Analytic optimization for holographic optical elements

E. Hasman and A. A. Friesem

Department of Electronics, Weizmann Institute of Science, Rehovot 76100, Israel

Received February 11, 1988; accepted August 11, 1988

A method is presented for designing optimal holographic optical elements. The method is based on an analytic ray-tracing procedure that uses the minimization of the mean-squared difference of the propagation vector components between the actual output wave fronts and the desired output wave fronts. The minimization yields integral equations for the grating vector components that can be solved analytically, in some cases without any approximation. This leads to a well-behaved grating function that defines a holographic optical element.

INTRODUCTION

In an optical system that is designed to operate with monochromatic or quasi-monochromatic illumination sources, it is possible to replace the conventional refractive elements with holographic optical elements (HOE's) that are based on diffractive optics.[1] In general, the HOE's transform a given set of waves into another set of waves, and their performance is ideal when the readout geometry and wavelength are identical to the recording geometry and wavelength. However, in general, the geometries and wavelengths differ so that severe aberrations occur.

In order to minimize the aberrations, it is necessary to use optimization procedures for designing and recording a holographic element having a complicated grating function. Several procedures have been proposed. These are based on numerical iterative ray-tracing techniques[2] or on minimizing the mean-squared difference of the phases of the actual output wave fronts and the desired output wave fronts.[3–5] In the iterative ray tracing, extensive calculations of ray directions are required, and the solutions often converge to local minima rather than to the desired absolute minimum. In the phase minimization, the phase must be defined up to an additive constant so that the optimization procedure becomes rather complicated and it is necessary to resort to approximate solutions. As a result, such optimization procedures do not yield an exact solution except in specific cases.[6]

In this paper we present a different optimization procedure in which the design is based on analytical ray tracing that minimizes the mean-squared difference of the propagation vector components between the actual output wave fronts and the desired output wave fronts. The mean-squared difference of the vector components is defined in such a way that the functions involved are continuous. Specifically, we define continuous input parameters that characterize the propagation vector components of each wave front. We then obtain integral equations for the optimal grating vector components that can be solved analytically, in some cases without any approximation.

To illustrate our method we have designed Fourier-transform and imaging lenses. The performance of the lenses is analyzed by ray tracing and then is evaluated with a merit function that characterizes the quality of the lenses. The results are then compared with those of quadratic HOE's and with conventional HOE's recorded with spherical wave fronts, to illustrate the superiority of the newly designed lenses.

BASIC RELATIONS FOR OPTIMIZATION

A HOE can be described as a diffractive grating that transforms the phase of an incoming wave front to another output phase. The phase of the output wave front, ϕ_o, is given by

$$\phi_o = \phi_i \pm n\phi_h, \tag{1}$$

where ϕ_i is the phase of the input wave front, ϕ_h is the grating function of the HOE, and n represents various diffractive orders from the holographic grating. In general, only the first diffracted order is of interest, so that $n = 1$. To proceed, we now use the normalized propagation vectors and the grating vector of the holographic element, rather than the phases. The normalized propagation vectors, which can be regarded as the direction cosines of the input and output rays, can be written as

$$\hat{\bar{K}}_o = \frac{\lambda_{\text{read}}}{2\pi}\bar{\nabla}\phi_o, \qquad \hat{\bar{K}}_i = \frac{\lambda_{\text{read}}}{2\pi}\bar{\nabla}\phi_i, \tag{2}$$

and the grating vector can be written as

$$\bar{K}_h = \frac{\lambda_{\text{rec}}}{2\pi}\bar{\nabla}\phi_h = \frac{\lambda_{\text{rec}}}{\Lambda_x}\hat{x} + \frac{\lambda_{\text{rec}}}{\Lambda_y}\hat{y}, \tag{3}$$

where λ_{read} is the readout wavelength, λ_{rec} is the recording wavelength, $\bar{\nabla}$ is the gradient operator, and Λ_x and Λ_y are the grating spacings in the x and y directions.[7] The diffraction relations can now be written as

$$\hat{K}_{x_o} = \hat{K}_{x_i} - \mu K_{x_h}, \tag{4}$$

$$\hat{K}_{y_o} = \hat{K}_{y_i} - \mu K_{y_h}, \tag{5}$$

$$\hat{K}_{z_o} = \pm(1 - \hat{K}_{x_o}^2 - \hat{K}_{y_o}^2)^{1/2}, \tag{6}$$

where $\mu = \lambda_{\text{read}}/\lambda_{\text{rec}}$. Note that $\hat{K}_{x_o}^2 + \hat{K}_{y_o}^2$ should be <1 so that evanescent wave fronts are not obtained.

The goal when one is designing HOE's is to transfer input

0740-3232/89/010062-11$02.00

rays into corresponding output rays that will be optimized for a given range of input parameters. The input parameter could, for example, be the direction cosine of the incoming waves or the location of the input point sources. For a single specific input parameter it is relative easy to form a HOE that will yield the exact output rays. However, for a range of input parameters it is necessary to optimize the grating vector so as to minimize the difference between the actual output rays and the desired rays. The optimization is achieved by minimizing the mean-squared difference between these two sets of rays. In general, the optimization procedure can include the effects of the pupil function of the HOE, different optimization weighting for each input parameter, and readout with broad spectral illumination.

To simplify the presentation of our optimization method, we first describe the method in one-dimensional notation and then extend it to two dimensions.

One-Dimensional Optimization

The mean-squared difference of the propagation vectors is defined as

$$E^2 = \int\int\int W_\mu(\mu)W(a)P(x,a,\mu)[\hat{K}_{x_d}(x,a,\mu) - \hat{K}_{x_o}(x,a,\mu)]^2 d\mu da dx, \quad (7)$$

where the direction cosines of the output and desired rays, $\hat{K}_{x_o}(x, a, \mu)$ and $\hat{K}_{x_d}(x, a, \mu)$, depend on μ and the input parameter a and where x is the space coordinate on the HOE. The pupil function denoted by $P(x, a, \mu)$ is generally a binary function, the optimization weighting function for each input parameter is given by $W(a)$, where $0 \le W(a) \le 1$, and $W_\mu(\mu)$ denotes the weighting function for the readout wavelengths, where $0 \le W_\mu(\mu) \le 1$. Inserting Eq. (4) into Eq. (7) yields

$$E^2 = \int\int\int W_\mu(\mu)W(a)P(x,a,\mu)[\hat{K}_{x_d}(x,a,\mu) - \hat{K}_{x_i}(x,a,\mu) + \mu K_{x_h}(x)]^2 d\mu da dx. \quad (8)$$

The optimal grating vector component $K_{x_h}(x)$ can be determined by minimizing E^2. It is sufficient, however, to minimize a simpler integral that we denote by $e^2(x_o)$:

$$e^2(x_o) = \int\int W_\mu(\mu)W(a)P(x_o,a,\mu)[\hat{K}_{x_d}(x_o,a,\mu) - \hat{K}_{x_i}(x_o,a,\mu) + \mu K_{x_h}(x_o)]^2 d\mu da, \quad (9)$$

where x_o represents an arbitrary coordinate x. Differentiating $e^2(x_o)$ with respect to $K_{x_h}(x_o)$ and setting the result to zero yields

$$\int\int W_\mu(\mu)W(a)P(x_o,a,\mu)[\hat{K}_{x_d}(x_o,a,\mu) - \hat{K}_{x_i}(x_o,a,\mu) + \mu K_{x_h}(x_o)]\mu d\mu da = 0. \quad (10)$$

Thus the optimal grating vector component is

$$K_{x_h}(x) = \frac{-\left\{\int\int W_\mu(\mu)W(a)P(x,a,\mu)\mu[\hat{K}_{x_d}(x,a,\mu) - \hat{K}_{x_i}(x,a,\mu)]da d\mu\right\}}{\left[\int\int W_\mu(\mu)W(a)P(x,a,\mu)\mu^2 da d\mu\right]}. \quad (11)$$

Since the second derivative of e^2 is greater than zero, the optimal grating vector yields minimum e^2. The corresponding optimal grating function can be found by using Eq. (3):

$$\phi_h(x) = \frac{2\pi}{\lambda_{rec}} \int K_{x_h}(x)dx. \quad (12)$$

Usually the readout illumination is monochromatic, so that $W_\mu(\mu) = \delta(\mu - \mu_o)$, where $\delta(\mu)$ is the delta function and $\mu_o = \lambda_{read}/\lambda_{rec}$ = constant. Thus Eq. (11) simplifies to

$$K_{x_h}(x) = \frac{-\int W(a)P(x,a)[\hat{K}_{x_d}(x,a) - \hat{K}_{x_i}(x,a)]da}{\mu_o \int W(a)P(x,a)da}. \quad (13)$$

The function $K_{x_h}(x)$ depends on the $W(a)$, $P(x, a)$, $\hat{K}_{x_d}(x, a)$, and $\hat{K}_{x_i}(x, a)$, all of which are given for a specific element design.

