

Calculation of the transmittance function of a multilevel diffractive optical element considering multiple internal reflections

Hwi Kim

Byoung-ho Lee, FELLOW SPIE
Seoul National University
School of Electrical Engineering
Kwanak-Gu Shinlim-Dong
Seoul 151-744
Korea
E-mail: byoung-ho@snu.ac.kr

Abstract. A new method for calculating the transmittance function of a multilevel diffractive optical element is described. Multiple internal reflections in the diffractive optical element induce phase and amplitude modulation errors that deviate from the conventional transmittance function model. A transmittance function model of a multilevel diffractive optical element is devised, taking these multiple internal reflections into account. The results show that a surface relief profile of a multilevel diffractive optical element without phase modulation errors can be constructed based on the proposed model. © 2004 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.1795817]

Subject terms: transmittance function; diffractive optical element; multiple internal reflections.

Paper 030561 received Nov. 10, 2003; revised manuscript received Mar. 26, 2004; accepted for publication Apr. 19, 2004.

1 Introduction

The transmittance function model is adopted in modeling thin surface-relief-type optical devices such as gratings and lenses in wave optics.¹ In the conventional transmittance function model, the local phase change in an incident optical wave at a spatial point is proportional to the local thickness of the optical device at that point. However, some studies have been conducted to improve the transmittance function models not only of thin, but of thick surface-relief optical devices.^{2–5} Previous studies have concentrated on the analysis and modeling of wavefront modulation and nonparaxial effects, but the effects of multiple reflections inside the devices were neglected in these previous works. For materials with a low refractive index, multiple internal reflections can be negligible, since the Fresnel reflection coefficient is very small, even for an abrupt change in refractive index. However, when the refractive index of a material is high or the incidence of the optical wave is somewhat oblique, the Fresnel reflection coefficient is not very small, and as a result, the effects of multiple internal reflections should be expected to be significant.

One of the most important optical devices, the multilevel diffractive optical element (DOE), usually has an optical function that can be described by the conventional transmittance function model. However, it is certain that multiple reflections occur inside multilevel DOEs in practice. Because of this, consideration of multiple internal reflections is necessary if a more accurate model of the transmittance function of a multilevel DOE is to be obtained in cases of high refractive index or substantial incidence obliquity.

In this paper, a transmittance function model of a multilevel DOE is investigated, taking multiple internal reflections into account. We search for a geometric wave solution for the field distribution on the surface of a multilevel DOE

with the use of the Fresnel reflection and transmission coefficients on the boundary between the air and the material. For an obliquely incident and linearly polarized plane wave, the geometric field distribution on the surface of the DOE can be calculated by simple algebraic manipulation. Based on this analysis, a transmittance function model is proposed, in which phase and amplitude modulation errors inherent in the conventional transmittance function model are removed, and an analytic expression is provided. Problems associated with the construction of the surface relief profile are then treated. We concentrate on the realization of the desired phase modulation distribution, since the phase distribution of the optical wave has a more significant effect on the function of the DOE than the amplitude distribution. This inverse problem can be solved analytically for the case of oblique incidence, but only numerically for the case of normal incidence. A comparison of the conventional and proposed approaches is made.

In Sec. 2, the reflection and transmission of an oblique incident plane wave on the boundary between two different index materials are reviewed. In Sec. 3, the transmittance function model of a multilevel DOE is described, taking multiple internal reflections into account. In Sec. 4, problems associated with the construction of a surface relief structure of a DOE are analyzed, and a relief construction algorithm is devised for obtaining the desired phase distribution.

2 Fresnel Reflection and Transmission Coefficients of Obliquely Incident Plane Wave

The Fresnel reflection and transmission coefficients for an obliquely incident plane wave are reviewed in preparation to solving the problem at hand. For an obliquely incident plane wave these coefficients can be derived with the aid of elementary electromagnetics.⁶ Let us assume that the incident plane wave has an incidence angle of θ , an azimuthal

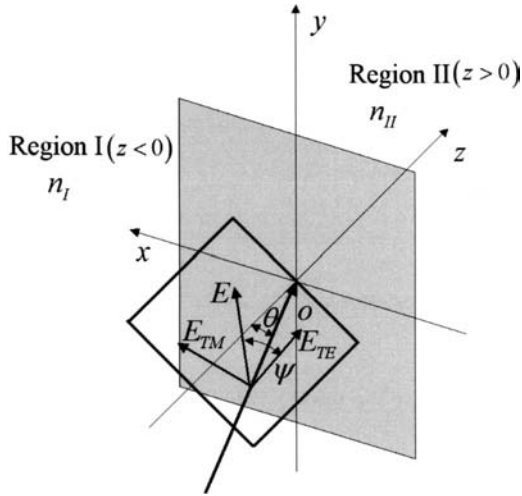


Fig. 1 An incident plane wave having incidence angle θ , azimuthal angle ϕ , and polarization angle ψ propagates from medium I with refractive index n_I to medium II with refractive index n_{II} .

angle of ϕ , and a polarization angle of ψ , as indicated in Fig. 1. The incident wave going from medium I to medium II can be represented as

$$\mathbf{E}_{\text{inc}} = (U_x \hat{\mathbf{x}} + U_y \hat{\mathbf{y}} + U_z \hat{\mathbf{z}}) \exp[jk_0 n_I (x \sin \theta \cos \phi + y \sin \theta \sin \phi + z \cos \theta)], \quad (1)$$

where U_x , U_y , and U_z are given by

$$(U_x, U_y, U_z) = E_0 (\cos \psi \cos \theta \cos \phi - \sin \psi \sin \phi, \cos \psi \cos \theta \sin \phi + \sin \psi \cos \phi, -\cos \psi \sin \theta). \quad (2)$$

Let the reflected and transmitted waves in region I and region II be, respectively, represented as

$$\mathbf{E}_I = \mathbf{E}_{\text{inc}} + (R_x \hat{\mathbf{x}} + R_y \hat{\mathbf{y}} + R_z \hat{\mathbf{z}}) \exp[j(k_{I,x}x + k_{I,y}y - k_{I,z}z)] \quad (3)$$

and

$$\mathbf{E}_{II} = (T_x \hat{\mathbf{x}} + T_y \hat{\mathbf{y}} + T_z \hat{\mathbf{z}}) \exp[j(k_{II,x}x + k_{II,y}y + k_{II,z}z)], \quad (4)$$

where R_x , R_y , and R_z are the reflected wave components and T_x , T_y , and T_z are the transmitted wave components. To satisfy the phase-matching conditions on the transverse plane ($z=0$), the following relations of the wave vector components must hold:

$$k_{I,x} = k_{II,x} = k_0 n_I \sin \theta \cos \phi, \quad (5)$$

$$k_{I,y} = k_{II,y} = k_0 n_I \sin \theta \sin \phi, \quad (6)$$

$$k_{I,z} = k_0 n_I \cos \theta, \quad (7)$$

and

$$k_{II,z} = (k_0^2 n_2^2 - k_0^2 n_1^2 \sin^2 \theta)^{1/2}, \quad (8)$$

where $(k_{I,x}, k_{I,y}, k_{I,z})$ and $(k_{II,x}, k_{II,y}, k_{II,z})$ are the wave vectors in medium I and medium II, respectively. With the use of the relations (5) to (8) and by means of some lengthy but straightforward manipulations, the reflection and transmission coefficients can be obtained as

$$\Gamma_x(\theta, \phi, \psi, n_I, n_{II}) = \frac{R_x}{U_x} = \frac{\Gamma_{TM} \cos \psi \cos \theta \cos \phi - \Gamma_{TE} \sin \psi \sin \phi}{\cos \psi \cos \theta \cos \phi - \sin \psi \sin \phi}, \quad (9)$$

$$\Gamma_y(\theta, \phi, \psi, n_I, n_{II}) = \frac{R_y}{U_y} = \frac{\Gamma_{TM} \cos \psi \cos \theta \sin \phi + \Gamma_{TE} \sin \psi \cos \phi}{\cos \psi \cos \theta \sin \phi + \sin \psi \cos \phi}, \quad (10)$$

