

5 Diffraction

Diffraction is a study of the light produced when an infinite plane wavefront is interrupted and light from only a finite region on the wavefront is allowed to propagate further. Using Huygens' principle we can treat each point where the light is not interrupted as the source of a new wavelet and investigate the net result of adding the light from all these sources.

We have already seen the simplest example of this, i.e. the case of two very narrow slits. We can generalise this by either having a larger number of discrete sources, or we can consider the effect of the finite width of the slits. The general example of the pattern formed in the latter case is shown in figure 33. In all cases we will consider the limit where the distance at which we view the diffraction pattern is much greater than either the size of the image or the limited region that the diffracted light originates from. This is the so-called Fraunhofer limit, and as shown in Section 3.3 in this limit we can assume all the interfering light rays are parallel to each other to a good approximation.

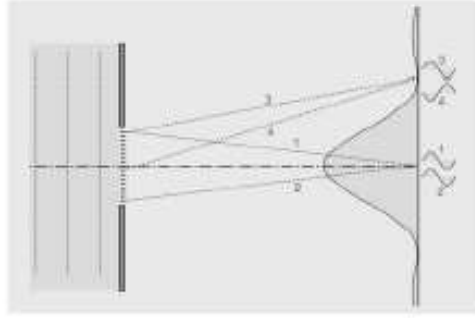


Figure 33: An example of diffraction, i.e. the pattern formed from transmission through a single slit.

5.1 Multiple Slits – Diffraction Grating

We first consider the example of an array of N equally spaced slits, with separation d forming a straight line. This is the generalisation of Young's double slits to cases where $N > 2$. The light rays from each slit, which will ultimately interfere with each other at a distant screen, is shown in figure 34. If we view the light at an angle θ to the normal of the line of slits, the path difference between successive rays is $d \sin \theta$. We can then calculate the total amplitude of light using the same approach as in Section 3.3. If the light received at the screen from the top slit is $E \exp(ikr_1)$, where E is the magnitude of the electric field and $k = 2\pi/\lambda$, then if each slit transmits the same amplitude of light the total amplitude is

$$A_{\text{Tot}} = E \exp(ikr_1) + E \exp(ikr_1 + i\delta) + E \exp(ikr_1 + 2i\delta) + \cdots + E \exp(ikr_1 + i(N-1)\delta), \quad (101)$$

