

L5

DIFFRACTION

OBJECTIVES

Aims

From this chapter you should gain an understanding of the process of diffraction and its role in producing distinctive interference patterns. You should also aim to understand how the effects of diffraction can affect and limit the formation of images. As the classical example of diffraction, you should be able to describe and explain the Rayleigh criterion and apply it to simple examples.

Minimum learning goals

1. Explain, interpret and use the terms:
diffraction, diffraction pattern, Fresnel diffraction, Fraunhofer diffraction, angular resolution, Rayleigh criterion, diffraction envelope, principal maximum, secondary maxima, double slit, diffraction grating.
2. Describe qualitatively the diffraction patterns produced in monochromatic light by single slits, rectangular apertures, circular apertures, double slits and diffraction gratings.
3. Describe the interference pattern produced by monochromatic light and a grating in terms of the interference pattern produced by a set of line sources modulated by a diffraction envelope.
4. State and apply the formulas for the angular widths of the central maxima in the Fraunhofer diffraction patterns of a single slit and a circular hole.
5. State the Rayleigh criterion, explain its purpose and apply it to simple examples.
6. Describe how wavelength, slit width, slit spacing and number of slits affect the Fraunhofer diffraction patterns produced by multiple slits and gratings.
7. State and apply the formula for the angular positions of maxima in the Fraunhofer diffraction patterns produced by multiple slits and gratings.

PRE-LECTURE

5-1 SHADOWS

In the ray model we suppose that when light travels through a homogeneous medium it moves along straight lines. That observation is often called the law of rectilinear propagation. The existence of shadows is good evidence for the ray model of light. When light from a small (or 'point') source goes past the edges of an opaque object it keeps going in a straight line, leaving the space behind the object dark (figure 5.1).

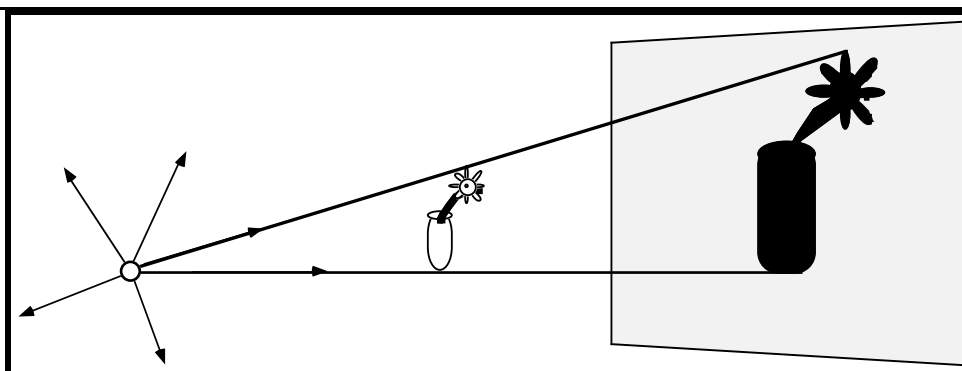


Figure 5.1. Straight line propagation of light

A small source of light produces sharp shadows.

When the light reaches some other surface the boundary between light and dark is quite sharp. (On the other hand if the light comes from an extended source the shadow is not so sharp - there is a region of partial shadow surrounding the total shadow.)

It was not until the about the beginning of the nineteenth century that scientists noticed that shadows are not really perfectly sharp. Looked at on a small enough scale the edge of a shadow is not just fuzzy, as you might expect for an extended source of light, but there are also light and dark striations or fringes around the edge of the shadow. Even more remarkable is the slightly later discovery that there is always a tiny bright spot right in the middle of the shadow cast by a circular object (figure 5.2). The fringes and the bright spot cannot be understood in terms of the ray model - the explanation lies in the wave model. According to the wave theory, the fringes are formed by the **diffraction** or bending of light waves around the edges of objects and the subsequent interference of the diffracted waves. The diffraction of water waves at a hole in a barrier was shown in the video lecture on interference (L4) and it is sketched in figure 4.4.

One of the effects of diffraction is the production of interference or fringe patterns. Although these patterns are actually interference patterns in the same sense as those we have already discussed, when they are produced by the bending of light around obstacles or apertures (holes) they are called **diffraction patterns**.

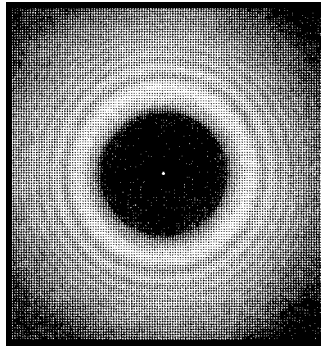


Figure 5.2. Shadow of a small circular object

The edges of the shadow contain fringes and there is a small bright spot in the middle.

Young's twin slits experiment shows that light does not always travel along straight lines. You can see that it must bend somewhere by looking at figure 5.3. Since there is light at the middle of the screen but no straight through path from the source to the screen, the light which gets there must somehow be going around corners. What must be happening is that after the light reaches the first slit it then spreads out so that some of it reaches the other two slits. Then, having passed through those slits it spreads out again in many directions until it reaches the screen. This behaviour is typical of waves, but not of particles.

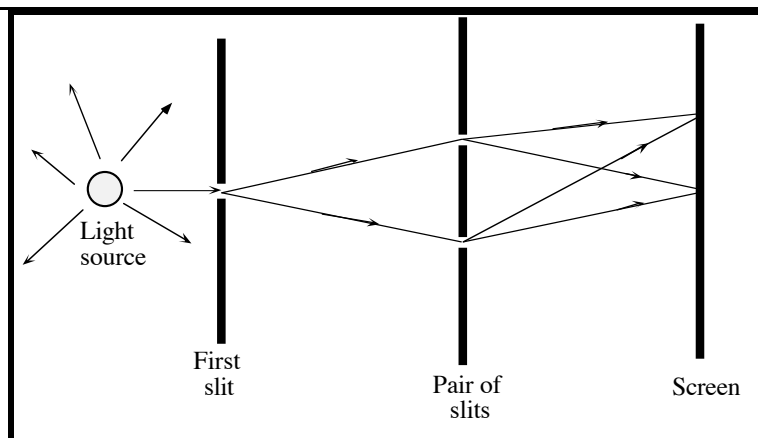


Figure 5.3. Arrangement for Young's experiment

Light cannot be travelling in a straight line all the way from the source to the screen.

The connection between waves and diffraction is much more noticeable for sound than it is for light. The observation that sound easily travels around corners is strong evidence for the wave nature of sound.

TEXT & LECTURE

5-2 HUYGENS' CONSTRUCTION

A way of describing how diffraction occurs was invented by Christian Huygens (1629-1695) in about 1679 and was modified much later into the form we now use by Augustin Fresnel (1788 - 1827). Huygens' construction is a method for locating the new position of a wave front. Starting from a known wavefront, we imagine each point on the wavefront to be a new source of secondary wavelets. The wavelet from each point then spreads out as a sphere (which appears as a circle in two-dimensional diagrams). After a certain time the new position of the original wavefront is defined by the boundary or envelope of all the secondary wavelets. Huygens construction for a plane wave going through a slit is shown in figure 5.4; after it has passed through the hole, the wavefront is no longer plane, but has curved edges. The result of Huygens' construction is significantly different from the ray model in that it shows light waves spreading into the region of the geometrical shadow. You should notice that opposite the slit the wavefront is still plane; it bends only at the edges. This bending effect is noticeable only on a scale comparable with the wavelength - for a very wide hole comparatively little of the wavefront bends around the edges.

