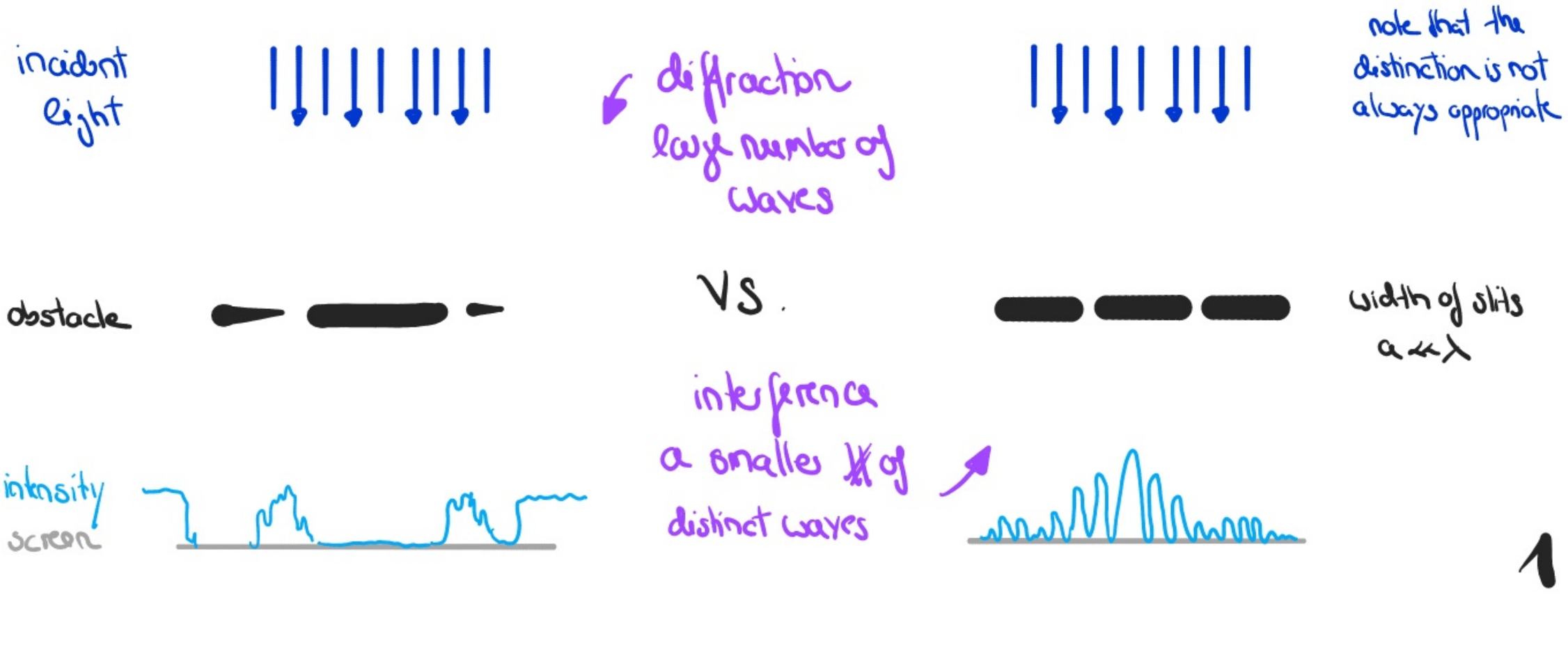
Phys 434 - Lecture 18

Diffraction through slits

1.) Introduction

Placing an opaque object between a point source and a screen creates a shadow composed of dark and bright regions that are very different from the rectilinear propagation one would expect from Geometric Optics. The phenomenon has been known since the 16th century and is generally referred to as 'diffraction'. As soon as an object blocks the path of a wavefront, the amplitude and/or phase are altered and the different segments of the wavefront moving beyond the obstacles interfere to create a distinct energy-density distribution referred to as the diffraction pattern. Interference and diffraction are thus not physically different; however the two terms are typically used in different situations:

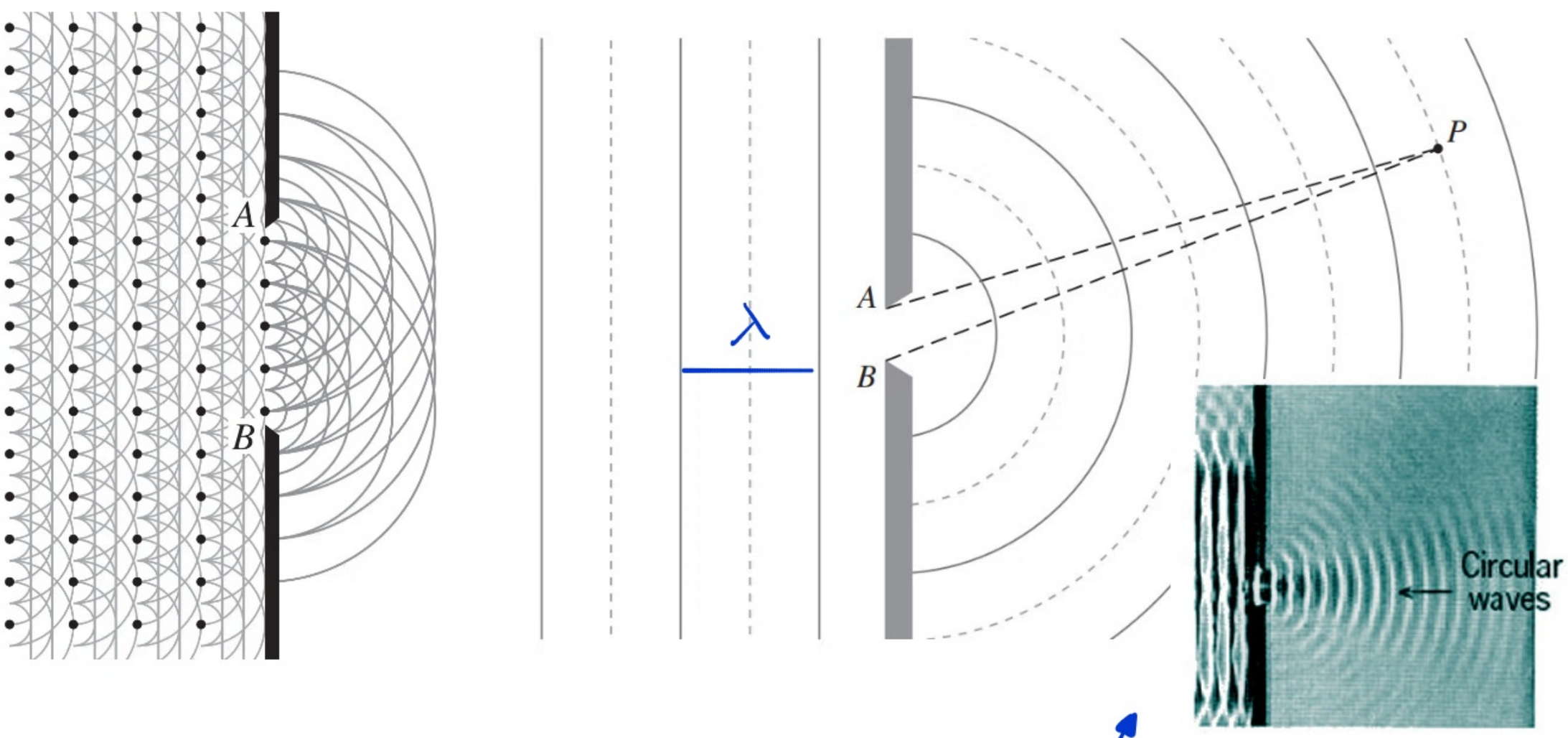


A useful tool (albeit not completely exact) to understand diffraction is Huygens' principle, i.e. each point of the wave-front serves as a source of secondary spherical wavelets. At any time, the wave-front is the envelope of these secondary wavelets. Note that this decomposition is wavelength independent and Huygens' principle cannot describe the diffraction process itself (i.e. soundwaves are barely affected by a tree, whereas light waves are strongly affected but Huygens' principle would provide the same wave-front pattern). This was resolved by Fresnel by adding the concept of interference: the amplitude of the optical field beyond the obstacle is given by the superposition of all the secondary wavelets.

Kirchhoff has shown that (with a few assumptions) it is possible to deduce the Huygens-Fresnel principle as a consequence of the scalar wave equation provided that the wavelength of light is small compared to the dimensions of the object or aperture through which it diffracts. An exact solution of the diffraction behaviour is not analytically tractable (and one of the most difficult problems in optics) but fortunately the Huygens-Fresnel principle agrees reasonably well with experiments and thus forms the basis for our discussion of diffraction.

Using the principle, we can understand the propagation of waves through an aperture (e.g. water waves moving through a barrier or light falling through

a wide slit): each unobstructed point on the incoming plane wave acts as a coherent secondary source with a maximum optical path length difference equal to $\Delta OPL_{max} = |\overline{AP} - \overline{BP}|$ (see figure below), corresponding to the source points at the edge of the aperture, with $\Delta OPL_{max} \leq \overline{AB}$.



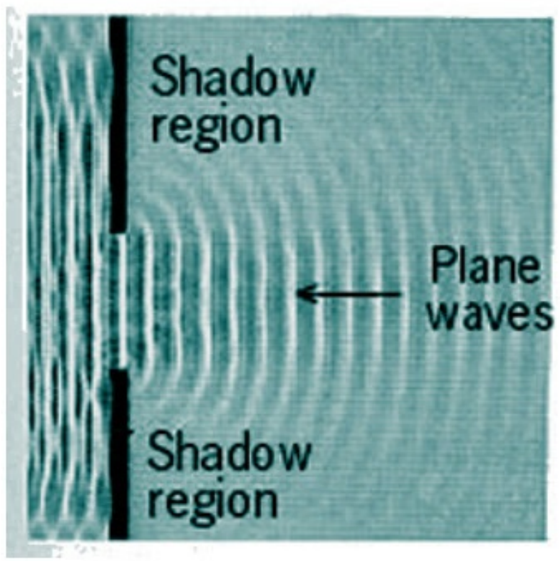
We can now distinguish two cases:

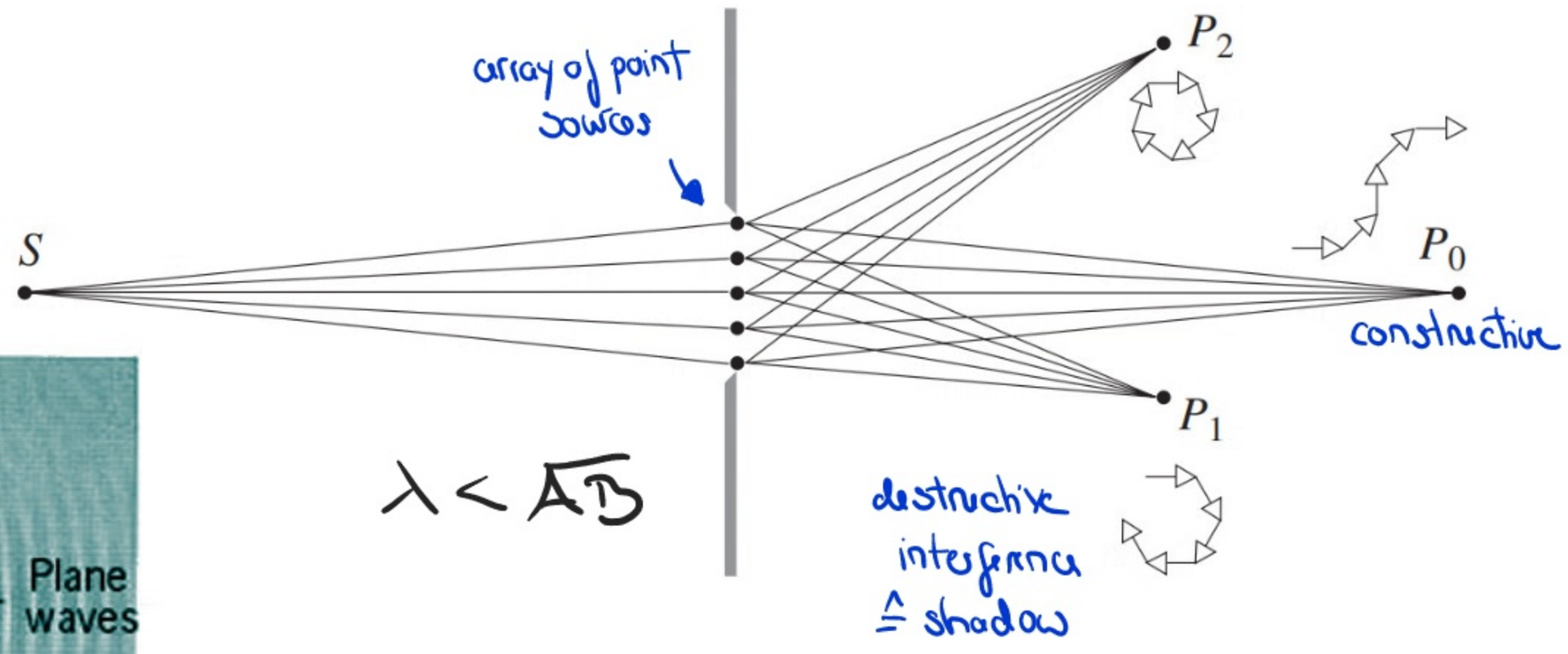
⇒ if $\lambda > \overline{AB}$, $\lambda > \Delta OPL_{max}$: all the in-phase wavelets interfere constructively and the wave spreads out at large angles beyond the obstruction ⇒ extended spherical wave pattern

not enough ΔOPL for destructive interf.

⇒ if $\lambda < \overline{AB}$, $\lambda < \Delta OPL_{max}$: only very close to the aperture can the waves interfere constructively with $\lambda > \Delta OPL$; beyond this region wavelets can interfere destructively ⇒ we observe the obstacle's shadow

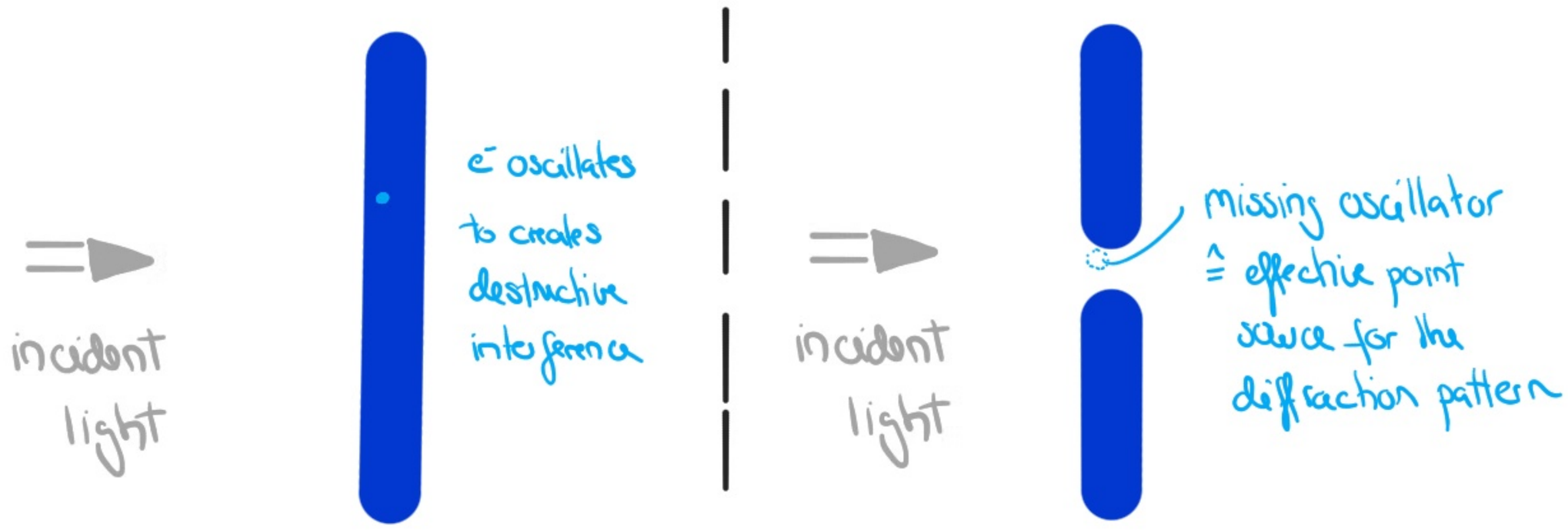
equivalent to an array of point sources





Microscopically, we can understand this by invoking the Lorentz model. An electromagnetic wave incident on an opaque material will drive the electronic oscillators within this material. These radiate waves of the same frequency which are superimposed onto the incident wave in such a way that no light is transmitted in forward direction, i.e. the wave is fully absorbed. Now we can instead remove a section of the material, so that light streams through the aperture. The oscillators that uniformly covered that section are thus removed as well and the remaining electrons in the screen are no longer affected by them. The field in the region beyond the screen is thus the original one (i.e. vanishing intensity) minus that of the removed oscillators. Apart from the sign this is equivalent to removing the incident light source and the screen but keeping the oscillators in the small section. The diffraction field can be thought of as arising exclusively from a set of fictious non-interacting point

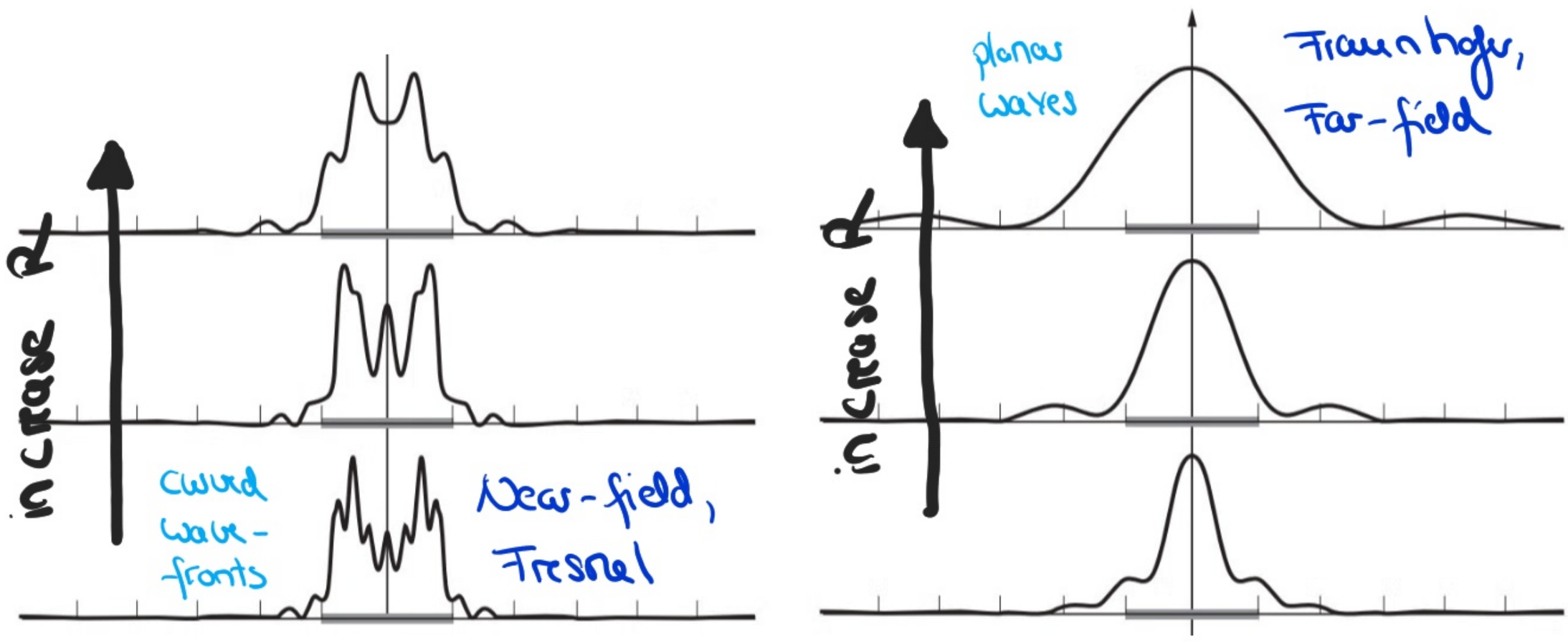
sources within the aperture, which is essentially the essence of the Huygens-Fresnel principle.



2.) Different types of diffraction

Imagine that we have an opaque screen Σ_1 containing a single small aperture of width b, which is illuminated by plane waves from a distant point source (wavelength λ). The plane of observation σ is a plane parallel to Σ_1, positioned at a distance R from the screen. Initially, if R is small, a clear image of the aperture is projected onto the screen (i.e. we see a clear shadow). As the screen is moved further away, the aperture is still recognisable but fringe structures start to appear on its periphery. For this so-called Fresnel or near-field diffraction phenomenon, the curvature of the wave fronts are very important. If σ is moved out even further the aperture is no longer visible as such and replaced by a characteristic diffraction pattern whose shape

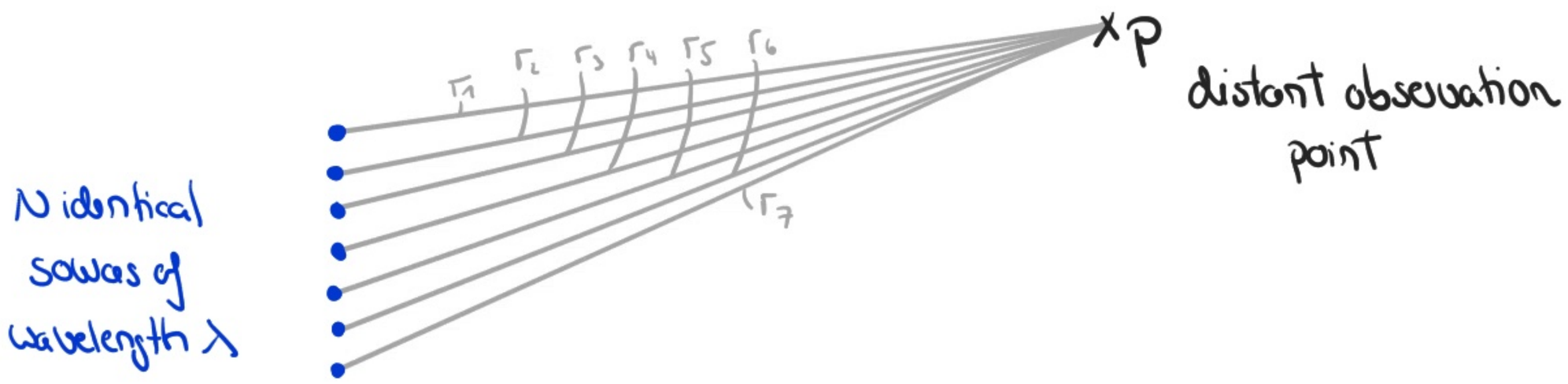
is independent of R, it only changes its size. This far-field or Fraunhofer diffraction pattern appears for $R > b^2/\lambda$ and is caused by almost planar wave fronts. We could rest back to the Fresnel case by sufficiently reducing the wavelength λ; decreasing λ even further the fringes would disappear and the aperture become visible again. The succession of diffraction patterns is illustrated below.



3.) Coherent oscillators

Understanding diffraction through an aperture (a slit in the simplest case) amounts to understanding the field produced by an array of N coherent point sources. If we focus on the far-field regime, the different rays interfering at a point P far away from the sources are almost parallel and will have travelled approximately the same distance. This implies that their amplitudes will be approximately equal, i.e.

$$E_0(r_1) = E_0(r_2) = \dots = E_0(r_N) = E_0(r) .$$



The sum of interfering spherical wavelets will thus be given as the real part of

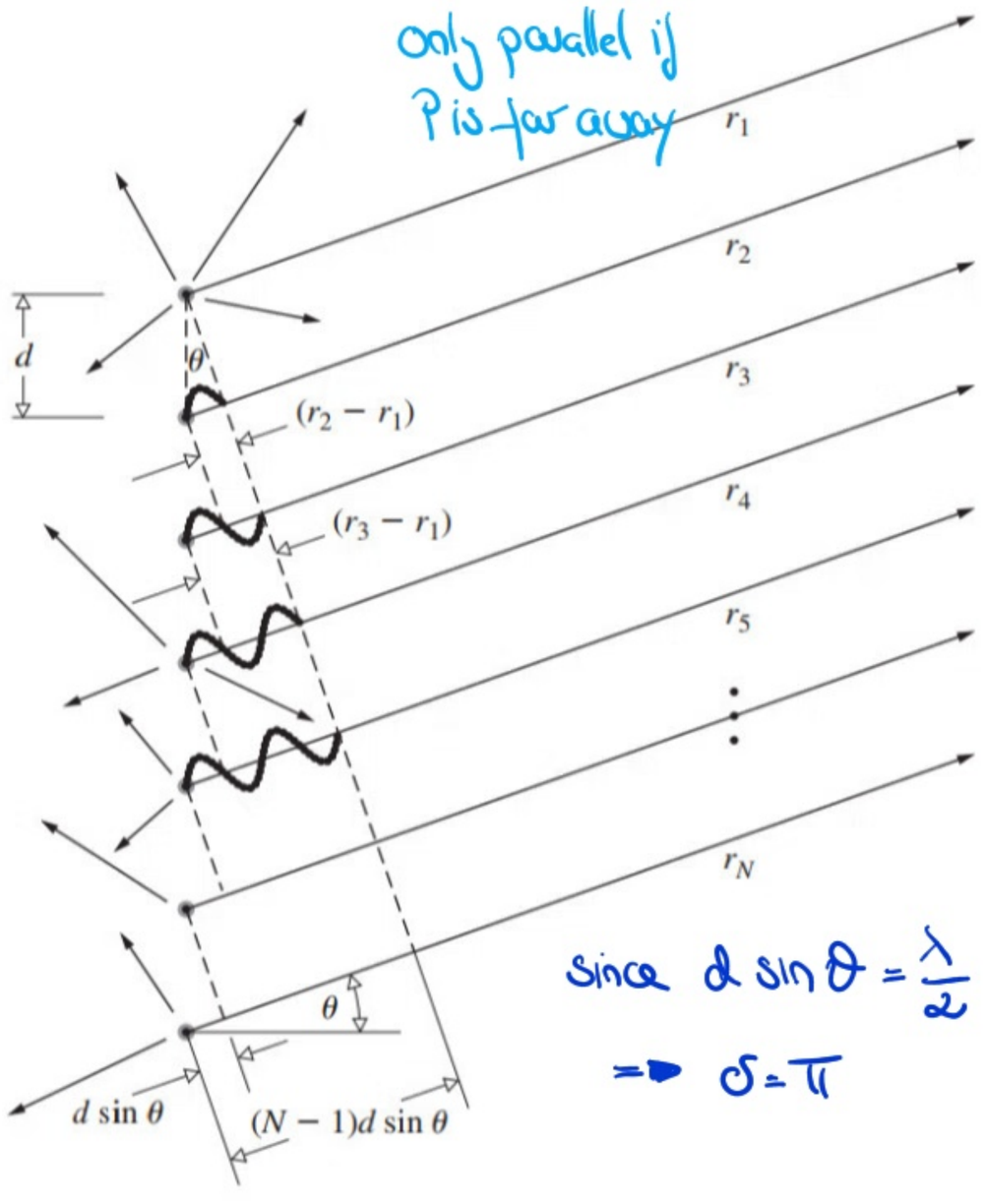
$$E = E_0(r)\, e^{i(kr_1 - \omega t)} + E_0(r)\, e^{i(kr_2 - \omega t} + \dots + E_0(r)\, e^{i(kr_N - \omega t)} .$$

From our discussion in Lectures 16 & 17 it should be clear that we can rewrite this expression taking advantage of the phase difference between adjacent beams resulting from the difference in optical path length. From the figure we deduce

$$\Delta OPL = n d \sin\theta ,$$

$$\Rightarrow \quad \delta = k_0 n d \sin\theta$$

$$= k d \sin\theta .$$



If the sources are in phase,

i.e. $\mathcal{E}_1 = \mathcal{E}_2 = \mathcal{E}_3 = \dots = \mathcal{E}_N$,

we can thus write

$$E = E_0(r)\, e^{-i\omega t}\, e^{ikr_1}\left(1 + e^{ik(r_2-r_1)} + \dots + e^{ik(r_N-r_1)}\right)$$

$$= E_0(r)\, e^{-i\omega t}\, e^{ikr_1}\left[1 + e^{i\delta} + (e^{i\delta})^2 + \dots + (e^{i\delta})^{N-1}\right].$$

The geometric series in square brackets converges because

$$\sum_{n=0}^{N-1} a^n = \frac{1-a^N}{1-a}, \qquad \text{for } a \neq 1.$$

So we obtain

$$E = E_0(r)\, e^{-i\omega t}\, e^{ikr_1}\, \frac{1-e^{i\delta N}}{1-e^{i\delta}}$$

$$= E_0(r)\, e^{-i\omega t}\, e^{ikr_1}\, \frac{e^{i\delta N/2}\left(e^{-i\delta N/2} - e^{i\delta N/2}\right)}{e^{i\delta/2}\left(e^{-i\delta/2} - e^{i\delta/2}\right)}$$

$$= E_0(r)\, e^{i[kr_1 - \omega t + (N-1)\delta/2]}\, \frac{\sin(N\delta/2)}{\sin(\delta/2)}.$$

Finally, if we define R as the distance from the centre of the oscillator array to the point P, we have $R = (N-1)/2 \cdot d\sin\vartheta + r_1$, so we have

$$E = E_0(r)\, e^{i(kR-\omega t)}\, \frac{\sin(N\delta/2)}{\sin(\delta/2)}.$$

We can thus deduce for the resulting intensity

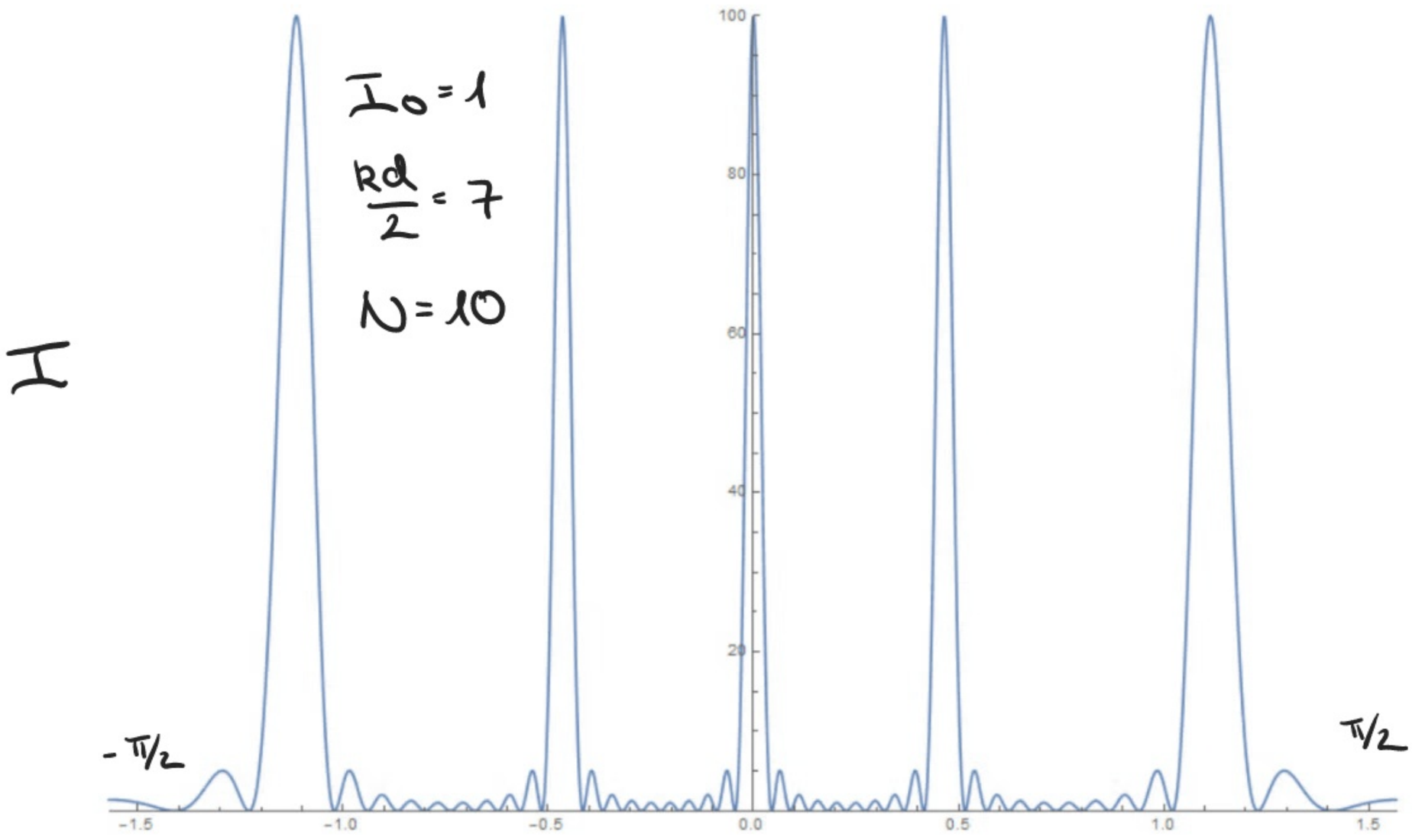
$$I = \tfrac{1}{2} E \cdot E^{*} = I_0 \frac{\sin^2(N\delta/2)}{\sin^2(\delta/2)} \quad .$$

$I_0 = \frac{1}{2} E_0^2$ intensity of a single oscillator at P

From this we deduce $I = 0$ if $N = 0$, $I = I_0$ for $N = 1$ and $I = 4 I_0 \cos^2(\delta/2)$ for $N = 2$ in agreement with our discussion in lecture 16. Note that the above result depends on θ (i.e. the distance of P from the centre of the array):

$$I = I_0 \cdot \frac{\sin^2(NKd/2 \cdot \sin\theta)}{\sin^2(Kd/2 \cdot \sin\theta)} \quad .$$

Note that the term in the numerator fluctuates rapidly, whereas the denominator varies rather slowly, so the combined expression gives rise to a series of sharp peaks separated by smaller secondary maxima:



The maxima occur in those directions ϑ_m such that $\delta = 2\pi m$, $m \in \mathbb{Z}$, which leads to

$$\delta = \cancel{2\pi} m = k d \sin\vartheta_m = \frac{\cancel{2\pi}}{\lambda} d \sin\vartheta_m ,$$

$$\Rightarrow \quad \underline{\underline{\lambda \cdot m = d \sin\vartheta_m}} \quad . \qquad (*)$$

Moreover using L'Hospital's Rule it is possible to show that

$$\frac{\sin^2(N k d/2 \cdot \sin\vartheta)}{\sin^2(k d/2 \cdot \sin\vartheta)} = N^2 \quad \text{for} \quad \delta \to 2\pi m ,$$

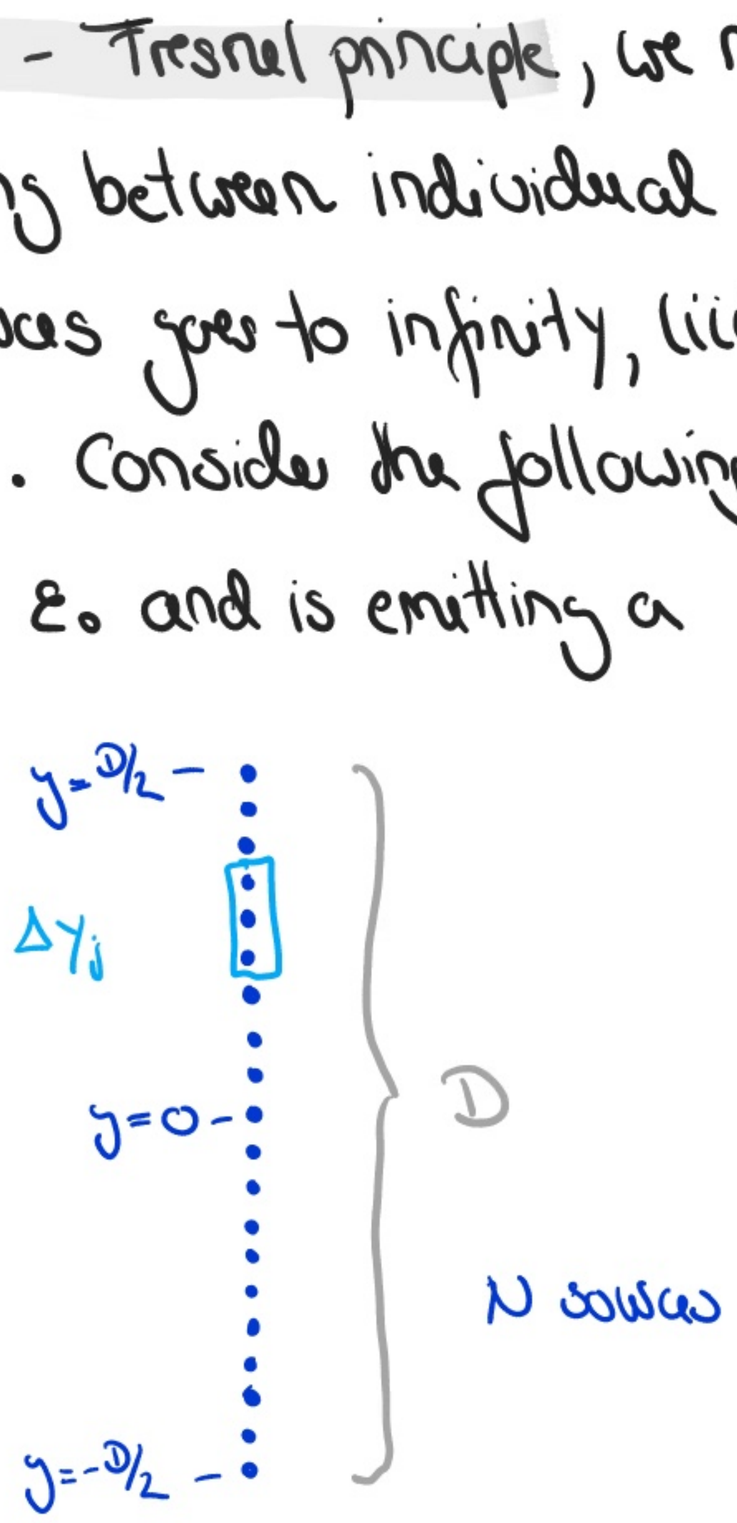
which implies that the principle maxima have intensities $I_{max} = N^2 I_0$, as we would expect for in-phase oscillators. The system will radiate a maximum in the direction perpendicular to the array, i.e. $\vartheta_0 = 0$. Note that if $d < \lambda$ only this $m = 0$ (zeroth-order maximum) exist as dictated by Eqn. (*). As ϑ increases, I falls off to zero at $N\delta/2 = \pi$, its first minimum.