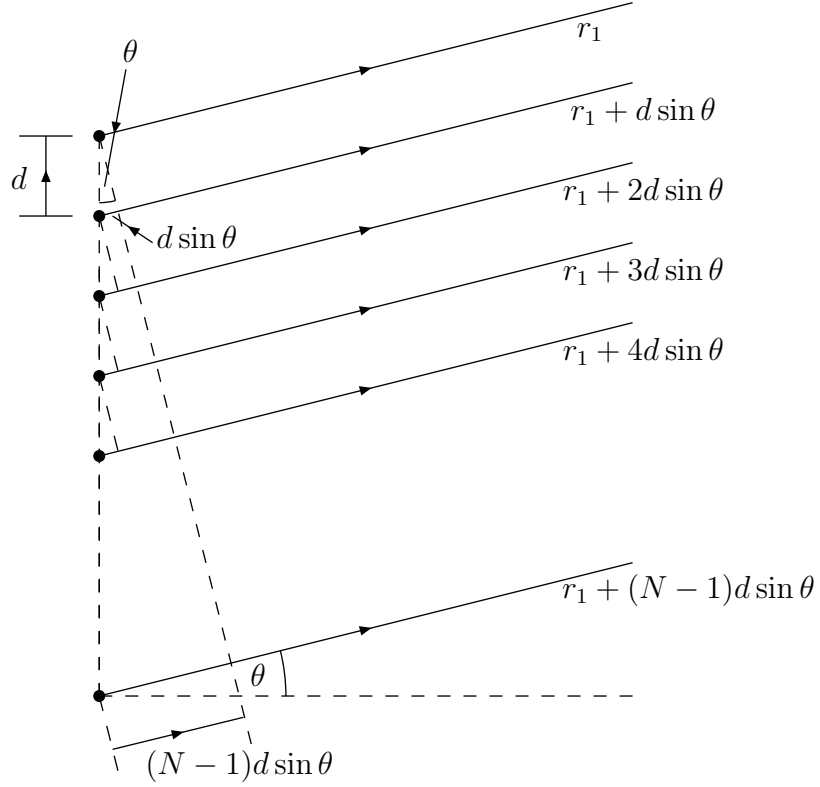


Figure 34: The path difference introduced from transmission through N very narrow slits.

where $\delta = (2\pi/\lambda)d \sin \theta$. We can pull out $E \exp(ikr_1)$ as a common factor and re-express as

$$\begin{aligned} A_{\text{Tot}} &= E \exp(ikr_1) (1 + \exp(i\delta) + \exp(2i\delta) + \cdots \exp((N-1)i\delta)) \\ &= E \exp(ikr_1) (1 + \exp(i\delta) + (\exp(i\delta))^2 + \cdots (\exp(i\delta))^{N-1}) \end{aligned} \quad (102)$$

The term within brackets is a geometric series, and can be summed using

$$1 + x + x^2 + x^{N-1} = \frac{1 - x^N}{1 - x}. \quad (103)$$

This results in

$$A_{\text{Tot}} = E \exp(ikr_1) \frac{1 - \exp(iN\delta)}{1 - \exp(i\delta)}. \quad (104)$$

Using the same trick as for the double slits and Michelson interferometer we pull out common factors giving

$$\begin{aligned} A_{\text{Tot}} &= E \exp(ikr_1) \frac{\exp(iN\delta/2)}{\exp(i\delta/2)} \left(\frac{\exp(-iN\delta/2) - \exp(iN\delta/2)}{\exp(-i\delta/2) - \exp(i\delta/2)} \right) \\ &= E \exp(ikr_1 + (N-1)\delta/2) \frac{\sin(N\delta/2)}{\sin(\delta/2)}. \end{aligned} \quad (105)$$

The intensity is proportional to the total amplitude squared so

$$I \propto |A_{\text{Tot}}|^2 = |E|^2 \frac{\sin^2(N\delta/2)}{\sin^2(\delta/2)}. \quad (106)$$

Introducing a normalisation constant I_0 , the intensity of the light from a diffraction grating is equal to

$$I(\theta) = I_0 \frac{\sin^2(N(\pi/\lambda)d \sin \theta)}{\sin^2((\pi/\lambda)d \sin \theta)}. \quad (107)$$

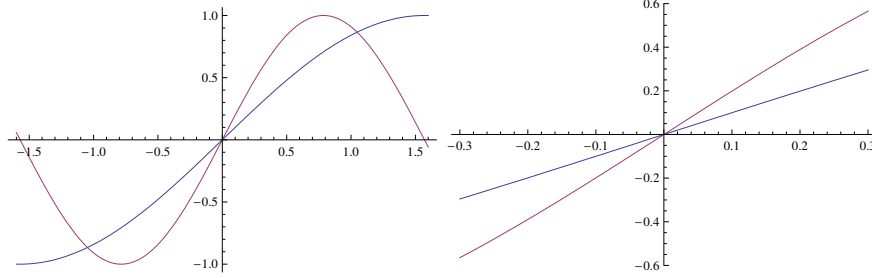


Figure 35: The relative x -dependence of $\sin x$ and $\sin 2x$ (left) and the same focusing on very small x (right).

