Optics-2

- **Reflection and Transmission**
- **Diffraction**

1 Reflection and Transmission

1A) Introduction

In this part we examine the issues of EM waves at the boundary between two media. When an EM wave with given amplitude, frequency, and direction is incident on a plane interface between two media, what will happen? We know that some part of the wave will be reflected, and some part will be transmitted. But in what direction will the reflected wave propagates? What is the amplitude of the reflected wave? What is the frequency of the reflected wave? The same questions can be asked for the transmitted wave also. In this part of the lecture, we will answer all these questions by simply using the boundary conditions of electric and magnetic fields that are presented in the lecture of EM-4.

To get started, let us first look at an interesting mathematics problem. Consider the equation

$$ax^2 + bx + c = 0. \qquad (1a)$$

Usually what is being implicitly asked is: at what value(s) of x in terms of a, b, c will Eq. 1a hold? The answer is, as we all know,

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \qquad (1b).$$

Now, if we ask: under what conditions will Eq. 1a hold *for all values of x*, then what will be the answer? The answer is $a = b = c = 0$.

Similarly, what is the condition for the following equation to be true *for all values of x and y*?

$$a_2 x^2 + a_1 x + a_0 + b_2 y^2 + b_1 y + b_0 = 0 \qquad (1c)$$

The answer is $a_2 = a_1 = b_2 = b_1 = a_0 + b_0 = 0$. We will meet similar equations when dealing with EM waves at a boundary.

1B) Basic Concept

Consider a boundary between medium-1 and medium-2. The refractive indexes of the two media are n_1 and n_2, respectively. The first refractive index is real, as the medium carries the incident EM wave. The second refractive index can be complex (recall the cases in lecture EM-4). The frequency of the incident wave ω_1 and the components of its wave vector $\vec{k}_1$ are all real. The two are related by the dispersion relation

$$k_1 = \frac{n_1\omega_1}{c} \qquad (2a)$$

Likewise, the EM wave in medium-2 must have

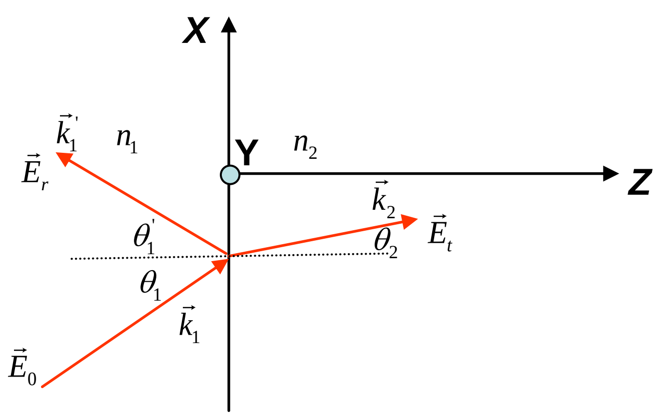
$$k_2 = \frac{n_2\omega_2}{c} \qquad (2b)$$

As is usually the case, it is assumed that there is no surface charge or surface current at the interface. The media are non-magnetic, so $\mu_1 = \mu_2 = 1$.

The boundary condition for the electric and magnetic fields are

$$\vec{E}_1^{//} = \vec{E}_2^{//} \qquad (3a)$$
$$D_1^{\perp} = D_2^{\perp} \qquad (3b)$$
$$B_1^{\perp} = B_2^{\perp} \qquad (3c)$$
$$\vec{H}_1^{//} = \vec{H}_2^{//} \qquad (3d)$$

The boundary conditions in Eqs. 3 must hold *everywhere on the interface and at all time*. Eqs. 2 and 3 are the basic physics concepts from which all wave phenomena at the boundary are derived.



Note that the arrows in the figure are for *indication of the propagation directions* of the respective EM waves only. They are not narrow light rays as in the geometric optics. The plane EM waves are infinitely wide. In fact, the incident wave and the reflected wave fill up the entire half space occupied by medium-1, and the transmitted wave fills up the whole half space of medium-2.

Define the plane formed by the wave vector $\vec{k}_1$ of the incident wave and the normal direction of the interface as the incidence plane, we choose the coordinate system such that the X-Z plane is the incidence plane, the X-axis and the Y-axis are parallel to the interface, and the Z-axis is perpendicular to the interface. The interface is at z = 0. The Y-axis is perpendicular to the incidence plane. The incident wave is linearly polarized.

1C) Polarization perpendicular to incidence plane

We first consider the case in which the electric field of the incident wave is along the Y-axis. The incident wave is

$$\vec{E}_0(\vec{r},t) = E_0\vec{y}_0 e^{i(k_{1x}x+k_{1z}z-\omega_1 t)} = E_0\vec{y}_0 e^{i(\vec{k}_1\cdot\vec{r}-\omega_1 t)} \qquad (4a).$$

$\vec{k}_1$ in Eq. 4a must satisfy the dispersion relation Eq. 2a. The incidence angle, defined as the angle between $\vec{k}_1$ and the interface normal, is given by $\tan\theta_1 = \frac{k_{1x}}{k_{1z}}$.

In order not to make things overly complicated, we assume for the time being that the electric fields of the reflected wave and the transmitted wave only have the Y-component. Likewise, we assume that the propagation directions of the reflected wave and the transmitted wave are still in the incidence plane. After going through the derivation below, one will see that one can easily do away with the possible presence of none Y-components, or the off-plane wave vector components, and show that they must all vanish. The reflected wave is then

$$\vec{E}_r(\vec{r},t) = E_r\vec{y}_0 e^{i(k'_{1x}x-k'_{1z}z-\omega'_1 t)} = E_r\vec{y}_0 e^{i(\vec{k}'_1\cdot\vec{r}-\omega'_1 t)} \qquad (4b).$$

$\vec{k}'_1$ in Eq. 4b must satisfy the dispersion relation Eq. 2a ($k'_1 = \frac{n_1\omega'_1}{c}$), since the reflected wave is also in medium-1. The reflection angle, defined as the angle between $\vec{k}'_1$ and the interface normal, is given by $\tan\theta_r = \frac{k'_{1x}}{k'_{1z}}$.