Question: Suppose you have an array of many closely-spaced sources (e.g. as used in interferometric radio antennas), such that $d < \lambda$ and that the sources have a relative phase shift ε between neighbouring antennas, where will you see the maxima?

In this case $\delta = kd \sin\vartheta + \varepsilon$ so that the various principle maxima will occur at the new angles $d \sin\vartheta_m = m\lambda - \varepsilon/k$. If we focus on the central maxima, we can vary its position by adjusting the value of ε. The principle of reversibility (without absorption, wave motion is reversible) suggests that the same pattern applies to an antenna array either used as a transmitter or a receiver. Hence, we can use the array e.g. as a radio telescope and 'point' it into a specific direction by combining the output from individual antennas with an appropriate phase shift between each of them. This so-called phased array radar is widely applied.

To connect this discussion to the Huygens - Fresnel principle, we need to consider the limit where (i) the spacing between individual sources goes to zero, (ii) the number of sources goes to infinity, (iii) each source has a vanishingly small intensity. Consider the following set-up, where each source has a strength ε_0 and is emitting a spherical wavelet of the form

$$E = \frac{\varepsilon_0}{r} e^{i(kr-\omega t)} .$$



Each segment Δy_j contains $\Delta y_j / D \cdot N$ sources and con-

tributs the electric field intensity

we assume Δy_j is small enough so that $r \approx r_j$ & no rel. phase difference

$$E_j = \frac{\varepsilon_0}{r} \cdot \frac{\Delta y_j}{\mathcal{D}} \cdot N \, e^{i(kr_j - \omega t)} .$$

In the limit $\varepsilon_0 \to 0$, $N \to \infty$ and $\mathcal{D} = \text{const}$, we can define a source strength per unit length given by

could be a function of y, if the phase changes across the array

$$\varepsilon_L \equiv \frac{1}{\mathcal{D}} \lim_{N \to \infty} (\varepsilon_0 N) .$$

The net field at a point P located a distance r_j from each segment can be obtained by summing all j contributions

$$E = \sum_{j=1}^{N} \frac{\varepsilon_L}{r_j} \Delta y_j \, e^{i(kr_j - \omega t)} .$$

mathematical idealisation

In the continuum limit (for a continuous line source), where $\Delta y_j \to 0$ and $N \to \infty$, the sum becomes an integral

$r = r(y)$ as the intensity at P depends on its position

$$E = \int_{-\mathcal{D}/2}^{\mathcal{D}/2} \frac{\varepsilon_L(y)}{r(y)} e^{i(kr(y) - \omega t)} \, dy .$$

This equation allows us to calculate the emission patton of a large number of tiny emitters or equivalently the light through an aperture.

Phys 434 - Lecture 19

Fraunhofer Diffraction

1.) Single slit

Continuing from the discussion of last's lecture, we need to evaluate

physical field corresponds to the real part

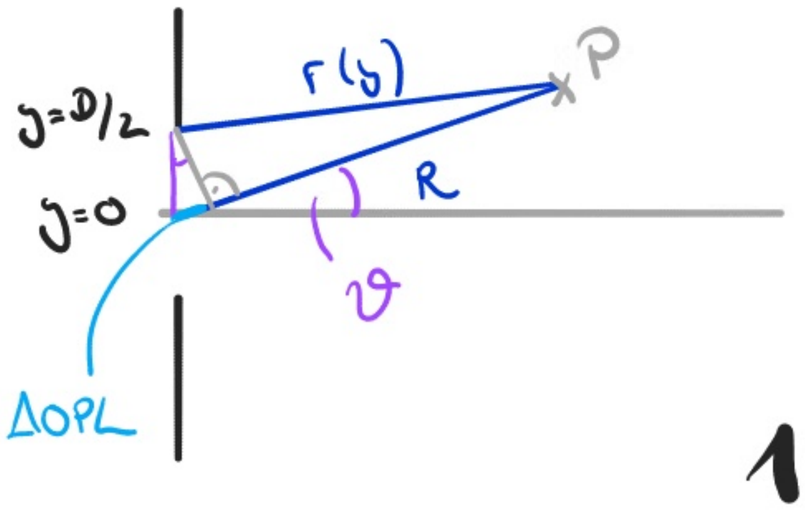
$$E = \int_{-D/2}^{D/2} \varepsilon_L(y)\, e^{i(kr(y)-\omega t)} \frac{1}{r(y)}\, dy \qquad (*)$$

in case of Fraunhofer's regime. Far away from the source, $R \gg D$, ε_L will be approximately constant and $r = r(y)$ will never deviate significantly from its midpoint value R. This implies that $1/r(y) \approx \text{const} \approx 1/R$. Retaining the r-dependence in the exponential as small changes in $r(y)$ can have a large effect on the resulting field, we can deduce that the contribution to the total electric field from the point sources within dy is

phase is very sensitive to r changes

$$dE = \varepsilon_L \frac{1}{R} e^{i(kr(y)-\omega t)}\, dy .$$

We can express r as a function of y by looking at the geometry of the system (equivalent to the discussion of the Michelson interferometer



from Lecture 16) to obtain the so-called Fraunhofer condition

$$R - r(y) \approx y \cdot \sin\vartheta \quad \rightarrow \quad \underline{\underline{r(y) \approx R - y\sin\vartheta}} \; .$$

r is linear in y

With this Eqn. (*) becomes

$$E = \frac{\varepsilon_L}{R} \int_{-D/2}^{D/2} e^{i(kR - yk\sin\vartheta - \omega t)} \, dy$$

$$= \frac{\varepsilon_L}{R} e^{i(kR-\omega t)} \int_{-D/2}^{D/2} e^{-iyk\sin\vartheta} \, dy$$

$$= \frac{\varepsilon_L}{R} e^{i(kR-\omega t)} \left[\frac{e^{-iyk\sin\vartheta}}{-ik\sin\vartheta} \right]_{-D/2}^{D/2}$$

$$= \frac{\varepsilon_L}{R} e^{i(kR-\omega t)} \frac{2}{R\sin\vartheta} \frac{(-1)}{2i} \left(e^{-i\,RD/2\sin\vartheta} - e^{+\dots} \right)$$

phase is the same as for a point source at the middle of the slit

$$= \frac{\varepsilon_L}{R} e^{i(kR-\omega t)} \left(R/2\sin\vartheta \right)^{-1} \cdot \sin\left(RD/2 \sin\vartheta \right) .$$

The resulting irradiance is thus (with $\beta \equiv RD/2 \cdot \sin\vartheta$)

$$I = \langle \mathrm{Re}[E^2] \rangle_T = 1/2 \; E \cdot E^*$$

$$= \underline{\underline{\frac{1}{2} \left(\frac{\varepsilon_L D}{R} \right)^2 \left(\frac{\sin\beta}{\beta} \right)^2 .}}$$

irradiance from an ideal line source in the far-field (Fraunhofer) limit

The principal maximum corresponds to $\theta = 0$, where $\sin\beta/\beta = 1$, so that $I(0) = 1/2\,(\varepsilon_L D/R)^2$. We can thus rewrite

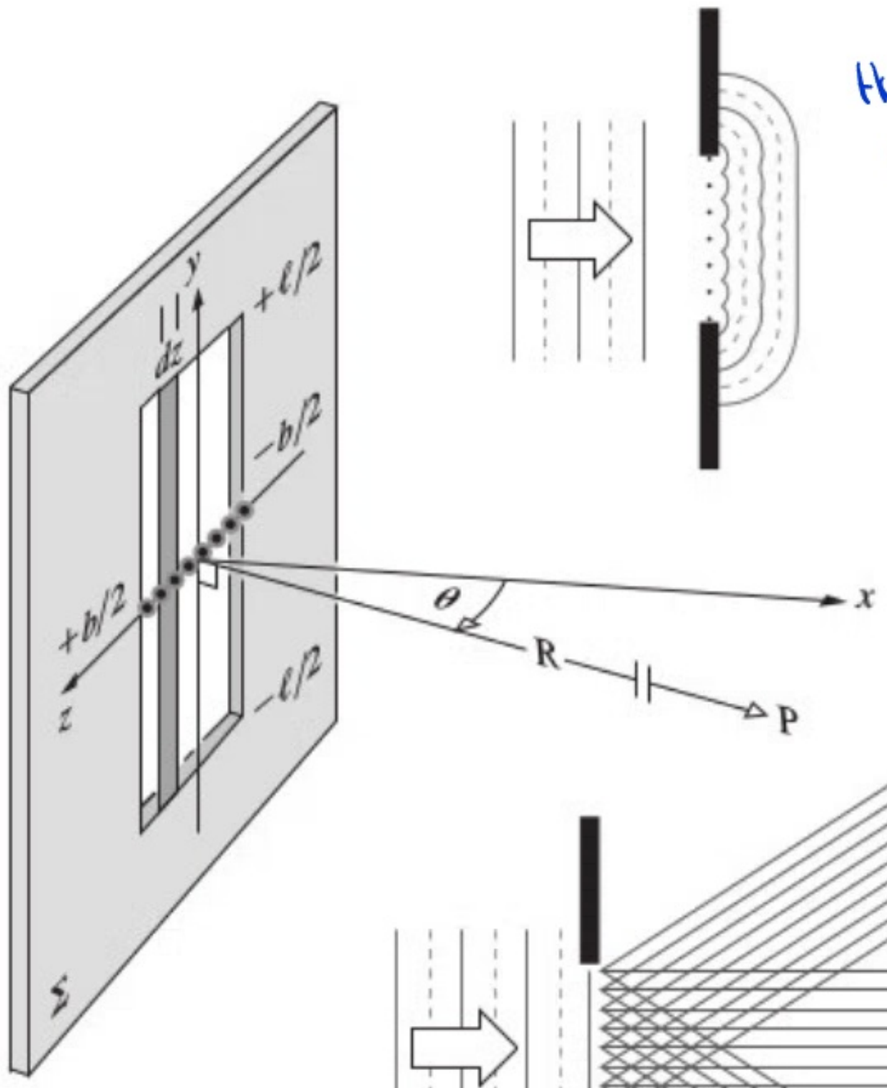
$$I(\theta) = I(0)\left(\frac{\sin\beta}{\beta}\right)^2 = I(0)\,\mathrm{sinc}^2\beta,$$

which is symmetric about the y axis. Note that since $\beta = \pi D/\lambda \cdot \sin\theta$, we can distinguish two cases:

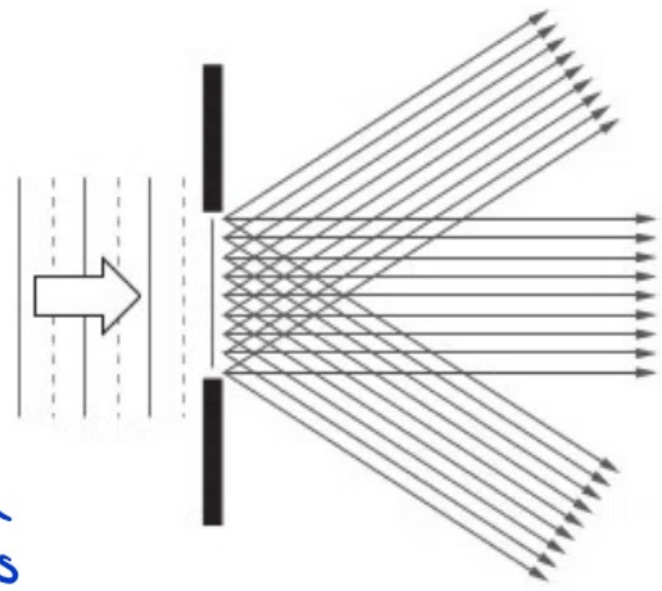
⇒ $D \gg \lambda$: the irradiance drops very rapidly as $\theta \neq 0$; long coherent light source can be envisaged as a single-point emitter radiating a circular wave mainly in the forward direction

⇒ $D \ll \lambda$ ($\beta \ll 1$): in this case $\sin\beta \approx \beta$ so that the irradiance is constant for all θ, $I(\theta) \approx I(0)$; the line source resembles a point source emitting spherical waves

While the previous discussion was concerned with the emission from a line source, realistic apertures will not be infinitely thin, but have a specific width b. To determine the resulting diffraction pattern, we devide the slit into long differential strips parallel to the y-axis (see next page for geometry). Motivated by the previous discussion, we can replace each of these line sources by a point emitter on the z-axis. Each of these will emit a circular wave in the $y=0$ plane since $D \gg \lambda$, that subsequently interfer, resulting in a



Huygens wavelets emitted across the aperture



Ray representation of the wavelets

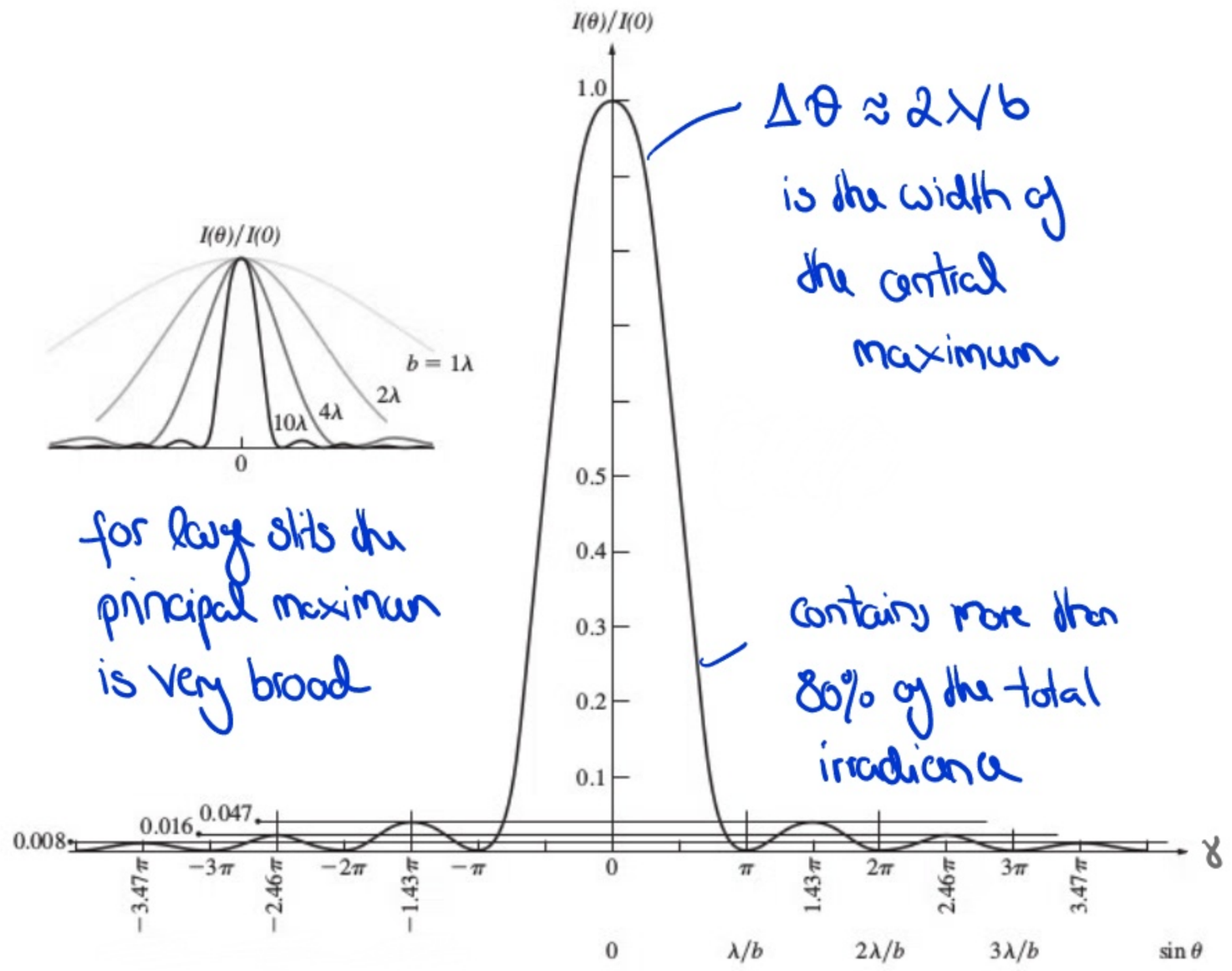
diffraction pattern is very narrow in y-direction, but stretched out in z-direction

diffraction pattern that is very narrow in the y-direction, i.e. there is very little diffraction parallel to the long edges of the slits. As these point sources in z-direction are equivalent to a line source in that direction, we readily deduce the corresponding Fraunhofer diffraction pattern to be

$$I(\theta) = I(0)\left(\frac{\sin\gamma}{\gamma}\right)^2,$$

where $\gamma \equiv Kb/2 \cdot \sin\theta$ and θ is measured from the xy plane (see geometry above). As $b \ll D$, γ is not large and thus higher-order maxima are observed. Extrema can be obtained by $dI/d\gamma \doteq 0$; the maxima are located at points where $\tan\gamma = \gamma$, while minima are lo-

cated at $\gamma = \pm\pi, \pm 2\pi, \ldots$ The characteristic single slit patten is illustrated below



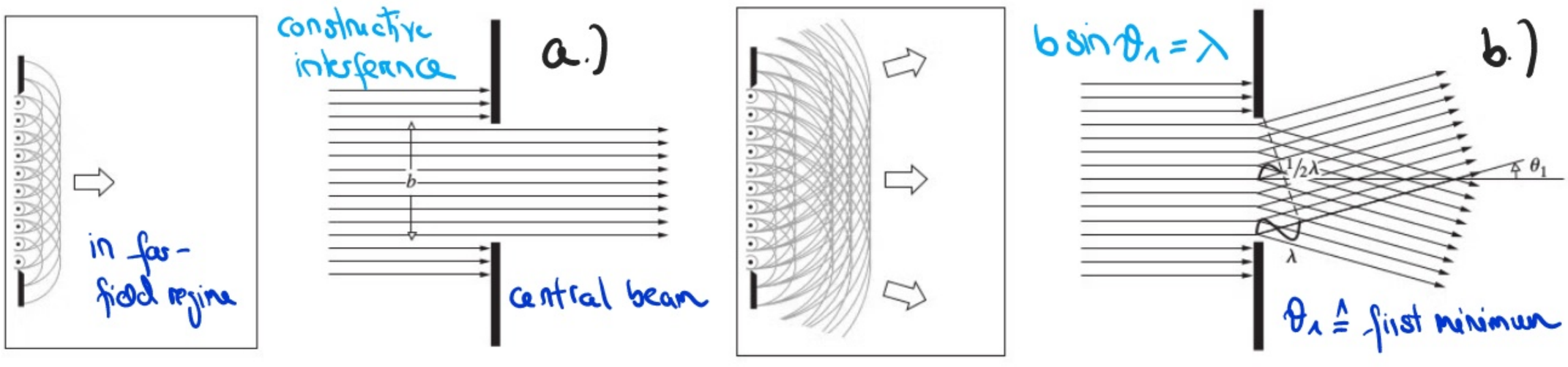
$\Delta\theta \approx 2\lambda/b$ is the width of the central maximum

for large slits the principal maximum is very broad

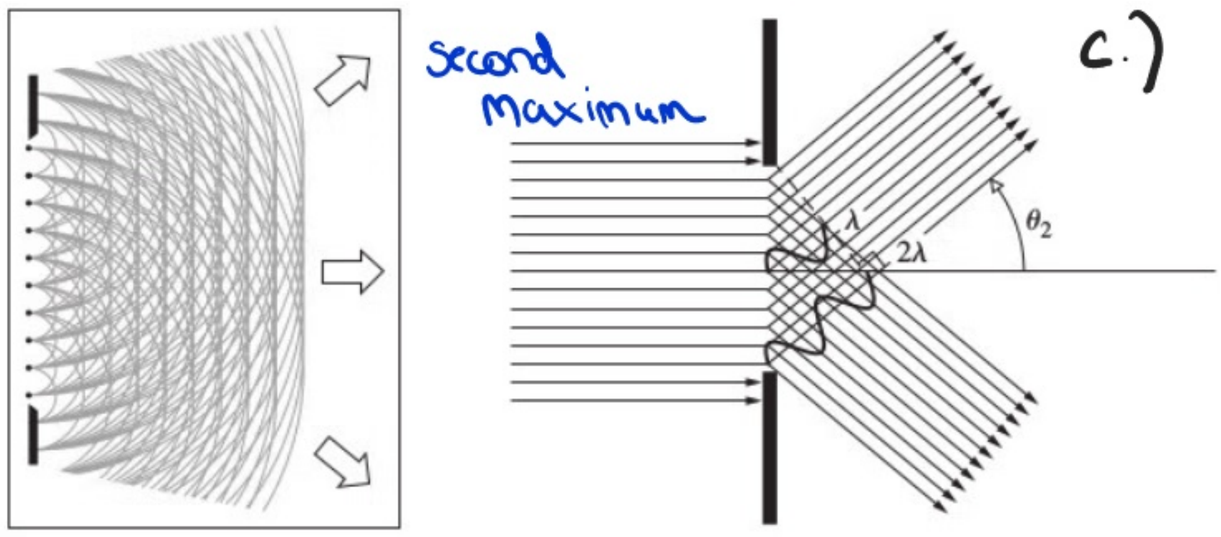
contains more than 80% of the total irradiance

to actually make the patten visible on a screen, we require a lens to convert the angles θ to a position on the screen; the diffraction patten is centred about the axis of the lens

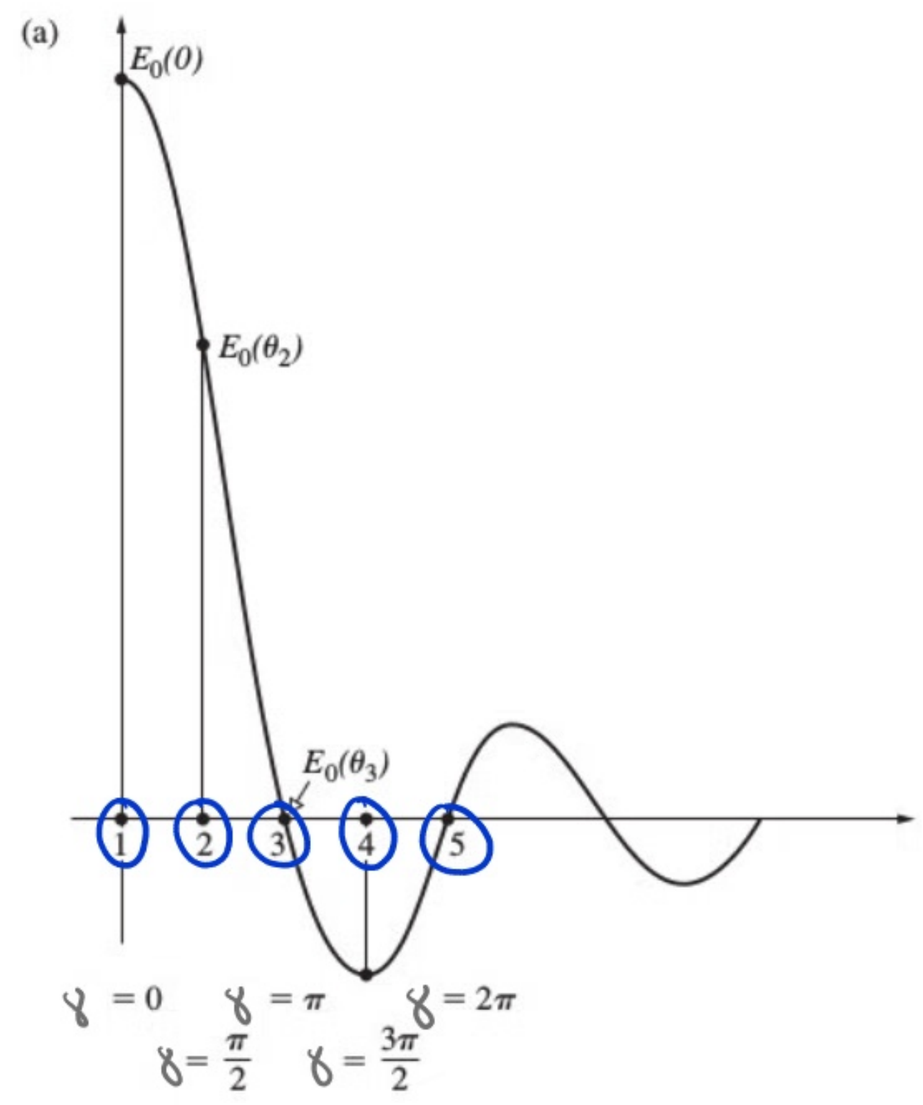
We can understand this characteristic pattern, by considering the individual Huygens' wavelets emitted from the aperture, which are all of same phase, wavelength and amplitude. The net wave propagating in forward direction (undiffracted beam) can be represented as the ray bundle in Figure a.). For the angle θ_1 illustrated in b.), the overall path length between the first and the Nth beam is equal to λ. Thus wavelets from the middle of the

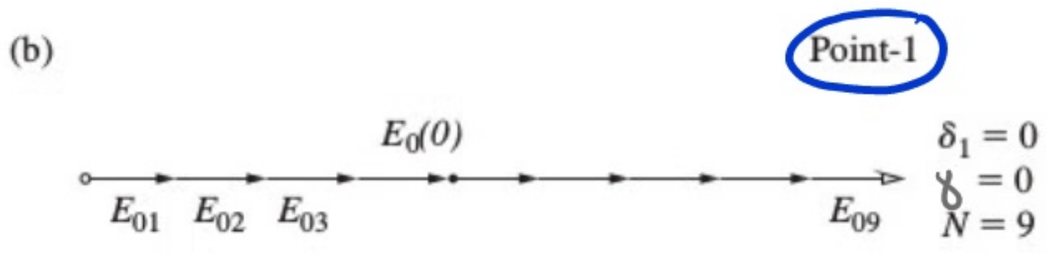


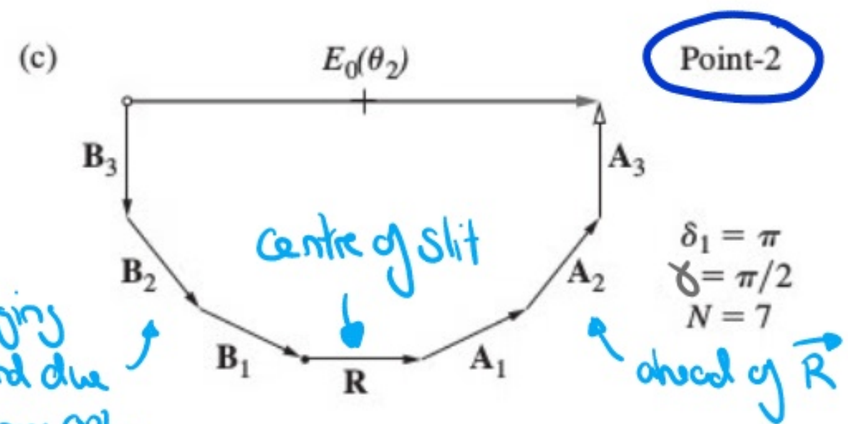
slit will arrive at the viewing screen with a path length difference of $\lambda/2$ with respect to those arriving from the top of the slit, and thus cancel each other. We can similarly pair all the other rays across the aperture so that they cancel; i.e. for the angle θ_1 the resultant E-field amplitude (and thus the irradiance) will be zero, corresponding to the first minimum. For a larger angle θ_2 (see Figure c.), the overall path difference will be 2λ, so that individual pairs of rays will have a path difference of λ and thus add constructively, resulting in the second maximum. Increasing θ further we would obtain another irradiance minimum, then maximum, and so on. The exact amplitude of the respective maxima can be understood by using phasor notation and assigning each wavelet a fixed E-field amplitude:



c.)

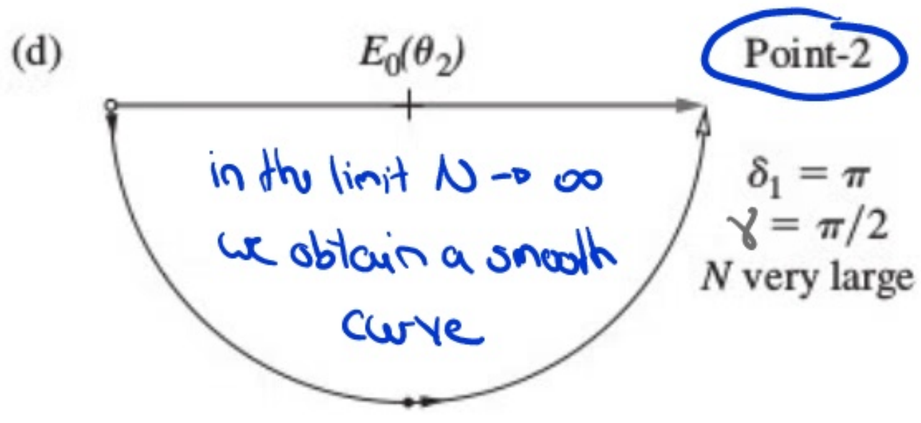






$\delta_1/(N-1)$ is their relative lag

$\vec{B}_i$ are the phasors for the beams below the middle of the slit, $\vec{A}_i$ those above



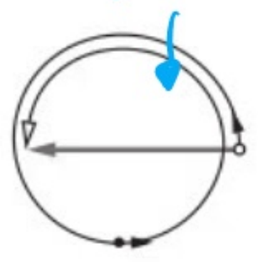
Point-2

$\delta_1 = \pi$
$\gamma = \pi/2$
N very large

Point-3

$\delta_1 = 2\pi$
$\gamma = \pi$
N very large

negative E



Point-4

$\delta_1 = 3\pi$
$\gamma = 3\pi/2$
N very large

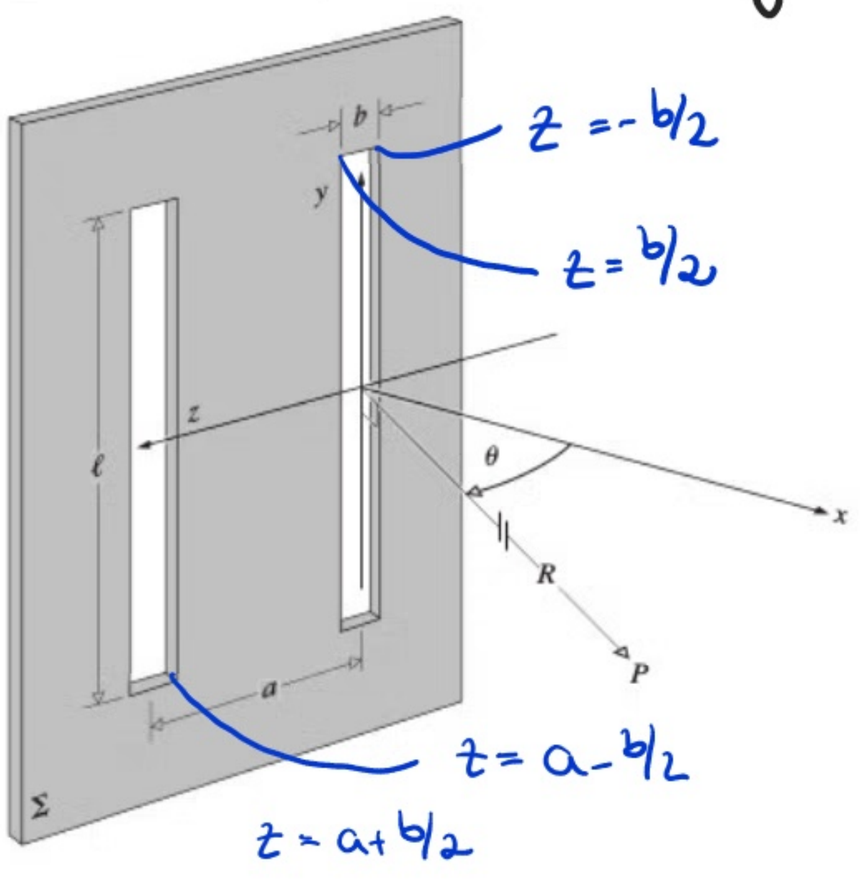
Point-5

$\delta_1 = 4\pi$
$\gamma = 2\pi$
N very large

Note that the continuous phasor curve obtained in the limit $N \to \infty$ is also referred to as the vibration curve. For each and every θ value there will be a specific configuration of this vibration curve; as θ increases, the OPL difference becomes larger and the relative phase angles between individual phasors increases so that the vibration curve gets tighter. The overall amplitude of subsequent maxima thus decreases the larger θ.

