



IRWIN AND JOAN JACOBS
CENTER FOR COMMUNICATION AND INFORMATION TECHNOLOGIES

Quadratic phase of tightly focused wavefronts

**Alexander Normatov, Boris
Spektor and Joseph Shamir**

CCIT Report # 716
January 2009

■ ■ ■ ■ ■ Electronics
■ ■ ■ ■ ■ Computers
■ ■ ■ ■ ■ Communications

DEPARTMENT OF ELECTRICAL ENGINEERING
TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA 32000, ISRAEL



Quadratic phase of tightly focused wavefronts

Alexander Normatov*, Boris Spektor, and Joseph Shamir

Department of Electrical Engineering, Technion – Israel Institute of Technology;
Technion City, Haifa 32000, Israel

*Corresponding author: alexn@tx.technion.ac.il

Abstract: In an aplanatic optical system, where the entrance pupil is not located at the front focal plane, a corresponding quadratic phase is expected at the back focal plane. However, tightly focused optical fields evaluated by methods based on the traditional Richards-Wolf approach, lacks this quadratic phase. In the current work we calculate the focused field, for both high and medium numerical apertures, based on the Stratton-Chu diffraction integral in a 2D system. We find that the quadratic phase factor depends on the numerical aperture, and it approaches the corresponding paraxial value for lower numerical apertures.

References and links

1. B. Richards and E. Wolf, "Electromagnetic diffraction in optical systems, II Structure of the image field in an optical system," *Proc. R. Soc. Lond. A.* **253**, 358-379 (1959).
2. A. Normatov, B. Spektor, and J. Shamir, "Tight Focusing of Wavefronts with Piecewise Constant Phase," CCIT Report #681, Electrical Engineering Faculty, Technion, January 2008.
3. J. A. Stratton and L. J. Chu, "Diffraction Theory of Electromagnetic Waves", *Phys. Rev.* **56**, 99-107 (1939).
4. E. Wolf and Y. Li, "Conditions for the validity of the Debye integral representation of focused fields," *Opt. Comm.* **39**, 205-210 (1981).
5. E. Wolf, "Electromagnetic diffraction in optical systems, I An integral representation of the image field," *Proc. R. Soc. Lond. A.* **253**, 349-357 (1959).
6. M. Born and E. Wolf, *Principles of Optics* (Cambridge University Press 2003).
7. A. Papoulis, *Systems and Transforms with Applications in Optics* (McGraw-Hill 1968).
8. P. Torok, P.R.T. Munro and E.E. Kriezis "Rigorous near- to far-field transformation for vectorial diffraction calculations and its numerical implementation," *J. Opt. Soc. Am. A.* **23**, 713-722 (2006).
9. P. Varga and P. Torok, "Focusing of electromagnetic waves by paraboloid mirrors. I. Theory," *J. Opt. Soc. Am. A.* **17**, 2081-2089 (2000).
10. J. W. Goodman, *Introduction To Fourier Optics* (McGraw-Hill 1968).
11. J. Shamir, *Optical Systems and Processes* (SPIE, 1999).

1. Introduction

Tight focusing of light in high numerical aperture (NA) systems is becoming increasingly important with recent advances in nanotechnology. The majority of methods suggested for the tight focusing of different types of illumination are based on the approach pioneered by Richards and Wolf in 1959 [1] (for a detailed review of these methods see [2]). Although this approach has a number of drawbacks it is the only one that offers reasonable 3D computation complexity for this class of diffraction problems. Among its drawbacks is the fact that the derived focused field at the back focal plane lacks the expected quadratic phase when the entrance pupil is not situated at the front focal plane.

The computational complexity can be substantially reduced when 2D diffraction problems are considered. Therefore these problems lend themselves to treatment by other rigorous diffraction approaches, which involve less assumptions and approximations than the Richards-Wolf (RW) approach. One such approach is the calculation of the diffracted field by means of the Stratton-Chu (SC) [3] diffraction integral.

In this work we perform a numerical investigation of the focused field of an ideal aplanatic 2D optical system, imaging a point source in infinity. We examine optical systems with NA values ranging from 0.2, which is high enough for the validity of the Debye approximation [4], to 0.96, which is about the limit for most immersion-less focusing optics. Our results show that in all the investigated range the RW approach produced a piecewise planar phase of the focused field on the back focal plane. In contrast, the SC approach produced a quadratic phase which approaches the quadratic phase produced by the paraxial approach, for decreasing NA values. It was found that the relation between the quadratic coefficients of the SC and paraxial cases was equal to $1-NA^2$. As explained later, this relation is not a consequence of optical aberrations. This interesting result was found valid also for 3D optical systems, for all the field components. In an analogy to the paraxial case, we believe that this quadratic phase is a property of an optical system. In other words, it is not dependent on the incident illumination, as long as this illumination satisfies the involved assumptions on the optical system.

