Lecture #5.4

Interference and Diffraction of EM Waves

During our previous lectures we have been talking about electromagnetic (EM) waves. As we know, harmonic waves of any type represent periodic process in time and in space. In the case of EM wave, it is a pattern where changing electric field causes appearance of the changing magnetic field which again causes appearance of changing electric field and so on. This process can be described mathematically with the help of Maxwell’s equations. It has been shown that EM wave is a transverse wave, where both electric and magnetic fields are oscillating in the direction perpendicular to the direction of the wave’s transmission. At the same time electric field is perpendicular to magnetic field. In harmonic sin-like electromagnetic wave both electric and magnetic fields are changing with time according to the sin-like law. These changes for both fields are in phase, so both electric and magnetic fields are reaching their maximum and minimum values simultaneously.

We saw that theoretical prediction about the speed of EM waves in vacuum made by Maxwell \( c = \sqrt{\frac{1}{\varepsilon_0\mu_0}} = 3 \times 10^8 \frac{m}{s} \) are confirmed in Fizeau’s experiment, measuring the speed of light. So, it is reasonable to believe that visible light is an electromagnetic wave. If we assume that light is EM wave then, having the same speed, light waves may have different wavelengths and frequencies. The wavelength, frequency and speed of light in vacuum are related according to

\[
c = \lambda f .
\]

So, the higher frequency means the lower wavelength. Visible part of light spectrum covers the range of wavelengths from \( 3.7 \times 10^{-7} \) m for violet color to \( 7.5 \times 10^{-7} \) m for red color. The waves with lower wavelengths represent ultraviolet light; the waves with higher wave lengths represent infrared light. As we can see those values of wavelengths are extremely small compared to the size of a typical every-day macroscopic objects. That is why we were not concerned with wave properties of light when we discussed geometrical optics and formation of images of macroscopic objects. We have explained all optical phenomena related to reflection and refraction without even bothering to think about structure of EM waves. The only property we used was that light wave propagates
along the straight path from the source. In fact, we could use all the same explanation without even assuming that light was a wave. It still will be working fine if we treat light as a stream of some particles.

So, until this moment we still do not have a concrete proof that light is, indeed, EM wave. The only phenomenon, we discussed which supports this point of view, is polarization of light. But being a wave, light should also demonstrate other wave properties. These properties include interference and diffraction. The reason why we could not observe them so far is because we were considering the objects, whose size is much larger than a typical wavelength of visible light. This, by the way, is the reason why we are able to see these objects at all. Now we shall turn our attention to the wave properties of light. The part of physics, which studies these wave properties, is called wave (or physical) optics.

1. Interference

Looking back at our discussion of mechanical waves, we know that two different mechanical waves arriving to the same point in space can interfere producing either constructive or destructive interferential pattern. If light is an electromagnetic wave then the similar phenomenon should occur for light as well. The difficulty is that, in order to observe interference, one has to have light waves with consistent phase difference. However, the light sources which we use in our everyday life, such as different light bulbs or even the Sun, all produce incoherent light. In general this light represents a mixture of waves of different wavelengths (different colors) coming from different microscopic light sources. Instead of having waves of the same frequency and with the certain value of the phase difference, we usually see a result of addition (superposition) of many different waves, which produce the light of averaged in time and in space intensity.

In order to obtain light’s interferential pattern, one has to have several monochromatic (of the same color or wavelength) and coherent (with constant value of the phase difference) light sources. Nowadays it is possible by using lasers. In older times some other methods had to be developed. One of the methods how this can be achieved is using light from the same source and spreading it in two rays. Those two rays then can be used to produce an interferential picture. This experiment is known, as the Young’s
double-slit experiment, since it was first performed by English physicist Thomas Young (1173-1829) in 1801. The light first passes one slit in a screen in order to obtain thin ray, then this ray passes through a double slit in the next screen producing two rays of consistent phase difference. Since those waves are originated from the same source, they have the same phase as they travel to the same distance in space or if the path difference between those waves is equal to the integer number of wavelengths. In those points in space one can see maxima of the light intensity (constructive interference). If the path difference is equal \((m + \frac{1}{2})\) wavelengths (where \(m\) is the integer number) then destructive interference occurs. The maxima and minima of interferential picture can be seen on a screen placed far enough from the slits.

