PHYSICAL REVIEW A VOLUME 11, NUMBER 4 APRIL 1975

From Maxwell to paraxial wave optics

Melvin Lax
Department of Physics, City College of New York, New York, New York 10031 and Bell Telephone Laboratories, Murray Hill, New Jersey 07974*

William H. Louisell† and William B. McKnight
Applied Physics Branch, US Army Missile Command, Redstone Arsenal, Alabama 35809
(Received 22 October 1974)

In this paper we are concerned with the propagation of a light beam through an inhomogeneous, isotropic medium with a possibly nonlinear index of refraction. The customary paraxial approximations of neglecting grad div$\mathcal{E}$ and seeking a plane-polarized solution are shown to be incompatible with the exact Maxwell equations. By starting from Maxwell's equations, and scaling transverse and longitudinal distances by the beam waist w_0 and diffraction length l, respectively, an expansion procedure in powers of w_0/l is developed. The exact equations obeyed by the zeroth-order fields are not Maxwell's equations but the customary paraxial approximation to Maxwell's equations. Equations for the first-, second-, and third-order fields are developed. The first-order field is found to be a longitudinal field. It is solved for explicitly in terms of the zeroth-order field which is transverse. Thus a precise knowledge of the meaning and accuracy of paraxial wave optics is obtained.

I. INTRODUCTION

In the study of the propagation of light in an inhomogeneous isotropic medium with variable (nonlinear) index of refraction[1] as well as in the study of modes in spherical laser resonators,[2] several approximations are made which lead to an apparent paradox. The first assumption is that the electric field is plane polarized in the x direction, for example. As we shall show below, it then follows from the exact Maxwell equations that the electric field must then be independent of the x coordinate. Then a paraxial approximation is made and the resulting equations are solved. Gaussian solutions are found in the transverse direction in spite of the fact that $\partial\mathcal{E}_x/\partial x$ must be zero from the exact equations.

In this paper we analyze the paraxial approximation in order to resolve this apparent paradox and present a systematic procedure for obtaining corrections when they are needed. The resulting equations have been solved numerically in an amplifying medium to display the combined effects of focusing (defocusing), gain, and diffraction. These numerical results will be presented in a separate paper. For simplicity we shall consider only monochromatic waves of frequency ω.

II. AN APPARENT PARADOX

The Maxwell equations for the complex fields $\vec{\mathcal{E}}$ and $\vec{\mathrm{H}}$ for a monochromatic wave varying as $e^{-i\omega t}$ are

$$\mathrm{curl}\vec{\mathcal{E}} = i\omega\mu_0\vec{\mathrm{H}}, \tag{2.1}$$

$$\mathrm{curl}\vec{\mathrm{H}} = -i\omega\epsilon_0\kappa\vec{\mathcal{E}}, \tag{2.2}$$

$$\mathrm{div}\,\epsilon\vec{\mathcal{E}} = \rho, \tag{2.3}$$

$$\mathrm{div}\vec{\mathrm{H}} = 0, \tag{2.4}$$

where

$$\kappa \equiv (\epsilon/\epsilon_0) + i(\sigma/\omega\epsilon_0). \tag{2.5}$$

This definition for κ ensures that the right-hand side of (2.2) includes conduction as well as displacement current. Here ϵ contains a linear (nonresonant) contribution due to the host background as well as a local nonlinear (resonant) contribution which has a real and imaginary part in general. From (2.1) and (2.2), it follows that

$$\mathrm{curl}\,\mathrm{curl}\vec{\mathcal{E}} = (\omega/c)^2\kappa\vec{\mathcal{E}}. \tag{2.6}$$

From (2.2) or (2.6) it follows that

$$\mathrm{div}\,\kappa\vec{\mathcal{E}} = 0. \tag{2.7}$$

If we look for a solution of (2.6) in which $\vec{\mathcal{E}} = [\mathcal{E}_x, 0, 0]$, it follows immediately from (2.1) that $H_x = 0$. From (2.2), it follows then that $\partial H_y/\partial x = 0$ and $\partial H_z/\partial x = 0$. Thus $\mathcal{E}_x$ must also be independent of x. This is quite satisfactory for a plane-wave solution of course. Under the present assumed form of solution it follows that $\mathrm{div}\vec{\mathcal{E}} = 0$ so that (2.6) reduces to the one-component wave equation

$$\nabla^2\mathcal{E}x = -(\omega/c)^2\kappa\mathcal{E}_x. \tag{2.8}$$

However, in the usual treatment of such problems, one assumes $\mathcal{E}_x = \mathcal{E}_x(x, y, z)$ so that $\mathrm{div}\mathcal{E} = \partial\mathcal{E}_x/\partial x \neq 0$. Nevertheless, grad div$\mathcal{E}$ is neglected. For solutions propagating mainly in the z direction, the ansatz is next made that

