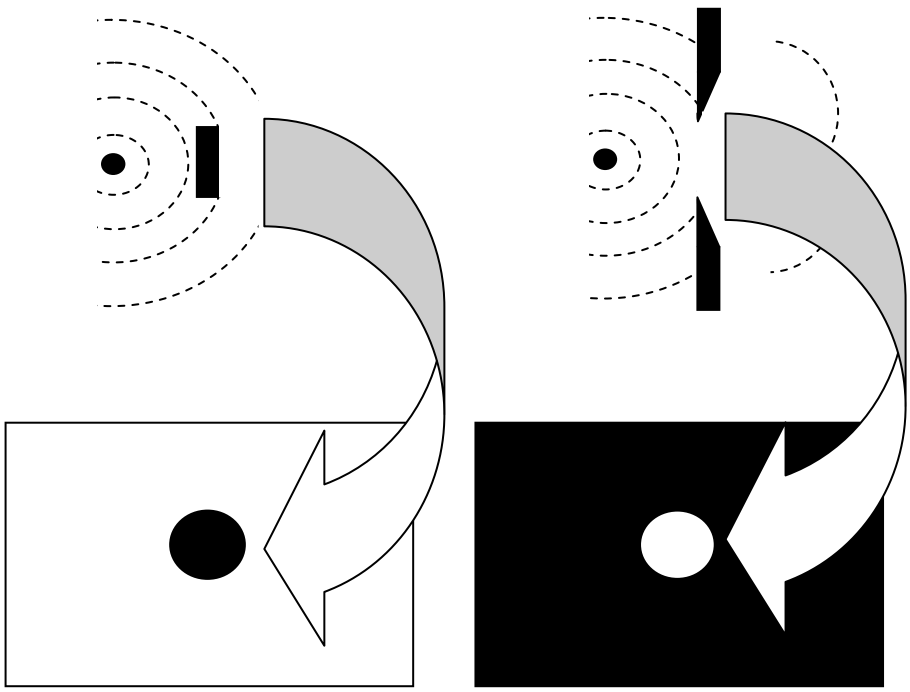
6. DIFFRACTION.

Introduction.

Diffraction and interference are approached in a very similar fashion and in fact there is very little difference between the two phenomena. Interference refers to the effects observed when two or more beams of light, separately derived from the same source, are brought together to show fringes. Diffraction, by comparison, refers to phenomena brought about by interference of light from point sources on a continuous portion of the same wavefront. For example, if we take a spherical or plane wave from a point source at its origin, O , we can easily find the field amplitudes and phases at some position, P, at any later time if we knew them at time t = 0 at O. If some opaque object or alternatively an aperture, is now placed between O and P there will be new field amplitudes and phases at P.

The form of the fields beyond the opaque object/aperture are complex and not purely suggestive of the sharp shadows predicted by a naïve application of ray optics. In general the intensity distribution at the edge of a shadow will show complex behavior. The new field beyond the opaque object is generated from the same waves that generated the original fields with the important exception that the elements from the area of the object are now missing. To find the new field/intensity distribution is the objective of **<u>diffraction theory</u>**.



The importance of these diffraction effects lies primarily in the fact that optical instruments frequently contain apertures and obstacles as part of their construction and we need to understand the action of these obstacles on the overall operation of the instrument. Such diffraction effects will for example place limits on the resolution of optical instruments. A good example is the circular aperture presented by a lens in many optical instruments.

Diffraction requires a reconsideration of **Huygen's Principle** which states that

Every point on a propagating wavefront acts as a new point source and serves as a source of spherical secondary waves. The amplitude of the new wavefront in advance is the product of the superposition of these secondary waves at the later time/advanced position.

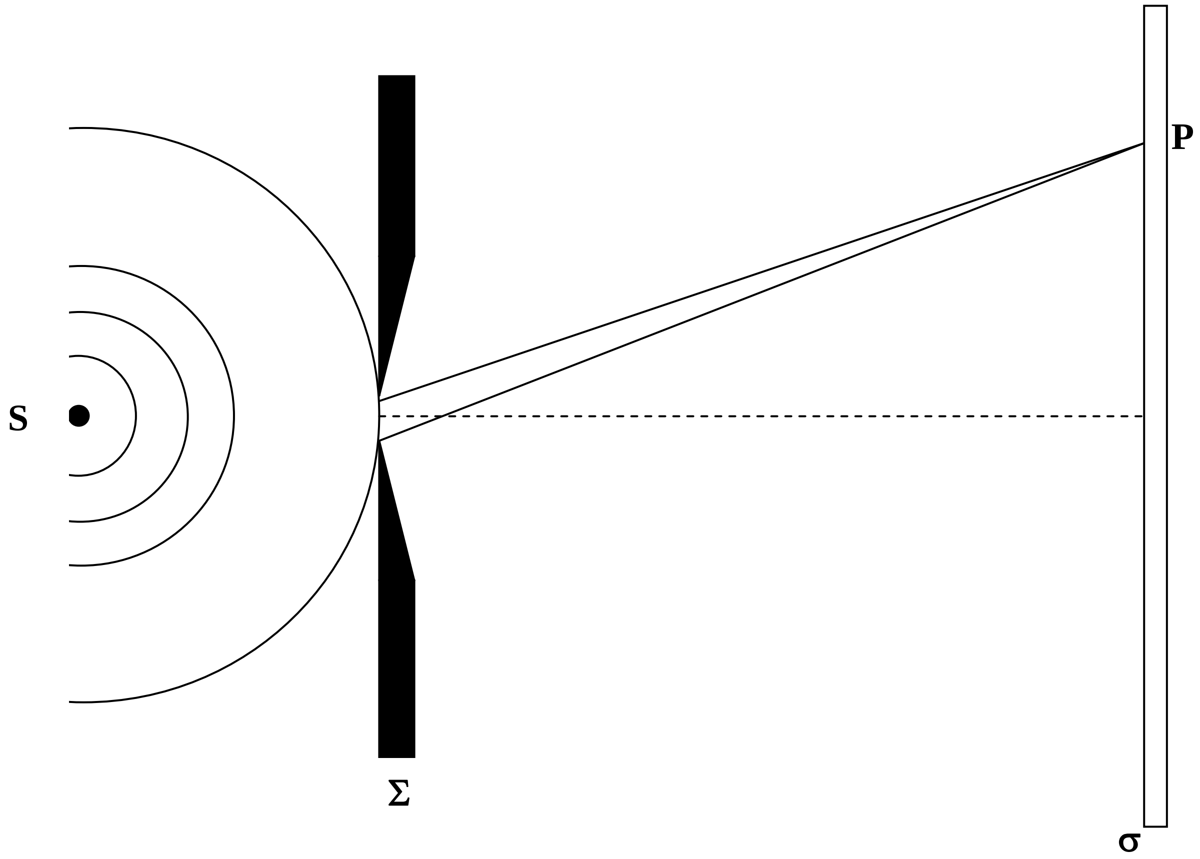
Much can be achieved following this principle but it is independent of any consideration of the wavelength involved and this becomes important when introducing the effect of obstacles on the advancing wavefront. It is well known that long wavelength radiation such as radio waves can advance around even large obstacles as long as the size of the obstacle is less than the wavelength under consideration eg. Trees usually act as no obstacle to radio of waves. Buildings on the other hand may introduce real obstacles that impede the progress of a radio wave and reception is only due to **reflection** around the building (ignoring wavelengths where reflection from the ionosphere or re-transmission from a satellite is important. These large objects also cast distinct shadows when illuminated by light.

Huygen's Principle has nothing to say about the phase of the advanced wave and to get a more accurate picture this needs to be accounted for in a revised version of Huygens Principle due to Fresnel, the **Huygens-Fresnel Principle** which includes phase and states that

Every unobstructed point on a propagating wavefront acts as a new point source and serves as a source of spherical secondary waves. The amplitude of the new wavefront in advance is the product of the superposition of these secondary waves at the later time/advanced position including the effects of both their amplitudes and relative phases.

Any approaches that attempt to incorporate the Huygens-Fresnel principle when calculating the effects of an aperture or any other restriction on a propagating wavefront are going to be difficult. Indeed, the problem of diffraction and calculations of its effects are among the most difficult (though not intractable) in optics. We will limit ourselves here to the examination of some simple but important diffraction problems that will also serve to illustrate the type of method that may be employed.

Before taking a detailed look at some simple problems it is useful to look at two limiting situations and to define their occurrence, namely near field and far field measurements.



In the above diagram a point source, S, illuminates an aperture in the aperture plane, Σ, after which the light passes through the aperture and impinges upon a screen in the image plane, σ. The precise details of what is observed in the image plane will depend on the wavelength of the light involved and the distance of the source from the aperture and of the aperture from the image plane or screen. Two types of diffraction theory are commonly distinguished depending on the approximations made. The simpler of these to work with is Fraunhoffer diffraction.

Fraunhofer diffraction

Fraunhofer diffraction requires

1. That the aperture plane is far from the point source, ie the distance is great compared with the aperture dimensions, then two simplifications in subsequent analysis result;

i) The wavefront at the aperture will approximate a plane wavefront (the radius of curvature is very large). That is, there is no relative phase difference, between the individual wavelets acting as the secondary sources, across the aperture.

Also

ii) The amplitude of the wavefront across the aperture will show no variation as all points across the aperture are approximately equidistant from the source (recall the amplitude falls as $\frac{1}{R}$ where R is the distance from the source to the aperture.

Both of these facts make the evaluation of the field at P on the screen much simpler when an integral across the aperture is evaluated.

If, further to this,

2. The screen is far enough from the aperture that rays arriving at the point P from across the aperture may be considered to propagate to P parallel to one another a further simplification in analysis is achieved and the relative phase difference of the rays on arrival at P compared with some ***arbitrarily chosen reference ray, commonly chosen as that arriving from a point in the centre of the aperture,*** will depend linearly on the distance from the reference point. This is called **the far field approximation** and **Fraunhofer diffraction** **is the simple diffraction theory used to study problems in this approximation.** Many elements of the approximation may be achieved by using highly coherent laser sources which, being highly collimated, are sources of plane waves.

Condition 1 may also be achieved by using a lens between the source and aperture with the lens acting to collimate the rays from the point source which is placed at the focus of the lens. Condition 2 may also be achieved by placement of a lens between the aperture and screen which acts to focus light from the aperture onto the screen (or observation point).

If, on the other hand;

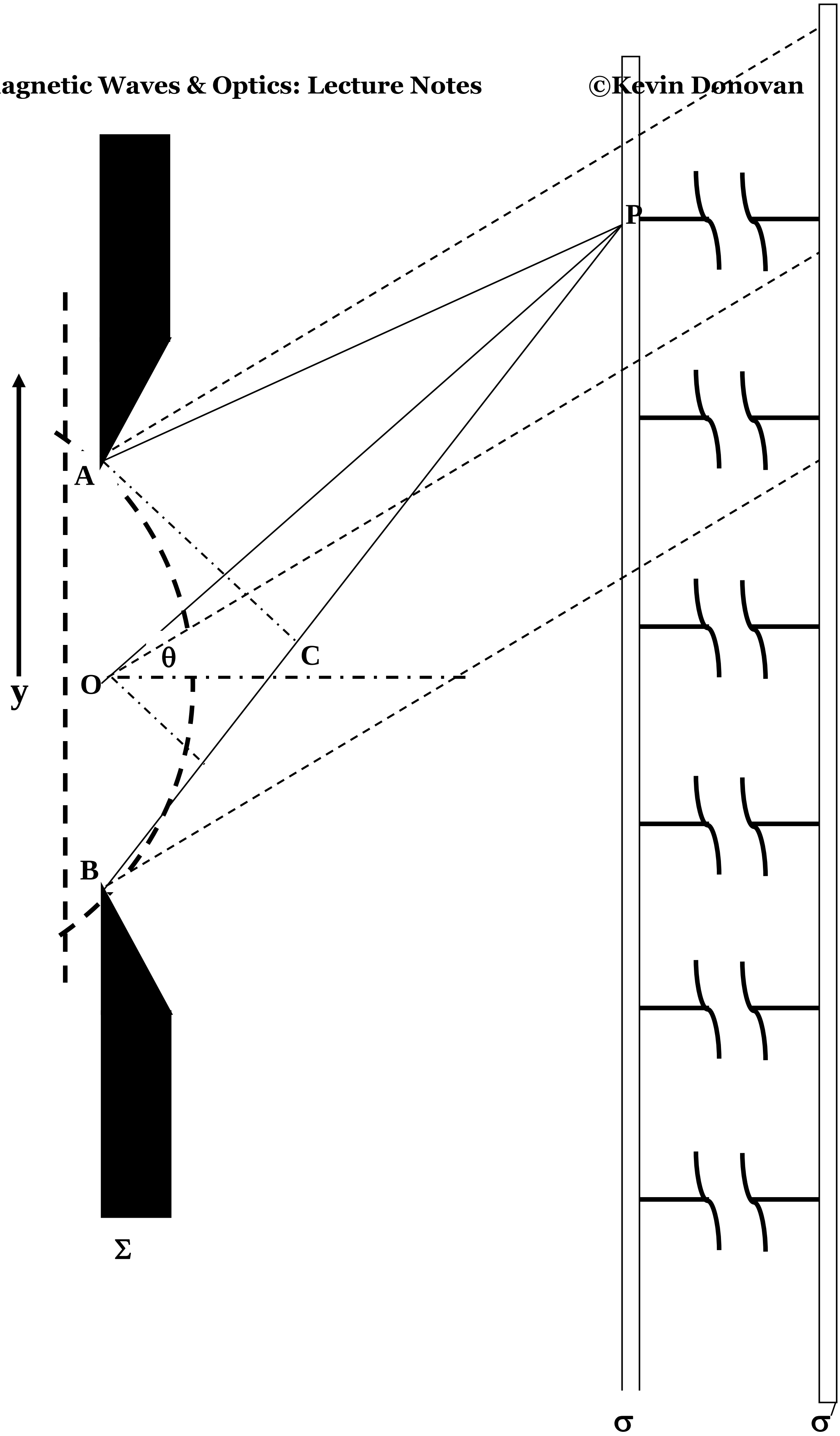
(i) The source is close to the aperture compared with its dimensions then the wavefront across the aperture will be spherical and the relative phase across the aperture compared to the phase at any particular position (eg. the centre of the aperture) will vary with position. This means that the relative phases of the secondary wavelets arriving at P will have a complex dependence on the position across the aperture from which they derive

or

(ii) The screen is close to the aperture in which case the phase difference on arrival at P depends in a non-linear way on the position from where the secondary wave emanated across the aperture

or if both conditions apply, the task of calculating the integral across the aperture of the fields arriving at P will be far more complex. **This is called the <u>near field condition</u> and <u>Fresnel diffraction</u> theory** is necessary to evaluate the fields and intensities at the screen.

These two limits are illustrated in the diagram below.



Illustrating the Fraunhoffer (parallel dashed lines and $\sigma^{/}$) , far field and Fresnel (solid lines and σ), near field situations.

The diagram shows an aperture and the spherical and plane wavefronts that may result from a source placed close to or far from the aperture respectively. If we are dealing with the plane wave the phase does not depend on y at the aperture then the phase difference at P due to the two extreme edges of the aperture is simply

$$\Delta\phi = k_0\left(BP - AP\right)$$

If the screen is far enough from the aperture and AP ≈ CP then this is simplified to

$$\Delta\phi = k_0\left(BC\right) \approx k_o b \sin\theta$$

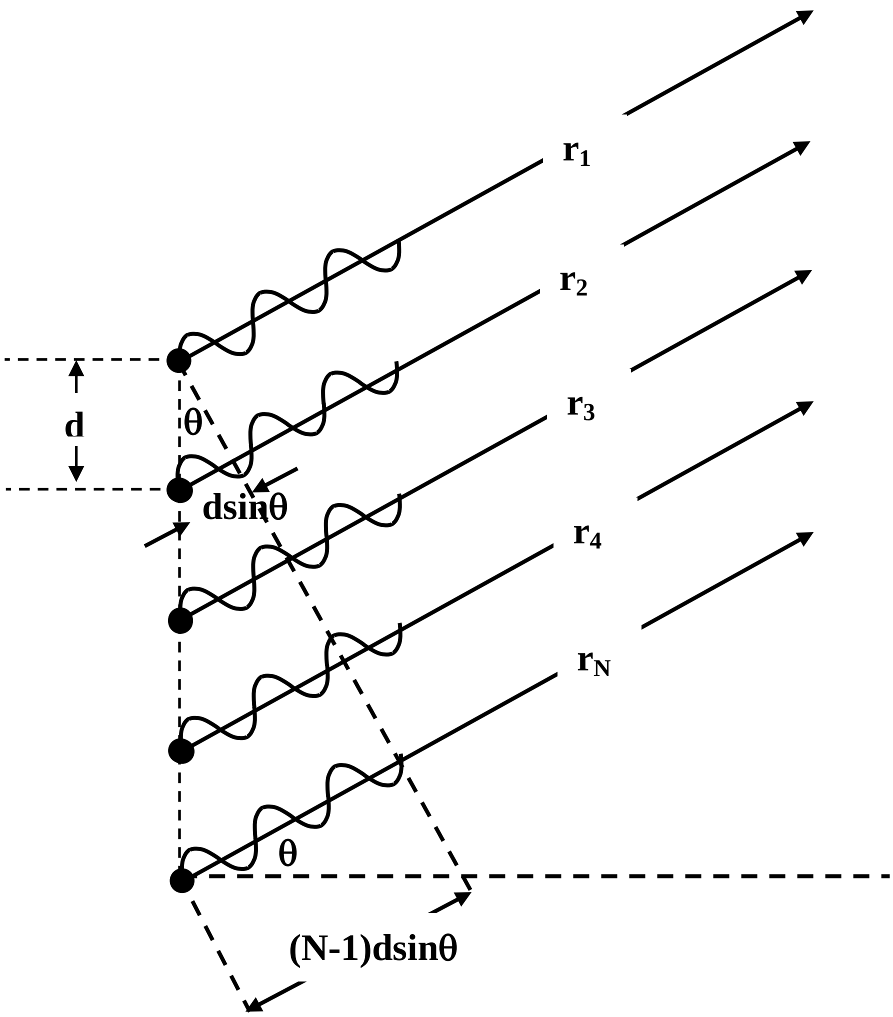
Whether we can use the far field approximation, AP ≈ CP, depends on the wavelength of the light. Clearly if AP – CP << λ_0 , then this extra distance will not add anything of consequence to the phase difference. On the other hand if AP – CP >> λ_0 there will be an important effect on the phase that will depend sinusoidally on the value of AP – CP.
If we can use the simplification then the phase difference for a wavelet emitted from a point on the aperture a distance y from O is simply

$$\Delta\phi \approx k_o y \sin\theta$$

and is linear in *y*.
Furthermore, if the screen is at a distance from the aperture such that all secondary wavelets have traveled approximately the same distance to reach a point P, then the amplitudes for the contributions from each of the secondary wavelets will be equal again greatly simplifying the problem.
This **far field or Fraunhofer approximation** is clearly far more straightforward than the near field or Fresnel diffraction that would need to be invoked were the spherical wavefront to describe the form of the phase variation at the aperture. We need to use the Fraunhoffer approximation to analyse several very important problems in diffractive optics.

a) Linear array of emitters.

To begin, a simple but useful problem is examined. We imagine a finite line of *N* individual, equally spaced oscillators, with spacing, *d*, with each one being in phase with the others and acting as a source of spherical waves.