Let us consider the form of the pattern this equation describes. It is easiest to start with the limit $\theta \rightarrow 0$. In this case all the light has travelled the same distance to the screen and there is no path difference between successive waves. Hence, this is the situation where we have perfect constructive interference. In order to obtain the expression for the intensity we need to examine the limit of the expression

$$I(\theta) = I_0 \frac{\sin^2 Nx}{\sin^2 x} \quad (108)$$

as $x \rightarrow 0$, where $x = (\pi/\lambda)d \sin \theta$. For small argument x we have the limit $\sin x \rightarrow x$ and $\sin Nx \rightarrow Nx$. Hence, both expressions $\rightarrow 0$, but $\sin x$ is smaller in this region than $\sin Nx$. This is shown in figure 35, where we consider $\sin 2x$ compared to $\sin x$. The former rises twice as quickly and reaches its maximum at $x = \pi/4$ whereas the latter takes until $\pi/2$. Regions nearer to 0 are shown in the right-hand figure and one can see that $\sin x$ is always smaller than $\sin 2x$ in this region. Using the expressions valid in the limit $x \rightarrow 0$ we obtain

$$I(0) = I_0 \frac{\sin^2 Nx}{\sin^2 x} \Big|_{x \rightarrow 0} \rightarrow I_0 \frac{N^2 x^2}{x^2} = I_0 N^2. \quad (109)$$

Since this is due to perfect constructive interference it is the maximum intensity and is known as a primary maximum. It will happen for any other value of $(\pi/\lambda)d \sin \theta$ where $\sin((\pi/\lambda)d \sin \theta) \rightarrow 0$, i.e. $(\pi/\lambda)d \sin \theta = m\pi$, since in all these cases both $\sin^2((\pi/\lambda)d \sin \theta)$ and $\sin^2(N(\pi/\lambda)d \sin \theta)$ will behave in the same way as for $\theta \rightarrow 0$. The condition for a primary maximum is therefore

$$(\pi/\lambda)d \sin \theta = m\pi \quad \rightarrow \quad d \sin \theta = m\lambda, \quad (110)$$

i.e. the expected result that the path difference between each successive ray is an integer number of wavelengths.

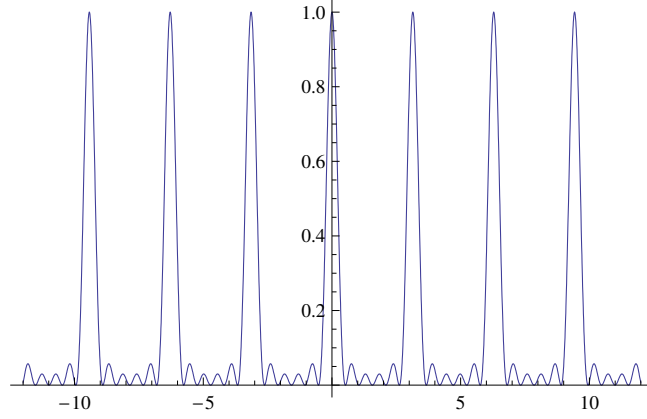


Figure 36: The diffraction pattern from a grating with 6 slits with intensity in units of $I_0 N^2$.

The intensity pattern will have zeros for all θ where $\sin(N(\pi/\lambda)d \sin \theta) = 0$, i.e. for $(\pi/\lambda)d \sin \theta = (k/N)\pi$ (k is an integer), except for the special case where $\sin((\pi/\lambda)d \sin \theta)$ is also zero, i.e. $(\pi/\lambda)d \sin \theta = m\pi$, where we have the primary maxima. This corresponds to the cases where $k = mN$. For given N there will be $N - 1$ values of k between two such conditions, so there are $N - 1$ zeros in the intensity between two primary maxima. The pattern is shown in figure 36. Between two successive zeros there must be another maximum. These will be approximately when $\sin^2(N(\pi/\lambda)d \sin \theta)$ is at its maximum value of 1. It is straightforward to check that the height of these maxima is much smaller than the height of the principal maxima. These maxima are called subsidiary maxima. There are $N - 2$ of them between the $N - 1$ zeros which exists between two principal maxima.

