Week 9: Diffraction

Diffraction is another manifestation of the wave property of light. Effects of diffraction can be seen in innumerable naturally occurring situations such as the blurring around shadows, the intensity redistribution when light falls on opaque obstacles or passes through slits. The usual intuition about light when it passes through a slit or is obstructed, fails somewhat because of the appearance of light in regions where we may not expect it to be there. For instance when light passes through a slit, it is expected that it would just illuminate the region which represents the region corresponding to the slit, on the screen. However, we see light beyond this region and we also note a regular pattern. For a circular obstacle we find that just behind its centre, sometimes there is a bright spot (known as Poisson's spot).

Loosely speaking, one says that light bends around obstacles. Such a thing occurs for sound or matter waves as well. How is this bending different from **refraction?**. In refraction, the shorter wavelengths refract more than the longer ones whereas in diffraction it is the longer wavelengths that deviate more than the shorter ones.

The study of diffraction essentially involves (like in interference) three elements: source (S), aperture (Σ) and screen (σ) . The distances between S, Σ and Σ , σ have a crucial role to play. Look at Figure 1, where we have a source very far away emitting spherical waves which have almost planar wavefronts as they reach the aperture plane Σ . The figure at the bottom, is that of the case when the screen is near the aperture plane. We then move the screen slowly away and observe the pattern on the screen. As we keep moving away, we note that some structure appears. When the screen is very far away we find that the pattern has a distinct structure and the structure goes well beyond the location of the aperture. This limit, when the source and the screen are very far away from Σ is called the **Fraunhofer** limit and a study of it is known as **Fraunhofer diffraction**. When the screen is closer to the aperture (i.e. it is at a smaller, finite distance) we are in the **near-field** or Fresnel zone and we observe **Fresnel diffraction**. We will mainly confine ourselves to the study of Fraunhofer diffraction.

It is true that even though diffraction does always occur, we must know the criterion under which we have **Fraunhofer diffraction**. A rule of thumb is

$$R > \frac{a^2}{\lambda} \tag{1}$$

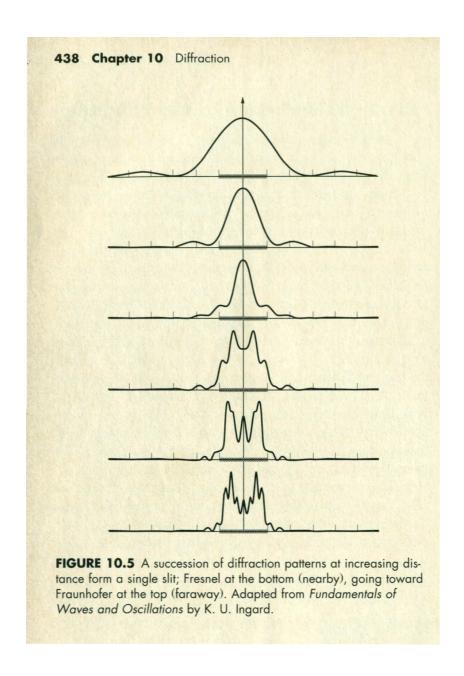


FIG. 1. Fresnel and Fraunhofer

where R is the smaller of the distances between S and Σ and Σ and σ . If R is infinity then a has no role. Increase in λ implies a shift towards the Fraunhofer limit, for fixed a.

