

FIBRE OPTICS. OPTICAL FIBRES

PACS numbers: 07.79.Fc; 41.81.Qb; 42.81.Dp
DOI: 10.1070/QE2002v032n08ABEH002280

Structure of light waves in a waveguide tapered to subwavelength transverse size

T.I. Kuznetsova, V.S. Lebedev

***Abstract*. The penetration of an *H* wave through a waveguide whose radius varies continuously from about the wavelength size to subwavelength dimensions (of the order of 1/10 wavelength) is calculated. The method for the field analysis is developed and the range of its applicability is studied by considering some numerical examples. The dependence of the transmission coefficient of a tapered waveguide on the radiation wavelength and on the profile steepness of the waveguide is studied. It is shown that a sharper decrease in the waveguide radius provides higher values of the field transmission coefficients.**

Keywords: *near-field optics, tapered waveguide, waveguide transmission coefficient*.

1. Introduction

The propagation of light through systems with a subwavelength aperture has been studied in recent years in a number of papers. The practical interest in this problem is associated with the development of ultrahigh-resolution optical microscopy, based on field localisation in the region of subwavelength dimensions. In the classical theoretical works on this subject [1 – 3], diffraction from a circular aperture was studied in detail. In the first works on the near-field scanning optical microscopy, the 'quasi-point' source of light was a hole in an opaque screen [4]. Extended structures, like micropipettes [5, 6], were also investigated. Tapered fibres [7, 8] turned out to be quite efficient, and it was found that it is expedient to deposit a layer of metal on the side surface of a micropipette or a fibre. Some theoretical works have also been devoted to this subject (see, for example, Refs [9 – 12] and references therein).

Bearing in mind the experimental studies using fibres with metal coatings, we will consider a dielectric waveguide with ideally reflecting metal walls. We will focus our attention on the variation of the waveguide diameter along its optical axis and its effect on the field structure. In analogy with the properties of cylindrical waveguides (see, for example, Ref. [13]), it can be naturally assumed that a tapered waveguide may exhibit the so-called cutoff, when the field channeling is terminated with decreasing the waveguide diameter, and a strong damping of the field is observed along the longitudinal coordinate.

It should be interesting to estimate this damping quantitatively. Such an estimate was performed earlier in Ref. [14], where the local damping factor was used to calculate the amplitude of the transmitted wave. The damping factor was introduced in accordance with the same rules as in a circular cylindrical waveguide; however, this factor was a function of the longitudinal coordinate in a tapered waveguide. The total attenuation was calculated by integrating the damping factor. In such an approach, the longitudinal variations in the waveguide parameters are considered in a too simplified form. The matter is that additional waves appearing in a tapered waveguide considerably alter the structure of the total field and hence cannot be neglected beforehand.

Here, it is appropriate to consider the analogy with the propagation of a light wave through an aperture in a reflecting screen [1]. It was found in Ref. [1] that the appearance of a reflected wave in front of the screen is a significant factor affecting the field leakage through a small aperture. This factor strongly changes the transmission coefficient compared to the value that would be obtained in calculations based on the Huygens principle.

In this work, we develop the theory of a tapered waveguide by taking into account the forward as well as the backward wave. The transformation of the initial field into higher-order waves is also estimated. The method is described as applied to tapered waveguides having a circular cross section. The method developed by us is similar to the so-called cross-section method developed in Refs [15, 16], but takes into account the specific features of the fibres used in a near-field optical microscope, such as small radii, considerable departures from the cylindrical profile, a sharp decrease in the radius at the fibre output up to values much smaller than the wavelength, and a strong reflection at the walls.

2. Formulation of the problem and basic equations

Let us present the initial equations and boundary conditions for the field in a circular tapered waveguide (Fig. 1). Let z be the coordinate along the waveguide axis; ρ is the distance from the waveguide axis; φ is the polar angle; and a is the inner radius of the waveguide that depends on the longitudinal coordinate. The medium in the waveguide is homogeneous and has the permittivity $\varepsilon = \text{const}$. The walls

T.I. Kuznetsova, V.S. Lebedev P.N. Lebedev Physics Institute, Russian Academy of Sciences, Leninskii prosp. 53, 119991 Moscow, Russia

Received 9 January 2002; revision received 31 May 2002
Kvantovaya Elektronika **32** (8) 727 – 737 (2002)
Translated by Ram Wadhwa

of the waveguide are assumed to be ideally reflecting. We confine our analysis to fields whose components do not depend on the angle φ. The time dependence of the components is chosen in the form $\exp(-i\omega t)$, and this time factor can be omitted in the subsequent expressions for the fields. The system of Maxwell's equations for magnetic type waves is reduced to the equation

$$\frac{\partial^2 E}{\partial z^2}+\frac{\partial}{\partial \rho}\left[\frac{1}{\rho}\frac{\partial}{\partial \rho}(\rho E)\right]+\varepsilon\frac{\omega^2}{c^2}E=0, \tag{1}$$

in which the nonzero field components are related by the expressions

$$E_\varphi = E,\quad H_\rho = i\frac{c}{\omega}\frac{\partial E}{\partial z},\quad H_z = -i\frac{c}{\omega}\frac{1}{\rho}\frac{\partial(\rho E)}{\partial \rho}. \tag{2}$$

The boundary condition has the form

$$E_\varphi|_{\rho=a(z)}=0. \tag{3}$$

The main feature of the problem formulated here is that the waveguide boundary is not a cylindrical surface, i.e., $a(z)\neq$ const.

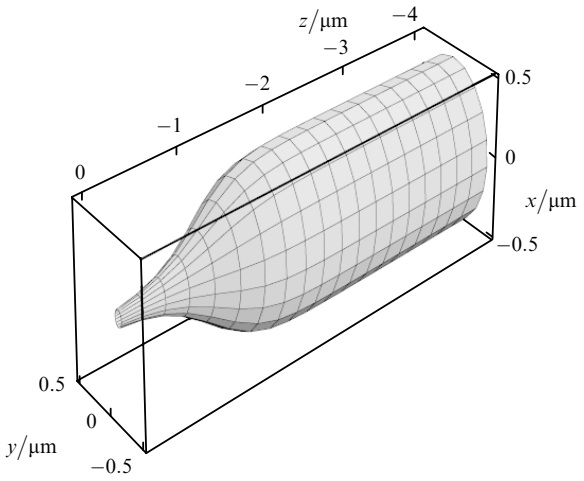

Figure 1. A hypergeometric tapered waveguide with input and output radii $a_{in}=a_{-\infty}=0.5$ μm, $a_{out}=a(z=0)=0.05$ μm, and the waveguide parameter $l=0.2$ μm.

We will construct the solution of the Helmholtz equation (1) for magnetic type waves, using the expansion in the complete system of functions at each cross section. To various values of longitudinal coordinate, their own eigenfunctions will correspond. We will work with functions $J_1(q_n\rho)$, where J_1 is the first-order Bessel function of the first kind, and q_n is a transverse wave number defined by the relations

$$q_n(z)=\frac{\xi_n}{a(z)},\quad J_1(\xi_n)=0\quad (n=1,2,\ldots), \tag{4}$$

the roots ξ_n being labelled in ascending order. Because of the dependence $a(z)$, the eigenfunctions prove to be dependent on z as on a parameter. They form a complete orthonormal system in each cross section z and satisfy the boundary condition (3). The expedience of using such a system of functions was indicated in Refs [15, 16] (see also Ref. [17], Section 9.2).

We present the solution of Eqn (1) in the form

$$E=\sum_{n=1}^{\infty}A_n(z)\Phi_n(\rho,z),\quad \Phi_n(\rho,z)=J_1[q_n(z)\rho]. \tag{5}$$

Note that the field in the expansion (5), is not divided into direct and opposite waves (in contrast to the approach used in Refs [11, 15, 16]). This means that each coefficient A_n in Eqn (5) contains both the direct and opposite waves. Hereafter, we will call the coefficients A_n the mode amplitudes (n = 1, 2, ...). The amplitudes A_n introduced in this way can take positive or negative values. These amplitudes depend on the longitudinal coordinate z. This section as well as the next two are devoted to finding the dependence $A_n(z)$.