We note that the condition for a principal maximum, $d \sin \theta = m\lambda$, means that a maximum of given order m occurs at smaller θ for smaller wavelength light. This means it may be possible to have more maxima (which must be for $\theta < 90^\circ$) for some lower wavelength (e.g. blue) than a longer wavelength (e.g. red). It also means that, as for the Fabry-Perot etalon, we can use a diffraction grating to distinguish between two closely separated spectral lines of slightly different wavelength.

5.1.1 Chromatic Resolution of a Diffraction Grating

We consider two wavelengths of light λ_1 and $\lambda_1 + \Delta\lambda$, where $\Delta\lambda \ll \lambda_1$. We wish to define the chromatic resolving power of a grating $|\lambda_1/(\Delta\lambda_{\min})|$. Two patterns are said to be just resolvable if the primary maximum of the light from one wavelength lies exactly at the position of the first zero of the other wavelength, as shown in figure 37.

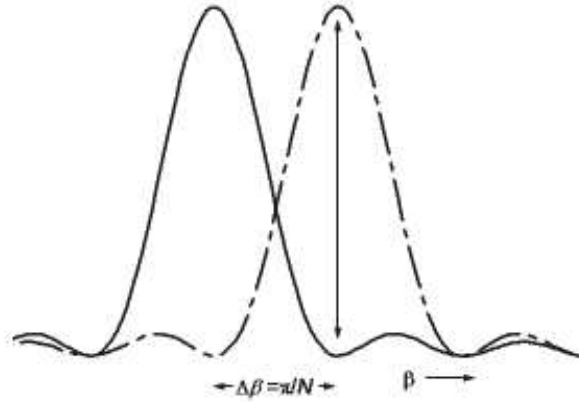


Figure 37: The criterion for telling apart two patterns from slightly different wavelengths.

Let us consider the pattern for the light with wavelength λ_1 . If the primary maximum is at $(\pi/\lambda_1)d \sin \theta = m\pi$, the first zero is at $m\pi + \pi/N$. Hence for fixed λ_1 the change in angle from the maximum to 0 may be expressed using

$$(\pi/\lambda_1)d \Delta \sin \theta = \pi/N \quad \rightarrow \quad \Delta \sin \theta = \frac{\lambda_1}{Nd}. \quad (111)$$

Considering instead how the position of the maxima move as λ changes, we have the maximum at $d \sin \theta = m\lambda$, which means

$$d \Delta \sin \theta = m \Delta \lambda \quad \rightarrow \quad \Delta \sin \theta = \frac{m \Delta \lambda}{d}. \quad (112)$$

Equating the two angular changes to obtain the minimum $\Delta \lambda$ we obtain

$$\frac{\lambda_1}{Nd} = \frac{m \Delta \lambda_{\min}}{d} \quad \rightarrow \quad \frac{\lambda_1}{\Delta \lambda_{\min}} = mN. \quad (113)$$

This defines the chromatic resolving power of the grating. As with the Fabry-Perot etalon the resolution increases with order of the fringe, since at wider angles for the grating there is more dispersion of wavelengths. For the grating the resolution also increases with the number of slits.

5.2 Diffraction from a Finite Width Slit

In order to fully derive the diffraction pattern from a single slit we have to use the technique of Fourier transforms, which you will cover in maths next year. Hence, here we will obtain the pattern using an argument based on the diffraction pattern we have already derived. We can think of a single slit with width a as a very large number of equally spaced narrow slits, each at the centre of a line of length d (so the separation between slits is also d). If there are N narrow slits, each occupying distance d the width of the single slit is $Nd = a$.

