4. Theory

4.1 Fraunhofer Diffraction

4.1.1 Fraunhofer Diffraction at a Single Slit

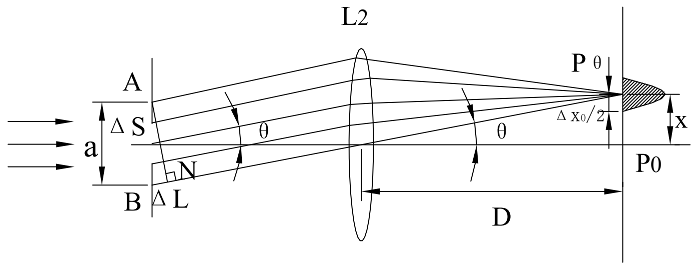

Figure 1 Schematic of Fraunhofer diffraction from single slit

Fraunhofer diffraction is the diffraction of parallel light rays, the so-called far-field diffraction, which can be treated and explained by using the Huygens' Principle. As a plane wave is incident on a long, narrow slit, secondary waves are emitted. These secondary waves are spherical, as if there were an infinite number of point sources across the aperture. From a particular observation point, each source has a different optical path that introduces a phase relationship between the waves emitted from the aperture. The total intensity of the waves emitted from the aperture bears a simple relationship with regard to the angle of diffraction, as described in many textbooks. In the observation plane, we may write:

$$I = I_0 \frac{\sin^2(u)}{u^2} \tag{1-1}$$

where $u = \frac{2\pi}{\lambda}\frac{a}{2}\sin\theta$, a: the slit-width, λ: the light wavelength, and θ: the angle of diffraction.

When $u = 0$, namely, $\theta = 0$, maximum diffraction occurs, so the relative intensity is $I/I_0 = 1$;

When $u = n\pi$, namely, $a\sin\theta = n\lambda$, where n is an integer, minimum diffraction occurs ($I = 0$).

Let $\frac{d}{du}(\frac{\sin u}{u}) = 0$, we get the conditions of secondary maxima:

$$u = \tan(u) \tag{1-2}$$

By calculating, we have the following table:

u	$\sin(\theta)$	I/I_0
$\pm1.43\pi$	$\pm1.43\lambda/a$	0.047
$\pm2.46\pi$	$\pm2.46\lambda/a$	0.017
$\pm3.47\pi$	$\pm3.47\lambda/a$	0.008
…	…	…

From the above table, the intensity of Fraunhofer diffraction at a single slit is illustrated in the following figure.

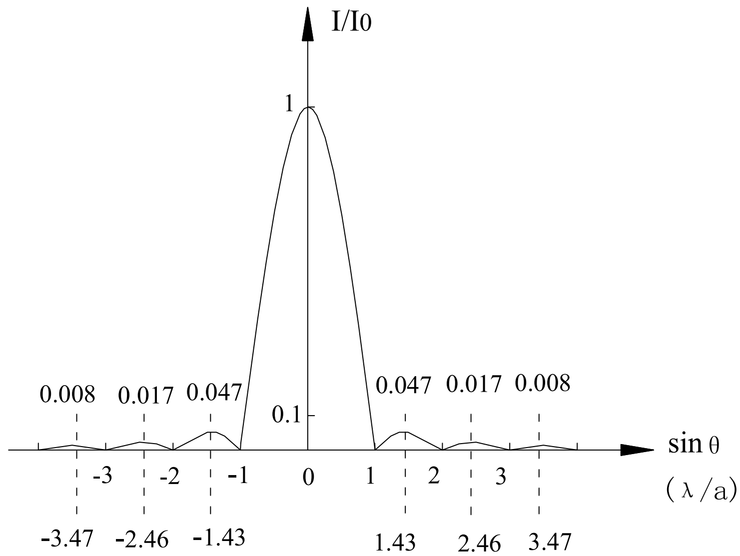

Figure 2 Intensity distribution of Fraunhofer diffraction at single slit

The divergence angle of a laser is very small, so the laser beam can be considered as a parallel beam. A laser can be used as the source for Fraunhofer diffraction as shown in Figure 3. If the distance between the detector and the slit is far enough so that the difference between AP_0 and OP_0 is far less than the laser wavelength, λ, the condition of Fraunhofer diffraction is satisfied. Let's prove it.