$$\Gamma_z(\theta, \phi, \psi, n_I, n_{II}) = \frac{R_z}{U_z} = -\Gamma_{TM}, \quad (11)$$

$$\tau_x(\theta, \phi, \psi, n_I, n_{II}) = \frac{T_x}{U_x} = \frac{\tau_{TM} \cos \psi \cos \theta \cos \phi - \tau_{TE} \sin \psi \sin \phi}{\cos \psi \cos \theta \cos \phi - \sin \psi \sin \phi}, \quad (12)$$

$$\tau_y(\theta, \phi, \psi, n_I, n_{II}) = \frac{T_y}{U_y} = \frac{\tau_{TM} \cos \psi \cos \theta \sin \phi + \tau_{TE} \sin \psi \cos \phi}{\cos \psi \cos \theta \sin \phi + \sin \psi \cos \phi}, \quad (13)$$

and

$$\tau_z(\theta, \phi, \psi, n_I, n_{II}) = \frac{T_z}{U_z} = \frac{\tau_{TM} \sin \theta_t}{\sin \theta}, \quad (14)$$

where $(\Gamma_x, \Gamma_y, \Gamma_z)$ and (τ_x, τ_y, τ_z) are the reflection and transmission coefficients for each polarization component. For the manipulation, Snell's law, $n_I \sin \theta = n_{II} \sin \theta_t$, is useful for deriving succinct expressions, so the refraction angle symbol θ_t is involved in the expressions (12) to (14). The coefficients Γ_{TE} , τ_{TE} , Γ_{TM} , and τ_{TM} are given by

$$\Gamma_{TE} = \frac{n_I \cos \theta - n_{II} \cos \theta_t}{n_I \cos \theta + n_{II} \cos \theta_t}, \quad (15)$$

$$\tau_{TE} = \frac{2n_I \cos \theta}{n_I \cos \theta + n_{II} \cos \theta_t}, \quad (16)$$

$$\Gamma_{TM} = \frac{n_I \cos \theta_t - n_{II} \cos \theta}{n_I \cos \theta_t + n_{II} \cos \theta}, \quad (17)$$

and

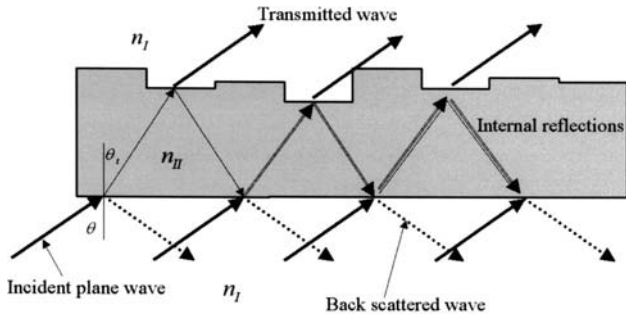


Fig. 2 Modeling of multiple internal reflections inside a multilevel DOE (two-dimensional view).

$$\tau_{TM} = \frac{2n_I \cos \theta}{n_I \cos \theta_t + n_{II} \cos \theta} \quad (18)$$

Let us symbolize the set of reflection and transmission coefficients as $(\Gamma_{1,x}, \Gamma_{1,y}, \Gamma_{1,z}, \tau_{1,x}, \tau_{1,y}, \tau_{1,z})$ when the wave propagates from medium I to medium II, and $(\Gamma_{2,x}, \Gamma_{2,y}, \Gamma_{2,z}, \tau_{2,x}, \tau_{2,y}, \tau_{2,z})$ when wave propagates from medium II to medium I. The latter set can easily be obtained by exchanging refractive index n_I for n_{II} and incidence angle θ for θ_t in Eqs. (9) to (18).

3 Calculation of the Transmittance Function of the Multilevel DOE

The conventional scheme for calculating the transmittance function of a multilevel DOE for an obliquely incident plane wave is as follows. Assuming that the incidence angle in medium I is θ and the refraction angle is θ_t , we note that the phase change along the normal direction (z direction) through the DOE is $k_{II,z}d$, where d is the normal thickness of the DOE at the point of the incidence. Since the relative phase difference of two waves propagating through medium I and medium II of thickness d is $(k_{II,z} - k_{I,z})d$, the maximum thickness d_{\max} of the surface relief structure to yield a 2π phase difference is given by

$$d_{\max} = \frac{\lambda}{n_{II} \cos \theta_t - n_I \cos \theta}, \quad (19)$$

and the thickness for inducing a relative phase change ϕ is linearly proportional to ϕ as

$$d_\phi = \frac{\phi \lambda}{2\pi(n_{II} \cos \theta_t - n_I \cos \theta)}. \quad (20)$$

In this conventional scheme, multiple internal reflections are regarded as being sufficiently weak to be negligible.

However, when the refractive index of a material is high or the incidence is rather oblique, multiple internal reflections are significant. Figure 2 shows that multiple internal reflections occur inside the DOE. In Fig. 2, n_I and n_{II} indicate the refractive indices of the surroundings (air) and the material of the DOE, respectively. A multilevel DOE consists of two parts—the substrate region and the modulation cell region—the refractive indices of which are the same: n_{II} . A modulation cell indicates one pixel and has a certain

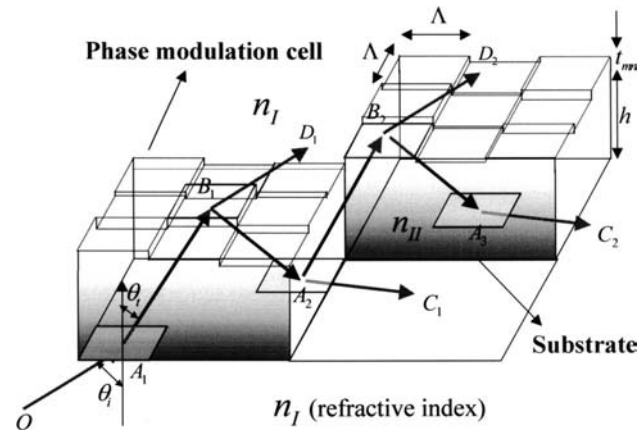


Fig. 3 Modeling of multiple internal reflections inside a multilevel DOE (three-dimensional view).

thickness on the surface relief structure. The incident wave enters the DOE at the bottom with incidence angle θ . Partial reflection and transmission (refraction) occur at this point. After the initial event, the refracted wave continues to propagate along a zigzag path inside the DOE. Whenever the running wave meets the boundary between air and material, partial reflection and transmission occur. As a result, the transmitted wave and the backscattered wave are superpositions of all the partially transmitted local waves at the ceiling plane and the bottom plane (see Fig. 2). It can be seen that many local plane waves contribute to the complex field distribution on the surface of the DOE. Multiple internal reflections would be expected to induce different phase and amplitude deviations from that predicted by the conventional transmittance model.

In this paper, a theoretical approach to multiple internal reflections is addressed in an attempt to obtain a more refined transmittance function model for multilevel DOEs. Before elaborating the theory, a restriction is placed on selecting the incidence angle of an incident plane wave. We introduce the so-called cell-to-cell matching condition to deal systematically with multiple reflections inside the DOE, with Fig. 3 representing a three-dimensional view of wave propagation inside the DOE. Let us define S as a set of wave vectors for propagating local plane waves inside the DOE as

$$S = \{\mathbf{k}_{II,mn}\}, \quad (21)$$

where the elementary wave vector $\mathbf{k}_{II,mn}$ is defined as

$$\begin{aligned} \mathbf{k}_{II,mn} &= (k_{II,x}, k_{II,y}, k_{II,z})_{mn} \\ &= \left(\frac{k_0 n_{II} \Lambda m}{[(\Lambda m)^2 + (\Lambda n)^2 + h^2]^{1/2}}, \right. \\ &\quad \frac{k_0 n_{II} \Lambda n}{[(\Lambda m)^2 + (\Lambda n)^2 + h^2]^{1/2}}, \\ &\quad \left. \pm \frac{k_0 n_{II} h}{[(\Lambda m)^2 + (\Lambda n)^2 + h^2]^{1/2}} \right), \end{aligned} \quad (22)$$

