535.80

INCREASE OF RESOLUTION OF OPTICAL SYSTEMS BY EFFECTIVE USE OF THE DEGREES OF FREEDOM ON THE OBJECT WAVE FIELD

A. N. KOROLEV

Usp. Fiz. Nauk 96, 261–289 (October, 1968)

I. INTRODUCTION

IT is known that any optical system has a limited resolving power, which is due to the wave nature of light. According to the classical theory of Abbe and Rayleigh, this limitation is connected with the diffraction and is determined by the wavelength of the employed radiation and by the numerical aperture of the optical system. The desire to increase the resolving power therefore has led naturally to an increase of the aperture of optical systems and to a decrease of the wavelength of the employed radiation.

According to Rayleigh, the resolving power is defined as the minimum distinguishable distance between two pointlike light sources of equal intensity. It should be noted that the resolving power defined in this manner can be greatly increased, and as shown by many authors, there are no limitations in principle on the distinction between two pointlike sources of light in image space[1-4]. This is connected with the fact that if the input signal employed consists of two δ-functions, there is an a priori limitation on the dimensions of the input action. In this case the resolving power is limited not by the diffraction, but by the noise of the optical system and of the radiation receiver. A number of methods have been proposed[5-8] for distinguishing between two pointlike sources of light beyond the limit of the Rayleigh resolution of the optical system, but such an "increase of the resolving power" has no influence whatever on the limiting spatial frequency transmitted by the optical system. This is due to the fact that consideration of such a narrow class of input signals excludes the possibility of describing the operation of the optical system under consideration in the case of a broader class of input actions. We shall therefore relate the concept of "resolving power" with the limiting spatial frequency transmitted by the optical system in image space.

There are objective possibilities of increasing the resolving power beyond the classical limit by redistributing the roles of the individual parameters of the optical system during the course of image formation. These parameters, which determine the information capacity of the optical communication channel, include the field of view, the angular aperture, and the spectrum of the temporal frequencies transmitted by the system.

These capabilities can be analyzed by using the concept of the degrees of freedom of the wave field in the transmission of an image with the aid of an optical system. To consider this concept, we must dwell on certain premises from information theory in the field of image transmission with the aid of optical systems. By regarding the optical system as a communication line that transmits information concerning the object in image space, we can draw an analogy between the formation of the optical image and the transmission of electrical signals. Whereas in the electrical communication technique the signal is a function of the time, in the transmission of the optical image the "signal" is a function of the spatial coordinate; the temporal frequencies of electric signals correspond to optical spatial frequencies. This analogy comes into play most fully in the transmission of images by television. In this case, element-by-element scanning of the image causes the spatial coordinates to be transformed into temporal coordinates, while the spatial frequencies of the object determine the frequency spectrum of the electric signal.

This analogy makes it possible to consider questions involved in the formation of an image with the aid of an optical system on the basis of general concepts of information theory, which was created for electrical communication lines[9].

In information theory there is a known sampling theorem, which determines the information capacity of the employed channel in communication engineering. It states that if a temporal function $f(t)$ with limiting frequency W, which is vanishingly small outside an interval $0 \leq t \leq T$, is received by a communication channel with bandwidth from 0 to W Hz, then only 2WT data will be transmitted in this case, and the message is determined fully and rigorously in terms of the values of the function at the "sampling points" $t = n/2W$ for all values $0 \leq n \leq 2WT$.

D. Gabor formulated an analog of this theorem for the transmission of an image with the aid of an optical system:

Assume that an object plane of area larger than the square of the wavelength is bounded by a black screen (Fig. 1). We assume also that there exists a similar limitation in the aperture plane of the optical system.

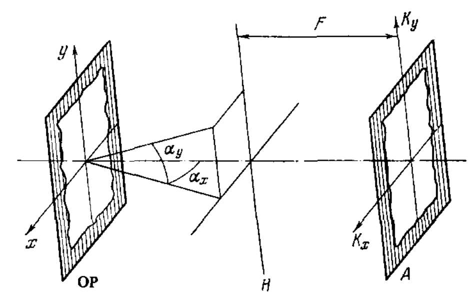

FIG. 1. Illustrating the theorem of D. Gabor. OP – object plane, H – principal plane of the lens, A – aperture plane, F – focal distance, x, y – coordinates of the object plane, α_x, α_y – aperture angles, K_x, K_y – coordinates of the Fourier plane.

Then, in the region bounded by these two black screens there exist n independent solutions of the wave equation $\Delta^2 u + (2\pi/\lambda)^2 u = 0$, with n determined by the expression

$$n = \lambda^{-2} \int\int\int\int dx\, dy\, d(\sin\alpha_x)\, d(\sin\alpha_y). \quad (1)$$

Thus, any light wave passing through the object plane and the aperture plane, can be resolved into n eigensolutions with n complex coefficients[10,11].

Let us obtain a quantitative estimate of the information capacity of an optical system from the point of view of the wave theory of light[12].

Let a plane monochromatic wave propagating in the z direction

$$u_0 = e^{2\pi i\left(\frac{z}{\lambda} - \nu t\right)}, \quad (2)$$

be incident on a two-dimensional object located in the plane $z = 0$; λ and ν are respectively the wavelength and frequency of the optical radiation. Directly behind the object, the amplitude is of the form

$$u(x, y, 0; t) = t(x, y)\, e^{-2\pi i \nu t}, \quad (3)$$

where $t(x, y)$ is the complex amplitude transmission of the object.

It is of interest to determine the amplitude of the wave for any value $z > 0$. It is known that on the basis of the Fourier theorem, any object can always be regarded as a superposition of a certain infinite system of spatial harmonics, each of which can be characterized by an orientation, spatial frequency, and amplitude.

We expand the function $t(x, y)$ in Fourier components by means of the formula

$$t(x, y) = \int_{-\infty}^{+\infty} T(f_x, f_y)\, e^{2\pi i(xf_x + yf_y)}\, df_x\, df_y. \quad (4)$$

Each component of the Fourier expansion is an infinite spatial wave with periods $1/f_x$ and $1/f_y$ in the x and y directions. The function $T(f_x, f_y)$ determine the amplitude of each of the Fourier components.

On the basis of the superposition principle, the amplitude at any point behind the object can be calculated by summing the individual waves corresponding to the transmission of each Fourier component. Owing to the phenomenon of diffraction by sinusoidal gratings whose orientation, frequency, and amplitude are determined by the Fourier expansion, the plane waves corresponding to each Fourier component propagate in directions determined by the direction cosines

$$\cos\alpha = \lambda f_x,$$
$$\cos\beta = \lambda f_y,$$
$$\cos\gamma = \lambda\left[\frac{1}{\lambda^2} - (f_x^2 + f_y^2)\right]^{1/2}. \quad (5)$$

Obviously, if $f_x^2 + f_y^2 > 1/\lambda^2$, we obtain damped waves, whose amplitudes decrease exponentially in the z direction, and vanish practically at the distance of several wavelengths. This leads to the well known conclusion that light with a wavelength λ, regardless of the circumstances, will not carry information concerning details of an object whose dimensions are smaller than $\lambda/2$.

Let us see now how information concerning an object, recorded in the light wave passing through the object, propagates when the distance z is increased. If the picture of the distribution of the illumination behand the object change very rapidly, so that the object becomes indistinguishable, the moduli of the Fourier transformation $T(f_x, f_y)$ do not change at all. This is understandable, since each point of the Fourier transform of a flat object with coordinates f_x and f_y corresponds to a definite direction of propagation of the light wave, which remains constant in free space. When $z \to \infty$, the entire similarity to the object is lost, and the distribution of the illumination becomes identical with the Fourier transform of the object.

If we place at a certain distance z from the object a lens with definite finite dimensions, then some of the information concerning the object will be lost; this lost information is contained in the Fourier components from which the light waves pass beyond the pupil of this lens. Therefore the lens transmits only those spatial frequencies for which

$$\arccos\left\{\lambda\left[\frac{1}{\lambda^2} - (f_x^2 + f_y)^2\right]^{1/2}\right\} \leqslant \frac{\alpha}{2}, \quad (6)$$

where α is the aperture angle of the lens.

The number of degrees of freedom of the wave field in the transmission of an image with the aid of an optical system can be defined as the number of independent really existing parameters necessary for a complete description of the wave field forming the image of the object.

For simplicity let us calculate the number N of the degrees of freedom of the wave field in the image space for an optical system with a rectangular aperture, assuming that the image surface is a rectangle with dimensions L_x and L_y[13]. For the calculation, it is convenient to make the following assumption: let the distribution of the amplitudes in the object plane be repeated with periodicity L_x and L_y in the directions x and y. This assumption should not influence the result, since obviously no information is added thereby. The spectrum of the spatial frequencies of the object then becomes discrete and the following spatial frequencies are conserved:

$$k_x = \frac{2\pi n_x}{L_x}, \quad k_y = \frac{2\pi n_y}{L_y}, \quad n_x, n_y = 0, \pm 1, \ldots \quad (7)$$

Therefore the number of degrees of freedom of the field of a monochromatic light wave, with allowance for the aperture of the system, will equal

$$\left(1 + L_x \frac{k_x'}{\pi}\right)\left(1 + L_y \frac{k_y'}{\pi}\right), \quad (8)$$

where k_x' and k_y' are the limiting spatial frequencies transmitted by the optical system in the directions x and y, respectively. In addition, for each degree of freedom of the wave field there are two independent states of polarization, which can be used for image transmission.

If the object is nonstationary, i.e., varying in time or moving, then it is necessary to consider also properties of the light. In this case each spatial degree of freedom of the wave field can be regarded as a separate independent temporal communication line, which has

$$N_t = 2(1 + \Delta\nu\Delta t) \quad (9)$$

degrees of freedom in the frequency interval $\Delta\nu$ in an observation time Δt (see the sampling theorem, p. 712).

