8. Diffraction

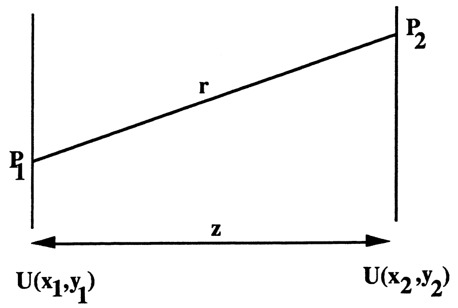
Here we are concerned with propagating an optical field from one plane to another. The question is "If I know the field in plane 1 can I find the field in plane 2?".

In order to see that this might not be trivial consider the simple experiment where we simply shine light through a circular aperture and look at the intensity on a screen some distance away. If the hole is large a geometrical image is formed on the screen. As the hole becomes smaller fringing appears and the pattern becomes smaller until as the hole becomes even smaller when the pattern starts to become larger! This is illustrated on the next page and indicates that diffraction is something we need to worry about if we put obstacles in the path of optical beams. We also need to understand diffraction if we want to analyse the focussing ability of lenses and the operation of simple imaging and image processing systems.

We shall take a very simple approach to describing diffraction phenomena via ***Huygens principle***. Other more mathematical theories are available – all involve approximations of one sort of another and essentially reduce to the simple Huygens formulation in many cases of practical interest. Huygen (1690) suggested (effectively) that

> *"Each point on a primary wavefront can be considered as the source of a secondary wave and a secondary wavefront can be constructed as the envelope of these secondary spherical wavefronts."*

Formally our problem is to find the field $U(x, y, z)$ if we know the field in some other plane, say $U(x, y, 0)$.



The field at a particular point P_1 *is* $U(x_1, y_1)$. It gives rise to a spherical wave with this field as strength. If we look for the field at P_2 a distance *r* from P_1 we can easily write it as

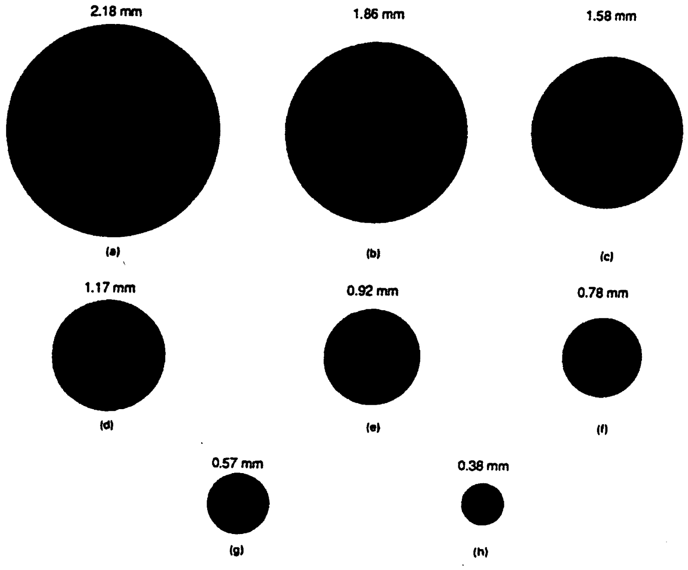
$$U(x_1, y_1)\frac{exp - jkr}{r}$$

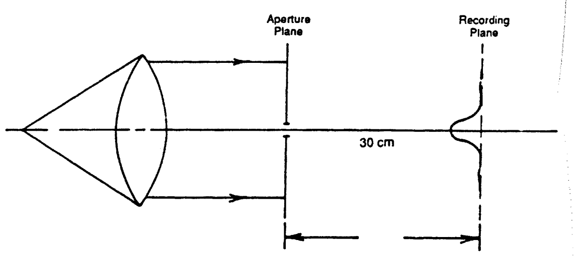
The field at P_2 due to the entire field in the (x_1, y_1) plane is now simply