where k_0 is the wave number in vacuum, and Λ and h are the length of a unit cell and the thickness of the substrate, as shown in Fig. 3. The integer parameter pair (m, n) discretizes the propagation direction of the local plane wave in Eq. (22). Hence, a local plane wave propagates from a sectioned cell A_1 at the bottom plane to another sectioned cell B_1 at the ceiling plane, and a portion of it is reflected to propagate to another cell A_2 at the bottom plane. The theory is unfolded using this setup. If the substrate thickness h is sufficiently large, the allowed propagation directions will be almost continuous. Thus to satisfy Snell's law, wave vectors of the incident wave, backscattered wave, and transmitted wave in medium I take the form

$$\mathbf{k}_{1,mn} = (k_{1,x}, k_{1,y}, k_{1,z}) = \left(\frac{k_0 n_{II} \Lambda m}{[(\Lambda m)^2 + (\Lambda n)^2 + h^2]^{1/2}}, \frac{k_0 n_{II} \Lambda n}{[(\Lambda m)^2 + (\Lambda n)^2 + h^2]^{1/2}}, \pm k_0 n_I \left[1 - \left(\frac{n_{II}}{n_I} \right)^2 \frac{(\Lambda m)^2 + (\Lambda n)^2}{(\Lambda m)^2 + (\Lambda n)^2 + h^2} \right]^{1/2} \right). \quad (23)$$

Considering Eqs. (5) to (8) and (22), θ , θ_t , ϕ , and ψ can be found for an index pair (m, n) . The reflection and transmission coefficients $(\Gamma_{1,x}, \Gamma_{1,y}, \Gamma_{1,z}, \tau_{1,x}, \tau_{1,y}, \tau_{1,z})$ and $(\Gamma_{2,x}, \Gamma_{2,y}, \Gamma_{2,z}, \tau_{2,x}, \tau_{2,y}, \tau_{2,z})$ are obtained by substituting the obtained θ , θ_t , ϕ , and ψ into Eqs. (9) to (14).

Let us consider the situation in which a local plane wave strikes the $(0,0)$ 'th cell area at the bottom plane of the substrate. The refracted wave with the wave vector \mathbf{k}_{mn} propagates inside the substrate, reaching the (m,n) 'th cell at the ceiling plane of the DOE as shown in Fig. 3. A portion of the local plane wave is reflected at the cell surface (the ceiling plane), and the other portion is transmitted to medium I. The reflected partial wave propagates to meet the $(2m, 2n)$ 'th cell at the bottom plane. Therefore, the multiple reflection process inside the DOE can be formulated as follows. The phase variation of the local plane wave running through the substrate region is given by

$$P_{\uparrow} = P_{\downarrow} = \exp[jk_0 n_{II} (\Lambda m \sin \theta_t \cos \phi + \Lambda n \sin \theta_t \sin \phi + h \cos \theta_t)], \quad (24)$$

where \uparrow and \downarrow indicate upward propagation and downward propagation, respectively. Let the thickness of the (m, n) 'th modulation cell be t_{mn} ; then the phase change by the surface relief piece appears as

$$P_{mn} = \gamma_{mn} \exp(jk_0 n_{II} \cos \theta_t t_{mn}), \quad (25)$$

where the loss induced by shadowing effects by thin vertical facets is not considered.

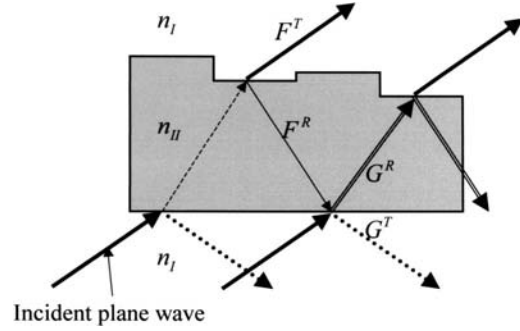


Fig. 4 The symbols F^R , F^T , G^R , and G^T denote the reflected and transmitted wave components at the bottom and ceiling, respectively.

Four symbols to be used in applying the theory are shown in Fig. 4. The symbols F^T and F^R , respectively, denote the wave transmitted to the surroundings and the wave reflected to the substrate at the ceiling plane. The symbols G^T and G^R denote the wave transmitted from the substrate to the surroundings and the wave reflected at the bottom plane. But it should be noted that G^R and G^T are also used to indicate the initial refracted and reflected waves of the starting incident wave at the bottom plane, respectively. With these four symbols, we proceed to describe a multiple reflection process inside the DOE, as shown in Fig. 3. Let the index pair (p, q) indicate the (p, q) 'th sectioned area at the bottom plane or the ceiling plane. The subscript integer pair (p, q) of a quantity implies that the quantity is located in the (p, q) 'th sectioned area. For convenience, the incident plane wave is taken to be linearly polarized in the x direction, that is, $U_x \neq 0$ and $U_y = U_z = 0$ in Eq. (1). Thus only x -direction components of waves are considered. However, to simplify the description, the subscript x is omitted hereafter in all used terms. At the sectioned area (p, q) , the incident plane wave is denoted by $E_{p,q}^{\text{inc}}$. The initial refracted wave $G_{p,q}^R$ and the reflected wave $G_{p,q}^T$ at the position then take the forms

$$G_{p,q}^R = E_{p,q}^{\text{inc}} \tau_1 \quad (26)$$

and

$$G_{p,q}^T = E_{p,q}^{\text{inc}} \Gamma_1. \quad (27)$$

The initial refracted local wave $G_{p,q}^R$ goes to the $(m + p, n + q)$ 'th modulation cell at the ceiling plane; the transmitted wave $F_{m+p,n+q}^T$ and the reflected wave $F_{m+p,n+q}^R$ at the ceiling plane are then represented as

$$\begin{aligned} F_{m+p,n+q}^R &= G_{0 \times m + p, 0 \times n + q}^R P_{\uparrow} P_{m+p,n+q} \Gamma_2 \\ &= E_{p,q}^{\text{inc}} \tau_1 P_{\uparrow} P_{m+p,n+q} \Gamma_2 \end{aligned} \quad (28)$$

and

$$F_{m+p,n+q}^T = G_{0 \times m+p, 0 \times n+q}^R P_{\uparrow} P_{m+p,n+q} \tau_2 \\ = E_{p,q}^{\text{inc}} \tau_1 P_{\uparrow} P_{m+p,n+q} \tau_2. \quad (29)$$

In the preceding equations, the product $P_{\uparrow} P_{m+p,n+q}$ represents the phase variation along the ray path. The transmitted wave $F_{m+p,n+q}^T$ will contribute to the resultant field distribution at the $(m+p, n+q)$ 'th cell at the ceiling plane, and the reflected wave $F_{m+p,n+q}^R$ contributes to the $(2m+p, 2n+q)$ 'th cell at the bottom plane. At the position $(2m+p, 2n+q)$ on the bottom plane, the relations between the reflected wave $G_{2m+p, 2n+q}^R$ and the transmitted wave $G_{2m+p, 2n+q}^T$ are

$$G_{2m+p, 2n+q}^R = F_{m+p,n+q}^R P_{m+p,n+q} P_{\downarrow} \Gamma_2 \quad (30)$$

and

$$G_{2m+p, 2n+q}^T = F_{m+p,n+q}^R P_{m+p,n+q} P_{\downarrow} \tau_2. \quad (31)$$

