ISSN 1028-3358, Doklady Physics, 2023, Vol. 68, No. 10, pp. 328–333. © Pleiades Publishing, Ltd., 2023.
Russian Text © The Author(s), 2023, published in Doklady Rossiiskoi Akademii Nauk. Fizika, Tekhnicheskie Nauki, 2023, Vol. 512, pp. 44–50.

PHYSICS

Effective Methods of Synthesis and Optimization of a Holographic Mask

V. V. Chernik[a,*]

Presented by Academician F.L. Chernous'ko April 3, 2023

Received April 18, 2023; revised April 18, 2023; accepted May 7, 2023

Abstract—Several statements of the problems of synthesis of a holographic mask in the form of optimization problems for quality of holographic images are presented. An effective algorithm for the synthesis of holographic masks based on FFT with the complexity $O(N\ln N)$, where N is the number of elements of the depicted object, is described. Based on this algorithm, a scalable software package has been developed and implemented that allows synthesizing holographic masks for various lithography applications, including the production of MEMS, MOEMS, and high-end chips. Experimental results are presented.

Keywords: diffraction, computer-generated holography, optimization problems, photolithography

DOI: 10.1134/S1028335823100038

Projection photolithography is today the main tool for the production of microelectronic products. Constantly increasing resolution requirements have made the photolithography process very sensitive to external disturbances and have increased the cost of manufacturing the projection mask. As the optical resolution limit is approached, diffraction effects increasingly negatively affect the quality of the projection image. This is due to the fact that historically the technology of projection photolithography developed basing on the principles of ray optics. Resolution enhancement techniques (RETs), such as OPC, PSM, and SMO, are essentially methods to partially neutralize diffraction effects. However, they significantly increase the cost of the mask and complicate the manufacturing process [1–3]. Modeling and designing RETs has become a very expensive theoretical and, most importantly, engineering task aimed at suppressing manifestations of the wave nature of light (scattering, proximity effects, the Gibbs phenomenon). The cost of projection masks and related optical equipment increases disproportionately as microelectronic elements become miniaturized.

In recent years, the holographic approach to obtaining high-resolution images has been studied intensively. A group of authors is developing the technology of subwavelength holographic lithography (SVGL) [1–3]. This technology addresses several challenges that have arisen in the microelectronics industry over the past few decades. The basic idea is to exploit the wave properties of light using holographic techniques, rather than suppress them. Many engineering problems are moving into the area of modeling and numerical calculations. Efficient algorithms for computer synthesis of holographic masks based on FFT with complexity $O(N\ln N)$ were developed, where N is the number of image elements, which made it possible to synthesize holographic masks for photolithography of layers of modern microelectronic devices.

A holographic mask is a physical implementation of the function of the amplitude, phase, or amplitude-phase modulation of a light wave. This function can be obtained as the intensity of the interference pattern of the reference and object waves, in accordance with the classical principle proposed by D. Gabor. According to this principle, an object is illuminated by some light beam, forming an object wave (Fig. 1a). At the same time, the reference light beam, mutually coherent with the object beam, forms a reference wave [4, 5]. The result of the interference of these waves is recorded on the plane H. The resulting interference pattern can be understood as a function of amplitude modulation $T : H \to [0;1]$. If in place of a plane H we install a diffraction element that implements the resulting amplitude modulation function for the reconstruction wave, which is inverse to the reference wave, then in place of the object an electromagnetic field will be restored that repeats the field of the object wave, i.e., the holographic image of an object (Fig. 1b). The simplest design of a holographic mask is an opaque screen with

[a] *Ishlinskii Institute for Problems in Mechanics, Russian Academy of Sciences, Moscow, Russia*