$$U(x_2, y_2) = \iint U(x_1, y_1)\frac{exp - jkr}{r}\, dx_1\, dy_1$$

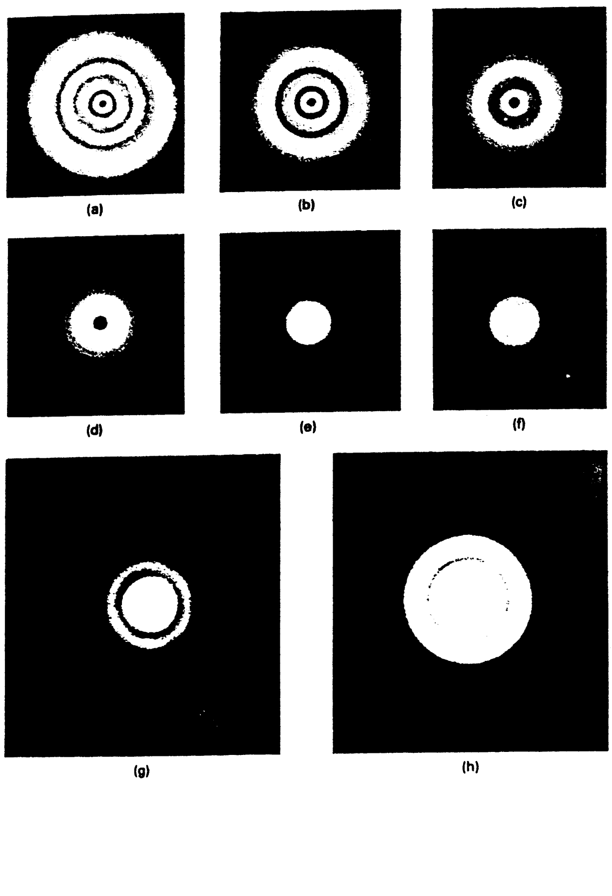
and

ZJ Laczik
T Wilson





Experimental arrangement for producing diffraction patterns of circular aper- tures placed in the aperture plane along with the circular apertures.



$$r = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + z^2}$$

This is the key integral which we will use to solve various diffraction problems. However, before we do so it will be useful to discuss an alternative approach to propagating the field from one plane to another. This is based on the idea of representing the input field as an ***angular spectrum of plane waves***, propagating each component separately, and then adding the propagated components together to find the total field. We regard the input field as $U(x, y, 0)$ and note that it can be written as

$$U(x, y, 0) = \int_{-\infty}^{\infty} \int A(m, n)\ exp - 2\pi j\,(mx + ny)\ dm\ dn$$

where $A(m, n)$ is given, formally, by

$$A(m, n) = \iint U(x, y, 0)\ exp\ 2\pi j\,(mx + ny)\ dx\ dy$$

These, of course, are Fourier Transform pairs.

Recall that a plane wave travelling in the $\boldsymbol{k}$ direction may be written

$$B(x, y, z) = \exp - j\,\boldsymbol{k}\cdot\boldsymbol{r} = exp - j\left(k_x x + k_y y + k_z z\right)$$

where

$$k = \frac{2\pi}{\lambda} = \sqrt{k_x^2 + k_y^2 + k_z^2}$$

Thus across the plane $z = 0$, a complex exponential function $exp - 2\pi j(mx + ny)$ may be regarded as a plane wave propagating with

$$k_x = 2\pi m\,;\, k_y = 2\pi n\,;\, k_z = \frac{1}{\lambda}\sqrt{1 - (\lambda m)^2 - (\lambda n)^2}$$

The complex amplitude of that plane wave component is simply

$$A(m, n) = A\left(\frac{kx}{2\pi}, \frac{ky}{2\pi}\right)$$

and

$$A\left(\frac{kx}{2\pi}, \frac{ky}{2\pi}\right) = \iint U(x, y, 0)\ exp\ j\left(k_x x + k_y y\right)\ d_x\, d_y$$

is called the ***angular spectrum.***

The angular spectrum now propagates as

$$A(m, n, z) = A(m, n)\, exp - \frac{j}{\lambda}\sqrt{1-(\lambda m)^2-(\lambda n)^2}\, z$$

and hence the field at (x, y, z) is given by

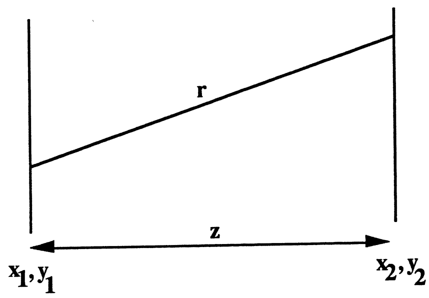
$$U(x, y, z) = \iint A(m, n, z)\, exp\, 2\pi j\,(mx + my)\, dm\, dn$$

or

$$U(x, y, z) = \iint A(m, n, 0)\, exp - \frac{j}{\lambda}\sqrt{1-(\lambda m)^2-(\lambda n)^2}\, z\, exp\, 2\pi j\,(mx + ny)\, dm\, dn$$

This final result is *sometimes* more convenient to use than the Huygens integral. It is pure mathematical and physical convenience. This expression can be shown to be equivalent to our Huygen integral. It's not too difficult but does involve an unfamiliar integral. It's done in Henry Stark – Application of Optical Fourier Transforms.

Fresnel and Fraunhofer Diffraction



$$U(x_2, y_2) = \frac{j}{\lambda}\iint U(x_1, y_1)\,\frac{exp - jkr}{r}\, dx_1\, dy_1$$
$$r = \sqrt{(x_2 - x_2)^2 + (y_1 - y_2)^2 + z^2}$$

The pre-multiplying factor j/λ has been introduced. This is the factor that appears if we derive the diffraction integral more rigorously and we include it here to make comparison of our formulae with those found in books more straightforward. We can now make certain approximations to make this integral easy to evaluate in specific regions of space.

(i) Assume z is sufficiently large that it is reasonable to set $r = z$ in the denominator.

