Light diffraction in thick films

N. M. Pomerantsev

Usp. Fiz. Nauk 111, 507–524 (November 1973)

CONTENTS

1. INTRODUCTION

The invention and refinement of holographic methods[1–3] has aroused interest in light diffraction, and in particular, in diffraction in thick films. The practical value of this phenomenon has grown considerably. The use of light diffraction in holography has its own features. Holography can become of practical importance only when the light output of the hologram is large enough, i.e., when a significant part of the light incident on it goes into the reconstructed image.

The light output of a hologram is characterized by its diffraction efficiency, which shows what part of the light radiation intensity goes into the first order of diffraction when the hologram is a sinusoidal diffraction grating. Such a grating can be obtained by action of two interfering waves on a light-sensitive film. One of the waves is considered the reference wave, and the other is the object wave. By varying the angle of incidence of the object wave, one can get a grating with a varying spatial frequency that corresponds to one of the spatial frequencies of the holographed object. Hence, such a grating can be treated as a recording of one of the components of the Fourier image of the object being holographed, while the result of light diffraction by such a structure is treated as a response of a linear system to one component of the Fourier image of the object (the diffraction phenomenon is linear, owing to the linearity of the original equations).

By treating the dependence of the diffraction efficiency on the spatial frequency, we can thus get the directional diagram of the hologram, since the expansion in terms of spatial frequencies is at the same time an expansion in terms of plane waves that proceed at different angles from the object.

The dependence of the diffraction efficiency on the spatial frequency arises mainly from two factors: the properties of the photosensitive film, and the properties of the hologram itself. We shall not consider the first factor, but shall treat only the effect on the diffraction efficiency of the properties of the hologram itself.

As we know, holograms can be classified as thin-film and thick-film. For thin-film holograms, all directions of plane waves into which one can resolve the wave emerging from the object are equivalent, and there is no need for a directional diagram of the hologram itself.

For thick-film holograms, as Fig. 1 illustrates, the effective path length that the light wave traverses within the medium depends on the direction of propagation of this wave, and a directional diagram exists.

I shall discuss below the criterion by which one can classify holograms as thin-film and thick-film.

Moreover, holograms can be amplitude and phase holograms (however, holograms are mainly prepared as mixed amplitude-phase holograms.[1]) Thick-layer phase holograms are the most promising from the standpoint of getting the maximum diffraction efficiency. Hence, we shall treat here mainly the diffraction of coherent light by a phase diffraction grating.

However, the diffraction efficiency does not fully characterize the properties of a hologram, since it neglects the phase relations between the components of the Fourier image of the holographed object. The phase relations play a substantial role in constructing the image from its Fourier components, and in determining the properties of a hologram we must also have its phase characteristics.

In order to answer the question of how the diffraction efficiency will be related to the spatial frequency of the grating, i.e., the directional diagram of a thick-film hologram, we must solve the problem of diffraction of coherent light in a medium having a spatially periodic distribution of refractive index.

Different theories exist for light diffraction in thick films. These theories have been mainly developed by studying diffraction of x-rays in crystals[4] and by studying diffraction of light by ultrasonic waves.[5]

Without taking up all of these theories here in detail, we shall only point out that there are two theories for diffraction of x-rays in crystals: the kinematical theory (the solution by the perturbation method)[6] and the dynamical theory[7] (the solution for the first orders of diffraction). The kinematical theory as applied to thick-film holograms has made it possible to establish the following features:[8] 1) directional selectivity; 2) color selectivity; 3) existence of only one reconstructed image. However, this theory does not give the true

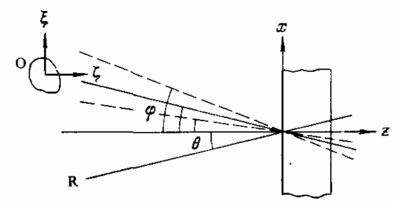

FIG. 1. Resolution of the wave $F(\xi, \zeta)$ proceeding from the object into plane waves. The dotted lines show some of the directions of the plane waves into which the resolution is being made.

values of the amplitude, and it seems impossible to calculate the diffraction efficiency by using it.

A theory of coupled waves[9,10] has been developed for diffraction of light by ultrasonic waves. This theory is of interest, in that it permits one to distinguish two limiting cases of diffraction. They will be treated below.

Moreover, the problem of light diffraction in a thick film has been solved with a computer[11] for certain special cases.

In this review, we shall first establish which parameters characterize the holographic diffraction grating used in the theory. Then we shall obtain and study exact solutions of the problem. In conclusion, we shall treat the approximate solutions, which permit one to present a physical picture of the phenomenon, and shall construct the directional diagram of thick-film holograms.

2. PARAMETERS OF A PHASE DIFFRACTION GRATING OBTAINED BY INTERFERENCE OF TWO PLANE WAVES

Let us consider a spatial (three-dimensional) interference pattern obtained by superposing two plane waves: the reference wave $J_1 = Ae^{i\rho\cdot r}$ and the object wave $J_2 = Be^{i\kappa\cdot r}$, where A and B are the amplitudes of the waves, and the vectors ρ, κ, and r have the components

$$\rho \equiv \{k \sin\theta,\ 0,\ k\cos\theta\},\qquad \kappa \equiv \{k\sin\varphi,\ 0,\ k\cos\varphi\},$$
$$r \equiv \{x,\ y,\ z\}.$$

Figure 2 shows the directions of propagation of the waves and the angles θ and φ.

The intensity $I = (J_1 + J_2)(J_1^* + J_2^*)$ of the total wave can be represented as follows:

$$I = A^2 + B^2 + 2AB\cos\{(\rho - \kappa)r\}. \qquad (1)$$

The first term in this expression is the square of the amplitude of the reference wave, and it does not depend on the angle θ.

The second term is the square of the amplitude of the object wave, for which we have taken only one Fourier component of the radiation emerging from the object. Since the different Fourier components have different values, depending on the angle φ, B^2 will be a function of the angle φ.

In phase holograms, the low-frequency component of the function $B^2 = B^2(\varphi)$ will superpose on the image during reconstruction to create highlights that diminish the image contrast. This is the essential defect of phase holograms. However, there are methods for diminishing interference of this type.[12] Discussion of these methods lies outside the scope of this article. However, we note that one of the conditions for getting an acceptable signal-noise ratio is that B^2 should be small in comparison with A^2. That is, we must have $A^2 \gg B^2$, a condition which we shall use later on.

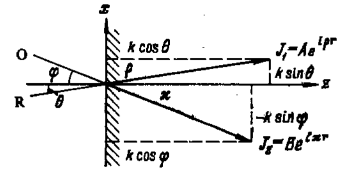

FIG. 2. Wave vectors of the reference (R) and object (O) waves, their coordinates and angles of incidence of the waves on the photosensitive film.