Let us expand the derivative $\partial\Phi_n/\partial z$ in functions Φ_m:

$$\frac{\partial\Phi_n}{\partial z}=\sum_{m=1}^{\infty}c_{nm}\Phi_m,\quad c_{nm}=\frac{1}{N}\int_0^a\frac{\partial\Phi_n}{\partial z}\Phi_m\rho\,d\rho,$$
$$N=\int_0^a\Phi_m^2\rho\,d\rho. \tag{6}$$

Taking (6) into account, we expand $\partial^2\Phi_n/\partial z^2$ in functions Φ_m:

$$\frac{\partial^2\Phi_n}{\partial z^2}=\sum_{m=1}^{\infty}\left(\Phi_m\frac{dc_{nm}}{dz}+c_{nm}\sum_{p=1}^{\infty}c_{mp}\Phi_p\right). \tag{7}$$

By substituting Eqns (4), (5) and (7) into Eqn (1) and equating the coefficients of each eigenfunction to zero in the resulting equation, we obtain

$$\frac{d^2A_p}{dz^2}+A_p\left[\varepsilon\frac{\omega^2}{c^2}-q_p^2(z)\right]+2\sum_{n=1}^{\infty}\frac{dA_n}{dz}c_{np}$$
$$+\sum_{n=1}^{\infty}A_n\left(\frac{dc_{np}}{dz}+\sum_{m=1}^{\infty}c_{nm}c_{mp}\right)=0. \tag{8}$$

The expression for the coefficients c_{nm} [see Eqn (6)] has the form

$$c_{nm}=\left(\frac{1}{a}\frac{da}{dz}\right)\tilde{c}_{nm},$$
$$\tilde{c}_{nm}=\begin{cases}-\dfrac{2\xi_n\xi_m}{\xi_n^2-\xi_m^2}\dfrac{J_0(\xi_n)}{J_0(\xi_m)}, & n\neq m,\\ 1, & n=m.\end{cases} \tag{9}$$

Here, $J_0(z)$ is the zero-order Bessel function of the first kind. Because all the coefficients c_{nm} have identical dependences on the coordinate z [see Eqns (9)], it seems most natural to introduce new wave amplitudes $Z_n(z)=$ const $\times\, a(z)A_n(z)$ for further analysis. Be setting the constant equal to $1/a_{-\infty}$, we obtain

$$A_n(z)=\frac{a_{-\infty}}{a(z)}Z_n(z), \tag{10}$$

and the functions Z_n and A_n coincide at the input ($z\to-\infty$).

By using Eqn (10), we can represent the system (8) in the form

$$\frac{\mathrm{d}^2 Z_p}{\mathrm{d}z^2} + Z_p\left[\varepsilon\frac{\omega^2}{c^2} - q_p^2(z)\right] + Z_p\left(\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}z}\right)^2 \sum_{m\neq p}^{\infty} \tilde{c}_{pm}\tilde{c}_{mp} = \psi_p(z), \tag{11}$$

where

$$\psi_p = -2\,\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}z}\sum_{n\neq p}^{\infty}\frac{\mathrm{d}Z_n}{\mathrm{d}z}\,\tilde{c}_{np} - \left[\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}z}\right)\right]\sum_{n\neq p}^{\infty} Z_n\tilde{c}_{np} - \left(\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}z}\right)^2 \sum_{n\neq p}^{\infty} Z_n \sum_{m\neq n,\,p}^{\infty} \tilde{c}_{nm}\tilde{c}_{mp}\,. \tag{12}$$

The transformations carried out in this section do not contain any approximations, and the system of equations (11), (12) is completely equivalent to the initial equation (1). Note that the coupling coefficients of waves with different indices are of the order of $a^{-1}\mathrm{d}a/\mathrm{d}z$, while the self-action of a wave caused by the irregularities of the waveguide profile is proportional to $(a^{-1}\mathrm{d}a/\mathrm{d}z)^2$. This was achieved due to the transformation (10). The transition from amplitudes A_n to new independent variables defined by relation (10) is also useful in that it clearly exhibits an increase in the electromagnetic energy density due to a decrease in the waveguide cross section.

3. Adiabatic approximation

The solution of the system of equations (11), (12) can be obtained using the perturbation theory in the parameter $a^{-1}(\mathrm{d}a/\mathrm{d}z)\,c\,(\sqrt{\varepsilon}\omega)^{-1}$. The smallness of this parameter means that the characteristic longitudinal size of variation in the waveguide radius is much larger than the light wavelength. We assume that the derivative $\mathrm{d}a/\mathrm{d}z$ can be arbitrarily small, i.e., the waveguide profile varies adiabatically slowly. We construct the solution first in the zero-order approximation by neglecting all the terms containing $a^{-1}(\mathrm{d}a/\mathrm{d}z)$ in Eqns (11), (12).

Assuming that the right-hand side of Eqn (11) is zero and omitting the last term in the left-hand side, we obtain

$$\frac{\mathrm{d}^2 Z_p}{\mathrm{d}z^2} + \left[\varepsilon\frac{\omega^2}{c^2} - q_p^2(z)\right]Z_p = 0. \tag{13}$$

This system of equations shows that in the zero-order approximation, all the waves with different subscripts p are independent, as would be the case for a regular waveguide. However, two distinctions from the regular waveguide can already be perceived. The first of these is the dependence of q_p on z, which leads to the interaction and coupling between two waves propagating in the directions $\pm z$, as well as to a variation in the amplitudes of these waves with changing z. The second distinction from a regular waveguide is that the amplitude of a wave with the subscript p is not the function Z_p, but the function A_p [see Eqn (10)], which is inversely proportional to the radius of the waveguide. Thus, even the zero-order approximation takes into account the important characteristics of a tapered waveguide.

Let us choose the dependence of the waveguide radius on the longitudinal coordinate z in the form

$$a(z) = \left[\frac{1+\exp(-z/l)}{(1/a_\infty)^2 + (1/a_{-\infty})^2\exp(-z/l)}\right]^{1/2}, \tag{14}$$

where l is a parameter characterising the region of tapering of the waveguide. This formula describes a gradually tapered fibre, whose radius varies from $a_{\mathrm{in}} \equiv a_{-\infty}$ at the input (i.e., for $z \to -\infty$) to a_∞ for $z \to +\infty$. Consider also the point $z = 0$, because the region $-\infty < z < 0$ correctly approximates a long fibre with a sharpened output end. The radius of the waveguide for $z = 0$ is assumed to be the output radius a_{out} determined by the expression

$$a_{\mathrm{out}} \equiv a(0) = a_\infty\left[\frac{2}{1+(a_\infty/a_{-\infty})^2}\right]^{1/2}. \tag{15}$$

For a significant difference between radii $a_{-\infty}$ and a_∞, the output radius of the fibre is defined by the relation $a_{\mathrm{out}} \approx \sqrt{2}a_\infty$. Note that it follows from (14) that the waveguide radius changes significantly along the longitudinal coordinate within the length $\sim 5l$.

We will call such a waveguide a hypergeometric waveguide, because the field in it is described by hypergeometric functions. Fig. 1 shows the surface of such a waveguide, described by expression (14), with the following parameters: $a_{\mathrm{in}} = 0.5$ μm, $a_{\mathrm{out}} = 0.05$ μm, $l = 0.2$ μm. The dependence (14) chosen by us is relevant because it provides an appropriate shape of a tapered waveguide, and allow us to obtain a solution of the basic equation (13) in an explicit form. Further, we introduce the notation

$$m = \left(\frac{c\xi_1}{\omega\sqrt{\varepsilon}a_{-\infty}}\right)^2, \quad M = \left(\frac{c\xi_1}{\omega\sqrt{\varepsilon}a_\infty}\right)^2. \tag{16}$$

Here, $\xi_1 \approx 3.832$ is the smallest nonzero root of the equation $J_1(\xi) = 0$, which defines the transverse wave number q_1 of the fundamental wave. We will perform most calculations in this work for radiation with a wavelength $\lambda = 2\pi c/\omega = 500$ nm, assuming that $\sqrt{\varepsilon} = 1.5$. Then, taking into account the above values of a_{in} and a_∞, we obtain the relations $m < 1$ and $M > 1$.

Substitution of relations (4), (14) and (16) into Eqn (13) leads to the expression:

$$\frac{\mathrm{d}^2 Z_p}{\mathrm{d}z^2} + \varepsilon\frac{\omega^2}{c^2}\left[1 - \frac{\xi_p^2}{\xi_1^2}\,\frac{M + m\exp(-z/l)}{1+\exp(-z/l)}\right]Z_p = 0. \tag{17}$$

The solution of this equation is expressed in terms of hypergeometric functions. It is presented, for example, in Ref. [18] where the problem of reflection of a particle from a one-dimensional potential barrier was discussed.

Consider first the lowest-order wave. In this case ($p = 1$), Eqn (17) has two linearly independent solutions:

$$U(z) = \exp(zX/l) \times {}_2F_1[-X + \mathrm{i}Y, -X - \mathrm{i}Y; 1 - 2X; -\exp(-z/l)], \tag{18}$$

$$V(z) = \exp(-zX/l) \times {}_2F_1[X + \mathrm{i}Y, X - \mathrm{i}Y; 1 + 2X; -\exp(-z/l)]. \tag{19}$$

Here, ${}_2F_1$ is a hypergeometric function;

$$X = \frac{\omega\sqrt{\varepsilon}\,l}{c}\,(M-1)^{1/2}; \quad Y = \frac{\omega\sqrt{\varepsilon}\,l}{c}\,(1-m)^{1/2}. \tag{20}$$

Note that as $z \to +\infty$, the function $U(z)$ increases, while the function $V(z)$ decreases. As a solution of Eqn (17) describing the amplitude of the fundamental wave ($p = 1$) in the absence of any emitting or amplifying sources at $+\infty$, we choose a function decreasing at $+\infty$:

$$Z_1(z) = V(z). \tag{21}$$

Let us present asymptotic expressions for function (19) for $z \to \pm\infty$. For $z \to \infty$, the function $V(z)$ takes the form

$$V(z) = \exp\left(-\frac{zX}{l}\right), \quad \exp\left(-\frac{z}{l}\right) \ll 1. \tag{22}$$