The reflected wave $G_{2m+p, 2n+q}^R$ then propagates to the $(3m+p, 3n+q)$ 'th cell at the ceiling plane. The relationship between the transmitted wave $F_{3m+p, 3n+q}^T$ and reflected wave $F_{3m+p, 3n+q}^R$ at the ceiling plane are analogous to the relations with the previous local plane wave components $F_{m+p,n+q}^T$ and $F_{m+p,n+q}^R$:

$$F_{3m+p, 3n+q}^R = G_{2m+p, 2n+q}^R P_{\uparrow} P_{3m+p, 3n+q} \Gamma_2 \\ = F_{m+p,n+q}^R P_{m+p,n+q} P_{\downarrow} \Gamma_2 P_{\uparrow} P_{3m+p, 3n+q} \Gamma_2 \\ = \alpha P_{m+p,n+q} P_{3m+p, 3n+q} F_{m+p,n+q}^R, \quad (32)$$

$$F_{3m+p, 3n+q}^T = \alpha P_{m+p,n+q} P_{3m+p, 3n+q} F_{m+p,n+q}^T, \quad (33)$$

where α is defined by $\alpha = P_{\downarrow} \Gamma_2 P_{\uparrow} \Gamma_2$. From Eqs. (32) and (33), we can obtain two coupled progressions by mathematical induction:

$$F_{(2r+1)m+p, (2r+1)n+q}^R \\ = \alpha P_{(2r+1)m+p, (2r+1)n+q} P_{(2r-1)m+p, (2r-1)n+q} \\ \times F_{(2r-1)m+p, (2r-1)n+q}^R \quad (34)$$

and

$$F_{(2r+1)m+p, (2r+1)n+q}^T \\ = \alpha P_{(2r+1)m+p, (2r+1)n+q} P_{(2r-1)m+p, (2r-1)n+q} \\ \times F_{(2r-1)m+p, (2r-1)n+q}^T, \quad (35)$$

having the respective initial terms

$$F_{m+p,n+q}^R = G_{0 \times m+p, 0 \times n+q}^R P_{\uparrow} P_{m+p,n+q} \Gamma_2 \\ = E_{p,q}^{\text{inc}} \tau_1 P_{\uparrow} P_{m+p,n+q} \Gamma_2 \quad (36)$$

and

$$F_{m+p,n+q}^T = G_{0 \times m+p, 0 \times n+q}^R P_{\uparrow} P_{m+p,n+q} \tau_2 \\ = E_{p,q}^{\text{inc}} \tau_1 P_{\uparrow} P_{m+p,n+q} \tau_2. \quad (37)$$

This results in two formulas,

$$F_{(2r+1)m+p, (2r+1)n+q}^R \\ = \alpha^r \frac{\prod_{l=0}^r (P_{(2l+1)m+p, (2l+1)n+q})^2}{P_{m+p,n+q} P_{(2r+1)m+p, (2r+1)n+q}} F_{m+p,n+q}^R \quad (38)$$

and

$$F_{(2r+1)m+p, (2r+1)n+q}^T \\ = \alpha^r \frac{\prod_{l=0}^r (P_{(2l+1)m+p, (2l+1)n+q})^2}{P_{m+p,n+q} P_{(2r+1)m+p, (2r+1)n+q}} F_{m+p,n+q}^T, \quad (39)$$

which are the reflected and transmitted wave components at the ceiling plane, respectively. Using Eq. (39), we can derive the total field distribution $F(u, v)$ on the ceiling plane, which corresponds to the input plane of the Fresnel diffraction transform in the design of the DOE.

Let us examine wave components that contribute to the field distribution at a certain position (u, v) at the ceiling plane. Observing the left-hand side of Eq. (39), it can be easily seen that the total field distribution is the sum of all contributing components

$$F(u, v) = \sum_{r,p,q} F_{(2r+1)m+p, (2r+1)n+q}^T, \quad (40)$$

where all r, g, q must satisfy the relation

$$(u, v) = ((2r+1)m+p, (2r+1)n+q). \quad (41)$$

The corresponding transmittance function of the DOE, $T(u, v)$, can be obtained as

$$T(u, v) = \exp(-jk_0 n_I \cos \theta t_{uv}) F(u, v) / E_{u,v}^{\text{inc}}. \quad (42)$$

For example, let us consider the case of normal incidence. Then the direction of propagation of the local plane wave is normal to the bottom and the ceiling plane, and the number of multiple internal reflections is infinite. So $m=0$, $n=0$, and $r \rightarrow \infty$. If these conditions are substituted into Eq. (39), the r 'th wave component $F_{p,q,r}^T$ takes the form

$$F_{p,q,r}^T = \alpha^r \frac{\prod_{l=0}^r (P_{(2l+1)0+p, (2l+1)0+q})^2}{P_{0+p,0+q} P_{(2k+1)0+p, (2k+1)0+q}} F_{p,q,0}^T \\ = \alpha^r (P_{p,q})^{2r} F_{p,q,0}^T. \quad (43)$$

The total transmitted field $F(p, q)$ at the (p, q) position is the sum of wave components $F_{p,q,r}^T$, so the following relation holds:

$$\begin{aligned}
 F(p, q) &= \sum_r F_{p, q, r}^T \\
 &= \frac{F_{p, q, 0}^T}{1 - \alpha(P_{p, q})^2} \\
 &= \frac{E_{p, q}^{\text{inc}} 2n_I n_{II}}{2n_I n_{II} \cos[k_{II, z}(t_{pq} + h)] - j(n_I^2 + n_{II}^2) \sin[k_{II, z}(t_{pq} + h)]}.
 \end{aligned}
 \quad (44)$$

Consequently, with the aid of Eq. (42), the transmittance function is given by

$$T(p, q) = \frac{2n_I n_{II} \exp(-jk_{I, z} t_{pq})}{2n_I n_{II} \cos[k_{II, z}(t_{pq} + h)] - j(n_I^2 + n_{II}^2) \sin[k_{II, z}(t_{pq} + h)]}. \quad (45)$$

The derived formula agrees with the transmission coefficient obtained in solving a textbook level problem. For convenience, we reduce the optical path length of the substrate thickness modulo 2π . So the phase modulation formula reads as

$$\psi(t_{pq}) = \arg \left[\frac{2n_I n_{II} \exp(-jk_{I, z} t_{pq})}{2n_I n_{II} \cos(k_{II, z} t_{pq}) - j(n_I^2 + n_{II}^2) \sin(k_{II, z} t_{pq})} \right]. \quad (46)$$

When $t_{pq} = 0$ and $\psi(0) = 0$, the smallest positive value of t_{pq} to make the phase change by 2π can be then determined by finding the first positive zero of the following function:

$$\begin{aligned}
 T(t_{pq}) &= 2n_I n_{II} \sin(k_{I, z} t_{pq}) \cos(k_{II, z} t_{pq}) \\
 &\quad - (n_I^2 + n_{II}^2) \sin(k_{II, z} t_{pq}) \cos(k_{I, z} t_{pq}).
 \end{aligned}
 \quad (47)$$

Let the thickness corresponding to a 2π phase change be denoted by $t_{2\pi}$. In addition, note that the phase modulation (46) can be viewed as a function of the thickness divided by the wavelength with two parameters, viz., the refractive indices n_I and n_{II} . That is, the thickness is scalable with the wavelength for a fixed phase modulation value. Therefore we only need to concern ourselves with the phase variation curve, nonlinear in the thickness variable, normalized by the wavelength, for several refractive indices n_{II} of the material and a fixed refractive index n_I of the surroundings. Our problem is independent of the value of the wavelength if the thickness is scaled appropriately.

Figure 5 shows the changes in phase and amplitude for a normally incident plane wave as functions of the thickness of the passed dielectric slab for three different material indices n_{II} . In Fig. 5, for convenience, the thickness normalized by the 2π -phase thickness $t_{2\pi}$ is used as the variable for the horizontal axis. In Fig. 5(a), the dashed and solid lines indicate, respectively, the phase variation curves in the range from 0 to 360 deg with the change in normalized thickness of the dielectric slab, considering multiple internal reflections, and those of the conventional transmittance model, for $n_{II} = 1.5, 2.5, 3.5$. It is seen that, as the refractive index becomes higher, the influence of multiple internal reflections on phase variation becomes more significant.

