

# Analyses of Vector Gaussian Beam Propagation and the Validity of Paraxial and Spherical Approximations

Carl G. Chen, Paul T. Konkola, Juan Ferrera, Ralf K. Heilmann and Mark L. Schattenburg

*Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*

Submitted to *Applied Optics: Optical Technology and Biomedical Optics* on June 20, 2000

## Abstract

In this paper, we present a full vector analysis of Gaussian beam propagation by using the well-known method of angular spectrum of plane waves. The Gaussian beam is assumed to traverse a charge-free, homogeneous, isotropic, linear and non-magnetic dielectric medium. The angular spectrum representation, in its vector form, is applied to a problem with a Gaussian intensity boundary condition. After some mathematical manipulations, each non-zero propagating electric field component is expressed in terms of a power series expansion. The usual paraxial result corresponds to the first term in the expansion of the transverse field component. The leading term in the longitudinal expansion is readily identified as the first non-fundamental Hermite-Gaussian (HG) mode. In spite of offering some straightforward mathematical analyticity, this approach has a notable drawback. Specifically, the higher order terms that we and others<sup>1</sup> have derived diverge at locations that are sufficiently far away from the initial boundary, yielding unphysical results. Hence, any meaningful usage of the expansion approach calls for a careful study of its range of applicability. By considering the transition of a Gaussian wave from the paraxial to the spherical regime, we are able to derive a simple expression for the range within which the series produce numerically satisfying answers.

## 1. Introduction

Thanks to its elegant physics and intuitive simplicity, the angular spectrum representation has been used to solve a variety of problems involving the propagation, transmission and reflection of Gaussian waves. Its theoretical foundation has been well anchored by a large number of authors. Traditionally however, the representation has been used in its scalar form, and the propagating field has been calculated in conjunction with the so-called paraxial approximation<sup>4-6</sup>. Though researchers such as Agrawal and Pattanayak<sup>1</sup> have extended the solution beyond the paraxial result, their approach to the problem remains scalar in nature.

Despite some added mathematical difficulties, the extension of the scalar solution to the vector is relatively straightforward. Rooting from the full vector description of the angular spectrum of plane waves, we adopt essentially the same scheme as that implemented by Agrawal and Pattanayak. First, a specific boundary field distribution with a Gaussian intensity profile is proposed, the use of which conveniently eliminates one of the two transverse electric field components,  $y$  in our case. We then proceed to calculate an exact solution of the propagating vector field,  $\vec{E}(\vec{r})$ , for any spatial displacement  $\vec{r}$ . The transverse  $x$ -component has been investigated by Agrawal and Pattanayak. The result is expressed in terms of a power series expansion, with a well-defined expansion parameter. The first term in the expansion corresponds to the usual paraxial result, while higher order terms represent non-Gaussian corrections. This paper focuses on deriving a similar expansion for the longitudinal  $z$ -component, with the leading term of the expansion corresponding to the first non-fundamental Hermite-Gaussian (HG) mode. The forms of these two series, and the relationships among the various terms, satisfy those postulated by Lax, Louisell and McKnight<sup>7</sup>. The main advantage of such an analytical approach, besides offering physical insight, is its considerable simplification over the often lengthy numerical calculations.

A limitation associated with this expansion approach is the rapid divergence of the higher order terms at distances far from the initial boundary plane. To quantify the range of applicability for the series, we examine two schemes that are often used to approximate a scalar

Gaussian beam: the paraxial approximation, valid for wave descriptions close to the axis of propagation, and the spherical approximation, valid at radial distances that are far greater than a wavelength. By studying the phase difference between these two approximations, we are able to derive an expression which quantifies the transition from the paraxial to the spherical regime. We show that beyond this transition plane, higher order terms contained in the aforementioned series become increasingly divergent with distance, thus producing unphysical results. To analytically study the beam behavior here, alternatives such as the method of stationary phase<sup>3</sup> must be used.

This work is motivated by a desire to understand the theoretical aspects of both scalar and vector Gaussian beam propagation, which are of great importance to the successful implementation of our scanning beam interference lithography system (SBIL)<sup>10</sup>. The precise characterization of propagating Gaussian waves is also vital in understanding and constructing fine metrology instruments<sup>11,12</sup>.

## 2. Vector Gaussian Beam

In his classic papers, Rhodes<sup>8,9</sup> derived the full vector representation of the angular spectrum of fields. Following Carter<sup>3</sup>, we rewrite Rhodes' results with a slightly more intuitive notation, which also coincides with that used by Agrawal and Pattanayak<sup>1</sup>:

$$E_x(\vec{r}) = \int \int_{-\infty}^{+\infty} A_x(p, q) e^{ik(px+qy+mz)} dp dq, \quad (1)$$

$$E_y(\vec{r}) = \int \int_{-\infty}^{+\infty} A_y(p, q) e^{ik(px+qy+mz)} dp dq, \quad (2)$$

$$E_z(\vec{r}) = - \int \int_{-\infty}^{+\infty} \left[ \frac{p}{m} A_x(p, q) + \frac{q}{m} A_y(p, q) \right] e^{ik(px+qy+mz)} dp dq. \quad (3)$$

Eqs. (1), (2) and (3) together represent the solution to the vector wave equation

$$(\nabla^2 + k^2)\vec{E}(\vec{r}) = 0 \quad (4)$$

in the  $z > 0$  half-space. An harmonic time dependence,  $e^{-i\omega t}$ , has been suppressed in the field expressions, and

$$m = \begin{cases} \sqrt{1 - p^2 - q^2} & \text{if } p^2 + q^2 \leq 1, \\ i\sqrt{p^2 + q^2 - 1} & \text{if } p^2 + q^2 > 1. \end{cases} \quad (5)$$

The wave vector has a magnitude  $k = \sqrt{\epsilon}\omega/c$ .

Physically, each of the three electric field components is made up of plane waves traveling in all directions, whose magnitudes are determined by the complex factors  $A_x(p, q)$  and  $A_y(p, q)$ . Imaginary  $m$  values signify the existence of evanescent waves, which propagate freely along the  $xy$  plane but decay exponentially along the positive  $z$  direction. Eqs. (1) to (3) demonstrate a simple physical principle: an electric field that satisfies the wave equation can be expressed in terms of, at most, two scalar functions,  $A_x$  and  $A_y$ .

To find these scalar functions, it is enough to specify the two transverse field components  $E_x$  and  $E_y$ , at a boundary plane  $z = 0$ . Our choice of boundary conditions evolves from that of Agrawal and Pattanayak. Specifically, we adopt an initial Gaussian intensity profile for the  $x$ -component,

$$E_x(x, y, 0) = \exp\left[-\frac{x^2 + y^2}{2\omega_0^2}\right], \quad (6)$$

where  $\omega_0$  is a measure of the beam width. For the  $y$ -component, we choose

$$E_y(x, y, 0) = 0. \quad (7)$$

By inverse Fourier transforming (1) and (2), substituting in (6) and (7), and performing the consequent integrations, we find

$$A_x(p, q) = \left(\frac{k}{2\pi}\right)^2 \iint_{-\infty}^{+\infty} E_x(x, y, 0) e^{-ik(px+qy)} dx dy = \frac{1}{2\pi f^2} \exp\left[-\frac{p^2 + q^2}{2f^2}\right], \quad (8)$$

$$A_y(p, q) = \left(\frac{k}{2\pi}\right)^2 \iint_{-\infty}^{+\infty} E_y(x, y, 0) e^{-ik(px+qy)} dx dy = 0, \quad (9)$$

where  $f = (k\omega_0)^{-1}$  is proportional to the ratio of the wavelength over the beam width. Combining (2) and (9), we obtain  $E_y(x, y, z) = 0$ . Thus the choice of a zero  $y$ -component at the boundary plane ensures the vanishing of the  $y$ -component throughout the  $z > 0$  half-space. We need now only to concern ourselves with the  $x$  and  $z$  components.

With  $A_x$  and  $A_y$  given by (8) and (9), the transverse ( $x$ ) and longitudinal ( $z$ ) components of the electric field are represented by two double integrals. Due to the cylindrical symmetry of the problem, the double integrals can be further reduced to single ones:

$$E_x(\vec{r}) = \int_0^\infty \frac{1}{f^2} e^{-\frac{b^2}{2f^2}} e^{ikmz} J_0(k\rho b) b db, \quad (10)$$

$$E_z(\vec{r}) = - \int_0^\infty \frac{ix}{\rho} \frac{1}{f^2} e^{-\frac{b^2}{2f^2}} e^{ikmz} J_1(k\rho b) \frac{b^2}{\sqrt{1-b^2}} db, \quad (11)$$

where  $b = \sqrt{p^2 + q^2}$ ,  $\rho = \sqrt{x^2 + y^2}$ , and  $J_0$  and  $J_1$  are the zeroth and first-order Bessel functions, respectively. When  $b > 1$ ,  $m$  is complex, and the integrands decay exponentially, becoming negligibly small as one retreats further than several wavelengths away from the  $z = 0$  plane. Hence, as a first step to evaluate (10) and (11) analytically, we drop the evanescent field components by lowering the upper limit of integration from  $\infty$  to 1:

$$E_x(\vec{r}) = \int_0^1 \frac{1}{f^2} e^{-\frac{b^2}{2f^2}} e^{ikmz} J_0(k\rho b) b db, \quad (12)$$

$$E_z(\vec{r}) = - \int_0^1 \frac{ix}{\rho} \frac{1}{f^2} e^{-\frac{b^2}{2f^2}} e^{ikmz} J_1(k\rho b) \frac{b^2}{\sqrt{1-b^2}} db. \quad (13)$$

In reality, we are most often interested in knowing the beam character tens if not thousands of wavelengths away from the beam waist. At that distance, the approximation made in (12) and (13) is for all practical purposes the same as the exact result.

