Diffraction

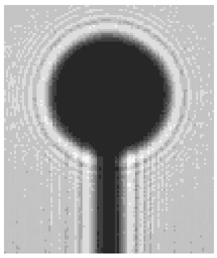
This chapter includes some mathematical detail that is intended for those who are taking Hons physics, or others who have taken level 1 maths. You will get the most out of this chapter if you already appreciate Huygens' Principle, the phenomenon of interference of waves and the underlying idea behind phasors.

Diffraction is a phenomenon that you have to look hard to find but which is always there. It exerts a big influence on what you can do with light, for example controlling the amount of detail that can be seen through a microscope, a telescope and, indeed, any well-made optical instrument. Diffraction is the phenomenon at the heart of viewing holograms and multi-coloured 'holographic' designs; it is central to the concept of optical image filtering and, as we have indirectly seen, makes diffraction gratings possible. It gives rise to natural phenomena like coronae around the moon, the colours of iridescent clouds, the Spectre of the Brocken and colouring on some butterfly wings.

Diffraction was discovered and named by Franceso Grimaldi, who lived in the first half of the seventeenth century. He experimented on light passing through very small holes, through structured objects like linen mesh and bird's feathers. He particularly examined the fine detail he saw at the edge of shadows. He published his results not long before his early death at the age of 45 in a book known in abbreviation as *de Lumine* (on Light). However, his results seemed to be hard won and the phenomena did not receive the attention they might have done from his successors, who found other more obviously interesting subjects. His results were, as one commentator put it, a rock upon which many great people's ideas foundered, or would have done so if they hadn't steered clear. For the modern scientist, the man who put diffraction on the chart was the perceptive and immensely able Frenchman, Augustin Fresnel. We've met him before.

What is diffraction?

Diffraction is the spreading out of light from its geometrically defined path. Diffraction happens for all kinds of wave phenomena and its presence in optics is a fundamental demonstration that light behaves like a wave. When water waves travel onto a projecting stone and you see circular waves radiating from the stone, that is diffraction. When waves come into a harbour mouth and bend around so as to rock boats that seem to be in the shelter of the harbour walls, that is diffraction (and perhaps some reflection). When you can hear a concert almost equally well from a seat behind an obstructing pillar, then at least part of the reason is diffraction, along with reflection from around the hall.

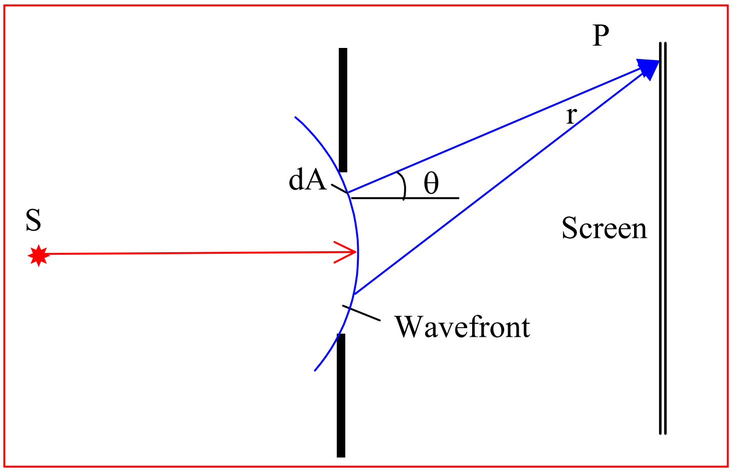


In optics, diffraction, like interference, tends to produce patterns of illumination with bands of bright and dark light. The picture alongside shows the diffraction pattern around the shadow of a pin-head. The reason for the bands is basically that diffraction and interference are much the same thing. Both phenomena arise from the coherent addition of waves coming from more than one effective source. The only difference is that diffraction patterns are what you get when waves from a continuous distribution of sources add up; interference is the term reserved for adding up waves from 2, 3, or at least a discrete number of sources.

The Huygens - Fresnel - Kirchhoff theory

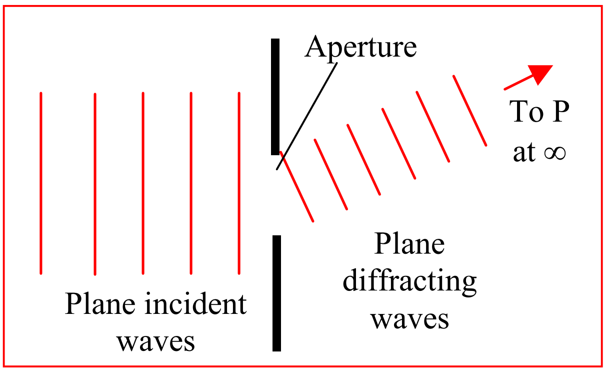
The name, if nothing else, is a bit of a mouthful. The basic idea that explains diffraction is based on Huygens' theory. However, you may remember that Huygens' theory makes no mention anywhere of the wavelength of light, and wavelength is central to diffraction. Fresnel took Huygens' underlying idea and described what was happening using the appropriate mathematics, which was quite a recent development in Fresnel's time. He introduced the wavelength of the light and was able to deduce the appearance of diffraction patterns in remarkably good agreement with some pains-taking observations he himself had made. Kirchhoff, later in the 19th century, realised that one or two of the terms in Fresnel's theory weren't quite right and there was some incompleteness in it. A modern student of optics would start with Kirchhoff's theory.

Diffraction occurs when the wavefront of advancing light is partially blocked. The simplest example is blocking by the edges of a slit or aperture. A slit is tall and thin; apertures are rectangular or circular in many cases. On a plane diagram, they are represented as a gap.



In the illustration, a source S produces a wavefront that reaches the aperture. Huygens said, you'll remember, that each point on the wavefront acts as a source of wavelets that determines where the wavefront will be later. Fresnel went along with this and added that at a point P on the screen you must therefore add up all the wavelets reaching P from every tiny area dA of the wavefront. [In 3 dimensions the wavefront covers an area]. Each infinitesimal area dA contributes a wave like $E\cos(kr - \omega t)$ whose amplitude, E, falls off as the distance from the small area and whose inclination is determined by the angle θ, namely how far off to the side P is from the straight-through position. The sum over the whole aperture is represented by an integral. It looks complicated. It is quite complicated. That Fresnel managed to get anywhere without any guidance from anyone else says a lot about his analytic ability.

The diffraction patterns that are predicted in general by this approach are called **Fresnel diffraction**. I hope you get the idea behind what is being done. We're not going to work out any Fresnel diffraction patterns. We're going to consider a special, simpler, but very important case - **Fraunhofer diffraction.**



Fraunhofer diffraction

Fraunhofer diffraction is named after the same optical instrument maker who discovered the dark lines in the solar spectrum. It is diffraction in the following circumstances:

1) The **source of light is a long way away** "at infinity", or effectively so. This means that the waves arriving at the aperture are plane waves. We usually take the plane wavefronts as parallel to the aperture, so the whole aperture is filled with a wavefront at once. This means that all the 'secondary sources' on the wavefront emit in-phase across the aperture.
2) The **screen is a long way away** "at infinity", or effectively so. This has two implications. First, the angle θ at which a ray travels to get to P is the same right across the aperture. Secondly, the difference in amplitudes of the waves coming from opposite ends or sides of the aperture is negligible.

The implication of these two special circumstances is that the **phase change in the received light from across the aperture varies linearly with the coordinates across the aperture**. So, for example, if y measures the distance up the aperture, then in Fraunhofer diffraction the phase change varies in proportion to y, not y^2 or y^3, just y.