To study diffraction quantitatively, we need to first learn how to construct the **diffraction** integral. Let us assume (as shown in Fig. 2) that the slit lies between $-\frac{D}{2}$ and $\frac{D}{2}$ along the y axis (see Figure 2). When light is incident on this slit from behind, by the Huygens-Fresnel principle each point on the illuminated slit will act as a source of secondary wavelets which will subsequently interfere. Now consider the case where there are N such sources inside

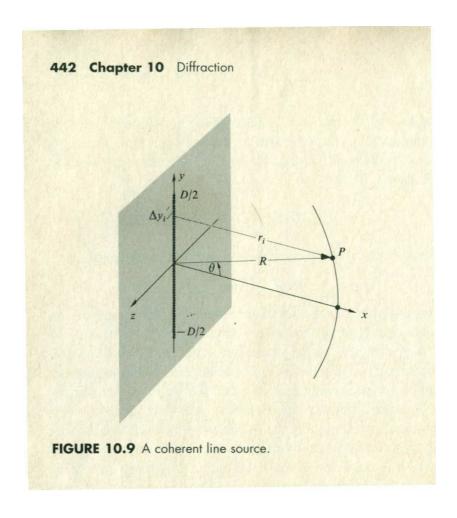


FIG. 2. Fraunhofer diffraction by a single slit: diffraction integral

the slit. Let us further consider a region of infinitesimal width Δy_i located at y_i in the slit. There will be $N\frac{\Delta y_i}{D}$ sources in Δy_i . The width of the slit is usually assumed to be smaller than λ or, to get a good diffraction pattern, they should be of comparable dimensions. We take the region of width D as divided into M segments of width Δy_i (i.e. i runs from 1 to M). For a spherical wavelet, we write its electric field as

$$E = \frac{\mathcal{E}_0}{r}\sin\left(\omega t - kr\right) \tag{2}$$

where \mathcal{E}_0 is the source strength. Therefore, for the *ith* segment, the net electric field, by linear superposition will be

$$E_i = \frac{\mathcal{E}_0}{r_i} \left[\sin \left(\omega t - k r_i \right) \right] \frac{N \Delta y_i}{D} \tag{3}$$

In the limit $N \to \infty$, $\mathcal{E}_0 \to 0$, we assume

$$\mathcal{E}_L = \frac{1}{D} \lim_{N \to \infty} \mathcal{E}_0 N \tag{4}$$

The net electric field for all the segments is thus obtained as

$$E = \sum_{i=1}^{M} \frac{\mathcal{E}_L}{r_i} \sin(\omega t - kr_i) \, \Delta y_i \tag{5}$$

In the continuum limit, when $M \to \infty$ and $\Delta y_i \to dy$ we get,

$$E = \mathcal{E}_L \int \frac{\sin(\omega t - kr)}{r} dy \tag{6}$$

This is known as the **diffraction integral**. Note that r(y) is a function of y here.

Single slit diffraction: Let us now turn to the simplest study of diffraction—i.e. diffraction by a single slit. We choose the coordinate y as shown in the Figure 3. With this choice, the slit lies between $-\frac{b}{2}$ to $\frac{b}{2}$ and therefore, the diffraction integral becomes:

$$E = \mathcal{E}_L \int_{-\frac{b}{2}}^{\frac{b}{2}} \frac{\sin(\omega t - kr)}{r} dy \tag{7}$$

We now make an approximation. It is easy to note (see the geometry in Figure 3) that with the definitions of y, θ and R, we have

$$r^2 = R^2 + y^2 - 2yR\sin\theta\tag{8}$$

In our case, R is large and the domain of y is small (since b is small). Therefore, we may write

$$r = R \left(1 + \frac{y^2}{R^2} - 2\frac{y}{R}\sin\theta \right)^{\frac{1}{2}}$$

$$\approx R \left(1 - 2\frac{y}{R}\sin\theta \right)^{\frac{1}{2}}$$

$$\approx R - y\sin\theta \tag{9}$$

We use this approximation in the diffraction integral to get

$$E = \mathcal{E}_L \int_{-\frac{b}{2}}^{\frac{b}{2}} \frac{\sin(\omega t - k(R - y\sin\theta))}{R} dy$$
 (10)

where we have retained the $y \sin \theta$ in the phase but removed it in the denominator. This is because, the $R\left(1-\frac{y}{R}\right)$ in the denominator, can be expanded and the $\left(1-\frac{y}{R}\right)$ can be brought to the numerator as $\left(1+\frac{y}{R}\right)$ to give an extra second piece proportional to $\frac{y}{R}$ —this is small (since y is small and R is large) and therefore, can be ignored in comparison to the other term.

