

# Comparison of different theories for focusing through a plane interface

S. H. Wiersma

*Department of Physics and Astronomy, Free University, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands*

P. Török

*Multi-Imaging Centre, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK*

T. D. Visser

*Department of Physics and Astronomy, Free University, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands*

P. Varga

*Research Institute for Materials Science, Budapest P.O. Box 49, H-1525, Hungary*

Received August 8, 1996; accepted January 6, 1997

We consider the problem of light focusing by a high-aperture lens through a plane interface between two media with different refractive indices. We compare two recently published diffraction theories and a new geometrical optics description. The two diffraction approaches exhibit axial distributions with little difference. The description based on geometrical optics is shown to agree well with the diffraction optics results. Also, some implications for three-dimensional imaging are discussed. © 1997 Optical Society of America [S0740-3232(97)02207-2]

## 1. INTRODUCTION

The effect of a dielectric interface on the electromagnetic field has been studied by several workers.<sup>1-4</sup> These theories are usually approximations of some rigorous solutions;<sup>2</sup> however, exact solutions of either Maxwell's equations or the wave equation have also been obtained.<sup>1,4</sup>

The subject of focusing of electromagnetic waves by a high-numerical-aperture lens into a homogeneous medium was described by Wolf<sup>5</sup> and Richards and Wolf.<sup>6</sup> Their work may be regarded as the vectorial generalization of the Debye diffraction formula.<sup>7</sup> The focused electromagnetic field is given as a superposition of plane waves, whose propagation vectors all fall inside the geometrical light cone. Although in their paper Richards and Wolf<sup>6</sup> regarded their solution as an approximation of a physical problem, Luneburg showed<sup>8</sup> that it is an exact solution of an abstract mathematical problem. In a recently published paper, Török *et al.*<sup>9</sup> gave a rigorous solution for the problem of focusing through a plane interface, which satisfies both Maxwell's equations and the homogeneous wave equation. This work may be considered as the extension of the Richards–Wolf theory to the case of focusing into an inhomogeneous medium. Another recently published study by Wiersma and Visser<sup>10</sup> also took the Richards–Wolf theory as a starting point to describe the effect of a plane dielectric interface behind the lens.

Török *et al.*<sup>9</sup> obtained the electric and magnetic vectors in the second medium by means of a matrix formalism and then applied a coherent superposition of plane waves

to obtain the diffraction pattern. It was shown independently by Wiersma and Visser<sup>10</sup> that it is also possible to obtain the field in the second medium by means of another vectorial diffraction theory. This so-called *m* theory<sup>11</sup> was introduced by several workers. Smythe<sup>12</sup> and Toraldo di Francia<sup>13</sup> both used it to describe diffraction by an aperture in a perfectly conducting screen. (The latter treatment is probably the clearest.) Severin<sup>14</sup> generalized the same formalism to dielectrics. His approach uses the idea that the field in a half-space is completely determined by the tangential component of either the electric or the magnetic field on the plane that bounds the half-space. It was his insight that this plane can be any mathematical plane, which need not coincide with a physical screen. The *m*-theory solutions satisfy Maxwell's equations, and the boundary values are reproduced when the observation point, where the field is calculated, is chosen on the plane.

The aim of the present study is to compare the two recently developed theories (Refs. 9 and 10), since they use two completely different methods to describe the effect of a plane interface on a converging spherical wave. Also, a geometrical analysis of this problem is presented, which provides a first approximation of the intensity distribution.

The organization of this paper is as follows. In Section 2 we derive the electric-field vector in the second medium. Next, following Refs. 9 and 10, we briefly show in Section 3 how the plane-wave and *m*-theory solutions are obtained. This section is concluded with a comparison of

the vectorial  $m$  theory and the scalar first Rayleigh–Sommerfeld integral. In Section 4 a geometrical optics approximation is presented that is capable of predicting some important features of the intensity distribution in the second medium. Numerical results obtained for the two diffraction theories and the geometrical optics approximation are compared for several examples in Section 5. In Section 6 some implications of the two theories for three-dimensional (3-D) imaging are discussed. We also present an appendix with a derivation of the  $m$  theory, as the original references are less accessible.

## 2. ELECTRIC VECTOR IN THE SECOND MEDIUM

The geometry of our problem is depicted in Fig. 1. It was shown in Ref. 9 that the electric vector amplitude in the second medium can be derived by successive application of certain coordinate transformations. These transformations handle the  $s$ - and  $p$ -polarized components separately. After introduction of the usual spherical polar coordinate system, with  $\phi$  denoting the azimuthal angle and  $\theta_j$  denoting the polar angle in the first ( $j = 1$ ) and the second ( $j = 2$ ) medium, the electric field in the second medium can be written as

$$\mathbf{E}_2 = \mathbf{R}^{(-1)}[\mathbf{P}^{(2)}]^{-1}\mathbf{IP}^{(1)}\mathbf{LRE}^{(0)}. \quad (1)$$

Here  $\mathbf{E}^{(0)} = (E_0, 0, 0)$  is the incident electric vector amplitude in front of the lens, which is taken as  $x$  polarized;

$$\mathbf{R} = \begin{bmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (2)$$

which describes a rotation of the coordinate system around the optical axis;

$$\mathbf{L} = \begin{bmatrix} \cos \theta_1 & 0 & \sin \theta_1 \\ 0 & 1 & 0 \\ -\sin \theta_1 & 0 & \cos \theta_1 \end{bmatrix}, \quad (3)$$

which describes the effect of the lens on ray propagation;

$$\mathbf{P}^{(n)} = \begin{bmatrix} \cos \theta_n & 0 & -\sin \theta_n \\ 0 & 1 & 0 \\ \sin \theta_n & 0 & \cos \theta_n \end{bmatrix}, \quad (4)$$

which describes a rotation of the coordinate system around one of the lateral directions; and

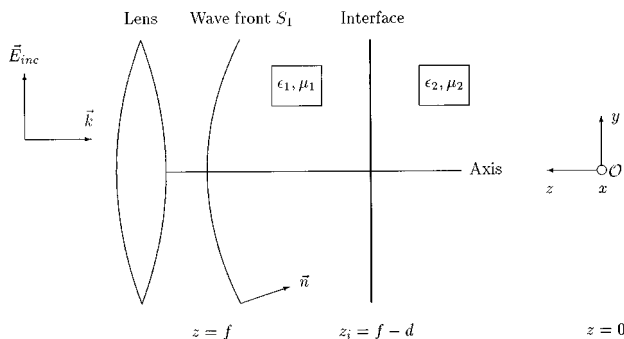


Fig. 1. Geometry of the system.