2.) Double slit

Suppose we have two slits of width b and a centre-to-centre separation of a. Each aperture would generate the same single-slit diffraction pattern, but the individual wavelets from both slits are coherent and will thus interfer depending on their relative path lengths. To mathematically describe this, we generalise the single-slit analysis in the following way



$$E = \frac{\mathcal{E}_L}{R} e^{i(kR-\omega t)} \left[\int_{-b/2}^{b/2} e^{-ikz\sin\theta}\, dz + \int_{a-b/2}^{a+b/2} e^{-ikz\sin\theta}\, dz \right]$$

$$= 11 \left[\frac{\sin(bk/2\sin\theta)}{k/2\sin\theta} + e^{-ika\sin\theta} \frac{\sin(bk/2\sin\theta)}{k/2\sin\theta} \right]$$

$$= 11 \left[\frac{b\cdot\sin\gamma}{\gamma} + e^{-ika\sin\theta} \frac{b\cdot\sin\gamma}{\gamma} \right] \qquad \gamma = \frac{bk}{2}\sin\theta$$

$$= \frac{\mathcal{E}_L b}{R} e^{i(kR-\omega t)} \frac{\sin\gamma}{\gamma} e^{-ika/2\sin\theta} \left[e^{ika/2\sin\theta} + e^{-ika/2\sin\theta} \right]$$

$\alpha \equiv ka/2\sin\theta$

finite width of both slits modifies the amplitude of the interference pattern

this looks like the 'simple' two-slit pattern for infinitely small slits (Lecture 16)

$$= \frac{2\mathcal{E}_L b}{R} e^{i(kR-\omega t)} e^{-i2} \cos\alpha\ \mathrm{sinc}\,\gamma \,.$$

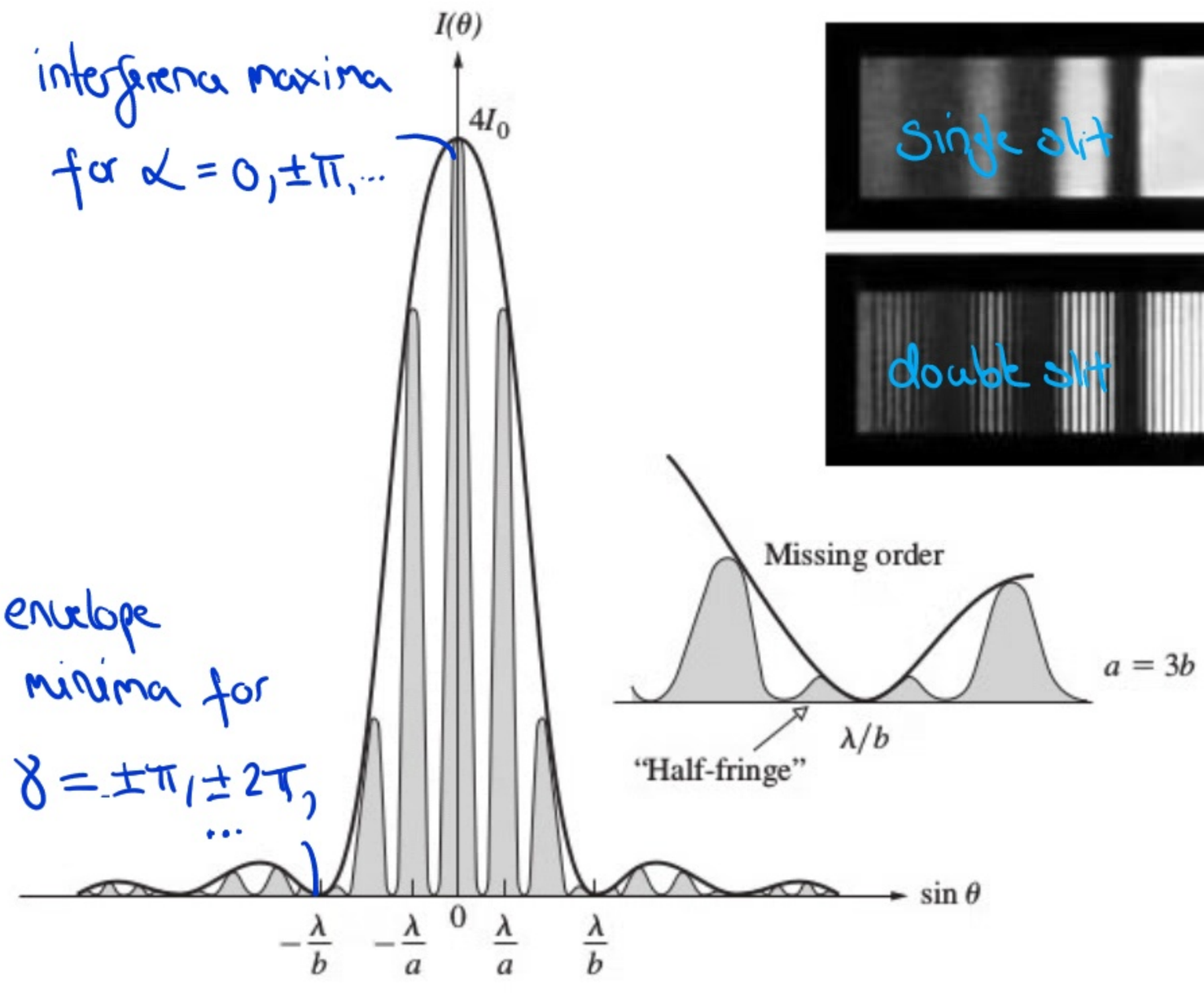
The intensity of the double slit is thus given by

$$I(\theta) = 1/2\, E\cdot E^* = 1/2 \left(\frac{\mathcal{E}_L b}{R}\right)^2 4\cos^2\alpha\ \mathrm{sinc}^2\gamma$$

$$= 4 I_0 \cos^2\alpha \cdot \mathrm{sinc}^2\gamma \,.$$

In the $\theta = 0$ direction ($\alpha = \gamma = 0$), we obtain the principal maximum with $I(0) = 4I_0$. In the limit b and $\gamma \to 0$, $\mathrm{sinc}\,\gamma = 1$ and the result reduces to that of Young's double slit experiment (Lecture 16). If

however, $a = 0$ and thus $\alpha = 0$, the two slits merge into one and we recover the single-slit diffraction pattern with the source strength doubled. We can thus imagine the resulting pattern as an interference term $\cos^2\alpha$ modulated by the $\text{sinc}^2\gamma$ diffraction term. Note that as a result of this we can obtain 'missing orders', i.e. points where a minimum of the diffraction envelope falls onto a maximum of the interference pattern.



3.) Multiple slits

We can generalise the discussion of the double slit to understand the effect of N slits on the diffraction process. Mathematically, we would describe this by summing all the individual contributions of individual slits, se-

parated a distance a, i.e. $\gamma = \frac{kb}{2}\sin\theta$

$$E = \frac{\mathcal{E}_L}{R} e^{i(kR-\omega t)} \sum_{n=0}^{N-1} \int_{na-b/2}^{na+b/2} e^{-ikz\sin\theta}\, dz \quad .$$

Instead of evaluating this, we can use our insight from the double slit to deduce E. We have seen that each of the slits behaves like a point source with amplitude $A = \mathcal{E}_L b \,\mathrm{sinc}\,\gamma$ located at the centre of the slit. At a given angle of observation, all slits will have the same γ, so the multi-slit system will act like an array of N point sources (spaced by a) with amplitude $A \propto \mathrm{sinc}\,\gamma$. These N point sources sum to a total field

measure R from the centre of the array of slits

$$E = \frac{A}{R} e^{i(kR-\omega t)} \frac{\sin(N\alpha)}{\sin(\alpha)} \,,$$

see phased-array discussion in Lecture 18

so that we find for the irradiance

$$I = \tfrac{1}{2}\left(\frac{\mathcal{E}_L b}{R}\right)^2 \mathrm{sinc}^2\gamma \cdot \frac{\sin^2(N\alpha)}{\sin^2\alpha}$$

$$= I_0 \,\mathrm{sinc}^2\gamma \cdot \left(\frac{\sin(N\alpha)}{\sin\alpha}\right)^2 .$$

I_0 is the intensity from one slit at the observation point

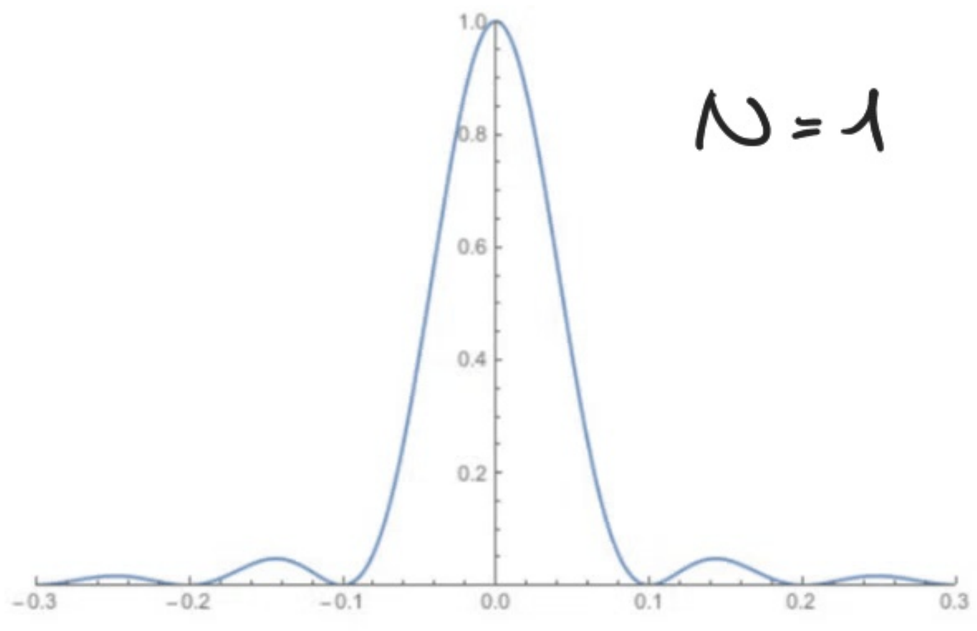
$\alpha = \frac{ka}{2}\sin\theta$

We observe that the single slit 'envelope' remains as the overall shape of the diffraction pattern in the far-field limit. The principal maxima occur

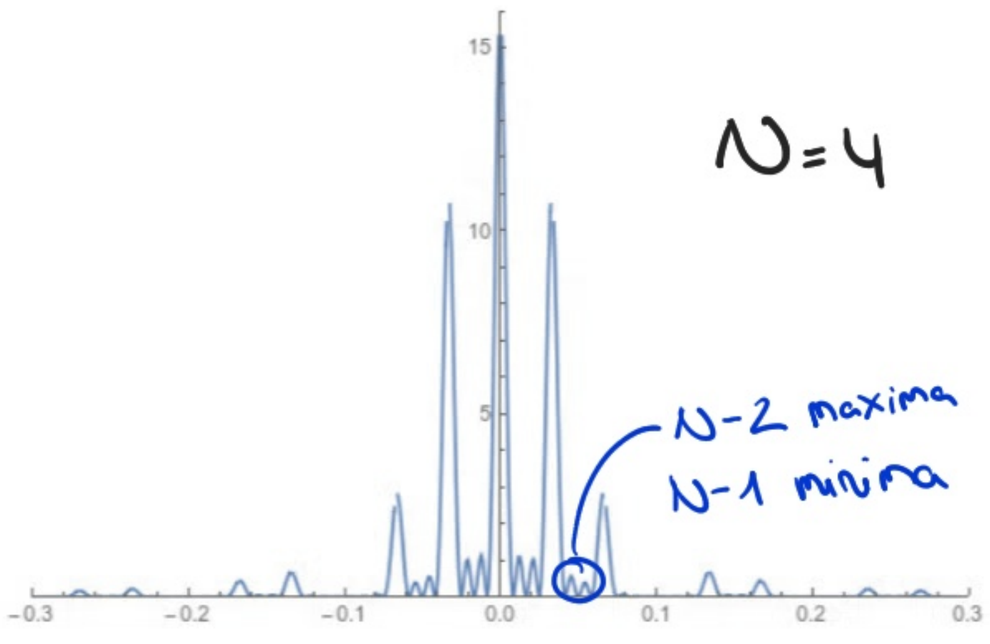
for $\alpha = 0, \pm\pi, \ldots$ or equivalently $a \sin\theta_m = m\lambda$ for $m \in \mathbb{Z}$, where m is also referred to as the **order of diffraction**. Note that the **positions** of the principal maxima **are fixed** but they **become increasingly narrower** as more and more slits are illuminated. This is due to $N-1$ **minima** occuring between principal maxima; these minima are determined by $(\sin(N\alpha)/\sin\alpha)^2 = 0$ or equivalently $\alpha = \pm\pi/N, \pm 2\pi/N, \ldots$ This implies that the **two nearest minima** to a principal maximum are separated a distance $\Delta\alpha = 2\pi/N$. Hence increasing N, decreases $\Delta\alpha$ and the peaks become narrower. Since $I(0) = I_0 \cdot N^2$ the intensity maxima also become brighter as more slits are illuminated. Finally note that between each pair of minima, we will also obtain a **subsidiary maximum** (for $N-1$ minima, we will have $N-2$ secondary maxima) whose intensity we could in principle understand again using **phasors**.

plotted for $I_0 = 1$

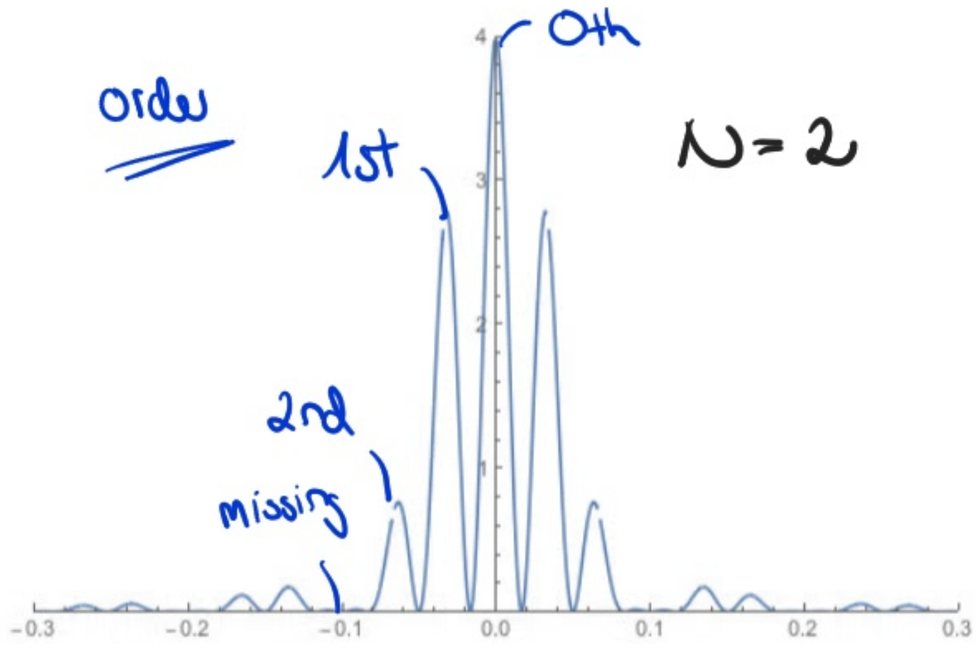
$b = 10\lambda$, $a = 30\lambda$



$N = 1$



$N = 4$



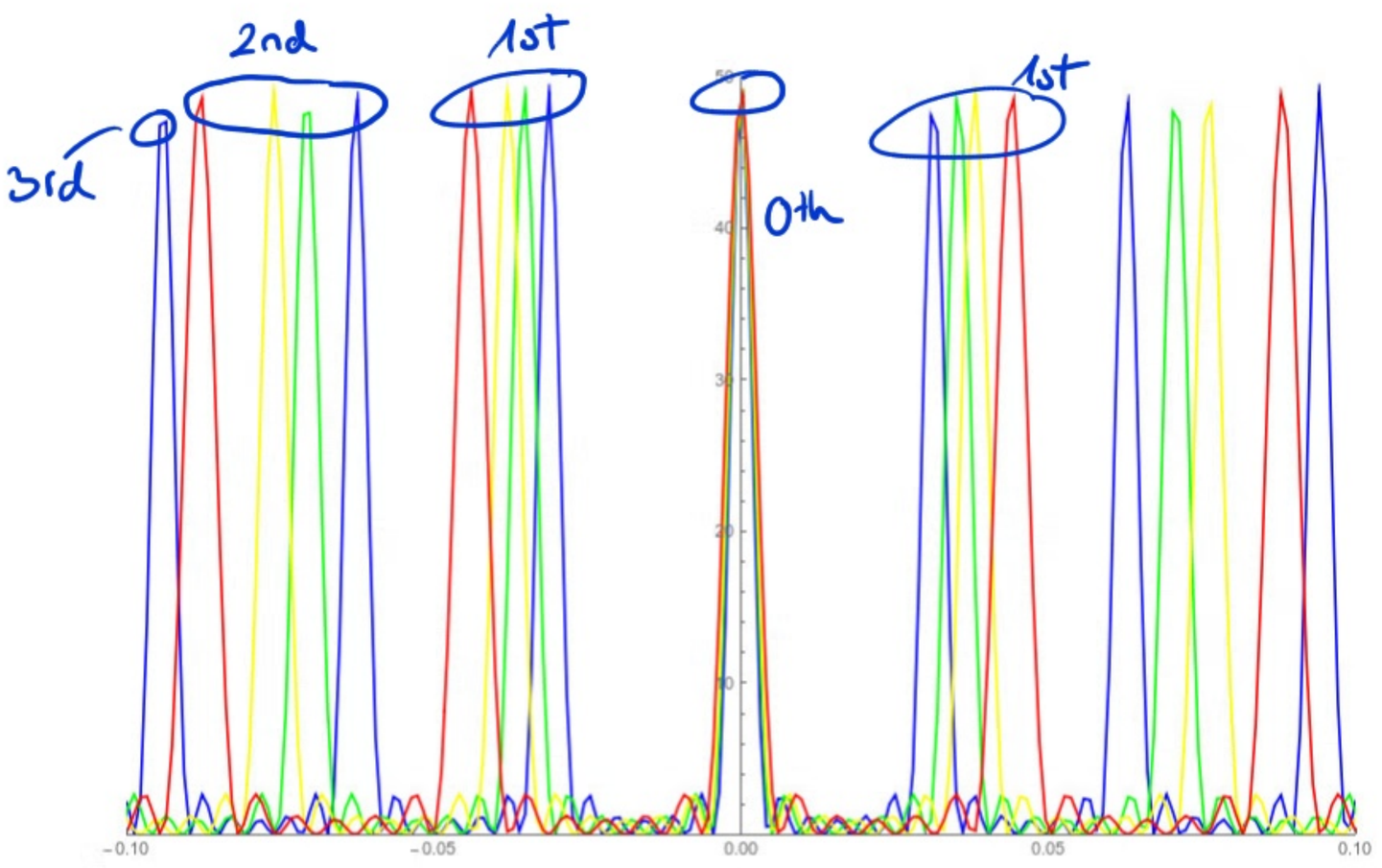
$N = 2$

4.) Diffraction grating

A repetitive array of diffracting elements (either apertures or other obstacles) that have the effect of producing periodic alterations in phase or amplitude are referred to as a diffraction grating. The simplest form is the multi-slit setup discussed in the last section. Since the amplitude is altered the wave propagates through the system, the configuration is referred to as a transmission amplitude grating. As discussed before the principal maxima occur at $a \sin\theta_m = m\lambda$, also known as the grating equation for normal incidence. Note that this depends on the wavelength λ, and thus suggests that we can analyse the characteristics of polychromatic light by studying the diffraction pattern. This is the main idea of grating spectroscopy. We can characterise the spreading-out of the colours by the dispersion

$$\mathcal{D} \equiv \frac{d\theta_m}{d\lambda} = \frac{m}{a\cos\theta_m} .$$

higher order & smaller a increases the dispersion



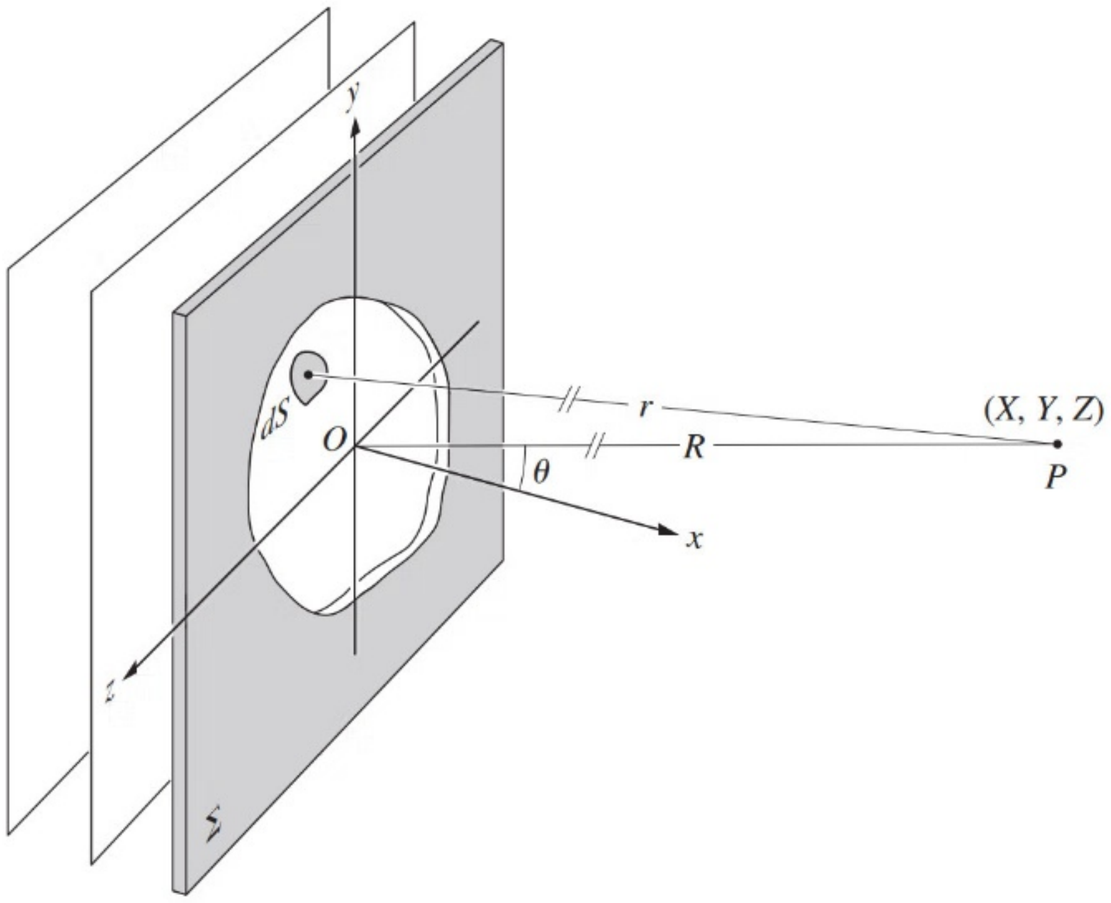
$N = 7$
$I_0 = 1$
$a = 15\mu m$

Phys 434 - Lecture 20

Fresnel Diffraction

1.) An arbitrary aperture in the far-field limit

Before we consider Fresnel diffraction in more detail, we are going to investigate the Fraunhofer diffraction pattern for an arbitrary aperture as illustrated below. According to the Huygens-Fresnel principle, we can determine the resulting flux density distribution at a point P, by envisaging a small area dS within the aperture covered by coherent point sources. Since dS is smaller than λ, all contributions at P remain in-phase and add constructively, so that we write



$\mathcal{E}_A$ is the source strength per unit area

$$d\mathcal{E} = \frac{\mathcal{E}_A}{r} e^{i(kr-\omega t)} dS, \qquad (**)$$

where $r = [X^2 + (Y-y)^2 + (Z-z)^2]^{1/2}$. As in Lecture 19, in the Fraunhofer limit we replace $1/r$ in $d\mathcal{E}$ by $1/R$ and expand for the phase

$$r = \left[\underbrace{x^2 + y^2 + z^2}_{= R^2} - (2yy + 2zz) + y^2 + z^2 \right]^{1/2}$$

$$= R\left[1 - 2(yy + zz)/R^2 + \underbrace{(y^2 + z^2)/R^2}_{\text{small for small aperture}}\right]^{1/2}$$

$$\approx R\left[1 - 2(yy + zz)/R^2\right]^{1/2}$$

$$\approx R\left(1 - \frac{yy + zz}{R^2}\right).$$

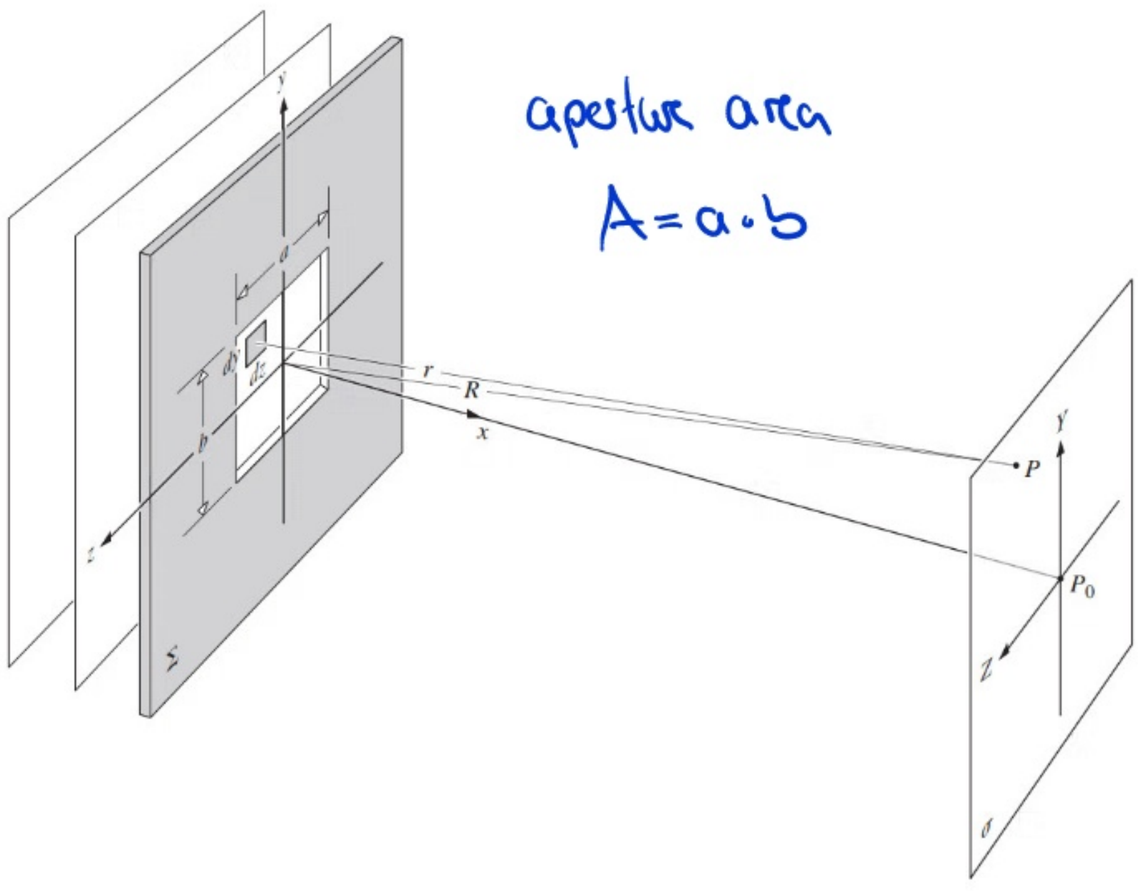
keeping the first order correction

The total disturbance is thus given by

integral over the aperture surface

$$E = \iint dE\, dS = \frac{\mathcal{E}_A}{R} e^{i(kR - \omega t)} \iint e^{-ik(yy + zz)/R}\, dS .$$

If we consider for example the rectangular case shown below, we have



aperture area

$A = a \cdot b$

$$E = \frac{\mathcal{E}_A}{R} e^{i(kR - \omega t)}$$

$$\int_{-b/2}^{b/2} e^{-iky\,y/R}\, dy \int_{-a/2}^{a/2} e^{-ikz\,z/R}\, dz$$

$$= \frac{\mathcal{E}_A}{R} e^{i(kR - \omega t)}$$

$$b \operatorname{sinc}\beta \cdot a \operatorname{sinc}\alpha ,$$

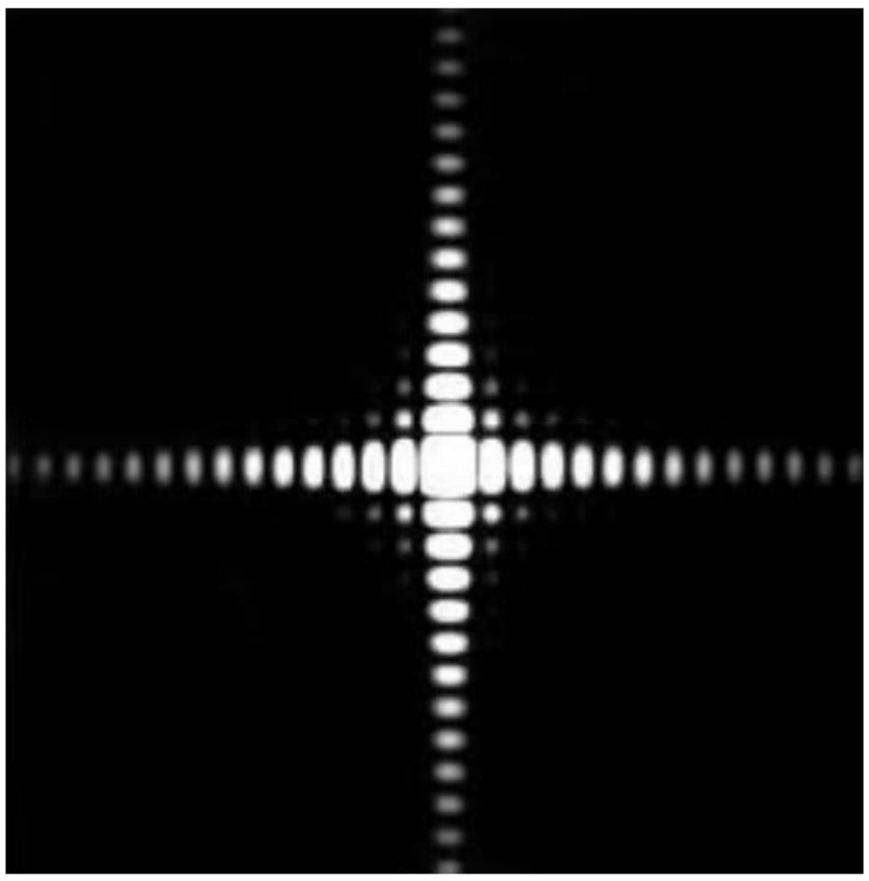
where we define $\beta \equiv ky\,b/2R$ and $\alpha \equiv kz\,a/2R$. With $A \equiv a \cdot b$

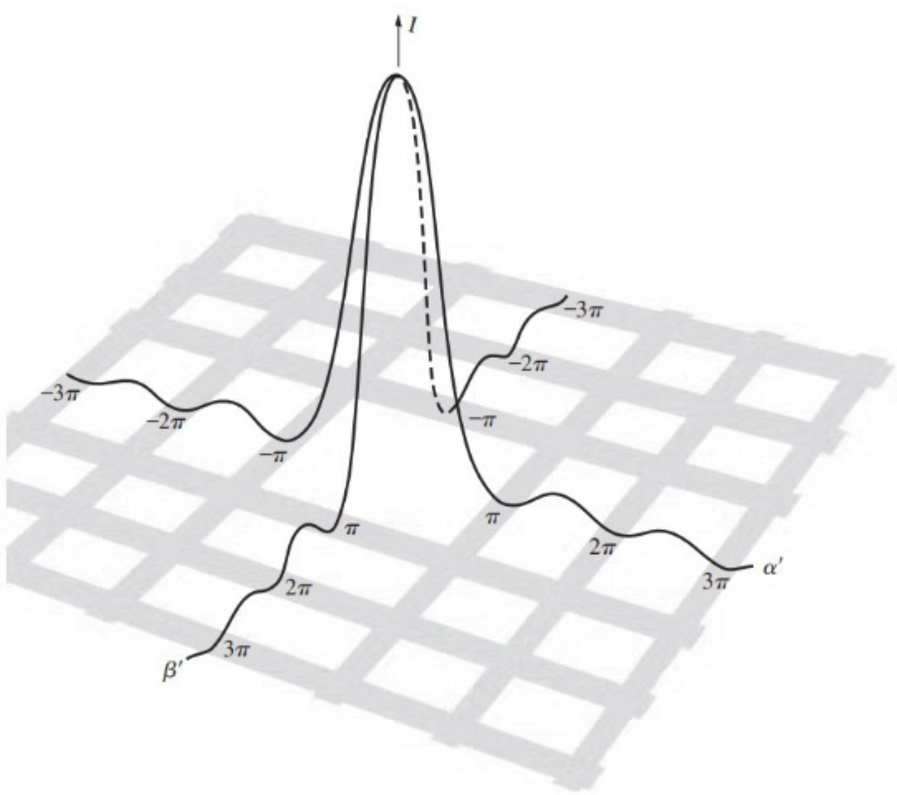
-the irradiance becomes with $I_0 \equiv \mathcal{E}_1^2 A^2 / 2R^2$

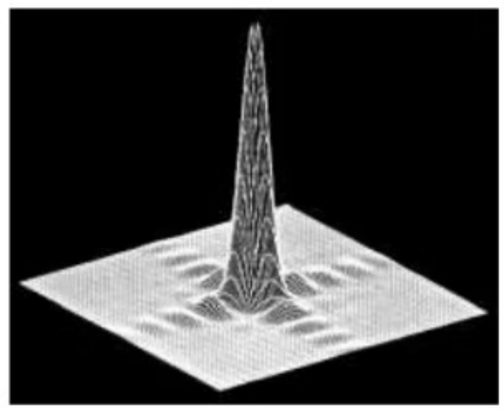
$$I(Y,Z) = I_0 \, \mathrm{sinc}^2\alpha \cdot \mathrm{sinc}^2\beta \; .$$

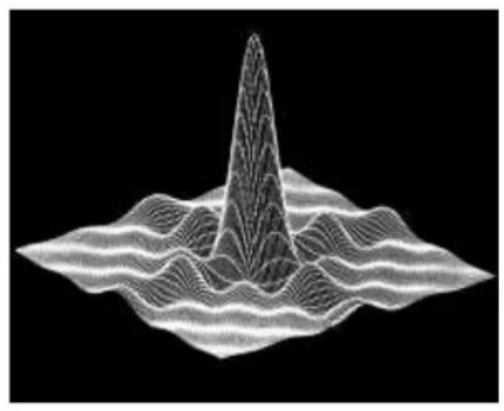
a narrow aperture in z-direction produces a wide diffraction pattern and vice versa !!