The transmitted wave is

$$\vec{E}_t(\vec{r},t) = E_t\vec{y}_0 e^{i(k_{2x}x+k_{2z}z-\omega_2 t)} = E_t\vec{y}_0 e^{i(\vec{k}_2\cdot\vec{r}-\omega_2 t)} \qquad (4c)$$

If $\vec{k}_2$ is real, then the refractive angle can be defined likewise $\tan\theta_2 = \frac{k_{2x}}{k_{2z}}$.

The total electric field in the space of $z \le 0$ is $\vec{E}_0(\vec{r},t)+\vec{E}_r(\vec{r},t)$, and that in the space $z \ge 0$ is $\vec{E}_t(\vec{r},t)$. Everything about the incident wave and the media are known. That means E_0, $\vec{k}_1$, ω_1, n_1, and n_2 are given. We need to find everything about the reflected and the transmitted waves. What we have are the boundary conditions of Eqs. 3 which must be satisfied at the interface z =

0, the dispersion relations of Eqs. 2, plus the form of the electric fields in Eqs. 4. You may wonder how many equations we must have in order to also the unknowns. But the equations we have are not ordinary ones. They must be satisfied *everywhere on the interface and at all time*.

As the electric fields have no components perpendicular to the interface, Eq. 3b is automatically satisfied. Applying Eq. 3a at $z = 0$, we get

$$E_0 e^{i(k_{1x}x-\omega_1 t)} + E_r e^{i(k_{1x}x-\omega_1' t)} = E_t e^{i(k_{2x}x-\omega_2 t)} \qquad \text{(5a)}$$

Equation 5a must be satisfied at all coordinate values of *x* and time *t*. Consider *t* first. To satisfy Eq. 5a at all time, the only possibility is that the frequencies are all equal. That is

$$\omega_1 = \omega_1' = \omega_2 \qquad \text{(5b).}$$

Now we see that the frequency of the wave does not change when it goes through or is reflected by the interface. We will simply refer to the frequency of the waves as ω from now on. Using the dispersion relation in medium-1, we then get

$$k_1 = k_1' \qquad \text{(5c)}$$

Likewise, Eq. 5a must be satisfied at all values of *x*. This leads to the second result

$$k_{1x} = k_{1x}' = k_{2x} \qquad \text{(5d).}$$

That is, the components of the wave vectors of the three waves parallel to the interface must be the same. Eqs. 5c and 5d then lead to $k_{1z}' = \pm k_{1z}$. $k_{1z}' = k_{1z}$ means that the incident wave dose not change direction at the boundary, which cannot be true. $k_{1z}' = -k_{1z}$ means the reflected wave is bounced back by the interface, which makes sense. Combining with Eq. 5d, one can verify that this is exactly the reflection law: The incidence angle is equal to the reflection angle $\theta_1' = -\theta_1$. So this so called reflection law can actually be derived from the general EM wave theory.

If $\vec{k}_2$ is real, using Eqs. 2 and Eq. 5d, one can get the Snell's law

$$n_1 \sin\theta_1 = n_2 \sin_2 \qquad \text{(5e).}$$

So Snell's law can also be derived from general EM wave theory. Note that

$$k_{2z} = \sqrt{k_2^2 - k_{2x}^2} \qquad \text{(5f).}$$

In the case of real refractive indexes,

$$k_{2z} = \frac{\omega}{c}\sqrt{n_2^2 - n_1^2 \sin^2\theta_1} \qquad \text{(5g).}$$

It can become imaginary if $n_2 < n_1 \sin\theta_1$. This is the condition of total reflection. Let $k_{2z} = i\frac{\omega}{c}\sqrt{n_1^2\sin^2\theta_1 - n_2^2} \equiv i\kappa$. The transmitted wave then takes the form

$$\vec{E}_t(\vec{r},t) = E_t\vec{y}_0 e^{i(k_{2x}x + k_{2z}z - \omega_2 t)} = E_t\vec{y}_0 e^{-\kappa z} e^{i(k_{1x}x - \omega_2 t)} \qquad \text{(5h)}$$

It propagates along the interface and its amplitude decays exponentially with the distance to the interface. This is the interface mode of the wave propagation. As will be shown later, $|E_r| = E_0$ when this happens.

Finally, Eq. 5a is reduced to

$$E_0 + E_r = E_t \qquad \text{(5i).}$$

Equation 5i alone cannot uniquely determine E_r and E_t. We need another equation to fulfill the requirement. We have used only two boundary conditions related to the electric field so far. Now we will use the two related to the magnetic field. Using the Faraday's Law in the Maxwell Equations $\vec{B} = \frac{\vec{k}\times\vec{E}}{\omega}$, we have the magnetic fields of the incident, reflected, and transmitted waves as

$$\vec{B}_0(\vec{r},t) = \frac{\vec{k}_1\times\vec{E}_0(\vec{r},t)}{\omega} = \frac{E_0}{\omega}(k_{1x}\vec{x}_0 + k_{1z}\vec{z}_0)\times\vec{y}_0 e^{i(\vec{k}_1\cdot\vec{r}-\omega t)} = \frac{E_0}{\omega}(k_{1x}\vec{z}_0 - k_{1z}\vec{x}_0)e^{i(\vec{k}_1\cdot\vec{r}-\omega t)} \qquad \text{(6a)}$$