Without considering the concrete properties of the photographic material, we shall assume that the illumination of the hologram by the two beams of rays and the subsequent treatment (whenever necessary) will result in changes in the dielectric constant that are proportional to the light intensity that was incident at each point of the the photomaterial. Hence, we can assume that the dielectric constant of the photomaterial after the diffraction grating has been recorded on it will depend as follows on the coordinates:

$$\varepsilon = \varepsilon_0 + \Delta\varepsilon\cos\{(\rho - \kappa)r\}. \qquad (2)$$

If we bear in mind the above-cited inequality, which implies that $(\Delta\epsilon/\epsilon_0) \ll 1$, then we get the following expression for the propagation constant $\beta = (2\pi/\lambda)\sqrt{\epsilon} = k\sqrt{\epsilon}$:

$$\beta = k\sqrt{1 + \frac{\Delta\varepsilon}{\varepsilon_0}\cos\{(\rho - \kappa)r\}} \approx k\left[1 + \frac{\Delta\varepsilon}{2\varepsilon_0}\cos\{(\rho - \kappa)r\}\right] \qquad (3)$$

Thus the characteristics of a phase diffraction grating are determined by three parameters: the angle of incidence θ of the reference ray, the angle of incidence φ of the object ray, and the maximum relative change in dielectric constant. When the relation of the dielectric constant to the coordinates has this form, we can also take account of absorption in the holographic film by replacing the quantities ϵ_0 and $\Delta\epsilon$ by their complex values, respectively. However, we shall consider ϵ_0 and $\Delta\epsilon$ to be real.

In some cases, the angle of incidence of the light on the grating in diffraction (i.e., in reconstructing the hologram) differs from the angle of incidence of the reference beam of rays during recording. Then a fourth parameter ϑ will arise, which is the angle of incidence of the light on the grating during diffraction.

3. METHODS FOR SOLVING THE PROBLEM

The steady-state Maxwell equations for a non-magnetic medium have the form

$$\mathrm{rot}\,\mathrm{rot}\,\mathbf{E} + \beta^2\mathbf{E} = 0. \qquad (4)$$

If we consider the relation between the dielectric constant and the coordinates that was established above, in this case we can represent the propagation constant β as follows:

$$\beta^2 = k^2 + k^2\alpha\,[\exp\{i(\rho - \kappa)r\} + \exp\{-i(\rho - \kappa)\mathbf{r}\}]$$

where ρ, κ, and r are the vectors defined above, and $\alpha = \Delta\epsilon/2\epsilon_0$.

Let us use the method of coupled waves, and seek a solution of Eqs. (4) in the form of the following Fourier series:

$$\mathbf{E} = \sum_{m=-\infty}^{\infty}\mathbf{E}_m(z)\exp(i\mu_m\mathbf{r}). \qquad (5)$$

The vectors $\mathbf{E}_m(z)$ (which are subject to the definition of the function) are assumed to depend only on the variable z, since we consider that the relation of E to x is purely periodic, owing to the periodicity of structure and unlimited dimensions of the diffraction grating in the direction of the x axis.

We can easily convince ourselves that we can derive from the Maxwell equations a system of equations with constant coefficients by using such a Fourier series. In order to do this, we must define the vectors μ_m as follows:

$$\mu_m = \sigma + m(\rho - \kappa), \qquad m = 0,\ \pm1,\ \pm2,\ \ldots, \qquad (6)$$

Here $\boldsymbol{\sigma}$ is the wave vector of the reconstructing wave incident on the grating:

$$\sigma \equiv \{k\sin\vartheta,\ 0,\ k\cos\vartheta\}.$$

If the reconstructing source wave is incident on the grating in the same direction as the reference wave had been during recording, then $\boldsymbol{\sigma}=\boldsymbol{\rho}$, and $\vartheta=\theta$.

For the values of μ_m defined above, the solution takes on the form

$$\mathbf{E}=\sum_{m=-\infty}^{\infty}\mathbf{E}_m(z)\exp\{i\sigma\mathbf{r}+im(\boldsymbol{\rho}-\boldsymbol{\varkappa})\mathbf{r}\}. \tag{7}$$

The physical meaning of this solution is that an infinite series of diffracted waves $\mathbf{E}_m(z)\exp\{im(\boldsymbol{\rho}-\boldsymbol{\kappa})\cdot\mathbf{r}\}$ lying on both sides of the source wave arises in addition to the wave $\mathbf{E}_0(z)e^{i\boldsymbol{\sigma}\cdot\mathbf{r}}$ that coincides in direction with the source wave within the medium (i.e., after the wave has passed through the surface).

The amplitudes and phases of all the waves depend on z, since the waves interact among themselves as they propagate within the medium. They interact because they are periodically reflected and refracted as they pass through the film structure. Owing to the periodicity of reflections and refractions, as well as interference, light is propagated in the medium only in certain directions. That is, the energy is "channelized," so that its transfers caused by reflection and refraction can occur only between adjacent channels. Figure 3 illustrates the arrangement of these channels. It shows several first wave vectors, and the case is shown in which the direction of the reconstructing wave coincides with the direction of the reference wave during recording.

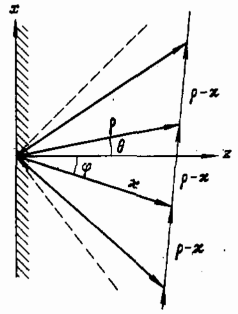

FIG. 3. Propagation directions of the diffracted waves within the medium.

In order to derive a system of equations in a form convenient for solution, we shall make one assumption that does not limit the generality of the solution. We shall assume that the polarization of the waves does not change during diffraction. This assumption is fully valid for isotropic media. Then we can take

$$\mathbf{E}_m(z)=\mathbf{e}_m E_m(z),$$

where $E_m(z)$ is a scalar function, and $\mathbf{e}_m$ is a unit vector that does not depend on z and lies in the plane of polarization and perpendicular to the direction of propagation of the wave. Consequently, the vectors $\mathbf{e}_m$ obey the relations

$$\mathbf{e}_m\mathbf{e}_m=1,\quad \mathbf{e}_m\boldsymbol{\rho}=0.$$

By substituting the Fourier series into the Maxwell equations, with account taken of all of the cited relationships, we get the following system of homogeneous equations with constant coefficients for $E_m(z)$:

$$\begin{aligned}\frac{d^2E_m}{dt^2}&+2i[\cos\vartheta+m(\cos\theta-\cos\varphi)]\frac{dE_m}{dt}\\&-4m\sin\frac{\varphi+\theta}{2}\left[\sin\left(\vartheta+\frac{\varphi-\theta}{2}\right)+m\sin\frac{\varphi+\theta}{2}\right]E_m\\&+\alpha(\mathbf{e}_m\mathbf{e}_{m-1}E_{m-1}+\mathbf{e}_m\mathbf{e}_{m+1}E_{m+1})=0,\end{aligned} \tag{8}$$

where we have introduced the new variable $t=kz$.