For the field of a monochromatic wave $N_t = 2$, since two independent parameters suffice to determine the amplitude and the phase of the wave.

Thus, the total number of degrees of freedom of the wave field in the formation of an image with the aid of an optical system can be written in the form

$$N = 2N_t\left(1 + L_x \frac{k_x'}{\pi}\right)\left(1 + L_y \frac{k_y'}{\pi}\right). \tag{10}$$

For $L_x k_x' \gg 1$ and $L_y k_y' \gg 1$, as happens in most cases,

$$N \approx 2N_t L_x L_y \frac{k_x' k_y'}{\pi^2} = 2N_t SW, \tag{11}$$

where $W = k_x' k_y'/\pi^2$ is the band width of the spatial frequencies transmitted by the system and $S = L_x L_y$ is the area of the object.

The quantity SW, which is the product of the area of the object by the maximum spatial frequency, obviously determines the number of spatial degrees of freedom of the wave field of the object which are used for the formation of the image (cf. Gabor's theorem, p. 712).

Inasmuch as in the formation of the image the wave field is the carrier of the information, the following statement holds true: the only invariant for the given system is the number N of the degrees of freedom of the wave field (N = const), and not the transmission bandwidth of the spatial frequencies.

Thus, it becomes possible to exceed the limiting transmitted spatial frequency f' for a given system by effectively using the degrees of freedom of the wave field of the object and suitably decreasing one of the other factors in formula (11). This premise is the basis for the image-formation methods that will be considered below.

In accordance with the concept of the degree of freedom of the wave field as the fundamental invariant of the optical system, it is possible to do the following:

1) Increase the bandwidth of the transmitted spatial frequencies by decreasing the bandwidth of the transmitted temporal frequencies (in the limiting case, at zero bandwidth of the spatial frequencies it is possible to use the temporal frequencies for the transmission of spatial information).

2) Increase the bandwidth in the x direction and suitably decrease the bandwidth in the y direction, in such a way that the initial two-dimensional band of the spatial frequencies W remains unchanged.

3) Increase the bandwidth of the transmitted spatial frequencies by decreasing the useful field of the object.

4) Double the transmission bandwidth of the optical system, using two independent states of the polarization of the light waves for the formation of the image.

To realize these possibilities, a number of original optical systems have been proposed and realized, in which the image can be formed and the classical resolving-power limit exceeded. A feature common to all these systems is that special devices, splitting the light beams, are placed in conjugate planes of the object space and of the image space (Fig. 2). The device M_1 in image space splits the incident light beam into several coherent beams, which illuminate the object at different angles of incidence. When these beams pass through the object, the light beam is diffracted at each

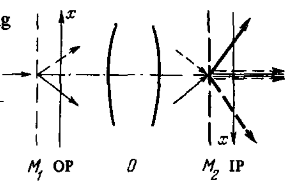

FIG. 2. Generalized scheme explaining the principle of increasing the revolving power of optical systems by using a device splitting the light beams. M_1 and M_2 – splitting devices installed in the object and image spaces, OP – object plane, IP – image plane, O – lens.

spatial frequency of the object. Thus, several waves corresponding to the same component of the spatial frequency of the object, appear in the image space of the system. But in an optical system in which the property of invariance in space is satisfied, there should be only one wave corresponding to one spatial frequency of the object. Therefore the device M_2, installed in the image space, acts on all the waves corresponding to one component of the spatial frequency in such a way, as to combine them in one direction. This, however, gives rise to waves with other undesirable directions. The elimination or neutralization of these waves is the main problem in the realization of such systems. Once this problem is solved, it is possible to obtain a spatially-invariant image of the object.

We shall describe later certain variants of devices which make it possible to increase the resolving power of optical systems in excess of the classical limit by effective utilization of the degrees of freedom of the object wave field. In addition, at the end of the article we shall consider methods of obtaining spatial information by using only temporal degrees of freedom.

II. INCREASE OF SPATIAL RESOLUTION BY REDUCING THE TEMPORAL RESOLUTION

According to the invariance theorem, the bandwidth of the spatial frequencies of an optical system can be increased from W to $\hat{W} > W$ by decreasing the bandwidth of the temporal frequencies from $\Delta\nu$ to $\Delta\hat{\nu} < \Delta\nu$ with $\hat{W}\Delta\hat{\nu} = W\Delta\nu$, all other components of N remaining unchanged.

The main idea of this method is to transmit in different temporal frequency bands of width $\Delta\hat{\nu}$, lying inside the band $\Delta\nu$ of the ordinary system, information concerning different spatial frequency bands. This is performed by illuminating the object by coherent waves that are incident on the object at different angles that have different temporal frequencies. Such waves are obtained when one wave is diffracted by a moving grating or by an ultrasonic traveling wave in a liquid or a solid (the Brillouin effect).

In[14,15] is described an optical system based on this principle, which makes it possible to increase the resolving power in one direction. This system employs optically conjugate gratings, located in the object plane and the image plane and moving in phase in a direction perpendicular to the lines of the grating. Without stopping in detail on this particular case, let us consider a more general variant in which two-dimensional gratings are used, making it possible to increase the resolving power of an optical system in two directions[16].

Assume that a plane light wave with a radiation frequency ν_0 is normally incident on a two-dimensional grating M_1 with constants d_x and d_y in the x and y directions. The amplitude transmission of the grating is

$$M_1(x, y) = \sum_{j,\, l=0,\, \pm 1,\, \ldots} m_{j,\, l} e^{2\pi i\left(\frac{jx}{d_x} + \frac{ly}{d_y}\right)}. \tag{12}$$

The grating M_1 moves with constant velocity $v(v_x, v_y)$ in the x, y plane. The amplitude of the wave directly behind the grating M_1 is

$$e^{-2\pi i \nu_0 t} M_1(x - v_x t,\, y - v_y t) =$$
$$= \sum_{j,\, l=0,\, \pm 1,\, \ldots} m_{j,\, l} \exp\left\{2\pi i\left[\frac{jv_x}{d_x} + \frac{lv_y}{d_y} - \left(\nu_0 + \frac{jv_x}{d_x} + \frac{lv_y}{d_y}\right)\right]\right\}. \tag{13}$$

Equation (13) shows that the initial light wave, after being diffracted by the spatial structure of the grating with frequencies j/d_x and l/d_y, will have wave-vector components

$$k_x^{(s)} = 2\pi \frac{j}{d_x}, \quad k_y^{(s)} = 2\pi \frac{l}{d_y} \tag{14}$$

and temporal frequencies

$$\nu_{j,\, l} = \nu_0 + \frac{jv_x}{d_x} + \frac{lv_y}{d_y}. \tag{15}$$

As is well known, the angle of incidence of the wave illuminating the object determines the band width of the spatial frequencies passing through the optical system. Therefore, by obtaining coherent waves with different wave-vector components and with different temporal frequencies, it becomes possible to transmit information concerning different bands of spatial frequencies with the aid of different bands of temporal frequencies. In order for these individual temporal-frequency bands not to become superimposed, they must differ from one another at least by $\Delta\hat{\nu}$. This condition determines the vector velocity of the grating $v(v_x, v_y)$:

$$\frac{v_x}{d_x} + \frac{v_y}{d_y} \gg \Delta\hat{\nu}, \tag{16}$$

and the temporal-frequency band width $\Delta\nu$ should be sufficient to transmit the temporal changes of the object.

The device for realizing this method (Fig. 3) consists of a grating M_1 moving with velocity v in the plane between the object and the optical system (in this particular case - in the object plane), and a grating M_2 moving in a conjugate plane of the image space with conjugate velocity v′. The constants of the gratings M_1 and M_2 should also be conjugate, and the lines of both gratings should be parallel to one another.

Assume that the described system produces an image of an object with spatial frequencies f_x and f_y and with a proper temporal frequency ν (for a stationary object, $\nu = 0$). The action of the grating M_1 on the initial light wave is equivalent to illumination of the object by coherent waves with wave-vector components $k_x^{(S)}$, k_y^S ($j, l = 0, \pm 1, \ldots$). After the interaction with the spatial frequencies of the object, the light waves will have wave-vector components $2\pi(f_x + j/d_x)$ and $2\pi(f_y + l/d_y)$, and a temporal frequency $\nu + \nu_0 + jv_x/d_x + lv_y/d_y$.

That part of these waves which is transmitted by the optical system with finite angular aperture, is diffracted in the image space at the spatial frequencies j'/d_x and l'/d_y of the grating M_2. This yields light waves with wave-vector components

$$2\pi\left[f_x + \frac{j+j'}{d_x}\right], \quad 2\pi\left[f_y + \frac{l+l'}{d_y}\right] \tag{17}$$

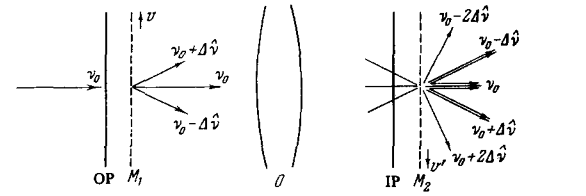

FIG. 3. Optical scheme of the method with synchronously moving masks. M_1 and M_2 – optical conjugate gratings in the object and image spaces, OP – object plane, IP – image plane, O – lens, v, v′– velocities of masks M_1 and M_2, respectively.

and with a temporal frequency

$$\nu + \nu_0 + \frac{(j+j')\,v_x}{d_x} + \frac{(l+l')\,v_y}{d_y}.$$

(It is assumed that the magnification between the object plane and the image plane is equal to unity.)

In order for the obtained image to be spatially invariant, it is necessary to use for the image formation only the wave-vector component with $j' = -j$ and $l' = -l$. In this case, only one wave with components $2\pi f_x$, $2\pi f_y$ and with temporal frequency $\nu + \nu_0$ will correspond in image space to the object under consideration.