(ii) Assume $|x_1 - x_2|$ *and* $|y_1 - y_2|$ are sufficiently small to permit us to write

$$r = z + \frac{(x_1 - x_2)^2}{2z} + \frac{(y_1 - y_2)^2}{2z}$$

i.e. the paraxial approximation, which makes the diffraction integral look like

$$U\left(X_2,\ y_2,\ z\right)=\frac{exp\ -\ jkz}{\lambda z}\cdot\iint U\left(X_1,\ y_1\right)\ exp\ -\ \frac{jk}{2z}\left\{\left(x_1-x_2\right)^2+\left(y_1-y_2\right)^2\right\}dx_1\,dy_1$$

when these assumptions hold – called the ***Fresnel*** approximation – we speak of ***Fresnel diffraction***, characterised by the above equation.

We can rewrite this equation as

$$U\left(x_{2,}\,y_{2,}\,z\right)\quad\sim\quad j\frac{exp-jkz}{\lambda z}\cdot exp-\frac{jk}{2z}\left(x_2^2+y_2^2\right)\cdot$$
$$\iint U\left(x_1,y_1\right)exp+\frac{jk}{z}\left(x_1x_2+y_1y_2\right)exp-\frac{jk}{2z}\left(x_1^2+y_1^2\right)dx_1\,dy_1$$

If our observation plane z is sufficiently far away such that

$$z>>\frac{k}{2}\left(x_1^2+y_1^2\right)_{maximum\ value}$$

Then

$$\frac{k}{2z}\left(x_1^2+y_1^2\right)_{max}<<1$$

and

$$exp\ -\ j\ \frac{k}{2z}\left(x_1^2+y_1^2\right)_{max}\sim 1$$

Thus the integral considerably simplifies to

$$U\left(x_2,\ y_2,\ z\right)=j\ \frac{exp\ -\ jkz}{\lambda z}\ exp\ -\ \frac{jk}{2z}\left(x_2^2+y_2^2\right)\cdot\iint U\left(x_1,\ y_1\right)\ exp\ \frac{jk}{z}\left(x_1x_2+y_1y_2\right)\ dx_1\,dy_1$$

In this case we speak of ***Fraunhofer diffraction***.

If we introduce a Fresnel Number, *F*, as

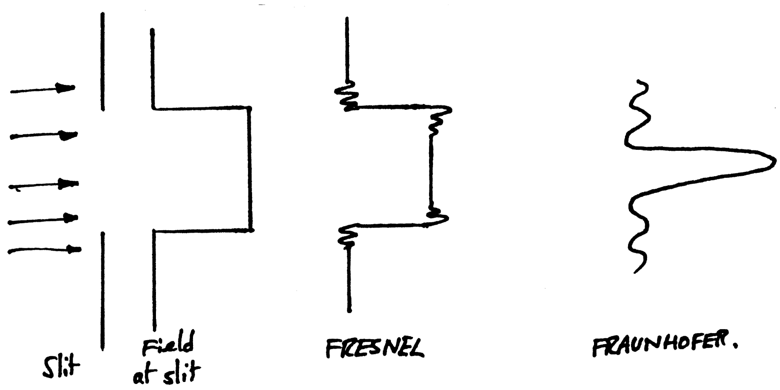
$$F=\frac{\left(x_1^2+y_1^2\right)_{max}}{\lambda z}$$

We see that the approximation is valid as long as

$$F<<1/\pi$$

We also note that in the Fraunhofer case the "far field" is ***proportional to the Fourier transform of*** $U\left(x_1,\ y_1\right)$.

We now consider the specific example of a slit aperture



We will only consider the Fraunhofer case here because the maths is easy – but how far away do we really have to be before it applies. Recall

$$z >> \frac{k}{2}\left(x_1^2 + y_1^2\right)_{max}^2 = \frac{\pi}{\lambda}\left(x_1^2 + y_1^2\right)_{max}$$

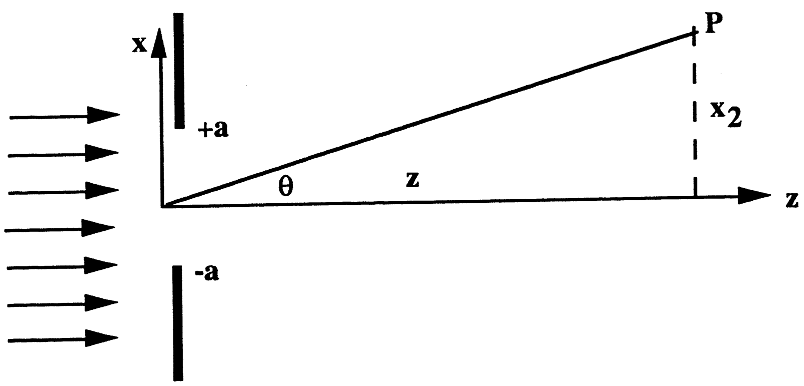
Assume $\lambda = 0.5\mu\text{m}$

(i) $\left(x_1^2 + y_1^2\right)_{max} = 1 \;\; inch$ $\qquad z >> 1.9\, km\,!!$

(ii) $\left(x_1^2 + y_1^2\right)_{max} = 10\,\lambda = 5\mu\text{m}$ $\qquad$ then $z >> 150\,\mu m = 0.15\; mm$

(iii) $\left(x_1^2 + y_1^2\right)_{max} = 1\, mm$ $\qquad z >> 6\text{m}\;(ish)$

Let us now consider the following geometry and calculate the far-field field.