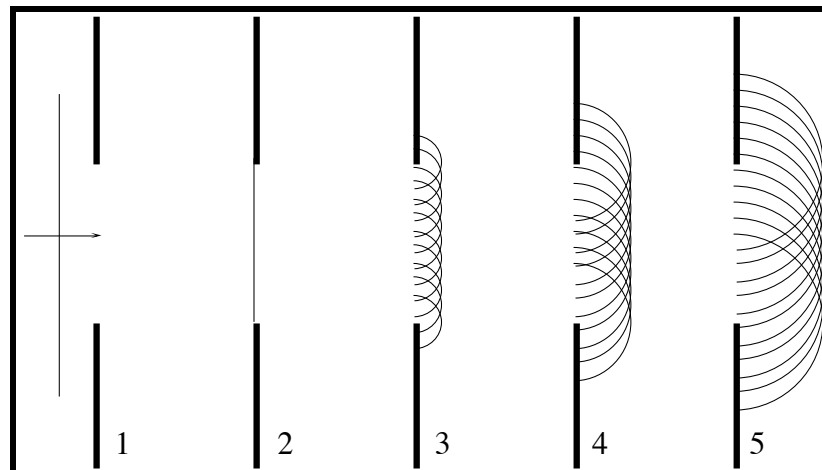


Figure 5.4. Huygens' construction

Five stages in the progress of a wavefront through a slit. The diagrams show the construction for three equal time intervals after a plane wave reaches the slit, travelling left to right. At any stage the new wavefront is the boundary of all the wavelets used in the construction.

Although the Huygens construction 'explains' how new wavefronts are formed, it does not predict the wave's amplitude; other methods are needed for that. The construction does however contain a clue about the strength of the waves. If you look at the straight-through wave in figure 5.4

you will see that there are many wavelets, but on the sides relatively few wavelets appear, which would seem to suggest, correctly, that the diffracted wave is weaker on the sides.

5-3 PRODUCING DIFFRACTION PATTERNS

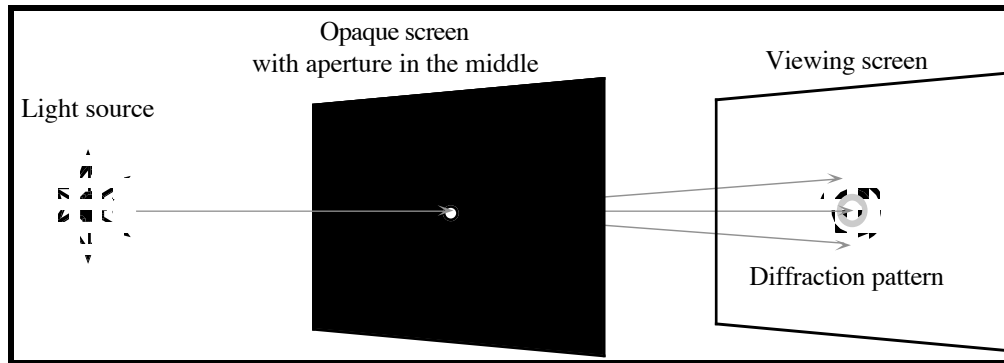


Figure 5.5. Producing a diffraction pattern
(Not to scale)

Details of the distribution of light over the diffraction pattern depend on the distances of both the source and the screen from the diffracting aperture. The general situation (figure 5.5), known as **Fresnel diffraction**, can be mathematically very complex. However, calculations are greatly simplified if both the source and the screen are at very large distances from the aperture (i.e. if those distances are much greater than the diameter of the aperture). We will deal quantitatively only with this situation, which is known as **Fraunhofer diffraction**.

Fraunhofer conditions can be achieved in practice by using two lenses (figure 5.6). The first lens ensures that the wavefronts arriving at the aperture will be plane (with parallel rays) and the second lens brings beams of diffracted light together to form an interference pattern on the screen.

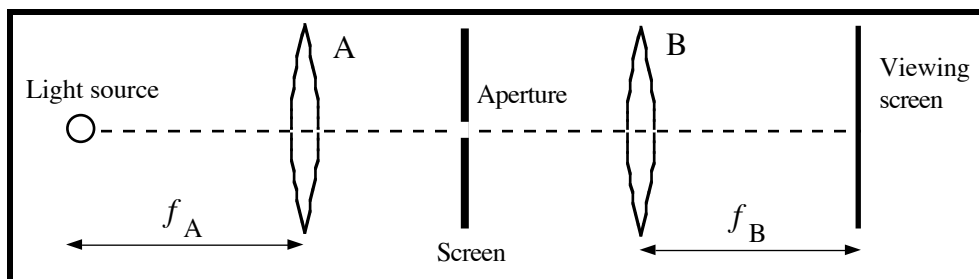


Figure 5.6. Producing Fraunhofer diffraction

Lens A ensures that the wavefronts which arrive at the aperture are plane. In order to see what the diffraction pattern at infinity would be like, lens B is used to produce its image on the screen.

5-4 DIFFRACTION AT A SINGLE SLIT

You will recall that light consists of a superposition of many elementary waves, with a wide range of frequencies. In the simple theory of diffraction we deal with one frequency at a time. The diffraction of a complex beam of light is then described in terms of what happens to each of the component waves with different frequencies.

The simplest case of Fraunhofer diffraction is that for a long narrow slit in an opaque screen. The interference pattern consists of a set of light and dark parallel fringes (figure 5.7).

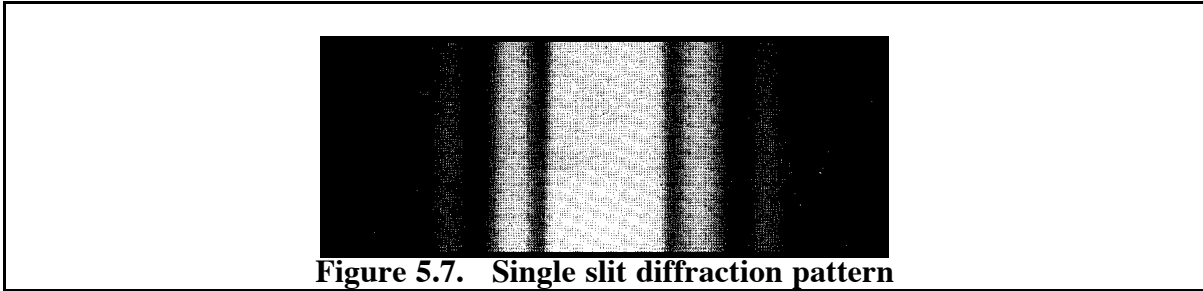


Figure 5.7. Single slit diffraction pattern

Figure 5.8 shows how the diffraction by a slit is studied using Fraunhofer conditions. The first lens ensures that the wavefronts arriving at the aperture will be plane and the second lens focusses the light onto the screen.

