# LIGHT DIFFRACTION

## 3.1. Huygens-Fresnel principle

The phenomenon of interference is a clear confirmation of the wave nature of light. However, review of the wave properties of light would be incomplete without consideration of phenomenon of light diffraction.

Since ancient times, observation of the distribution of light on the boundary between light and shadow of objects of different shapes forced researchers to think about the possible “rounding” obstacles by light. The first mention of it was found in the works of the famous artist, scientist and researcher Leonardo da Vinci (1452–1519).

In 1665, Grimaldi described similar phenomena in detail. Prominent German physicist of the XX century Sommerfeld (1868–1951) defined the phenomenon of diffraction of light as deviation of light rays from a straight line that cannot be explained by reflection, refraction or distortion of rays in media with variable index of refraction.

Diffraction of light, like interference, is explained by the wave nature of light and leads to light penetration into the zone of geometrical shadow. Phenomenon of diffraction is associated with interference.

If light propagation in a homogeneous medium is free (unobstructed), interference of coherent secondary waves, which are generated by every elementary section of the wave surface, causes rectilinear propagation of light. Conversely, light waves meeting with obstacles cannot form full shade and bend around the obstacles, getting into the zone of the geometrical shadow.

There are two types of diffraction. If an obstacle is placed near the light source and the screen, then falling or diffracted waves have curved (usually spherical) surface.

This case is called Fresnel diffraction. If falling and diffracted waves can be considered as plane waves, this phenomenon is called Fraunhofer diffraction.

Plane waves, necessary for the Fraunhofer diffraction, are obtained by distancing of the light source and screen from the obstacle or by using lenses.

There is no fundamental physical difference between diffraction and interference of light. Both phenomena are caused by the light flux redistribution as a result of superposition (overlapping) of coherent waves.

Historically, redistribution of the light intensity arising as a result of superposition of waves excited by finite number of coherent sources is called interference. Redistribution of the light intensity arising from the superposition of waves excited by infinite number of coherent sources disposed continuously (for example, elementary sections of the wave surface) is called diffraction.

Famous Dutch physicist Huygens was the first who proved the wave nature of light in 1690. The basis of the theory of light propagation is a principle named after him. Huygens principle states that *every point to which a light wave comes can be regarded as a source of secondary coherent waves*. To determine the wave front at later moment of time, we should build a surface that encircles these secondary waves. Using this principle Huygens explained the straightness of light propagation and laws of reflection and refraction. However, the amplitude (intensity) redistribution and rounding the obstacles by waves (light deviations from linearity) were not mentioned.

The phenomenon of diffraction within the wave theory was explained by French physicist Fresnel in 1818; he completed the Huygens principle idea of interference of secondary waves. Considering the amplitudes and phases of the secondary waves gives possibility of finding amplitude and intensity of the resultant wave at any point in space. Modified Huygens principle is called the principle of Huygens-Fresnel; it is the basic principle of wave optics. The principle allows us to consider the intensity of the resultant wave in different directions and solve the problem of diffraction of light. In accordance with the Huygens-Fresnel principle, secondary hemispherical waves are coherent. Hence, intensity of the resultant wave at some point of the screen is a result of interference of all secondary waves coming to the point.

### 3.2. Fresnel zones

Consider a monochromatic light which wavelength is $\lambda$; it spreads in a homogeneous environment. For simplicity, we assume a point source; therefore, a closed surface $S$ around the source is a sphere with the radius $a$, the elementary areas $ds$ of the surface are mutually coherent (Fig. 3.1).

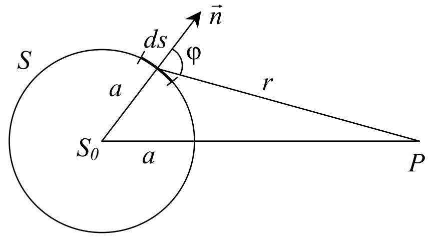



Fig. 3.1

In accordance with the Huygens-Fresnel principle, each elementary area *ds* of the luminous surface *S* (wavefront) is regarded as the center of the secondary source (secondary waves). Oscillations of light waves coming to the point *P* from each elementary area *ds* are described by the equation:

$$dE = K(\varphi)\frac{a_0 ds}{r}\cos(\omega t - kr + \alpha), \tag{3.1}$$

where $A = K(\varphi)\frac{a_0 ds}{r}$ is the amplitude of oscillations in the point *P* caused by the action of the surface element *ds*; $a_0$ is a factor which is determined by the amplitude of the light wave at the location of the area *ds*; $(\omega t + \alpha)$ is the light wave phase at the location of the wave surface *S*; $k = 2\pi/\lambda$ is the wave number; *r* is the distance from the element *ds* to the point *P*. Fresnel assumed that the slope coefficient $K(\varphi)$ depends on the angle $\varphi$ between the normal $\vec{n}$ to the surface element *ds* and the direction to the point *P*; it ranges from 1 ($\varphi = 0$) to 0 ($\varphi \geq \pi/2$). This means that the secondary sources do not emit in the opposite direction, it is why Fresnel secondary waves are hemispherical.

Kirchhoff gave mathematical reasoning and refinement of the Huygens–Fresnel principle in 1882.

He showed that $K(\varphi) = (1 + \cos\varphi)/\lambda$. This means that the amplitude of the secondary waves is zero not when $\varphi \geq \pi/2$, as Fresnel thought, but only if $\varphi = \pi$. However, in most cases of diffraction observation, the angle $\varphi$ is small (close to zero), so this explanation does not affect the final result.

The resulting oscillation in the point *P* is the superposition of oscillations *dE* of all elements *ds* of the surface *S*:

$$E = \int_S K(\varphi)\frac{a_0}{r}\cos(\omega t - kr + \alpha)ds\,. \quad (3.2)$$

If a closed surface $S$ is arbitrary taken, calculation of the integral (3.2) is difficult. However, as shown by Fresnel, the problem is greatly simplified, and the integration can be replaced by a simple algebraic or graphical addition, if the wave front is spherical. This method of approximate calculation of integrals (3.2) was named the *method of Fresnel zones*.

In accordance with this method, the spherical wave front is divided into annular zones centered at the point $O$ so that the lengths of straight lines that connect edges of the annular zones with the observation point $P$ are differed by $\lambda/2$. Fig. 3.2 shows that $a$ is a distance from the source to the top of the wave surface $O$; $b$ is a distance from the top to the observation point $P$; $b + m\lambda/2$ is a distance from the outer edge of the $m$-th zone to the point $R$; $m$ is the number of Fresnel zone.

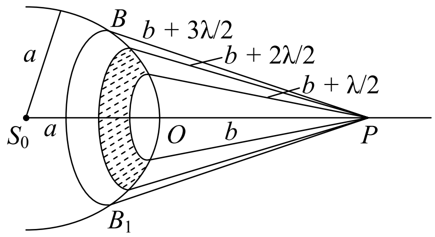



Fig. 3.2

This separation of the wave surface to Fresnel zones leads to the fact that oscillations of the waves coming to the point $P$ from the neighboring Fresnel zones are opposite by phases, because the path difference between two adjacent zones to the point $P$ is $\lambda/2$. Therefore, the phase difference between neighboring Fresnel zones equals $\pi$. Thus, the resulting oscillation amplitude at point the $P$ can be represented as a series of amplitudes with alternating signs:

$$A = A_1 - A_2 + A_3 - A_4 + \ldots \pm A_m\,, \quad (3.3)$$

where $A_m$ is the amplitude of the oscillations from the $m$-th Fresnel zone in the point $P$. Sign «+» corresponds to the odd and «–» to the even-numbered zones.

As it follows from equation (3.1), the amplitude of oscillations generated by $m$-th Fresnel zone in the point $P$ depends on the area of

$m$-th zone, the angle between the outer normal to the surface area and the direction to the point $P$, and the distance between the $m$-th zone and the observation point $P$. To define an area $\Delta S_m$ of $m$-th Fresnel zone, consider Fig. 3.3.

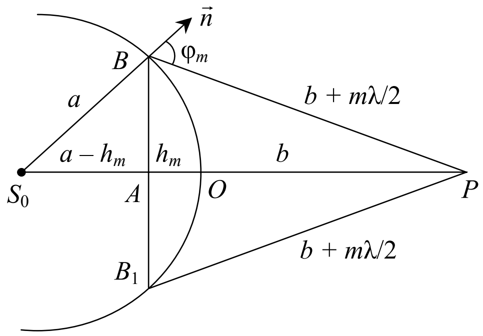



Fig. 3.3

The points $B$ and $B_1$ of the spherical segment $BOB_1$ show the outside border of the Fresnel zone with the index $m$, the radius of the zone is $AB = AB_1 = r_m$. From the triangles $PBA$ and $S_0BA$ we get:

$$r_m^2 = a^2 - (a - h_m)^2 = \left(b + m\frac{\lambda}{2}\right)^2 - (b + h_m)^2, \tag{3.4}$$

from (3.4) we define the height of the spherical segment

$$h_m = \frac{bm\lambda + (m\lambda/2)^2}{2(a+b)}. \tag{3.5}$$

Consider that $\lambda << b$, therefore, the second term in the numerator of the equation (3.5) can be neglected for small values of $m$. Finally, the height of the $m$-th segment is:

$$h_m = \frac{bm\lambda}{2(a+b)}. \tag{3.6}$$

Considering that $h_m << a$, we get from (3.4) that the outer radius of the $m$-th Fresnel zone is $r_m^2 = 2ah_m$, taking into account the equation (3.6), we get the final result:

$$r_m = \sqrt{2ah_m} = \sqrt{\frac{ab}{a+b}m\lambda}. \tag{3.7}$$

If a wave that propagates from the source is plane, then $a \to \infty$ and

$$r_m = \sqrt{m\lambda b}\ . \tag{3.7a}$$

Inside of the spherical segment $BOB_1$ (see Fig. 3.2 and 3.3) *m* Fresnel zones are placed and the segment area is $S_m = 2\pi a h_m$ (*a* is the sphere radius, $h_m$ is the height of the spherical segment); the segment area is equal to the sum of areas of *m* Fresnel zones:

$$S_m = 2\pi a h_m = \frac{\pi ab}{a+b} m\lambda = \Delta S_1 + \Delta S_2 + ... + \Delta S_m\ .$$

The area of the first ($m=1$) zone is: $\Delta S_1 = \dfrac{\pi ab\lambda}{a+b}$, the area of *m*-th zone is:

$$\Delta S_m = S_m - S_{m-1} = \frac{\pi ab\lambda}{a+b}\ .$$

Thus, the areas of all Fresnel zones for small *m* ($m<10$) can be considered as approximately equal: $\Delta S_1 = \Delta S_2 = ... = \Delta S_m$. But the distance $b_m$ from a zone to the point *P* slowly increases with the number of the zone *m*. The angle $\varphi_m$ between the normal to an area and the direction to the point *P* also increases with *m*; hence, slope coefficient $K(\varphi_m)$ decreases with increasing number of area. All this leads to the fact that the amplitude $A_m = K(\varphi_m)\dfrac{E_0 \Delta S_m}{b_m}$ of the oscillation from *m*-th zone in the point *P* gradually decreases with the increase of the number of zone. This means that the amplitudes of the oscillations in the point *P* monotonically (very slowly) decrease creating a sequence:

$$A_1 > A_2 > ... > A_{m-1} > A_m > A_{m+1} > ...$$

Let us return to the alternating series (3.3) and represent it, for example for $m=3$, as

$$A = \frac{A_1}{2} + \left(\frac{A_1}{2} - A_2 + \frac{A_3}{2}\right) + \frac{A_3}{2}\ ,$$

and for $m=4$ as

$$A = \frac{A_1}{2} + \left(\frac{A_1}{2} - A_2 + \frac{A_3}{2}\right) + \left(\frac{A_3}{2} - \frac{A_4}{2}\right) - \frac{A_4}{2}.$$

Because of the monotonic decrease of the amplitudes, it can be assumed that $A_2 \approx A_1$, $A_2 \approx A_3$, $A_2 \approx A_1/2 + A_3/2$ (in the general case ($A_m \approx \approx (A_{m-1} + A_{m+1})/2$). Consequently, the equation in brackets is equal to zero, so finally for $m = 3$ we get $A = \frac{A_1}{2} + \frac{A_3}{2}$. Similarly, for $m = 4$, we get: $A = \frac{A_1}{2} - \frac{A_4}{2}$. Summarizing these results, we obtain:

$$A = \frac{A_1}{2} \pm \frac{A_m}{2}, \qquad (3.8)$$

where the sign «+» corresponds to the odd and «–» to the even number of *m*. If the wave front is completely open ($m \to \infty$) and $A_m \to 0$, we obtain:

$$A = \frac{A_1}{2}. \qquad (3.9)$$

The equation (3.9) shows that the light wave amplitude emitted by a completely open source in a point *P* is equal to half of the amplitude of the oscillations excited in this point by only the first one (central) Fresnel zone. If the opaque barrier with a hole that leaves open only the central Fresnel zone is placed on the path of spherical wave, the amplitude in the point *P*, according to the equation (3.8), equals $A_1$, that exceeds twice the amplitude *A* (3.9). This means that the light intensity in the point *P* if the obstacle is present, four times greater than the intensity if the obstacle is absent. At first glance, this result is paradoxical, but it is well confirmed by the experiments (see. chapter 3.4).

