INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

H4.SMR/845-3

Second Winter College on Optics

20 February - 10 March 1995

Diffractive Optics

J. Turunen

University of Joensuu
Finland

DIFFRACTIVE OPTICS

Jari Turunen
University of Joensuu, Finland

The range of optical functions that can be performed by traditional optical elements such as lenses, mirrors, and prisms is limited by the restrictions on the shapes of reflective and refractive surfaces that can be fabricated by grinding and polishing techniques. These techniques readily permit the manufacture of spherical and cylindrical surfaces only, and the generation of even slight aspherics greatly increases the cost. Modern diamond-turning techniques permit the fabrication of strongly aspheric optical surfaces, but still leave much to be desired in the manufacture of rotationally asymmetric surfaces, miniature elements, and arrayed or multiplexed optics.

There is, however, an alternative: in place of a macroscopic surface-modulation structure one may use microstructured surface or index modulation profiles. These can be fabricated on a flat substrate by methods such as selective etching or material deposition, or by interferometric recording (holography) if the substrate is coated with a layer of photosensitive material. Then the optical function of the element (e.g., focusing of an incident plane wave) may be achieved by a proper design of the microstructure, which has a thickness of the order of one optical wavelength. In the design process of such microstructured element one typically has to abandon geometrical optics, the standard tool in the design of traditional macrostructured optical elements, or at least to complement its results by considerations of interference and diffraction within the microstructure.

These preliminaries lead us naturally to the concept of diffractive optics, which may be defined as *the realization of a given optical function by means of microstructured media.* Diffractive optics is a novel technology with exceptional promise in modern optical design, optoelectronics, optical communication, metrology, materials processing, and various other application areas. It provides the means to fabricate, e.g., flat optical elements with wavefront-transformation capabilities equivalent to those of arbitrarily profiled smooth surfaces. Diffractive elements can be fabricated by standard microlithographic techniques and are therefore naturally suitable for the realization of arrayed and multiplexed micro-optical components. They can be copied cheaply in large quantities by techniques such as embossing. It is therefore clear that diffractive optics, as a distinct extension of traditional optics, will be one of the key enabling optical technologies of the future.

Diffractive optics, which employs microstructured surface or index modulation, will obviously demand an entire set of new optical-design techniques, most often based on the wave theory of light. In many cases the response of a diffractive element may be described adequately my means of a scalar treatment of electromagnetic wave propagation in homogeneous and microstructured media. However, the dimensions of the microstructures that can be fabricated keep reducing at a steady pace: it is already possible to manufacture details with subwavelength dimensions. Hence there is an increasing need to treat the analysis and design of diffractive elements by rigorous electromagnetic diffraction theory, i.e., by exact solution of Maxwell's equations.

In these lectures some basic principles of diffractive optics are first covered. Wave propagation in homogeneous and structured media are then briefly discussed with the aid of plane-wave representations of the electromagnetic field. Some of the most prominent synthesis methods of diffractive elements are described together with certain chosen applications of diffractive optics. The manufacture of diffractive elements is also covered in some detail, with the emphasis on modern microlithographic techniques. Finally,

the integration of several diffractive and other optoelectronic elements into a solid system is discussed.

1 Fundamentals of diffractive optics

The basic task of any optical element is to transform a given incident field into an output field, which should possess certain properties determined by the specified optical function. At a quite general level, we may state that the optical function can be defined by some function s (the signal), which has non-zero values only within some bounded spatial or spatial-frequency region, called the signal window W (see Fig. 1). For example, the optical function of a focusing element may be defined as the maximum concentration of energy into a focal spot. Then W is a single spatial point and s is the energy density at this point. As another example, a beam splitter converts an incident plane wave into two (or more) angularly separated output waves; now W is a spatial-frequency window of some predetermined width, while s defines the directions and the relative intensities of the desired plane waves.

Often either the illumination wave or the optical function of the element (or both) is multiplexed. The element may, e.g., be required to produce diffraction-limited focal spots, perhaps in predetermined locations, for a set of illuminating plane waves of different wavelengths or angles of incidence. Also in these cases it is possible to define s and W, although the definition of one or both may depend on some parameter associated with the incident field or the geometry of the element. For example, we may define the optical function of a polarizing beam splitter such that it should transmit a TE-polarized plane wave completely, while simultaneously deflecting any TM-polarized light away from the region of interest. Now W may again be seen as a spatial-frequency window, whereas s fixes the propagation directions and intensities, being dependent upon the state of polarization of the incident field.

In view of the above considerations it appears clear that the concepts

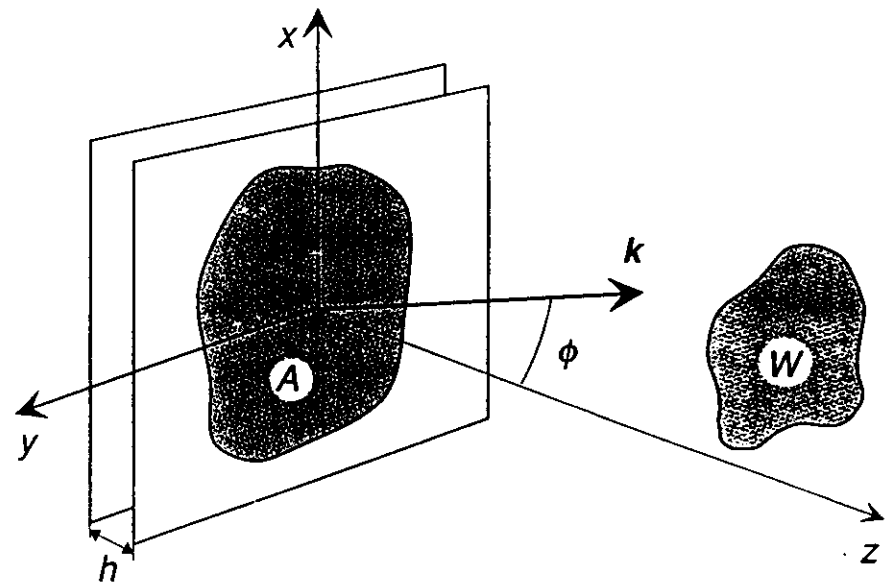

Figure 1: Illustration of the operation of an optical element confined between the planes $z = z_0 - h$ and $z = z_0$. The field in the exit plane $z = z_0$ is assumed to be insignificant outside the domain A, and we specify some properties of the field (the signal s) within a region W (the signal window).

of signal s and the signal window W can always be defined in a meaningful manner. In the simple examples considered, which were drawn out of classical optics, the introduction of s and W may appear somewhat artificial. However, for more exotic optical functions that can be realized using the methods of diffractive optics, these concepts turn out to be very useful. Consider, e.g., the problem of pattern projection: the objective is to convert an incident field (such as a Gaussian laser beam) into a field with a predetermined intensity distribution (signal s) in a bounded section of a transverse plane (signal window W), located some distance away from the element. Since s specifies only the field intensity inside W, the phase is left as a degree of freedom. By definition of W, the field outside it is of no concern and thus provides an additional degree of freedom. Moreover, the proportion of the incident energy that passes through W may be treated as a degree of freedom. Obviously, however, one would like to maximize this quantity (the diffraction efficiency η) as long as the pattern fidelity is not

sacrificed.

The diffraction efficiency η is indeed an important characteristic of any optical element or system. For example, in the theory of diffraction gratings for application in spectroscopy it is customary to compute spectral efficiency curves $\eta(\lambda)$ for the main diffraction order (Petit, 1980). Very often, in diffractive optics, the available degrees of freedom such as the phase freedom inside W and the complex-amplitude freedom outside it are used to increase η (Bryngdahl and Wyrowski, 1990, Wyrowski and Bryngdahl, 1991). Practical methods of achieving this will be discussed below. However, we note at this stage the remarkable fact that when the field within W is known we can calculate an upper bound for η without actually designing the element (Wyrowski, 1991; Wyrowski, 1993).

Once the signal s and the signal window W have been fixed by consideration of the requirements of the particular application, it remains to find a material structure that can generate s within W. This is a quite general design problem in optics, and our intention is to describe practical methods for its solution with the aid of diffractive structures. Before considering such methods in detail, it is necessary to provide the theoretical means of treating electromagnetic wave propagation in uniform and structured media. If we disregard quantum effects, which become important only at low light levels, the analysis methods may be divided into three classes: electromagnetic wave theory, scalar wave theory, and geometrical optics. The complexity of the mathematical formalism decreases in this sequence.

2 Wave propagation in uniform media

It follows from Maxwell's equations that, in a dielectric material of refractive index n each component U of a monochromatic electromagnetic field (wavelength λ) satisfies a Helmholtz equation

$$\left[\nabla^2 + (kn)^2\right] U(x,y,z) = 0, \tag{1}$$

where $k = 2\pi/\lambda$ and we have assumed a time dependence $\exp(-i\omega t)$, with $\omega = kc$, c being the speed of light in vacuum.

If we denote by $\hat{U}$ the Fourier transform of U with respect to x and y, i.e.,

$$\hat{U}(\alpha,\beta,z) = \frac{1}{(2\pi)^2}\int\int_{-\infty}^{\infty} U(x,y,z)\exp\left[-i\left(\alpha x+\beta y\right)\right]\mathrm{d}x\mathrm{d}y, \tag{2}$$

Eq. (1) is transformed into an ordinary second-order differential equation

$$\frac{\mathrm{d}^2}{\mathrm{d}z^2}\hat{U}(\alpha,\beta,z) + w^2(\alpha,\beta)\hat{U}(\alpha,\beta,z) = 0, \tag{3}$$

where

$$w(\alpha,\beta) = \begin{cases} \left[(kn)^2-(\alpha^2+\beta^2)\right]^{1/2} & \text{if } \alpha^2+\beta^2 \le (kn)^2 \\ i\left[(\alpha^2+\beta^2)-(kn)^2\right]^{1/2} & \text{otherwise.} \end{cases} \tag{4}$$

The solution of Eq. (3) may be written in the form

$$\hat{U}(\alpha,\beta,z) = T(\alpha,\beta)\exp\left[iw(\alpha,\beta)\left(z-z_0\right)\right] + R(\alpha,\beta)\exp\left[-iw(\alpha,\beta)\left(z-z_0\right)\right] \tag{5}$$

where T and R are arbitrary functions and $z = z_0$ is an arbitrary reference plane. Taking an inverse Fourier transform, we obtain

$$\begin{aligned} U(x,y,z) &= \int\int_{-\infty}^{\infty} T(\alpha,\beta)\exp\left\{i\left[\alpha x+\beta y+w(\alpha,\beta)\left(z-z_0\right)\right]\right\}\mathrm{d}\alpha\mathrm{d}\beta \\ &+ \int\int_{-\infty}^{\infty} R(\alpha,\beta)\exp\left\{i\left[\alpha x+\beta y-w(\alpha,\beta)\left(z-z_0\right)\right]\right\}\mathrm{d}\alpha\mathrm{d}\beta, \end{aligned} \tag{6}$$

which is a superposition of waves that propagate in directions determined by wave vectors $\boldsymbol{k} = (\alpha,\beta,w)$ or $\boldsymbol{k} = (\alpha,\beta,-w)$. Waves with real values of w represent homogeneous (propagating) plane waves, whereas waves with imaginary values of w either grow of decay exponentially with z. The latter are known as evanescent waves.