This paper is organized as follows. The next section derives the expressions for the 2D focusing of piecewise quasi constant phase wavefronts at the entrance pupil, which is based on the RW approach. The third section presents the expressions for the 2D SC integral. The investigated problem is discussed in the fourth section, which is followed by the numerical results and their analysis in the fifth section. Finally, the conclusions are drawn.

2. Two dimensional focusing based on Richards-Wolf approach

2.1 Problem statement and system description

The original RW approach was formulated for an aplanatic optical system having rotational symmetry. In this section we reformulate the original approach and adapt it for a 2D system. In the course of this reformulation we leave a provision for wavefronts with piecewise quasi constant phase at the entrance pupil as opposed to a plane wave in the original approach. The involved 2D system is a degraded case of a 3D system with the field distribution along the y axis being infinitely uniform. Thus without loss of generality we consider the xz plane, where the z axis is also the axis of the optical system.

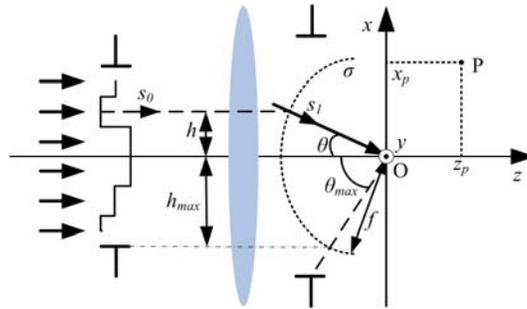


Fig. 1. Schematic illustration of the investigated problem.

The involved optical system is schematically depicted in Fig. 1. In the object space, the incident wavefront, at the entrance pupil, is formed by rays propagating along the optical axis z . Each ray is associated with the unit vector s_0 , as shown in the figure. A distance between a ray and the z axis is denoted h . After passing the optical system, the ray leaves through the exit pupil in a direction of an associated unit vector s_1 . All the rays can be thought of as emerging from an imaginary Gaussian cylinder σ with radius f , the focal length of the optical system. The axis of the Gaussian cylinder is situated on the y axis. The rays and their associated unit vectors s_1 emerging from the Gaussian cylinder are directed towards the geometrical focal point O . Each vector s_1 forms an angle θ with the negative direction of the z axis. Angles θ situated in the second quadrant are defined as positive and those in the third

quadrant are defined as negative. The numerical aperture of the optical system defines a maximum angle θ_{max} . The fields are evaluated at a point $P(x,z)$ in the vicinity of the geometrical focus.

According to [5] these fields in an angular spectrum representation, for a 2D case, are given by the following expressions, where \mathbf{e} denotes the electric field and \mathbf{h} denotes the magnetic field:

$$\mathbf{e}(x, z) = -\frac{i}{\sqrt{\lambda}} \int_{\Omega} \frac{\mathbf{a}(s_x)}{s_z} \exp\{ik[\Phi(s_x) + s_x x + s_z z]\} ds_x \quad (1)$$

$$\mathbf{h}(x, z) = -\frac{i}{\sqrt{\lambda}} \int_{\Omega} \frac{\mathbf{b}(s_x)}{s_z} \exp\{ik[\Phi(s_x) + s_x x + s_z z]\} ds_x \quad (2)$$

The variables s_x, s_z are the respective x and z components of the unit vector \mathbf{s}_l . These unit vectors span the integration surface angle Ω . Ray strength factors are denoted \mathbf{a} and \mathbf{b} . Wavelength is denoted by λ and wave-number by k . The considered optical system is aplanatic. Thus for a case of imaging a point source at infinity the wavefronts in the image space are cylindrical with the common center at the Gaussian image point, which is the geometrical focal point. Therefore the aberration function Φ is assumed to be free of terms representing both primary and higher order optical aberrations. Yet, it reflects phase deviations of the incident wavefront from a plane wave case as they appear at the entrance pupil. The square root of wavelength factor is due to the 2D nature of the considered problem.