### 3.3. Graphical calculation of the resultant amplitude. Zone plate

The resulting oscillations amplitude can be obtained using a graphical method of adding oscillations. Harmonic oscillations can be represented as a vector which length is equal to the oscillations amplitude and the angle between the vector and the direction of the oscillations propagation is the oscillations phase. It is supposed that the amplitude of the vector rotates counterclockwise around the axis passing through its beginning. Its angular velocity is considered as cyclic frequency of the oscillations.

When we add oscillations, the resulting oscillations vector is equal to the vector sum of the components vectors and the angle that the resultant vector makes with the direction of oscillations propagation creates the resulting phase.

To represent graphically the whole first Fresnel zone, it should be divided from the center into equal subzones so narrow that the phase of every subzone could be considered constant.

Then the oscillations caused by the action of the first subzone at the point *P* can be represented by a vector which length is proportional to the amplitude and direction is determined by the phase of this subzone. As a result of increasing the distance *r* and the reduction of the slope coefficient $K(\varphi)$ (see Fig. 3.1), the vector of oscillations of each consequent subzone decreases in magnitude and has the phase lag from the oscillations generated by the previous subzone. So, the second (adjacent) subzone can be represented by a vector a little turned relatively to the first vector.

The length of this vector does not almost differ from the first one, since the value of *r* changes very little. Thus, the vector diagram that defines the action of a number of subzones, forming a Fresnel zone, is shown by a broken line.

This line is a chain of vectors $\Delta\vec{A}_i$, the sum of which will be a resulting oscillation amplitude $\vec{A}_P = \sum \Delta\vec{A}_i$ of a Fresnel zone at the point *P*.

Fig. 3.4 *a*, *b*, *c* shows vector diagrams where the values of the resulting vectors $\vec{A}_1$, $\vec{A}_2$ and $\vec{A}_3$ are amplitudes of oscillations excited in the point *P* respectively by the first, the first two and the first three Fresnel zones.

Fig. 3.4 shows that the first Fresnel zone is divided into six subzones $i = 6$. Since the width of each Fresnel zone corresponds to a change in the phase by π, then each subzone vector creates the angle $\delta = \pi/6$ with the previous one.

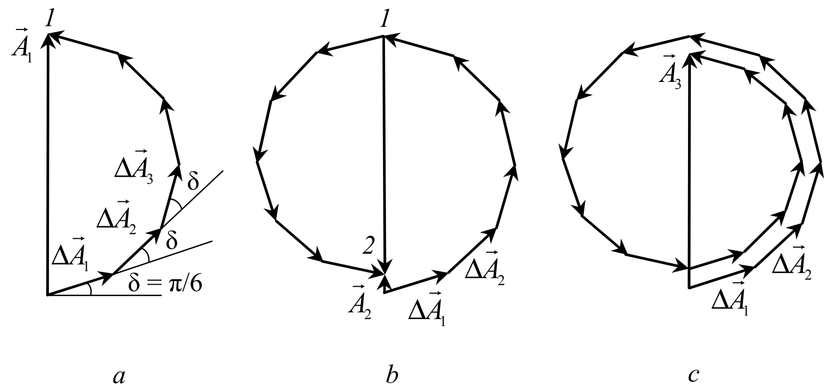



Fig. 3.4

If we divide each Fresnel zone into an infinite number of subzones, the broken line becomes a curve and each Fresnel zone will be shown as one semicircle. Obviously, if all Fresnel zones are open (entirely open wavefront), we obtain a spiral with a focus at the point $A_\infty$ (Fig. 3.5) which is called *Fresnel spiral*. Fig. 3.5 shows that $A_\infty = A_1/2$, it coincides with the result of algebraic addition of (3.9).

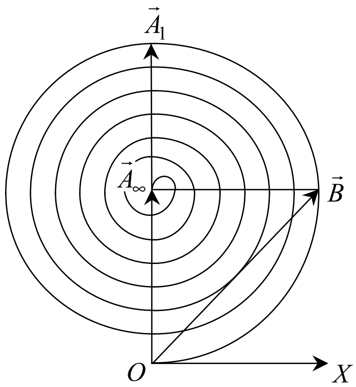



Fig. 3.5

Oscillations of light waves of even and odd Fresnel zones in the point *P* have opposite phases and therefore mutually weak each other. If on the way of a light wave we put an obstacle in the form of small plates that would overlap all even or odd areas, the light intensity in the point *P* will increase significantly.

This obstacle is called *amplitude zone plate* that can be made using Newton's rings pattern. Amplitude zone plate acts like a convex lens.

Greater effect can be achieved if we change the phase of oscillations of even or odd Fresnel zones to the opposite one, i.e. by π. This plate is called *phase zone plate*.

The first phase zone plate was made by Wood. He covered the glass with a thin layer of paint and engraved phase zone plate so that the optical thickness of the odd zones differed from even ones by the thickness $\lambda/2$.

A phase plate increases the resulting amplitude twice and light intensity four times in comparison with an amplitude plate.

### 3.4. Fresnel diffraction by simple obstacles

**Diffraction by round hole.** Let us place an opaque screen with a circular aperture of the radius $r_0$ towards a spherical wave from the point source $S_0$. The screen $E$ is placed perpendicularly to the line passing through the point source $S_0$ and the center of the hole $O$ (Fig. 3.6).

If $r_0$ is much less than the lengths $a$ and $b$, the length $a$ can be considered as the distance from the source to the obstacle and the length $b$ can be considered as distance from the obstacle to the point $P$.

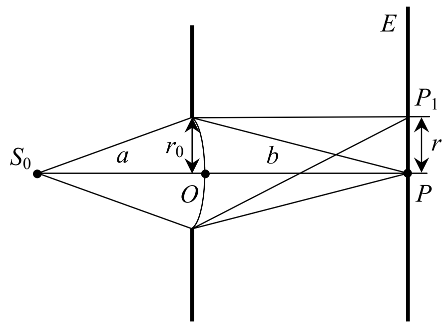



Fig. 3.6

The method of Fresnel zones makes it easy to determine the intensity of light in the point $P$ at the center of the diffraction pattern. If the distances $a$ and $b$ satisfy the relation (3.7)

$$r_0 = r_m = \sqrt{\frac{ab}{a+b} m\lambda} \qquad (3.10)$$

where *m* is integer, the hole opens exactly *m* of the first Fresnel zones.

Thus, the intensity in the point *P* is determined by the number of Fresnel zones *m*, which fit into the hole. If *m* is an even number, the intensity in the point *P* is minimal and for the small values of *m* is almost zero because the oscillations of the light waves in the point *P*, which are generated by adjacent Fresnel zones, are antiphase and mutually destroyed.

Accordingly, the intensity at the point *P* is maximal if *m* is an odd number, because the oscillations from one zone will not compensate. Moreover, as it can be seen from the equation (3.3) for small *m,* amplitude $A_m$ differs a little from $A_1$. Thus, for any odd *m,* the resultant amplitude in the point P is approximately equal to $A_1$. From the equation (3.10) we can see that the number of open Fresnel zones is:

$$m = \frac{r_0^2}{\lambda}\left(\frac{1}{a} + \frac{1}{b}\right).$$

Calculation of light intensity in the other points of the screen is much more complex since the relevant Fresnel zones are partially closed by the opaque obstacle.

But for reasons of axial symmetry and according to the law of conservation of energy it is obvious that the total diffraction pattern around the point *P* in monochromatic light must be in the form of concentric light and dark alternating rings transferred smoothly into each other.

If the distance from the point *P* increases, intensity maxima decrease. If the light is white, the rings are rainbow colored.

Diffraction patterns for the three cases of open Fresnel zones and appropriate distribution of light intensity along the diameter of the diffraction pattern are shown in Fig. 3.7.

If the hole $r_0$ opens the first $(m = 1)$ Fresnel zone or its part, a bright spot appears in the point *P* of the screen; the spot intensity is maximal in the center of the screen.

It gradually decreases from the center, alternating light and dark rings do not occur in this case.

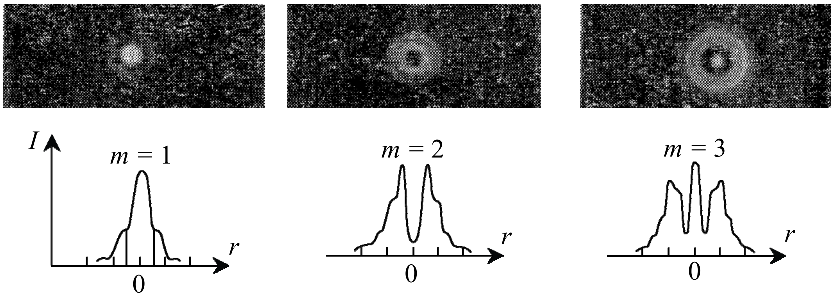



Fig. 3.7

If the hole opens the first two Fresnel zones ($m = 2$), a dark spot appears in the center of the screen; a bright ring appears around the dark spot. If $m = 3$, on the contrary, there is a bright spot in the center, and a dark ring around it. The increase of the number $m$ of open Fresnel zones leads to the increase of the number of the light and dark rings. If the number $m$ is odd, the center is bright. If the number $m$ is even, the center is dark. If the hole includes a large number of Fresnel zones, the intensity in the center becomes almost uniform and just on the edge near the geometrical shadow can be seen the narrow alternating light and dark rings.

**Diffraction by round disk.** Let us place an opaque obstacle in the form of a small circular disk of radius $r_0$ perpendicular to the direction of propagation of the spherical waves (Fig. 3.8).