Let us assume that, e.g., the half-space $z > z_0$ is source-free and non-structured. Then, obviously, homogeneous waves propagating in negative z-directions cannot exist here, and evanescent waves that grow exponentially

with increasing z must also vanish. Hence $R(\alpha, \beta) = 0$ and we are left with

$$U(x, y, z) = \int\int_{-\infty}^{\infty} T(\alpha, \beta) P(\alpha, \beta, z) \exp\left[i\left(\alpha x + \beta y\right)\right] \mathrm{d}\alpha \mathrm{d}\beta, \qquad (7)$$

where

$$P(\alpha, \beta, z) = \exp\left[iw(\alpha, \beta)\left(z - z_0\right)\right], \qquad (8)$$

and by Fourier transformation at the plane $z = z_0$ we obtain

$$T(\alpha, \beta) = \frac{1}{(2\pi)^2} \int\int_{-\infty}^{\infty} U(x, y, z_0) \exp\left[-i\left(\alpha x + \beta y\right)\right] \mathrm{d}x \mathrm{d}y. \qquad (9)$$

Expression (7)–(9) are known as the angular-spectrum representation of an optical field, and $T(\alpha, \beta)$ is known as the angular spectrum of plane waves associated with the field $U(x, y, z_0)$. Had we assumed the half-space $z < z_0$ source-free and non-structured, the term $T(\alpha, \beta)$ instead of $R(\alpha, \beta)$ would have vanished.

In the form given above, the angular spectrum representation is a rigorous solution of the wave propagation problem in a uniform medium, but only for a scalar field. To obtain the solution for a vectorial electromagnetic field, we write a solution of this form for all six scalar components of the electric and magnetic field vectors. In view of Maxwell's equations, only two of these six components (for example, the x- and y-components of the electric field) are independent. The other scalar components can be related to them by a straightforward application of Maxwell's equations.

Let us next assume that the angular spectrum $T(\alpha, \beta)$ has appreciable values only at low spatial frequencies, i.e., if

$$\alpha^2 + \beta^2 \ll (kn)^2 . \qquad (10)$$

In other words, all plane-wave components of the field propagate in directions close to the z-axis, and evanescent waves are excluded. This condition, which may be called the paraxial approximation, permits us to write

$$w(\alpha, \beta) \approx kn - \left(\alpha^2 + \beta^2\right) / (2kn). \qquad (11)$$

Substitution into Eq. (7) gives

$$U(x,y,z) = \exp(ikn\Delta z) \int\int_{-\infty}^{\infty} T(\alpha,\beta) \times \exp\left[-i\left(\alpha^2+\beta^2\right)\Delta z/(2kn)\right] \exp\left[i\left(\alpha x+\beta y\right)\right] \mathrm{d}\alpha\mathrm{d}\beta, \tag{12}$$

where we have abbreviated $\Delta z = z - z_0$. Inserting Eq. (9) into Eq. (12), we obtain

$$U(x,y,z) = (2\pi)^{-2} \exp(ikn\Delta z) \int\int_{-\infty}^{\infty} U(x',y',z_0)\mathrm{d}x'\mathrm{d}y' \times \int\int_{-\infty}^{\infty} \exp\left\{-i\left(\alpha^2+\beta^2\right)\Delta z/(2kn) + i\left[(x-x')\,\alpha + (y-y')\,\beta\right]\right\} \mathrm{d}\alpha\mathrm{d}\beta. \tag{13}$$

If we now perform the (α,β)-integrations using the formula

$$\int_{-\infty}^{\infty} \exp\left(-ax^2+bx\right)\mathrm{d}x = (\pi/a)^{1/2} \exp\left(b^2/4a\right), \tag{14}$$

we obtain the well-known Fresnel propagation integral

$$U(x,y,z) = \frac{n}{i\lambda\Delta z} \exp(ikn\Delta z) \times \int\int_{-\infty}^{\infty} U(x',y',z_0) \exp\left\{\frac{i\pi n}{\lambda\Delta z}\left[(x-x')^2+(y-y')^2\right]\right\} \mathrm{d}x'\mathrm{d}y'. \tag{15}$$

Note that in the derivation of the Fresnel formula from the exact angular spectrum of plane waves the only approximation made is that of Eq. (10). Hence, whenever this assumption is valid, we may use the Fresnel formula even if the signal window W is located in the immediate vicinity of the plane $z = z_0$. This fact is not easily seen if Eq. (15) is derived by the more conventional method (Goodman, 1968).

Let us next assume that W recedes into the far zone, i.e., if $\rho_{\max}$ denotes the radius of a circle that encloses A in Fig. 1,

$$\Delta z \gg (\pi/\lambda)\,\rho_{\max}^2. \tag{16}$$

Expanding the squares in Eq. (15), we get

$$U(x,y,z) = \frac{n}{i\lambda\Delta z}\exp\left(ikn\Delta z\right)\exp\left[\frac{i\pi n}{\lambda\Delta z}\left(x^2+y^2\right)\right]\int\int_{-\infty}^{\infty} U(x',y',z_0) \times \exp\left[\frac{i\pi n}{\lambda\Delta z}\left(x'^2+y'^2\right)\right]\exp\left[-\frac{i2\pi n}{\lambda\Delta z}\left(xx'+yy'\right)\right]\mathrm{d}x'\mathrm{d}y', \tag{17}$$

and if the condition (16) is valid, the quadratic phase term inside the integral is negligible, which enables us to write

$$U(x,y,z) = \frac{n}{i\lambda\Delta z}\exp\left(ikn\Delta z\right)\exp\left[\frac{i\pi n}{\lambda\Delta z}\left(x^2+y^2\right)\right] \times\int\int_{-\infty}^{\infty} U(x',y',z_0)\exp\left[-\frac{i2\pi n}{\lambda\Delta z}\left(xx'+yy'\right)\right]\mathrm{d}x'\mathrm{d}y'. \tag{18}$$

This is recognized as the Fraunhofer propagation formula.

3 Scattering by non-periodic structures: a groove in a perfectly conducting substrate

We now have at our possession the mathematical means to propagate an optical field from the exit face of the element to the signal window W. To be able to analyze the entire geometry shown in Fig. 1, we need methods to obtain the field in the exit plane from the knowledge of the incident field and the structure of the element. A complete description of the available methods would fill more than one book because of the great diversity of possible structures and methods of solution (Beckmann and Spizzichino, 1963; Gaylord and Moharam, 1985; Korpel, 1988; Maystre, 1984; Petit, 1980; Solymar and Cooke, 1981). We must therefore restrict the discussion to some special geometries and methods of approach, which will hopefully provide some insight into the general methodology.

Consider the geometry of Fig. 2, where a y-invariant groove of width c and depth h in a perfectly conducting substrate is illuminated by a two-dimensional wave that propagates in the xz-plane towards the groove. We

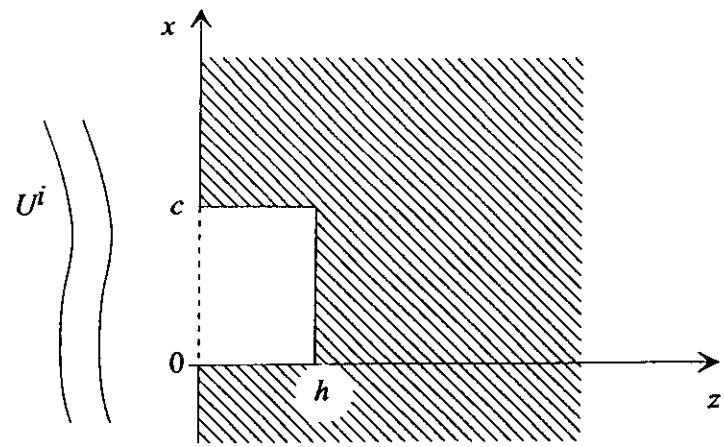

Figure 2: Diffraction of a two-dimensional electromagnetic field by a groove in a perfectly conducting substrate: geometry and notation.

denote the (known) angular spectrum of the incident field by $A(\alpha)$ and, in view of the y-invariance of the diffraction problem, write the incident wave in the form

$$U^i(x,z) = \int_{-\infty}^{\infty} A(\alpha) \exp\{i[\alpha x + w(\alpha)z]\}\, \mathrm{d}\alpha \tag{19}$$

in the region $z < 0$. Diffraction by the groove produces in the same region a diffracted wave, which may be written in the form

$$U^d(x,z) = \int_{-\infty}^{\infty} R(\alpha) \exp\{i[\alpha x - w(\alpha)z]\}\, \mathrm{d}\alpha, \tag{20}$$

where the minus sign in front of w indicates [c.f., Eq. (6)] that this diffracted field will consist of homogeneous waves that propagate in negative z-directions and of evanescent waves that decay exponentially when $z \to -\infty$.

From now onwards, we assume for simplicity that U denotes the y-component of the electric field whenever the electromagnetic nature of the optical field is of concern. When the diffracting object is y-independent, we also assume that this is the only non-vanishing component of the electric field, i.e., we consider the case of TE polarization. The case of TM polarization, where the electric field has x and z-components, but the magnetic

field has only a y-component, can also be treated but the mathematical formalism typically becomes slightly more complicated.

In general, Eq. (20) does not apply inside the groove, although there has been some controversy over this question (Maystre, 1984). To obtain an appropriate field representation inside the groove we use techniques that are familiar from any first course on electromagnetics. We separate the variables in the Helmholtz equation, writing $U(x,z) = X(x)Z(z)$, to obtain two ordinary second-order differential equations, whose solutions can be expressed in terms of exponential functions. The electromagnetic boundary conditions require that the electric field be zero at the perfectly conducting groove boundaries. Thus the solution may be written as a superposition of waveguide modes, in the form

$$U^g(x,z) = \sum_{m=1}^{\infty} X_m(x) \left\{ \exp\left(i\gamma_m z\right) - \exp\left[i\gamma_m \left(2h - z\right)\right] \right\} a_m, \tag{21}$$

where

$$X_m(x) = (2/c)^{1/2} \sin\left(m\pi x/c\right), \tag{22}$$

and

$$\gamma_m = \begin{cases} \left[(kn)^2 - (m\pi/c)^2\right]^{1/2} & \text{if } |m| \leq knc/\pi \\ i\left[(m\pi/c)^2 - (kn)^2\right]^{1/2} & \text{otherwise.} \end{cases} \tag{23}$$

Here a_m are as-yet undetermined parameters.

We now have expressions for the fields in the region $z < 0$ (where the total field is the sum of U^i and U^d) and inside the groove. It remains to apply the electromagnetic boundary conditions at $z = 0$ to solve the parameters a_m and to obtain the angular spectrum $R(\alpha)$ of the diffracted field. Requiring that the electric field vanishes at the perfectly conducting boundaries $x < 0$ and $x > c$, and is continuous if $0 < x < c$ we obtain

$$\int_{-\infty}^{\infty} \left[A(\alpha) + R(\alpha)\right] \exp\left(i\alpha x\right) \mathrm{d}\alpha \tag{24}$$

$$= \begin{cases} \sum_{m=1}^{\infty} X_m(x) \left[1 - \exp\left(i2\gamma_m h\right)\right] a_m & \text{if } 0 < x < c \\ 0 & \text{otherwise.} \end{cases} \tag{25}$$

Multiplication by $\exp(-i\alpha' x)$, integration from $-\infty$ to ∞, and use of the Fourier-integral definition of Dirac's delta function then leads to

$$R(\alpha') = \frac{1}{2\pi} \sum_{m=1}^{\infty} I_m(\alpha') \left[1 - \exp(i2\gamma_m h)\right] a_m - A(\alpha'), \tag{26}$$

where the integrals

$$I_m(\alpha) = \int_0^c \exp(-i\alpha x)\, X_m(x) \mathrm{d}x \tag{27}$$

can be evaluated analytically. Thus we have an expression for $R(\alpha)$ in terms of the modal coefficients a_m.