$$\mathcal{E}_x = \psi e^{ikz}. \tag{2.9}$$

Then it is customary to make the approximation that

$$\left|\frac{\partial^2\psi}{\partial z^2}\right| \ll k\left|\frac{\partial\psi}{\partial z}\right| , \tag{2.10}$$

which is essentially a paraxial approximation. At this stage, however, various authors[1,2] appropriately (as we shall show) forget that $\partial\psi/\partial x$ should equal zero. They then write (2.8) as

$$\nabla_T^2\psi + 2ik\left(\frac{\partial\psi}{\partial z}\right) - k^2\psi = -\left(\frac{\omega}{c}\right)^2 \kappa\psi, \tag{2.11}$$

where

$$\nabla_T^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} . \tag{2.12}$$

In the case of the empty spherical laser resonator for which $\kappa = 1$ and $\omega = ck$, (2.11) is solved subject to boundary conditions and the modes are found. The lowest mode is Gaussian[2] in x and y and not independent of x as the starting approximations suggest. Experimentally, the laser-oscillator modes found in this apparently inconsistent way agree extremely well with those predicted by this theory.

We proceed to study the paraxial approximation in an effort to resolve this paradox and present a systematic procedure to obtain higher order corrections when needed.

III. THE PARAXIAL APPROXIMATION

The assumptions discussed in the previous section are discarded. We shall start afresh with the full Maxwell equations and show how an appropriate scaling of Maxwell's equations leads in a natural way to the usual approximations. The advantage of a systematic procedure, however, is that we shall be expanding in terms of a small parameter, and the corrections to the standard paraxial results are readily obtained. The usual discussion merely makes the uncontrolled approximation of dropping grad div$\vec{\mathcal{E}}$ and $\partial^2\psi/\partial z^2$ terms.

Since we shall be interested in waves propagating mainly in the z direction, we write our field as

$$\vec{\mathcal{E}} = \vec{\mathcal{E}}_T + \hat{a}_z \mathcal{E}_z \equiv e^{ikz}(\vec{\mathbf{F}}_T + \hat{a}_z F_z) , \tag{3.1}$$

where $\hat{a}_z$ is a unit vector in the z direction and T stands for the transverse part of the field. Also we let

$$\nabla = \nabla_T + \hat{a}_z \frac{\partial}{\partial z} , \tag{3.2}$$

where ∇_T is the transverse gradient. When we use these in (2.6) we obtain for the transverse and longitudinal components

$$\nabla_T\left(\nabla_T \cdot \vec{\mathbf{F}}_T + \frac{\partial F_z}{\partial z} + ikF_z\right) - \nabla_T^2 \vec{\mathbf{F}}_T - \frac{\partial^2 \vec{\mathbf{F}}_T}{\partial z^2} - 2ik\frac{\partial \vec{\mathbf{F}}_T}{\partial z} + k^2\vec{\mathbf{F}}_T = \left(\frac{\omega}{c}\right)^2 \kappa \vec{\mathbf{F}}_T , \tag{3.3}$$

$$\frac{\partial}{\partial z}(\nabla_T \cdot \vec{\mathbf{F}}_T) + ik\nabla_T \cdot \vec{\mathbf{F}}_T - \nabla_T^2 F_z = \left(\frac{\omega}{c}\right)^2 \kappa F_z . \tag{3.4}$$

If we are considering Gaussian-like beams, then there is some characteristic width w_0 in the transverse dimension. For such a beam, there is also a diffraction length

$$l = kw_0^2 = 2\pi w_0^2/\lambda \tag{3.5}$$

which is a characteristic length in the longitudinal direction associated naturally with such a beam. Accordingly, we proceed to scale (3.3) and (3.4) and let

$$x = w_0\xi, \quad y = w_0\eta; \quad z = l\zeta . \tag{3.6}$$

Since it is anticipated that for all problems of interest $w_0 \ll l$, we shall let

$$f \equiv \frac{w_0}{l} = \frac{1}{kw_0} . \tag{3.7}$$

When we use (3.6) and (3.7) in (3.3) and (3.4) we obtain with no approximations as yet

$$\nabla_T\left(f\nabla_T \cdot \vec{\mathbf{F}}_T + f^2\frac{\partial F_\zeta}{\partial \zeta} + iF_\zeta\right) - f\nabla_T^2 \vec{\mathbf{F}}_T - f^3\frac{\partial^2\vec{\mathbf{F}}_T}{\partial\zeta^2} - 2if\frac{\partial\vec{\mathbf{F}}_T}{\partial\zeta} = f\left[\left(\omega\frac{w_0}{c}\right)^2\kappa - (kw_0)^2\right]\vec{\mathbf{F}}_T , \tag{3.8}$$

$$f^3\frac{\partial}{\partial\zeta}(\nabla_T \cdot \vec{\mathbf{F}}_T) + if\nabla_T \cdot \vec{\mathbf{F}}_T - f^2\nabla_T^2 F_\zeta = f^2\left(\omega\frac{w_0}{c}\right)^2 \kappa F_\zeta , \tag{3.9}$$

where we have let

$$\vec{\mathbf{F}}_T(\vec{\mathbf{r}}_T, z) \rightarrow \vec{\mathbf{F}}_T(\vec{\rho}, \zeta), \qquad F_z(\vec{\mathbf{r}}_T, z) \rightarrow F_\zeta(\vec{\rho}, \zeta) . \tag{3.10}$$