In order to do this we need to find an expression for the "input" field $U(X_1, y_1)$ at the screen. We find this field using the so-called Kirchhoff boundary condition. This states that the field within the aperture is not affected by the aperture and that the field outside the aperture, $|x| > a$ is identically zero. This is clearly only an approximation, but does give results which generally agree with what is observed. Thus

$$
\begin{aligned}
U(X_1) &= 1 \qquad |x| < a \\
&= 0 \qquad otherwise
\end{aligned}
$$

Thus

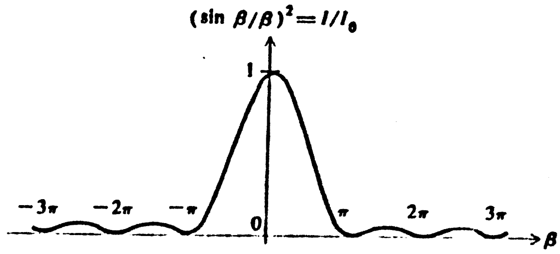
$$U(x_2) \sim \int_{-a}^{a} exp\ j\ \frac{kx_1\ x_2}{z}\ dx_1$$

which leads to

$$I(x_2) \sim |U(x_2)|^2 \sim \left[\frac{sin\left(\frac{kax_2}{z}\right)}{\frac{kax_2}{z}} \right]^2$$

or

$$I(x_2) = sinc^2\left(\frac{kax_2}{z}\right)$$



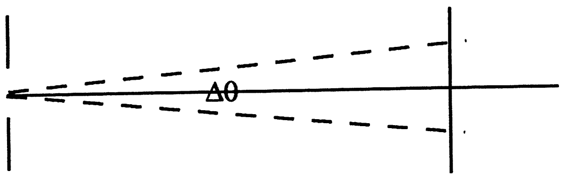
Fraunhofer diffraction pattern of a single slit.

If we introduce an angle θ

$$\theta = \frac{x_2}{z}$$

Then $I(\theta) = sinc^2(ka\theta)$

Thus we can think of diffraction as causing the beam to spread



If we denote the beam spreading $\Delta\theta$ as shown – to the first zero of the diffraction pattern we can easily see that

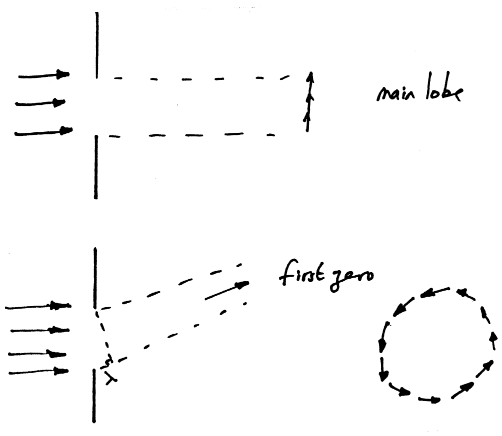
$$ka\left(\frac{\Delta\theta}{2}\right) = \pi$$

$$\therefore\ \Delta\theta = \frac{2\pi}{ka} = \frac{\lambda}{a}$$

i.e. beam spreading will increase as the aperture size (measured in wavelengths) decreases.

This is important and is exactly what was observed earlier.

The fact that the diffraction pattern has zeros has a very simple explanation based on interference effects. The total field at a particular point on the screen will result from interference between all the imagined point sources in the aperture. In the direction of the zeros the phase difference across the aperture is $2\pi\ (or\ 2m\pi)$. If we split the aperture up into a small number of elements and calculate their relative phase differences we can add the final fields in a vector diagram. In the case of destructive interference (zeros) we find a zero resultant



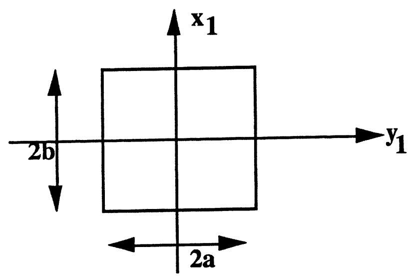
This physical interpretation is discussed more fully by Hutley in his book on Diffraction Gratings. He also makes the important point:

> *"However, it is most important to realize that the reason the irradiance in the shadow is often very small is not that the waves fail to propagate there but that when they do they interfere in such a way that the contributions from most parts of the aperture cancel each other out. If the phase or the amplitude (or both) of the contributions from different parts of the aperture are modified it is possible to reduce the extent to which destructive interference takes place and the diffraction*

pattern may be modified in such a way that there is a substantial resultant amplitude in what would otherwise have been the shadow."