The value of the scalar product $\mathbf{e}_m\cdot\mathbf{e}_{m+1}$ depends on the orientation of the vector $\mathbf{E}_0$ of the source wave. If this vector is perpendicular to the plane of incidence (i.e., parallel to the y axis), then as we can easily see, all the vectors $\mathbf{e}_m$ are parallel to one another, and we have $\mathbf{e}_m\cdot\mathbf{e}_{m+1}=1$.

However, if the vector $\mathbf{E}_0$ does not form a right angle with the plane of incidence, then $\mathbf{e}_m\cdot\mathbf{e}_{m+1}\le 1$. Thus, rotation of the vector $\mathbf{E}_0$ from a position parallel to the y axis will decrease the coupling between the waves. We shall treat only the case of greatest practical importance of maximum coupling, and we shall assume that $\mathbf{e}_m\cdot\mathbf{e}_{m+1}=1$. Then $\mathbf{e}_m\cdot\boldsymbol{\kappa}=0$.

The coefficients of the system of equations acquire a simpler form when the direction of the source (reconstructing) wave, i.e., the wave incident on the grating during diffraction exactly coincides with the direction of the reference wave during recording. The equations will then be

$$\begin{aligned}\frac{d^2E_m}{dt^2}&+2i[(m+1)\cos\theta-m\cos\varphi]\frac{dE_m}{dt}\\&-4m(m+1)\sin^2\frac{\varphi+\theta}{2}E_m+\alpha(E_{m-1}+E_{m+1})=0.\end{aligned} \tag{9}$$

4. STUDY OF THE SYSTEM OF EQUATIONS. THE EXACT SOLUTION

Regardless of the direction of the source wave, we can represent the system of equations that describe diffraction in the form:

$$\frac{d^2E_m}{dt^2}+2i\beta_m\frac{dE_m}{dt}-\gamma_m E_m+\alpha(E_{m-1}+E_{m+1})=0. \tag{10}$$

The form of this system is such that we can interpret it as a system of equations describing oscillatory processes in a system of coupled resonators. Let us call them quasiresonators. However, we should remember that the comparison with resonators is purely formal, and the oscillation of the quasiresonators occurs in space, rather than in time.

In addition to the system (10), we can also write a system for the coupled quantities:

$$\frac{d^2E_m^*}{dt^2}-2i\beta_m\frac{dE_m^*}{dt}-\gamma_m E_m^*+\alpha(E_{m-1}^*+E_{m+1}^*)=0. \tag{11}$$

The boundary conditions for solving the problem must be chosen by starting with the assumption that in the absence of coupling ($\alpha=0$) there will be no diffraction, and the amplitudes of oscillation of all the quasiresonators but E_0 will be zero, while only a constant component exists for E_0. The amplitude of this constant component equals the amplitude of the source wave. We shall take it to be unity. Thus we have

$$E_m|_{t=0}=\delta_{m0},\quad \left.\frac{dE_m}{dt}\right|_{t=0}=0, \tag{12}$$

where

$$\delta_{m0}=\begin{cases}1 \text{ when } m=0,\\ 0 \text{ when } m\neq 0.\end{cases}$$

The diffracted waves arise only because of the mutual coupling of the waves. Hence, the boundary conditions for $\alpha\neq 0$ should not differ from those that have just been stated. A difference in them would imply that we had introduced other causes for appearance of diffracted waves besides the coupling between the waves.

Let us consider the conservation laws. The system of equations for the coupled waves E_m has two integrals that we can obtain as follows. Let us multiply the systems for E_m and E_m^* by their conjugate quantities, and sum over all m and subtract the one sum from the other. We get

$$\sum_{m=-\infty}^{\infty}\left(\frac{d^2E_m}{dt^2}E_m^*-\frac{d^2E_m^*}{dt^2}E_m+2i\beta_m\frac{d}{dt}E_mE_m^*\right)=0.$$

Since the solution of the system of equations with constant coefficients is the exponential function $\exp(i\omega t)$, the first two terms of the obtained relationship compensate one another, and we have

$$\sum_{m=-\infty}^{\infty}\beta_mE_m(z)E_m(z)=\text{const}. \tag{13}$$

This is the first law of conservation in the system of coupled quasiresonators.

As we can easily see, the obtained expression is the flux of energy passing through a surface perpendicular to the z axis lying at an arbitrary site within the medium. Thus, it is not the amount of energy overall that is conserved in diffraction, but its flux (in the direction of the z axis) through the surface of the film, and energy can "spread" in the direction of the x axis, i.e., it can enter higher orders of diffraction.

The boundary conditions (12) imply that

$$\sum_{m=-\infty}^{\infty}\beta_mE_m(z)E_m^*(z)=\cos\theta. \tag{14}$$

In order to obtain the other integral of the system of equations, let us multiply the system for E_m by dE_m^*/dt, and correspondingly, the system for E_m^* by dE_m/dt. After summing and adding the two systems analogously to what we did before, we get the expression

$$\sum_{m=-\infty}^{\infty}\left[\frac{dE_m}{dt}\frac{dE_m^*}{dt}-\gamma_mE_mE_m^*+\alpha(E_{m-1}+E_{m+1})E_m^*\right]=0,$$

which we have written with account taken of the conditions (12).

This is the second law of conservation of energy in the system of coupled quasiresonators. We can interpret the first term $(dE_m/dt)(dE_m^*/dt)$ in the expression for the second law as being the kinetic energy ("rate" of growth of the wave along the x axis), the term $-\gamma_mE_mE_m^*$ as the potential energy, and the term $\alpha(E_{m-1}+E_{m+1})E_m^*$ as the coupling energy. Thus, the integral (15) characterizes the movement of energy in the system of quasiresonators, and it shows that the gradient of energy that characterizes its transfer into different orders of diffraction depends on the coupling coefficient α and the constant γ_m.