A number of other assumptions, used in [1], must be stated. Both the object space and the image space have $\varepsilon \approx \mu \approx 1$. The refraction in the optical system involves small angles that do not significantly affect the direction that the vibration of the electric (and magnetic) field makes with the xz plane. As all the rays are contained in the xz plane, which also contains the optical axis, this plane has the same significance as a meridional plane in a 3D optical system of rotation. The integrals (1) and (2) imply Kirchhoff boundary conditions, which can be justified only for cases when the exit pupil dimensions and the propagation distance to the focal point are large compared to the wavelength.

2.2 Derivation of an explicit expression for the focused field

In the preceding subsection it was assumed that a piecewise quasi constant wavefront at the entrance pupil was adequately described by rays parallel to the optical axis. This assumption implies that each segment with a quasi constant phase satisfies the conditions for such a representation. Moreover, it is assumed that only a small fraction of the energy in the incident field is contained in regions where non-negligible phase variations occur, in particular in the boundaries between neighboring segments. Under these conditions it is possible to use the principle of superposition to evaluate the field in the focal region as a superposition of the contributions obtained from each quasi constant phase segment of the incident field. It should be noted that no assumptions are made about the incident optical field amplitude or polarization.

The electric fields \mathbf{e}_0 and \mathbf{e}_1 , associated with the incident and emerging rays, in the object and image space accordingly can be represented by the following expressions:

$$\mathbf{e}_0 = l_0 \exp(ik\psi_0) \exp(i\gamma_0) \mathbf{e}_{u0} \quad (3)$$

$$\mathbf{e}_1 = l_1 \exp(ik\psi_1) \exp(i\gamma_1) \mathbf{e}_{u1} \quad (4)$$

The l_0 and l_1 are the amplitude factors, ψ_0 and ψ_1 are the values of the eikonal, γ_0 is the phase deviation of the incident wavefront from a plane wave, γ_1 is the phase deviation of the

emerging wavefront from the Gaussian cylinder, \mathbf{e}_{u0} and \mathbf{e}_{u1} are unit vectors in the direction of the fields \mathbf{e}_θ and \mathbf{e}_I , respectively. Similarly to [1], using the sine condition

$$h = f \sin(\theta) \quad (5)$$

and the geometrical optics intensity law we obtain:

$$l_1 = l_0 \sqrt{\cos(\theta)} \quad (6)$$

The expression is the same as for the 3D case. The reason is that the area factor responsible for the square root lies, in case of 3D optical system, in the meridional plane which corresponds to the xz plane as discussed above.

A converging cylindrical wavefront at the image space corresponds to a plane wave in the object space. Thus the phase deviations from them for the same ray are equal, i.e.:

$$\gamma_0 = \gamma_1 \quad (7)$$

From (3), (4), (6) and (7) the expression for the electrical field of the emerging ray becomes:

$$\mathbf{e}_I = l_0 \sqrt{\cos(\theta)} \exp(i\gamma_0) \exp(ik\psi_1) \mathbf{e}_{u1} \quad (8)$$

When expressed with the ray strength factor \mathbf{a} the electrical field of the emerging ray on the Gaussian cylinder surface is given as:

$$\mathbf{e}_I = \mathbf{a} \exp(i\gamma_0) \exp(ik\psi_1) / \sqrt{f} \quad (9)$$

Consequently, for the ray strength factor \mathbf{a} we get the following notation:

$$\mathbf{a} = l_0 \sqrt{f \cos(\theta)} \mathbf{e}_{u1} \quad (10)$$

A couple of notes must be made at this point. First, there is a small difference in \mathbf{a} from the classical 3D case as the amplitude is proportional to the square root of the focal distance, rather than to the distance itself. Second, the phase deviation γ , which is neither in \mathbf{a} nor in the eikonal is accounted for in the aberration function Φ . Third, the distinction between the positive and the negative θ values gives a different meaning to the θ angle as opposed to the 3D case.

In order to obtain the direction of \mathbf{e}_{u1} we define a Cartesian xyz system with origin at the geometrical focal point as shown in Fig. 1. Similarly to [1] we introduce two unit vectors \mathbf{g}_0 and \mathbf{g}_I . These vectors lie in the xz plane such that the vector \mathbf{g}_0 is perpendicular to the ray in the object space and \mathbf{g}_I is perpendicular to the ray in the image space, while both are directed away from the optical axis z . As we assumed above (similarly to [1]) the angle that the electric and magnetic fields make with the xz plane does not change between the object space and image space. Thus the corresponding projections of these fields are equal, as shown for the electric field case:

$$\mathbf{g}_0 \cdot \mathbf{e}_{u0} = \mathbf{g}_I \cdot \mathbf{e}_{u1} \quad (11)$$

$$(\mathbf{g}_0 \times \mathbf{s}_0) \cdot \mathbf{e}_{u0} = (\mathbf{g}_I \times \mathbf{s}_1) \cdot \mathbf{e}_{u1} \quad (12)$$

