Modern Optics
From diffraction theory to Fourier optics

0.1 Introduction

One of the first reference to diffraction phenomena appeared with Leonardo da Vinci (1452 – 1519) but were only accurately described in 1665 by Grimaldi. At the time the corpuscular theory, widely used at this period, could not explain the diffraction. In particular at the boundaries of shadows one could observe the appearance of dark and bright bands: the *diffraction fringes*. It is important to note that the problems linked with diffraction problems are amongst the most difficult ones encountered in optics and rigorous solution are in fact very rare.

As we will see the first real explanation of such observation relies on the principle enunciated by the Dutch physicist Christiaan Huygens (1629 – 1695): "*every point upon which the luminous disturbance falls may be regarded as the center of a new disturbance propagated in the form of spherical waves*". However such hypothesis was totally ignored for more than a century until Augustin-Jean Fresnel exposed his theory in 1818 using Huygens's construction. And it took another sixty years to be mathematically formulated by Kirchhoff in 1882!

0.1.1 Fresnel's calculation from Huygens construction

The initial step to explain the observation of diffraction is the Huygens construction. In this approach each point on the wavefront of the disturbance can be considered to be a new source of a "*secondary*" spherical disturbance (Fig. 1). Then the wavefront at a

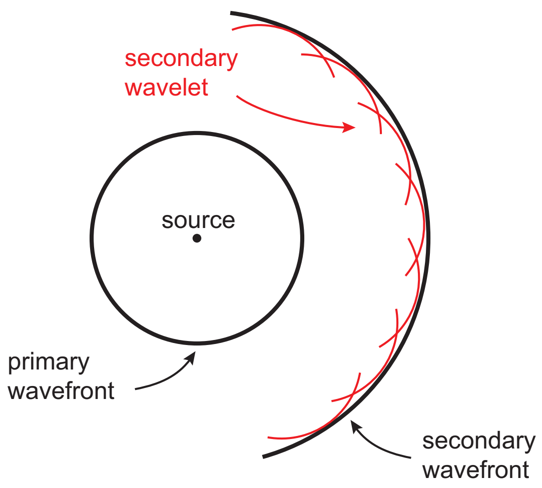

Figure 1: Huygens construction of the propagation of the light

later instant is found by constructing the *envelope* of the secondary wavelets. As we saw previously the spherical wave is mathematically described by

$$A(r) = A_0 \frac{e^{ik(r-r_0)}}{|\mathbf{r} - \mathbf{r_0}|} \tag{1}$$

Fresnel's idea is to use this approach to evaluate the amount of field at an observation point P when the considered source is a source point P_0 (Fig. 2). Considering that the

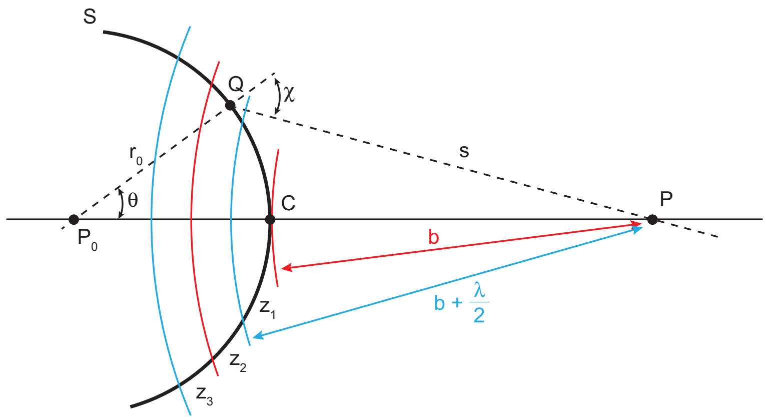

Figure 2: Fresnel construction in order to evaluate the field in an observation point P from a point source P_0. The Fresnel zones are indicated by z_i.

point source P_0 emits a spherical wave, the disturbance at the point Q is represented by

$$A(Q) = A_0 \frac{e^{ikr_0}}{r_0} \tag{2}$$

and therefore the disturbance at the point P coming from the point Q is[1]

$$A(P) = A(Q) \frac{e^{iks}}{s} \tag{3}$$

As a consequence the contribution $dU(P)$ due to the element dS around the point Q is

$$dU(P) = K(\chi)\ \frac{Ae^{ikr_0}}{r_0} \cdot \frac{e^{iks}}{s}\ dS \tag{4}$$

with $s = QP$. The factor $K(\chi)$ is the so-called "*inclination factor*". It describes the variation of the amplitude of the secondary wave with the direction of the secondary wave. The "*angle of diffraction*" (χ) is the angle between the normal to the wavefront S at the point Q and the direction QP. Of course the total disturbance at the observation point P is the integration of $dU(P)$ over the whole wavefront S:

$$U(P) = \iint_S dU(P) = \frac{Ae^{ikr_0}}{r_0} \iint_S \frac{e^{iks}}{s} K(\chi)\ dS \tag{5}$$

The trick that Augustin Fresnel used to evaluate the integral eq. (25) is to divide the problem into zones, the so-called *Fresnel's zones*. These are spheres that are centered about the observation point P and with radii $b = CP$; $b + {}^{\lambda}/_{2}$; $b + {}^{2\lambda}/_{2}$; $b + {}^{3\lambda}/_{2}$; ... ; $b + {}^{j\lambda}/_{2}$; As we see on Fig. 2, these spheres divide the wavefront sphere S into zones, indicated on the figure by z_1, z_2, ..., z_j. Moreover, if we assume the $r_0 \gg \lambda$ and $b \gg \lambda$, then we can assume that $K(\chi)$ has the same value K_j for all points within the same Fresnel's zone. Moreover to calculate the integral we need to express dS:

$$dS = r_0^2 \sin\theta\ d\theta\ d\varphi \tag{6}$$

[1] The secondary wave is also a spherical wave.

where φ is the azimuthal angle. Note however that we can eliminate the $\theta-$dependence by noting that $s^2 = r_0^2 + (r_0+b)^2 - 2r_0(r_0+b)\cos\theta$ and $s \cdot ds = r_0\ (r_0+b)\sin\theta\ d\theta$. The contribution from the j^{th} zone to $U(P)$ is then

$$U_j(P) = \frac{Ae^{ikr_0}}{r_0} \iint \frac{e^{iks}}{s} K_j\ \frac{r_0}{r_0+b} \cdot sdsd\varphi = 2\pi \frac{Ae^{ikr_0}}{r_0+b} K_j \underbrace{\int_{b+\frac{j-1}{2}\lambda}^{b+j\frac{\lambda}{2}} e^{iks}\ ds}_{I} \tag{7}$$

where

$$\begin{aligned} I = \frac{1}{ik} e^{iks}\Big|_{b+\frac{j-1}{2}\lambda}^{b+j\frac{\lambda}{2}} &= \frac{-i}{k}\left\{\exp\left[ik\left(b+j\frac{\lambda}{2}\right)\right] - \exp\left[ik\left(b+\frac{j-1}{2}\lambda\right)\right]\right\} \\ &= \frac{-i}{k}\exp\left(+ikb\right)\exp\left(+ikj\frac{\lambda}{2}\right)\left[1-\exp\left(-ik\frac{\lambda}{2}\right)\right] \end{aligned} \tag{8}$$

therefore

$$U_j(P) = \frac{-2\pi i}{k} K_j \frac{Ae^{ik(r_0+b)}}{r_0+b} e^{ikj\frac{\lambda}{2}}\left(1-e^{ik\frac{\lambda}{2}}\right) \tag{9}$$

and since $k\lambda = 2\pi$ then

$$e^{ikj\frac{\lambda}{2}}\left(1-e^{-ik\frac{\lambda}{2}}\right) = e^{i\pi j}\left(1-e^{-i\pi}\right) = (-1)^j \times 2 \tag{10}$$

which yields

$$U_j(P) = 2i\lambda(-1)^{j+1}\ K_j\ \frac{Ae^{ik(r_0+b)}}{r_0+b} \tag{11}$$

The total disturbance at the observation point P is the sum of all the contribution for each dS and therefore we can write that

$$U(P) = 2i\lambda\ \frac{Ae^{ik(r_0+b)}}{r_0+b} \sum_{j=1}^{n}(-1)^{j+1}\ K_j \tag{12}$$