Two-Dimensional Extension

The mean-squared difference of the propagation vector components, as given in Eq. (7), can be extended to two dimensions. For monochromatic readout, it includes two scalar equations, one for E_x^2 and the other for E_y^2, and is written as

$$E_\perp{}^2 = \int\int\int\int W(a)W(b)P(x,y,a,b)[\hat{K}_{\perp_d}(x,y,a,b) - \hat{K}_{\perp_o}(x,y,a,b)]^2 da db dx dy, \quad (14)$$

where $\perp$ denotes the transverse vector components $\hat{x}$ and $\hat{y}$ and where a and b are the input parameters for the x and y coordinates, respectively. Equation (14) can be expanded, by using Eqs. (4) and (5), to

$$E_\perp{}^2 = \int\int\int\int W(a)W(b)P(x,y,a,b)[\hat{K}_{\perp_d}(x,y,a,b) - \hat{K}_{\perp_i}(x,y,a,b) + \mu_o K_{\perp_h}(x,y)]^2 da db dx dy. \quad (15)$$

As was done for the one-dimensional optimization, the optimal two-dimensional grating vector, $\overline{K}_h(x, y)$, is obtained by minimizing $E_\perp{}^2$ to yield

$$K_{\perp_h}(x,y) = \frac{-\int\int W(a)W(b)P(x,y,a,b)[\hat{K}_{\perp_d}(x,y,a,b) - \hat{K}_{\perp_i}(x,y,a,b)]da db}{\mu_o \int\int W(a)W(b)P(x,y,a,b)da db}. \quad (16)$$

Then, the two-dimensional grating function, $\phi_h(x, y)$, is found by integrating along some arbitrary path to yield

$$\phi_h(x, y) - \phi_h(0, 0) = \frac{2\pi}{\lambda_{\text{rec}}} \int_c \bar{K}_h \cdot d\bar{r} = \frac{2\pi}{\lambda_{\text{rec}}} \int_0^{(x,y)} K_{x_h}(x, y)dx + K_{y_h}(x, y)dy, \tag{17}$$

where $\phi_h(0, 0)$ can be defined to be zero.

For a unique solution, the condition of $\nabla_\perp \times \bar{K}_h = 0$ must be fulfilled, where the gradient $\nabla_\perp$ denotes $(\partial/\partial x)\hat{x} + (\partial/\partial y)\hat{y}$. This condition can be written explicitly as

$$\frac{\partial K_{x_h}(x, y)}{\partial y} = \frac{\partial K_{y_h}(x, y)}{\partial x}. \tag{18}$$

For an on-axis holographic element having circular symmetry, this condition is fulfilled; however, in general, for off-axis elements this condition cannot be fulfilled, so that an exact solution for the grating function $\phi_h(x, y)$ cannot be found. Nevertheless, it is possible to obtain approximate solutions. For example, when the off-axis angle is relatively low, it is possible to approximate the grating function by simply adding a linear phase term to the one-dimensional on-axis design,[4,6]

$$\phi_h(x, y) = [\phi_h(r)]_{\text{on-axis}} + \frac{2\pi}{\lambda_{\text{rec}}} (\sin \theta_r)x. \tag{19}$$

In Eq. (19) θ_r is the off-axis angle and

$$[\phi_h(r)]_{\text{on-axis}} = \frac{2\pi}{\lambda_{\text{rec}}} \int K_{r_h}(r)dr, \tag{20}$$

where $K_{r_h}(r)$ is essentially the same as given by Eq. (13) after replacement of x with r, where $r = (x^2 + y^2)^{1/2}$. Alternatively, since in many cases $K_{x_h}(x, y)$ has weak dependence on the y coordinate and $K_{y_h}(x, y)$ has weak dependence on the x coordinate over the whole holographic element's area, these two-dimensional grating vector components can be approximated by one-dimensional components $K_{x_h}(x, y = 0)$ and $K_{y_h}(x = 0, y)$. Consequently, the two-dimensional grating function, $\phi_h(x, y)$, can be approximated by two separate one-dimensional functions, as

$$\phi_h(x_o, y_o) = [\phi_h(x_o)]_{\text{off-axis}} + [\phi_h(y_o)]_{\text{on-axis}} = \frac{2\pi}{\lambda_{\text{rec}}} \left[\int_{x=0}^{x_o} K_{x_h}(x, y = 0)dx + \int_{y=0}^{y_o} K_{y_h}(x = 0, y)dy \right], \tag{21}$$

where the one-dimensional grating vector components are found according to Eq. (13). The approximation of Eq. (21) is more general than that of Eq. (19), since it is also valid for higher off-axis angles.

Analysis of Aberrations

In order to determine the aberrations for a holographic element, it is necessary to determine the difference between the phase of the output wave front and the phase of the desired wave front, given in one dimension as

$$\psi(x) = \phi_d(x) - \phi_o(x). \tag{22}$$

In our notation the relevant angular transverse aberrations, $\Upsilon(x, a)$, are thus given by[8]

$$\Upsilon(x, a) = \frac{\lambda_{\text{read}}}{2\pi} \frac{\partial \psi}{\partial x} = \hat{K}_{x_d}(x, a) - \hat{K}_{x_o}(x, a) \simeq D + Ax + Cx^2 + Sx^3 + \ldots, \tag{23}$$

where D denotes the distortion, A denotes the coefficient of the field curvature and astigmatism, C denotes the coefficient of the coma, and S denotes the coefficient of the spherical aberrations.

In general, the optimization procedures reduce all aberrations simultaneously. It is, however, possible to obtain some trade-off among the various aberrations.[5] For example, in some applications it may be more important to reduce the coma and astigmatism at the expense of the distortion. Such a trade-off can be incorporated into our optimization method by choosing a different desired propagation vector component $\hat{K}_{x_d}'$. Specifically,

$$\hat{K}_{x_d}' = \hat{K}_{x_d} + \sum_{i=1}^{m} \sum_{j=0}^{l} c_{ij} a^i x^j = \hat{K}_{x_d} + \sum_{i=1}^{m} c_{i0} a^i + \sum_{i=1}^{m} c_{i1} a^i x + \ldots. \tag{24}$$

If, for example, we optimize the design for all aberrations but allow the distortion to be arbitrary, then l is zero; thus only the c_{i0} coefficients must be optimized. The optimal c_{i0} coefficients can be found by minimizing the mean-squared difference of the propagation vector components given in Eq. (8); i.e., differentiating with respect to c_{i0} and setting the result to zero,

$$\frac{dE^2}{dc_{i0}} = 0. \tag{25}$$

Alternatively, we could allow additional or other aberrations to be arbitrary, so that different coefficients must be optimized.

FOURIER-TRANSFORM LENS

The operation of a holographic Fourier-transform lens (FTL) is described with the aid of the one-dimensional representation in Fig. 1. The input object is a transparency that is illuminated with a coherent beam; each spatial-frequency component can be represented by a plane wave emerging from the transparency at a certain angle, θ_i. The FTL focuses each of these plane waves to a point at the output plane corresponding to the angular direction of the waves.[9] The width of the input transparency aperture is $2W$, and it extends from coordinate W_2 to W_1, whereas the holographic lens aperture extends from coordinate D_1 to D_2. Finally, d_o and d_i are the distances from the holographic element to the object and output plane, respectively.