The Square Aperture

Having considered the slit case it is now very easy to treat the square aperture.



This gives

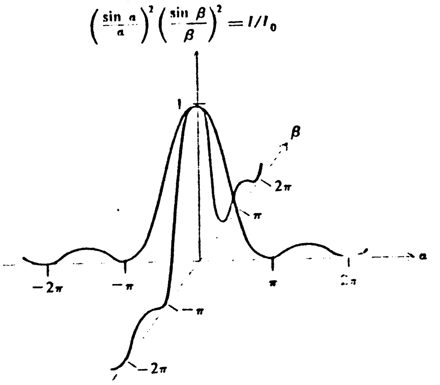
$$I(x_2, y_2) = I_0\, sinc^2\left(\frac{kax_2}{z}\right) \cdot sinc^2\left(\frac{kby_2}{z}\right)$$

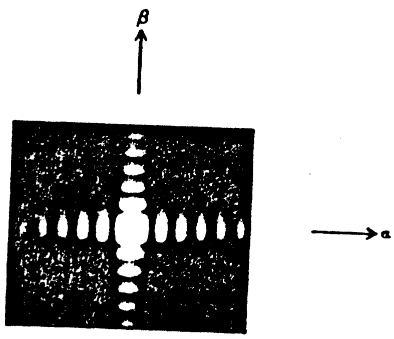
or
$$I_0\, sinc^2(ka\theta)\, sinc^2(ka\varphi)$$

where the angles θ *and* φ are defined as

$$\theta = \frac{x_2}{z}; \varphi = \frac{y_2}{z}$$

As before the scale of the diffraction pattern bears an inverse relationship to the scale of the aperture.





The Circular Aperture

We can perform exactly the same calculation for the circular aperture. As it will prove useful later we will derive the expression now.

$$U(x_2, y_2) \sim \iint U(x_1, y_1)\, exp\; j\frac{k}{z}(x_1\, x_2 + y_1\, y_2)\, dx_1\, dy_1$$

As $U(x_1, y_1)$ will be radially symmetric – so will $U(x_2, y_2)$ and so it is best to work in polar co-ordinates. Thus let

$$x_1 = r\cos\theta \quad x_2 = \rho\cos\varphi$$
$$y_2 = r\sin\theta \quad y_2 = \rho\sin\varphi$$

which leads to

$$U(\rho) = \iint U(r)\cdot exp\ j\frac{k\rho r}{z}\cos(\theta - \varphi)\, rdrd\theta$$

We now make use of the integral

$$\frac{1}{2\pi}\int_0^{2\pi} exp\ \ jz\ cos(\theta = \ \phi)\ d\theta = J_0(z)$$

where $J_0(z)$ is a *Bessel function* of the first kind of zero order. This enables us to write

$$U(\rho) \sim \int U(r)\, J_0\left(\frac{k\rho r}{z}\right) r\ dr$$

This is the form of the Fourier Transform for circularly symmetric functions and is called the ***Fourier Bessel transform*** or ***Hankel transform***.

If our circular aperture has a radius *a* the field becomes

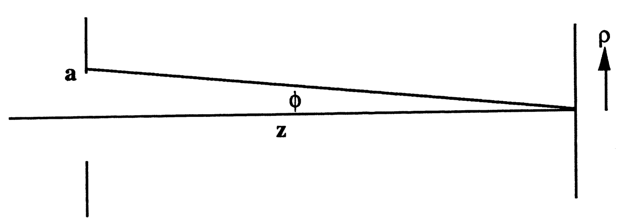
$$U(\rho) \sim \int_0^a 1.\ J_0\left(\frac{k\rho r}{z}\right) r\ dr$$

We now need to know the identity

$$\int_0^x \xi J_0(\xi)\, d\xi = x\ J_1(x)$$

where J_1 is a Bessel function of the first kind of the first order. Thus we can finally write

$$I(\rho) \sim |U(\rho)|^2 = \left[2\frac{J_1(k\rho\ a/z)}{k\rho\ a/z}\right]^2$$



Now we introduce an angle $\sin \varphi = a/z$ such that

$$I(\rho) = \left[2 \frac{J_1(k\rho \sin\varphi)}{k\rho \sin\varphi} \right]^2$$

where a factor of two has been introduced to normalise $I(0)$ to unity. If we finally introduce a new co-ordinate, v, related to ρ by