It is remarkable that the contributions from successive zones are alternatively positive and negative. The task is now to evaluate the series

$$\Sigma = \sum_{j=1}^{n}(-1)^{j+1}K_j = K_1 - K_2 + K_3 - \cdots + (-1)^{n+1}K_n \tag{13}$$

A trick to evaluate eq. (13) is to write it as

$$\Sigma = \frac{K_1}{2} + \left(\frac{K_1}{2} - K_2 + \frac{K_3}{2}\right) + \left(\frac{K_3}{2} - K_4 + \frac{K_5}{2}\right) + \cdots + \ ? \tag{14}$$

The last term of the series actually depends if n is even or odd. In n is odd the last term is $^{+K_n}/_2$. On the other hand, if n is even, the last term is $(^1/_2)\, K_{n-1} - K_n$. If we now assume that $\forall j | K_j > {^1/_2}\left(K_{j-1} + k_{j+1}\right)$, the means of the first neighbours then each contributions to the series Σ in bracket is negative. Therefore we have

$$\begin{cases} \Sigma\ <\ \dfrac{K_1}{2} + \dfrac{K_n}{2} & \text{for } n \text{ odd} \\ \Sigma\ <\ \dfrac{K_1}{2} + \dfrac{K_{n-1}}{2} - K_n & \text{for } n \text{ even} \end{cases} \tag{15}$$

Another way to evaluate the series Σ is to write it as

$$\Sigma = K_1 - \frac{K_2}{2} - \left(\frac{K_2}{2} - K_3 + \frac{K_4}{2}\right) - \left(\frac{K_4}{2} - K_5 + \frac{K_6}{2}\right) - \cdots + ? \tag{16}$$

and this time the last term is either $-1/2K_{n-1} + K_n$ for n odd or $-1/2K_n$ for n even. As previously since we assumed that $\forall j|1/2\,(K_{j-1} + k_{j+1})$, the terms in brackets are negative. As a consequence we can say that the series

$$\begin{cases} \Sigma \;>\; K_1 - \dfrac{K_2}{2} - \dfrac{K_{n-1}}{2} + K_n & \text{for } n \text{ odd} \\ \Sigma \;>\; K_1 - \dfrac{K_2}{2} - \dfrac{K_n}{2} & \text{for } n \text{ even} \end{cases} \tag{17}$$

The contributions K_j are such that K_j differs only slightly from its neighbours ($K_j - 1/2\; K_{j+1} \simeq 1/2\; K_j$). Merging the relations from eq. (15) and eq. (17) we can write:

- For n odd

$$\begin{cases} \Sigma \;<\; \dfrac{K_1}{2} + \dfrac{K_n}{2} & \text{(1}^{\text{st}}\text{ way to evaluate the series)} \\ \Sigma \;>\; K_1 - \dfrac{K_2}{2} - \dfrac{K_{n-1}}{2} + K_n \simeq \dfrac{K_1}{2} + \dfrac{K_n}{2} \end{cases}$$

  Therefore for n odd we have

$$\boxed{\Sigma = \frac{K_1}{2} + \frac{K_n}{2}} \tag{18}$$

- For n even

$$\begin{cases} \Sigma \;<\; \dfrac{K_1}{2} + \dfrac{K_{n-1}}{2} - K_n \simeq \dfrac{K_1}{2} - \dfrac{K_n}{2} \\ \Sigma \;>\; K_1 - \dfrac{K_2}{2} - \dfrac{K_n}{2} \simeq \dfrac{K_1}{2} - \dfrac{K_n}{2} \end{cases}$$

  Therefore for n even we have

$$\boxed{\Sigma = \frac{K_1}{2} - \frac{K_n}{2}} \tag{19}$$

and we can now fully evaluate the disturbance at the observation point P:

$$\boxed{U(P) = i\lambda \frac{Ae^{ik(r_0+b)}}{r_0+b}\,(K_1 \pm K_n)} \;\text{ with }\; \begin{cases} \text{sign + for n odd} \\ \text{sign - for n even} \end{cases} \tag{20}$$

Note that

$$U_1(P) = 2i\lambda \frac{Ae^{ik(r_0+b)}}{r_0+b} K_1 \tag{21a}$$

$$U_n(P) = 2i\lambda \frac{Ae^{ik(r_0+b)}}{r_0+b}\,(-1)^{n+1}\,K_n = \begin{cases} 2i\lambda \dfrac{Ae^{ik(r_0+b)}}{r_0+b} K_n & \text{if n is odd} \\ -2i\lambda \dfrac{Ae^{ik(r_0+b)}}{r_0+b} K_n & \text{if n is even} \end{cases} \tag{21b}$$

And if we combine eq. (21) with eq. (20) we see that the total disturbance at the observation point P simply becomes:

$$U(P) = \frac{1}{2}\left[U_1(P) + U_n(P)\right] \tag{22}$$

Moreover since the last Fresnel zone QP (see Fig. 2) is tangent to the wave, then $\chi = (\pi/2)$ and $K(\chi)$ is assume to be zero. Therefore $U_n(P) = 0$ and the total disturbance at the observation point P reduces in

$$\boxed{U(P) = \frac{1}{2}U_1(P) = i\lambda K_1 \frac{Ae^{ik(r_0+b)}}{r_0 + b}} \tag{23}$$

Of course at this stage we still do not know what is K_1. But actually, we can notice that if $i\lambda K_1$ was equal to unity, we would have the disturbance given at the obgservation point simply by the application of the propagation of a spherical wave over the distance $r_0 + b$. This expression is then in agreement with the effect of a spherical wave propagating in free space if $i\lambda K_1 = 1$ *i.e.*

$$K_1 = -\frac{i}{\lambda} = \frac{e^{-i\frac{\pi}{2}}}{\lambda} \tag{24}$$

The numerator of this expression assumes that the secondary waves is out of phase with the primary wave by a quarter of a period, whilst the denominator indicates that the amplitude of the secondary vibrations are reduced by a factor λ to those of the primary wave. This assumption shows that the Huygens-Fresnel principle leads to the correct expression for the propagation of a spherical wave in free space! Another surprising consequence of the Huygens-Fresnel principle is expressed by the eq. (23): The total disturbance at P is equal to half the disturbance due to the first Fresnel zone! ... but there is more to come...

0.1.2 Obstruction of certain Fresnel zones

Everything masked except half of the first Fresnel zone

Here we consider that we place a mask between the point source S and the observation point P. The mask has a circular aperture corresponding to exactly half the first Fresnel zone (Fig. 3).