Choosing to examine the waves traveling off at an angle θ towards some distant point the diagram shows a collection of effectively parallel rays with a path difference of $\Lambda = d\sin\theta$ between adjacent rays and a consequent phase difference of $nk_0\Lambda = k_0\Lambda$ between each. (In what follows n = 1 as the problem involves waves propagating in air) Because they each travel the same distance to the observation point the amplitudes will be equal, $E_0(r_1) = E_0(r_2) = \ldots = E_0(r_N) = E_0(r)$. To find the field at the distant point is a simple matter of summing each of the individual fields including phase.

$$E = E_0(r)\left(e^{j(k_0 r_1 - \omega t)} + e^{j(k_0 r_2 - \omega t)} + e^{j(k_0 r_3 - \omega t)} + \ldots . e^{j(k_0 r_N - \omega t)}\right) \qquad (6.1)$$

Which can be re-expressed as

$$E = E_0(r)e^{-j\omega t}e^{jk_0 r_1}\left(1 + e^{jk_0(r_2 - r_1)} + e^{jk_0(r_3 - r_1)} + ...e^{jk_0(r_N - r_1)}\right) \tag{6.2}$$

Noting that the phase difference between adjacent waves is $\Delta\phi = k_0\Lambda = k_0 d\sin\theta$

Using the diagram we note

$$\Delta\phi = k_0 d\sin\theta = k_0(r_2 - r_1) \tag{6.3}$$

The total field is then

$$E = E_0(r)e^{-j\omega t}e^{jk_0 r_1}\left(1 + e^{j\Delta\phi} + e^{2j\Delta\phi} + ...e^{(N-1)j\Delta\phi}\right) \tag{6.4}$$

$$E = E_0(r)e^{-j\omega t}e^{jk_0 r_1}\left(1 + e^{j\Delta\phi} + \left(e^{j\Delta\phi}\right)^2 + ...\left(e^{j\Delta\phi}\right)^{(N-1)}\right) = E_0(r)e^{-j\omega t}e^{jk_0 r_1}\sum_{n=0}^{N-1} e^{jn\Delta\phi} \tag{6.5}$$

The ***finite*** geometric progression on the RHS can be easily evaluated

$$E = E_0(r)e^{-j\omega t}e^{jk_0 r_1}\frac{e^{jN\Delta\phi} - 1}{e^{j\Delta\phi} - 1} \tag{6.6}$$

The quotient on the RHS can be re-arranged using a manipulation that is frequently used in problems of this type to obtain the expression in a form in which de Moivres theorem may be applied, thus

$$E = E_0(r)e^{-j\omega t}e^{jk_0 r_1}\frac{e^{jN\Delta\phi/2}\left(e^{jN\Delta\phi/2} - e^{-jN\Delta\phi/2}\right)}{e^{jk\Delta\phi/2}\left(e^{j\Delta\phi/2} - e^{-j\Delta\phi/2}\right)} \tag{6.7}$$

And finally applying De Moivre's theorem to obtain

$$E = E_0(r)e^{-j\omega t}e^{jk_0r_1}e^{j(N-1)\Delta\phi/2}\left(\frac{\sin\frac{N\Delta\phi}{2}}{\sin\frac{\Delta\phi}{2}}\right) = E_0(r)e^{-j\omega t}e^{j\left[k_0r_1+(N-1)\frac{\Delta\phi}{2}\right]}\left(\frac{\sin\frac{N\Delta\phi}{2}}{\sin\frac{\Delta\phi}{2}}\right)$$

(6.8)

Defining R as the distance of the centre of the line of oscillators to the observation point we have

$$R = \frac{1}{2}(N-1)d\sin\theta + r_1 \tag{6.9}$$

Using this along with the expression for $\Delta\phi$, $\Delta\phi = k_0 d\sin\theta = k_0(r_2 - r_1)$, we can simplify the field further

$$E = E_0(r)e^{j(k_0R-\omega t)}\left(\frac{\sin\frac{N\Delta\phi}{2}}{\sin\frac{\Delta\phi}{2}}\right) \tag{6.10}$$

Using $I = \frac{\langle E^2\rangle}{\eta}$ and the time average introducing a factor $1/2$ we may obtain the intensity in the usual way

And therefore by noting that the intensity of a single oscillator at the centre of the line of oscillators, I_0 , is given by

$$I_0 = \frac{E_0^2}{2\eta}e^{2jk_0R} \tag{6.11}$$

The total intensity at the distant point due to N in phase oscillators is then given by;

$$I = I_0\frac{\sin^2\frac{N\Delta\phi}{2}}{\sin^2\frac{\Delta\phi}{2}} = I_0\frac{\sin^2\left[N\frac{k_0d}{2}\sin\theta\right]}{\sin^2\left[\frac{k_0d}{2}\sin\theta\right]} \tag{6.12}$$

The directional dependence of *I* is contained in 6.12. As $\theta \to m\pi$, $sin\theta \to \theta$ and 6.12 approximates to

$$I = I_0 \frac{\left[N \frac{k_0 d}{2} \theta\right]^2}{\left[\frac{k_0 d}{2} \theta\right]^2} = N^2 I_0 \qquad (6.13)$$

The numerator oscillates *N* times more rapidly than the denominator as θ is gradually increased but it is the denominator that determines the positions of subsequent peaks when it goes to zero. As the numerator is modulated by the more slowly varying denominator each major peak is accompanied by a set of satellite peaks.

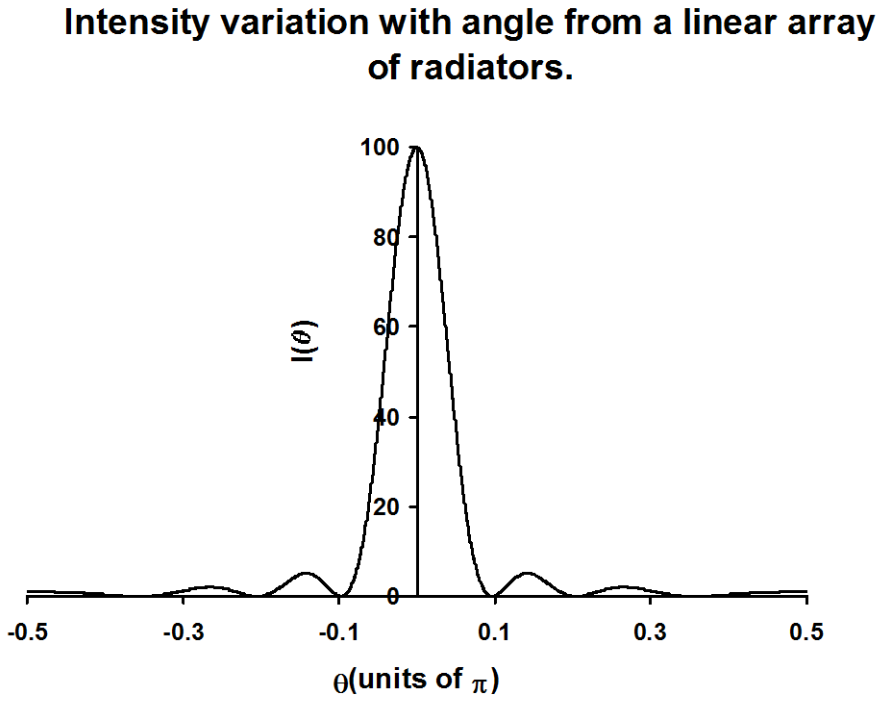
The major peaks therefore occur for angles θ_M when

$$\frac{\Delta\phi}{2} = \frac{k_0 d}{2} \sin\theta_M = m\pi \qquad (6.14a)$$

or equivalently

$$d \sin\theta_M = m\lambda_0 \qquad (6.14b)$$

First thing to note is that in the forward direction the intensity is N^2 times the intensity of an individual oscillator.



Intensity distribution from a linear array of ten radiators with separation d = λ/3.

Recalling the reversibility of propagation that is a requirement of optics, this also means that an aerial structured as a linear array of detectors will behave similarly to the linear array emitter described above and will also have highly directional receiving properties and indeed such structured aerials are commonly found acting as TV aerials at longer wavelengths.

An interesting proposition is the phased array antenna which can be engineered using these concepts, where with longer wavelengths, radio or microwaves, each of a linear array of transmitters/receivers may be engineered with the possibility of being able to control the relative intrinsic phase difference, δ, between adjacent transmitters. Now the total phase difference between adjacent emitters is

$$\Delta\phi = k_0 d \sin\theta + \delta \qquad (6.15)$$

This moderates 6.14 for the maxima to appear where

$$k_0 d \sin\theta + \delta = m\pi \qquad (6.16)$$

Or

$$\sin\theta_M = \frac{m\pi - \delta}{k_0} \tag{6.17}$$

This enables the directionality of the emission to be finely controlled by an adjustment of δ.

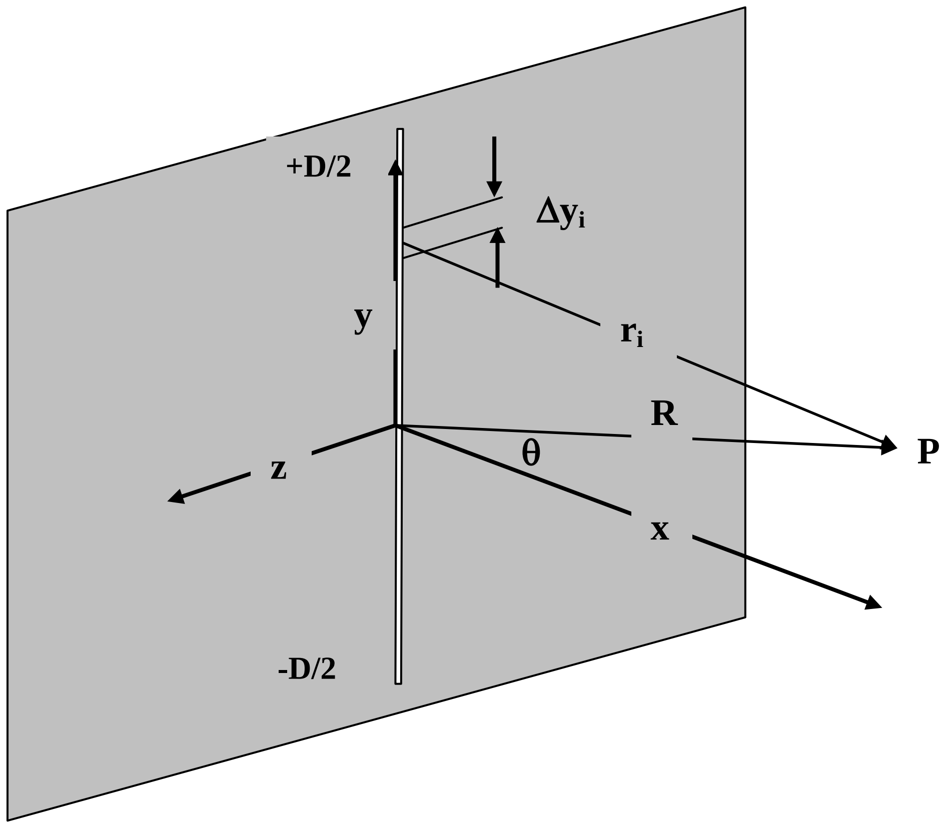
The direction that an aerial looks in can similarly be finely controlled. **Phased array antennae** are now used with two dimensional arrays of receivers or emitters to introduce this controllable directionality and such phased array antennae are widely used as receivers in astronomy or emitter/receivers in radar work.

Fraunhoffer (Far Field)Diffraction.

a) The line source

The last topic of a linear array of emitters could have just as easily been brought in for discussion under the general heading of interference. It appears here because we soon move on to describe Fraunhoffer diffraction from a single slit of finite but narrow width. To do this we need to first study **the line source where instead of a discrete array of transmitters we have a continuous line of transmitters and instead of a discrete summation an integral will be used.** Otherwise the two systems are very similar.

The line source itself is a physical fiction but it serves as an elemental emitter and building block when studying more realistic problems such as diffraction from a single slit (double slit, multiple slits).



The above diagram shows the idealized line source of length,*D*, that we now consider. It is illuminated from behind by plane waves. The width of the line is much less than the wavelength. Each point along the line emits a spherical wave whose electric field at a distance *r* from the slit is written

$$E(r) = \left(\frac{\varepsilon_0}{r}\right)\sin(\omega t - k_0 r) \tag{6.18}$$

Where a sinusoidal representation is used ***(although a cosine or imaginary exponential representation would give the same ultimate result as far as intensity is concerned)*** and **the pre-factor includes the inverse r dependence which is necessary to describe the drop in field amplitude with r as the spherical wave expands and advances.**

It should be noted that ε_0 is not an electric field as the unit of the pre-factor, $\frac{\varepsilon_0}{r}$ needs to be that of an electric field. It is called the source strength. There are an enormous number, *N*, of oscillator sources along the line. The line is divided into finite number, *M*, of extremely small segments of length Δy whose separation is vanishingly small and each of which contains $\left(\Delta y / D\right)N$ sources. At a distant point P the ith segment at a distance r_i from P contributes an amount

$$E_i = \left(\frac{\varepsilon_0}{r_i}\right)\left(\frac{\Delta y_i}{D} N\right) sin(\omega t - kr_i) \tag{6.19}$$

to the electric field, where ***we assume that Δy_i is so small that all of the sources within it are in phase at P*** (r_i = constant). Now we can allow *N* to tend to infinity to enable calculus to be used in evaluations. As *N* becomes large ε_0 will become infinitesimally small and we define ε_L as the source strength per unit length.

$$\varepsilon_L = \frac{1}{D}\lim_{N\to\infty}(\varepsilon_0 N) \tag{6.20}$$

The net field at P from all *M* segments is

$$E = \sum_{i=1}^{M}\frac{\varepsilon_L}{r_i}\Delta y_i \sin(\omega t - k_0 r_i) \tag{6.21}$$

And for a continuous line source we replace the summation by an integral

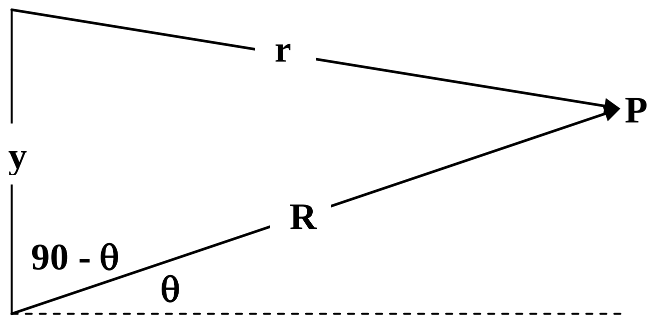
$$E = \varepsilon_L \int_{-D/2}^{+D/2} \frac{\sin(\omega t - kr)}{r} dy \qquad (6.22)$$

$r = r(y)$ and we need to find r in terms of y in order to proceed with the integral over dy. We may assert that $r \approx R$ when $R >> D$, ie we use the Fraunhoffer, far field approximation. The infinitesimal element dy contributes dE to the total field where

$$dE = \frac{\varepsilon_L}{R} \sin(\omega t - k_0 r) dy \qquad (6.33)$$

We have been careful about what to do with r inside the sinusoid as the phase is much more sensitive to r than is the amplitude. Using the same procedure that was used in the analysis of Young's slits we can find a relation between y and r as follows;

From the previous diagram and using the law of cosines, $c^2 = a^2 + b^2 - 2ab\cos C$, on the triangle



$$\left. \begin{aligned} r^2 &= R^2 + y^2 - 2Ry\cos(90-\theta) = R^2 + y^2 - 2Ry\sin\theta \\ r^2 - R^2 &= (r+R)(r-R) \approx 2R(r-R) = y^2 - 2Ry\sin\theta \\ r - R &= \frac{y^2}{2R} - y\sin\theta \approx -y\sin\theta \end{aligned} \right\}$$

$$r \approx R - y\sin\theta \qquad (6.34)$$

NB. θ is measured from the *xz* plane

We can use this in 6.33 and integrate to find the electric field at P

$$E = \frac{\varepsilon_L}{R} \int_{-D/2}^{+D/2} \sin[\omega t - k_0(R - y\sin\theta)]dy \tag{6.35}$$

Concentrating on evaluation of the integral in 6.35

$$\int \; = \int_{-D/2}^{+D/2} sin[(\omega t - k_0 R) + k_0 y\, sin\,\theta)]dy \tag{6.36}$$

To evaluate the integral we begin by using the trigonometric identity

$$\sin(u+v) = \sin u \cos v + \cos u \sin v$$

The integral is then written as

$$\int \; = sin(\omega t - k_o R)\int_{-D/2}^{+D/2} cos(k_0 y\, sin\,\theta)dy + cos(\omega t - k_0 R)\int_{-D/2}^{+D/2} sin(k_0 y\, sin\,\theta)dy$$

(6.37)

The second integral on the RHS must vanish as it is an odd function and we are left with

$$\int \; = sin(\omega t - k_o R)\int_{-D/2}^{+D/2} cos(k_0 y\, sin\,\theta)dy = sin(\omega t - k_0 R)\left[\frac{1}{k_0\, sin\,\theta} sin(k_0 y\, sin\,\theta)\right]_{-D/2}^{+D/2}$$

(6.38)

$$\int \; = sin(\omega t - k_0 R)\frac{1}{k_0\, sin\,\theta}\left[sin\left(k_0 \frac{D}{2} sin\,\theta\right) - sin\left(-k_0 \frac{D}{2} sin\,\theta\right)\right] \tag{6.39}$$

And finally

$$\int \; = sin(\omega t - k_0 R)\frac{2}{k_0\, sin\,\theta} sin\left(k_0 \frac{D}{2} sin\,\theta\right) \tag{6.40}$$

Using this integral in 6.35 we have for the electric field at P

$$E = \frac{\varepsilon_L D}{R}\frac{1}{k_0 \frac{D}{2}\sin\theta}\sin\left(k_0 \frac{D}{2}\sin\theta\right)\sin(\omega t - k_0 R) \tag{6.41}$$

6.41 has been phrased in a way suggestive of making the substitution

$$\beta = k_0 \frac{D}{2}\sin\theta \tag{6.42}$$

Leading to the compact form

$$E = \sin(\omega t - k_0 R)\frac{\varepsilon_L D}{R}\frac{\sin\beta}{\beta} \tag{6.43}$$

It is the intensity that is measured and we find this in the normal way by taking the time average of the electric field squared and dividing by impedance, η

$$I(\theta) = \frac{1}{2\eta}\left(\frac{\varepsilon_L D}{R}\right)^2\left(\frac{\sin\beta}{\beta}\right)^2 \tag{6.44}$$

$$I_0 = \frac{1}{2\eta}\left(\frac{\varepsilon_L D}{R}\right)^2 \tag{6.45}$$

and

$$I(\theta) = I_0\left(\frac{\sin\beta}{\beta}\right)^2 = I_0 \mathrm{sinc}^2\beta \tag{6.46}$$

For $\beta \to 0$ (θ = 0) , $\dfrac{\sin\beta}{\beta} \to 1$ and therefore $I(\theta = 0) = I_0$.

i) If the line length, D, is of any great length compared to λ then β, $\left(= \pi\dfrac{D}{\lambda}\sin\theta\right)$, becomes very large compared with sinβ which is no greater than 1 and therefore the intensity drops off very rapidly for any angle away from the straight through angle. In these circumstances, **with *D* >> λ , *the emission of the line source is symmetric about the y axis and it behaves as a single point emitter radiating in the forward direction with circular wavefronts*** (as opposed to spherical) propagating in the xz plane.

ii) If, on the other hand λ >> *D*, then β is very small and $\frac{\sin\beta}{\beta} \approx 1$ independent of θ according to 6.46 and the line source behaves as a point source or emitter of spherical waves.