The resulting diffraction pattern for a square aperture is illustrated as









2.) General diffraction problem

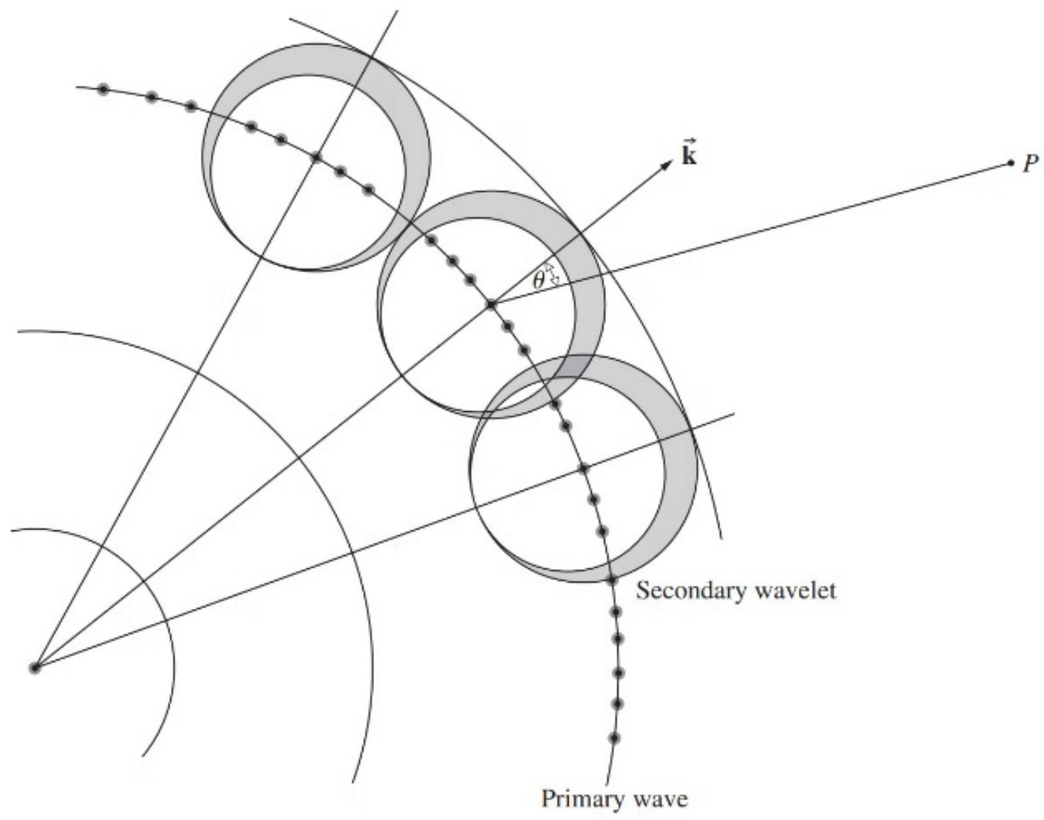
The approximations that led to the Fraunhofer diffraction pattern are no longer applicable, when we study the near-field regime. While it is still reasonable to assume that dE will be some generalised form of Eqn. (*), we have to account for shortcomings of the HF principle, namely that it predicts a backward propagating wave that is not observed. To modify the radiation pattern of a secondary emitter accordingly, we introduce the obliquity or inclination factor $K(\theta)$, which accounts for the wavelets' directionality:

$$K(\theta) = \frac{1}{2}(1+\cos\theta),$$

this can be derived within Kirchhoff's scalar theory, which is beyond the scope here

where θ is the angle between the normal of the primary wave front and the direction to P as shown here.

$K(\theta)$ is maximal in forward direction with $K(0)=1$ and vanishes in the opposit direction, $K(\pi)=0$.



We can now apply this concept to study the spherical wave emitted by a point source S positioned at the origin, which encounters a barrier or aperture and is subsequently observed at a point P. We obtain

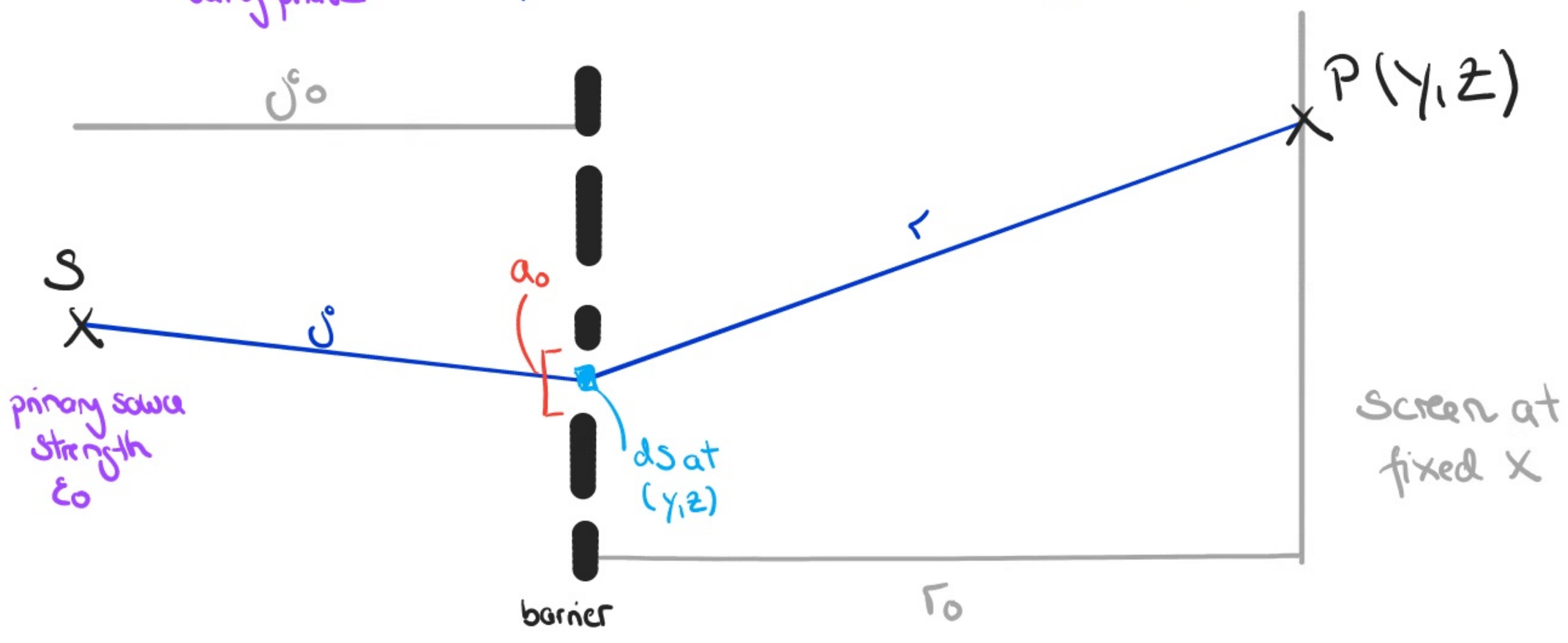
$\frac{\mathcal{E}_0}{\delta}$ ≙ primary's strength at dS

$$dE_P = -\frac{i\,\mathcal{E}_0}{\delta\lambda}\,\frac{K(\theta)}{r}\,e^{i(k(\rho+r)-\omega t)}\,dS. \qquad (**)$$

i → secondary and primary are $\pi/2$ out of phase

$\frac{\mathcal{E}_0}{\delta\lambda}$ source strength per unit area in dS

λ → required to reach correct free propagation amplitude



Nok that Eqn. (***) can only be applied if polarisation effects are negli- and is not applicable to describe the diffraction within a few wavelengths of the aperture. Nonetheless, the relation is very useful to analyse the resulting pattern. In most physical cases we are further typically interested in situations, where $r_0 \gg a_0$ and $f_0 \gg a_0$, where a_0 is the aperture's radial dimension. Assuming that we are interested in small ϑ, we can approximate

(i) $K(\vartheta) = 1/2(1+\cos\vartheta) \approx 1$,

(ii) $$f = [f_0^2 + y^2 + z^2]^{1/2} \approx \underbrace{f_0 [1 + (y^2+z^2)/2f_0^2]}_{\text{for the exponent of exponential}} \approx \overbrace{f_0}^{\text{for the denominator}},$$

(iii) $$r = [r_0^2 + (Y-y)^2 + (Z-z)^2]^{1/2}$$

$$\approx r_0 [1 + (Y-y)^2/2r_0^2 + (Z-z)^2/2r_0^2] \quad \text{(exponential)}$$

$$= r_0 - (Yy+Zz)/r_0 + (y^2+z^2+y^2+z^2)/2r_0$$

$$\approx r_0 \,. \quad \text{(denominator)}$$

We thus obtain for the oscillators within dS

$$dE_p = \underbrace{-\frac{i\mathcal{E}_0}{\lambda f_0 r_0} e^{i(k(r_0+f_0)-\omega t)} e^{ik(y^2+z^2)/2r_0} e^{-ik(Yy+Zz)/r_0}}_{\text{independent of } y,z;\ \text{constant prefactor}} \cdot e^{ik/2(y^2+z^2)(1/r_0+1/f_0)} \; dy\, dz \,.$$

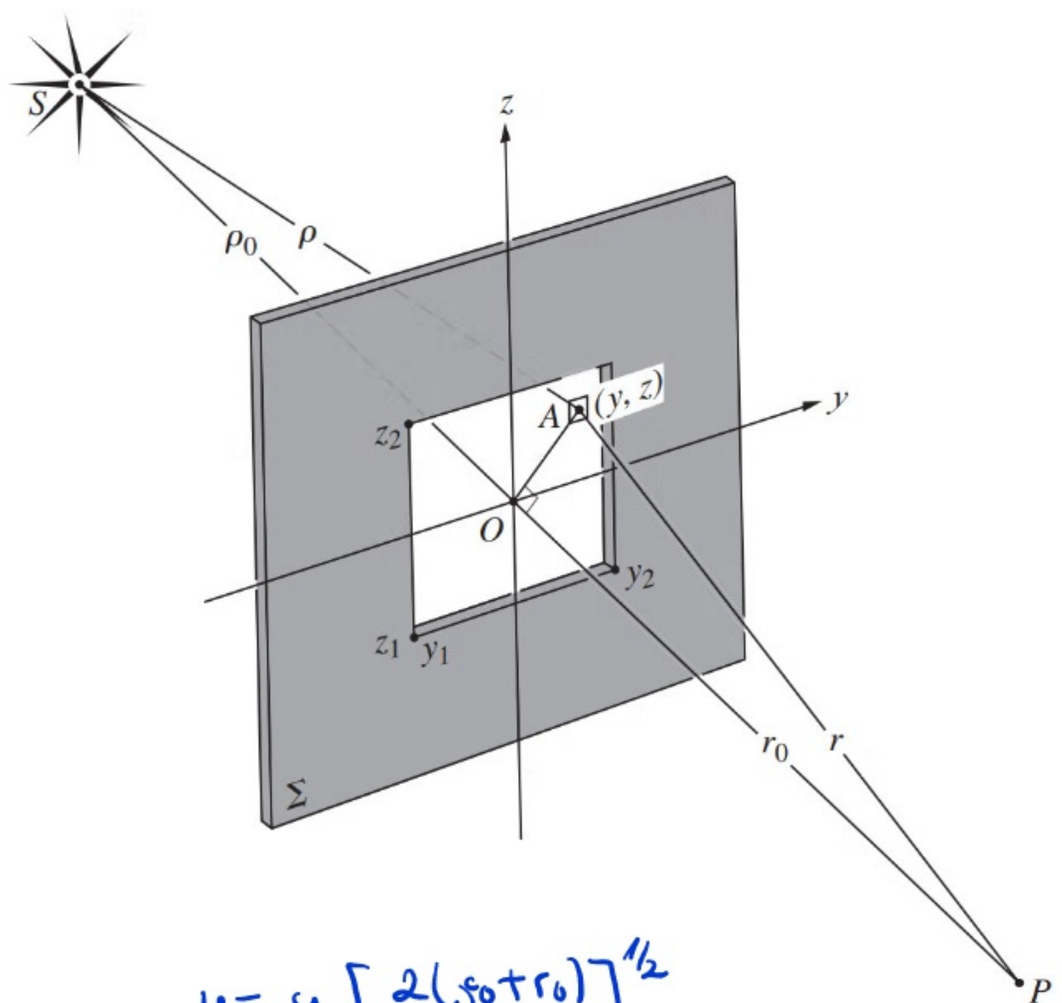
linear dependence on y,z → Fraunhofer term

quadratic dependence → Fresnel diffraction

Not that while the integrals associated with Fraunhofer diffraction are often analytically tractable, the quadratic dependence of Fresnel diffraction generally implies that the integrals over the aperture variables are more difficult and typically have to be evaluated numerically. Hence, to discuss behaviour of Fresnel diffraction analytically one would typically take $Y = Z = 0$ to focus on the observation point along the optical axis.

3.) The rectangular aperture

Consider the geometry on the right. Focusing on the observation point P along the x-axis, we obtain



$$E = -\frac{i\mathcal{E}_0}{\lambda s_0 r_0} e^{i(k(r_0+s_0)-\omega t)} \cdot \int_{y_1}^{y_2} dy \int_{z_1}^{z_2} dz \; e^{ik(y^2+z^2)\frac{s_0+r_0}{2s_0r_0}}$$

$$u \equiv y\left[\frac{2(s_0+r_0)}{\lambda r_0 s_0}\right]^{1/2}, \qquad v \equiv z\left[\frac{2(s_0+r_0)}{\lambda r_0 s_0}\right]^{1/2}$$

$$= \underbrace{-\frac{i\mathcal{E}_0}{2(s_0+r_0)} e^{i(k(r_0+s_0)-\omega t)}}_{\text{half the unobstructed amplitude } A_0 \text{ (phase doesn't matter for } I)} \int_{u_1}^{u_2} du\, e^{i\pi u^2/2} \int_{v_1}^{v_2} dv\, e^{i\pi v^2/2}.$$

This expression can be rewritten in terms of the so-called Fresnel integrals

$$C(x) \equiv \int_0^x \cos\left(\frac{\pi w^2}{2}\right) dw, \qquad S(x) \equiv \int_0^x \sin\left(\frac{\pi w^2}{2}\right) dw.$$

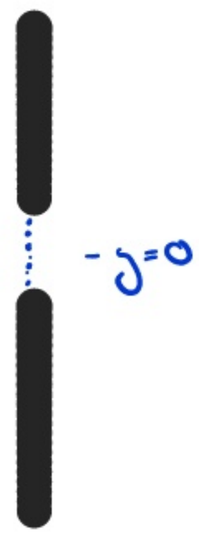
We find

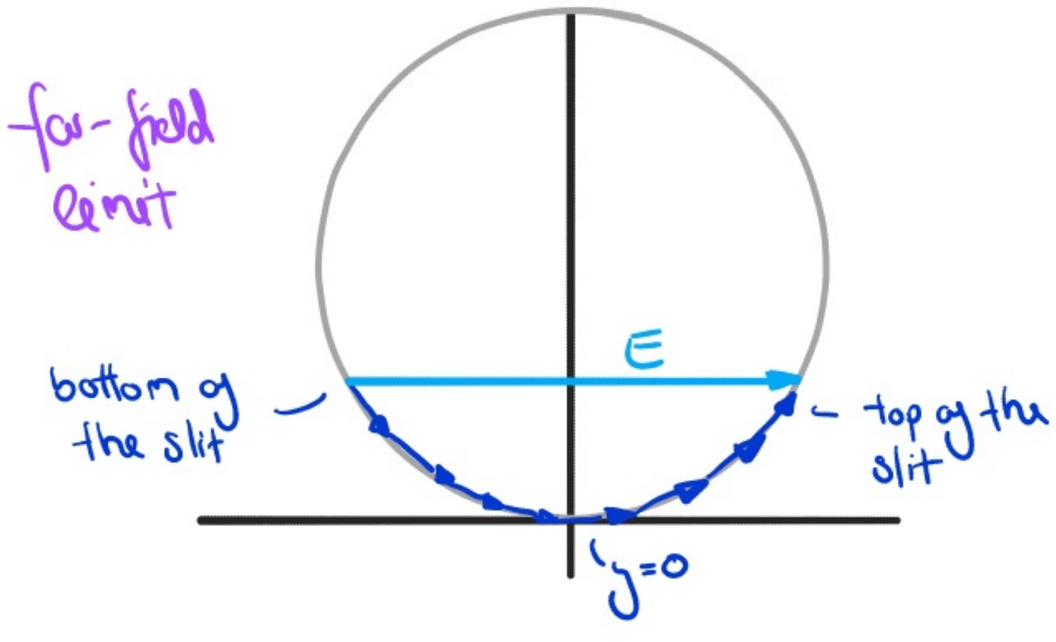
$$E = -\frac{iA_0}{2}\Big[C(x)+iS(x)\Big]_{u_1}^{u_2}\Big[C(x)+iS(x)\Big]_{v_1}^{v_2},$$

where $C(x)$ and $S(x)$ are tabulated special functions. As the maths is much more complicated when $Y, Z \neq 0$, we can instead build the overall diffraction pattern by keeping P fixed but moving the aperture around, resulting in different boundary values y_1, y_2, z_1, z_2 and thus u_1, u_2, v_1, v_2. The error in doing so is small provided that the aperture is displaced by distances $< \lambda_0$.

4.) Cornu spiral

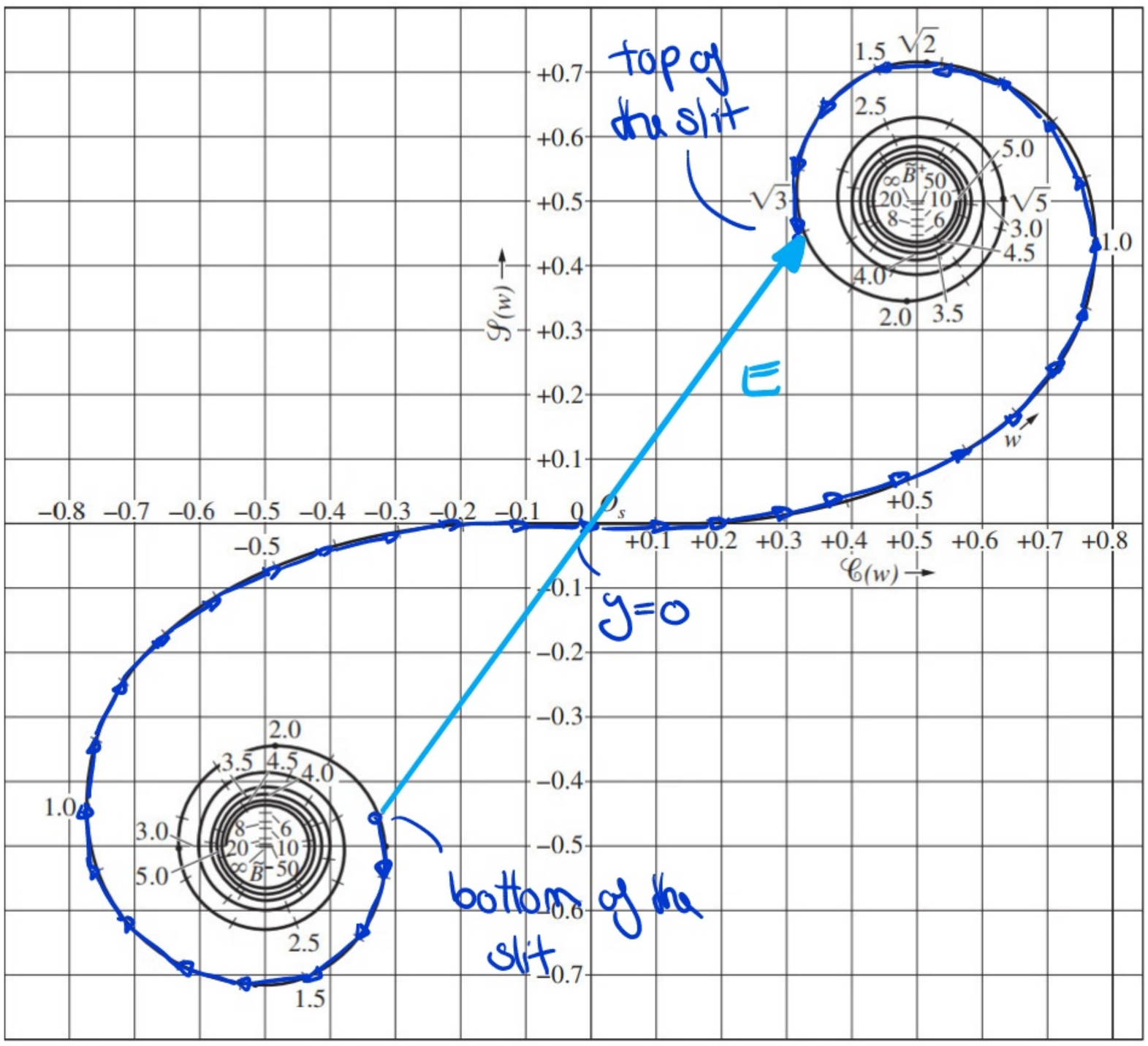
So far, we have analysed the Fresnel diffraction from a very mathematical point of view, but it is possible to gain some more insight using a geometrical approach, very similar to that of the vibration curve discussed in the last lecture. If we consider just one dimension of the aperture, e.g. the y-direction, the diffraction integral is essentially a sum of all the spherical wavelets from all the point sources in the aperture. In the Fraunhofer limit, the phase of each source varies linearly with y and there is a constant relative phase between two adjacent wavelets interfering at point P. Drawing the corresponding phasors resulted in a circle:





Nok that the start and end points as well as the radius of the circle depend on the aperture width and position of P.

In the Fresnel case however, the phase increases quadratically with y (faster than linear), so that the phasors will lie on a spiral shape, the so-called Cornu spiral. For an on-axis point P and a slit centred around $y=0$, we obtain

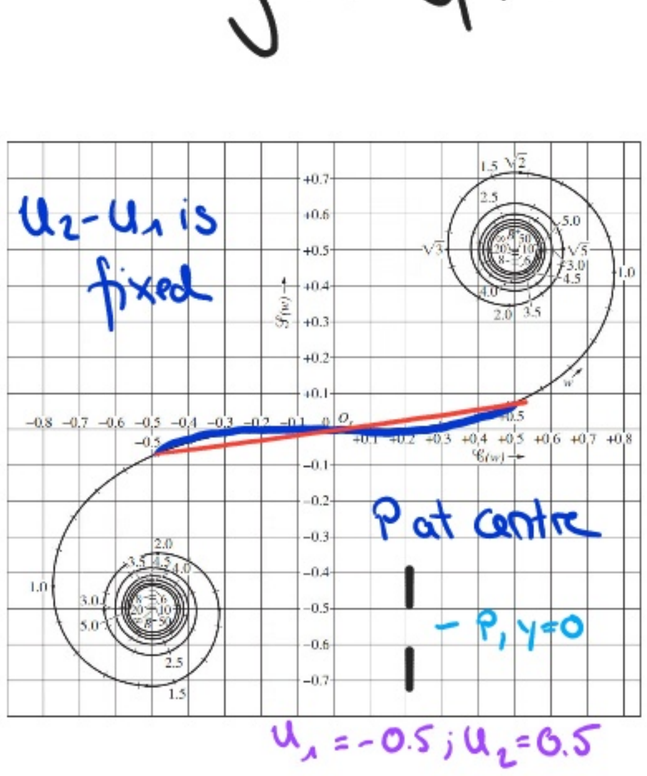


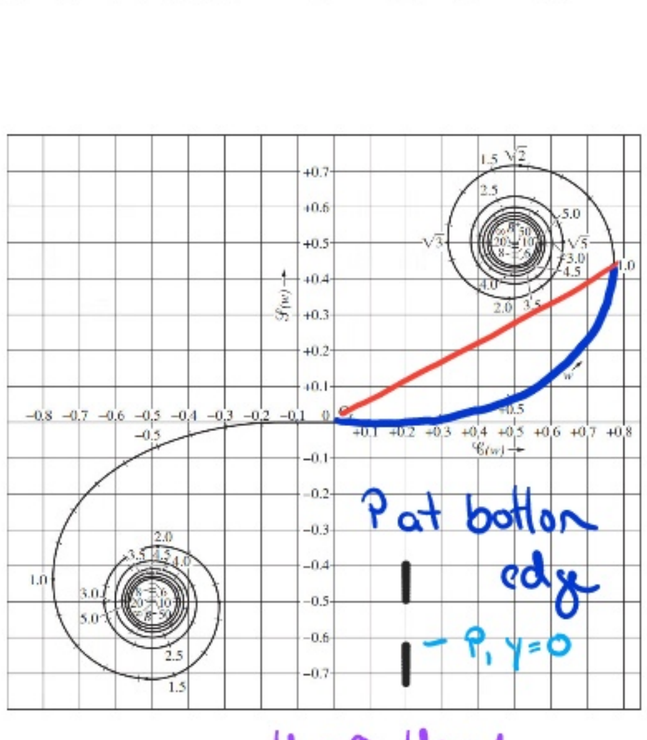
The wider the slit, the further up/down the phasors extend.

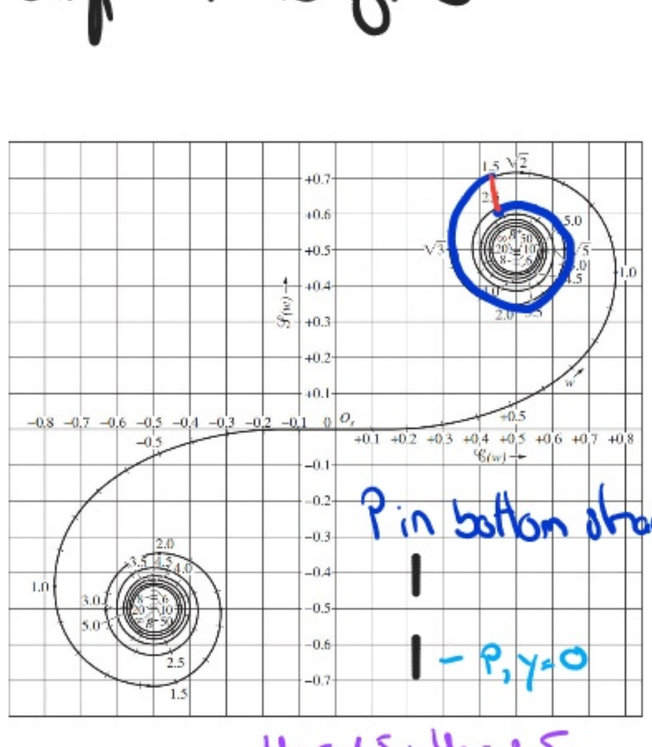
Moreover, we can get a quantitative prediction about the Fresnel pattern, because this spiral is essentially a plot of $\{C(x), S(x)\}$ for different values of x. Remember that the contribution to E from the y-direction was given by $[C(x)+i\,S(x)]_{u_1}^{u_2}$ so that we can define the corresponding phasor $\overrightarrow{D}_{12}(u)$ as the vector connecting the points on the spiral with $x=u_1$ and $x=u_2$; i.e. we can read off its magnitude/length using a rules. If we similarly define $\overrightarrow{D}_{12}(v) \equiv [C(x)+i\,S(x)]_{v_1}^{v_2}$, we can multiply the magnitudes of the two phasors together to obtain the total field amplitude and thus the overall intensity (where $I_0 = A_0^2/2$)

$$I(u,v) = \frac{I_0}{4}\,|\overrightarrow{D}_{12}(u)|^2\,|\overrightarrow{D}_{12}(v)|^2 \,.$$

As we examine different regions of the diffraction pattern by moving the aperture but keeping P fixed, we can visualise the diffraction integral as sliding the aperture end points along the Cornu spiral; i.e. for a fixed aperture size, we trace out arcs of the same length but positioned along different sections of the spiral. The vectors $\overrightarrow{D}_{12}$ thus have different lengths.

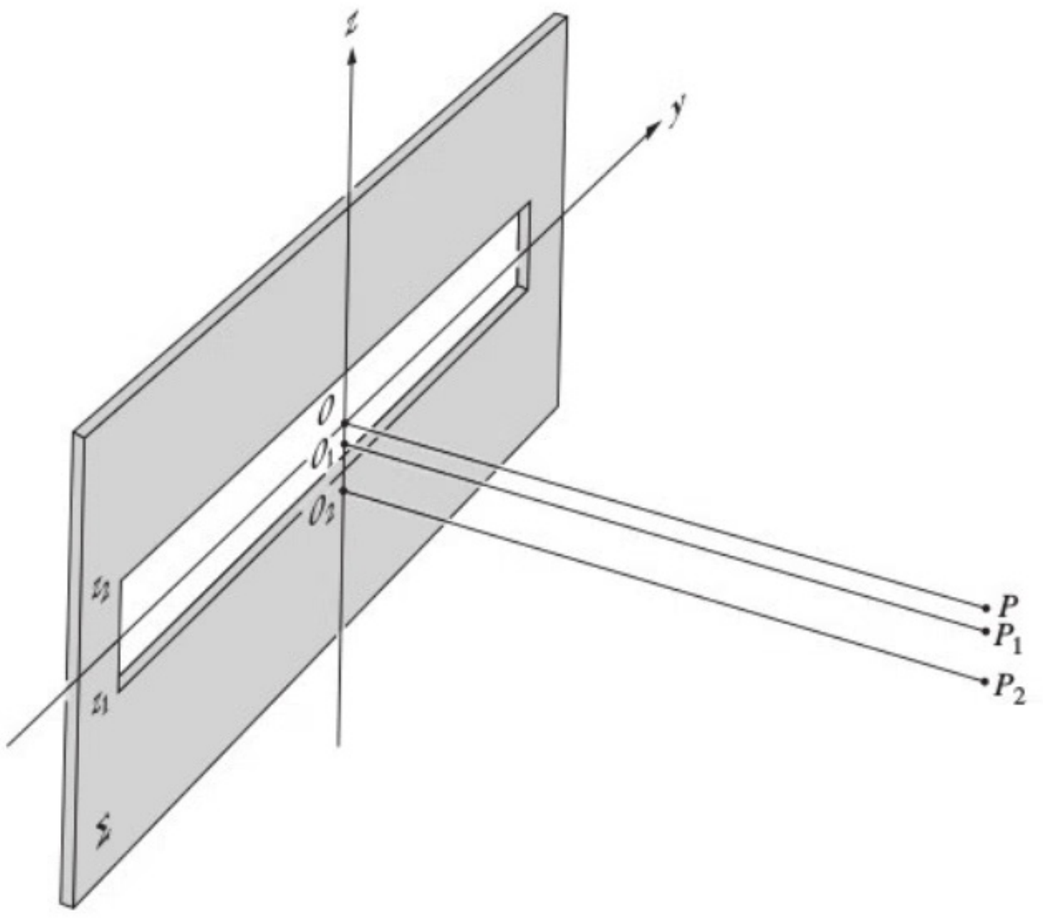






Note also that if we were to increase the aperture width, the end points of $\vec{B}_{12}$ would spiral counterclockwise toward their limiting values B^-, B^+. The phasors thus pass through a series of extrema, so that the intensity changes continuously in brightness. We can use this concept to illustrate the Fresnel diffraction pattern of a slit, i.e. for $u_2 \to \infty$, $u_1 \to -\infty$, we have

$$I(v) = \frac{I_0}{2} |\vec{B}_{12}(v)|^2 .$$

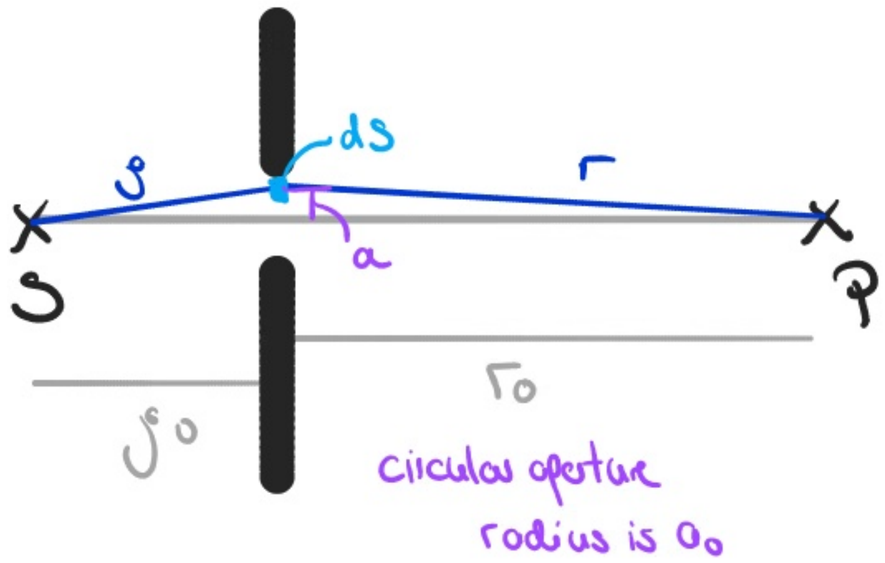


5.) Circular aperture

Now consider a point source observed through a spherical aperture from a point P on the axis. The contribution to the total electric field at P coming from the coherent point sources within the aperture will depend on their phase, determined by the optical path length $s+r$. Using the