It also follows from the electromagnetic boundary conditions that the x-component of the magnetic field, i.e., the normal derivative of U, must be continuous across the plane $z = 0$ in the groove region $0 < x < c$. This implies that

$$\int_{-\infty}^{\infty} r(\alpha) \left[A(\alpha) - R(\alpha)\right] \exp(i\alpha x)\, \mathrm{d}\alpha = \sum_{m=1}^{\infty} \gamma_m X_m(x) \left[1 + \exp(i2\gamma_m h)\right] a_m \tag{28}$$

when $0 < x < c$. Multiplication by $X_p(x)$, integration over the interval $0 < x < c$, and use of the orthonormality of $X_m(x)$ leads to

$$\int_{-\infty}^{\infty} I_p^*(\alpha) r(\alpha) \left[A(\alpha) - R(\alpha)\right] \mathrm{d}\alpha = \sum_{m=1}^{\infty} \gamma_m \left[1 + \exp(i2\gamma_m h)\right] \delta_{pm} a_m, \tag{29}$$

where the asterisk denotes complex conjugation and δ is the Kronecker delta function. Insertion of Eq. (26) into Eq. (29) yields a set of linear equations

$$\sum_{m=1}^{\infty} \{K_{pm}[1-\exp(i2\gamma_m h)] + \gamma_m [1+\exp(i2\gamma_m h)]\,\delta_{pm}\}\, a_m$$
$$= 2\int_{-\infty}^{\infty} r(\alpha) I_p^*(\alpha) A(\alpha) \mathrm{d}\alpha, \qquad (30)$$

where we have denoted

$$K_{pm} = \frac{1}{2\pi}\int_{-\infty}^{\infty} r(\alpha) I_m(\alpha) I_p^*(\alpha) \mathrm{d}\alpha. \qquad (31)$$

If the summations in Eqs. (30) are truncated and the integrals K_{mp} are evaluated numerically, the modal coefficients a_m are obtained from the linear system (30) and the angular spectrum $R(\alpha)$ is finally obtained from Eq. (26). Thus the diffraction problem of a TE-polarized field by a groove is solved completely.

The treatment of the groove-diffraction problem given above is a completely rigorous solution of Maxwell's equations and the electromagnetic boundary conditions, and it requires numerical computations of some complexity. If the incident field is paraxial, one may envisage a far simpler method of treating the diffraction problem: when traversing to the bottom of the groove and back into the plane $z = 0$ the incident field could be expected to acquire a phase delay of $\phi/(2\pi) = 2h/\lambda$ compared to the field incident upon regions outside the groove. This would give a diffracted field of the form

$$U^d(x,0) = U^i(x,0)\exp[i\phi(x)] \qquad (32)$$

where $\phi(x)$ represent the binary-valued spatial variation of the phase delay. The angular spectrum $R(\alpha)$ is then be given by the y-invariant version of Eq. (9). In general, we may state that this approximate model gives acceptable results for $R(\alpha)$ if the groove width c is several times greater than the optical wavelength λ (and $h < \lambda$).

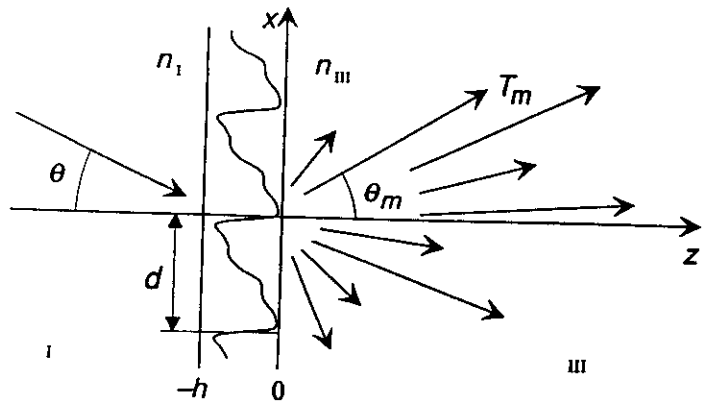

Figure 3: Diffraction of a plane wave by a grating. Region I is dielectric, but regions II and III may be metallic, with a finite conductivity.

4 Diffraction by a periodic structure

In the example given above we assumed the substrate to be perfectly conducting, which greatly simplifies the theory but is not an entirely appropriate assumption in the visible region of the electromagnetic spectrum. Moreover, the geometry considered was rather special.

We proceed to present a method of solution applicable to a wide class of periodic media, i.e., gratings (see Fig. 3). The modulated structure of period d in the x-direction is assumed y-independent, z-independent in the range $0 < z < h$, and illuminated by a plane wave

$$U^{\rm i}(x,z) = \exp\left[i\left(\alpha_0 x + r_0 z\right)\right], \tag{33}$$

where $\alpha_0 = kn_{\rm I}\sin\theta$ and $r_0 = kn_{\rm I}\cos\theta$.

In view of the Floquet-Bloch theorem, the periodicity of the material forces the field to be of a pseudoperiodic form $U(x+d,z) = U(x,z)\exp(i\psi)$, and the incident field fixes $\psi = \alpha_0 d$. If we apply the pseudoperiodicity to Eq. (20), we obtain a condition

$$\int_{-\infty}^{\infty} R(\alpha)\exp\left\{i\left[\alpha x - w(\alpha)z\right]\right\}\left[\exp(i\alpha d) - \exp(i\alpha_0 d)\right]\mathrm{d}\alpha = 0, \tag{34}$$

which can only hold if

$$\alpha = \alpha_l = \alpha_0 + 2\pi l/d, \tag{35}$$

l integer. Hence

$$R(\alpha) = \sum_{l=-\infty}^{\infty} R_l \delta(\alpha - \alpha_l) \tag{36}$$

and we obtain

$$U_{\mathrm{I}}^d(x,z) = \sum_{l=-\infty}^{\infty} R_l \exp\left[i\left(\alpha_l x - r_l z\right)\right], \tag{37}$$

with $r_l^2 = w^2(\alpha_l) = (kn_{\mathrm{I}})^2 - \alpha_l^2$. Similarly, in region III,

$$U_{\mathrm{III}}^d(x,z) = \sum_{l=-\infty}^{\infty} T_l \exp\left\{i\left[\alpha_l x - t_l(z-h)\right]\right\}, \tag{38}$$

where $t_l^2 = (kn_{\mathrm{III}})^2 - \alpha_l^2$ and when taking square roots we assume a sign convention $\Re\{t_l\} + \Im\{t_l\} > 0$ (for a metallic grating n_{III} is complex-valued). Expressions (37) and (38) are known as Rayleigh expansions.

Let us denote by $\epsilon_r(x)$ the (possibly complex) permittivity distribution of the grating. It then follows from Maxwell's equations that the electric field in region II satisfies a non-constant coefficient Helmholtz equation

$$\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} + k^2\epsilon_r(x)\right] U_{\mathrm{II}}(x,z) = 0. \tag{39}$$

To obtain a field representation in region II, we again separate the variables, which yields

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} X(x) + \left[k^2\epsilon_r(x) - \gamma^2\right] X(x) = 0 \tag{40}$$

and

$$\frac{\mathrm{d}^2}{\mathrm{d}z^2} Z(z) + \gamma^2 Z(z) = 0. \tag{41}$$

To solve Eq. (40), we express the permittivity in the form of a Fourier series

$$\epsilon_r(x) = \sum_{p=-\infty}^{\infty} \varepsilon_p \exp\left(i2\pi p x/d\right), \tag{42}$$

where the Fourier coefficients are given by

$$\varepsilon_p = \frac{1}{d}\int_0^d \epsilon_r(x) \exp\left(-i2\pi p x/d\right) \mathrm{d}x, \tag{43}$$

and attempt a pseudoperiodic expansion of the form

$$X(x) = \sum_{m=-\infty}^{\infty} P_m \exp(i\alpha_m x). \tag{44}$$

This leads to

$$\sum_{m=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} \left[(\alpha_m^2 + \gamma^2) \exp(i\alpha_p x)\, \delta_{mp} - k^2 \varepsilon_p \exp(i\alpha_{m+p} x)\right] P_m = 0. \tag{45}$$

If we multiply Eq. (45) by $\exp(-i\alpha_l x)$ and integrate over the period, we obtain a system of eigenvalue equations

$$\sum_{m=-\infty}^{\infty} \left(k^2 \varepsilon_{l-m} - \alpha_m^2 \delta_{lm}\right) P_m = \gamma^2 P_l, \tag{46}$$

which can be solved by standard numerical algorithms if the summation is truncated. This yields a set of discrete eigenvalues γ_n^2 and eigenvector-elements P_{ln}. The solution for $U(x,z)$ may now be expressed in the form

$$E_y(x,z) = \sum_{n=1}^{\infty} \left\{a_n \exp(i\gamma_n z) + b_n \exp\left[-i\gamma_n (z-h)\right]\right\} \sum_{l=-\infty}^{\infty} P_{ln} \exp(i\alpha_l x), \tag{47}$$

which also contains the (obvious) solution of Eq. (41). We adopt a sign convention $\Re\{\gamma_n\} + \Im\{\gamma_n\} > 0$ when taking square roots.

It again remains to apply the electromagnetic boundary conditions to obtain a complete solution of the diffraction problem. Now the boundary conditions require the continuity of U and its normal derivative at $z = 0$ and $z = h$. The continuity of U at $z = 0$ yields

$$\begin{aligned} &\exp(i\alpha_0 x) + \sum_{l=-\infty}^{\infty} R_l \exp(i\alpha_l x) \\ &\quad = \sum_{n=1}^{\infty} \left[a_n + b_n \exp(i\gamma_n h)\right] \sum_{l=-\infty}^{\infty} P_{ln} \exp(i\alpha_l x), \end{aligned} \tag{48}$$

or

$$R_l = \sum_{n=1}^{\infty} \left[a_n + b_n \exp(i\gamma_n h)\right] P_{ln} - \delta_{l0}. \tag{49}$$

The other boundary conditions give

$$T_l = \sum_{n=1}^{\infty} \left[a_n \exp\left(i\gamma_n h\right) + b_n \right] P_{ln}, \tag{50}$$

$$r_l \left(\delta_{l0} - R_l\right) = \sum_{n=1}^{\infty} \gamma_n \left[a_n - b_n \exp\left(i\gamma_n h\right) \right] P_{ln}, \tag{51}$$

and

$$t_l T_l = \sum_{n=1}^{\infty} \gamma_n \left[a_n \exp\left(i\gamma_n h\right) - b_n \right] P_{ln}. \tag{52}$$

We insert Eq. (49) into Eq. (51) and Eq. (50) into Eq. (52) to obtain a system of linear equations

$$\sum_{n=1}^{\infty} \left(r_l + \gamma_n\right) P_{ln} a_n + \sum_{n=1}^{\infty} \left(r_l - \gamma_n\right) \exp\left(i\gamma_n h\right) P_{ln} b_n = 2 r_l \delta_{l0}, \tag{53}$$

$$\sum_{n=1}^{\infty} \left(t_l - \gamma_n\right) \exp\left(i\gamma_n h\right) P_{ln} a_n + \sum_{n=1}^{\infty} \left(t_l + \gamma_n\right) P_{ln} b_n = 0, \tag{54}$$

from which a_n and b_n may be solved by truncating the summations. Then the complex amplitudes of the diffraction orders are calculated from Eqs. (49) and (50).