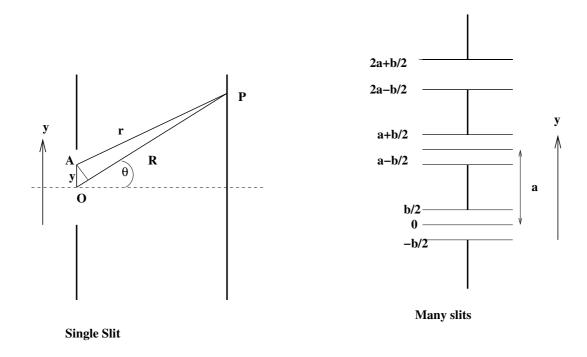


FIG. 3. Single slit and many slits

We now need to evaluate this integral to get the electric field at the observation point P. We have

$$E = \mathcal{E}_L \int_{-\frac{b}{2}}^{\frac{b}{2}} \frac{\sin\left(\omega t - k\left(R - y\sin\theta\right)\right)}{R} dy \tag{11}$$

$$= \frac{\mathcal{E}_L}{R} Im \left[\int_{-\frac{b}{2}}^{\frac{b}{2}} e^{i(\omega t - kR - ky\sin\theta)} dy \right]$$
 (12)

$$= \frac{\mathcal{E}_L}{R} Im \left[e^{i(\omega t - kR)} \left(\frac{e^{iky\sin\theta}}{ik\sin\theta} \right)_{-\frac{b}{2}}^{\frac{b}{2}} \right]$$
 (13)

$$= \frac{\mathcal{E}_L b}{R} Im \left[\left(\frac{\sin \beta}{\beta} \right) e^{i(\omega t - kR)} \right]$$
 (14)

$$= \frac{\mathcal{E}_L b}{R} \left(\frac{\sin \beta}{\beta} \right) \sin \left(\omega t - kR \right) \tag{15}$$

Therefore, the intensity, given by $\langle EE^*\rangle_T$ is given as

$$I = I_0 \left(\frac{\sin \beta}{\beta}\right)^2 \tag{16}$$

where

$$\beta = \frac{kb\sin\theta}{2} \tag{17}$$

and $I_0 = \frac{1}{2} \left(\frac{\mathcal{E}_L b}{R} \right)^2$.

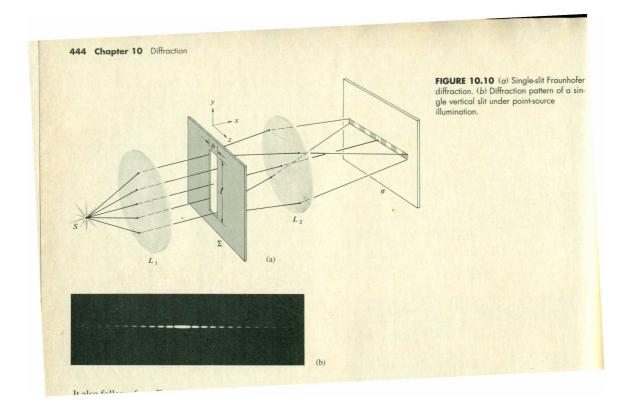


FIG. 4. Single slit

The intensity distribution for the single slit is shown in Figure 4. Where do the maxima and minima occur? To see this, let us differentiate the intensity expression w.r.t. β . This gives, as the condition for extrema the following equation:

$$\frac{dI}{d\beta} = 2I_0 \frac{\sin \beta}{\beta} \left(\frac{\cos \beta}{\beta} - \frac{\sin \beta}{\beta^2} \right) = 0 \tag{18}$$

Thus the extrema occur at $\beta=m\pi$ $(m=0,\pm 1,\pm 2...)$ and $\tan\beta=\beta$. It is easy to see that $\beta=0$ is the central maximum. $\beta=m\pi$ $(m=\pm 1,\pm 2,\pm 3...)$ are minima. Further, one can solve the transcendental equation $\tan\beta=\beta$ to obtain the maxima at $\beta=\pm 1.43\pi,\pm 2.46\pi,\pm 3.47\pi,...$.