For $z \to -\infty$, the function $V(z)$ represents two travelling waves propagating towards each other:

$$V(z) = \frac{\Gamma(1+2X)\Gamma(-\mathrm{i}2Y)}{\Gamma(X-\mathrm{i}Y)\Gamma(1+X-\mathrm{i}Y)}\exp\left(\frac{\mathrm{i}zY}{l}\right) + \frac{\Gamma(1+2X)\Gamma(\mathrm{i}2Y)}{\Gamma(X+\mathrm{i}Y)\Gamma(1+X+\mathrm{i}Y)}\exp\left(-\frac{\mathrm{i}zY}{l}\right), \quad \exp\left(\frac{z}{l}\right) \ll 1. \tag{23}$$

Here, $\Gamma(x)$ is the complete gamma function. By introducing the notation

$$A_{\mathrm{in}} = \frac{\Gamma(1+2X)\Gamma(-\mathrm{i}2Y)}{\Gamma(X-\mathrm{i}Y)\Gamma(1+X-\mathrm{i}Y)}, \tag{24}$$

we can write instead of (23)

$$V(z) = A_{\mathrm{in}}\exp\left(\frac{\mathrm{i}zY}{l}\right) + A_{\mathrm{in}}^{*}\exp\left(-\frac{\mathrm{i}zY}{l}\right). \tag{25}$$

The moduli of the direct and reflected waves are equal because the electromagnetic energy flux vanishes in a narrow region of the waveguide [$z \to \infty$; see the asymptotic expression (22)]. For $z \to -\infty$, the field has the form of a standing wave. The field characteristics for intermediate values of z, obtained from expressions (19) and (21), will be discussed in the next section

4. Numerical analysis of the field characteristics

For a more accurate description of the dependence of the field in a waveguide on the longitudinal coordinate z, we turn to the amplitude A_n, which is expressed in terms of the solution obtained above and the radius $a(z)$ (14). For the first wave ($n = 1$), expression (10) has the form

$$A_1(z) = \frac{a_{\mathrm{in}}}{a(z)}\,V(z). \tag{26}$$

Figs 2–4 show the results of calculations for this amplitude normalised to the modulus of the amplitude A_{in} of the incident wave (24), i.e., for $A_1(z)/|A_{\mathrm{in}}|$.

Fig. 2 shows the shape of the normalised amplitude in the range $-3.25 < z/\lambda < 0.25$ of variation in the longitudinal coordinate. The solid curve corresponds to the field amplitude calculated in the zero approximation [see (19), (24), (25)] from the rate of variation in the waveguide profile. These calculations visually demonstrate the main properties of the behaviour of field in a tapered waveguide. They are manifested in the presence of oscillations for negative values of z and a considerable field damping with increasing z in the region where the radius of the waveguide decreases sharply. The same oscillations are preserved further in the negative region $z < -3.25\lambda$, and the damping still occurs at higher positive values of $z > 0.25$ nm. Calculations whose results are presented in Fig. 2 were performed for $\lambda = 500$ nm and $l = 50$ nm. The results of similar calculations of the dependence of the field amplitude on the longitudinal coordinate z are shown by the solid curve in Fig. 3 for a sharper variation ($l = 25$ nm) of the waveguide profile. Apart from the variants shown in Figs 2 and 3, calculations were also made for $l = 75$, 100 and 250 nm. In all these cases, the dependences of the field amplitude on the longitudinal coordinate were similar.

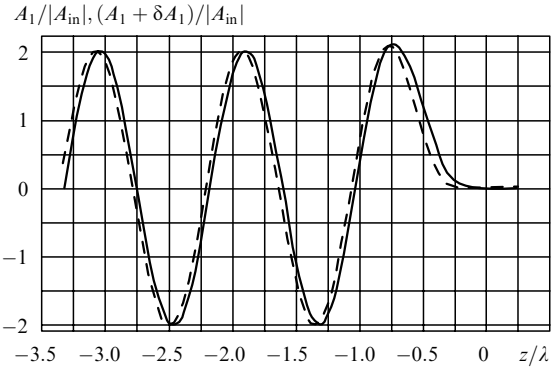

Figure 2. Dependences of the normalised amplitudes $A_1/|A_{\mathrm{in}}|$ and $(A_1 + \delta A_1)|A_{\mathrm{in}}|^{-1}$ of the electric field of a magnetic type fundamental wave ($\lambda = 500$ nm, $\xi_1^{\mathrm{m}} = 3.832$) on the longitudinal coordinate z/λ in a waveguide with parameters $a_{\mathrm{in}} = \lambda/2$, $a_{\infty} = \lambda/20$ and $l = \lambda/10$. The solid curve is the field amplitude calculated in the adiabatic approximation over the rate of variation in the waveguide profile; the dashed curve is the calculation taking into account the perturbation δA_1 (30) determined by the self-action A_1 of the wave.

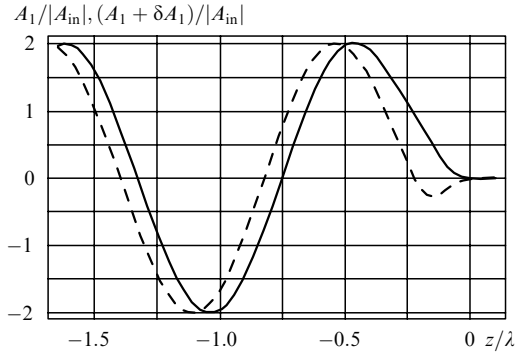

Figure 3. Same as in Fig. 2, but for the waveguide parameter $l = \lambda/20$.

Fig. 4 shows the behaviour of the field of the fundamental magnetic type wave ($\lambda = 500$ nm, $\xi_1^{\mathrm{m}} = 3.832$) in the region of variation of the longitudinal coordinate where it exhibits a sharp damping. Calculations were made for four cases corresponding to different rates of variation in the

waveguide radius in the longitudinal direction. The obtained results show an exceptionally strong dependence of the field amplitude on the parameter l. For example, the normalised field amplitude $A_1(z=0)/|A_{\rm in}|$ at the point $z=0$ is equal to 1.8×10^{-5} and 6.4×10^{-3} for $l=\lambda/10$ and $l=\lambda/20$, respectively. Thus, for a smoother variation in the waveguide profile (larger values of l), the field attenuation is found to be stronger. This is due to the fact that for a smaller length of the transition region, attenuation begins later (i.e., the critical radius is attained for larger values of z). Therefore, a small length of the transition region (i.e., obtuse angles at the output) is always preferable for obtaining large amplitudes of output fields.

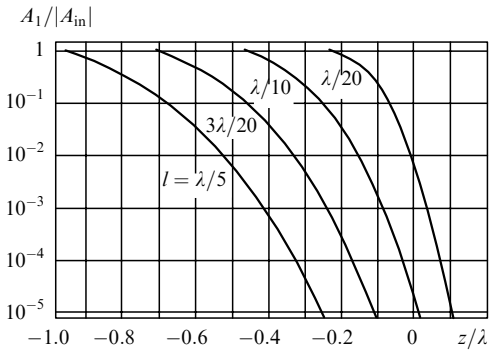

Figure 4. Damping of magnetic type waves ($\lambda=500$ nm, $\xi_1^{\rm m}=3.832$) with increasing the longitudinal coordinate z in the region of a significant decrease in the waveguide radius for $a_{\rm in}=\lambda/2$ and $a_\infty=\lambda/20$. The curves are the normalised field amplitudes $A_1/|A_{\rm in}|$ calculated for different l.

5. Analysis of the applicability of the zero-order approximation

We have obtained above the dependence of the amplitude of the first wave on the longitudinal coordinate $Z_1(z)$ using Eqn (17) for $p=1$. Let us assume that there are no waves with $p\neq1$ in the zero approximation [$Z_p(z)=0$]. Based on this assumption and the obtained dependences $Z_1(z)$, we determine the correction to the solution associated with the terms containing $a^{-1}{\rm d}a/{\rm d}z$ that were omitted in (11) and (12). Assuming that the amplitude Z_1 is larger than all the corrections, we obtain from (11) and (12) inhomogeneous equations for higher-order waves (for Z_p with $p\neq1$) and for the correction δZ_1 to the fundamental wave amplitude. These equations are not given here because of the limited volume of this paper and will be presented elsewhere. Note that they can be solved with the help of the Green function. As the fundamental pair of solutions, we choose functions (18) and (19) for finding δZ_1, while for finding the amplitude Z_p we choose the functions

$$W_p(z)=\exp(zY_p/l)\times{}_2F_1[X_p+Y_p,-X_p+Y_p;2Y_p+1;-\exp(z/l)], \quad (27)$$

$$V_p(z)=\exp(-zY_p/l)\times{}_2F_1[X_p+Y_p,X_p-Y_p;2X_p+1;-\exp(-z/l)], \quad (28)$$

where

$$X_p=\frac{\omega\sqrt{\varepsilon}\,l}{c}\left(\frac{\xi_p^2}{\xi_1^2}M-1\right)^{1/2};\qquad Y_p=\frac{\omega\sqrt{\varepsilon}\,l}{c}\left(\frac{\xi_p^2}{\xi_1^2}m-1\right)^{1/2}. \quad (29)$$

The obvious constraints at $\pm\infty$ are imposed on the solutions Z_p. Moreover, the correction δZ_1 was also subjected to the condition that the wave arriving from $-\infty$ and appearing in the sum $Z_1+\delta Z_1$ had a modulus of the amplitude coinciding with the modulus of the amplitude of a wave travelling in the same direction and appearing in Z_1 (this condition is equivalent to the invariance of the incoming flux). We will present the final formulas for the quantities $A_p=Z_p(a_{-\infty}/a(z))$ and $\delta A_1=\delta Z_1(a_{-\infty}/a(z))$.

