DS 12:
Fraunhofer Diffraction

Philip Cherian
April 15, 2022

1 Introduction

Let us start by considering the setup of a diffraction experiment: an incoming plane wave, an obstacle (usually a screen with slits or holes in it) which "diffracts" this wave, and a screen on which the result is observed. The standard description of this phenomenon is usually based on "Huygens' Principle": diffracted light is the sum of secondary wavelets from the unblocked incoming light.

When Huygen's proposed this principle, he was able to qualitatively describe the propagation of linear and spherical waves. However, the mathematical tool that he would have needed to make it more quantitative – integral calculus – was not yet well-established at the time (Huygens' proposed his theory in 1678).[1] The rigorous mathematical description of diffraction would have to wait until 1818, when Augustin-Jean Fresnel began to take an interest in the diffraction of light.

In what follows, we will more-or-less be using Fresnel's approach to diffraction. We will take the polarisation of the incoming wave to be fixed (the first of many approximation we will both make and criticise later), and we will concentrate on the amplitude and phase. As a result, we will be describing our electromagnetic wave using a complex *scalar* ψ. We set up our coordinate axes as shown in Figure (1). We consider a monochromatic incident wave (of wavelength λ) propagating along the $z'z-$axis. Points on the diffracting plane (i.e., the plane containing the obstacle or aperture) will be denoted by primed coordinates (x', y'), while those on the screen will be denoted by unprimed coordinates (x, y). Additionally, we choose our z and $z'-$axis such that the diffracting plane lies at $z' = z = 0$, while the observing plane lines at some z.

We now make our most important claim: the wave at any point $P(x, y, z)$ on the screen receives contributions from the wave at every point $P'(x', y', 0)$ on the diffracting plane. Since we are in the linear regime, the principle of superposition holds, and so we can say that

$$\psi(x, y, z) \propto \iint_\Gamma \mathrm{d}x'\mathrm{d}y'\ \psi(x', y', 0)e^{ikr}, \tag{1}$$

where $\psi(x', y', 0)$ denotes the field at some point $P(x', y')$ on the diffracting plane, and we have integrated over all these points to get the resulting contribution at the point $P(x, y)$. The integral is computed over the area of the aperture, which we denote by Γ. The quantity r is the distance between the points (x, y, z) and $(x', y', 0)$, and this shows that the phase that goes into the integral is also a function of the coordinates x' and y'. As we shall see, this is crucial to our results.

In many interesting situations, this distance r is often very close to z. This is equivalent to saying that the angle between PP' and the $z-$axis is very small. This is because of a combination of factors: the

[1] Coincidentally, however, in 1672 – only 6 years before he proposed this theory – Huygens' had met with the famous polymath Leibniz and convinced him to devote a significant amount of his time to Mathematics. Leibniz would later go on to develop much of calculus as we know it today.

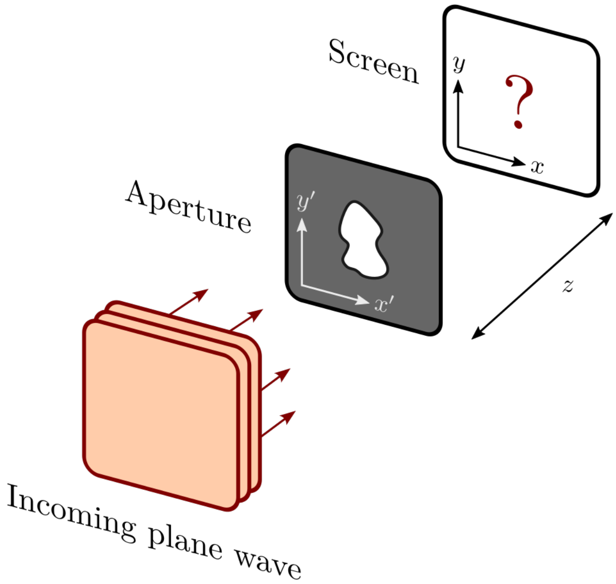

Figure 1: A basic diffraction setup. The incoming wave is represented by plane wavefronts that impinge on the aperture in the diffracting plane. The result is a pattern that is then observed on the screen. The diffracting and observation planes are denoted by primed and unprimed coordinates respectively, and the coordinates are chosen such that the diffracting plane occurs at $z = z' = 0$.