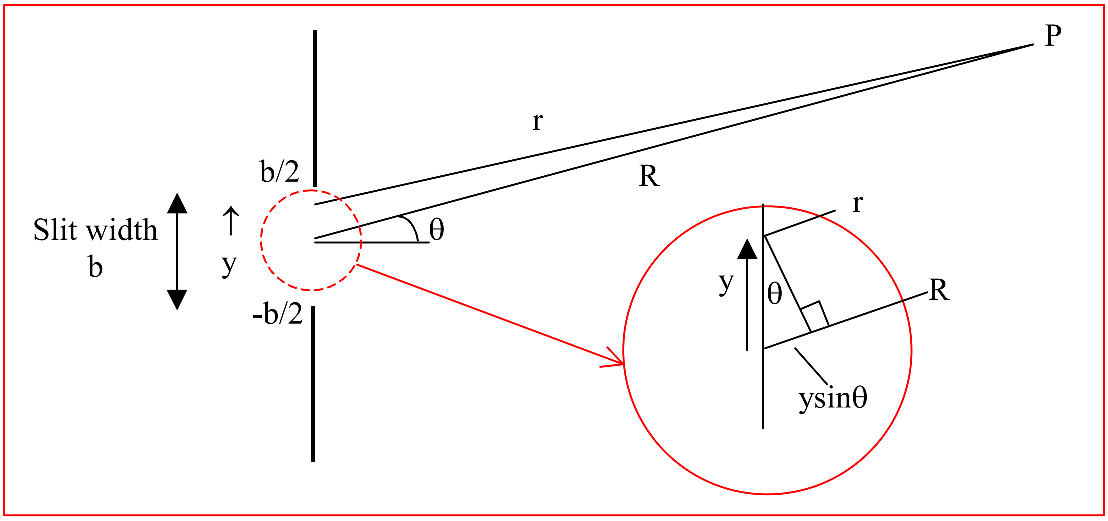
Mathematically, the total electric field, E_P, at the point P on the screen is given by

$$\boxed{E_P \propto \int_{aperture} \cos(kr - wt)\, dA} \quad .$$

The solution to working out how much Fraunhofer diffraction is produced with a given aperture is to work out this integral. If you have a good mathematical nose, you'll smell the word Fourier at this point. Working out the Fraunhofer diffraction pattern is 'just' a matter of working out a Fourier transform. Remember that, as always, the irradiance, I, depends on the square of the field:

$$\boxed{I \propto \left\langle E_P^2 \right\rangle} \quad .$$

Origin of Fraunhofer diffraction from a slit



What does the diffraction pattern look like? It consists of a brightly

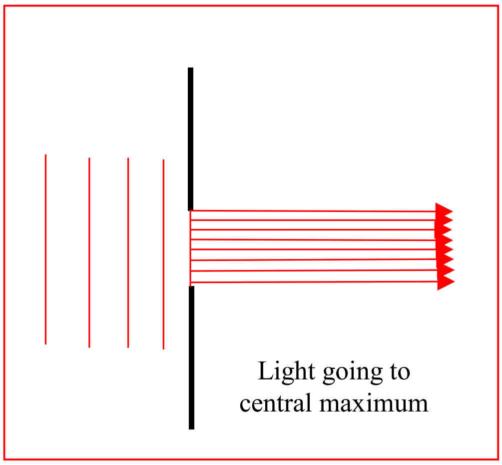
slit

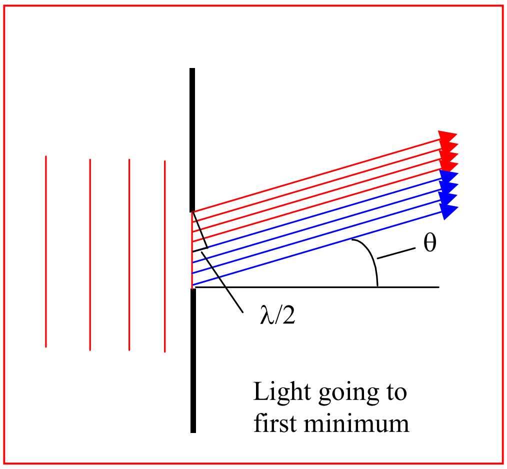
Diffraction pattern

illuminated central part with some much fainter bands on either side. The narrower the slit, the broader but fainter is the central band. To see what's going on we have to look at the phase shift of the light coming out of various parts of the slit, as measured perpendicular to the diffracted pencil of light.

The slit has height b. y is the coordinate up the slit, measured from the middle. In the full width diagram on the previous page, what is happening in the vicinity of the slit is shown in the enlarged region within the red circle. The picture to have in mind is that each point within the slit is radiating light. We look in detail at all the light reaching P. Look across the slit at the light coming out at angle θ. Compared to the light coming from a point on the slit at height y, the light from the centre of the slit has to travel the small extra distance of $y\sin\theta$. Put more naturally, compared with light from the centre of the slit, light coming from point y has an extra path difference of $-y\sin\theta$ (see the diagram on the previous page). The effect of this has to be added up right across the slit. In mathematical terms, the waves reaching P must be integrated (i.e. summed) for all values of y from $-b/2$ to $+b/2$. Those taking PX2014 may be able to do the integration. However, we don't need to pluck a mathmatical result out of thin air. There is an easier way to see more or less what's going to happen.

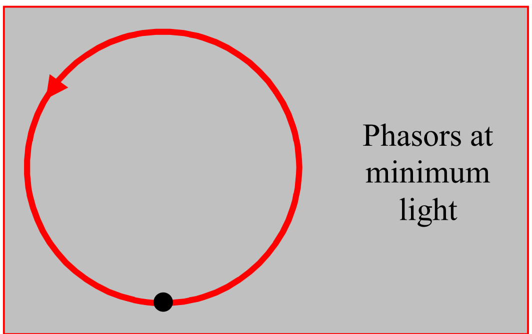
In the straight-through direction, all the light coming through the slit is in-phase. It all constructively interferes and the total amplitude depends on the width of the slit, b. If you'd like to think in terms of phasors, imagine the slit divided into very many equal portions, say 100 or even many more. The phasors representing the electric field from each portion will all point in the same direction and when you add them end to end you will naturally get the





maximum amplitude possible.

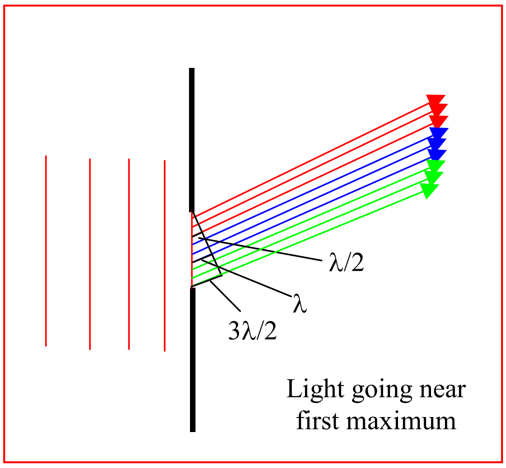
As you look away from the straight-through direction, θ gets bigger and there comes a point where the light from the centre of the slit is exactly half-a-wavelength, λ/2, out of phase with light coming from the top of the slit. It therefore interferes destructively. Likewise, light from a little below halfway interferes destructively with light from a little below the top, and so on right the way down to the bottom of the slit. In effect, light from the lower half of the slit cancels out the light from the top half of the slit and the sum total is zero. This gives a dark minimum (of zero) in the diffraction pattern. Therefore, you can deduce that the *angle at which the first minimum occurs* is given when:



$$\boxed{\begin{gathered}(b/2)\sin\theta = \lambda/2 \\ or \\ b\sin\theta = \lambda \quad .\end{gathered}}$$

If you like the phasor picture, you can see what is happening in terms of the phasors that represent the sum of the waves coming out of the slit. The total phase difference from top to bottom of the slit is 2π. The corresponding phasor diagram has curled up into a complete circle, with no resultant value of E.

At a larger angle, light from the first third of the slit will be cancelled out by light from the second third of the slit, leaving uncancelled light from the final third. This angle gives more or less the next maximum beyond the first minimum. The colours in the adjacent diagram are just there to highlight the thirds involved.



Light going near first maximum

A second minimum occurs when light from the first quarter is cancelled by light from the second quarter, and similarly for light for the next two quarters. The phasors have now gone around a circle twice. You can see how a succession of zeros is produced with small maxima about half-way in between. The minima occur when:

$$\boxed{b\sin\theta = n\lambda} \quad ,$$

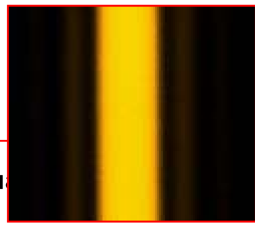
where n is an integer.

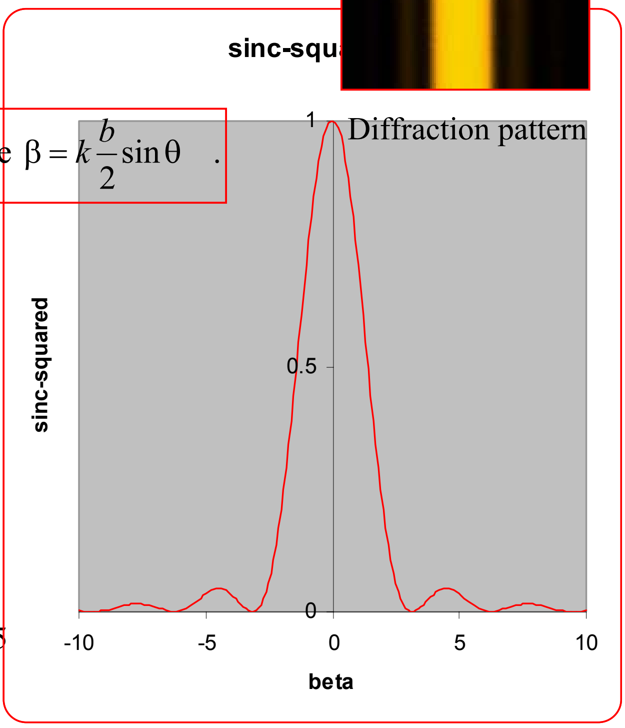
For any angle, the amplitude of the result will scale as the slit width. Hence the irradiance of a diffraction pattern depends on the slit width squared, i.e. b^2.