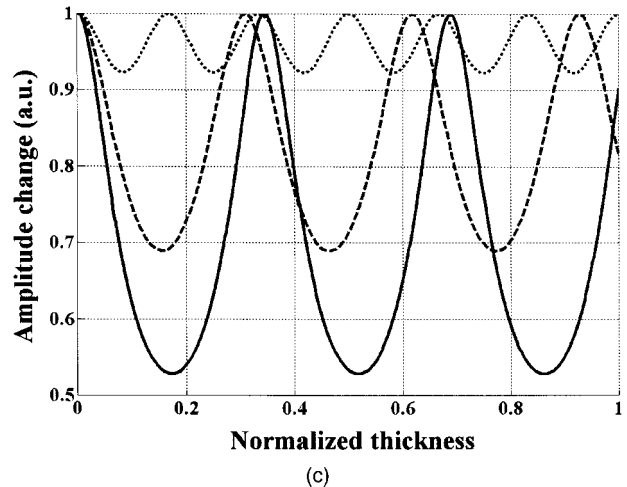
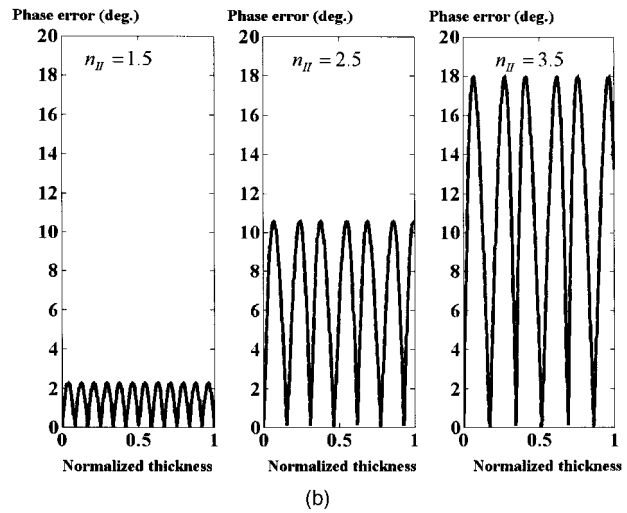
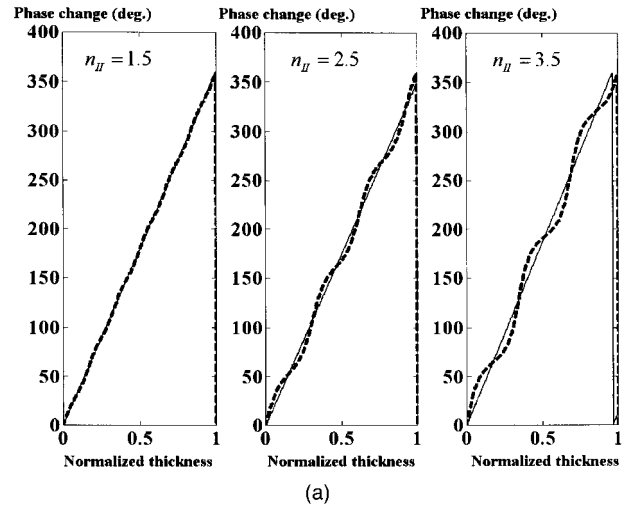


Fig. 5 (a) Comparison of the phase change curves of a normal incident plane wave through a single dielectric slab as a function of slab thickness for several material refractive indices ($n_{II} = 1.5, 2.5, 3.5$). The thick dashed lines are the phase change curves with multiple internal reflections considered, while the thin solid lines are those obtained by the conventional transmittance model. (b) Comparison of phase errors as a function of the slab thickness for several material refractive indices. (c) Comparison of the change in amplitude of the transmitted wave as a function of the slab thickness for several material indices. Solid line, dashed line, and dotted line indicate the cases $n_{II} = 3.5$, $n_{II} = 2.5$, and $n_{II} = 1.5$, respectively.

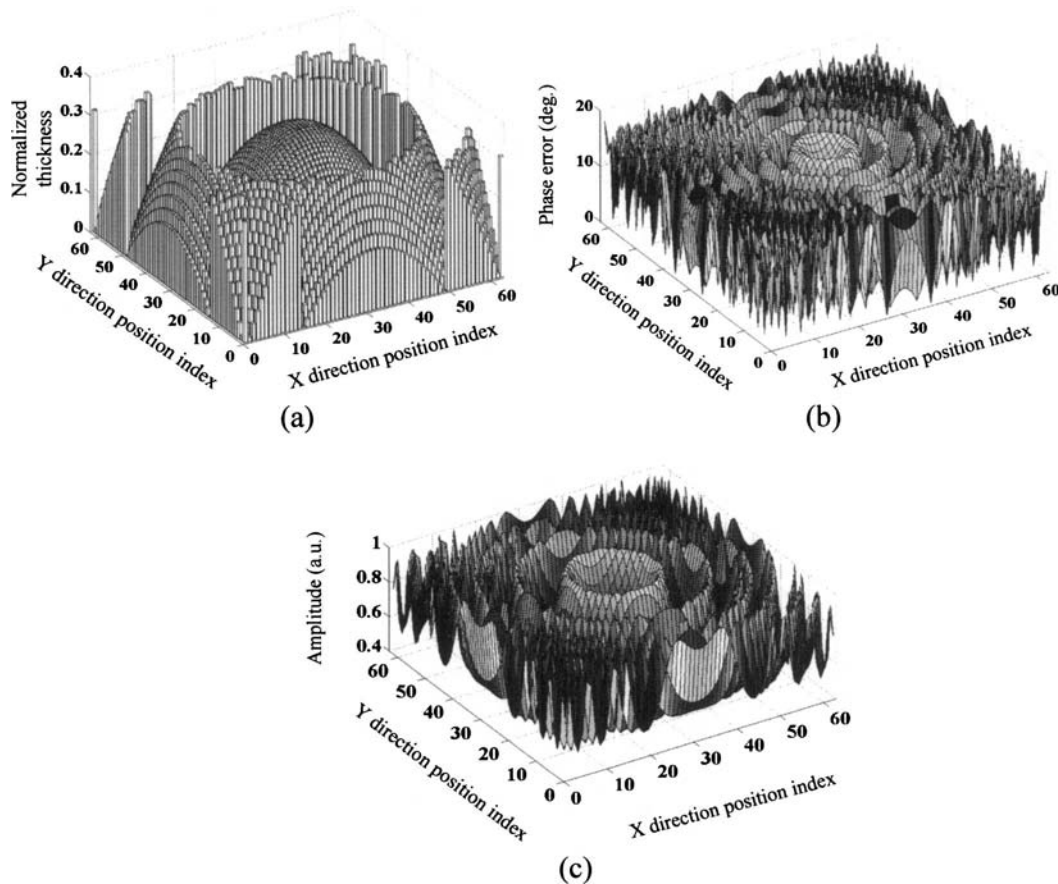


Fig. 6 (a) The conventional construction of the surface relief structure of a multilevel DOE generating a lenslike phase modulation profile, (b) the phase error distribution induced by multiple internal reflections, and (c) the amplitude distribution of the transmitted wave in the case of normal incidence.

For a clear comparison of the three cases, the magnitude of the phase errors for the conventional transmittance model is shown in Fig. 5(b). The maximum phase error values are 2, 10, and 18 deg, respectively. Since the refractive index of a semiconductor material such as GaAs is about 3.6, a consideration of multiple internal reflections is of practical importance in the design of DOEs using a high-index material. Figure 5(c) shows the variation in the amplitude of the transmittance function (45) as a function of normalized thickness for the three cases. The dotted and dashed lines indicate the cases of $n_{II}=1.5$ and 2.5, respectively. In the case of $n_{II}=3.5$ (solid line), remarkable amplitude variations down to 50% transmittance are observed, which correspond to only a 25% transmission in terms of intensity. Therefore the dependence of the phase and amplitude variations on slab thickness for a local wave transmitted through a dielectric slab should be considered in the design of DOEs. It is desirable that a proper cell thickness be determined to generate a desired phase modulation in the construction of a surface relief structure, so as to realize the optimized DOE phase profile, and the amplitude variation data as shown in Fig. 5(c) should be used appropriately in the step involving the optimization of the DOE phase profile using the iterative Fourier transform algorithm.