It is probably clear to the reader by now that the complex amplitudes R_l and T_l correspond to the amplitudes of the diffraction orders of the grating in regions I and III, respectively. The propagation directions θ_l of these orders are given by the well-known grating equations, which can be obtained from Eq. (35) if we write $\alpha_l = k n_{\mathrm{I}} \sin\theta_l$ in region I and $\alpha_l = k n_{\mathrm{III}} \sin\theta_l$ in III.

We are normally interested in the power that propagates with a particular diffracted wave. The diffraction efficiency η_l of lth diffacted order is obtained if we divide the z-component of the Poynting vector associated with this order by that of the incident wave. For reflected orders we get

$$\eta_l = \left(r_l / r_0\right) \left|R_l\right|^2, \tag{55}$$

and for transmitted orders

$$\eta_l = \left(t_l / r_0\right) \left|T_l\right|^2. \tag{56}$$

For a dielectric grating the sum of the efficiencies of all reflected and transmitted orders is unity, which is an expression of energy conservation.

The rigorous method given above for the analysis of diffraction gratings again requires some straightforward programming, and yet again we may use optical path considerations to obtain approximate results. If the incident wave is paraxial ($\sin\theta \approx \theta$) and the materials in regions I–III are dielectric, we obtain a spatially variable, periodic phase-delay profile

$$\phi(x) = kh\sqrt{\epsilon_r(x)} = khn(x), \tag{57}$$

and the complex amplitudes T_m of the transmitted orders are then obtained, by inverting Eq. (38) in the plane $z = h$, from

$$T_m = \frac{1}{d}\int_0^d \exp\left[i\phi(x)\right]\exp\left(-i2\pi x/d\right)\mathrm{d}x. \tag{58}$$

We compare the predictions of the rigorous and the approximate approaches in Fig. 4 for a binary profile with $n_{\mathrm{I}} = 1.5$, $\epsilon_r = 2.25$ when $0 < x < d/2$, $\epsilon_r = 1$ when $d/2 < x < d$, $n_{\mathrm{III}} = 1$, $\theta = 0$, and $h = \lambda$. The prediction of the approximate model is that $\eta_1 = 40.5\%$ and $\eta_0 = 0$ irrespective of d/λ, and we see that the approximate model is adequate when the length of the grating period is at least a few wavelengths.

5 Amplitude transmittance method

The rigorously calculated results given in Fig. 4 show that in many instances it is appropriate to use approximate methods based on optical path calculations. In general, if the (possibly complex) refractive index of the modulated region is given by $n(x, y, z)$, $0 < z < h$, we may introduce a complex-amplitude transmission function

$$t(x, y) = \exp\left[ik\int_0^h n(x, y, z)\mathrm{d}z\right] \tag{59}$$

such that

$$U^d_{\mathrm{III}}(x, y, h) = t(x, y)U^{\mathrm{i}}(x, y, 0). \tag{60}$$

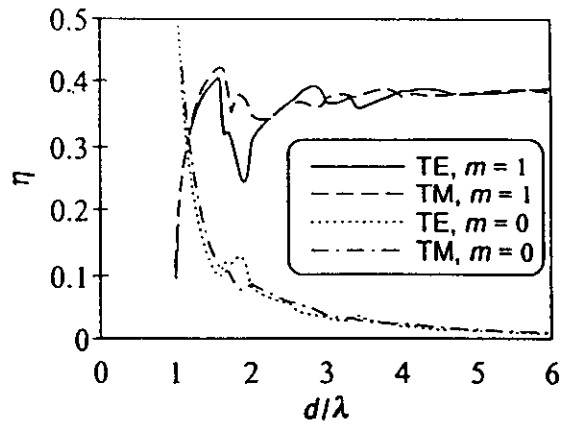

Figure 4: Rigorously calculated diffraction efficiencies of orders $m = 0$ and $m = 1$ for a binary grating as a function of the grating period d (normalized to wavelength λ.

If the element is sufficiently thin and its features are sufficiently large compared to the wavelength, this amplitude transmittance approach gives useful results for periodic and non-periodic elements alike.

It should be stressed, however, that the range of validity of this approach depends very much on the type of profile investigated, and that the binary case considered in Fig. 4 is among the least critical in this respect. Let us consider a Q-level dielectric surface-relief grating with $n(x,z) = n$ if $z \leq h(x)$, $n(x,z) = 1$ elsewhere, and $h(x) = h(q-1)/Q$ when $d(q-1)/Q \leq x < dq/Q$. A straightforward calculation gives

$$\eta_m = \mathrm{sinc}^2(m/Q) \left| \frac{1}{Q} \frac{1 - \exp\{-i2\pi[h(n-1)/\lambda + m]\}}{\exp\{i2\pi[h(n-1)/\lambda + m]/Q\} - 1} \right|^2, \qquad (61)$$

where $\mathrm{sinc}(x) = \sin(\pi x)/(\pi x)$. The ratio that appears in Eq. (61) is ≤ 1, and it approaches unity when $2\pi[h(n-1)/\lambda + m] \to 0$. Hence the maximum (minus) first-order efficiency of a Q-level staircase grating is

$$\eta_{-1} = \mathrm{sinc}^2(1/Q). \qquad (62)$$

For a binary grating we have $\eta_{-1} \approx 40.5\%$, for a four-level grating $\eta_{-1} \approx 81\%$, for an eight-level grating $\eta_{-1} \approx 95\%$, etc. When $Q \to \infty$, i.e., we obtain

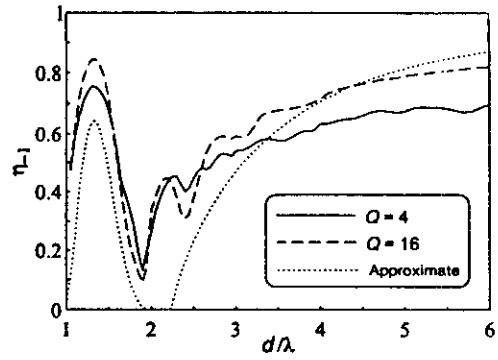

Figure 5: First-order diffraction efficiencies of stair-step gratings with four (solid line) and 16 (dashed line) levels, calculated with rigorous diffraction theory, and approximation given by a Beckmann-type model.

a triangular profile, $\eta_{-1} \to 1$. Figure 6 gives the rigorously calculated results, which show a complete failure of the approximate model when $d/\lambda \approx 2$, and a relatively slow convergence towards the complex-amplitude transmittance results when d/λ is increased.

6 Synthesis approaches

Let us denote by U_s a field that satisfies the signal function s in W at least within the tolerances of the application. If W is a section of some plane $z = z_s$ and we know U_s in it, the methods of sect. 2 may be applied to determine the field distribution $U_s(x, y, 0)$ across the exit plane of the element (confined between $z = -h$ and $z = 0$). It appears that the remaining task is to find a diffractive element that transforms the incident field $U_i(x, y, -h)$ into $U_s(x, y, 0)$ at least within a proportionality factor. Hence, in view of the complex-amplitude transmittance approach, we should realize a transmittance

$$t(x, y) = \alpha \frac{U_s(x, y, 0)}{U_i(x, y, -h)} \tag{63}$$

where α is chosen such that $|t(x,y)| \leq 1$ for all x and y (in a passive diffractive element we have no amplification).

We note that it is, in general, not a trivial matter to determine U_s from s if only the intensity is fixed in W by s. A constant phase could be chosen, but then $U_s(x,y,0)$ will have a high intensity near the axis, and weak off-axis components that nevertheless contain most of the important information about $U_s(x,y,z_s)$. This situation is well known in holography, and is highly unsatisfactory in view of Eq. (63), because the factor α is determined by the dc peak and a lot of attenuation is needed to obtain a linear recording of the important off-axis information.

In holography one typically employs diffuse illumination to smooth the intensity of $U_s(x,y,0)$. Following this lead, we could introduce a random phase in $U_s(x,y,z_s)$. However, then a significant portion of $U_s(x,y,0)$ may fall outside the finite aperture of the diffractive element and we obtain speckle-like intensity variations, also well known in holography. What we therefore need is a signal-dependent phase that makes $U_s(x,y,z_s)$ essentially band-limited, i.e., $|U_s(x,y,0)|$ is confined within the permitted aperture of the element. Practical methods of designing the phase of $U_s(x,y,z_s)$ have been described by Wyrowski and Bryngdahl (1989), and by Bräuer, Wyrowski, and Bryngdahl (1991).

In view of Eq. (63), we need in general an element with continuous attenuation and phase-delay profiles. It is possible to fabricate such elements (Chu, Fienup, and Goodman, 1973), but it is difficult to reach a sufficient accuracy. One would prefer elements with more restricted structure. First, in view of fabrication considerations, it would be convenient to have no attenuation at all, because then a surface-relief profile alone would suffice. Second, one would often like to quantize the surface-relief profile, permitting perhaps only two levels.

Any constraint on the structure of the element will undoubtedly lead to a diffracted field

$$U_d(x,y,0) = t(x,y)U_i(x,y,-h) \neq U_s(x,y,0) \tag{64}$$

If we now propagate U_d into the plane $z = z_s$, we obtain a field of the form

$$U_d(x, y, z_s) = U_s(x, y, z_s) + U_f(x, y, z_s) \tag{65}$$

where U_f represents noise. This noise is of no concern if it vanishes inside W. Therefore our objective must be to achieve this situation by modification of the diffractive structure. This procedure is known as coding.

An effective approach to coding is the use of a carrier grating to divide the incident field into several diffraction orders, of which we utilize one (typically the strongest diffracted order) and therefore move W off-axis. We encode the amplitude and phase of the ratio $U_s(x, y, 0)/U_i(x, y, -h)$ into local variations of the diffraction efficiency and phase of the chosen carrier order, respectively.

It is not difficult to prove analytically, or to see by geometrical construction, that a shift of the carrier-grating structure by an amount Δx in the x-direction will modify the phase of mth diffraction order according to

$$\phi'_m = \phi_m - 2\pi m \Delta x/d, \tag{66}$$

where d is the period of the carrier grating. This result holds quite rigorously, irrespective of the structure of the grating period, and it is known as Lohmann's detour-phase principle for Lohmann was the first to use it in coding (Brown and Lohmann, 1966; Lohmann and Paris, 1967). Equation (66) shows that we can modulate the phase of, e.g, the first diffraction order at will by local modulation of the shift Δx. For example, a binary surface-relief grating with a first-order efficiency of 40.5% can be used as a carrier, and coding of the amplitude information is achieved through a local variation of the aspect ratio (groove width/period) since

$$\eta_{\pm 1} = \frac{4}{\pi^2} \sin^2(\pi c/d), \tag{67}$$

where we have assumed a groove depth $h = \lambda/(n-1)$. In two-dimensional modulation of the carrier grating it is convenient to divide the element into stripes and apply phase and amplitude modulation to each stripe separately.