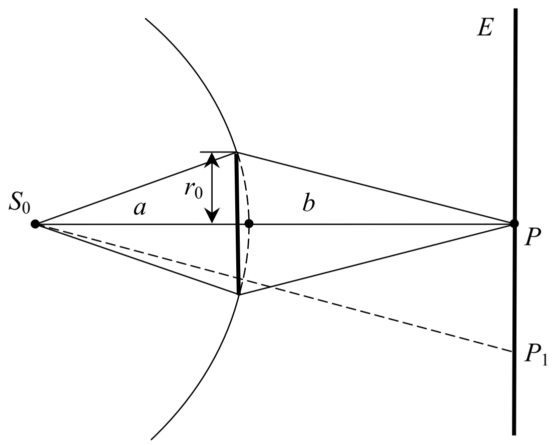



Fig. 3.8

The diffraction pattern on the screen is determined by the number of Fresnel zones that remain open because the disc covers the first Fresnel zones. If the disc closes *m* the first Fresnel zones, we can use the equation (3.8) to determine the amplitude of the light wave in the screen central point *P*, but instead of the amplitude of the first zone $A_1$, we have to substitute the amplitude of the first open zone $A_{m+1}$; instead of $A_m$ we have to substitute the amplitude of the last open zone $A_n$:

$$A = \frac{A_{m+1}}{2} \pm \frac{A_n}{2}.$$

If the value of *m* is small and $n \to \infty$, the amplitude of the last open zone tends to zero $A_n \to 0$; therefore,

$$A = \frac{A_{m+1}}{2}.$$

Thus, in contrast to diffraction by round hole, if we observe diffraction by round disk, diffraction maximum (bright spot) is always observed in the center of the diffraction pattern. Its amplitude equals half of the first open Fresnel zone amplitude.

Light intensity in an arbitrary point (Fig. 3.8) of the geometric shadow of the disc is much more difficult for calculation, but it is clear that the diffraction pattern is axially symmetric around the point *P*. Accurate calculation and the experiment show that the diffraction pattern consists of the light and dark rings that change smoothly into each other. The number of rings is determined by the number of closed Fresnel zones. Diffraction patterns created by the discs that cover one, two, and three Fresnel zones are shown on Fig. 3.9. In all cases, a light spot is in the center of the diffraction pattern.

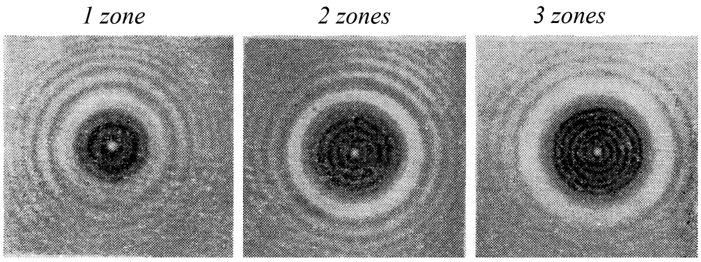



Fig. 3.9

If a disc covers only a small part of the first Fresnel zone, it does not create any shadows; light intensity on the screen remains the same as in the case without obstacles. If a disc radius is big and the disc covers a large number of Fresnel zones ($A_{m+1} << A_1$), a geometrical shadow is observed behind the disc.

There is an interesting historical fact. In 1918, Fresnel put forward his theory of diffraction for the award of French Académie. A member of the premiums Poisson (a supporter of the corpuscular theory of light) had proved on the basis of Fresnel theory that a bright spot has to be observed in the geometric center of the shadow of a small disk; the spot was called Poisson's spot. But the first experiments did not confirm the Poisson prediction. On this basis, Poisson concluded that the Fresnel theory is false. However, another member of the committee, Arago prepared more detailed experiment and proved the existence of the light spot in the center of the diffraction pattern. That was a victory of the wave theory of light.

**Diffraction by straight edge of a half plane**. Using the principle of Huygens-Fresnel, spherical diffraction by round hole and a disc was studied; axial symmetry prompted the choice of shape of division of the wave surface areas in the form of the circular Fresnel zones. Plane obstacles require division of the open part of the wave surface by Fresnel zones in the form of *straight strips* parallel to the edge of the plane.

Let a plane monochromatic wave of length $\lambda$ fall on a plane opaque obstacle with the straight edge. The obstacle $B$ is installed perpendicular to the direction of the wave propagation. The screen $E$ is placed at the distance $l$ behind the obstacle parallel to the plane. Then, the half plane $B$ is a part of the plane wave surface $S$, shown by dotted line (Fig. 3.10).

Using the vector diagram, determine the oscillations amplitude at the point $P_0$ on the geometric shadow edge. We divide the open part of the wave surface $S$ into zones as narrow strips. Let us select the width of the zones $d_1$, $d_2$, $d_3$, … so that the path difference from the edges of the adjacent zones is $\Delta = \lambda/2$ (Fig. 3.10, *a*). Oscillations in the point $P_0$ created by waves from adjacent zones will differ in phase by $\pi$. The amplitudes of the oscillations of the respective zones are determined by areas of these zones, hence, by the dimensions of $d_1$, $d_2$, $d_3$, … . We can deduce from Fig. 3.10 that the total width of the first $m$ zones is:

$$d_1 + d_2 + ... + d_m = \sqrt{(l + m\lambda/2)^2 - l^2} = \sqrt{lm\lambda + m^2(\lambda/2)^2}\ .$$

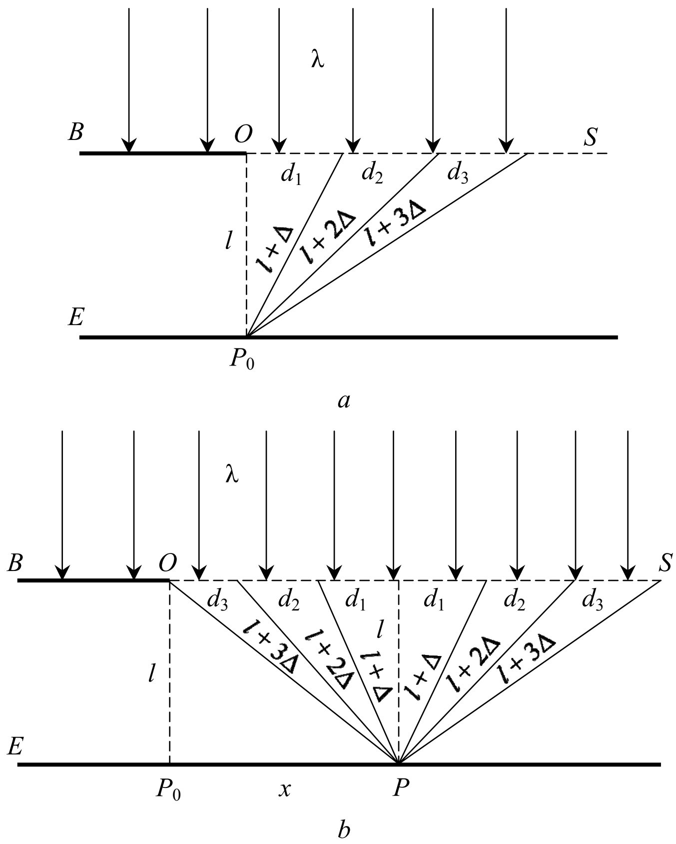


Fig. 3.10

Since the zones are narrow, $\lambda/2 << l$; therefore, for the small $m$ the second term under the root can be neglected. Then

$$d_1 + d_2 + ... + d_m = \sqrt{lm\lambda}\,. \tag{3.11}$$

The equation (3.11) shows that $d_1 = \sqrt{l\lambda}$, so $d_1 + d_2 + ... + d_m = d_1\sqrt{m}$, then

$$d_m = d_1\left(\sqrt{m} - \sqrt{m-1}\right). \tag{3.12}$$

The calculation of $d_m$ by the equation (3.12) gives:

$$d_1 : d_2 : d_3 : d_4 : d_5 :... = 1 : 0.41 : 0.32 : 0.27 : 0.23 :... . \tag{3.13}$$

Obviously, the areas of the zones as narrow straight strips are in the same ratio.

According to the series (3.13), the areas and, hence, the amplitudes of waves generated from the respective zones in the point $P_0$ will initially (for the first zones) rapidly decrease, then this decrease becomes very slow. For this reason, a broken line formed by the oscillations graphical adding is initially flat and then goes into a spiral.

Fig. 3.11 shows comparison of two vector diagrams for the case of circular zones (zones of approximately equal areas) and straight lines (areas decreasing); we see that the amplitudes for circular areas are about the same but the amplitudes for the straight strips are reduced according to the proportion (3.13). In both cases, the lag phase for each subsequent vector is the same.

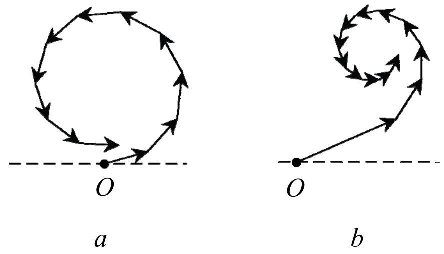



Fig. 3.11

If the width of the zones goes to zero, the broken line on Fig. 3.11 turns into a smooth curve, which is the right half of a Cornu spiral (Fig. 3.12).

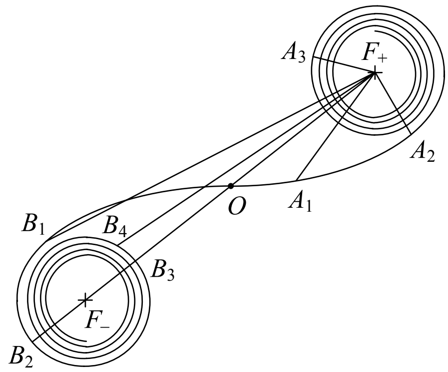
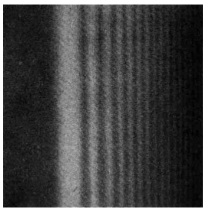



Fig. 3.12

Along with the right part, the Cornu spiral has the left side symmetric relatively to the point *O*. This part corresponds to the action of oscillations coming from the same points of the zones located to the left from the point *O* that means the absence of half-plane *B* (Fig. 3.10). The amplitude in the point $P_0$ from the completely open wave surface is a segment $F_-F_+$ that connects two points (spiral focuses); the spiral approaches asymptotically to the focuses. If the half-plane *B* is present (see Fig. 3.10, *a*), the light wave amplitude in the point $P_0$ on the geometric shadow boundary is a segment $OF_+$ that is twice smaller than the segment $F_-F_+$ (Fig. 3.12). This means that the light intensity in the point $P_0$ equals ¼ of the intensity if the wave surface is completely open $I(P_0) = 0,25I_0$. The value $I_0$ corresponds to the intensity if the wave surface is completely open (absence of the half-plane). The diffraction pattern created by the half-plane is shown on Fig. 3.13.



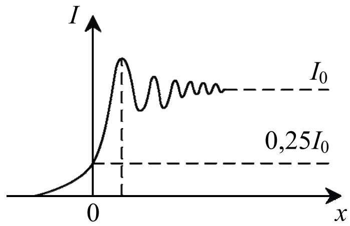



Fig. 3.13

**Diffraction by narrow slit.** Let us consider diffraction by narrow slit (Fig. 3.14) using the Cornu spiral

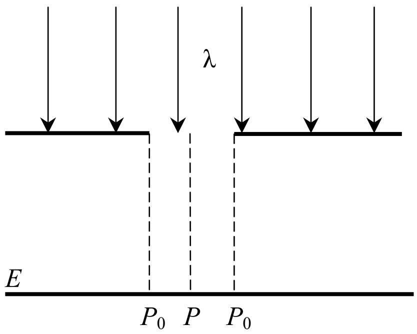



Fig. 3.14

The diffraction pattern created by a narrow slit is a result of superposition of diffraction patterns from the two half-planes. The resulting amplitude $\vec{A}$ in the point $P$ equals a distance between the symmetrical points of the spiral (Fig. 3.15).