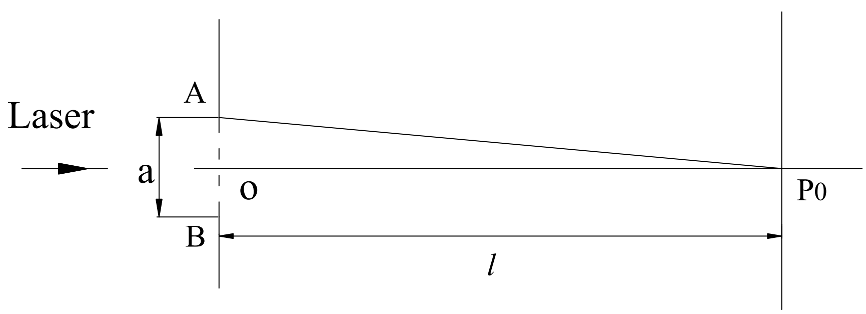

Figure 3 Schematic of far-field Fraunhofer diffraction

If the slit-width, *a*, is considered far less than the observation distance from the slit, l, then

$$AP_0\text{-}OP_0=\sqrt{l^2+\frac{a^2}{4}}-l\approx l\left(1+\frac{a^2}{8l^2}\right)-l=\frac{a^2}{8l} \qquad (1\text{-}3)$$

In the experiment, if l = 0.8 m, a = 0.1 mm, then

$\frac{a^2}{8l}$=15.6 nm, which is far less than λ (632.8 nm). Hence, the far-field condition for Fraunhofer diffraction is satisfied.

4.1.2 Fraunhofer Diffraction at a Multi-Slit

When the single slit is replaced by a 'multi-slit' that has N evenly-spaced slits with a width of a for each slit and a slit-to-slit separation of d, the resulted intensity of diffraction is the intensity distribution from N coherent sources multiplied by the diffraction pattern from a single slit. The interference pattern is determined by the phase angle $\beta=(\pi\times d\times\sin\theta)/\lambda$, where λ is the wavelength of light, and θ the angle of diffraction or angle of observation. Because the diffraction pattern is determined by $u=(2\pi\times a\times\sin\theta)/2\lambda$, the net result is

$$I = I_0\left(\frac{\sin u}{u}\right)^2\left(\frac{\sin N\beta}{\sin\beta}\right)^2 \qquad (1\text{-}4)$$

where $(\sin u)/u$ or $(\sin u)^2/u^2$ is a diffraction function and $(\sin N\beta)/(\sin\beta)$ or $(\sin N\beta)^2/(\sin\beta)^2$ is an interference function.

Based on equation (1-4), the following picture shows the normalized interference function with different N values,

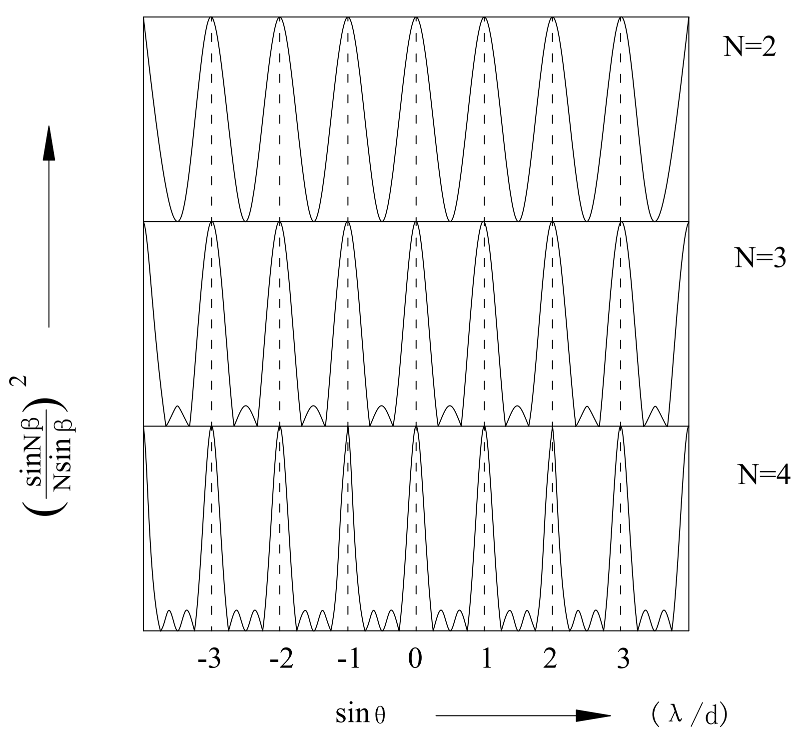

Figure 4 Normalized interference intensity distribution of N coherent sources

1) *Peak values, positions, and numbers of the principal maxima*

When $\beta = k\pi$, namely $\sin\theta = k\dfrac{\lambda}{d}$ $(k = 0, \pm 1, \pm 2, \pm 3, \cdots)$, then

$$I = I_0 (\frac{\sin u}{u})^2 (\frac{\sin N\beta}{\sin\beta})^2 = N^2 I_0 (\frac{\sin u}{u})^2 \qquad (1\text{-}5)$$

Obviously, the principal maximum diffraction intensity from a multi-slit with N evenly-spaced slits is N^2 times the diffraction intensity at a single slit.