The separation of the necessary component of the wave vector from the aggregate (17) can be realized with the aid of a temporal frequency filter with a transmission band $\Delta\hat{\nu}$, tuned to a frequency $\nu + \nu_0$. However, small shifts of the temporal frequency ν_0, obtained during the motion of the gratings, make it necessary to use a narrow band $\Delta\hat{\nu}$. This makes the use of temporal frequency filters undesirable in practice. Therefore image formation with the aid of a temporal frequency filter can be replaced by image formation with a time constant $\tau = (\Delta\hat{\nu})^{-1}$.

It is of interest to consider the transfer function and the scattering function of such a modified system and compare it with analogous functions of an ordinary optical system. If we use for the purpose of obtaining the superresolution effect gratings with an amplitude transmission:

$$\left.\begin{aligned} M_1(x, y) &= \sum_{j,\, l=0,\, \pm 1,\, \ldots} m_{j,\, l} e^{2\pi i\left(\frac{jx}{d_x} + \frac{lx}{d_y}\right)}, \\ M_2(x, y) &= \sum_{j,\, l=0,\, \pm 1} m'_{j,\, l} e^{2\pi i\left(\frac{jx}{d_x} + \frac{lx}{d_y}\right)}, \end{aligned}\right\} \tag{18}$$

then we can show that temporal frequency filtration of the output signal in accordance with the indicated conditions yields a spatially-invariant system with a transfer function

$$D(f_x, f_y) = \sum_{j,\, l=0,\, \pm 1,\, \ldots} \hat{m}_{j,\, l} D_L\left(f_x - \frac{j}{d_x},\, f_y - \frac{l}{d_y}\right), \tag{19}$$

where $\hat{m}_{j,l} = m_{j,l} m_{-j,-l}$ and $D_L(f_x, f_y)$ is the transfer function of the lens. The scattering function is determined by the expression

$$\hat{F}(x, y) = F(x, y)\,\hat{M}(x, y), \tag{20}$$

where $F(x, y)$ is the scattering function of the ordinary system and

$$\hat{M}(x, y) = \sum_{j,\, l=0,\, \pm 1,\, \ldots} \hat{m}_{j,\, l} \exp\left[2\pi i\left(\frac{jx}{d_x} + \frac{l_y}{d_y}\right)\right] \tag{21}$$

is the cross-modulation function of the amplitude transmission of the gratings M_1 and M_2.

In the subsequent analysis of the various methods of increasing the resolving power, we shall not stop to discuss the transfer function and the scattering function, since they are similar in character.

III. INCREASE OF THE BANDWIDTH OF THE SPATIAL FREQUENCIES IN THE DIRECTION BY A SUITABLE DECREASE IN THE y DIRECTION

This method can be effectively used for objects whose spatial-frequency spectrum has a large width in the x direction and a small width in the y direction (extreme case - one-dimensional object)[13,17]. It is obvious that the high frequencies of such an object, in the x direction beyond the limit of the bandwidth of the system will not be transmitted or reproduced in the image, whereas the frequency spectrum in the y direction will not fill the band width of the optical system (Fig. 4).

It is necessary to transform the spectrum of the spatial frequencies so as to make it suitable for the transmission of the spatial frequencies of the optical system. After passing through the optical system, the spectrum of the spatial frequencies should be again reconstructed in the initial form, so as to obtain the spatially-invariant image of the object.

For simplicity let us consider a one-dimensional object with amplitude transmission

$$T(x) = \cos 2\pi f x, \tag{22}$$

which is illuminated by a normally incident monochromatic wave. The frequency f is not transmitted by the optical system, since $f > f'$, where f' is the limiting spatial frequency transmitted by the optical system.

In the object plane, we place a mask M_1, comprising a grating with a cosinusoidal transmission distribution, turned through a small angle φ relative to the y axis. Its transmission can be written in the form

$$M_1(x, y) = \cos [2\pi f_0 (x \cos \varphi + y \sin \varphi)]. \tag{23}$$

The object and the mask together form a joint object with a complicated transmission

$$u(x, y) = \frac{1}{2} \cos \{2\pi [x (f + f_0 \cos \varphi) + y f_0 \sin \varphi]\} + \frac{1}{2} \cos \{2\pi [x (f - f_0 \cos \varphi) - y f_0 \sin \varphi]\}. \tag{24}$$

Such a superposition of the object and the mask produces moire strips with spatial frequencies $f + f_0 \cos \varphi$, $f - f_0 \cos \varphi$, and $f_0 \sin \varphi$. At definite values of f_0 and φ, the spatial frequency $f - f_0 \cos \varphi$ can be transmitted by the optical system. In this case we obtain in the image space an illumination distribution

$$u'(x, y) = \frac{1}{2} \cos \{2\pi [x (f - f_0 \cos \varphi) - y f_0 \sin \varphi]\}. \tag{25}$$

To separate the spatial frequency f of the object, an identical mask M_2, optically conjugate with the mask M_1, is placed in the image plane, and produces (for the case $\Gamma = 1\times$)

$$T'(x, y) = \frac{1}{4} \cos 2\pi f x + \frac{1}{4} \cos \{2\pi [x (f - 2f_0 \cos \varphi) - 2y f_0 \sin \varphi]\}. \tag{26}$$

In Eq. (26), the term $(1/4) \cos 2\pi f x$, which does not depend on the coordinate y, represents the image of the initial object, whose spatial frequency f lies beyond the limit of the band width of the optical system. To attain similarity between the object and the image, it is necessary to destroy the terms that depend on y. This can be realized with the aid of a vertical cylindrical lens or by spatial filtration of the spectrum of the image frequencies with the aid of a horizontal slit filter.

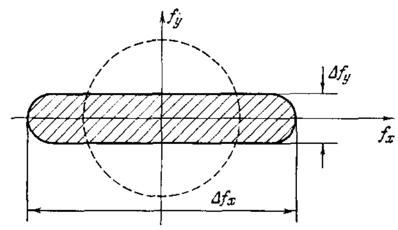

FIG. 4. Spectrum of spatial frequencies transmitted by the lens, and frequency spectrum of the employed object. Δf_x and Δf_y – widths of the object spectrum in the x and y directions.

Let us consider the transformation of spatial-frequency spectra in all stages of the process of image formation in the described method. These transformations are best observed with the aid of the scheme shown in Fig. 5.

If a one-dimensional aperiodic object is located in the plane O, then the corresponding spatial-frequency spectrum, shown in Fig. 6a, is obtained in the Fourier plane A. The periodic mask M_1 with a cosinusoidal transmission distribution will be represented in the Fourier plane A by two points corresponding to its spatial frequency and orientation, and by a point corresponding to the zero frequency (Fig. 6b). When the object and mask are superimposed, a complicated spectrum is obtained, constituting the result of the action of the mask on the object (Fig. 6c). The aperture of the lens (assumed for convenience to be square) (Fig. 6d) transmits only part of the obtained complicated spectrum (Fig. 6e) and, as can be seen from the figure, this transmitted part contains all the frequencies of the object, although their relative position is somewhat disturbed. The optical system L_1L_2 transfers the frequency distribution in the Fourier plane A into the plane B. The mask M_2, which is identical with M_1 and is mounted in the image plane, deforms the spectrum of the spatial frequencies. This deformation of the spectrum has the same character as the result of the superposition of the object and the mask M_1 (Fig. 6f). This results in a reconstructed spectrum of the object with certain additions to the spectrum, located in the direction f_y, which distort the image of the object. They can be illuminated with the aid of a

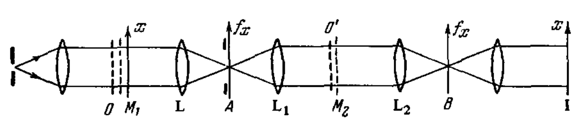

FIG. 5. Optical scheme demonstrating the transformation of the spectrum of the spatial frequencies of an object. M_1 and M_2 – optically conjugated gratings in the object and image spaces, O – object plane, I – image plane, L – lens, A – Fourier plane of lens L, L_1, L_2 – optical system that transfers the frequency spectrum from plane A to plane B.

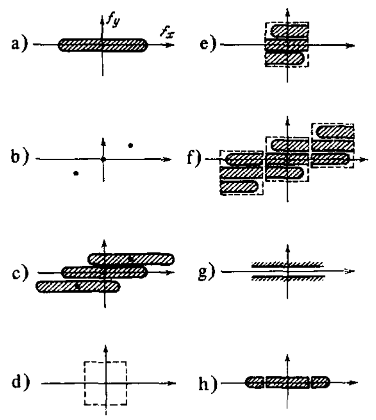

FIG. 6. Spatial-frequency spectra: a) of the object, b) of the mask M_1, c) of the effective object consisting of O and M_1, d) transmission region of the lens L, e) frequency spectrum transmitted by the lens L, f) spectrum obtained as the result of interaction between the intermediate image of O and the mask M_2, g) transmission spectrum of spatial-frequency filter, h) spectrum of final image.

narrow slit-type frequency filter whose transmission is shown in Fig. 6g, and which moves in the plane B. This filter will pass only the frequencies (Fig. 6h) corresponding to the spectrum of the object. As a result, a spatially-invariant image of the object O, having spatial frequencies hitherto lying beyond the transmission limit of the optical system, is obtained in the I plane.

Thus, if the spectrum of the object fills, for example, not more than 1/k of the frequency region of the lens in the f_y direction, it is possible to obtain an image of details of the object with spatial frequency f_x, exceeding by a factor k the limiting frequency transmitted by the lens. However, the dimension of the frequency region of the obtained image $(\Delta f_x \cdot \Delta f_y)$ can never exceed the frequency region $(\Delta f'_x \cdot \Delta f'_y)$ transmitted by the lens. This is clearly seen in Fig. 6e.