screen is often quite far away from the diffracting plane, compared to the size of the aperture, and the diffraction pattern obtained often falls off in intensity quite fast as you move away from $(x, y) = (0, 0)$. This "small-angle" approximation for wave propagation, known as the **paraxial approximation**, can simplify the exponential in the above integral, since we can write

$$\begin{aligned} r &= \sqrt{(x-x')^2 + (y-y')^2 + z^2}, \\ &= z\sqrt{1 + \frac{(x-x')^2 + (y-y')^2}{z^2}}, \\ &\approx z\left(1 + \frac{1}{2}\frac{(x-x')^2 + (y-y')^2}{z^2}\right), \\ &\approx z + \frac{(x-x')^2 + (y-y')^2}{2z}, \end{aligned} \tag{2}$$

where we use the binomial approximation to simplify the square-root. As can be seen, the paraxial approximation implies that we are working in the regime where the second term in the last equation above is small compared to z. This approximation can thus be used to simplify the exponential in our original equation, since now:

$$\exp(ikr) = \exp(ikz) \times \exp\left(ik\left(\frac{(x-x')^2 + (y-y')^2}{2z}\right)\right) \tag{3}$$

The first of the two exponentials to the right represents an overall phase-change of kz, due the distance of the screen from the aperture.

Using this fact, we can simplify Equation (1) to

$$\psi(x,y,z) = \mathcal{N} \iint_\Gamma \mathrm{d}x'\mathrm{d}y'\ \psi(x',y',0)\ e^{ikz}\ \exp\left(ik\frac{(x-x')^2}{2z}\right)\exp\left(ik\frac{(y-y')^2}{2z}\right), \tag{4}$$

where $\mathcal{N}$ is a normalisation constant.

1.1 Calculating the normalisation constant $\mathcal{N}$

In order to compute $\mathcal{N}$, we use the following thought-experiment. Let us assume that we have an aperture in the diffracting plane, Γ. Now, imagine that we made this aperture infinitely large: this would be equivalent to saying that there was no aperture at all! What would we expect to see on the screen in this case? Well, if there were no aperture, and we had an incident plane wave, then we should have

$$\psi(x',y',0) = A\ e^{ikz}\Big|_{z=0} = A, \tag{5}$$

where A is some overall constant. Notice that since we had already chosen our axes such that $z = 0 = z'$, our plane wave is just the constant A over the entire diffracting plane. Now, what would we expect to happen at the screen? Well, the as the plane wave propagates along the $z-$axis, the only thing that happens is that it picks up an additional phase of $\exp(ikz)$. As a result, we would expect,

$$\psi(x,y,z) = Ae^{ikz}. \tag{6}$$

We can plug this into Equation (4) to find the normalisation constant $\mathcal{N}$. This gives us

$$Ae^{ikz} = \mathcal{N}\int_{-\infty}^{\infty}\mathrm{d}x'\int_{-\infty}^{\infty}\mathrm{d}y'\ A\ e^{ikz}\ \exp\left(ik\frac{(x-x')^2}{2z}\right)\exp\left(ik\frac{(y-y')^2}{2z}\right). \tag{7}$$

Notice that since we have assumed an aperture that is infinitely large (i.e. Γ is the entire $x'y'-$plane), the integral runs over the entire two-dimensional plane. This makes the integral (slightly) easier to calculate. Removing the factors of $A\exp(ikz)$ from both sides of the equation and rearranging it, you should be able to see that

$$\mathcal{N}\int_{-\infty}^{\infty}\mathrm{d}x'\ \exp\left(ik\frac{(x-x')^2}{2z}\right)\int_{-\infty}^{\infty}\mathrm{d}y'\ \exp\left(ik\frac{(y-y')^2}{2z}\right) = 1. \tag{8}$$