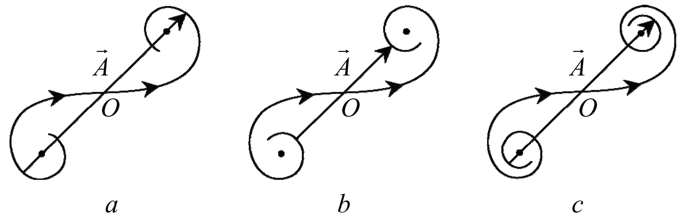



Fig. 3.15

If the slit width increases, the amplitude $\vec{A}$ in the point $P$ changes from maximums (Fig. 3.15, *a*, *c*) to minimums (Fig. 3.15, *b*) and vice versa.

If the slit width is big, the diffraction patterns are observed only near the slit edges; the light intensity $I_0$ in the center is constant (Fig. 3.16).

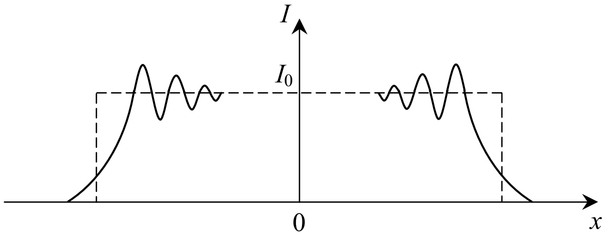



Fig. 3.16

### 3.5. Fraunhofer Diffraction

The examples of Fresnel diffraction are observed without any optical devices. A screen is placed at a short distance from the obstacle; therefore, the diffraction patterns are produced by divergent secondary spherical waves. This diffraction was scrupulously researched by Fresnel; thus, it is called Fresnel diffraction or diffraction at divergent rays.

Fraunhofer proposed another way of diffraction observing. A direct beam of parallel rays (plane wave) is directed to the opening or the slit, and the diffraction pattern is observed on the screen placed far from the obstacles. Therefore, it can be considered that the screen is placed at infinity, and the diffraction pattern is observed in parallel rays. As a result, the diffracted wave is also flat. Since the incident and diffracted rays are parallel, this is called diffraction in parallel rays or Fraunhofer diffraction. In practice, the screen is not placed at a long distance; instead, the diffraction pattern is watching using a lens or a telescope fixed at infinity.

Thus, in the case of Fraunhofer diffraction, two conditions have to be realized: incident and diffracted waves are flat. If at least one of two conditions is not occurred, we have the case of Fresnel diffraction.

A device for Fraunhofer diffraction observations is shown on Fig. 3.17. A point source is placed in the lens main focal plane $L_1$.

After the lens, a parallel beam goes to an opaque obstacle *B* with an opening. The diffraction pattern is observed in the focal plane of the second lens $L_2$.

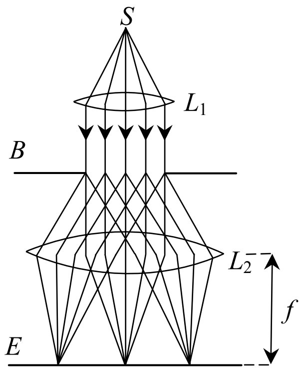


Fig. 3.17

**The criterion that characterizes the type of diffraction.** Although there is no fundamental physical difference between Fraunhofer and Fresnel diffraction, but there are geometrical conditions under which it is possible to observe a particular type of diffraction.

Set the approximate quantitative criterion that allows determining the type of diffraction. For this purpose, we use the formula (3.7a)

$$r_m = \sqrt{m\lambda b}.$$

This formula describes the case when a flat wave ($a \to \infty$) normally falls to the opening of radius $r_m$; *m* is the number of Fresnel zones that are fitted in the opening for the observation of the point *P*, distanced from the opening at a distance *b*. Using (3.7a), we get $m = r_m^2 / \lambda b$. Since the nature of the diffraction pattern is determined exclusively by the number of the open Fresnel zones, the last equation can be taken as a criterion *p* for determining the diffraction type. Substituting $r_m$ for a characteristic size *d* of the opening, and *b* for *l*, we obtain

$$p = d^2 / l\lambda ,$$

where *d* is the radius of the opening or disc or, for example, the width of the slit and so on.

The value of this dimensionless parameter determines the nature of the diffraction:

if $p \sim 1$, this is Fresnel diffraction;

if $p >> 1$, this is approximation of geometrical optics;

if $p << 1$, this is Fraunhofer diffraction.

In fact, if $p \sim 1$ (a small number of Fresnel zones is open), we have a classic case of Fresnel diffraction when minima and maxima of intensity are observed on the screen. Depending on the number of the open Fresnel zones, it can be either a maximum or a minimum in the center of the diffraction pattern.

The case $p >> 1$ corresponds to a large number of open Fresnel zones. Under this condition, the central part of the screen is equally lit and a microscopic diffraction pattern is observed only on the boundary of the geometrical shadow. This case is classified as an approximation of geometrical optics when you can ignore the phenomenon of diffraction and use the ordinary laws of geometrical optics. The criterion for the application of geometrical optics is not just the smallness of the wavelength compared with the size of the obstacles (eg, slit width), and the value *p* that should be $p >> 1$. For example, $d / \lambda = 1000$ ($\lambda << d$) i $l / d = 1000$. But, $p = d^2 / l\lambda = 1$. So, we have a perfect Fresnel diffraction in this case.

**Fraunhofer diffraction by slit.** Let us consider the example of Fraunhofer diffraction of a plane monochromatic wave passing through a narrow infinitely long slit $OO'$ of the width *b* in the opaque obstacle *B*. The incident plane wave front, the slit, and the screen are parallel to each other (Fig. 3.18). Fraunhofer Diffraction by narrow slit is a system of interference maxima (blur images of the light source) separated by dark interference minima.

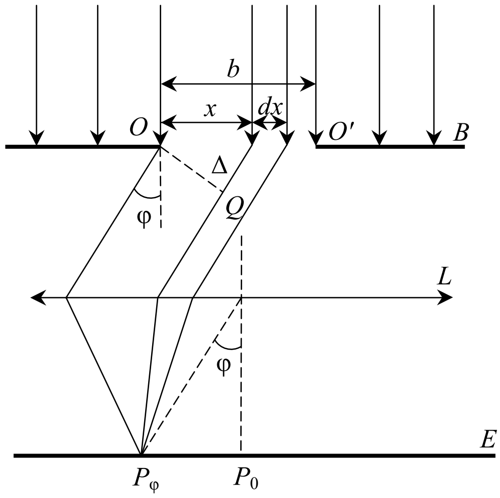



Fig. 3.18

Let us divide the open part of the wave front into an infinite number of parallel to the slit edges elementary zones of the width $dx$. One of these zones is shown at a distance $x$ from the slit left edge (point $O$). Secondary waves propagated from the zones at the angle φ are collected by the lens $L$ at the point $P_\varphi$. The lens collects flat (non-spherical) waves in the focal plane, so factor $1/r$ in the equation (3.1) is absent in the case of Fraunhofer diffraction. If the angles φ are not very large, the factor $K(\varphi)$ in (3.1) is constant. Then the amplitude depends only on the area of the zone. The area is proportional to $dx$; therefore, the amplitude $dA = Cdx$, where $C$ is a constant. The oscillations from all elementary zones of the slit come to the point $P_0$ with the same phases and mutually reinforce each other:

$$A_0 = \int dA = \int_0^b Cdx = Cb, \text{ hence } C = \frac{A_0}{b} \text{ i } dA = \frac{A_0}{b} dx .$$

To determine the resultant amplitude in any point $P_\varphi$ on the screen, it is necessary to take into account the phase difference between the waves coming from the different zones of the slit to the observation point. If the phase of oscillations produced by the elementary zone located near the left edge of the slit (point *A*) is taken as zero, the phase of oscillations from the zone $dx$ (Fig. 3.18) is

$$\delta = k \cdot \Delta = \frac{2\pi}{\lambda} x \sin\varphi,$$

Where $k = 2\pi/\lambda$ is the wavenumber; $\Delta = x\sin\varphi$ is the optical path difference.

Thus, the oscillations caused by the elementary zone of coordinate $x$ at the point $P_\varphi$ (its position is determined by the angle of diffraction $\varphi$) can be represented as

$$dE_\varphi = \frac{A_0 dx}{b}\cos\left(\omega t - \frac{2\pi}{\lambda} x \sin\varphi\right),$$

which is the real part of the equation:

$$dE_\varphi = \frac{A_0 dx}{b} e^{i\left(\omega t - \frac{2\pi}{\lambda} x \sin\varphi\right)},$$

where the initial phase can be regarded as equal to zero.

Oscillations of all the elementary zones are coherent (because they are parts of a plane wave surface).

So, finding the resultant amplitude in an arbitrary point comes down to the interference problem or the summation of oscillations. This is expressed by an integral over the entire width of the slit for all values of $x$ from zero to $b$:

$$E_\varphi = \int_0^b \frac{A_0}{b} e^{i\left(\omega t - \frac{2\pi}{\lambda} x \sin\varphi\right)} dx .$$

Let us put before the integral sign factors that do not depend on $x$, and take $\gamma = \frac{\pi}{\lambda}\sin\varphi$, then

$$E_\varphi = \frac{A_0}{b} e^{i\omega t} \int_0^b e^{-2i\gamma x} dx = \frac{A_0}{2i\gamma b} e^{i\omega t}(1 - e^{-2i\gamma b}),$$

or

$$E_\varphi = \frac{A_0}{\gamma b} e^{i\omega t} e^{-i\gamma b}\left(\frac{e^{i\gamma b} - e^{-i\gamma b}}{2i}\right) = \frac{A_0 \sin(\gamma b)}{\gamma b} e^{i(\omega t - \gamma b)}. \quad (3.14)$$

The real part of (3.14) expresses the resulting oscillations of the light wave in the point $P_\varphi$

$$E_\varphi = \frac{A_0 \sin(\gamma b)}{\gamma b} \cos(\omega t - \gamma b),$$

which amplitude is

$$A_\varphi = \frac{A_0 \sin(\gamma b)}{\gamma b} = A_0 \frac{\sin(\pi b \sin\varphi / \lambda)}{\pi b \sin\varphi / \lambda}.$$

Light intensity is proportional to the square of the amplitude, therefore,

$$I_\varphi = I_0 \frac{\sin^2(\pi b \sin\varphi / \lambda)}{(\pi b \sin\varphi / \lambda)^2} = I_0 \frac{\sin^2\alpha}{\alpha^2}, \quad (3.15)$$

Where $\alpha = \pi b \sin\varphi / \lambda$; $I_0$ is the intensity in the center of the diffraction pattern. Using (3.15) we see that $I_\varphi = I_{-\varphi}$. This means that the diffraction pattern is symmetrical relatively to the center of the lens. If the slit is displaced parallel to the screen (along the *x*, Fig. 3.18), the diffraction pattern on the screen remains fixed (its middle lies across the center of the lens). Instead, the shift of the lens at the fixed slit is accompanied by the same shift of the diffraction pattern on the screen.

**The intensity distribution.** The intensity distribution according to formula (3.15) is shown on Fig. 3.19. Let us analyze it.

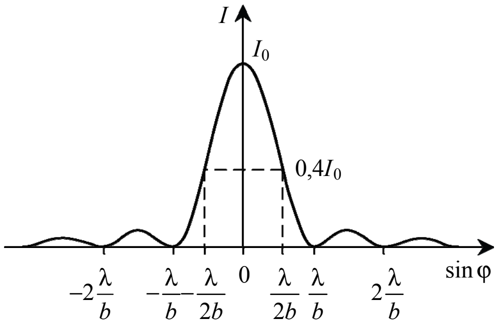



Fig. 3.19

It is known that the limit $\lim_{\alpha \to 0} \frac{\sin\alpha}{\alpha}$ goes to one. Then, from (3.15) it follows that the main central maximum is in the center of the diffraction pattern. Obviously, the condition for the existence of the main peak does not depend on $\lambda$, and its position remains the same for all wavelengths. Thus, in case of Fraunhofer diffraction, maximum is always observed in the center of the diffraction pattern, which has the form of the light and dark alternating bands. Remember that in Fresnel diffraction, central diffraction band can be both bright and dark.