IV. INCREASING THE RESOLVING POWER OF OPTICAL SYSTEMS BY DECREASING THE OBJECT FIELD[13,18]

The optical scheme of a setup realizing this method is shown in Fig. 7. Both masks, M_1 and M_2, are optically conjugate lattices; mask M_1 is placed in the object space, and mask M_2 in the image space, but not in the object plane and in the image plane. It is assumed that the masks are line gratings, which produce only two diffraction orders for transmitted light. (We have in mind phase or amplitude-phase gratings, which produce no zeroth-order diffraction.)

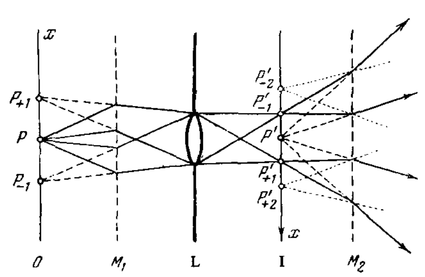

FIG. 7. Diagram for method of increasing the resolving power of an optical system by reducing the field of the object. M_1 and M_2, – optically conjugate gratings in the object and the image spaces, O – object plane, I – image plane, L – lens.

A monochromatic spherical wave emitted from the point P in the object plane is split by the mask M_1 into two spherical waves, which seem to emerge from the points P_{+1} and P_{-1}, and their lateral shift from the point P is equal to

$$\pm \Delta x = z_0 \frac{\lambda}{d}, \tag{27}$$

where z_0 is the distance between the plane of the object O and the mask M_1, d is the constant of the grating M_1, and λ is the wavelength of the employed radiation.

The points P_{+1} and P_{-1} are transmitted by the optical system to the image plane as respectively P'_{+1} and P'_{-1}. The spherical waves emerging from P'_{+1} and P'_{-1} split as a result of diffraction by the mask M_2 into corresponding pairs of spherical waves, seemingly emerging from the points P', P'_{+2} and P', P'_{-2}. As seen from Fig. 7, the image of the point P is obtained in this case with an angular aperture in the x direction, exceeding the angular aperture of the optical system without the masks. The angular aperture in the y direction remains unchanged.

Let us determine the effective angular aperture of the system using masks M_1 and M_2 having two orders of diffraction.

For the optical system without the masks, the angular aperture is determined obviously by the quantity

$$\sin\frac{\theta}{2} = \frac{D}{2a}, \tag{28}$$

where θ' is the angular aperture of the lens, a is the distance from the object to the lens, and D is the diameter of the lens.

When the mask M_1 with a grating constant d is used, the angular aperture of the system can be determined by starting from the conditions for the diffraction of light in oblique incidence:

$$\sin\frac{\theta'}{2} - \sin\varphi = \frac{\lambda}{d}, \tag{29}$$

where θ' is the effective angular aperture of the system, φ is the diffraction angle of light incident on the mask M_1 at an angle $\theta'/2$; hence

$$\sin\frac{\theta'}{2} = \frac{D}{2a} + \frac{\lambda}{d}\left(1 - \frac{z_0}{a}\right). \tag{30}$$

Thus, when masks M_1 and M_2 are used we obtain a magnification of the aperture of the system, determined by the term $(\lambda/d)[1 - (z_0/a)]$ of Eq. (30). Obviously, we can double the angular aperture of the optical system and reproduce in the image space object frequencies from zero to $2f'_x$, where f'_x is the limiting spatial frequency transmitted by the optical system in the x direction when used in the ordinary manner. This corresponds to a relation

$$\frac{\lambda}{d}(a - z_0) = \frac{D}{2}. \tag{31}$$

If the relation

$$\frac{\lambda}{d}(a - z_0) > \frac{D}{2} \tag{32}$$

holds, then the light rays emerging from the object in a direction of the optical axis or at small angles to it do not enter the optical system after being diffracted by the mask M_1. As a result, information is lost concerning the lower spatial frequencies of the object, corresponding to these rays, but the optical system will transmit higher spatial frequencies, exceeding $2f'_x$.

It should be noted that when the aperture of an optical system is doubled by this method, it is necessary to decrease by one half the useful area of the object that can be transferred without damage to the image plane. This is connected with the fact that when spherical waves from points P'_{+1} and P'_{-1} are diffracted by the mask M_2, there appear, besides the point P', also two additional images of the point P, namely P'_{+2} and P'_{-2}. If the object occupies the zone $L_x > 2\Delta x$, then the described optical system cannot transmit it completely to the image space; to obtain the image one can use only zones of the object with width $2\Delta x$ (Δx from Eq. (25)) parallel to the y axis. These zones of the object should be separated by non-transparent bands, the transmission of which is equal to zero and whose width is also equal to $2\Delta x$. In this case, the disturbing points P'_{+2} and P'_{-2} will be located in the zone of the dark bands and will not distort the image of the object. This means, that, in accordance with the concept of the invariance of the number of degrees of freedom of the wave field for the given optical system, when the bandwidth of the spatial frequencies is doubled it is necessary to decrease the useful area of the object by one-half.

If line gratings producing more than two diffraction orders are used in the system, or else if crossed gratings are used, then the more general theory, which was considered in[18], must be used to describe the operation of such a system. Obviously, when gratings with n diffraction orders are used, it is possible to increase the bandwidth of the spatial frequencies by a factor n, but then it is necessary to decrease the useful field of the object to 1/n of the initial value. In this case the screen covering the object will have opaque bands that are n times broader than the transparent bands. When crossed gratings are used, it is necessary to use a suitable screen with crossed opaque bands. As a result the decrease of the useful field of the object is equivalent to the broadening of the two-dimensional transmission band of the spatial frequencies.

V. DOUBLING OF THE RESOLVING POWER OF OPTICAL SYSTEMS BY USING LIGHT WAVES WITH TWO POLARIZATION DIRECTIONS TO FORM THE IMAGE

As indicated in the generalized analysis of the described method, a common feature of all such systems is the use of beam-splitting devices placed in conjugate planes of the object and image spaces. We have considered so far image-forming methods in which such devices are amplitude, phase, and amplitude-phase gratings. In this case the effective aperture of the optical system is increased as a result of the diffraction of the light by these gratings.

One can use for this purpose also birefringent prisms (Rochon, Senarmont, Wollaston, etc.). In this case the increase of the effective aperture of the system is connected with the change of the direction of the light for two mutually perpendicular components of the polarization after the passage through such prisms[19-22].

If the objects do not have birefringence of dichroism, then the change of the amplitude and phase of the light on passing through the object does not depend on the state of polarization of the incident light. For these objects, a single polarization component suffices to produce the image; the other polarization component can be used to transmit to the image space such object parameters which do not pass through the optical system under normal conditions. In this case the two mutually perpendicular components represent two independent information carriers.

A simple scheme illustrating the gist of this method is shown in Fig. 8. In conjugate planes of the image and object space are placed Wollaston prisms which, as is well known, deflect two mutually perpendicular components of polarized light symmetrically away from the optical axis. The action of the Wollaston prism W_1 on the illuminating radiation causes the object to be illuminated, as it were, by light from two sources equally shifted in opposite directions away from the employed source. The polarizer P, installed ahead of the prism, is oriented at 45° to the directions of the oscillations separated by the prism, making it possible to obtain two illuminating light beams of equal intensity. The radiation from both virtual sources of light is polarized in mutually perpendicular planes.

As a result, the illumination of the object at normal incidence of the light wave is replaced, as it were, by oblique illumination from two sides. Such an illumination, from two coherent sources, is frequently used to increase the resolving power when working with an ordinary microscope. There, however, a system of interference fringes from the zeroth diffraction orders is observed for both sources, and has no bearing on the object. This shortcoming is eliminated in the described method with the aid of a second Wollaston prism, W_2, placed ahead of the image. The prism W_2 again returns the light beams of the two polarization components to the common direction. This means, that both spectra of the spatial frequencies of the object, which occupy in the Fourier plane two different positions, again are brought to one place. An analyzer A, oriented at an angle 45° to the polarization directions separated by the prism W_2, is used to enable both parts of this aggregate spectrum, which are polarized in mutually perpendicular directions, to interfere with each other and to produce a single image. If the splitting of the illuminating rays by the prism W_1 is equal to the aperture angle of the optical system, then a doubling

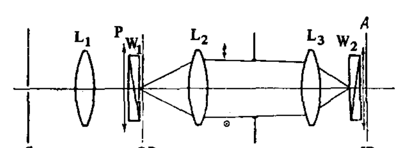

FIG. 8. Diagram of method in which berefringent prisms are used to split the light beams. S – light source, OP – object plane, IP – image plane, L_1 – lens, L_2L_3 – optical system transferring the image from OP to IP. W_1 and W_2 – optically conjugate Wollaston prisms, installed in the object and image spaces, P – polarizer, A – analyzer.

of the resolving power of the optical system is obtained compared with the usual coherent illumination at normal incidence.

VI. INCREASING THE RESOLVING POWER BY USING ADDITIONAL PARAMETERS, IN THE FORM OF THE WAVELENGTH OF THE LIGHT RADIATION OR OF THE PHASE OF THE WAVE, TO TRANSMIT THE SPATIAL INFORMATION

If the diffraction spot in the image plane corresponds to a certain finite section A in the object plane, then no information whatever can be obtained in an ordinary system concerning the elements of the object belonging to this section.

However, as shown in[23], it is possible to resolve the individual elements of the object in the region of the diffraction spot by artificially imparting to each element a certain additional parameter, that distinguishes it from another neighboring element. Such a parameter can be the wavelength of the light radiation or the phase of the wave.

The gist of the method is that the object is illuminated by a wave field that has in the object plane a definite radiation frequency distribution or else a wave field with a fixed ratio of the phases of the illuminating beams in the object plane. In the former case the illumination of the object can be effected with the aid of dispersion systems that resolve the optical radiation with respect to wavelength. For the latter variant one could use interference devices which make it possible to obtain a spatial phase gradient of the optical radiation. Such additional parameters, which are spatially connected with the object plane, constitute degrees of freedom of the wave field, with the aid of which it is possible to transmit spatial information. The general invariance theorem does not reflect the informational significance of these degrees of freedom in explicit form, and obviously must be expanded. The quantitative increase of the information capacity of the optical communication channel when these additional parameters are used will be considered below in the concrete discussion of the methods.