**e-mail: gungho424@gmail.com*

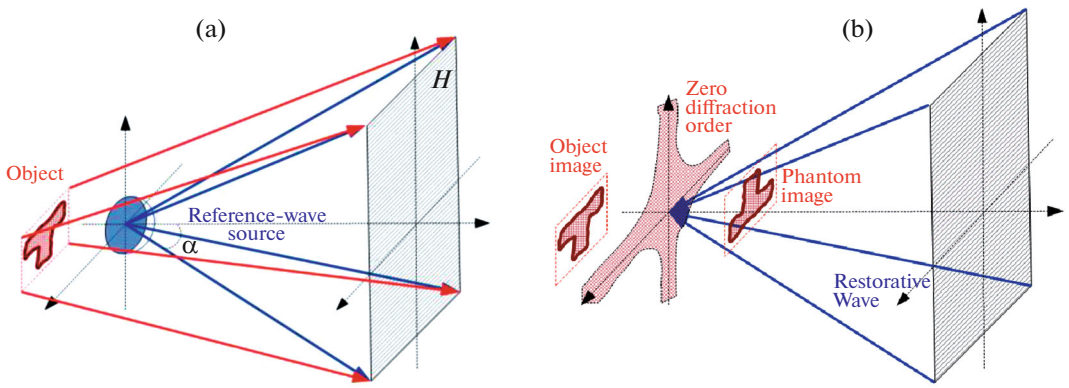

Fig. 1. Holographic approach to obtaining an image of an object.

holes, where the amplitude modulation function is approximated by the density and size of the holes. This design, however, has a number of disadvantages, the main one of which is the presence of a high-intensity zero order of diffraction. Let us describe some types of hologram mask designs and point out their features and advantages.

Let $O : H \to \mathbb{C}$ be the complex amplitude of the reference wave on the hologram plane, and let $\Pi : H \to \mathbb{C}$ be the complex amplitude of the object wave. Then consider the function

$$T = \frac{|O + \Pi|^2}{|O|^2}, \tag{1}$$

taking real non-negative values on the hologram area H as a function of amplitude modulation. Applying it to the backward wave O^*, we get

$$TO^* = \frac{\left(|O|^2 + |\Pi|^2\right)O^*}{|O|^2} + \frac{\Pi O^{*2}}{|O|^2} + \Pi^*. \tag{2}$$

The resulting field is decomposed into three components:

(1) Π^* is the field of the inverted object wave, which creates an image of the object;

(2) $\dfrac{\Pi O^{*2}}{|O|^2}$ is the field of a distorted wave, pseudo-symmetrical to the object wave relative to the optical axis, creating a distorted image of the object (phantom image);

(3) $\dfrac{\left(|O|^2 + |\Pi|^2\right)O^*}{|O|^2}$ is the zero order diffraction, focusing in the vicinity of the focus in the case of a diverging reference wave, or creating background noise in the case of a plane reference wave.

Note that, for the physical implementation of a holographic mask, it will also be necessary to introduce a normalization coefficient M such that $\dfrac{T}{M} : H \to [0;1]$.

The zero order diffraction concentrates a significant portion of the energy of the reconstruction wave, which negatively affects the diffraction efficiency of the setup. In addition, zero-order radiation scattered by elements of the optical circuit is a significant source of noise, which, due to the coherence of the radiation used, reveals itself in the form of phase speckle noise, which destroys the image quality [5]. To reduce the degree of influence of the zero order of diffraction on the image, optical setups with a convergent reconstruction beam are used, in which the zero order is focused and the useful image can be separated spatially from the main flow of the zero order components. The advantage of such a hologram is the ease of manufacturing its physical implementation in the form of a binary amplitude mask, in which amplitude modulation is implemented using holes of variable size in an opaque screen (Fig. 2a). The size of the holes is selected in such a way as to implement locally the required amplitude modulation in a given area.