We have already calculated the total disturbance at the point P:

$$U(P) = 2i\lambda \frac{Ae^{ik(r_0+b)}}{r_0 + b} \sum_{j=1}^{n} (-1)^{j+1} K_j \tag{25}$$

However this time $n = 1$ and we need to multiply the result by $(1/2)$ since we only consider half the first Fresnel zone. We then find that the disturbance at the observation point P is now

$$U(P) = i\lambda K_1 \frac{A^{ik(r_0+b)}}{r_0 + b} \tag{26}$$

Comparing this result with eq. (23) we see that the presence of the mask does not modify anything: the disturbance if the same than when we do not block any light from the point source!

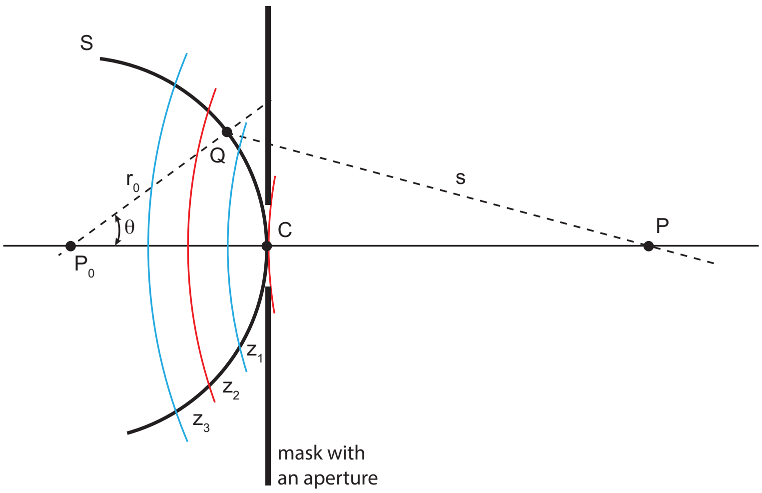

Figure 3: Fresnel construction when a mask with an aperture is placed between the point source S and the observation point P.

Everything masked except the first Fresnel zone

As we did previously we place a mask between the point source S and the observation point P but this time the whole first Fresnel zone is left open. Using the eq. (25) we obtain that the disturbance at P is

$$U(P) = 2i\lambda K_1 \frac{A^{ik(r_0+b)}}{r_0+b} \tag{27}$$

That is very remarkable. If we calculate the intensity $I(P) = |U(P)|^2$, we see that blocking everything but the first Fresnel zone yield **four times more** light at P that without any aperture !!!

Everything masked except the first two Fresnel zone

Once again we use a mask but this time the circular aperture is such that the first two Fresnel zone are considered in the eq. (25). The result is immediate:

$$U(P) = 2i\lambda \frac{Ae^{ik(r_0+b)}}{r_0+b}(K_1 - K_2) \tag{28}$$

but since $K_j \simeq K_{j+1}$ we find that there is almost no light at the observation point P although the aperture is now twice more opened than previously! This is obviously due to the fact that the contributions from consecutive Fresnel zones yield destructive interference.

The first Fresnel zone is masked

What happens if instead of having an aperture allowing n Fresnel zone(s) to be taken into account we now use a mask that has the size of the first Fresnel zone?

The disturbance at P is now

$$U(P) = 2i\lambda \frac{Ae^{ik(r_0+b)}}{r_0+b} \sum_{j=2}^{n} (-1)^{j+1} K_j = 2i\lambda \frac{Ae^{ik(r_0+b)}}{r_0+b} (-K_2 + K_3 - K_4 + \dots) \tag{29}$$

We can use the same trick as before to evaluate the results of the series and find that

$$-K_2 + K_3 - K_4 + \dots \simeq \frac{-K_2}{2} \simeq \frac{-K_1}{2}$$

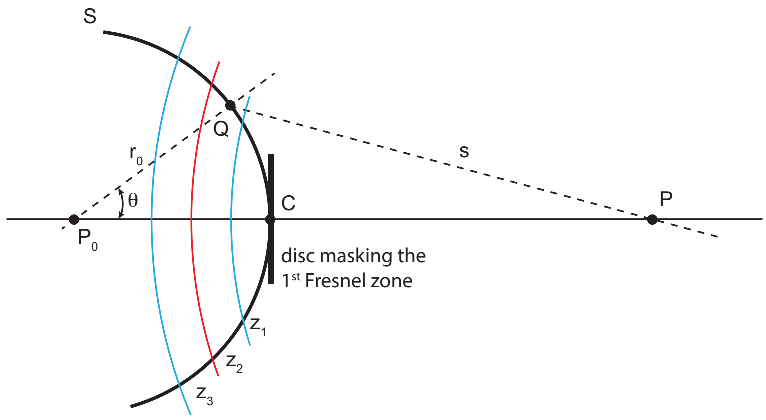

Figure 4: Fresnel construction when a circular disc masks the first Fresnel zone.

since $K_j \simeq K_{j+1}$. This leads to the disturbance at the observation point P

$$U(P) = -i\lambda k_2 K_2 \frac{Ae^{ik(r_0+b)}}{r_0+b} \simeq \frac{-Ae^{ik(r_0+b)}}{r_0+b} \tag{30}$$

Not only there is light at P, although it is in the shadow of the mask placed between S and P, but the intensity is the same as if there were no disc at all!!!

It is important to replace this calculation in its historical context. At the beginning of the 19^{th} century the French Academy of Science organised a contest in order to explain the strange phenomenon of diffraction. Thomas Young has just demonstrated his famous *double-slit* experiment and the idea that the light can propagate not always in straight line was gaining ground although most scientists were still in favor of the corpuscular theory of light introduced by Newton about a hundred years earlier. When the civil engineer Augustin-Jean Fresnel presents his theory in 1818 he is only 30 years old. His approach, based on waves rather that on corpuscular is totally new and he has to face several a jury composed by strong supporters of the Newton's theory of light: Jean-Batiste Biot, Pierre Simon Laplace and the French mathematician Siméon Denis Poisson[2]. Poisson, as a member of the jury for the contest tried to refute Fresnel's approach by inferring that the presence of a disc covering the first Fresnel zone would not yield total darkness in the shadow of the mask as it is commonly seen. Arago who had actually supported Fresnel and pushed him to present his wave theory[3] performed the experiment and found the bright spot in the middle of the shadow as suggested by eq. (29). This proves that Fresnel was right and that the wave approach to deal with light was correct. This bright spot is now either called the *Arago spot* or the *Fresnel bright spot*, but most commonly as the *Poisson spot*.

In any cases what is certain is that *diffraction phenomenon is odd and certainly not trivial*.

Number of Fresnel zones

Let assume that we have an aperture centered around C. We assume that the radius a of the aperture is much smaller that the distance from the screen to the observation point P. The screen is perpendicular to PP_0 (Fig. 5) The N^{th} Fresnel zone contained with the

[2] The other two members of the jury are Louis Joseph Gay-Lussac and Dominique-François-Jean Arago, who chairs the committee

[3] Another important scientist to push Fresnel to present his work at the contest of the French Academy of Science is André-Marie Ampère.