It is convenient for FTL design to let the input parameter a be a direction cosine of the plane waves emerging from the transparency, so

$$a = \alpha = \sin \theta_i. \tag{26}$$

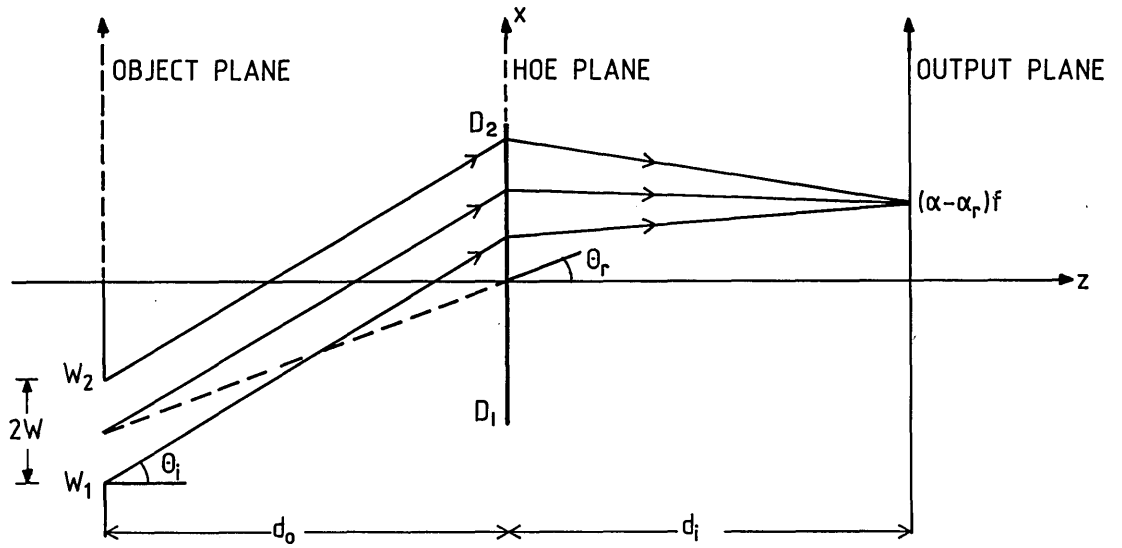

Fig. 1. Readout geometry for an off-axis FTL.

Consequently the normalized propagation vector of the input rays is

$$\hat{K}_{x_i}(x, a) = \hat{K}_{x_i}(\alpha) = \alpha. \tag{27}$$

Now, an input plane wave having a direction cosine α must be transformed at a distance d_i into a spherical wave converging to a point $(\alpha - \alpha_r)f$, where $\alpha_r = \sin\theta_r$, θ_r is the off-axis angle, and f is a proportionality constant. The direction cosines of the desired output rays then become

$$\hat{K}_{x_d}(x, a) = \hat{K}_{x_d}(x, \alpha) = \frac{-[x - (\alpha f - \alpha_r f)]}{\{[x - (\alpha f - \alpha_r f)]^2 + d_i^2\}^{1/2}}. \tag{28}$$

It is often appropriate to modify $\hat{K}_{x_d}$ by permitting arbitrary distortion that varies with the input parameters. Thus using Eq. (24) yields

$$\hat{K}_{x_d}' = \hat{K}_{x_d} + \sum_{i=1}^{m} c_{i0}\alpha^i. \tag{29}$$

For simplicity, we retain only the first coefficient c_{10}, so Eq. (29) simplifies to

$$\hat{K}_{x_d}' \simeq \frac{-[x - (\alpha f - \alpha_r f)]}{\{[x - (\alpha f - \alpha_r f)]^2 + d_i^2\}^{1/2}} + c_{10}\alpha. \tag{30}$$

Substituting $\hat{K}_{x_i}$ from Eq. (27) and the modified $\hat{K}_{x_d}$ from relation (30) into Eq. (13), and for $W(a) = 1$ and $\mu_0 = 1$, we obtain

$$K_{x_h}(x, c_{10}) = -\frac{1}{[\alpha_2(x) - \alpha_1(x)]} \times \int_{\alpha_1(x)}^{\alpha_2(x)} \left(\frac{-[x - (\alpha f - \alpha_r f)]}{\{[x - (\alpha f - \alpha_r f)]^2 + d_i^2\}^{1/2}} + c_{10}\alpha - \alpha \right) d\alpha, \tag{31}$$

where the pupil function, $P(x, a)$ in Eq. (13), is expressed by the upper $[\alpha_2(x)]$ and lower $[\alpha_1(x)]$ direction cosines of the input plane waves that intercept the holographic lens at a point x. The detailed expressions for $\alpha_1(x)$ and $\alpha_2(x)$ are given in Appendix A.

The solution for Eq. (31) is

$$K_{x_h}(x, c_{10}) = \frac{(1 - c_{10})[\alpha_1(x) + \alpha_2(x)]}{2} - \frac{1}{[\alpha_2(x) - \alpha_1(x)]f} \times (\{[x + \alpha_r f - \alpha_2(x)f]^2 + d_i^2\}^{1/2} - \{[x + \alpha_r f - \alpha_1(x)f]^2 + d_i^2\}^{1/2}). \tag{32}$$

The coefficient c_{10} is found from the simplified relation of Eq. (8),

$$E^2 = \int_{D_1}^{D_2} \int_{\alpha_1(x)}^{\alpha_2(x)} [\hat{K}_{x_d} + c_{10}\alpha - \hat{K}_{x_i} + K_{x_h}(x, c_{10})]^2 d\alpha dx, \tag{33}$$

and, with Eq. (25), yields

$$c_{10} = \frac{-\int_{D_1}^{D_2} \int_{\alpha_1(x)}^{\alpha_2(x)} [\hat{K}_{x_d} - \hat{K}_{x_i} + K_{x_h}(c_{10} = 0)] \left[\alpha - \frac{\alpha_1(x) + \alpha_2(x)}{2} \right] d\alpha dx}{\int_{D_1}^{D_2} \int_{\alpha_1(x)}^{\alpha_2(x)} \left[\alpha - \frac{\alpha_1(x) + \alpha_2(x)}{2} \right]^2 d\alpha dx}. \tag{34}$$

The solution given by Eq. (32) is rather general and can be simplified to obtain approximate optimal designs. For example, it is possible to expand Eq. (32) by assuming the paraxial approximation for large f/x and a small off-axis angle and setting $d_o = di = f$ and $c_{10} = 0$ to obtain the approximate solution of the quadratic off-axis FTL,[6] given by

$$[K_{x_h}(x)]_{\text{simplified}} \simeq \frac{x}{f} + \alpha_r = [K_{x_h}(x)]_q. \tag{35}$$

As an illustration, we evaluated the performance of the on-axis and off-axis one-dimensional optimal FTL's given by Eqs. (32) and (34), by using a ray-tracing analysis[7] that is based on Eqs. (4) and (6). The specified parameters for the on-axis FTL were $\theta_r = 0°$, $f = d_i = 60$ mm, $d_o = 60$ mm, $D_2 =$

$-D_1 = 30$ mm, $W = 10$ mm, and $c_{10} = 0.008$, whereas for the off-axis FTL they were $\theta_r = 45°, f = d_i = 60$ mm, $d_o = f\cos\theta_r = 42.5$ mm, $D_2 = -D_1 = 30$ mm, $W = 10$ mm, and $c_{10} = 0.0122$. For comparison, we also performed a ray-tracing analysis for a quadratic FTL [Eq. (35)] as well as for a spherical FTL for which

$$[K_{x_h}(x)]_{\text{sph}} = \frac{x}{(x^2 + f^2)^{1/2}} + \alpha_r. \tag{36}$$

The specified parameters for these were the same as those described above.