5.1 Correction to the fundamental wave amplitude

For the correction δA_1 to the fundamental wave amplitude ($p=1$), the results of calculations can be presented in the following form:

$$\delta A_1=\left\{CV(z)-\frac{l}{2X}\left[U(z)\int_z^\infty V(\zeta)f_1(\zeta){\rm d}\zeta+V(z)\int_{-\infty}^z U(\zeta)f_1(\zeta){\rm d}\zeta\right]\right\}\left(\frac{a_{-\infty}}{a(z)}\right), \quad (30)$$

where

$$f_1(z)=-\left(\frac{1}{a}\frac{{\rm d}a}{{\rm d}z}\right)^2\sigma_1 Z_1;\quad \sigma_1=\sum_{k\neq1}^{\infty}\tilde{c}_{1k}\tilde{c}_{k1}; \quad (31)$$

$$C=\left[1+B^2\left(\frac{\beta\gamma^*-\gamma\beta^*}{2\beta\beta^*}\right)^2\right]^{1/2}-B\,\frac{\beta\gamma^*+\gamma\beta^*}{2\beta\beta^*}-1; \quad (32)$$

$$B=-\frac{l}{2X}\int_{-\infty}^{\infty}V(\zeta)f_1(\zeta){\rm d}\zeta; \quad (33)$$

$$\beta=\frac{\Gamma(2X+1)\Gamma(-{\rm i}2Y)}{\Gamma(X-{\rm i}Y)\Gamma(X+1-{\rm i}Y)};\qquad \gamma=\frac{\Gamma(-2X+1)\Gamma(-{\rm i}2Y)}{\Gamma(-X-{\rm i}Y)\Gamma(-X+1-{\rm i}Y)}. \quad (34)$$

Corrections to the solution obtained in the adiabatic approximation were calculated from expressions (30)–(34) for a waveguide with radii $a_{\rm in}=250$ nm, $a_\infty=25$ nm, and a wide range of the parameter l (250, 100, 75, 50 and 25 nm). As expected, the corrections were found to be very small for large values of the parameter l. The dashed curve in Fig. 2 shows the results of calculation of the amplitude $A_1+\delta A_1$ of the magnetic type fundamental wave for $l=50$ nm ($\lambda=500$ nm, $\xi_1^{\rm m}=3.832$) obtained from expression (30), and their comparison with the amplitude A_1 (solid curve) calculated in the adiabatic approximation [see (14), (19), (26)]. One can see that, when the perturbation δA_1 is taken into account, the standing wave pattern is displaced slightly to the left of the origin, while the damping of the field amplitude begins for slightly lower values of z compared to those obtained in calculations in the zero-order approximation. The results of calculations presented in Fig. 2 show that for $l\gtrsim50$ nm, the introduction of the correction δA_1 (30) leads to slight variations in the wave amplitude A_1

obtained in the zero approximation. Thus, the zero-order (adiabatic) approximation provides a reasonably good quantitative description of the structure of the fundamental wave in the cases considered.

Our aim was to find the values of l for which the correction to the adiabatic approximation becomes noticeable. In this connection, Fig. 3 shows the results of similar calculations of the amplitude $A_1 + \delta A_1$ for $l = 25$ nm (dashed curve). In this case, a comparison with the results of calculations of the field amplitude in the zero-order approximation (solid curve) shows a large shift of the standing wave pattern towards negative values of z. Moreover, an additional peak is formed for $l = 25$ nm in the longitudinal coordinate region of the waveguide where the field begins to attenuate noticeably. A similar extremum is also observed in the first version of the calculations ($l = 50$ nm), but it has a very small amplitude and can hardly be seen in Fig. 2. The value $l = 25$ nm defines the boundary beginning from which the adiabatic approximation gives only a rough estimate for the field in a hypergeometric waveguide. Note, however, that our recent calculations of a cone (an exactly solvable problem) and their comparison with the above results show that the adiabatic approximation provides a reasonable quantitative estimate even for $l = 25$ nm.

5.2 Transformation of a field into higher-order waves

Let us now discuss the transformation of a field into higher-order waves. The expression for the amplitudes A_p ($p \geqslant 2$) has the form

$$A_p = \left(\frac{a_{-\infty}}{a(z)}\right)\left\{-\frac{l}{2X_p}\frac{\Gamma(X_p+Y_p)\Gamma(X_p+Y_p+1)}{\Gamma(2Y_p+1)\Gamma(2X_p)}\right.$$

$$\left.\times\left[W_p\int_z^{\infty} V_p(\zeta)f_p(\zeta)\mathrm{d}\zeta + V_p\int_{-\infty}^{z} W_p(\zeta)f_p(\zeta)\mathrm{d}\zeta\right]\right\}, \qquad (35)$$

where

$$f_p(z) = -2\left(\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}z}\right)\tilde{c}_{1p}\frac{\mathrm{d}Z_1}{\mathrm{d}z}$$

$$-\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}z}\right)\tilde{c}_{1p}Z_1 - \left(\frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}z}\right)^2\sigma_p Z_1; \qquad (36)$$

$$\sigma_p = \sum_{k\neq 1,p}^{\infty}\tilde{c}_{1k}\tilde{c}_{kp}\,. \qquad (37)$$

Expression (35) was used to calculate the dependence of the amplitudes $A_2(z)$ of the magnetic type second wave ($p = 2$, $\xi_2^{\mathrm{m}} = 7.016$) and $A_3(z)$ of the magnetic type third wave ($p = 3$, $\xi_3^{\mathrm{m}} = 10.710$) on the longitudinal coordinate z. These dependences are shown in Figs 5 and 6 for $l = 50$ and 25 nm ($\lambda = 500$ nm). In both cases, the amplitudes A_2 and A_3 are normalised to the modulus of the input amplitude of the fundamental wave ($p = 1$, $\xi_1^{\mathrm{m}} = 3.832$), i.e., to the quantity $|A_{\mathrm{in}}|$.

One can see from Figs 5 and 6 that the amplitudes of the second and third waves are much smaller than the amplitude of the first wave. They differ significantly from zero only over a small part of the z axis (where the waveguide is considerably tapered), and are virtually equal to zero outside this region. These dependences lead to the conclusion that the unperturbed solution for $l \gtrsim 50$ nm (wave

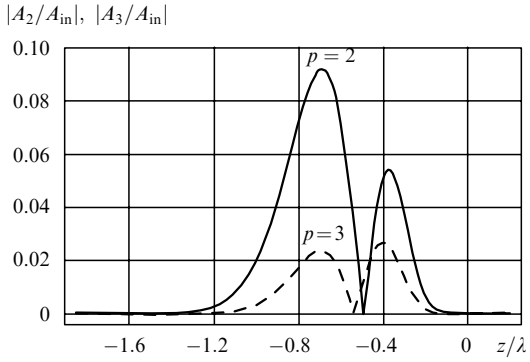

Figure 5. Dependences of the normalised amplitudes A_2 ($p = 2$, $\xi_2^{\mathrm{m}} = 7.016$) and A_3 ($p = 3$, $\xi_3^{\mathrm{m}} = 10.170$) on the longitudinal coordinate z/λ. The solid and dashed curves are calculated from expressions (10) and (35), respectively, for $a_{\mathrm{in}} = \lambda/2$, $a_{\infty} = \lambda/20$ and $l = \lambda/10$.

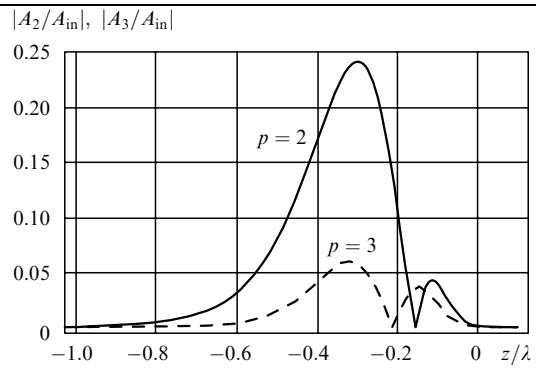

Figure 6. Same as in Fig. 5, but for the waveguide parameter $l = \lambda/20$.

A_1) is a good approximation to the exact solution of the problem.

6. Transmission coefficient of the waveguide

We have considered above the properties of electromagnetic waves in an infinite waveguide. The optical scheme for a finite waveguide looks as follows. The radiation propagates in vacuum at first, and falls on a metal layer containing a narrow channel (waveguide), passes through the finite length of the channel and then emerges in open space where it is transformed in accordance with the wave propagation laws. The result of this transformation is different in the cases of the near-field and far-field zones.