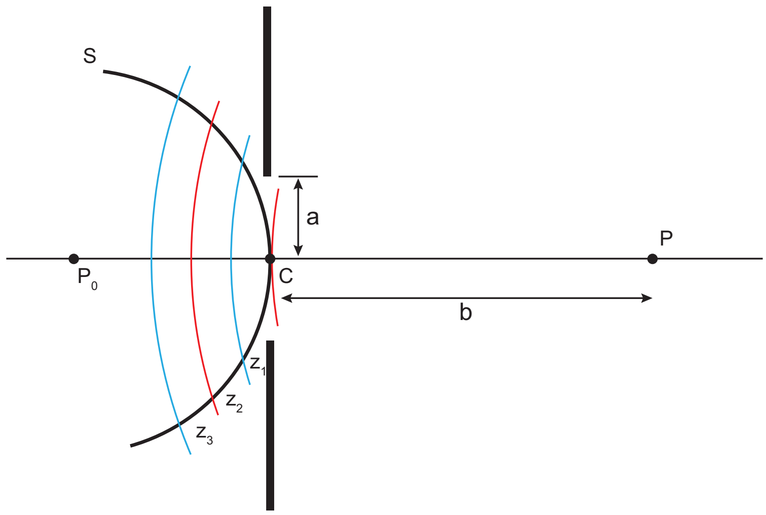

Figure 5: Calculation of the number of Fresnel zone in an optical system.

aperture has a radius roughly equal to the distance from P to the edge of the aperture. Therefore using the Pythagorean theorem we have

$$b^2 + a^2 = \left(b + N\frac{\lambda}{2}\right)^2 \tag{31}$$

And since we assumed that $a \ll b$ we can write that

$$\sqrt{b^2 + a^2} \simeq b\left[1 + \frac{1}{2}\left(\frac{a}{b}\right)^2\right]$$

which in combination with eq. (31) yields

$$N = \frac{a^2}{b\lambda} \tag{32}$$

0.2 Mathematical formulation(s) of diffraction

The idea from Augustin Fresnel that the light disturbance at any observations point P results from the superposition of all contributions from secondary was was put in a more rigorous form by Gustav Kirchhoff (1883). Mathematically Kirchhoff used the Green's theorem, which he applied to the Helmholtz wave equation. Considering a monochromatic scalar wave

$$V(x, y, z, t) = U(x, y, z)e^{-i\omega t} \tag{33}$$

we already know that such a wave is a solution of the Helmholtz equation

$$\left(\nabla^2 + k^2\right) U = 0 \tag{34}$$

where $k = (^{\omega}/_{c})$. The Green's theorem says that: "let V be a volume bounded by a closed surface S and let P be any point within it. If U as well as its 1^st^ and 2^nd^ derivative are single-valued and continuous within and on the surface S, then any other function satisfying the came continuity properties as U (Let call this function G) will also satisfy

$$\iiint_V \left(U\nabla^2 G - G\nabla^2 U\right) dV = -\iint_S \left(U\frac{\partial G}{\partial n} - G\frac{\partial U}{\partial n}\right) dS \tag{35}$$

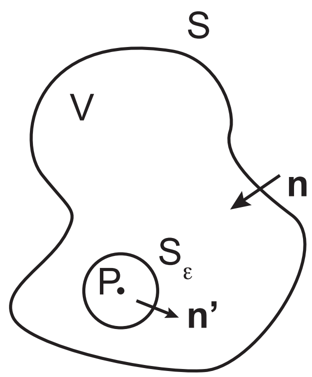

Figure 6: Definition of the vectors for the application of Green's theorem to the Helmholtz equation.

where ∂_n is a partial derivative in the inward normal direction at each point on S The difficulty in using the Green's theorem[4] in the problem of diffraction is to find a suitable Green's function G, as well as the surface S. In the present case let assume that G is also a solution of the wave equation

$$\left(\nabla^2 + k\right) G = 0 \tag{36}$$

Then the LHS of eq. (35) becomes

$$\iiint_V \left(U\nabla^2 G - G\nabla^2 U\right) dV = \iint_V -Uk^2 G + Gk^2 U dV = 0 \tag{37}$$

In this case the Green's theorem simply becomes

$$\iint_S \left(U\frac{\partial G}{\partial n} - G\frac{\partial U}{\partial n}\right) dS = 0 \tag{38}$$

To apply the Green's theorem to the diffraction problem we choose a spherical wave as the Green's function

$$G(s) = \frac{1}{s} e^{iks} \tag{39}$$

where s is the distance from a source point $P(x, y, z)$. Of course such function does not fulfill the requirement of continuity at $s = 0$ but we can circumvent this problem by excluding P from the domain of integration. The surface S becomes $S + S_\epsilon$ where S_ϵ is a sphere surrounding P (Fig. 6). Using this new surface area the eq. (38) becomes

$$\iint_S U\ \partial_n \left(\frac{1}{s} e^{iks}\right) - \frac{1}{s} e^{iks}\ \partial_n U\ dS + \iint_{S_\epsilon} U\ \partial_n \left(\frac{1}{s} e^{iks}\right) - \frac{1}{s} e^{iks}\ \partial_n U\ dS_\epsilon = 0 \tag{40}$$

[4]To understand the meaning of the Green's function, let suppose that we wish to solve an inhomogeneous linear differential equation

$$a_2 \frac{d^2 U}{dx^2} + a_1 \frac{dU}{dx} + a_0 U = V(x)$$

where $V(x)$ is a driving function. Additionally we impose a set of boundaries conditions on $U(x)$. It can be shown that if $G(x)$ is the solution of the same differential where $V(x)$ is replaced by an impulsive driving force $\delta(x - x')$ but with the same boundary conditions, then the general solution $U(x)$ can be expressed as the convolution

$$U(x) = \int G(x - x') V(x') dx'$$

It is then clear that the Green's function $G(x)$ had the form of an impulse response, which is a very similar approach as the one commonly used in electronics.

At any point (x, y, z) the derivative ∂_n yields

$$\partial_n \left(\frac{e^{iks}}{s} \right) = \mathbf{n} \cdot \nabla \left(\frac{e^{iks}}{s} \right) = \mathbf{n} \cdot \mathbf{s} \frac{\partial}{\partial s} \left(\frac{e^{iks}}{s} \right)$$
$$= \mathbf{n} \cdot \mathbf{s} \frac{e^{iks}}{s} \left(ik - \frac{1}{s} \right) = \frac{e^{iks}}{s} \left(ik - \frac{1}{s} \right) \cos(\mathbf{n}, \mathbf{s}) \tag{41}$$

where $\mathbf{n}$ is the unity vector normal to the surface and pointing inwards. Since $\mathbf{s}$ is the vector joining P to (x, y, z) on the surface S_ϵ, which is a sphere surrounding P, we simply have $\cos(\mathbf{n}, \mathbf{s}) = +1$. Using this in eq. (40) gives

$$\iint_S U \, \partial_n \left(\frac{e^{iks}}{s} \right) - \frac{e^{iks}}{s} \partial_n U \, dS = - \iint_{S_\epsilon} U \underbrace{\cos(\mathbf{n}, \mathbf{s})}_{=+1} \frac{e^{iks}}{s} \left(ik - \frac{1}{s} \right) - \frac{e^{iks}}{s} \partial_n U \, dS_\epsilon$$
$$\Leftrightarrow \iint_S U \, \partial_n \left(\frac{e^{iks}}{s} \right) - \frac{e^{iks}}{s} \partial_n U \, dS = - \iint_\Omega U \frac{e^{ik\epsilon}}{\epsilon} \left(ik - \frac{1}{\epsilon} \right) - \frac{e^{ik\epsilon}}{\epsilon} \partial_n U \epsilon^2 d\Omega \tag{42}$$

where Ω is the solid angle (Fig. 7).