Appearance of Fraunhofer diffraction from a slit

The result of doing the integration on page 3 over a slit (it is only an integral in the one variable, y) is this:

$$\boxed{E_P \propto b\frac{\sin\beta}{\beta}\ ,\ \text{where}\ \beta = k\frac{b}{2}\sin\theta \quad .}$$





Only the term sinβ/β, written as sincβ and pronounced "sink beta", gives the change in diffraction pattern with θ. Now, (sinβ/β) = 1, when β→0, therefore, the irradiance

$$\boxed{I = I(0)\left(\frac{\sin\beta}{\beta}\right)^2 = I(0)\mathrm{sinc}^2\beta}$$

is given by:

I(0) is the irradiance at the centre of the pattern. The function $\text{sinc}^2\beta$ has a strong central peak and, as you can see from its definition, equi-spaced zeros when $\beta = n\pi$ ($n \neq 0$). This is just the sort of variation we deduced earlier.

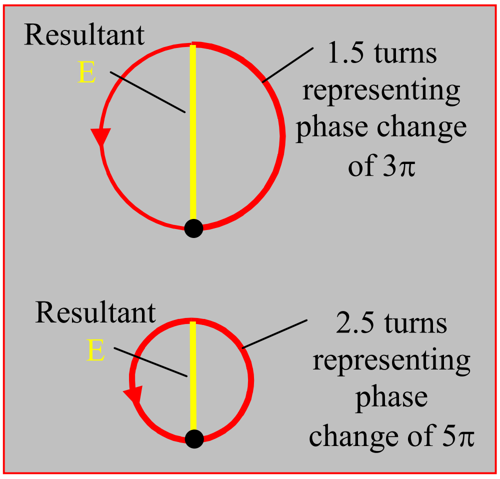
The *zeros of the diffraction pattern occur when*:

$$n\pi = \beta$$
$$= (kb\sin\theta)/2$$
$$= \pi b\sin\theta/\lambda$$

i.e. $$\boxed{b\sin\theta = n\lambda}$$.

This is exactly the expression we got from the 'intuitive' argument in the previous pages. The first zero is particularly significant because it defines the width of the central maximum, where most of the intensity is located. The first zero occurs when $b\sin\theta = \lambda$. For slits of width > 10λ, this occurs at small angles given by $\theta \approx \lambda/b$, with θ in radians.

The first subsidiary maximum is ~4.7 % of the main maximum and the next maximum ~1.7%. The subsidiary maxima occur about halfway between the minima and are half as wide as the main maxima (as a function of β).

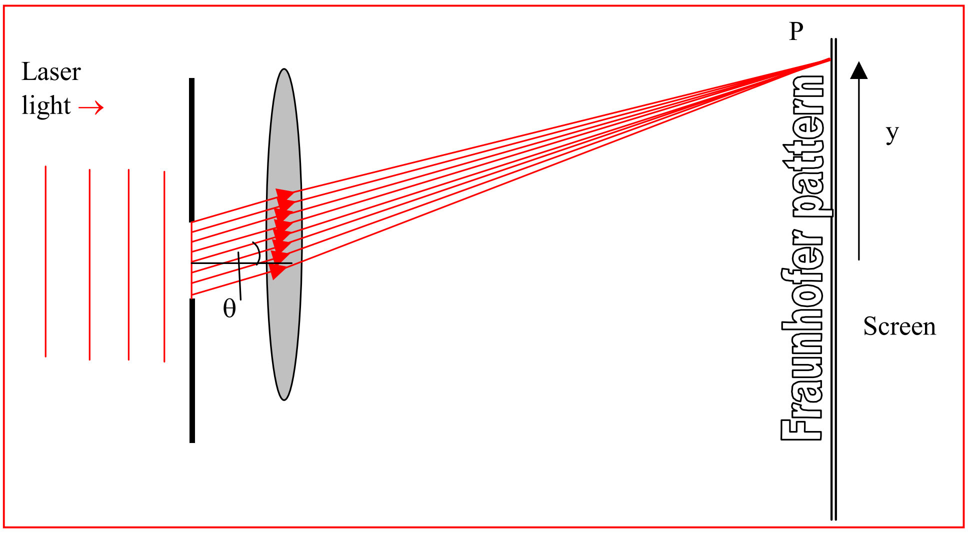


You don't need to be an expert in trigonometry to see that these figures make sense, if you have got hold of the idea of phasors. When the path difference across the slit is 3λ/2, as in the earlier figure, then the phase difference is 3π. The phasor diagram for the addition of waves across the slit has changed from being a straight line in the straight-through direction to curling around in a circle one-and-a-half times. Let's say the circle has radius r. The amplitude is represented by $2r$ instead of the complete length of $3\pi r$. Hence the relative irradiance is $(2r/3\pi r)^2 = 0.044$, or 4.4%, close to the value found by locating the first secondary maximum of the sinc^2 function. Likewise, you'd expect the second maximum to have an irradiance of $(2r/5\pi r)^2 = 0.016$, or 1.6%. Compare these figures with the accurate ones given above. They are very close.

We'll come back to consequences of the sinc^2 shape later. Notice one very important feature at the moment. The narrower the slit, the wider is the diffraction pattern. It is just this effect that makes a CD glint all over with diffraction colours. The little pits on the surface that code the information are very small indeed and diffract the light over very wide angles. An old-fashioned vinyl record where the grooves are much farther apart shows much narrower diffraction and you have to sight along the record edgewise to get reasonable diffraction colours.

Making Fraunhofer diffraction happen

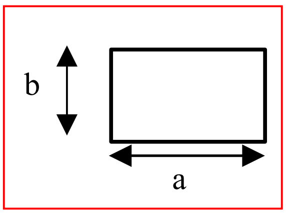
Those who will be taking the practical course PX2505 will get a chance to set up and take photographs of Fraunhofer diffraction patterns from a range of objects. How can you get any reasonable irradiance on a screen if the source and screen are both at infinity? The answer is that we use an optical 'trick'. A single lens converges a parallel pencil of light to a point in its rear focal plane; it also creates a parallel pencil from a point source in its front focal plane. Hence to see



Fraunhofer diffraction using a point source you need to place the source in the front focal plane of a lens, then put the aperture in the beam, then place a second lens after the aperture and look in its rear focal plane. In the lab, you'll use an expanded laser beam so the diagram is simplified to that shown above.

Rectangular aperture

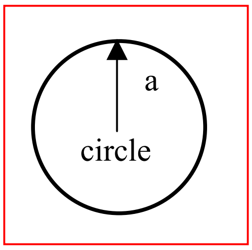
The rectangular aperture of dimensions a×b is essentially a set of slits stacked up at right angles to their length. The aperture looks like a slit if you take a section parallel to any side and you will not be surprised to find that the diffraction pattern is given by the product of two slit diffraction patterns, arranged at right angles to each other. Put another way, the phase change vertically (y) is the same for all point across (x) the slit. Hence, the vertical and horizontal contributions to the diffraction pattern are independent.

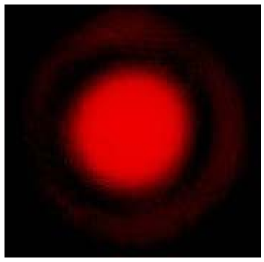


Again, *for the mathematical*, a point P on the observing screen is now located by two angles, θ in the direction of b and ϕ in the direction of a. Hence, defining α in a similar way to β as $\alpha = k\frac{a}{2}\sin\phi$, the diffraction from an aperture is given as $I = I(0)\left(\frac{\sin\alpha}{\alpha}\right)^2\left(\frac{\sin\beta}{\beta}\right)^2$.

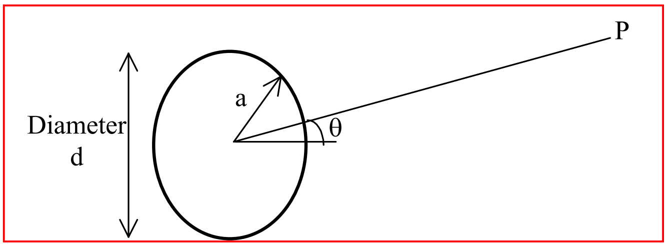
Notice again the reciprocal relationship between the size of the aperture and the spread of the diffraction pattern. Small slits have widely spreading diffraction patterns, and *vice-versa*. The scale against which small and large are measured is the wavelength of the radiation.