$$\mathbf{I} = \begin{bmatrix} \tau_p & 0 & 0 \\ 0 & \tau_s & 0 \\ 0 & 0 & \tau_p \end{bmatrix}, \quad (5)$$

which describes the effect of the plane dielectric interface, with  $\tau_p$  and  $\tau_s$  as the Fresnel transmission coefficients. From Eq. (1) it follows that the electric field immediately to the right of the interface is given by

$$\mathbf{E}^{(2)} = \sqrt{\cos \theta_1} \begin{pmatrix} \tau_p \cos \theta_2 \cos^2 \phi + \tau_s \sin^2 \phi \\ \tau_p \cos \theta_2 \sin \phi \cos \phi - \tau_s \sin \phi \cos \phi \\ -\tau_p \sin \theta_2 \cos \phi \end{pmatrix}. \quad (6)$$

Note that Eq. (6) can also be obtained from vectorial considerations [Ref. 10, Eq. (21)]. For the special case  $\epsilon_1 = \epsilon_2$ , Eq. (6) reduces to the expression for the electric field given by Richards and Wolf<sup>6</sup> for a single homogeneous medium.

## 3. DIFFRACTION OPTICS SOLUTIONS

### A. Plane-Wave Solution

The basis of this solution is that the electromagnetic field just before the interface can be expressed as a superposition integral that sums up all possible plane waves propagating within the divergence angle of the high-aperture lens. Each plane wave is transmitted through the interface. Then we write a similar expression for the field in the second medium, just after the interface. These two expressions must give the same field at the interface, or, in other words, the first integral is used as a boundary condition for the second integral. In the first medium and at the interface  $z_i = \lim_{\delta \rightarrow 0} f - d + \delta$  (see Fig. 1), the incident electric field in an angular spectrum representation is given by<sup>6</sup>

$$\begin{aligned} \mathbf{E}_1(x, y, z_i) = & -\frac{ik_1}{2\pi} \iint_{\Omega_1} \frac{\mathbf{a}(s_{1x}, s_{1y})}{s_{1z}} \\ & \times \exp[ik_1(s_{1x}x + s_{1y}y \\ & + s_{1z}z_i)] ds_{1x} ds_{1y}. \end{aligned} \quad (7)$$

The transmitted field in the second material, at the close vicinity ( $z_i = \lim_{\delta \rightarrow 0} f - d - \delta$ ) of the interface, is given by

$$\begin{aligned} \mathbf{E}_2(x, y, z_i) = & -\frac{ik_1}{2\pi} \iint_{\Omega_1} \mathbf{M} \frac{\mathbf{a}(s_{1x}, s_{1y})}{s_{1z}} \\ & \times \exp[ik_1(s_{1x}x + s_{1y}y \\ & + s_{1z}z_i)] ds_{1x} ds_{1y}, \end{aligned} \quad (8)$$

where  $\mathbf{M}$  is an operator describing the transition of the strength vector  $\mathbf{a}$  through the interface,  $k_j$  is the wave number,  $\hat{s}_j = (s_{jx}, s_{jy}, s_{jz})$  is the unit vector along a typical ray in the first ( $j = 1$ ) and the second ( $j = 2$ ) medium, and  $\Omega_1$  is the semiaperture angle of the lens. We represent the field inside the second medium again as a superposition of plane waves. This representation is a solution of the time-independent wave equation and Maxwell's equations and can be written as

$$\mathbf{E}_2(x, y, z) = -\frac{ik_2}{2\pi} \iint_{\Omega_2} \mathbf{F}(\hat{s}_2) \times \exp[ik_2(s_{2x}x + s_{2y}y - s_{2z}z)] ds_{2x} ds_{2y}. \quad (9)$$

Here  $\mathbf{F}(\hat{s}_2)$  is a function determined by the boundary condition (8). By expanding Eq. (9) and using spherical polar coordinates, we find that the axial distribution of the linearly polarized electric field in the second medium is given by

$$E_{2x}(z) = \frac{ik_1 f l_0}{2} \int_0^{\Omega_1} \sqrt{\cos \theta_1} \sin \theta_1 \times \exp[i(f-d)(k_1 \cos \theta_1 - k_2 \cos \theta_2)] \times (\tau_s + \tau_p \cos \theta_2) \exp(-ik_2 z \cos \theta_2) d\theta_1, \quad (10)$$

where  $f$  is the focal length of the lens and  $l_0$  is an amplitude factor. It is emphasized that when off-axis points are computed, the expression for  $\mathbf{E}_2 = (E_{2x}, E_{2y}, E_{2z})$  consists of a linear combination of three integral functions, each containing only a single integral. This makes the numerical evaluation much faster, compared with the case in which multiple integrals are to be evaluated.

It is worthwhile analyzing Eq. (10). The function  $\exp[i(f-d)(k_1 \cos \theta_1 - k_2 \cos \theta_2)]$  introduces a phase term in the integral and thus represents an aberration. The amplitude factor  $(\tau_s + \tau_p \cos \theta_2)$  may be regarded as a polarization-dependent apodization function. The term  $\sqrt{\cos \theta_1}$  is introduced because the lens is assumed to obey the sine condition, and, finally,  $\exp(-ik_2 z \cos \theta_2)$  is the well-known defocus phase factor. It is important to point out that the integration is carried out with the parameters of the first medium, but the integrand is a mixture of quantities of the first and second media, which also means that when irregular (evanescent) waves are computed,<sup>15</sup> it is not necessary to introduce complex contour integration.

## B. *m*-Theory Solution

As derived in Appendix A, the *m* theory requires knowledge of only the tangential component of the electric field on the integration surface. The electric field within the integration volume, i.e., to the right of the interface in Fig. 1, is then completely determined. This is in agreement with the uniqueness theorem.<sup>16</sup> As the integration surface  $\Sigma$ , we take the plane immediately to the right of the interface, described by  $z = \lim_{\delta \rightarrow 0} f - d - \delta$ . In principle,  $\Sigma$  extends to infinity. In the following the integration area is limited to the intersection of the geometrical light cone with the interface. So, for a high-Fresnel-number lens obeying the sine condition, this approximation is reasonable as long as the interface does not lie too close to the geometrical focus of the lens at  $z = 0$ . The field that is incident on the interface is taken as the field in the absence of the second medium. The electric field on the integration surface just after the interface is then given by Eq. (6).