The diffraction efficiency and also the signal-to-noise ratio within W can be improved significantly by iterative coding schemes, which make use of the design freedoms introduced in sect. 1. There are two general approaches to iterative solution of the synthesis problem. We call these the direct approach and the inverse approach. In both we make use of a merit function, which measures the quality of the solution.

In the direct approach we first define a type of microstructure that we are able to manufacture. This structure contains a set of parameters that may each take at least two different values. We then try to obtain s in W by making modifications in this set of parameters and by observing the effects of such modifications in the merit function. Typically we first choose a random parameter set and evaluate the merit function associated with this configuration. Then we begin an optimization process, in which the parameters are modified according to some coherent strategy until the merit function reaches a satisfactory value or the process stagnates. Any optimization scheme (gradient algorithms, simulated annealing, genetic algorithms, etc.) may be used. Of course, the numerical efficiency may depend critically on the choice of the optimization scheme and also upon the skill of numerical implementation. However, from a fundamental point of view, these are secondary considerations compared to the choice of the merit function, which should simultaneously account for the available degrees of freedom and the constraints of the synthesis problem. Since the merit function is defined only inside W, the amplitude freedom is effectively used throughout the design procedure. The scale freedom can be accounted for by including in the merit function a variable related to efficiency η and planning the optimization strategy in such a manner that η is ultimately maximized.

In the inverse approach, we start from the signal wave U_s in W. Once $U_s(x, y, 0)$ has been calculated, we determine a transmittance $t(x, y)$ by requiring that the constraints be satisfied. Then we calculate $U_d(x, y, 0) = t(x, y)U_i(x, y, -h)$ and compute the field in W. If this new field satisfies the signal function within the prescribed accuracy, the process is stopped. If

not, we modify it such that s is satisfied within W, using some or all of the available freedoms to avoid exceedingly drastic changes (for example, if only the intensity is fixed, we leave the phase unchanged). Then we return to the plane $z = 0$ and continue the iteration until a satisfactory solution is found or the process stagnates. We stress that it is not necessary to demand that the constraints in $z = 0$ be fully satisfied in the beginning stages of the procedure. For example, if we wish to design an element with a binary-valued $t(x, y)$, it has proved beneficial to introduce the constraints gradually during the iteration (Wyrowski, 1989).

If we compare the direct and the inverse approaches, the strength of the direct scheme is that it is available whenever we are able to determine the response of the diffractive structure either analytically or numerically. The inverse approach, on the other hand, is applicable only if the complex-amplitude transmittance approach is valid. However, within its domain of applicability, the inverse approach is numerically much more efficient than the direct approach, not least because the Fast Fourier Transform (FFT) algorithm may be used to evaluate the Fresnel and Fraunhofer propagation integrals in a highly efficient parallel fashion.

We now proceed to apply these synthesis methods to some selected applications.

7 Design of diffractive lenses

It is sometimes possible to ignore the amplitude modulation altogether without causing error in the signal. For example, if we wish to focus $U_i(x, y, -h)$ into a spot, the essential point is to transform its phase profile into a perfect spherical profile with a radius of curvature equal to z_s. Hence, in view of Fig. 6, the phase of $t(x, y)$ should be

$$\phi(x, y) = k\left(z_s - \sqrt{z_s^2 + x^2 + y^2}\right) - \arg\left[U_i(x, y, -h)\right] \qquad (68)$$

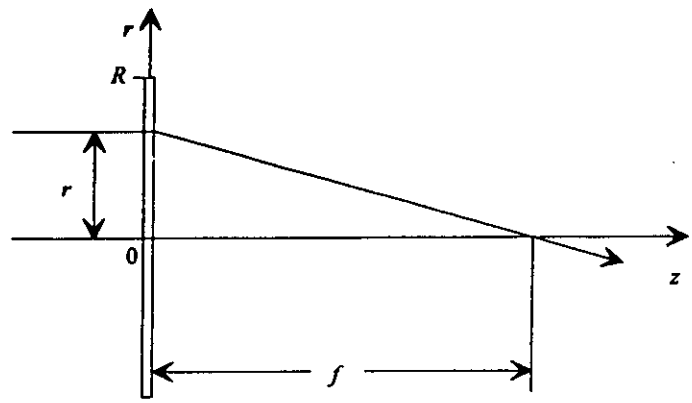

Figure 6: Focusing of an incident wave into a spot by a diffractive lens of focal length $f = z_s$.

Note that we have permitted here an aberrated incident wave, and that a diffractive element can straightforwardly compensate for the aberrations. Very often the incident wave is a plane wave, and then we have

$$\phi(r) = k\left(z_s - \sqrt{z_s^2 + r^2}\right), \tag{69}$$

where $r = \sqrt{x^2 + y^2}$. In what follows, we assume this to be the case for simplicity.

An element that generates the required phase transformation can be constructed, e.g., by means of Lohmann's detour-phase principle, Eq. (66), if we consider a circular grating instead of a linear one and vary the radial shift Δr in a monotonic manner. This implies that the local grating period becomes a simple function of r:

$$d(r) = \frac{\lambda}{\sqrt{1 + (z_s/r)^2}}. \tag{70}$$

In general, the phase profile $\phi(x, y)$ and the local grating period are related through

$$d(x, y) = \frac{2\pi}{|\nabla\phi(x, y)|}. \tag{71}$$

The local efficiency of a Q-level quantized circular grating is given by Eq. (61) as long as $d(r) \gg \lambda$, but if this is not the case (a high numerical aperture) we can estimate the local efficiency from Fig. 5. Now, because we consider a circular grating with a variable period, the directions of the diffraction orders will vary with position r and generate a set of converging and diverging spherical waves (only the zeroth order generates a plane wave). The undesired waves will, of course, interfere with the desired spot at $z = z_s$, but their influence is greatly reduced by the fact that the radii of curvature are different. Therefore, in the case of diffractive lenses, it is not usually necessary to introduce an additional linear carrier frequency to avoid the overlap.

Very often one is interested in the performance of a diffractive lens over some finite field of view. We may apply the grating equation and Eq. (71) to obtain ray-tracing equations for diffractive surfaces:

$$n_2 \sin \theta'_{m,x} = n_1 \sin \theta_x + \frac{m}{k} \frac{\partial}{\partial x} \phi(x, y), \tag{72}$$

$$n_2 \sin \theta'_{m,y} = n_1 \sin \theta_y + \frac{m}{k} \frac{\partial}{\partial y} \phi(x, y), \tag{73}$$

where n_2 and n_1 are the refractive indices of the media behind and in front of the diffractive surface, respectively, the primed angles are related to the diffracted ray, the unprimed angles to the incident ray, and m is the index of the diffraction order. Note that with $m = 0$, or if $\phi(x, y)$ is constant, we have Snell's law. The ray-tracing equations can be used to calculate spot diagrams for diffractive lenses even if the diffractive surface is part of some hybrid system, which also contains refractive and reflective surfaces. Point spread functions of diffractive lenses can be computed by application of the angular-spectrum approach to the transmitted field, or by the Fresnel propagation formula in the paraxial case.

If the performance of a diffractive lens (or a hybrid system) is not satisfactory over the desired field of view, we may attempt to improve the

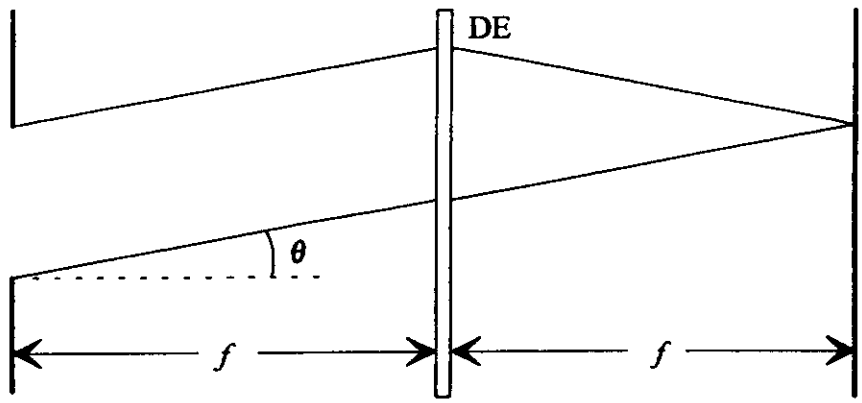

Figure 7: Diffractive lens in an exact Fourier-transforming configuration, suitable for, e.g., laser scanning.

situation by adding to the phase function $\phi(x,y)$ a correction term

$$\phi_c(x,y) = k\sum_{p,q} c_{pq}x^p y^q, \tag{74}$$

where c_{pq} are parameters that may be optimized using a lens design program that contains the diffractive ray-tracing equations (most commercial programs do). From a fabrication point of view, the addition of $\phi_c(x,y)$ is no complication. Therefore a diffractive surface can act much like a refractive surface with arbitrary aspheric coefficients, and this provides an idea of the improvements possible by inclusion of diffractive surfaces in an optical system.

There is an alternative approach to the design of diffractive lenses, which is somewhat more restricted in scope (it is difficult to treat hybrid systems) but has considerable theoretical appeal (Winick and Fienup, 1983; Cederquist and Fienup, 1987). We proceed to describe this method in a general level, using the configuration in Fig. 7 as an illustrative example.

In lens design we typically face a number of competing demands, and look for a balance between them. These demands can be formally indexed by a parameter p (the angle of incidence in Fig. 7). It is typically straightforward

to find a configuration, which is at least in some sense ideal for a single value of p, and we denote this by $\phi(x,y;p)$. Typically the incident field at $z=-h$ is a function of p, and we therefore introduce the notation $P(x,y;p) = |U_i(x,y,-h;p)|$. Furthermore, we define a weight function $W(p)$, which permits us to emphasize some values of p.

Let $\phi(x,y)$ denote the compromise phase delay that we wish to optimize by minimization of the norm (merit function)

$$E = \int\int_{-\infty}^{\infty} W(p)P(x,y;p)\left[\phi(x,y) - \phi(x,y;p)\right]^2 \mathrm{d}x\mathrm{d}y\mathrm{d}p. \tag{75}$$

The variation of E,

$$\delta E = 2\int\int_{-\infty}^{\infty} W(p)P(x,y;p)\left[\phi(x,y) - \phi(x,y;p)\right]\delta\phi(x,y)\mathrm{d}x\mathrm{d}y\mathrm{d}p, \tag{76}$$

reaches an extremum when $\delta E = 0$. For an arbitrary $\delta\phi(x,y)$, this yields

$$\int_{-\infty}^{\infty} W(p)P(x,y;p)\left[\phi(x,y) - \phi(x,y;p)\right]\mathrm{d}p \tag{77}$$

Hence

$$\phi(x,y) = \frac{\int_{-\infty}^{\infty} W(p)P(x,y;p)\phi(x,y;p)\mathrm{d}p}{\int_{-\infty}^{\infty} W(p)P(x,y;p)\mathrm{d}p} \tag{78}$$

represents the best compromise between $\phi(x,y;p)$ in terms of the norm E and the weight function $W(p)$.