For a DOE designed to generate a lenslike phase modulation profile, we show an example of the analysis of phase and amplitude errors. In the case of normal incidence, the

numerical results are presented in Fig. 6. The conventional transmittance model gives the thickness of a cell generating a phase modulation value ψ as

$$t(\psi) = \frac{\lambda \psi}{2\pi(n_{II} - n_I)}. \quad (48)$$

The value for amplitude modulation is assumed to be a constant $A(\psi)=1$. Using the conventional formula (48), the surface relief structure of a DOE is constructed to generate a lenslike phase modulation profile. Figure 6(a) shows the constructed surface relief profile of the DOE. The vertical axis in Fig. 6(a) represents the thickness normalized by wavelength. Assuming a material refractive index $n_{II}=3.5$, the phase errors are distributed over nearly all the positions on the DOE, as shown in Fig. 6(b). As previously seen in Fig. 5(c), the maximum phase error is about 18 deg. Figure 6(c) shows the amplitude distribution of the transmitted wave. The amplitude varies from 0.53 to 1 with thickness t_{pq} , as seen in Fig. 5(b). This confirms that the conventional formula (48) does not produce the correct thickness to give the desired phase change ψ and that the assumption of constant amplitude modulation fails.

Next, we consider the case of oblique incidence ($m \neq 0$ or $n \neq 0$). Substituting Eqs. (25), (37) into (39), the r 'th transmitted wave component originating from the initial position (p, q) at the bottom plane takes the form

$$F_{(2r+1)m+p, (2r+1)n+q}^T = P_{\uparrow} E_{p,q}^{\text{inc}} \alpha^r \tau_1 \tau_2 \exp \left[jk_0 n_{\text{II}} \right. \\ \left. \times \cos \theta_t \left(t_{[(2r+1)m+p][(2r+1)n+q]} \right. \right. \\ \left. \left. + 2 \sum_{l=0}^{r-1} t_{[(2l+1)m+p][(2l+1)n+q]} \right) \right]. \quad (49)$$

Substituting Eq. (49) into Eq. (40), the total field distribution $F(u, v)$ at the position (u, v) can be described as

$$F(u, v) = \sum_{r,p,q} F_{(2r+1)m+p, (2r+1)n+q}^T \\ = \sum_{r,p,q} P_{\uparrow} E_{p,q}^{\text{inc}} \alpha^r \tau_1 \tau_2 \exp \left[jk_0 n_{\text{II}} \right. \\ \left. \times \cos \theta_t \left(t_{[(2r+1)m+p][(2r+1)n+q]} \right. \right. \\ \left. \left. + 2 \sum_{l=0}^{r-1} t_{[(2l+1)m+p][(2l+1)n+q]} \right) \right] \\ = \exp(jk_0 n_{\text{II}} \cos \theta_t t_{uv}) P_{\uparrow} \tau_1 \tau_2 \sum_{r,p,q} E_{p,q}^{\text{inc}} \alpha^r \\ \times \exp \left(jk_0 n_{\text{II}} \cos \theta_t 2 \right. \\ \left. \times \sum_{l=0}^{r-1} t_{[(2l+1)m+p][(2l+1)n+q]} \right). \quad (50)$$

A systematic method for calculating $F(u, v)$ is as follows. First, let us assume that the concerned DOE has a rectangular aperture, on which an M -by- N matrix of phase modulation cells is built. The position index pair (u, v) of a cell is in the range from $(0, 0)$ to $(M-1, N-1)$. However, the other index pair (p, q) indicating a sectioned area in the bottom plane is allowed to be valued in the range from $(-\infty, -\infty)$ to (∞, ∞) . Without loss of generality, it can be assumed that both m and n of the index pair (m, n) indicating the direction of the propagating wave are positive. Let us begin the calculation of $F(u, v)$ in ascending order of the total number of wave components superposed to make the final field distribution $F(u, v)$. If the starting incidence position is $(p, q) = (-m, -n)$ in the bottom plane, the field calculation starts at the position $(u, v) = (0, 0)$ that the refracted wave reaches. At that position, only the refracted wave component contributes to the final field distribution. After this, the position of the second field distribution to be

calculated is $(u, v) = (2m, 2m)$. At the second position $(u, v) = (2m, 2m)$, the superposition of one wave component originating at the incidence position $(p, q) = (-m, -n)$ and the other originating at the incidence position $(p, q) = (0, 0)$ yields the final field distribution. For general cases, the calculation details are as follows. For an initial position index pair (p, q) , the first equation takes a form with one term,

$$F(p+m, q+n) = \exp(jk_0 n_{\text{II}} \cos \theta_t t_{(p+m)(q+n)}) P_{\uparrow} \tau_1 \tau_2 E_{p,q}^{\text{inc}} \\ = \exp(jk_z t_{(p+m)(q+n)}) P_{\uparrow} \tau_1 \tau_2 E_{p,q}^{\text{inc}}. \quad (51)$$

The second equation is expressed with two terms as

$$F(p+3m, q+3n) \\ = P_{\uparrow} E_{p+2m, q+2n}^{\text{inc}} \tau_1 \tau_2 \exp(jk_z t_{(p+3m)(q+3n)}) \\ + P_{\uparrow} E_{p,q}^{\text{inc}} \tau_1 \tau_2 \alpha \exp[jk_z (t_{(p+3m)(q+3n)} \\ + 2t_{(p+m)(q+n)})] \\ = \exp(jk_z t_{(p+3m)(q+3n)}) P_{\uparrow} \tau_1 \tau_2 \alpha \\ \times \exp[jk_z (2t_{(p+m)(q+n)})] \{E_{p+2m, q+2n}^{\text{inc}} \alpha^{-1} \\ \times \exp[-jk_z (2t_{(p+m)(q+n)})] + E_{p,q}^{\text{inc}}\}. \quad (52)$$

Similarly, the third equations can be obtained as

$$F(p+5m, q+5n) \\ = \exp(jk_z t_{(p+5m)(q+5n)}) P_{\uparrow} \tau_1 \tau_2 \alpha^2 \\ \times \exp[jk_z 2(t_{(p+3m)(q+3n)} + t_{(p+m)(q+n)})] \\ \times \{E_{p+4m, q+4n}^{\text{inc}} \alpha^{-2} \exp[-jk_z 2(t_{(p+3m)(q+3n)} \\ + t_{(p+m)(q+n)})] + E_{p+2m, q+2n}^{\text{inc}} \alpha^{-1} \\ \times \exp(-jk_z 2t_{(p+m)(q+n)}) + E_{p,q}^{\text{inc}}\}. \quad (53)$$

With the aid of mathematical induction, the general form of the field distribution can be obtained as

$$F(p+(2r+1)m, q+(2r+1)n) \\ = \exp(jk_z t_{[p+(2r+1)m][q+(2r+1)n]}) P_{\uparrow} \tau_1 \tau_2 \alpha^r \\ \times \exp \left(j2k_z \sum_{l=0}^{r-1} t_{[p+(2l+1)m][q+(2l+1)n]} \right) \\ \times \sum_{l=0}^r E_{p+2lm, q+2ln}^{\text{inc}} \alpha^{-l} \\ \times \exp \left(-j2k_z \sum_{c=0}^{l-1} t_{[p+(2c+1)m][q+(2c+1)n]} \right), \quad (54)$$