We recall the expression for the *m*-theory integral, Eq. (A17) of Appendix A:

$$\mathbf{E}(\mathbf{Q}) = 2 \int_{\Sigma} [\mathbf{m} \times \mathbf{E}(\mathbf{P})] \times \nabla G(\mathbf{P}, \mathbf{Q}) d\Sigma, \quad (11)$$

with  $\mathbf{E}(\mathbf{P})$  given by Eq. (6). The normal  $\mathbf{m}$  on the interface  $\Sigma$  points in the positive  $z$  direction (see Fig. 1). The Green's function  $G$  is defined as

$$G(\mathbf{P}, \mathbf{Q}) = \frac{\exp(ik_2 s)}{4\pi s}; \quad (12)$$

thus

$$\nabla G = \left( \frac{1}{s} - ik_2 \right) G \frac{\mathbf{Q} - \mathbf{P}}{|\mathbf{Q} - \mathbf{P}|}, \quad (13)$$

where  $\mathbf{P}$  and  $\mathbf{Q}$  denote the position vectors of the points  $P$  and  $Q$ , respectively. Also,  $s = |\mathbf{P} - \mathbf{Q}|$ , so

$$s = [t^2(\theta_1) + z^2 - 2z(f-d)]^{1/2}, \quad (14)$$

with

$$t(\theta_1) = (f-d)/\cos \theta_1. \quad (15)$$

Hence  $f - t(\theta_1)$  is the path that a ray, traveling at angle  $\theta_1$  from the initial spherical wave front  $S$  to the interface, traverses. A phase factor  $F(\theta_1)$  and an amplitude factor  $A(\theta_1)$  account for the phase and amplitude changes that a ray undergoes along this path:

$$F(\theta_1) = \exp\{ik_1[f - t(\theta_1)]\}, \quad (16)$$

$$A(\theta_1) = \frac{f \cos^{3/2} \theta_1}{f - d}. \quad (17)$$

From these quantities, following Wiersma and Visser,<sup>10</sup> the electric field along the optical axis is given by

$$E_{2x}(z) = \frac{f(f-d)}{2} [z - (f-d)] \int_0^{\Omega_1} (\tau_s + \tau_p \cos \theta_2) \times \exp\{i[k_2 s + k_1(f-t)]\} \times \left( \frac{1}{s^3} - \frac{ik_2}{s^2} \right) \tan \theta_1 d\theta_1. \quad (18)$$

It is interesting to compare the axial distributions given by the *m* theory and the first Rayleigh–Sommerfeld diffraction integral<sup>17</sup> for the general case of diffraction of a wave by a circular aperture  $\mathcal{A}$ . Assume that the aperture is centered on the  $z$  axis at  $z = f - d$ . The first Rayleigh–Sommerfeld integral for a vector field  $\mathbf{E}$  reads as

$$\mathbf{E}_{RS_1}(x, y, z) = -\frac{1}{2\pi} \int_{\mathcal{A}} \mathbf{E}_0(x_P, y_P, f-d) \times \frac{\partial}{\partial z} \left[ \frac{\exp(ik_2 R)}{R} \right] d\mathcal{A}, \quad (19)$$

with  $x_P$  and  $y_P$  in the aperture  $\mathcal{A}$ , and

$$R = [(x - x_P)^2 + (y - y_P)^2 + (z - f + d)^2]^{1/2}. \quad (20)$$

As above, the field in the aperture,  $\mathbf{E}_0$ , is taken as the incident field. We have

$$\frac{\partial}{\partial z} \left[ \frac{\exp(ik_2 R)}{R} \right] = (z - f + d) \left( \frac{ik_2}{R^2} - \frac{1}{R^3} \right) \exp(ik_2 R), \quad (21)$$

from which

$$\begin{aligned} \mathbf{E}_{RS_1}(x, y, z) = & -\frac{z - f + d}{2\pi} \int_{\mathcal{A}} \begin{pmatrix} \mathbf{E}_x(P) \\ \mathbf{E}_y(P) \\ \mathbf{E}_z(P) \end{pmatrix} \\ & \times \left( \frac{ik_2}{R^2} - \frac{1}{R^3} \right) \exp(ik_2 R) d\mathcal{A}. \end{aligned} \quad (22)$$

In order to compare this with the corresponding result from the  $m$  theory [Eq. (11)], we express  $(\mathbf{m} \times \mathbf{E}) \times \nabla G$  with the help of Eqs. (13) and (14) in terms of

$$\begin{aligned} \mathbf{P} - \mathbf{Q} = & (x_P - x, y_P - y, f - d - z), \\ s = R = |\mathbf{P} - \mathbf{Q}| = & [(x_P - x)^2 + (y_P - y)^2 \\ & + (z - f + d)^2]^{1/2} \end{aligned}$$

and obtain

$$\begin{aligned} (\mathbf{m} \times \mathbf{E}) \times \nabla G = & -\frac{1}{4\pi} \left( \frac{1}{R^3} - \frac{ik_2}{R^2} \right) \exp(ik_2 R) \\ & \times \begin{pmatrix} (z - f + d)E_x(P) \\ (z - f + d)E_y(P) \\ (y - y_P)E_y(P) + (x - x_P)E_x(P) \end{pmatrix}. \end{aligned} \quad (23)$$

If we confine the calculation to the  $z$  axis ( $x = y = 0$ ), then the axial field according to the  $m$  theory is given by

$$\begin{aligned} \mathbf{E}_{m \text{ theory}}(0, 0, z) = & -\frac{z - f + d}{2\pi} \\ & \times \int_{\mathcal{A}} \begin{pmatrix} E_x(P) \\ E_y(P) \\ [y_P E_y(P) + x_P E_x(P)] / (z - f + d) \end{pmatrix} \\ & \times \left( \frac{ik_2}{R^2} - \frac{1}{R^3} \right) \exp(ik_2 R) d\mathcal{A}. \end{aligned} \quad (24)$$

From a comparison of Eqs. (22) and (24), it is seen that the first Rayleigh–Sommerfeld integral and the  $m$  theory predict the same axial field distribution, provided that in Eq. (22) the boundary condition for  $(E_x, E_y, E_z)$  in the aperture  $\mathcal{A}$  can be correctly specified, i.e., is taken as the column vector in the integrand of Eq. (24). This is an important finding, as the direct application of known diffraction theories is not possible because of the discontinuity at the interface. When, however, the field in the second medium, just after the interface, is specified on the plane of integration,  $\mathcal{A}$ , any diffraction theory can, in principle, be applied.

#### 4. GEOMETRICAL OPTICS APPROXIMATION

In the previous sections the focusing of converging electromagnetic waves through an interface has been studied

by using two different diffraction theories. In this section we show that a much simpler geometrical approach gives a surprisingly good approximation to the intensity distribution.