Positions of diffraction maxima and minima of each number except of the first depend on the wavelength. Therefore, if a slit is illuminated by a white light, the central maximum position remains unchanged but it will have a rainbow color around the edges. Minima of the light will not be observed in any point of the screen, as both the maxima and minima of light with different $\lambda$ overlap.

Equation (3.15) shows that the minimum position is defined by

$$\pi b \sin\varphi / \lambda = \pm n\pi \Rightarrow \sin\varphi = \pm n\lambda / b \quad (n = 1, 2, 3, ...), \qquad (3.16)$$

where $n$ is the order of diffraction, $n \neq 0$ while the central maximum is formed for $n = 0$; $\Delta = b \sin\varphi$ is the path difference between the beams that spread from the edges of the slit (Fig. 3.18).

Equation (3.16) shows that the decrease of the slit width $b$ is accompanied by expansion of the diffraction pattern.

Between these minima, the *secondary* maxima are placed (Fig. 3.19). The angles of diffraction maxima can be found through graphic solving of the transcendental equation $\mathrm{tg}\alpha = \alpha$, where $\alpha = \pi b \sin\varphi / \lambda$. These equations are based on the conditions for an extremum of (3.15), i.e. the extremum of $\sin\alpha/\alpha$. It is enough to take the derivative of $\sin\alpha/\alpha$ and equate it to zero. Graphical solution of transcendent equation:

$$\alpha_1 = \pi b \sin\varphi_1 / \lambda = \pm 1{,}43\pi; \quad \alpha_2 = \pi b \sin\varphi_2 / \lambda = \pm 2{,}46\pi;$$

$$\alpha_3 = \pi b \sin\varphi_3 / \lambda = \pm 3{,}47\pi; \quad \alpha_4 = \pi b \sin\varphi_4 / \lambda = \pm 4{,}48\pi \text{ etc.}$$

Substituting them into equation (3.15) and taking into account that for $\alpha_0 = 0$ the intensity equals $I_0$, we get:

$$I_0 : I_1 : I_2 : I_3 : ... = 1 : 0.045 : 0.016 : 0.008 : ....$$

It shows that the intensity of the secondary maxima rapidly decrease, in particular, the intensity of the first peak does not exceed 5 % of the intensity of the central peak. This means that most of the light flux that has passed through the slit is concentrated in the central maximum.

**The method of Fresnel zones.** Considered calculation of the diffraction pattern is mathematically accurate. Let us show that using the less accurate method of Fresnel zones leads to the same results, but by much simpler way.

Division of the slit on the Fresnel zones of width $\lambda/(2\sin\varphi)$ provides that the optical path difference from the edges of each Fresnel zone is $\Delta = \lambda/2$. All zones emit exactly identical waves. So, interference from each pair of adjacent zones gives zero resulting oscillation amplitude, since these zones are the sources of oscillations with the same amplitude but with the opposite phases.

Thus, the result of interference in a point is determined by how many Fresnel zones are fitted into the slit.

If the number of zones is even:

$$b\sin\varphi = \pm 2n\frac{\lambda}{2}, \ (n = 1,\ 2,\ 3,\ldots),$$

diffraction minimum is observed. If the number of zones is odd:

$$b\sin\varphi = \pm(2n+1)\frac{\lambda}{2} \quad (n = 1,\ 2,\ 3,\ldots), \tag{3.17}$$

diffraction maximum is observed due to a non-compensated Fresnel zone.

Using the formula for the maxima (3.17), we can define the variable $\alpha = \pi b\sin\varphi/\lambda = \pm\pi(n+1/2)$:

$$\alpha_1 = \pm 1.5\pi,\ \alpha_2 = \pm 2.5\pi,\ \alpha_3 = \pm 3.5\pi,\ \alpha_4 = \pm 4.5\pi \text{ etc.}$$

Comparing these values with previous ones, we can see that the difference between them is small; so, in practice the position of the maxima is convenient to calculate by the formula (3.17).

**The method of graphical addition of amplitudes.** Let us divide the slit (the open part of the wave surface) into $N$ zones of equal width. In general, the amplitude $\Delta A$ of each zone depends on the coefficient $K(\varphi)$.

However, we can neglect this dependence for the small angles and assume that the oscillations of each zone have the same amplitude. Different values of the phase difference $\delta = k\Delta = 2\pi b\sin\varphi/\lambda$ give different curvature of the broken line.

Hence, using the graphical images we can get a chain of equal in magnitude vectors $\Delta\vec{A}$ and they are rotated relatively to each other at the same angle; the resultant amplitude $\vec{A}$ is the sum of vectors $\Delta\vec{A}$.

For $\varphi = 0$, the phase difference $\delta$ is zero and the vector diagram has the form shown on Fig. 3.20, *a*. The amplitude of the resulting oscillations is equal to the sum of amplitudes $N\Delta A = A_0$.

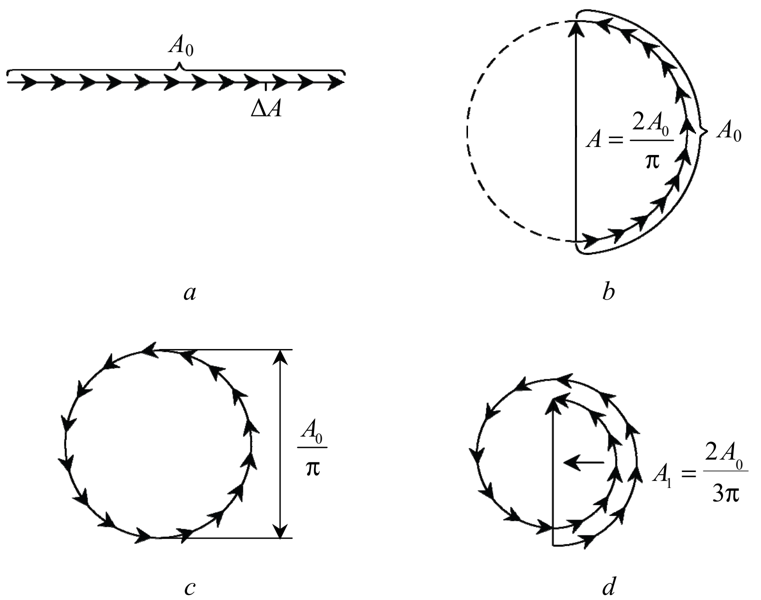



Fig. 3.20

If $\delta = \pi$ ($\Delta = b\sin\varphi = \lambda/2$), the oscillations from the edges of the slit are antiphase. The vectors $\Delta\vec{A}$ form a semi-circle of the length $A_0$ (Fig. 3.20, *b*). Thus, the resulting amplitude is equal to $2A_0/\pi$. The intensity is proportional to the amplitude squared; so, the intensity at these points is $4I_0/\pi^2 \approx 0.4I_0$ (see Fig. 3.19).

If $\delta = 2\pi$ ($\Delta = b\sin\varphi = \lambda$), the broken line is being closed forming a circle of the length $A_0$ and diameter $A_0/\pi n$ (Fig. 3.20, *c*).

The first maximum is observed if $\delta = 3\pi$ ($\Delta = b\sin\varphi = 3\lambda/2$); then, $A_1 \approx 2A_0/3\pi$, (Fig. 3.20, *d*).

The intensity of this peak is $I_1 = 4I_0 / 9\pi^2 \approx 0.045 I_0$.

Similarly, we can find the relative intensity of other secondary maxima. As a result, we obtain the known relation:

$$I_0 : I_1 : I_2 : I_3 : ... = 1 : \left(\frac{4}{9\pi^2}\right) : \left(\frac{4}{25\pi^2}\right) : \left(\frac{4}{49\pi^2}\right) : ... =$$

$$= 1 : 0.045 : 0.016 : 0.008 : ...$$

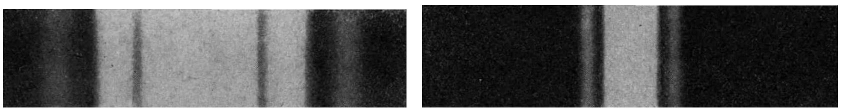
**Width of the slit.** Let us consider the influence of the slit width onto the Fraunhofer diffraction pattern. Fig. 3.19 shows that the edges of the central peak fall on the first minima, which correspond to the values of $\sin\varphi = \pm\lambda/b$. So, the angular width of the central maximum on the screen is:

$$\Delta\varphi = 2\arcsin(\lambda/b) . \qquad (3.18)$$

It shows that with the decreasing size of the slit $b$ the angular width of the central maximum increases. This means that the central maximum (all secondary maxima as well) expands. If $b = \lambda$, the angular width of the central peak is $\Delta\varphi = \pi$ (the central maximum extends to infinity), and there are no minimums.

If $b/\lambda$ is small (very narrow slit), the diffraction peaks are broad, and the whole picture is fuzzy.

With the further increase in the slit size, the diffraction pattern becomes clearer with the brighter and sharper peaks. Diffraction patterns for the narrow and wide slits are shown on Fig. 3.21, *a*, *b*.



*a* *б*

Fig. 3.21

Finally, if $b/\lambda$ is large ($b >> \lambda$), the central maximum becomes very narrow and bright.

It is a screen image formed by the lens according to the laws of geometrical optics. In this case, the value of $\sin\varphi = \pm\lambda/b$ can be replaced by $\varphi \approx \pm\lambda/b$, then, formula (3.18) for the angular width of the central maximum is simplified to the form:

$$\Delta\varphi \approx 2\lambda/b .$$

### 3.6. Diffraction grating

The diffraction grating is an optical instrument designed for the decomposition of light into a spectrum and wavelengths measuring. The simplest one-dimensional grating is a collection of parallel pitches drawn on a glass plate and placed at equal distances from each other; the transparent parts between pitches act as slits. This grating is called the amplitude grating; it has different transparency in different places that changes the amplitude (intensity) of the transmitted light.

The diffraction grating is shown on Fig. 3.22. The slit width is $b$, the width of the opaque barrier between the slits is $a$. The value $d = a + b$ is called a *period or diffraction grating constant*. The collecting lens $L$ is placed parallel to the grating; the screen $E$ is placed in the lens focal plane. The light source is a brightly illuminated slit placed in the focal plane of another lens placed before the diffraction grating. Under this condition, the diffraction pattern is observed in parallel rays (Fraunhofer diffraction).

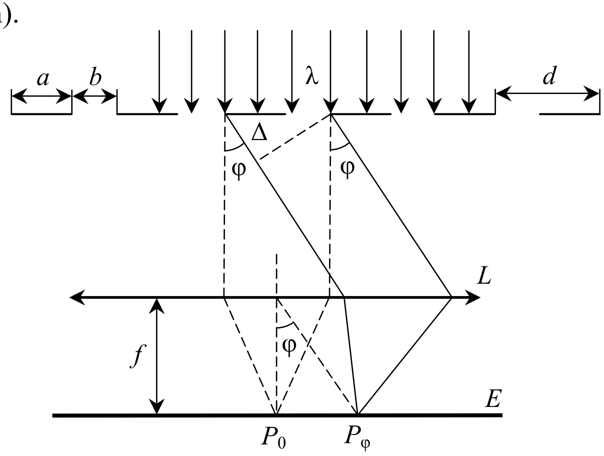


Fig. 3.22

Each slit separately gives a diffraction pattern shown on Fig. 3.19. However, oscillations from all slits are coherent as they all belong to the same wave surface. Hence, the rays from different slits interfere in all points of the screen. So, the resulting oscillation in a point $P_\varphi$ is the vector sum of $N$ amplitudes ($N$ is the number of slits in a diffraction grating); the amplitudes $A_\varphi$ are the same and the phases differ by the same value:

$$\delta = k\Delta = \frac{2\pi}{\lambda} d \sin\varphi , \tag{3.19}$$

where $\Delta = d\sin\varphi$ is the optical path difference between two rays from the adjacent slits (Fig. 3.22).