Please note that the function component T, which is responsible for the zero order of diffraction, is a real-valued function. Consider the following modulation function:

$$T_1 = \frac{|O + \Pi|^2 - (|O|^2 + |\Pi|^2)}{M_1 |O|^2} = \frac{O^* \Pi + O\Pi^*}{M_1 |O|^2}. \tag{3}$$

Here T_1 takes real values, and the normalization coefficient M_1 is chosen so that $T_1 : H \to [-1;1]$. Applying such a modulation function to O^* similarly to (2), we get

$$T_1 O^* = \frac{\Pi O^{*2}}{|O|^2} + \Pi^*,$$

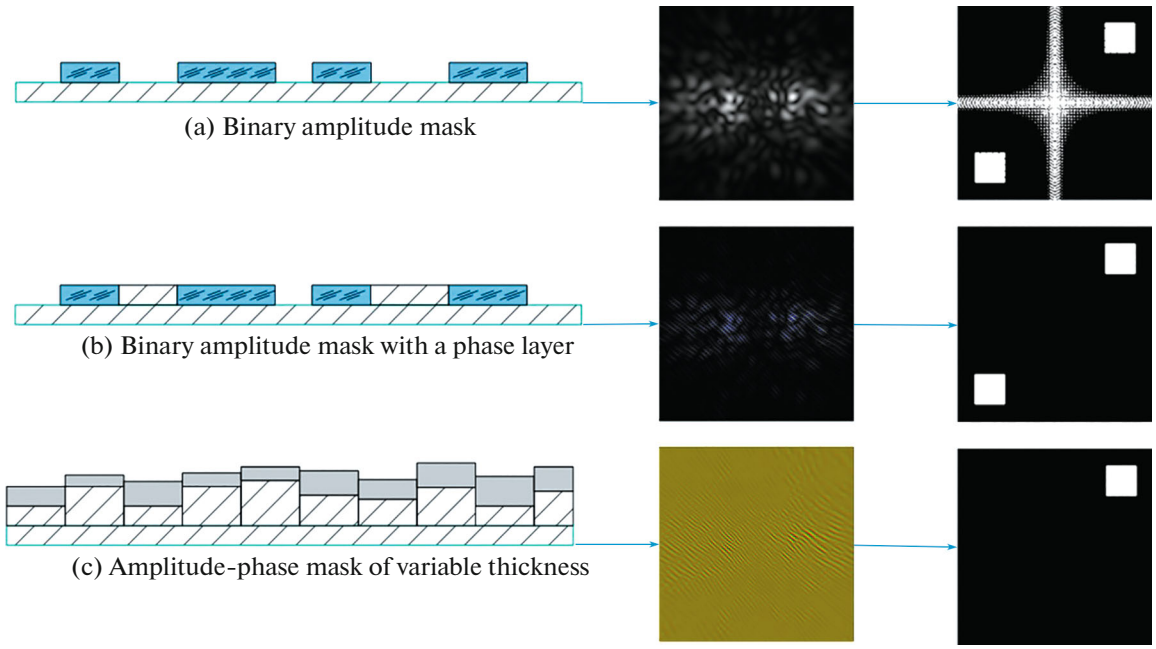

Fig. 2. Three types of holographic mask designs. The corresponding diffraction pattern during reconstruction: (a) using an amplitude mask consists of the zero order of diffraction, an object image, and a phantom image; (b) using an amplitude mask with a phase layer, it consists of an object image and a phantom image; and (c) in the case of an amplitude-phase mask, this is only an image of the object.

and thus, the reconstruction wave under the influence of such a modulation function will have only two components corresponding to the object image and the phantom image. For physical implementation, it is necessary to apply an additional phase layer that rotates the wave phase by 180° on the mask elements corresponding to negative values of the modulation function (Fig. 2b). A significant disadvantage is also the shift in the average transmittance level of the mask to zero. Thus, if the average value of the function T after normalization is equal to 0.5, the average value of the function T_1 turns out to be zero, and most of the mask area turns out to be dark. Masks of this type have a serious advantage due to the almost complete absence of a zero diffraction order. The diffraction efficiency of such a mask is significantly higher than the amplitude efficiency (up to four times). The image quality is also noticeably improved due to a better signal-to-noise ratio. However, even taking into account additional clearing procedures, the manufacturing of such a mask requires a large number of small elements in the range $(0;0.25\lambda]$, where λ is the wavelength of reference and reconstruction radiation [3]. This affects not only the cost of the mask, but also the resistance to modeling errors and manufacturing defects of such elements. To model the passage of light through such holes reliably, an accurate diffraction model is required.