The result 6.46 for the line of emitters can be compared with the result found earlier for a linear array of discrete emitters, 6.12 .

$$I = I_0 \frac{\sin^2 \frac{N\Delta\phi}{2}}{\sin^2 \frac{\Delta\phi}{2}} = I_0 \frac{\sin^2\left[N \frac{k_0 d}{2} \sin\theta\right]}{\sin^2\left[\frac{k_0 d}{2} \sin\theta\right]} \tag{6.12}$$

If we define

$$\beta = \frac{k_0 d}{2} \sin\theta$$

Then 6.12 for the line of discrete emitters may be written

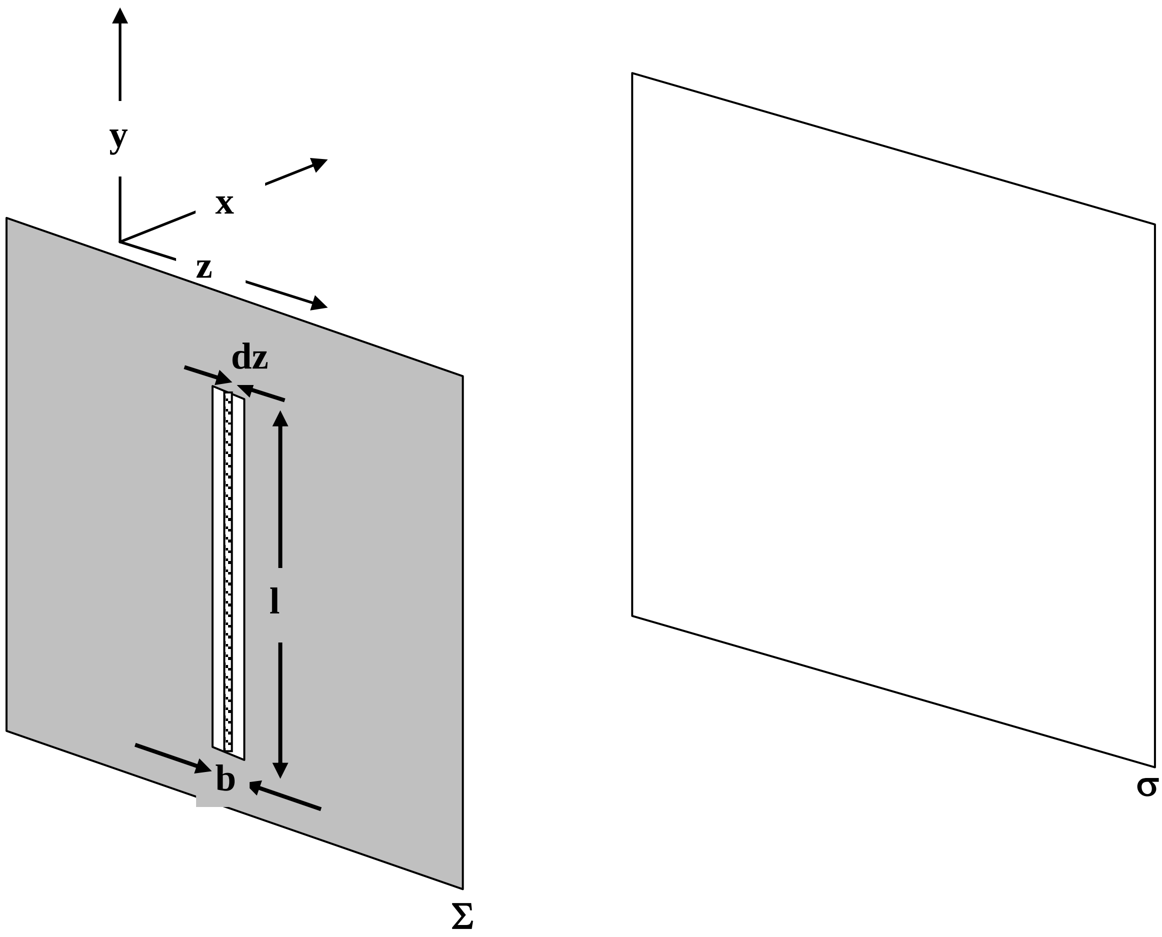
$$I = I_0 \frac{\sin^2(N\beta)}{\sin^2 \beta} \tag{6.12b}$$

We can compare this with the continuous line of emitters described by 6.46 whilst recalling that the βs in each equation are ***slightly different where one uses D, the length of the line whilst the other uses d, the separation of adjacent emitters***.

This analysis of a line source has given us the tools to study the behavior of a slit of finite width.

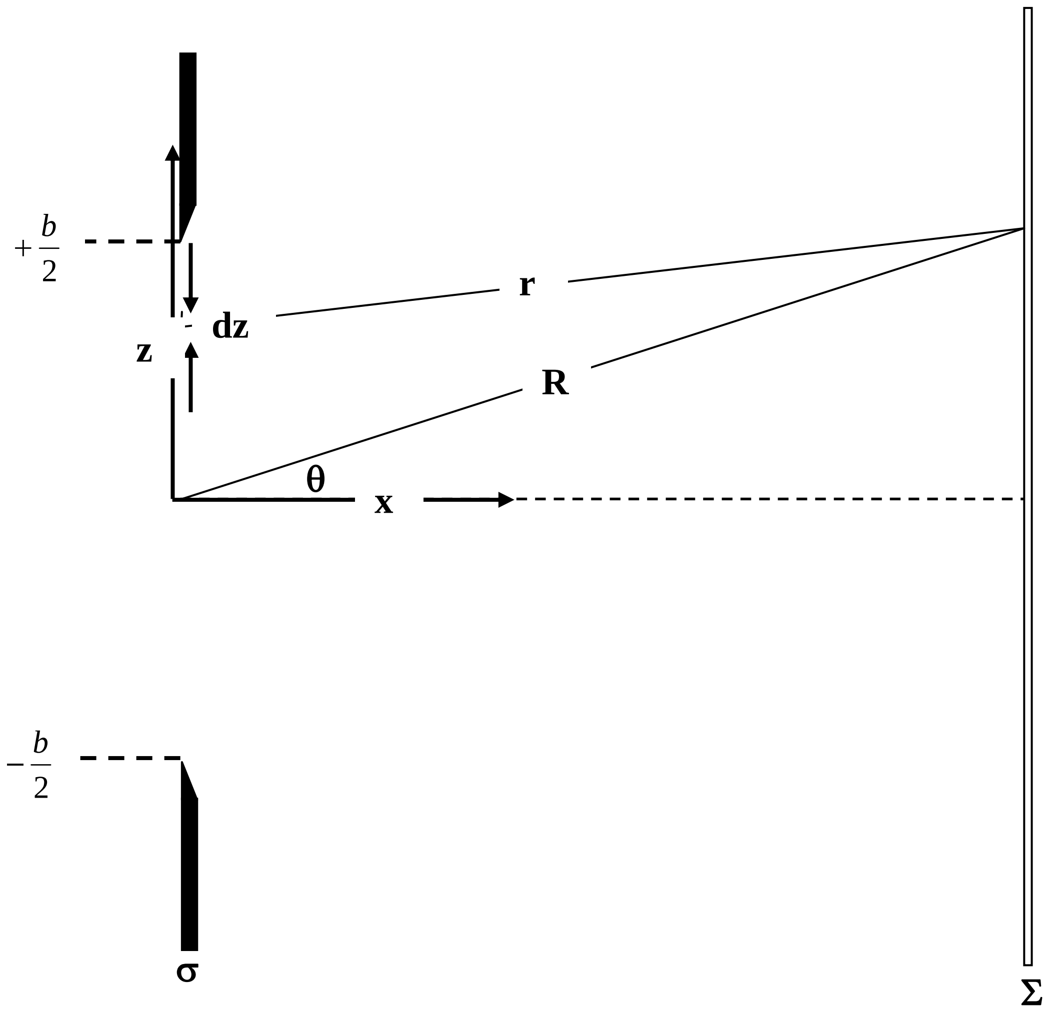
The single slit

The single slit or rectangular aperture with a finite but narrow width, b, of several hundred wavelengths and a length, l, of several centimeters is depicted in the diagram below.



The single slit of width b in the object plane Σ, is illuminated from the left by plane waves derived **from either a laser or from a point source at the focus of a collimating lens**. The radiation that propagates beyond the slit to the right falls upon a screen in the image plane, σ , **either far enough from Σ or brought to a focus at σ, such that the Fraunhofer diffraction limit applies.** To find this field distribution we divide the rectangular aperture into strips of width dz as shown and length l, **each strip acting as a coherent line source which can be replaced by a point emitter on the z**

axis that emits circular waves in the xz plane. The problem is then to find the field variation in the xz plane of an infinite number of point emitters along the z axis. Looking down on top of the slit in the xz plane, the coordinate system origin is in the centre of the slit where we measure all phases with respect to



$$r = R - z\sin\theta$$

The integral to be evaluated is then equivalent to that in 6.35

$$E = \frac{\varepsilon_L}{R}\int_{-b/2}^{+b/2}\sin[\omega t - k_0 r]dz \tag{6.47}$$

As previously we show that $r = R - z\sin\theta$

$$E = \frac{\varepsilon_L}{R} \int_{-b/2}^{+b/2} \sin[\omega t - k_0 (R - z\sin\theta)] dz \qquad (6.48)$$

NB The line of emitters is now in the* z *direction across the width of the slit and each emitter is an infinitesimal slit or line source acting as a point source emitter of circular waves.

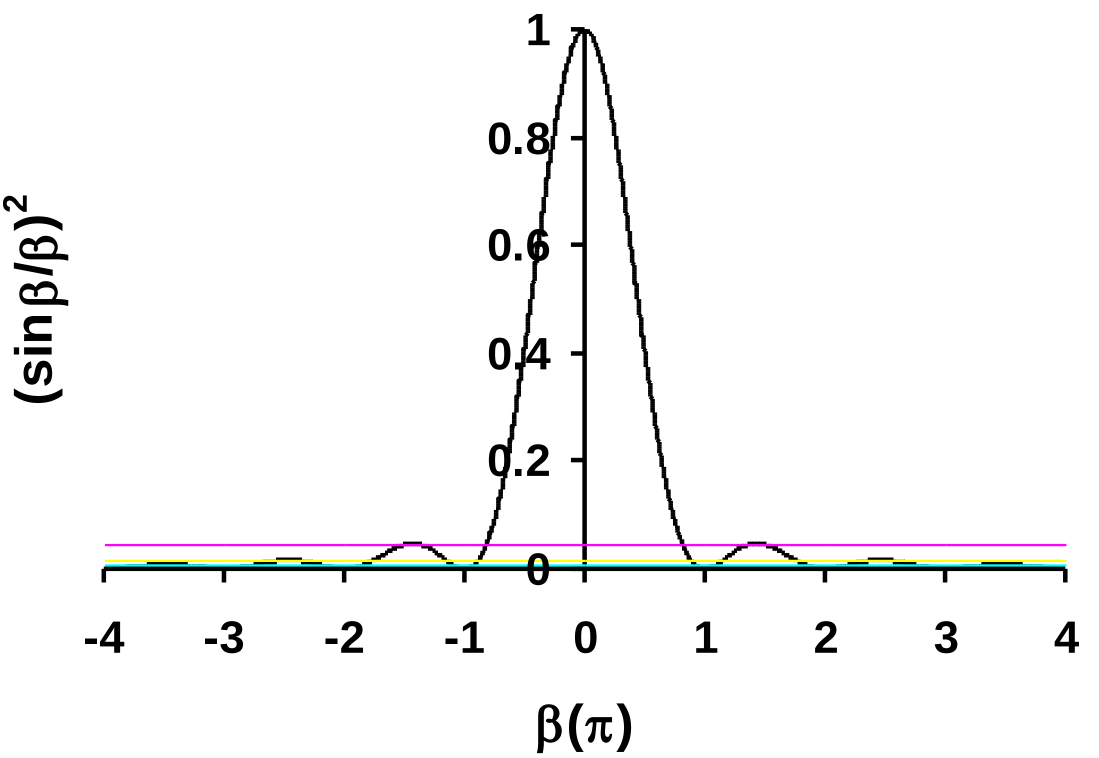
We know from the previous problem how this integral works out and the solution is again

$$I(\theta) = I(0)\left(\frac{\sin\beta}{\beta}\right)^2 \qquad (6.49)$$

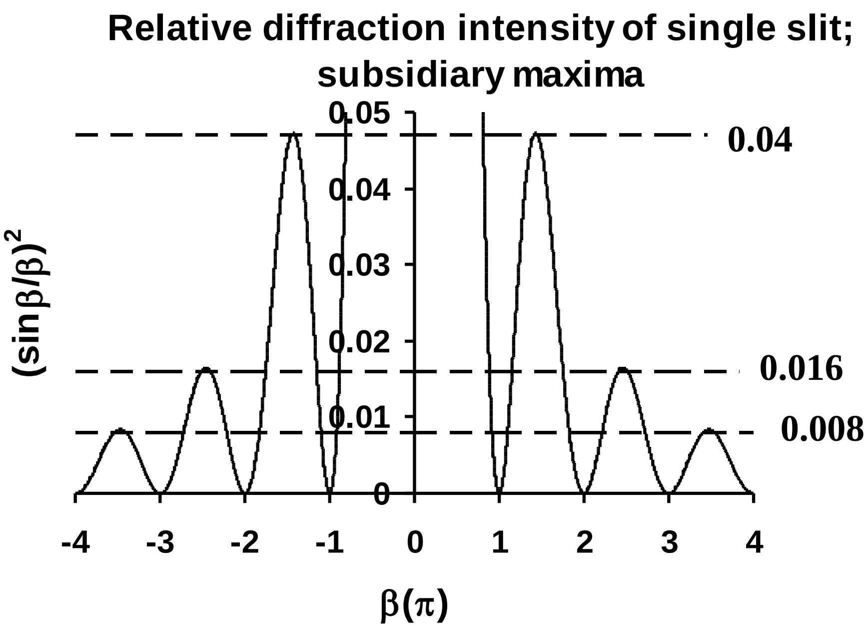
Where now

$$\beta = k_0 \frac{b}{2}\sin\theta \qquad (6.50)$$

Relative diffraction intensity of single slit



In this case the line source is short and it is b and not D that appears in the factor β and β is small enough that whilst the irradiance varies rapidly with θ, there will be higher order subsidiary maxima as emphasized in the graph below where the central maximum has amplitude 1.



We find the values of β where the extrema in $I(\theta)$ occur by differentiating 6.49 wrt β and equating the differential to zero.

Writing

$$u = \sin^2 \beta \qquad\qquad v = \frac{1}{\beta^2}$$

$$\frac{duv}{d\beta} = u\frac{dv}{d\beta} + v\frac{du}{d\beta} = \frac{2\sin\beta\cos\beta}{\beta^2} - \frac{2\sin^2\beta}{\beta^3} \tag{6.51}$$

$$\frac{dI}{d\beta} = I(0)\frac{2\sin\beta(\beta\cos\beta - \sin\beta)}{\beta^3} \tag{6.52}$$

Therefore extrema occur when

$$\sin\beta = 0 \tag{6.53a}$$

or

$$\beta\cos\beta - \sin\beta = 0 \qquad \text{ie} \qquad \beta = \tan\beta \qquad (6.53b)$$

6.53a and 6.53b are both satisfied for β = 0, this is a special case and is a maximum representing the undiffracted light. Looking at 6.47 we can see that when β = 0, the undiffracted light intensity is

$$I(\theta = 0) = I_0$$

And this represents a maximum in intensity.

MINIMA

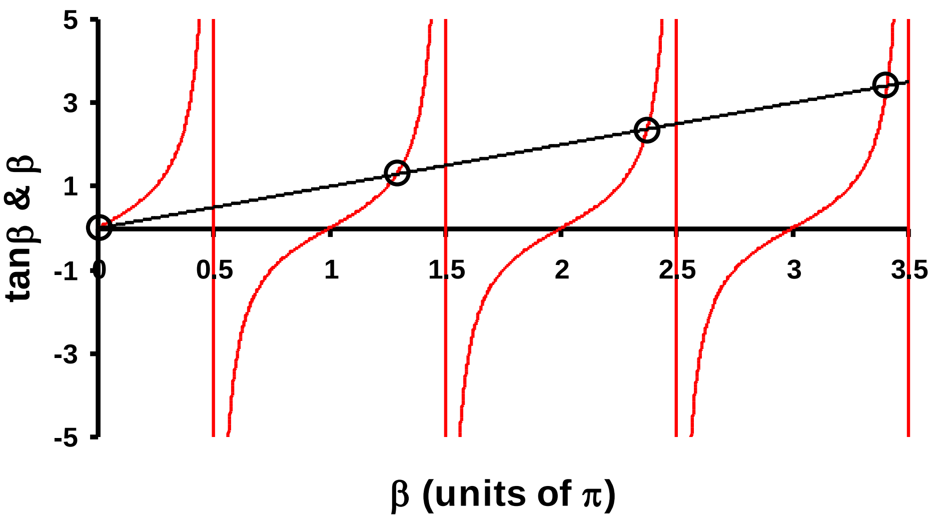
Further analysis of 6.51a in the light of 6.49 leads to the conclusion that **minima occur for** $\sin\beta = 0$ (except for the particular situation discussed above of $\beta = 0$) and in fact the intensity is zero at these minima where

$$\beta = \frac{\pi}{\lambda_0} b\sin\theta = \pm m\pi \qquad b\sin\theta = m\lambda_0 \qquad (6.54)$$

MAXIMA

6.53b is a transcendental equation with no analytic solution. It will need to be solved graphically to find the values of β where I(θ) has maxima.

Plots of β and tanβ against β



The graphical solution of 6.51b is shown in the above graph with four of the solutions circled. The crossing points are β = 1.43π, 2.46π, 3.47π ……. where the first three **subsidiary maxima will occur.** We note first that between each pair of minima where I = 0 there is one maximum. If we concentrate on one of these, the first for example, β = 1.43π. We know that the first subsidiary maximum occurs when

$$\beta_1 = \frac{\pi}{\lambda_0} b \sin\theta_1 = 1.43\pi \tag{6.55}$$

Looking at 6.55 it is clear that if we know both the slit width and the wavelength of light we can find the angle at which the first (or any other) subsidiary maximum occurs. We may deduce the following important behaviour concerning the angle at which subsidiary maxima are to be found

i) ***If we have a monochromatic source and keep the wavelength constant increasing the slit width, b, means that the angle must get smaller.***

ii) ***On the other hand, if we try to restrict the beam spatially at the aperture plane, Σ, by reducing the slit width then the diffraction pattern will be spread out further as θ will need to increase in order to keep β constant. These considerations demonstrate for us the effect of varying the slit width.***

iii) ***If now, we keep the slit width constant and increase the wavelength then the angle must also increase in order to keep β constant satisfying 6.55. This means for example, that red light will diffract to greater angles than blue light.***

If we are using a white light source then ***the first diffraction order (and the other higher orders) will contain short wavelengths at smaller angles and longer wavelengths at larger angles.*** The diffracted beam will gradually change colour from blue through green and yellow to red as we move along a screen from the central peak to larger angles before starting the sequence again as we come to the second subsidiary maximum. The central peak will contain all colours and will therefore be white light.