Thus, the condition for minima turns out to be

$$b\sin\theta = m\lambda\tag{19}$$

where $m = \pm 1, \pm 2....$

Exercise: Check all the above-stated maxima minima of the single slit intensity distribution by calculating the second derivative of the intensity distribution function.

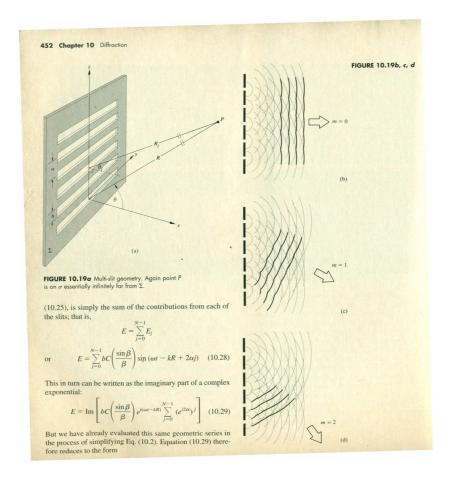


FIG. 5. Multiple slits

Multiple slits: We will now move towards obtaining the intensity distribution for many slits placed along y between $-\frac{b}{2}, \frac{b}{2}$; $a - \frac{b}{2}, a + \frac{b}{2}$; $2a - \frac{b}{2}, 2a + \frac{b}{2}$;; $(N-1)a - \frac{b}{2}$, $(N-1)a + \frac{b}{2}$ (see Figure 3, right side figure and also Figure 5). The total electric field, in this case is given as

$$E = C \int_{-\frac{b}{2}}^{\frac{b}{2}} F(y) dy + C \int_{a-\frac{b}{2}}^{a+\frac{b}{2}} F(y) dy = C \int_{2a-\frac{b}{2}}^{2a+\frac{b}{2}} F(y) dy + \dots + C \int_{(N-1)a-\frac{b}{2}}^{(N-1)a+\frac{b}{2}} F(y) dy$$
 (20)

where

$$F(y) = \sin(\omega t - k(R - y\sin\theta)) \tag{21}$$

Notice that we have kept θ as the same for all the integrals which is actually not the case. This assumption is based on the fact that the slits are all close to the origin and therefore $\theta_j \approx \theta$. Let us now consider the electric field at P due to the jth slit. This is given as

$$E_{j} = \frac{C}{k \sin \theta} \left[\sin \left(\omega t - kR \right) \sin \left(ky \sin \theta \right) - \cos \left(\omega t - kR \right) \cos \left(ky \sin \theta \right) \right]_{ja - \frac{b}{2}}^{ja + \frac{b}{2}}$$
 (22)

After some straighforward algebra, we get

$$E_{j} = bC \frac{\sin \beta}{\beta} \sin (\omega t - kR + 2\alpha j)$$
(23)

where $\alpha = \frac{ka\sin\theta}{2}$.

Exercise: Work through the algebra to arrive at the above expression for E_i .

To get the total electric field one now needs to sum over all the j. Therefore, we have

$$E = bC \frac{\sin \beta}{\beta} \sum_{j=0}^{N-1} \sin (\omega t - kR + 2\alpha j)$$
 (24)

$$=bC\frac{\sin\beta}{\beta}Im\left[e^{i(\omega t-kR)}\sum_{j=0}^{N-1}\left(e^{2i\alpha}\right)^{j}\right] \tag{25}$$