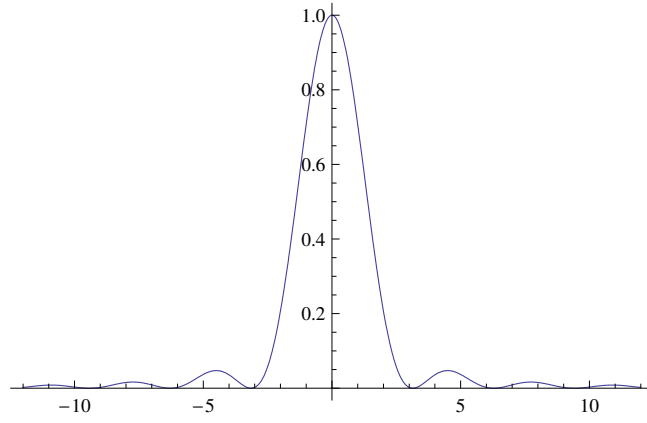


Figure 38: The diffraction pattern from a single slit of finite width.

We can then let $N \rightarrow \infty$ and $d \rightarrow 0$ such that $Nd = a$ is a finite constant. Doing this we know that the intensity of the array of narrow slits is

$$\begin{aligned} I(\theta) &= I_0 \frac{\sin^2(N(\pi/\lambda)d \sin \theta)}{\sin^2((\pi/\lambda)d \sin \theta)} \\ &= I_0 \frac{\sin^2((\pi/\lambda)a \sin \theta)}{\sin^2((\pi/\lambda)(a/N) \sin \theta)}. \end{aligned} \quad (114)$$

But as $N \rightarrow \infty$, $a/N \rightarrow 0$ and

$$\sin((\pi/\lambda)(a/N) \sin \theta) \rightarrow (\pi/\lambda)(a/N) \sin \theta. \quad (115)$$

This means that

$$I = I_0 \frac{N^2 \sin^2((\pi/\lambda)a \sin \theta)}{((\pi/\lambda)a \sin \theta)^2}. \quad (116)$$

However, we have implicitly always been talking about energy density of light which is proportional to $|E|^2$. As we let $N \rightarrow \infty$ the amount of light transmitted from each region of the slit shrinks like $1/N$. Hence, the effective amplitude of the light undergoing interference should be divided by N to reflect this, and therefore the intensity should be divided by N^2 . Doing this we obtain the correct answer for a slit of finite width a ,

$$I = I_0 \frac{\sin^2((\pi/\lambda)a \sin \theta)}{((\pi/\lambda)a \sin \theta)^2}. \quad (117)$$

This is of the form $\frac{\sin^2 x}{x^2}$ where $x = (\pi/\lambda)a \sin \theta$. This is maximum as $x \rightarrow 0$ and we get

$$\frac{\sin^2 x}{x^2} \rightarrow \frac{x^2}{x^2} = 1. \quad (118)$$

It is a minimum for $x \equiv (\pi/\lambda)a \sin \theta = m\pi$ for integer m , i.e. for $a \sin \theta = m\lambda$. There are further maximum very near to where $\sin^2((\pi/\lambda)a \sin \theta)$ is a maximum, i.e. at $(\pi/\lambda)a \sin \theta = (m + \frac{1}{2})\pi$, i.e. at $a \sin \theta = (m + \frac{1}{2})\lambda$. The diffraction pattern from a single slit is shown in figure 38. One can see that the central maximum is much higher than the other maxima.

5.2.1 Resolution of Images – Rayleigh Criterion.

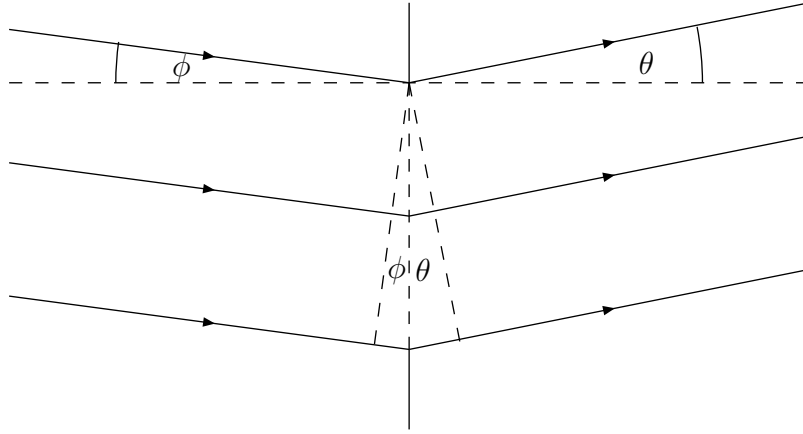


Figure 39: Light incident at a small angle ϕ to a slit.