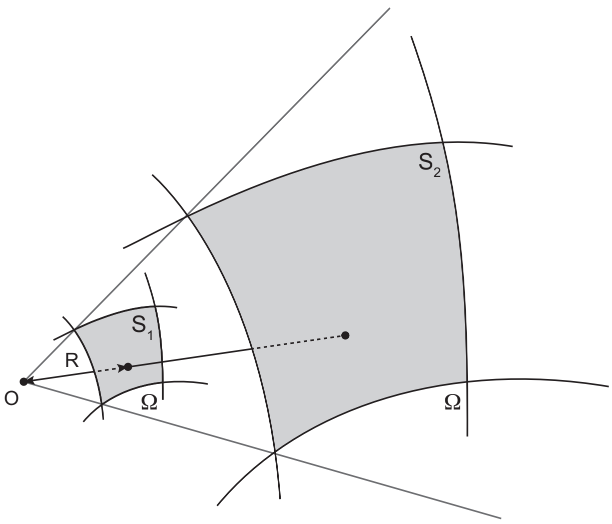

Figure 7: Definition of the solid angle Ω.

As shown on the figure although the areas of the surfaces S_1 and S_2 are different the have the same solid angle which is defined as

$$\Omega = \frac{S(R)}{R^2} \tag{43}$$

Moreover since we only want to exclude the point P from the calculation because of the singularity of the Green function at $s = 0$ we need to take the limit

$$\lim_{\epsilon \to 0} \iint_\Omega U \frac{e^{ik\epsilon}}{\epsilon} \left(ik - \frac{1}{\epsilon} \right) - \frac{e^{ik\epsilon}}{\epsilon} (\partial_n U) \, \epsilon^2 d\Omega$$
$$= \lim_{\epsilon \to 0} \iint_\Omega \underbrace{U \epsilon e^{ik\epsilon} ik}_{\hookrightarrow 0} - \underbrace{U e^{ik\epsilon}}_{\hookrightarrow U} - \underbrace{\epsilon e^{ik\epsilon} \partial_n U}_{\hookrightarrow 0} \, d\Omega = - \iint U d\Omega = -4\pi U \tag{44}$$

And using on eq. (41) yields the disturbance at the observation point P:

$$\boxed{U(P) = \frac{1}{4\pi} \iint_S \left[U \partial_n \left(\frac{e^{iks}}{s} \right) - \frac{e^{iks}}{s} \partial_n U \right] dS} \tag{45}$$

This is the *Helmholtz-Kirchhoff integral theorem.*

0.2.1 Fresnel-Kirchhoff's diffraction formula

The previously derived Helmholtz-Kirchhoff integral formula (HKiF) embodies the idea from the Huygens-Fresnel formulation of the diffraction problem, but in this form the different surface element are more complicated to evaluate that Fresnel originally assumed. Kirchhoff that actually in many cases we can make approximations to reduce the HKiF into a much simpler formulation. Let's assume that we have a screen between the source point P_0 and the observation point P.

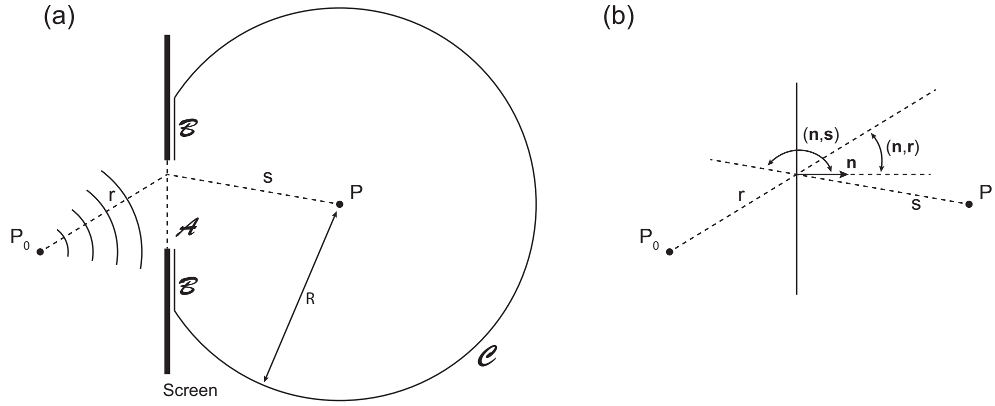

Figure 8: Construction for calculating the Fresnel-Kirchhoff diffraction integral. As previously $\mathbf{n}$ is the normal to the surface $\mathcal{C}$ and it is defined inwards.

According to the HKiF (eq. (45)) the disturbance at the observation point P results from the contributions for the three surfaces $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$.

$$U(P) = \frac{1}{4\pi}\left\{\iint_{\mathcal{A}}\left[U\partial_n\left(\frac{e^{iks}}{s}\right) - \frac{e^{iks}}{s}\partial_n U\right]dS + \iint_{\mathcal{B}}(\dots)\,dS + \iint_{\mathcal{C}}(\dots)\,dS\right\} \tag{46}$$

The problem to calculate these integrals is that the values of U and $\partial_n U$ on the respective areas $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are not know exactly! We then need assumptions which are the following ones:

- on $\mathcal{A}$, the value is assumed to be the same as if the were no screen.
- on $\mathcal{B}$, the value is null.
- since $\mathcal{C}$ is chosen arbitrarily, we can take a sphere with a radius of curvature that is so large that light has not yet reached $\mathcal{C}$ at the time of interest, *i.e.* when the light is at P, since $\mathcal{C}$ is much farther away from the aperture than P is.

In these conditions we have

- on $\mathcal{A}$:

$$U = U^{(i)} = \frac{Ae^{ikr}}{r} \quad \text{since this is a spherical wave coming from } P_0$$
$$\frac{\partial U}{\partial n} = \frac{\partial U^{(i)}}{\partial n} = \frac{Ae^{ikr}}{r}\left(-\frac{1}{r} + ik\right)\cos(\mathbf{n},\mathbf{r}) \quad \text{as was done before}$$

- on $\mathcal{B}$, $U = 0$ and $\partial_n U = 0$
- same on $\mathcal{C}$.

As a consequence the integral (eq. (46)) becomes

$$U(P) = \frac{1}{4\pi}\iint_{\mathcal{A}} \frac{Ae^{ikr}}{r}\frac{e^{iks}}{s}\left(ik - \frac{1}{s}\right)\cos(\mathbf{n},\mathbf{s}) - \frac{e^{iks}}{s}\left(ik - \frac{1}{r}\right)\frac{Ae^{ikr}}{r}\cos(\mathbf{n},\mathbf{r})\,dS \tag{47}$$

and since the wave-number is $k = ({}^{2\pi}/_{\lambda})$ is λ is small with respect to s and r we have

$$ik - \frac{1}{s} \simeq ik$$
$$ik - \frac{1}{r} \simeq ik$$

and finally we obtain the **Fresnel-Kirchhoff** formula:

$$\boxed{U(P) = \frac{iA}{2\lambda} \iint_{\mathcal{A}} \frac{e^{ik(r+s)}}{rs} \left[\cos(\mathbf{n}, \mathbf{s}) - \cos(\mathbf{n}, \mathbf{r})\right] \; dS} \tag{48}$$

We should point that this expression can only be applied when the source is a single point. Moreover the derivation assumes that the aperture is illuminated by one expanding spherical wave. Such restriction can, fortunately, be removed by the *Rayleigh-Sommerfeld* theory.