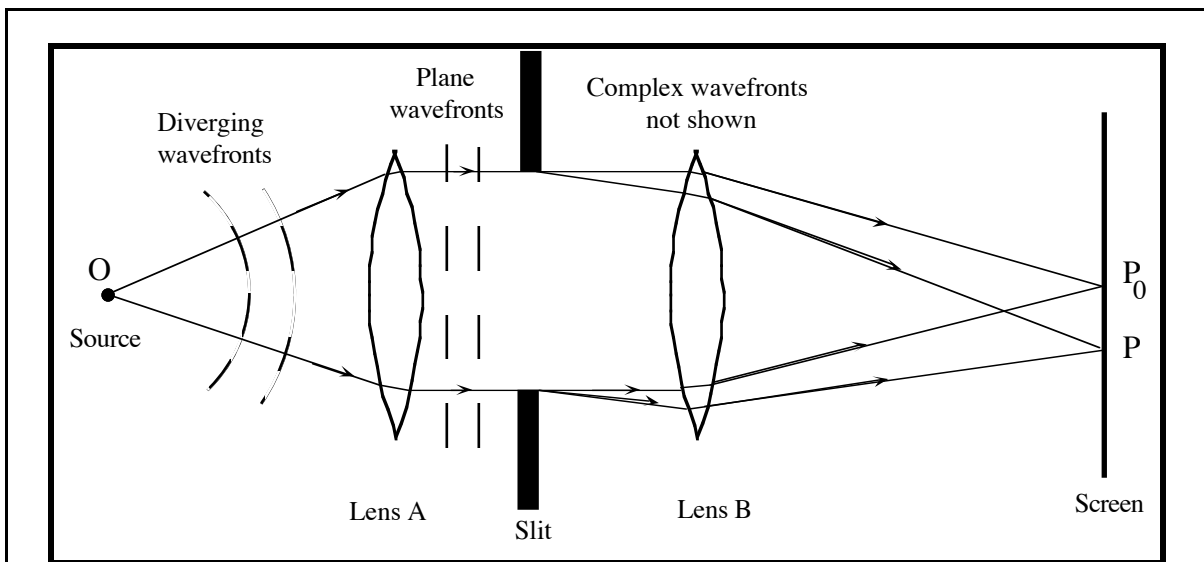
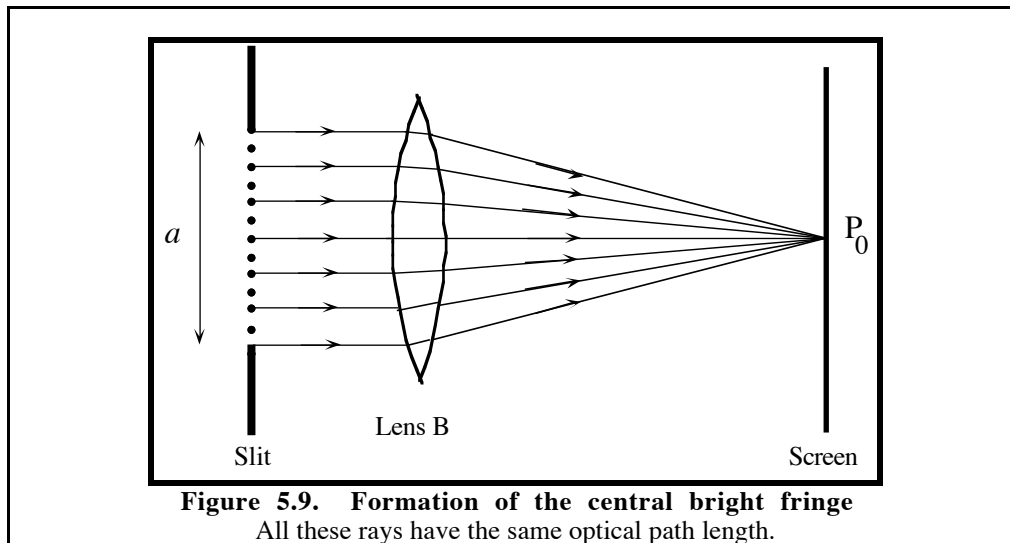


Figure 5.8. Fraunhofer diffraction by a single slit

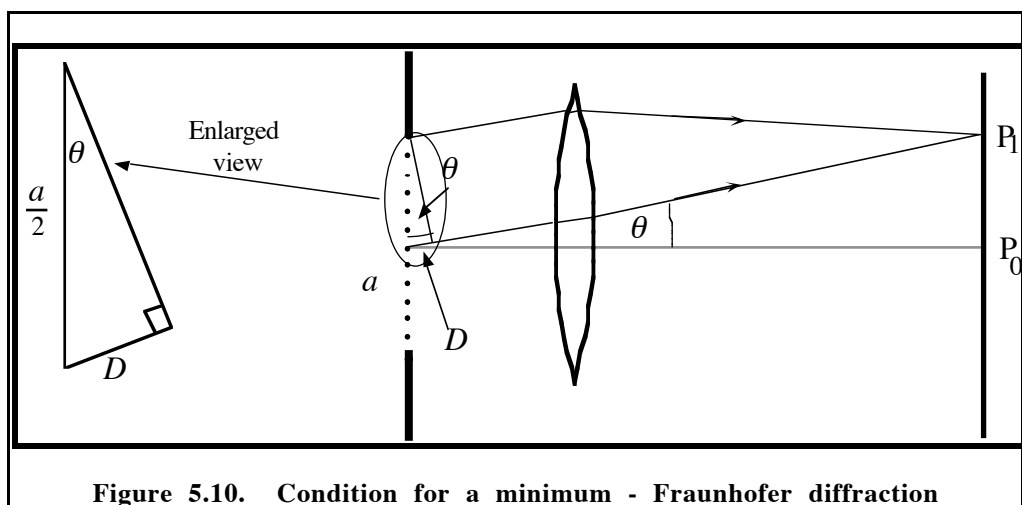
After passing through the slit light waves spread out in all directions. The second lens brings parallel groups of rays to a focus on the screen. If there were no diffraction there would be only one such focus (P_0) for each source point O.

The wavefronts arriving at the slit are plane, but because of diffraction, the wavefronts on the other side will not be plane. Light arriving at any point in the aperture has a fixed phase relationship to light from the same part of the source arriving at any other point in the aperture. The simplest case is to imagine an elementary plane wavefront arriving parallel to the slit as shown in figure 5.8. Since all points on a wavefront have the same phase we can imagine the aperture filled with many tiny coherent sources. When the light from all of these sources comes together at various points on the screen an interference pattern will be seen. The brightness at any point will depend on the phase differences among all the secondary waves arriving there and those phase differences will depend on the optical paths travelled by the different waves. Path differences can be calculated using rays which leave the slit parallel to each other, so that they would eventually meet at infinity were it not for the presence of the second lens. Since the lens itself introduces no additional optical path difference, calculations can be done on the assumption that the rays meet at infinity.

The following argument, given in the video lecture, shows how to work out the condition for a minimum in the interference pattern. Consider rays coming from various coherent points spread across the slit (figure 5.9).



All the rays parallel to the axis will be focussed at P_0 . Although the geometrical path lengths of the rays are obviously different, the optical paths from different points across the slit to the point P_0 are all equal. That is so because the longer paths outside the lens are compensated by shorter optical paths of the rays within the lens. Rays from the outside of the slit go through more air but less glass than rays near the middle of the lens. So all the light arriving at P_0 is in phase, giving a bright region there.



For other points on the screen the phases for light arriving along a parallel group of rays are all different, but it is fairly straightforward to work out where there will be complete cancellation. Pick two parallel rays, one leaving from the top of the slit and the other from a point halfway across it (figure 5.10). Both rays leave the slit at an angle θ to the axis. If the angle θ is chosen so that the path difference between these two rays is some odd multiple of half a wavelength ($D = \lambda/2, 3\lambda/2, \dots$ etc) they will interfere to produce a minimum in the irradiance at the screen.

Any other pair of rays half a slit width apart and leaving at angle θ will also interfere to give a minimum. We can choose similar pairs of rays until the whole aperture has been covered. Since all pairs produce a minimum there will be a dark region at P_1 . From the small triangle in the diagram it can be seen that the path difference for every pair of rays is equal to $(a/2)\sin\theta$. Finally equate this expression with the value of D (above) needed for a minimum.

The condition for an interference minimum is that

$$\sin\theta = \frac{m\lambda}{a} \quad \text{for } m = \pm 1, \pm 2, \pm 3 \dots \dots \dots (5.1)$$

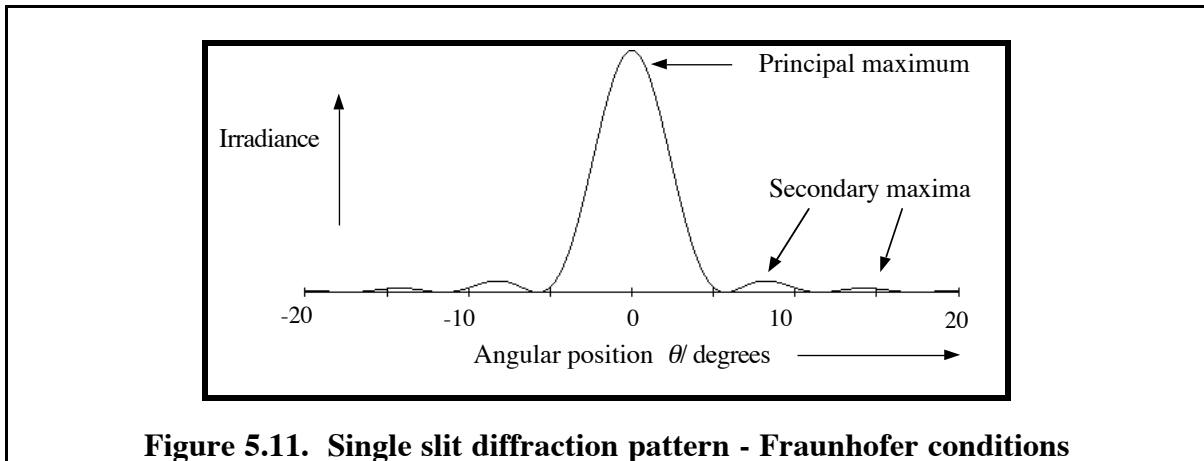


Figure 5.11. Single slit diffraction pattern - Fraunhofer conditions

A plot of the intensity of the light against position on the screen is shown in figure 5.11. Several features of the diffraction pattern are worth noting.