Also

$$\nabla_T \equiv \hat{a}_x\frac{\partial}{\partial\xi} + \hat{a}_y\frac{\partial}{\partial\eta} . \tag{3.11}$$

where $\hat{a}_x$ and $\hat{a}_y$ are unit vectors in the x and y

directions.

Before we can proceed to make the paraxial approximation, we must know the size of the nonlinear part of κ. Let us consider the case of a homogeneously broadened laser amplifier medium in which we let $\sigma = 0$.

The dielectric response of a medium with n atoms per unit volume, and a homogeneous linewidth γ_{ab} can be written[3]

$$\begin{aligned}\kappa &= \kappa_L + \kappa'_{NL} + \kappa''_{NL}\\ &= \kappa_L + \frac{(\Omega - i)[\sqrt{\kappa_L}(cg/\omega)]}{1+\Omega^2+I},\\ &\equiv \kappa_L + [\sqrt{\kappa_L}(cg/\omega)]m(\omega, |\vec{\mathbf{F}}|^2),\end{aligned} \tag{3.12}$$

where $\Omega = (\omega - \omega_{ab})/\gamma_{ab}$, κ_L is the background linear dielectric constant (assumed real), g is the on-resonance small signal gain per meter, and $I = |\vec{\mathbf{F}}|^2/F_s^2$ is the field intensity in units of the saturation intensity. The gain g should be taken from experiment. In our calculation the gain is defined by

$$g = -\frac{\omega}{c}\,\frac{\kappa''(\omega=\omega_{ab}, \vec{\mathbf{F}}=0)}{\sqrt{\kappa_L}} = \frac{np^2D_w\omega}{\sqrt{\kappa_L}\,\hbar\gamma_{ab}\epsilon_0 c}, \tag{3.13}$$

where D_w is the unsaturated population inversion (called $\mathfrak{N}$ by Lamb[3]), n is the concentration of active atoms

$$D_w \equiv \frac{\lambda_a}{\gamma_a} - \frac{\lambda_b}{\gamma_b}, \tag{3.14}$$

where λ_a and λ_b are the pump rates (per atom) into the upper and lower states, respectively, γ_a and γ_b are the corresponding decay rates (due to spontaneous emission), $\omega_{ab} = \omega_a - \omega_b$ is the atomic frequency difference and γ_{ab} is the associated homogeneous broadening linewidth. In the above equation p is the transition atomic dipole moment and $F_s = h\gamma_{ab}/p$ is the saturation electric field. We also choose

$$k^2 = (\omega/c)^2\kappa_L\,. \tag{3.15}$$

For this model (3.8) and (3.9) reduce to

$$\nabla_\tau\left(f\nabla_\tau\cdot\vec{\mathbf{F}}_\tau + f^2\frac{\partial F_\zeta}{\partial\zeta} + iF_\zeta\right) - f\nabla_\tau^2\,\vec{\mathbf{F}}_\tau - f^3\frac{\partial^2\vec{\mathbf{F}}_\tau}{\partial\zeta^2} - 2if\,\frac{\partial\vec{\mathbf{F}}_\tau}{\partial\zeta} = (fglm)\,\vec{\mathbf{F}}_\tau\,, \tag{3.16}$$

$$f^3\frac{\partial}{\partial\zeta}(\nabla_\tau\cdot\vec{\mathbf{F}}_\tau) + if\nabla_\tau\cdot\vec{\mathbf{F}}_\tau - f^2\nabla_\tau^2\,F_\zeta = [1+(f^2glm)]\,F_\zeta\,. \tag{3.17}$$

We note that $f^2gl = g/k =$ the gain in a distance k^{-1} is always small even when $gl \gg 1$. Thus we can always obtain a consistent solution[4] if we expand the field in powers of f. Only alternate powers are found necessary:

$$\begin{aligned}\vec{\mathbf{F}}_\tau &= \vec{\mathbf{F}}_\tau^{(0)} + f^2\,\vec{\mathbf{F}}_\tau^{(2)} + \cdots\,,\\ F_\zeta &= fF_\zeta^{(1)} + f^3\,F_\zeta^{(3)} + \cdots\,.\end{aligned} \tag{3.18}$$