$$\vec{B}_r(\vec{r},t) = \frac{\vec{k}_1'\times\vec{E}_r(\vec{r},t)}{\omega} = \frac{E_r}{\omega}(k_{1x}\vec{x}_0 - k_{1z}\vec{z}_0)\times\vec{y}_0 e^{i(\vec{k}_1'\cdot\vec{r}-\omega t)} = \frac{E_r}{\omega}(k_{1x}\vec{z}_0 + k_{1z}\vec{x}_0)e^{i(\vec{k}_1'\cdot\vec{r}-\omega t)} \qquad \text{(6b)}$$

$$\vec{B}_t(\vec{r},t) = \frac{\vec{k}_2\times\vec{E}_t(\vec{r},t)}{\omega} = \frac{E_t}{\omega}(k_{2x}\vec{x}_0 + k_{2z}\vec{z}_0)\times\vec{y}_0 e^{i(\vec{k}_2\cdot\vec{r}-\omega t)} = \frac{E_t}{\omega}(k_{1x}\vec{z}_0 - k_{2z}\vec{x}_0)e^{i(\vec{k}_2\cdot\vec{r}-\omega t)} \qquad \text{(6c)}$$

Note that the frequency in the above equations is the same, and in Eq. 6c k_{2x} has been replaced by k_{1x} according to Eq. 5d.

The total magnetic field in the $z \le 0$ region is $\vec{B}_0(\vec{r},t) + \vec{B}_r(\vec{r},t)$, while in the $z \ge 0$ region it is $\vec{B}_t(\vec{r},t)$. The components perpendicular to the interface is along the Z-axis, and the parallel components are along the X-axis. One can verify that Eq. 3c is automatically satisfied, and Eq. 3d leads to, after eliminating some common factors on both sides of the equation,

$$E_0 k_{1z} - E_r k_{1z} = E_t k_{2z} \qquad \text{(6d)}$$

Solving Eqs. 5i and 6d, we finally get the reflection and transmission of the EM waves.

$$r_\perp \equiv \frac{E_r}{E_0} = \frac{1-\beta}{1+\beta} \qquad (6e)$$

$$t_\perp = \frac{E_t}{E_0} = \frac{2}{1+\beta} \qquad (6f)$$

$$\beta \equiv k_{2z} / k_{1z} \qquad (6g)$$

The $\perp$ sign indicates that what we have obtained are for the cases where the polarization of the incident wave is perpendicular to the incidence plane. When all wave vector components are real, which means total reflection cases and cases with complex n_2 are *excluded*, Eq. 6g becomes

$$\beta = \frac{n_2 \cos\theta_2}{n_1 \cos\theta_1} \qquad (6h)$$

$r_\perp$ and $t_\perp$ in Eqs. 6e and 6f can then be expressed in terms of incident and refraction angles. Such formulae are commonly seen in many textbooks. One should remember, however, that expressing $r_\perp$ and $t_\perp$ in such angles is only valid when all wave vector components are real. The $r_\perp$ and $t_\perp$ expressed in terms of $\beta \equiv k_{2z} / k_{1z}$, on the other hand, is valid for ALL situations. This is because complex wave vectors have well defined physical meanings. Complex wave vectors mean exponential decay or decrease of the amplitude of the wave with distance. Complex angles, on the other hand, have no physical meanings. When all wave vector components are real, the propagation direction of the wave can be specified by the angle. If any one of the components is complex, no such direction can be specified and the description of the wave propagation mode has to be modified accordingly.

In the case of total reflection $n_2 < n_1 \sin\theta_1$, $k_{2z} = \frac{\omega}{c}\sqrt{n_2^2 - n_1^2 \sin^2\theta_1} = i\frac{\omega}{c}\sqrt{n_1^2 \sin^2\theta_1 - n_2^2} = i\kappa$ is purely imaginary and so is β. Expressing β as $\beta = iA$ where A is a real number, one then see that $r_\perp = \frac{1-iA}{1+iA}$, and $|r_\perp| = 1$, i. e., total reflection occurs.

1D) Polarization in the incidence plane

In the case the electric field of the incident wave is in the incidence plane, the magnetic field is then perpendicular to the incidence plane. From Faraday's law, the amplitude of the magnetic field and that of the electric field is related by $B = \frac{n}{c}E$. The magnetic fields of the incident, reflected, and transmitted waves are

$$\vec{B}_0(\vec{r},t) = B_0 \vec{y}_0 e^{i(k_{1x}x + k_{1z}z - \omega_1 t)}$$

$$\vec{B}_r(\vec{r},t) = B_r \vec{y}_0 e^{i(k_{1x}^{'}x - k_{1z}^{'}z - \omega_1^{'} t)} \qquad (7a)$$

$$\vec{B}_t(\vec{r},t) = B_t \vec{y}_0 e^{i(k_{2x}x + k_{2z}z - \omega_2 t)}$$

The total magnetic field in the space of $z \le 0$ is $\vec{B}_0(\vec{r},t)+\vec{B}_r(\vec{r},t)$, and that in the space $z \ge 0$ is $\vec{B}_t(\vec{r},t)$. One can see that since there is no magnetic field component perpendicular to the interface, boundary condition Eq. 3c is automatically satisfied. Equation 3d leads to the same results as stated in Eqs. 5b and 5d, and the reflection law that the reflection angle is equal to the incidence angle. The remaining equation becomes

$$B_0 + B_r = B_t \qquad \text{(7b).}$$

From Faraday's law, the amplitude of the magnetic field and that of the electric field is related by $B = \frac{n}{c}E$. Replacing magnetic fields by electric fields in Eq. 7b, we get

$$n_1(E_0 + E_r) = n_2 E_t \qquad \text{(7b').}$$

Using Ampere's Law $\vec{k}\times\vec{B} = -\frac{n^2}{c^2}\omega\vec{E}$, we get