The detailed results for the spot sizes as a function of the input angles for the optimal, the quadratic, and the spherical FTL's are shown in Figs. 2 and 3; these results do not take into account the diffraction from the aperture. The spot sizes were determined by calculating the standard deviation of the location of the rays at the output plane as a function of the angular directions for each input plane wave. Figure 2 shows the results for the on-axis format; the results shown in

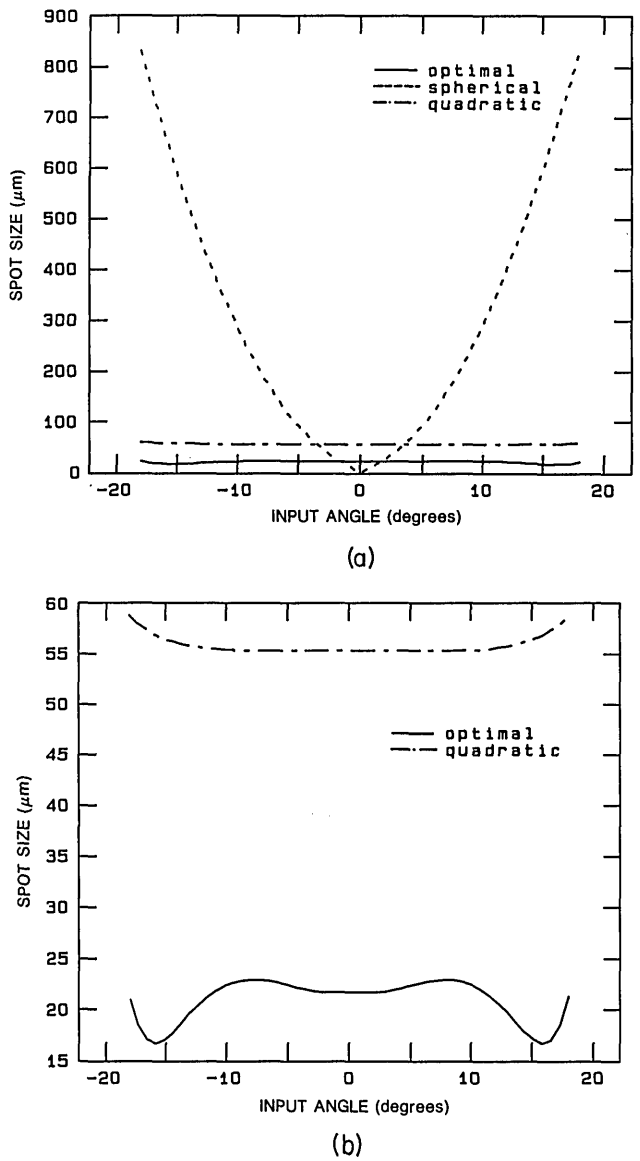

Fig. 2. Spot size as a function of the input angle for an on-axis FTL: (a) spherical, quadratic, and optimal grating functions; (b) magnified region of optimal and quadratic grating functions.

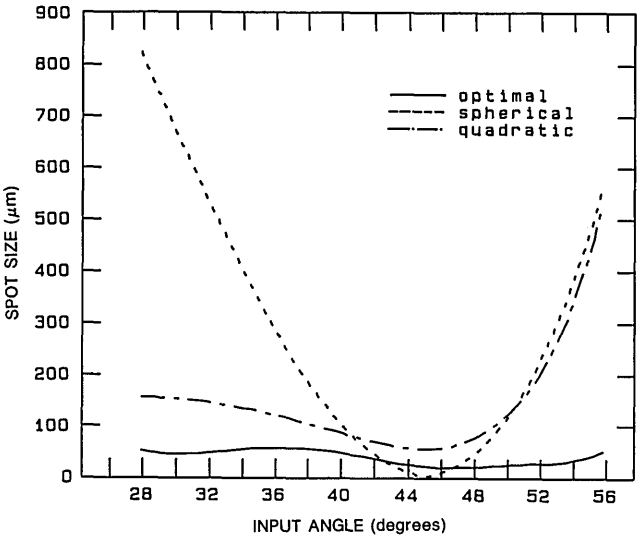

Fig. 3. Spot size as a function of the input angle for an off-axis FTL: spherical, quadratic, and optimal grating functions.

Fig. 2(b) are merely a magnification of the results in the lower portion of Fig. 2(a). As shown, the spot sizes for the optimal FTL are somewhat bettter than those for the quadratic lens over the entire range of input angles. They are significantly better than those for the spherical lens. Figure 3 shows the results for the off-axis format. The spot sizes for the optimal FTL are uniform over the entire range of input angles, and they are obviously better than for the other two lenses, especially at larger input angles. In the off-axis format, the performance of the quadratic lens approaches the poor performance of the spherical lens as the off-axis angle increases.

Analysis of Aberrations

In order to identify and evaluate the various aberrations for the FTL's, we begin with Eq. (4) and relation (23) to obtain the angular transverse aberration as

$$\Upsilon = \hat{K}_{x_d} - \hat{K}_{x_i} + \mu K_{x_h}. \tag{37}$$

Considering only third-order aberrations,[8] the normalized propagation vectors $\hat{K}_{x_d}$ and $\hat{K}_{x_i}$ for the on-axis FTL with $d_i = d_o = f$ become

$$\hat{K}_{x_d} = \frac{-(x - \alpha f)}{[(x - \alpha f)^2 + f^2]^{1/2}}$$

$$\simeq \alpha - \frac{1}{2}\alpha^3 + \left(\frac{3}{2}\alpha^2 - 1\right)\frac{x}{f} - \frac{3}{2}\alpha\left(\frac{x}{f}\right)^2 + \frac{1}{2}\left(\frac{x}{f}\right)^3, \tag{38}$$

and $\hat{K}_{x_i}$ is given by Eq. (27). The grating vector for the optimal FTL is the same as for the quadratic FTL,

$$(K_{x_h})_{\text{opt}} \simeq \frac{x}{f}, \tag{39}$$

and the grating vector for the spherical FTL is

$$(K_{x_h})_{\text{sph}} = \frac{x}{(x^2 + f^2)^{1/2}}$$

$$\simeq \frac{x}{f} - \frac{1}{2}\left(\frac{x}{f}\right)^3. \tag{40}$$

Now, for the FTL format as shown in Fig. 1, we need not find the aberrations for the entire hologram but can resort to subareas that are referred to as local holograms. The center for each such local hologram, $x_o(\theta_i)$, is

$$x_o(\theta_i) = f \tan\theta_i = f\frac{\alpha}{(1-\alpha^2)^{1/2}}. \tag{41}$$

Consequently at each local hologram the overall aberrations for the optimal FTL can be separated[8] into distortion (D), astigmatism and field curvature (A), coma (C), and spherical aberration (S) as

$$\Upsilon_{\text{opt}}{}^{D} = -\frac{1}{2}\alpha^3 + \left(\frac{3}{2}\alpha^2 + \mu - 1\right)\frac{\alpha}{(1-\alpha^2)^{1/2}} - \frac{3}{2}\frac{\alpha^3}{(1-\alpha^2)} + \frac{1}{2}\frac{\alpha^3}{(1-\alpha^2)^{3/2}}, \tag{42}$$

$$\Upsilon_{\text{opt}}{}^{A} = \left[\left(\frac{3}{2}\alpha^2 + \mu - 1\right) - \frac{3\alpha^2}{(1-\alpha^2)^{1/2}} + \frac{3}{2}\frac{\alpha^2}{(1-\alpha^2)}\right]\frac{(x-x_o)}{f}, \tag{43}$$

$$\Upsilon_{\text{opt}}{}^{C} = \left[\frac{3}{2}\frac{\alpha}{(1-\alpha^2)^{1/2}} - \frac{3}{2}\alpha\right]\left(\frac{x-x_o}{f}\right)^2, \tag{44}$$

$$\Upsilon_{\text{opt}}{}^{S} = \frac{1}{2}\left(\frac{x-x_o}{f}\right)^3. \tag{45}$$