where r is in the range

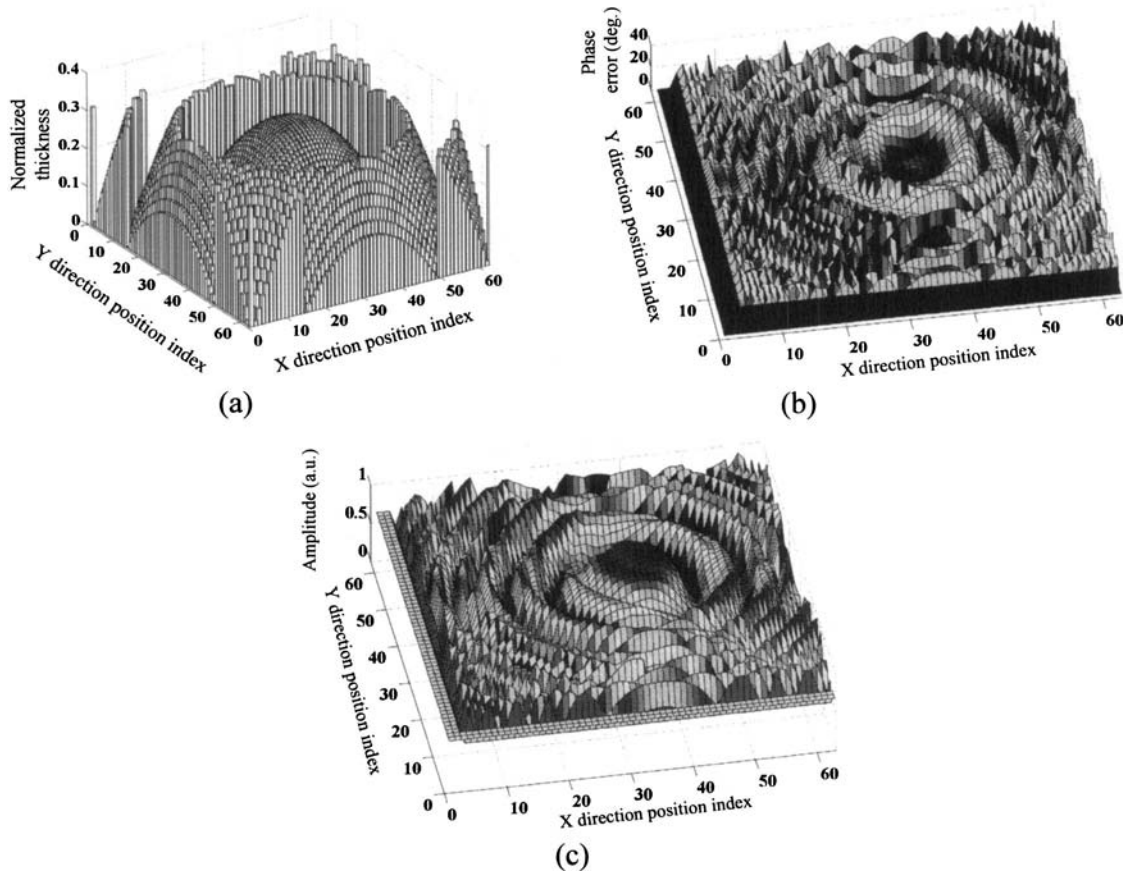


Fig. 7 (a) The conventional construction of the surface relief structure of a multilevel DOE generating a lenslike phase modulation profile, (b) the phase error distribution induced by multiple internal reflections, and (c) the amplitude distribution of the transmitted wave in the case of oblique incidence.

$$0 \leq r \leq r_{\max} = \min \left(\left\lceil \frac{M-p-m-1}{2m} \right\rceil, \left\lceil \frac{N-q-n-1}{2n} \right\rceil \right),$$

in which $\lceil \cdot \rceil$ is Gauss's symbol.

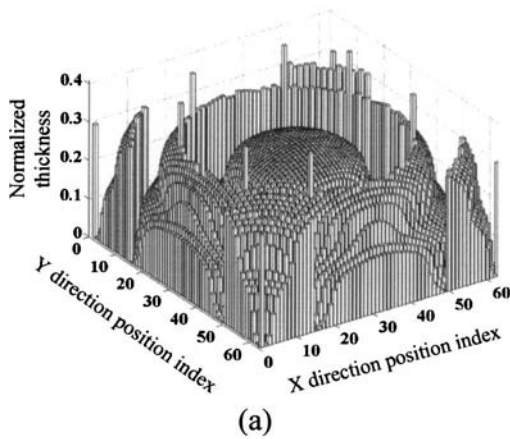
In the other hand, for (u, v) to cover any integer pair of the cell matrix representing the entire aperture of the DOE, p and q of the index pair of the incidence position should be in the ranges $-m \leq p \leq m-1$ and $-n \leq q \leq n-1$, respectively. Therefore the number of initial position index pairs is $4mn + 2n(M-2m) + 2m(N-2n)$. For an incidence position (p, q) , one sequential equation system of Eq. (54) can be determined. Thus, all the equations to calculate the total field distribution $F(u, v)$ are classified into $4mn + 2n(M-2m) + 2m(N-2n)$ sets. Each equation system is parametrized by a starting incidence position index (p, q) , and its sequential procedure has a variable r , indicating the number of superposed wave components. Knowing the total field distribution $F(u, v)$, and using Eq. (42), the corresponding transmittance function can be obtained. The mentioned sequential formulation is useful for dealing with the construction of the DOE surface relief profile in the next section.

For the same DOE used in the simulation of the case of normal incidence, phase and amplitude errors can be analyzed in the case of oblique incidence with an incidence angle of 11.5° deg $[(m, n) = (2, 2)]$. The numerical results

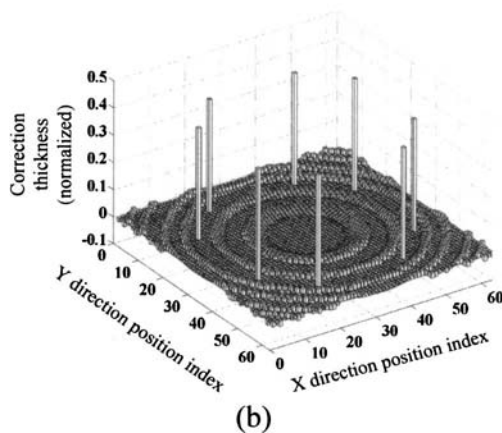
are presented in Fig. 7. The conventional transmittance model [Eq. (20)] produces the surface relief structure of the DOE, as shown in Fig. 7(a). The vertical axis in Fig. 7(a) represents the thickness normalized by the wavelength. With a refractive index of the material of $n_{II} = 3.5$, the phase errors are distributed, as shown in Fig. 7(b). The maximum phase error is about 26° , which is larger than that in the normal incidence case, and the distribution appears quite asymmetric. In general, as the incidence angle increases, phase errors become larger, a generality that was confirmed by several simulations. Figure 7(c) shows the amplitude distribution of the transmitted wave. The amplitude varies from 0.387 to 0.979 with the variation in thickness—a larger variation than that in the case of the normal incidence.

4 Construction of the Surface Relief Structure of the DOE to Generate the Desired Phase Modulation Profile

The surface relief structure of a DOE for generating the desired phase distribution can be constructed by inverting the equation for the transmittance function derived in the previous section. The newly obtained surface relief structure can improve on that of the conventional transmittance model. The case of normal incidence has an aspect that is different from that of oblique incidence. In the case of nor-



(a)



(b)

Fig. 8 (a) The surface relief structure constructed by the proposed method, and (b) the distribution of the thickness correction, in the case of normal incidence.

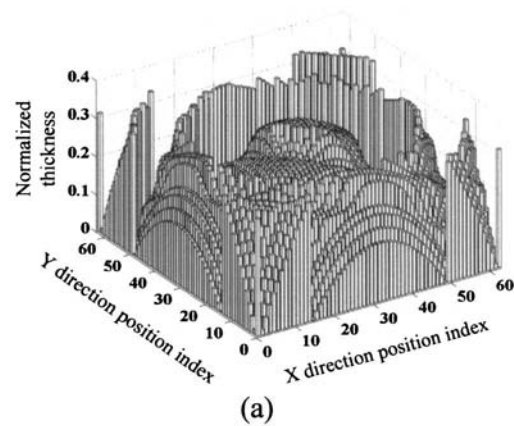
mal incidence, only a numerical solution can be obtained, while in the case of oblique incidence, the problem can be treated analytically. The reason for this is that an infinite number of reflections are involved in the normal incidence case, while a finite number of reflections (within some lateral spatial range) are involved in the oblique incidence case. In the case of normal incidence, the transmittance function of the DOE is given by

$$T(p, q) = \frac{2n_1 n_2 \exp(-jk_{1,z} t_{pq})}{2n_1 n_2 \cos[k_{11,z}(t_{pq} + h)] - j(n_1^2 + n_2^2) \sin[k_{11,z}(t_{pq} + h)]}. \quad (55)$$