Note on the inclination factor from Fresnel: Let assume that instead of the aperture $\mathcal{A}$ we consider a portion $\mathcal{W}$ of the incident wave. We are then in the same construction as Fresnel did prior the insertion of a screen. If the radius of curvature of the wave is sufficiently large the contribution from $\mathcal{C}$ – the part that comes from P_0 – is negligible. Moreover on $\mathcal{W}$ we have $\cos(\mathbf{n}, \mathbf{r_0}) = 1$. Introducing the inclination factor as Fresnel did by $\chi = \pi - (\mathbf{r_0}, \mathbf{s})$ leads to

$$\cos(\mathbf{n}, \mathbf{s}) - \cos(\mathbf{n}, \mathbf{r}) = -(1 + \cos\chi) \tag{49}$$

and the Fresnel-Kirchhoff integral becomes

$$U(P) = \frac{-i}{2\lambda} \frac{Ae^{ikr_0}}{r_0} \iint_{\mathcal{W}} \frac{e^{iks}}{s} (1 + \cos\chi) \; dS \tag{50}$$

This is in agreement with Fresnel's original formulation if the contribution from the element dW of the wavefront is

$$\frac{-i}{2\lambda} \frac{Ae^{ikr_0}}{r_0} \frac{e^{iks}}{s} (1 + \cos\chi) \tag{51}$$

which yields the inclination factor

$$K(\chi) = \frac{-i}{2\lambda} (1 + \cos\chi) \tag{52}$$

What is remarkable is that for the central zone, where $\chi = 0$, $K(\chi) = K_1 = ({}^{-i}/_{\lambda})$. This is exactly what Augustin Fresnel assumed. Note on the other hand that his second assumption $K\left({}^{\pi}/_{2}\right) = 0$ is not true.

Physical interpretation The Fresnel-Kirchhoff integral can be written slightly differently by[5]

$$U(P) = \frac{1}{i\lambda} \iint_{\mathcal{A}} U_\Sigma \frac{e^{iks}}{s} \psi \; dS = \iint_{\mathcal{A}} U'_\Sigma \frac{e^{iks}}{s} \; dS \tag{53}$$

[5] If we compare this expression to the eq. (48) we readily see that $2\psi = \cos(\mathbf{n}, \mathbf{s}) - \cos(\mathbf{n}, \mathbf{r})$ which is directly linked with the inclination factor $K(\chi)$ introduced by Fresnel.

where U_Σ is the disturbance at the aperture and ψ is directly related to the inclination factor. Since the

$$U'_\Sigma = \psi U_\Sigma = \left(\frac{1}{i\lambda}\psi\right)\frac{Ae^{ikr}}{r} \tag{54}$$

The interpretation of the Fresnel-Kirchhoff integral is now obvious: the disturbance at P results from an infinite number of secondary waves (U'_Σ) located at the aperture itself and which are propagating towards P.

0.2.2 Rayleigh-Sommerfeld theory

For the Fresnel-Kirchhoff theory of diffraction, the chosen Green-function was a spherical wave emitted from the source point P_0. As we saw, this give a reasonable result and it was actually used a lot as it gave accurate results and explained many experimental observation. However it also contains a few inconsistencies that motivate other scientist for find a more suitable form of the diffraction integral. One important inconsistency if that on the screen the field and its derivative was null (See conditions on p. 11). Mathematically such condition yields that the filed is null *everywhere*, which is obviously inconsistent with the rest of the calculation. The idea from Sommerfeld was to use a different Green-function which would be null on the screen but different from zero elsewhere. His idea was to imagine a Green-function generated by the source point P_0 but also from a point source $\tilde{P}_0$, which is the image of P_0 mirrored by the plane of the aperture (Fig. 9b).

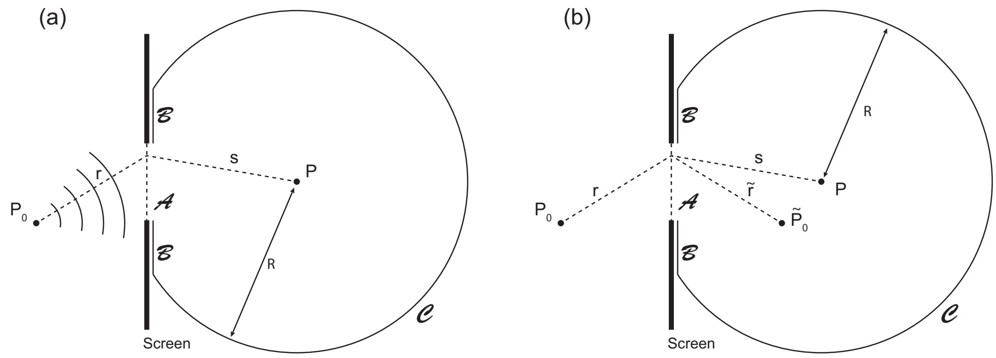

Figure 9: Construction for calculating the Fresnel-Kirchhoff as was shown before (a) and the Sommerfeld diffraction integral.

Since the Green function now consists of two spherical waves, one emitted by P_0 and the other from $\tilde{P}_0$ it can be written as two from $G_\pm$:

$$G_- = \frac{1}{r}\,e^{ikr} - \frac{1}{\tilde{r}}e^{ik\tilde{r}} \tag{55a}$$

$$G_+ = \frac{1}{r}\,e^{ikr} + \frac{1}{\tilde{r}}e^{ik\tilde{r}} \tag{55b}$$

Using G_-, the field vanishes on the plane of aperture and the conditions p. 11 only need to be applied for U. Using the general formulation of the disturbance at the observation point P_0 (eq. (23))

$$U(P) = \frac{1}{4\pi}\iint_S \left[U\partial_n G - G\partial_n U\right]\ dS \tag{56}$$

the disturbance at the observation point becomes

$$U_I(P) = \frac{-1}{4\pi} \iint_S G_- \partial_n U \; dS \tag{57}$$

This is called the *first Rayleigh-Sommerfeld* solution. If on the other hand we use the other version of the Green function G_+ we obtain

$$U_{II}(P) = \frac{1}{4\pi} \iint_S U \partial_n G_+ \; dS \tag{58}$$

From these two versions of the disturbance at the observation point, we can do the calculation as we did for the Fresnel-Kirchhoff integral (section0.2.1) we would obtain

$$U_I(P) = \frac{iA}{\lambda} \iint_{\mathcal{A}} \frac{e^{ik(r+s)}}{rs} \cos(\mathbf{n}, \mathbf{s}) \tag{59a}$$

$$U_{II}(P) = -\frac{iA}{\lambda} \iint_{\mathcal{A}} \frac{e^{ik(r+s)}}{rs} \cos(\mathbf{n}, \mathbf{r}) \tag{59b}$$

As we see from these equation and the Fresnel-Kirchhoff integral we can always write the disturbance as the form

$$U(P) = \frac{iA}{\lambda} \iint_{\mathcal{A}} \frac{e^{ik(r+s)}}{rs} \psi dS \tag{60}$$

where the inclination factor is given by

$$K(\chi) = \begin{cases} (1/2)\left[\cos(\mathbf{n}, \mathbf{s}) - \cos(\mathbf{n}, \mathbf{r})\right] & \text{Kirchhoff approach} \\ \cos(\mathbf{n}, \mathbf{s}) & 1^{st} \text{ Rayleigh-Sommerfeld approach} \\ -\cos(\mathbf{n}, \mathbf{r}) & 2^{nd} \text{ Rayleigh-Sommerfeld approach} \end{cases} \tag{61}$$

Note that the Kirchhoff solution is the arithmetic average of the two Rayleigh-Sommerfeld solutions. At this stage we can wonder which expression to use. Actually, they all give very similar results as long as the aperture is large with respect to the wavelength. However, when we consider that the point source is very far from the aperture the angle $(\mathbf{n}, \mathbf{r} \approx \pi$ and therefore the inclination factor becomes:

$$K(\chi) = \begin{cases} (1/2)\left[\cos\theta + 1\right] & \text{Kirchhoff approach} \\ \cos\theta & 1^{st} \text{ Rayleigh-Sommerfeld approach} \\ 1 & 2^{nd} \text{ Rayleigh-Sommerfeld approach} \end{cases} \tag{62}$$

where $(\mathbf{n}, \mathbf{s}) = \theta$. For simplicity the 1^{st} Rayleigh-Sommerfeld approach is usually taken.