### 3. Power Series Expansion

Integral (12) has been thoroughly investigated by Agrawal and Pattanayak<sup>1</sup>. Integral (13) is the subject of study of this paper. Though at times we have to build new tools to tackle this seemingly different problem, our basic approach closely resembles that of Agrawal and Pattanayak's.

The following expansion of  $e^{ikmz}$ , identical to Eq. (13) in Ref. 1, is valid for  $b < 1$ :

$$e^{ikmz} = \exp(ikz\sqrt{1-b^2}) = \left(\frac{\pi kz}{2}\right)^{\frac{1}{2}} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{kzb^2}{2}\right)^n H_{n-\frac{1}{2}}^{(1)}(kz), \quad (14)$$

where

$$\begin{aligned}
H_{-\frac{1}{2}}^{(1)}(kz) &= \left(\frac{2}{\pi kz}\right)^{\frac{1}{2}} e^{ikz}, \\
H_{n-\frac{1}{2}}^{(1)}(kz) &= \left(\frac{2}{\pi kz}\right)^{\frac{1}{2}} e^{ikz} i^{-n} \sum_{m=0}^{n-1} (-1)^m \frac{(n+m-1)!}{m!(n-m-1)!} \left(\frac{1}{2ikz}\right)^m, \quad n \geq 1,
\end{aligned} \tag{15}$$

is the Hankel function of the first kind<sup>13</sup>. (Note the critical factor  $m!$  which is missing in Agrawal and Pattanayak's paper due to a probable typographical error.)

Substituting (14) into (13), we can switch freely the order of summation and integration because the region of integration falls within the radius of convergence of the series. We then obtain

$$E_z(\vec{r}) = -\frac{ix}{\rho} \frac{1}{f^2} \left(\frac{\pi kz}{2}\right)^{\frac{1}{2}} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{kz}{2}\right)^n H_{n-\frac{1}{2}}^{(1)}(kz) T_n(\rho), \tag{16}$$

where

$$T_n(\rho) = \int_0^1 e^{-\frac{b^2}{2f^2}} \frac{b^{2n+2}}{\sqrt{1-b^2}} J_1(k\rho b) db. \tag{17}$$

An analytical evaluation of (17) is difficult. We opt for an alternative that expands the integrand first. For  $b < 1$ , we have a convergent expansion

$$\frac{1}{\sqrt{1-b^2}} = 1 + \frac{1}{2}b^2 + \frac{3}{8}b^4 + \dots \tag{18}$$

Direct replacement leads to

$$\begin{aligned}
T_n(\rho) &= \int_0^1 e^{-\frac{b^2}{2f^2}} b^{2n+2} \left(1 + \frac{1}{2}b^2 + \frac{3}{8}b^4 + \dots\right) J_1(k\rho b) db \\
&\equiv T_n^{(1)}(\rho) + T_n^{(2)}(\rho) + \dots
\end{aligned} \tag{19}$$

where

$$T_n^{(1)}(\rho) = \int_0^1 e^{-\frac{b^2}{2f^2}} b^{2n+2} J_1(k\rho b) db, \tag{20}$$

$$T_n^{(2)}(\rho) = \frac{1}{2} \int_0^1 e^{-\frac{b^2}{2f^2}} b^{2n+4} J_1(k\rho b) db, \tag{21}$$

etc.

Though the transformed integrals seem easier to handle, their evaluations remain thorny till we notice that for small  $f$  ratios, i.e.,  $f < 0.4$ , or equivalently  $\omega_0 > 0.4\lambda$ , values of the

integrals will stay nearly the same as the upper limit of integration extends to  $\infty$ . In other words, the contribution to the integral from 1 to  $\infty$  becomes increasingly negligible as  $f$  decreases. Calculations<sup>14</sup> then yield

$$T_n^{(1)}(\rho) = \int_0^\infty e^{-\frac{b^2}{2f^2}} b^{2n+2} J_1(k\rho b) db = (\sqrt{2}f)^{2n+3} \frac{n!}{2} \left(\frac{\rho^2}{2\omega_0^2}\right)^{\frac{1}{2}} e^{-\frac{\rho^2}{2\omega_0^2}} L_n^1\left(\frac{\rho^2}{2\omega_0^2}\right), \quad (22)$$

$$\begin{aligned} T_n^{(2)}(\rho) &= \frac{1}{2} \int_0^\infty e^{-\frac{b^2}{2f^2}} b^{2n+4} J_1(k\rho b) db \\ &= \frac{1}{2} (\sqrt{2}f)^{2n+5} \frac{(n+1)!}{2} \left(\frac{\rho^2}{2\omega_0^2}\right)^{\frac{1}{2}} e^{-\frac{\rho^2}{2\omega_0^2}} L_{n+1}^1\left(\frac{\rho^2}{2\omega_0^2}\right), \end{aligned} \quad (23)$$

etc., where  $L_b^a$  is the associated Laguerre polynomial.

By now, it already seems qualitatively plausible that (16) can be expressed in terms of a power series expansion, with some  $f$ -power being the expansion parameter. To proceed with the quantitative derivations, we plug (19) into (16) and define

$$E_z(\vec{r}) \equiv E_z^i(\vec{r}) + E_z^{ii}(\vec{r}) + \dots \quad (24)$$

where

$$E_z^i(\vec{r}) = -\frac{ix}{\rho} \frac{1}{f^2} \left(\frac{\pi kz}{2}\right)^{\frac{1}{2}} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{kz}{2}\right)^n H_{n-\frac{1}{2}}^{(1)}(kz) T_n^{(1)}(\rho), \quad (25)$$

$$E_z^{ii}(\vec{r}) = -\frac{ix}{\rho} \frac{1}{f^2} \left(\frac{\pi kz}{2}\right)^{\frac{1}{2}} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{kz}{2}\right)^n H_{n-\frac{1}{2}}^{(1)}(kz) T_n^{(2)}(\rho), \quad (26)$$

etc.

First, focus on (25). Substituting in (15) and (22), we obtain, after persevering through some heavy algebra,

$$E_z^i(\vec{r}) = -\frac{ixf}{\omega_0} e^{-\frac{\rho^2}{2\omega_0^2}} e^{ikz} \left\{ 1 + \sum_{n=1}^{\infty} \left(\frac{-iz}{l}\right)^n L_n^1\left(\frac{\rho^2}{2\omega_0^2}\right) \sum_{m=0}^{n-1} \frac{(n+m-1)!}{m!(n-m-1)!} \left(\frac{-1}{2ikz}\right)^m \right\}, \quad (27)$$

where  $l = k\omega_0^2 = \omega_0/f$  is the so-called diffraction length, which characterizes the longitudinal direction associated with the Gaussian beam. To further demonstrate the physics wrapped in (27), we regroup the terms according to their  $m$  values and perform the summations to obtain

$$E_z^i(\vec{r}) \equiv f E_z^{(1)}(\vec{r}) + f^3 \frac{-\frac{iz}{l}}{\left(1 + \frac{iz}{l}\right)^2} L_2^1 \left( \frac{\rho^2}{2\omega_0^2 \left(1 + \frac{iz}{l}\right)} \right) E_z^{(1)}(\vec{r}) + \mathcal{O}(f^5), \quad (28)$$

where

$$E_z^{(1)}(\vec{r}) = -\frac{ix}{\omega_0} \frac{e^{ikz}}{\left(1 + \frac{iz}{l}\right)^2} \exp \left[ -\frac{\rho^2}{2\omega_0^2 \left(1 + \frac{iz}{l}\right)} \right]. \quad (29)$$

In so doing, we take note of Endnote 11 in Ref. 1. Similarly, we find for the second term

$$\begin{aligned} E_z^{ii}(\vec{r}) &= -\frac{ixf^3}{\omega_0} e^{-\frac{\rho^2}{2\omega_0^2}} e^{ikz} \left\{ L_1^1 \left( \frac{\rho^2}{2\omega_0^2} \right) + \sum_{n=1}^{\infty} (n+1) \left( \frac{-iz}{l} \right)^n L_{n+1}^1 \left( \frac{\rho^2}{2\omega_0^2} \right) \times \right. \\ &\quad \left. \sum_{m=0}^{n-1} \frac{(n+m-1)!}{m!(n-m-1)!} \left( \frac{-1}{2ikz} \right)^m \right\} \\ &= f^3 \frac{1}{1 + \frac{iz}{l}} L_1^1 \left( \frac{\rho^2}{2\omega_0^2 \left(1 + \frac{iz}{l}\right)} \right) E_z^{(1)}(\vec{r}) + \mathcal{O}(f^5). \end{aligned} \quad (30)$$

Clearly,  $E_z^{iii}(\vec{r})$  and higher order terms will weigh in with  $\mathcal{O}(f^5)$  or higher.