The process of forming the image with the use of the wavelength or the phase can be explained in the following manner. Assume that when the object is illuminated, each of its elements within the confines of a section A corresponds to a definite value of the wavelength λ (or phase φ). The aggregate of the elements of the object in the x-axis direction can be regarded as a certain distribution of illumination $I = f_1(x)$. The use of a dispersion system for the illumination of the object introduces an additional condition $\lambda = f_2(x)$. If the dispersion region of the illuminating system coincides in the object plane with the section A, then each value of λ will correspond to a definite value of the illumination $I = F(\lambda)$. This distribution of the illumination is a function of the wavelength and is independent of the coordinate. Using in image space a second dispersion system, for which $\lambda = f_3(x')$, where x' is the coordinate in the image plane, we obtain at $I = F(\lambda)$ a distribution of the illumination in the image of the section A in the form $I = f_4(x')$.

Let us consider as image-forming system using dispersion systems that effect wavelength resolution

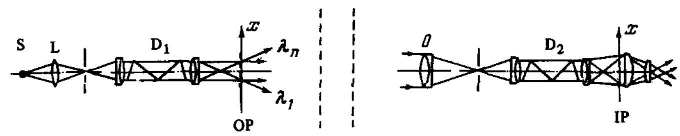

FIG. 9. Diagram of an image-forming method using dispersion systems that resolve the optical radiation with respect to wavelength. S – light source with broad radiation spectrum, L – illuminating lens, D_1 and D_2 – dispersion systems, OP – object plane, IP – image plane, O – objective.

of the optical radiation. Figure 9 shows the optical diagram of such a system, which can be used to form images of objects with one-dimensional optical-density distribution. The broad continuous spectrum of the radiation source, obtained with the aid of a dispersion system D_1 (any spectral instrument that produces a flat spectrum with a definite dispersion and resolution) is projected in the object plane OP. To each zone of the one-dimensional object in the x direction, corresponding to the direction of the dispersion system, there corresponds a definite value of the wavelength λ. The object transforms the distribution of the illumination in the spectrum, changing the intensity of the radiation for each wavelength in accordance with the distribution of the optical density in the x direction. To reproduce the image of the object, it not sufficient to obtain the spectrum of the optical radiation passing through the object, and to register it with the aid of a nonselective receiver. It is possible to use for the transmission of the optical radiation an optical system with a spatial resolution equal to zero, e.g., a light pipe.

The resolving power of the described method is determined only by the linear dispersion of the employed dispersion systems and by their spectral resolving power. Therefore, for the given case we can determine the number of spatial degrees of freedom in the form

$$N_x = \frac{\Delta\lambda}{\delta\lambda}, \tag{33}$$

where $\Delta\lambda$ is the width of the employed spectrum of the optical radiation and $\delta\lambda$ is the minimum resolved spectral interval.

Obviously, it is impossible to approach such a method from the usual point of view. In the language of communication theory, we can say that the information concerning each element of the object is transmitted through a separate frequency channel, and the process of the image transmission reduces to frequency modulation in the object space and to frequency demodulation in the image space. In this connection, the system has interesting advantages: first, the diffraction does not impose any limitations on the resolving power, and second, phase distortions of the light waves behind the object do not disturb the image-formation process. Therefore, without affecting the image transmission, it is possible to place between the object and the receiving system even an inhomogeneous medium or a diffusely-scattering medium, in which the scattering does not depend on the wavelength, provided only that a light flux strong enough to be registered reaches the receiving system.

We now consider another method, in which phase dispersion systems are used in the object and image planes. In this case the object is illuminated in such a way, that the object plane coincides with the interference field, in which there is a certain gradient of path difference (or phase difference) in a direction coinciding with the direction of variation of the optical density in the one-dimensional object.

Figure 10 shows an optical system realizing this method. The phase dispersion systems employed are Fabry-Perot interference wedges; the illumination is with white light. Under a suitable choice of the direction and angle of the wedge W_1, each zone of the object in the x direction is illuminated by light rays with a definite path difference Δl, determined by the thickness of the corresponding zone of the interference wedge. The path differences of the rays, defined as

$$\Delta\varphi = \frac{\Delta l}{\lambda} \tag{34}$$

(where λ is the wavelength of the light), is periodic in the x direction, reaching equal values in those places where the path difference in the wedge changes by an integer number λ. The linear pitch of the phase variation is

$$l = \frac{\lambda}{2}\gamma\Gamma, \tag{35}$$

where γ is the angle of the wedge W_1 and Γ is the magnification of the optical system L_1L_2.

Let an object illuminated in this way be projected by an optical system O_1O_2 in the image plane. To prevent two neighboring interference orders from falling in the same diffraction maximum, it is necessary that the linear resolving power of this system be equal to the interval l in the object space. Then the sections of the object illuminated by light beams with equal phase differences but with path differences differing by not less than λ will fall in different diffraction maxima.

To determine the distribution of the illumination within the limits of the diffraction spot in the image of the object, i.e., beyond the limits of the resolving power of the optical system, it is sufficient to install in the image plane an interference wedge K_2. If the corresponding sections of the wedges W_1 and W_2 producing the different beam path differences are conjugate with each other, it is possible to observe the interference in white light. Since the image of each individual element (zone of the one-dimensional object) is now determined by the width of the interference maximum and not by the diffraction maximum, this increases the resolving power of the system by a factor $l'/\Delta p$, where l' is the half-width of the diffraction maximum and Δp is the width of the interference maximum (Fig. 11). The width of the latter, as is well known, is determined by the reflection coefficient R of the mirror coatings on the wedge surfaces. Since the ratio $l'/\Delta p$ can reach 10–20 and more, this produces a corresponding increase of the resolving power by 10–20 times compared with the usual systems.

For this system, we obtain for the number of spatial degrees of freedom

$$N_x = (N_x)_L (N_x)_{ph} = (N_x)_L \frac{\pi}{\delta\varphi}, \tag{36}$$

where $(N_x)_L$ is the number of spatial degrees of freedom for the employed optical system, $(N_x)_{ph}$ are the additional degrees of freedom connected with the use of the phase dispersion systems, and $\delta\varphi = (1 - R)/\sqrt{R}$ is the width of the interference maximum expressed in terms of the phase. The quantity $\pi/\delta\varphi$ has a definite physical meaning and corresponds, in the theory of multi-ray interference, to the concept of the "number of effective beams."

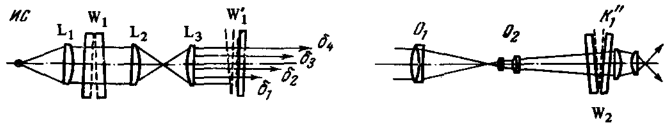

FIG. 10. Image-formation method using optical conjugate phase dispersion systems. S – light source with broad continuous radiation spectrum, L_1 – illuminating lens, W_1, W_1', and W_1'',– Fabry-Perot interference wedge and its images, L_2L_3 – optical system projecting the image of wedge W_1 in the object plane, O_1O_2 – optical system forming the image of the object, K_2 Fabry-Perot interference wedge located in the image plane.

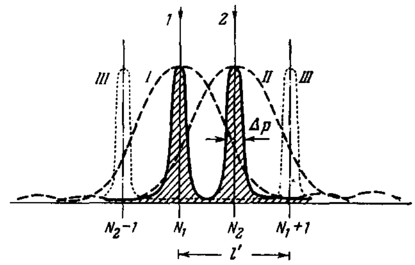

FIG. 11. Diffraction distribution (curves I and II) and distribution of illumination in the interference pattern for multiple reflections (curve III). l' – half-width of diffraction distribution and distance between neighboring interference orders, Δ_p – half-width of interference distribution.

Let us see now the results of such an increase of the resolving power of the system in the x direction. To this end, we must analyze the change of the illumination of the object when the interference wedge W_1 is introduced. Let us assume that the interference wedge W_1 is illuminated by a parallel beam of white light. If a plane wave front is incident on the interference wedge, then, as is well known, multiple reflections from the wedge mirrors produce ultimately a "set" of wave fronts turned through an angle $2\gamma n$ relative to the incident beam, where γ is the wedge angle and n is the serial number of the reflection of the incident light wave from the interference-wedge mirrors. The light waves obtained as a result of the reflections are coherent, are shifted in phase, and decrease in intensity. The illumination of the object by such coherent light waves, directed at different angles to the object plane, increases the band width of the spatial frequencies transmitted by the system in the x direction. The phase differences of the illuminating beam are compensated for with the aid of wedge W_2.

It should be noted that the "set" of wave fronts obtained with the interference wedge W_1 and illuminating the object, occupies a certain space along the z axis, determined approximately by

$$\Delta z = 2tN_{eff}, \tag{37}$$

where t is the thickness of the wedge W_1, N_{eff} is the number of effective beams for the mirrors of wedge

W_1. Since the light waves determined by these fronts are coherent, it is possible in principle to distinguish in this case between object elements located in the z direction within the limits of the zone Δz.

We must mention here that when we determined the number of degrees of freedom of the wave field in the case of transmission of an object with the aid of an optical system, we referred to transmission of the images of two-dimensional objects. However, an optical system delivers to the image space three-dimensional information concerning the object, and in the general case one should speak of the three-dimensional transfer function of the optical system and to determine the corresponding number of degrees of freedom of the wave field[24]. Taking this into consideration, we can conclude that the described method is also in full agreement with the concept of invariance of the number of degrees of freedom of the wave field during the image formation. A consequence of the increase of the resolving power in the x direction is in this case the decrease of the resolving power in the z direction, and the number of degrees of freedom of the wave field, corresponding to the transmission of three-dimensional information concerning the object, obviously, remains constant.