In contrast to diffraction theory, geometrical optics predicts that the intensity distribution is confined to a finite part of the  $z$  axis (see Fig. 2). Let  $\rho$  denote the distance from the  $z$  axis at which a ray incident under an angle  $\theta_1$  crosses the interface. We then have

$$\tan \theta_1 = \frac{\rho}{f - d}. \quad (25)$$

If the refracted ray makes an angle  $\theta_2 = \arcsin[n_1(\sin \theta_1)/n_2]$  with the  $z$  axis, then

$$\tan \theta_2 = \rho/h(\theta_1). \quad (26)$$

Eliminating  $\rho$  gives

$$\begin{aligned} h(\theta_1) = (f - d) \frac{\tan \theta_1}{\tan \theta_2} = (f - d) \frac{n_2 \cos \theta_2}{n_1 \cos \theta_1}, \\ 0 < \theta_1 \leq \Omega_1. \end{aligned} \quad (27)$$

We note that this expression does not hold for the ray incident at  $\theta_1 = 0$ . However, since this ray corresponds to an infinitely small area of the incident beam, its contribution to the intensity distribution will be negligible. Equation (27) defines two shadow boundaries on the  $z$  axis between which the intensity is concentrated. These boundaries, a marginal one,  $z_m$ , and a paraxial one,  $z_p$ , are at

$$z_m = f - d - h(\Omega_1), \quad (28)$$

$$z_p = f - d - \lim_{\theta_1 \rightarrow 0} h(\theta_1) = (f - d)(1 - n_2/n_1). \quad (29)$$

The above derivation is valid only if no total internal reflection takes place, and we may hence use Snell's law.

Every point on the optical axis within the shadow boundaries corresponds to a value of  $\theta_1$ . In order to determine this inverse function, we square Eq. (27) to obtain

$$\begin{aligned} h^2(1 - \sin^2 \theta_1) = (f - d)^2 (n_2/n_1)^2 \\ \times [1 - (n_1/n_2)^2 \sin^2 \theta_1]. \end{aligned} \quad (30)$$

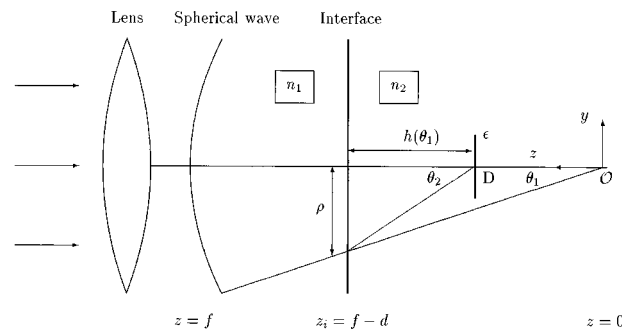


Fig. 2. Ray tracing for focusing through an interface. A lens with focal length  $f$  and semiaperture angle  $\Omega_1$  is placed at a distance  $d$  in front of an interface.

Hence

$$\sin^2[\theta_1(h)] = [(f-d)^2(n_2/n_1)^2 - h^2]/[(f-d)^2 - h^2]. \quad (31)$$

Using the relation  $h = f - d - z$  (see Fig. 2), we finally find that

$$\theta_1 = \arcsin\left(\left\{\frac{(f-d)^2[(n_2/n_1)^2 - 1]}{2z(f-d) - z^2} + 1\right\}^{1/2}\right), \quad n_1 \neq n_2. \quad (32)$$

In order to calculate the intensity on the  $z$  axis, we make use of the geometrical law of intensities.<sup>18</sup> It should be realized that intensity within the framework of a vectorial diffraction theory is a scalar quantity, whereas in geometrical optics it is an energy flux through a surface.

Consider a disk-shaped detector  $D$ , with a radius  $\epsilon$ , placed perpendicular to the optical axis at a distance  $h$  from the interface (see Fig. 2). After we let the diameter of the detector become very small, the flux through its surface as a function of position can be compared with the axial intensity distribution that is predicted by the two vectorial theories.

Let a refracted ray traveling at an angle  $\theta_2$  hit the center of the detector. Also, the rays that intersect the optical axis at a distance  $\Delta h$  on either side of the detector are intercepted. In approximation we have, for small detectors,

$$\Delta h \approx \epsilon/\tan \theta_2. \quad (33)$$

All these rays lie in the  $\sin \theta_1$  interval in the first medium between  $\sin[\theta_1(h + \Delta h)]$  and  $\sin[\theta_1(h - \Delta h)]$ , where  $\sin[\theta_1(h)]$  is given by Eq. (31). As the lens is assumed to obey the sine condition

$$r = f \sin \theta_1, \quad (34)$$

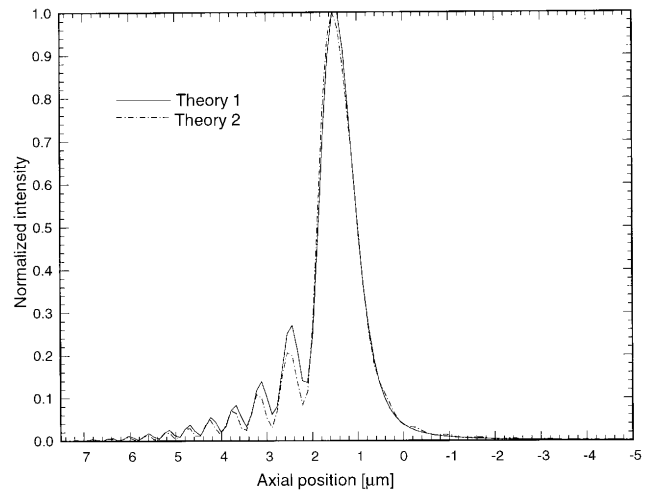
where  $r = \sqrt{x^2 + y^2}$  denotes the distance from the axis at which an incident ray enters the lens, all rays between  $r_I = f \sin[\theta_1(h + \Delta h)]$  and  $r_{II} = f \sin[\theta_1(h - \Delta h)]$  will be detected. If the incident beam has a homogeneous intensity distribution, then this implies that the relative total intensity at the detector plane equals

$$I(h; \epsilon) = \pi|r_I^2 - r_{II}^2| \cos[\theta_2(h)]. \quad (35)$$

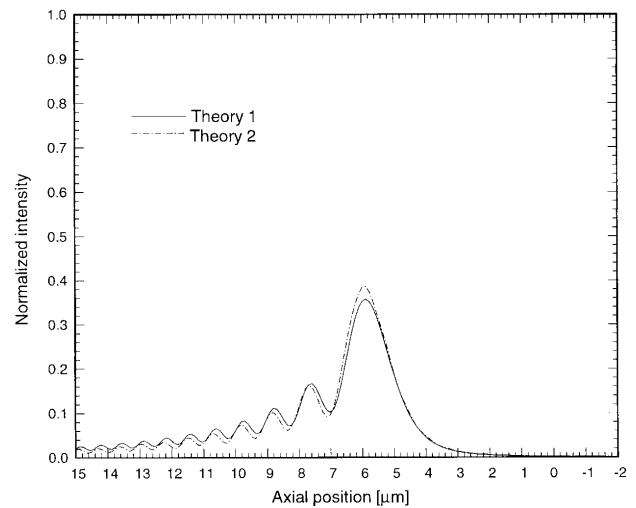
The factor  $\cos \theta_2$  describes the usual flux dependence on the direction of propagation and the orientation of the detector surface.