Combining (24), (28) and (30), we obtain for the longitudinal component

$$E_z(\vec{r}) = f E_z^{(1)}(\vec{r}) + f^3 E_z^{(3)}(\vec{r}) + \mathcal{O}(f^5), \quad (31)$$

where

$$E_z^{(3)}(\vec{r}) = \left\{ \frac{1}{1 + iz/l} L_1^1 \left( \frac{\rho^2}{2\omega_0^2(1 + iz/l)} \right) - \frac{iz/l}{(1 + iz/l)^2} L_2^1 \left( \frac{\rho^2}{2\omega_0^2(1 + iz/l)} \right) \right\} E_z^{(1)}(\vec{r}). \quad (32)$$

Hitherto, (16) has been transformed into a power series expansion with an expansion parameter  $f^2$ . Compare (31) with Agrawal and Pattanayak's derivation for the transverse component which has been further expanded to include the fourth-order term,

$$E_x(\vec{r}) = E_x^{(0)}(\vec{r}) + f^2 E_x^{(2)}(\vec{r}) + f^4 E_x^{(4)}(\vec{r}) + \mathcal{O}(f^6), \quad (33)$$

where

$$E_x^{(0)}(\vec{r}) = \frac{e^{ikz}}{1 + iz/l} \exp \left[ -\frac{\rho^2}{2\omega_0^2 (1 + iz/l)} \right], \quad (34)$$

$$E_x^{(2)}(\vec{r}) = -\frac{iz/l}{(1 + iz/l)^2} L_2 \left( \frac{\rho^2}{2\omega_0^2 (1 + iz/l)} \right) E_x^{(0)}(\vec{r}), \quad (35)$$

$$E_x^{(4)}(\vec{r}) = -\frac{3iz/l}{(1 + iz/l)^4} \left[ L_4 \left( \frac{\rho^2}{2\omega_0^2 (1 + iz/l)} \right) + \frac{\rho^2}{8\omega_0^2} L_3^1 \left( \frac{\rho^2}{2\omega_0^2 (1 + iz/l)} \right) \right] E_x^{(0)}(\vec{r}), \quad (36)$$

and where  $L_b \equiv L_b^0$ . As may be intuitively evident, the longitudinal component is much weaker in strength than the transverse, due to an additional multiplicative factor  $f$  and a quadratic power dependence on  $1/z$ . It is also out of phase with the transverse component. Furthermore, unlike the transverse component whose lowest-order term maintains a Gaussian intensity profile along  $z$ , the longitudinal component possesses no such behavior due to its intrinsic  $x$  dependence. In fact, the leading term in the longitudinal expansion corresponds to the first non-fundamental HG mode, as will be discussed in the next section.

The term  $E_x^{(0)}(\vec{r})$  is commonly referred to as the paraxial result, also known as the fundamental HG mode. It is the “traditional” scalar Gaussian solution. The  $|E_x^{(0)}(\vec{r})|^2$  intensity has a  $1/e$  radius

$$\omega(z) = \omega_0 \sqrt{1 + z^2/l^2}. \quad (37)$$

We refer to Eqs. (31) and (33) together as the vector solution and Eq. (33) alone the scalar. One may have noticed that  $E_z$  at the boundary  $z = 0$  does not vanish. This is consistent with the theory of electrodynamics: if two boundary field components are known, the third will be fixed automatically. For this reason, unlike a scalar beam, the propagation of a vector beam does not allow a pure scalar Gaussian intensity specification at the boundary.

#### 4. Discussion

One can show explicitly that Eq. (29) represents the first non-fundamental HG mode with  $x$  mirror symmetry<sup>6</sup>. Fig. 1 shows a plot of  $|E_z^{(1)}|^2$ . The total absence of the  $y$  mirror-symmetric term can be attributed to the simple boundary field we considered. If instead of (6) and (7), we had adopted a more complicated boundary field condition, for instance,

$$E_x(x, y, 0) = \frac{1}{\sqrt{2}} \exp\left[-\frac{x^2 + y^2}{2\omega_0^2}\right], \quad (38)$$

$$E_y(x, y, 0) = \frac{1}{\sqrt{2}} \exp\left[-\frac{x^2 + y^2}{2\omega_0^2}\right], \quad (39)$$

then the analytical procedure outlined in the prior sections would have led to an additional  $y$  mirror-symmetric field in  $E_z^{(1)}$ .

Eqs. (31) and (33) fit the forms of the two power series proposed by Lax, Louisell and McKnight<sup>7</sup> (see Eq. (3.18)). Together the two render an explicit description of a vector Gaussian beam propagation through space. Lax *et al.* also derived four sets of differential equations (Eqs. (3.20), (3.19), (3.24) and (3.25), with  $g = 0$ ) that relate the various terms in (31) and (33):

$$\left(2ik\frac{\partial}{\partial z} + \nabla_T^2\right)\mathcal{E}_x^{(0)}(\vec{r}) = 0, \quad (40)$$

$$\mathcal{E}_z^{(1)}(\vec{r}) = i\omega_0\frac{\partial}{\partial x}\mathcal{E}_x^{(0)}(\vec{r}), \quad (41)$$

$$\left(2ik\frac{\partial}{\partial z} + \nabla_T^2\right)\mathcal{E}_x^{(2)}(\vec{r}) = -kl\frac{\partial^2\mathcal{E}_x^{(0)}(\vec{r})}{\partial z^2}, \quad (42)$$

$$\mathcal{E}_z^{(3)}(\vec{r}) = i\omega_0\frac{\partial\mathcal{E}_x^{(2)}(\vec{r})}{\partial x} + il\frac{\partial\mathcal{E}_z^{(1)}(\vec{r})}{\partial z}, \quad (43)$$

where  $\nabla_T^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$  is the transverse portion of the Laplacian. Note, however, by the choice of their definition, the  $\mathcal{E}$ 's in (40) through (43) differ from our  $E$ 's by an absence of the factor  $e^{ikz}$ . (See Eq. (3.1) in Ref. 7.)

Eq. (40) is the well-known paraxial equation<sup>4-6</sup>. Its solution,  $\mathcal{E}_x^{(0)}$ , is the paraxial result, which describes the fundamental HG mode. While  $\mathcal{E}_z^{(1)}$  represents the first non-fundamental HG mode, terms  $\mathcal{E}_x^{(2)}$ ,  $\mathcal{E}_z^{(3)}$  and higher correspond to vector non-Gaussian corrections to the paraxial optics. It can be quickly verified that our expressions (29), (32), (34) and (35) satisfy the above four differential equations. Hence the mathematics employed thus far is indeed plausible.

As an illustration, the ratio of the vector field intensity to its scalar counterpart is plotted in Figs. 2 and 3 for two particular sets of parameters. Since  $f$  is small in both cases, the first two terms in the  $E_x$  expansion well approximate the exact integral result (10), and the same holds for  $E_z$  (11). Hence the plots have been generated using only these four terms. Note that in Fig. 2, where  $f = 0.043$ , the calculation indicates a 1% difference between the

scalar and vector intensities at the edge of the region, while in Fig. 3, where  $f = 0.28$ , a 40% difference is indicated. This illustrates the general trend of the reduction in magnitude of higher-order Gaussian terms (both scalar and vector) as  $f$  is reduced.

## 5. Range of Applicability

Astute readers may wonder why, in Figs. 2 and 3, we have chosen propagation distances so close to the  $z = 0$  boundary plane. Fig. 4 shows a plot of the scalar field intensity at a distance  $z = 20$  mm, for  $\omega_0 = 1.3 \mu\text{m}$  and  $\lambda = 351.1$  nm ( $f = 0.043$ ): the solid line represents  $|E_x^{(0)}|^2$ , i.e., the intensity due to the first term in Eq. (33), the dotted line  $|E_x^{(0)} + f^2 E_x^{(2)}|^2$  and the dashed line  $|E_x^{(0)} + f^2 E_x^{(2)} + f^4 E_x^{(4)}|^2$ . As the trend clearly indicates, at this  $z$ , the higher-order terms in the expansion (33) becomes increasingly divergent. Following the remarks prior to Eq. (22) in Ref.<sup>1</sup>, we speculate that the behavior arises from an interchange of summations. A similar divergence in the longitudinal component (31) is also observed. Therefore, the series expansion approach, at least in its truncated form, has a limited range of applicability in the  $z$  direction, within which the approach offers sensible corrections to the paraxial optics, but beyond which it gives unrealistic results, as Fig. 4 clearly demonstrates. To quantify this range, we start by considering the evolution of a scalar Gaussian wave (10, 33) from the paraxial to the spherical regime. The key is to realize that beyond a certain transition point  $z = z_t$ , paraxial optics will no longer render a fitting description of the wave propagation. Hence the series expansion, which provides higher-order corrections to the paraxial result, will correspondingly break down.

Fig. 5 shows a schematic of the coordinates used in deriving  $z_t$ . For  $\rho = \sqrt{x^2 + y^2}$  sufficiently small, we are very close to the  $z$  axis. We term this the paraxial regime. Inside, the first term in (33), i.e., Eq. (34), gives an accurate description of the scalar Gaussian beam. The paraxial phase is,

$$\phi_{par}(\rho, z) = kz - \arctan\left(\frac{z}{l}\right) + \frac{\rho^2 z}{2\omega_0^2 l (1 + z^2/l^2)}. \quad (44)$$

On the other hand, for  $\rho$  large such that  $kr \gg 1$ , where  $r = \sqrt{\rho^2 + z^2}$  is the spherical radius,

Carter<sup>3</sup> showed, via the method of stationary phase, that (10) can be approximated as

$$E_x(\vec{r}) = -i \frac{lz}{r^2} \exp\left(-\frac{kl\rho^2}{2r^2}\right) e^{ikr}. \quad (45)$$

Hence the beam carries a spherical phase,

$$\phi_{sph}(\rho, z) = -\frac{\pi}{2} + kr = -\frac{\pi}{2} + k\sqrt{\rho^2 + z^2}. \quad (46)$$

We term this the spherical regime. The difference between  $\phi_{par}$  and  $\phi_{sph}$  is denoted

$$\Delta\phi(\rho, z) = \phi_{par}(\rho, z) - \phi_{sph}(\rho, z). \quad (47)$$

The transition  $\rho = \rho_t$  from the paraxial to the spherical regime is found by minimizing  $\Delta\phi$  with respect to  $\rho$ , yielding

$$\rho_t(z) = l\sqrt{2 + \frac{l^2}{z^2}}. \quad (48)$$

Intuitively, Eq. (48) makes good sense. As  $z$  decreases, the paraxial regime should incorporate a greater circular region in the  $xy$ -plane.