In this work, we are mainly interested in the near-field study. Therefore, we will deal with the energy field densities rather than the energy flux when considering the waveguide transmission coefficient. This is caused by the fact that the dominant role in near-field devices is played by non-propagating waves, which are not connected with the energy flux. We introduce a number of quantities which represent the electric field energy density integrated over the cross section in the appropriate regions of the optical scheme (for definiteness, we consider below the magnetic type waves). For the initial radiation (for $z = -\infty$), we introduce the quantity

$$P_{-\infty} = 2\pi\int_0^{a_{\mathrm{beam}}}\left|E_{\varphi}^{(-\infty)}\right|^2\rho\mathrm{d}\rho, \qquad (38)$$

where a_{beam} is the radius of the input beam; $E_{\varphi}^{(-\infty)}$ is the intensity of the field of a single wave propagating in the

direction of the waveguide [in contrast to the formulas presented below, expression (38) does not contain the reflected waves]. Further, we introduce the integrated quantity

$$P(z_0) = 2\pi \int_0^{a(z_0)} |E_\varphi(\rho, z_0)|^2 \rho \mathrm{d}\rho, \tag{39}$$

which corresponds to the input cross section $z = z_0$ ($z_0 < 0$) of a waveguide of radius $a(z_0)$. Accordingly,

$$P(z = 0) = 2\pi \int_0^{a_{\mathrm{out}}} |E_\varphi(\rho, z = 0)|^2 \rho \mathrm{d}\rho \tag{40}$$

is the integral of the energy density in the output cross section $z = 0$ of a waveguide of radius $a_{\mathrm{out}} \equiv a(0)$. Expression (40) contains the field E_φ in the waveguide calculated neglecting the effect of the interface with free space ($z = 0$). In the expression

$$P_{\mathrm{out}} = 2\pi \int_0^{a_{\mathrm{out}}} \left|E_\varphi^{\mathrm{out}}(\rho, z = 0)\right|^2 \rho \mathrm{d}\rho \tag{41}$$

the quantity $E_\varphi^{\mathrm{out}}(\rho, z = 0)$ is the field $E_\varphi(\rho, z = 0)$ modified after reflection from the output face of the waveguide ($z = 0$).

If we are also interested in the far-field zone, we should introduce another integrated quantity

$$P_\infty = \int_S |E_\varphi(\rho, z)|^2 \mathrm{d}S, \tag{42}$$

where the integral is taken over the surface S of a hemisphere whose radius tends to infinity.

The field transformation at each part of the optical scheme is characterised by the appropriate transfer coefficients

$$\begin{aligned} T_{-\infty\,\mathrm{w}} &= \frac{P(z_0)}{P_{-\infty}}, \quad T_{\mathrm{w}} = \frac{P(0)}{P(z_0)}, \\ T_{\mathrm{w}\,0} &= \frac{P_{\mathrm{out}}}{P(0)}, \quad T_{0\,\infty} = \frac{P_\infty}{P_{\mathrm{out}}}. \end{aligned} \tag{43}$$

Here, $T_{-\infty\,\mathrm{w}}$ is the transfer coefficient from free space to the waveguide, T_{w} is the transmission coefficient of the waveguide itself, $T_{\mathrm{w}0}$ is the transformation coefficient caused by the effect of the waveguide output face, and $T_{0\infty}$ is the transfer coefficient of the free space from the near-field to the far-field zone.

The integrated near-field energy density P_{out} (41) at the waveguide output is expressed in terms of the integrated density $P_{-\infty}$ of the input radiation and the total transfer coefficient T taking into account the three stages of transformation:

$$P_{\mathrm{out}} = P_{-\infty} T, \quad T = T_{-\infty\,\mathrm{w}} T_{\mathrm{w}} T_{\mathrm{w}\,0}. \tag{44}$$

To find the integrated far-field energy density P_∞ (42), the expression for the resultant transmission coefficient $T_{-\infty\,\infty}$ should be supplemented by one more cofactor $T_{0\,\infty}$ corresponding to the transformation of the field in the free space:

$$T_{-\infty\,\infty} = T_{-\infty\,\mathrm{w}} T_{\mathrm{w}} T_{\mathrm{w}\,0} T_{0\,\infty}, \quad P_\infty = P_{\mathrm{out}} T_{0\,\infty}. \tag{45}$$

Let us explain the meaning of the transformation of the field in the free space. Suppose that the field distribution in a certain plane has the form of a spot of radius a ($a \ll \lambda$), the field vanishing for $\rho > a$. The Fourier spectrum of such a spot occupies the region $k_\rho \sim 1/a$ of transverse wave numbers. Only a small fraction of these harmonics corresponds to the propagating waves. Waves with $k_\rho < \omega/c = 2\pi/\lambda$ propagate while the remaining waves are damped with distance from the plane under consideration. Only nondamped waves, which constitute a fraction of the order of $(a/\lambda)^2$ of the entire spatial spectrum, contribute to the energy flux and the far-field intensity. The contribution to the near-field intensity comes from all field harmonics and the damped waves play a dominant role in this case. That is why the integrated density (and not the energy flux) characterises the near-field intensity.

Consider now a situation when the field is inside a waveguide with a subwavelength diameter. The transmission coefficient T_{w} connects the integrated energy densities in the input and output cross sections of the waveguide. This coefficient, being the most interesting parameter in the near-field optics, is the main object of our investigations. One should bear in mind that the coefficient $T_{\mathrm{w}\,0} \approx 1$ in the case under study. The energy flux in the absence of reflection from the output end is equal to zero.

If we take into account the reflection at the waveguide boundary with free space, the field intensity at the output end will be proportional to $|1 + R|^2$, where R is the amplitude reflection coefficient. The electromagnetic energy flux is determined by the quantity $\mathrm{Im}R$. If the reflection coefficient R is small, it has an insignificant effect on the output field intensity. At the same time, this coefficient (or, rather, its imaginary part) completely determines the energy flux. This circumstance also points towards the need to analyse the near field in terms of the energy density rather than the energy flux.

The problems associated with the field transformation at the input and output cross sections will be discussed in detail in Appendix 1 and Appendix 2.

In the following, our calculations will be directed at studying the behaviour of the transmission coefficient T_{w} of the waveguide itself. Consider a semi-infinite ($-\infty < z \leqslant 0$) tapered hypergeometric waveguide with parameters $z_0 \to -\infty$, $a_{\mathrm{in}} \equiv a(z_0) \sim \lambda$ and $a_{\mathrm{out}} \equiv a(0) \ll \lambda$, and determine the spectral dependence of the transmission coefficient T_{w}. We assume that the quantities a_{in}, a_{out} and l are fixed, and will vary the wavelength of the input radiation. The amplitude $A_1(0)$ will give the field amplitude at the output $z = 0$, i.e., $A_{\mathrm{out}} = A_1(0)$.

In this section, we consider the magnetic type waves, as well as the electric type waves. Naturally, a rigorous analysis of the electric type waves should take into account the dependence of the amplitude on the varying radius, as well as the dependence of the boundary condition (which generally connects the radial and longitudinal components of the electric field) on the varying radius. However, for a very slow variation in the profile, we can use the eigenfunctions of a regular waveguide and introduce in them the parametric dependence on z. In the examples considered by us, we used namely such an approach by replacing the eigenvalue $\xi_1^{\mathrm{m}} = 3.832$ of the wave number (used for the magnetic type waves) by the quantity $\xi_1^{\mathrm{e}} = 2.405$, which is a root of the equation $J_0(\xi) = 0$.

Fig. 7 shows the results of calculation of the transmission coefficients T_{w} obtained by us as functions of the

light wavelength λ. These coefficients are connected with the ratio $A_{out}/|A_{in}|$ of the field amplitude at the waveguide output to the modulus of the input amplitude through the relation $T_w = [a_{out}A_{out}/(a_{in}|A_{in}|)]^2$. All calculations were performed for the fundamental ($p = 1$) magnetic and electric type waves in the zero-order approximation in the rate of waveguide profile variation for input and output radii $a_{in} = 250$ nm, $a_{out} = 35$ nm, and the parameter $l = 40$ and 60 nm.