Oscillations of light waves from the first slit in the point $P_\varphi$ have the form $E_1 = A_\varphi e^{i\omega t}$. Then, the phases of oscillations from the other slits lag behind (or outstrip) by $\delta$ and the other oscillations can be described by the equations:

$$E_2 = A_\varphi e^{i(\omega t - \delta)} = E_1 e^{-i\delta},\ E_3 = E_1 e^{-2i\delta},\ \ldots\ E_N = E_1 e^{-(N-1)i\delta},\ldots .$$

The sum of these oscillations is the resulting oscillation in the direction $\varphi$ relatively to the diffraction grating:

$$(E_\varphi)_{\text{gr}} = E_1[1 + e^{-i\delta} + e^{-2i\delta} + \ldots + e^{-(N-1)i\delta}] .$$

The equation in brackets is the sum of $N$ exponentially terms with the first term, which is equal to one. Using the formula for the sum of geometric progression terms, we obtain

$$(E_\varphi)_{\text{gr}} = E_1 \frac{1 - e^{-iN\delta}}{1 - e^{-i\delta}} = A_\varphi \frac{1 - e^{-iN\delta}}{1 - e^{-i\delta}} e^{i\omega t} . \tag{3.20}$$

Square of the amplitude is the intensity of the light wave in the focal plane of the lens. Multiplying the left and the right sides of (3.20) by the complex conjugate values, we obtain:

$$(I_\varphi)_{\text{gr}} = (E_\varphi)_{\text{gr}} (E_\varphi)^*_{\text{gr}} = A_\varphi^2 \frac{(1 - e^{-iN\delta})(1 - e^{iN\delta})}{(1 - e^{-i\delta})(1 - e^{i\delta})} e^{i\omega t} e^{-i\omega t} ,$$

where

$$(I_\varphi)_{\text{гр}} = A_\varphi^2 \frac{2 - (e^{iN\delta} - e^{-iN\delta})}{2 - (e^{i\delta} - e^{-i\delta})} = A_\varphi^2 \frac{1 - \cos N\delta}{1 - \cos\delta} = A_\varphi^2 \frac{\sin^2(N\delta/2)}{\sin^2(\delta/2)}$$

[we have used the well-known transformation formula: $e^{ix} - e^{-ix} = 2\cos x$, $1 - \cos x = 2\sin^2(x/2)$]. Substituting the values $I_\varphi = A_\varphi^2$ from one slit from the equation (3.15):

$$(I_\varphi)_{\text{gr}} = I_0 \frac{\sin^2\alpha}{\alpha^2} \cdot \frac{\sin^2(N\delta/2)}{\sin^2(\delta/2)} , \tag{3.21}$$

or taking into account formulas (3.19) and $\alpha = \pi b \sin\varphi/\lambda$

$$(I_\varphi)_{\text{gr}} = I_0 \frac{\sin^2(\pi b \sin\varphi/\lambda)}{(\pi b \sin\varphi/\lambda)^2} \cdot \frac{\sin^2(N\pi d \sin\varphi/\lambda)}{\sin^2(\pi d \sin\varphi/\lambda)}, \qquad (3.22)$$

where $I_0 = A_0^2$ is the intensity created by one slit across the center of the lens at the point $P_0$.

The first factor in (3.22) is responsible for the diffraction intensity distribution from each slit. It is zero at the points for which

$$b \sin\varphi = \pm n\lambda \quad (n = 1, 2, 3, ...). \qquad (3.23)$$

At these points, the intensity of the wave created by each individual slit is zero [see condition (3.16)]. Since all of the diffraction grating slits are identical, the minimum condition (3.23) is true for all the other slits and for the whole grating.

The second factor in equation (3.22), which is responsible for the interference of waves from the slits, gives the value $N^2$ in the points that satisfy the condition:

$$d \sin\varphi = \pm m\lambda, \quad (m = 0, 1, 2, 3, ...). \qquad (3.24)$$

The equation (3.24) represents *condition of the main maximum* of the diffraction grating; *m* is the order of the main maximum. As it will be shown below, the greater the diffraction grating numbers of slits *N*, the narrower and sharper the maxima.

In the intervals between adjacent maxima, the additional minima arise for the areas where the oscillations of the slits mutually extinguish each other. Interference from slits is described by the second factor in (3.21); so, the additional interference minima appear when $\sin(N\delta/2) = 0$, however $\sin(\delta/2) \neq 0$. That is $N\pi d \sin\varphi/\lambda = h\pi$ ($\delta = 2h\pi/N$), and

$$d \sin\varphi = \pm \frac{h}{N}\lambda \qquad (3.25)$$

($h = 1, 2, ..., N-1, N+1, ..., 2N-1, 2N+1, ...$).

The value $h$ in the formula (3.25) takes all integer values except of 0, *N*, 2*N* ... If $h = 0, N, 2N, ...$ the ratio $h/N$ is an integer ($m = h/N$) and minimum additional condition (3.25) becomes the condition of maximum (3.24). The function (3.22) graph for $N = 4$ and $b/d = 1/3$ is shown on Fig. 3.23.

The dotted curve that goes around the top of the main maxima represents the intensity of light from one slit multiplied by $N^2$. It is seen that three additional minima and two secondary maxima are located between every two main maxima. If $b/d = 1/3$, the main maxima of the 3rd, 6th, etc. orders fall on minimum intensity from one slit (their positions are given in brackets), hence, these peaks disappear.

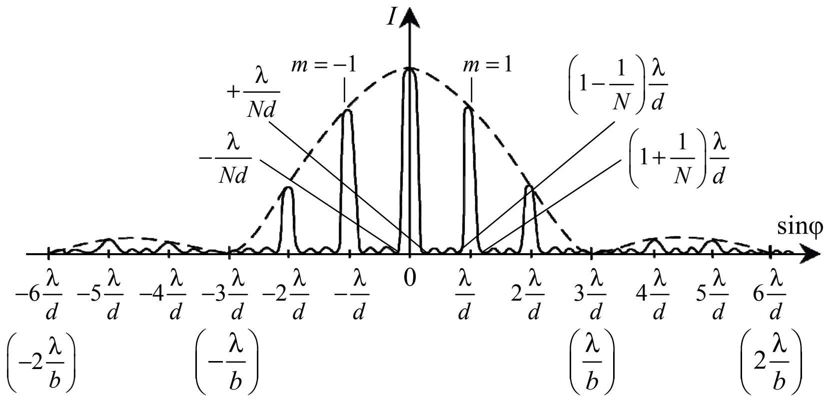



Fig. 3.23

Indeed, if the condition (3.24) takes place, the formula (3.22) takes the form

$$I_m = I_0 \frac{\sin^2(\pi b \sin\varphi/\lambda)}{(\pi b \sin\varphi/\lambda)^2} N^2 .$$

If $\sin\varphi = m\lambda/d$, we obtain

$$I_m = \frac{I_0 N^2 d^2}{\pi^2 m^2 b^2} \sin^2 \frac{\pi m b}{d} . \tag{3.26}$$

We have two conclusions from the equation (3.26). First, with the increase of the order of diffraction $m$ the intensity of the corresponding maximum sharply decreases inversely proportional to the square of the order of diffraction $I_m \sim 1/m^2$. Second, the light intensity of the $m$-th maximum depends on the relationship $b/d$ and when $mb/d$ is an integer, $\sin(\pi md/d) = 0$ and the intensity of the corresponding main maximum becomes zero $I_m = 0$.

The number of main maxima is determined by the ratio of the grating period $d$ and the wavelength $\lambda$. Since the module of $\sin\varphi$ cannot be greater than one, then the formula (3.24) gives

$$m \le d/\lambda . \qquad (3.27)$$

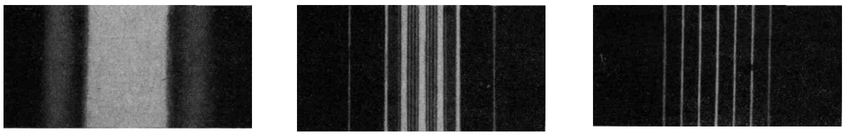
Let us estimate the angular width of the main maxima. The equation (3.25) shows that the main maximum of $m$-th order occurs when $h = mN$. So, the next adjacent minima occur if $h = mN - 1$ and $h = mN + 1$. According to the equation (3.25) $\sin\varphi'_m = \frac{mN+1}{Nd}\lambda$, $\sin\varphi''_m = \frac{mN-1}{Nd}\lambda$, and the difference of the sines of the angles is:

$$\sin\varphi'_m - \sin\varphi''_m = 2\cos\frac{\varphi'_m + \varphi''_m}{2}\sin\frac{\varphi'_m - \varphi''_m}{2} = \frac{2\lambda}{Nd} .$$

For the large values of $N$, the difference $\varphi'_m - \varphi''_m$ is small; hence, $\sin\frac{\varphi'_m - \varphi''_m}{2} \approx \frac{\delta\varphi_m}{2}$ and $\varphi'_m + \varphi''_m \approx 2\varphi_m$. Then, taking into account the formula (3.24) we obtain:

$$\delta\varphi_m = \frac{2\lambda}{Nd\cos\varphi_m} = \frac{2\lambda}{Nd\sqrt{1-\sin\varphi_m^2}} = \frac{2\lambda}{Nd\sqrt{1-m^2\lambda^2/d^2}} . \qquad (3.28)$$

As it can be seen from the equation (3.28), the angular width of the maxima is directly proportional to the wavelength $\lambda$ and inversely proportional to the total width of the diffraction grating $Nd$. The greater the number of the diffraction grating slits $N$, the narrower and sharper the main maxima. Diffraction patterns, which are formed respectively by one, five, and twenty slots, are shown on Fig. 3.24.



$N = 1$ $\qquad$ $N = 5$ $\qquad$ $N = 20$

Fig. 3.24

If the sun (white) light is falling to the diffraction grating, maxima for different wavelengths do not coincide with each other [except the central maximum ($m = 0$)].

Therefore, the central maximum is white, and the other maxima look like colored stripes, that are spectra of the first, second, third, etc. orders. The color of each strip changes from violet on the inner edge (the closest to the zero-order maximum) to red on the outer edge.

The property of diffraction grating is widely used to study the spectral composition of light, determination of wavelengths and intensities of all its monochromatic components. Measurement methods of Angstrom (1868) and especially Rowland (1888) gave possibility to create a detailed atlas of the spectra of sunlight. They were able to measure wavelengths up to the sixth decimal place.

Apparatus for the spectral composition of light study on the base of diffraction grating is called diffraction spectrograph. The main characteristics of any spectral instrument are the *angular dispersion, spectral resolution and dispersion region.*

**The angular dispersion *D*.** Using angular dispersion *D*, we can determine the degree of angular separation of different wave lengths.

$$D = \frac{\delta\varphi}{\delta\lambda},$$

where $\delta\varphi$ is the angular distance between the spectral lines that differ in the wavelength by $\delta\lambda$.