- The angle between the central peak and the first minimum ($m = 1$) is given by

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

So the widest diffraction patterns are produced by the narrowest slits. If the slit is only one wavelength wide the angular position of the first minimum would be 90° giving an angular separation of 180° between the zero intensity positions - so the diffracted light spreads out in all directions. On the other hand if the slit is wide, the angle of the first minimum is small and diffraction effects may be hard to see.

- The principal maximum is roughly twice the width of the secondary maxima. You can see that by putting $\sin\theta \approx \theta$ in equation 5.1 and comparing the change in angular position going from $m = 1$ to $m = -1$ with the change in going from $m = 1$ to $m = 2$, for example.
- The width of the diffraction pattern depends on wavelength. The pattern for red light is wider than that for blue light.
- The intensities of the secondary maxima are very much less than the intensity of the principal maximum.

5-5 DIFFRACTION BY A CIRCULAR APERTURE

In practice the most important example of diffraction produced by a single aperture is that for a circular hole. Most optical instruments have circular apertures or lenses that act as circular apertures, so whether we wish it or not we will get diffraction effects.

For Fraunhofer diffraction at a circular hole, the plot of intensity against position on the screen (figure 5.12) is similar in its general shape to that for a single rectangular slit. However it differs from the single slit pattern in the following ways.

- The diffraction pattern is a circular patch of light (called the Airy disc) surrounded by rings of light.
- The angle between the principal maximum and the first minimum is about 20% greater than that for a slit of the same width:

$$\sin\theta = 1.22 \frac{\lambda}{a} . \quad \dots (5.2)$$

- The spacings between adjacent minima are not as uniform as those for the slit.
- The intensities of the secondary maxima are smaller.

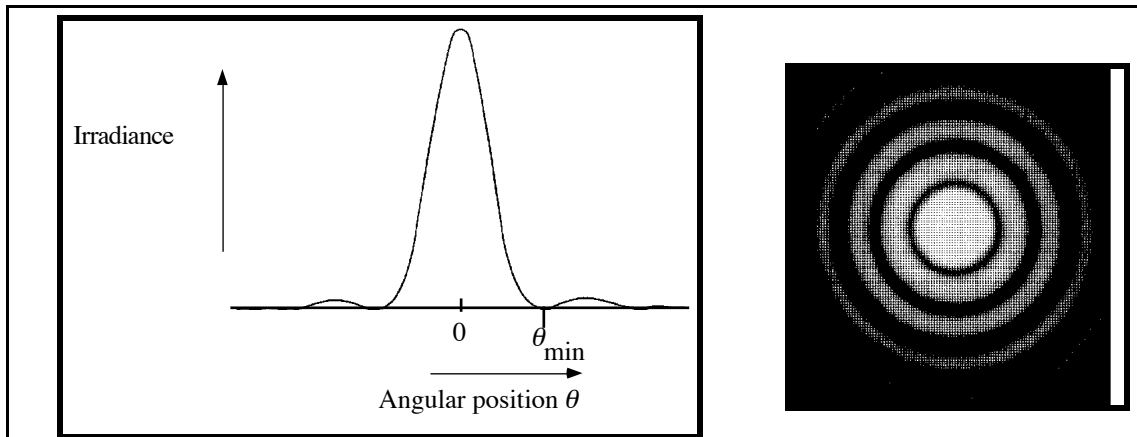


Figure 5.12. Diffraction pattern for a circular aperture

At a large distance from the hole (Fraunhofer conditions) the first minimum in the irradiance occurs at an angular position $\sin \theta_{\min} = 1.22 \lambda/D$. The secondary maxima are less bright than those for a slit.

Example

It is important to realise the size of the diffraction pattern for an average sort of lens, say 50 mm in diameter, which acts as a circular aperture whenever it is used to form images. Consider a typical visible wavelength of 500 nm and calculate the angle between the central maximum and the first minimum.

$$\begin{aligned} \sin \theta &= 1.2 \frac{\lambda}{D} \\ &= \frac{1.2 \times 500 \times 10^{-9} \text{ m}}{50 \times 10^{-3} \text{ m}} = 1.2 \times 10^{-5}. \end{aligned}$$

Since θ is small, $\sin \theta \approx \theta$ and $\theta = 1.2 \times 10^{-5}$ rad or 0.0007° .

5-6 YOUNG'S DOUBLE SLIT

Young's classic interference experiment with the two slits has already been described (§§4-2, 4-4) and quoted as evidence for the diffraction of light (§5-2). Detailed descriptions of the two-slit interference pattern involve the width of a slit, a , and the separation between the slits, d (figure 5.13). Note that $d > a$.

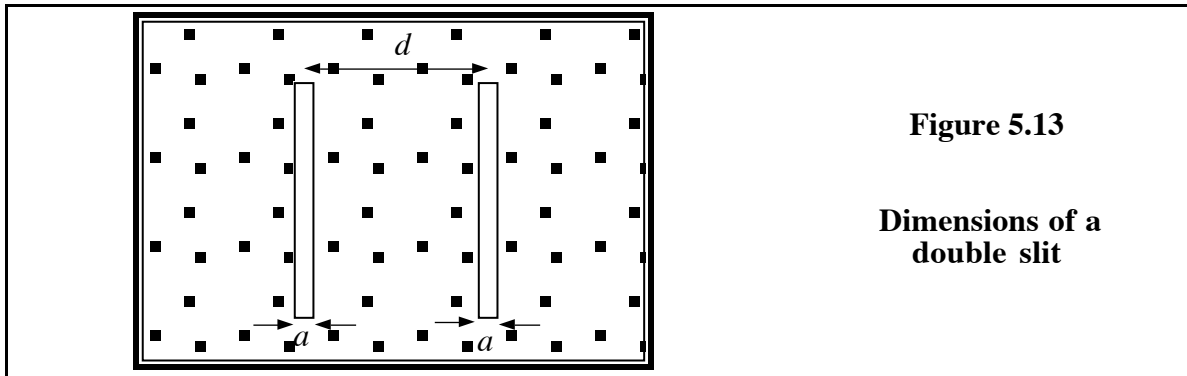


Figure 5.13

Dimensions of a double slit

A simple analysis of the experiment assumed that both the sources were very narrow - much less than a wavelength wide ($a \ll \lambda$). In practice however real slits are not that narrow and the interference pattern depends on the width of the slits as well as their separation. The kind of interference pattern produced is illustrated in figures 5.14 and 5.15.

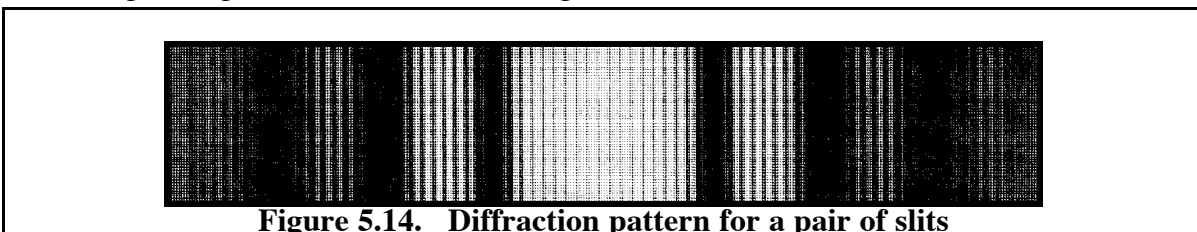


Figure 5.14. Diffraction pattern for a pair of slits

For Fraunhofer conditions the plot of irradiance against angular position on the screen shows a pattern of almost equally spaced fringes with different brightnesses (figure 5.15). The maximum irradiance occurs in the central bright fringe, at position $\theta = 0$. Notice how the fringes occur in groups, with the brightest fringes in the middle group.