Circular aperture





A circular aperture is defined by a single parameter, its radius *a*. The diffraction pattern will be circularly symmetric, meaning that it will be a set of concentric rings. The resulting intensity pattern is called the Airy disk, after George Airy, a famous 19th century Astronomer Royal who first worked it out.



For the mathematical

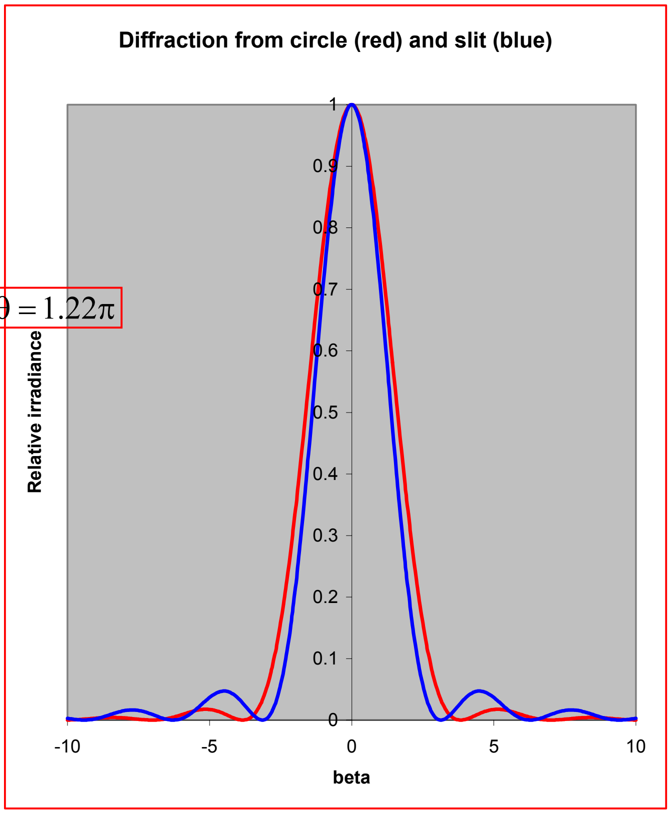
Let θ be the angle away from the straight-through direction. The mathematical will not be surprised that the diffraction pattern is given in terms of one of the Bessel functions, since they occur very frequently in problems involving circular symmetry.

J_1 is the Bessel function of the first kind of order 1. It is well known, in the sense that it can be calculated by well-known means such as with an Excel spreadsheet or even looked up in tables. The accompanying graph was evaluated using Excel to compare the diffraction from a circle and a slit of the same width.

$$I = I(0)\left(2\frac{J_1(ka\sin\theta)}{ka\sin\theta}\right)^2$$
$$= I(0)\left(2\frac{J_1(\beta)}{\beta}\right)^2 \text{, where } \beta = ka\sin\theta \ .$$

The intensity function looks pretty like $\mathrm{sinc}^2\beta$. Because a circle of radius *a* is smaller in area than a square of half-side b/2, you might quite correctly expect that the circle diffraction pattern is a little more spread out. The first minimum of the Airy pattern occurs when:

$$ka\sin\theta = 1.22\pi$$

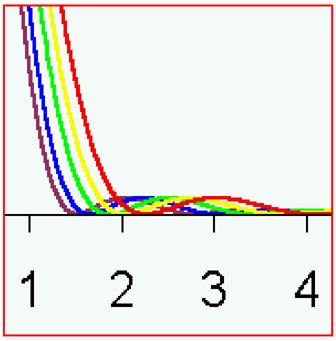


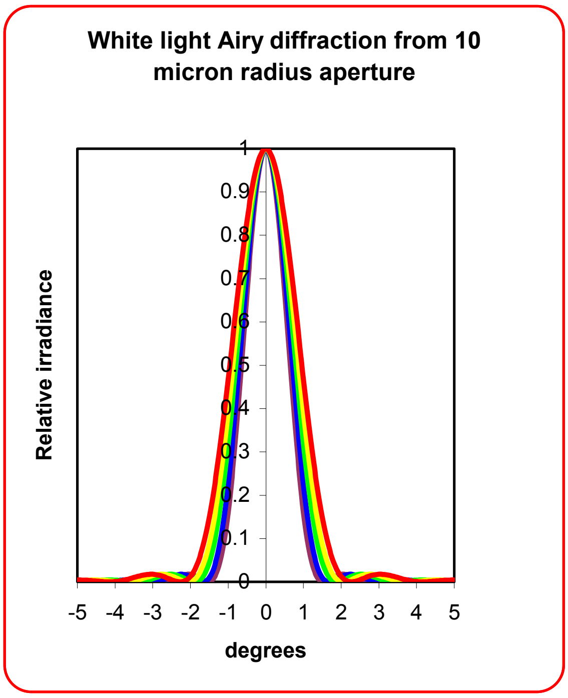
i.e. at $\sin\theta = 1.22\lambda/2a = 1.22\lambda/d$, where d is the diameter of the circle. *Hecht* has good pictures.

How big is the pattern? Take a representative wavelength as 550 nm (green light). For a hole width of 1 mm, the angle to the first minimum is $\sin^{-1}(1.22\times550\times10^{-9}/1\times10^{-3}) = 0.038°$ or 2.3 minutes of arc, not very big at all. Now consider another example. If you measure a circular diffraction pattern and find the first minimum is at 2°, then the aperture diameter must be $d = (1.22\times550\times10^{-9}/\sin(2°) = 1.92\times10^{-5}$ m = 19.2 μm.

Even though diffraction patterns from large openings are very small, they aren't necessarily unimportant. In most optical instruments, it is the diffraction pattern from a circular aperture that sets the ultimate limit to the resolution. Hence, the size of the central maximum from a circular aperture is the most useful of all information to remember.

Diffraction in white light

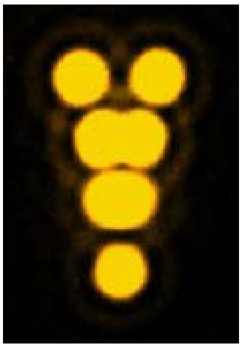




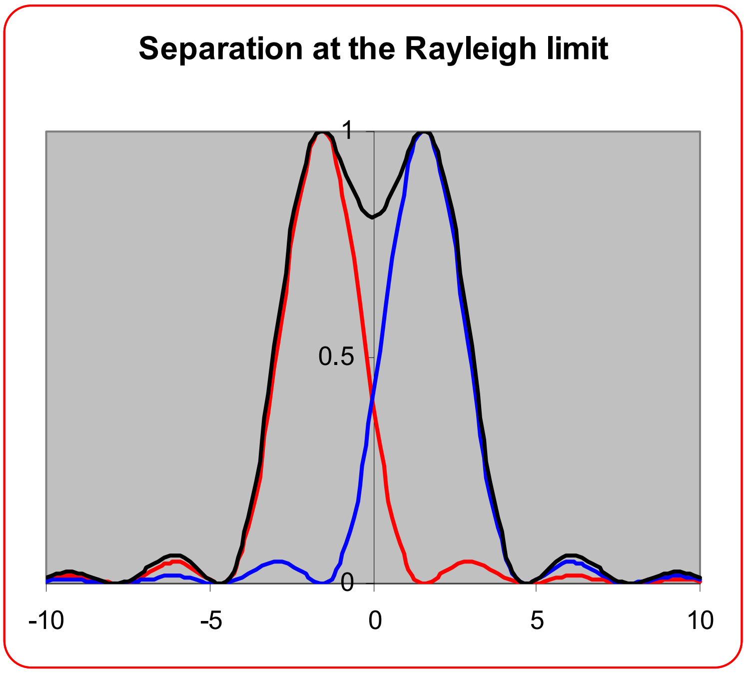
Finally, notice another very important effect. Violet light has a shorter wavelength than red light and hence its diffraction pattern will be more compact. If you look at the diffraction pattern in white light, containing all of the spectrum, then you expect to see the spectrum overlapping in the middle to create white. The violet goes to zero leaving the middle and top end of the spectrum visible, before rising to its first maximum as the red goes to zero. The accompanying illustrations show an actual picture and the corresponding irradiance patterns. You shouldn't be surprised to see a ring of almost rainbow colours around the centre white patch. This has relevance to the corona phenomenon discussed shortly.

Rayleigh criterion for diffraction-limited resolution

Lord Rayleigh, a century ago, gave a simple argument to estimate the diffraction-limited resolution of an optical instrument limited by Fraunhofer diffraction. This is the case for most observing instruments with the best quality lenses. His argument is still a very useful, quick, estimate that gives almost the right answer. You can do a bit better than the Rayleigh limit if you know precisely how your instrument spreads out light from a point and you have some knowledge of what you are looking at.