Similarly for the spherical FTL,

$$\Upsilon_{\text{sph}}{}^{D} = -\frac{1}{2}\alpha^3 + \left(\frac{3}{2}\alpha^2 + \mu - 1\right)\frac{\alpha}{(1-\alpha^2)^{1/2}} - \frac{3}{2}\frac{\alpha^3}{(1-\alpha^2)} + \frac{1}{2}(1-\mu)\frac{\alpha^3}{(1-\alpha^2)^{3/2}}, \tag{46}$$

$$\Upsilon_{\text{sph}}{}^{A} = \left[\left(\frac{3}{2}\alpha^2 + \mu - 1\right) - \frac{3\alpha^2}{(1-\alpha^2)^{1/2}} + \frac{3}{2}(1-\mu)\frac{\alpha^2}{(1-\alpha^2)}\right]\frac{(x-x_o)}{f}, \tag{47}$$

$$\Upsilon_{\text{sph}}{}^{C} = \left[-\frac{3}{2}\alpha + \frac{3}{2}(1-\mu)\frac{\alpha}{(1-\alpha^2)^{1/2}}\right]\left(\frac{x-x_o}{f}\right)^2, \tag{48}$$

$$\Upsilon_{\text{sph}}{}^{S} = \left[\frac{1}{2}(1-\mu)\right]\left(\frac{x-x_o}{f}\right)^3. \tag{49}$$

As an illustration, we calculated the various aberrations for the specific FTL having $D_2 = -D_1 = 30$ mm, $f = d_i = d_o = 60$ mm, $W = 10$ mm, and $\mu = 1$. The results of the aberrations (in units of milliradians), as a function of the incident input angle, are shown in Fig. 4. As shown by the results for the spherical FTL in Fig. 4(a), the spherical aberrations are zero, but the other aberrations increase rapidly as the input angle increases. The results for the optimal FTL, shown in Fig. 4(b), indicate that, although the spherical aberrations increase, all the other aberrations decrease significantly. The sum total of the aberrations for the optimal design is evidently much better than for the spherical design.

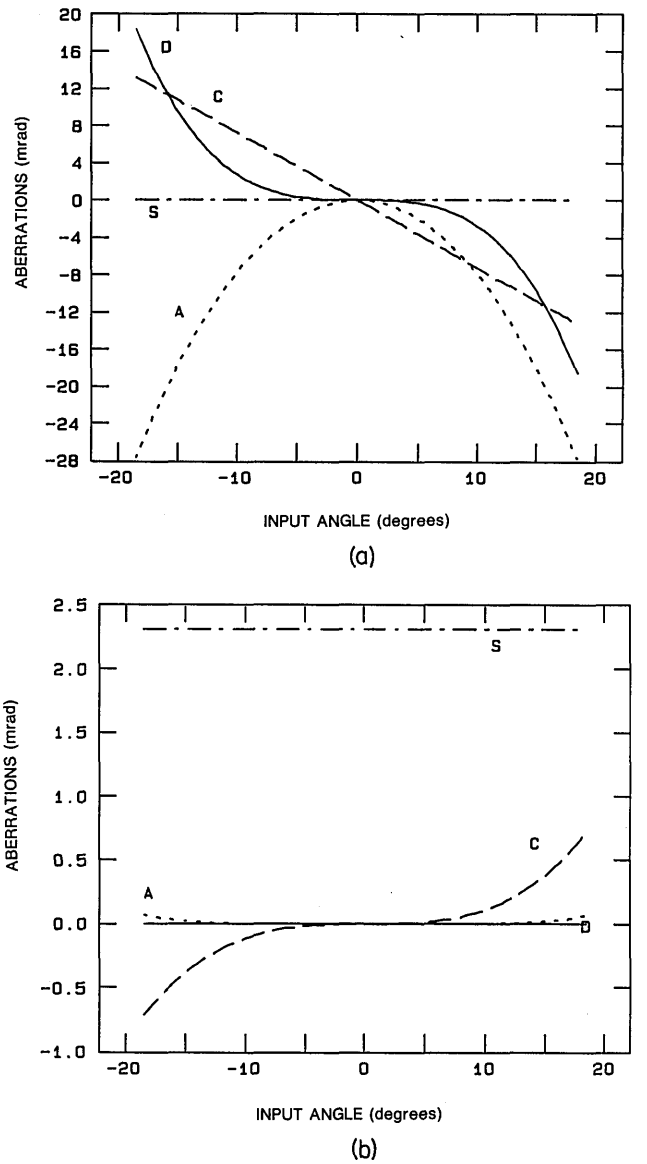

Fig. 4. Aberrations as a function of the input angle for the on-axis FTL. D, distortion; A, astigmatism and field curvature; C, coma; S, spherical aberration. (a) Spherical grating function; (b) optimal grating function.

Extension to a Two-Dimensional Fourier-Transform Lens

In order to extend the design to two dimensions, we used the general approximation given by Eq. (21). The grating vector components K_{x_h} and K_{y_h} were found first by using Eq. (32) and Appendix A, where, for K_{y_h}, we let $\alpha_r = 0$, replaced the variable x by the variable y, and used $\alpha_1(y)$ and $\alpha_2(y)$ according to the on-axis format. For the specific lens having $(\theta_r)_x = 40°$ and $(\theta_r)_y = 0°$, $f = d_i = 100$ mm, $d_o = f\cos 40° = 76.6$ mm, $D_{2_y} = -D_{1_y} = 20$ mm, $D_{2_x} = 19$ mm, $D_{1_x} = -27$ mm and $W = 5$ mm, we found that the optimal first distortion coefficients are $(c_{10})_x = 0.0023$ and $(c_{10})_y = 0.0011$. We then performed ray-tracing analysis[7] by using Eqs. (4)–(6) for the optimal, as well as for the quadratic and spherical, two-dimensional FTL.

The results, which do not take into account the diffraction from the aperture, are shown in Fig. 5, which shows the spot diagrams for the three lenses as a function of nine discrete input angles of $[\theta_{x_i}, \theta_{y_i}]$, where θ_{x_i} and θ_{y_i} are the complements

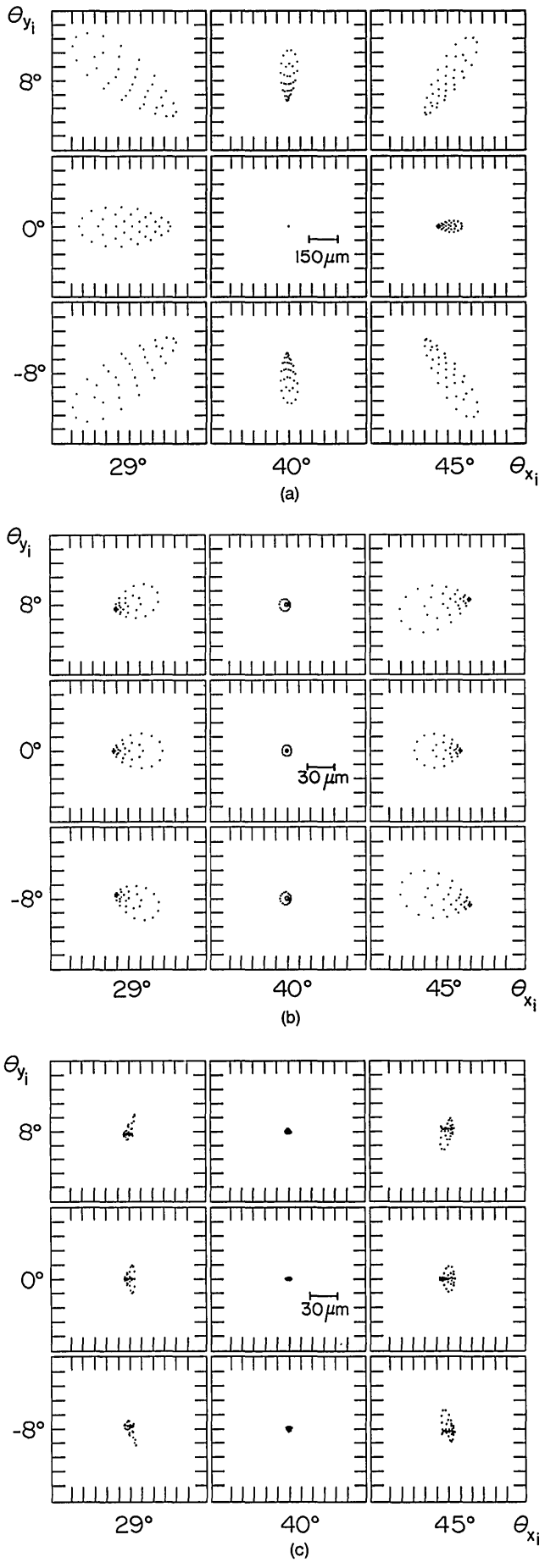

Fig. 5 (left). Spot diagrams for the off-axis FTL. (a) Spherical grating function; (b) quadratic grating function; (c) optimal grating function.