This conclusion allows us to draw an analogy between the described method of image formation and interference microscopy, for which there is a known uncertainty relation by Ingelstam for the determination of the transverse and longitudinal coordinates of the object element[25,26]. From this point of view it can be assumed that the object in the described method is supplemented by an artificially produced phase structure, which, varying in the z direction in accordance with a definite law, can be used for the formation of the image of object elements with increased resolving power in the x direction.

VII. METHODS OF IMAGE FORMATION WHICH MAKE IT POSSIBLE TO INCREASE THE RESOLVING POWER IN TWO MUTUALLY PERPENDICULAR DIRECTIONS

We have considered a number of methods which make it possible to increase the resolving power of an optical system in one direction by effectively using the degrees of freedom of the wave field of the object. Here, as seen from the description of the methods, it is necessary to have a priori certain information concerning the object, so as to be able to use the particular method of image formation with increased resolving power. Each of these methods can be used only for a limited class of objects, for example, for objects that do not depend on the time, have no birefringence, etc.

We now must stop to discuss the possibility of using the described methods of image formation to increase the resolving power in two directions and their effective use for a wide range of objects with two-dimensional optical-density distribution.

Three variants have been proposed for this purpose[27]:

1) Use of superposition masks that split the light beam in identical manner, but have different orientations.

2) Rotation of the devices used for the splitting of the light beam around the optical axis.

3) Use of two different types of devices to increase the resolving power in two mutually perpendicular directions.

The possibility of using the first variant was already mentioned in Secs. II and IV. Variants in which synchronously-moving one-dimensional gratings are used to form images of two-dimensional objects with increased resolving power are analyzed in[28], for the case when the gratings are located in the object and image planes. Obviously, the employed gratings must be moved in such a way, that the system remains spatially-invariant during the time of image registration. If we used crossed gratings, which move in such a way that the direction of the displacement bisects the angle between the grating lines, then different illumination conditions are obtained in the image plane for zones that are parallel to the direction of the grating motion, and the transmission of the spatial harmonics will depend on their orientation. Such a system can therefore not be effectively used to produce an image of a two-dimensional object.

The same paper[28] discusses a second variant of the use of synchronously moving one-dimensional gratings in conjunction with their linear and rotational motions, and the transfer function is determined for such an image-formation method. Figure 12 shows how the band width of the transmitted spatial frequencies is increased in this case if the gratings occupy successively the positions $\varphi = 0°$, 45, 90, and 135°.

If sinusoidal gratings are used with a spatial frequency f_0 and the gratings are rotated continuously, the transfer function of the lens acquires a certain addition in the form of a function that is symmetrical with respect to f_0 and has a width $2f'$ (where f' is the limiting spatial frequency transmitted by the lens). Figure 13 shows the transfer function of a lens and the additional terms obtained by using gratings with frequencies $f_0 = f'$, $1.5f'$, $2f'$, and $2.5f'$. It is stated that it is possible to obtain a modulation of about 15% at the point with limiting spatial frequency f'. This variant of two-dimensional image formation with the aid of an optical system with increased resolving power in one

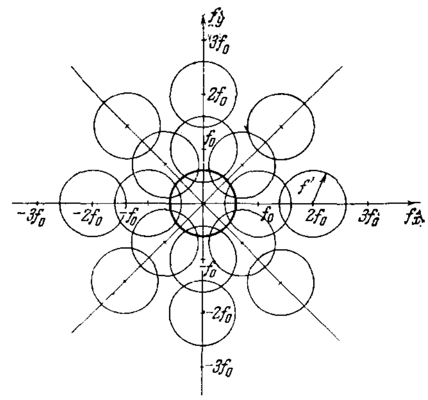

FIG. 12. Increase of the bandwidth of the transmitted spatial frequencies in successive rotation of the gratings through 45°.

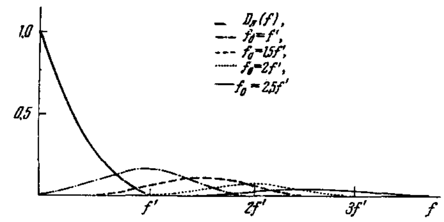

FIG. 13. Transfer function of optical system using linear and rotational grating motion.

direction was investigated by the author together with A. I. Kartashev[29]. As the system with increased resolution in one direction, we used an objective with a slit diaphragm in the plane of the entrance pupil. The ratio of the resolving powers in the mutually perpendicular directions (parallel and perpendicular to the slit) was 50 : 1. Such an optical system can be regarded as the analog of the systems described above, in which the resolving power depends on the direction. The image was produced by multiple successive superposition of images at different orientations of the direction of the maximum resolving power of the optical system: images corresponding to 36 positions of the slit diaphragm, rotating about the optical axis of the image (in steps of 5°) were recorded in succession on a single photographic plate.

Figure 14a shows the image of an object, obtained with an objective having a slit diaphragm in a fixed position, while Fig. 14b shows an image of the same object obtained by rotating the diaphragm around the optical axis by the method indicated above. (The object is a set of luminous points.)

The distribution of the illumination in the image of the point make it possible to assess the resolving power of the optical system. For the described image-formation method, the character of this distribution can be determined by taking as the basis the distribution of the illumination in the image of a point for a direction perpendicular to the slit, and assuming that the relative illumination for any coordinate of the distribution is reduced upon rotation of the slit diaphragm in proportion to the area of the annular zones, having equal width in the radial direction and having a radius corresponding to the coordinate of the determined distribution. (This assumption is valid for the case when the resolving power of the optical system in one direction greatly exceeds the resolving power in the remaining directions, and the width of the annular zones is determined by the distribution of the illumination in the image of the point for the direction of the maximum resolution.)

Figure 15 shows the diffraction distribution of the illumination in the image of a point (scattering function) for an objective with 1.2 mm diameter, corresponding to the width of the slit diaphragm (curve 1), and the calculated curve (2) for the described image-formation method. From a comparison of curves 1 and 2 we can conclude that the diameter of the spot in the point image, bounded by the illumination level amounting to 50% of the maximum, is smaller by a factor of 15 for the described method.

In[29] we discuss also the possibility of recording a two-dimensional object with the system rotating about the optical axis, using a two-step negative-positive method. In this method, one first obtains negative images of the objects for individual positions of the slit diaphragm; these are then summed by successive superposition printing on a photographic plate or paper. Such a method makes it possible to exclude the tails of the illumination distributions, which produce a strong halo and decrease the resolving power in the considered negative method.

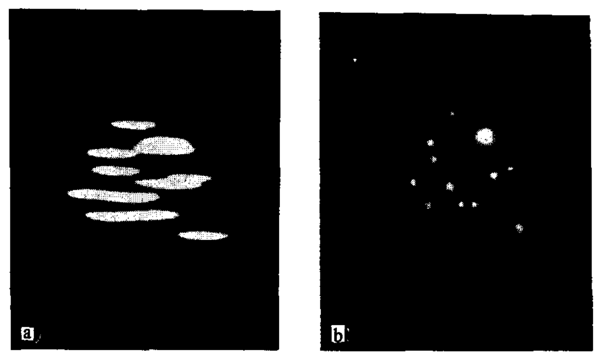

FIG. 14. a) Image obtained with the aid of an optical system with increased resolving power in one direction; b) image of the same object obtained by rotating the optical system with increased resolving power in one direction around the optical axis.

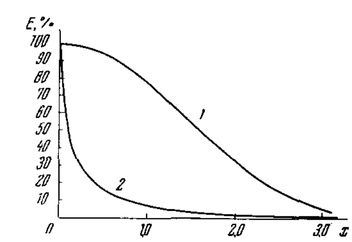

FIG. 15. Theoretical distribution of the illumination in the image of a point. 1 – objective diameter 1.2 mm, corresponding to the width of the slit, 2 – using the described method of image formation (x – coordinate in image plane, E – relative illumination);

Thus, an optical system with increased resolving power in one direction can be used to obtain images of objects not only of linear structure, but also objects with arbitrary two-dimensional brightness distribution. In the described image-formation method, the effective resolving power becomes the same for all directions and greatly exceeds the resolving power of the initial system, but does not reach the maximum.

VIII. TRANSMISSION OF SPATIAL INFORMATION WITH THE AID OF TEMPORAL DEGREES OF FREEDOM WITH REDUCTION OF THE BAND WIDTH OF THE TRANSMITTED SPATIAL FREQUENCIES TO PRACTICALLY ZERO

In spite of the fact that this method cannot be regarded as a purely optical one, since the image is produced with the aid of an additional radio-electronic device, it is of interest to consider its information capacity with respect to the transmission of spatial information.

If the method reduces to element by element scanning of the image of the object, then the objective-receiver system has only one spatial degree of freedom. For example, in a television transmitter the light-sensitive target of the television tube receives at each instant of time only one element of the image, in spite of the fact that the lens itself, which projects the image on the target of the tube, can transmit a large number of spatial elements.

Another variant of image transmission is also possible, in which the object is scanned with the aid of a moving diaphragm[19] or an illuminating beam (traveling-beam television system). In this case it is possible to dispense with the optical system, and to feed the optical energy transmitted through each element of the object (or reflected from it) directly to the optical-radiation receiver.

Obviously, the image transmitted with the aid of a temporal communication channel can have

$$N_{x,y} = N_t = 2\Delta\nu \cdot \Delta T \tag{38}$$

degrees of freedom, where $\Delta\nu$ is the bandwidth of the temporal frequencies, used for the image transmission, ΔT is the time during which the image is transmitted (see the sampling theorem).

The quantities $\Delta\nu$ and ΔT are chosen to fit the employed communication channel. Thus, for phototelegraphic transmission the bandwidth $\Delta\nu$ is of the order of several kHz and the time ΔT of the order of several minutes, and the transmission of an image from outer space is effected at the corresponding bandwidth within several hours. In the television transmission method one obtains (for the Russian standard) $(625)^2$ spatial degrees of freedom at a frame transmission time 1/25 sec and a temporal-frequency bandwidth on the order of 6 MHz.