Now, take  $z \rightarrow 0$ . We have  $\rho_t$  going to  $\infty$  and the  $1/e$  beam radius  $\omega(z)$  (Eq. (37)) nearing  $\omega_0$ . Since the entire beam is confined within  $\rho_t$ , the paraxial regime dictates the field character. As  $z$  moves away from  $z = 0$ ,  $\rho_t$  decreases, shrinking down to 0 as  $z \rightarrow \infty$ , and  $\omega(z)$  increases, approaching  $\infty$ . The beam now falls wholly outside of  $\rho_t$ , where the spherical regime reigns. Hence, to find boundary  $z = z_t$  that signals the transition from one regime to the other, we equate  $\rho_t$  to the  $1/e$  beam radius  $\omega(z)$ ,  $\rho_t(z_t) = \omega(z_t)$ , or

$$l\sqrt{2 + \frac{l^2}{z_t^2}} = \omega_0\sqrt{1 + \frac{z_t^2}{l^2}}. \quad (49)$$

Solving for  $z_t$  in the limit  $f \ll 1$ , we obtain

$$z_t = l\sqrt{\frac{(2 - f^2) + \sqrt{4 + f^4}}{2f^2}} \approx \sqrt{2} \frac{l}{f} = \sqrt{2} k^2 \omega_0^3, \quad (50)$$

which shows that  $z_t$  scales as the third power of  $\omega_0$ . In other words, the transformation from the paraxial to the spherical regime happens faster for smaller  $\omega_0$ . Physically, a Gaussian

beam with a tiny initial radius will diverge very quickly. Its phase front starts to resemble a sphere after traveling only a very short distance. The examples depicted in Figs. 2 and 3 were chosen at the position  $z = z_t$ , with  $z_t = 1$  mm in Fig. 2 and  $z_t = 3.6 \mu\text{m}$  in Fig. 3. Beyond these values, the paraxial optics quickly breaks down and the series expansion approach becomes invalid.

The validity of Eq. (50) is quantitatively demonstrated in Figs. 6, 7 and 8. In Fig. 6, a numerical integration<sup>15</sup> of the scalar Gaussian field (10) is carried out for  $\lambda = 351.1$  nm,  $\omega_0 = 1.3 \mu\text{m}$  and  $z = z_t/2 = 0.5$  mm, to obtain the exact field magnitude and phase information. Magnitudes and phases due to the first three terms in the expansion (33), i.e.,  $E_x^{(0)}$ ,  $E_x^{(0)} + f^2 E_x^{(2)}$  and  $E_x^{(0)} + f^2 E_x^{(2)} + f^4 E_x^{(4)}$ , are then calculated in sequence. Their deviations from the numerical results are plotted. The deviations in magnitude and phase due to the spherical approximation are also shown. As expected, within the beam radius, the results clearly show that the paraxial regime dominates and the inclusion of higher order terms offer increasingly accurate approximations. However, outside of the beam radius, the spherical approximation takes over and provides a much better fit. Figs. 7 and 8 were done at the same  $\lambda$  and  $\omega_0$ , but for  $z = z_t = 1$  mm and  $z = 2z_t = 2$  mm, respectively. The trend indicated by the curves clearly casts doubt on the validity of the series expansion approach for  $z > z_t$ . If the series were well behaved, we would expect higher-order terms to converge uniformly towards the numerical result. However, as shown in Fig. 8(b), the diverging behavior of the second-order term is evident at  $z = 2z_t$ . For  $z \gg z_t$ , the divergence of higher-order terms only gets worse (Fig. 4). Therefore, beyond the range of applicability demarcated by  $z_t$ , our power series method is unreliable and should not be used. Spherical approximation (45), on the other hand, becomes increasingly accurate for larger  $z$ . Note that Eq. (45) is attained via the method of stationary phase<sup>3</sup>, and corresponds to the first term in an asymptotic expansion<sup>2</sup>. Thus in general, if one seeks to describe the field outside of  $z_t$ , other analytical tools such as asymptotic approximations can be used.

## 6. Vector Corrections

In deriving the range of applicability for the series expansion approach, we considered only the transverse component of the electric field, i.e., Eq. (33). Since our goal is to study vector Gaussian beam propagation, it is also essential to investigate the contribution from the longitudinal component.

We know from Sec. 3 that the leading term in the longitudinal expansion,  $fE_z^{(1)}$  in Eq. (31), is significantly weaker than  $E_x^{(0)}$ , which explains why Gaussian beams have traditionally been expressed by the scalar paraxial approximation. To study whether vector corrections are needed, we have to compare  $fE_z^{(1)}$  to  $f^2E_x^{(2)}$ , the next higher order term in the transverse expansion.

As shown in Fig. 1, for a given  $z$ , due to asymmetry, the intensity

$$\left|fE_z^{(1)}(x, y, z)\right|^2 = f^2 \frac{x^2}{\omega_0^2(1+z^2/l^2)^2} \exp\left[-\frac{x^2+y^2}{\omega_0^2(1+z^2/l^2)}\right] \quad (51)$$

takes on maximum values at two mirror-opposite points,  $(x, y=0, z)$  and  $(-x, y=0, z)$ . By maximizing  $\left|fE_z^{(1)}(x, y=0, z)\right|^2$  with respect to  $x$ , we obtain the corresponding maximum intensity

$$\left|fE_z^{(1)}\right|_{max}^2 = \frac{f^2}{e(1+z^2/l^2)} \quad (52)$$

at

$$x_{max} = \omega_0 \sqrt{1+z^2/l^2}. \quad (53)$$

Eqs. (52) and (53) show that the location of the intensity maximum coincides with the  $1/e$  beam waist for  $\left|E_x^{(0)}\right|^2$ , and that the maximum intensity trails  $\left|E_x^{(0)}\right|_{max}^2$  by a factor of  $f^2/e$ . Furthermore,  $\left|fE_z^{(1)}\right|_{max}^2$  is also a monotonic decreasing function of  $z$ , having a maximum value of  $f^2/e$  at  $z=0$ . At the transition range  $z_t = \sqrt{2}l/f$ , assuming  $f \ll 1$ , i.e.,  $z_t \gg l$ , we have

$$\left|fE_z^{(1)}\right|_{max}^2 \approx \frac{f^2}{e z_t^2/l^2} = \frac{f^4}{2e} \approx 0.184 f^4. \quad (54)$$

Similarly, calculations aided by Mathematica show that the intensity

$$\left| f^2 E_x^{(2)}(\rho, z_t) \right|^2 = f^4 \frac{z_t^2/l^2}{(1 + z_t^2/l^2)^3} \left| L_2 \left( \frac{\rho^2}{2\omega_0^2 (1 + iz_t/l)} \right) \right|^2 \exp \left[ -\frac{\rho^2}{\omega_0^2 (1 + z_t^2/l^2)} \right] \quad (55)$$

has a maximum of

$$\left| f^2 E_x^{(2)} \right|_{max}^2 \approx \frac{4}{e^4} f^4 \approx 0.073 f^4 \quad (56)$$

at

$$\rho_{max} \approx 2\omega_0 z_t/l = 2\sqrt{2} \frac{\omega_0}{f}. \quad (57)$$

At  $z = 0$ , we find that the term  $fE_z^{(1)}$  is finite and the term  $f^2E_x^{(2)}$  is zero, while at  $z = z_t$ , Eqs. (54) and (56) demonstrate that these terms are comparable in magnitude. Therefore, vector corrections must be included in any study that wants to go beyond the scalar paraxial approximation  $E_x^{(0)}$ .

## 7. Conclusion

An angular spectrum analysis of vector Gaussian beam propagation is first presented. Starting with a particular boundary condition, a Gaussian beam is propagated in the positive  $z$ -direction and expressions for the transverse and longitudinal electric fields in the form of power series expansions are obtained. The usual paraxial result represents a dominant term in the expansion of the transverse field, while the foremost term in the longitudinal expansion represents the first non-fundamental Hermite-Gaussian mode. All higher terms, both in the transverse and longitudinal directions, represent non-Gaussian corrections to the paraxial optics. In general, higher-order terms, both transverse and longitudinal, increase as the expansion parameter  $f = (k\omega_0)^{-1}$  increases. Finally, the series expansion is accurate only within the region  $z_t \leq \sqrt{2} \frac{l}{f}$ . To study the beam analytically for  $z \gg z_t$ , alternative approaches such the method of stationary phase must be sought.

## **8. Acknowledgments**

This work was supported by the Defense Advanced Research Projects Agency (DARPA) under Grant DAAG55-98-1-0130, and the National Aeronautics and Space Administration (NASA) under Grant NAG5-5105.