Same approximations as before, we write

$$\rho + r \approx \rho_0 + r_0 + \frac{a^2}{2\rho_0} + \frac{a^2}{2r_0} .$$



The phase for all wavelets originating from a ring with radius a is thus

$$\varphi = k(\rho_0 + r_0) + k a^2/2R_0 , \text{ where } R_0 \equiv \frac{\rho_0 r_0}{\rho_0 + r_0} .$$

Taking again $K(\theta) \approx 1$, we obtain with $dS = 2\pi a\, da$

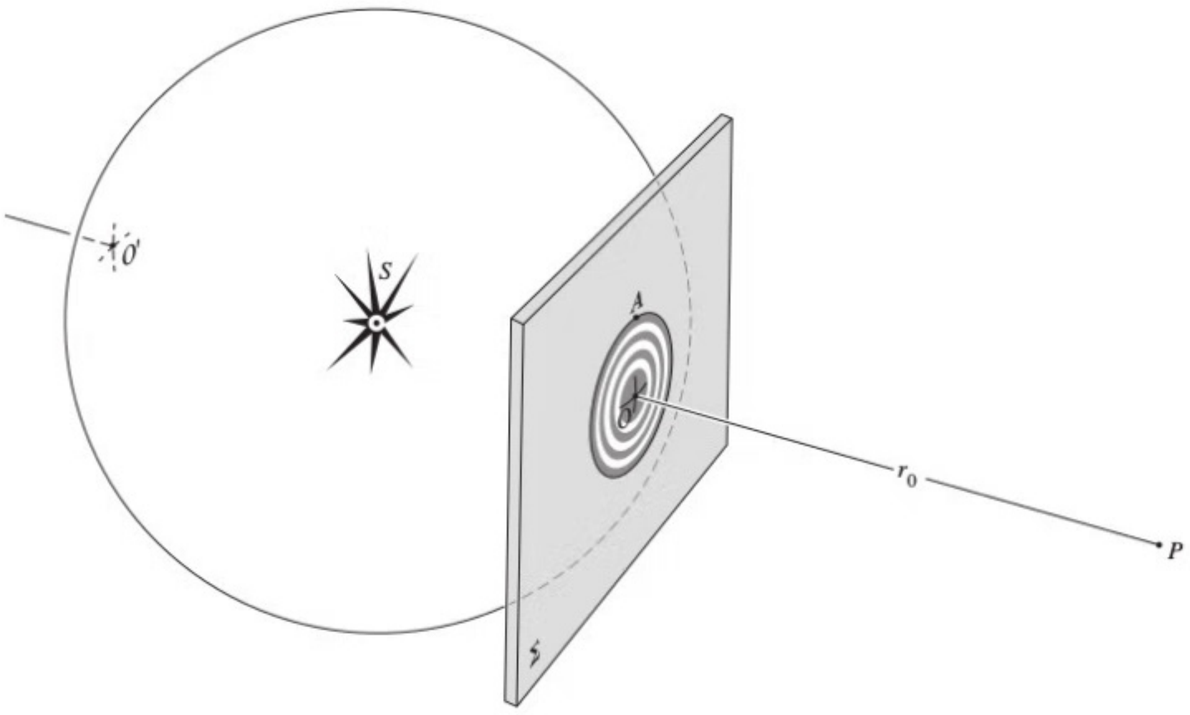
$$E = \frac{-i\mathcal{E}_0}{\rho_0 r_0 \lambda} e^{i(k(\rho_0 + r_0) - \omega t)} \int_0^{a_0} 2\pi a\, e^{ia^2 k/2R_0}\, da$$

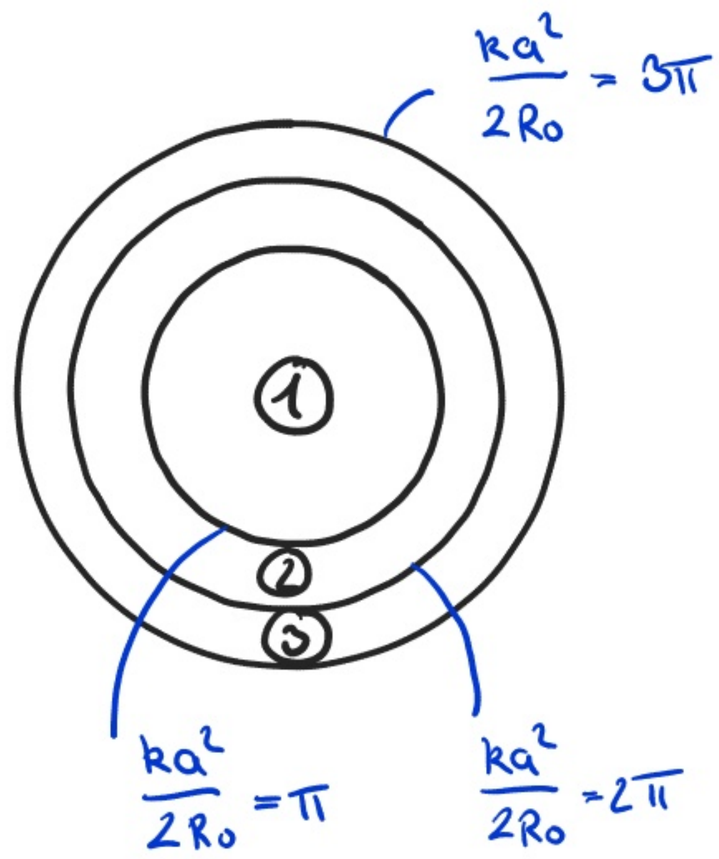
$$= i\frac{-i\mathcal{E}_0}{\rho_0 r_0 \lambda} e^{i(k(\rho_0 + r_0) - \omega t)} \frac{2\pi}{k} R_0 \left[1 - e^{ia_0^2 k/2R_0}\right]$$

$$= \frac{\mathcal{E}_0}{\rho_0 + r_0} e^{i(k(\rho_0 + r_0) - \omega t)} \left[1 - e^{ia_0^2 k/2R_0}\right] .$$

This expression oscillates around 0 as the aperture diameter or the distance r_0 is increased, corresponding to a series of bright fringes observed behind a circular aperture. We can intuitively understand the above integral as a sum of annular regions with different phase. These rings are conveniently separated into Fresnel zones, different regions over which

the phase changes by π, which we can illustrate as



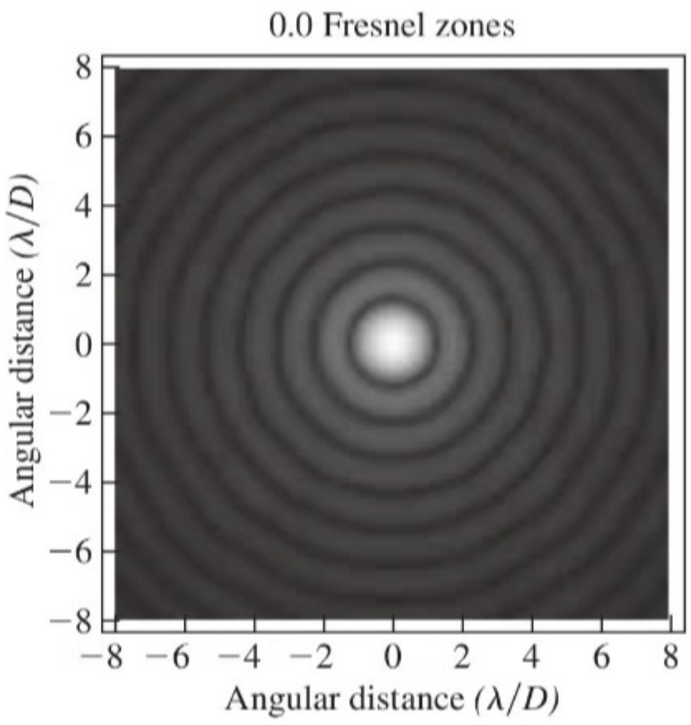


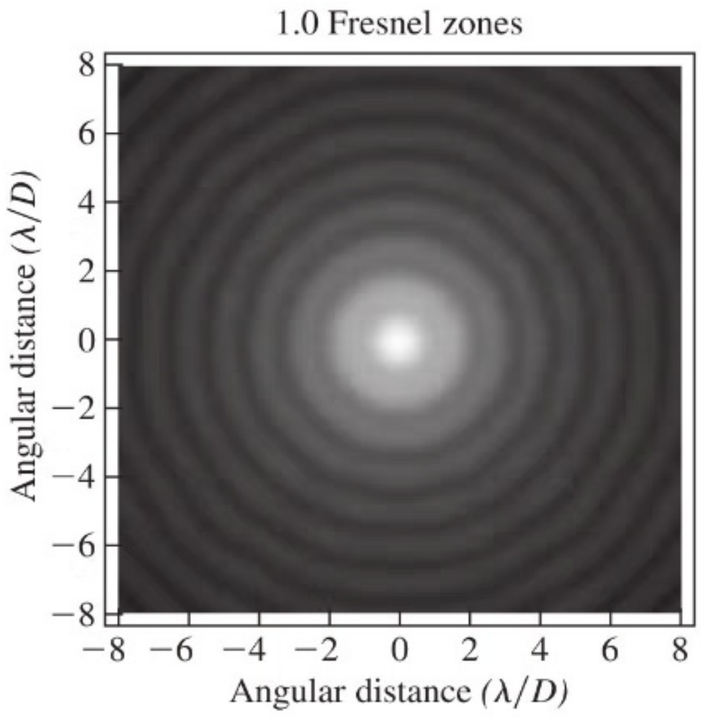
The outer radius of the mth Fresnel zone is given by

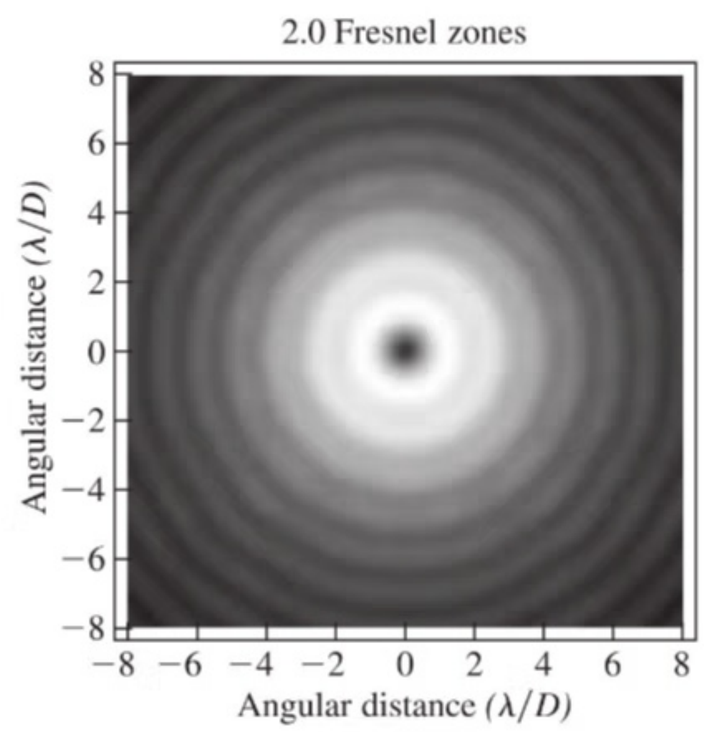
approximate because of our expansion for r

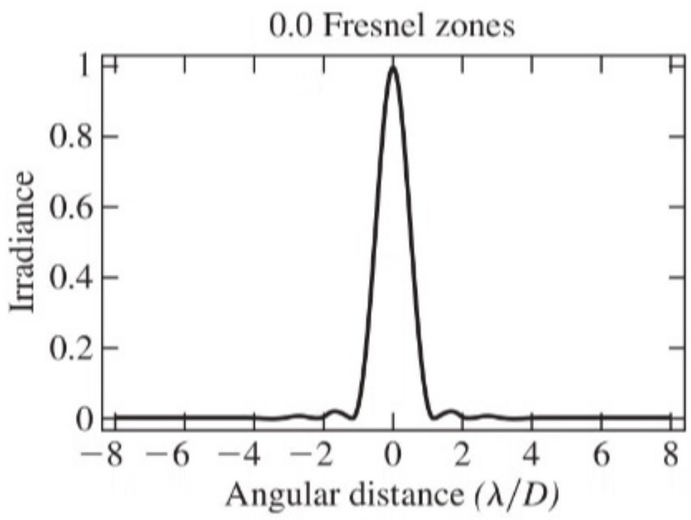
$$\frac{ka_m^2}{2R_0} \approx m\pi \quad \rightarrow \quad \underline{\underline{a_m \approx \sqrt{mR_0\lambda}}},$$

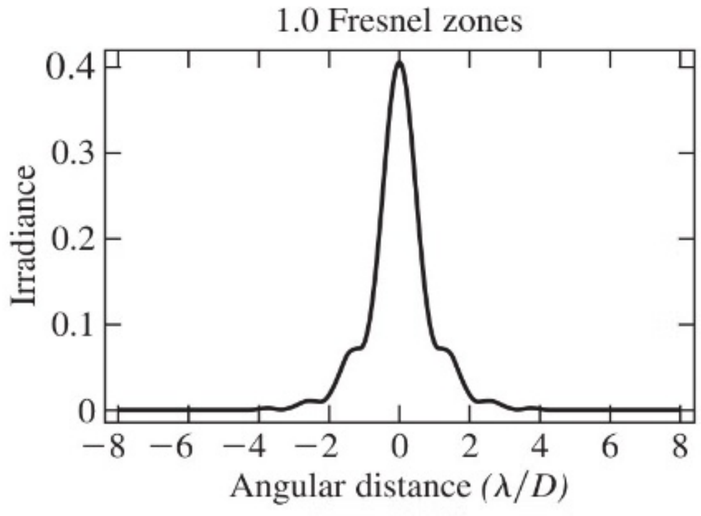
which implies that each zone has approximately the same area $\underline{\underline{A_F \approx \pi R_0 \lambda}}$. Moreover due to the relative phase shift, the amplitude from successive Fresnel zones as observed at point P approximately cancel. Hence, as you open up the aperture (make a_0 larger) we allow in successive Fresnel zones, leading to the observed oscillation in intensity at P. Note that the irradiance distribution for $m=0$ corresponds to the Fraunhofer pattern (see PS #6). The behaviour for $m=0, 1, 2$ is shown on the following page. Note that we can in principle understand the resulting irradiance pattern again by focusing on phasors; however instead of focusing on individual point sources, we use contributions from annular regions to construct the vibration curve.

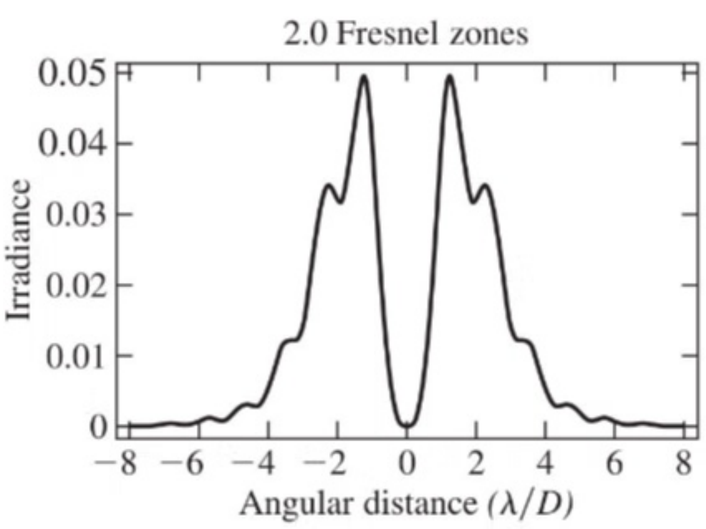

0.0 Fresnel zones
Angular distance (λ/D)
Angular distance (λ/D)


1.0 Fresnel zones
Angular distance (λ/D)
Angular distance (λ/D)


2.0 Fresnel zones
Angular distance (λ/D)
Angular distance (λ/D)


0.0 Fresnel zones
Irradiance
Angular distance (λ/D)


1.0 Fresnel zones
Irradiance
Angular distance (λ/D)


2.0 Fresnel zones
Irradiance
Angular distance (λ/D)

Phys 434 - Lecture 21

Fourier Transforms / Optics

1.) Introduction

In this lecture, we will extend the discussion of Lecture 11, where we have introduced Fourier analysis as a way to decompose a signal into its constituent harmonics. It, however, has a much wider application in Optics as Fourier analysis can be used to understand how optical systems process light to form images, i.e. how they alter amplitudes and phases.

We previously defined the Fourier and inverse Fourier transforms in one dimension as

$$F(k) = \mathcal{F}\{f(x)\} = \int_{-\infty}^{\infty} f(x)\, e^{ikx}\, dx,$$

or in terms of t, ω by substituting $x \to t$, $k \to \omega$

$$f(x) = \mathcal{F}^{-1}\{F(k)\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(k)\, e^{-ikx}\, dk.$$

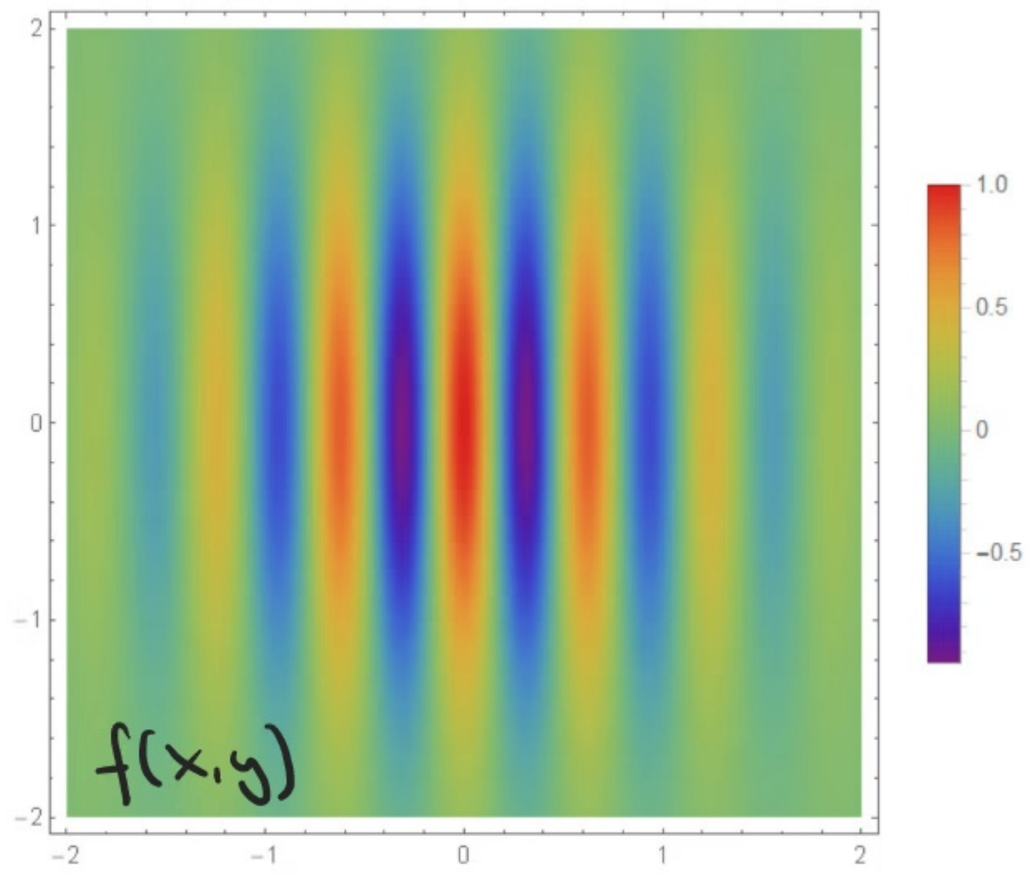
These definitions result in $\mathcal{F}\{F(k)\} = 2\pi f(-x)$ as a result of the prefactor and the different sign in the exponent (which is essentially related to the fact that two lenses will create an inverted image; see below).

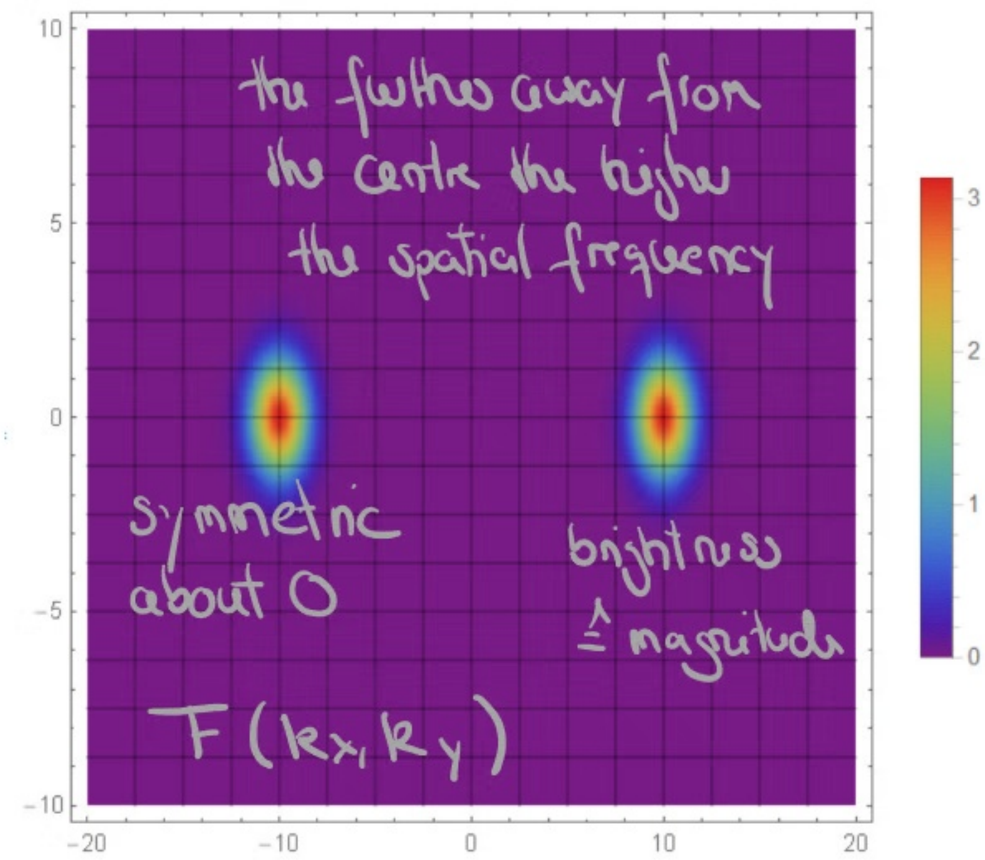
In optics, we typically deal with two-dimensional signals (such as the field across an aperture), so it is important to extend the notion of Fourier transforms to two dimensions, where we have

$$f(x,y) = \frac{1}{(2\pi)^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(K_x, K_y)\, e^{-i(K_x x + K_y y)}\, dK_x\, dK_y,$$

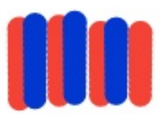
$$F(K_x, K_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, e^{i(K_x x + K_y y)}\, dx\, dy.$$

We can illustrate the two-dimensional transform by plotting the function $f(x,y) = \cos(10x)\, e^{-(x^2+y^2)/2}$ (cosine multiplied by a Gaussian) and its Fourier transform $F(K_x, K_y)$:





As we expect, we recover peaks in $F(k_x, k_y)$ concentrated around the spatial frequencies $K_y = 0$ and $K_x = \pm 10$ due to the cosine function in $f(x,y)$. Note that if the $f(x,y)$ pattern was sharper



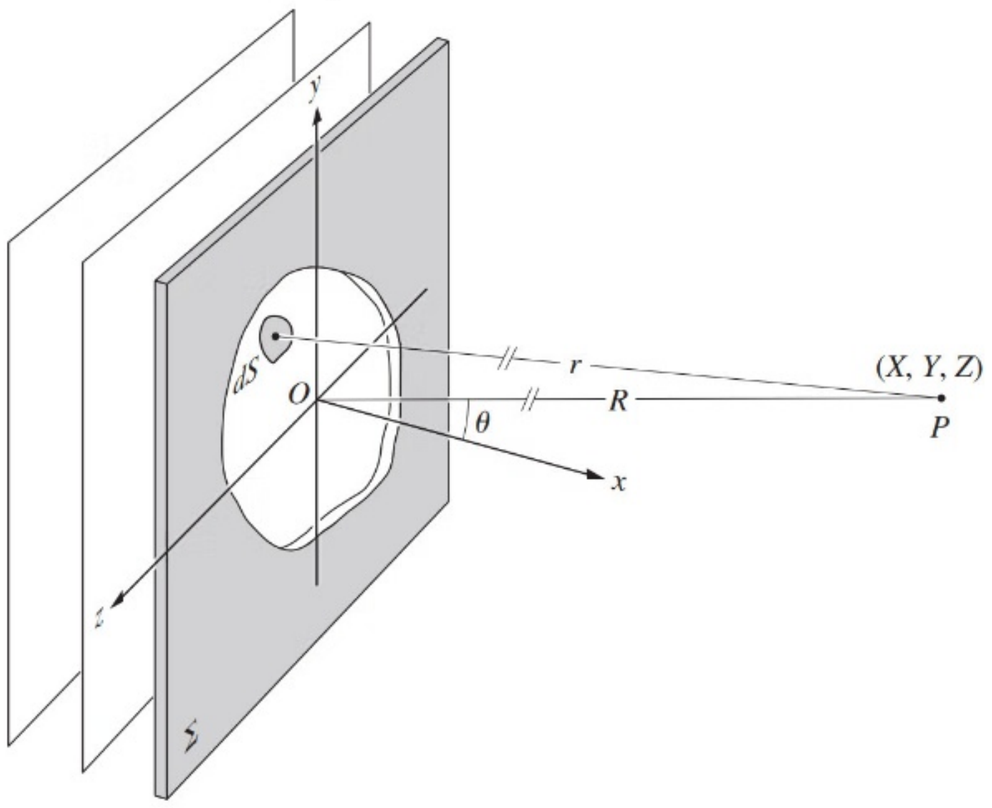
2

we would require higher spatial frequencies in the spectrum to adequately resolve the edges of the pattern. As in the one-dimensional case, $f(x,y)$ can essentially be constructed as a linear combination of plane waves, $e^{-i(K_x x + K_y y)}$, weighted in amplitude and phase by a complex factor $F(K_x, K_y)$.

2.) Fourier transforms & Fraunhofer diffraction

The idea that a signal can be understood in terms of its individual constituents is something we used to derive the Fraunhofer diffraction pattern, i.e. the emerging wave-front behind an aperture is a superposition of plane waves moving in different directions. The different 'Fourier' components make up the diffracted field as it emerges from an aperture. In Lecture 20, we found for the field generated by an arbitrarily shaped aperture

$$E = \frac{\mathcal{E}_A}{R} e^{i(kR-\omega t)} \cdot \iint_{\text{Aperture}} e^{-ik(Yy+Zz)/R}\, dy\, dz, \qquad (*)$$



where we assumed the source strength $\mathcal{E}_A$ to be constant over the entire aperture. However, in principle, $\mathcal{E}_A$ could vary as a function of y, z or even be complex if the barrier was made out of a dielectric or partially transmissive.

To account for such position-dependent optical path lengths (which would certainly affect the diffraction pattern), we combine $\mathcal{E}_A$ and the additional prefactors into a single complex quantity, the aperture function

$$\underline{A(y,z) = A_0(y,z)\, e^{i\phi(y,z)}},$$

where the amplitude of the field over the entire aperture is given by A_0 and point-to-point variations are captured by the $e^{i\phi}$. We can thus rewrite Eqn. (*) as

A is only non-zero over the aperture region

$$E(Y,Z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(y,z)\, e^{ik(Yy+Zz)/R}\, dy\, dz .$$

To see how this relates to the Fourier picture, we define the spatial frequencies $K_Y \equiv kY/R$ and $K_Z \equiv kZ/R$, which essentially implies that each point on the image plane corresponds to a specific spatial frequency. Moreover, now $E(Y,Z)$ (the total wave moving in the direction of Y,Z coming from all the possible locations $dy\,dz$) reads

$$E(K_Y,K_Z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(y,z)\, e^{i(K_Y y + K_Z z)}\, dy\, dz \quad . \quad (**)$$

The field distribution in the Fraunhofer diffraction pattern is the Fourier transform of the aperture function, $E(K_Y,K_Z) = \mathcal{F}\{A(y,z)\}$.

Or equivalently, the field distribution in the image plane is the spatial-frequency spectrum of the aperture function. The inverse transform is

$$A(y,z) = \mathcal{F}^{-1}\{E(K_y,K_z)\} = \frac{1}{(2\pi)^2}\iint_{-\infty}^{\infty} E(K_y,K_z)\, e^{-i(K_y y + K_z z)}\, dK_y dK_z .$$

This explains again that a narrow aperture corresponds to a spread-out diffraction pattern, because a narrow signal has a broad transform and vice versa.

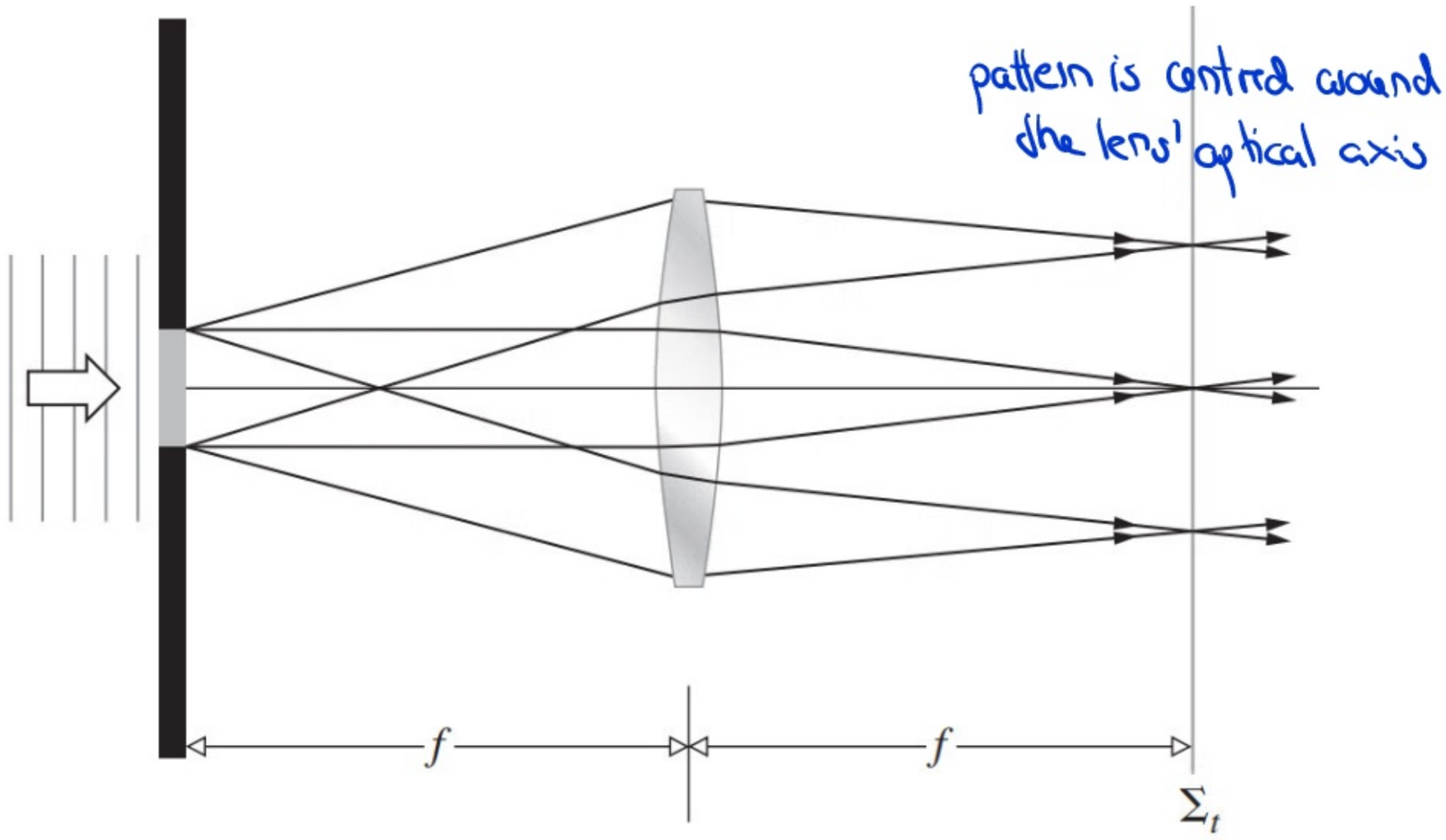
3.) The lens as a Fourier analyser

We have discussed before that the Fraunhofer pattern, in principle, only becomes visible for parallel rays, converging at infinity. To actually observe the resulting diffraction pattern on a screen, we require a converging lens that collects parallel bundles of rays and combines them in the back focal plane. If we place a screen at this position, the so-called 'transform plane', we observe the far-field diffraction pattern of the obstacle spread across it. The lens thus acts as a Fourier transformer in 2D.

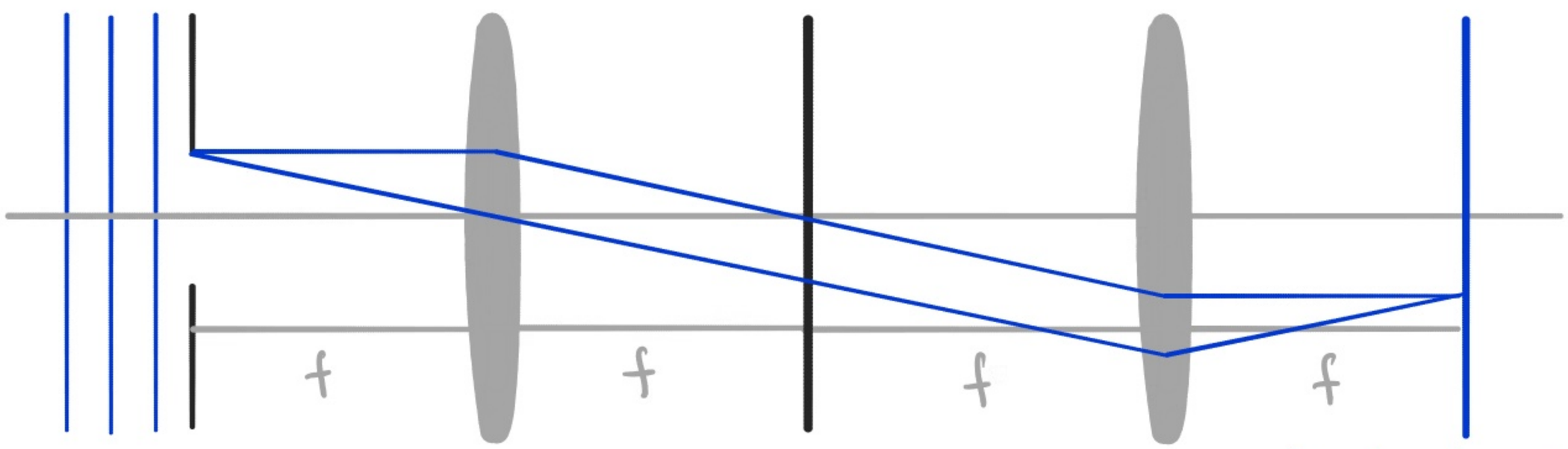
Note that it is important that the screen is positioned exactly a focal length f away from the lens, otherwise the Fraunhofer pattern will not be visible. If we are only interested in the intensity on the screen then the distance between the object and the lens does not matter. If however we want to

obtain the correct phase (i.e. $\vec{E}$-field), then the obstacle needs to be positioned in the lens' front focal plane; only then do the different diffracted waves maintain their phase relationships, travelling equal optical path lengths to the transform plane as illustrated below.

equal path lengths (phase preservation) if the lens is positioned a distance f from the aperture and the screen



Question: What happens in this scenario?



aperture screen
$A(y,z)$

transform plane
$\mathcal{F}\{A(y,z)\}$

transform plane of the transform plane
$\mathcal{F}\{\mathcal{F}\{A(y,z)\}\} \propto A(-y,-z)$

As we would obtain from geometric optics, the image of the aperture would be inverted, i.e. $\pi_T = -1$.