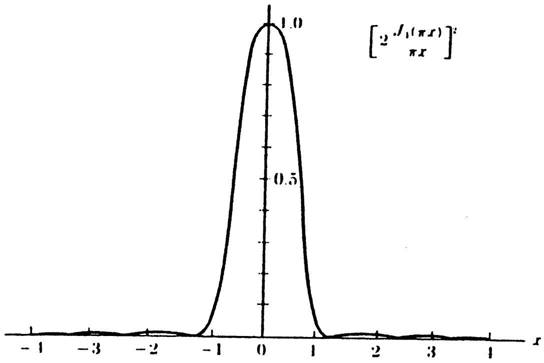
$$v = k\rho \sin \varphi$$

permits us to write the intensity as

$$I(v) = \left[\frac{2 \; J_1(v)}{v} \right]^2$$

which, we will see later, is a very important function. It turns out to be related to the intensity distribution of the focus of a lens. This is why we have bothered to work it out! The pattern is called an ***Airy*** pattern or ***Airy disk***. It is plotted on the next page together with a table of useful values.

x	$\left[2 \frac{J_1(\pi x)}{\pi x} \right]^2$	max or min
0	1	max
1.220	0	min
1.635	0.0175	max
2.233	0	min
2.679	0.0042	max
3.238	0	min
3.699	0.0016	max



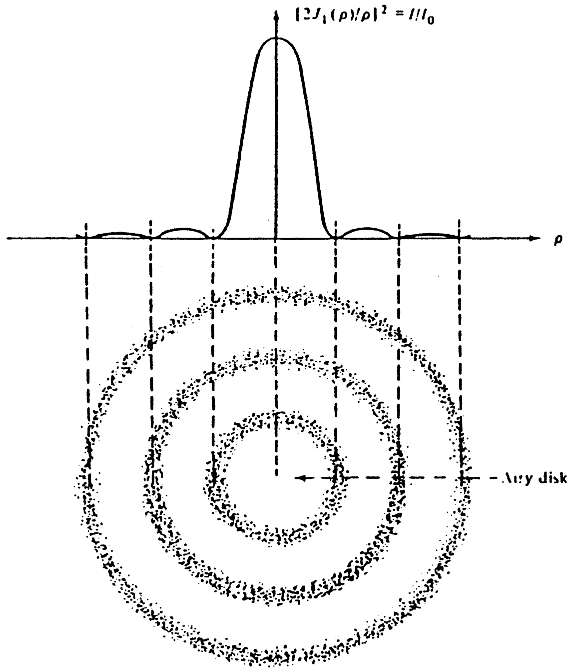
We note that the first zero is given by $v_0 = 1.22\pi$ which corresponds to

$$\rho_0 = \frac{1.22\,\lambda}{2 \sin \varphi}$$

This has great implications, as we shall see later, in the resolving power of optical instruments.

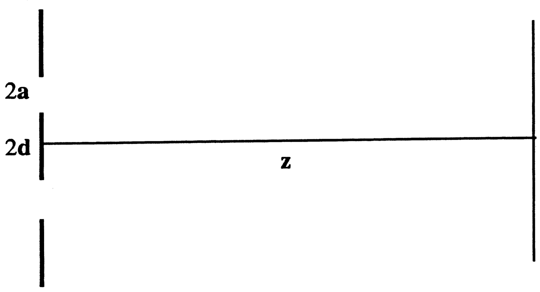
The Airy pattern is shown on the next page. Note, of course, that it is circularly symmetric and compare it with the pattern found for the square aperture.

Caution: Maths is easier in *x-y* co-ordinates and a square is a bit like a circle but the diffraction patterns turn out to be very different. However if we compare $sinc^2\, x$ with $\left(J_1(x)/x\right)^2$ then along one of the axes the difference in the patterns is not so great. So we need to be careful!



Double Slit

We now consider diffraction by two slits, each of width, $2a$, and separation $2d$.

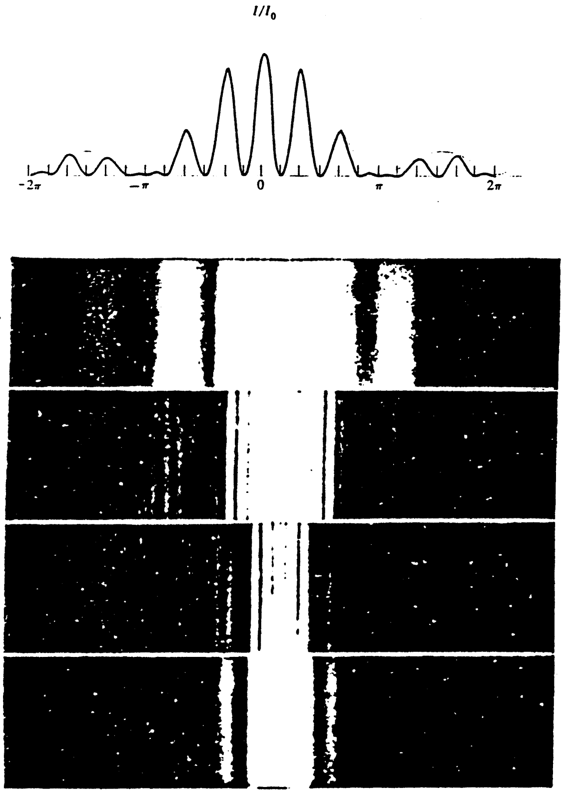


$$U(x_2,\ z) \sim \int_{d-a}^{d+a} exp\ j\frac{kx_1\ x_2}{z}\ dx_1 + \int_{-d-a}^{-d+a} exp\ j\frac{kx_1\ x_2}{z}\ dx_1$$