Exercise: Notice how the two integrals above are nearly identical. Show, using appropriate substitutions, that you can write the above equation as

$$\mathcal{N} g^2 = 1, \tag{9}$$

where

$$g = \int_{-\infty}^{\infty} \mathrm{d}u \, \exp\left(i \frac{ku^2}{2z} \right) \tag{10}$$

Exercise: The integral that defines g looks very much like a Gaussian integral. You should all remember how to perform a Gaussian integral, but just in case you don't, show that

$$\int_{-\infty}^{\infty} \mathrm{d}u \, e^{-\alpha u^2} = \sqrt{\frac{\pi}{\alpha}}. \tag{11}$$

Exercise: The Gaussian integral is usually defined for real values of α. However, it can be shown that it also works for complex α, provided that $\mathrm{Re}(\alpha) > 0$. In our case,

$$\alpha = -i \left(\frac{k}{2z} \right), \tag{12}$$

and so $\mathrm{Re}(\alpha) = 0$. We are thus pushing the identity a little beyond its limit of validity. However, it turns out that we can do this. Ignore this issue and show that

$$g = \sqrt{\frac{2\pi i z}{k}}, \tag{13}$$

and therefore that

$$\mathcal{N} = -\frac{ik}{2\pi z} \tag{14}$$

1.2 The inverse-square law

Let's take a couple of minutes to examine the factor of z in the denominator of $\mathcal{N}$. This factor essentially means that – from Equation (4) – the amplitude of the field at $P(x, y, z)$ falls off as

$$\psi(x, y, z) \sim \frac{1}{z} \qquad \Longleftrightarrow \qquad \text{intensity} \equiv |\psi(x, y, z)|^2 \sim \frac{1}{z^2}. \tag{15}$$

This shows that if we ignore the effects of the term in the integral, the intensity of light on the screen varies as the inverse-square of the distance to the screen. This should be a familiar result, and we will describe why this occurs. To see why, remember that we're working in the paraxial approximation. We will rewrite approximation in terms of the characteristic size D of the aperture. Intuitively, you could think of D as being the radius of a circle that just contains the entire aperture. If you're more mathematically inclined, you could define it as something like

$$D^2 = \max(x'^2 + y'^2), \qquad x', y' \in \Gamma, \tag{16}$$

but don't get too bogged down by the details. Suffice it to say that D defines the size of the aperture, and is an order-of-magnitude quantity, so factors of 2 or π or so on are unimportant. Another way of describing the paraxial approximation is to say that

$$\frac{D}{z} \ll 1 \qquad \Longleftrightarrow \qquad D \ll z, \tag{17}$$

the size of the aperture is much smaller than the distance to the screen. In this case, if we go sufficiently far enough, our aperture would start to behave like a point source, which is well known to follow the inverse square law for intensity in three-dimensions.

2 The far-field regime

Using our result for $\mathscr{N}$, Equation (4) becomes

$$\psi(x,y,z) = -\frac{ik}{2\pi z}\iint_\Gamma \mathrm{d}x'\mathrm{d}y'\; A\, e^{ikz}\, \exp\left(ik\frac{(x-x')^2}{2z}\right)\exp\left(ik\frac{(y-y')^2}{2z}\right). \tag{18}$$

(Note how we have used $\psi(x',y',0) = A$, since our incident wave is a plane wave.) This integral is still very hard to compute since it is over the domain Γ, and thus we need to make one further assumption to simplify the terms in the exponential. Expanding out the exponentials can collecting the terms, we have:

$$\exp\left(ik\frac{(x-x')^2}{2z}\right)\exp\left(ik\frac{(y-y')^2}{2z}\right) = \underbrace{\exp\left(ik\frac{(x^2+y^2)}{2z}\right)}_{(*)}\underbrace{\exp\left(ik\frac{(x'^2+y'^2)}{2z}\right)}_{(**)}\underbrace{\exp\left(-ik\frac{(xx'+yy')}{z}\right)}_{(***)}. \tag{19}$$

Let us now examine these three terms one after the other. The first term depends only on the position on the screen, $P(x,y)$, and therefore does not affect the overall integral. In fact, it is possible to show that if the screen were the surface of a sphere of radius z, this term would not appear at all. We will therefore ignore its contribution.