It is also possible to consider the modulation function T_2, in which only the component responsible for the image of the object is present, i.e., complex-valued modulation function

$$T_2 = \frac{O\Pi^*}{M_2|O|^2}. \quad (4)$$

To implement such a function on a hologram mask physically, amplitude and phase layers of variable thickness are required (Fig. 2c). A rigorous diffraction model is also needed to simulate the passage of a light wave through a layered medium with a variable refractive index.

Let us consider the issue of the quality of the created image. The main objects of the methods described above for restoring a holographic image, based on the principle of D. Gabor, are the reference wave O, object or object wave Π, and a hologram, the mathematical representation of which is the modulation function T, T_1, or T_2 depending on the intended structure of the mask. We will further denote the modulation function by the letter T, implying one of the options. Let us introduce the operator D, for which modulation functions T match the reconstructed holographic image I, i.e., $I = D(T)$. Image of the object is a result of exposure of the holographic mask by restoring wave. The image produces an imprint on a photosensitive material. Image defects usually manifest themselves in the form of uneven brightness of different elements, fluctuations in brightness within one element, as well as insufficient contrast in the pro-

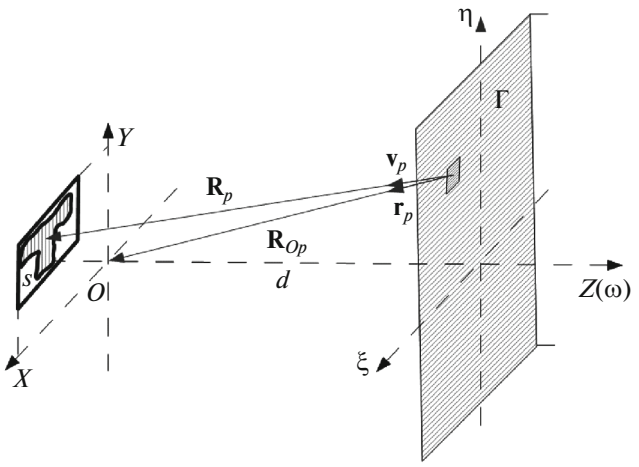

Fig. 3. Scheme of the holographic approach to image acquisition.

duction of microelectronic components, which leads to uneven edges of elements (LER, LWR [6]) imprints in the photoresist layer. The degree of defectiveness of the holographic image I can be expressed through the difference from some ideal image A according to some metric σ, which can reflect various aspects of quality from the uniformity of the illumination intensity across the entire image field to local defects in the final print in the photoresist. Let us call this metric the quality functional. The optimization task is to minimize these differences and image defects. Let us consider the formulation of the problem of optimizing the image quality due to variations in the modulation function T, i.e.,

$$\sigma(A, D(T)) \underset{T}{\rightarrow} \min. \tag{5}$$

To solve this problem using numerical methods, we divide the hologram with a rectangular grid into small cells and represent the modulation function in the form $T(\xi,\eta) = \sum_p \alpha_p \psi_p(\xi,\eta)$ (Fig. 3). Here α_p is the average modulation coefficient of pth cell, $\psi_p(\xi,\eta)$ is the indicator function of pth cell, equal to zero outside the cell and to one inside, and (ξ,η) are the coordinates on the hologram plane.

Operator D, based on the Fresnel–Kirchhoff approach, can be represented as an integral

$$D(T) = \iint_\Gamma K(x-\xi, y-\eta, d) O(\xi,\eta) T(\xi,\eta) d\xi d\eta \approx \sum_p \alpha_p D_p(x,y), \tag{6}$$

where $O(\xi,\eta)$ is the complex amplitude of the reconstruction wave field on the hologram plane, $K(x-\xi, y-\eta, d) = \frac{de^{-ikR_p}}{R_p^2}$, $k = \frac{2\pi}{\lambda}$, $\mathbf{R}_p = (x-\xi, y-\eta, d)$, and Γ is the hologram area. Given the modulation function representation $T(\xi,\eta)$ in the form of a linear combination of characteristic functions of a cell, this integral decomposes into a corresponding linear combination of integrals over cells, each of which, using the far-field approximation [7] and under the assumption of a spherical restoring wave, can be represented as