For the single slit it is interesting to consider the resolution of images of sources at slightly different points. Let us consider that we have one object which emits light which arrives exactly perpendicular to the slit, and another where the light arrives at a small angle ϕ , i.e. the two sources are separated by the angle ϕ . The light arriving at the slit from first source produces the diffraction pattern we have just derived. That arriving from the second appears as in figure 39. In this case the path difference between waves from different parts of the slit is proportional to $a \sin \theta$ from the path difference after travelling through the slit, and there is an additional difference of $a \sin \phi$ in the light arriving at the slit. This second path difference could be $-\sin \phi$ if the small angle ϕ is the other side of the normal to the slit. Hence, for the second source the path difference is

$$\sin \theta \pm \sin \phi \approx \sin(\theta \pm \phi), \quad (119)$$

where we have used the approximation that both angles are small, which will be true if we are looking at the limits of resolution.

This means that the diffraction pattern from the second source is

$$I = I_0 \frac{\sin^2((\pi/\lambda)a \sin(\theta \pm \phi))}{((\pi/\lambda)a \sin(\theta \pm \phi))^2}, \quad (120)$$

which is the same as before but shifted by an angle of $\pm \phi$. We can distinguish the two sources if we can see that there are two separate patterns. The criterion for this is called Rayleigh's criterion, and is that the two patterns are said to be distinguishable if the maximum of one is at the minimum of the other. The minimum of the first pattern is at $(\pi/\lambda)a \sin \theta = \pi$. This is the maximum of

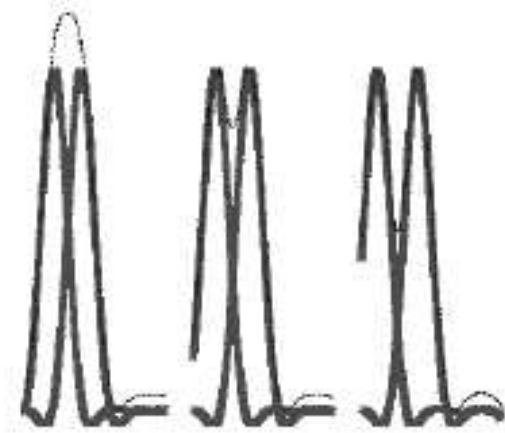


Figure 40: The Rayleigh criterion for telling apart the images from two closely separated sources.

the second pattern if $\theta = \pm\phi$. This means the two objects are distinguishable if

$$|(\pi/\lambda)a \sin \phi| = \pi \quad \rightarrow \quad |\sin \phi| = \lambda/a = \sin \theta_R \approx \theta_R. \quad (121)$$

So the minimum resolvable angle using Rayleigh's criterion, θ_R is (λ/a) . The patterns from the two objects are shown in figure 40, where we move from separations less than Rayleigh's criterion on the left, to exactly at minimum separation, to greater than minimum separation. When the two patterns are separated by the criterion amount the minimum intensity in between the two peaks can be found easily by noting that in the middle both patterns have argument $\pi/2$. Hence the intensity is

$$I = 2 \times I_0 \frac{\sin^2(\pi/2)}{(\pi/2)^2} = 2 \times \frac{1}{\pi^2/4} = \frac{8}{\pi^2} \approx 0.8. \quad (122)$$

If we have a circular aperture rather than a 1-dimensional slit, the pattern formed and argument for the limit of resolutions is very similar, but the diffraction pattern is a Bessel function rather than depending on the sin function. The features are much the same, but a detailed study shows that in this case the limit of resolution is $\theta_R = 1.22\lambda/D$ where D is the diameter of the circle.