As we see, the quasiresonators in our interpretation substantially differ from ordinary resonators. There are two forms of potential energy in the system of quasiresonators that resemble coupled diffracted waves:

$$\mathscr{E}_1=\sum_{m=-\infty}^{\infty}\beta_mE_mE_m^* \text{ and } \mathscr{E}_2=-\sum_{m=-\infty}^{\infty}\gamma_mE_mE_m^*, \tag{15}$$

This closely involves the existence of two partial frequencies that differ in value (rather than in sign, as in the usual case). These frequencies will be calculated below. In the approximate solution that we shall treat in Sec. 5, the term that corresponds to the energy gradient in (15) is lacking, and there remains one conservation law (13) and one partial frequency for each quasiresonator.

The convergence of the series that express the first and second conservation laws is implied by the theorem of existence of a unique solution of a system of differential equations. The convergence of these series implies that $E_m \to 0$ as $m \to \pm\infty$. Thus, the waves of the zero and first orders play the major role in diffraction. It is easy to see that we have $\gamma_0=\gamma_1=0$ when the reconstructing and reference beams coincide, and the energy $\mathscr{E}_2$ does not exist for these orders. Hence, the potential energy $\mathscr{E}_2$ that accumulates in the higher orders plays the role of a restraining factor in the system that hinders transfer of energy into the higher diffraction orders.

We interpret the spatial distribution of light as it passes through the medium along the z axis as being the oscillation of the quasiresonators in time.

During the diffraction process, energy is transferred from the zero-order wave E_0 (which exists alone in the medium immediately after the light has passed through the surface) to the higher-order waves. In a system of coupled resonators, the ability for energy to be transferred from one resonator to another depends on synchronization of the oscillations at the partial frequencies, i.e., the frequencies that the single resonators would have if separated into individual partial systems in the absence of coupling.

Let us find the partial frequencies. If we set the coupling coefficient α to be zero, we get the system of equations

$$\frac{d^2E_m}{dt^2}+2i\beta_m\frac{dE_m}{dt}-\gamma_mE_m=0, \tag{16}$$

that describes the oscillations of mutually uncoupled quasiresonators.

Each quasiresonator that is described by one of the equations of the system has (in contrast to ordinary resonators) two different partial frequencies that differ in value:

$$(\omega_{1,2})_m=-\beta_m\pm\sqrt{\beta_m^2-\gamma_m}. \tag{17}$$

The oscillations of the partial system corresponding to the quasiresonator will not be harmonic, since the ratio of the two partial frequencies ω_1/ω_2 is not a rational number.

Let us write out the values of the partial frequencies for several quasiresonators close to E_0 (we are treating the case in which the reconstructing ray and the reference ray coincide):

$$\begin{array}{l|l}
E_2 & -3\cos\theta+2\cos\varphi\pm\sqrt{(-3\cos\theta+2\cos\varphi)^2-24\sin^2\frac{\varphi+\theta}{2}} \\
E_1 & -2\cos\theta+\cos\varphi\pm\sqrt{(-2\cos\theta+\cos\varphi)^2-8\sin^2\frac{\varphi+\theta}{2}} \\
E_0 & 0, \qquad \cos\theta \\
E_{-1} & 0, \qquad \cos\varphi \\
E_{-2} & \cos\theta-2\cos\varphi\pm\sqrt{(\cos\theta-2\cos\varphi)^2-8\sin^2\frac{\varphi+\theta}{2}}.
\end{array}$$

Examination of the values of the partial frequencies shows that the most favorable conditions for energy exchange are between the zero and minus-first quasiresonators. However, if the angles θ and φ are small, then the partial frequencies of all the quasiresonators are close to zero. Hence, decrease in the angles θ and φ improves the conditions for transfer of energy into the higher diffraction orders. Conversely, if we increase the angles θ and φ, then the energy will mainly be transferred from the zero-order wave only into the minus-first-order wave.

It is precisely this condition, which ensures the existence of waves of only two orders, that is necessary for getting a high diffraction efficiency. However, knowledge of the partial frequencies makes it possible to get only a qualitative picture of the energy distribution over the diffraction orders.

Let us proceed to the system of equations in the presence of coupling:

$$\frac{d^2E_m}{dt^2}+2i\beta_m\frac{dE_m}{dt}-\gamma_m E_m+\alpha\,(E_{m-1}+E_{m+1})=0. \tag{18}$$

Its natural frequencies are defined as the roots of the characteristic equation

$$\begin{vmatrix} \cdots & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdots & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdots & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdots & 0 & \alpha & D_1 & \alpha & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & \alpha & D_0 & \alpha & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & \alpha & D_{-1} & \alpha & 0 & \cdots \\ \cdots & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdots & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdots & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \end{vmatrix}=0, \tag{19}$$

Here $D_m=-(\omega^2+2\omega\beta_m+\gamma_m)$, and the general solution has the form

$$E_m=\sum_{l=-\infty}^{\infty}C_{ml}e^{i\omega_l t}. \tag{20}$$

The coefficients C_{ml} are found from the above-mentioned boundary conditions. The total solution, which is the sum of all the diffracted waves, now has the form

$$\mathbf{E}=\sum_{m=-\infty}^{\infty}\mathbf{e}_m\sum_{l=-\infty}^{\infty}C_{ml}\exp\,[i\boldsymbol{\sigma}\mathbf{r}+im\,(\boldsymbol{\rho}-\boldsymbol{\varkappa})\,\mathbf{r}+i\omega_l kz]. \tag{21}$$

As we see from this solution, there are two series of diffracted waves. The first series is a distribution over the individual channels (which are shown in Fig. 3), in each of which an entire group of waves is propagated. Thus, each wave of the m-th order $E_m(z)$ consists in turn of a series (generally infinite) of closely distributed waves. These second series of waves are represented in Fig. 4 by their wave vectors. For clarity of illustration, the angles between the vectors have been made much greater than they actually are, while the number of waves has been restricted to four. While propagating in almost the same direction, these waves interfere. Consequently, as the diffracted waves propagate within the medium, their intensity varies according to a complicated law.

The natural frequencies ω_l are not multiples of one another, and the oscillation of the entire system cannot be expanded into a series of harmonic oscillations. The functions $E_m(z)$ become sums of periodic functions of non-multiple periods. That is, they are almost-periodic functions of the variable z, as represented by their Fourier series. The theory of such functions and operations on them has been well developed.[13] According to this theory, the sum of oscillations of the form $C_{ml}\exp(i\omega_l t)$ (which in the terminology of H. A. Bohr are commonly called "pure") can be combined into a series (the Fourier series of the almost-periodic function). The convergence of the series (20), which is composed of pure oscillations arranged in order of increasing ω_l is implied by the theorem of existence of a unique solution of a system of differential equations.

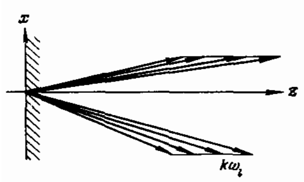

FIG. 4. The groups of waves that propagate in the direction of a diffraction order. The value of $k\omega_l$ is half the distance between the ends of the mean vectors.