4.) A few useful properties

Motivated by what we have seen so far, we will discuss a few more relevant examples and properties of Fourier transforms:

- **Dirac delta function** (allows us to represent a point source as well as build up images of many point sources forming on object):

$$\delta(x) = \begin{cases} 0 & x \neq 0 \\ \infty & x = 0 \end{cases} \quad \text{or} \quad \int_{-\infty}^{\infty} \delta(x - x_0) f(x)\, dx = f(x_0),$$

($f(x)$: some continuous function; 'sifting' property)

$\Rightarrow$ Fourier representation: $\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ikx}\, dk,$

which suggests $\delta(x) = \mathcal{F}^{-1}\{1\}$ and so $\mathcal{F}\{\delta(x)\} = 1$; i.e. the sharp peak in δ requires all spatial frequencies in the spectrum.

- **Displacements and phase shifts**: Suppose that $F(k) = \mathcal{F}\{f(x)\}$, so what is $\mathcal{F}\{f(x - x_0)\}$ (shifted by x_0 to the right)? Calculate

$$\mathcal{F}\{f(x - x_0)\} = \int_{-\infty}^{\infty} f(x - x_0)\, e^{+ikx}\, dx$$

(substitute $x - x_0 = s$)

$$= \int_{-\infty}^{\infty} f(s)\, e^{+ik(x_0 + s)}\, ds = e^{+ikx_0} \int_{-\infty}^{\infty} f(s)\, e^{+iks}\, ds,$$

which gives $\mathcal{F}\{f(x-x_0)\} = e^{\pm ikx_0}\mathcal{F}\{f(x)\}$; hence the Fourier transform of a function displaced in space (or time) is the transform of the undisplaced function multiplied by an exponential linear in x_0. This will be useful to study the image formed by many point sources.

- **Composite functions**: Suppose that $f(x,y) = g_1(x,y) + g_2(x,y)$, since the Fourier transform is linear in f, we obtain $F(k_x, k_y) = G_1(k_x, k_y) + G_2(k_x, k_y)$. This is particularly useful to determine the shape of a Fourier transform if you can decompose f into functions whose Fourier transforms you already know.

- **Convolution theorem**: Suppose we have two functions $f(x)$ and $g(x)$ with the Fourier transforms $F(k) = \mathcal{F}\{f(x)\}$ and $G(k) = \mathcal{F}\{g(x)\}$, then $G(k)F(k) = \mathcal{F}\{f \circledast g\}$, where $f \circledast g$ denotes the convolution integral between f and g

transform of the convolution is the product of transforms

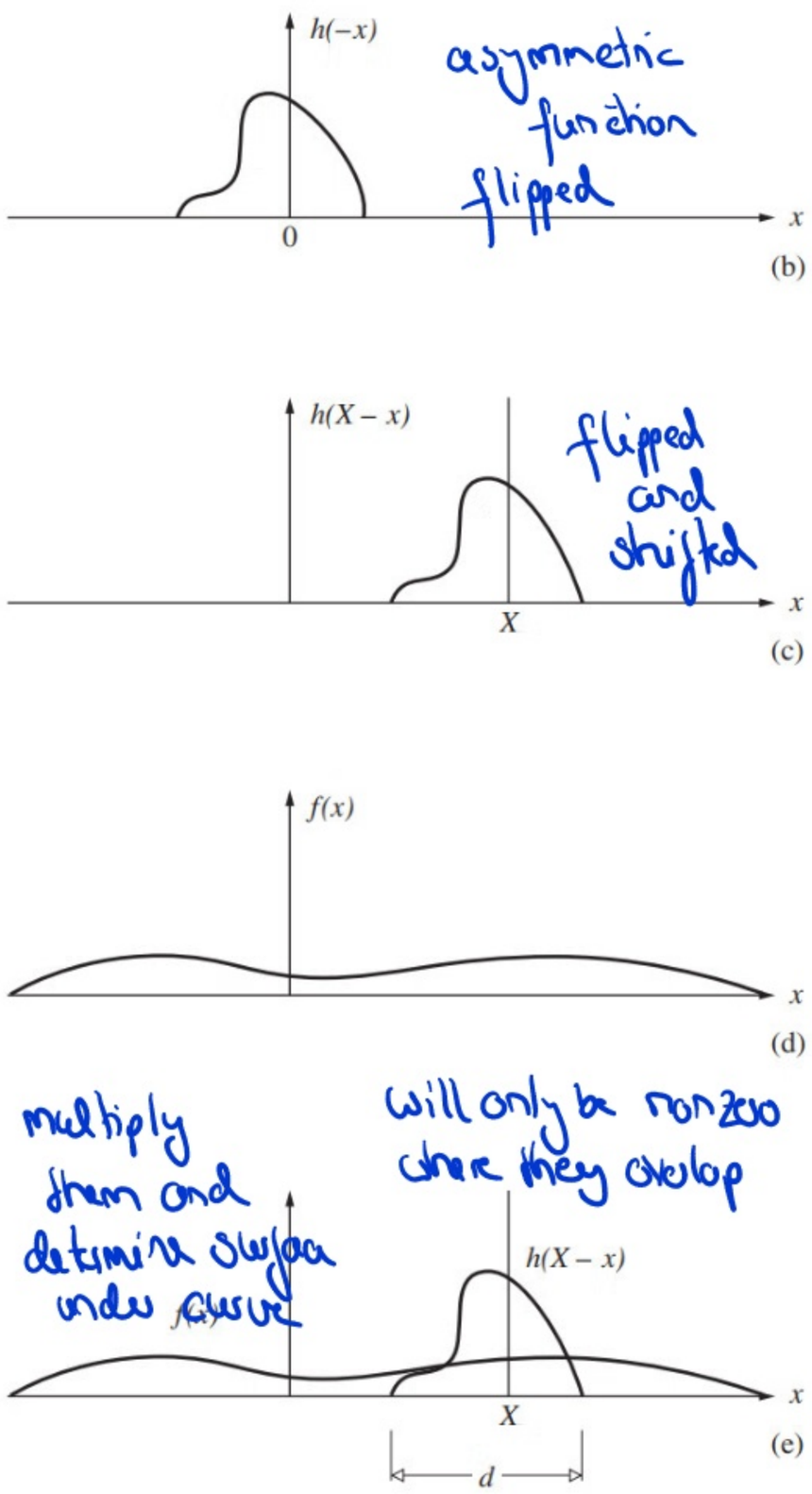
$$f \circledast g \equiv \int_{-\infty}^{\infty} f(\eta)\, g(x-\eta)\, d\eta .$$

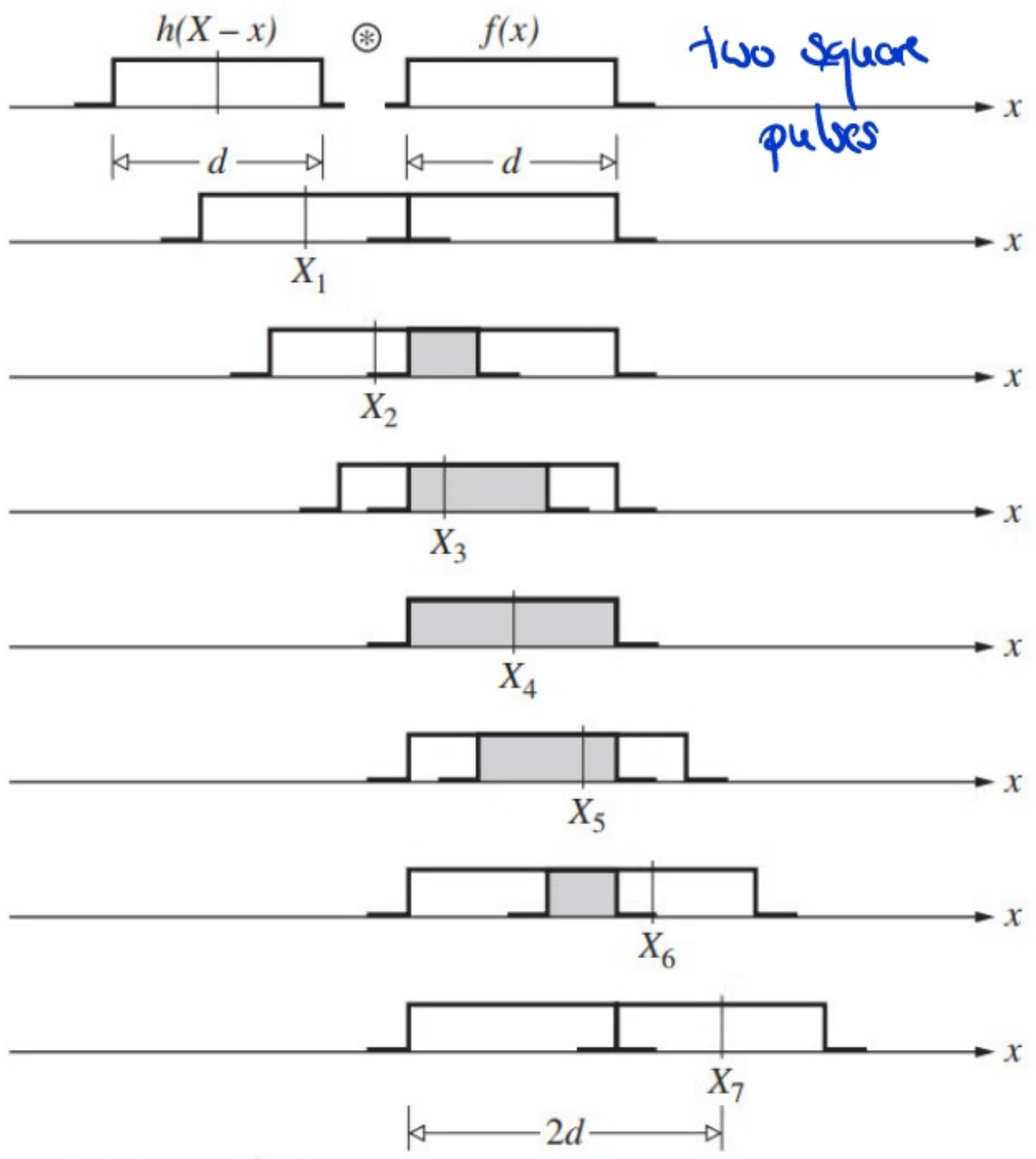
Conversely, we have $\mathcal{F}\{f(x)\,g(x)\} = 1/2\pi\, F \circledast G$, where

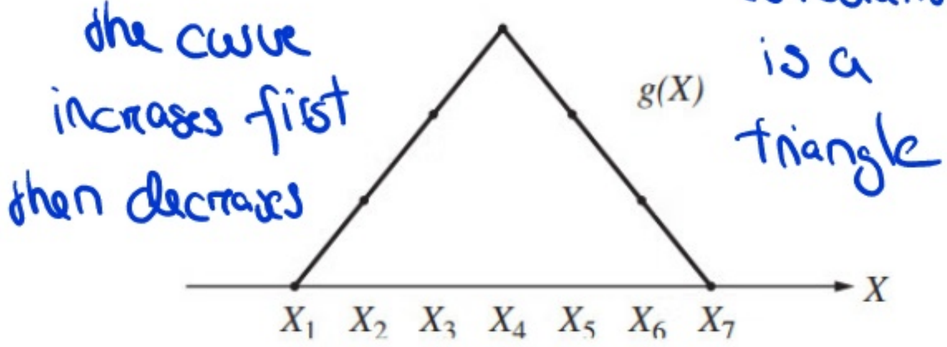
transform of the product is the convolution of the transforms

$$F \circledast G \equiv \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\gamma)\, G(k-\gamma)\, d\gamma .$$

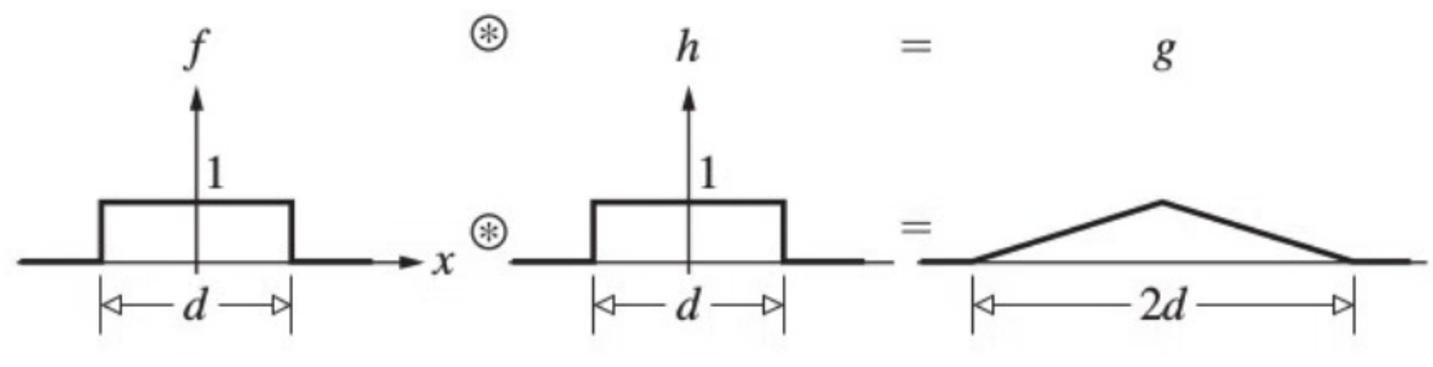
The convolution process can be illustrated as

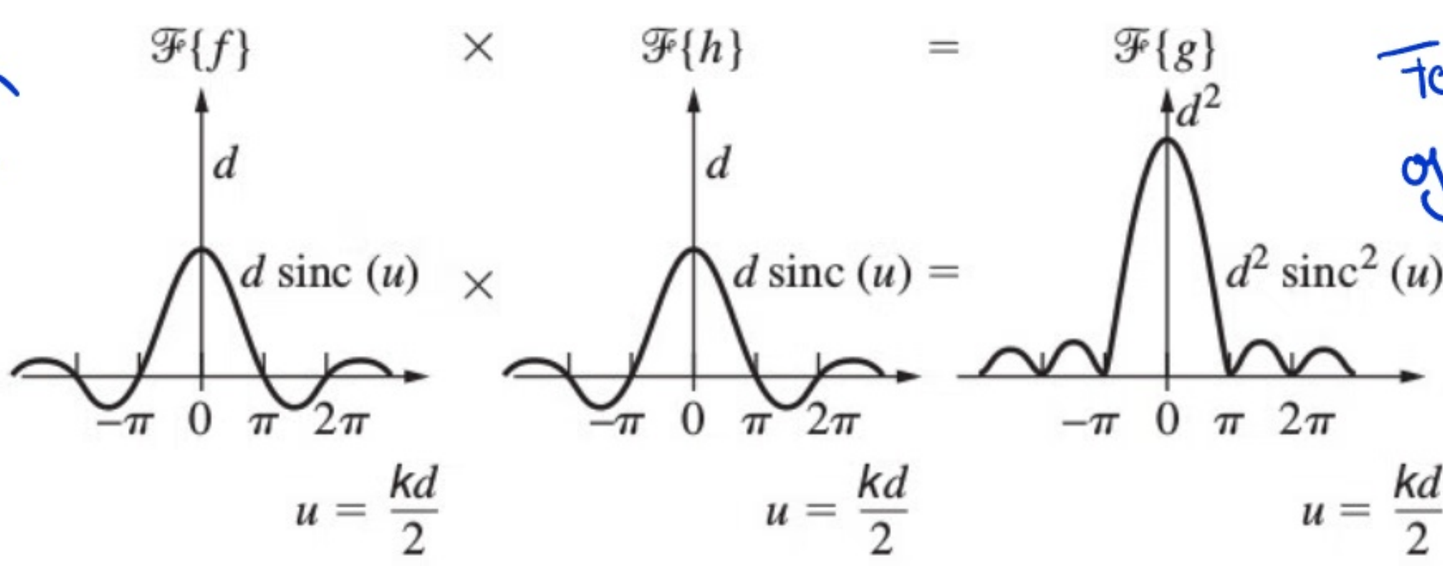






The convolution theorem thus becomes





5.) Array theorem

With our mathematical knowledge, we can now revisit all the Fraunhofer diffraction patterns, we calculated before and interpret them in terms of the Fourier picture. Consider a long slit illuminated by a plane wave, where

$$A_1(z) = \begin{cases} A_0 & \text{for } |z| \leq b/2, \\ 0 & \text{otherwise.} \end{cases}$$

rectangular opening with width b

Eqn. (**) then leads to (assuming a 1D problem)

$K_z = \frac{kz}{R}$

$$E(K_z) = \mathcal{F}\{A_1(z)\} = A_0 \int_{-b/2}^{b/2} e^{iK_z z}\, dz$$

$$= \underline{\underline{A_0 b\, \mathrm{sinc}(K_z b/2)}}.$$

The Fourier perspective can provide us with new insight into why the single-slit diffraction pattern remains as the envelope function as we add more slits. In the case of two slits, the aperture function can be written as the sum of two single slits separated by a:

$$A_2(z) = A_1(z - a/2) + A_1(z + a/2),$$

shift property

$$\Rightarrow \mathcal{F}\{A_2(z)\} = e^{-iK_z a/2}\, \mathcal{F}\{A_1(z)\} + e^{+iK_z a/2}\, \mathcal{F}\{A_1(z)\}$$

single-slit envelope stays

$$= 2\cos(K_z a/2)\, \mathcal{F}\{A_1(z)\}.$$

Alternatively, we can also write the two-slit aperture-function as a convolution of the single slit with two delta-functions, i.e.

$$A_2(z) = A_1(z) \circledast \left[\delta(z'-a/2) + \delta(z'+a/2)\right]$$

$$= \int_{-\infty}^{\infty} A_1(z') \left[\delta(z-z'-a/2) + \delta(z-z'+a/2)\right] dz'.$$

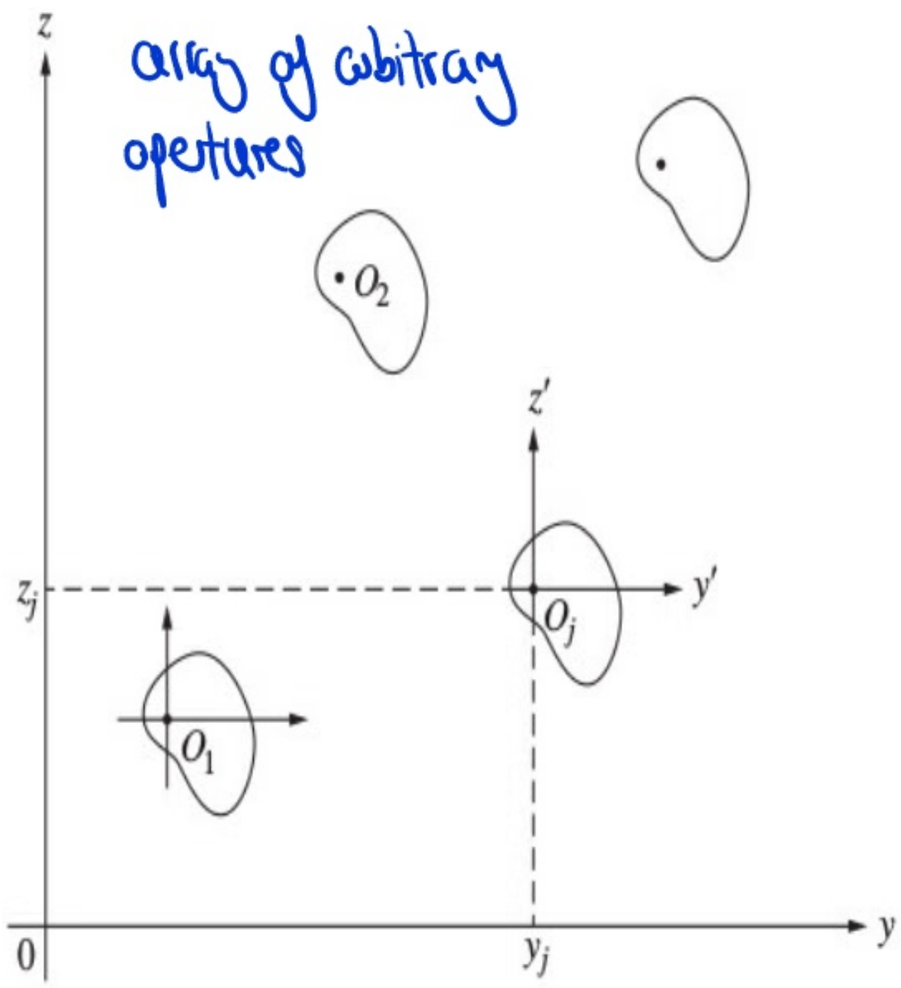
The convolution theorem now dictates

$$\mathcal{F}\{A_2(z)\} = \mathcal{F}\{A_1(z)\} \cdot \mathcal{F}\{\left[\delta(z-a/2) + \delta(z+a/2)\right]\}.$$

$\mathcal{F}\{A_2(z)\}$: two-slit diffraction pattern

$\mathcal{F}\{A_1(z)\}$: single-slit pattern

$\mathcal{F}\{\left[\delta(z-a/2) + \delta(z+a/2)\right]\}$: diffraction pattern for two infinitessimally thin slits



We can generalise this idea to an arbitrary array of apertures in two dimensions: Consider an aperture whose aperture-function can be written in the form $A(\vec{r}) = A_0 \circledast A_d$ where $A_d(\vec{r}) = \sum_{j=1}^{N} \delta(\vec{r} - \vec{r}_j)$, where $\vec{r}_j = (y_j, z_j)$ denotes the centre of the ith (identical) aperture, then the overall diffraction pattern will be

Array theorem

$$\mathcal{F}\{A\} = \mathcal{F}\{A_0\} \cdot \mathcal{F}\{A_d\},$$

$\mathcal{F}\{A_d\} = \sum_{j=1}^{N} e^{-i\vec{K}\vec{r}_j}$

i.e. the diffraction pattern of one aperture acts as the envelope, multiplied by the diffraction pattern of infinitessimally small holes at $\vec{r}_j$.

6.) Image formation

Fourier theory also plays an important role in image formation, provided that the imaging system is linear. Linearity is satisfied if for two input signals (e.g. an aperture function) $f_{1,2}(y,z)$ the corresponding output signals (e.g. diffraction pattern) $g_{1,2}(y,z)$ satisfy

for coherent sources $\vec{E}$ fields are linear, for incoh. sources intensities

$$a f_1(y,z) + b f_2(y,z) \rightarrow a g_1(y,z) + b g_2(y,z).$$

Moreover, we typically deal with stationary systems for which $f(y,z) \rightarrow g(y,z)$ means that $f(y-y_0, z-z_0) \rightarrow g(y-y_0, z-z_0)$, i.e. by displacing the input position, we only change the output position.

In the following, we will denote the effect of the linear system on some input as $g(y,z) = \mathcal{L}\{f(y,z)\}$. Then if we know how a linear system responds to a δ-function input (i.e. a point source), we can reconstruct the response to an arbitrary input as

linear comb. of δ functions weighted by $f(y',z')$

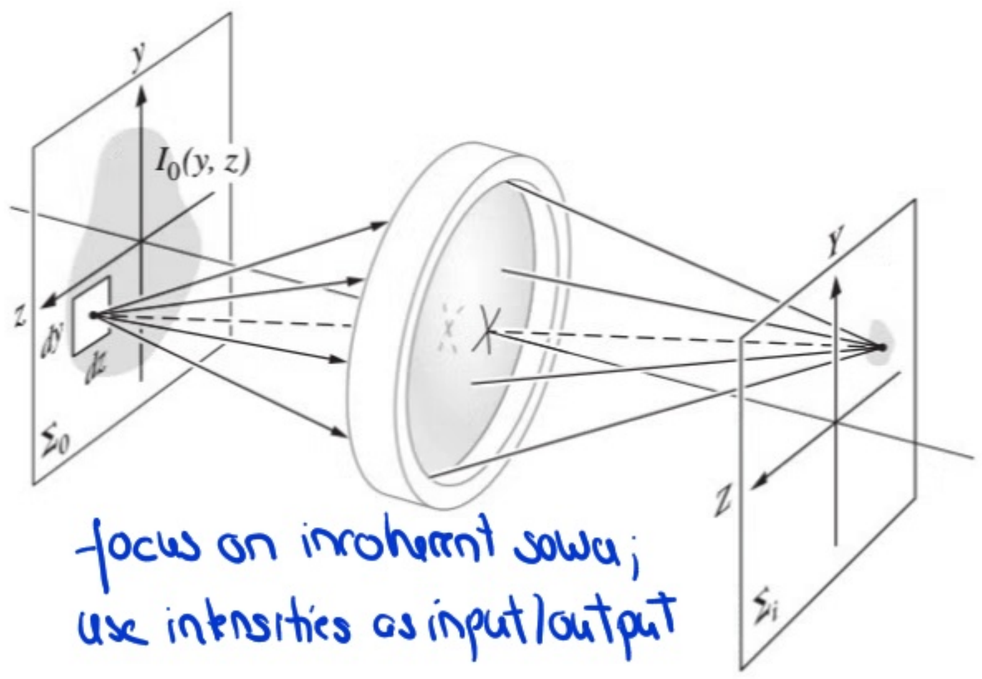
sifting property

$$g(y,z) = \mathcal{L}\left\{\iint f(y',z')\,\delta(y-y')\,\delta(z-z')\,dy'\,dz'\right\}$$

$$= \iint f(y',z')\, \underbrace{\mathcal{L}\{\delta(y-y')\delta(z-z')\}}_{\triangleq \text{ "impulse response"}}\, dy'dz'$$

Now if we consider a lens forming an image, a small element of intensity at position (y,z), namely $I_o(y,z)\,dy\,dz$ will correspond to a blob of intensity on the image plane, contributing the following to the total intensity

$$dI_i(y,z) = S(y,z;Y,Z) \cdot I_o(y,z)\,dy\,dz,$$



focus on incoherent sources; use intensities as input/output

where $S(y,z;Y,Z)$ is the so-called point-spread function (as the name says it describes how an optical system transforms a single object point). As the system is linear, we can integrate over y,z to get the total output intensity

$$I_i(Y,Z) = \int_{-\infty}^{\infty}\int I_o(y,z)\, S(y,z;Y,Z)\, dy\, dz.$$

S describes how a system acts on a δ-function input; shifting δ around changes position of S so it has to depend on $Y-y, Z-z$

$$= \int_{-\infty}^{\infty}\int I_o(y,z)\, S(Y-y;Z-z)\, dy\, dz$$

$$= \underline{\underline{I_o \circledast S}}.$$

This is exactly the convolution of I_0 and the point-spread function, which implies that for each point on the object located at a specific position (y_0, z_0) a 'blob' $S(y-y_0, z-z_0)$ appears on the screen. The actual image is then obtained by summing all the 'blobs' for all points on the object. The point-spread function thus tells us everything we need to know about the imaging system. Note that for a perfect (diffraction-limited) system, the Airy function describes S.

Phys 434 - Lecture 24

Gaussian Beams, Lens Transformations

1.) The paraxial wave equation

In our discussion of optics so far, we have used ray tracing and the Huygens-Fresnel principle to understand the propagation of light through optical systems. However, we have not tried to mathematically describe any propagation modes beyond plane waves. Nevertheless, having used lasers in your demos you will have observed that the light should be represented as a wave (ie. the solution to the wave equation) with a well-defined direction and finite transverse dimension.

To determine the modes of lasers, we will focus on the scalar wave equation $\nabla^2 E = 1/c^2 \partial_t^2 E$ (satisfied by each vector component of E). For a (i) monochromatic wave, we have $E(\vec{r}, t) = E(\vec{r})\, e^{-i\omega t}$ and thus

$$\nabla^2 E - 1/c^2 \partial_t^2 \vec{E} = \nabla^2 E + \omega^2/c^2 \vec{E},$$

$$\Rightarrow \underline{\underline{\nabla^2 E + k^2 E = 0}}. \quad \text{Helmholtz equation}$$

Assuming also that (ii) the wave has a well-defined direction (referred to as

the paraxial approximation), we consider beams of the form $E(\vec{r}) = A(x,y,z)\, e^{ikz}$ (in analogy with plane waves where $e^{i\vec{k}\vec{r}}$ captures propagation with a well-defined direction $\vec{k}$). For this ansatz, the wave will mostly propagate along the z-direction with an 'envelope' $A(x,y,z)$ that should vary slowly on the order of a wavelength λ, so that $|\partial A| / |A| \ll 1$. Substituting into the Helmholtz equation gives

$$\nabla^2 E = \underbrace{\left(\partial_x^2 + \partial_y^2\right)}_{=\partial_\perp^2} + \partial_z^2 \Big) A(x,y,z)\, e^{ikz}$$

$$= \partial_\perp^2 A\, e^{ikz} + \partial_z \left(\partial_z A\, e^{ikz} + A\, ik\, e^{ikz}\right)$$

$$= \partial_\perp^2 A\, e^{ikz} + \Big(\underbrace{\partial_z^2 A}_{\text{small compared to}} + \underbrace{2ik\,\partial_z A}_{\uparrow} - k^2 A\Big) e^{ikz},$$

$$\Rightarrow \quad \nabla^2 E + k^2 E \approx \left(\partial_\perp^2 A + 2ik\,\partial_z A\right) e^{ikz} \overset{!}{=} 0 .$$

In the paraxial (or 'slowly-varying' envelope) approximation, the Helmholtz equation reads

$$\boxed{\nabla_\perp^2 A + 2ik\,\partial_z A \approx 0 .}$$

looks like the free particle Schrödinger equation in 2D where z is playing the role of time

Solutions to this equation provide an excellent approximation to the modes of lasers. To find an ansatz for A, we look at spherical

waves, which can be approximated as

$$E(r) = \frac{A_0}{r} e^{ikr} = \frac{A_0}{r} e^{ik(x^2+y^2+z^2)^{1/2}}$$

$$\approx \frac{A_0}{r} e^{ikz} e^{ik(x^2+y^2)/2z} .$$

$(x^2+y^2+z^2)^{1/2}$
$= z(1+\frac{x^2+y^2}{z^2})^{1/2}$
$\approx z(1+\frac{x^2+y^2}{2z^2})$

From this we can essentially read off $A(\vec{r}')$ for a Gaussian beam

$$A(\vec{r}') = \frac{A_0}{z - iz_R} e^{ik \frac{x^2+y^2}{2(z-iz_R)}} .$$

you may recognise this solution from QM as the free particle solution that starts off with a narrow Gaussian and then broadens

It is straightforward to show that this is a solution to the paraxial Helmholtz equation for any value z_R. Note that it is not the only possible solution as we will discuss below, but it is the only solution with no nodes in the transverse direction.

2.) Gaussian beams

Gaussian beams are crucial for any optics experiment using lasers, so we will discuss their properties in some more detail. As a first step, we calculate the intensity

$$I(\vec{r}') \propto |A(\vec{r}')|^2 = \frac{|A_0|^2}{z^2+z_R^2} e^{ik \frac{x^2+y^2}{2(z-iz_R)}} e^{-ik \frac{x^2+y^2}{2(z+iz_R)}}$$

these won't cancel because of the complex denominator

$$= \frac{|A_0|^2}{z^2+z_R^2}\, e^{ikr^2/2\left[(z-iz_R)^{-1}-(z+iz_R)^{-1}\right]}$$

$$= \frac{|A_0|^2}{z^2+z_R^2}\, e^{ikr^2/2\left[\frac{z+iz_R-(z-iz_R)}{z^2+z_R^2}\right]}$$

$$= \frac{|A_0|^2}{z^2+z_R^2}\, e^{-\frac{kr^2 z_R}{z^2+z_R^2}}$$

This is a Gaussian profile and the name 'Gaussian beam'.