In all the above cases, a qualitative analogy is observed

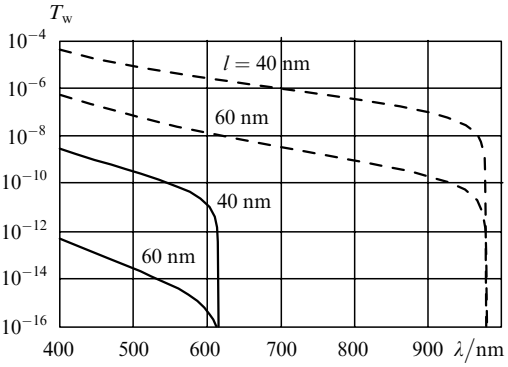

Figure 7. Transmission coefficient T_w of a waveguide as a function of the wavelength λ for $a_{in} = 250$ nm, $a_{out} = 35$ nm and for different values of l. The solid and dashed curves are the results of calculations for the fundamental waves ($p = 1$) of magnetic ($\xi_1^m = 3.832$) and electric ($\xi_1^e = 2.405$) types for $\lambda < \lambda_c$ (corresponding to the cutoff wavelengths $\lambda_c^m = 614.9$ and $\lambda_c^e = 979.8$, respectively).

in the behaviour of the waveguide transmission coefficient T_w as a function of the wavelength. An increase in the wavelength λ leads to a sharp decrease in the coefficient T_w, the decrease becoming exceptionally strong as the wavelength approaches the cutoff value $\lambda_c = 2\pi\sqrt{\varepsilon}\,a_{in}/\xi_1$ (beyond which the waves are no longer propagating even at the input to the waveguide). For the radius a_{in} and permittivity ε of the fibre chosen by us ($\sqrt{\varepsilon} = 1.5$), the cutoff wavelengths λ_c^m and λ_c^e for the fundamental magnetic and electric waves are found to be 614.92 and 979.78 nm, respectively. A comparison of the values of T_w for the magnetic type fundamental wave (for $l = 40$ nm) for different wavelengths shows that T_w decreases by a factor of about 220 as the wavelength λ increases from 400 to 600 nm. A similar comparison for the electric type fundamental wave shows that T_w decreases by a factor of about 400 as the wavelength λ increases from 400 to 900 nm.

At the same time, large quantitative differences in the coefficients T_w for different types of waves are worth noting. For example, for $\lambda = 500$ nm and $l = 40$ nm, the transmission coefficient T_w for an electric type wave is about 2.6×10^4 times higher than the corresponding value for a magnetic type wave. The results shown in Fig. 7 also clearly demonstrate a strong dependence of the output field on the parameter l characterising the steepness of the waveguide profile.

7. Conclusions

The technique developed by us for wave investigations allowed us to establish the spatial structure of the field in a tapered waveguide. Our studies were carried out in the region of geometrical parameters close to the experimental values that are realised by using an optical fibre as a light source in a near-field microscope.

For the conditions chosen by us here, the fundamental wave excited in the waveguide remained dominant along the length of the waveguide. Only an insignificant transfer of the fundamental wave energy to the higher-order waves occurs. Our calculations show a strong dependence of the field characteristics on the length of the transition region where the tapering of waveguide occurs. Shorter transition regions (i.e., steeper waveguide profiles) ensure a higher transmission due to a decrease in the contribution to damping from the transition region. An increase in the steepness of the waveguide profile also leads to another effect: the energy transfer to higher-order waves increases, thus leading to a violation of the applicability of the computational technique used here.

Most calculations were performed under conditions $l \geqslant \lambda/10$ for which the transformation to higher-order modes is insignificant. It is clear qualitatively that it is expedient to use waveguides with a sharper variation of the radius to ensure the best transmission. It was found in this work that for $l \approx \lambda/10$, the transmission T_w of the waveguide attains values $\sim10^{-5} - 10^{-4}$ for electric type waves. Going over to $l \approx \lambda/20$, we find that, according to the estimate obtained on the basis of adiabatic approximation, the transmission of the waveguide increases by an order of magnitude. The validity of this estimate is confirmed by our recent calculations for a cone where an exact solution exists. The results of calculations for a cone are in fair agreement with the results obtained in the adiabatic approximation for a segment of a hypergeometric waveguide whose walls have the appropriate slopes.

Calculations made by us in this work show that the dependence of the transmission coefficient of a tapered waveguide on the wavelength is quite strong. Naturally, short waves are always preferable, and this is qualitatively in full agreement with the properties of regular waveguides. The same is true for the dependence of the transmission coefficient on the transverse structure of the exciting wave: the advantages of waves with small transverse wave numbers are obvious. Therefore, electric and magnetic waves with nonzero azimuthal numbers are of considerable interest. The extension of the approach developed by us in this work to a wider range of problems taking into account the specific nature of these waves is of fundamental importance.

Acknowledgements. This work was supported by the Russian Foundation for Basic Research (Grant No. 00-02-17245) and the programme 'Optics. Laser Physics' of the Ministry of Industry, Science and Technology of the Russian Federation.

Appendix 1. Reflection of waves at the inlet to a waveguide

Consider the incidence of a wave from free space on a circular cylindrical waveguide of radius a. We assume that at the input cross section $z = z_0$ ($z_0 < 0$), the waveguide is transformed into an ideally conducting infinite plane extending from $\rho = a$ to $\rho = \infty$. This picture reflects the fact that the waveguide walls are not thin and are thicker than the inner diameter of the waveguide.

The matching of fields in the free space and in the waveguide depends radically on the transverse structure of

the incident wave. To simplify the calculations, we choose the field in a form close to the waveguide mode, namely, we consider a Bessel beam

$$E_\varphi = AJ_1(\alpha\rho)\left\{\exp\left[\mathrm{i}(z-z_0)\left(\frac{\omega^2}{c^2}-\alpha^2\right)^{1/2}\right]\right.$$
$$\left.-\exp\left[-\mathrm{i}(z-z_0)\left(\frac{\omega^2}{c^2}-\alpha^2\right)^{1/2}\right]\right\}, \quad \text{(A1.1)}$$

where A is the amplitude of the incident wave. This expression contains both the incident and reflected waves, and hence the field is equal to zero at the end in the plane $z = z_0$. The transverse component of the wave vector α is a preset characteristic of the initial wave. We assume that the input Bessel beam is limited at $\rho = a_{\text{beam}}$ (i.e., a_{beam} is the beam radius), and set $J_1(\alpha a_{\text{beam}}) = 0$ for convenience. Thus, $\alpha a_{\text{beam}} = \xi_n$, where ξ_n is one of the zeros of the Bessel function with $n \neq 1$. Because the input radius of the waveguide $a \equiv a(z_0) \ll a_{\text{beam}}$, the inequality $\alpha a \ll 1$ is satisfied.

The field of the type (A1.1) is supplemented by a continuous set of reflected waves with other wave vectors. However, it is convenient to write at first the expression for the field in the waveguide:

$$E_\varphi^{\text{w}} = \sum_{s=1}^{\infty} B_s \Phi_s(\rho)\exp\left[\mathrm{i}(z-z_0)\left(\frac{\omega^2}{c^2}\varepsilon - q_s^2\right)^{1/2}\right], \quad \text{(A1.2)}$$

where B_s are constant coefficients, and

$$\Phi_s(\rho) \equiv J_1(q_s\rho); \quad J_1(q_s a) = 0. \quad \text{(A1.3)}$$

The phase for fields in (A1.2) and in analogous expressions is chosen in the form:

$$\arg\left(\frac{\omega^2}{c^2}\varepsilon - q_s^2\right)^{1/2} = \begin{cases} 0, & q_s^2 < \dfrac{\omega^2}{c^2}\varepsilon, \\ \dfrac{\pi}{2}, & q_s^2 > \dfrac{\omega^2}{c^2}\varepsilon. \end{cases} \quad \text{(A1.4)}$$

Taking expression (A1.2) into account and using the field continuity relation at $z = z_0$, $\rho < a$, we can write the expression for the total field in free space as:

$$E_\varphi^{(0)} = AJ_1(\alpha\rho)\left\{\exp\left[\mathrm{i}(z-z_0)\left(\frac{\omega^2}{c^2}-\alpha^2\right)^{1/2}\right]\right.$$
$$\left.-\exp\left[-\mathrm{i}(z-z_0)\left(\frac{\omega^2}{c^2}-\alpha^2\right)^{1/2}\right]\right\} + \sum_{s=1}^{\infty} B_s g_s(\rho, z), \quad \text{(A1.5)}$$

$$g_s(\rho, z) = \int_0^\infty \tilde{g}_s(\varkappa) J_1(\varkappa\rho)$$
$$\times \exp\left[-\mathrm{i}(z-z_0)\left(\frac{\omega^2}{c^2}-\varkappa^2\right)^{1/2}\right]\varkappa \mathrm{d}\varkappa, \quad \text{(A1.6)}$$

where the sign of the root is chosen in accordance with (A1.4), $\varkappa$ is the transverse wave number in free space, and

$$\tilde{g}_s(\varkappa) = \int_0^a \Phi_s(\rho')J_1(\varkappa\rho')\rho' \mathrm{d}\rho'. \quad \text{(A1.7)}$$

Taking (A1.3) into account, we can calculate the function $\tilde{g}_s(\varkappa)$ in an explicit form:

$$\tilde{g}_s(\varkappa) = \frac{q_s a}{\varkappa^2 - q_s^2} J_1(\varkappa a)J_0(q_s a). \quad \text{(A1.8)}$$

For $\varkappa = q_s$, this expression takes the form

$$\tilde{g}_s(q_s) = \frac{a^2}{2} J_0^2(q_s a). \quad \text{(A1.9)}$$

The radial component of the magnetic field is equal to $\mathrm{i}(c/\omega)\,\partial E_\varphi^{(0)}/\partial z$ in free space and $\mathrm{i}(c/\omega)\,\partial E_\varphi^{\text{w}}/\partial z$ in the waveguide. To ensure the continuity of magnetic field at the aperture, the equality

$$\left.\frac{\partial E_\varphi^{(0)}}{\partial z}\right|_{z=z_0} = \left.\frac{\partial E_\varphi^{\text{w}}}{\partial z}\right|_{z=z_0} \quad \text{(A1.10)}$$

should be satisfied for $z = z_0$ and $\rho < a$.