8 Design of array illuminators

An array illuminator, or a multiple beam splitter, is an element capable of dividing an incident wave into a specified number of output waves with a specified power distribution. These output waves propagate into different directions and separate spatially after propagating a sufficient distance away from the element. Now the signal s is the distribution of spot intensities in some signal window W. Since the spots are required to be separate, no speckle problems will arise and therefore phase freedom is completely available in the design. Because s is a discrete set of spots, we are tempted

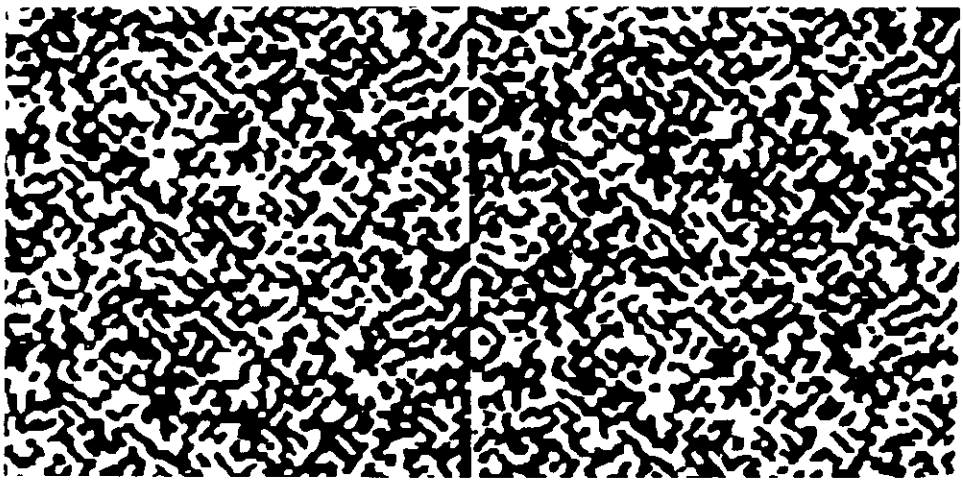

Figure 8: The structure of one period of an array illuminator that can generate 64×32 equal-efficiency diffraction orders with a diffraction efficiency $\eta \approx 75\%$.

to use periodic diffractive elements in this application: a periodic element will generate a set of diffraction orders, which will separate spatially upon propagation in free space if at least a few periods of the element are illuminated by a wave with a reasonably planar phase front. When a lens is placed behind such a periodic element, the diffraction orders give rise to spots in the back focal plane of the lens. The exact intensity distribution of the incident field will not influence the relative intensities of the spots, given by the efficiencies η_m of the diffraction orders of the grating. The efficiencies are determined by the internal structure of the grating period, and its optimization is therefore the central problem in the design of array illuminators of this type (for a discussion of other types of array illuminator, see Streibl, 1989). It is more than appropriate to cite the contributions of Dammann (Dammann and Görtler, 1971; Dammann and Klotz, 1978) to the design of array illuminators at this point.

In the paraxial domain, we may use either direct or inverse synthesis schemes to design array illuminators. Here the use of design freedoms is

perhaps more explicit than in any other application discussed above or below. As already mentioned, the phase freedom is completely available. This permits us to achieve high diffraction efficiencies η, i.e., to minimize the use of scale freedom. The use of amplitude freedom remains necessary, and it leads to non-zero values of the efficiencies of higher diffraction orders outside W. Figure 8 shows the structure of a binary surface-relief type array illuminator that can generate an array of 64×32 equal-intensity diffraction orders in the paraxial domain (Vasara *et al.*, 1992). The black and white areas distinguish phase-delay values of zero and π. This particular structure was obtained by a direct approach based on optimization of parameters that define the trapezoidal features in the profile by the method of simulated annealing. The diffraction efficiency of the structure is $\eta \approx 75\%$. Rather similar structures can be designed by iterative inverse methods at a greatly reduced computational cost.

We chose not to present a picture of the pattern generated by the element depicted in Fig. 8, fearing that the reproduction would not do justice to the actual fidelity. Instead, we show in Fig. 9 the 32×16 array of spots generated by another element with a somewhat simpler structure. Note the higher orders outside the array (i.e., the signal window W), which together contain some 25% of the incident optical energy. Note also the central zeroth-order spot, which is a results of a slight relief depth error.

The efficiency η into the array may, as expected in view of the considerations given above, be increased if we employ multilevel instead of binary surface-relief profiles. This is illustrated in Fig. 10. The replacement of a binary profile in (a) by a 16-level profile in (b) greatly reduces the noise outside W but is seen to simultaneously degrade the signal fidelity in W. The latter, however, if only a consequence of a more complicated fabrication procedure, which involves alignment of masks between adjacent lithographic steps (see the fabrication section below). The efficiency of the 16-level element is theoretically some 95% and in experimental realization $\sim$ 90% (Miller *et al.*, 1993).

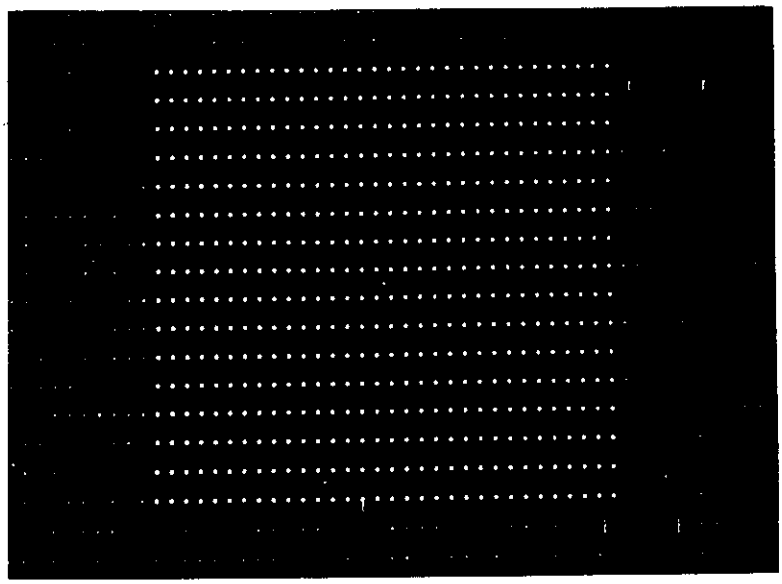

Figure 9: An array of 32 × 16 spots generated by a binary surface-relief element of the type shown in the previous figure.

The results given above apply to elements that operate in the paraxial domain of diffractive optics, i.e., the diffraction angles involved must be small. However, one increasingly wishes to have multiple beam splitters with large deflection angles, for example to divide the incident beam into two or more diffracted beams that can propagate within a substrate material along zig-zag paths by means of total internal reflection. This requires deflection angles of ~ 45 degrees, and therefore the application of rigorous diffraction theory is required (Noponen and Turunen, 1994a).

In the non-paraxial domain the application of rigorous theory can break some significant boundaries. For example, it is possible to design array illuminators with $\eta = 100\%$ (Noponen *et al.*, 1992) at least if the substrate is made of a perfectly conducting material and the period is chosen such that all propagating orders are included in the signal. In this case phase freedom and the amplitude freedom of evanescent waves (which do not carry energy) are used in the design procedure. The upper bound for diffraction efficiency (Wyrowski, 1991) is exceeded, but this is not a surprise because we no longer

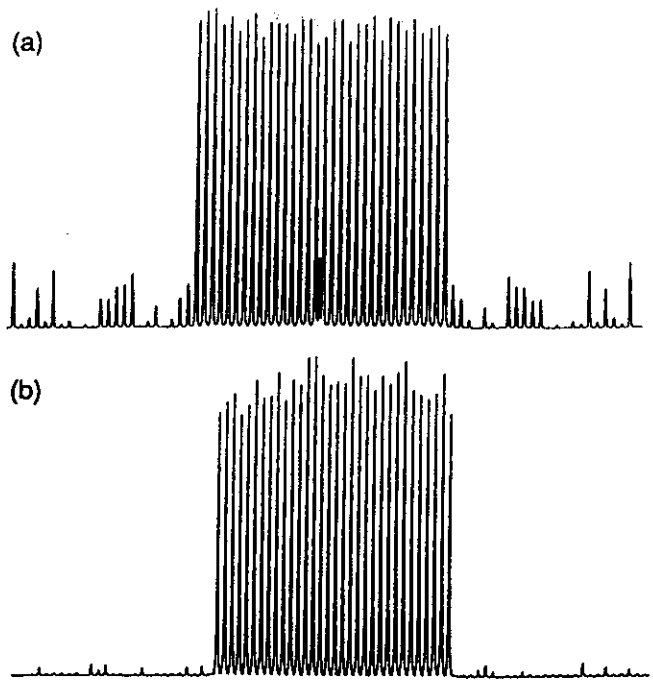

Figure 10: An experimental cross section of a 32–beam array generated by an element with (a) a binary and (b) 16-level surface-relief profile.

operate in the paraxial domain and the complex-amplitude transmittance method is not applicable.

9 Synthesis of subwavelength-period diffractive elements

Let us consider a grating with a period d that is smaller than the optical wavelength in the substrate material. Then, in view of the grating equation, a normally incident plane wave can not generate any diffracted orders but is divided into forward and backward-propagating zeroth-order plane waves. The amplitudes of these two waves depend critically on the structure of the subwavelength-period grating and on the state of polarization of the incident plane wave. In fact, the effect of the subwavelength grating is somewhat similar to that of a multilayer thin film deposited on the surface.

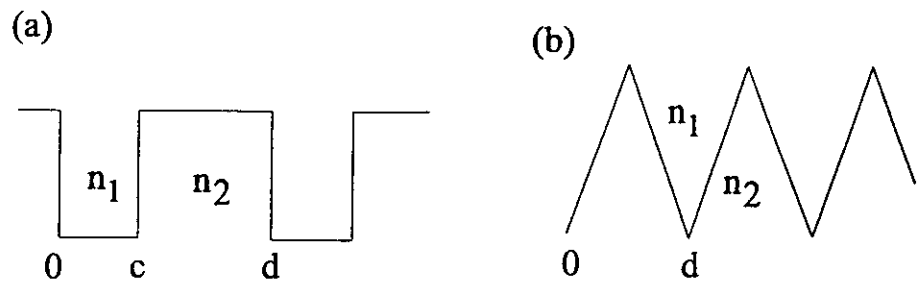

Figure 11: Subwavelength-period gratings. (a) Binary grating. (b) Triangular grating.

If $d \ll \lambda$, the properties of the subwavelength-period grating can be predicted rather accurately by means of the lowest-order theory of form birefringence (Born and Wolf, 1980, pp. 705–708). Rather elementary, yet ingenious considerations of Maxwell's equations lead to the result that in the geometry of Fig. 11(a) the effective refractive index within the grating is given by

$$n_{\parallel} = \left[n_1^2 c/d + n_2^2(1 - c/d)\right]^{1/2}, \tag{79}$$

whereas for TM polarization it is given by

$$n_{\perp} = \frac{n_1 n_2}{\left[n_2^2 c/d + n_1^2(1 - c/d)\right]^{1/2}}. \tag{80}$$

This means that the subwavelength-period grating will act like a uniaxial anisotropic crystal although the grating material is strictly isotropic.