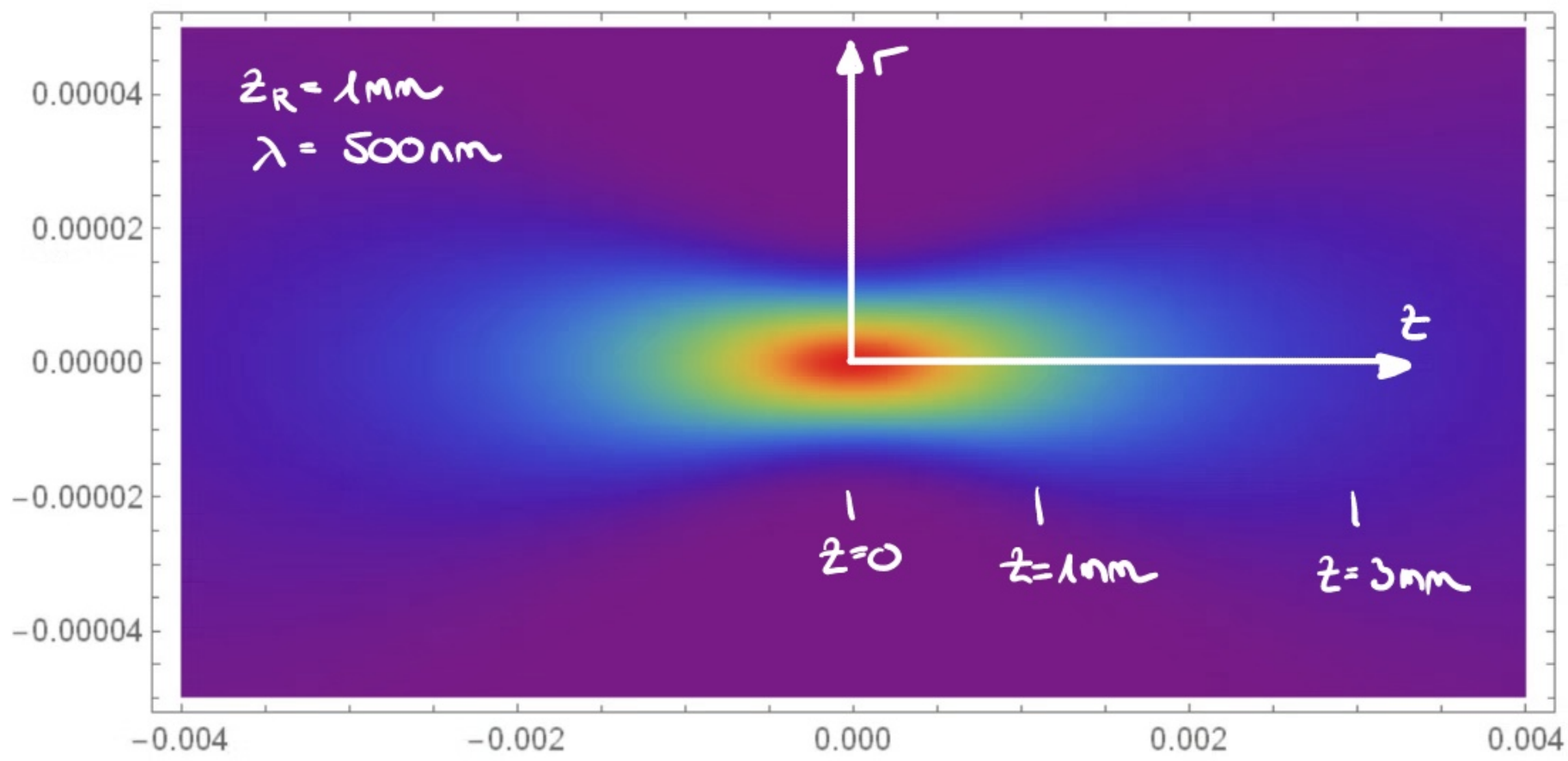
The irradiance profile is typically written in the form

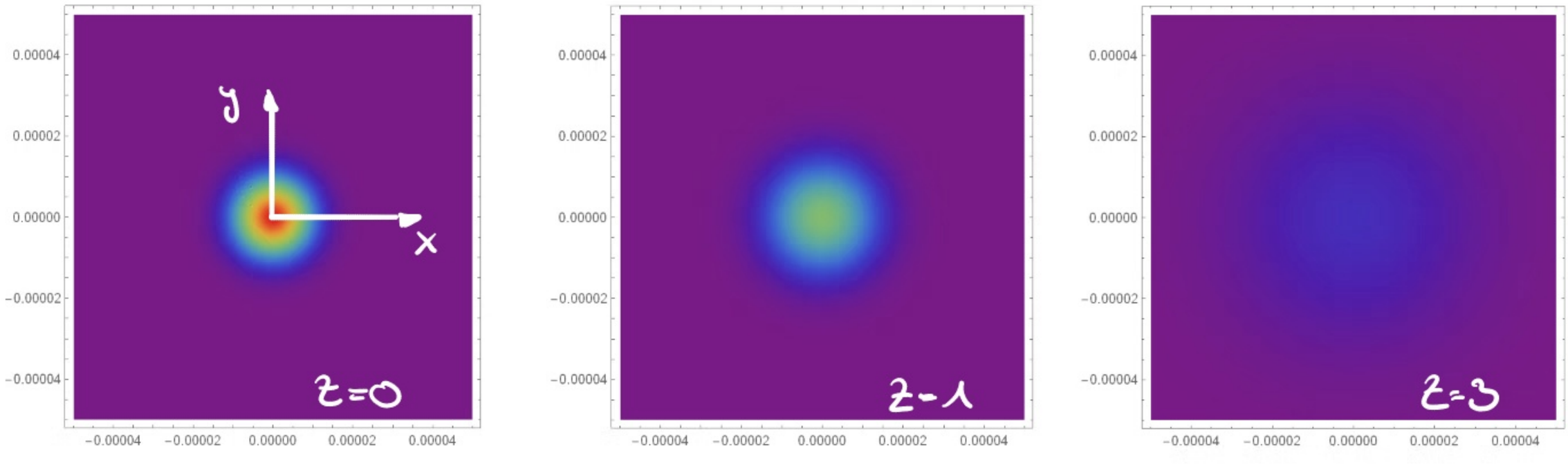
$$I(r,z) = \frac{I_0}{1+z^2/z_R^2}\, e^{-2r^2/w^2},$$

where

$$w(z) = \left[\frac{2(z^2+z_R^2)}{kz_R}\right]^{1/2} = \left[\frac{\lambda z_R}{\pi}\left(1+\frac{z^2}{z_R^2}\right)\right]^{1/2}.$$

We can plot the longitudinal and transverse profile as





We observe that the beam looks Gaussian in the transverse direction but its peak intensity and width vary with distance. We further note

i) $w(z)$ is the '$1/e^2$-radius', ie. the radius at which the intensity drops to $1/e^2$ of its maximum in the transverse direction; $w_0 \equiv w(z=0)$ is typically referred to as the 'waist' of the beam

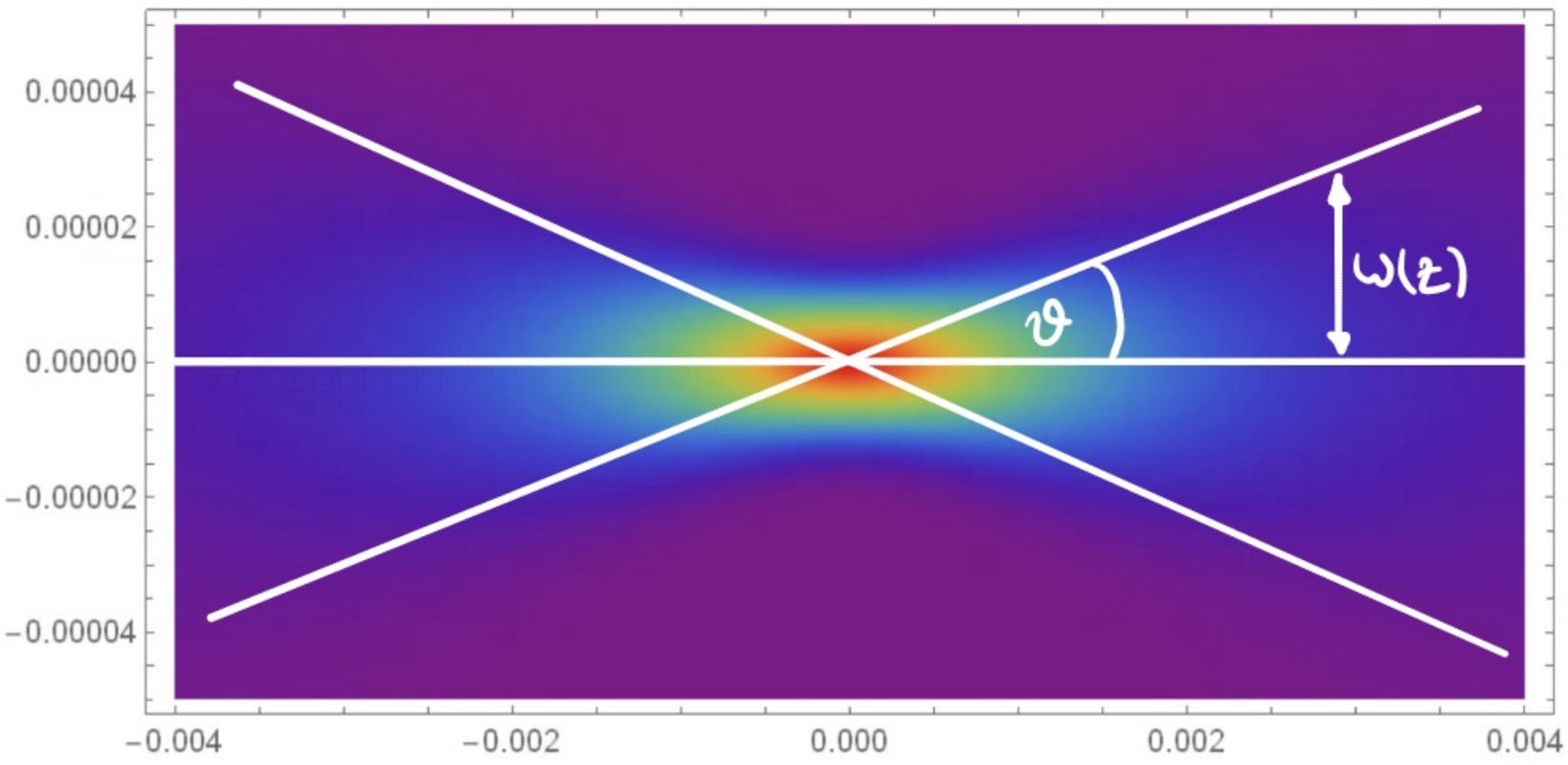
ii) z_R is called the Rayleigh range; ie. when $z = z_R$, the intensity at the centre of the beam drops to half its value at the focus at $z=0$; the parameter characterises how tight the focus is along the axial direction

iii) w_0 and z_R are related as $\underline{z_R = \pi w_0^2/\lambda}$, so that

$$w(z) = w_0 \left[1 + \left(\frac{\lambda z}{\pi w_0^2}\right)^2\right]^{1/2} .$$

The tighter the focus (smaller w_0), the shorter the distance over which the beam is nearly focused (shorter z_R).

(iv) We determine the divergence angle ϑ:



For away from the focus, we have

$$\tan\vartheta \overset{\text{paraxial}}{\approx} \vartheta = \frac{\omega(z)}{z} \overset{z\gg z_R}{\approx} \frac{1}{z}\left(\frac{\lambda}{\pi z_R}\right)^{1/2} z = \left(\frac{\lambda}{\pi z_R}\right)^{1/2} = \frac{1}{\omega_0}.$$

The tighter the focus, the larger the divergence.

So far, we have only focused on the intensity distribution of the Gaussian beam, but what do the wavefronts actually look like?
To answers that, we return to the electric field amplitude and separate out the amplitude and phase of the complex field:

$$E(\vec{r}) = \frac{A_0}{z - iz_R} e^{ikz} e^{ik\frac{x^2+y^2}{2(z-iz_R)}}$$

$$= \frac{A_0}{z - iz_R} e^{ikz} \underbrace{e^{-\frac{kz_R(x^2+y^2)}{2(z_R^2+z^2)}}}_{\text{amplitude}} \underbrace{e^{i\frac{kz(x^2+y^2)}{2(z_R^2+z^2)}}}_{\text{phase}}.$$

Recall that in the paraxial approximation, a spherical wave at radius $R = (x^2+y^2+z^2)^{1/2}$ with $R \approx z$ looks like $e^{ikz}\, e^{ik(x^2+y^2)/2R}$.

Using this analogy, we can write the second (phase) term of $E(\vec{r})$ in terms of the curvature of the wave fronts, ie.

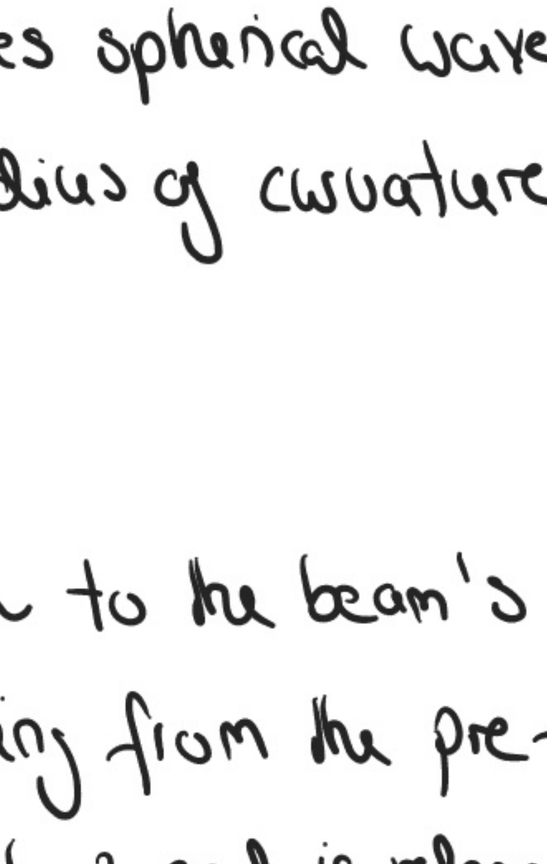
$$\frac{ik(x^2+y^2)\,z}{2(z^2+z_R^2)} = \frac{ik(x^2+y^2)}{2R(z)} \quad \text{where} \quad \underline{\underline{R(z) \equiv z\left(1+\frac{z_R^2}{z^2}\right)}}.$$

This suggests that the wavefronts of a Gaussian beam are curved, with a radius of curvature that changes along the optical axis. We recover flat wave fronts (corresponding to $R \to \infty$) in the limits $z \to 0$ and $z \to \infty$, where the latter closely resembles spherical wave behaviour. Moreover, the smallest (or tightest) radius of curvature corresponds to $R(z = z_R) = 2z_R$.

Note however that this is not the only contribution to the beam's phase. There is an additional contribution coming from the pre-factor $1/z - iz_R$. This phase will also change with z and is referred to as the Gouy phase. It can be calculated as

$$(z - iz_R)^{-1} = \frac{z + iz_R}{z^2 + z_R^2} = \frac{i(z_R - iz)}{z^2 + z_R^2}$$

$$= \frac{i\, e^{-i\zeta(z)}}{(z^2+z_R)^{1/2}} = \underbrace{\frac{i}{z_R}\frac{\omega_0}{\omega(z)}}_{\text{add to amplitude}} e^{-i\zeta(z)},$$



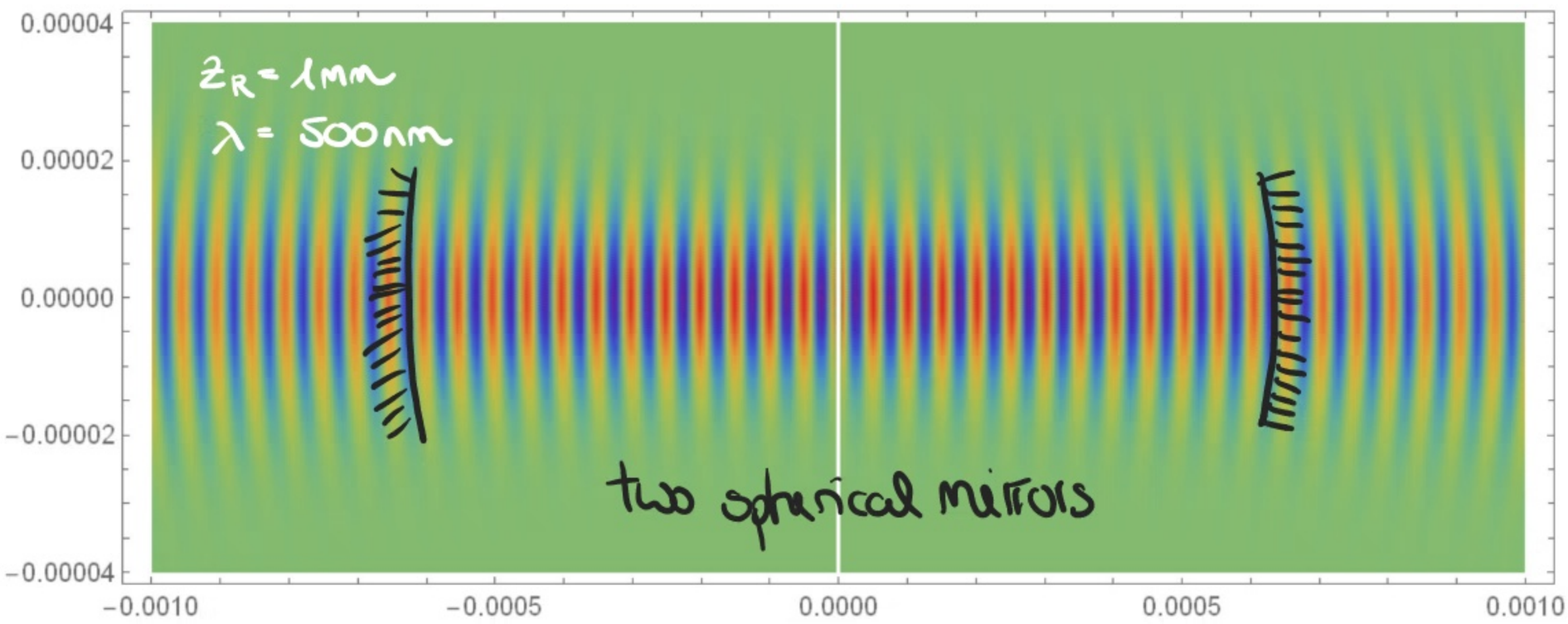
where $$\zeta(z) = \arctan\left(\frac{z}{z_R}\right) .$$

We can combine these results to write a complete expression for the electric field of a Gaussian beam as

$$E(\vec{r}) = A_0^{raw} \frac{\omega_0}{\omega(z)} e^{-r^2/\omega^2(z)} e^{i\overbrace{[kz - \zeta(z) + k\, r^2/2R(z)]}^{\text{total phase } \phi(r,z)}} .$$

- $\frac{\omega_0}{\omega(z)}$: intensity falls off away from the focus
- $e^{-r^2/\omega^2(z)}$: Gaussian transverse profile with variable width
- $\zeta(z)$: Gouy phase modifies the wavefront spacing
- $k\, r^2/2R(z)$: wavefront curvature

We can plot this as follows

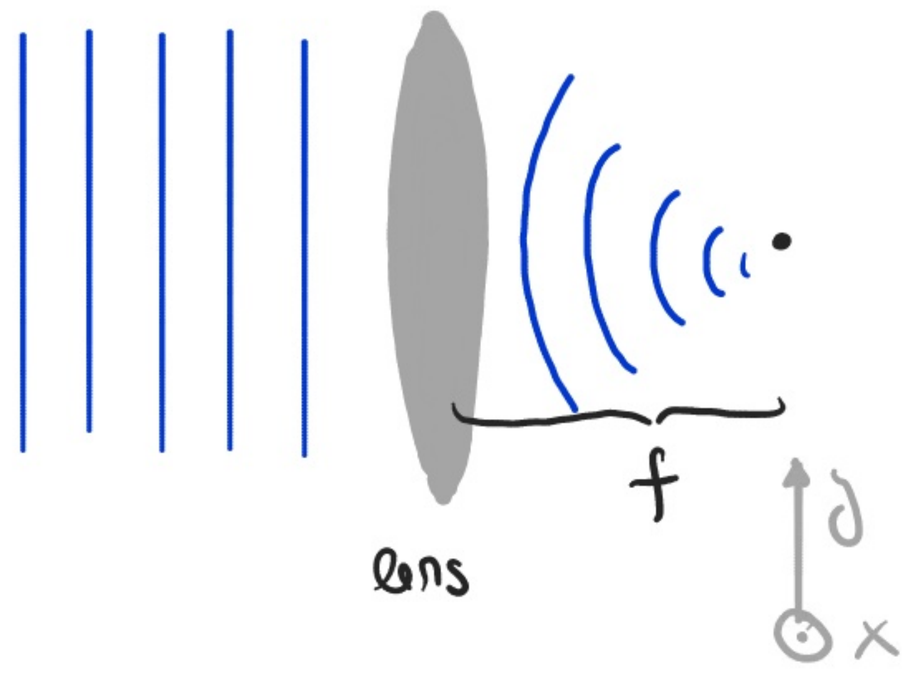


This picture is closely connected to the concept of a laser cavity as adding two spherical mirrors with $R(z)$ (ie. they exactly follow the wave fronts) would not affect the field $E(\vec{r})$, ie. the resulting field

would still be a solution to the Helmholtz equation. For the cavity the overall phase of the Gaussian beam would also have to satisfy $\phi = n \cdot \pi$ so that a multiple of $\lambda/2$ fits in between the mirrors.

3.) Propagation of a Gaussian beam

To understand how a Gaussian beam is affected by a lens, we need to consider how the lens affects wavefronts (not rays). Generally speaking, a lens transforms a plane wave into a focusing (or defocusing) wave, which is essentially a section of a spherical wave. We thus expect the lens to imprint a curvature onto the wavefront. In the paraxial approximation, the effect of a thin lens is to modify the wave amplitude by

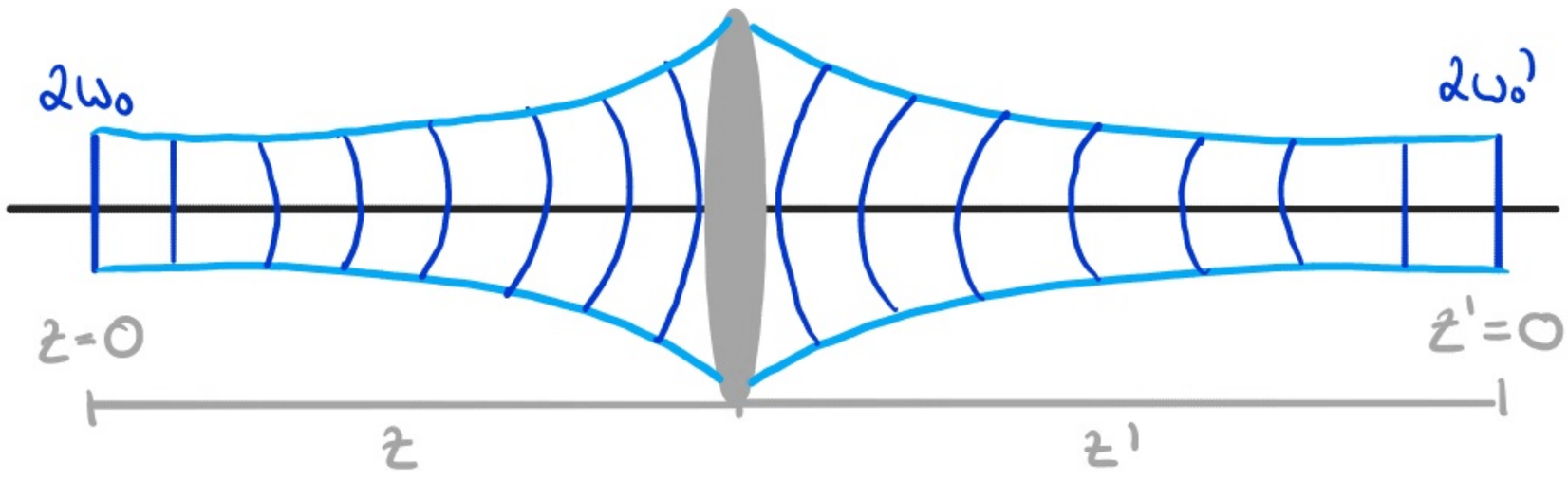


$$t(x,y) = e^{-ik(x^2+y^2)/2f} \quad .$$

lens transmission amplitude

While this can be quantitatively derived by calculating the optical path length difference for different wavefronts (and thus the resulting phase), the result can be understood qualitatively because the lens is thickest in the middle. As x, y increase the optical path length (and hence the

phase) decrease. So the effect of a lens on a Gaussian beam is to alter its curvature without changing its intensity:



$$\text{phase } \phi = kz - \zeta + \frac{kr^2}{2R} \;;\; \phi' = \phi - k\frac{r^2}{2f} = kz - \zeta + \frac{kr^2}{2R'} .$$

The transmitted wave is also a Gaussian beam but has a new radius of curvature R' determined by $1/R' = 1/R - 1/f$. The width of both beams at the lens has to be the same, so $w'(z') = w(z)$. Using these two relations, we can solve for w_0' and z' to find

$$\Rightarrow \quad w_0' = M w_0 = \frac{|f|}{[(z-f)^2 + z_R^2]^{1/2}} w_0 ,$$

$$\Rightarrow \quad z' = M^2 z + (1-M^2) f ,$$

convention: $z' > 0$ if second beam waist after the lens

where M is the magnification. From the above sketch, we would want to identify the first waist w_0 with the object plane and the second waist with the image plane. This is indeed possible as we can rewrite the equation for z' in the following form

$$\frac{1}{z+\frac{z_R^2}{z-f}} + \frac{1}{z'} = \frac{1}{f} \quad .$$

This matches the lens makes equation in the limit $|z_R^2/z-f| \ll z$. This is satisfied if the beam is tightly focused (z_R is small) and the waist is not positioned close to f. In the same limit, we have for R

$$w_0' \approx \frac{f}{f-z} w_0 = \frac{z'}{z} w_0 ,$$

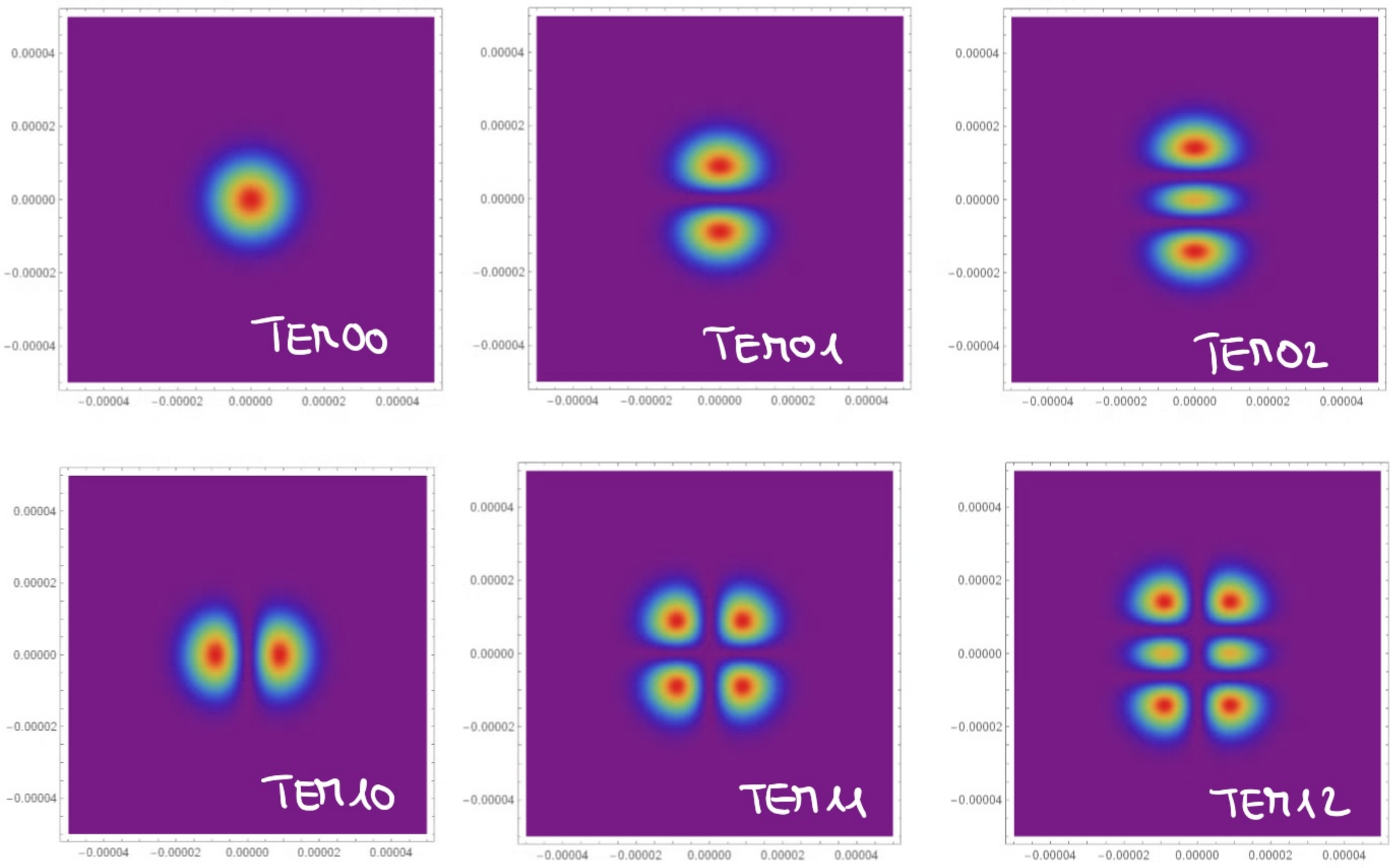
which is the usual transverse magnification in Geometric Optics. Note that analogously, the Gaussian beam can be considered within diffraction and Fourier theory.

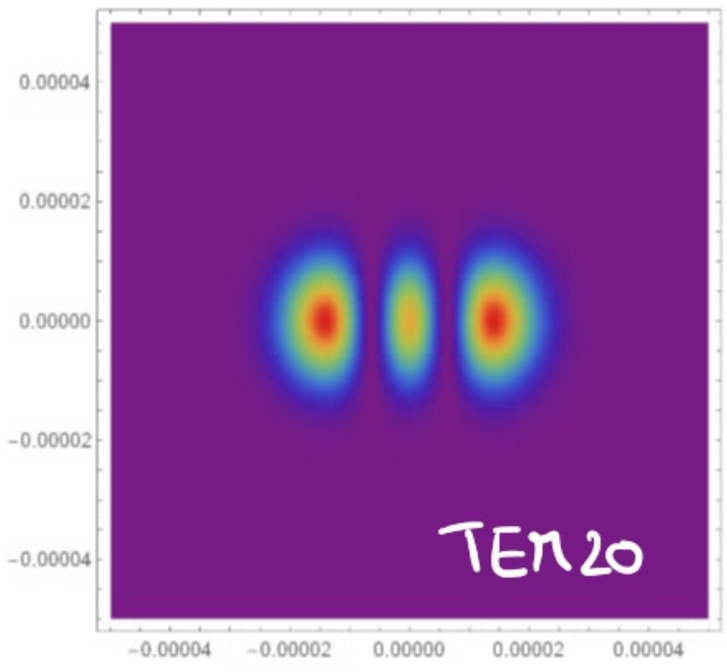
4.) Higher-order modes

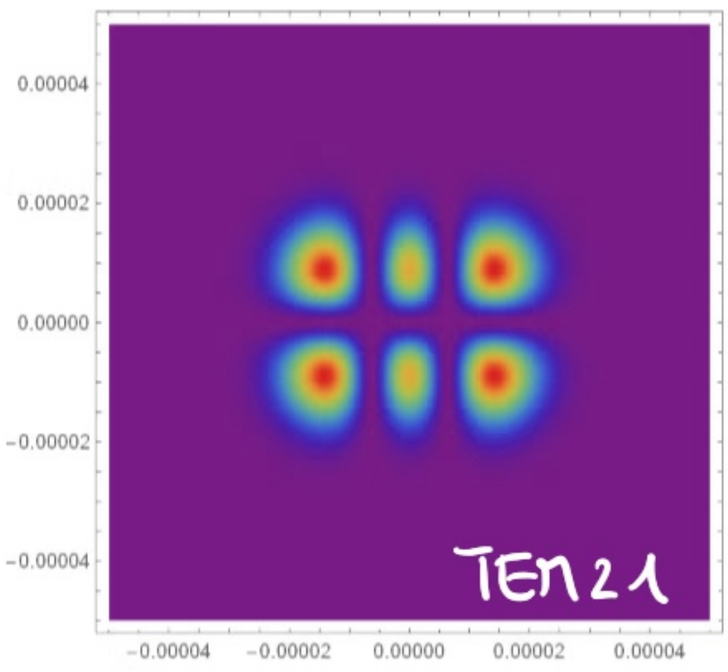
Gaussian beams are not the only solution to the paraxial Helmholtz equation, but they share an underlying structure with a family of solutions known as Hermite Gaussian modes:

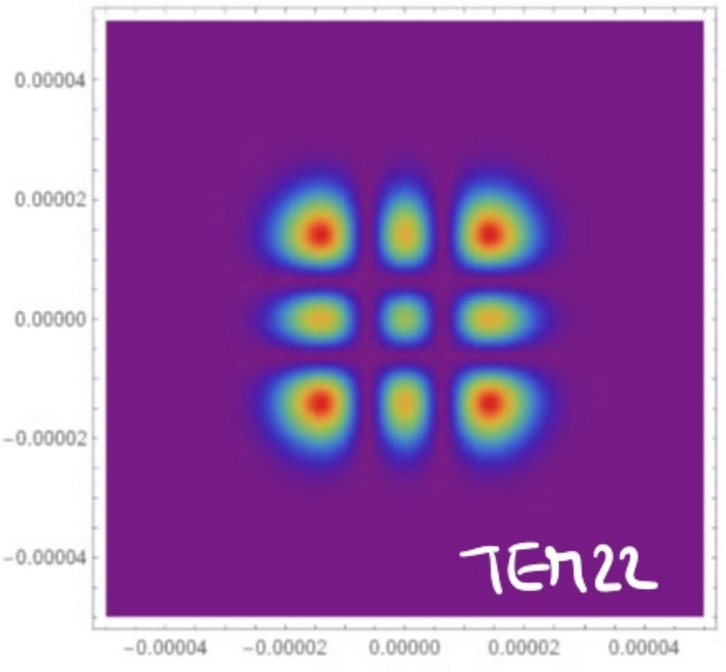
$$E_{lm}(\vec{r}) = A_{l,m} \frac{w_0}{w(z)} G_l\left[\frac{\Omega x}{w(z)}\right] G_m\left[\frac{\Omega y}{w(z)}\right] \cdot e^{ikz + ik(x^2+y^2)/2R(z) - i(l+m+1)\zeta(z)} ,$$

where $G_\ell(s) = H_\ell(s) e^{-s^2/2}$ with the Hermite polynomials H_ℓ. Note that $w(z)$, $R(z)$ and $\zeta(z)$ are the same parameters as for the Gaussian beams (some z-dependence). Moreover, the wave fronts of the higher-order modes are parallel to the wave-fronts of the Gaussian beams (same curvature and thus same response to lens transformations). Differences however become obvious, when we plot the intensity profile of the modes typically labelled $TEM_{\ell m}$. As ℓ, m increase the number of nodes in the intensity profile as well as the radial size increase.