That F_ζ has no zero-order term follows from (3.17). Before equating various powers of f in (3.16) and (3.17), we must take into account the fact that m is a function of $|\vec{\mathbf{F}}|^2$ and it must be expanded in powers of f. When we use (3.18); then from (3.17) it follows for terms of order f that

$$F_\zeta^{(1)} = i\nabla_\tau\cdot\vec{\mathbf{F}}_\tau^{(0)}\,, \tag{3.19}$$

while from (3.17), the lowest-order nonvanishing terms are

$$\nabla_\tau^2\,\vec{\mathbf{F}}_\tau^{(0)} + 2i\left(\frac{\partial\vec{\mathbf{F}}_\tau^{(0)}}{\partial\zeta}\right) = -(gl)m_0\vec{\mathbf{F}}_\tau^{(0)}\,, \tag{3.20}$$

where

$$m_0 \equiv m(|\vec{\mathbf{F}}^0|^2) = \frac{\Omega - i}{1+\Omega^2+(I^0)^2} \tag{3.21}$$

is the form appropriate to homogeneous broadening.

We have thus resolved the paradox, since to lowest order the field is purely transverse and may depend on the transverse coordinate. However, in next order a small longitudinal component of the field must be present and its size depends on $(w_0/l) = f$. Furthermore, we have a procedure to obtain the next-order correction to the transverse components.

When we use (3.19), the terms of order f^3 in (3.16) and (3.17) become, respectively,

$$\nabla_\tau\left[\nabla_\tau\cdot\vec{\mathbf{F}}_\tau^{(2)} + i\frac{\partial}{\partial\zeta}(\nabla_\tau\cdot F_\tau^{(0)}) + iF_\zeta^{(3)}\right] - \nabla_\tau^2\vec{\mathbf{F}}_\tau^{(2)} - \frac{\partial^2\vec{\mathbf{F}}_\tau^{(0)}}{\partial\zeta^2} - 2i\frac{\partial\vec{\mathbf{F}}_\tau^{(2)}}{\partial\zeta} = (gl)(m_0\vec{\mathbf{F}}_\tau^{(2)} + m_2\vec{\mathbf{F}}_\tau^{(0)}) \tag{3.22}$$

and

$$\frac{\partial}{\partial\zeta}(\nabla_\tau\cdot\vec{\mathbf{F}}_\tau^{(0)}) + i\nabla_\tau\cdot\vec{\mathbf{F}}_\tau^{(2)} - i\nabla_\tau^2(\nabla_\tau\cdot\vec{\mathbf{F}}_\tau^{(0)}) = F_\zeta^{(3)} + (gl)m_0 i\nabla_\tau\cdot\vec{\mathbf{F}}_\tau^{(0)}\,, \tag{3.23}$$

where m_2 is the term of order f^2 in the expansion of m. If we solve (3.23) for $F_\zeta^{(3)}$ and substitute into (3.22) and use (3.20), we obtain

$$\nabla_\tau^2 F_\tau^{(2)} + 2i\frac{\partial\vec{\mathbf{F}}_\tau^{(2)}}{\partial\zeta} + glm_0\vec{\mathbf{F}}_\tau^{(2)} = -(gl)m_2\vec{\mathbf{F}}_\tau^{(0)} - \frac{\partial^2\vec{\mathbf{F}}_\tau^{(0)}}{\partial\zeta^2}, \tag{3.24}$$

while

$$F_\zeta^{(3)} = i\nabla_T \cdot \vec{F}_T^{(2)} - \nabla_T \cdot \left(\frac{\partial \vec{F}_T^{(0)}}{\partial \zeta}\right)$$

$$= i\nabla_T \cdot \vec{F}_T^{(2)} + i\frac{\partial F_\zeta^{(1)}}{\partial \zeta}. \qquad (3.25)$$

Equation (3.24) shows that if $gl \lesssim 1$, then our original expansion parameter $f^2 = (w_0/l)^2$ is descriptive of the ratio of $\mathcal{E}_T^{(2)}/\mathcal{E}_T^{(0)} = f^2 F_T^{(2)}/F_T^{(0)}$. When $gl \gg 1$, our theory remains valid, but the presence of a factor gl on the right-hand side shows that $\mathcal{E}_T^{(2)}/\mathcal{E}_T^{(0)} \sim f^2 gl = g/k$. Since g/k is the gain in a distance $\lambda/2\pi$, this ratio is always small, and our theory is clearly still valid even if $gl \gg 1$.

The ratio of $\mathcal{E}_z^{(1)}/\mathcal{E}_T^{(0)}$ is of order $f = (kw_0)^{-1}$ regardless of the size of gl. By (3.25) we see that $\mathcal{E}_z^{(3)}/\mathcal{E}_T^{(0)} \sim f^3 gl \sim f(g/k)$ since $\mathcal{E}_T^{(2)}$ is of order $gl\mathcal{E}_T^{(0)}$. We therefore see that $\mathcal{E}_z^{(3)}$ is completely negligible regardless of the size of gl.