$$\begin{aligned}
\vec{E}_0(\vec{r},t) &= \frac{c}{n_1 k_1} B_0\left(k_{1z}\vec{x}_0 - k_{1x}\vec{z}_0\right)e^{i(k_{1x}x+k_{1z}z-\omega t)} \\
\vec{E}_r(\vec{r},t) &= \frac{c}{n_1 k_1} B_r\left(-k_{1z}\vec{x}_0 - k_{1x}\vec{z}_0\right)e^{i(k_{1x}x-k_{1z}z-\omega t)} \qquad \text{(7c)} \\
\vec{E}_t(\vec{r},t) &= \frac{c}{n_2 k_2} B_t\left(k_{2z}\vec{x}_0 - k_{1x}\vec{z}_0\right)e^{i(k_{1x}x+k_{2z}z-\omega t)}
\end{aligned}$$

One can verify that Eq. 3b leads to the same equation 7b, and Eq. 3a leads to

$$\frac{k_{1z}}{n_1}(E_0 - E_r) = \frac{k_{2z}}{n_2}E_t \qquad \text{(7d)}$$

Solving Eqs. 7b and 7d, we get

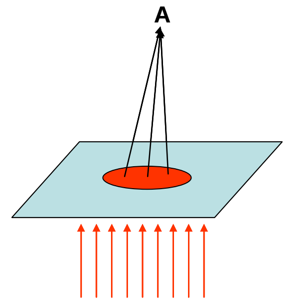
$$r_{//} = \frac{\frac{n_2}{n_1} - \frac{n_1}{n_2}\beta}{\frac{n_2}{n_1} + \frac{n_1}{n_2}\beta} = \frac{n_2^2 - \beta n_1^2}{n_2^2 + \beta n_1^2} \qquad \text{(7e),}$$

$$t_{//} = \frac{2}{\frac{n_2}{n_1} + \frac{n_1}{n_2}\beta} = \frac{2n_1 n_2}{n_2^2 + \beta n_1^2} \qquad \text{(7f).}$$

The definition of β is still the same as Eq. 6g. Equations 6e, 6f, 7e, 7f are called Fresnel's equations.

2 **Diffraction**

2A) Introduction



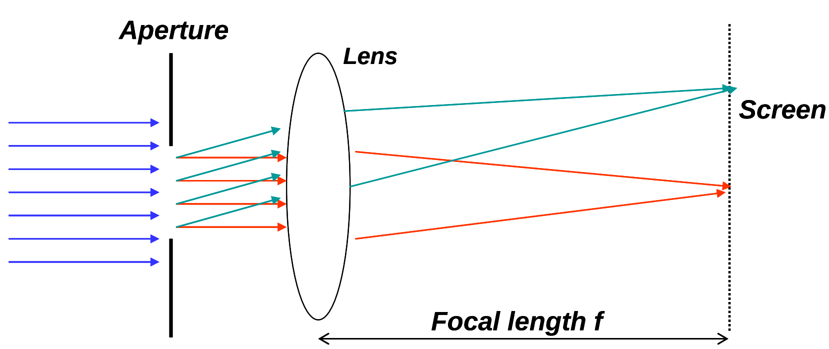
So far we have studied the propagation of EM waves in infinitely large space, and at the boundary between two media. The diffraction of EM waves deals with the problem of EM waves encountering an obstacle, such as a small hole or slit. As shown in the figure, an opening (aperture) is illuminated by a plane wave from below, and one is to find the EM wave above the aperture. The rest of the infinitely large screen is opaque. The exact solutions to such problems require solving the Maxwell equations with the given boundary conditions, and usually there are no analytical solutions so numerical ones are sought instead. A good approximation, however, exists and that is the theory of diffraction of EM waves. According to such theory, each point on the aperture is a source of EM wave. The EM waves above the aperture is the combined contributions of all these sources. The problem of diffraction is then converted to the relatively simple calculation of these contributions at every point in space behind the screen.

For simplicity, we ignore the factor $\frac{1}{2}c\varepsilon_0$ from now on and set the intensity of light equal to the amplitude square of the electric field, $I = |E|^2$. The opening is on a flat plane. In part-2B through part-2E we take the approximation that the observation screen is far away from the aperture. Diffraction under such condition is called Fraunhofer diffraction. In part-2F we deal with the problems that such approximation no longer holds. These are called Fresnel diffraction.

2B) Single long slit

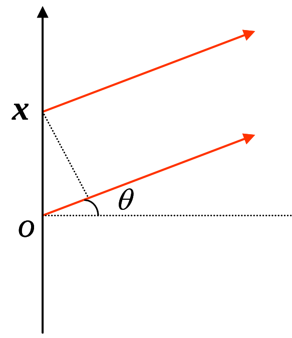
(i) Normal incidence

Consider the simplest case, the long slit of width *a*. The figure below shows the side view of the diffraction configuration. A plane wave of amplitude E_0 and wave vector *k* is normally incident onto the slit from the left side. A focusing lens is placed behind the aperture, and a viewing screen is placed on the focal plane of the lens.