The series sum which we need to evaluate is straightforward. We have

$$\sum_{i=0}^{N-1} \left(e^{2i\alpha} \right)^j = \frac{e^{2iN\alpha} - 1}{e^{2i\alpha} - 1} = e^{i(N-1)\alpha} \frac{\sin N\alpha}{\sin \alpha}$$
 (26)

Thus the total electric field is

$$E = bC \frac{\sin \beta}{\beta} \frac{\sin N\alpha}{\sin \alpha} Im \left[e^{i(\omega t - kR + \alpha(N-1)))} \right]$$
 (27)

$$=bC\frac{\sin\beta}{\beta}\frac{\sin N\alpha}{\sin\alpha}\sin\left(\omega t - kR + \alpha(N-1)\right))$$
(28)

Hence, the intensity distribution at P, due to N identical slits, obtained after calculating $\langle EE^*\rangle_T$, is given as

$$I = I_0 \left(\frac{\sin \beta}{\beta}\right)^2 \left(\frac{\sin N\alpha}{\sin \alpha}\right)^2 \tag{29}$$

where $I_0 = \frac{b^2 C^2}{2}$. For N = 2, i.e. a double slit, we get

$$I = 4I_0 \left(\frac{\sin \beta}{\beta}\right)^2 \cos^2 \alpha \tag{30}$$

The intensity distribution for N slits can be written exclusively as a function of β , using the fact that

$$\alpha = -\frac{a}{b}\beta \tag{31}$$

We then have

$$I = I_0 \left(\frac{\sin \beta}{\beta}\right)^2 \left(\frac{\sin N \frac{a}{b} \beta}{\sin \frac{a}{b} \beta}\right)^2 \tag{32}$$

The above expression, analysed as a function of β , will give the location of all the extrema.

Double slit: Notice that the double slit intensity distribution given as

$$I = 4I_0 \left(\frac{\sin \beta}{\beta}\right)^2 \cos^2 \alpha \tag{33}$$

is the Young's double slit intensity distribution **modulated** by the term $\left(\frac{\sin \beta}{\beta}\right)^2$. Figure 6 shows the double slit and its intensity distribution.

We will discuss the maxima minima structure of the double and multiple slits in the next lecture. This will lead us (for many slits) to a study of the diffraction grating and its properties.

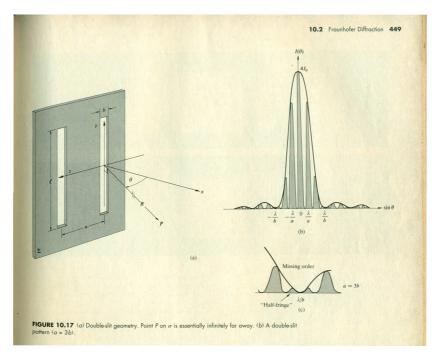


FIG. 6. Double slit

We have obtained the intensity distribution for a single slit and then, for N regularly spaced identical slits. Let us recall the N-slit intensity distribution:

$$I = I_0 \left(\frac{\sin \beta}{\beta}\right)^2 \left(\frac{\sin N\alpha}{\sin \alpha}\right)^2 \tag{34}$$

where $\beta = \frac{kb \sin \theta}{2}$ and $\alpha = \frac{ka \sin \theta}{2}$ (b is the slit width and a is the inter-slit distance). θ denotes the location of the point on the screen. We will now focus for a while on the double slit. The intensity distribution for a double slit (N=2) is given as:

$$I = 4I_0 \left[\left(\frac{\sin \beta}{\beta} \right)^2 \right] \cos^2 \alpha \tag{35}$$

The term in square brackets is the **modulation** which incorporates the intra-slit diffraction effects in the usual Young's double slit interference pattern $(4I_0\cos^2\alpha)$. It is useful to learn about the location of the maxima–minima for the above pattern.