In our original unscaled variables, (3.18) and (3.17) yield in the two lowest orders

$$\nabla_T^2 \vec{F}_T^{(0)} + 2ik\frac{\partial \vec{F}_T^{(0)}}{\partial z} = -g\sqrt{\kappa_L}\left(\frac{\omega}{c}\right) m_0 \vec{F}_T^{(0)}, \qquad (3.26)$$

$$kF_z^{(1)} = i\nabla_T \cdot \vec{F}_T^{(0)} . \qquad (3.27)$$

For numerical calculations, as usual, the scaled version seems generally preferable. (Note that $\boldsymbol{F}_z^{(1)} = fF_\zeta^{(1)}$.)

One surprising result is that the longitudinal component of the field is out of phase with the transverse components.

IV. RAY EQUATIONS

With no loss of generality we may look for plane polarized solutions of (3.20) of the form[5]

$$\vec{F}_T^{(0)} = \vec{E}e^{iS}, \qquad (4.1)$$

where $\vec{E}$ and S are real. When we use this in (3.20) and equate real and imaginary parts, we obtain

$$\nabla_T^2 \vec{E} - (\nabla_T S)^2 \vec{E} - 2\left(\frac{\partial S}{\partial \zeta}\right)\vec{E} = -(gl)(\mathrm{Re}\, m_0)\vec{E}, \qquad (4.2)$$

$$2(\nabla_T S \cdot \nabla_T)\vec{E} + \nabla_T^2 S \vec{E} + 2\left(\frac{\partial \vec{E}}{\partial \zeta}\right) = -(gl)(\mathrm{Im}\, m_0)\vec{E}, \qquad (4.3)$$

where by (3.21), $\mathrm{Im}\, m_0 < 0$. If we take the scalar product of both sides of (4.2) with $\vec{E}$, we obtain

$$(\nabla_T S)^2 + 2\left(\frac{\partial S}{\partial \zeta}\right) = (gl)(\mathrm{Re}\, m_0) + (\vec{E}\cdot\nabla_T^2\vec{E})E^{-2}. \qquad (4.4)$$

This is the analog of the eikonal equation of geometrical optics in which diffraction is included in the last term. If we next take the scalar product of both sides of (4.3) with $\vec{E}$, it may be rewritten

$$\nabla_T \cdot (E^2 \nabla_T S) + \left(\frac{\partial E^2}{\partial \zeta}\right) = -(gl)(\mathrm{Im}\, m_0)E^2. \qquad (4.5)$$

This is an energy transport equation.

Rays normal to surfaces of constant phase obey the equations

$$\frac{d\vec{r}_T}{ds} = \frac{\nabla_T S}{|\nabla(S + kz)|}, \qquad (4.6)$$

$$\frac{dz}{ds} = \frac{k + (\partial S/\partial z)}{|\nabla(S + kz)|}, \qquad (4.7)$$

so that dividing (4.6) by (4.7) we obtain

$$\frac{d\vec{r}_T}{dz} = \frac{\nabla_T S}{k + (\partial S/\partial z)}. \qquad (4.8)$$

In the scaled variables $\vec{r}_T = w_0\vec{\rho}$ and $z = \zeta l$, this becomes

$$\frac{d\vec{\rho}}{d\zeta} = \frac{\nabla_T S}{1 + f^2(\partial S/\partial \zeta)}. \qquad (4.9)$$

However, we are talking about the surfaces of constant phase to lowest order in f so that

$$\frac{d\vec{\rho}}{d\zeta} = \nabla_T S. \qquad (4.10)$$

If U is any function of $\vec{\rho}$ and ζ, we have by (4.10)

$$\frac{dU}{d\zeta} = \nabla_T U \cdot \left(\frac{d\vec{\rho}}{\partial \zeta}\right) + \frac{\partial U}{\partial \zeta}$$

$$= (\nabla_T S \cdot \nabla_T) U + \frac{\partial U}{\partial \zeta}. \qquad (4.11)$$

Accordingly, along a ray, (4.3) may be written

$$\frac{d\vec{E}}{d\zeta} + \frac{1}{2}[\nabla_T^2 S + (gl)\,\mathrm{Im}\, m_0]\,\vec{E} = 0 \qquad (4.12)$$

or

$$\frac{d(\ln E^2)}{d\zeta} = -[\nabla_T^2 S + (gl)\,\mathrm{Im}\, m_0]. \qquad (4.13)$$

Of course, one must not forget that $\mathrm{Im}\, m_0$ is a function of E^2.