0.2.3 Fresnel and Fraunhofer diffraction

Fresnel diffraction and limit

As we previously derived the Fresnel-Kirchhoff diffraction integral is

$$U(P) = \frac{iA}{2\lambda} \iint_{\mathcal{A}} \frac{e^{ik(r+s)}}{rs} \left[\cos(\mathbf{n}, \mathbf{s}) - \cos(\mathbf{n}, \mathbf{r})\right] \; dS \tag{63}$$

Obviously numerical solution of this integral are possible but it would be better if we could find analytical solutions. The figure 10 presents a possible configuration in which we need to use the Fresnel-Kirchhoff integral in order to evaluate the disturbance at the observation point P. The source is placed at P_0 at a distance r' from the aperture. $Q(\xi, \eta)$ is a point of the aperture.
Note that

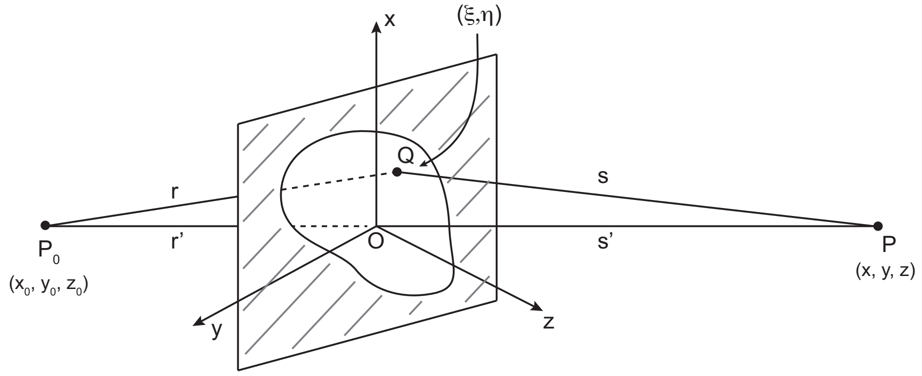

Figure 10: Construction to calculate the Fresnel-Kirchhoff integral. Note that the normal to the screen does not necessarily coincide with the axis defined by P_0P.

- The dimension of the aperture is much larger than the considered wavelength. As a consequence as dS explores the whole domain of integration the term $(r+s)$ varies a lot more than the wavelength, and therefore $\exp[i(r+s)]$ oscillates rapidly.
- If P_0 and P are far from the aperture – compared to the dimension of the aperture – then $\cos(\mathbf{n},\mathbf{s}) - \cos(\mathbf{n},\mathbf{r})$ does not vary much. Moreover is P_0 and P are far from the aperture, we can choose any point of the aperture O such that the angle (P_0O, P_0P) and (OP, P_0P) remain small. Within this assumption $\cos(\mathbf{n},\mathbf{s}) - \cos(\mathbf{n},\mathbf{r}) \approx -2\cos\delta$, where δ is the angle between P_0P and the normal to the aperture.
- If P_0 and P are far from the aperture, then $(^1/_{rs})$ can be replaced by $(^1/_{r's'})$

With these conditions the Fresnel-Kirchhoff integral becomes:

$$U(P) = \frac{-iA}{\lambda}\frac{\cos\delta}{r's'}\iint_{\mathcal{A}} \exp\left[ik(r+s)\right]\ dS \tag{64}$$

Let now express the quantities r, r', s and s' in Cartesian coordinates. The point Q has the coordinates $(\xi,\eta,0)$. Therefore:

$$|P_0Q|^2 = |(\xi,\eta,0)-(x_0,y_0,z_0)|^2 = r^2 = (x_0-\xi)^2+(y_0-\eta)^2+z_0^2 \tag{65a}$$
$$r'^2 = x_0^2+y_0^2+z_0^2 \tag{65b}$$
$$s^2 = (x-\xi)^2+(y-\eta)^2+z^2 \tag{65c}$$
$$s'^2 = x^2+y^2+z^2 \tag{65d}$$

$$\Rightarrow \begin{cases} r^2 &= r'^2 - 2(x_0\xi+y_0\eta)+\xi^2+\eta^2 \\ s^2 &= s'^2 - 2(x\xi+y\eta)+\xi^2+\eta^2 \end{cases} \tag{66}$$

and if we assume that the dimension of the aperture are small compare to r' and s' then we can do a Taylor expansion as

$$\begin{cases} r^2 &\approx r'^2 - \dfrac{x_0\xi+y_0\eta}{r'} + \dfrac{\xi^2+\eta^2}{2r'} - \dfrac{(x_0\xi y_0\eta)^2}{2r'^3} - \dots \\ s^2 &\approx s'^2 - \dfrac{x\xi+y\eta}{s'} + \dfrac{\xi^2+\eta^2}{2s'} - \dfrac{(x\xi y\eta)^2}{2s'^3} - \dots \end{cases} \tag{67}$$

Then we have the disturbance at the observation point

$$U(P) = \frac{-i\cos\delta}{\lambda}\frac{Ae^{ik(r'+s')}}{r's'}\iint_{\mathcal{A}} e^{ikf(\xi,\eta)}\, d\xi d\eta \tag{68}$$

$$\text{with}\quad f(\xi,\eta) = -\frac{x_0\xi + y_0\eta}{r'} - \frac{x\xi + y\eta}{s'} + \frac{\xi^2+\eta^2}{2r'} + \frac{\xi^2+\eta^2}{2s'} - \frac{(x_0\xi y_0\eta)^2}{2r'^3} - \frac{(x\xi y\eta)^2}{2s'^3}\ \dots$$

If we introduce the direction cosines

$$\ell_0 = -\frac{x_0}{r'};\quad \ell = \frac{x}{s} \tag{69a}$$

$$m_0 = -\frac{y_0}{r'};\quad m = \frac{y}{s}$$

then the function $f(\xi,\eta)$ becomes

$$\begin{aligned} f(\xi,\eta) &= (\ell_0-\ell)\xi + (m_0-m)\eta \\ &\quad + \underbrace{\frac{1}{2}\left[\left(\frac{1}{r'}+\frac{1}{s'}\right)\left(\xi^2+\eta^2\right) - \frac{\left(\ell_0\xi+m_0\eta\right)^2}{r'} - \frac{\left(\ell\xi+m\eta\right)^2}{s'}\right] + \dots}_{\text{if negligible, we have Fraunhofer diffraction, otherwise Fresnel diffration}} \end{aligned} \tag{70}$$

Obviously the integral (eq. (68)) is much easier to evaluate in the case where the quadratic term in ξ^2 and η^2 (as well as the higher-order term) are negligible Rigorously this situation only happens when $r' \to \infty$ and $s' \to \infty$ corresponding to a source and the observation screen located at an infinite distance from the diffracting object. This situation is referred as *Fraunhofer diffraction*. Let's look at the quadratic term. In the Fresnel-Kirchhoff integral (eq. (68)) this appears in the complex exponential. This contribution of this term is therefore negligible if

$$\frac{1}{2}k\left[\left(\frac{1}{r'}+\frac{1}{s'}\right)\left(\xi^2+\eta^2\right) - \frac{\left(\ell_0\xi+m_0\eta\right)^2}{r'} - \frac{\left(\ell\xi+m\eta\right)^2}{s'}\right] \ll 2\pi \tag{71}$$

It can be shown that this is equivalent to the following conditions:

$$\frac{(\xi^2+\eta^2)_{\text{max.}}}{r'} \ll \lambda \tag{72a}$$

$$\frac{(\xi^2+\eta^2)_{\text{max.}}}{s'} \ll \lambda \tag{72b}$$

By contrast with the Fraunhofer diffraction if we cannot neglect the higher-order terms of the expansion in $f(\xi,\eta)$ we speak of *Fresnel diffraction*.