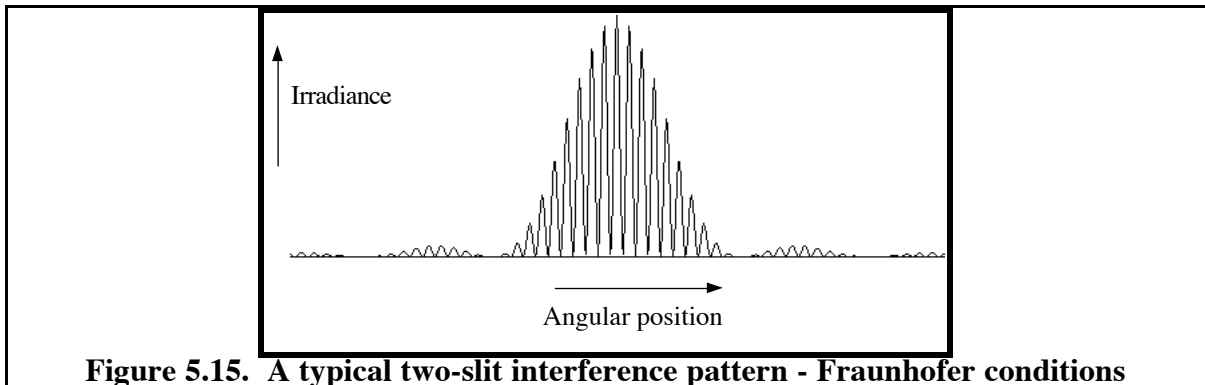


Figure 5.15. A typical two-slit interference pattern - Fraunhofer conditions

Figure 5.16 shows a comparison between the interference patterns for one and two slits. You may recall from chapter L1 that the maximum intensity for two identical coherent sources should be four times (not twice) the intensity for one source. That result posed a puzzle about conservation of energy. Energy or power is distributed differently in the two cases. Although the peak of the two-slit curve is four times higher than that of the single-slit curve, there are places where there is no light at all. The total power is represented by the areas under the graphs and, as expected from conservation of energy, the area under the two-slit curve is twice the area under the one-slit curve.

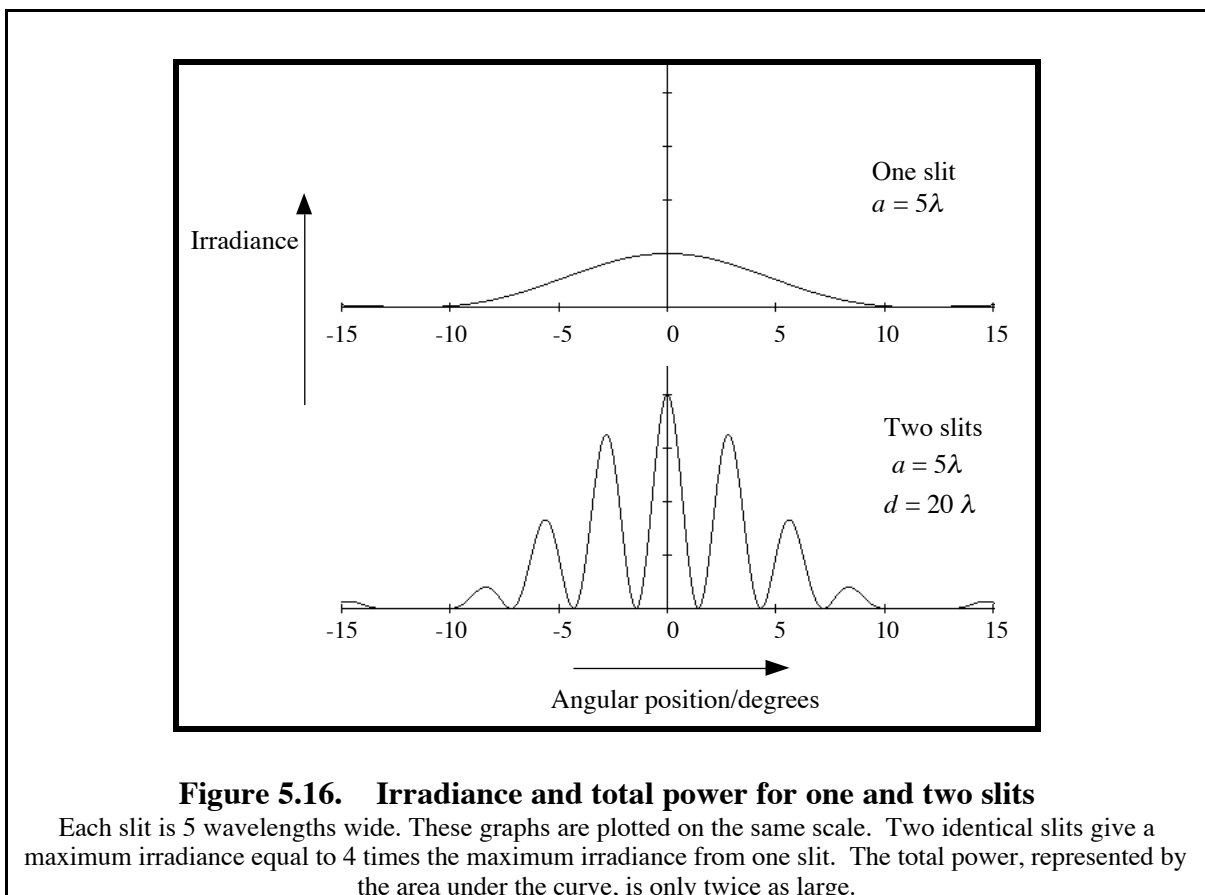
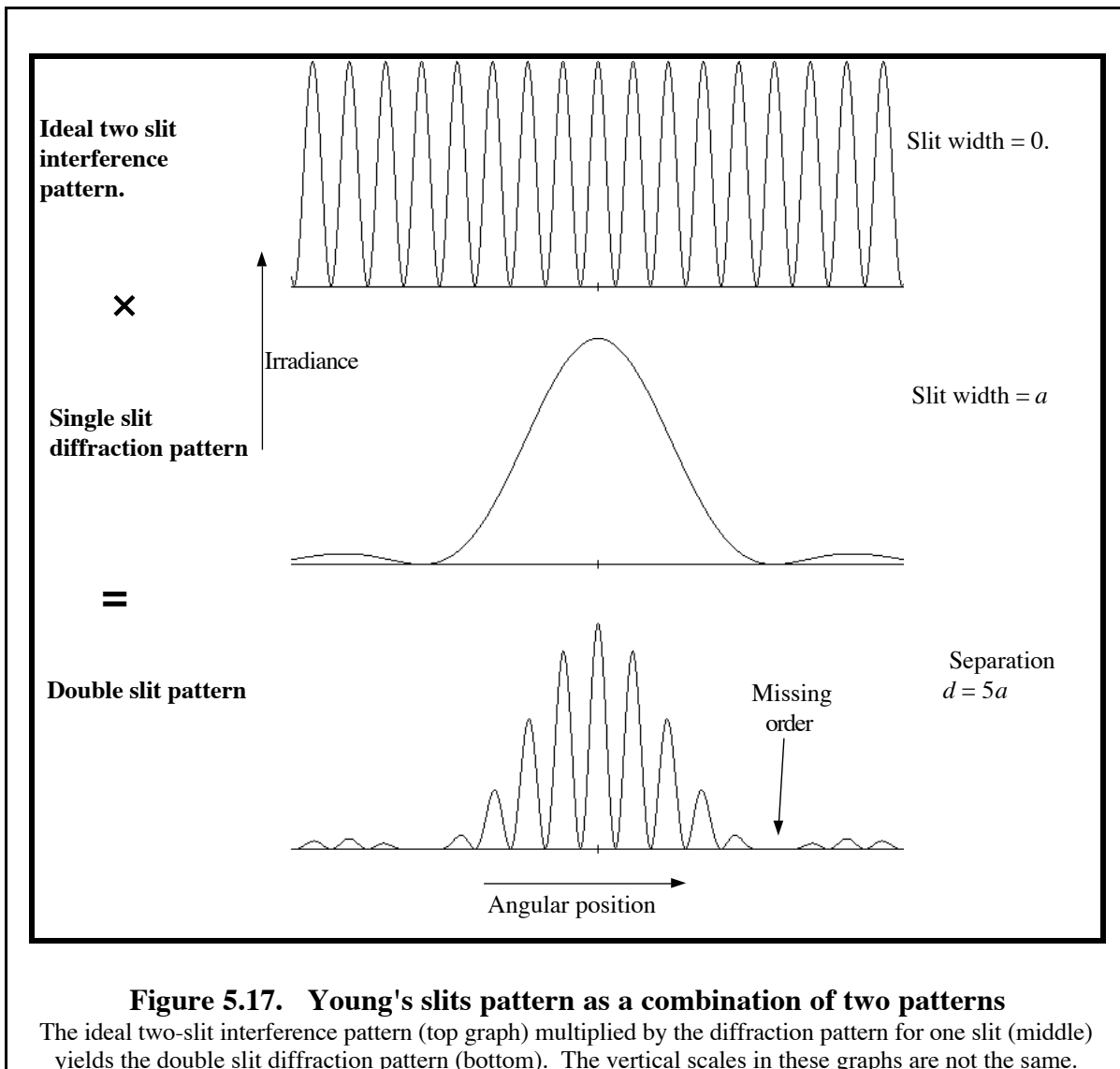