The aforementioned conditions are the general conditions necessary for existence of maxima and minima of interferential picture, which is

\[
\Delta l = m\lambda, m = 0, 1, ...
\]

\[
\Delta \phi = 2\pi m, m = 0, \pm 1, ...
\]

for constructive interference and

\[
\Delta l = \left( m + \frac{1}{2} \right)\lambda, m = 0, 1, ...
\]

\[
\Delta \phi = 2\pi \left( m + \frac{1}{2} \right), m = 0, \pm 1, ...
\]

for destructive interference. At the same time as electric and magnetic fields are added or subtracted the intensity of light, being proportional to the square of the field amplitude changes according to

\[
I = CA^2 = C(A_1 \pm A_2)^2 = C(A_1^2 \pm 2A_1A_2 + A_2^2) = I_1 + I_2 \pm 2\sqrt{I_1I_2},
\]

where “plus” sign takes place for constructive and “minus” sign for destructive interference.

Another example of interference takes place in thin liquid films. You can observe it by looking at soap bubbles having different colors on the surface of the film. The reason for this phenomenon is the reflection of light from the surface of the thin film. The light reflects from both outer and inner surfaces of the film. Two waves, coming out as a result of this reflection, have a consistent phase difference (being produced by the same
original light source). So, they can interfere, producing maxima and minima of intensity for some particular color.

Phase difference may change during the reflection of EM wave from the boundary of the other medium. This situation is very similar to the one which we studied when talking about standing waves in strings. If a string has its end tied to a solid support then the phase of this wave changes to the opposite during the reflection. On the other hand, if a string has a loose end then it will move in a same pattern as the incoming wave and there is no phase difference during the reflection. A similar situation occurs with light wave hitting the boundary between the two substances. If the index of refraction of the second medium is lower than the index of refraction of the first medium then there is no phase change. If, however, light reflects from the substance with higher index of refraction, it changes its phase to the opposite.

Taking this into account, we can consider interference of light, which occurs as light passes through a thin air wedge between the two glass plates. In this situation the incoming light wave is partly reflected from the boundary between the upper glass plate and air wedge and it does not change its phase. Another part of the wave penetrates into the air wedge and then is reflected from the second glass plate. In this case it changes its phase by 180 degrees (or $\lambda/2$ in terms of wavelength). Both reflected rays may interfere with each other producing an interferential pattern. The path difference between the two waves is going to be twice the thickness, $d$, of the air wedge. Therefore, the constructive interference occurs if

$$\frac{\lambda}{2} + 2d = m\lambda, m = 0,1,...$$

$$\frac{1}{2} + \frac{2d}{\lambda} = m, m = 0,1,...$$

The destructive interference takes place if

$$\frac{\lambda}{2} + 2d = \left( m + \frac{1}{2} \right)\lambda, m = 0,1,...$$

$$\frac{1}{2} + \frac{2d}{\lambda} = m + \frac{1}{2}, m = 0,1,...$$

$$\frac{2d}{\lambda} = m, m = 0,1...$$
A similar situation takes place in the famous experiment, which was originally conducted by Newton and now referred to as the “Newton’s rings”-experiment. In this experiment the upper glass plate is replaced by a curved piece (convex lens). As a result a series of circular interferential fringes (rings) can be seen.

Finally let us consider what happens, when interference occurs in a thin liquid or soup film. This time the light changes its phase, when it is reflected from the outside surface of the film, and it does not change its phase when it is reflected from the inside surface. In order to find the path difference for the two waves, we have to take into account the fact that the wavelength \( \lambda_n = \frac{\lambda}{n} \) of light inside of the film with refraction index \( n \) is different compared to its value \( \lambda \) in vacuum. For the film of thickness \( d \), the total phase difference between the reflected rays is \( \frac{2d}{\lambda_n} - \frac{1}{2} = \frac{2dn}{\lambda} - \frac{1}{2} \), so the constructive interference occurs if

\[
\frac{2dn}{\lambda} - \frac{1}{2} = m, m = 0, 1, \ldots
\]

and destructive interference will take place if

\[
\frac{2dn}{\lambda} = m - \frac{1}{2}, m = 0, 1, \ldots
\]

These general principles form the basis of CD technology, where laser’s ray is reflected from the thin film on the surface of a CD, forming interferential pattern, which can be used for reading of information of the CD.

2. Diffraction

The ability of light, which is originally emitted from one light source, to travel around obstacles or through apertures producing secondary light rays is called diffraction. Those diffracted light rays can then interfere to produce interferential patterns.