The predictions of the zeroth-order theory of form birefringence may be refined by a number of techniques; ultimately, only the rigorous theory of gratings can give correct results. It is noteworthy, however, that the zeroth-order results hold quite well provided that $d < \lambda$ by only a small margin. This leads to several important applications that are discussed below.

Let us first consider the effective birefringence revealed by Eqs. (79) and (80). If we ignore boundary effects (which we can show to be a good

approximation by means of rigorous diffraction theory), this difference in TE and TM effective indices permits a straightforward generation of polarization components such as quarter-wave and half-wave plates by a suitable choice of the grating depth h (Cescato *et al.*, 1990).

Let us next consider a single state of polarization and notice that, in view of Eq. (79) or Eq. (80), we can achieve any desired effective refractive index by a control of the grating's aspect ratio c/d. Remembering from the theory of thin dielectric films the fact that a layer of thickness $\lambda/(4n)$ with refractive index $n = \sqrt{n_1 n_2}$ will yield zero reflectivity, we can identify an antireflection property of binary subwavelength-period gratings for a given state of linear polarization. This prediction is again verified, to a good approximation, by rigorous diffraction theory.

Wide-band antireflection coatings with a wide angular acceptance can be achieved by subwavelength-period gratings if we employ triangular (or nearly so) profiles instead of binary profiles (see Fig. 11b). To explain this phenomenon, we imagine that the grating be sliced into segments in the depth direction. The effective index of each segment is then determined by Eqs. (79) and (80), being close to n_1 and n_2 near the top and the bottom of the structure, respectively. Hence we achieve a gradual index transition from n_1 to n_2 through the effect of the microstructure. Such a tapered index modulation will lead to low reflectance across wide angular and wavelength bands (Raguin and Morris, 1993).

Equations (79) and (80) also indicate an interesting possibility to modulate the phase of the transmitted zeroth diffraction order by control of the local aspect ratio of a subwavelength-period grating. This technique was introduced independently by Stork *et al.* (1991) and Farn (1992). In principle, it permits the realization of any phase transformation in the paraxial domain (see Noponen and Turunen, 1994b, for an investigation of the limits of applicability evaluated by rigorous diffraction theory). The drawback of this approach is the high ratio of the profile depth and the smallest lateral feature size.

10 Fabrication methods

Thus far we have considered only the design of diffractive elements. However, the fabrication is an equally important issue.

Currently there are two major approaches for the fabrication of diffractive elements: optical interferometry (holography) and the microlithographic technology adopted from the manufacture of integrated electronic circuits. These schemes may also be combined such that an element fabricated by lithography is used, with spatial filtering, as an object for interferometric recording (Bartelt and Case, 1982). We concentrate on the lithographic scheme, mainly because the holographic method has been discussed in great detail elsewhere (Collier, Burckhardt, and Lin, 1971). We emphasize, however, that the holographic scheme remains the only feasible method in some instances, e.g., if the complexity of the signal is such that the diffractive profile can not be synthesized mathematically because of the high computational cost.

In the early stages of the development of the mathematical synthesis of diffractive elements the most common methods of realization were based on computer-controlled plotters capable of submillimeter resolution (hence the term computer-generated hologram) and photoreduction to provide a resolution of some 100 line pairs per millimeter. Therefore the elements were restricted strictly to the paraxial domain.

The commonly available laser printers are today specified to have a resolution of 300 or 600 dots per inch. Unfortunately, these resolution figures are of rather cosmetic nature from the viewpoint of diffractive optics, because they bear no relationship to the actual resolution of small features, measured in lines per millimeter; only the edges of larger features can in practice be defined at an accuracy approaching these figures.

The methods described above are capable of creating black-and-white (and perhaps grey-tone) transparencies. If used as diffractive elements, these have low efficiencies η. To improve η, we need to convert the amplitude

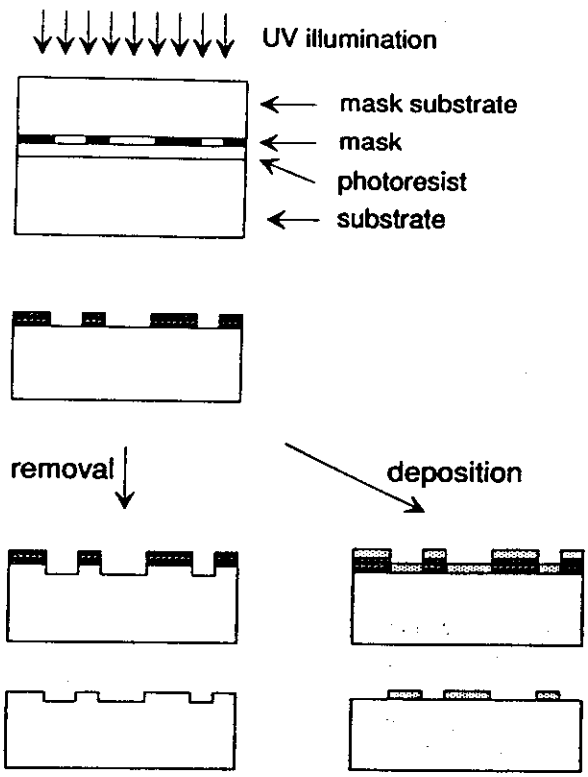

Figure 12: Material removal and material deposition techniques for fabricating binary surface-relief profiles.

transmittance into a phase transmittance, which is traditionally done by bleaching a silver-halide emulsion as is customary in holography. A nearly four-fold increase in η is easily achieved, but the problem of restricted resolution remains.

The great advances that have taken place in the manufacture of diffractive elements since the beginning of the era are largely due to the introduction of lithographic microfabrication technology into the generation of surface profiles. These techniques can be divided into two broad classes, i.e., material removal and material deposition, which are illustrated in Fig. 12.

The techniques illustrated in Fig. 12 are based on the use of binary amplitude masks with a structure that corresponds to the desired binary

surface-relief profile. Such masks can be fabricated by optical or electron-beam pattern generators that are in standard use in electronic circuit manufacturing technology. Once the mask is fabricated, it is placed in contact with a photoresist-coated substrate, or imaged onto it by a high-quality stepper lens. The resist is exposed by ultraviolet illumination and subsequently developed to reveal the substrate in the exposed regions.

In the material-removal scheme the selectively resist-coated substrate is placed in a vacuum chamber and etched, e.g., by means of ion bombardment. Once the required depth is achieved, the resist is removed and we have a binary surface-relief structure. In material deposition, thin-film manufacturing technology is used to grow a film of the required thickness onto the substrate. As long as the resist thickness is greater than that of the film, the resist removal process will also "lift off" those parts of the layer grown on top of the resist and we again have a binary structure.

The lithographic process can be repeated a desired number of times to create multilevel surface-relief profiles as illustrated in Fig. 13. The existing profile is resist-coated and a new mask with a different structure is introduced. In general, Q lithography steps will give a profile with 2^Q discrete levels. The main difficulty is the optical alignment of the masks between the lithography steps, in which an accuracy better than ~ 0.5 μm can only be achieved with difficulty.

The fabrication methods described above appear attractive, because they utilize lithographic technology known to be cost-effective in electronic circuit fabrication. There are also direct-write methods capable of generating continuous surface-relief profiles by controlled-dose laser beam exposure (Gale *et al.*, 1992) and electron-beam exposure (Nishihara and Suhara, 1987). Such direct-write techniques appear quite expensive at first sight. However, they can be used to generate a master element that may be copied in large quantities by methods such as embossing. For more details on diffractive-optics fabrication technology, see Herzig *et al.* (1993).

Thus far we have only discussed passive diffractive elements, i.e., ele-

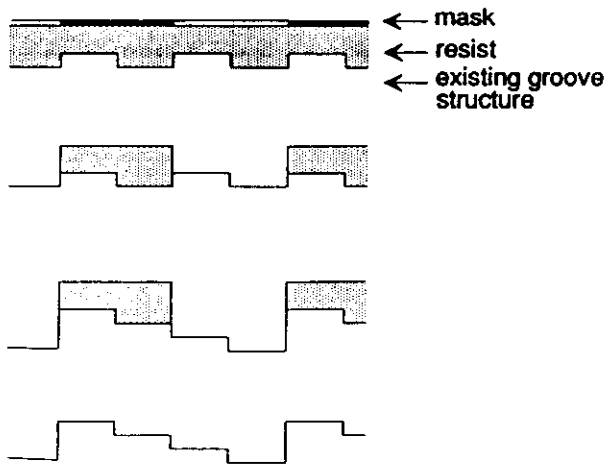

Figure 13: Fabrication of a four-level diffractive element.

ments with fixed signals. It would, or course, be of great interest to have the capability of modifying s in real time. This can be achieved by means of spatial light modulators, i.e., devices that modulate some property of the incident field (amplitude, phase, or state of polarization). In diffractive optics, phase modulation is of greatest interest, and it can be achieved, e.g., by liquid crystal devices (Amako and Sonehara, 1992). Unfortunately, amplitude modulation remains commercially more important because of display applications, and therefore phase-modulating liquid-crustal spatial light modulators are still at an experimental stage. Compared to fixed diffractive elements, the main drawback of these devices is the large pixel size ($\sim 50 \times 50\ \mu\text{m}^2$).

11 Integration of diffractive elements

The great technological impact of electronics in the latter half of this century is largely due to integration of many electronic components into a single device capable of performing many different functions. It is not inconceivable that optical components could be integrated in a similar fashion, and this

line of thought led to the concept of guided-wave integrated optics some 25 years ago (Miller, 1969). In guided-wave integrated optics, optical components control the propagation of guided waves that are confined in a thin high-index dielectric layer fabricated on top of a low-index substrate. It is straightforward to fabricate diffractive elements for waveguide optics: the effective refractive index is modulated, e.g., by means of a suitably shaped cover layer. Elements such as diffractive lenses can be realized in this manner (Nishihara and Suhara, 1987).

The greatest difference between "free-space" and guided-wave diffractive elements is that in the latter the refractive-index modulation is only $\Delta n \sim 0.01 - 0.1$, or even less. Hence volume diffraction effects can be quite significant and sophisticated design of the local grating profile is needed to design high-efficiency, high numerical aperture diffractive lenses (Huttunen, Turunen, and Saarinen, 1994).

Active diffractive elements for guided-wave optics can be realized most conveniently by launching a surface acoustic wave across the beam path by means of an interdigital transducer (Tsai, 1990). This permits deflection and scanning of guided waves, reconfigurable beam splitting, and various routing operations for optical switching. A review of the use of gratings in guided-wave optics is given by Suhara and Nishihara (1986).

The main disadvantage of guided-wave optics in integration of optical components is that it restricts optics to two dimensions. However, there are methods of integration that avoid this loss of the third dimension.

In stacked optics (Iga *et al.*, 1982) arrays of optical elements such as lenses are fabricated and the substrates are stacked together to form a solid block. Originally this scheme was introduced as an application of graded-index microlens arrays to perform operations such as fiber-array to fiber-array coupling (Fig. 14) but diffractive lenses could be used instead of gradient-index lenses and various other configurations and applications are possible.