$$D_p(x,y) \approx \frac{2ik}{\pi} \frac{\exp(ik(R_p + R_{Op}))}{R_p^2} \times \frac{\sin(k(r_{p1} + \nu_{p1})\Delta)}{k(r_{p1} + \nu_{p1})} \frac{\sin(k(r_{p2} + \nu_{p2})\Delta)}{k(r_{p2} + \nu_{p2})},$$

where Δ is the grid pitch on the hologram, (ν_{p1}, ν_{p2}) and (r_{p1}, r_{p2}) are the coordinates of unit direction vectors $\boldsymbol{\nu}$ and $\mathbf{r}$ along the axes X and Y, and the origin of the system OXY is the focus of the restoring spherical wave.

Let us also introduce a grid on the image plane. Then the intensity of the ideal image of the object can be represented as a set of values $\{a_q \in [0;1], q = 1,\ldots,N\}$ at the nodes of this grid. Let us name the number N of grid nodes in the image sufficient to convey information about the imaged object during sampling, the complexity of the object. Then the problem in discrete form can be represented in the following form:

$$\sigma = \sum_q (b_q b_q^* - c a_q)^2 \xrightarrow[\{\alpha_p\}, c]{} \min, \quad b_q = \sum_p K_D(p,q)\alpha_p. \tag{7}$$

Here, the deviation of the intensity of the reconstructed image from the ideal intensity distribution is l_2-norm of the normalized difference at the nodes of the computational grid, coefficient c has the meaning of a normalization coefficient, which is responsible for the fact that the reconstructed image should be close to the ideal shape. Coefficient c is also the subject of variation during the optimization process. Matrix $K_D(p,q)$ is a matrix of local responses of the system, i.e., $K_D(p,q) = D_p(x_q, y_q)$, where (x_q, y_q) are the coordinates of the corresponding node on the image plane. This matrix has the dimension $N*M$, where N is the complexity of the depicted object and M is the number of cells in the hologram, and it is not sparse; i.e, the values of its elements are evenly distributed throughout the matrix. The complexity of the hologram mask corresponds to the complexity of the image, so the asymptotically computational complexity of such a problem is $O(N^2)$. The optimization algorithm for this problem statement was implemented [8] basing on the gradient descent method, with the Gabor hologram (1) being used as the first approximation. It was shown that the quality of the image restored using a computer-synthesized hologram can be significantly improved. As a result of ten iterations

of the gradient descent method, l_2-norm of deviation of the intensity of some images from the ideal, was improved by more than two times compared to the first approximation. However, due to the quadratic dependence of the computational complexity of the optimization problem on the complexity of the imaged object in formulation (7), such an optimization method can be implemented only for small objects. Real objects, the photolithography of which is of industrial interest, are layers of microelectronic components and can have a complexity of ~10^{12} or more.

Let us introduce the concept of a virtual object $V : S \to \Omega$, where S is the area in which the object is defined (Fig. 3) and Ω is a circle of unit radius on the complex plane. We also introduce the operator G, which to each virtual object V assigns modulation function T a hologram mask obtained by one of the methods described above, i.e., $T = G(V)$. The operator G is the composition of a non-local operator of type (6), which associates the complex amplitude of the object wave over the area of a hologram to a virtual object with highlighting V, which is the distribution of the object wave on the object, and one of three operators (1), (3), or (4), assigning to each object wave a certain modulation function of the hologram mask. Operator T the same as operator D, introduced earlier and each modulation function T which matches the reconstructed holographic image, has a non-local character. Thus, each element of the virtual object contributes to the definition of the modulation function T at every point of the hologram. Consider the composition of operators $H(V) = D(G(V))$. The integral representation of the operator H will look like