The resolution of an instrument is a measure of how close together two objects are when their blurred images can just be distinguished. Assume that the optics are good enough that the lens imperfections do not spread out the images significantly and you are just left with diffraction spreading.



The two sinc^2 curves in the figure represent the diffraction images from two neighbouring point objects. They are separated such that the maximum of one curve lies over the first minimum of the other. The third (black) curve is the sum of the two intensities. You can see that there is a discernible dip in the middle, allowing you to distinguish the two objects. *This configuration is the Rayleigh criterion for resolution*. Since the diffraction pattern from lenses will be small, the Rayleigh criterion becomes a limit in the angular separation $\Delta\theta$ of two adjacent sources in terms of the lens diameter *d*:

$$\boxed{\Delta\theta = 1.22\,\lambda / d}$$

For example, the entrance pupil of a telescope limits the resolution observable for closely spaced objects. The entrance pupil should be the diameter of the objective and hence '*d*' will then be this diameter. How big a resolution does this typically produce? A 100 mm diameter objective will have a diffraction-limited resolution at a wavelength of 500 nm of $\Delta\theta = 1.22\times5\times10^{-7}/1\times10^{-1} = 6.1\times10^{-6}$ radians = 1.26″ arc.

The corona around the moon

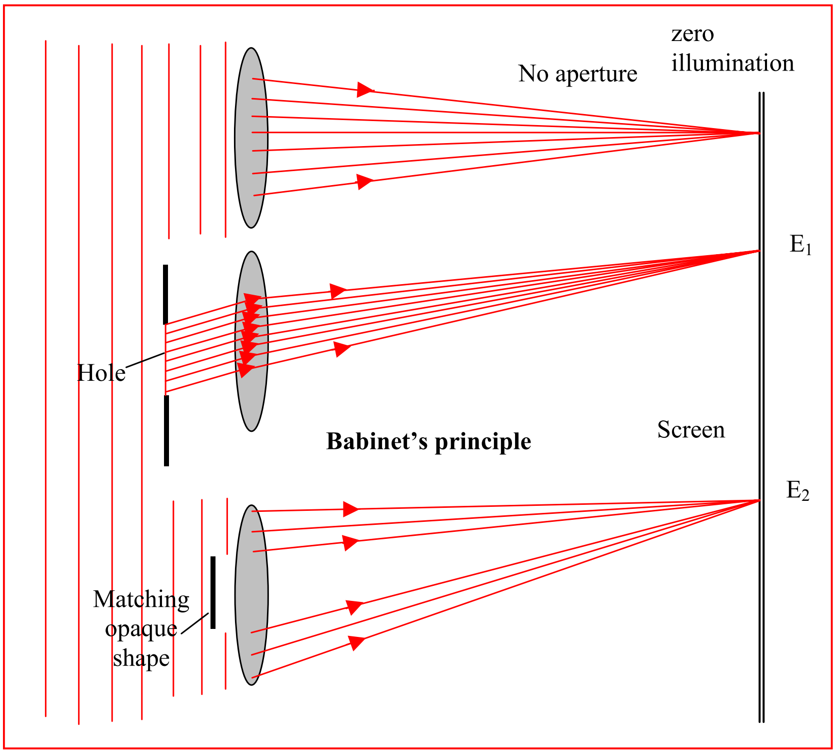
When you look up at the moon showing slightly hazily through altostratus clouds, you may well see a coloured corona that fluctuates a bit as the clouds drift across the moon. It has been described loosely as looking like a 'target in the sky'. The colours are quite delicate, the outer part usually appearing a brownish red. If you are lucky, you may see two or even three rings. This pattern is essentially an Airy disc, writ large in the sky.

Why do we see a diffraction pattern on such a big scale? Why is it like an Airy disk? The answer to the second question lies in the optics; the answer to the first question lies in the geometry of the circumstances.

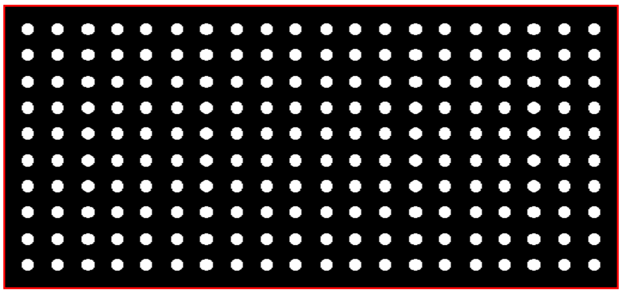
The diffraction pattern is produced by myriads of tiny, spherical, water droplets in the cloud. In cross-section, they are circular. Due to *refraction,* a water droplet deviates virtually all the light shining on it out of its original path. Water droplets are therefore essentially opaque to the direct passage of light and as potential diffraction objects might just as well be a cloud of circular dust particles. If they are small, they will diffract light conspicuously. Now for the clever bit. There is a simple argument, due to Babinet, that says the Fraunhofer diffraction pattern from a circular opaque disk in a clear background will be just the same as the pattern from a matching circular aperture in an opaque background. The argument is quite general and is called Babinet's principle.

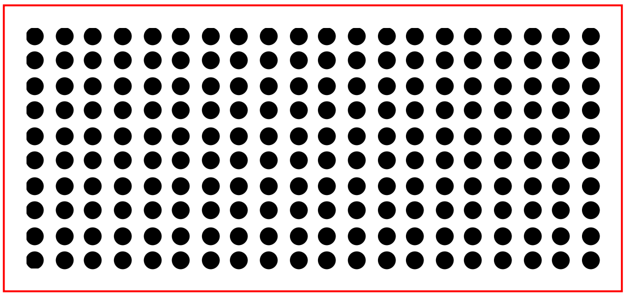
Babinet's principle

The diagram alongside repeats the set-up for forming a diffraction pattern. It shows three separate figures: the illumination when there is no obstruction; the illumination due to diffraction through a hole; the illumination from an opaque shape that just fits the hole.



Babinet's argument is short. In the absence of any obstruction, there is zero illumination on the screen except for a single point in the centre. Hence, for any other point on the screen, all the waves that pass through the hole (producing a total E_1)





+ all the rays that pass around the obstruction (producing a total E_2) must add up to zero. In symbols:

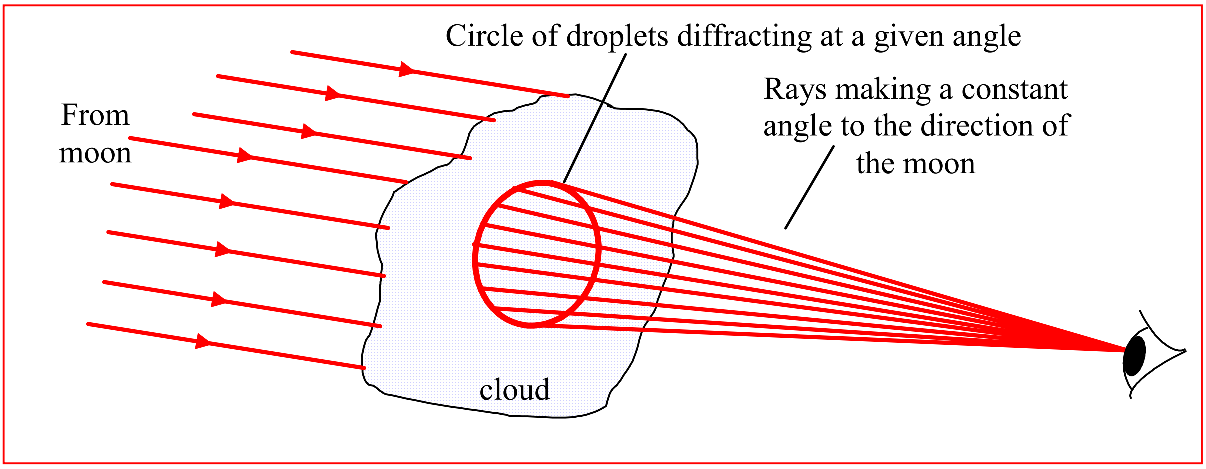
$$\boxed{E_1 + E_2 = 0}$$

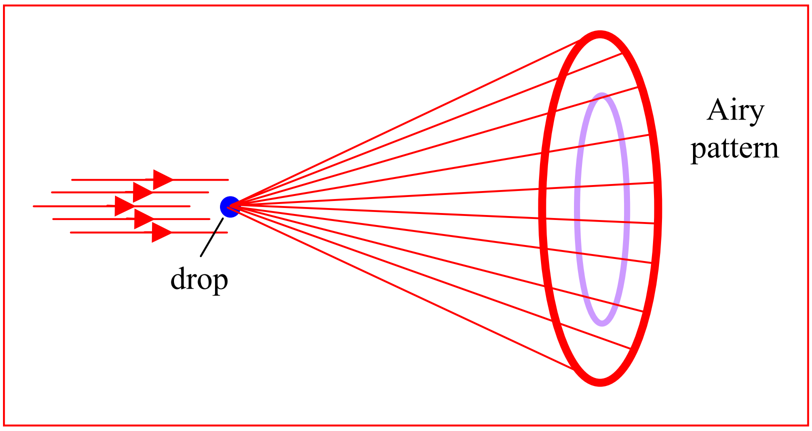
Hence $E_2 = -E_1$.