of the angles between the incident ray and the x and y axes, respectively; the ranges of angles θ_{x_i} and θ_{y_i} were $29° < \theta_{x_i} < 45°$ and $-8° < \theta_{y_i} < 8°$, so $\Delta\theta_{x_i} = 16°$ and $\Delta\theta_{y_i} = 16°$. As shown in Fig. 5(a), the small central spot diagram for the spherical lens is essentially ideal because the recording and readout geometries are identical. However, as the readout input angles differ from the recording angles, the spot diagrams spread substantially. Figures 5(b) and 5(c) show the spot diagrams for the quadratic and optimal lenses, respectively; note that the scale is now magnified by a factor of 5. For the quadratic lens, shown in Fig. 5(b), the spread in the spot diagram is generally much smaller than for the spherical lens; the spot diagrams are largest at the extreme angles of $\theta_{x_i} = 45°$. Finally, as shown in Fig. 5(c), it is evident that the optimal lens is uniformly superior to the other lenses, with relatively small spot diagrams even at the extreme angles.

It should be noted that in our optimal design example we introduced some distortions so as to minimize the spot size. Had we desired to reduce the distortions as well, the spot sizes would have increased by ~10%. The overall spot sizes for the optimal element would still be, of course, significantly better than for the other designs. We also calculated the exact distortions for the example of $\theta_{x_i} = 29°$, $\theta_{y_i} = -8°$, and $(c_{10})_x = 0$, $(c_{10})_y = 0$ and found that, for the optimal design, the distortions would be 1 μm in the x direction and 7 μm in the y direction. The corresponding distortions for the spherical design are 631 and 380 μm, respectively.

IMAGING LENS

The operation of a holographic imaging lens (HIL) is described with the aid of one-dimensional representation in Fig. 6. The object with width $2A$ is centered at A_c and extends from coordinates A_1 to A_2. The holographic lens aperture is $2D$, and it is centered about the optical axis Z. The distances from the holographic element to the object and image planes are d_o and d_i, respectively. Finally, a stop aperture is inserted between the object and the imaging lens, at such a distance as to satisfy the relation $d_{o_1}/d_{o_2} = A/D$; the width of the stop aperture is $2W$, and it extends from coordinates W_2 to W_1. For imaging, it is necessary that spherical waves emerging from the object points β intercept the HIL and be transformed into spherical waves that converge to points x_i at the imaging plane. The locations of x_i can be expressed by the general relation

$$x_i = -(\beta + \eta)M, \tag{50}$$

where M denotes the magnification of the imaging lens and η denotes a constant off-axis displacement of the image.

It is convenient in designing the HIL to let the input parameter a be an object point β. Consequently, the normalized propagation vectors of the input rays can be written as

$$\hat{K}_{x_i}(x, a) = \hat{K}_{x_i}(x, \beta) = \frac{(x - \beta)}{[(x - \beta)^2 + d_o^2]^{1/2}}, \tag{51}$$

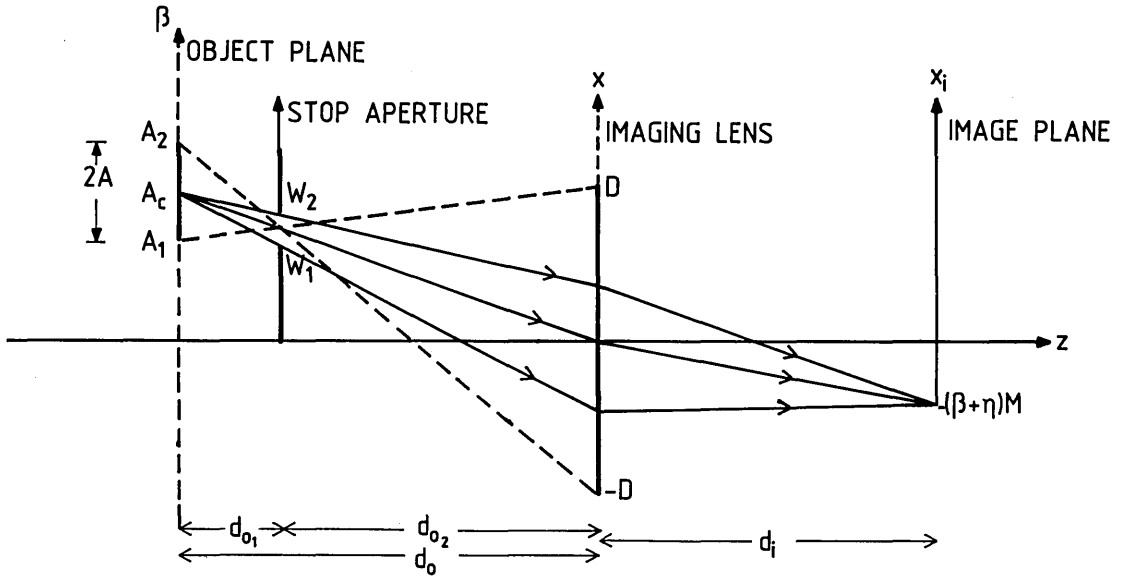

Fig. 6. Readout geometry for an off-axis HIL.

and the normalized propagation vector of the desired output rays becomes

$$\hat{K}_{x_d}(x, a) = \hat{K}_{x_d}(x, \beta) = \frac{-[x + M(\beta + \eta)]}{\{[x + M(\beta + \eta)]^2 + d_i^2\}^{1/2}}. \quad (52)$$

Substituting $\hat{K}_{x_i}$ from Eq. (51) and $\hat{K}_{x_d}$ from Eq. (52) into Eq. (13), and for $W(a) = 1$ and $\mu_o = 1$, we obtain

$$K_{x_h}(x) = -\frac{1}{[\beta_2(x) - \beta_1(x)]} \int_{\beta_1(x)}^{\beta_2(x)} \left(\frac{-[x + M(\beta + \eta)]}{\{[x + M(\beta + \eta)]^2 + d_i^2\}^{1/2}} - \frac{(x - \beta)}{[(x - \beta)^2 + d_o^2]^{1/2}} \right) d\beta, \quad (53)$$

where the pupil function, $P(x, a)$ in Eq. (13), is expressed by the upper $[\beta_2(x)]$ and lower $[\beta_1(x)]$ locations of object points whose rays intercept the holographic lens at a point x. The expressions for $\beta_2(x)$ and $\beta_1(x)$ are given in Appendix B. The solution of Eq. (53) yields

$$K_{x_h}(x) = \frac{1}{[\beta_2(x) - \beta_1(x)]} \left(\frac{1}{M} \{[x + M(\beta_2(x) + \eta)]^2 + d_i^2\}^{1/2} - \frac{1}{M} \{[x + M(\beta_1(x) + \eta)]^2 + d_i^2\}^{1/2} + \{[x - \beta_1(x)]^2 + d_o^2\}^{1/2} - \{[x - \beta_2(x)]^2 + d_o^2\}^{1/2} \right). \quad (54)$$