A clear-cut example of the use of temporal degrees of freedom for the transmission of spatial information is a high-resolution coherent radar lateral-scanning system with synthesized aperture. In such systems, the source of information concerning the objects is the interference field, which is the result of the space-time interaction of electromagnetic waves reflected from the object and the coherent radiation of the antenna. The spatial recording of this interference field on a film is equivalent to a hologram, and the information 14 contains concerning the object can be separated by subsequent processing. Such a radar system ensures a spatial resolution even when the aperture of the antenna is close to zero, and the system is a one-dimensional channel used to transmit a time-dependent signal.

Let a radar antenna be mounted on an airplane flying horizontally with velocity v_x and let it sound the earth's surface with electromagnetic waves with frequency ν_0[16]. Let us see how information is obtained concerning the structure of the terrain in the x direction along the flight trajectory (Fig. 16). The amplitude of the wave reflected by the point object $O(x_0, y_0, z_0 = 0)$ and received by the antenna at the point

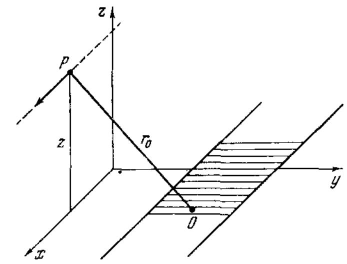

FIG. 16. Diagram illustrating the action of a radar with synthesized aperture. P – radar antenna mounted on an airplane flying at an altitude z with velocity v_x, O – pointlike object on the terrain.

$P(x = v_x t, y = 0, z)$ is

$$u_s(t) = u_0 \exp[-i\omega_0(t - t_{p0})], \tag{39}$$

where u_0 is a constant that depends on the reflectance of the object, $\omega_0 = 2\pi\nu_0$, and t_{p0} is the time delay determined by the length of the path from the antenna to the object and back to the antenna, using the expression

$$t_{p0} = 2c^{-1}[(x - x_0)^2 + y_0^2 + z^2]^{1/2} = \frac{2r_0}{c} + \frac{(x - x_0)^2}{r_0 c}, \tag{40}$$

where $r_0 = (y_0^2 + z^2)^{1/2}$, and c is the velocity of the electromagnetic waves.

The received signal is mixed with the coherent initial signal of frequency ν_0 in the receiver:

$$u_r(t) = u_r \exp(-i\omega_0 t). \tag{41}$$

The resultant temporal signal $u_r^*(t)u_s(t) + u_r(t)u_s^*(t)$ is amplified and is transformed into a spatial signal by recording on a moving film. If the film velocity is $\bar{v}_x$, then the position $\bar{x}$ on the film corresponds to the instant of observation t and to the airplane position x:

$$\bar{x} = \bar{v}_x t = \frac{\bar{v}_x}{v_x} x. \tag{42}$$

The signal corresponding to the pointlike object O, is

$$u_r^*(\bar{x})\, u_s(\bar{x}) + u_r(\bar{x}) u_s^*(\bar{x}) = u_r^* u_0 \exp i\varphi(\bar{x}), \tag{43}$$

where

$$\varphi(\bar{x}) = 2k_0 r_0 + \left(\frac{v_x}{\bar{v}_x}\right)^2 k_0 \frac{(\bar{x} - \bar{x}_0)^2}{r_0}, \tag{44}$$

and $k_0 = 2\pi\nu_0/c$. This signal is the equivalent of the usual Fresnel type hologram, transformed in a scale $\bar{v}_x/v_x$.

The phase $\varphi(x)$ of the recorded signal is expanded in a Taylor series about the point x':

$$\varphi(x) = \varphi(x') + \left.\frac{d\varphi}{dx}\right|_{x=x'} (x - x') + \frac{1}{2} \left.\frac{d^2\varphi}{dx^2}\right|_{x=x'} (x - x')^2 + \dots \tag{45}$$

The first derivative of the phase at the point $x = x'$ is the local spatial frequency at this point:

$$k_x(x') = \left.\frac{d\varphi}{dx}\right|_{x=x'}. \tag{46}$$

Therefore the local spatial frequencies represented in the resultant signal (43) can be obtained by differentiating (44)

$$k_x(x) = \frac{2k_0(x - x_0)}{r_0}. \tag{47}$$

The earth's surface covered simultaneously by the radar beam has a linear dimension in the x direction:

$$L_x = r_0 \frac{\lambda_0}{D_x}, \tag{48}$$

where λ_0 is the wavelength of the electromagnetic radiation for the radar antenna, and D_x is the antenna diameter. The received signal therefore covers a band of spatial frequencies with width

$$\Delta k_x = \frac{2k_0 L_x}{r_0}. \tag{49}$$

It is known that for any wave process there is an uncertainty relation that limits the accuracy of the simultaneous determination of the coordinate x and the corresponding component k_x of the wave vector:

$$\delta x \cdot \delta k_x \geqslant 2\pi. \tag{50}$$

Therefore the resolving power is defined by the relation

$$\Delta x = \frac{2\pi}{\Delta k_x} = \frac{\lambda_0 r_0}{2L_0} = \frac{D_x}{2}, \tag{51}$$

whereas for an ordinary radar system $\Delta x = L_x$.

In order to obtain and record a signal corresponding to the bandwidth of the local spatial frequencies Δk_x, the required bandwidths for the temporal channel and for the spatial frequencies on the film are $\Delta\omega_0$ and $\Delta\bar{k}_x$, such that

$$v_x \Delta k_x = \Delta\omega_0 = \bar{v}_x \Delta k_x. \tag{52}$$

Within the time interval ΔT, the airplane covers a distance $s_x = v_x \Delta T$, and the received signal is recorded on a film of length $\bar{s}_x = \bar{v}_x \Delta T$. The number of degrees of freedom of the signal obtained from the field of the object s_x is transmitted by the receiving channel in the time interval ΔT within a temporal-frequency band width $\Delta\nu_0$, and is ultimately recorded on a film segment $\bar{s}_x$ within a spatial-frequency band width $\Delta\bar{k}_x$. We can write

$$\left.\begin{aligned} N_x &= \frac{2s_x \Delta k_x}{2\pi}, \\ N_t &= 2\Delta\nu_0 \Delta T, \\ N_{\bar{x}} &= \frac{2\bar{s}_x \Delta\bar{k}_x}{2\pi}. \end{aligned}\right\} \tag{53}$$

Thus, the spatial information obtained with the aid of such a radar system has N_x degrees of freedom, with

$$N_x = N_t = N_{\bar{x}}. \tag{54}$$

The manner of obtaining information concerning the structure of the object in the y direction, perpendicular to the flight direction, is essentially the same from the point of view of the discussed question, and reduces to registration of the interference field equivalent to the Fraunhofer hologram.

The close analogy between the described coherent radar and optical holography makes it necessary to determine the information capacity of holograms and to analyze the factors limiting it. The importance of such an analysis is determined by the extensive practical utilization of holography at the present time.

IX. NUMBER OF WAVE-FIELD DEGREES OF FREEDOM RECORDED IN A HOLOGRAM

As is well known, a hologram is obtained by interference between the light waves scattered by the object and the incident light wave. The information of the object is carried in this case by the phase of the optical radiation.

In discussing the question of the effective utilization of the degrees of freedom of a wave field in image formation, it is of interest to determine how many degrees of freedom of the wave field of the object are fixed in the holographic process in a hologram of a given dimension[16]. We shall regard the photographic plate used to record the hologram as a system for the recording of the two-dimensional distribution of a wave-field intensity having a limiting spatial frequency.

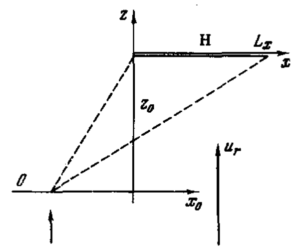

FIG. 17. Production of Fresnel hologram. O – object ($x_0 \leqslant 0$), H – holographic plate ($x \geqslant 0$), u_r – reference beam.

Let us determine the information capacity of a Fresnel-type hologram. Let the light wave illuminating the object and the reference wave be monochromatic, with a radiation frequency ν_0; the object is independent of the time. The scheme for obtaining the hologram is shown in Fig. 17. A spherical wave from the object point (x_0, y_0) in the plane $z = -z_0$ has at the point (x, y) in the hologram plane $z = 0$ an amplitude

$$u_s(x, y; t) = u_0 \exp\{-i[\omega_0 t - \varphi(x, y)]\}, \tag{55}$$

where

$$\varphi(x, y) = k_0[(x-x_0)^2 + (y-y_0)^2 + z_0^2]^{1/2} \tag{56}$$

and $k_0 = 2\pi\nu_0/c$. It is assumed that the reference wave is plane and that it is superimposed on the hologram plane with amplitude

$$u_r(x, y; t) = u_r \exp(-i\omega_0 t). \tag{57}$$

Since the photographic plate is sensitive to the intensity distribution, the photograph records the picture of the Fresnel zone

$$u_r^*(x, y; t)\, u_s(x, y; t) = u_r^* u_0 \exp[i\varphi(x, y)], \tag{58}$$

corresponding to the given point of the object.

From (56) we see that the local spatial frequencies at the point (x, y) in the hologram plane are equal to

$$\begin{aligned} k_x &= \frac{d\varphi}{dx} = \frac{k_0(x-x_0)}{[(x-x_0)^2 + (y-y_0)^2 + z_0^2]^{1/2}} \approx \frac{k_0(x-x_0)}{z_0}, \\ k_y &= \frac{d\varphi}{dy} = \frac{k_0(y-y_0)}{[(x-x_0)^2 + (y-y_0)^2 + z_0^2]^{1/2}} \approx \frac{k_0(y-y_0)}{z_0} \end{aligned} \tag{59}$$

under the condition

$$\frac{|x-x_0|}{z_0} \ll 1 \text{ and } \frac{|y-y_0|}{z_0} \ll 1.$$

If the object is limited to $x_0 \leq 0$, and the hologram plate to $x \geq 0$, then all the spatial frequencies of the object lie in the right half of the (k_x, k_y) plane, i.e., we are dealing with holograms having a single side band.