$$H(V) = \iint_S K_H(x - x_0, y - y_0, x_0, y_0)V(x_0, y_0)dx_0dy_0, \quad (8)$$

where (x_0, y_0) and (x, y) are Cartesian coordinates in the image area and $K_H(\Delta x, \Delta y, x_0, y_0)$ is the operator kernel that describes the effect of changing the value of the complex amplitude of the illumination wave of a virtual object at a point (x_0, y_0) to the complex amplitude of the reconstruction wave at the point (x, y). Let us describe some qualitative properties of the kernel K_H. It depends on the optical design and the choice of the reconstruction wave. Its absolute value decreases rapidly with increasing $r = \sqrt{\Delta x^2 + \Delta y^2}$, i.e., operator H, unlike operators D and G, of which it is a composition, has a local character. For computational problems it makes sense to limit the support of the function K_H by the $(\Delta x, \Delta y)$ region $r < 5\lambda$, and neglect the values K_H at $r \geq 5\lambda$. Finally, if the restoring wave $O(\xi, \eta)$ is ideally flat, and the hologram area is infinite in all directions, then $K_H = K_H(\Delta x, \Delta y)$ does not depend on the point (x_0, y_0). In the practically significant case of a finite hologram and a converging reconstruction wave, the core K_H depends on the point (x_0, y_0); however, this dependence manifests itself only at distances on the order of $10^4\lambda$. Thus, for computational purposes within a subdomain of order size $10^3\lambda$, the core K_H depending only on $(\Delta x, \Delta y)$ can be considered. This observation makes it possible to divide the region of a virtual object into subregions A_n, inside each of which the operator H is represented as a convolution operator with some kernel K_H^n, which is an approximation of the kernel K_H in this subarea.

Let us now consider the problem of optimizing the image quality similarly to (5), but we will consider a virtual object as the object of variations:

$$\sigma(A, H(V)) \xrightarrow[V]{} \min.$$

Let us introduce a uniform rectangular grid with a step Δ on each subarea S_n. Then the ideal image of the object A can be represented as a set of values $\{a_{pq} \in [0;1]\}$ at the nodes of this grid. Virtual object V can also be represented as a two-dimensional array $\{v_{pq}\}$, where each value v_{pq} is some complex number from the unit circle with the center at zero, and the quality functional as the sum of the restrictions of this functional on each of the subdomains $\sigma = \sum \sigma_n$. In practice, there is no need to find the global minimum of the quality functional. It is sufficient to achieve a level of quality such that the predetermined tolerances are met. Therefore, in each subdomain the problem can be formulated as follows. Find $\{v_{pq} \in \mathbb{C} : |v_{pq}| \leq 1\}$ and a real constant c such that

$$\sigma_n = \sum_{p,q} (b_{pq}b^*_{pq} - ca_{pq})^2 < \varepsilon,$$

where $b_{pq} = \sum_{lm} K_H((p - l)\Delta, (q - m)\Delta)v_{lm}$.

To solve the problem of hologram synthesis in C++ using the MPI and OpenMP libraries, a software package for cluster systems was developed. It implements the optimization algorithms described above, as well as similar algorithms for the vector diffraction model [9]. The calculations are parallelized in such a way that each subdomain S_n is processed on a separate set of computing nodes using the gradient descent method. All gradients and fields are represented using convolution operators, so the fast Fourier transform and the fftw3 library are used to speed up calculations. The computational complexity of the algorithms is thus equal to $O(N^* \ln N)$.

The independence of the definitions of a virtual object and an ideal image in the proposed approach provides additional flexibility and allows more complex problems of hologram synthesis and optimization to be posed in a universal manner. For example, one