The second term is more problematic. We know how the compute quadratic exponentials when the domain of integration is over the entire real line, but when the domain is of some finite size, it is not possible to find a closed-form solution. We would thus like to eliminate this term, and to do this, we invoke another approximation, that we are working in the **far-field** or **Fraunhofer** regime, i.e. that

$$\frac{kD^2}{2z} \ll 1 \qquad \Longleftrightarrow \qquad \frac{D}{z} \ll \frac{2}{kD} = \frac{\lambda}{\pi D}. \tag{20}$$

Notice how this approximation (unlike the paraxial one) also includes the wavelength of the incident light λ.

Thus, the only term that remains is the third one. We will now try to simplify this term. First, recall that the wavevectors $\mathbf{k}$ from each different point on the aperture have the same magnitude, but different unit vectors. Thus,

$$\mathbf{k} = k\hat{\mathbf{n}}, \qquad \text{where} \qquad \hat{\mathbf{n}} = \frac{x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}}{r} \quad \approx \quad \left(\frac{x}{z}\right)\hat{\mathbf{x}} + \left(\frac{y}{z}\right)\hat{\mathbf{y}} + \hat{\mathbf{z}}, \tag{21}$$

where in the last step we have used the fact that $r \approx z$. From the above result, it should be clear that since

$$\mathbf{k} = k_x\hat{\mathbf{x}} + k_y\hat{\mathbf{y}} + k_z\hat{\mathbf{z}}, \tag{22}$$

we have

$$k_x \approx \frac{kx}{z}, \quad k_y \approx \frac{ky}{z}, \quad k_z \approx k. \tag{23}$$

Thus, by simplifying the exponentials in this far-field regime, we can write the field on the screen $\psi(x, y, z)$ as

$$\psi(x, y, z) = \left(\frac{k}{2\pi i}\right)\left(\frac{Ae^{ikz}}{z}\right)\iint_\Gamma \mathrm{d}x'\mathrm{d}y'\, e^{-ik_x x'} e^{-ik_y y'} \tag{24}$$

3 Putting it all together

To summarise what we have done so far, we have considered a standard diffraction setup with some sort of aperture of shape Γ and characteristic size D through which an incident plane wave passes. We then assume that each point within this aperture contributes to the the field at an arbitrary point $P(x, y)$ on the screen. In order to simplify the resulting integral, we made two approximations: the paraxial approximation, and the far-field or Fraunhofer approximation.

Paraxial approximation: For an aperture of characteristic size D, at a distance z from the screen,

$$\frac{D}{z} \ll 1, \qquad \Longleftrightarrow \qquad z \gg D. \tag{25}$$

Far-field approximation: For an aperture of characteristic size D, at a distance z from the screen, and light of wavelength λ,

$$\frac{D}{z} \ll \frac{\lambda}{\pi D} \qquad \Longleftrightarrow \qquad z \gg \frac{\pi D^2}{\lambda}. \tag{26}$$

Note: Consider visible (green) light of wavelength $\lambda = 500$ nm, below are the conditions required for two different aperture sizes, 1 mm and 10 microns. Most laboratory diffraction experiments involves sizes within this range.

D	Paraxial approximation	Far-field approximation
1 mm	$z \gg 10^{-3}$ m	$z \gg 6$ m
10 μm	$z \gg 10^{-5}$ m	$z \gg 6 \times 10^{-4}$ m

As can be seen, for sizes close to 10 to 100 microns, both requirements are easily satisfied.