5.2.2 The Limit of Resolution of the Eye

Typically the human eye has a pupil diameter (for normal light intensity) of about 2mm. Therefore the limit of resolution is

$$\Delta\theta = \frac{1.22 \times 500 \times 10^{-9}}{2 \times 10^{-3}} = 3 \times 10^{-4} \text{ rads} = 0.02^\circ. \quad (123)$$

At 100m, using $\tan \theta \approx \theta$, which is an excellent approximation for these small angles, this corresponds to a separation of two objects of about 3cm. This is

about the best an eye can possibly do. The image of any object is formed on the retina at the back of the eye, and will subtend the angle formed by the object. Since one cannot resolve angles smaller than 0.02° this means there is no point having light receptors in the retina closer together than this angle. Indeed, this turns out to be true. The density of receptors is no greater than required by this limit.

5.3 Gratings with Finite Width Slits

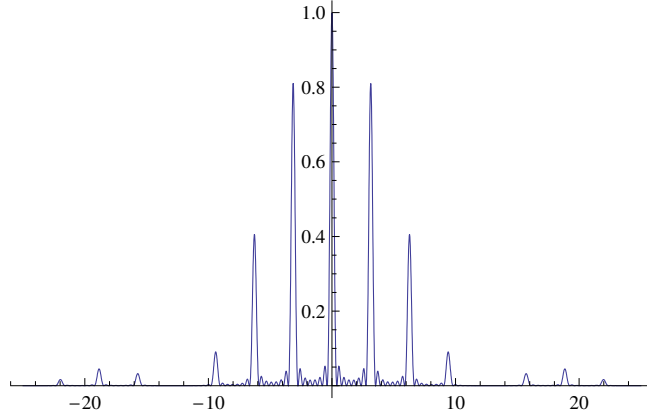


Figure 41: The diffraction pattern from a diffraction grating with slits of finite width. The x axis is in units of $(\pi/\lambda)d \sin \theta$.

We have seen that the diffraction pattern from a grating with N very narrow slits of separation d is

$$I(\theta) = I_0 \frac{\sin^2(N(\pi/\lambda)d \sin \theta)}{\sin^2((\pi/\lambda)d \sin \theta)}, \quad (124)$$

and that from a single slit with width a is

$$I(\theta) = I_0 \frac{\sin^2((\pi/\lambda)a \sin \theta)}{((\pi/\lambda)a \sin \theta)^2}. \quad (125)$$

When we consider a grating to have finite width slits the resulting pattern is just the product of that for the grating and the shape of that for the single slit, i.e.

$$I(\theta) = I_0 \frac{\sin^2(N(\pi/\lambda)d \sin \theta)}{\sin^2((\pi/\lambda)d \sin \theta)} \times \frac{\sin^2((\pi/\lambda)a \sin \theta)}{((\pi/\lambda)a \sin \theta)^2}. \quad (126)$$

It must be the case that $d > a$, so the first minimum of the single slit pattern must be at a greater angle than the first primary maximum away from the centre. The general form of the pattern is shown in figure 41. One sees that the primary maxima fall until the first single slit minimum, then rise and fall again, with much smaller size. This example has $d = 4a$, but good gratings will have d/a greater and so will have slower fall of the primary maxima intensity. Note

that the primary maxima are at $d \sin \theta = m\lambda$ and the diffraction minima for the single slit are at $a \sin \theta = k\lambda$, where m and k are integers. If the ratio d/a is an integer then primary maxima and diffraction minima will coincide, and when $d/a = m/k$ the order m grating maxima will be missing. Indeed, in figure 41 there are no maxima for $m = 4$ or $m = 8$, or indeed if we plotted further for any multiple of 4.