Differentiation of the equation (3.24) with the constants *m* gives $d\cos\varphi\delta\varphi = m\delta\lambda$, where

$$D = \frac{\delta\varphi}{\delta\lambda} = \frac{m}{d\cos\varphi}.$$

One can see that for some order of the spectrum *m*, the smaller the diffraction grating period *d*, the greater the angular dispersion. In addition, the angular dispersion increases with *m*, i.e. the higher the order of spectrum *m*, the greater the angular dispersion. For small angles $\cos\varphi \approx 1$, so:

$$D \approx \frac{m}{d}.$$

**Linear dispersion.** Since the diffraction lines are often observed on a screen or a photographic plate, then, it is conveniently to replace the angular distance between the lines on the linear distance $\delta l$. Obviously that $\delta l = f\delta\varphi$, so the linear dispersion is:

$$D_{\text{lin}} = \frac{\delta l}{\delta\lambda} = fD \ ,$$

or for the small angles

$$D_{\text{lin}} = f\frac{m}{d} \ .$$

Linear dispersion is measured in millimeters per angstroms.

**Spectral resolution.** Spectral resolution shows the possibility to distinguish two close spectral lines with the wavelengths $\lambda$ and $\lambda + \delta\lambda$. If the diffraction maxima are blurred, they mix into each other and you cannot distinguish them. The narrower maxima require the smaller angle between them to resolute them in the space.

Spectral resolution of the diffraction grating is a dimensionless value

$$R = \frac{\lambda}{\delta\lambda} ,$$

where $\delta\lambda$ is the smallest difference of two wavelengths of the spectral lines when these lines are observed separately.

According to the Rayleigh criterion, two spectral lines of the same intensity with the similar wavelengths $\lambda$ and $\lambda + \delta\lambda$ are resoluted if maximum of the wavelength $\lambda$ coincides with minimum for the wavelength $\lambda + \delta\lambda$ (Fig. 3.25). Under this condition, the light intensity between maxima is not more than 80 % of the maximum. This is enough to see these two maxima separately.

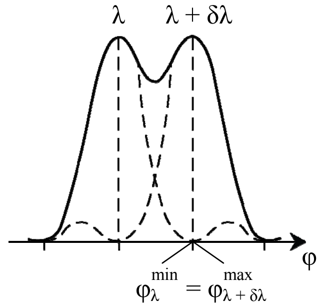


Fig. 3.25

According to (3.25), the condition of $m$-th order first minimum is $h = mN + 1$; hence, for the wavelength $\lambda$:

$$d \sin\varphi_{\lambda}^{\min} = (mN + 1)\lambda / N = (m + 1/N)\lambda \, .$$

According to the Rayleigh criterion, this minimum coincides with the $m$-th order maximum for the wavelength $\lambda + \delta\lambda$:

$$d \sin\varphi_{\lambda+\delta\lambda}^{\max} = m(\lambda + \delta\lambda) \, .$$

Since $\varphi_{\lambda}^{\min} = \varphi_{\lambda+\delta\lambda}^{\max}$, then

$$m(\lambda + \delta\lambda) = (m + 1/N)\lambda \, ,$$

where

$$R = \frac{\lambda}{\delta\lambda} = mN.$$

According to equation (3.27), the maximum order of diffraction is $m = d/\lambda$. Accordingly, the maximum value of spectral resolution is

$$R_{max} = \frac{Nd}{\lambda},$$

where the product $Nd$ is equal to the total width of the diffraction grating.

**Dispersion region**. Spectral instrument is not suitable for the study of certain areas of the spectrum if the spectra of adjacent orders overlap. Let us find the width of the spectral interval in which there is no overlap of the spectra of the adjacent orders. The long-wavelength edge of the *m*-th order spectrum coincides with the edge of the shortwave spectrum ($m$ +1)-th order if $m(\lambda + \Delta\lambda) = (m+1)\lambda$ ; so, the dispersion region is:

$$\Delta\lambda = \frac{\lambda}{m}.$$

Spectra of the first, second or third orders are usually observed. Accordingly, the dispersion regions $\Delta\lambda = \lambda$, $\Delta\lambda = \lambda/2$ or $\Delta\lambda = \lambda/3$ are sufficiently large. In particular, the first-order diffraction grating dispersion region coincides with the whole area of the visible spectrum; so, you can analyze even white light. This is a huge advantage of the diffraction grating compared to interferential spectral instruments, including Fabry-Perot interferometer.

### *3.6.1. Resolution of the lens objective*

The main component of optical instruments such as telescopes, cameras, etc., is a lens that mainly provides precision of this optical system.

Resolution is directly related to the wave nature of light, whereas it was believed in geometrical optics that the ideal optical system can show any point of the object as a point of the screen image. However, the diffraction of light limits images resolution. Therefore, we must take into account Fraunhofer diffraction for determining the resolution of optical instruments.

Due to diffraction by round lens aperture, not the point source image but the combined intensity distribution is formed in its focal plane (Fig. 3.26). It has the form of a central maximum surrounded by the concentric dark and light rings. When the light is white, the rings have rainbow colors, but the central maximum always remains white.

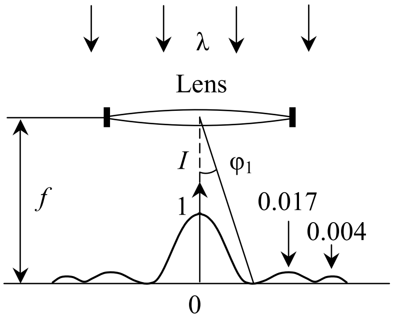



Fig. 3.26

The calculation shows that the angular size of the first dark ring is determined by:

$$\sin\varphi_1 = 1.22\lambda / D,$$

where $D$ is the lens frame diameter. If $D >> \lambda$, we can assume that

$$\varphi_1 \approx 1.22\lambda / D. \qquad (3.29)$$

Approximately 84 % of the luminous flux passing through the frame of the lens falls on the central maximum. The intensity of the first bright ring is 1.75 % and for second one is 0.4 % of the intensity of the central maximum. In the first approximation, we can assume that the diffraction pattern consists of only one central maximum; its angle radius is determined by the formula (3.29). This maximum is a blur image (due to diffraction of light) of infinitely distant point source. Hence, the wave nature of light leads to the fact that even the most perfect lens cannot provide perfect optical image.

Let us consider two incoherent point sources, for example, two close stars, which are observed by the telescope. If the distance between the centers of their images is compared with the size of the central maxima, the resulting image does not differ from the image of one point source. Then, the optical device does not recognize the two points. Starting from a certain distance between the sources, a hollow appears between the centers of the two maxima, and it will be observed as separate images of two point sources (Fig. 3.27).

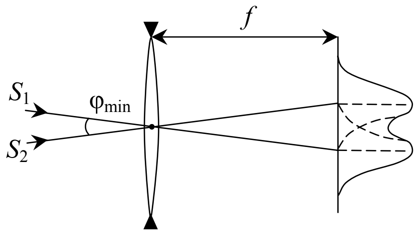


Fig. 3.27

According to the Rayleigh criterion, two incoherent point sources are observed separately if the center of the diffraction maximum of one of them coincides with the nearest minimum. Fig. 3.26 shows that it occurs when the angular distance between the point sources is minimal, that is defined by the formula (3.29)

$$\varphi_{\min} = 1.22\lambda / D .$$

The value inversed to the minimum angle is called *resolution of an optical device*:

$$R = 1/\varphi_{\min} = D/1.22\lambda . \qquad (3.30)$$

Formula (3.30) shows that the larger the diameter, the greater the resolution of the optical device. We can consider the pupil of the eye as the lens diameter. Assuming that the diameter of the pupil is 4 mm, we obtain that minimum angular distance between two incoherent point sources that are perceived separately by our eyes is:

$$\varphi_{\min} = 1.22\frac{0.55\cdot 10^{-3}}{4} \approx 0.17\cdot 10^{-3}\ \text{rad} \approx 35'' .$$

The concept of resolution is especially important for telescopes. A telescope with one of the world's largest mirror of the 5 m diameter could theoretically provide angular resolution

$$\varphi_{\min} = 1.22\frac{0.55\cdot 10^{-6}}{5} \approx 1.3\cdot 10^{-7}\ \text{rad} \approx 0.03''$$

and resolution $R \sim 10^7$ .

Note that the large size of astronomical telescopes is a result of both the size of a mirror and the lens. The great lens increases the light flux incoming into the telescope from celestial objects. It is proportional to the lens diameter squared; so, large telescopes can detect and photograph celestial objects of low brightness.

To increase the resolution of astronomical telescopes, we should also get rid of the negative effects of the atmosphere. The removal of telescopes into space is one of the solutions of the problem. An example is the American telescope «Hubble» with a mirror diameter of 2.4 m, which is on the Earth orbit (the altitude is 589 km) since 1990. Due to the absence of turbulent flows in the atmosphere, resolution of «Hubble» is 7–10 times greater than the analogous telescope located on the Earth. In addition, «Hubble» is able to record the electromagnetic radiation in the infrared and ultraviolet (including relict soft X-ray emission) wavelengths, where the absorption of these radiations by the Earth's atmosphere is very large.

Due to «Hubble», science has unique information about the formation and existence of the Universe. In particular, as the results of observations of quasars, the modern cosmological model is built in which the Universe filled with dark energy expands with acceleration. Also, the specified age of the Universe is 13.7 billion years.

## 3.7. Diffraction by two-dimensional and three-dimensional gratings. X-ray diffraction

Conditions (3.24) for the diffraction maxima formation is written for the case of perpendicular incidence of the plane wave on the grating (see Fig. 3.18).

Determine the condition of the maxima if a plane wave falls incline (Fig. 3.28). In this case, we should consider that the interfering parallel rays $1'$ and $2'$ from the adjacent grating slits, except the path difference $\Delta_1$ must have an additional path difference $\Delta_2$ caused by the incline fall of the beams.

Fig. 3.28 shows that the total path difference of the interfering rays is:

$$\Delta = \Delta_1 - \Delta_2 = d(\sin\varphi - \sin\psi) .$$

Thus, the condition for the diffraction maxima for the plane wave incline fall is:

$$d(\sin\varphi - \sin\psi) = \pm m\lambda . \qquad (3.31)$$

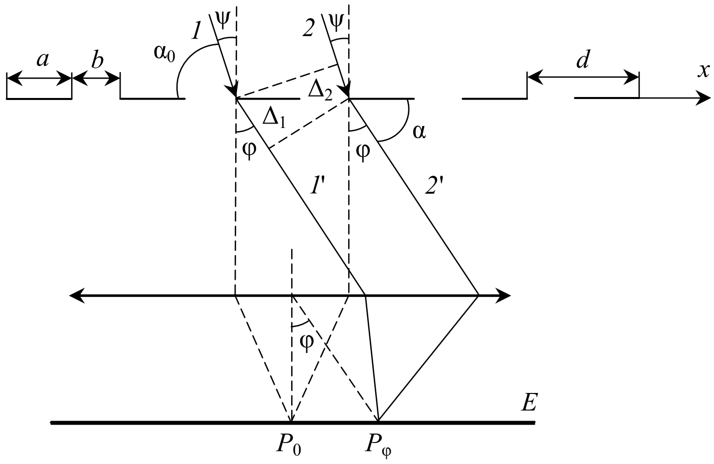



Fig. 3.28

It is more convenient to characterize the direction of the incident wave through the angles $\alpha_0$ and $\alpha$ between the beams and the axis $x$ (slide angles), (Fig. 3.28). It is obvious that the equation (3.31) turns into the form

$$d(\cos\alpha - \cos\alpha_0) = \pm m\lambda . \quad (3.32)$$

If two one-dimensional diffraction gratings are placed one after the other so that their strokes intersect, then we obtain a flat two-dimensional periodic structure. Diffraction of such structures is overlapping of diffraction patterns from the corresponding one-dimensional gratings. The maxima and minima of the gratings are placed mutually perpendicular. Let the first grating create several maxima determined by the condition:

$$d_1 \sin\varphi_1 = \pm m_1\lambda ; \quad (m_1 = 0,\ 1,\ 2,\ldots). \quad (3.33)$$

The second grating divides the first maxima according to the condition:

$$d_2 \sin\varphi_2 = \pm m_2\lambda ; \quad (m_2 = 0,\ 1,\ 2,\ldots). \quad (3.34)$$

As a result, the diffraction pattern will have the form of the properly placed spots in the plane. Each spot is described by two indices of diffraction $m_1$ and $m_2$ (Fig. 3.29).