In the Young’s double-slit experiment the conditions of constructive and destructive interference were considered in equations 5.4.2 and 5.4.3. The following picture illustrates light leaving the slits and shows how \( \Delta l \) is related to geometry of the system
In this picture $d$ is the distance between the two slits, $L$ is the distance between the slits and the screen, $y$ is the distance from the central interferential maximum on the screen to either maximum or minimum of interferential pattern. The fact whether or not it is maximum or minimum is determined by conditions 5.4.2 or 5.4.3. It is easy to see from the picture that equations 5.4.2 and 5.4.3 can be rewritten in terms of distance $d$ and angle $\theta$ as

$$d \sin \theta = n \lambda, n = 0, 1, ... \quad (5.4.9)$$

for constructive interference and

$$d \sin \theta = \left( n + \frac{1}{2} \right) \lambda, n = 0, 1, ... \quad (5.4.10)$$

for destructive interference. At the same time, one can relate angle $\theta$ to other two variables as

$$y = L \tan \theta \quad (5.4.11)$$

Note that if the angle $\theta$ is small enough then there is no difference between the sine of this angle, the tangent of this angle and the angle itself. In this situation, known as “the small angle approximation”, all of the equations become linear and the problem becomes extremely simple.

We have referred to diffraction as the ability for EM waves to travel around the obstacles. The reason why it becomes possible is because, when the EM hits the obstacle or the slit, each boundary point of this obstacle or each point of the slit becomes the source of the secondary spherical EM wave. These secondary waves can then interact,
producing interferential patterns on the screen. So, we can see that it is not even necessary to have two slits. It is possible to obtain interferential picture from one slit as long as the size of the slit is comparable to the wavelength. This problem is, of course, much more complicated compared to the two slit experiment. Now it has to do with the interference not of the two but of the infinite number of secondary EM waves from all the different points of the single slit. Without going in complicated technical details, we can divide all the waves in pairs such as the distance between the two sources emitting those waves is $w/2$, where $w$ is the width of the slit. It is easy to see that maximum of light intensity occurs right in front of the center of the slit. If we want to find positions of the minima of interferential picture, we can consider superposition of light waves in each pair, mentioned above, and apply the same condition 5.4.2 as for the double slit experiment. For the first minimum it will give

$$\frac{w}{2} \sin \theta = \frac{\lambda}{2},$$

$$w \sin \theta = \lambda.$$  \hspace{1cm} (5.4.12)

The next minimum occurs due to the interference of the secondary light source separated by a distance $w/4$ and so on. So, the general condition for having minimum of interferential pattern is

$$w \sin \theta = n\lambda, n = 0,1...$$  \hspace{1cm} (5.4.13)

The central fringe for the single slit diffraction is going to be the brightest one. It carries most of light’s intensity and has the angular size which can be calculated as $2\theta$ from equation 5.4.12 and gives (in the limit of small angles) $2\frac{\lambda}{w}$.

Surely, not only slits of rectangular shape may produce diffraction pictures but aperture of any shape if it is small enough. In particular, a circular opening will produce a diffraction pattern in the form of the bright and dark circles surrounding the central bright spot. This puts the limits on resolution of human vision, since a human eye has the aperture of the finite size. Exact mathematical calculations show that for a circular aperture of diameter $D$, the angular position of the first dark fringe can be found according to equation similar to 5.4.12 which is

$$\sin \theta = 1.22 \frac{\lambda}{D}$$  \hspace{1cm} (5.4.14)
This condition puts limits on the ability to distinguish two light sources as separate sources or as one source. If a distance between the two sources is too small and their diffraction pictures overlap in such a way that the dark fringe of the first diffraction pattern goes through the center of the second source, then those sources look like one single source of light.

If diffraction picture can be produced by the two or even single slit, it surely can be produced by a series of slits separated by a very small distances $d$, known as diffraction grating. Diffraction grating works almost as two-slit screen, except the intensity of all the maxima of interferential picture is going to be the same in the limit of very large number of slits per unit of length on the grating. Positions of these maxima are still determined by equation 5.4.2. Since those positions depend on wavelength of light, diffraction grating works as a glass prism in the Newton’s experiment allowing to expand white light into rays of different colors.

Not only visible light, but any type of EM wave should go through diffraction as it passes through diffraction grating of appropriate scale. This should not necessarily be one-dimensional grating as we just discuss. For instance, one can consider diffraction of X-rays on 3-dimesional crystalline structure of the solid substance. In fact, the study of these 3-dimesional diffraction pictures is the only way for us to learn about geometrical shape of those crystals.