## REFERENCES

1. G. P. Agrawal and D. N. Pattanayak, “Gaussian beam propagation beyond the paraxial approximation,” *J. Opt. Soc. Am.* **69**, 575–578 (1979).
2. M. Born and E. Wolf, *Principles of Optics*, 6th ed. (Cambridge University Press, 1980), p.752.
3. W. H. Carter, “Electromagnetic Field of a Gaussian Beam with an Elliptical Cross Section,” *J. Opt. Soc. Am.* **62**, 1195–1201 (1972).
4. H. Kogelnik, “On the Propagation of Gaussian Beams of Light Through Lenslike Media Including those with a Loss and Gain Variation,” *Appl. Opt.* **4**, 1562–1569 (1965).
5. D. C. O’Shea, *Elements of Modern Optical Design* (Wiley-Interscience, 1985), pp.247–252.
6. A. Yariv, *Optical Electronics in Modern Communications*, 5th ed. (Oxford, 1997), Chapter 2.
7. M. Lax, W. H. Louisell and W. B. McKnight, “From Maxwell to paraxial wave optics,” *Phys. Rev. A* **11**, 1365–1370 (1975).
8. D. R. Rhodes, “On the Stored Energy of Planar Apertures,” *IEEE Transactions on Antennas and Propagation* **AP 14**, 676–683 (1966).
9. D. R. Rhodes, “On a Fundamental Principle in the Theory of Planar Antennas,” *Proceedings of the IEEE* **52**, 1013–1021 (1964).
10. M. L. Schattenburg, C. Chen, P. N. Everett, J. Ferrera, P. Konkola and H. I. Smith, “Sub-100 nm metrology using interferometrically produced fiducials,” *Journal of Vacuum Science & Technology B* **17**, 2692–2697 (1999).
11. Michael H. Lim, Juan Ferrera, K. P. Pipe, and Henry I. Smith, “A holographic phase-shifting interferometer technique to measure in-plane distortion,” *Journal of Vacuum*

- Science & Technology B **17**, 2703–2706 (1999).
12. G. E. Sommargren, “Phase shifting diffraction interferometry for measuring extreme ultraviolet optics,” D. R. Kania and G. D. Kubiak, eds., 108–112, *Extreme Ultraviolet Lithography*, Vol. 4 of OSA Trends in Optics and Photonics (Optical Society of America, Washington, D.C., 1996).
  13. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 5th ed. (Academic Press, 1994), p.978.
  14. I. S. Gradshteyn and I. M. Ryzhik, *ibid.*, p.737:6.631 (1), p.1087:9.220 (2) and p.1062:8.972 (1).
  15. W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in Fortran 77: The Art of Scientific Computing, (Vol. 1 of Fortran Numerical Recipes)*, 2nd ed. (Cambridge University Press, 1999), Chapter 16.

Fig. 1. A plot of  $|E_z^{(1)}|^2$  as a function of  $x$  and  $y$  for  $\omega_0 = 1.3 \mu\text{m}$ ,  $\lambda = 351.1 \text{ nm}$  and  $z = 1 \text{ mm}$ .  $E_z^{(1)}$  represents the first non-fundamental Hermite-Gaussian mode. Note the mirror symmetry of the two humps about the  $x$  axis.

Fig. 2. A plot of the ratio of the vector field intensity,  $|E_x|^2 + |E_z|^2$ , to the scalar intensity,  $|E_x|^2$ , for  $z = 1 \text{ mm}$ ,  $\omega_0 = 1.3 \mu\text{m}$  and  $\lambda = 351.1 \text{ nm}$  ( $f = 0.043$ ). A 1% deviation is registered near the edge. The inset shows the scalar intensity profile  $|E_x|^2$ , sampled over the same  $0.2 \text{ mm} \times 0.2 \text{ mm}$  region.

Fig. 3. A plot of the ratio of the vector field intensity,  $|E_x|^2 + |E_z|^2$ , to the scalar intensity,  $|E_x|^2$ , for  $z = 3.6 \mu\text{m}$ ,  $\omega_0 = 0.2 \mu\text{m}$  and  $\lambda = 351.1 \text{ nm}$  ( $f = 0.28$ ). A  $> 40\%$  deviation is registered near the edge. The inset shows the scalar intensity profile  $|E_x|^2$ , sampled over the same  $5 \mu\text{m} \times 5 \mu\text{m}$  region.

Fig. 4. Plots for  $z = 20 \text{ mm}$ ,  $\omega_0 = 1.3 \mu\text{m}$  and  $\lambda = 351.1 \text{ nm}$  ( $f = 0.043$ ), the solid line representing  $|E_x^{(0)}|^2$ , the dotted line  $|E_x^{(0)} + f^2 E_x^{(2)}|^2$  and the dashed line  $|E_x^{(0)} + f^2 E_x^{(2)} + f^4 E_x^{(4)}|^2$ .

Fig. 5. A schematic showing the various coordinate measures used in deriving the range of applicability  $z_t$  for the series expansion. Beyond the range, paraxial optics breaks down and the series expansion approach becomes invalid.

Fig. 6. Plots for  $\omega_0 = 1.3 \mu\text{m}$ ,  $\lambda = 351.1 \text{ nm}$  and  $z = z_t/2 = 0.5 \text{ mm}$ . (a) The numerically obtained scalar field intensity  $|E_x|^2$ . (b) The thin solid, dotted, dashed, and thick solid lines represent differences from the numerical intensity by  $|E_x^{(0)}|^2$ ,  $|E_x^{(0)} + f^2 E_x^{(2)}|^2$ ,  $|E_x^{(0)} + f^2 E_x^{(2)} + f^4 E_x^{(4)}|^2$  and the spherical result, respectively. Results are normalized to  $|E_x|_{max}^2$ . (c) The thin solid, dotted, dashed, and thick solid lines represent deviations from the numerical phase by  $E_x^{(0)}$ ,  $E_x^{(0)} + f^2 E_x^{(2)}$ ,  $E_x^{(0)} + f^2 E_x^{(2)} + f^4 E_x^{(4)}$  and the spherical result, respectively.

Fig. 7. Plots for  $\omega_0 = 1.3 \mu\text{m}$ ,  $\lambda = 351.1 \text{ nm}$  and  $z = z_t = 1 \text{ mm}$ . (a) The numerically obtained scalar field intensity  $|E_x|^2$ . (b) The thin solid, dotted, dashed, and thick solid lines represent differences from the numerical intensity by  $|E_x^{(0)}|^2$ ,  $|E_x^{(0)} + f^2 E_x^{(2)}|^2$ ,  $|E_x^{(0)} + f^2 E_x^{(2)} + f^4 E_x^{(4)}|^2$  and the spherical result, respectively. Results are normalized to  $|E_x|_{max}^2$ . (c) The thin solid, dotted, dashed, and thick solid lines represent deviations from the numerical phase by  $E_x^{(0)}$ ,  $E_x^{(0)} + f^2 E_x^{(2)}$ ,  $E_x^{(0)} + f^2 E_x^{(2)} + f^4 E_x^{(4)}$  and the spherical result, respectively.

Fig. 8. Plots for  $\omega_0 = 1.3 \mu\text{m}$ ,  $\lambda = 351.1 \text{ nm}$  and  $z = 2z_t = 2 \text{ mm}$ . (a) The numerically obtained scalar field intensity  $|E_x|^2$ . (b) The thin solid, dotted, dashed, and thick solid lines represent differences from the numerical intensity by  $|E_x^{(0)}|^2$ ,  $|E_x^{(0)} + f^2 E_x^{(2)}|^2$ ,  $|E_x^{(0)} + f^2 E_x^{(2)} + f^4 E_x^{(4)}|^2$  and the spherical result, respectively. Results are normalized to  $|E_x|_{max}^2$ . (c) The thin solid, dotted, dashed, and thick solid lines represent deviations from the numerical phase by  $E_x^{(0)}$ ,  $E_x^{(0)} + f^2 E_x^{(2)}$ ,  $E_x^{(0)} + f^2 E_x^{(2)} + f^4 E_x^{(4)}$  and the spherical result, respectively.