How can we make the distinction between both regimes? Assuming a circular aperture with a radius a, the conditions eq. (72a) and eq. (72b) are equivalent to

$$|s'| \gg \frac{(\xi^2+\eta^2)_{\text{max.}}}{\lambda} \ \Leftrightarrow\ |s'| \gg \frac{a^2}{\lambda} \tag{73}$$

At this stage we should remind that when we evaluated the number of Fresnel zones (see sec. 0.1.2, p. 7), the parameter s' was then called b. Calculating the number of Fresnel zones is then a good way to evaluate whether we are in Fresnel of in the Fraunhofer diffraction:

$$N_F = \frac{a^2}{b\lambda} \ll 1 \quad \text{Fraunhofer diffraction} \tag{74a}$$

$$N_F \gtrsim 1 \quad \text{Fresnel diffraction} \tag{74b}$$

On Figure 11, we evaluated the full form of the integral (eq. (68)) for different value of the Fresnel number N_F to calculate the diffraction pattern from a single slit. For this figure we keep the product $b\lambda$ constant and vary the size of the slit a. The size of the slit is indicated by the shaded area on each sub-graph. As we see the Fresnel number influences significantly the final result of the calculation! We should also note that the *transition* does not appear around $N_F = 1$ but actually before this. This is the reason why the condition for Fraunhofer diffraction corresponds to $b \gg (a^2/\lambda)$ and not simply $b > (a^2/\lambda)$. As an example, let's assume that we have a 1 mm diameter aperture illuminated with

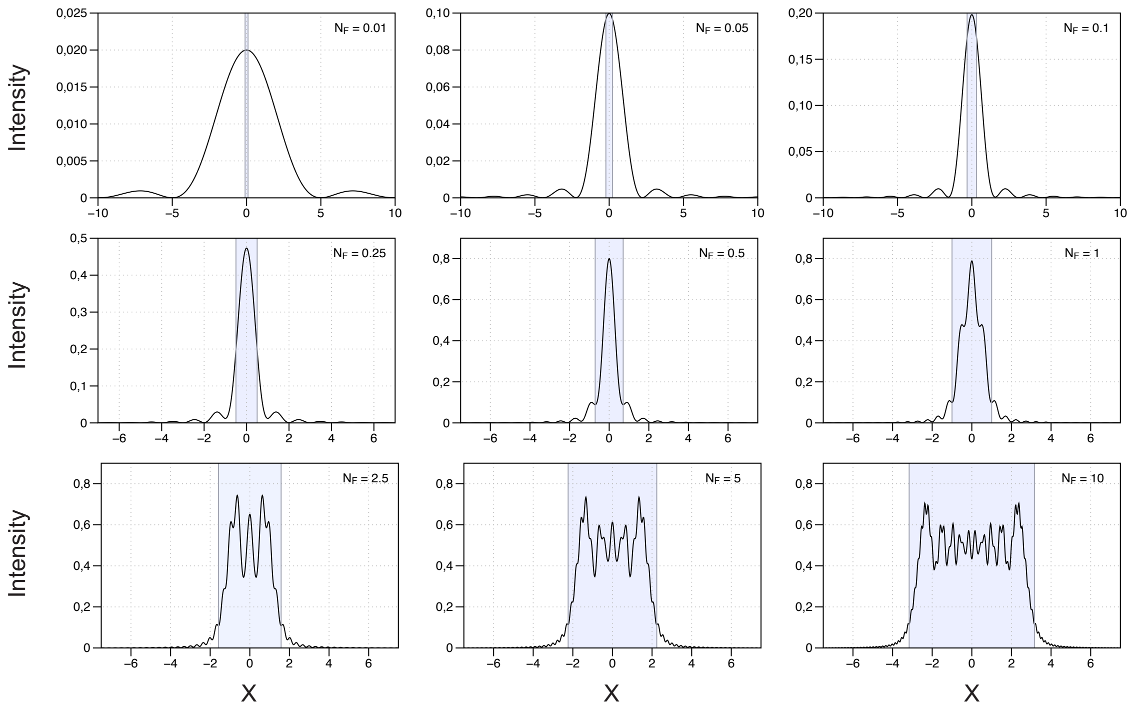

Figure 11: Evolution of the diffraction pattern depending on the Fresnel number N_F. On each graphs the shaded area represents the size of the slit. Note that the X-axis is not the same on the graphs on the top row.

a laser working at $\lambda = 532$ nm. Where to put the screen to be sure that we are in the Fraunhofer diffraction?
We need to calculate $(a^2/\lambda) = \left((0.5 \times 10^{-3})^2/532 \times 10^{-9}\right) = 47$ *cm. The screen has to be a distance much larger than this!*

Fraunhofer diffraction

Let call $p = \ell - \ell_0$ and $q = m - m_0$. By doing so, the effect of a change of the aperture will not affect the diffraction pattern as long as this shift is in the plane of the aperture.

Then the disturbance at the observation P calculated from eq. (68)) becomes[6]

$$U(P) = U(p,q) = C \iint_{\mathcal{A}} e^{-ik(p\xi + q\eta)} \, d\xi d\eta \tag{75}$$

and C is a constant that takes into account the pre-factor in eq. (68). If we consider that the process of diffraction is lossless then we can calculate the constant C from the intensity at the center of the aperture $I_0 = |U(0,0)|^2$. and obviously

$$U(0,0) = C \iint_{\mathcal{A}} d\xi d\eta = A \times C. \tag{76}$$

where A is the area of the aperture. The constant in the Fraunhofer integral (eq. (75)) is then simply

$$C = \frac{\sqrt{I_0}}{\text{Area of the aperture}} \tag{77}$$

We can also notice that besides the boundaries of the Fraunhofer integral (boundaries of the aperture), the expression for $U(p,q)$ looks very much like a 2−dimension Fourier transform. If is actually possible to slightly modify this expression to have a perfect analogy with a Fourier integral:

$$U(p,q) = C \iint_{-\infty}^{+\infty} G(\xi,\eta) e^{-\frac{i2\pi}{\lambda}(p\xi + q\eta)} \, d\xi d\eta \tag{78}$$

$$\text{where} \begin{cases} G(\xi,\eta) = 1 & \text{everywhere in the opening} \\ G(\xi,\eta) = 0 & \text{elsewhere} \end{cases}$$

The function $G(\xi,\eta)$ is called the **pupil function**.

[6]Of course if we assume that we are in the conditions for Fraunhofer diffraction, only the first term of $f(\xi,\eta)$ is conserved. Note moreover that the minus sign in the exponential comes from the definition of p and q and the form of the function $f(\xi,\eta)$ given at eq. (70).

Bibliography

[1] Max Born and Emil Wolf. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light (7th Edition).* Cambridge University Press, 7th edition, 1999.

[2] J.W. Goodman. *Introduction to Fourier Optics.* McGraw-Hill physical and quantum electronics series. W. H. Freeman, 2005.

[3] John David Jackson. *Classical electrodynamics.* Wiley, New York, NY, 3rd ed. edition, 1999.