According to the theorem of approximation of almost-periodic functions,[13] the functions $E_m(z)$ can be approximated by a finite series composed of pure oscillations of the form $C_{ml}\exp(i\omega_l t)$:

We can make the infinite characteristic determinant (19) absolutely convergent by dividing each of its rows by $-(\omega^2+2\omega\beta_m+\gamma_m)$. If we restrict ourselves to the numbers $m_{max}=M_1$ and $|-m_{max}|=M_2$, and solve the characteristic equation, we get $2(M_1=M_2+1)$ eigenfrequencies of the system, and thus, we can represent the functions $E_m(z)$ as sums of $2(M_1+M_2+1)$ pure oscillations. Thus we can get a solution of the system of equations for $E_m(z)$ with any given accuracy.

As an example, we shall give the solution for the following values of the parameters: $\alpha=0,\ 1$; $\vartheta=\theta=\varphi=10°$. Figure 5 shows graphically the spectra of the functions $E_m(z)$ for $M_1=M_2=2$ as calculated with a computer.[14] We see from this diagram that the waves of orders E_0 and E_{-1} are mainly represented by oscillations at two frequencies. This gives us grounds for solving the problem in the two-wave approximation, as will be discussed below.

Figure 6 shows the relation of the wave intensities $|E_m|^2$ to the variable t = kz. As we see from this diagram, the intensities of all the waves oscillate strongly as they propagate within the medium. The intensities of the waves E_0 and E_{-1} attain maxima close to unity, and they decrease to values close to zero. The maximum intensities of the other waves are smaller. We also can see from this diagram that waves having $|m|\geq 3$ no longer play a substantial role, and we can neglect them.

We must note that in this problem we have neglected reflection and refraction of waves at the surfaces of the medium. However, these phenomena have no relation to diffraction within the medium, and it makes no sense to

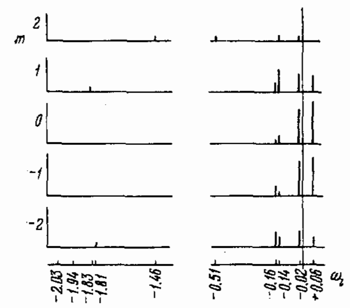

FIG. 5. The spectrum of eigenfrequencies of the quasi-resonators for the diffracted waves of orders 2, 1, 0, −1, −2, for α = 0, 1, θ = φ = 10°.

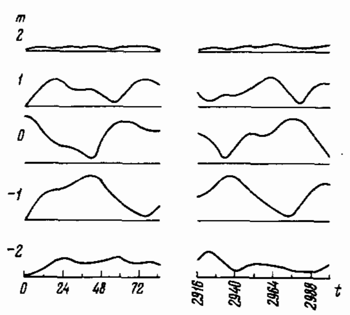

FIG. 6. The oscillations of the light intensity in the diffraction orders 2, 1, 0, −1, −2 as the light passes through the thickness of the medium for α = 0, 1, θ = φ = 10°. The two groups of presented segments of curves are two parts of the very same curves.

complicate the problem by taking them into account. We must bear in mind the fact that the angles θ and φ are the angles at which the reference and object beams of rays propagate within the medium, rather than outside.

Re-reflection of light from the two surfaces within the medium can also be treated separately.[15]

5. APPROXIMATE SOLUTIONS

By using a set of simplifying assumptions, we can get an analytic solution for two limiting cases of diffraction, which we shall now examine. Let us assume that $\theta=\varphi$ (symmetrical incidence of the object and reference waves on the hologram), and introduce into the equations a small parameter, for which we shall make the substitution of variable $t=(\xi\cos\theta)/\alpha$. Then the system of equations describing the diffracted waves will be

$$\frac{\alpha^2}{\cos^2\theta}\frac{d^2E_m}{d\xi^2}+2i\alpha\frac{dE_m}{d\xi}-4m(m+1)\sin^2\theta E_m+\alpha(E_{m-1}+E_{m+1})=0. \quad (22)$$

If the parameter $\alpha^2/\cos^2\theta$ is smaller than α, then the term containing the second derivative can be omitted, and the system of equations is considerably simplified:

$$2i\frac{dE_m}{d\xi}-4m(m+1)\frac{\sin^2\theta}{\alpha}E_m+E_{m-1}+E_{m+} \quad (23)$$

In studying these equations, let us first assume that there is no coupling between the waves ($\alpha=0$), and study the partial frequencies. We get a system of equations

$$2i\,dE_m/d\tau-4m(m+1)\sin^2\theta\;E_m=0, \quad (24)$$

where $\tau=kz/\cos\theta$.

This system describes the oscillations of quasiresonators that are not mutually coupled (now they are harmonic). Each of the resonators will oscillate with its own partial frequency, which is a multiple of $2\sin^2\theta$.

As we know, the condition for best energy transfer in a system of coupled resonators is synchronization of the partial systems, i.e., their ability to oscillate in phase for a prolonged time. We shall assume that the oscillations of the quasiresonators are synchronous enough when the phase of oscillations having a frequency equal to the partial frequency varies little throughout the path of the light ray in the medium. That is, the value of the phase $4\sin^2\theta\cdot\tau|_{z=d}$, which we shall denote as Q, is small (here d is the thickness of the film).

Under these conditions, energy transfer from one quasiresonator to another is simplified, and the maximum intensities of a large number of diffraction orders (starting with the zero order) will differ little from one another. We can estimate quantitatively the smallness of the parameter Q by comparing theory with experiment. Such a comparison, which has been made many times in studying diffraction of light by ultrasound, has shown that excitation of a large number of symmetrically arranged diffraction orders is observed at values of Q of the order of $\pi/12$ or less.

Hence we have $Q\le\pi/12$, or if we bear in mind the fact that $Q=4kd\sin\theta\tan\theta$, we get

$$kd\leqslant\frac{\pi}{48\sin\theta\,\mathrm{tg}\,\theta}. \quad (25)$$

The region of kd values that satisfy this inequality is called the region of Raman-Nath diffraction, named after the investigators who studied this type of diffraction.[16]

If we turn to the simplified system of equations that describes the oscillation of coupled quasiresonators,

$$2i\frac{dE_m}{d\xi}+E_{m-1}+E_{m+1}=m(m+1)\frac{Q\cos\theta}{\alpha kd}E_m, \quad (26)$$

and if we assume therein that $Q=0$, we get the relationship

$$2i\,dE_m/d\xi+E_{m-1}+E_{m+1}=0, \quad (27)$$

which is satisfied by the Bessel functions, if we assume that

$$E_m(\xi)=i^mJ_m(\xi). \quad (28)$$

This is the solution of the posed problem in the Raman-Nath approximation.