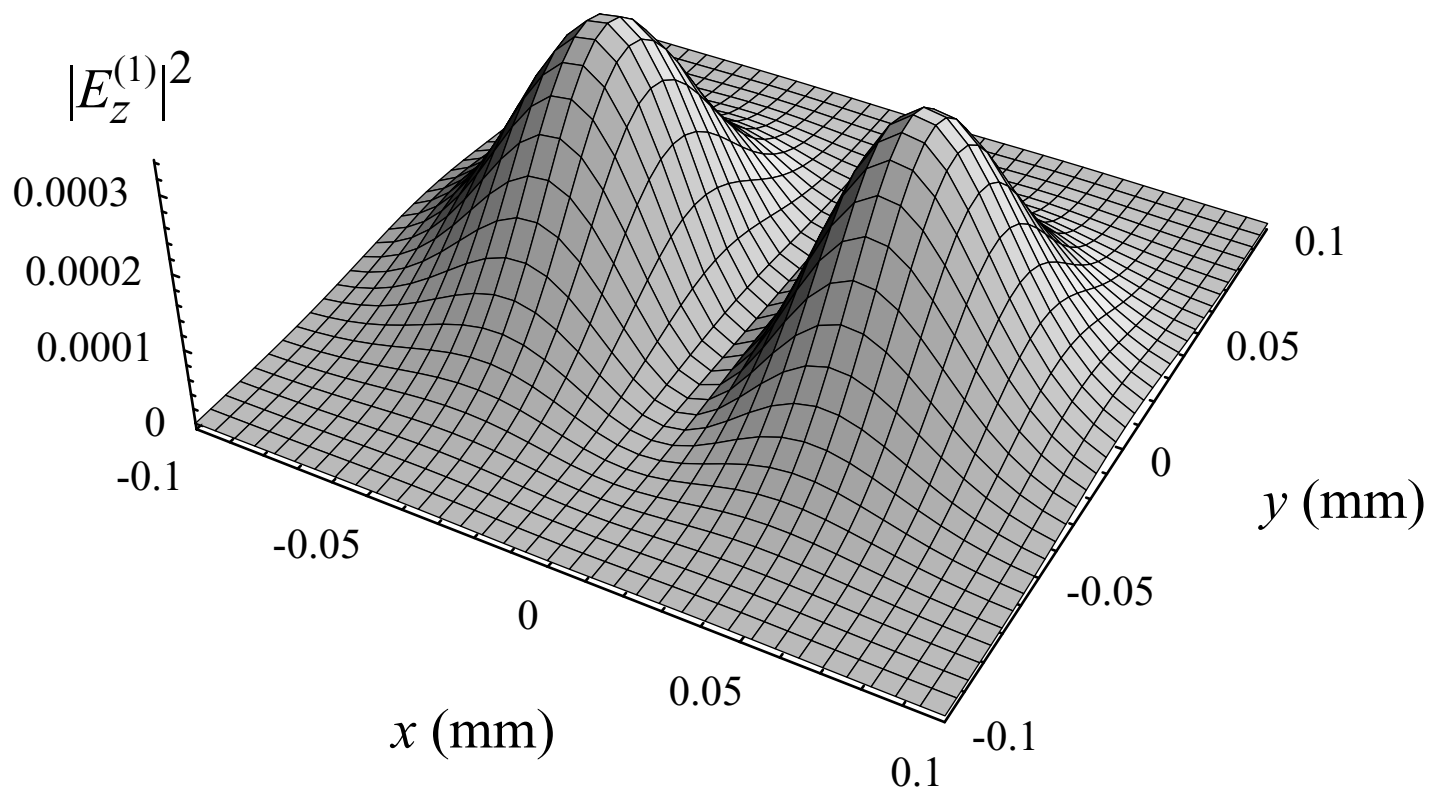


Fig. 1

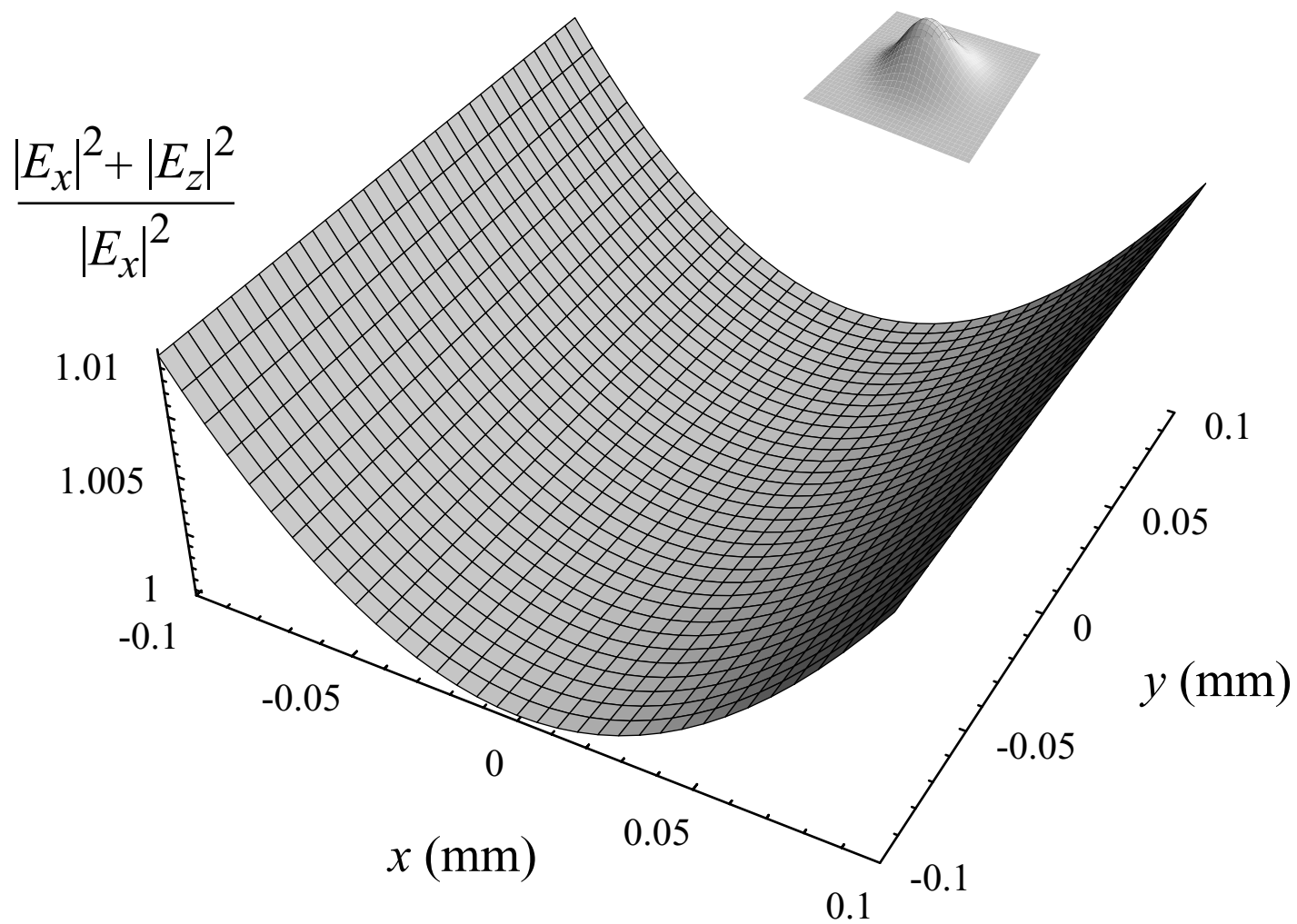


Fig. 2

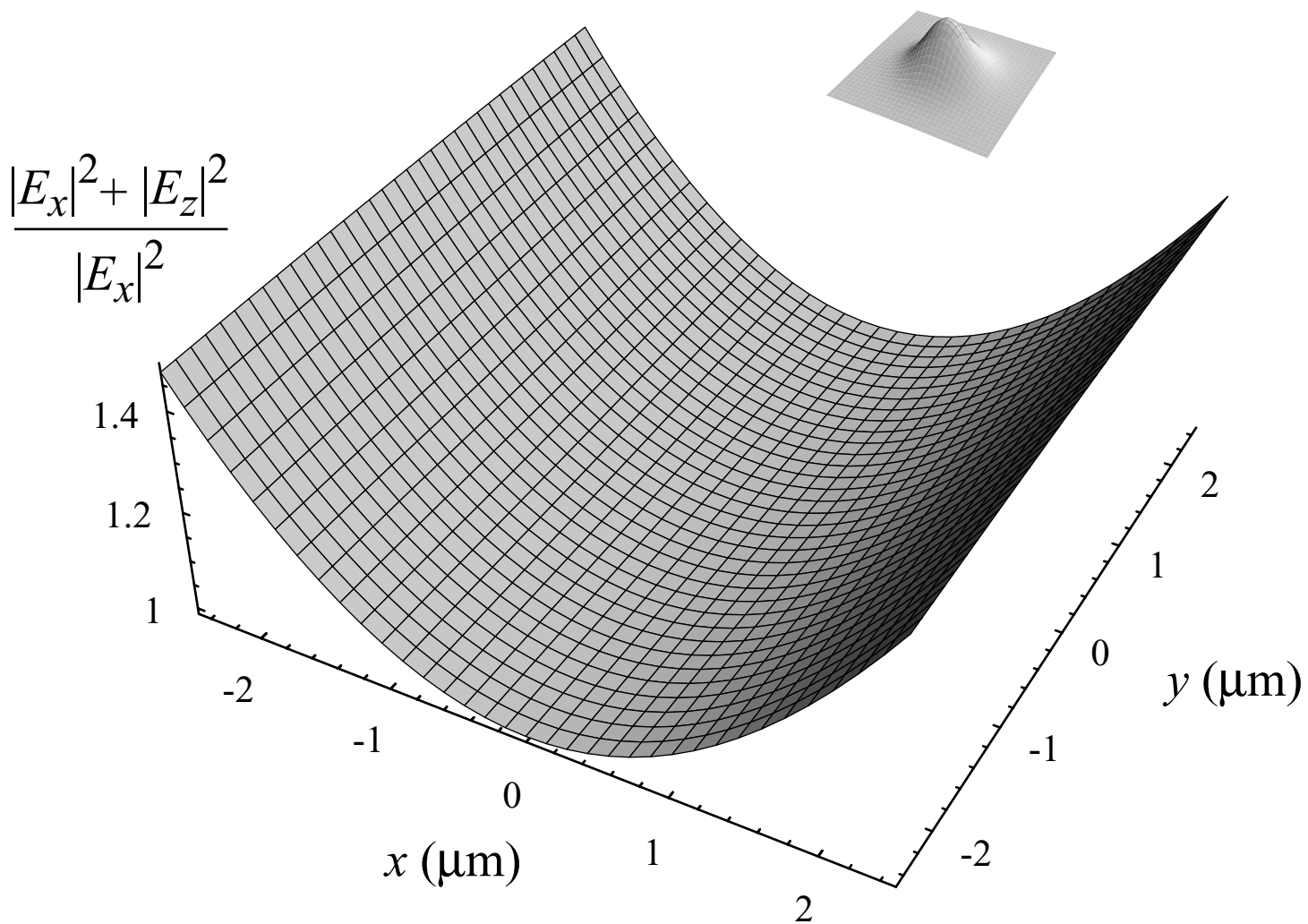


Fig. 3