## 5. NUMERICAL RESULTS

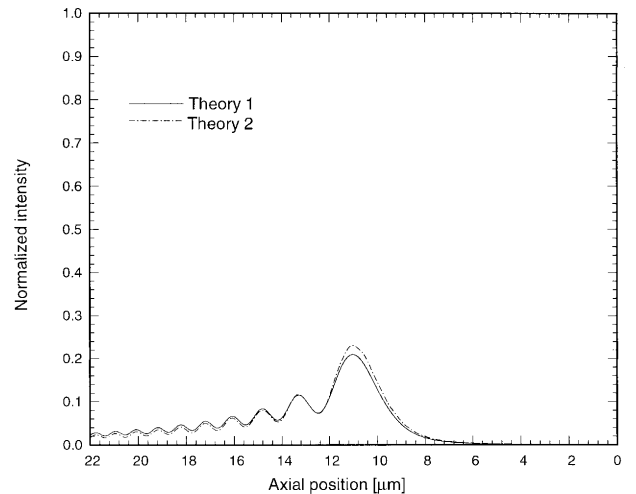
The distribution of the intensity along the optical axis was computed from Eqs. (10) (for theory 1, the plane-wave solution) and (18) (for theory 2, the  $m$  theory). A program to evaluate these expressions was written in FORTRAN by using the Numerical Algorithm Group subroutine package. The numerical computations were performed on an IBM 486DX4 computer. Results were directly visualized by using the TECPLOT program. We present results for a lens with NA = 1.4 and a glass/water interface ( $n_1 = 1.54$ ,  $n_2 = 1.33$ ). The focusing depths ( $f - d$ ) were 10, 50, and 100  $\mu\text{m}$ . The wavelength in vacuum was  $\lambda_0 = 488 \text{ nm}$ . The numerical re-



(a)



(b)



(c)

Fig. 3. Comparison of theories 1 and 2. The lens numerical aperture was NA=1.4, and computations were performed for a glass/water interface ( $n_1 = 1.54$ ,  $n_2 = 1.33$ ) with focusing depths  $f - d$  of (a) 10, (b) 50, and (c) 100  $\mu\text{m}$ . Individual images are normalized to the intensity obtained for the 10- $\mu\text{m}$  focusing depth.

sults are shown in Fig. 3, where individual images are normalized to the intensity obtained for  $f - d = 10 \mu\text{m}$ . As the figure shows, the two theories predict the axial location of the focus (main peak) with excellent agreement. The decrease in peak intensity throughout the computed range agrees well; somewhat greater differences (<9%) are found at greater focusing depths. The axial location of the sidelobes is the same according to the two theories. It is interesting to note that theory 1 gives less lobe structure on the negative side of the distributions, but initially it predicts higher lobes on the positive side. The agreement between the two theories is better at smaller focusing depths for the main peak, but as the focusing depth increases, the agreement becomes better for the sidelobe structure and worse for the peak intensity. This stems from the different approximations made in theories 1 and 2, as mentioned above.

Results of theory 2 are plotted in Fig. 4 for a lens with  $\text{NA} = 1.32$  and a wavelength of  $632.8 \text{ nm}$  (HeNe laser). The focusing depth was  $50 \mu\text{m}$ . The axial intensity distribution is shown for (curve a)  $n_1 = 1.51$  and  $n_2 = 1.33$ , (curve b)  $n_1 = n_2 = 1.51$ , and (curve c)  $n_1 = 1.33$  and  $n_2 = 1.51$ . As this figure shows, the induced aberration has a profound broadening effect on the intensity distribution. Also, the global appearance of curves a and c are mirror imaged with respect to the  $z = 0$  plane.

The geometrical intensity, given by Eq. (35), is plotted in Fig. 5. The lens and media parameters for curves a and b correspond to those for curves a and c of Fig. 4, respectively. Apart from differences such as (1) a sudden jump from zero to a finite intensity at the geometrical shadow boundary instead of a gradually rising peak and (2) a smooth instead of a jagged distribution, the overall form of the intensity shows the main characteristics of the electromagnetic diffraction pattern. Notice that the FWHM of the intensity peak depends through relation (33) on the radius  $\epsilon$  of the detector.

## 6. CONSEQUENCES FOR THREE-DIMENSIONAL IMAGING

Recently, it has been pointed out that in 3-D microscopy the size of objects may be severely overestimated.<sup>19</sup> This is also found to be the case when pointlike fluorescent objects are imaged.<sup>20</sup> This elongation effect occurs if there is a refractive-index mismatch between the immersion fluid of the objective lens and the cover glass on the one hand, and the medium in which the object is embedded on the other. Typically, the refractive index of the immersion oil is  $n_{\text{oil}} = 1.51$ . For watery biological objects, however, the refractive index is  $n_{\text{water}} = 1.33$ . As we show below, the apparent elongation factor may be quite large under these circumstances.

When a refractive-index mismatch occurs, there are two separate causes that contribute to the elongation effect:

1. The width of the illumination peak along the  $z$  axis becomes comparable with the size of the object. If we denote the width of the peak by  $a$  and the true size of the object by  $b$ , then there is a distance  $a + b$ , rather than

just  $b$ , where the two overlap and hence a light signal (e.g., fluorescence) is generated.

2. The shift of the object stage during  $z$  scanning is frequently mistaken to be equal to the shift of the point that is imaged. In practice, however, the latter will be smaller ( $n_1 > n_2$ ) or larger ( $n_2 > n_1$ ).