Which can easily be integrated to give the intensity as

$$I \sim cos^2\left(\frac{kdx_2}{z}\right) \cdot sinc^2\left(\frac{kax_2}{z}\right)$$

i.e. the result as for diffraction from a single slit, $sinc^2(kax_2/z)$, but multiplied by a factor $cos^2(kdx_2/z)$ which is determined by the spacing of the slits. This gives diffraction patterns of the form



The fact that the slits have finite width corresponds to the *envelope* function

$$sinc^2\left(\frac{kax_2}{z}\right)$$

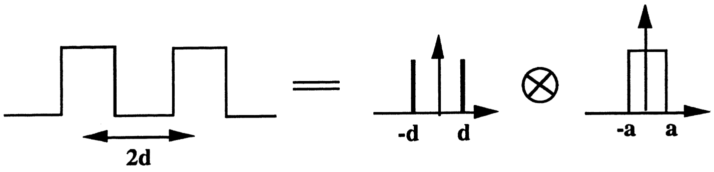
which, essentially, determines how the intensity is distributed among the various diffraction orders. We note that in the case of $a = 0$, $sinc^2\left(kax_2/z\right) = 1$, and so the intensity becomes simply

$$I \sim cos^2\left(\frac{kdx_2}{a}\right)$$

which, naturally, is precisely the result we obtained earlier for the idealised Young's slits experiment.

We can easily extend the analysis to the case where we have *N* equally spaced slits rather than just two. We would do this by just evaluating the integral but it is more useful to do it by using our knowledge of the properties of Fourier transforms.

To illustrate this let us return to our earlier case of two slits. We can easily write the transmission function in terms of the *convolution* between the transmission function of a single slit and two suitably spaced spikes (delta functions). These delta functions are an example of an array function. We will denote the convolution operation by the symbol $\otimes$.



Recall that

$$F\{a \otimes b\} = F\{a\} . F\{b\}$$

i.e. the Fourier transform of the convolution of two functions, $F\{a \otimes b\}$ is equal to the *product* of the Fourier transforms of the individual function.

We now use this result in our two-slit case. If $b(x_1)$ represents the extent of the slit, then we know from before that the Fourier transforms is given by

$$sinc\left(\frac{kax_2}{z}\right)$$

If $a(x_1)$ represents the array factor, i.e. two delta functions, then

$$a(x_1) = \delta(x_1 + d) + \delta(x_1 - d)$$

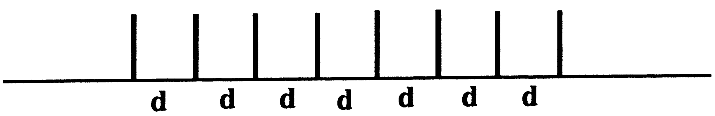
$$F\{a(x_1)\} = \int [\delta(x_1 + d) + \delta(x_1 - d)] exp\ j\frac{kx_1 x_2}{z}\ dx_1$$
$$= exp\ j\frac{kdx_2}{z} + exp - j\frac{kdx_2}{z}$$
$$= 2 cos\left(\frac{kdx_2}{z}\right)$$

Thus

$$F\ (a \otimes b) = F(a) \cdot F(b) = cos\ \frac{kdx_2}{z} \cdot sinc\left(\frac{kax_2}{z}\right)$$

i.e. the same result as before, but now it is clear where the two multiplicative factors come from.

Let us now extend this to the case of N slits - This is, essentially the ***diffraction grating***. The array factor now consists of N spikes (delta functions) each separated by a distance d.



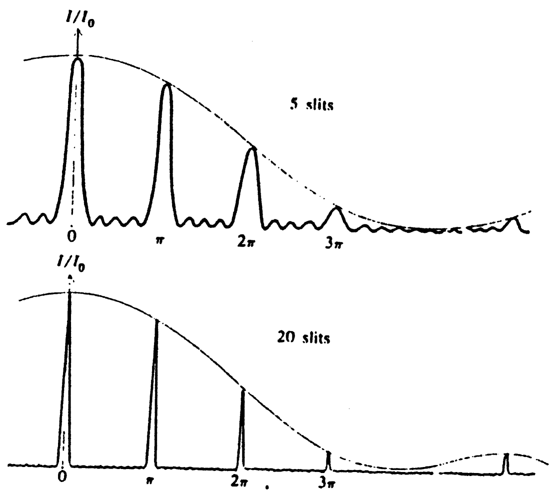
It is straightforward to show that the Fourier transform of this function, obtained by summing a geometric progression may be written as

$$F(a) \sim \frac{\sin N\left(\dfrac{kdx_2}{z}\right)}{N\,\sin\left(\dfrac{kdx_2}{z}\right)}$$

and hence we can write the intensity as

$$I(x_2) = sinc^2\left(\frac{kax_2}{z}\right)\cdot\left[\frac{\sin N\left(\dfrac{kdx_2}{z}\right)}{N\,\sin\left(\dfrac{kdx_2}{z}\right)}\right]^2$$

which, of course, reduces to the double slit expression when $N = 2$. The effect of increasing N on the diffraction patterns is shown on the next page.