One of the most important questions that diffraction theory can answer if we are designing optical instruments is how the existence of apertures affect the quality of images as they are transferred through the optical system and the most important effect is that ***a finite size parallel beam will spread as it propagates due to diffraction***. We can answer the question "by how much" with the results we now have. The important quantity in this consideration is ***the angular spread of the image***. Looking at the result and ***the position of the first minima the angular spread of the central maximum is between*** $\boldsymbol{\beta = \pm \pi}$**.** For the small angles that we are dealing with in the far field (Fraunhoffer) approximation we can approximate $\sin\theta \approx \theta$

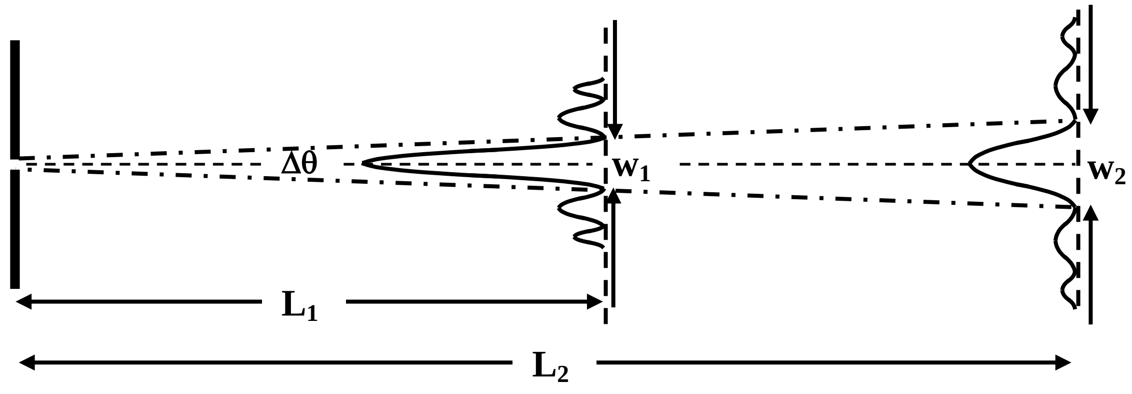
$$\beta = \frac{\pi}{\lambda_0} b \sin\theta = \frac{\pi}{\lambda_0} b\theta = \pi \tag{6.56}$$

at the first minimum and therefore the angular spread, $\Delta\theta$

$$\Delta\theta = 2\frac{\lambda_0}{b} \tag{6.57}$$

At a distance L from the aperture with $L >> b$ the width, W, of the beam is

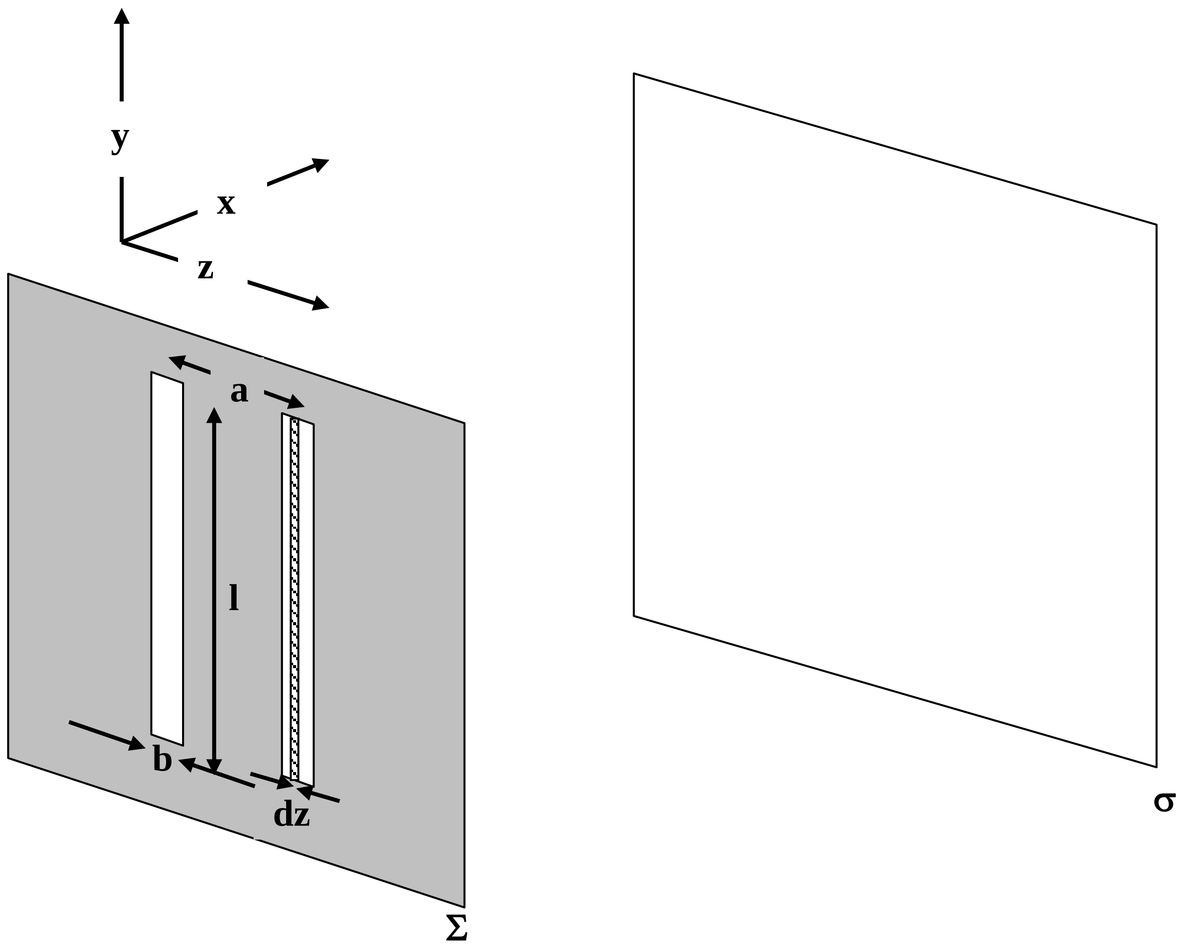
$$W = L\Delta\theta = 2\frac{\lambda_0}{b} L \tag{6.58}$$



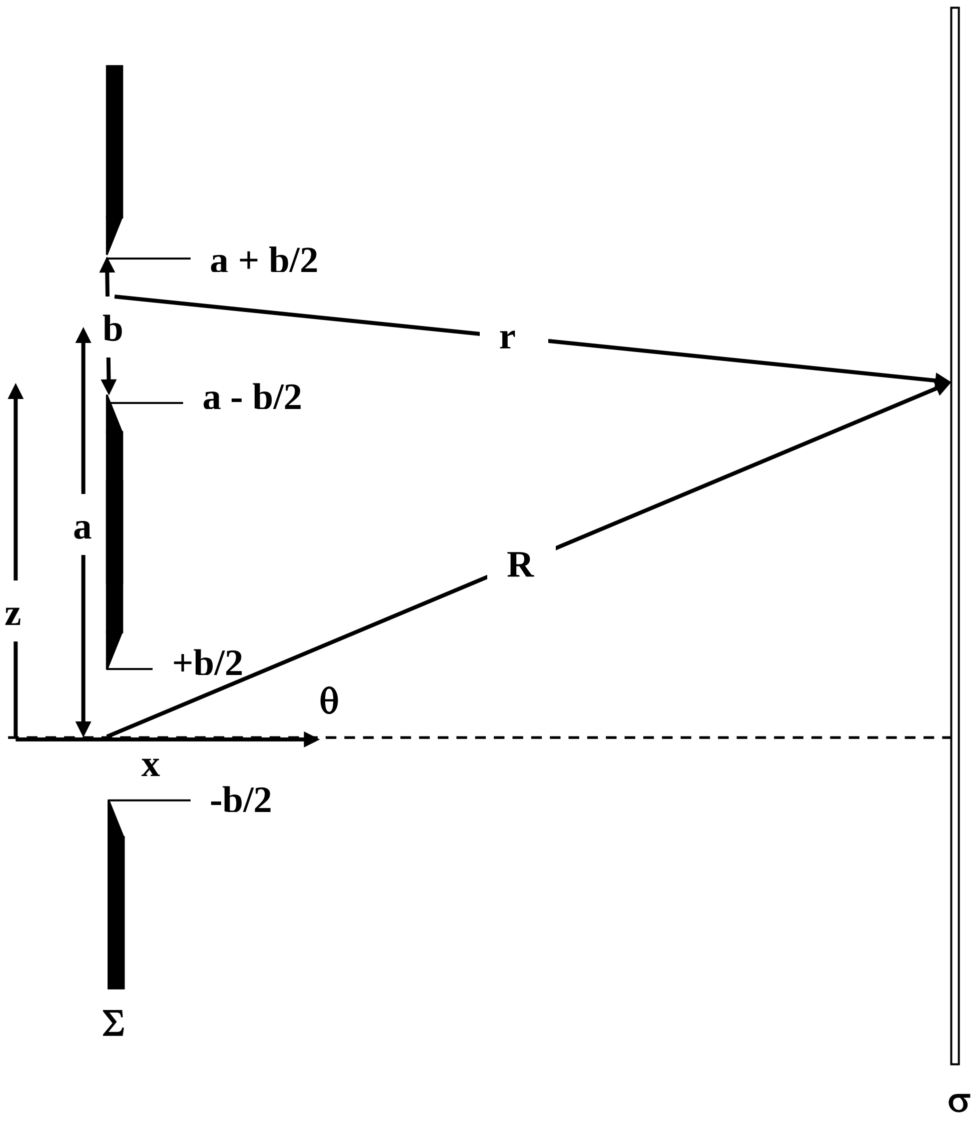
The **half angle**, $\theta_{1/2} = \frac{\Delta\theta}{2}$, is called **the beam divergence** and represents the rate at which the beam spreads.

From 2.54, $\theta = \frac{\lambda_0}{b}$, but we should remember that **this is a lower limit on divergence** and is known as **the diffraction limited beam divergence**. It is very important to note at this point that the previous analysis is independent of how the beam at Σ has a limited width *b*. Any plane wave limited to some size, *b*, by collimation with lenses or as the output from a highly collimated laser will spread according to 2.54 and 2.55. We note that **the beam spreads less the larger the limiting size, *b*. As we try to confine a beam more at some plane, Σ, the more it will spread as it propagates away from this plane.** Therefore, if for example, you wish to get a laser beam to the moon and back in order to determine it's distance, and you require that the beam spread is minimal on the way there then it is appropriate to start with a beam that is not confined at the start of the journey.

The double slit



To analyze the intensity distribution on a screen in the plane, σ, from a double slit in the aperture plane, Σ, illuminated from the left, as shown above, the apertures are again divided into infinitesimal strips of width dz and length l which act as line sources and therefore point sources. The figure shown below is a plan view allowing the geometry to be seen. The origin of the coordinate system is in the centre of the first slit



The integral performed is similar to 6.47

$$E = \frac{\varepsilon_L}{R}\left[\int_{-b/2}^{+b/2}\sin[\omega t - k_0(R - z\sin\theta)]dz + \int_{a-b/2}^{a+b/2}\sin[\omega t - k_0(R - z\sin\theta)]dz\right] \qquad (6.59)$$

We have solved the integrals already for the line source and the solution of each of the integrals in 6.59 is the same giving for the electric field

$$E = \underbrace{b\frac{\varepsilon_L}{R}\left(\frac{\sin\beta}{\beta}\right)}_{\textit{Diffraction term}}\underbrace{[\sin(\omega t - k_0 R) + \sin(\omega t - k_0 R + 2\alpha)]}_{\textit{Interference term}} \qquad (6.60)$$

Where

$$\alpha = \frac{k_0 a}{2}\sin\theta \qquad (6.61)$$

is the additional part to the second of the integrals in 6.59 due to the phase difference between the two slits.

Previously, for the line source and single slit, when it came to finding the intensity the time averaged squared electric field yielded only a factor 0.5 from a single sinusoid $\sin(\omega t - k_0 R)$. There is now a summation of two sinusoids with the time dependence and this will change the resulting intensity significantly. Using the trigonometric identity

$$\sin u + \sin v = 2\sin\left(\frac{u+v}{2}\right)\cos\left(\frac{u-v}{2}\right)$$

6.60 may be simplified

$$E = 2b\frac{\varepsilon_L}{R}\left(\frac{\sin\beta}{\beta}\right)[\sin(\omega t - k_0 R + \alpha)\cos\alpha] \qquad (6.62)$$

We can now find the intensity by squaring the electric field and taking a time average and dividing by the impedance, η

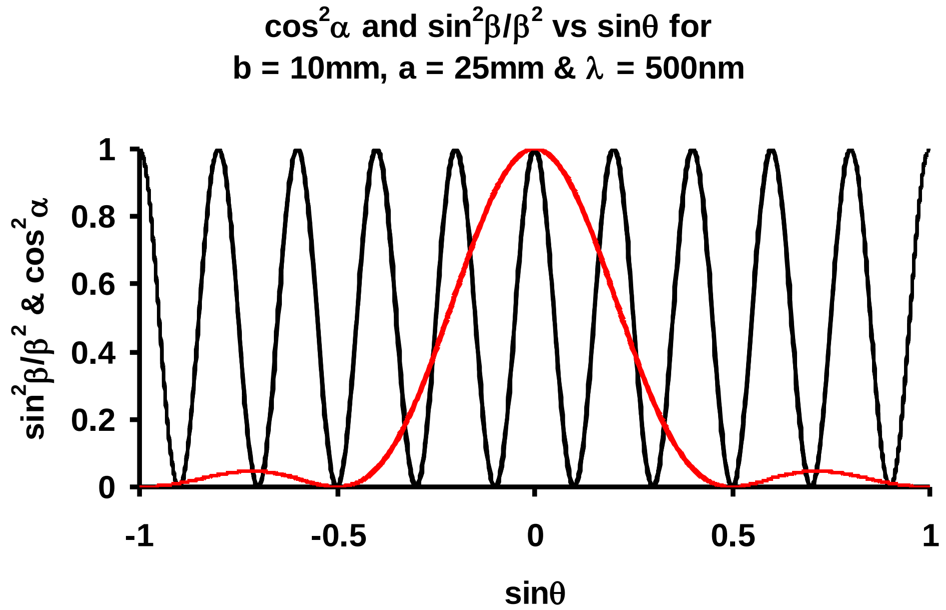
$$I(\theta) = \left(2b\frac{\varepsilon_L}{R}\right)^2 \frac{1}{2\eta}\left(\frac{\sin\beta}{\beta}\right)^2\cos^2\alpha = 4I_0\underbrace{\left(\frac{\sin\beta}{\beta}\right)^2}_{\textit{Diffraction term}}\ \&\ \underbrace{\cos^2\alpha}_{\textit{Interference term}} \qquad (6.63)$$

In the direction $\theta = 0$, $\beta = \alpha = 0$ and

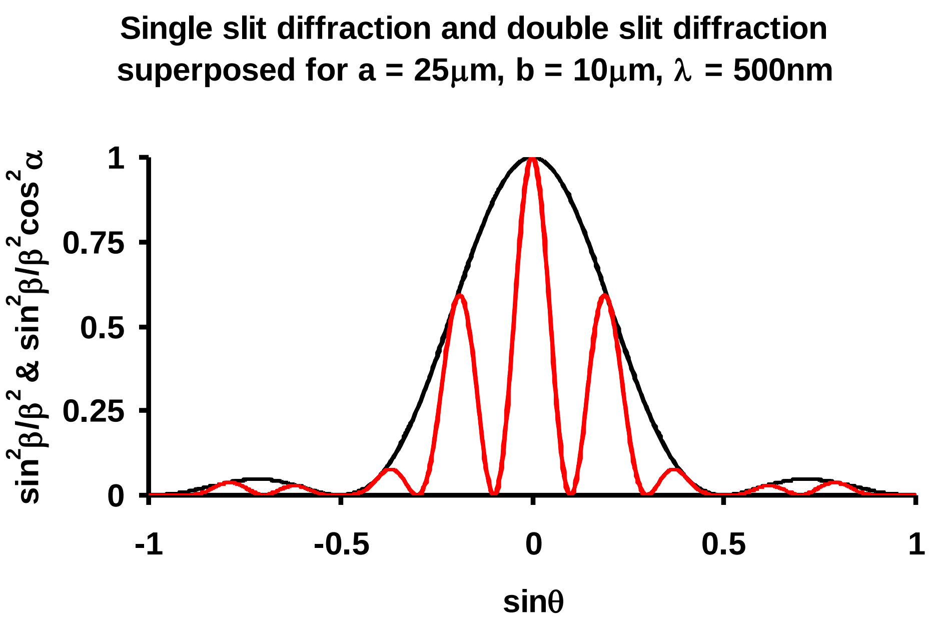
$$I(\theta = 0) = 4I_0 \qquad (6.63a)$$

where I_0 is the intensity from either one of the slits and the factor 4 is due to the fact that at the straight through point the electric field is twice what it would be with one of the slits covered and the intensity is proportional to the square of the electric field..

The intensity distribution of 6.63 is the single slit distribution, $\frac{\sin^2\beta}{\beta^2}$ modulated by the oscillatory function, $\cos^2\alpha$. **The first is a diffraction term** and **the second an interference term** from interference between the two slits. Below, are shown the two functions whose product produces the double slit diffraction distribution plotted against sinθ.



The graph below shows how the double slit intensity appears for some realistic values of a, b and λ in the mid visible.



Bear in mind that $\sin\theta = \pm 0.5$ represents light propagating at 90^o to the system axis (x direction).