Considering $|\sin\theta| \le 1$, so $|k| \le d/\lambda$

Let $\lambda = 0.4d$, then k can only take $0, \pm 1, \pm 2$.
If $\lambda \ge d$, then $k = 0$, it means there is only zero order.

2) Positions of zero intensity points, numbers of the secondary maxima

Between these principal maxima there are weak secondary maxima. Let $\sin N\beta = 0$ and $\sin\beta \ne 0$, we can get these positions of zero intensity points.

$$\beta = (k + \frac{m}{N})\pi \text{ , namely } \sin\theta = (k + \frac{m}{N})\frac{\lambda}{d} \qquad (1\text{-}6)$$

where $k = 0, \pm 1, \pm 2, \cdots$; $m = 1, 2, \cdots, N-1$.

Because there are (N–1) zero points between two principal maxima, and there is a secondary maximum between the consecutive zero points, the number of secondary maxima is N–2. The following figure shows the influence of the diffraction function (N = 5, $d = 3a$).

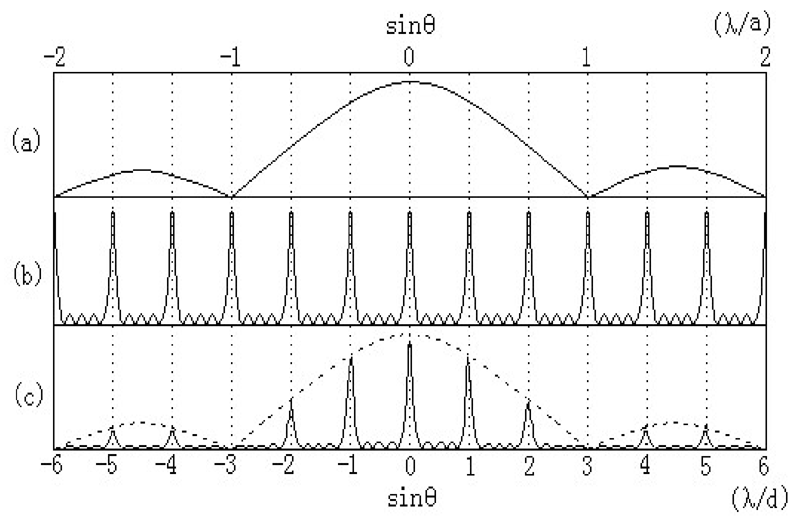

Figure 5 Diffraction intensity (c) of 5-slit (N = 5, $d = 3a$)

It shows that the interference function (b) is modulated by the diffraction function (a), yielding the net result shown in (c). For $d = 3a$, the third order ($k = 3$) principal maximum of the interference function coincides with the first zero position of single slit diffraction. This is the so-called missing order phenomenon.

4.1.3 Fraunhofer Diffraction at a Single Circular Aperture

When light from a point source passes through a small circular aperture, it produces a diffuse circular disc known as the Airy's disc. The diffraction pattern is given by:

$$I = I_0[\frac{2J_1(x)}{x}]^2,\ (x = \frac{\pi D}{\lambda}\sin\theta) \tag{1-7}$$

where D is the diameter of the circular aperture, $J_1(x)$ is the first-order Bessel function, and λ is wavelength of the source. The following table gives some values of the Bessel function:

x	0	1.220π	1.635π	2.233π	2.679π	3.238π
$[2J_1(x)/x]^2$	1	0	0.0175	0	0.0042	0

So the first minimum (dark ring) condition is:

$$a\sin\theta = 1.22\lambda \tag{1-8}$$

and the Rayleigh criterion, or the equation for angular resolution is

$$\theta_{\min} = 1.22\frac{\lambda}{d} \tag{1-9}$$

4.2 Fresnel Diffraction

Fresnel diffraction or near-field diffraction is a process of diffraction that occurs when a wave passes through an aperture or slit, and diffracts in the near field. It occurs when the source and the observation distance are relatively close to the obstacle forming the diffraction pattern. If the source, the obstacle, and the observation distance are far enough away that both incident and diffracted waves are considered parallel, then Fraunhofer diffraction occurs.