To understand the maxima-minima locations we will adopt a somewhat general approach. Let us write the intensity distribution as

$$I = I_0 I_1^2 I_2^2 (36)$$

where $I_1 = \frac{\sin \beta}{\beta}$ and $I_2 = 2\cos \alpha$. We recall that $\alpha = \frac{a}{b}\beta$. Thus, both I_1 and I_2 are functions of β (or α). Let us now construct I' and I'' where the prime denotes differentiation w.r.t. β . We find them to be

$$I' = 2I_0 I_1 I_2 \left(I_1 I_2' + I_1' I_2 \right) \tag{37}$$

$$I'' = 2I_0 \left(I_1^2 I_2'^2 + I_2^2 I_1'^2 + 4I_1 I_1' I_2 I_2' + I_1^2 I_2 I_2'' + I_2^2 I_1 I_1'' \right)$$
(38)

Exercise: Obtain the above expressions for I' and I''

From the expression for I' (set equal to zero) we get the extrema conditions as

$$I_1 = 0 ; I_2 = 0 ; I_1 I_2' + I_1' I_2 = 0 (39)$$

It is easy to see that if $I_1=0$ or $I_2=0$ then I'' is always positive and therefore we have minima for these conditions. If $I_1=0$ we have $\beta=m\pi$ $(m=\pm 1,\pm 2,...)$. If $I_2=0$ we have $\frac{a}{b}\beta=\frac{2m+1}{2}\pi$ or, equivalently $\alpha=\frac{2m+1}{2}\pi$ $(m=0,\pm 1,....)$.

What about the condition $I_1I'_2 + I_2I'_1 = 0$? Using the expressions given above for I_1 and I_2 we get (after some simple adjustments)

$$\left(1 - \frac{\tan \beta}{\beta}\right) = \tan \beta \left(\frac{a}{b} \tan \frac{a}{b}\beta\right) \tag{40}$$

Therefore, given the relation between a and b we can solve the transcendental equation above to obtain the β . The trivial solution of the above equation for all a, b is $\beta = 0$. To see this put $\beta \to 0$ in the expressions for I_1 , I_2 and their derivatives to get $I_1'' \to -1$ and $I_2'' \to -2\left(\frac{a}{b}\right)^2$. Hence, as $\beta \to 0$, we have

$$I'' = -8I_0 \left[1 + \left(\frac{a}{b} \right)^2 \right] \tag{41}$$

and $\beta \to 0$ is a maximum.

Exercise: Work out the above stated result.

The other extrema can be found from the solutions of the transcendental equation given above. For example with a=2b one finds that the first maximum on either side appear at $\beta=\pm 1.433^c$.

Exercise: Show that $\alpha = m\pi$ $(m \neq 0)$ is an approximate solution for the other maxima which works pretty well for all practical purposes.

The generic features of the double slit pattern are as follows.

- Central maximum at $\beta = 0$.
- Minima at $\beta = m\pi \ (m = \pm 1, \pm 2, ...)$
- Subsidiary maxima at those $\beta \neq 0$ which solve the transcendental equation given above.
- If a = nb then there are 2n subsidiary maxima within a diffraction envelope. This 2n counts the half fringes at the edge minima.
- The diffraction minima at $\beta = m\pi(m \neq 0)$ coincides with the interference maxima when a = nb and $\alpha = mn\pi$. These correspond to the **missing orders** in the double slit pattern.

One can attempt to analyse the intensity distribution for N slits in the same way as given above for the double slit. This however will be a bit technically demanding, particularly the transcendental equation and its analysis. We therefore switch over to the variable α and state the results for the N slits. Note that here, the modulation of the N slit interference pattern is by the same diffraction term.

We state below the results for N slits.

• Principal maxima occur near $\alpha = m\pi$ $(m = 0, \pm 1,)$ i.e. wherever $\frac{\sin N\alpha}{\sin \alpha} = N$). The central maximum has an intensity N^2I_0 . The other principal maxima have **lower** intensities essentially due to the modulation by the $sinc^2\beta$. The maxima condition stated above is equivalent to

$$a\sin\theta_m = m\lambda\tag{42}$$

which is known as the diffraction grating equation.