If the width of the slits becomes vanishingly small then $\dfrac{\sin^2\beta}{\beta^2} \to 1$ and 6.63 tends to

$$I(\theta) = 4I_0\cos^2\alpha$$

(6.64)

Which is the expression we found for the interference problem of Young's slits, 5.113

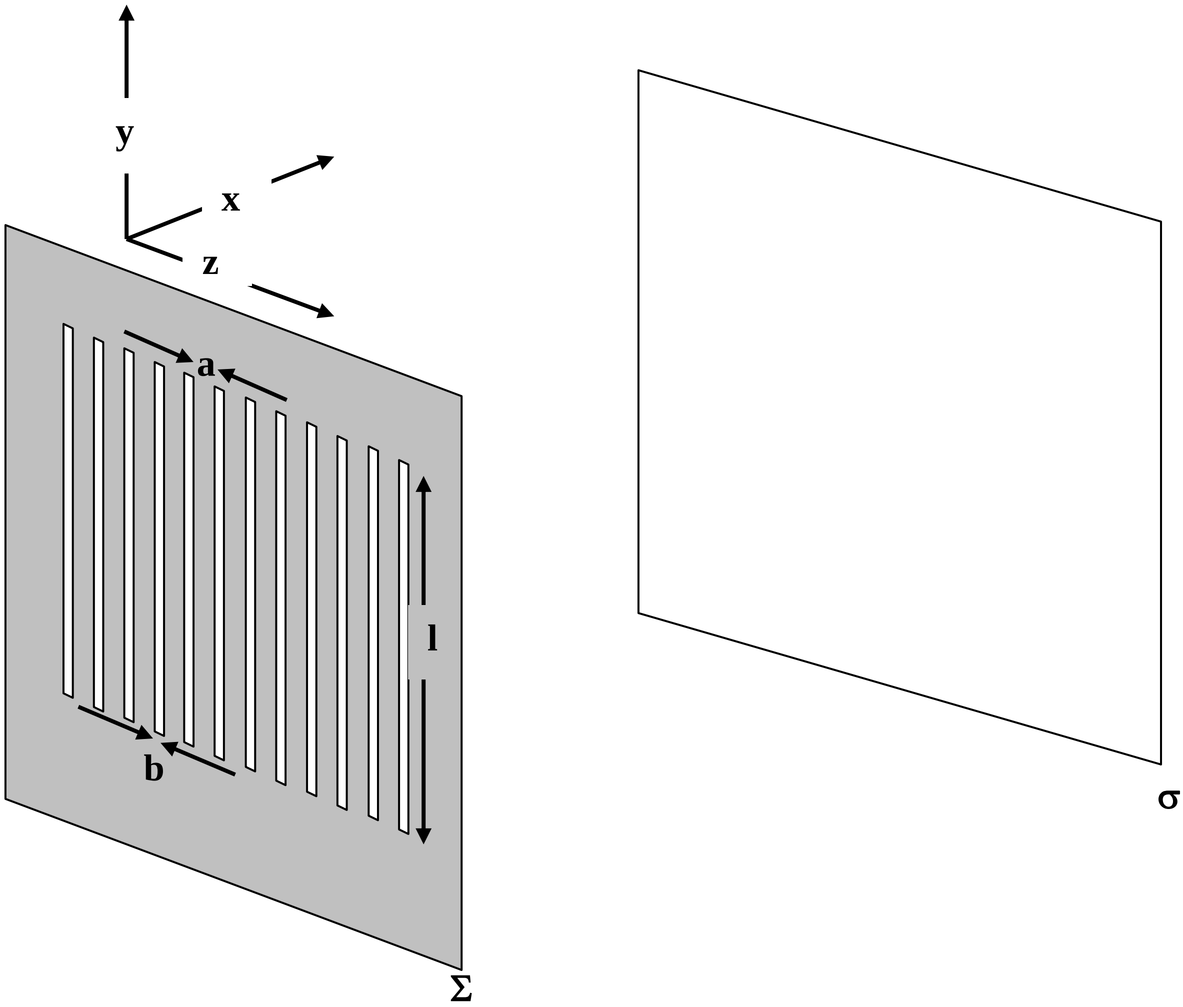
Alternately, if we reduce the separation of the slits, making a smaller then $\alpha \to 0$ and 6.63 tends to

$$I(\theta) = 4I_0\left(\frac{\sin\beta}{\beta}\right)^2 \qquad (6.65)$$

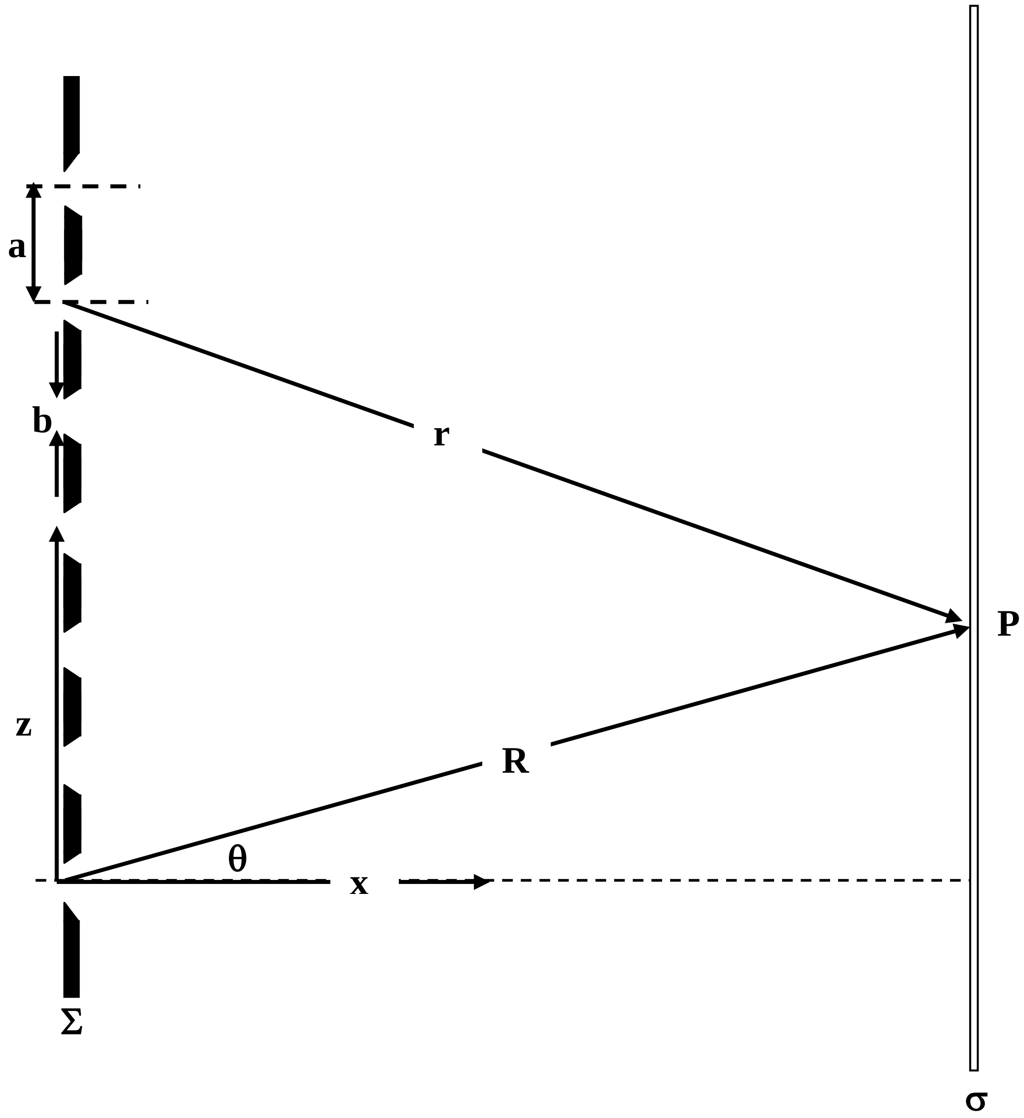
which is the diffraction from a single slit, 6.49, where the extra factor 4 appears in the intensity because the electric field is twice as large due to the two slits.

The above demonstrates that with two slits we have single slit diffraction modulated by double slit interference.

Multiple slit diffraction/interference



The diagram shows the multiple slit setup, and it is to be treated in the same way as the double slit set up with an integration along z of line sources acting in the xz plane as sources of circular waves. There are of course many more integrals to be carried out but they are all similar and the work has been done already. The slits are again of width b and length l (the length is considered much longer than λ and never features in the calculations) and separated by a distance a. The diagram below shows the plan view again and we have a similar arrangement to that seen previously.



The geometry is as shown, where the first slit is is labelled as zero, with the origin of the coordinate system in the center of the first slit. The integration across all N sources in the z direction is now

$$E = \frac{\varepsilon_L}{R}\left[\int_{-b/2}^{+b/2} F(z)dz + \int_{a-b/2}^{a+b/2} F(z)dz + \int_{2a-b/2}^{2a+b/2} F(z)dz + ... \int_{(N-1)a-b/2}^{(N-1)a+b/2} F(z)dz\right]$$

(6.66)

As before, $F(z)=\sin[\omega t-k_0 r]$ where r is the distance from the position z in the aperture plane and we use the approximation, $r=R-z\sin\theta$, as before so $F(z)=\sin[\omega t-k_0(R-z\sin\theta)]$.

The n[th] integral in 6.66 gives the n[th] contribution to the field

$$E_n=\frac{\varepsilon_L}{R}\int_{na-b/2}^{na+b/2}\sin[\omega t-k_0(R-z\sin\theta)]dz \tag{6.67}$$

Evaluating 6.67 in the same way that has been done previously

$$E_n=\frac{\varepsilon_L}{R}\frac{1}{k_0\sin\theta}\left[\sin(\omega t-k_0R)\sin(k_0z\sin\theta)-\cos(\omega t-k_0R)\cos(k_0z\sin\theta\right]_{na-b/2}^{na+b/2} \tag{6.68}$$

$$E_n=\frac{\varepsilon_L}{R}\frac{1}{k_0\sin\theta}\left[\sin(\omega t-k_0R)\left\{\sin\left(k_0\left(na+b/2\right)\sin\theta\right)-\sin\left(k_0\left(na-b/2\right)\sin\theta\right)\right\}\right]$$

$$-\frac{\varepsilon_L}{R}\frac{1}{k_0\sin\theta}\left[\cos(\omega t-k_0R)\left\{\cos\left(k_0\left(na+b/2\right)\sin\theta\right)-\cos\left(k_0\left(na-b/2\right)\sin\theta\right)\right\}\right] \tag{6.69}$$

Using the definitions of $\alpha=\frac{k_0a}{2}\sin\theta$ and of $\beta=k_0\frac{b}{2}\sin\theta$ to tidy 6.69

$$E_\mathrm{n}=\left\{\frac{\varepsilon_L}{R}\frac{b}{2\beta}\left[\sin(\omega t-k_0R)\{\sin(2n\alpha+\beta)-\sin(2n\alpha-\beta)\}\right]-\right.$$

$$-$$

$$\left.\frac{\varepsilon_L}{R}\frac{b}{2\beta}\left[\cos(\omega t-k_0R)\{\cos(2n\alpha+\beta)-\cos(2n\alpha-\beta)\}\right]\right\} \tag{6.70}$$

Using

$$\sin u - \sin v = 2\cos\left(\frac{u+v}{2}\right)\sin\left(\frac{u-v}{2}\right)$$

and

$$\cos u - \cos v = -2\sin\left(\frac{u+v}{2}\right)\sin\left(\frac{u-v}{2}\right)$$

We may tidy up further

$$E_n = \frac{\varepsilon_L}{R}\frac{b}{2\beta}\left[\sin(\omega t - k_0 R)\{2\cos(2n\alpha)\sin\beta\} - \cos(\omega t - k_0 R)\{-2\sin(2n\alpha)\sin\beta\}\right] \quad (6.71)$$

$$E_n = \frac{\varepsilon_L}{R}\frac{b}{\beta}\sin\beta\left[\sin(\omega t - k_0 R)\cos(2n\alpha) + \cos(\omega t - k_0 R)\sin(2n\alpha)\right] \quad (6.72)$$

The square bracket can be simplified using the trigonometric identity

$$\sin u \cos v = \frac{1}{2}\left[\sin(u+v) + \sin(u-v)\right]$$

$$\sin u \cos v + \cos u \sin v = \frac{1}{2}\left[2\sin(u+v) + \sin(u-v) + \sin(v-u)\right] = \sin(u+v)$$

$$E_n = \frac{\varepsilon_L}{R} b \frac{\sin\beta}{\beta}\sin(\omega t - k_0 R + 2n\alpha) \quad (6.73)$$

We need to sum the disturbances from all of the slits to find the total field at the observation point

$$E = \sum_{n=0}^{N-1} E_j = \frac{\varepsilon_L}{R} b \frac{\sin\beta}{\beta}\sum_{n=0}^{N-1}\sin(\omega t - k_0 R + 2n\alpha) \quad (6.74)$$

It is convenient to now write the sinusoid as the imaginary part of an exponential

$$E = \mathrm{Im}\left[\frac{\varepsilon_L}{R} b \frac{\sin\beta}{\beta} e^{j(\omega t - k_0 R)}\sum_{n=0}^{N-1} e^{(j2\alpha)n}\right] \quad (6.75)$$

We have evaluated this geometric series before when examining an array of emitters in equations 6.5 and 6.6 where

$$\sum_{n=0}^{N-1} e^{jn\Delta\phi} = \frac{e^{jN\Delta\phi} - 1}{e^{j\Delta\phi} - 1}$$

$$E = \frac{\varepsilon_L}{R} b \frac{\sin\beta}{\beta} e^{j(\omega t - k_0 R)} \operatorname{Im}\left[\frac{e^{j2N\alpha} - 1}{e^{j2\alpha} - 1}\right] \tag{6.76}$$

We use the same method as previously, re-arranging the quotient on the RHS of 6.76

$$E = \frac{\varepsilon_L}{R} b \frac{\sin\beta}{\beta} \operatorname{Im}\left\{e^{j(\omega t - k_0 R)}\left[\frac{e^{jN\alpha}\left(e^{jN\alpha} - e^{-jN\alpha}\right)}{e^{j\alpha}\left(e^{j\alpha} - e^{-j\alpha}\right)}\right]\right\} \tag{6.77}$$

And using De Moivre

$$E = \frac{\varepsilon_L}{R} b \frac{\sin\beta}{\beta} \operatorname{Im}\left\{e^{j(\omega t - k_0 R + (N-1)\alpha)}\left[\frac{\sin(N\alpha)}{\sin\alpha}\right]\right\} \tag{6.78}$$

The sin of the exponential represents the imaginary part and so

$$\boxed{E = \frac{\varepsilon_L}{R} b \frac{\sin\beta}{\beta} \sin(\omega t - k_0 R + (N-1)\alpha) \frac{\sin N\alpha}{\sin\alpha}} \tag{6.79}$$

Finally we get to the intensity (flux) distribution by the usual route

$$\boxed{I(\theta) = I_0 \left(\frac{\sin^2\beta}{\beta^2}\right)\left(\frac{\sin^2 N\alpha}{\sin^2\alpha}\right)} \tag{6.80}$$

I_0 is the intensity emitted in the forward direction by any one of the slits and at $\theta = 0$ (where $\alpha = \beta = 0$) the intensity of the system is

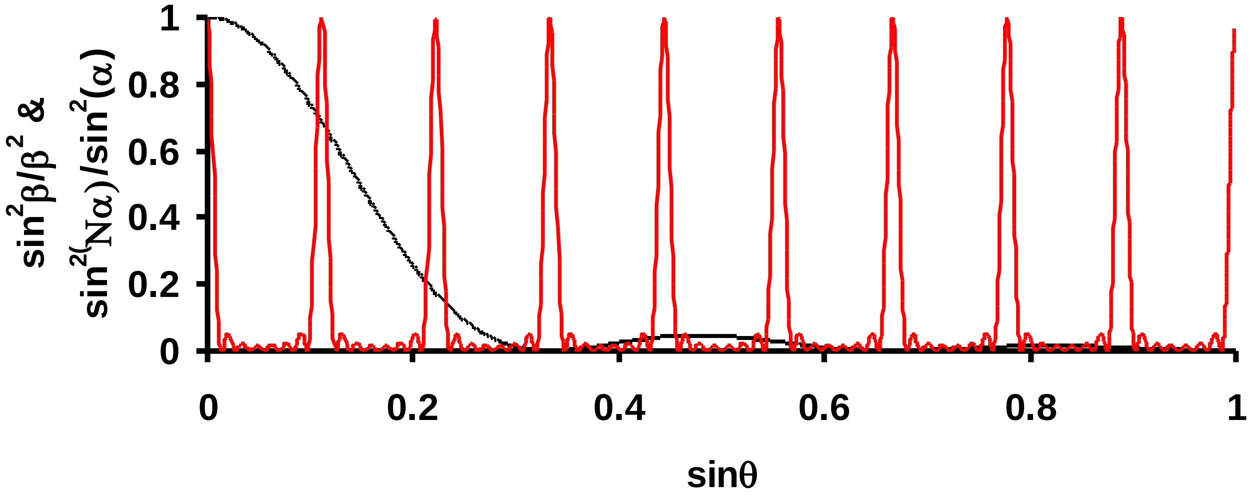
$$I(\theta = 0) = N^2 I_o \tag{6.81}$$

This is in agreement with our result 6.63a for two slits with N = 2.

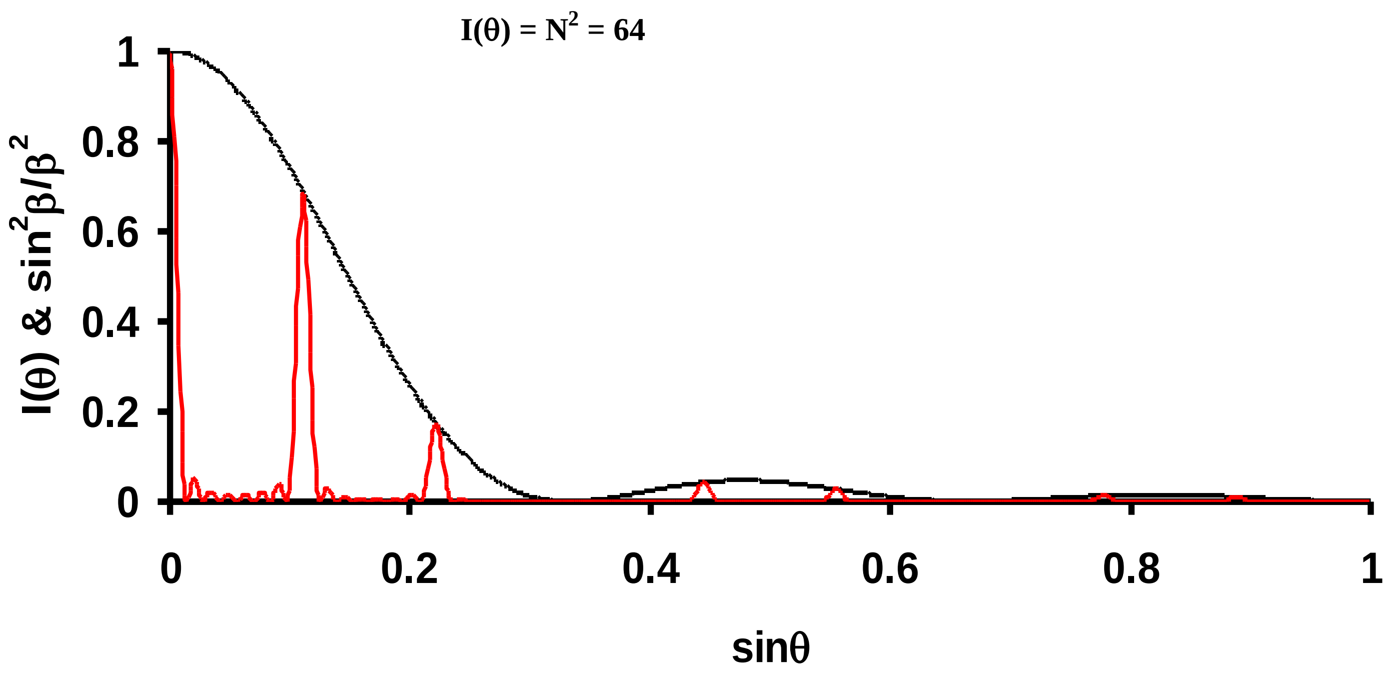
If the width of the slit was shrunk to approach zero and $\frac{\sin^2 \beta}{\beta^2} \to 1$ then equation **6.80 would become the equation previously obtained for a linear array of N emitters 6.12b.**

There is again a diffraction term, $\frac{\sin^2 \beta}{\beta^2}$ **, modulated by an interference term,** $\frac{\sin^2 N\alpha}{\sin^2 \alpha}$ **as indicated by the graph shown below.**

Multiple slit interference and the single slit diffraction envelope, a = 3b , b = 3λ, N = 8



Multiple slit diffraction and the single slit envelope, a = 3b , b = 3λ, N = 8



i) The straight through intensity or zeroth order.

We need to consider the angular dependence of the intensity distribution for multiple slits very carefully;

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

The straight through intensity is straightforward as we have the fact that at θ = 0, (α = 0, β = 0)

The diffraction term is then;

$$\frac{\sin^2\beta}{\beta^2} = 1.$$

The interference term

$$\frac{\sin^2 N\alpha}{\sin^2\alpha} \approx \frac{N^2\alpha^2}{\alpha^2} = N^2$$

as θ ⇒ 0 and here we get the direct transmission of 6.81.

$$I(\theta = 0) = N^2 I_o$$

ii) Principle maxima,

The diffraction term needs careful attention as $\alpha \to m\pi$. The denominator of the quotient goes to zero and we could naively expect to have the principle maxima where this condition is satisfied. Ie.;

$$\alpha = \frac{k_0 a}{2}\sin\theta = m\pi$$

However N in the numerator of the interference term is also an integer and the numerator will therefore also be an integer multiple of π that will go to zero at the same time. There is therefore a competition between the numerator and the denominator which both go to zero as $\alpha \to m\pi$!