Taking the point source at $x = 0$ as the reference point, the phase difference between the source at $x = 0$ and any point at position x is

$$\delta = kx\sin\theta \cong kx\frac{X}{f} \qquad \text{(8a),}$$



where X is the coordinate on the screen in the vertical direction. Each element dx at position x contributes $E_0 e^{ikx\sin\theta} dx / f$ to the total field on the screen. The total electric field due to the contributions from all the point sources on the opening is the sum of each field,

$$E(\theta) = \frac{E_0}{f}\int_{-a/2}^{a/2} e^{ikx\sin\theta} dx = \frac{E_0}{ikf\sin\theta}\left(e^{i\frac{ka\sin\theta}{2}} - e^{-i\frac{ka\sin\theta}{2}}\right) = \frac{2aE_0}{f}\left(\frac{\sin\alpha}{\alpha}\right) \qquad \text{(8b),}$$

$$\text{where } \alpha = \frac{1}{2}ka\sin\theta \qquad \text{(8c)}$$

Using

$$\sin\theta \cong \theta \cong \frac{X}{f} \qquad \text{(8d),}$$

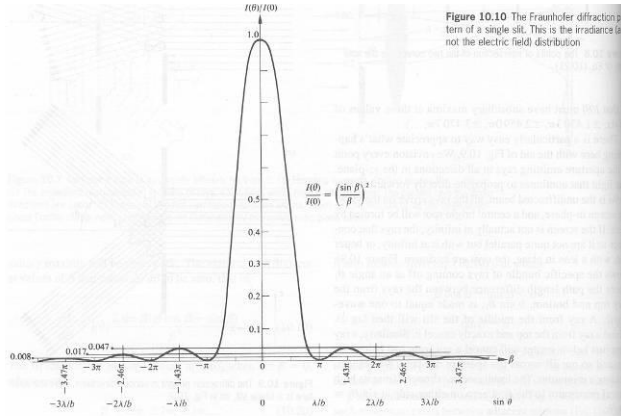
we can convert the angular distribution $E(\theta)$ to the spatial distribution on the screen $E(X)$. The intensity distribution is simply

$$I(\theta) = \frac{1}{2}|E|^2 = \frac{2a^2E_0^2}{f^2}\left(\frac{\sin\alpha}{\alpha}\right)^2 = I_0\left(\frac{\sin\alpha}{\alpha}\right)^2 \qquad \text{(8e)}$$

In the case where the lens is absent, and the distance from the screen to the aperture is D, then according to Eq. 6a in the lecture notes Optics-1 where we are dealing with the double-slit interference, the phase difference is

$$\delta = kx\sin\theta \cong kx\frac{X}{D} \qquad \text{(8f).}$$

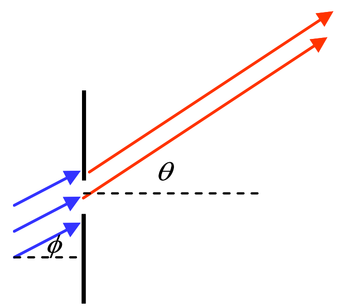
We should simply do a replacement using $\sin\theta \cong \theta \cong \frac{X}{D}$ in Eq. 8e to obtain the spatial distribution of intensity on the screen.



The major maximum of intensity occurs at $\alpha = 0$, or at $X = 0$. The minor maxima occur at $\alpha = \tan\alpha$. The roots of this equation have to be found numerically. The minima occur at $\beta = \pm m\pi$, $m = 1, 2, \ldots$, as shown above. For an infinitely large slit, only the forward propagation ($\alpha = 0$) remains. This is consistent with the fact that in infinite space the plane waves propagate in a straight direction. On the other hand, when the slit is very small, $\alpha \to 0$, $\dfrac{\sin\alpha}{\alpha} \to 1$, $I(\theta) = I_0$. The slit then behaves like a point source.

(ii) Oblique incidence

So far the point sources on the opening are all in phase. In the case of oblique incidence with angle ϕ, there is a phase difference which depends on the position of the point source $\delta(x) = kx\sin\phi$ among the sources. A point source at position x emits wave $E_0 e^{i(kx\sin\theta - \delta(x))}dx$. As an exercise, show that the intensity distribution maintains the form given by Eq. 8e but with

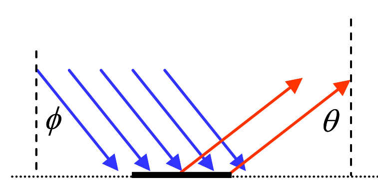


$$\alpha = \frac{1}{2}ka(\sin\theta - \sin\phi) \qquad \text{(8g).}$$

Maximum intensity occurs at $\theta = \phi$.

(iii) Reflective 'slit'.

Diffraction also occurs when an EM wave is incident on a long strip of reflective surface instead of a slit. Each point on the reflective surface is a source of EM wave. The situation is much the same as in (ii). As an exercise, show that show that the intensity distribution maintains the form given by Eq. 8e but with



$$\alpha = \frac{1}{2}ka(\sin\theta + \sin\phi) \qquad \text{(8h).}$$

Maximum intensity occurs at $\theta = -\phi$, i. e., following the reflection law off an infinitely large surface. This is also true when $a \to \infty$. The only maximum occurs at $\alpha = 0$ and $\theta = -\phi$.

2C) Double slits and many slits

(i) Double slits

Consider two identical slits of width *a* and separate by a distance *b* (> *a*) measured from the centers of the slits. The wave is normally incident on the aperture. All others are the same as in the single slit case. Put the centers of the slits at $x = \pm \frac{b}{2}$, respectively, the electric field on the screen is then (keep in mind that $\sin\theta = \frac{X}{f}$)

$$E(\theta) = \frac{E_0}{f}\left(\int_{b/2-a/2}^{b/2+a/2} e^{ikx\sin\theta}dx + \int_{-b/2-a/2}^{-b/2+a/2} e^{ikx\sin\theta}dx\right) = \frac{4aE_0}{f}\frac{\sin\alpha}{\alpha}\cos(\frac{1}{2}kb\sin\theta) \qquad (9a)$$

The intensity is

$$I(\theta) = \frac{1}{2}|E|^2 = I_0\left(\frac{\sin\alpha}{\alpha}\right)^2\cos^2\beta \qquad (9b)$$

The definition of α follows Eq. 8c, and $\beta = \frac{1}{2}kb\sin\theta$. Comparing with the double-slit interference result presented in Optics-1, we find that there is an extra factor $\left(\frac{\sin\alpha}{\alpha}\right)^2$ in Eq. 9b, which is due to the diffraction effect. The factor $\cos^2\beta$ is due to the interference effect between the two slits.