If we next differentiate both sides of (4.10) and use (4.11) and (4.4), we obtain

$$\frac{d^2\vec{\rho}}{d\zeta^2} = (\nabla_T S \cdot \nabla_T)\nabla_T S + \frac{\partial(\nabla_T S)}{\partial \zeta}$$

$$= \frac{1}{2}\left(\nabla_T(\nabla_T S)^2 + 2\nabla_T\frac{\partial S}{\partial \zeta}\right)$$

$$= \frac{1}{2}\nabla_T\left((\nabla_T S)^2 + 2\frac{\partial S}{\partial \zeta}\right)$$

$$= \frac{1}{2}\nabla_T[(gl)\,\mathrm{Re}\, m_0 + E^{-2}\vec{E}\cdot\nabla_T^2\vec{E}]. \qquad (4.14)$$

This is the form used in Ref. 1 obtained by the analogy of (4.4) to the Hamiltonian-Jacobi equation.

In cylindrical coordinates the "eikonal" equation (4.4) becomes

$$\left(\frac{\partial S}{\partial \rho}\right)^2+\frac{1}{\rho^2}\left(\frac{\partial S}{\partial \varphi}\right)^2+2\frac{\partial S}{\partial \zeta}=gl\,\mathrm{Re}\,m_0(I)+\frac{1}{\rho\sqrt{I}}\frac{\partial}{\partial \rho}\left(\rho\frac{\partial\sqrt{I}}{\partial\rho}\right)+\frac{1}{\rho^2\sqrt{I}}\frac{\partial^2\sqrt{I}}{\partial\varphi^2}\,, \qquad (4.15)$$

while the energy-transport equation (4.5) becomes

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left[I\rho\frac{\partial S}{\partial \rho}\right]+\frac{1}{\rho^2}\frac{\partial}{\partial\varphi}\left(I\frac{\partial S}{\partial\varphi}\right)+\frac{\partial I}{\partial\zeta}=-gl\,\mathrm{Im}\,m_0(I)\cdot I\,. \qquad (4.16)$$

The ray equations (4.10) become

$$\frac{d\rho}{d\zeta}=\frac{\partial S}{\partial\rho}\,, \qquad (4.17)$$

$$\rho^2\frac{d\varphi}{d\zeta}=\frac{\partial S}{\partial\varphi}\,, \qquad (4.18)$$

while Eqs. (4.14) become

$$\frac{d^2\rho}{d\zeta^2}-\left(\frac{d\varphi}{d\zeta}\right)^2\rho=\frac{1}{2}\frac{\partial}{\partial\rho}\left[gl\,\mathrm{Re}\,m_0(I)+\frac{1}{\rho\sqrt{I}}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial\sqrt{I}}{\partial\rho}\right)+\frac{1}{\rho^2\sqrt{I}}\frac{\partial^2\sqrt{I}}{\partial\varphi^2}\right], \qquad (4.19)$$

$$\frac{d}{d\zeta}\left(\rho^2\frac{d\varphi}{d\zeta}\right)=\frac{1}{2}\frac{\partial}{\partial\varphi}\left[gl\,\mathrm{Re}\,m_0(I)+\frac{1}{\rho\sqrt{I}}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial\sqrt{I}}{\partial\rho}\right)+\frac{1}{\rho^2\sqrt{I}}\frac{\partial^2\sqrt{I}}{\partial\varphi^2}\right]. \qquad (4.20)$$

If we seek a solution of Eqs. (4.15) through (4.20) which has rotational invariance, it is convenient to introduce a quantity related to the energy flux between two concentric circles:

$$U(\zeta)=\int_{\rho_1(\zeta)}^{\rho_2(\zeta)} I(\rho,\zeta)\,2\pi\rho\,d\rho\,. \qquad (4.21)$$

The effects of refraction and diffraction in $U(\zeta)$ are compensated for by the use of rings bounded by the rays $\rho_1(\zeta)$ and $\rho_2(\zeta)$. The $U(\zeta)$ will change only because of true gains and losses. The result takes the form

$$\frac{\partial U}{\partial\zeta}=+2\pi gl\int_{\rho_1(\zeta)}^{\rho_2(\zeta)}\frac{I\rho\,d\rho}{1+\Omega^2+I^2}\,, \qquad (4.22)$$

where $I=I(\rho,\zeta)$. To prove this result we differentiate (4.21) to obtain

$$\frac{\partial U}{\partial\zeta}=2\pi\left[\int_{\rho_1}^{\rho_2}\frac{\partial I}{\partial\zeta}\rho\,d\rho+\rho_2\frac{d\rho_2}{d\zeta}I(\rho_2,\zeta)-\rho_1\left(\frac{d\rho_1}{d\zeta}\right)I(\rho_1,\zeta)\right]. \qquad (4.23)$$

We next eliminate $\partial I/\partial\zeta$ using (4.16) (with angular derivatives omitted) and obtain on integrating

$$\frac{\partial U}{\partial\zeta}=2\pi\left[-\rho_2\left(\frac{\partial S}{\partial\rho_2}\right)I(\rho_2,\zeta)+\rho_1\left(\frac{\partial S}{\partial\rho_1}\right)I(\rho_1,\zeta)+\rho_2\left(\frac{d\rho_2}{d\zeta}\right)I(\rho_2,\zeta)-\rho_1\left(\frac{d\rho_1}{d\zeta}\right)I(\rho_1,\zeta)-(gl)\int_{\rho_1(\zeta)}^{\rho_2(\zeta)}\mathrm{Im}\,m_0(I)\,I\rho\,d\rho\right]. \qquad (4.24)$$

All terms on the right cancel except for the last in view of (4.18) and we obtain (4.22) when we use (3.21).