To understand what happens when , $m \neq 0$ (m = 0 is the straight through case previously considered) we need to use l'Hopitals rule concerning what happens to a quotient of two functions of x as they both tend to zero at the same value of x.

l'Hopitals rule states that

$$\lim_{x \to a} \frac{f(x)}{g(x)} = \lim_{x \to a} \frac{\frac{d}{dx} f(x)}{\frac{d}{dx} g(x)}$$

Therefore applying this rule

$$\lim_{\alpha \to m\pi} \frac{\sin N\alpha}{\sin \alpha} = \lim_{\alpha \to m\pi} \frac{N \cos N\alpha}{\cos \alpha} = \pm N$$

Ie. both cosines are equal to 1.

Therefore the **principle maxima** which occur where the denominator is zero at $\alpha = m\pi$, will have, according to l'Hopitals rule, the value

$$I_0 \frac{\sin^2 \beta}{\beta^2} \left(\frac{\sin N\alpha}{\sin \alpha} \right)^2 = NI_0 \frac{\sin^2 \beta}{\beta^2} \qquad (6.82a)$$

for

$$\alpha = 0, \pm\pi, \pm 2\pi ...... \pm m\pi \qquad (6.82b)$$

Since $\alpha = \frac{k_0 a}{2} \sin\theta = \frac{\pi a}{\lambda_0} \sin\theta$, the condition can be expressed equivalently as a condition on λ

$$a \sin\theta = m\lambda_0 \qquad (6.82c)$$

iii) Minima

Zeroes or minima exist whenever the numerator of the interference term, $\frac{\sin^2 N\alpha}{\sin^2 \alpha}$, is zero **whilst the denominator is not.** So generally the condition for a zero or minimum is

$$\frac{\sin N\alpha}{\sin \alpha} = 0 \qquad (6.83a)$$

or

$$N\alpha = \pm m\pi \qquad (6.83b)$$

But if $\frac{m}{N}$ is an integer, then the denominator of the interference term will tend to zero as well then

$$\alpha = \pm\frac{m}{N}\pi = \pm q\pi$$

Where q is also an integer and the value of the interference term will not be zero.
NB. We have discussed the m = 0 solution where θ = 0 and in this case the denominator would be zero as well and as seen previously this corresponds to a maximum.
Generally we have zeroes where

$$\alpha = \pm\left(m+\frac{1}{N}\right)\pi, \pm\left(m+\frac{1}{N}\right)2\pi, \ldots\ldots \pm\left(m+\frac{1}{N}\right)(N-1)\pi, \pm\left(m+\frac{1}{N}\right)(N+1)\pi, \ldots. \tag{6.83c}$$

NB the $\pm\left(m+\frac{1}{N}\right)N$ term in the 6.83c sequence is missing as it equals mN + 1 which is an integer and therefore the denominator goes to zero. Otherwise the numerator becomes $\sin\left(N\left(m+\frac{1}{N}\right)\pi\right) = \sin((mN+1)\pi) = 0$. In other words there are (N-1) minima between consecutive principle maxima.

The criterion for a zero may also be written in terms of λ

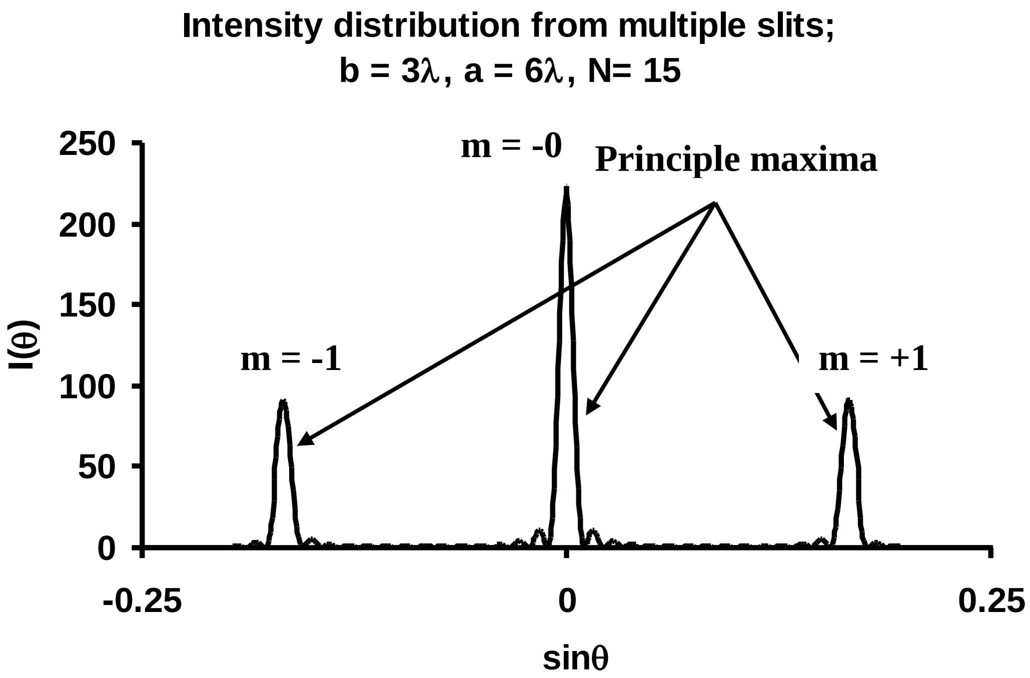
$$a\sin\theta = \pm\left(m+\frac{1}{N}\right)\lambda_0 \tag{6.83d}$$

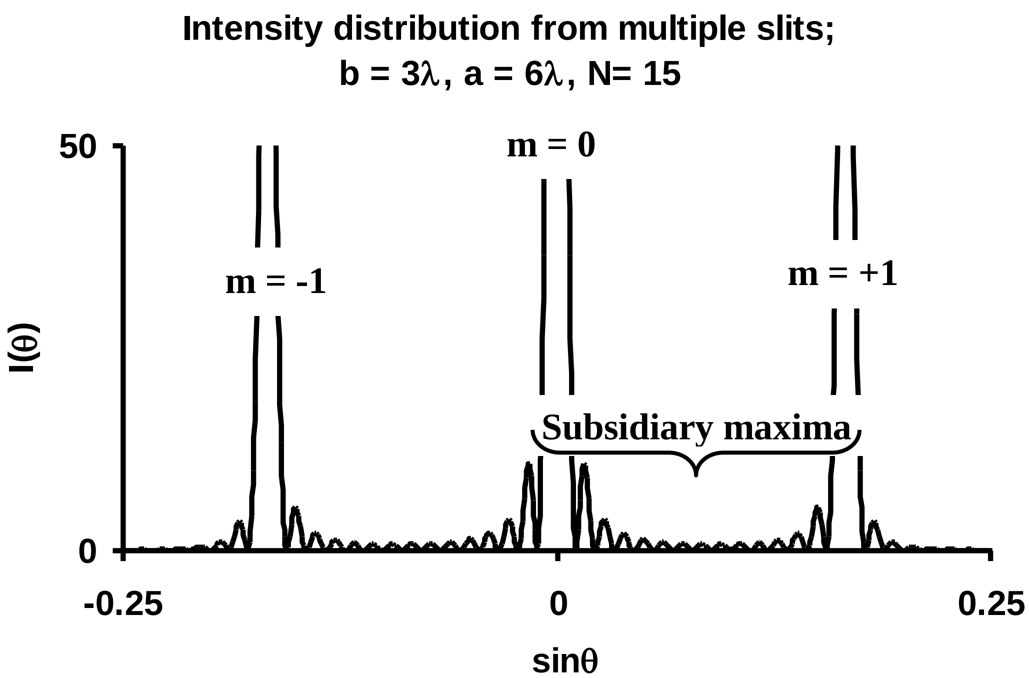
And between each of these minima there will be **<u>subsidiary maxima</u>** located ***approximately*** at the points where $\sin N\alpha$ has its greatest values.

$$\alpha = \pm\frac{3\pi}{2N}, \pm\frac{5\pi}{2N}\ldots.$$

There are (N-2) subsidiary maxima between each principle maximum related to the existence of (N-1) minima between principle maxima.

The graphs below show these properties with the second graph emphasizing the subsidiary maxima. In that example N = 15 and there are 13 subsidiary maxima.



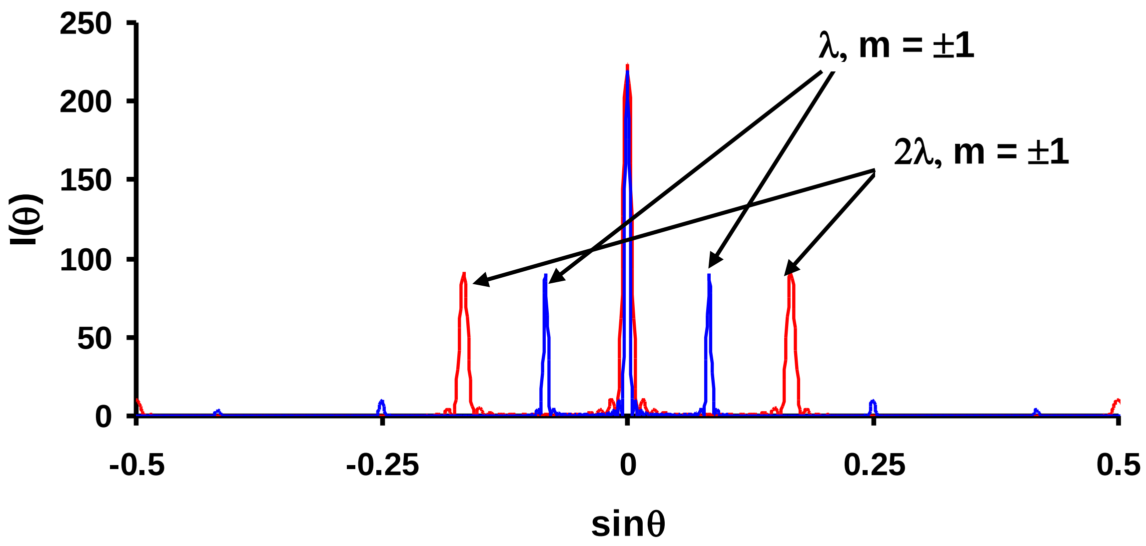


The upper graph shows the m = 0 and ±1 principle maxima whilst the lower graph shows the many much smaller subsidiary maxima for this set of slit parameters.

b) Diffraction grating.

The multiple slit aperture offers a method for obtaining wavelength separation as different wavelengths will have $m = \pm 1$ etc. principle maxima diffracted at different angles as expressed by 6.82c, $a \sin\theta = m\lambda_0$, where for constant order m if λ is larger then θ must be bigger to maintain the equality. ie. blue wavelengths will be diffracted to smaller angles than red wavelengths in any given order m. In fact as we can see from the graphs the intensity in higher orders diminishes rapidly

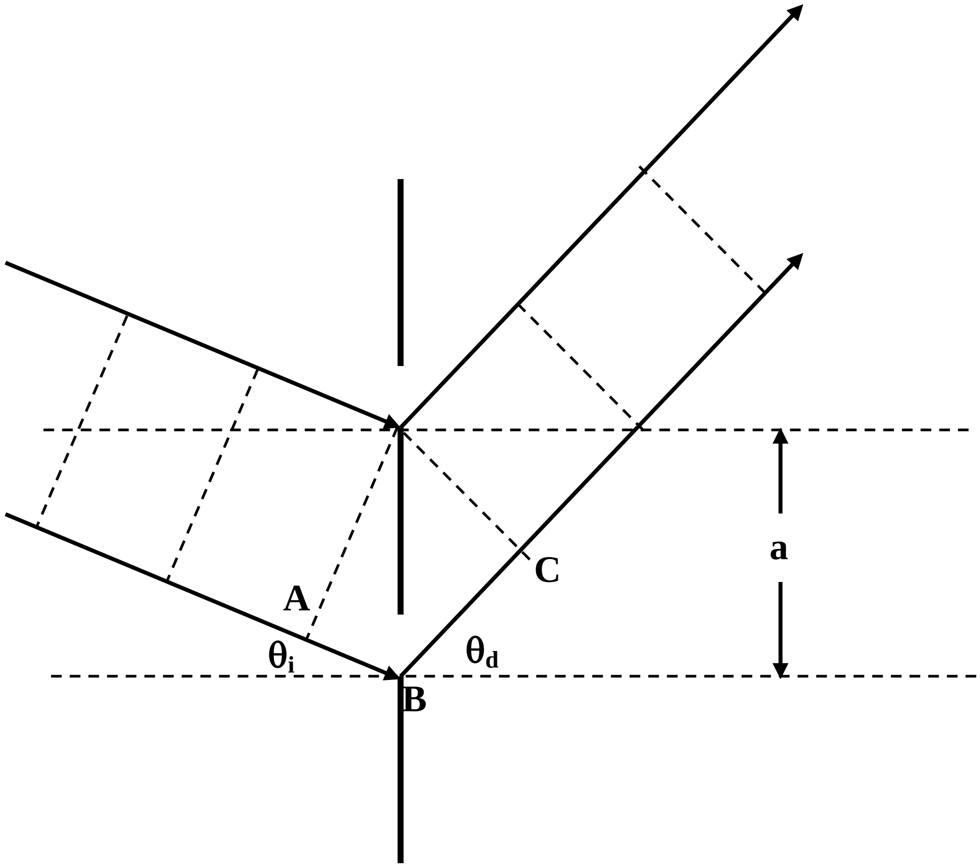
λ dependence of intensity distribution;



In the preceding analysis of multiple slits that lead to 6.80 the intensity distribution was obtained as well as the direction at which maxima were observed in 6.82c. When studying diffraction gratings we are just going to be concerned with the angular directions in which the principle maxima occur as described by 6.82c

$$a \sin\theta = m\lambda_0 \qquad (6.82c)$$

This can be obtained by a more simple analysis where the relative intensities of the principle maxima are ignored and we just seek the condition for effective diffraction with the aid of geometry.



The above diagram shows two slits from the multiple slit previously discussed. It shows incoming waves (solid arrows) and plane wavefronts (dashed lines) incident at an angle θ_i and outgoing waves and plane wavefronts. Diffracted at an angle θ_d. To find the directions in which there is strong diffraction we need the optical path difference, Λ, to be some integer multiple, m, of the wavelength, λ_0.

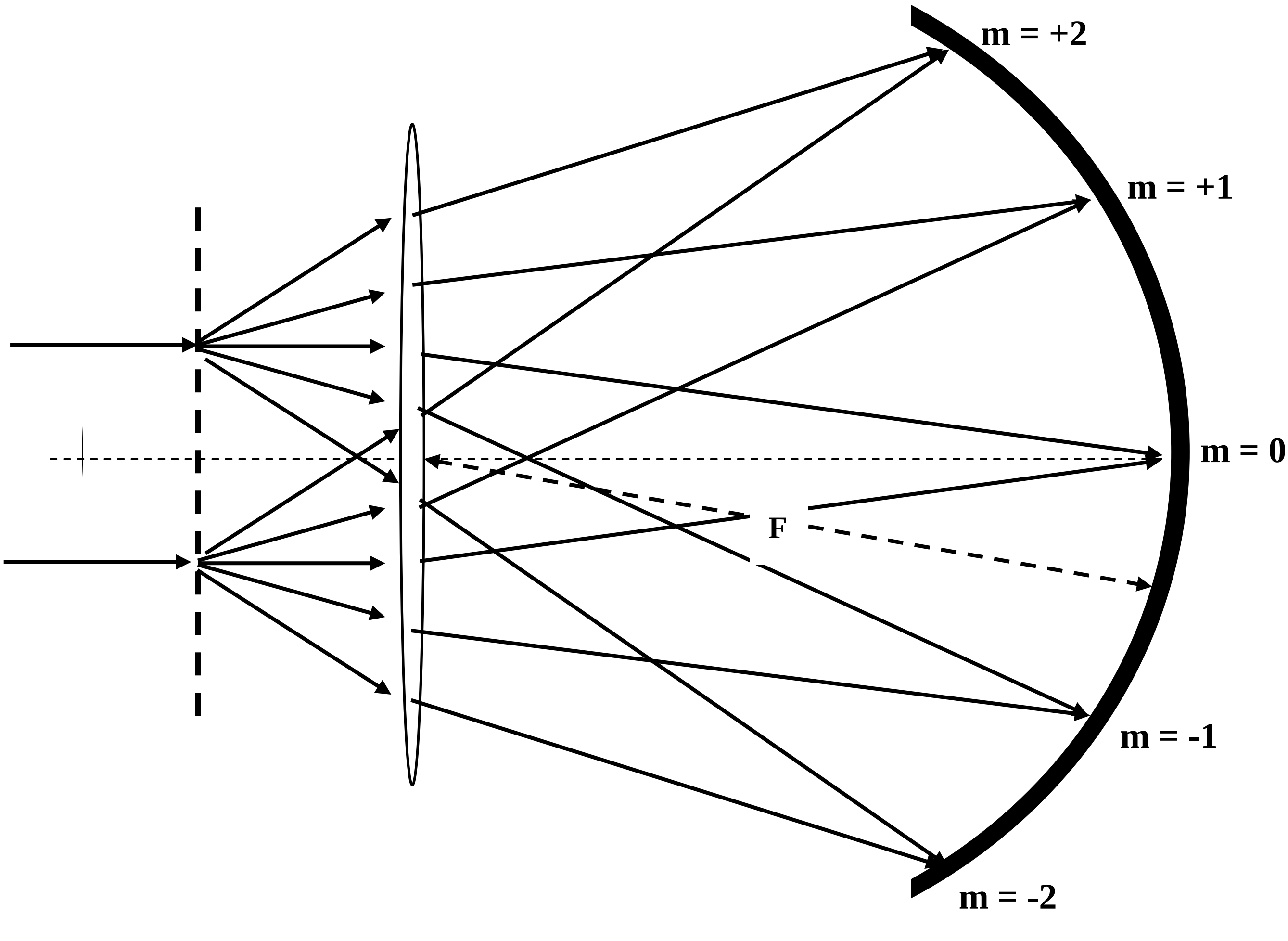
$$\Lambda = AB + BC = a(\sin\theta_i + \sin\theta_d) = \pm m\lambda_0 \tag{6.84}$$

NB we are again assuming propagation in air with n = 1 as we have throughout this section of diffraction and the previous on interference. This is why the wavelength and wavevector are being written as λ_0 and k_0 respectively with the subscript 0 to remind that this is in free space.

In the previous analysis, it was assumed that the incident plane wave was perpendicular to the slits and that therefore $\theta_i = 0$. In this case 6.84 becomes

$$a \sin \theta_d = \pm m\lambda_0 \qquad (6.84a)$$

which is 6.82c again as required.



The above diagram shows how the different diffraction orders may be displayed using a lens to bring the parallel rays of each order from each slit to a focus at a screen placed at the focus of the lens. Note that the screen is curved in order to maintain the distance F.

As we have seen the longer (red) wavelengths will be diffracted at greater angles than the shorter (blue) wavelengths. This means that the longer wavelengths of the m^{th} order will overlap with the shorter wavelengths of the $(m+1)^{th}$ order. ie if there are two wavelengths where the condition

$$(m+1)\lambda_{m+1} = m\lambda_m \tag{6.85}$$

applies, they will be diffracted at the same angle and similarly to the Fabry Perot studied previously it is necessary to define a free spectral range when deciding on how useful a diffraction grating will be in operation. Ie. if we are using the diffraction grating to separate wavelengths before performing spectral analysis it is undesirable to have the red light of one order to overlap with the blue light of the higher order.