Using these approximations, we arrive at Equation (24) for the field on the screen. This is an interesting equation. From our previous Discussion Session, you should remember the definition of the (one-dimensional) Fourier transform $F(k)$ of a function $f(x)$

$$F(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{d}x\, f(x) e^{-ikx}. \tag{27}$$

Looking at Equation (24), you should see that (modulo some overall constant factors) it looks very much like a two-dimensional Fourier Transform. However, there is one important difference: while the Fourier Transform is defined as an integral over all space, the integral in Equation (24) is defined only over the region Γ. As a result, this is not *truly* a two-dimensional Fourier Transform. However, it turns out that this is quite easily remedied.

The way we remedy this is to consider an **aperture function**, $W(x', y')$, defined as follows

$$W(x', y') = \begin{cases} 1, & x \in \Gamma \\ 0 & \text{otherwise.} \end{cases} \tag{28}$$

The aperture function can be thought of as the "transparency" on the diffraction plane: it is completely transparent within the aperture, and completely opaque outside it. Using this definition, it should be trivial to show that we can write Equation (24) as

$$\psi(x, y, z) = \left(\frac{k}{2\pi i}\right)\left(\frac{Ae^{ikz}}{z}\right)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{d}x'\mathrm{d}y'\, e^{-ik_x x'} e^{-ik_y y'} W(x', y'). \tag{29}$$

We thus reach a very important conclusion: The field on the screen is proportional to the two-dimensional Fourier Transform of the aperture function!

4 Applications

4.1 The single slit

We can now use this remarkable result to compute the far-field diffraction pattern for a number of different apertures. Let us begin by considering perhaps the simplest aperture, a single slit. Let us assume the slit to have a width of $2a$ and a height of $2b$.The aperture function of such slit is quite simple:

$$W(x', y') = \begin{cases} 1, & |x'| < a \text{ and } |y'| < b, \\ 0, & \text{otherwise.} \end{cases} \tag{30}$$

Equation (29) tells us that the diffraction pattern obtained on the screen is

$$\psi(x, y) = C\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{d}x'\mathrm{d}y'\, e^{-ik_x x'} e^{-ik_y y'} W(x', y'), \tag{31}$$

where C is some constant of proportionality.

Exercise: Using the definition of the aperture function, show that

$$\psi(x, y) = C\int_{-a}^{a}\int_{-b}^{b} \mathrm{d}x'\mathrm{d}y'\, e^{-ik_x x'} e^{-ik_y y'}. \tag{32}$$

Exercise: Evaluate this simple integral to show that

$$\psi(x, y) = 4C \times \left(\frac{\sin(k_x a)}{k_x}\right) \times \left(\frac{\sin(k_y b)}{k_y}\right) = 4Cab \times \left(\frac{\sin(k_x a)}{k_x a}\right) \times \left(\frac{\sin(k_y b)}{k_y b}\right). \tag{33}$$

Exercise: What is $\psi(0, 0)$, i.e. what is the value of the field directly in front of the single slit?

The diffraction pattern on the screen is thus a two dimensional sinc function. Of course, this is the *field* on the screen; what we actually see is the intensity, which is the mod-square of this quantity, i.e. $|\psi(x, y)|^2$.

We can use this fact to arrive at a famous result that you should all be familiar with from school: the positions of the minima in the diffraction pattern of a single slit.

The zeros of $|\psi(x,y)|^2$ occur whenever $\sin(k_x a) = 0$, and $\sin(k_y b) = 0$. This occurs whenever the arguments of the sines is an integral multiple of π, and since we know that

$$k_x = \frac{kx}{r} = k\sin\theta_x \qquad \text{and} \qquad k_y = \frac{ky}{r} = k\sin\theta_y, \tag{34}$$

we can say that the zeros of the intensity pattern occur when

$$\begin{aligned} k\sin\theta_x a = n\pi \quad &\Longleftrightarrow \quad 2a\sin\theta_x = n\lambda \quad \text{in the } x\text{–direction},\\ k\sin\theta_y b = n\pi \quad &\Longleftrightarrow \quad 2b\sin\theta_y = n\lambda \quad \text{in the } y\text{–direction}. \end{aligned} \tag{35}$$

The angles θ_x and θ_y are marked in Figure (2).