Therefore, the only difference between E_1 and E_2 is that they are 180° out-of-phase. Since the phase of a wave makes no difference to its irradiance, the diffraction pattern of an obstruction that just fits a hole is the same as that of the hole. Elegant, really.

Babinet's principle is why you can treat cloud droplets the same as holes in a screen, as far as their diffraction pattern is concerned.

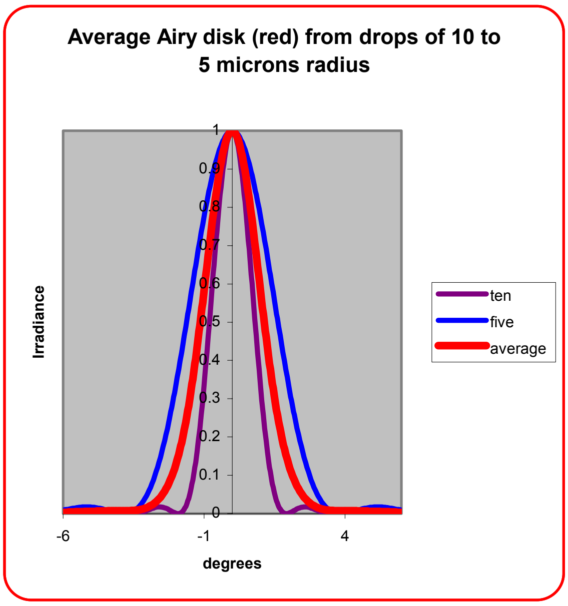
Why the corona is the same shape as the diffraction pattern





That was most of the optics: now for the geometry. When you look up at a cloud a little away from the direction of the moon, all droplets lying at a constant angle from the moon (e.g. 2°) lie on a circle. The Airy diffraction pattern is also circular. Therefore, all the drops contributing illumination at a given angle (say 2°) are doing so from the very same angular part of the diffraction pattern. The illumination at that angle therefore just represents the illumination of the Airy pattern.

The final real-life complication is that cloud drops vary in size somewhat. The corresponding Airy patterns are different angular sizes. If the cloud drops are particularly uniform, you see the corona very well, sometimes even two or three rings of the Airy pattern. If there is a small range of cloud droplet sizes, then you tend to see the first ring with some muted colour and the outer ones are smeared out. If there is a big range of droplet sizes, then you see nothing but a central bright patch of light, called the aureole.



Mother-of-pearl clouds, sometimes called iridescent clouds, show diffraction colours over a much bigger area of the sky. Aberdeen is a good place to look out for them. They arise from very high clouds, often stratospheric, made with particularly small droplets, typically less tha 1 micron in radius.

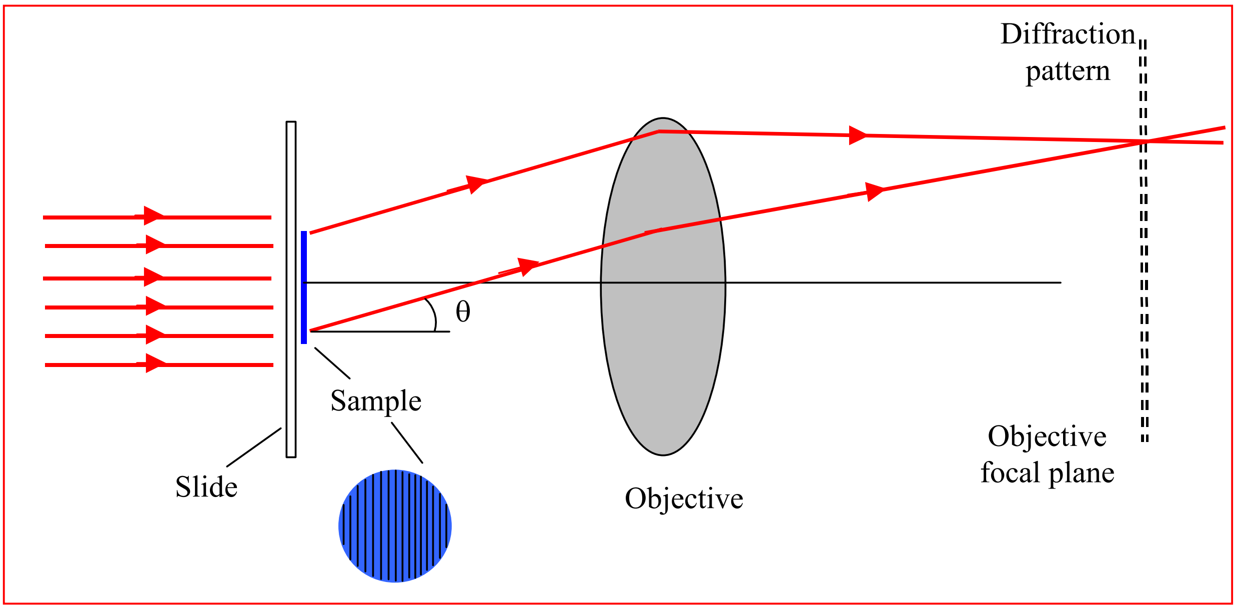
The appearance of diffraction rings isn't just limited to those produced by cloud droplets. Thomas Young (of "Young's slits" and musch more) invented a device for measuring blood corpuscle size called an *eriometer* that really consisted of nothing more than smearing a blood sample over a microscope slide. When you look through the sample at a monochromatic source, a source with one dominant spectral line, you see the Airy diffraction rings characteristic of the blood cells. The angular size of the ring is larger the smaller the size of the cells; the ring is very well defined when the cells are all similar in size and pretty circular in cross-section. The same idea can be used to look at the size of fungal spores and other seeds. The 'classic' example is the ring pattern produced by lycopodium powder, where the spores are particularly uniform in size. On a different scale of wavelength, a very similar effect is produced by the diffraction pattern observed from a gas of molecules at X-ray wavelengths.

Getting away from electromagnetic radiation, G. P. Thomson observed and measured the characteristic ring pattern diffraction given by low energy electrons diffracted by thin metal foils. This was the world's first experimental evidence that De Broglie's ideas on particles possessing a wavelength was indeed true. I have to mention it here, because G. P. Thomson did his work in 1927/28 while he was Professor of Natural Philosophy at Aberdeen University, and he won the Nobel Prize in Physics for it 10 years later.

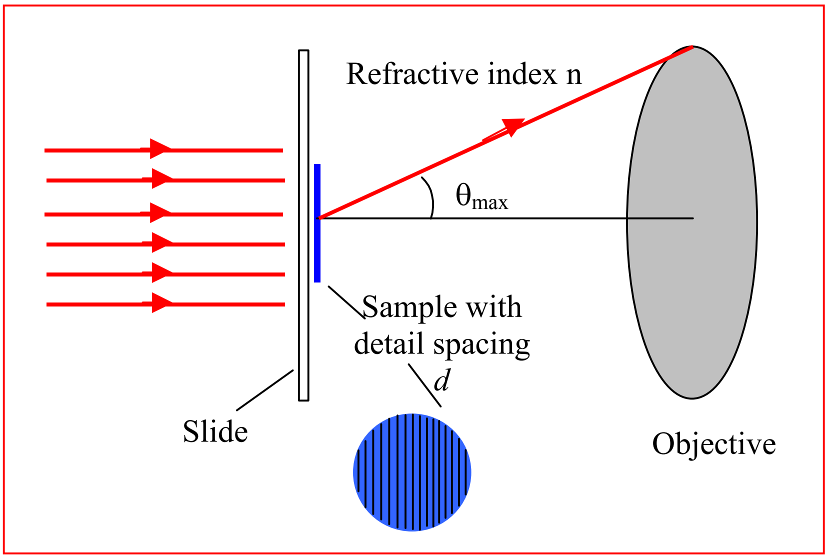
Diffraction and the microscope

If cloud droplets illuminated by the moon, or the sun, produce diffraction patterns, surely everything does? Yes, that's true. Diffraction patterns are mainly on a small angular scale and the microscope is conspicuously where consideration of diffraction patterns becomes important.