The functions $E_m(\xi)$ satisfy the boundary conditions

$$E_m(0)=\delta_{m0},$$

in full accord with the conditions of the problem.

The nature of the variation of the amplitudes of the diffracted waves indicates that, in the Raman-Nath diffraction region, energy is gradually transferred from the zero-order wave symmetrically to all the higher-order waves, with an insignificant backward return to the zero-order wave. Owing to this process, the energy finally is transferred to the high-order waves, and the amplitude of the zero-order wave is considerably diminished. This is well illustrated by Fig. 7, where we see how the intensities of the waves in the first three orders of diffraction oscillate as the light penetrates into the interior of the medium.

However, such a picture is an idealization, and it presupposes synchronization of the oscillation of the partial systems. Actually, as the thickness of the medium increases (i.e., with increasing z), the synchronization of the diffracted waves breaks down ever more, and this hinders energy transfer into the higher diffraction orders.

The general solution is the sum of all the waves

$$E(z)=e^{i\sigma\tau}\sum_{m=-\infty}^{\infty}i^mJ_m(\xi)\exp\{im(\rho-\varkappa)\mathbf{r}\}.$$

We can easily sum the series over the Bessel functions, and we get the wave

$$E(z)=\exp\left[i\sigma\mathbf{r}+i\frac{\alpha kz}{\cos\theta}\cos\{(\rho-\varkappa)\mathbf{r}\}\right],$$

which corresponds to the source wave, but modulated in phase.

Now we can easily see that in diffraction in the Raman-Nath region only the phase of the source wave shifts, and the shift is proportional to the index of refraction in the medium.

Let us consider the other limiting case in which Q is large. Now there will be no synchronization of the

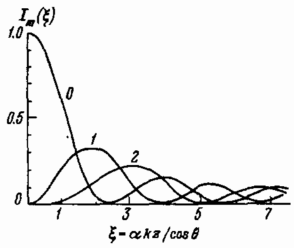

FIG. 7. The intensitives (I_m) of the waves propagating in the zero, first, and second diffraction orders in diffraction in the Raman-Nath region.

partial systems, and energy transfer into the higher diffraction orders will be hindered. However, we can easily see that the partial frequencies of the quasiresonators of indices 0 and −1 are zero, while the partial frequencies of all the rest of the quasiresonators now differ considerably from zero. This gives us grounds for treating oscillations in a system that consists only of two coupled quasiresonators and for considering it to be isolated. That is, we assume that energy is not transferred to other quasiresonators.

Such a situation arises when the Bragg condition is satisfied:[17]

$$\lambda/n\Lambda = 2\sin\vartheta,$$

where Λ is the period of the diffraction grating, and n is the refractive index.

As we know, the period of a grating recorded by using two plane waves is

$$\Lambda = \lambda/2n\sin\theta.$$

If the wave incident on the grating during diffraction exactly coincides with the reference wave used in recording, then $\vartheta = \theta$, and the Bragg conditions are automatically satisfied.

Now we have a system of equations in the two-wave approximation:

$$2i\frac{dE_0}{d\xi} + E_{-1} = 0, \quad 2i\frac{dE_{-1}}{d\xi} + E_0 = 0. \tag{29}$$

This system has the non-zero eigenfrequencies $\alpha k/2\cos\theta$ and $-\alpha k/2\cos\theta$. Under the boundary conditions given above, its solutions will be the harmonic oscillations

$$E_0 = \cos\left(\frac{\alpha k}{2\cos\theta}z\right), \quad E_{-1} = i\sin\left(\frac{\alpha k}{2\cos\theta}z\right) \tag{30}$$

The intensities of the propagating waves are determined by the expressions

$$|E_0|^2 = \frac{1+\cos(\alpha kz/\cos\theta)}{2}, \quad |E_{-1}|^2 = \frac{1-\cos(\alpha kz/\cos\theta)}{2} \tag{31}$$

We must find the Q values for which we can assume the two waves E_0 and E_{-1} to be isolated from the other waves. Experimental studies have shown that this happens when $Q \geq 4\pi$, or

$$kd \gg \frac{\pi}{\sin\theta\,\mathrm{tg}\,\theta} \tag{32}$$

The region of kd values that satisfies this inequality is called the region of Bragg diffraction. The region of ks values lying between the limits

$$\frac{\pi}{48\sin\theta\,\mathrm{tg}\,\theta} \leq kd \leq \frac{\pi}{\sin\theta\,\mathrm{tg}\,\theta}, \tag{33}$$

can be considered to be the intermediate region.

We can conclude from the expressions for the intensities of the diffracted waves that energy in the Bragg diffraction region is periodically transferred back and forth between the zero-order wave to the adjacent (minus-first) order. Here the diffraction process is analogous to the oscillations of a system of two coupled pendulums, where under certain initial conditions the energy is transferred from one pendulum to the other in a similar way.

We can also treat diffraction in the Bragg region as resulting from "beating" that arises from adding the two waves that propagate in each of the two channels that correspond to the diffraction orders:

$$E_{ml} \sim \exp\left[i\sigma\tau + im(\rho - \varkappa)\mathbf{r} + il(\alpha k/2\cos\theta)z\right];$$

where $m = 0, -1$ are the diffraction orders, and $l = 1, -1$ is the index of the wave.

Both the zero-order wave E_0 and the minus-first-order wave E_{-1} consist of two waves, as is shown in Fig. 8. Superposition of these waves leads to beats.

If the Bragg conditions are not exactly satisfied ($\vartheta \neq \theta$), then the partial frequency of the minus-first-order wave differs somewhat from zero. So that there will be no complete synchronization of the waves of orders zero and minus one. As in the general case, the oscillations of the set of quasiresonators become nonharmonic. However, we can distinguish cases in which the angles ϑ and θ are close to one another, and the partial frequency of the minus-first order quasiresonator is much closer to zero than the partial frequencies of all the others. Then energy transfer again mainly occurs between the waves of orders zero and minus one.

The approximations that we have discussed give a correct picture of diffraction only if the coefficient of the second derivative $\alpha^2/\cos^2\theta$ is much smaller than the coefficient α of the first derivative, i.e., when the following inequality is satisfied:

$$\cos^2\theta \gg \alpha. \tag{34}$$

This is well illustrated by Fig. 9, which shows the approximate solution by the dotted curve,[18] and the exact solutions for different values of α by the solid curves.[14]

The inequality (34) is always satisfied at small angles θ (since α cannot be larger than 1/2). However, Q decreases with decreasing angle θ, and the diffraction changes from the Bragg region into the intermediate region. At large angles θ, the discussed approximation will be valid only if α is very small.