The longitudinal component of the field by (3.19) and (4.1) is

$$\begin{aligned}F_\zeta^{(1)}&=i\nabla_T\cdot(\vec{\mathrm{E}}e^{iS})\\&=ie^{iS}[\nabla_T\cdot\vec{\mathrm{E}}+i\nabla_T S\cdot\vec{\mathrm{E}}]\\&=[(\nabla_T\cdot\vec{\mathrm{E}})^2+(\nabla_T S\cdot\vec{\mathrm{E}})^2]^{1/2}e^{i(S+\psi+\pi/2)}\,,\end{aligned} \qquad (4.25)$$

where

$$\tan\psi=\nabla_T S\cdot\vec{\mathrm{E}}/\nabla_T\cdot\vec{\mathrm{E}}\,. \qquad (4.26)$$

Equation (4.25) describes a longitudinal component of the electric field whose ratio to the dominant transverse component is of first order in the expansion parameter $f=1/kw_0$. Moreover, an exact relation has been supplied between the first-order correction field and the zeroth-order field.

V. AN EXAMPLE

We shall consider the special case of modes in an an empty spherical cavity resonator[2] and determine the first correction to the normal modes. We assume the field is polarized in the x direction so that

$$\vec{\mathrm{F}}_T^{(0)}=(Ee^{iS},0)\,. \qquad (5.1)$$

In cylindrical coordinates (ρ,φ,ζ), (4.4) becomes

$$\left(\frac{\partial S}{\partial\rho}\right)^2+\frac{1}{\rho^2}\left(\frac{\partial S}{\partial\varphi}\right)^2+2\frac{\partial S}{\partial\zeta}=\frac{1}{E_x}\left\{\frac{\partial^2E_x}{\partial\rho^2}+\frac{1}{\rho}\frac{\partial E_x}{\partial\rho}+\frac{1}{\rho^2}\frac{\partial^2E_x}{\partial\varphi^2}\right\}\,, \qquad (5.2)$$

while (4.3) reduces to

$$\frac{\partial S}{\partial\rho}\frac{\partial E_x}{\partial\rho}+\frac{1}{\rho^2}\frac{\partial S}{\partial\varphi}\frac{\partial E_x}{\partial\varphi}+\frac{1}{2}\left(\frac{\partial^2S}{\partial\rho^2}+\frac{1}{\rho}\frac{\partial S}{\partial\rho}+\frac{1}{\rho^2}\frac{\partial^2S}{\partial\varphi^2}\right)E_x+i\frac{\partial E_x}{\partial\zeta}=0\,. \qquad (5.3)$$

It is straightforward to verify that exact solutions of these equations may be written as[2]

$$S = -m\varphi + \frac{\rho^2}{R(\zeta)} - (2l + m + 1)\tan^{-1}2\zeta , \qquad (5.4)$$

$$E = w^{-1}(\zeta)\, e^{-\xi/2}\, \xi^{m/2}\, L_l^m(\xi) , \qquad (5.5)$$

where

$$w^2(\zeta) = 1 + (2\zeta)^2 ,$$

$$R(\zeta) = \left(\frac{1 + (2\zeta)^2}{2\zeta}\right) , \qquad (5.6)$$

$$\xi = 2\rho^2/w^2(\zeta) .$$

$L_l^m(\xi)$ are associated Laguerre polynominals. m and l are integers and l here should not be confused with the diffraction length. These solutions are valid when the beam waist size w_0 at $z = 0$ is small compared with the diffraction length $l = kw_0^2$.

The ray equations (4.18) become

$$\frac{d\rho}{d\zeta} = \frac{\partial S}{\partial \rho} = \frac{4\zeta\rho}{1 + (2\zeta)^2} , \qquad (5.7)$$

$$\rho\frac{d\varphi}{d\zeta} = \frac{1}{\rho}\frac{\partial S}{\partial \varphi} = -\frac{m}{\rho} , \qquad (5.8)$$

where we used (5.4) through (5.6). These may be integrated to yield the hyperbolic rays

$$\rho = \rho_0\,(1 + 4\zeta^2)^{1/2} , \qquad (5.9)$$

where $r_0 \equiv w_0\rho_0$ is the radius of the ray at $\zeta = 0$. Also

$$\varphi = -(m/2\rho_0^2)\tan^{-1}2\zeta + \varphi_0 . \qquad (5.10)$$

Note from (5.8) that $\rho^2(d\varphi/d\zeta) = -m$, a constant.