Based on the geometry of the problem we get the following expressions for the vectors, where the \mathbf{x}_u , \mathbf{y}_u and \mathbf{z}_u are the unit vectors in the direction of the axes x , y and z :

$$\mathbf{g}_0 = \text{sign}(x) \mathbf{x}_u = \text{sign}(\theta) \mathbf{x}_u \quad (13)$$

$$\mathbf{g}_1 = \mathbf{x}_u \text{sign}(\theta) \cos \theta + \mathbf{z}_u \sin \theta \quad (14)$$

$$\mathbf{s}_0 = \mathbf{z}_u \quad (15)$$

$$\mathbf{s}_1 = -\mathbf{x}_u \sin \theta + \mathbf{z}_u \cos \theta \quad (16)$$

The direction of \mathbf{e}_{u0} depends on the polarization of the incident light. Here we consider both x and y polarizations. For the incident x polarization, after some algebra we get the direction of \mathbf{e}_{u1} and consequently the value of the ray strength factor \mathbf{a}_x :

$$\mathbf{a}_x = l_{0x} \sqrt{f \cos \theta} (\mathbf{x}_u \cos \theta + \mathbf{z}_u \sin \theta) \quad (17)$$

The l_{0x} is the amplitude factor of the incident x polarization. For the incident y polarization and the corresponding amplitude factor l_{0y} the value of the ray strength factor \mathbf{a}_y is:

$$\mathbf{a}_y = l_{0y} \sqrt{f \cos \theta} \mathbf{y}_u \quad (18)$$

Writing down similar expressions for the strength factors of the magnetic field is facilitated by the relation of the magnetic and electric fields for a plane wave [6]:

$$\mathbf{h} = (\sqrt{\varepsilon/\mu}) \mathbf{s} \times \mathbf{e} \quad (19)$$

Thus we get for incident electric field polarizations x and y :

$$\mathbf{b}_x = \mathbf{y}_u l_{0x} \sqrt{f \cos \theta \varepsilon/\mu} \quad (20)$$

$$\mathbf{b}_y = -l_{0y} \sqrt{f \cos \theta \varepsilon/\mu} (\mathbf{x}_u \cos \theta + \mathbf{z}_u \sin \theta) \quad (21)$$

Assembly of the results (16), (17), (18), (20) and (21) into (1) and (2) produces the following expressions for the fields at a point $P(x_p, z_p)$:

$$\mathbf{e}(x_p, z_p) = -i \sqrt{\frac{f}{\lambda}} \int_{-\theta_{\max}}^{\theta_{\max}} \left[l_{0x}(\theta) e^{i\gamma_{0x}(\theta)} \mathbf{M} + l_{0y}(\theta) e^{i\gamma_{0y}(\theta)} \mathbf{N} \right] d\theta \quad (22)$$

$$\mathbf{h}(x_p, z_p) = -i \sqrt{\frac{f \varepsilon}{\lambda \mu}} \int_{-\theta_{\max}}^{\theta_{\max}} \left[l_{0x}(\theta) e^{i\gamma_{0x}(\theta)} \mathbf{N} - l_{0y}(\theta) e^{i\gamma_{0y}(\theta)} \mathbf{M} \right] d\theta \quad (23)$$

The definitions of \mathbf{M} and \mathbf{N} are:

$$\mathbf{M} = \begin{pmatrix} m_x \\ m_y \\ m_z \end{pmatrix} = \sqrt{\cos \theta} \begin{pmatrix} \cos \theta \\ 0 \\ \sin \theta \end{pmatrix} e^{ik(-\sin \theta x_p + \cos \theta z_p)} \quad (24)$$

$$\mathbf{N} = \begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix} = \sqrt{\cos \theta} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} e^{ik(-\sin \theta x_p + \cos \theta z_p)} \quad (25)$$

The integrals (22), (23) imply the following variable change as related to (1) and (2):

$$d\Omega = \frac{ds_x}{s_z} = d\theta \quad (26)$$

The integrals (22) and (23) are a reformulation for an original RW approach with a provision for a piecewise quasi constant phase of the incident wavefront. When this provision is suppressed by setting $\gamma_{0x}=\gamma_{0y}=0$, it can be verified that the original RW expressions reduce to the above 2D expressions by setting $\varphi=0$ and letting θ have negative values. Therefore when the integrals (22) and (23) are used below for the evaluation of the fields in the vicinity of the geometrical focal point they represent the result based on the RW approach.