Figure 5.16. Irradiance and total power for one and two slits

Each slit is 5 wavelengths wide. These graphs are plotted on the same scale. Two identical slits give a maximum irradiance equal to 4 times the maximum irradiance from one slit. The total power, represented by the area under the curve, is only twice as large.

The interference pattern in Young's experiment can be described as a combination of the interference pattern for two very narrow slits and the pattern produced by one slit of finite width. When the mathematical forms of these two patterns are multiplied together we get the shape of the observed two-slit pattern. The spacing of the fringes is determined by the separation of the slits but their brightness is influenced by the width of a slit. Figure 5.17 shows how the interference pattern of two slits with zero width is modulated by the pattern of a single slit.



Since the description of the two-slit interference is just a combination of two simpler cases that we have already considered, it is to be expected that the results we got earlier should carry over. The locations of the interference maxima are still given by equation 4.3. For the

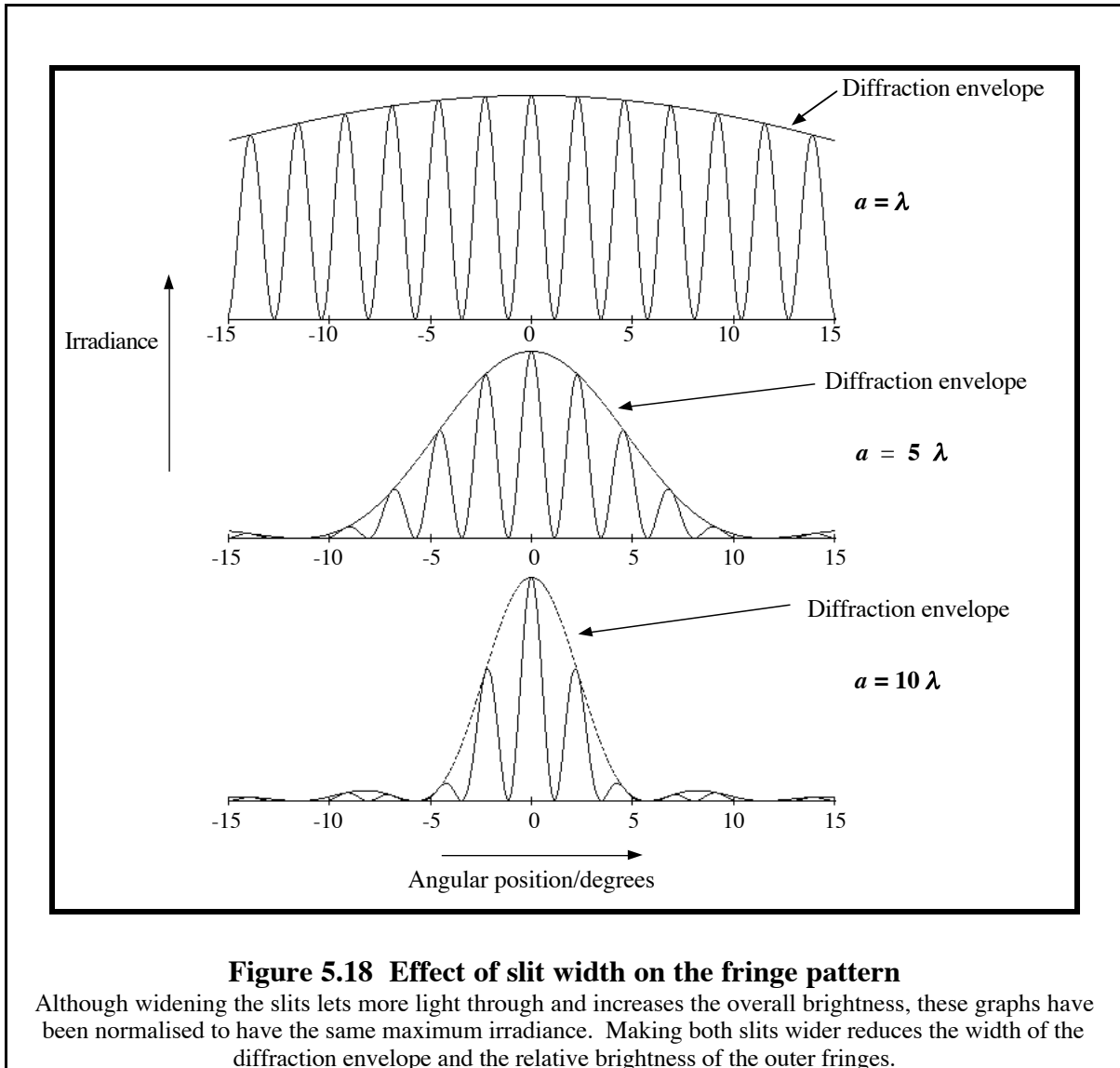
$$m\text{th order maximum:} \quad \sin\theta = \frac{m\lambda}{d} ; \quad m = 0, \pm 1, \pm 2, \dots \quad \dots (5.3)$$

And the places where there would be a zero in the single slit pattern are still dark; for a non-zero integer value of n , there is a minimum when

$$\sin\theta = \frac{n\lambda}{a} ; \quad n = \pm 1, \pm 2, \dots \quad \dots (5.4)$$

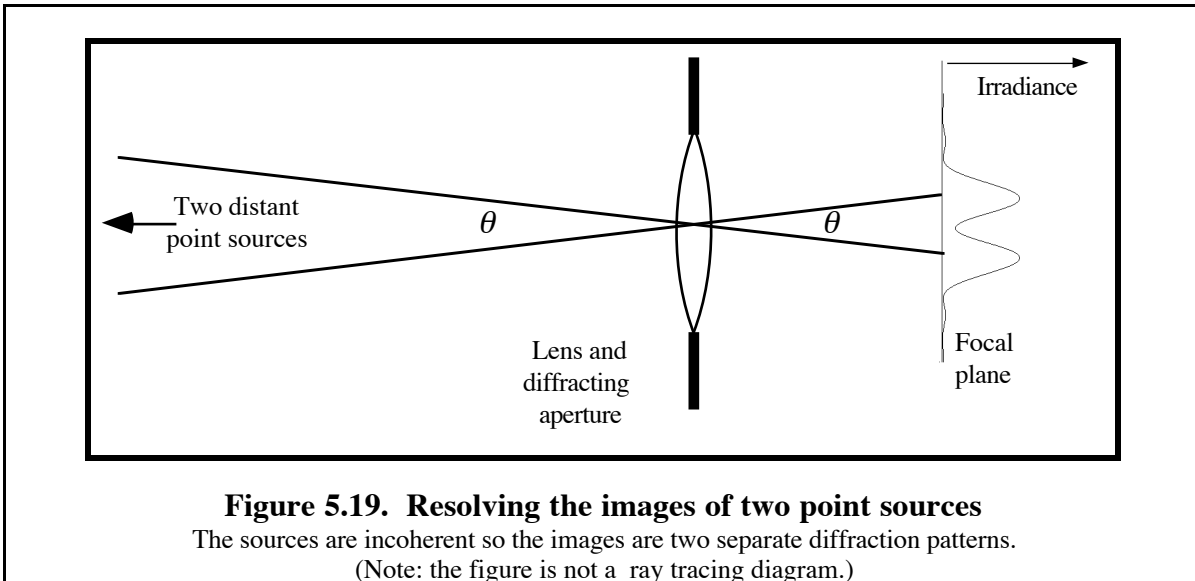
If you look again at figure 5.17 you can see that in the place where the fifth order fringe ($m = 5$) ought to be there is nothing (remember to count the central fringe as number 0). That happened because the condition for the first zero ($n = 1$) in the single slit pattern coincided with the position of the fifth bright fringe in the two-slit pattern. By comparing the two conditions (equations 5.3 and 5.4) you can see that the missing fringe was caused by the fact that the slit spacing in the example was chosen to be exactly five times the slit width. Other missing orders may be caused by similar coincidences.