(ii) *N* slits

Consider *N* identical slits of width *a* under normal incidence. The central positions of the slits are at $x = mb$, $m = 0, 1, \ldots N-1$. The electric field due to the *m*-th slit is

$$E_m(\theta) = \frac{E_0}{f}\int_{mb-a/2}^{mb+a/2} e^{ikx\sin\theta}dx = \frac{E_0}{f}\left(e^{i(mb+\frac{a}{2})k\sin\theta} - e^{i(mb-\frac{a}{2})k\sin\theta}\right) = E_1(\theta)e^{imkb} \qquad (10a)$$

Here $E_1(\theta)$ is the electric field due to a single slit given by Eq. 8c. It is then straightforward to show that the total electric field due to all slits is

$$E(\theta) = E_1(\theta)(1 + e^{ikb\sin\theta} + e^{2ikb\sin\theta} + ... + e^{ikb(N-1)\sin\theta}) = E_1(\theta)\left(\frac{e^{iNkb\sin\theta} - 1}{e^{ikb\sin\theta} - 1}\right) \qquad (10b).$$

The intensity is

$$I(\theta) = I_0 \left(\frac{\sin\alpha}{\alpha}\right)^2 \left(\frac{\sin(N\beta)}{\sin\beta}\right)^2 \qquad (10c).$$

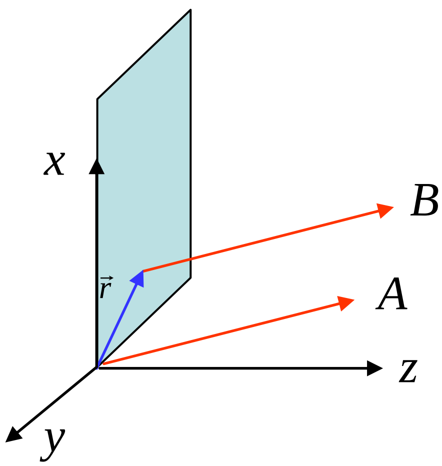
(iii) Gratings

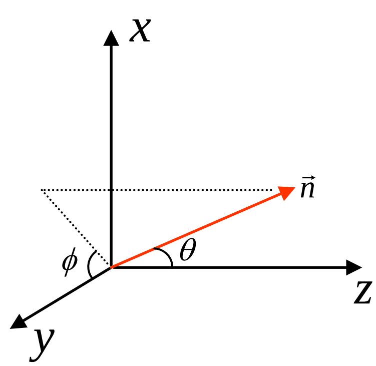
Gratings can be viewed as multiple slits in reflection geometry. Instead of *N* slits in transmission mode, we have *N* reflecting stripes. Considering the simplest case of normal incidence, the diffraction intensity is then given by Eq. 10c also. The resolution of the gratings follows the same Eq. 10d as in the transmission geometry.

2D) Rectangular apertures

(i) Single aperture

Take a coordinate system as shown in the left figure below. The planer aperture is in the *x*-*y* plane. The emission direction is given by $\vec{n}$, as shown in the right figure.





According to the above figure,

$$\vec{n} = \sin\theta\sin\phi\vec{x}_0 + \sin\theta\cos\phi\vec{y}_0 + \cos\theta\vec{z}_0 \qquad (11a)$$

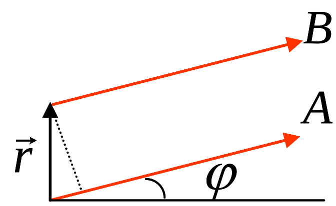
If a focusing lens of focal length *f* is placed behind the aperture, the wave emitted in such a direction will fall onto the position of

$$X = f\sin\theta\sin\phi \qquad (11b)$$

$$Y = f \sin\theta \cos\phi \qquad (11c)$$

of the screen on the focal plane of the lens. If there is no lens and the screen is at a (far) distance of D from the aperture, replace f in the above equations by D.

Suppose the aperture is uniformly illuminated. What is the phase difference between the wave A emitted from the point source at the origin and wave B emitted by the point at the position (x, y) as marked by the position vector $\vec{r} = x\vec{x}_0 + y\vec{y}_0$? Let us take the plane made by $\vec{r}$ and $\vec{n}$, as shown in the figure. It is obvious that the path difference is



$$\delta d = r\sin\varphi = r\cos(\frac{\pi}{2} - \varphi) = \vec{r}\cdot\vec{n} \qquad (11d).$$

Therefore each point source of area $dxdy$ at $\vec{r} = x\vec{x}_0 + y\vec{y}_0$ emits wave

$$dE = \frac{E_0}{f^2} e^{ik\vec{r}\cdot\vec{n}} dxdy = \frac{E_0}{f^2} e^{ik(x\sin\theta\sin\phi + y\sin\theta\cos\phi)} dxdy = \frac{E_0}{f^2} e^{ik(\frac{xX}{f} + \frac{yY}{f})} dxdy \qquad (11e).$$

$$\begin{aligned} E &= \frac{E_0}{f^2} \int_{-a/2}^{a/2}\int_{-b/2}^{b/2} e^{ik(x\sin\theta\sin\phi + y\sin\theta\cos\phi)} dxdy = \frac{E_0}{f^2}\left(\int_{-a/2}^{a/2} e^{ikx\sin\theta\sin\phi} dx\right)\left(\int_{-b/2}^{b/2} e^{iky\sin\theta\cos\phi} dy\right) \\ &= 4E_0\left(\frac{\sin\alpha}{\alpha}\right)\left(\frac{\sin\beta}{\beta}\right) \end{aligned} \qquad (11f)$$

Here

$$\alpha = \frac{1}{2} ka\sin\theta\sin\phi = \frac{1}{2}\frac{X}{f} ka \qquad (11g)$$

$$\beta = \frac{1}{2} kb\sin\theta\cos\phi = \frac{1}{2}\frac{Y}{f} kb \qquad (11h)$$

Finally, the intensity distribution is given by

$$I = I_0\left(\frac{\sin\alpha}{\alpha}\right)^2\left(\frac{\sin\beta}{\beta}\right)^2 \qquad (11i).$$

A typical diffraction pattern is shown in the figure.