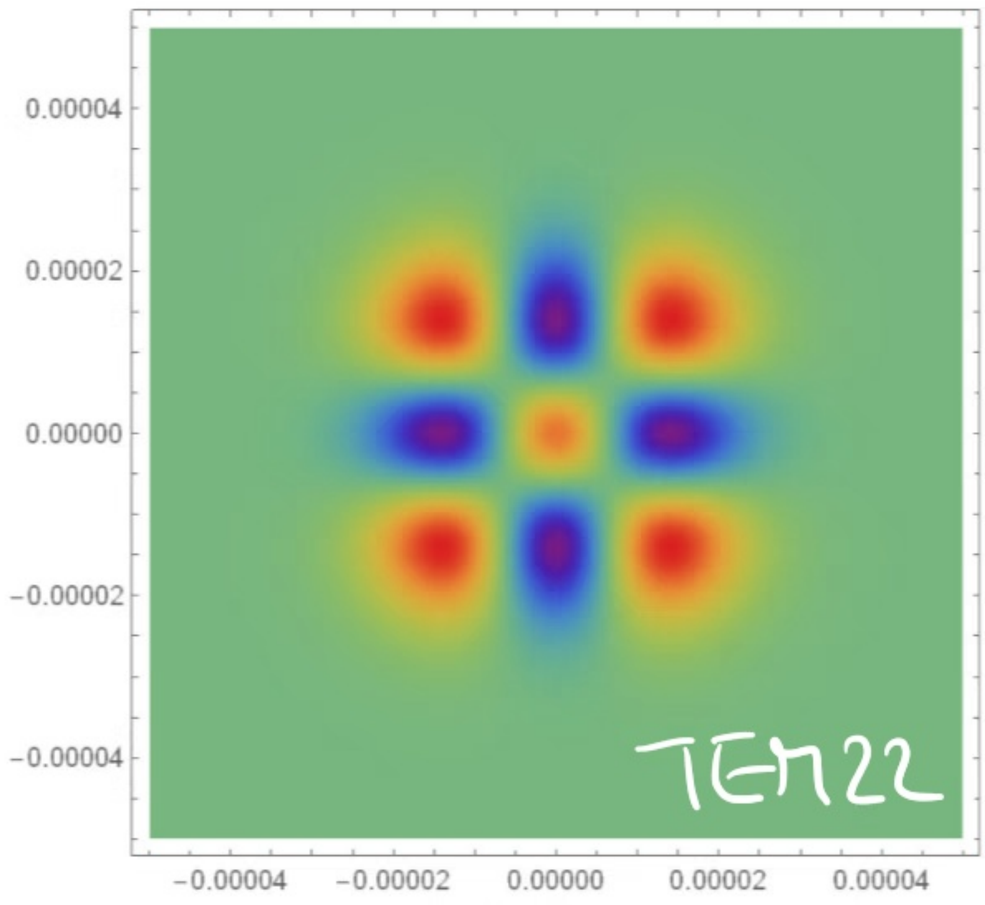








Note that the electric field takes on both positive and negative values as illustrated on the right for the TEM_{22} mode (blue $\hat{=}$ negative and red $\hat{=}$ positive field).



One final note: since all these modes share the same values for $R(z)$, $w(z)$ and $\zeta(z)$ they can all coexist in a laser cavity (if two spherical mirrors support a Gaussian beam, they will also support the TEM_{lm} modes). However, the transverse modes will not have the same resonance frequency as the fundamental TEM_{00} mode due to the $(l+m+1)$ prefactor of the Gouy phase. A resonant cavity will thus emit in the fundamental mode.

Phys 434 - Lecture 25

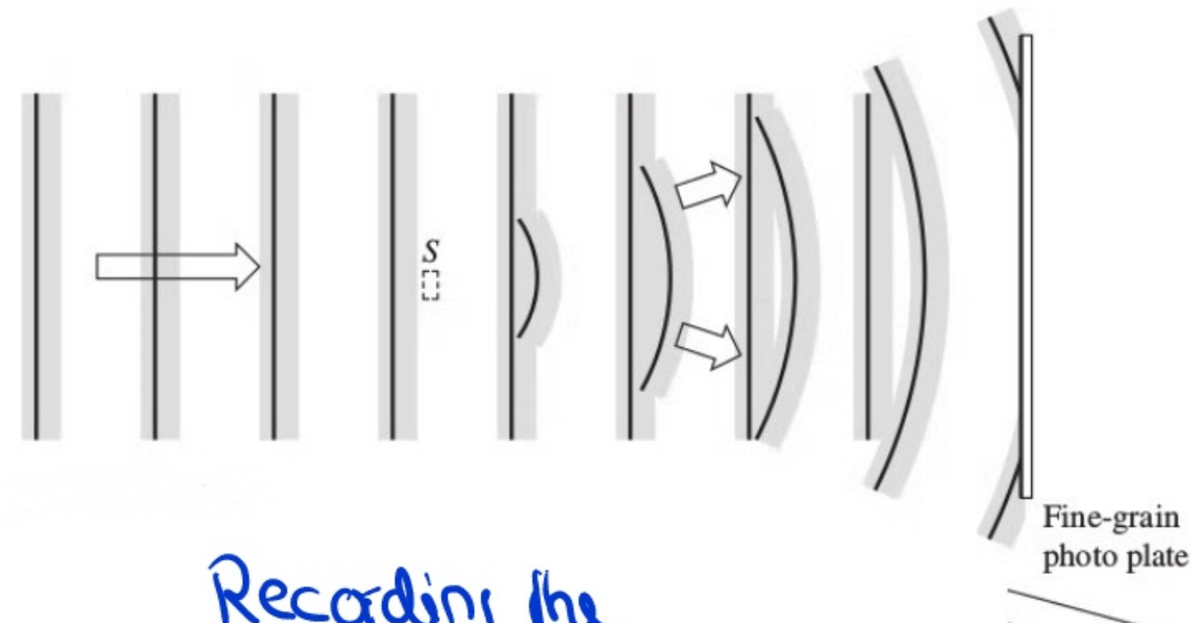
Holography

1.) General concept

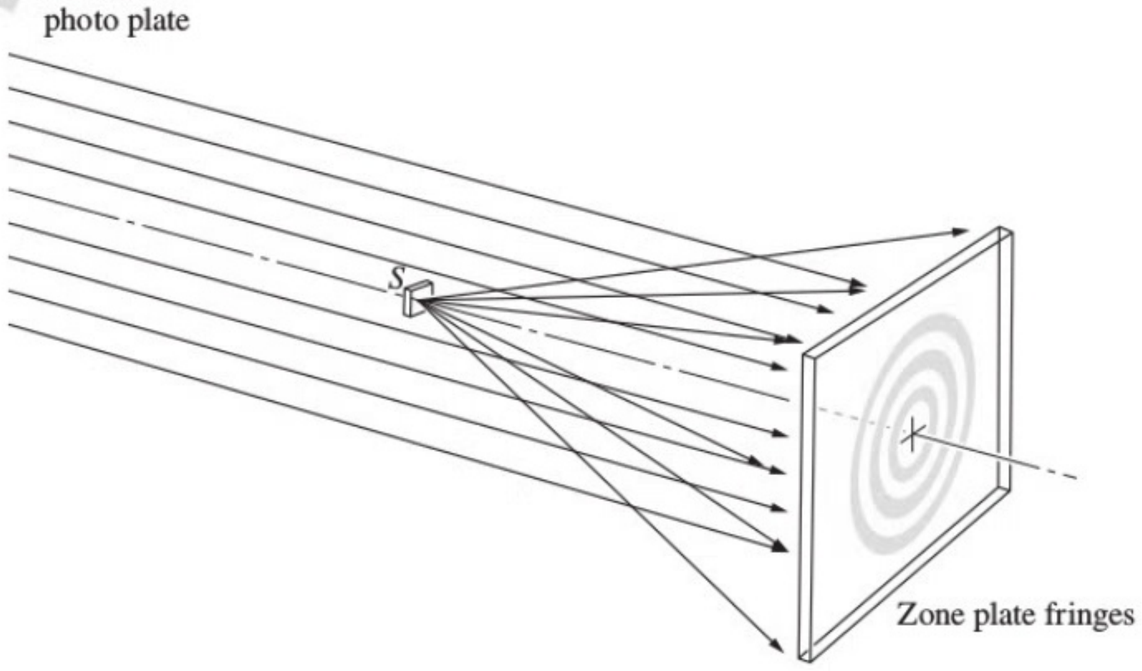
The standard way to record images is via photography, where we compress the three-dimensional world into a two-dimensional mapping of the irradiance distribution. This implies that we do not see an accurate reproduction of the light field emanating from the object but instead a point-by-point recording of the square of the electric field amplitude. The light reflected off a film does thus not carry any information about the phase of the wave emerging from the object. If we were able to record the phase as well as the amplitude and somehow reconstruct both, the resulting light field would be indistinguishable from the original and you would be able to see the reformed image in all three dimensions. This is the central idea of holography, which we will address in this lecture.

The first person to conduct experiments on holography was Dennis Gabor in the 1940s. His original set-up (patented in 1947) is illustrated below and was based on a two-step lensless imaging process. In the first step, the recording, he photographically recorded an interference pattern

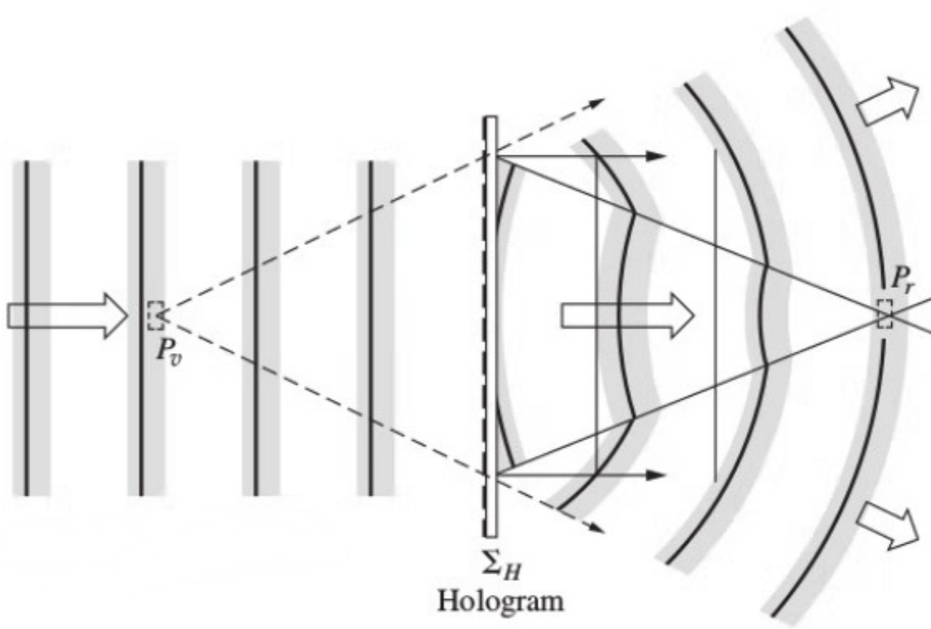
generated by the interaction of a scattered quasimonochromatic light wave (also referred to as the object beam) and a coherent reference beam. He called the resulting pattern a hologram (from Greek 'holos' ≙ whole & 'grapho' ≙ drawing). This hologram essentially contains the fringe pattern of the interference process and thus information about the phase and amplitude of the wave. If the object were very small, the scattered waves would be almost spherical and the interference pattern a series of concentric rings, very much like a Fresnel zone plate but with gradually varying intensity. Remember that this plate somewhat acts like a lens in that it diffracts collimated light into a beam converging to a focal point Pr as well as producing a diverging wave, which appears to come from Pv and constitutes a virtual image.



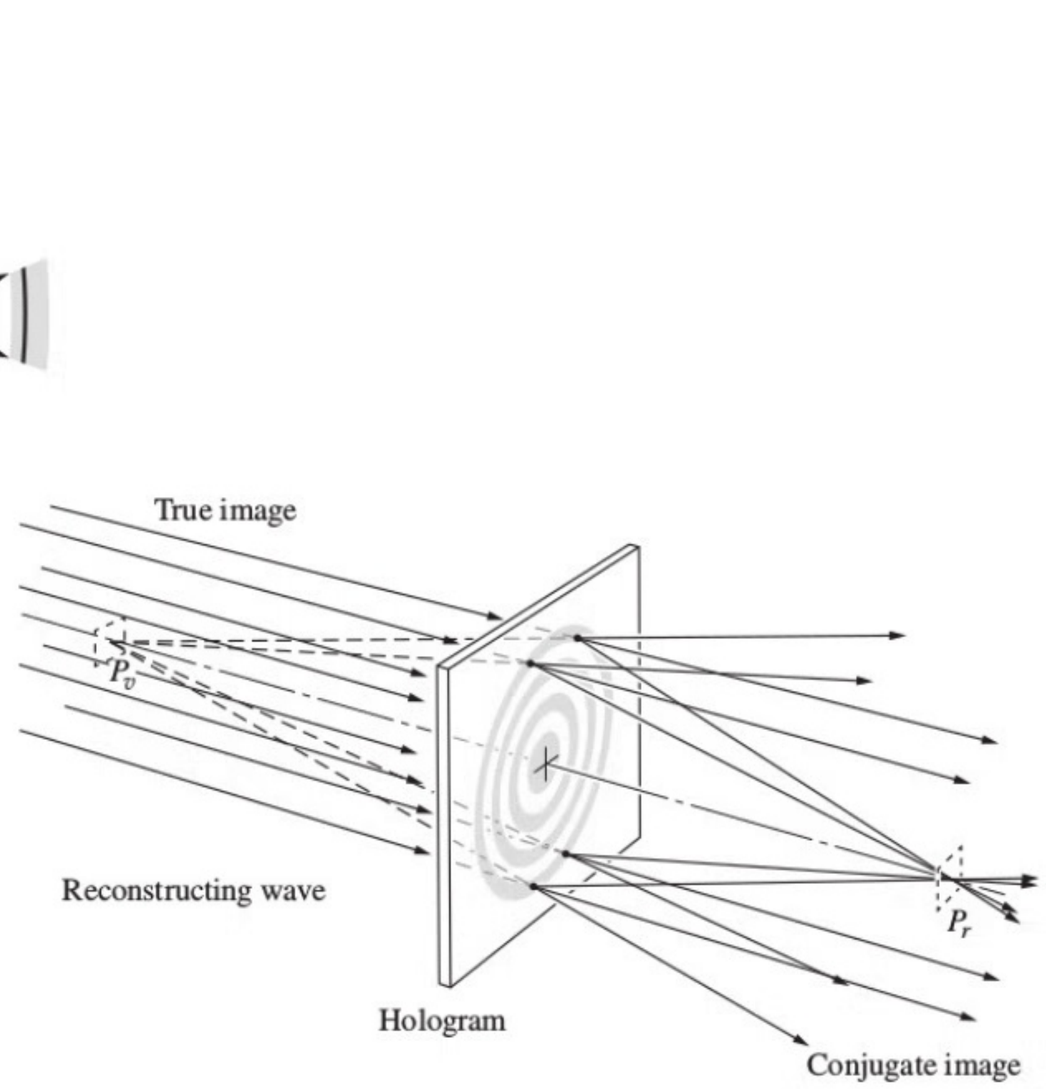
Recording the hologram (interference fringes of coherent background and scattered object wave)



With this in mind, the second step of the procedure is the reconstruction of the light field or image by diffracting a coherent beam by a transparency, which is the developed hologram. If the original object is not small but extended, we can imagine that each point on the object generates its own zone plate displaced relative to the other plates. The entire ensemble thus forms the hologram. During the reconstruction process each zone plate then forms both a real and virtual image of a single object point and we can recover the full object.

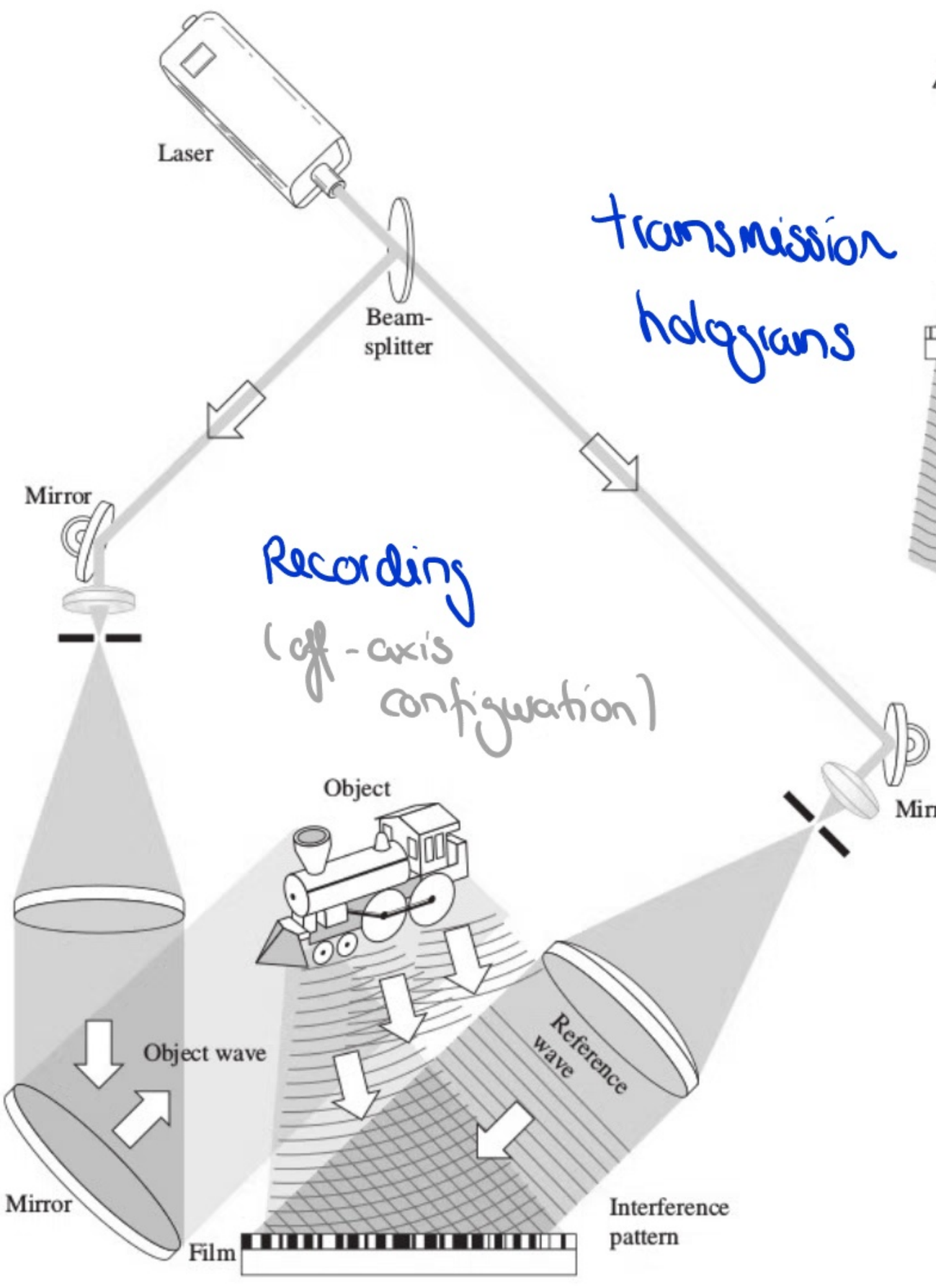


Hologram reconstruction

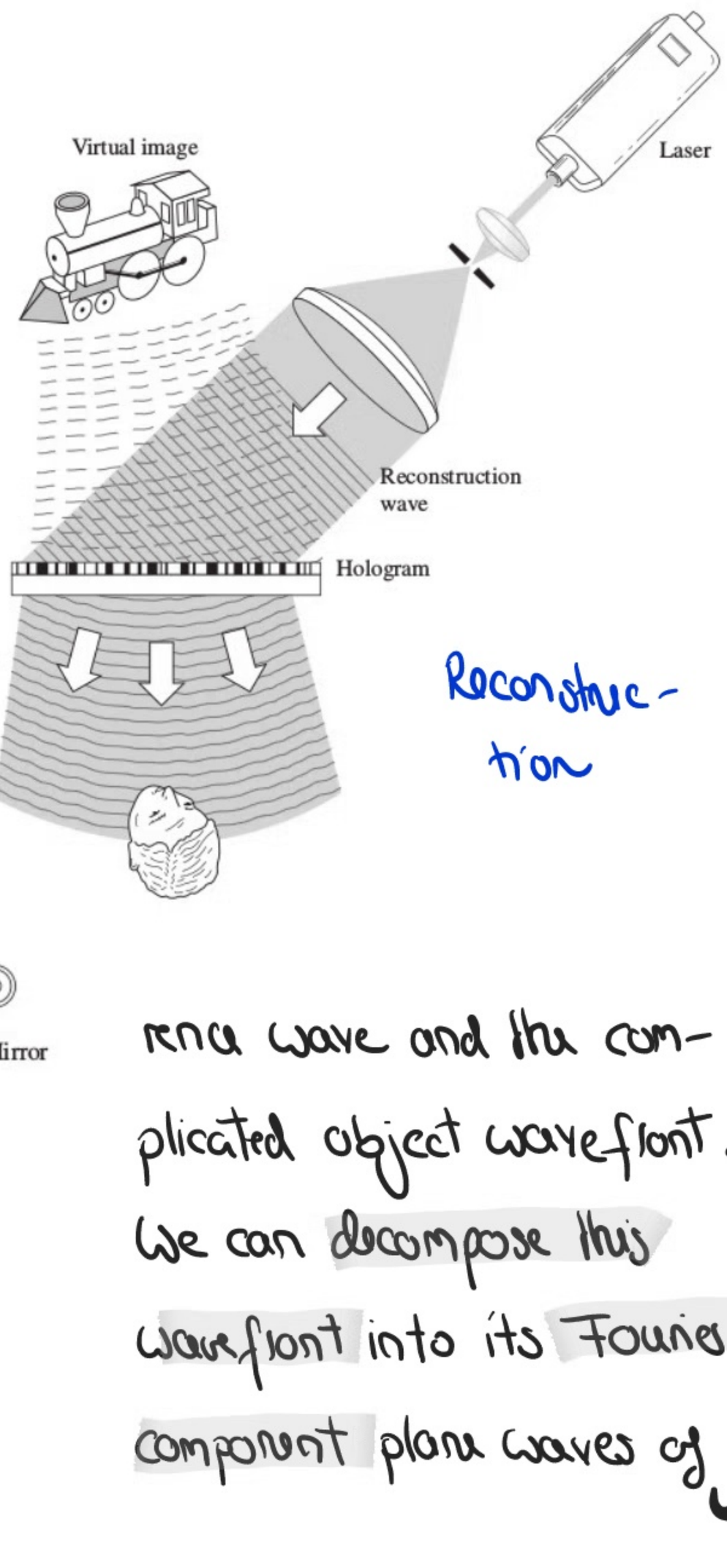


If the reconstruction beam has the same wavelength as the recording beam the virtual image is undistorted and appears at the location originally occupied by the object; so the virtual image is the 'true' image of the object (has the same light field), while the real image is typically referred to as the conjugate image.

Gabor's work which earned him the 1971 Nobel Prize in Physics, remained essentially unnoticed for about 15 years until in the 1960s the first lasers were developed. This enabled the first practical realisation of holographic recordings of 3D objects, which no longer used Gabor's in-line configuration shown above (the true image is inconveniently located behind the conjugate image) but the set-up below (which is characteristic for so-called transmission holograms). As for the in-line set-up the hologram is the interference pattern generated by the refe-



Recording (off-axis configuration)



transmission holograms

Reconstruction

rence wave and the complicated object wavefront. We can decompose this wavefront into its Fourier-component plane waves of

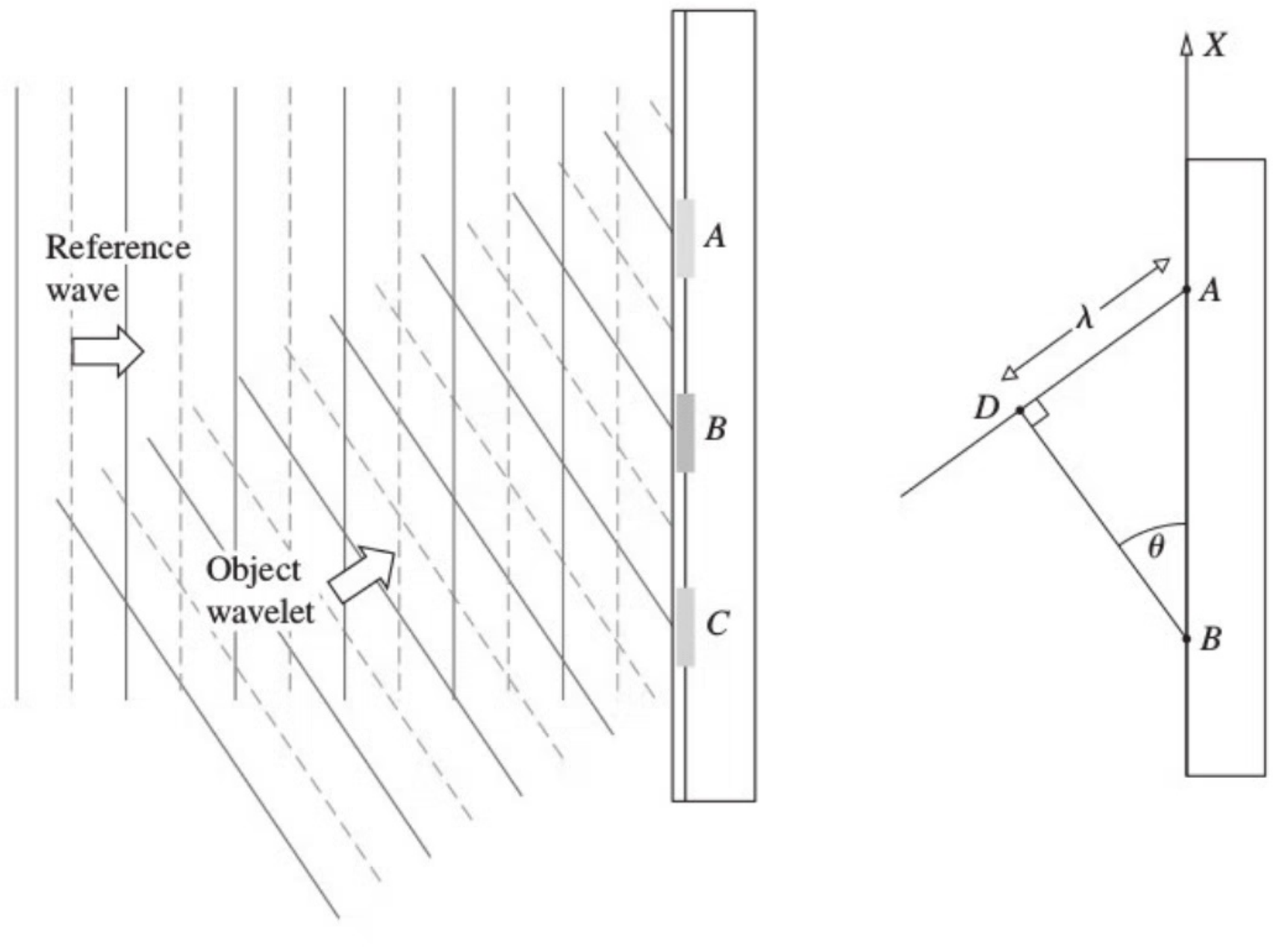
distinct spatial frequencies. Each of these plane waves interferes with the reference wave on the photographic plate thus preserving information about its specific frequency in form of a characteristic finge pattern.

Imagine for example the simplified scenario of two waves illustrated below: for this specific instance the reference wave and object wave coming in at an angle ϑ, have overlapping wave crests at the points A, D and C corresponding to interference maxima. As the waves continue to propagate to the right they will remain in phase so that the maxima remain fixed. The relative phase between the two waves is thus

for $x = \overline{AB}$
$\phi = 2\pi$

$$\phi(x) = 2\pi \frac{x}{\overline{AB}} = \frac{2\pi}{\lambda} x \cdot \sin\vartheta .$$

$\sin\vartheta = \frac{x}{\overline{AB}}$



If the two waves have the same amplitude E_0 the resultant field is

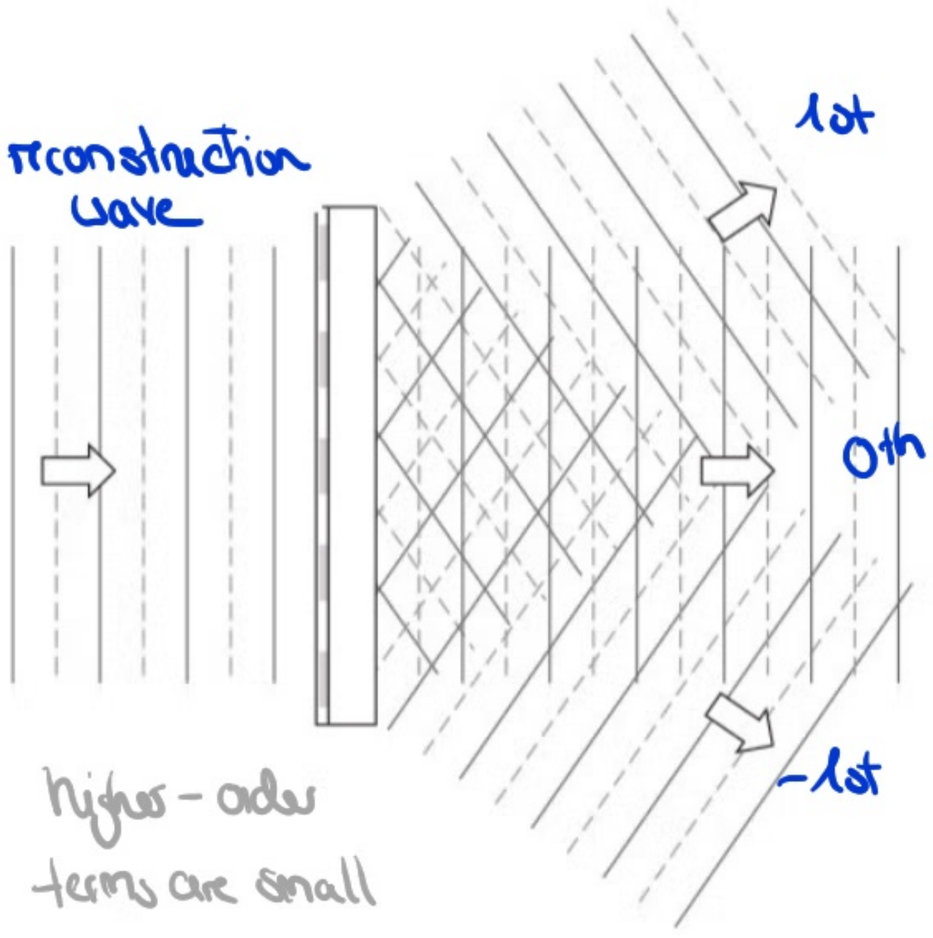
$$E = 2E_0 \cos \phi/2 \sin(\omega t - kx - \phi/2),$$

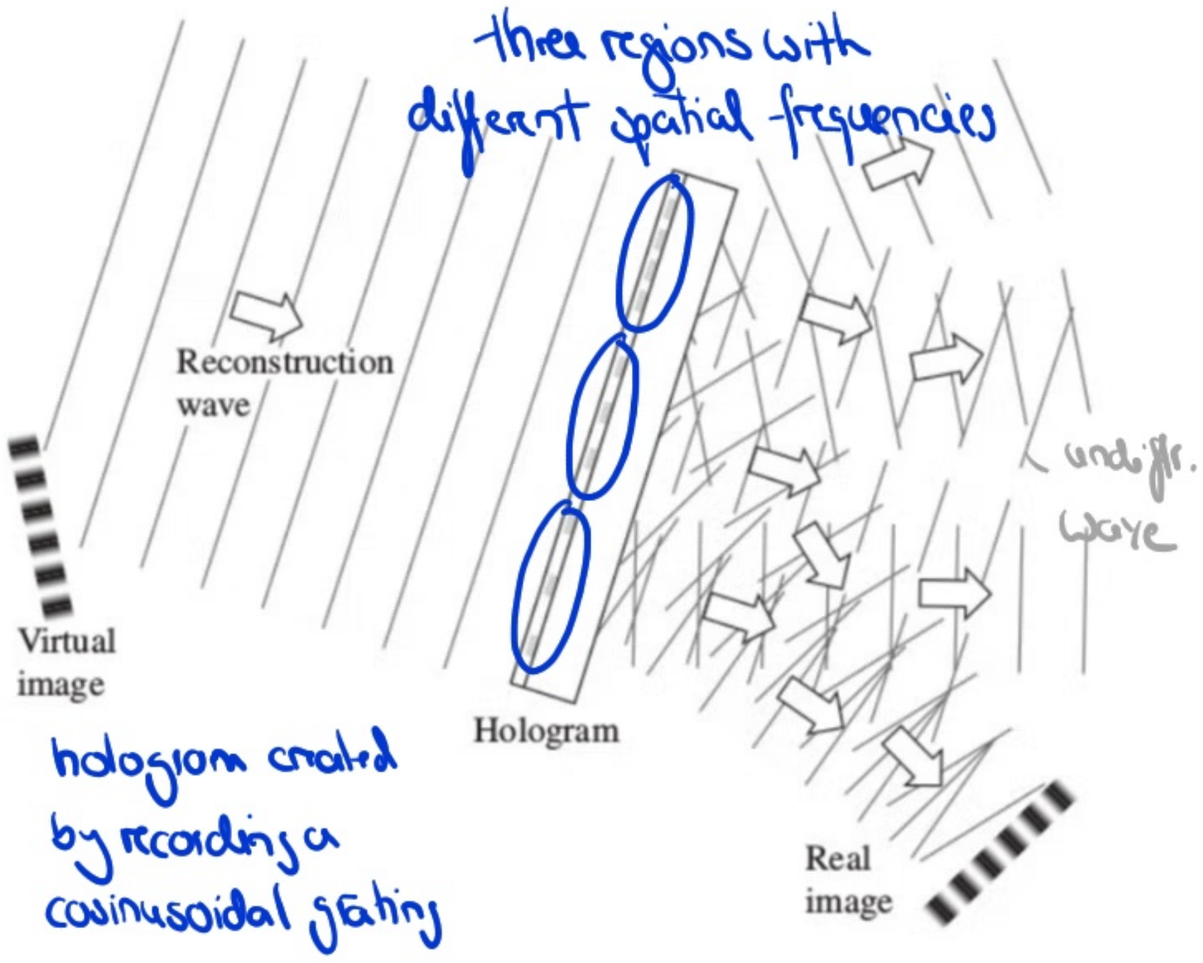
see Lecture 12

with a cosinusoidal irradiance distribution

$$\underline{\underline{I = 2E_0^2(1 + \cos\phi)}}.$$

Upon processing the film so that the amplitude transmission profile corresponds to I, we essentially obtain a cosinusoidal diffraction grating. Upon illumination with a plane wave identical to original reference wave, different beams will emerge ie. the different diffraction orders. One of the first order beams will continue in the direction of the original object wave and thus correspond to its reconstructed wavefront.





More complex objects will subsequently result in a more complicated diffraction pattern, encoding the relative phase angle ϕ between the object and the reference waves. Moreover if their amplitudes are different, the irradiance of the fringes would have been altered accordingly. Shining the reconstruction beam onto the resulting pattern again leads to a diffraction process that produces a virtual 3D image. If you were looking 'into' the hologram (like looking through a window), you could 'see' the object like it were sitting there, ie. it has complete three-dimensionality; e.g. as you move the image moves (you see different image depending on where you stand), there is motion parallax (things in the foreground move faster than those in the background) and you have 'real' stereoscopic view / depth perception (both eyes see different images, turned into a 3D picture). Another interesting aspect is that each part of the hologram contains information about the entire object. Thus each fragment of the hologram can produce the entire image albeit with different resolution.

So far, we have focused on transmission holography (object and reference wave traverse the film from the same side), but something similar happens if both waves traverse the film from opposite sides. The resulting fringe pattern is referred to as reflection hologram and has to be viewed from the side of the reference beam (as if looking into a mirror).