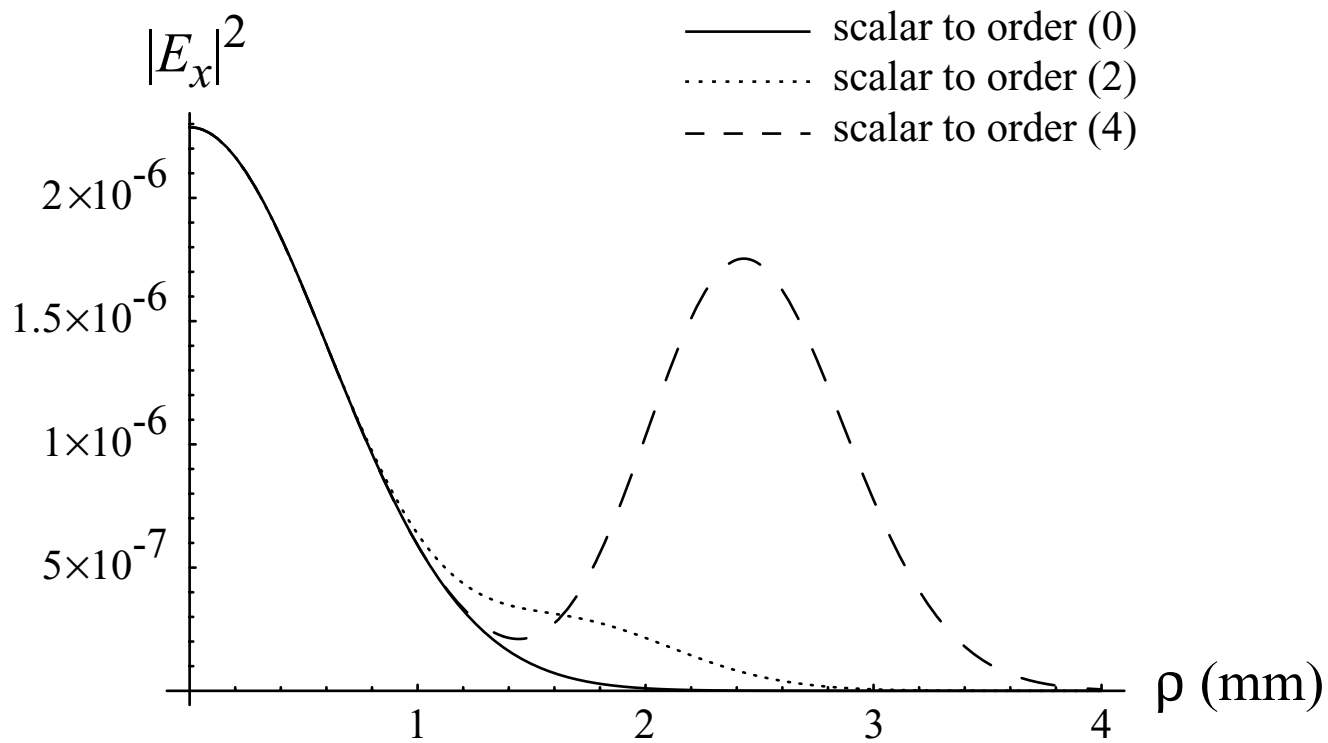


Fig. 4

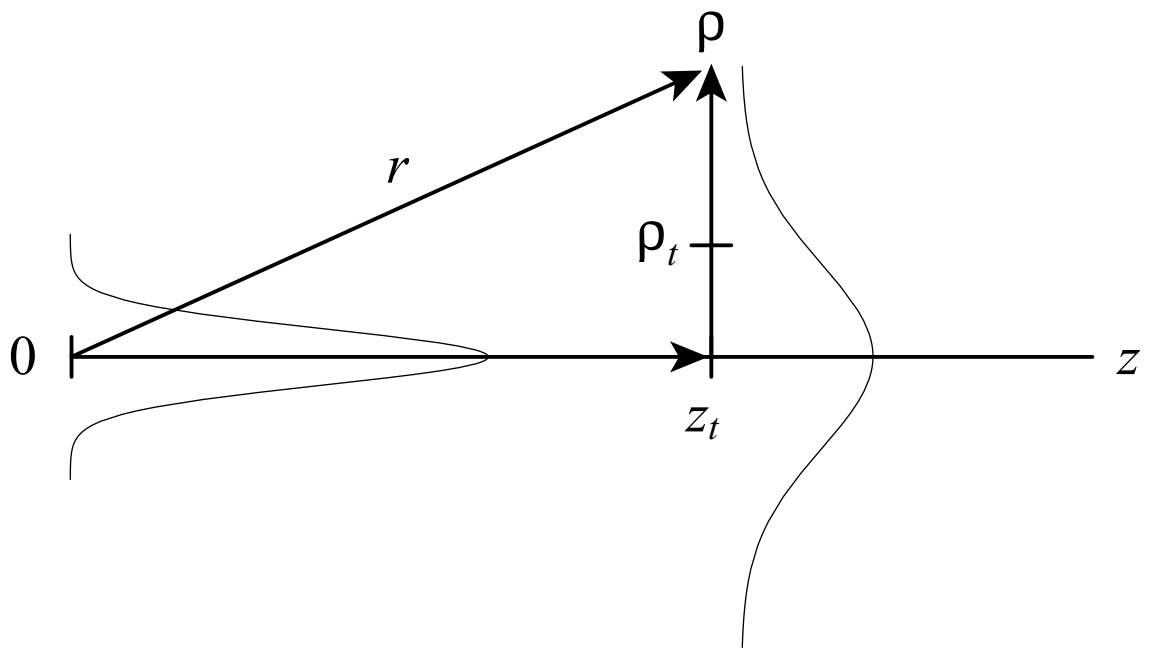


Fig. 5

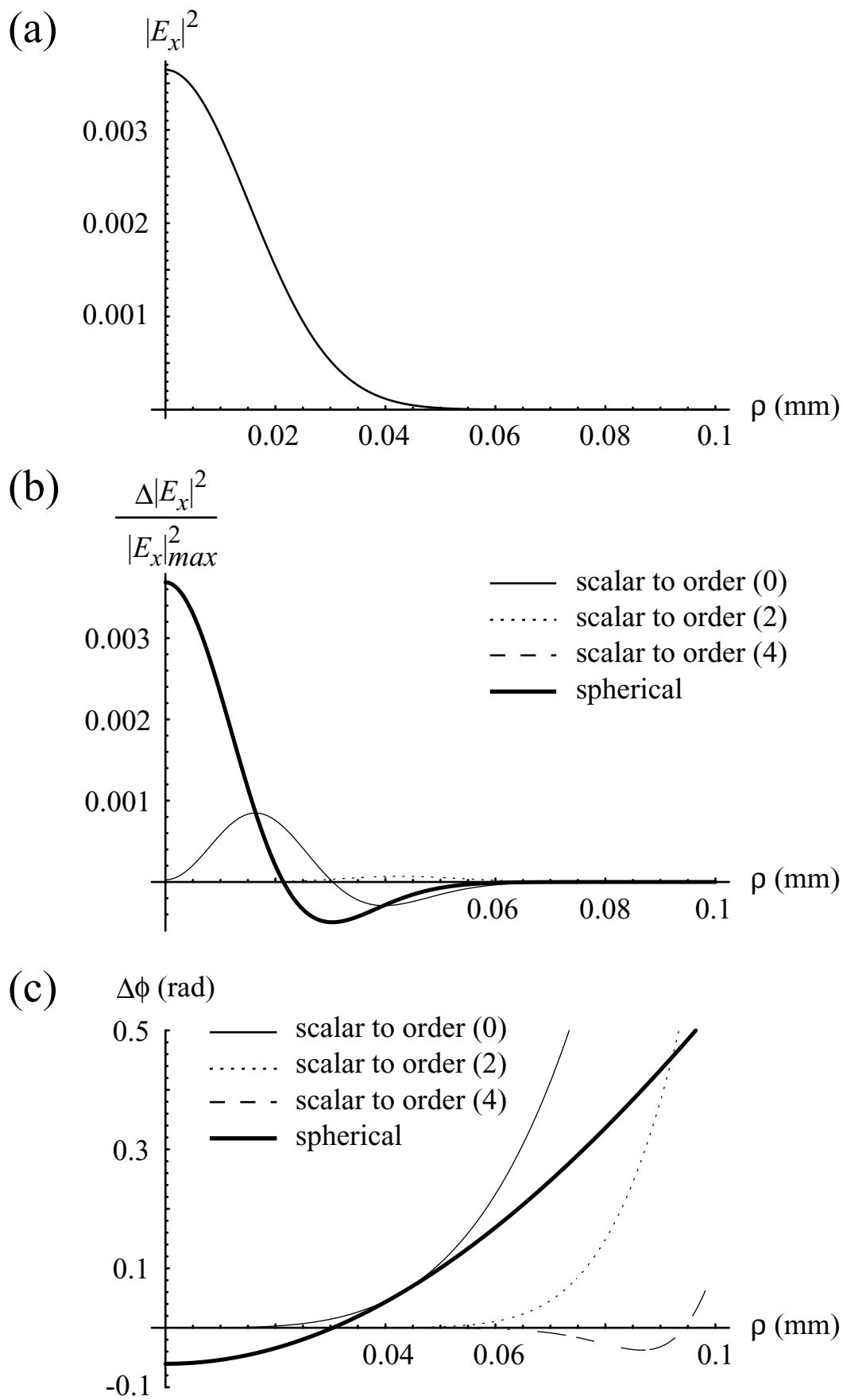


Fig. 6