In curve a of Fig. 4, the first effect is elucidated. Here the vertical illumination distribution is depicted for the case of a watery object imaged by an immersion-oil objective with  $\text{NA} = 1.32$ . As can be seen, the full width at the first minimum of the main peak is  $3.4 \mu\text{m}$ . This is comparable with the typical size of cells ( $\sim 10 \mu\text{m}$ ). So, when the probe is scanned over the object along the  $z$  axis, there is a trajectory of  $10 + 3.4 \mu\text{m}$  over which a

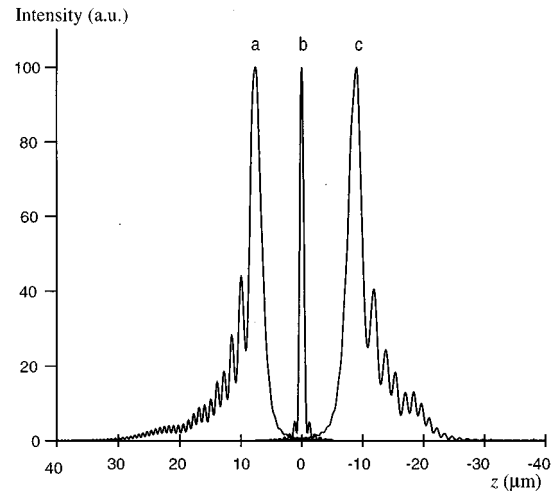


Fig. 4. Results of theory 2 for a lens with semiaperture angle  $60^\circ$  and for a wavelength of  $632.8 \text{ nm}$  (HeNe laser). The focusing depth  $f - d = 50 \mu\text{m}$ , the focal length  $f = 1 \times 10^{-2} \text{ m}$ , and  $\mu_1 = \mu_2 = \mu_0$ . The axial intensity distribution is shown for (curve a)  $n_1 = 1.51$  and  $n_2 = 1.33$ , (curve b)  $n_1 = n_2 = 1.51$ , and (curve c)  $n_1 = 1.33$  and  $n_2 = 1.51$ .

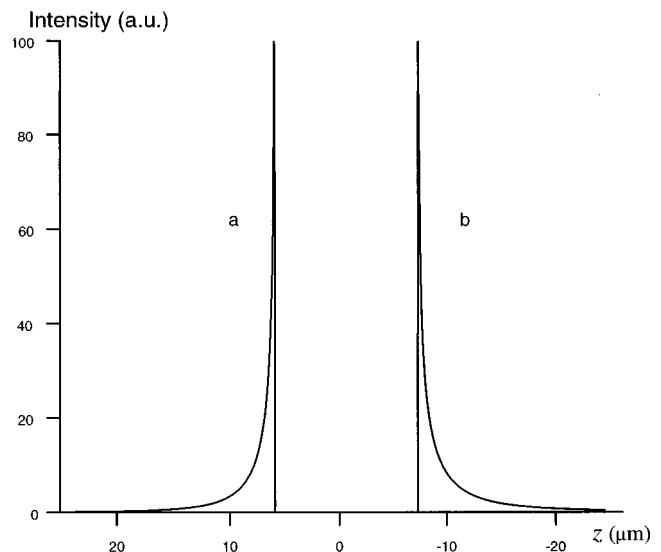


Fig. 5. Intensity distributions according to geometrical optics. Curves a and b correspond to curves a and c, respectively, of Fig. 4. The detector radius  $\epsilon = 1 \times 10^{-8} \text{ m}$ .

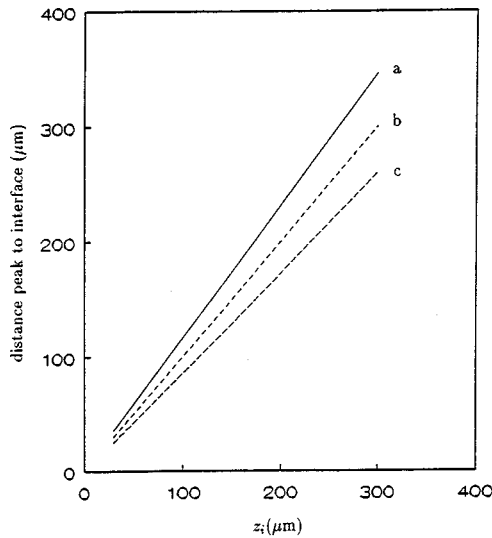


Fig. 6. Distance between the peak and the interface plotted versus the position of the lens. Only if  $n_1 = n_2$  (curve b) does the peak precisely follow the movement of the lens. If  $n_1 > n_2$  (curve c), the peak position shifts less than that of the lens. For  $n_1 < n_2$  (curve a), the opposite holds. In all cases  $\Omega = 60^\circ$ ,  $\mu_1 = \mu_2 = \mu_0$ ,  $f = 10^{-2}$  cm, and  $\lambda = 632.8$  nm. In curve a,  $n_1 = 1.33$  and  $n_2 = 1.51$ ; in curve b,  $n_1 = n_2 = 1.33$ ; and in curve c,  $n_1 = 1.51$  and  $n_2 = 1.33$ .

fluorescent signal is generated. Notice that if we approximate a confocal imaging process by taking the square of the axial intensity distribution, then the full width at the first minimum does *not* change.

The second effect is shown in Fig. 6. It follows, in a good approximation, that a  $z$  shift  $\Delta_z$  of the object stage (or equivalently the lens) results in a shift  $\Delta_z n_1/n_2$  of the object point that is imaged. Effects 1 and 2 together result in the following observation:

An object with true  $z$  dimension  $b$ , imaged with an illumination peak of width  $a$ , will appear to have the following size:

$$\text{apparent size} = \frac{n_1}{n_2} (\text{true size} + \text{peak width}). \quad (36)$$

In Ref. 19, cells of  $7.7 \mu\text{m}$  were studied at  $\lambda_0 = 514$  nm,  $n_1 = 1.518$ ,  $n_2 = 1.33$ , and  $\text{NA} = 1.32$ . Expression (36) then yields an elongation factor (apparent size divided by true size) of 1.64. This prediction agrees very well with the measured factor of  $1.74 \pm 0.30$ .

There is another interesting consequence of the first effect. If an object with a size much smaller than the size of the illumination distribution is imaged, then the axial scan produces the shape of the distribution rather than the shape of the object. This has been shown for scanning infrared microscopy by Török *et al.*<sup>21</sup> In this way so-called nanospheres may be used to produce information on the diffraction pattern.

## 7. CONCLUSIONS

In this paper we have compared two different diffraction theories that describe the effect of a plane dielectric interface on a converging electromagnetic wave.

For shallow focusing depths our numerical results

show that, on comparison of theories 1 and 2, the main axial peaks of the distribution show an excellent agreement. For greater focusing depths the agreement was found to be excellent for the higher-order axial lobe structure.

We have also shown that the  $m$  theory and the first Rayleigh–Sommerfeld integral produce the same axial distribution if the boundary conditions are specified correctly.