The free spectral range for the order m is then defined in the same way as for the Fabry Perot as

$$\lambda_{FSR} = \lambda_m - \lambda_{m+1} = \lambda_m - \frac{m}{m+1}\lambda_m = \frac{\lambda_m}{m+1} \tag{6.86}$$

The free spectral range is the wavelength range over which there is no spectral overlap between orders and it depends on the order we are considering, furthermore it decreases with increasing order. This is the same result as found for the Fabry Perot.

The angular dispersion of the grating is another important property describing how well the different wavelengths within a particular order are separated

$$D_{Ang} = \frac{d\theta_m}{d\lambda} \tag{6.87}$$

$$m\lambda = a\sin\theta_m \tag{6.82b}$$

$$D_{Ang} = \frac{m}{a\cos\theta_m} \tag{6.88}$$

If a lens is used to focus the different wavelengths onto a photographic sheet (or other detector) in it's focal plane as in the diagram above, then we can define the linear dispersion

$$D_{Lin} = \frac{dy}{d\lambda} \tag{6.89}$$

Now, we have $dy = Fd\theta$ and so

$$D_{Lin} = F\frac{d\theta}{d\lambda} = FD_{Ang} = F\frac{m}{a\cos\theta_m} \tag{6.90}$$

The resolution of a grating is a measure of the sharpness of the peaks in any particular order. We are now going to need our previous analysis which gave the intensity distribution of a multiple slit aperture.

If $\Delta\lambda_{\min}$ is the minimum wavelength interval that is just resolvable by Rayleighs criterion then the resolution, $\mathcal{R}$, is

$$\mathcal{R} = \frac{\lambda}{\Delta\lambda_{\min}} \tag{6.91}$$

Rayleighs criterion states that two peaks at different wavelengths, λ and $\lambda+\Delta\lambda$, are resolvable if the peak of one coincides (same angle) with the first minimum of the other. We found previously that the principle maxima (peaks) occur at

$$a\sin\theta = \pm m\lambda_0 .... \tag{6.82b}$$

And the minima at

$$a\sin\theta = \pm\left(m + \frac{1}{N}\right)\lambda_0 \tag{6.83d}$$

Therefore Rayleighs criterion is that

$$\left(m + \frac{1}{N}\right)\frac{\lambda}{a} = \frac{m(\lambda + \Delta\lambda)}{a} \tag{6.92a}$$

$$\left(m+\frac{1}{N}\right)\lambda = m(\lambda + \Delta\lambda) \qquad (6.92b)$$

Re arranging to find $\mathcal{R} = \dfrac{\lambda}{\Delta\lambda}$

$$\lambda + \Delta\lambda = \left(1+\frac{1}{mN}\right)\lambda \qquad (6.92c)$$

$$\Delta\lambda = \left(1+\frac{1}{mN}-1\right)\lambda \qquad (6.93)$$

$$\frac{\Delta\lambda}{\lambda} = \frac{1}{mN} \qquad (6.94)$$

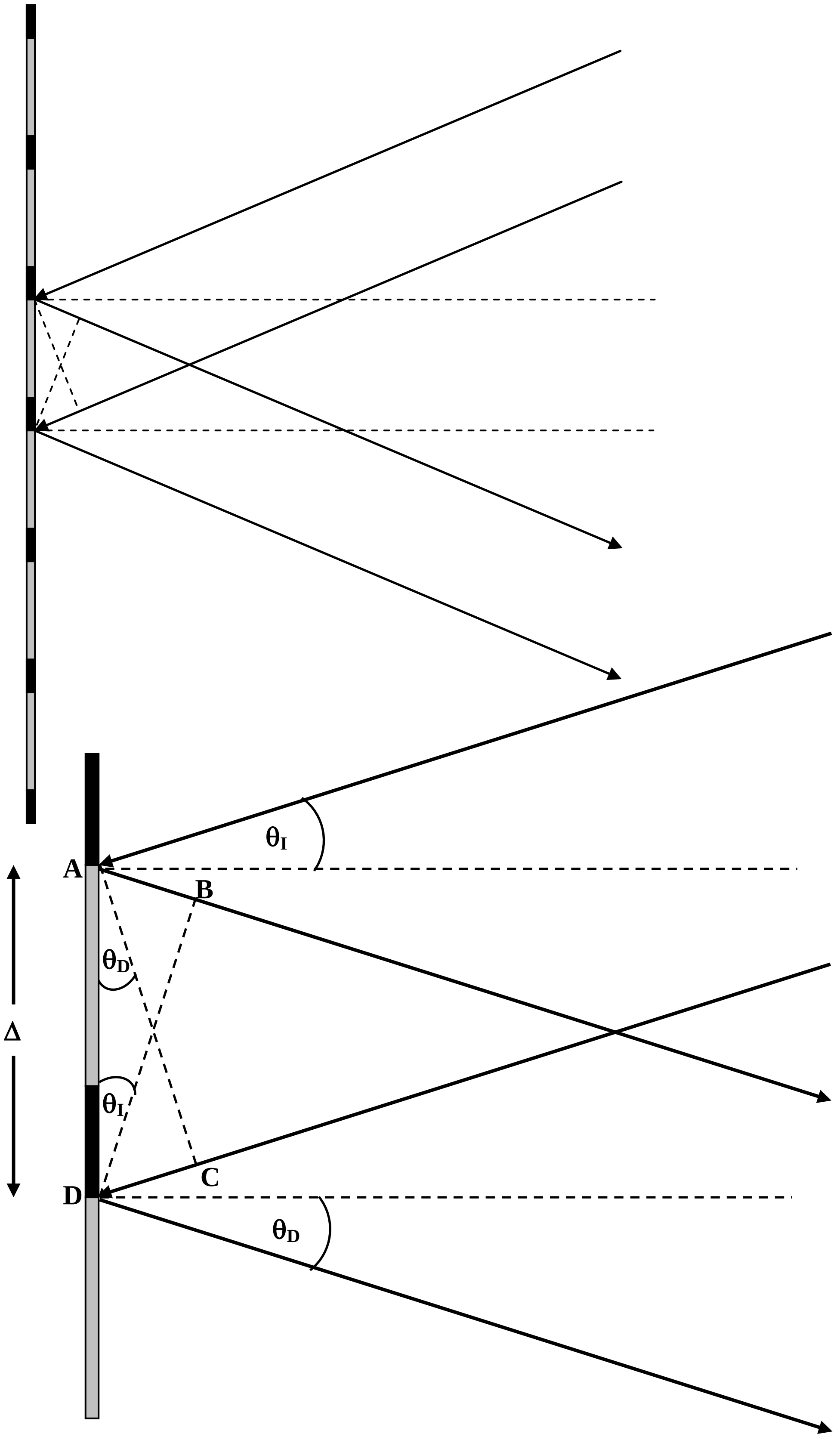
The resolution of the grating is then given by

$$\mathcal{R} = \frac{\lambda}{\Delta\lambda} = mN \qquad (6.95)$$

Higher orders have higher resolution although the intensity available in higher orders is much reduced. Increasing the number of lines will also increase the resolution. Interestingly, neither the slit width or separation have any effect on the resolution other than indirectly inasmuch as decreasing a or b will increase N in the area illuminated if this area remains fixed.

We have been examining the **transmission grating** in the preceding discussion. It is also possible to make a grating to act as a dispersing element which operates in a reflecting mode.

The reflection grating operates under the same principles as the transmission grating with the difference being that the grating consists of lines of high reflectivity separated by lines of low reflectivity. Such a grating with a repeat distance, Δ , and its operation is shown in the diagram below.


θI
A
B
θD
Δ
θI
C
D
θD

A plane wave with a wavefront AC is incident on the reflection grating at an angle of incidence θ_I and is reflected at a diffraction angle θ_D. For strong diffraction the diffracted wavefront BD must also be a plane wave. This requirement is that the optical path difference of the two rays shown is equal to an integral number of wavelengths.

$$\Lambda = AB - CD = m\lambda$$

The laws of reflection require that the angle of incidence is equal to the angle of reflection and this allows us to construct the triangles ABD and ACD

$$\Lambda = \Delta \sin\theta_I - \Delta \sin\theta_D = m\lambda \tag{6.96}$$

A convention requires that when a diffracted ray is on the opposite side of the normal to the surface (grating) as the incident ray it is negative.

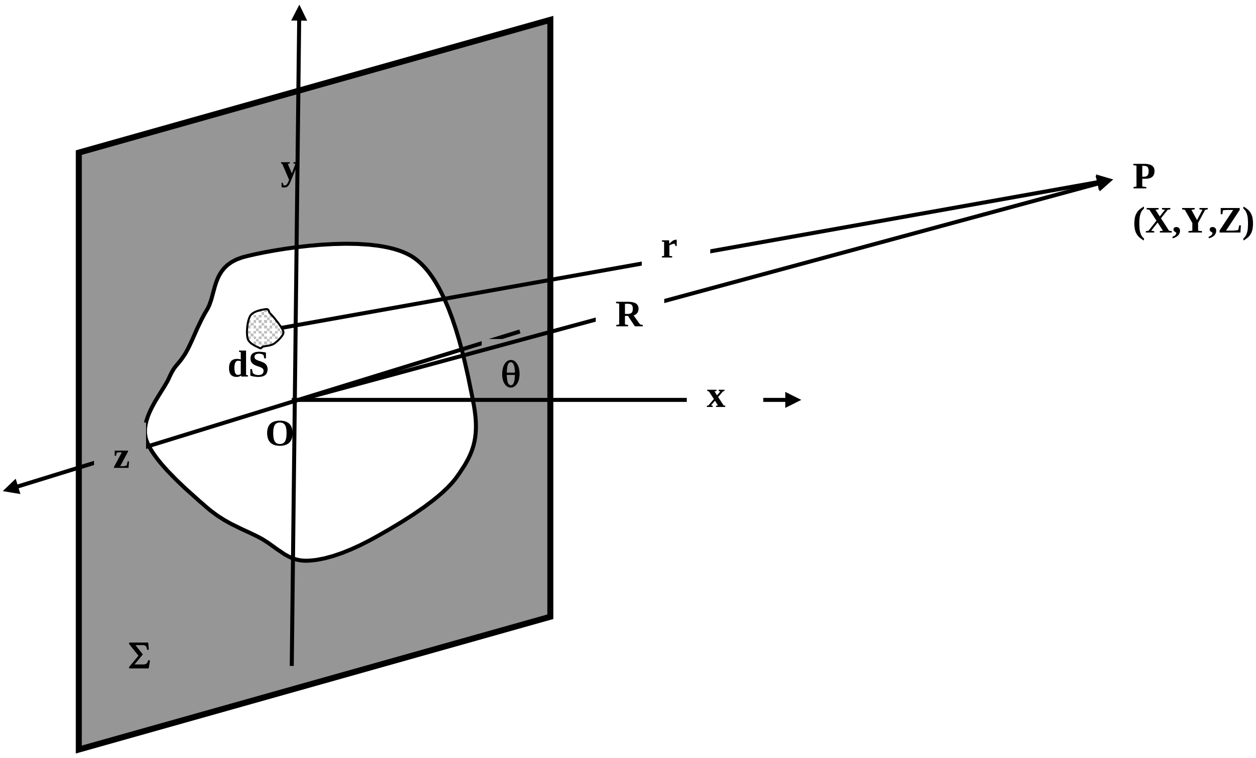
$$\theta_D = -\theta_I \tag{6.97}$$

$$\Lambda = 2\Delta \sin\theta_D = m\lambda \tag{6.98}$$

for strong diffraction. This is the same as the Bragg diffraction condition for X-ray diffraction from crystal lattices

Noting that Δ is the separation of the centre of adjacent reflecting regions which was a in the transmission grating

c) Rectangular aperture.



The above diagram shows a generalized aperture in the plane, Σ, illuminated from behind by plane waves. The problem is to find the electric field/intensity in the far field at some point P. The phase is considered to be constant over the aperture as is the amplitude. We consider a differential area *dS* in this aperture whose dimensions are much smaller than λ. This means that all contributions to the field from *dS* remain in phase at P and therefore interfere constructively. *dS* is considered to emit a spherical wave and therefore all of these considerations remain true independent of θ. *R* is the distance of the point P from the origin of axes, O, in the aperture plane and *r* is the distance of the element from P. We consider the aperture to have a source strength per unit area, ε_A . The contribution to the field from dS at P is

$$dE = \left(\frac{\varepsilon_A}{r}\right)e^{j(\omega t - kr)}dS \tag{6.99}$$

Using the exponential notation means we need to choose either the real or the imaginary part depending on whether we prefer cosines or sines respectively. The length of *r* is

$$r = \sqrt{X^2 + (Y-y)^2 + (Z-z)^2} \tag{6.70}$$

And this distance must approach infinity for the Fraunhofer approximation to remain valid. This means we can replace *r* by *R* in the amplitude factor but need to be more circumspect with the r that appears in the phase factor (the exponential).
Using

$$R = \sqrt{X^2 + Y^2 + Z^2} \tag{6.71}$$

Whence

$$r = R\left[1 + \frac{y^2 + z^2}{R^2} - 2\frac{Yy + Zz}{R^2}\right]^{1/2} \tag{6.72}$$

In the far field *R* is much larger than the aperture dimensions leaving

$$r = R\left[1 - 2\frac{Yy + Zz}{R^2}\right]^{1/2} \tag{6.73}$$

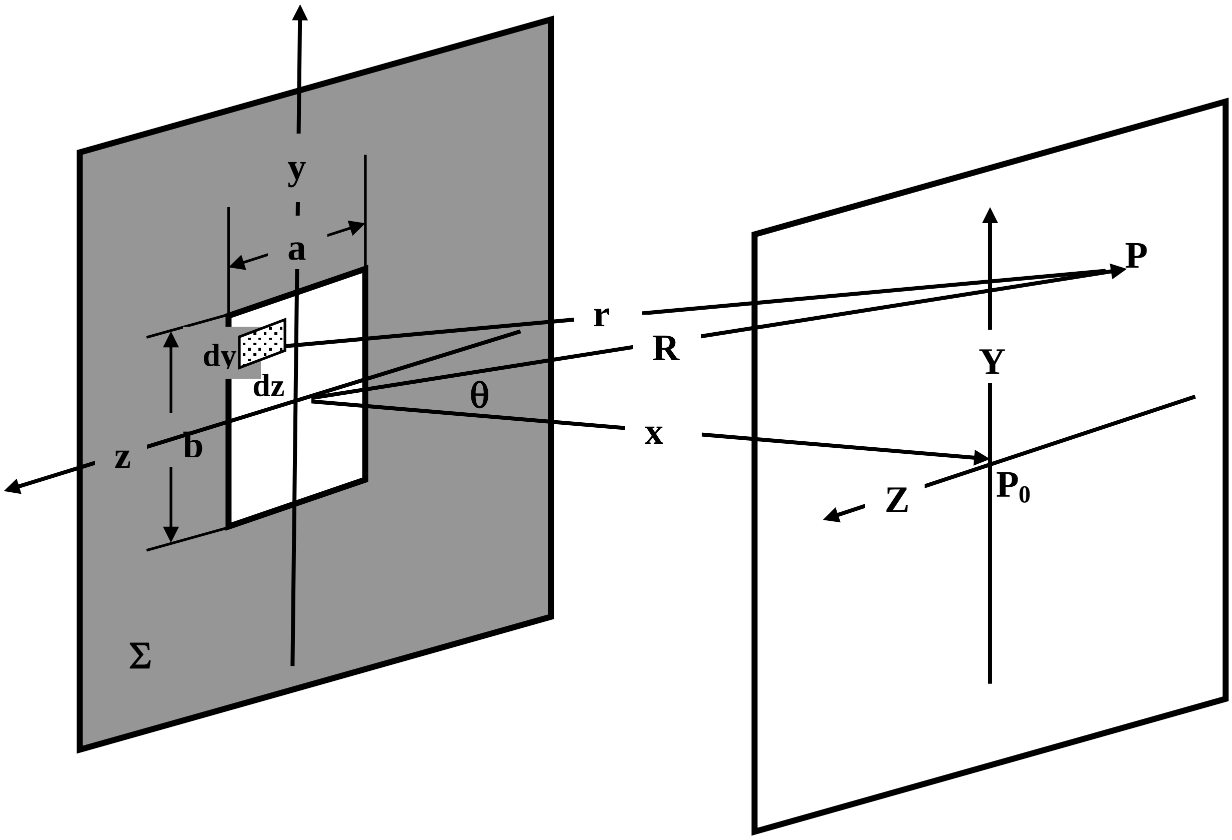
We can expand this binomially and use only the first terms to obtain

$$r = R\left[1 - \frac{Yy + Zz}{R^2}\right] \tag{6.74}$$

Then the total field, *E*(P) arriving at P is

$$E = \frac{\varepsilon_A e^{j(\omega t - kR)}}{R} \iint_{Aperture} e^{jk(Yy+Zz)/R} dS \tag{6.75}$$

Now consider a specific configuration, the rectangular aperture as shown below.