- The **minima** occur wherever $\sin N\alpha = 0$, i.e. when $\alpha = \pm \frac{\pi}{N}, \pm \frac{2\pi}{N}, \dots$. There are N-1 **minima** between any two principal maxima.
- The **subsidiary maxima** occur according to the solutions of a transcendental equation. Approximately, they appear at $\alpha = \pm \frac{3\pi}{2N}, \pm \frac{5\pi}{2N}, \dots$ There are N-2 subsidiary maxima between two principal maxima. The intensity at a subsidiary maximum can be obtained by

substituting, say $\alpha = \frac{3\pi}{2N}$ in the intensity distribution function. This gives

$$I_{sub} = N^2 I_0 \left(\frac{\sin \beta}{\beta}\right)^2 \left(\frac{2}{3\pi}\right)^2 \approx \left(\frac{2}{3\pi}\right)^2 N^2 I_0 = \frac{1}{22} N^2 I_0$$
 (43)

where we have assumed β , α as small. It is clear that the intensity drop from the central maximum (N^2I_0) is quite substantial.

• The intensity at the principal interference maxima can be found by using $\alpha = m\pi$ ($m \neq 0$) and $\beta = \frac{b}{a}m\pi$. Substituting these expressions in the intensity function for N slits one obtains

$$I = I_0 \frac{N^2 a^2}{b^2 m^2 \pi^2} \sin^2 \frac{m\pi b}{a} \tag{44}$$

It is clear that when $m\pi^{\frac{b}{a}}=n\pi$ (i.e. $m^{\frac{b}{a}}=n,\ n\neq 0$, an integer) the intensity drops to zero. Assuming a=pb, this will happen if m=pn. If p=3 say, then the principal maxima of orders 3, 6 ... will be missing in the spectrum. Note that if p is irrational, there are no missing orders.

Features of the intensity distribution for N=6 are shown in Figure 7.

The N slit case is the example of the simplest **diffraction grating** known as a **transmission grating**. Such a device is useful in finding wavelengths, wavelength differences and other quantities which arise in spectroscopy. We shall now discuss three important quantities which are associated with a diffraction grating.

The first of these is the **angular width** of a principal maximum. To find this we write the following equation

$$a\sin\left(\theta_m + \Delta\theta_m\right) = \left(m + \frac{1}{N}\right)\lambda\tag{45}$$

where $\theta_m + \Delta \theta_m$ is the location of the minimum adjacent to the m th principal maximum. Assuming $\Delta \theta_m$ as a small quantity (i.e. assuming $\sin \Delta \theta_m \approx \Delta \theta_m$ and $\cos \Delta \theta_m \approx 1$) we get

$$\Delta\theta_m = \frac{\lambda}{Na\cos\theta_m} \tag{46}$$

This is known as the **angular width**. The smaller this width is, the sharper the diffraction pattern. A smaller width occurs for larger N.

A second quantity is known as the **dispersive power** defined as $D = \frac{d\theta}{d\lambda}$. This turns out to be (on differentiating the maxima condition w.r.t. λ)