Let us consider what band of the object spatial frequencies in the x direction is recorded by a hologram of given dimension L_x with limiting spatial frequency k'_x. From (59) we find that for the object field near $x_0 = 0$ the hologram records the spatial frequencies

$$0 \leqslant k_x \leqslant \frac{k_0 L_x}{z_0}. \tag{60}$$

It is assumed that $L_x \leq L_m$, where

$$L_m = \frac{z_0 k'_x}{k_0} \quad (61)$$

is the maximal useful length of the hologram plate. The spatial frequencies obtained when $x > L_m$, above the limiting frequency of the photographic plate, are not recorded.

Relation (60) can be expressed in the form

$$0 \leqslant k_x \leqslant k_0 \sin \alpha_x, \quad (62)$$

where α_x is the angle at which the plane of the hologram is seen from the optic point $x_0 = 0$. The maximum useful angular dimension of the hologram is determined from the condition

$$\sin \alpha_m = \frac{k'_x}{k_0}. \quad (63)$$

(For emulsions with limiting frequency $f'_x = k'_x/2\pi = 1000$ lines/mm, used for the visible part of the spectrum, $\alpha_m \approx 30°$.)

Let us consider the bandwidth of the spatial frequencies of the object, recorded in the hologram, as a function of the position of the object point in the (x_0, y_0) plane. For the object field in the interval $0 \leq x_0 \leq L_m - L_x$, the recorded spatial frequencies are

$$\frac{k_0(-x_0)}{z_0} \leqslant k_x \leqslant \frac{k_0(L_x - x_0)}{z_0}. \quad (64)$$

The bandwidth of the spatial frequencies is in this case

$$\Delta k_x(x_0) = \frac{k_0 L_x}{z_0}. \quad (65)$$

For the object field in the interval $L_m - L_x \leq -x_0 \leq L_m$, the recorded spatial frequencies are

$$\frac{k_0(-x_0)}{z_0} \leqslant k_x \leqslant k'_x = \frac{k_0 L_m}{z_0},$$

i.e., the bandwidth of the spatial frequencies is

$$\Delta k_x(x_0) = \frac{k_0(L_m + x_0)}{z_0} \quad (66)$$

and decreases linearly to zero at the edge of the object field at $x_0 - L_m$.

A photographic plate of length L_x with limiting spatial frequency k_x can record a one-dimensional spatial signal with

$$(N_x)_{\max} = \frac{L_x k'_x}{\pi} = L_x L_m \frac{k_0}{\pi z_0} \quad (67)$$

degrees of freedom. If the plate records a hologram, then $(N_x)_{\max}$ is the upper limit of the wave-field degrees of freedom that can be recorded.

We shall show that the capabilities of the photographic plate are not fully utilized in a hologram with a single side band, i.e., that $N_x < (N_x)_{\max}$.

The number of wave-field degrees of freedom recorded in the interval Δx of the hologram is $2\Delta x \Delta k_x(x)/2\pi$, where $\Delta k_x(x)$ is the spatial-frequency band width present in this interval and capable of being recorded by the emulsion. From the condition that one sideband is present $(x \geq 0,\ x_0 \leq 0)$ it follows that the spatial frequencies $k_x(x) \leq k_0 x/z_0$ are missing from the hologram, and therefore

$$\Delta k_x(x) = k'_x - \frac{k_0 x}{z_0}. \quad (68)$$

The total number of degrees of freedom of the spatial signal $u_r^* u_s$, recorded in the entire plate, is

$$N_x = 2(2\pi)^{-1} \int_0^{L_x} \Delta k_x(x)\,dx = 2L_x\left(L_m - \frac{1}{2}L_x\right)\frac{k_0}{2\pi z_0}. \quad (69)$$

It can be shown that in the integration over the entire field of the object L_m the number of degrees of freedom

$$N_x = 2(2\pi)^{-1} \int_0^{L_m} \Delta k_x(x_0)\,dx_0 \quad (70)$$

with allowance for relations (65) and (66), turns out to equal the value given by formula (69).

The efficiency with which the information capacity of the photographic plate is employed for Fresnel holograms is

$$\eta = \frac{N_x}{(N_x)_{\max}} = 1 - \frac{1}{2}\frac{L_x}{L_m}. \quad (71)$$

For a plate of length L_x, the minimum distance $(z_0)_{\min}$ at which the plate has the maximum useful can be determined from the formula (see Eq. (61))

$$(z_0)_{\min} = L_x \frac{k_0}{k'_x}. \quad (72)$$

Then Eq. (71) can be written in the form

$$\eta = 1 - \frac{1}{2}\frac{(z_0)_{\min}}{z_0}. \quad (73)$$

We see therefore that the efficiency of using the photographic plate, η, is not constant but depends on the distance z_0 between the object and the hologram. It is obvious that the information capacity of the plate is used more efficiently if the object is moved farther away. If $z_0 \to \infty$ for Fraunhofer holograms with a single side band, then an efficiency $\eta = 1$ is reached.

However, in view of the fact that the resolving power of holograms[30-32] is determined by the expression

$$\Delta x = \lambda_0 \frac{z_0}{2L_x}, \quad (74)$$

the resolution for closer objects will be better than for remote ones.

A rectangular photoplate with dimensions L_x and L_y in the directions of x and y respectively can record a spatial wave field by means of

$$(N_{x,y})_{\max} = \left(\frac{L_x k'_x}{\pi}\right)\left(\frac{L_y k'_y}{\pi}\right) \quad (75)$$

degrees of freedom. Assuming that the emulsion is isotropic, we set $k'_x = k'_y$. Since the condition of the presence of a single side band does not limit the component k_y of the spatial frequencies, we can show that

$$N_y = (N_y)_{\max} = \frac{L_y k'_y}{2\pi}.$$

It follows therefore that the efficiency with which the information capacity of the two-dimensional plate is employed in holography with single side band is

$$\eta = \frac{N_{x,y}}{(N_{x,y})_{\max}} = \frac{N_x}{(N_x)_{\max}}.$$

X. CONCLUSION

The main parameter that determines the information capacity of a system for image formation is the number of degrees of freedom of the wave field, N, which can be transmitted by the system into the image space. This number N is fixed for a given system. Therefore the bandwidth of the spatial frequencies transmitted by the system can be increased, but only at the expense of decreasing one of the other factors that determine N.

In conclusion, the author is grateful to A. I. Kartashev and N. R. Batarchukova for taking part in a discussion of the materials and for valuable hints.

[1]H. Wolter, Ann. d. Phys. 7, 341 (1950).
[2]G. Toraldo di Francia, JOSA 45, 497 (1955).
[3]I. Harris, JOSA 54, 931 (1964).
[4]Ya. I. Khurgin and V. P. Yakovlev, Metody teorii tselykh funktsiĭ v radiofizike, teorii svyazi i optiki (Methods of the Theory of Entire Functions in Radiophysics, Communication Theory, and Optics), Fizmatgiz, 1962.
[5]E. Lau, Phys. Zs. 38, 446 (1937).
[6]A. Lohmann and D. Paris, Optik 22, 226 (1965).
[7]P. Lacomme, Optica Acta 1, 33 (1954).
[8]K. D. Sinel'nikov and A. N. Ryazanov, Opt. Spektrosk. 5, 184 (1958).
[9]H. Wolter, Physica 6, 457 (1958).
[10]D. Gabor, Proc. of the Simposium on Astronomical Optics, Amsterdam, 1956, p. 17.
[11]K. Miyamoto, JOSA 50, 856 (1960).
[12]D. Gabor, Progr. Optics 1, 109 (1961).
[13]W. Lukosz, JOSA 56, 1463 (1966).
[14]W. Lukosz, Zs. Naturforsch. 18a, 436 (1963).
[15]W. Lukosz and M. Marchand, Optica Acta 10, 241 (1963).
[16]W. Lukosz, JOSA 57, 932 (1967).
[17]M. Grimm and A. Lohmann, JOSA 56, 1151 (1966).
[18]A. Bachl and W. Lukosz, JOSA 57, 163 (1967).
[19]M. Franson, Nuovo Cimento, Suppl. 9, 283 (1952).
[20]A. Lohmann, Optica Acta 3, 97 (1957).
[21]W. Gartner and A. Lohmann, Zs. Phys. 174, 18 (1963).
[22]A. Lohmann and D. Paris, Appl. Optics 3, 1037 (1964).
[23]A. I. Kartashev, Opt. Spektrosk. 9, 394 (1960).
[24]R. Frieden, JOSA 57, 56 (1967).
[25]E. Ingelstam, Ann. Phys. 7, 309 (1954).
[26]E. Menzel, Optik und Spektroskopie aller Wellenlänge, Berlin, 1962, p. 18.
[27]D. Paris and B. Morgenstern, JOSA 53, 1342 (1963).
[28]D. Paris and B. Morgenstern, JOSA 54, 1282 (1964).
[29]A. I. Kartashev and A. N. Korolev, Opt. Spektrosk. 23, 450 (1967).
[30]G. W. Stroke, Introduction to Coherent Optics and Holography, Academic Press, 1966.
[31]A. A. Pistol'kors, Dokl. Akad. Nauk SSSR 172, 334 (1967) [Sov. Phys.-Dokl. 12, 79 (1967)].
[32]B. P. Konstantinov, A. N. Zaĭdel', V. B. Konstantinov, and Yu. I. Ostrovskiĭ, Zh. Tekh. Fiz. 36, 1718 (1966) [Sov. Phys.-Tech. Phys. 11, 1279 (1967)].

Translated by J. G. Adashko