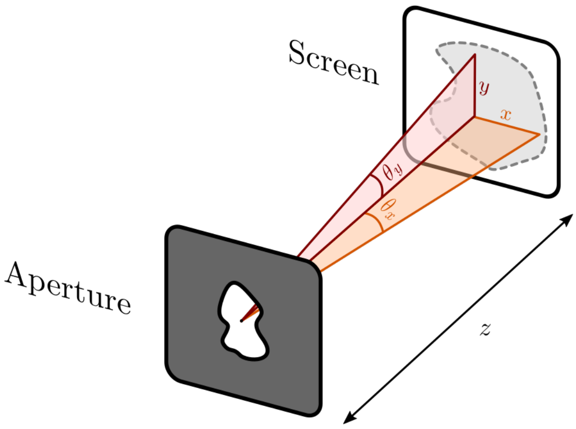

Figure 2: Coordinates on a the screen. Given a value of x (or y) on the screen, the angle θ_x (or θ_y) is defined as the angle subtended by x (or y) and the $z'z$–axis that connects the centres of both planes. This parametrisation can be used for any diffraction pattern, not just the single slit; however in the case of the single slit diffraction pattern, the zeros occur independently on the x– and y–directions, because of the form of the intensity given in Equation (33).

In school, you must have seen that the minimas of a single slit of some width d occur at the angles θ which satisfy

$$d\sin\theta = n\lambda, \qquad n \in \mathbb{Z}. \tag{36}$$

Since in our case $d = 2a$ in the horizontal direction, and $d = 2b$ in the vertical direction, it should be easy for you to see that Equation (35) is precisely the same thing that you saw in school!

Exercise: Here we have mentioned that the minima occur at a single point. However, when actually performing the experiment, you will find that the dark regions are not single points; they are extended regions, despite being smaller in width than the bright patches. Can you think of why this is the case?

4.2 The circular aperture

We can now consider the diffraction pattern due to an aperture that is *circular* with some radius a.

$$W(x', y') = \begin{cases} 1, & x'^2 + y'^2 < a^2, \\ 0, & \text{otherwise}. \end{cases} \tag{37}$$

Unlike the single-slit, this configuration mixes the x' and y' coordinates, meaning that we can no longer perform the integrals over these variables independently. Given the symmetry of the problem, it is much more useful to move to polar coordinates.

Exercise: Using polar coordinates, and Equation (23), show that

$$\begin{aligned} \psi(x, y) &= C \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{d}x' \mathrm{d}y' \, e^{-ik_x x'} e^{-ik_y y'} W(x', y') \\ &= C \int_0^a r' \mathrm{d}r' \int_0^{2\pi} \mathrm{d}\theta' \, \exp\left(-i \frac{krr'}{z} \cos(\theta' - \theta)\right), \\ &= C \int_0^a r' \mathrm{d}r' \int_0^{2\pi} \mathrm{d}\theta' \, \exp\left(-i \frac{krr'}{z} \cos\theta'\right), \end{aligned} \tag{38}$$

where $k = \sqrt{k_x^2 + k_y^2}$ is the norm of the $\mathbf{k}$ vector, (r', θ') is the coordinate on the aperture, and (r, θ) is the coordinate on the screen. Since our problem has cylindrical symmetry, we would expect our solution to be independent of θ. Thus, in the last step we have chosen $\theta = 0$, i.e., a horizontal line on the screen, without loss of generality.

The above integrals are sadly quite messy to calculate, and require two special identities related to Bessel functions. The first identity below allows us to calculate the integral of θ':

$$\int_0^{2\pi} \mathrm{d}\theta' \, e^{-ix\cos\theta'} = 2\pi J_0(x), \tag{39}$$

where $J_0(x)$ is the zeroth-order Bessel function of the first kind.