Taking (A1.2) and (A1.5) into account, equality (A1.10) assumes the form

$$A2\mathrm{i}\left(\frac{\omega^2}{c^2}-\alpha^2\right)^{1/2} J_1(\alpha\rho)$$
$$+\sum_{s=1}^{\infty} B_s \int_0^\infty \left[-\mathrm{i}\left(\frac{\omega^2}{c^2}-\varkappa^2\right)^{1/2}\right]\tilde{g}_s(\varkappa)J_1(\varkappa\rho)\varkappa \mathrm{d}\varkappa$$
$$= \sum_{s=1}^{\infty} B_s \mathrm{i}\left(\frac{\omega^2}{c^2}\varepsilon - q_s^2\right)^{1/2} \Phi_s(\rho). \quad \text{(A1.11)}$$

By multiplying the right- and left-hand sides of (A1.11) by $\Phi_n(\rho)$ and integrating with respect to $\rho \mathrm{d}\rho$, we obtain

$$2\mathrm{i}\left(\frac{\omega^2}{c^2}-\alpha^2\right)^{1/2} A \int_0^a J_1(\alpha\rho)\Phi_n(\rho)\rho \mathrm{d}\rho$$
$$+\sum_{s=1}^{\infty} B_s \int_0^a \Phi_n(\rho)\rho \mathrm{d}\rho \int_0^\infty \left[-\mathrm{i}\left(\frac{\omega^2}{c^2}-\varkappa^2\right)^{1/2}\right] \quad \text{(A1.12)}$$
$$\times \tilde{g}_s(\varkappa)J_1(\varkappa\rho)\varkappa \mathrm{d}\varkappa = \mathrm{i}\left(\frac{\omega^2}{c^2}\varepsilon - q_s^2\right)^{1/2} B_n \int_0^a \Phi_n^2(\rho)\rho \mathrm{d}\rho.$$

Taking (A1.8) into account and passing to dimensionless quantities $x = \varkappa a$ and $\xi_n = q_n a$, we can reduce (A1.12) to the form

$$2\mathrm{i}a\left(\frac{\omega^2}{c^2}-\alpha^2\right)^{1/2} AG(\xi_n, \alpha a) + \sum_{s=1}^{\infty} B_s b_{sn}$$
$$= \mathrm{i}\left(\frac{\omega^2 a^2}{c^2}\varepsilon - \xi_n^2\right)^{1/2} B_n, \quad \text{(A1.13)}$$

where

$$G(\xi_n, \alpha a) = \frac{1}{N_n}\int_0^a J_1(\alpha\rho)J_1(q_n\rho)\rho \mathrm{d}\rho;$$
$$N_n = \int_0^a J_1^2(q_n\rho)\rho \mathrm{d}\rho; \quad \text{(A1.14)}$$

$$b_{sn} = \frac{J_0^2(\xi_n)}{2}\int_0^\infty \left[-\mathrm{i}\left(\frac{\omega^2 a^2}{c^2} - x^2\right)^{1/2}\right] \times G(\xi_n, x)G(\xi_s, x)x\mathrm{d}x; \quad \text{(A1.15)}$$

$$G(\xi_n, x) = \frac{2\xi_n}{x^2 - \xi_n^2}\frac{J_1(x)}{J_0(\xi_n)}. \quad \text{(A1.16)}$$

Recall that $J_1(\xi_n) = 0$, and hence the function $G(\xi_n, x)$ becomes equal to unity for $x = \xi_n$.

The system of equations (A1.13) allows us to determine the amplitudes B_n of all the waves entering the waveguide. We confine ourselves to an approximate solution for the amplitude B_1 of the lowest mode. The justification for using this approximation is the fact that the modules of non-diagonal elements b_{sn} are much smaller than the diagonal elements. Putting $n = 1$ in (A1.13) and assuming that $b_{s1} = 0$ for $s \neq 1$, we obtain the amplitude transmission coefficient upon a transition from free space to the waveguide:

$$D_{-\infty\,\mathrm{w}} \equiv \frac{B_1}{A} = 2\mathrm{i}\left(\frac{\omega^2 a^2}{c^2} - \alpha^2 a^2\right)^{1/2} \times G(\xi_1, \alpha a)\left[b_{11} + \mathrm{i}\left(\frac{\omega^2 a^2}{c^2}\varepsilon - \xi_1^2\right)^{1/2}\right]^{-1}. \quad \text{(A1.17)}$$

By calculating the coefficient b_{11} appearing in (A1.17), we obtain numerical estimates for the amplitude transmission coefficient.

Taking into account the relations $\alpha a \ll 1$ and $\alpha a_{\mathrm{beam}} = \xi_n$, we can represent (A1.17) in the form

$$D_{-\infty\,\mathrm{w}} = \frac{-2\mathrm{i}\xi_n}{\xi_1 J_0(\xi_1)}\left(\frac{\omega a}{c}\right)\left(\frac{a}{a_{\mathrm{beam}}}\right) f\left(\frac{\omega a}{c}\right),$$
$$f\left(\frac{\omega a}{c}\right) = \left[b_{11} + \mathrm{i}\left(\frac{\omega^2 a^2}{c^2}\varepsilon - \xi_1^2\right)^{1/2}\right]^{-1}. \quad \text{(A1.18)}$$

The energy coefficient $T_{-\infty\,\mathrm{w}}$ of transfer from free space to the waveguide [see Eqn (43)] is related with the coefficient $D_{-\infty\,\mathrm{w}}$ by the expression

$$T_{-\infty\,\mathrm{w}} = \left[\int_0^a J_1^2(q_1\rho)\rho\mathrm{d}\rho \Big/ \int_0^{a_{\mathrm{beam}}} J_1^2(\alpha\rho)\rho\mathrm{d}\rho\right]\left|\frac{B_1}{A}\right|^2 = \frac{J_0^2(\xi_1)}{J_0^2(\xi_n)}\left|\frac{B_1}{A}\right|^2\frac{a^2}{a_{\mathrm{beam}}^2}. \quad \text{(A1.19)}$$

By substituting (A1.18) into (A1.19), we obtain

$$T_{-\infty\,\mathrm{w}} = \eta_n\left(\frac{\omega}{c}\right)^2\left(\frac{a^6}{a_{\mathrm{beam}}^4}\right)\mathscr{F}\left(\frac{\omega a}{c}\right), \quad \text{(A1.20)}$$

where $\eta_n = \{2\xi_n/[\xi_1 J_0(\xi_n)]\}^2$ are constant coefficients, and $\mathscr{F}(\omega a/c) = |f(\omega a/c)|^2$ is a dimensionless function. Note that formula (A1.20) contains the radius $a(z_0)$ which virtually coincides with the radius $a_{\mathrm{in}} \equiv a(-\infty)$ if the input cross section $z = z_0 < 0$ of the hypergeometric waveguide is chosen at a fairly gently sloping (cylindrical) region. Thus, expressions (A1.15)–(A1.20) determine the behaviour and magnitude of the transfer coefficient $T_{-\infty\,\mathrm{w}}$ from free space to the waveguide for different values of the parameters a/a_{beam}, ε and $\omega a/c$.

Appendix 2. Reflection of waves from the output of a waveguide

Consider now the distortion of a damped wave in a cylindrical waveguide caused by the presence of the interface with the free space. We choose the waveguide mode in the form

$$E_\varphi = A_1\Phi_1(\rho)\exp\left[-z\left(q_1^2 - \frac{\omega^2}{c^2}\varepsilon\right)^{1/2}\right], \quad \text{(A2.1)}$$

which corresponds to an infinite waveguide (A_1 is the amplitude of the lowest mode). We assume that the waveguide is truncated at the cross section $z = 0$ and is transformed into an ideally conducting plane extending from $\rho = a$ (i.e., $a = a_{\mathrm{out}}$) to infinity (in the same way as it is assumed for the input end). The field modified due to the interface effect has the form

$$E_\varphi^{\mathrm{w}} = A_1\Phi_1(\rho)\exp\left[-z\left(q_1^2 - \frac{\omega^2}{c^2}\varepsilon\right)^{1/2}\right] + \sum_{s=1}^\infty B_s\Phi_s(\rho)\exp\left[z\left(q_s^2 - \frac{\omega^2}{c^2}\varepsilon\right)^{1/2}\right]. \quad \text{(A2.2)}$$

The field in free space is represented in the form of a Fourier–Bessel integral. Note that for $z = 0$ and $\rho < a$, it should coincide with the field in the waveguide, while for $z = 0$ and $\rho > a$, it should vanish. This field [taking into account Eqn (1)] should have the form

$$E_\varphi^{(0)} = A_1 g_1(\rho, z) + \sum_{s=1}^\infty B_s g_s(\rho, z), \quad \text{(A2.3)}$$

$$g_s(\rho, z) = \int_0^\infty \tilde{g}_s(\varkappa)J_1(\varkappa\rho) \times \exp\left[\mathrm{i}z\left(\frac{\omega^2}{c^2} - \varkappa^2\right)^{1/2}\right]\varkappa\mathrm{d}\varkappa. \quad \text{(A2.4)}$$

Note that for the problems on reflection from the input and output ends of the waveguide, the functions g_s differ only in their dependence on z, while their dependence on ρ is the same [see Eqns (A1.6) and (A2.4)]. Therefore, the functions $\tilde{g}_s(\varkappa)$ [see Eqn (A1.7)] are identical in both cases. To ensure the continuity of the magnetic field, the equality

$$\left.\frac{\partial E_\varphi^{(0)}}{\partial z}\right|_{z=0} = \left.\frac{\partial E_\varphi^{\mathrm{w}}}{\partial z}\right|_{z=0} \quad \text{(A2.5)}$$

should be satisfied for $\rho < a$.