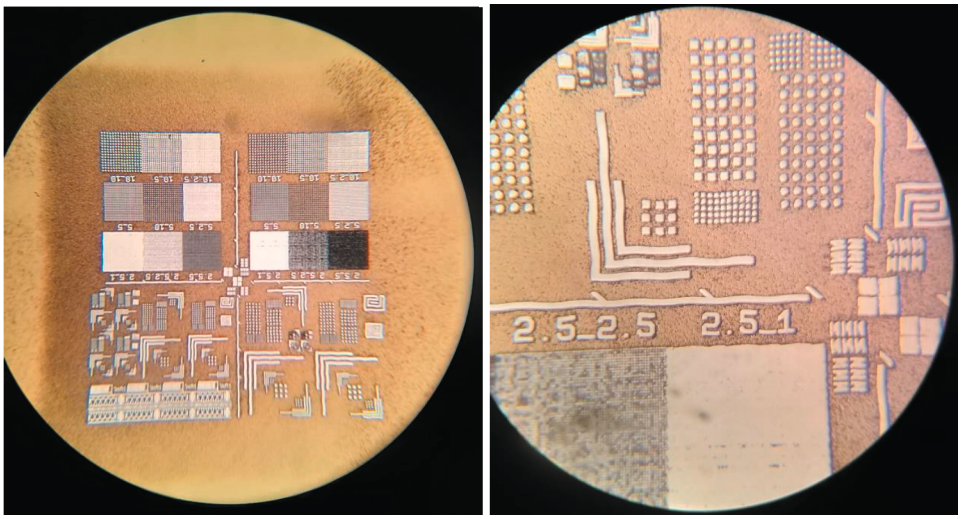

Fig. 4. The result of developing the exposure of a test holographic image 2.5 × 2.5 mm^2 in a photosensitive material 30 μm thick (NA = 0.24; λ = 441.6 nm; resolution 5 μm).

can set the task of optimizing image quality in several parallel planes. Thus,

$$\sigma(A, H(V)) = \sum_n \sigma_n(A_n, H(V)) \underset{V}{\to} \min,$$

where A_n are the sections of an ideal image by planes and σ_n is the restriction of the quality functional to a given plane. Numerical solutions to optimization problems in some of these formulations were implemented in a software package. This makes it possible to solve the problem of synthesizing hologram masks for complex three-dimensional images, as well as the problem of optimizing the depth of focus of an optical system. To verify the software package and the mathematical models used on its basis, a series of experiments were carried out [10, 11], during which hologram masks were synthesized with its help, manufactured on an electronic lithograph, and installed in an optical circuit with optics and an illuminator predesigned for illumination. One of the latest series of experiments was devoted to obtaining images in thick (10–50 μm) photoresists at a reconstruction radiation wavelength of 441.6 nm (Fig. 4).

FUNDING

This work was supported by ongoing institutional funding. No additional grants to carry out or direct this particular research were obtained.

CONFLICT OF INTEREST

The author of this work declares that he has no conflicts of interest.

REFERENCES

1. M. V. Borisov, V. A. Borovikov, A. A. Gavrikov, D. Yu. Knyaz'kov, V. I. Rakhovskii, D. A. Chelyubeev, and A. S. Shamaev, Dokl. Phys. **55**, 436 (2010).
2. V. Rakhovsky, D. Knyazkov, A. Shamaev, V. Chernik, A. Gavrikov, D. Chelyubeev, P. Mikheev, and M. Borisov, in *Proceedings of the European Mask and Lithography Conference,* Proc. SPIE **8352**, 83520P (2012).
3. V. V. Chernik, Zh. Radioelektron., No. 1, 1 (2017).
4. D. A. Gabor, Nature (London, U.K.) No. 161, 777 (1948).
5. H. J. Caulfield, *Handbook of Optical Holography* (Academic, New York, 1979).
6. C. A. Mack, *Fundamental Principles of Optical Lithography: The Science of Microfabrication* (Wiley, New York, 2007).
7. F. G. Bass and I. M. Fuks, *Wave Scattering at a Statistically Rough Surface* (Pergamon, New York, 1979).
8. V. V. Chernik, in *Proceedings of the 6th International Conference on Parallel Computations and Control Problems, 2012* (2012), Vol. 2.
9. P. A. Mikheev, Mosc. Univ. Comput. Math. Cybern. **38**, 8 (2014).
10. M. Borisov, V. Chernik, L. Merkushov, A. Shamaev, V. Rakhovski, and D. Chelubeev, Proc. SPIE **11324**, 113241J (2020). https://doi.org/10.1117/12.2552013
11. M. Borisov, V. Chernik, A. Shamaev, V. Rakhovski, and D. Chelubeev, Proc. SPIE **11324**, 113241 (2020). https://doi.org/10.1117/12.2551936