If, instead of coupling light into a thin waveguide, we let it propagate in-

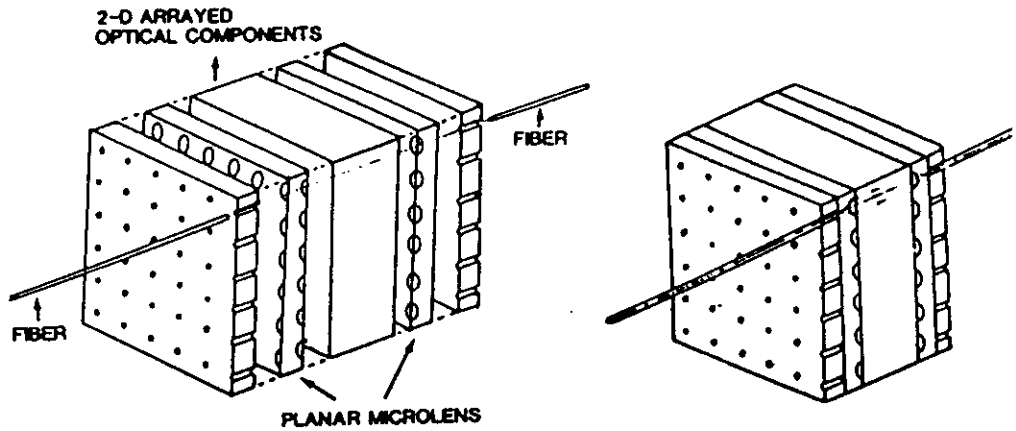

Figure 14: Coupling of light between two fiber arrays by means of stacked optics (Iga *et al.*, 1982).

side a substrate of thickness ~ 1 mm as illustrated in Fig. 15 and control the propagation by elements fabricated on top of the substrate, we may speak of substrate-mode optics (Brenner and Sauer, 1988; Kostuk *et al.*, 1989) or planar-integrated optics (Jahns and Huang, 1989). This relatively new but very promising approach has been used, e.g., to demonstrate an integrated optical array illuminator (Downs and Jahns, 1990), a split-and-shift module (Jahns and Brumback, 1990), and integrated optical interconnection modules (Sauer, 1989; Walker *et al.*, 1993).

12 Conclusions

In this brief tutorial we have been able cover only some of the issues important in diffractive optics, and have completely ignored most of its applications. However, it is hoped that the great scientific, technical, and commercial potential of diffractive optics has become clear: since diffractive optics is a significant extension of classical refractive and reflective optics, its role in modern technology can only expand in the future.

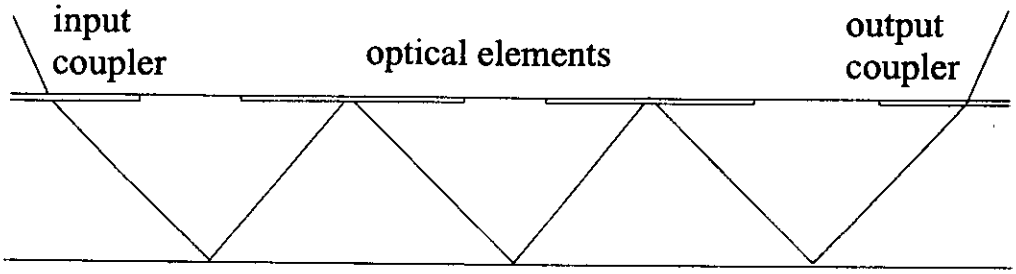

Figure 15: Principle of planar-integrated optics: light propagates inside a substrate and is controlled by diffractive elements fabricated on the surface.

Acknowledgments

I have benefited greatly from collaboration with many colleagues over the past ten years. In particular, I would like to thank F. Wyrowski for numerous stimulating discussions. Furthermore, I would like to express my gratitude to the Emil Aaltonen Foundation for a grant that has substantially promoted my writing activities.

References

Amako, J., and Sonehara, T., 1991, Kinoform using an electrically controlled birefringent spatial light modulator, Appl. Opt. **30**, 4622–4628.

Bartelt, H., and Case, S. K., 1982, High-efficiency hybrid computer-generated holograms, Appl. Opt. **21**, 2886–2890.

Beckmann, P. and Spizzichino, A., 1963, *The Scattering of Electromagnetic Waves from Rough Surfaces*, Oxford: Pergamon.

Born, M., and Wolf, E., 1980, *Principles of Optics* 6th Ed., Pergamon: Oxfrd,

Bräuer, R., Wyrowski, F., and Bryngdahl, O., 1991, Diffusers in digital holography, J. Opt. Soc. Am. A **8**, 572–578.

Brenner, K.-H., and Sauer, F., 1988, Diffractive-reflective optical interconnects, Appl. Opt. **27**, 4251–4254.

Brown, B. R., and Lohmann, A. W., 1966, Complex spatial filtering with binary masks, Appl. Opt. **5**, 967–969.

Bryngdahl, O. and Wyrowski, F., 1990, Digital holography — Computer-generated holograms, in E. Wolf (Ed.) *Progress in Optics* XXVIII, pp. 1–86, Amsterdam: North-Holland.

Cederquist, J. N., and Fienup, J. R., 1987, Analytic design of optimum holographic optical elements, J. Opt. Soc. Am. A **4**, 699–705.

Cescato, L. H., Gluch, E., and Streibl, N., 1990, Holographic quarterwave plates, Appl. Opt. **29**, 3286–3290.

Chu, D. C., Fienup, J. R., and Goodman, J. W., 1973, Multi-emulsion, on-axis, computer-generated holograms, Appl. Opt. **12**, 1386–1388.

Collier, R. J., Burckhardt, C. B., and Lin, L. H., 1971, *Optical Holography*, Academic Press: New York.

Dammann, H., and Görtler, K., 1971, High-efficiency in-line multiple imaging by means of multiple phase holograms, Opt. Commun. **3**, 312–315.

Dammann, H., and Klotz, E., 1978, Coherent optical generation and inspection of two-dimensional periodic structures, Opt. Acta **24**, 505–515.

Downs, M. M., and Jahns J., 1990, Integrated-optical array generator, Opt. Lett. **15**, 769–770.

Farn, M. W., 1992, Binary gratings with increased efficiency, Appl. Opt. **31**, 4453–4458.

Gale, M. T., Lang, G. K., Raynor, J. M., Shütz, H., and Prongué, D., 1992, Fabrication of kinoform structures for optical computing, Appl. Opt. **31,** 5712–5715.

Gaylord, T. K. and Moharam, M. G., 1985, Analysis and applications of optical diffraction by gratings, *Proc. IEEE* **73,** 894–937.

Goodman, J. W., 1968, *Introduction to Fourier Optics* (MacGraw-Hill, San Francisco, 1968).

Herzig, H. P., Gale, M. T., Lehmann, H. W., and Morf, R., 1993, Diffractive components: computer-generated elements, in *Perspectives for Parallel Optical Interconnects*, Ph. Lalanne and P. Chavel, Eds., Springer: Berlin.

Huttunen, J., Turunen, J., and Saarinen, J., 1994, High-efficiency diffractive lenses by parametric optimization, Appl. Opt. **33,** 1715–1725.

Iga, K., Oikawa, M., Misawa, S., Banno, J., and Kokobun, Y., 1982, Stacked planar optics: an application of the planar microlens, Appl. Opt. **25,** 3388–2296.

Jahns, J., and Brumback, B. A., 1990, Integrated-optical split-and-shift module based on planar optics, Opt. Commun. **76,** 318–320.

Jahns, J., and Huang, A., 1989, Planar integration of free-space optical components, Appl. Opt. **28,** 1602–1605.

Korpel, A., 1988, *Acousto-Optics*, New York: Marcel Dekker.

Lohmann, A. W. and Paris, D. P., 1967, Binary Fraunhofer holograms, generated by computer, *Appl. Opt.* **6,** 1739–48.

Maystre, D., 1984, Rigorous vector theories of diffraction gratings, in E. Wolf (Ed.), *Progress in Optics*, Vol. XXI, pp. 1–67, Amsterdam: North-Holland.

Miller, S. E., 1969, Integrated optics: an introduction, Bell Syst. Techn. J. **48,** 2059–2069.

Miller, J. M., Taghizadeh, M. R., Turunen, J., and Ross, N., Multi-level-

grating array generators: fabrication analysis and experiments, Appl. Opt. **32,** 2519–2525.

Nishihara, H. and Suhara, T., 1987, Micro Fresnel lenses, in E. Wolf (Ed.), *Progress in Optics*, Vol. XXIV, pp. 1–40, Amsterdam: North-Holland.

Noponen, E., and Turunen, J., 1994a, Eigenmode method for electromagnetic synthesis of diffractive elements with three-dimensional profiles, J. Opt. Soc. Am. A, **11,** 2494–2502.

Noponen, E., and Turunen, J., 1994b, Binary high-frequency-carrier diffractive optical elements: electromagnetic theory, J. Opt. Soc. Am. A **11,** 1097–1109.

Petit, R. (Ed.), 1980, *Electromagnetic Theory of Gratings*, Berlin: Springer-Verlag.

Raguin, D. H., and Morris, G. M., 1993, Analysis of antireflection-structured surfaces with continuous one-dimensional surface profiles, Appl. Opt. **32,** 2582–2598.

Sauer, F., 1989, Fabrication of diffractive-refractive optical interconnects for infrared operation based on total internal reflection, Appl. Opt. **28,** 386–388.

Solymar, L. and Cooke, D. J., 1981, *Volume Holography and Volume Gratings*, London: Academic Press.

Streibl, N., 1989, Beam shaping with optical array generators, J. Mod. Opt. **36,** 1559–1573.

Stork, W., Streibl, N., Haidner, H., and Kipfer, P., 1991, Artificial distributed-index media fabricated by zero-order gratings, Opt. Lett. **16,** 1921–1923.

Suhara, T., and Nishihara, H., 1986, Integrated optics components and devices using periodic structures, IEEE J. Quant. Electr. **QE–22,** 845–867.

Tsai, C. S., Ed., 1990, *Guided-wave Acousto-optics. Interactions, Devices, and Applications*, Prentice-Hall: Englewood Cliffs.

Vasara, A., Taghizadeh, M. R., Turunen, J., Westerholm, J., Noponen, E., Ichikawa, H., Miller, J. M., Jaakkola, T., and Kuisma, S., 1992, Binary surface-relied gratings for array illumination in digital optics, Appl. Opt. **31**, 3320–3336.

Winick, K. A., and Fienup, J. R., 1983, Optimum holographic elements recorded with non-spherical wave fronts, J. Opt. Soc. Am. **73**, 208–217.

Wyrowski, F., 1989, Iterative quantization of digital amplitude holograms, Appl. Opt. **28**, 3864–3870.

Wyrowski, F., 1991, Upper bound of the diffraction efficiency of diffractive phase elements, Opt. Lett. **16**, 1915–1917.

Wyrowski, F., 1993, Design theory of diffractive phase elements in the paraxial domain, J. Opt. Soc. Am. A **10**, 1553–1561.

Wyrowski, F., and Bryngdahl, O., 1989, Speckle-free reconstruction in digital holography, J. Opt. Soc. Am. A **6**, 1171–1174.

Wyrowski, F. and Bryngdahl, O., 1991, Digital holography as part of diffractive optics, *Rep. Progr. Phys.* **54**, 1481–1571.