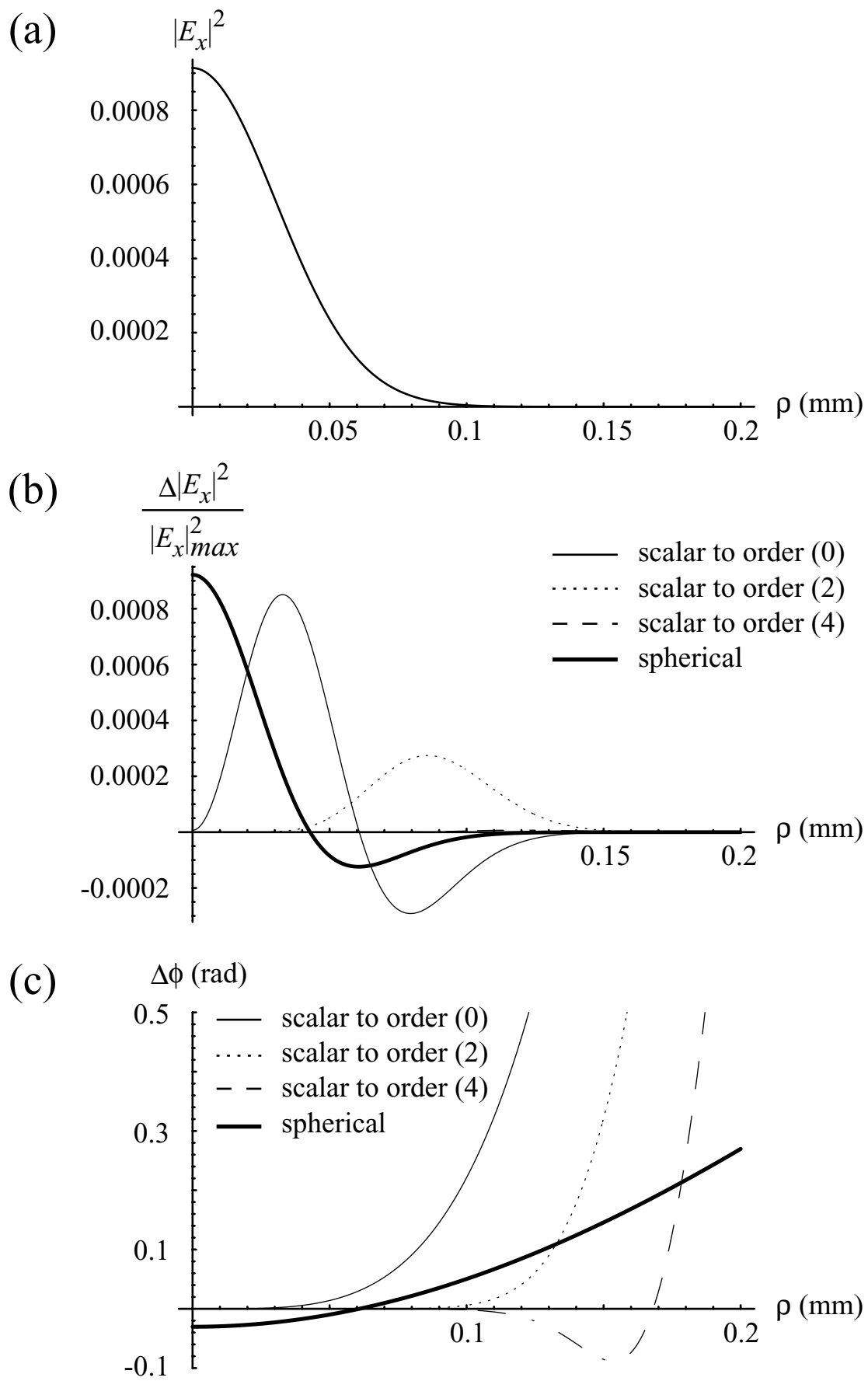


Fig. 7

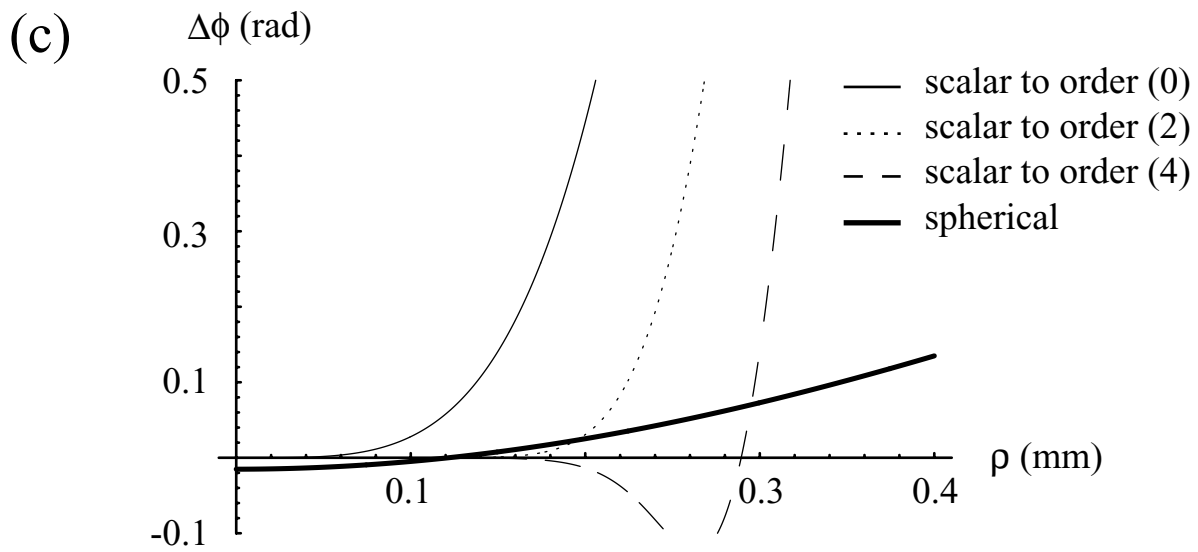
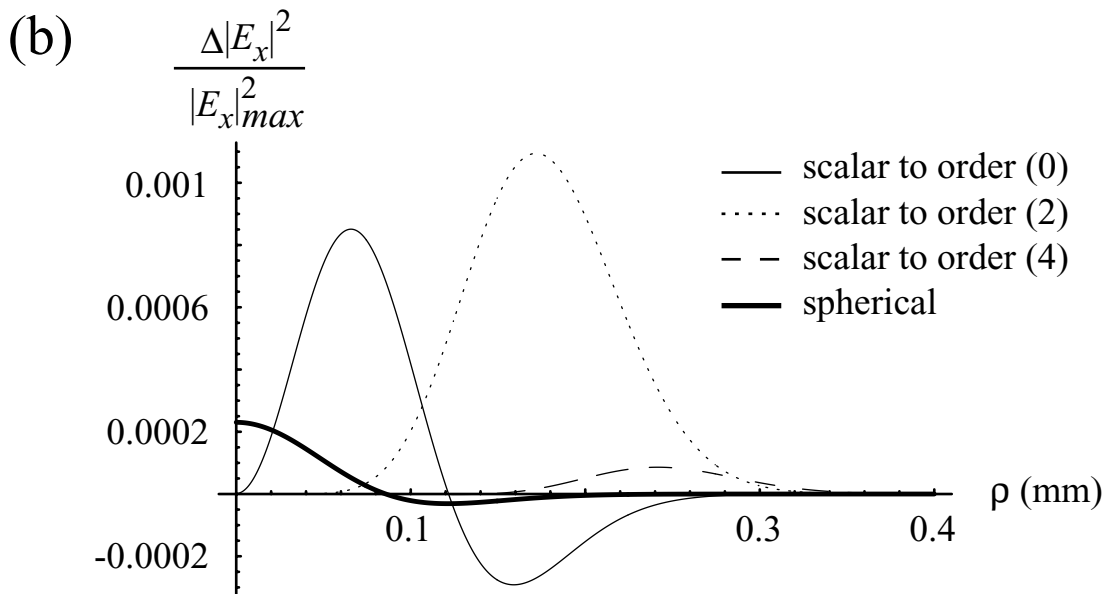
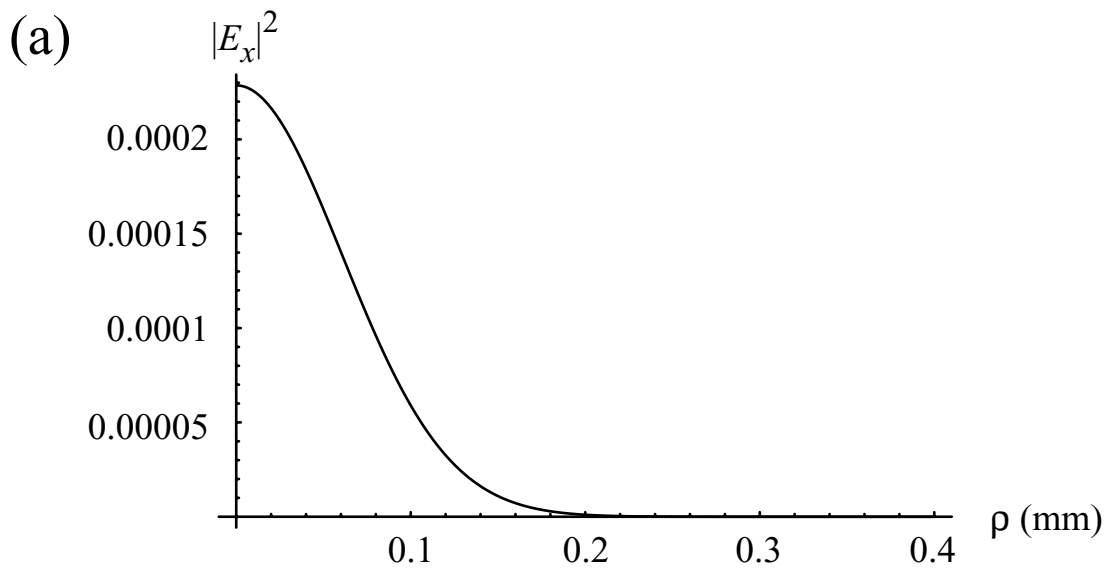


Fig. 8