Note that for an aperture of arbitrary shape the total electric field is

$$E = \frac{E_0}{f^2}\iint_S e^{ik(x\sin\theta\sin\phi + y\sin\theta\cos\phi)} dxdy \qquad (12).$$

Only when S is rectangular can the surface integral be carried out as in Eq. 11f. Some shapes can be considered as combination of multiple rectangles. So they can be dealt with using multiple-aperture diffraction approach.

(ii) Multiple apertures

Before considering multiple apertures, let us first consider an aperture which is centered at (x_0, y_0) instead of (0, 0). Still taking (0, 0) as the reference point for phase difference, the total electric field then becomes

$$\begin{aligned} E &= \frac{E_0}{f^2}\left(\int_{x_0-a/2}^{x_0+a/2} e^{ikx\sin\theta\sin\phi}dx\right)\left(\int_{y_0-b/2}^{y_0+b/2} e^{iky\sin\theta\cos\phi}dy\right) \\ &= 4E_0\left(\frac{\sin\alpha}{\alpha}\right)\left(\frac{\sin\beta}{\beta}\right)e^{ik(x_0\sin\phi+y_0\cos\phi)\sin\theta} \end{aligned} \qquad (13a).$$

The intensity distribution is still the same as Eq. 11i.

Now, for two identical apertures centered at (x_1, y_1) and (x_2, y_2), respectively, the total field is then

$$E = 4E_0\left(\frac{\sin\alpha}{\alpha}\right)\left(\frac{\sin\beta}{\beta}\right)(e^{ik(x_1\sin\phi+y_1\cos\phi)\sin\theta} + e^{ik(x_2\sin\phi+y_2\cos\phi)\sin\theta}) \qquad (13b)$$

The extra factor $(e^{ik(x_1\sin\phi+y_1\cos\phi)\sin\theta} + e^{ik(x_2\sin\phi+y_2\cos\phi)\sin\theta})$ leads to the interference effect between the two apertures. Using the same approach, multiple apertures can be dealt with accordingly.

2E) Circular aperture

For a circular aperture of radius a one should simply use Eq. 12 to carry out the integral. Due to cylindrical symmetry of the geometry, the diffraction pattern depends only on θ. The pattern on the screen is a series of concentric circular bands, as shown in the figure. The angular width of the central peak (spot) is $I(\theta) = I_0\left(\frac{2J_1(ka\sin\theta)}{ka\sin\theta}\right)^2$. $J_n(x)$ is a special function called the Bessel Function of the first kind. They can be calculated numerically. The first root of $J_1(x) = 0$ is x = 3.83. Let $ka\sin\theta = 3.83$, we get the angular width of the first maximum

$$\theta_{\max} = 0.61\frac{\lambda}{a} \qquad (14c)$$

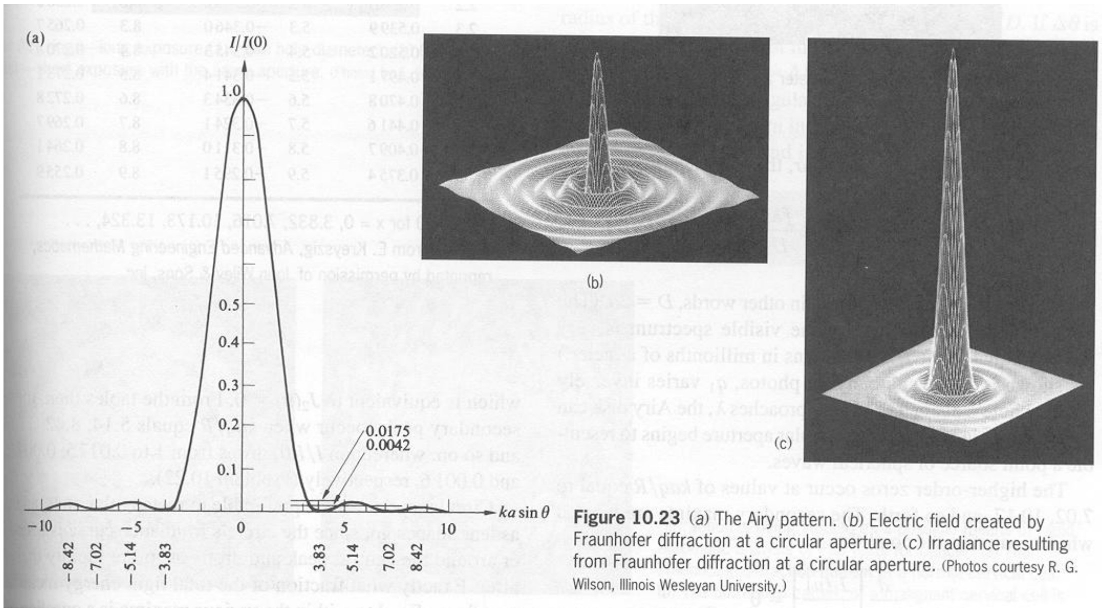


Figure 10.23 (*a*) The Airy pattern. (*b*) Electric field created by Fraunhofer diffraction at a circular aperture. (*c*) Irradiance resulting from Fraunhofer diffraction at a circular aperture. (Photos courtesy R. G. Wilson, Illinois Wesleyan University.)