A microscope slide with a thin sample on it is, in effect, nothing more than a complicated aperture with very fine detail that can diffract light over large angles. This light must go into the microscope objective, otherwise it won't be seen. The Fraunhofer diffraction pattern



appears in the rear focal plane of the objective. The finer the detail on the slide, the bigger the diffraction angle and the further from the central position is the corresponding diffraction illumination.



These ideas have a direct implication for the detail that can be seen in a microscope. Very fine detail diffracts at very large angles and the light simply misses the microscope objective. You can imagine regular detail to be much like a fine diffraction grating with spacing *d*. The 'first-order' of the diffraction pattern goes sideways at an angle θ given by $d \sin\theta = \lambda$. If $\lambda = 550$ nm, then if the spacing is, say, 1 μm the angle θ is 33°. If the objective isn't wide enough to catch the light, then this detail isn't seen. Hence the width of the objective determines the finest detail a microscope can see. To be more exact, it is $\sin\theta_{max}$ that determines the detail. Microscopists call $n \times \sin\theta_{max}$ the **numerical aperture** of the microscope, where n is the refractive index of the medium that is between the slide and the objective. The refractive index is a minor complication that needn't concern us here. In air, $n \approx 1$, as you know.

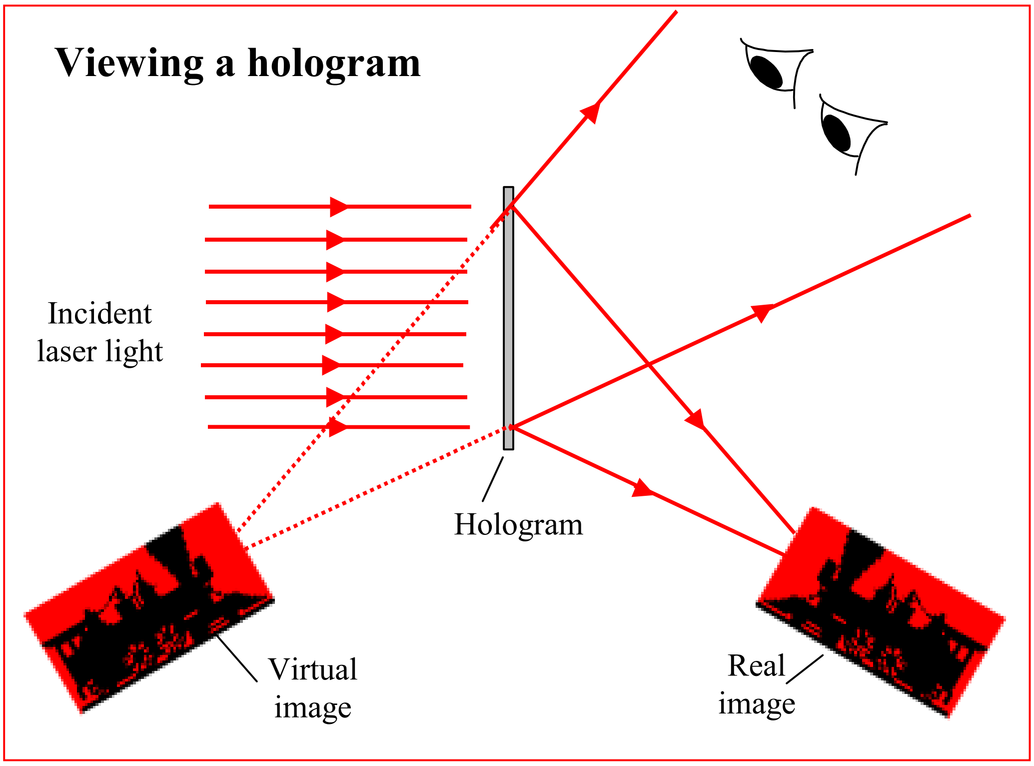
In conclusion:

1. Because of diffraction, the numerical aperture of a microscope objective limits the fineness of detail that a microscope can see.
2. It is a consequence of diffraction being linked to the wavelength of light that the wavelength limits the detail an optical microscope can see, even with the most favourable objective. A microscope cannot see detail closer together than the wavelength of the viewing radiation. (When $d = \lambda$, $\sin\theta = 1$).
3. By blocking out some of the light in the rear focal plane of the objective, you can selectively remove detail of a particular spacing and hence *optically filter* the image.

Looking at microscope imaging in this way was first done by the famous Ernst Abbe, who was employed at the Leitz optical works. His insights put German microscope manufacture at the forefront of the world market, when English makes had previously been there. Nowadays, the top manufacturers are, perhaps, Olympus, Nikon, Leitz, Zeiss and Leica. German makers are still there but British manufacturers are nowhere to be seen. A familiar story?

You could argue that the Fraunhofer diffraction pattern in the rear focal plane of the objective is what the eyepiece processes. The eyepiece, too, forms a diffraction pattern and a relaxed eye looking through it should be focused at infinity. Hence you could say that what is seen in the eyepiece is the diffraction pattern of the diffraction pattern of an object, namely the original object within a magnification factor. This result arises from the powerful mathematical result that the Fourier transform of a Fourier transform of a function is the original function itself. Early in this chapter I hinted that Fraunhofer diffraction patterns were found by a Fourier transform process. All this might seem a very roundabout way of looking at the imaging process but it helps to explain why real microscopes don't form perfect images of what they are looking at. It also helps understand our next topic: how a hologram gives us an image.

Viewing a hologram



You may remember how a hologram is made. It is a photograph of the interference pattern formed between a reference laser beam and the light given off by an object illuminated by the same laser beam. What is being given off by the object is essentially the complex diffraction wave produced by the object under laser illumination.

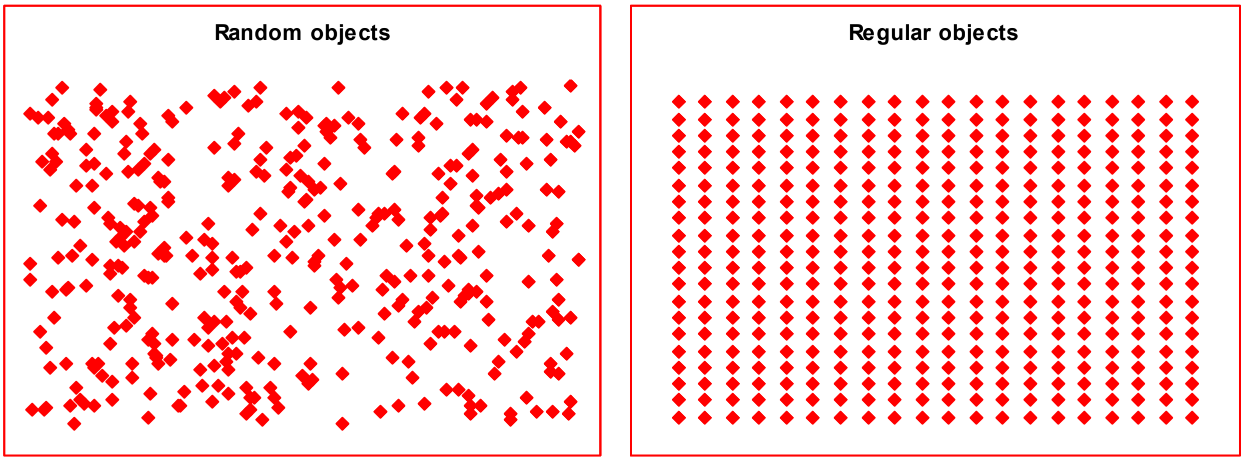
I know of no simple story that 'explains' how the image of the original object is re-created from the hologram. It certainly isn't obvious and has always seemed to me a complex business, so let's acknowledge the perception of Dennis Gabor, who invented the idea in 1947. He certainly deserved the Nobel Prize for Physics that he won in 1971. Seeing a hologram is a pure diffraction process. In the basic arrangement sketched above, the hologram is illuminated by a laser beam. The hologram diffracts, just like a diffraction grating, sending out the laser light into two diffraction orders on either side. What isn't at all obvious, but is true, is that one of these orders is a virtual image of the object and the other forms a real image. You might just have suspected that the diffraction patterns would represent the original object because the hologram is a record of the diffraction from the object. When the laser light shines through the hologram, what we see is a diffraction pattern of a diffraction pattern, i.e. something that represents the original object. Remember the discussion at the end of the previous section. The pattern we look at is the virtual image, which occupies the same spot behind the plate as the original object. The real image is 'thrown away'.