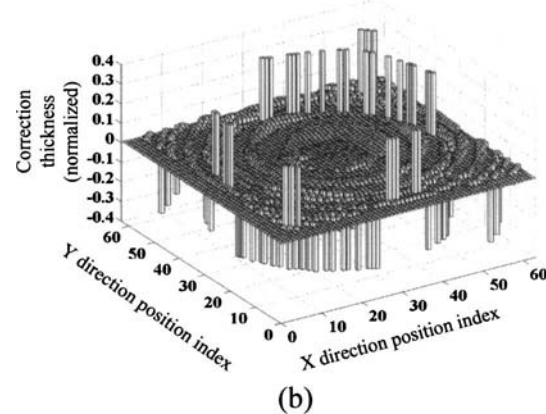
Therefore a simple numerical zero-finding algorithm can be used to obtain the appropriate thickness t_{pq} to satisfy the following relation:

$$\psi = \arg \left\{ \frac{2n_1 n_2 \exp(-jk_{1,z} t_{pq})}{2n_1 n_2 \cos[k_{11,z}(t_{pq} + h)] - j(n_1^2 + n_2^2) \sin[k_{11,z}(t_{pq} + h)]} \right\}, \quad (56)$$

where ψ is the desired phase change. The surface relief structure is constructed by the proposed method to generate the same lenslike phase modulation profile that appears in Figs. 6 and 7. The results are shown in Fig. 8. Figure 8(a)



(a)



(b)

Fig. 9 (a) The surface relief structure constructed by the proposed method, and (b) the distribution of thickness correction, in the case of oblique incidence.

shows the surface relief structure obtained. In comparing Figs. 6(a) and 8(a), there are significant differences in appearance between the structure obtained by the proposed method and that obtained by the conventional method. The difference in thickness between the surface relief structure of the proposed method and that using the conventional method, that is, the correction relief profile, is shown in Fig. 8(b). Phase modulation errors can be corrected by the addition of the correction relief profile to the structure obtained by the conventional method. The significant thickness variations at a few of the pixels shown in Fig. 8(b) are understandable if it is recognized that the thickness $t_{pq}(>0)$ inducing a specific phase modulation has an equivalent thickness $t_{pq} - t_{2\pi} (<0)$ inducing the same phase modulation. This is similar to understanding that an arbitrary phase $\theta (>0)$ is effectively equal to $\theta - 2\pi (<0)$. Let us refer to $t_{pq} - t_{2\pi}$ as an equivalent negative thickness of t_{pq} . The noticeably large corrected thickness at a few positions has the same effect on the phase modulation as does its equivalent negative thickness $t_{pq} - t_{2\pi}$. Therefore, if negative thickness of a cell were allowable, such a sudden difference in thickness would not be seen. However in our scheme, for convenience, the cell thickness is set to a positive value. Because the DOE consists of a substrate with constant thickness and a surface relief structure, a

negative value of the thickness of the surface relief structure is possible while maintaining the total thickness as a positive value.

In the case of oblique incidence, the surface relief structure of the DOE generating a desired phase modulation profile can be constructed by solving the independent sequential equation systems introduced in the previous section. The number of sequential equation systems is $4mn + 2n(M - 2m) + 2m(N - 2n)$. Let us consider the structure of an equation system with a starting incidence position index (p, q) . The equation system can be solved using

the sequential process. The construction of the surface relief structure proceeds in ascending order of r .

In the first step, using Eq. (51), we can directly obtain

$t_{(p+m)(q+n)}$ as

$$t_{(p+m)(q+n)} = \frac{1}{k_z} \arg \left[\frac{F(p+m, q+n)}{P_{\uparrow} \tau_1 \tau_2 E_{p,q}^{\text{inc}}} \right]. \quad (57)$$

Knowing the exact value of $t_{(p+m)(q+n)}$, we can continue to find the second solution, $t_{(p+3m)(q+3n)}$, by inversely solving Eq. (52), as

$$t_{(p+3m)(q+3n)} = \frac{1}{k_z} \arg \left(\frac{F(p+3m, q+3n)}{P_{\uparrow} \tau_1 \tau_2 \alpha \exp[jk_z(2t_{(p+m)(q+n)})] \{E_{p+2m, q+2n}^{\text{inc}} \alpha^{-1} \exp[-jk_z(2t_{p+m, q+n})] + E_{p,q}^{\text{inc}}\}} \right). \quad (58)$$

Based on mathematical induction, it can be seen that the general formula of the construction equation system takes the form

$$t_{[p+(2r+1)m][q+(2r+1)n]} = \frac{1}{k_z} \arg \left[\frac{F(p+(2r+1)m, q+(2r+1)n)}{P_{\uparrow} \tau_1 \tau_2 \alpha^r \exp(j2k_z \sum_{l=0}^{r-1} t_{p+(2l+1)m, q+(2l+1)n}) [\sum_{l=0}^r E_{p+2lm, q+2ln}^{\text{inc}} \alpha^{-1} \exp(-j2k_z \sum_{c=0}^{l-1} t_{p+(2c+1)m, q+(2c+1)n})]} \right]. \quad (59)$$

We can find, analytically, all the surface profile thicknesses of the DOE, t_{uv} , by sequentially solving Eq. (59) in the same way for all starting incidence position indices (p, q) .

The surface relief structure is constructed by the proposed method to generate the same lenslike phase modulation profile as was used in the case of normal incidence. The results are shown in Fig. 9. Figure 9(a) shows the surface relief structure obtained. There are significant differences in appearance between the surface relief structure obtained and that obtained by the conventional method, as can be seen by comparing Figs. 7(a) and 9(a). The surface relief structure obtained is quite asymmetric. The correction relief profile is shown in Fig. 9(b). Phase modulation errors can be corrected by the addition of a correction relief profile to the structure obtained by the conventional method.

5 Conclusions

A new transmittance function model of a multilevel DOE, taking multiple internal reflections into account, is proposed. Since the conventional transmittance function model does not consider multiple reflection effects inside the DOE, the effects generate both phase and amplitude modulation errors from the desired phase and amplitude modulation profile when the DOE is synthesized based on the conventional transmittance function model. To construct an appropriate surface relief structure for a DOE without inducing phase modulation errors by multiple internal reflections, a simple and reliable method is described, based on the proposed transmittance function model. The surface relief profile is found analytically for the case of oblique incidence, but only numerically for the case of normal incidence. The proposed method for calculating the transmit-

tance function of a multilevel DOE is purely algebraic, and as a result can be easily combined with the design algorithm for DOEs such as the iterative Fourier transform algorithm.

Acknowledgment

The authors acknowledge the financial support from the Ministry of Science and Technology of Korea through the National Research Laboratory Program.

References

1. J. Goodman, *Introduction to Fourier Optics*, 2nd ed., McGraw-Hill, New York (1981).
2. W. Singer and H. Tiziani, "Born approximation for the nonparaxial scalar treatment of thick phase gratings," *Appl. Opt.* **37**(7), 1249–1255 (1998).
3. A. Rohbach and W. Singer, "Scattering of a scalar field at dielectric surfaces by Born series expansion," *J. Opt. Soc. Am. A* **40**(10), 2651–2659 (1998).
4. W. Singer and K. Brenner, "Transition of the scalar field at a refracting surface in the generalized Kirchhoff diffraction theory," *J. Opt. Soc. Am. A* **12**(9), 1913–1919 (1995).
5. M. A. Golub, "Generalized conversion from the phase function to the blazed surface-relief profile of diffractive optical element," *J. Opt. Soc. Am. A* **16**(5), 1194–1201 (1999).
6. M. Born and E. Wolf, *Principles of Optics*, 7th ed., Cambridge University Press, New York (1999).



Hwi Kim received the BS and MS degrees from the School of Electrical Engineering at Seoul National University, Korea, in 2001 and 2003, respectively. He is currently a candidate for the PhD degree in that School. His primary research interests are in diffractive optics and photonic crystals.

has served as a committee member for various international conferences. He has authored or coauthored more than 110 papers in international journals and more than 170 international conference papers. In 1999, his laboratory was honored as a National Research Laboratory by the Ministry of Science and Technology of Korea. In 2002 he received the Presidential Young Scientist Award of Korea. His research fields are hologram applications, three-dimensional displays, and optical fiber gratings.



Byoungcho Lee received the BS and MS degrees in 1987 and 1989, respectively, from Seoul National University, Korea, in electronics engineering. He received a PhD degree in 1993 from the University of California at Berkeley in electrical engineering and computer science. In 1994, he joined the faculty of the School of Electrical Engineering, Seoul National University, where he is now an associate professor. He became a Fellow of the SPIE in 2002.

He is a senior member of the IEEE and a member of the OSA. He