Figure 5.18 shows how the single slit pattern forms an **envelope** to the double slit pattern for some different values of the slit width but the same slit spacing. You could say that the two-slit interference pattern has to be squeezed inside the single slit pattern (figure 5.18).



5-7 RESOLUTION OF IMAGES

The theory of diffraction shows that a point source of light or an object point cannot possibly produce a perfect point image even in an optical system which is free of aberrations. The wave nature of light places a fundamental limitation on the quality of an image. Resolving power is a somewhat loose term which refers to the ability of an optical system to distinguish fine detail in an image. It can be illustrated by considering how one might distinguish between the images of two stars in a telescope. Stars are so far away that they can be considered as point objects, and the fact that they often appear to have finite sizes is due to diffraction. Although a refracting telescope normally contains several lenses, the concept of resolving power can be understood by representing the system as one lens (the objective lens), and an aperture as shown in figure 5.19. The centres of the real images of the stars are formed in the focal plane of the objective lens and are separated by an angle θ which is equal to the angle subtended by the two stars at the objective. (The images are viewed using the telescope's eyepiece.)



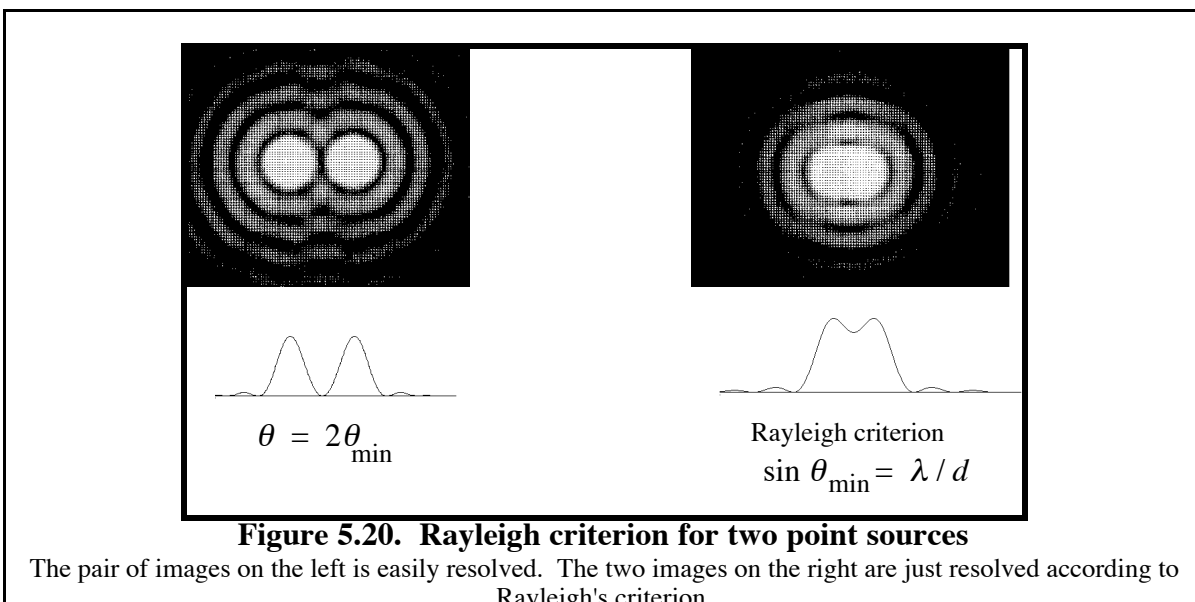
If the angular separation θ is so small that the images are effectively on top of each other, it will not be possible to recognise that there are two separate images. The **angular resolution** of an instrument is the smallest angular separation that can be distinguished. Although the decision about the actual value of an angular resolution may depend on the skill of the observer, the relative brightness of the sources and other factors, it is useful to have a generally agreed definition of when the two images of two incoherent point sources can be resolved. The criterion generally used was proposed by Lord Rayleigh (1821-1894): the two point sources are just resolved if the central maximum of one image coincides with the first minimum of the other.

For a circular aperture and Fraunhofer diffraction, you can see from equation 5.2 that the **Rayleigh criterion** is satisfied when the angular separation of the two point sources has the value θ_{\min} given by

$$\sin\theta_{\min} = \frac{1.22\lambda}{a}$$

where a is the diameter of the aperture. In practice, since the angles involved here are always very small we can put $\sin\theta = \theta$ and since 1.22 is near enough to 1, the Rayleigh criterion for a circular aperture is that

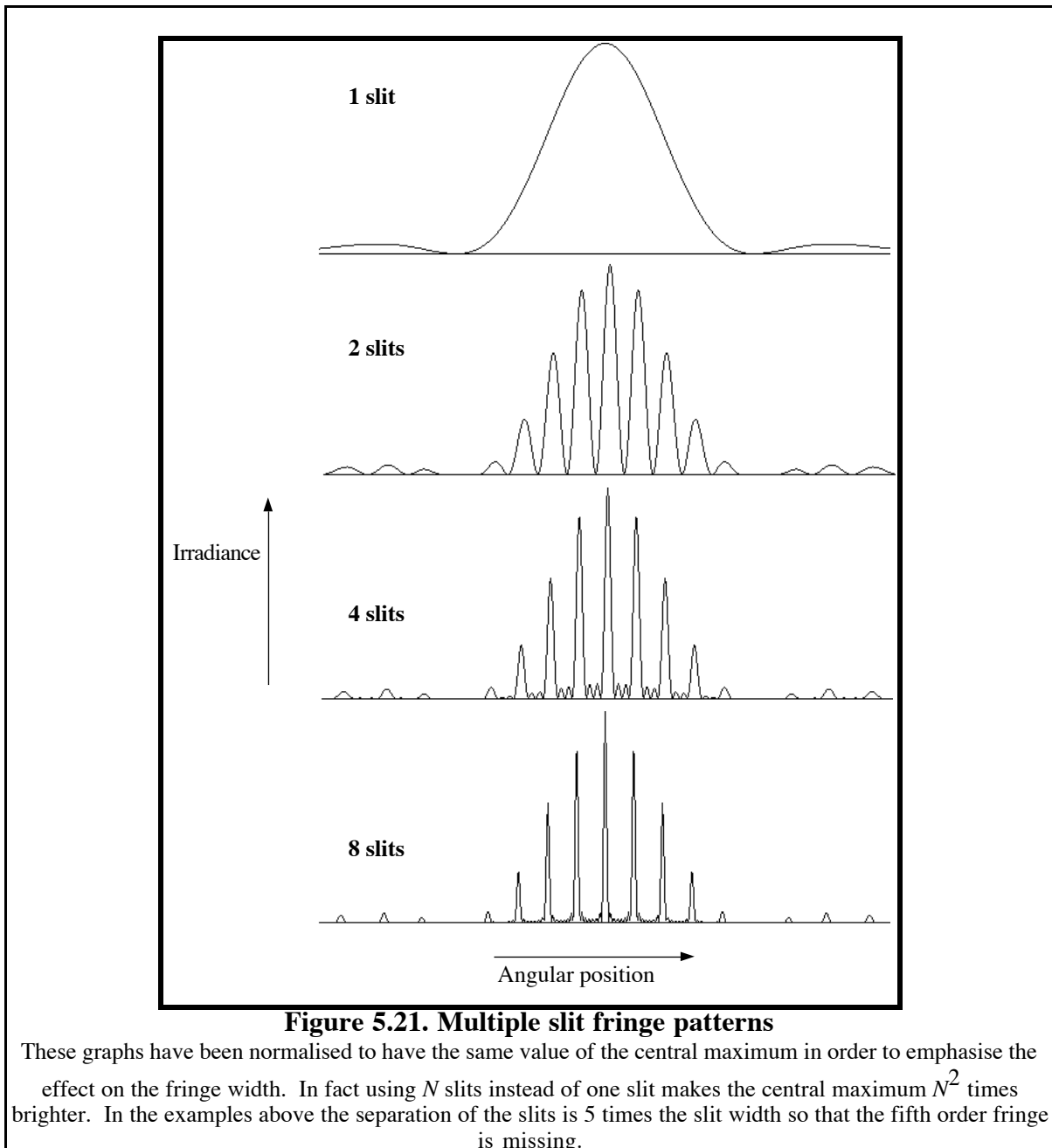
$$\text{angular resolution} = \frac{\lambda}{a} .$$