If we use (5.4)–(5.6) and (3.16) we see that the longitudinal correction is

$$F_\zeta^{(1)} = \left[i\cos\varphi \left(\frac{d(\ln L_l^m)}{d\xi} + \frac{1}{2}(m\xi^{-1} - 1)\right)\left(\frac{4\rho}{w^2}\right) - \left(\frac{2\rho\cos\varphi}{R}\right) - \left(\frac{m\sin\varphi}{\rho}\right)\right] E e^{iS} . \qquad (5.11)$$

For the fundamental mode, $m = l = 0$,

$$F_{\zeta 00}^{(1)} = -(iw^{-2} + R^{-1})2\rho\cos\varphi\, E_{00}\exp(iS_{00}) \qquad (5.12)$$

or

$$\frac{|F_{\zeta 00}^{(1)}|^2}{|\vec{F}_{T00}|^2} = \frac{4\rho^2\cos^2\varphi}{1 + (2\zeta)^2} . \qquad (5.13)$$

In unscaled variables

$$\frac{|F_{z00}^{(1)}|^2}{|\vec{F}_{T00}|^2} = \frac{4(x/kw_0)^2}{w^2(z)} , \qquad (5.14)$$

where $w^2(z) = w_0^2[1 + (2z/kw_0^2)^2]$. Since w_0 is usually significantly larger than k^{-1}, the conventional solution is accurate for x appreciably larger than the local beam diameter, $w(z)$.

VI. SUMMARY

The paraxial approximation makes the inconsistent assumptions that one can have a plane polarized electromagnetic wave whose electric vector depends on the transverse distance. We have established that this result is a consistent zeroth-order solution to Maxwell's equations obtained by expanding all fields as a power series in the ratio of beam diameter to diffraction length. The expansion is shown to contain only alternate orders in the expansion coefficient and thus converges rapidly. The first correction is evaluated and is shown to be a first-order component of the electric field along the beam direction. Equations which yield higher-order corrections are also presented.

---

*Partially supported by AROD and ONR.

†Permanent address: Departments of Physics and Electrical Engineering, University of Southern California, Los Angeles, Calif. 90007.

[1]See, e.g., W. G. Wagner, H. A. Haus, and J. H. Marburger, Phys. Rev. 175, 256 (1968); V. I. Talanov, Zh. Eksp. Teor. Fiz. Pis'ma Red. 2, 218 (1965) [JETP Lett. 2, 138 (1965)]; Yu P. Raiser, Zh. Eksp. Teor. Fiz. Pis'ma Red. [JETP Lett. 4, 1 (1966)]; S. K. Akhamanov, A. P. Sukhorukov, and R. V. Khokhlov, Zh. Eksp. Teor. Fiz. 50, 1537 (1966) [Sov. Phys.—JETP 23, 1025 (1966)]; H. Kogelnik, Appl. Opt. 4, 1562 (1965); P. L. Kelley, Phys. Rev. Lett. 15, 1005 (1965).

[2]H. Kogelnik and T. Li, Proc. IEEE, 54, 1312 (1966). See also W. H. Louisell, *Quantum Statistical Properties of Radiation* (Wiley, New York, 1973), Chap. 8; G. D. Boyd and J. P. Gordon, Bell Syst. Tech. J. 40, 489 (1961); L. A. Vainshtein, Zh. Tekhn. Fiz. 34, 183 (1964) [Sov. Phys.—Tech. Phys. 9, 157 (1964)].

[3]W. E. Lamb, Jr., Phys. Rev. 134, 1429 (1964); M. Lax, in *Statistical Physics*, edited by M. Chretien (Gordon and Breach, New York, 1968), Vol. 2, pp. 270–478.

[4]It would appear that our expansion procedure would fail when $gl \gg 1$. If, however, one uses the gain length as the appropriate length in the z direction, the diffraction terms are lost in the zeroth approximation. We shall show below that it will always be appropriate to use the diffraction length as the longitudinal scaling parameter. The effective expansion parameter relating alternate order will no longer be f^2, but the larger value $f^2gl = g/k$ which is still small. These remarks will be justified below.

[5]Since m is a function of $|\vec{F}_T^{(0)}|^2$, (3.20) is rotationally invariant. Thus, if $(F_x^{(0)}, 0, 0)$ is a solution of (3.20), so is $[0, F_x^{(0)}, 0]$ and also $(aF_x^{(0)}, bF_x^{(0)}, 0)$ where a and b are arbitrary (complex) constants with $|a|^2 + |b|^2 = 1$. We may thus construct solutions with elliptic polarization.