With a rectangular aperture of dimensions *ab* and an elemental area *dS* = *dydz* we can write the integral in a tractable form as two integrals using separation of variables

$$\frac{\varepsilon_A e^{j(\omega t-kR)}}{R}\iint_{Aperture} e^{jk(Yy+Zz)/R}dS = \frac{\varepsilon_A e^{j(\omega t-kR)}}{R}\int_{-b/2}^{+b/2} e^{jkYy/R}dy\int_{-a/2}^{+a/2} e^{jkZz/R}dz$$

(6.76)

The integrals can be solved by similar techniques to those used for a slit by defining

$$\beta' \equiv \frac{k_0 bY}{2R} \qquad \alpha' \equiv \frac{k_0 aZ}{2R} \qquad (6.77)$$

The two integrals can be written

$$\int_{-b/2}^{+b/2} e^{jk_0 Yy/R} dy = b\left(\frac{e^{j\beta^/} - e^{-i\beta^/}}{2j\beta^/}\right) = b\left(\frac{\sin\beta^/}{\beta^/}\right) \tag{6.78}$$

and

$$\int_{-a/2}^{+a/2} e^{jk_0 Zz/R} dz = a\left(\frac{e^{j\alpha^/} - e^{-j\alpha^/}}{2j\alpha^/}\right) = a\left(\frac{\sin\alpha^/}{\alpha^/}\right) \tag{6.79}$$

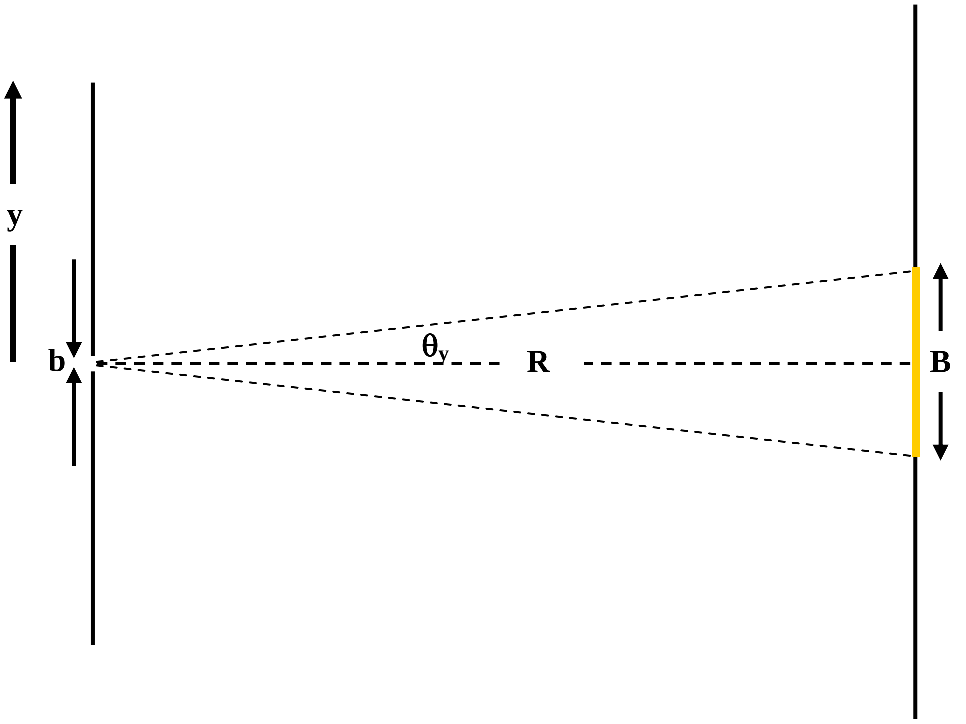
The electric field is then

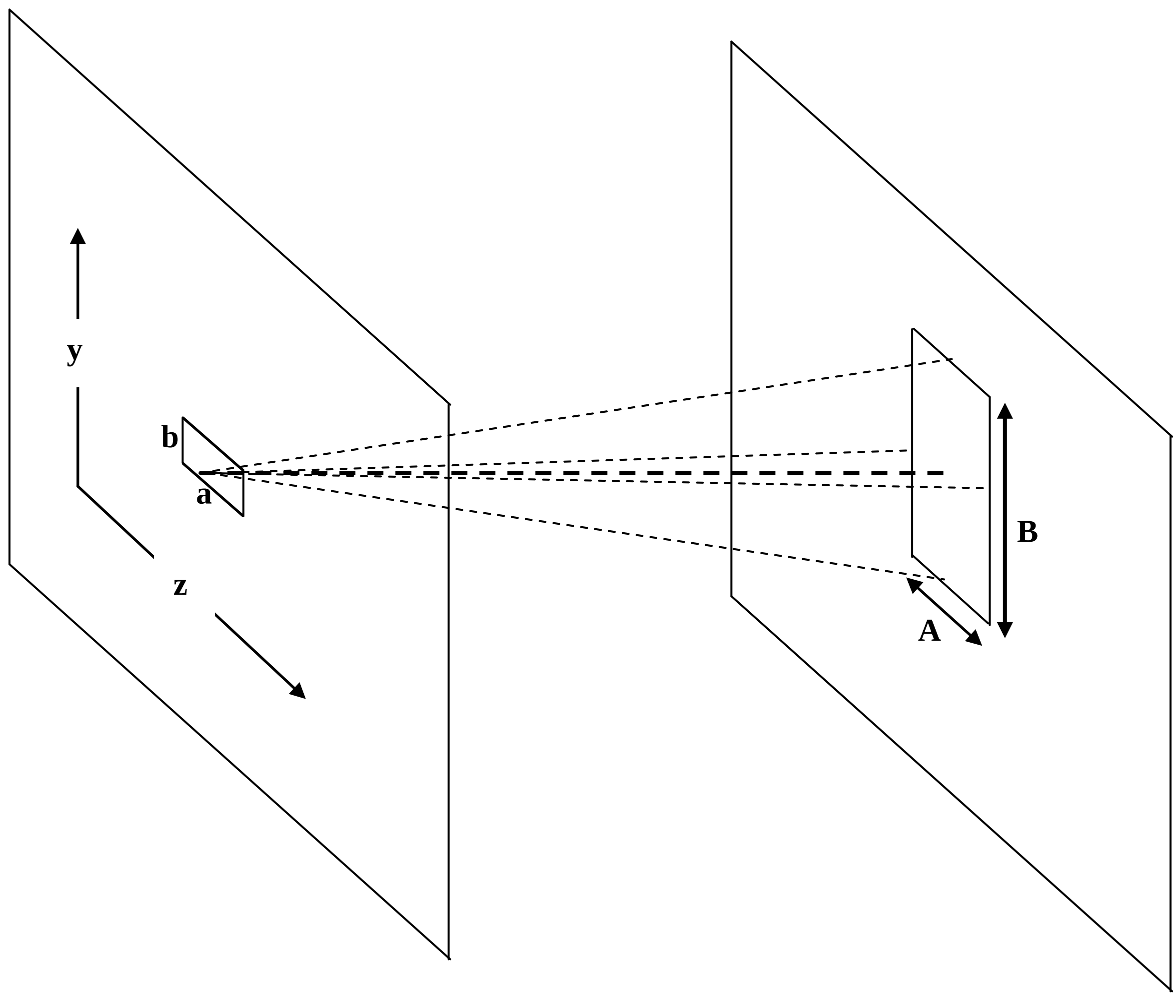
$$E = \frac{A\varepsilon_A e^{j(\omega t - k_0 R)}}{R}\left(\frac{\sin\alpha^/}{\alpha^/}\right)\left(\frac{\sin\beta^/}{\beta^/}\right) \tag{6.80}$$

and the intensity follows as

$$I = I_0\left(\frac{\sin\alpha^/}{\alpha^/}\right)^2\left(\frac{\sin\beta^/}{\beta^/}\right)^2 \tag{6.81}$$

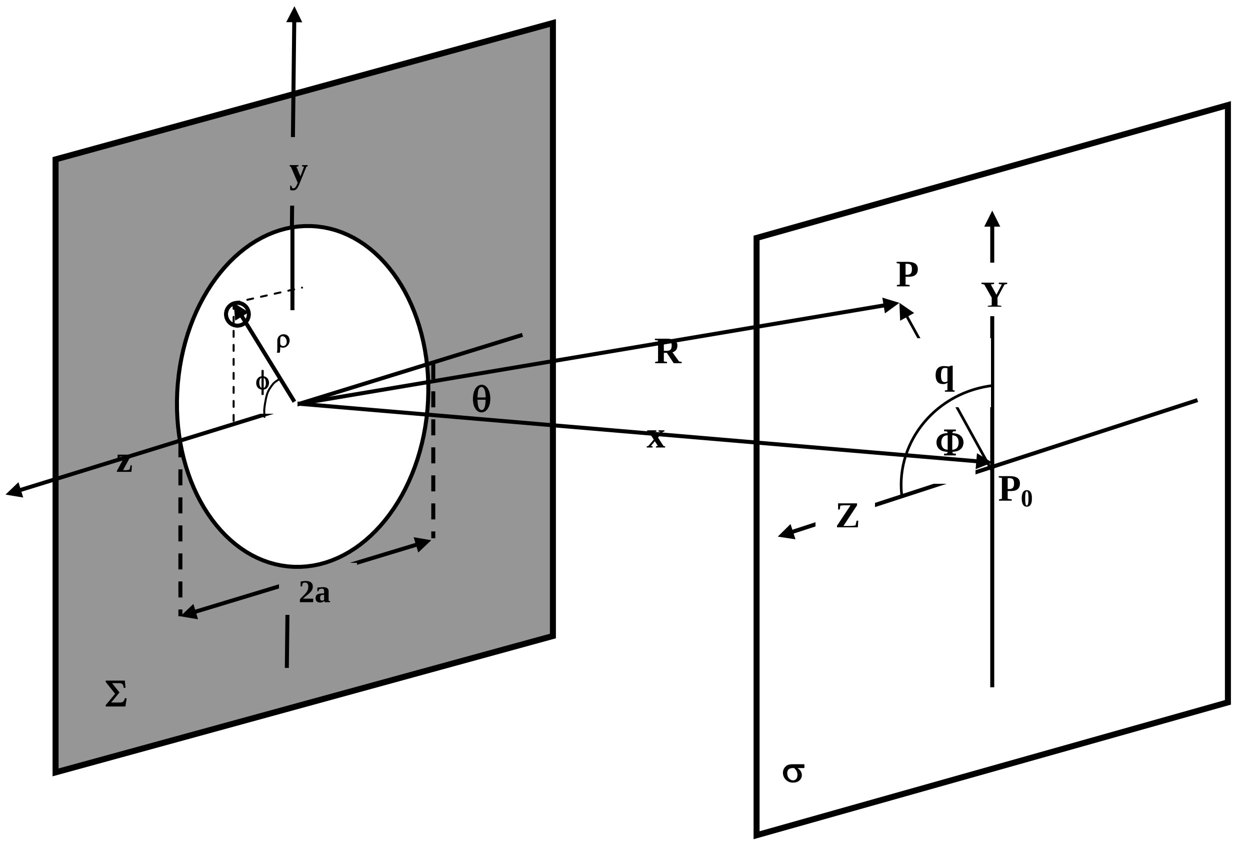
where I_0 is the intensity at P_0 ($Z = Y = 0$). The first sinc^2 term gives the variation with Z and the second the variation with Y, Z and Y being the coordinates on the far field screen. This is the familiar intensity variation of a slit now with a two dimensional character. Note that the central image is that of the aperture turned through 90^0. ie. the long and short axes of the aperture are swapped over in the image plane. Light is diffracted to greater angles for narrower slits and this is the reason that we see this inversion. Note also that the diffraction angle $\sin\theta_z \approx \theta_z = \frac{Z}{R}$ or $\theta_y = \frac{Y}{R}$


y
b
θy
R
B


y
b
a
z
B
A

Circular aperture.

Whilst more difficult to deal with than the rectangular aperture, the circular aperture is of greater significance in optics as such apertures appear in many applications. The behaviour of the circular aperture will have great bearing on the resolution limits of instruments containing such apertures. In the diagram below (and others like it that we have used previously) the viewing screen at σ is to be essentially at infinity in order that the Fraunhofer diffraction approximation may be used. Usually this is impractical and a converging lens would be used to collimate the light from the aperture which is equivalent to focusing at infinity. The screen could then be placed anywhere after the lens with the same pattern projected



The way in which circular apertures are to be understood do not differ in principle from the rectangular aperture. The above diagram sets out the geometry of the circular aperture.

We already have the general expression for obtaining the far field pattern from an arbitrary aperture in 6.75. In the case of a circular aperture with its symmetry the use of spherical co-ordinates is suggested for the problem as indicated in the diagram;

$$z = \rho\cos\phi \qquad y = \rho\sin\phi$$
$$Z = q\cos\Phi \qquad Y = q\sin\Phi \tag{6.82}$$

The differential area is

$$dS = \rho d\rho d\phi \tag{6.83}$$

We can now adapt 6.75, $E = \frac{\varepsilon_A e^{j(\omega t - kR)}}{R} \iint_{Aperture} e^{jk(Yy+Zz)/R} dS$ to obtain;

$$E = \frac{\varepsilon_A e^{j(\omega t - kR)}}{R} \int_{\rho=0}^{a} \int_{\phi=0}^{2\pi} e^{j(k\rho q/R)\cos(\phi-\Phi)} \rho d\rho d\phi \tag{6.84}$$

Because of the axial symmetry the electric field must be independent of Φ and because of this we have the freedom to set $\Phi = 0$ in evaluating the integral in order to gain a slight simplification.

The part of the integral involving ϕ is one that is frequently encountered in physical problems from solutions of atomic wavefunctions to waves on cylindrical drums and intensity distributions in optical fibres. Wherever cylindrical symmetry turns up this type of integral may be encountered and has been solved with its own set of tabulated functions.

The solution to the integral

$$\int_0^{2\pi} \exp\left(\frac{jk\rho q}{R}\cos\phi\right) d\phi = \int_0^{2\pi} \exp(u\cos\phi) d\phi \tag{6.85}$$

is not representable in terms of functions that we are readily familiar with (cos, sin exp etc.) and a new function is defined is defined as

$$\frac{1}{2\pi}\int_0^{2\pi}\exp(u\cos\upsilon)d\upsilon \equiv J_0(u) \tag{6.86}$$

Where $J_0(u)$ is the Bessel function of the first kind of zero order (see notes on optical fibres).

And more generally

$$\frac{i^{-m}}{2\pi}\int_0^{2\pi}\exp(m\upsilon + u\cos\upsilon)d\upsilon \equiv J_m(u) \tag{6.87}$$

Where $J_m(u)$ is the Bessel function of the first kind of mth order.
Bessel functions are tabulated in the same way as sines or cosines and are available in software such as Excel. We just need to know that they are the solution to the integral we are interested in.
Using the Bessel function, 6.84 can be re-written

$$E = \frac{\varepsilon_A e^{j(\omega t - kR)}}{R} 2\pi \int_{\rho=0}^{a} J_0(k\rho q / R)\rho d\rho \tag{6.88}$$

Another property of the Bessel functions, equivalent to the rules of differentiation for sines and cosines, is the recurrence relation

$$\frac{d}{du}\left[u^m J_m(u)\right] = u^m J_{m-1}(u) \tag{6.89}$$

Therefore

$$\int_0^u u' J_0(u')du' = uJ_1(u) \tag{6.90}$$

Where u' is a dummy variable

6.88 can now be integrated by changing the variable, $w = {}^{k\rho q}\!/_{R}$ and $d\rho = {}^{R}\!/_{kq}\,dw$ leaving 6.88 as

$$\int_{\rho=0}^{a} J_0(k\rho q/R)\rho d\rho = \left(\frac{R}{kq}\right)^2 \int_{w=0}^{w={}^{kaq}\!/_{R}} J_0(w)wdw \tag{6.91}$$

and using the integral form of the recurrence relation given above

$$E = \frac{\varepsilon_A e^{j(\omega t-kR)}}{R} 2\pi a^2 \frac{R}{kaq} J_1\left({}^{kaq}\!/_{R}\right) = \frac{\varepsilon_A e^{j(\omega t-kR)}}{R} A \frac{R}{kaq} J_1\left({}^{kaq}\!/_{R}\right)$$

(6.92)

The irradiance is found from the time average square of the field or ${}^{1}\!/_{2}\,EE^*$

$$I = \frac{2\varepsilon_A^2 A^2}{R^2}\left[\frac{J_1\left({}^{kaq}\!/_{R}\right)}{{}^{kaq}\!/_{R}}\right]^2 \tag{6.93}$$

A is the area of the circular aperture. To find the intensity at the centre of the pattern we set $q = 0$ and therefore $u = 0$.

$$J_0(u) = \frac{d}{du}J_1(u) + \frac{J_1(u)}{u} \tag{6.94}$$

From 6.86 and 6.87, $J_0(u = 0) = 1$ and $J_1(u=0) = \frac{1}{2\pi i}\int_0^{2\pi} e^{\upsilon} d\upsilon = 0$

Using l'Hopitals rule

$$\lim_{u\to 0}\frac{J_1(u)}{u} = \lim_{u\to 0}\frac{{}^{dJ_1}\!/_{du}}{{}^{du}\!/_{du}} = \lim_{u\to 0}\frac{dJ_1}{du} \tag{6.95}$$

and the RHS of 6.94 = $2\lim_{u\to 0} J_1(u)/u = J_0(u=0) = 1$

$$\lim_{u\to 0} J_1(u)/u = \frac{1}{2}$$

The intensity at P_0 is then

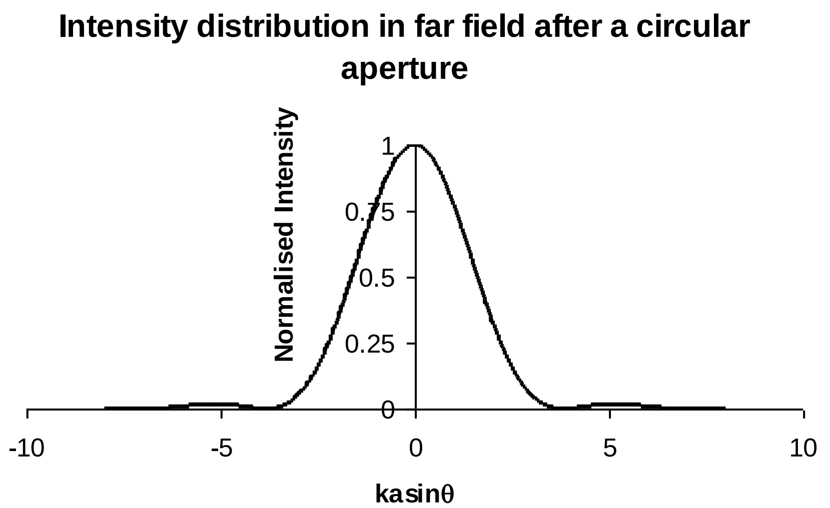
$$I(0) = \frac{\varepsilon_A^2 A^2}{2R^2} \qquad (6.96)$$

Allowing 6.93 to be rewritten

$$I = I(0)\left[\frac{2J_1\left(kaq/R\right)}{kaq/R}\right]^2 \qquad (6.97)$$

and finally, as $\sin\theta = q/R$

$$I(\theta) = I(0)\left[\frac{2J_1(ka\sin\theta)}{ka\sin\theta}\right]^2 \qquad (6.98)$$



The graph above shows the intensity distribution in the far field. The intensity does not depend on the azimuthal angle , Φ, as there is circular symmetry. Therefore this intensity distribution represents a bright central disc surrounded by a series of concentric dark and bright rings **of much reduced** amplitude. The bright central disc is important in many optics applications where there are circular apertures present (most lenses for example) and it plays a role in defining the resolution of telescopes for instance where the image is collected in the far field, the object is at infinity and the focus of the lens is adjusted to account for this. To find the diameter of the central disc **we need to find the first zero in the intensity distribution**. This occurs for $J_1(ka\sin\theta) = 0$

Looking at the properties of the Bessel functions we find that this occurs when the argument of the Bessel function is equal to 3.83

$$ka\sin\theta = 3.83$$

We may couch this in terms of wavelength, λ, and diameter of the aperture $D = 2a$

$$\frac{2\pi}{\lambda}\frac{D}{2}\sin\theta = 3.83$$

$$\sin\theta = \frac{\lambda}{\pi D} \times 3.83 = 1.22\frac{\lambda}{D}$$

The projected disc by a lens of focal length F on a screen will have a diameter d

$$\frac{d}{2F} = \sin\theta = 1.22\frac{\lambda}{D}$$

The disc has a radius, $R = \frac{d}{2}$ given by

$$R = 1.22\frac{F\lambda}{D}$$

This radius corresponds to a subtended angle of

$$\Delta\theta = \frac{R}{F} = 1.22\frac{\lambda}{D}$$

This disc is known as the Airy disc after an Astronomer Royal and sets the diffraction limits to the resolution of instruments such as telescopes. Rayleighs criteria states that two stars are just resolved when the bright centre of one disc (star) falls at the first zero of the second star. This gives the diffraction limited minimum angular resolution of a telescope as

$$\Delta\theta_{Min} = 1.22\frac{\lambda}{D}$$

Where D is the minimum aperture diameter in the optical system and λ is the wavelength of light being imaged. The larger the minimum diameter of the optical elements forming the telescope (or any other instrument) the better the angular resolution. We see now the need for large telescopes. Of course diffraction may not be the only limit on resolution and nowadays the platform for the telescope is likely to be above the Earths atmosphere in outer space.

One final point to note about this limit on resolution imposed by diffraction from circular apertures. Should it be required to take a collimated laser beam of diameter D and focus it to a spot size of radius w there is a diffraction limit on how small this spot may be or in other words how tightly focused the laser may be.
This diffraction limit is given by

$$w = 1.22\frac{F\lambda}{D}$$