The general solution given by Eq. (54) can be simplified by expanding and assuming paraxial approximation, and removing the stop aperture, to obtain

$$[K_{x_h}(x)]_{\text{simplified}} \simeq \frac{x}{f} - \left[\frac{A_c}{d_o} + \frac{-M(A_c + \eta)}{d_i} \right], \quad (55)$$

where

$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i}. \quad (56)$$

The corresponding grating vector for the spherical imaging lens design, where the recording of the holographic lens is done with two spherical waves, is

$$[K_{x_h}(x)]_{\text{sph}} = \frac{(x - A_c)}{[(x - A_c)^2 + d_o^2]^{1/2}} + \frac{[x + M(A_c + \eta)]}{\{[x + M(A_c + \eta)]^2 + d_i^2\}^{1/2}}. \quad (57)$$

The grating vector for the quadratic imaging lens design is

$$[K_{x_h}(x)]_q = \frac{x}{f} + \alpha_{r_o} + \alpha_{r_i}, \quad (58)$$

where

$$\alpha_{r_o} = \sin \theta_{r_o} = -\frac{A_c}{(A_c^2 + d_o^2)^{1/2}} \quad (59)$$

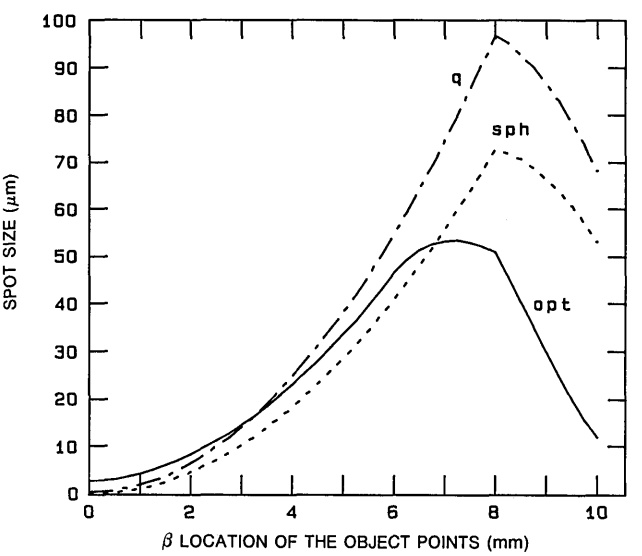

Fig. 7. Spot size as a function of the location of the object points (β) for a HIL with a stop aperture of 2 mm (W = 1 mm): spherical (sph), quadratic (q), and optimal (opt) grating functions.

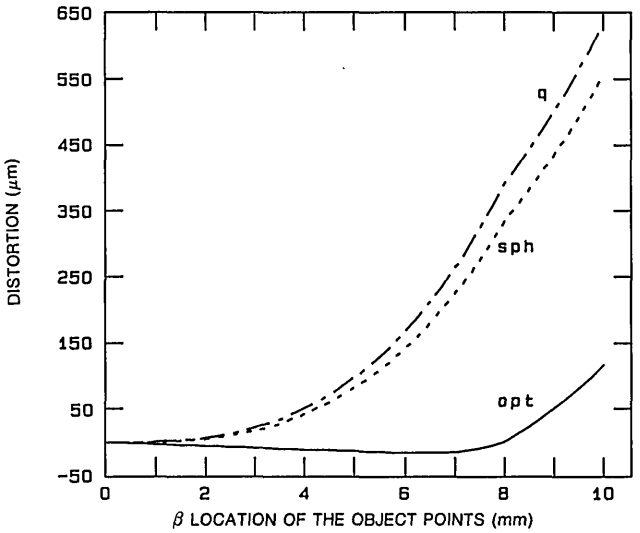

Fig. 8. Distortions as a function of the location of the object points (β) for a HIL with a stop aperture of 2 mm (W = 1 mm): spherical (sph), quadratic (q), and optimal (opt) grating functions.

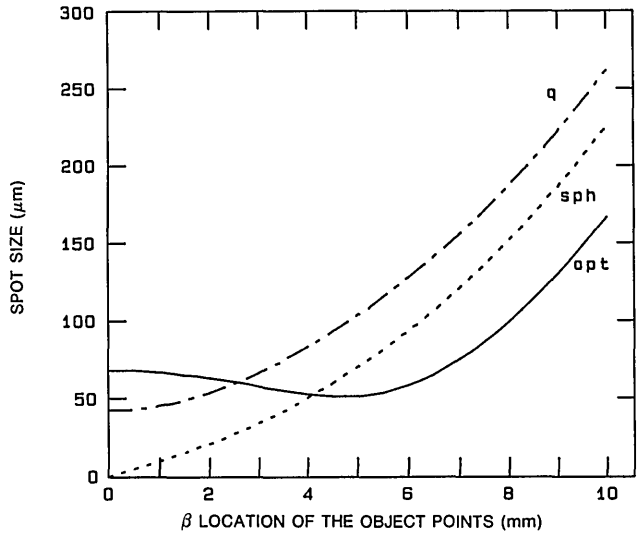

Fig. 9. Spot size as a function of the location of the object points (β), for a HIL without a stop-aperture configuration: spherical (sph), quadratic (q), and optimal (opt) grating functions.

and

$$\alpha_{r_i} = \sin\theta_{r_i} = \frac{M(A_c + \eta)}{\{[M(A_c + \eta)]^2 + d_i^2\}^{1/2}}. \tag{60}$$

It is evident that, for paraxial approximation and small off-axis angles, the optimal, quadratic, and spherical designs yield identical grating vectors.

As an illustration, we calculated the grating vectors and determined the performance of a specific one-dimensional, on-axis imaging lens by using the optimal [Eq. (54)], the spherical [Eq. (57)], and the quadratic [Eq. (58)] designs. The specified parameters of the lens were A_2 = 10 mm, A_1 = −10 mm, D = 10 mm, d_o = 220 mm, d_{o_1} = 110 mm, M = 10, η = 0, and $d_i = Md_o$ = 2200 mm. To determine the spot sizes at the image plane d_i, we performed a ray-tracing analysis, usings Eqs. (4) and (6), and calculated the standard deviation for the locations of the converging rays at the image plane as a function of the location of each object point. We also calculated the amount of distortions by subtracting the actual (average) location of each spot from the desired location as given in Eq. (50). These calculations were performed with a stop aperture of 2 mm (W = 1 mm) as well as without a stop aperture (W = 10 mm).

The calculated results of the spot sizes and the distortions as functions of the object point locations for the optimal, quadratic, and spherical imaging lenses are shown in Figs. 7–10; these results are symmetric about the center of the object plane, and only half the data are shown. Figures 7 and 8 show the results obtained with a stop aperture of 2 mm. As shown in Fig. 7, the spot sizes are comparable for all the designs when the location of the object points is less than 6 mm from the center. Beyond 6 mm the spot sizes are smallest for the optimal design, they deteriorate for the spherical design, and they deteriorate even more for the quadratic

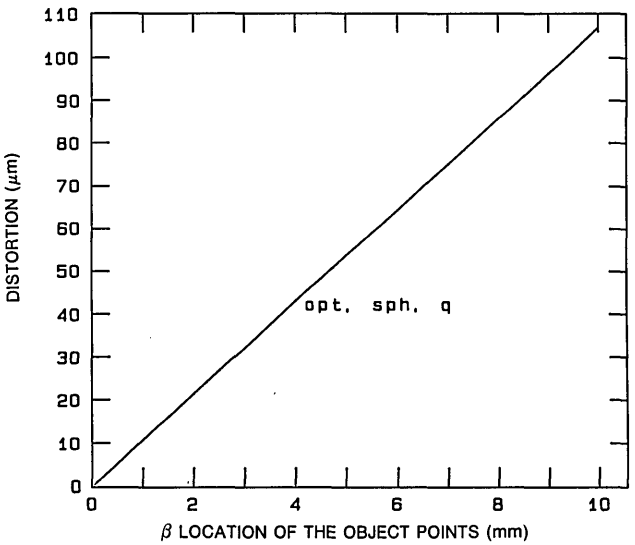

Fig. 10. Distortions as a function of the location of the object points (β), for a HIL without a stop-aperture configuration; spherical (sph), quadratic (q), and optimal (opt) grating functions.