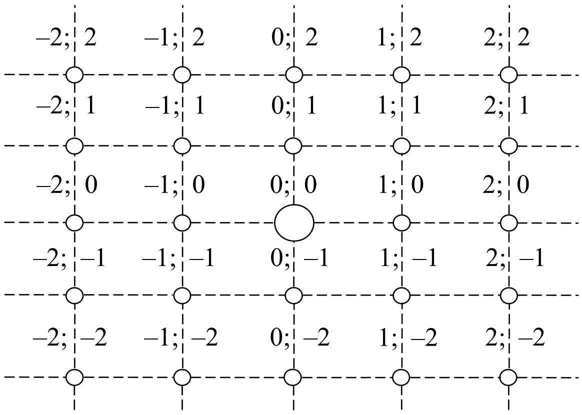



Fig. 3.29

By measuring the angles $\varphi_1$ and $\varphi_2$ that determine the positions of the maxima and knowing wavelength $\lambda$, we can find the two-dimensional diffraction grating periods $d_1$ and $d_2$ by formulas (3.33) and (3.34).

Diffraction is also observed in the three-dimensional structures (the crystal grating of solids). Atoms (molecules) play a role of centers that coherently scatter the incident light.

Assume that $d_1$, $d_2$, $d_3$ are periods of a rectangular grating along the three axes *X, Y, Z*, parallel to the three edges of the grating. Then the principal Fraunhofer diffraction maxima must satisfy three conditions arising from the condition (3.32) for the maxima of a one-dimensional diffraction grating. These relations are called *Laue conditions*:

$$
\begin{aligned}
d_1(\cos\alpha - \cos\alpha_0) &= \pm m_1\lambda; \\
d_2(\cos\beta - \cos\beta_0) &= \pm m_2\lambda; \qquad (3.35)\\
d_3(\cos\gamma - \cos\gamma_0) &= \pm m_3\lambda,
\end{aligned}
$$

where, $\alpha_0$, $\beta_0$, $\gamma_0$ and $\alpha$, $\beta$, $\gamma$ are the angles between the axes *X*, *Y*, *Z* and the directions of the incident and diffracted waves; $m_1$, $m_2$, $m_3$ are the integers presenting the maximum order (diffraction indices); $\lambda$ is the wavelength.

Obviously, there is a condition for a rectangular coordinate system:

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1. \qquad (3.36)$$

The angles $\alpha$, $\beta$, $\gamma$, which determine the diffraction maxima directions, can be found by solving the equations (3.35) and (3.36).

In particular, for a rectangular three-dimensional grating, it follows that the wavelength should be:

$$\lambda = -2\frac{\frac{m_1}{d_1}\cos\alpha_0 + \frac{m_2}{d_2}\cos\beta_0 + \frac{m_3}{d_3}\cos\gamma_0}{(m_1/d_1)^2 + (m_2/d_2)^2 + (m_3/d_3)^2}.$$

A grating dimension ($d$ ~ 0.5 nm) is much smaller than the wavelengths of visible light ($\lambda$ ~ 500 nm), so the condition $\lambda < 2d_{\min}$ is not satisfied for the visible light. However, X-rays with the wavelengths (0.1–0.01 nm) satisfy this condition. This means that crystals are natural diffraction gratings for X-ray.

In 1912, X-ray diffraction was first recorded on a film by Laue, Friedrich and Knipplinh. At that time, this event was of great scientific importance because:

a) it was definitively proved that X-rays have the same electromagnetic nature as visible light but differ from it by the significantly smaller wavelength that leads to their great penetrating power. At the scale of the wavelength, X-rays take place between ultraviolet light and gamma radiation;

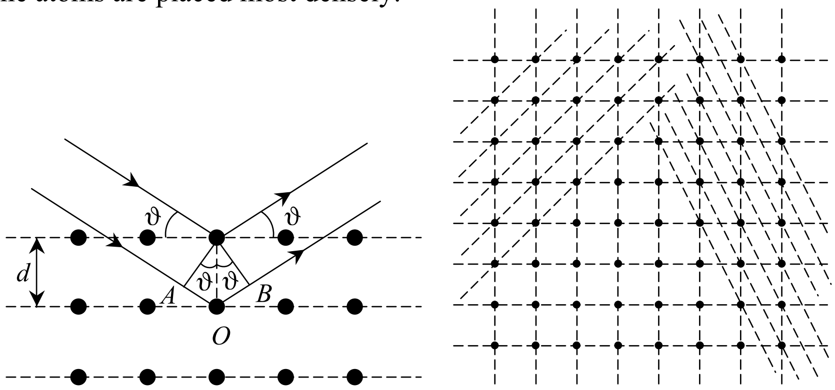
b) all scientists accepted the idea of a discrete and periodic structure of crystalline substances.

Visual explanation of the phenomenon of diffraction of X-rays passing through a crystal is much simpler, if we consider the diffraction of X-rays as a result of reflection on crystal parallel planes, which consist of atoms (molecules) of a crystal grating. This explanation was independently proposed by Bragg and Wolf.

Secondary waves, reflected on different atomic planes, are coherent; so, as a result of interference, we get maxima, if the path differences between neighboring waves are multiple $\lambda$. The refractive index of all substances for X-rays is practically equal to one; so, the path difference of the two waves reflected on the adjacent crystal planes is $AO + OB = 2d\sin\vartheta$ (Fig. 3.30, *a*), where $d$ is the interplanar distance; $\vartheta$ is the angle of sliding. Fraunhofer diffraction maxima are determined by the Bragg–Wulff formula:

$$2d\sin\vartheta = \pm m\lambda, \quad (m = 0,\ 1,\ 2, \ldots). \qquad (3.37)$$

A lot of parallel atomic planes in different directions can be drawn in a crystal (Fig. 3.30, *b*). Each of these systems can create some diffraction maxima. However, the most effective are the planes where the atoms are placed most densely.




*a* *b*

Fig. 3.30

Note that the Bragg–Wolf formula (3.37) can be obtained as a consequence of the Laue conditions (3.35).

Diffraction of X-rays by crystals was developed in two main directions: *X-ray spectroscopy* (the study of the spectral composition of radiation) and *X-ray structure analysis* (the study of the structure of crystals). These two ways are based on the use of formula (3.37).

Knowing the grating parameters, we can define the directions of diffraction maxima that perform spectral analysis of X-rays. Conversely, knowing the wavelength and the type of diffraction pattern, we can find the crystal structure, i.e. to do X-ray analysis. In particular, the Lauye method is based on diffraction of a narrow beam of «white» X-rays (a continuous spectrum of the different wavelengths) by fixed monocrystalline sample. As a result, a system of diffraction spots is formed on the photographic plate. We can determine the symmetry type of the crystal by the location of the diffraction spots.

These methods have a significant disadvantage – a small amount of diffraction spots (several dozen) recorded on a film. Powerful X-ray analysis can be performed using modern diffractometers equipped with computers. Such devices use very powerful monochromatic X-rays; it is possible to rotate a single-crystal sample arbitrarily relative to the direction of the incident X-ray beam.

This allows focusing a single-crystal sample that condition (3.37) can be established for each of the crystal planes, and then measuring the intensity of every maximum. As a result, we can observe hundreds or thousands of diffraction maxima.

This increase in diffraction data allows to perform complete decryption of the crystal structure (establish types of atoms and their coordinates in the unit cell of the crystal).

If we know the atoms coordinates, we can calculate any geometrical parameters of molecules (distance and angles between atoms, etc.) and the crystal grating.

This is why X-ray analysis method is considered as the most informative of all existing physical and chemical methods of three-dimensional structure of crystal bodies.

## ? Test Questions

1. What is diffraction of light?
2. Formulate the Huygens-Fresnel principle.
3. What is the Fresnel zone?
4. Build the scheme of Fresnel diffraction by round hole.
5. Build the scheme of Fresnel diffraction by round disk.
6. Explain the Fresnel spiral principle.
7. Build the scheme of Fresnel diffraction by straight edge of a half plane.
8. Build the scheme of Fresnel diffraction by slit in the opaque screen.
9. Explain the Cornu spiral principle.
10. What is the criterion of transition from Fresnel diffraction to Fraunhofer diffraction and to geometric optics?
11. Build the scheme of the experiment of Fraunhofer diffraction by slit in an opaque screen.
12. What is a diffraction grating?
13. Write the condition of the main maxima for a diffraction grating.
14. Give a definition of the angular dispersion.
15. Give a definition of the spectral resolution.
16. What is the Rayleigh criterion for the observation of two close spectral lines?
17. What is the resolution of the diffraction grating?

## Sample Problems

***Problem 1.*** A beam of light from a discharge tube falls normally onto a diffraction grating. What should the constant of the diffraction grating be for the maxima of the two lines $\lambda_1 = 6563$ Å and $\lambda_2 = 4102$ Å to coincide in the direction $\varphi = 41°$?

| *Data*: | *Solution* |
|---|---|
| $\lambda_1 = 6563 \cdot 10^{-10}$ m<br>$\lambda_2 = 4102 \cdot 10^{-10}$ m<br>$\varphi = 41°$ | In our case $d\sin\varphi = k_1\lambda_1 = k_2\lambda_2$. Hence $\frac{k_1}{k_2} = \frac{\lambda_1}{\lambda_2} = \frac{6563}{4102} = 1.6$. Since $k_1$ and $k_2$ must be integer, obviously, the values $k_1 = 5$ and $k_2 = 8$ satisfy the above-mentioned condition. So |
| $d = ?$ | $$d = \frac{k_1\lambda_1}{\sin\varphi} = \frac{5 \cdot 6563 \cdot 10^{-10}}{0.656} = 5 \cdot 10^{-6} \text{ m.}$$ |

## Problems

**1.** Find the number of Fresnel zones $m$, which fill a hole of the radius $r$ for a point that is remote at the distance $b$ from the hole center, if the incident wave is flat. ($m = r^2/b\lambda$)

**2.** A beam of light falls normally on the diffraction grating. The light wavelength of 589 nm forms diffraction angle of the first order 17° in the spectrum. Some other line forms the angle 24° in the spectrum of the second order Find the wavelength of the second line and the number of grooves per unit length of the diffraction grating. (4010 nm, 500 mm$^{-1}$)

**3.** Find the largest order of spectrum for the yellow sodium line ($\lambda = 589$ nm), if the diffraction grating constant is 2μm. (3)

**4.** What is the least diffraction grating constant if two potassium lines in the first order spectrum ($\lambda_1 = 404.4$ nm and $\lambda_2 = 404.7$ nm) are distinguishable? The diffraction grating width: 3 cm, 6 cm (22 microns; 44 microns).

**5.** Find the number of grooves per unit length of a diffraction grating if the mercury green line ($\lambda = 546.1$ nm) in the first order spectrum is observed at the angle of 19°. (600 mm$^{-1}$)

6. What is a diffraction grating constant if the maxima for $\lambda_1 = 656.3$ nm and $\lambda_2 = 410.2$ nm coincide at the angle of 41°? (5μm)

**7.** The minimum angular dispersion of a diffraction grating is $1.266 \cdot 10^{-3}$ rad/nm. Find the angular distance $\Delta\varphi$ between the diffraction grating spectrum lines with $\lambda_1 = 480$ nm and $\lambda_2 = 680$ nm. ($\Delta\varphi = 22°$)