A geometrical optics approximation describing the main features of the aberrated intensity distribution has been developed. This provides a rapid method to estimate the location of the peak and gives an indication of whether the axial lobe structure is more pronounced on the positive or the negative side of the main peak.

The two diffraction theories can predict the value of the elongation factor that is observed in 3-D microscopy. This calculated elongation factor agrees well with experimental results.

## APPENDIX A: $m$ THEORY

The derivation of the  $m$  theory presented in this appendix essentially follows Toraldo di Francia.<sup>13</sup> It is well known that a solution  $f(Q)$  of the inhomogeneous wave equation

$$\nabla^2 f + k^2 f = h(Q), \quad (A1)$$

with  $k$  as the wave number, can be obtained in a form

$$f(Q) = - \int_V G(P, Q) h(P) dP, \quad (A2)$$

where  $G(P, Q)$  represents the Green's function pertaining to the Helmholtz equation and  $V$  is the 3-D support of the function  $h$ . Hence

$$\nabla^2 G(P, Q) + k^2 G(P, Q) = -\delta(Q - P). \quad (A3)$$

Next, following Sommerfeld,<sup>22</sup> we construct the point  $Q'$  that is the mirror image of  $Q$  with respect to the plane  $\Sigma$  (see Fig. 7). Two new Green's functions can then be defined as

$$G^\pm(P, Q) = G(P, Q) \pm G(P, Q'). \quad (A4)$$

It follows from Eq. (A3) that they satisfy

$$\begin{aligned} \nabla_P^2 G^\pm(P, Q) + k^2 G^\pm(P, Q) \\ = -\delta(Q - P) \mp \delta(Q' - P). \end{aligned} \quad (A5)$$

Consider an arbitrary constant vector  $\mathbf{c}$ , with components  $\mathbf{c}_1$  parallel and  $\mathbf{c}_2$  perpendicular to the plane  $\Sigma$ . Define the vector function  $\Gamma$  as

$$\Gamma(P, Q) = \mathbf{c}_1 G^-(P, Q) + \mathbf{c}_2 G^+(P, Q). \quad (A6)$$

Thus Eq. (A5) yields

$$\begin{aligned} \nabla_P^2 \Gamma(P, Q) + k^2 \Gamma(P, Q) = -(\mathbf{c}_1 + \mathbf{c}_2) \delta(Q - P) \\ + (\mathbf{c}_1 - \mathbf{c}_2) \delta(Q' - P). \end{aligned} \quad (A7)$$

Next, we derive an identity that will be of use below (cf. Ref. 23). Let  $\mathbf{A}$  and  $\mathbf{B}$  be two vector functions of position, which, together with their first and second derivatives,

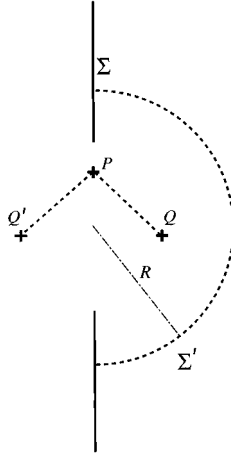


Fig. 7. Geometry used for the construction of the Green's functions  $G^\pm$ .

are continuous throughout  $V$  and on the surface  $S$  that bounds  $V$ . The divergence theorem is applied to the vector  $\mathbf{A} \times (\nabla \times \mathbf{B})$ , giving

$$\int_V \nabla \cdot [\mathbf{A} \times (\nabla \times \mathbf{B})] dV = \int_S [\mathbf{A} \times (\nabla \times \mathbf{B})] \cdot \mathbf{m} dS, \quad (\text{A8})$$

where  $\mathbf{m}$  is a unit normal vector directed outward from  $S$ . Upon expansion of the integrand of the volume integral, a vector analog of Green's first identity is obtained:

$$\begin{aligned} \int_V [\nabla \times \mathbf{A} \cdot \nabla \times \mathbf{B} - \mathbf{A} \cdot \nabla \times (\nabla \times \mathbf{B})] dV \\ = \int_S [\mathbf{A} \times (\nabla \times \mathbf{B})] \cdot \mathbf{m} dS. \end{aligned} \quad (\text{A9})$$

The vector analog of Green's second identity (Green's theorem) is obtained by reversing the roles of  $\mathbf{A}$  and  $\mathbf{B}$  in Eq. (A9) and subtracting one expression from the other:

$$\begin{aligned} \int_V [\mathbf{B} \cdot \nabla \times (\nabla \times \mathbf{A}) - \mathbf{A} \cdot \nabla \times (\nabla \times \mathbf{B})] dV \\ = \int_S [\mathbf{A} \times (\nabla \times \mathbf{B}) - \mathbf{B} \times (\nabla \times \mathbf{A})] \cdot \mathbf{m} dS. \end{aligned} \quad (\text{A10})$$

Specializing the left-hand side of Eq. (A10) to the case  $\mathbf{A} = \mathbf{E}$  and  $\mathbf{B} = \mathbf{\Gamma}$ , we can rewrite it as

$$\int_V [\mathbf{\Gamma} \cdot \nabla_P \times (\nabla_P \times \mathbf{E}) - \mathbf{E} \cdot \nabla_P \times (\nabla_P \times \mathbf{\Gamma})] dP. \quad (\text{A11})$$

Working out the triple products while using the relations  $\nabla^2 \mathbf{E} = -k^2 \mathbf{E}$  and  $\nabla \cdot \mathbf{E} = 0$ , together with Eq. (A7), yields

$$-\int_V \mathbf{E} \cdot [\nabla_P (\nabla_P \cdot \mathbf{\Gamma})] dP - \mathbf{E}(Q) \cdot \mathbf{c} = -\mathbf{E}(Q) \cdot \mathbf{c}, \quad (\text{A12})$$

where we used the result that  $\nabla_P \cdot \mathbf{\Gamma} = 0$  if  $P$  is on the plane  $\Sigma$ .