The second relation is an integral relation between Bessel functions of different orders. If $n > -1$, we can write

$$\int_0^a x^{n+1} J_n(\kappa x) \, \mathrm{d}x = \left(\frac{a^{n+1}}{\kappa}\right) J_{n+1}(\kappa a). \tag{40}$$

Exercise: Given the above results, show that

$$\psi(r, \theta) = C \times \left(\frac{2\pi a z}{kr}\right) \times J_1\left(\frac{kar}{z}\right). \tag{41}$$

Exercise: Using the fact that

$$\lim_{x \to 0} \left(\frac{J_1(x)}{x}\right) = \frac{1}{2}, \tag{42}$$

write the above equation in terms of $\psi(0, 0)$, the amplitude directly in front of the aperture.

As with the single slit, we can now find the positions of the zeros of this diffraction pattern. The zeros of $|\psi(r,\theta)|^2$ are the zeros of J_1. Unlike sines and cosines, the zeros of the Bessel functions are not regularly spaced, and must be computed numerically. If we denote these zeros by α_n, then the zeros of our intensity pattern occur at different radial distances r_n, such that

$$\frac{kar_n}{z} = \alpha_n \quad \Longleftrightarrow \quad \frac{r_n}{z} \approx \sin\theta_n = \frac{\alpha_n}{2\pi}\frac{\lambda}{a}.$$

Exercise: Look up the first zero of $J_1(x)$ and use the above relation to show that if D is the diameter of the aperture, the first dark circle occurs when

$$\theta_1 \approx 1.22\frac{\lambda}{D}. \tag{43}$$

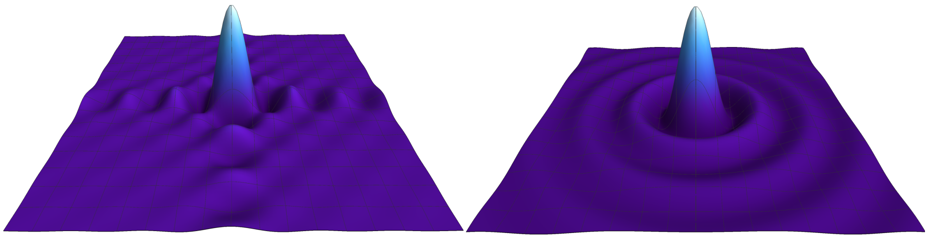

Figure 3: Amplitude profiles. The single slit (left) and circular aperture (right) amplitudes are plotted as a two-dimensional surface plot. In both cases the amplitude rapidly falls off away from the central maxima.

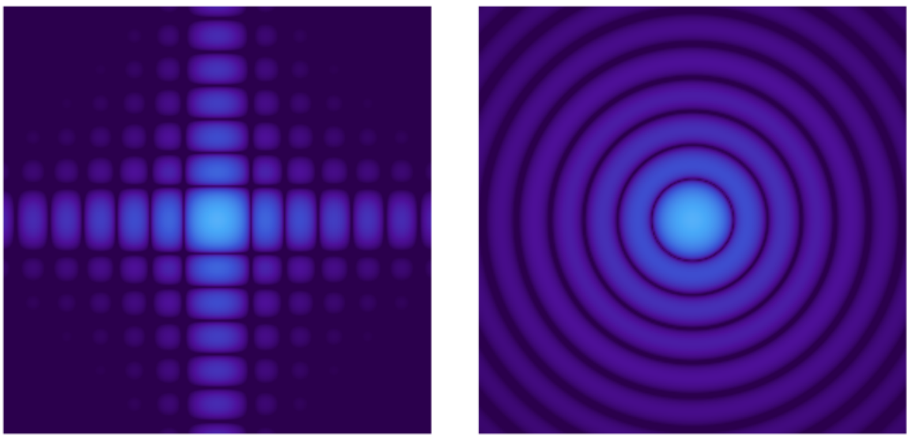

Figure 4: Intensity profiles. The single slit (left) and circular aperture (right) *intensity* profiles are plotted as a density plot. Additionally, the plots have been calibrated so that the colour gradient is *logarithmically* scaled. This is done for two reasons: firstly, the actual intensity falls off very quickly and becomes hard to distinguish; secondly, the response of the human eye is logarithmic, and therefore this is a much more faithful rendition of what one would see on the screen when this experiment is conducted in a lab.