The single slit factor appears again as an envelope function. Principle maxima occur within the envelope when $kdx_2/z = n\pi,\ n = 0, 1, 2 \cdots,$ or introducing an angle, $\sin\theta = x_2/z$, the condition becomes

$$2\mathrm{d}\ \sin\theta = n\lambda$$

which is, of course, ***Bragg's Law*** – The grating equation for normal incidence – It relates the wavelength to the angle of diffraction. The integer n is called the order of diffraction.

Resolving Power of a Grating

The angular width of the principle fringe may be determined by considering when the argument $N(kdx_2/z) = Nkd\ \sin\theta$ changes by π.

i.e. $$\Delta(Nkd\ \sin\theta) = \pi$$

or

$$\Delta\theta = \frac{\lambda}{N2d\ \cos\ \theta}$$

For a specific order – fixed n – we can relate $\Delta\theta\ to\ \Delta\lambda$ via Bragg's Law

$$2\text{d}\ \sin\theta = n\lambda$$

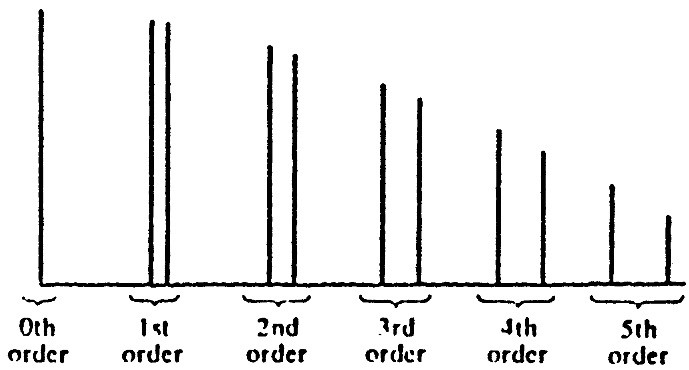
$$\therefore\ \Delta\theta = \frac{n\Delta\lambda}{2\text{d}\ \cos\ \theta}$$

From the two equations for $\Delta\theta$ it is easy to define a ***resolving power*** of the grating as

$$RP = \frac{\lambda}{\Delta\lambda} = nN$$

i.e. the product of the number of grooves, N, multiplied by the order number n.

The figure below shows the pattern for a many line grating illuminated with two wavelengths.



We have only discussed amplitude gratings here where all that the grating does is to block off part of the incident beam. This is often wasteful of intensity and so often the ruled grooves are shaped, usually as a sawtooth profile, to make most of the light appear in one order and hence increase the efficiency of the grating. This is the process of ***blazing*** (see, e.g. Guenther).

Errors

Our Fourier approach is particularly useful in determining what really affects the performance of a grating. As an example let us suppose that the groove profile of a grating was not properly rules, but the spacing between grooves was correct. The only effect this would have would be to modify the envelope function – i.e. the distribution of energy between orders – but it would have *no effect* on the resolution or position of the diffracted maxima. On the other hand if there were errors in the groove position then resolution would be affected. There is a good discussion on these important practical problems in M.C. Hutley – Diffraction Gratings.

The gratings we have discussed have been of the transmission type which only alter the amplitude of the incident wave. Another important type is the *phase* grating which only alters the phase of the incident wave. This may be described by

$$t(x) = exp\ \ j\varphi\ (x)$$

The Fraunhofer diffraction pattern is given by

$$U(x_2) \sim \int exp\ \ j\varphi(x_1)\ exp\ \ j\frac{kx_1\, x_2}{z} \cdot dx_1$$

Suppose $\varphi(x_1) = b\ cos\ \upsilon x_1$

$$U(x_2) \sim \int exp\ \ j\ \ b\ cos\ \upsilon\ x_1\ \ exp\ \ j\ \ \frac{kx_1\, x_2}{z}\ dx_1$$

which is, in general, very difficult to evaluate. If, however, *b* is sufficiently small that we can expand $exp\ x = 1 + x$ then the diffracted field will look like

$$\begin{aligned} &U(x_2) \sim \int (1 + \ jb\ cos\ \upsilon\ x_1)\ exp + \ j\ \ k\ \ sin\theta \cdot x_1\ dx_1 \\ &= \delta(x_1) + j\,\frac{b}{2}\ \delta\ (k\ \ sin\ \theta + b) + j\,\frac{b}{2}\ \delta\ (k\ sin\ \theta - b) \end{aligned}$$

i.e., a d.c. term (straight through beam) and two diffracted orders. The number of diffracted orders increases, of course, as *b* increases. (For more details see Goodman. He works out the integral for all values of *b*.)