We have already noted that waves can be diffracted at narrow gaps in an obstacle, and so spread out on passing through the gap. It is now time to study this phenomenon more closely, and we do so for light passing through a narrow slit. If light from a pin-hole source (so chosen for reasons of coherence) is passed through a narrow slit, what is observed is not a sharp shadow, but a series of fringes which diminish in intensity as one moves further away into the region where geometrical (ray) optics would lead one to expect the shadow to lie. The narrower the slit, the wider the spacing between the fringes. The centre of the fringe pattern (in the middle of the place where geometrical optics would lead you to expect it to be illuminated) is bright, and to either side there are symmetrically located bright fringes half the width of the central bright fringe.

All this can be explained by use of Huygens' construction. We may suppose that the slit is illuminated by normally incident coherent light. Remembering that Huygens suggested that each point on the unobstructed wave front is to be thought of as the source of secondary wavelets, imagine the width of the slit divided in two, and then consider a pair of points, the first in one half of the slit, and the other distance half the width of the slit away in the other. Light coming from these two points (considered as secondary sources as Huygens would teach us) starts out in phase. But in reaching a distant observer, there is a path difference for the two points which depends on the angle θ that the observation is made with respect to the direction of the incident beam. Simple trigonometry gives (to good approximation) for this path difference $(W/2)\sin\theta$, where W is the width of the slit. So the two contributions will interfere destructively whenever this differs by half a wavelength from an integer number of wavelengths. This is independent of which pair of points was chosen. The result then is that at angles satisfying this condition, there will be a dark fringe. The mth occurs when

$$W\sin\theta = m\lambda$$

, and between them there will be bright fringes. (The explanation for the diminution in brightness of the bright fringes as one goes away from the centre of the pattern is more complicated and we will not consider it here.) If the light diffracted through the slit is allowed to fall on a screen so as to observe the diffraction pattern there will be seen an alternation of light and dark fringes around the central bright fringe. A graph of the intensity is given at figure 18.3C on page 351 of the text-book.

When there are two slits, the resultant pattern obtained from their mutual interference is in fact a combination of the two effects we have discussed, each slit having a diffraction pattern and the two interfering where they overlap so as to produce what is illustrated in figure 18.3E. Increasing the number of equally spaced slits modifies the pattern further, with the consequence that the double slit pattern we first discussed dominates more and more over the single slit diffraction effect, and the result is to give sharper and sharper lines of brightness. This is what is observed with a *diffraction grating.*

Diffraction gratings are made by making regularly spaced rulings on a transparent or reflective surface. The closer together and the more regular the spacing of the rulings, the better the grating. Diffraction gratings are used to obtain high quality spectra. They make use of the fact that the angles at which the maxima in the diffraction pattern occur are different for different wavelengths. In fact, just as for the Young's slit experiment, they are at angles θ_m with the incident beam which satisfy

$$d \sin \theta_m = m\lambda.$$

The label m is the *order* of the line or of the spectrum in which the line occurs, and d is the spacing between the rulings or lines on the grating.

Spectroscopy is one of the most pervasive techniques in experimental physics. The details of spectra provide information about the structure of atoms and molecules. Spectra of stars provide information about their chemical composition, and (through the Doppler effect) their temperature motion.

Prepared by John M Charap on 04/03/02, updated 08/03/04