Taking Eqns (A2.1)–(A2.3) into account, this equality takes the form

$$\sum_{s=1}^\infty\left(q_s^2 - \frac{\omega^2}{c^2}\varepsilon\right)^{1/2}(B_s - \delta_{s1}A_1)\Phi_s(\rho) = \sum_{s=1}^\infty(B_s + \delta_{s1}A_1) \times \int_0^\infty\left[\mathrm{i}\left(\frac{\omega^2}{c^2} - \varkappa^2\right)^{1/2}\right]\tilde{g}_s(\varkappa)J_1(\varkappa\rho)\varkappa\mathrm{d}\varkappa, \quad \text{(A2.6)}$$

where δ_{s1} is the Kronecker delta. By multiplying both sides of the above equation by $\Phi_n(\rho)$ and integrating with respect to $\rho\mathrm{d}\rho$ within the limits $0 \leqslant \rho \leqslant a$, we obtain

$$(B_n - \delta_{n1}A_1)\left(q_n^2 - \frac{\omega^2}{c^2}\,\varepsilon\right)^{1/2}\int_0^a \Phi_n^2(\rho)\rho\mathrm{d}\rho$$

$$= \sum_{s=1}^{\infty}(B_s + \delta_{s1}A_1)\int_0^{\infty}\mathrm{i}\left(\frac{\omega^2}{c^2} - \varkappa^2\right)^{1/2}\tilde{g}_s(\varkappa)\tilde{g}_n(\varkappa)\varkappa\mathrm{d}\varkappa. \quad \text{(A2.7)}$$

Taking into account expression (A1.8) for $\tilde{g}_s(\varkappa)$ and passing to dimensionless variables $x = \varkappa a$ and $\xi_n = q_n a$, we represent (A2.7) in the form

$$(B_n - \delta_{n1}A_1)\left(\xi_n^2 - \frac{\omega^2 a^2}{c^2}\,\varepsilon\right)^{1/2} = -\sum_{s=1}^{\infty}(B_s + \delta_{s1}A_1)b_{ns}, \quad \text{(A2.8)}$$

where b_{sn} is defined by expressions (A1.15) and (A1.16). The system of equations (A2.8) allows us to determine the amplitudes B_n of all the waves reflected backwards in the waveguide. For simplicity, we confine ourselves to the approximate solution for B_1. As in Appendix 1, the solution is constructed using the fact that the nondiagonal elements b_{sn} are much smaller than the diagonal elements. Putting $n = 1$ in (A2.8) and assuming that $b_{s1} = 0$ for $s \neq 1$, we obtain

$$(B_1 - A_1)\left(\xi_1^2 - \frac{\omega^2 a^2}{c^2}\,\varepsilon\right)^{1/2} = -(B_1 + A_1)b_{11}, \quad \text{(A2.9)}$$

whence

$$R \equiv \frac{B_1}{A_1} = \frac{1-\mu}{1+\mu}, \quad \mu = b_{11}\left(\xi_1^2 - \frac{\omega^2 a^2}{c^2}\,\varepsilon\right)^{-1/2}. \quad \text{(A2.10)}$$

The quantity R is the amplitude coefficient of reflection from the open end of a waveguide of radius $a \equiv a_{\mathrm{out}}$. By using the fact that the relations

$$\left|\mathrm{Re}\, b_{11}\left(\xi_1^2 - \frac{\omega^2 a^2}{c^2}\right)^{-1/2} - 1\right| \ll 1,$$
$$\mathrm{Im}\, b_{11} \approx -\frac{1}{15\xi_1^2}\,\frac{\omega^5 a^5}{c^5} \quad \text{(A2.11)}$$

are valid for $\omega a/c \ll 1$, we obtain

$$\mathrm{Re}R = \frac{1}{2}\left[1 - \mathrm{Re}\, b_{11}\left(\xi_1^2 - \frac{\omega^2 a^2}{c^2}\right)^{-1/2}\right] - \frac{1}{2}\,\mathrm{Re}\, b_{11}\,\frac{\omega^2 a^2}{c^2}(\varepsilon - 1)\left(\xi_1^2 - \frac{\omega^2 a^2}{c^2}\,\varepsilon\right)^{-3/2}, \quad \text{(A2.12)}$$

$$\mathrm{Im}R = \frac{\omega^5 a^5}{c^5}\left[30\xi_1^2\left(\xi_1^2 - \frac{\omega^2 a^2}{c^2}\,\varepsilon\right)^{1/2}\right]^{-1}. \quad \text{(A2.13)}$$

Specific calculations performed by expressions (A2.11)–(A2.13) show that for $a \equiv a_{\mathrm{out}} \ll \lambda$, the real part $\mathrm{Re}R$ of the reflection coefficient is small and is $\sim 0.04 - O(a_{\mathrm{out}}^2/\lambda^2)$, while its imaginary part $\mathrm{Im}R$ is an even smaller quantity. Thus, the transfer coefficient $T_{\mathrm{w}0}$ (see Section 6) upon a transition from the waveguide of subwavelength cross section to the near-field zone of the free space is close to unity to a high degree of accuracy. This means that the field at the output of a finite waveguide can be calculated by neglecting reflection at the exit aperture.

Note that, by using expression (A2.13), we can obtain the expression for the coefficient $T_{0\infty}$ of energy transfer of free space from near field to far field:

$$T_{0\infty} \approx \frac{2}{15\xi_1^2}\,\frac{\omega^4 a^4}{c^4}. \quad \text{(A2.14)}$$

Thus, for $a_{\mathrm{out}}/\lambda \ll 1$, we have obtained the coefficient $T_{0\infty}$, which is proportional to $(a_{\mathrm{out}}/\lambda)^4$.

References

1. Bethe H.A. *Phys. Rev.*, **66**, 163 (1944).
2. Bouwkamp C.J. *Rep. Prog. Phys.*, **17**, 35 (1954).
3. Marchand E.W., Wolf E. *J. Opt. Soc. Am.*, **59**, 79 (1969).
4. Pohl D.W., Denk W., Lanz M. *Appl. Phys. Lett.*, **44**, 651 (1984).
5. Kopelman R., Tan W., in *Microscopy and Spectroscopy Images of the Chemical State* (New York, Basel, Hon Kong: Marcel Dekker, 1993).
6. Harootunian A., Betzig E., Isaacson M.S., Lewis A. *Appl. Phys. Lett.*, **49**, 674 (1986).
7. Meixner A.J., Zeisel D., Bopp M.A., Tarrach G. *Opt. Eng.*, **34**, 2324 (1995).
8. Tang M., Cai S.M., Liu Z.F. *Opt. Commun.*, **146**, 21 (1998).
9. Klimov V.V., Letokhov V.S. *Opt. Commun.*, **106**, 151 (1994).
10. Zuev V.S., Kuznetsova T.I. *Kvantovaya Electron.*, **24**, 462 (1997) [*Quantum Electron.*, **27**, 450 (1997)].
11. Knoll B., Keilmann F. *Opt. Commun.*, **162**, 177 (1999).
12. Klimov V.V., Perventsev Ya.A. *Kvantovaya Electron.*, **29**, 9 (1999) [*Quantum Electron.*, **29**, 847 (1999)].
13. Vainshtein L.A. *Elektromagnitnye volny* (Electromagnetic waves) (Moscow: Radio i Svyaz', 1988).
14. Palanker D.V., Simanovskii D.M., Huie P., Smith T.I. *J. Appl. Phys.*, **88**, 6808 (2000).
15. Katsenelenbaum B.Z. *Teoriya neregulyarnykh volnovodov s medlenno menyayushchimisya parametrami* (Theory of Irregular Waveguides with Slowly Varying Parameters) (Moscow: Izd. Akad. Nauk SSSR, 1961).
16. Shevchenko V.V. *Plavnye perekhody v otkrytytkh volnovodakh* (Continuous Transitions in Open Waveguides) (Moscow: Nauka, 1969).
17. Marcuse D. *Light Transmission Optics* (New York: Van Nostrand Reinhold Company, 1972; Moscow: Mir, 1974).
18. Landau L.D., Lifshitz E.M. *Quantum Mechanics: Non-Relativistic Theory* (Oxford, New York: Pergamon Press, 1977; Moscow: Nauka, 1974).