It follows that to improve resolving power one can use a shorter wavelength, which is usually not possible, or a larger lens which gives a larger aperture. That is one of the reasons that modern astronomers need large telescopes.

The discussion above applies directly to telescopes and other optical instruments which are used to look at incoherent sources. However the details do not necessarily carry over to all kinds of microscopy, because the illumination of adjacent parts of a specimen may be at least partially coherent. The analysis of such cases is more difficult, but the general idea that resolution can be improved by using a larger aperture or a shorter wavelength remains valid.

5-8 DIFFRACTION GRATINGS



Now that you have seen how a double slit pattern is formed, it is interesting to ask what happens if the number of equally spaced slits in a Young's experiment is increased. There are two main effects. Firstly, and fairly obviously, the whole diffraction pattern becomes brighter because more light gets through. The second effect is more surprising - the bright fringes get sharper! Although the positions of maximum irradiance remain unaltered the width of each bright fringe decreases as the number of slits increases. (A third, but less noticeable, effect is that some new weak fringes appear in between the principal maxima.)

The reason that the principal fringes stay in the same locations is that the conditions for maximum irradiance are still essentially the same - the optical path difference from adjacent slits to the point on the screen needs to be a whole number of wavelengths and that condition does not depend on how many pairs of slits you have. On the other hand, if there are many slits then the condition that all possible adjacent pairs of slits must give the same path difference is much tighter. Consider a point on the screen just a little bit away from a peak. The corresponding points on the wavefront at two adjacent slits do not have the exactly right phase difference for a maximum and the phases of the waves from other slits will be even more out of step. The more slits you have the worse the matching of the phases will be.

Although the fringes are much sharper their positions are still described by the same Young's slits equation (5.3).

The sharpening of the fringes is exploited in the **diffraction grating**, a device which consists of a very large number of narrow, uniformly spaced, parallel slits (typically 1200 slits per millimetre). There are two kinds of grating. A **transmission grating** is made by cutting grooves in a material such as glass; the grooves are effectively opaque strips and the unruled portions are the slits. A **reflection grating** works by reflecting light from many parallel mirror-like strips.

Since the fringe spacing depends on wavelength and because the bright fringes are very sharply defined, the diffraction grating can be used in a spectroscope or spectrograph to spread a beam of light into components with different frequencies. The grating is far superior to the prism because it can be made to give a much greater angular separation of the spectrum. If the slit separation is very small (say 10^{-6} m) and if there are many slits (say 10^6), then a line spectrum will have very sharp, clearly-separated interference maxima.

THINGS TO DO

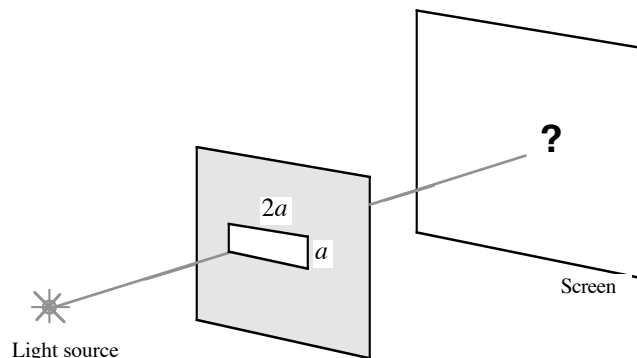
- The grooves on an LP record or (better) a compact disc can function as a reflection diffraction grating. Look at various sources of light reflected from the surface of a disc. How many orders of the diffraction pattern can you see? Can you estimate the spacing of the grooves?
- You can use a piece of finely woven cloth as a kind of two-dimensional transmission grating. Look through the cloth at a mercury or sodium street light.

QUESTIONS

Exercises

Q5.1 Estimate the width of the central maximum of a single slit diffraction pattern which appears on a screen 1.0 m away from a slit of width 0.10 mm, illuminated with light of $\lambda = 500$ nm.

Q5.2



Using the result that the angle between the centre and the first minimum of a single long slit diffraction pattern is $\sin \theta = \frac{\lambda}{a}$, can you predict what the diffraction pattern of the rectangular aperture in the figure will look like?

- Q5.3** Estimate the angular width of the central maximum of the diffraction pattern produced by a circular aperture 2.0 mm in diameter. Take $\lambda = 500$ nm.
- Q5.4** The two headlights of a distant approaching car are 1 m apart. Make a rough estimate of the distance at which the eye can resolve them. Take a pupil diameter of 5 mm and use a typical visible wavelength, say $\lambda = 500$ nm.
- Q5.5** A grating with 400 grooves per millimetre is used to examine the spectrum of a source of light.
- Calculate the angle between the first order maxima for red light with $\lambda = 700$ nm and violet light with $\lambda = 400$ nm.
 - Do the same for the two components of the sodium-D line which have wavelengths 589.0 nm and 589.6 nm.

Discussion questions

- Q5.6** Can you explain why the centre of the diffraction pattern for a circular obstacle always contains a bright spot?
- Q5.7** A grating will not be of much use for producing a clear spectrum if the first order fringes of some part of the spectrum overlap the second order fringes of some other part. Discuss. How should the grating be constructed in order to avoid the problem?
- Q5.8** Why is diffraction more noticeable for sound than for light?
- Q5.9** Light waves do not bend noticeably around buildings, but radio waves which are also electromagnetic waves do diffract around buildings. Discuss.
- Q5.10** What do you think of the claim that you can't see an interference pattern in a Young's experiment if the separation of the slits is less than half a wavelength?
- Q5.11** What would you see in a Young's two-slit experiment with white light?
- Q5.12** Why is a grating better than just two slits in an experiment to measure wavelength?

APPENDIX

MATHEMATICAL DESCRIPTION OF THE DIFFRACTION PATTERNS

The shape of the Fraunhofer intensity distribution curve for a grating with N slits is described by the equation:

$$I = A \left(\frac{\sin \alpha}{\alpha} \right)^2 \left(\frac{\sin(N\beta)}{\sin \beta} \right)^2$$

where

$$\alpha = \frac{\pi a}{\lambda} \sin \theta$$

which is half the phase difference between one edge and the middle of a slit

and

$$\beta = \frac{\pi d}{\lambda} \sin \theta$$

which is half the phase difference between corresponding points on adjacent slits. A is a constant which represents the amplitude of the incoming wave - assumed to be the same at all slits.

The single slit diffraction pattern ($N = 1$) and the Young's double slit case ($N = 2$) are also described by the same equation.

The envelope (single slit pattern) is described by the term $\left(\frac{\sin \alpha}{\alpha} \right)^2$ while the idealised pattern for N slits of zero width is given by the term $\left(\frac{\sin(N\beta)}{\sin \beta} \right)^2$.

The results for the zeros in the diffraction envelope (equation 5.1) and the condition for the maxima (equation 5.3) can be obtained from the equation above.