The three dimensional aspect of the image is not so hard to understand. Each eye intercepts the reconstructed wave that is the same, or similar, to the wave that would have come from the object if the object had been there instead. Each eye gets a slightly different view, as usual, and the 3D image is made in our brain by the normal visual process (whatever that is!). The three dimensionality of it is no different really from the three dimensionality given by a binocular microscope that has two objectives and two eyepieces. The difference in the two cases is that when we look at a hologram we can move our eyes about and see how the image changes as we see around corners of the object to previously hidden detail. With the binocular microscope, the objectives and the two tubes are fixed. Of course in the microscope you can always move the slide.

The kind of hologram we've talked about is a **transmission hologram**. The ones we usually see are white-light reflection holograms. I'll leave you to read how these work. It's not part of the course.

Holograms are used for much more than taking pictures of model engines, or works of art, or putting security markers on credit cards or banknotes. Holographic techniques can be used to create optical components, from diffraction gratings that channel the energy into one order of spectra only to components that act like lenses. They can be used for the volume storage of data and by making holograms with fringe patterns that are calculated rather than actually observed, 3D images of items that don't exist in reality can be generated. Perhaps the 'holodeck' is not as far off as you think.

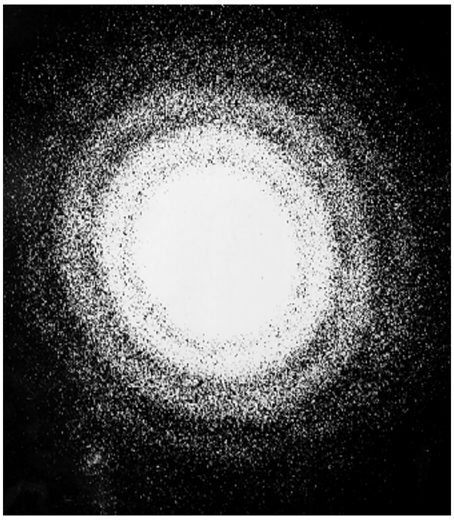
Fraunhofer Diffraction from Multiple Apertures



Multiple apertures are a common occurrence. We have already met *randomly distributed* multiple 'apertures' in the diffraction from cloud droplets. We have also met diffraction from a *regular pattern* of apertures in the diffraction grating, where identical diffracting slits are repeated at a regular interval along a line. In this final section I want to draw a few general lessons about these two circumstances. The results are not only of great importance in optics but they underpin the whole analysis of molecular structures and crystal structures, a topic that will be the subject of another course. There the arguments will be applied at the wavelengths of X-rays. There's little fundamental difference between X-rays and light apart from wavelength, which for X-rays is about 1000 – 10,000 times smaller than light.

To make sense of it all, you need to know just one result. The diffraction pattern from multiple identical apertures is the **product** of the **diffraction pattern of a single aperture** and the **interference pattern of a set of points** situated at the positions of each aperture in the pattern. That's it. How does it work?

Random distribution of identical objects



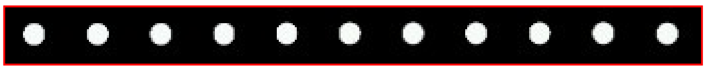
A random series of diffracting points in a plane has a 'spotty' interference pattern, with a central peak and intensity in the diffraction plane that shows random fluctuations on a general background. As the number, N, of points distributed at random gets large, the intensity anywhere becomes N times the intensity contributed by one aperture and the random fluctuations become closer together and hardly noticeable.

The picture alongside shows the diffraction pattern of a random array of notionally circular dots recorded by a student in PX2505. The number of dots was not particularly large, a few hundred at most, and you can still see the spottiness of the pattern. When you look at the diffraction pattern caused by moonlight shining through cloud droplets, or X-rays diffracted by a gas, then the number of 'objects' (droplets or gas molecules) contributing to the pattern is enormous and the spottiness is on too small a scale to be seen.

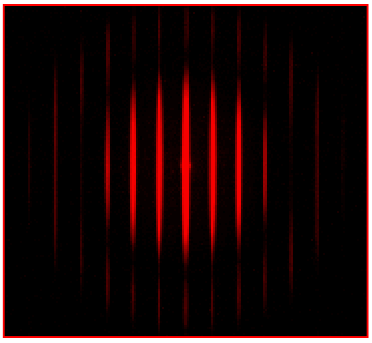
Notice a few other points.

1. The picture shown isn't truly circular, from which you can deduce that the dots were probably distorted in the reproduction process. They are **shorter in diameter in the direction that the diffraction ellipse is long**, and vice-versa.
2. If there were 200 apertures then the diffraction pattern would be 200 times as intense as from a single aperture. If you want to find out what the diffraction produced by a particular shape is, and it can be a complicated shape, don't just put a single copy of it into the laser beam but put a random distribution of many repeated copies.
3. You can't tell if the objects that made the diffraction pattern were a random array of dots on a clear background or a random array of holes in an opaque screen, thanks to Babinet's principle.

Regularly repeated identical objects

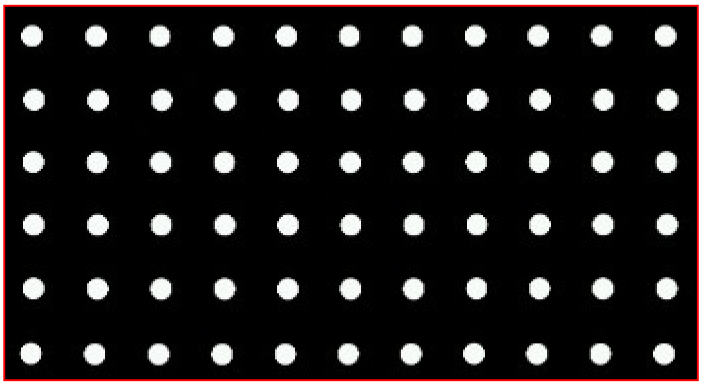


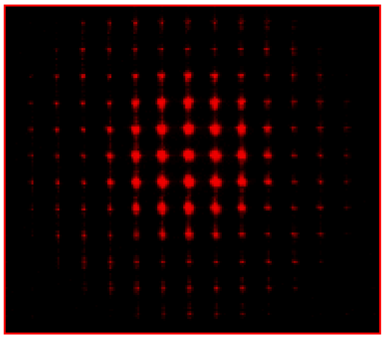
The interference pattern from a regular repetition of sites consists of narrow localised peaks. Broadly speaking, if there are N repetitions along a line, the peaks of intensity increase in height as N^2 and narrow in proportion to N. The diffraction is therefore strongly concentrated at the peaks of the interference pattern and when N becomes large, the diffraction really just shows up at



the location of these peaks. The height of any peak is determined by the strength of the diffraction pattern of a single aperture in that direction.

The accompanying illustrations show the effect for regularly repeated circular apertures. The first diffraction pattern is formed by a regular repetition of 50 sources in a line (only some are shown). In this case the Airy disc is sampled by what is effectively a set of coarse spectral lines.





The second pattern is formed by a rectangular grid of the same apertures. You can just see that the underlying diffraction is an Airy disk, now sampled in two dimensions at the diffraction points given by the interference pattern produced by a two dimensional set of points. We didn't discuss this interference pattern in the previous section but the result looks plausible. What you can't see is that the peaks are strong in proportion to the square of the number of points in the grid. For example, if the grid were 50 repetitions in each dimension, the peaks would be 2500 times brighter than the irradiance from a single aperture in each dimension. However, the irradiance would fall very quickly to zero and what you gain at the peak, you lose elsewhere. For example, it is not quite so easy to see at a glance if the pattern shown above is slightly elliptical, because it is only being sampled at a relatively few discrete points.

When you shine X-rays at a 3D crystal structure, the very same physics takes place. What counts in terms of the sharpness of diffraction peaks is the number of planes of atoms contributing to the peak You may have 10^4 planes, giving an enhancement by a factor of 10^8. The extremely weak X-ray diffraction from each molecule is therefore multiplied up enormously, but your loss is that you can only sample it at this high level of intensity around the interference maxima, which are the "Bragg peaks" in X-ray jargon. However, the gain at the Bragg peaks is worth it and makes the whole of X-ray structural analysis possible. If you want to find the structure of a very large protein such as haemoglobin, or even a virus DNA of mega daltons molecular weight, then it can be done if you form a crystal structure of the protein. Any crystal structure will do, because you're not interested in the actual arrangement in the crystal of the molecules themselves. It may be pretty simple. You're only interested in the diffraction from the molecule itself. The crystalline arrangement allows this to be sampled at thousands of relatively intense Bragg peaks and hence the structure of the original molecule gradually deduced.

I hope this section has illustrated that diffraction is far from simply a quirky phenomenon that produces fringes along the edges of the shadows of very small objects. Diffraction plays a central part in the propagation of light. The understanding of diffraction has helped explain natural phenomena and enabled sophisticated use to be made of light, in ways not imagined a century ago.

End of diffraction

JSR