The light spot on the screen on the focal plane of the lens then has the radius of

$$r_{\max} = f\theta_{\max} = 0.61\frac{\lambda f}{a} \qquad (14d)$$

If two light beams with the incidence angles differing only by $\delta\varphi < \theta_{\max}$, then the two spots on the screen will overlap on top of one another and cannot be resolved. This sets the limit to distinguish two nearly parallel beams, or the resolution, of the lens. The smaller the $\theta_{\max}$ is, the better the resolution will be. Therefore, one should use lenses as large as possible to achieve high resolution. The same is true for mirrors. That is one of the two major reasons that the primary lenses or mirrors of astronomy telescopes are always very large. The other reason is to collect as much light from distant objects as possible. The collecting power is obviously proportional to the area of the primary mirror.

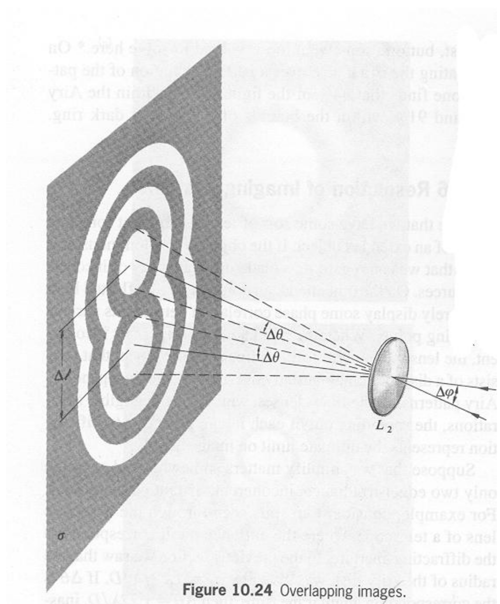
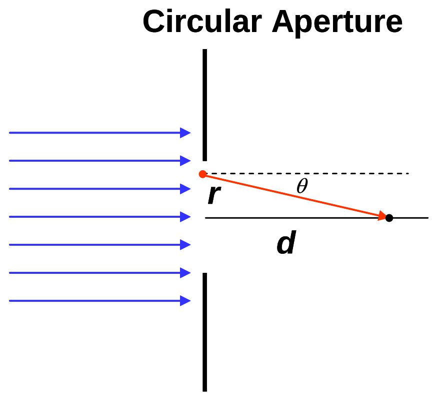


Figure 10.24 Overlapping images.

2F) Fresnel Diffraction

The problems we have studied so far are under the condition that the observation distance is much larger than the aperture size, i. e., far field diffraction. When such condition no longer holds, the diffraction theory has to be implemented with its exact form. Rather than giving a general formula, we study an example to illustrate how it works. Consider a circular aperture illuminated by a plane wave at normal incidence, as shown in the figure. We want to find the electric field at a position on the central axis of the aperture at distance *d* from the aperture. Each point on the aperture is still a new point source of EM waves. The contribution of a concentric ring of radius *r* and width *dr* on the aperture is, according to Fraunhofer diffraction,



$$dE = \frac{2\pi r dr}{\sqrt{d^2+r^2}}\cos(k\sqrt{d^2+r^2}) \qquad (15a).$$

Here $2\pi r dr$ is the area of the ring, $\sqrt{d^2+r^2}$ is the distance from the source to the observation point, $\cos(k\sqrt{d^2+r^2})$ is the phase, and we take the real form for convenience. In the exact form of diffraction theory, however, there should also be a direction factor $\cos\theta$, where θ is the angle between the normal direction of the aperture and the source-to-point distance vector. In the present case,

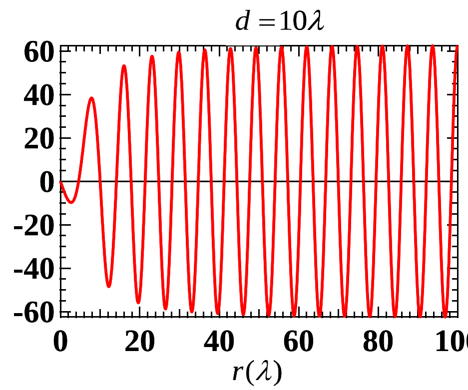
$$\cos\theta = \frac{d}{\sqrt{d^2+r^2}} \qquad (15b).$$

So the total field is

$$E = E_0\int_0^R 2\pi r\frac{d}{d^2+r^2}\cos(\frac{2\pi}{\lambda}\sqrt{d^2+r^2})\mathrm{d}r \qquad (15c).$$

The integration in Eq. 15c cannot be carried out analytically. We therefore examine the equation numerically instead. The integrant function in Eq. 15c is

$$f(r) = 2\pi r\frac{d}{d^2+r^2}\cos(\frac{2\pi}{\lambda}\sqrt{d^2+r^2}) \qquad (15d).$$



Assuming $R = 100\lambda$ and $d = 10\lambda$, the integrant function is plotted in the right figure. One can see that the function turns to positive and negative alternatively as *r* increases. If the aperture is not a clear circular hole, but is decorated with a series of concentric opaque rings. The radii and widths of the rings are such that they correspond to the regions of *r* when *f(r)* is positive (or negative). The positive regions of *f(r)* then become zero in value and only the negative ones are left. This will lead to the integration to become very large because the negative ones are no longer canceled by the positive ones. The effect is like the plane wave is focused by the aperture. Such a device is called zone plate. It can focus a plane wave without curved surface like lenses. This brings certain advantages in opto-electronics in telecommunications. The disadvantage is that it only works for a particular wavelength, while lenses work for a broad range of wavelengths.