Figure 10 shows in a graph the classification of diffraction in thick films into three regions by showing curves relating the quantity kd to the angle θ along the boundaries of the regions.

As we see from this diagram, when we vary the angle θ from 10° to 30°, the thickness of the hologram can be

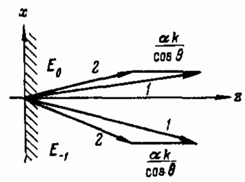

FIG. 8. The pairs of waves that propagate in the zero and minus-first diffraction orders in diffraction in the Bragg region.

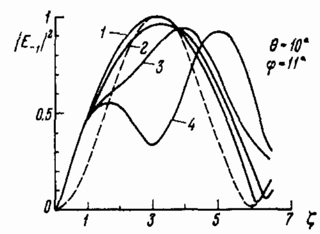

FIG. 9. Comparison of the exact and approximate solutions. Dotted curve—approximate solution. The values of α are the following for the other curves (exact solutions): 1—0.01; 2—0.05; 3—0.1; 4—0.2. The values of ζ = αkz are plotted as the abscissa.

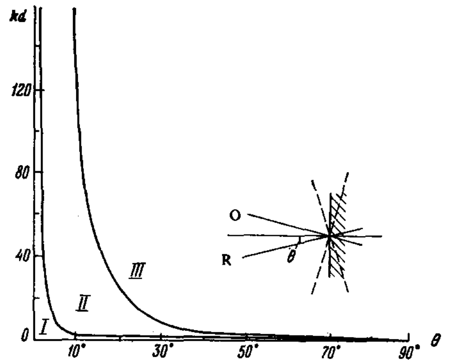

FIG. 10. The regions of light diffraction in thick films: (I)—the region of Raman-Nath diffraction; (II)—the intermediate region; (III)—the Bragg diffraction region.

reduced by an order of magnitude, but the hologram retains thick-film properties. The dependence of the properties of a hologram on the angle θ can explain how one can record, on material having the same thickness, holograms that have both thick-film and thin-film properties, depending on the angle of incidence of the reference beam of rays.

At small angles θ, it takes a very large thickness of film to get thick-film holograms. At angles θ close to 90°, the required value of kd becomes small, and in the limit ($\theta = 90°$), it vanishes. We can easily understand this from the diagram of incidence of rays shown in Fig. 10. When a ray is incident on the grating at a large angle (as shown in the diagram by the dotted lines), the path of the light in the medium greatly exceeds the film thickness. Hence, the dimensions of the region in which diffraction occurs are considerably increased. Thus, the thickness alone of the film does not yet suffice for characterizing holograms. In classifying them into thin-film and thick-film holograms, it is quite insufficient to compare only the film thickness with the wavelength, and one can get holograms having different properties on a film of the same thickness.

Moreover, we must consider the fact that Bragg diffraction cannot be observed in practice at angles of incidence of the rays less than 10°, and holograms with thick-film properties cannot be obtained here. We see from Fig. 10 that the curve sharply rises in the region $\theta = 8-10°$. Thus, even when the thickness of the film is increased greatly, for angles of incidence of the rays, e.g., of 10°, the point representing the value of kd will still lie in the immediate vicinity of the boundary of the Bragg region with the intermediate region, where the conditions for appearance of this type of diffraction are unfavorable. The graphs of the exact solution (see Fig. 6) demonstrate the same thing. Here we see that at $\theta = 10°$ the orders adjacent to the zero and minus-first orders are excited for such values of kd, for which Q is known to be large.

6. THE DIRECTIVITY PATTERN OF A THICK-FILM HOLOGRAM

The diffraction grating considered above was obtained by recording on the hologram the amplitude and phase of a plane wave corresponding to one spatial frequency of the object being holographed. By using such a grating, we can study the relation of the radiation intensity in the first diffraction order to the direction of the plane wave, and thus get the relation between the diffraction efficiency of the hologram and the spatial frequency of the object being holographed. We can call this relationship the directivity pattern of the thick-film hologram.

The diffraction efficiency is defined as the ratio of the radiation energy in the first (minus-first) order of diffraction to the radiation energy of the source wave. As we established from the law of conservation of energy discussed above, the flux of radiation energy through a plane perpendicular to the z axis is conserved in diffraction. Hence, in order to determine the diffraction efficiency, we must find the ratio of the corresponding energy fluxes. The expression for the diffraction efficiency (with account taken of the fact that $|E_0(0)|^2 = 1$) will have the form

$$\eta = (\cos\varphi/\cos\theta)\,|E_{-1}|^2. \tag{35}$$

We assume that the reconstructing wave coincides with the reference wave, i.e., $\vartheta = \theta$, and diffraction occurs in the Bragg region.

In order to find E_{-1} when $\varphi \neq \theta$, we must solve the system of equations

$$\begin{aligned} 2i\cos\theta\frac{dE_0}{dt} + \alpha E_{-1} &= 0,\\ 2i\cos\varphi\frac{dE_{-1}}{dt} + \alpha E_0 &= 0. \end{aligned} \tag{36}$$

Calculation of the natural frequencies of this system gives

$$\omega_{1,2} = \pm\frac{\alpha}{2\sqrt{\cos\theta\cos\varphi}},$$

and the solution for the initial conditions $E_0(0) = 1$, $E_{-1}(0) = 0$ will be

$$E_0 = \cos\left(\frac{\alpha kz}{2\sqrt{\cos\theta\cos\varphi}}\right),\quad E_{-1} = i\sqrt{\frac{\cos\theta}{\cos\varphi}}\sin\left(\frac{\alpha kz}{2\sqrt{\cos\theta\cos\varphi}}\right).$$

The diffraction efficiency is determined from the above expression as follows:

$$\eta = \sin^2\left(\frac{\alpha kz}{2\sqrt{\cos\theta\cos\varphi}}\right). \tag{37}$$

As is well known[18] and as the derived expression shows, the diffraction efficiency is a periodic function of the film thickness, and it varies from 0 to 1. The period of oscillation of the diffraction efficiency depends on the amount of coupling α and the angle of incidence of the rays on the hologram. As α increases, the period of oscillation of the diffraction efficiency decreases in such a way that a change in z over a small range will lead to substantial changes in the diffraction efficiency. On the other hand, increase in the angle θ also decreases the period of oscillation of the diffraction efficiency, and hence, it leads to the same results.