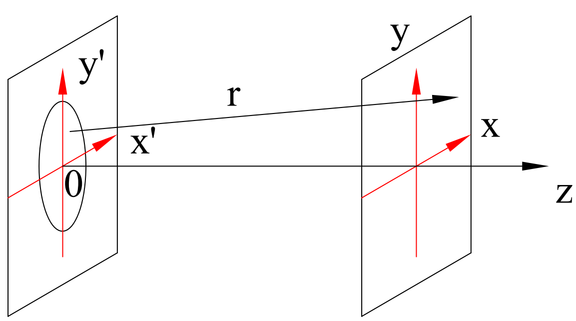

Figure 6 Diffraction geometry - aperture and image planes with coordinate systems

The Fresnel Diffraction Integral

The diffraction pattern of the electric field at a point (x, y, z) is given by:

$$E(x,y,z) = -\frac{i}{\lambda}\iint E(x',y',0)\frac{e^{ikr}}{r}\cos\theta dx'dy' \tag{1-10}$$

Where $r = \sqrt{(x-x')^2+(y-y')^2+z^2}$, i is the imaginary unit, and $\cos\theta = z/r$ is the cosine of the angle between z and r.

Analytical solution of this integral is impossible for all but the simplest diffraction geometries. Therefore, it is usually calculated numerically.

The Fresnel approximation

The main problem for solving the above integral is the expression of *r*. First, we can simplify the algebra by introducing the substitution: $\rho^2 = (x-x')^2+(y-y')^2$.

Substituting it into the expression for *r*, we get

$$r = \sqrt{\rho^2+z^2} = z\sqrt{1+\frac{\rho^2}{z^2}} \tag{1-11}$$

Next, using the Taylor series expansion $\sqrt{1+u} = (1+u)^{1/2} = 1+\frac{u}{2}-\frac{u^2}{8}+...$

We can express *r* as

$$\begin{aligned} r &= z\sqrt{1+\frac{\rho^2}{z^2}} \\ &= z\left[1+\frac{\rho^2}{2z^2}-\frac{1}{8}\left(\frac{\rho^2}{z^2}\right)^2+...\right] \\ &= z+\frac{\rho^2}{2z}-\frac{\rho^4}{8z^3}+... \end{aligned} \tag{1-12}$$

The key to the Fresnel approximation is to assume that the third element in the Taylor series in equation (1-12) is very small and therefore can be ignored. In order to make this assumption, this third element has to contribute to the variation of the exponential as an almost null term. In other words, it has to be much smaller than the period of the complex exponential, i.e. 2π:

$$k\frac{\rho^4}{8z^3} \ll 2\pi \tag{1-13}$$

By expressing *k* in terms of the wavelength, $k = 2\pi/\lambda$, we get the following relationship:

$$\frac{\rho^4}{z^3\lambda} << 8 \tag{1-14}$$

Multiplying both sides of equation (1-14) by z/λ, we have

$$\frac{\rho^4}{z^2\lambda^2} << 8\frac{z}{\lambda} \tag{1-15}$$

Or substituting the earlier expression for ρ^2

$$\frac{\left[(x-x')^2+(y-y')^2\right]^2}{z^2\lambda^2} << 8\frac{z}{\lambda} \tag{1-16}$$

If this condition holds true for all values of x, x', y, and y', then we can ignore the third term in the Taylor expression. Furthermore, if the third term is negligible, then all terms of higher order will be even smaller, so we can ignore them as well. We can then approximate the expression with only the first two terms, as:

$$r \approx z + \frac{(x-x')^2+(y-y')^2}{2z} \tag{1-17}$$

Thus, equation (1-17) is the Fresnel approximation, and the inequality (1-16) is the condition for the Fresnel approximation.

Fresnel Diffraction

The condition stated in inequality (1-16) for Fresnel approximation is fairly weak, since it allows all length parameters to take comparable values, provided the aperture is much smaller than the path length. Moreover, if we are interested in the field only in a small area close to the origin, i.e. for values of x and y much smaller than z, then we can assume $\theta \approx 0$, which means $\cos\theta \approx 1$ and r in the denominator of the Fresnel integral can be approximated by $r \approx z$

Unlike Fraunhofer diffraction, Fresnel diffraction has to consider the curvature of the wavefront, in order to correctly calculate the relative phase of interfering waves.

For Fresnel diffraction, the electric field at point (x,y,z) is given by:

$$E(x,y,z) = -\frac{i}{\lambda}\frac{e^{ikr}}{z}\iint E(x',y',0)e^{\frac{ik}{2z}\left[(x-x')^2+(y-y')^2\right]^2}dx'dy' \tag{1-18}$$

This is the Fresnel diffraction integral; it means that, if the Fresnel approximation is valid, the propagating field is a spherical wave that originates at the aperture and moves along z direction. The integral modulates the amplitude and phase of the spherical wave. Analytical solution of this expression is only possible in rare cases.