$$D = \frac{d\theta}{d\lambda} = \frac{m}{a\cos\theta_m} \tag{47}$$

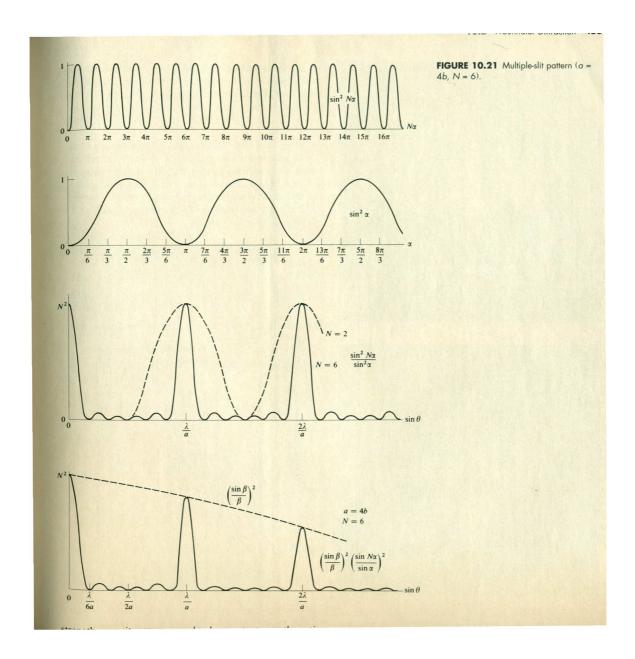


FIG. 7. Intensity distribution for N=6 slits with a=4b, figure from Hecht, Optics

Finally we obtain an expression for the **resolving power** of a diffraction grating. Let us assume that the source has two closely spaced wavelengths λ and $\lambda + \Delta \lambda$. If we want our grating to be able to resolve the two wavelengths, then the overlapping diffraction maxima must distinguish between them. The limit of resolution is set by the **Rayleigh criterion** where the minimum for λ falls on the maximum for $\lambda + \Delta \lambda$ (see Figure 8). If $\theta_m + \Delta \theta_m$ is the location of the minimum for λ then $\Delta \theta_m$ is equal to the angular width obtained just

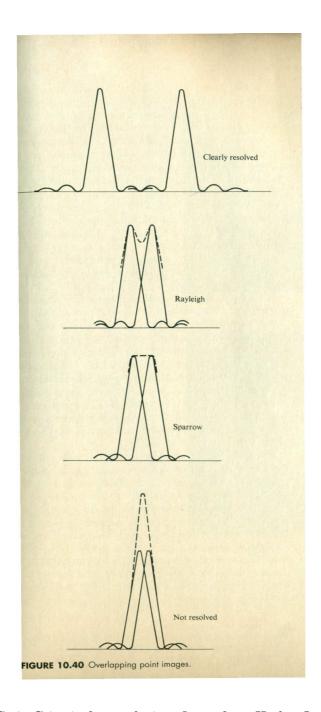


FIG. 8. Criteria for resolution, figure from Hecht, Optics

above. Further, the maximum for $\lambda + \Delta \lambda$ occurs such that

$$a\sin(\theta_m + \Delta\theta) = m(\lambda + \Delta\lambda) \tag{48}$$

The above equation, assuming $\Delta\theta$ as small gives

$$\Delta\theta = \frac{m\Delta\lambda}{a\cos\theta_m} \tag{49}$$

Equating $\Delta \theta = \Delta \theta_m$ results in

$$R = \frac{\lambda}{(\Delta \lambda)_{min}} = Nm \tag{50}$$

which gives the **chromatic resolving power** of a diffraction grating. Note that for large orders we have a larger resolution—but for large orders the intensity also drops. Recall the resolving power of the Fabry-Perot interferometer which was $0.97\mathcal{F}m$. Thus, the number of slits N, in a sense is the finesse (!) of a grating.

Figure 8 shows various types of criteria employed to obtain the resolving power. We have used the Rayleigh criterion.

In general, apertures could be two dimensional as well-like a rectangle, a circle or any other shape as such. The diffraction pattern will always reflect the symmetry aspects of the aperture. Let us assume that the aperture is specified by an **aperture function** $\mathcal{A}(y,z)$ where y,z are the coordinates whose domain specifies the aperture. The electric field at a point (Y,Z) on the screen is

$$E(Y,Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{A}(y,z)e^{ik(yY+zZ)}dydz$$
 (51)

Thus, the electric field is the **Fourier transform** of the aperture function. For a single slit, there will be just one integral, say, over z. The aperture function will be $\mathcal{A} = \mathcal{A}_0$ for $|z| < \frac{b}{2}$ and zero otherwise. On integrating one will obtain the electric field distribution for single slit diffraction.