Proceeding with the right-hand side of Eq. (A10), we get

$$\begin{aligned} \int_\Sigma [\mathbf{E} \times (\nabla_P \times \mathbf{\Gamma}) - \mathbf{\Gamma} \times (\nabla_P \times \mathbf{E})] \cdot \mathbf{m} d\Sigma \\ = \int_\Sigma \mathbf{E} \cdot (\nabla_P \times \mathbf{\Gamma}) \times \mathbf{m} d\Sigma, \end{aligned} \quad (\text{A13})$$

where the contribution of the integration over the hemisphere  $\Sigma'$  has been omitted, as it can be made arbitrarily small by letting  $R \rightarrow \infty$  (see Fig. 7). Also, we used the fact that the second term on the left-hand side of Eq. (A13) is zero, since  $\mathbf{\Gamma}$  is perpendicular to  $\Sigma$  for  $P$  on  $\Sigma$ . In addition,

$$(\nabla_P \times \mathbf{\Gamma}) \times \mathbf{m} = (\nabla_P \times \mathbf{\Gamma})_{\parallel} \times \mathbf{m}, \quad (\text{A14})$$

where the subscript  $\parallel$  denotes the component parallel to the plane  $\Sigma$ . It is seen from Eq. (A6) that

$$(\nabla \times \mathbf{\Gamma})_{\parallel} = 2\nabla_P G \times \mathbf{c}. \quad (\text{A15})$$

Using Eqs. (A14) and (A15) on the right-hand side of Eq. (A13) gives

$$\begin{aligned} 2 \int_\Sigma [-(\mathbf{E} \cdot \nabla G)(\mathbf{c} \cdot \mathbf{m}) + (\mathbf{E} \cdot \mathbf{c})(\nabla G \cdot \mathbf{m})] d\Sigma \\ = 2 \int_\Sigma \mathbf{c} \cdot (\mathbf{m} \times \mathbf{E}) \times \nabla G d\Sigma. \end{aligned} \quad (\text{A16})$$

Equating Eqs. (A12) and (A16) [by virtue of Eq. (A10)] and noticing that  $\mathbf{c}$  is an arbitrary vector, we finally find an expression for the field in a point  $Q$  in terms of its tangential component along the plane  $\Sigma$  only:

$$\mathbf{E}(Q) = 2 \int_\Sigma (\mathbf{m} \times \mathbf{E}) \times \nabla G d\Sigma, \quad (\text{A17})$$

where now, in agreement with Toraldo di Francia's notation, the vector  $\mathbf{m}$  is the *inward* normal to  $\Sigma$ . Note that, as  $Q \rightarrow \Sigma$ , i.e., when the observation point approaches the plane of integration, the assumed boundary condition is regained.

## ACKNOWLEDGMENTS

P. Török is on leave from the Central Research Institute for Physics of the Hungarian Academy of Sciences, Budapest. P. Török acknowledges partial support from Wellcome Trust, UK. His present address is University of Oxford, Department of Engineering Science, Parks Road, Oxford OX1 3PF UK. The authors are grateful to Emil Wolf for his comments on this manuscript.

Correspondence should be addressed to T. D. Visser at the address on the title page.

## REFERENCES AND NOTES

1. J. Gasper, G. C. Sherman, and J. J. Stamnes, "Reflection and refraction of an arbitrary electromagnetic wave at a plane interface," *J. Opt. Soc. Am.* **66**, 955–961 (1976).
2. H. Ling and S.-W. Lee, "Focusing of electromagnetic waves through a dielectric interface," *J. Opt. Soc. Am. A* **1**, 965–973 (1984).
3. J. J. Stamnes, *Waves in Focal Regions* (Hilger, Bristol, UK, 1986), Chap. 16, pp. 482–500.
4. S. Chang, J. H. Jo, and S. S. Lee, "Theoretical calculations

- of optical force exerted on a dielectric sphere in the evanescent field generated with a totally-reflected focused Gaussian beam," *Opt. Commun.* **108**, 133–143 (1994).
5. E. Wolf, "Electromagnetic diffraction in optical systems. I. An integral representation of the image field," *Proc. R. Soc. London, Ser. A* **253**, 349–357 (1959).
  6. B. Richards and E. Wolf, "Electromagnetic diffraction in optical systems. II. Structure of the image field in an aplanatic system," *Proc. R. Soc. London, Ser. A* **253**, 358–379 (1959).
  7. P. Debye, "Das Verhalten von Lichtwellen in der Nähe eines Brennpunktes oder einer Brennnlinie," *Ann. Phys. (Leipzig)* **30**, 755–776 (1909).
  8. R. K. Luneburg, *Mathematical Theory of Optics*, 2nd ed. (U. of California Press, Berkeley, Calif., 1964), Chap. VI, pp. 319–323.
  9. P. Török, P. Varga, Z. Laczik, and G. R. Booker, "Electromagnetic diffraction of light focused through a planar interface between materials of mismatched refractive indices: an integral representation," *J. Opt. Soc. Am. A* **12**, 325–332 (1995).
  10. S. H. Wiersma and T. D. Visser, "Defocusing of a converging electromagnetic wave by a plane dielectric interface," *J. Opt. Soc. Am. A* **13**, 320–325 (1996).
  11. The name *m* theory was coined by Karczewski and Wolf in two papers: B. Karczewski and E. Wolf, "Comparison of three theories of electromagnetic diffraction at an aperture. Part I: Coherence matrices," *J. Opt. Soc. Am.* **56**, 1207–1214 (1966); "Comparison of three theories of electromagnetic diffraction at an aperture. Part II: The far field," *J. Opt. Soc. Am.* **56**, 1214–1219 (1966).
  12. W. R. Smythe, "The double current sheet in diffraction," *Phys. Rev.* **72**, 1066–1070 (1947).
  13. G. Toraldo di Francia, *Electromagnetic Waves* (Interscience, New York, 1955), Chap. 10, pp. 213–223.
  14. H. Severin, "Zur Theorie der Beugung elektromagnetischer Wellen," *Z. Phys.* **129**, 426–439 (1951).
  15. P. Török, C. J. R. Sheppard, and P. Varga, "Study of evanescent waves for transmission near-field optical microscopy," *J. Mod. Opt.* **43**, 1167–1183 (1996).
  16. J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941), pp. 486–488.
  17. L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge U. Press, Cambridge, 1995), Chap. 3, pp. 125–127.
  18. M. Born and E. Wolf, *Principles of Optics*, 6th ed. (Pergamon, Oxford, 1991), Chap. 3, pp. 113–117.
  19. T. D. Visser and J. L. Oud, "Volume measurements in 3-D microscopy," *Scanning* **16**, 198–200 (1994).
  20. C. J. R. Sheppard and P. Török, "Effects of specimen refractive index on confocal imaging," *J. Microsc.* **185**, 366–374 (1997).
  21. P. Török, G. R. Booker, Z. Laczik, and R. Falster, "A new confocal SIRM incorporating reflection, transmission and double-pass modes either with or without differential phase contrast imaging," *Inst. Phys. Conf. Ser.* **134**, 771–774 (1993).
  22. A. Sommerfeld, *Optics* (Academic, New York, 1954), Chap. V, p. 200.
  23. J. A. Stratton and L. J. Chu, "Diffraction theory of electromagnetic waves," *Phys. Rev.* **56**, 99–107 (1939).