Study of the relation of η to the angle φ makes it possible to determine the capability of thick-film holograms for reproducing a spectrum of spatial frequencies[19] (directivity pattern). Thick-film holograms substantially differ in this capability from thin-film holograms.

For a pictorial representation of the relationship between the diffraction efficiency and the spatial frequency, we must construct graphs of the relation of η to φ for different values of z, which will now be the parameter. We can select a set of z values for which it is convenient to construct graphs from the condition that η must be 1 for $\varphi = \theta$ for the chosen value of z. The following values of the argument fulfill this condition:

$$\frac{\alpha kz}{2\cos\theta} = \left(m + \frac{1}{2}\right)\pi,$$

$$\text{or } z = \frac{(2m+1)\pi\cos\theta}{\alpha k},$$

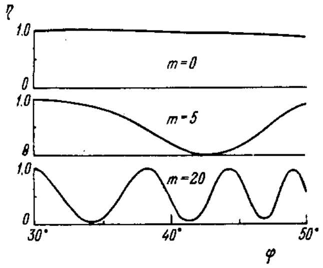

FIG. 11. The relation of the diffraction efficiency of a thick-film hologram to the angle φ in rectilinear coordinates ($\theta = 30°$).

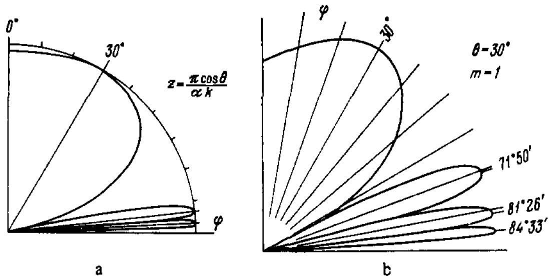

FIG. 12. The directivity patterns of thick-film holograms in polar coordinates for: a) m = 0; b) m = 1.

Here m = 0, 1, 2 ... is the number of periods of "beats" of the two waves that propagate in one given channel.

If we substitute this z value into the expression for η, we find

$$\eta = \sin^2\left[\left(m + \frac{1}{2}\right)\pi \sqrt{\cos\theta/\cos\varphi}\right]. \qquad (38)$$

Hence we can conclude that the optimum thickness of the hologram will be the one that gives m = 0. This gives

$$z_{\text{opt}} = \frac{\pi \cos\theta}{\alpha k}.$$

We can also see this from Fig. 11, which shows the relation of η to φ for several values of m.

Figures 12a and b show the directivity patterns in polar coordinates for m values of 0 and 1. The value $\theta = 30°$ was adopted for constructing the graphs.

As we see from the graphs, increase in the thickness of a hologram leads to a considerable non-uniformity of the directivity pattern, owing to the increase in the number of lobes that it has. As α increases, the non-uniformity of the directivity pattern is manifest at ever smaller film thicknesses. Thus, if m = 5, then the pattern has the form shown in Fig. 11 when $\alpha = 0.01$ and z = 65 μm, or when $\alpha = 0.1$ and z = 6.5 μm, etc.

As we have said, the relation of the diffraction efficiency to the spatial frequency (the directivity pattern) does not fully characterize a thick-film hologram. We must also study the phase relationships between the components of the spectrum of spatial frequencies. However, we can easily see that the phase will not depend on the angle φ in our case. The expression for E_{-1} is purely imaginary, and hence, the phase of all the components of the spectrum is shifted in comparison with the phase of the wave E_0 by the very same angle $\pi/2$.

In conclusion, we note that the graphs of directivity patterns given here are idealized because we have used the approximate solutions, which are valid only when $\alpha \ll \cos^2\theta$. We can envision the degree of idealization by comparing the exact solutions given above with the approximate solutions. This has been done in [14].

1) Of course, one can also classify them according to other traits. For example, there are permanent-recording and re-recording holograms, but there is no need for such a classification in treating diffraction.

[1] D. Gabor, Proc. Roy. Soc. **197**, 454 (1949) (reprinted as appendix, p. 166, in G. W. Stroke, An Introduction to Coherent Optics and Holography, Academic Press, New York, 1966) (Russ. Transl., in transl. of Stroke's book, "Mir", M., 1967).
[2] Yu. N. Denisyuk, Dokl. Akad. Nauk SSSR **144**, 1275 (1962) [Sov. Phys.—Doklady **7**, 543 (1962)].
[3] E. N. Leith and J. Upatnieks, J. Opt. Soc. Am. **52**, 1123 (1962).
[4] C. G. Darwin, Phil. Mag. **27**, 315 (1914).
[5] M. Born and E. Wolf, Principles of Optics, 3rd Edn., Pergamon Press, Oxford, 1965 (Russ. Transl., "Nauka", M., 1970).
[6] B. W. Batterman and H. Cole, Revs. Mod. Phys. **36**, 681 (1964).
[7] P. P. Ewald, ibid. 37, 46 (1965) (Russ. Transl., Usp. Fiz. Nauk **89**, 287 (1966)).
[8] D. Gabor and G. W. Stroke, Proc. Roy. Soc. A304, 275 (1968).
[9] W. R. Klein, B. D. Cook, and W. G. Mayer, Acustica **15**, 67 (1965).
[10] W. R. Klein and B. D. Cook, IEEE Trans. Sonics and Ultrasonics SU-14, 123 (1967).
[11] C. G. Burckhardt, J. Opt. Soc. Am. **56**, 1502 (1966).
[12] R. L. Lamberts and C. N. Kurtz, Appl. Opt. **10**, 1342 (1971).
[13] H. A. Bohr, Almost Periodic Functions, Chelsea Publ. Co., New York, 1947 (Russ. Transl., OGIZ, M., 1934); B. M. Levitan, Pochti-periodicheskie funktsii (Almost Periodic Functions), Fizmatgiz, M., 1958.
[14] N. M. Pomerantsev, Kvantovaya élektronika, No. 5 (11), 122 (1972).
[15] H. Kogelnik, J. Opt. Soc. Am. **57**, 431 (1967).
[16] C. V. Raman and N. S. Nagendra, Nat. Proc. Indian Acad. Sci., Sec. A, **2**, 406 (I), 413 (II), **3**, 75 (III), 119 (IV), 459 (V); **4**, 222 (generalized theory)(1935).
[17] A. J. W. Sommerfeld, Optik (Optics), Dieterich'sche Verlagsbuchhandlung, Wiesbaden, 1950 (Engl. Transl. Academic Press, New York, 1954; Russ. Trans., IL, M., 1953).
[18] H. Kogelnik, Bell Syst. Tech. J. **48**, 2909 (1969).
[19] N. M. Pomerantsev, Kvantovaya élektronika, No. 5 (11), 118 (1972).

Translated by M. V. King