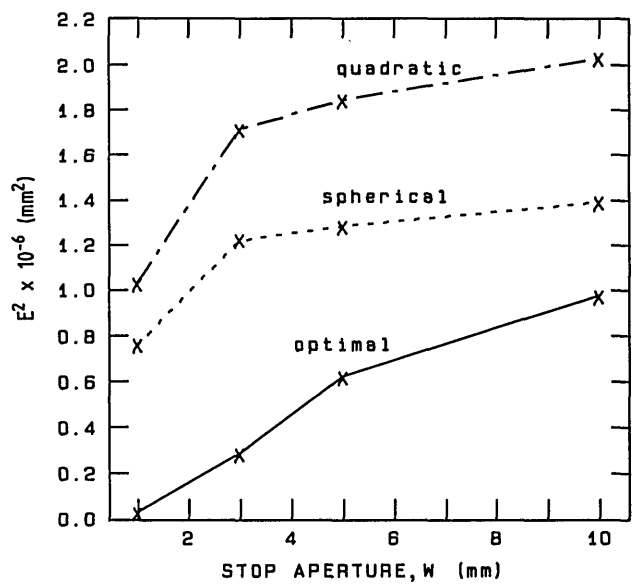

Fig. 11. Mean-squared difference E^2 as a function of the stop aperture (W), for spherical, quadratic, and optimal HIL's.

design. Figure 8 shows that the distortions for the optimal design are significantly smaller than those for the quadratic and spherical designs. Figures 9 and 10 show the results obtained without a stop aperture. As is shown in Fig. 9, there is no significant difference in the spot sizes for any of the three designs. The distortions are identical for all three designs and fall along the same line, as shown in Fig. 10.

It is also possible to evaluate the performance of the designs with the mean-squared difference of the propagation vectors, E^2, from Eq. (8). As an illustration, we calculated the mean-squared difference as a function of the width of the stop aperture. The results for the three designs are shown in Fig. 11. The value of E^2 for the optimal design is lower than for the spherical and quadratic designs, indicating a better lens performance. The improvement in performance is greatest for the smaller stop apertures. For example, when the stop aperture is 2 mm (i.e., $W = 1$ mm) the improvement between the optimal and spherical design, $E_{\text{opt}}^2/E_{\text{sph}}^2$, is 26.3, whereas for no stop aperture (i.e., $W = 10$ mm) the improvement is 1.42. The reason is that, as the stop aperture decreases, the rays from the object intercept a smaller local hologram so that the optimization for each is better.

As in the case of the FTL design, it is possible to introduce some trade-off among the various aberrations [as noted in Eq. (24)] so as to obtain a specific desired performance. Finally, the two-dimensional extension for the on-axis configuration is given by Eq. (16). For the off-axis configuration the general procedure given by Eq. (21) can be followed.

CONCLUSION

We have presented a new method for designing optimal HOE's. It is based on analytic ray tracing and relies on the propagation vectors of the waves and the grating vector rather than on the phases of the waves and the grating function. Our optimization method provides an analytic solution for the optimal grating vector usually without any approximation. Thus we were able to obtain an optimal solution for the grating function for any one-dimensional element and for on-axis two-dimensional elements having circular symmetry; for two-dimensional off-axis elements some approximation is needed to obtain the necessary grating function. The necessary arbitrary grating functions can be realized by resorting to computer-generated or computer-originated holograms.

To illustrate the optimization procedures, FTL's and HIL's were designed and evaluated. The results revealed that lenses designed with our optimization method perform far better than spherical and quadratic holographic lenses. The improvement is even greater when the sizes of the local holograms are reduced with respect to the overall holographic element.

APPENDIX A: EXPRESSIONS FOR THE EXTREME DIRECTION COSINES $\alpha_1(x)$ AND $\alpha_2(x)$

The expressions for the lower $[\alpha_1(x)]$ and the upper $[\alpha_2(x)]$ direction cosines that represent the pupil function are given by using the geometry and the notation of Fig. 1.

The lower $[\alpha_1(x)]$ direction cosine is

$$\alpha_1(x) = \frac{(x - W_2)}{[(x - W_2)^2 + (\Delta y)^2 + d_o^{\,2}]^{1/2}} \tag{A1}$$

when

$$\alpha_1(x) > \alpha_{\min};$$

otherwise

$$\alpha_1(x) = \alpha_{\min}. \tag{A2}$$

In Eq. (A1), Δy is defined by $\Delta y \equiv y_{\text{obj}} - y$, where y_{obj} represents the transparency coordinate at the object plane and y represents the coordinate at the hologram plane. For a one-dimensional grating function, $\Delta y = 0$, whereas, for a two-dimensional grating function (approximated by two separated one-dimensional grating functions), Δy is chosen to yield a minimum $\alpha_1(x)$. The direction cosine $\alpha_{\min}$ is for the plane wave having the lowest angular direction while still completely intercepting the hologram.

The upper $[\alpha_2(x)]$ direction cosine is

$$\alpha_2(x) = \frac{(x - W_1)}{[(x - W_1)^2 + (\Delta y)^2 + d_o^{\,2}]^{1/2}} \tag{A3}$$

when

$$\alpha_2(x) < \alpha_{\max};$$

otherwise

$$\alpha_2(x) = \alpha_{\max}. \tag{A4}$$

For a one-dimensional grating function, $\Delta y = 0$, whereas for a two-dimensional grating function Δy is chosen to yield the maximum $\alpha_2(x)$. The direction cosine $\alpha_{\max}$ is for the plane wave having the highest angular direction while still completely intercepting the hologram.

APPENDIX B: EXPRESSIONS FOR THE EXTREME LOCATIONS OF OBJECT POINTS $\beta_1(x)$ AND $\beta_2(x)$

The expressions for the upper $[\beta_2(x)]$ and the lower $[\beta_1(x)]$ locations of object points that represent the pupil function are given by using the geometry and the notation of Fig. 6.

The lower $[\beta_1(x)]$ location is

$$\beta_1(x) = \frac{d_{o_1}}{d_{o_2}}\left[-W + A_c\frac{d_{o_2}}{(d_{o_1} + d_{o_2})} - x\right] - W + \frac{A_c d_{o_2}}{(d_{o_1} + d_{o_2})} \tag{B1}$$

when

$$\beta_1(x) \geq A_1;$$

otherwise

$$\beta_1(x) = A_1. \tag{B2}$$

The upper $[\beta_2(x)]$ location is

$$\beta_2(x) = \frac{d_{o_1}}{d_{o_2}}\left[W + A_c\frac{d_{o_2}}{(d_{o_1} + d_{o_2})} - x\right] + W + \frac{A_c d_{o_2}}{(d_{o_1} + d_{o_2})} \tag{B3}$$

when

$$\beta_2(x) \leq A_2;$$

otherwise

$$\beta_2(x) = A_2. \tag{B4}$$

ACKNOWLEDGMENTS

We wish to thank Joseph Kedmi and Nir Davidson for valuable discussions.

REFERENCES

1. D. H. Close, "Holographic optical elements," Opt. Eng. **14,** 408–419 (1975).
2. R. C. Fairchild and R. J. Fienup, "Computer-originated aspheric holographic optical elements," Opt. Eng. **21,** 133–140 (1982).
3. K. A. Winick and J. R. Fienup, "Optimum holographic elements recorded with nonspherical wave fronts," J. Opt. Soc. Am. **73,** 208–217 (1983).
4. J. Kedmi and A. A. Friesem, "Optimized holographic optical elements," J. Opt. Soc. Am. A **3,** 2011–2018 (1986).
5. J. N. Cederquist and J. R. Fienup, "Analytic design of optimum holographic optical elements," J. Opt. Soc. Am. A **4,** 699–705 (1987).
6. J. Kedmi and A. A. Friesem, "Optimal holographic Fourier-transform lens," Appl. Opt. **23,** 4015–4019 (1984).
7. J. N. Latta, "Computer-based analysis of holography using ray tracing," Appl. Opt. **10,** 2698–2710 (1971).
8. J. N. Latta, "Computer-based analysis of hologram imagery and aberrations. I. Hologram types and their nonchromatic aberrations," Appl. Opt. **10,** 599–608 (1971).
9. J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1968).