3. Two dimensional focusing based on Stratton-Chu integral

The Stratton-Chu (SC) diffraction integral [3] is an exact and rigorous expression for the electric and magnetic fields in the investigated point. The only condition which is imposed in the original formulation is that the volume V upon which the integration of the incident field is performed does not contain any sources. It is also assumed that the volume is homogenous, isotropic and has zero conductivity. As it is implied in the original formulation, the Kirchhoff boundary conditions that are applied to the field on the surface of the volume reduce the accuracy of this approach. Still, this approach is believed to be more accurate than the RW approach as it does not make all the assumptions that are required for RW formulation.

The expression of the SC integral for the electric field is given in [3] as:

$$\mathbf{E}(x_p, y_p, z_p) = \frac{1}{4\pi} \int_S [i\omega\mu(\mathbf{n} \times \mathbf{H})\psi + (\mathbf{n} \times \mathbf{E}) \times \nabla\psi + (\mathbf{n} \cdot \mathbf{E})\nabla\psi] da \quad (27)$$

where $P(x_p, y_p, z_p)$ is a point where the field is evaluated in terms of the electric and magnetic fields, incident on the surface S , which bounds the volume V , in which P is located. The unit normal vector \mathbf{n} points outwards from S . The Green's function is denoted ψ . The integration is performed over the entire surface S . Application of this integral to a 2D problem requires a number of changes relative to the 3D case. Expression (27) implies that the amplitude is inverse proportional to the wavelength, thus multiplication by a square root of the wavelength is required. Similarly, the Green's function expression in 2D is [7]:

$$\psi = \pi H_0^1(kr) \xrightarrow{kr \gg 1} \sqrt{\frac{\lambda}{r}} e^{i(kr + \pi/4)} \quad (28)$$

It must be noted that the absolute value r of the distance vector \mathbf{r} , between the source point and P , is independent of the y coordinate since it is assumed to be zero elsewhere in the 2D problem. Accordingly, the gradient of the Green's function does not have elements in y direction:

$$\nabla\psi = \frac{e^{ikr}}{\sqrt{r}} \frac{ikr - 1}{r} (-\mathbf{r}_u) \quad (29)$$

Where \mathbf{r}_u is a unit vector in the direction of \mathbf{r} . For a monochromatic wave in vacuum we can restate the coefficient of the first term with help of intrinsic impedance of free space η_0 :

$$i\omega\mu = \eta_0 ik \approx 120\pi ik \quad (30)$$

Substitution of (28), (29) and (30) into (27) gives:

$$\begin{aligned}
\mathbf{E}(x_p, 0, z_p) &= \\
&= \frac{-\sqrt{\lambda}}{4\pi} \int_S \left\{ ik\eta_0 (\mathbf{n} \times \mathbf{H}) - [(\mathbf{n} \times \mathbf{E}) \times \mathbf{r}_u + (\mathbf{n} \cdot \mathbf{E}) \mathbf{r}_u] \frac{ikr-1}{r} \right\} \frac{e^{ikr}}{\sqrt{r}} da \quad (31)
\end{aligned}$$

In a similar way an expression for the magnetic field is (a starting expression can be found in [8]):

$$\begin{aligned}
\mathbf{H}(x_p, 0, z_p) &= \\
&= \frac{\sqrt{\lambda}}{4\pi} \int_S \left\{ \frac{ik}{\eta_0} (\mathbf{n} \times \mathbf{E}) + [(\mathbf{n} \times \mathbf{H}) \times \mathbf{r}_u + (\mathbf{n} \cdot \mathbf{H}) \mathbf{r}_u] \frac{ikr-1}{r} \right\} \frac{e^{ikr}}{\sqrt{r}} da \quad (32)
\end{aligned}$$

These integrals (31) and (32) can be used for the evaluation of the focused field from the known wavefront converging in the image space of the optical system.

We turn now to the definition of this converging wavefront. Here we deal with the same physical problem that is described in the previous section. In this problem the only source is located at infinity in the negative direction of the z axis. As previously, making the same assumptions on the incident illumination and the optical system, we can use the Gaussian cylinder for the description of light emerging from the optical system to its image space. We can use the complete geometrical form of the Gaussian cylinder as the integration volume V . The only field that is incident on S , the external surface of V , is the light emerging from the optical system. As the considered problem is infinite in the y direction we can analyze it on a 2D plane perpendicular to the y axis, which is chosen as xz plane, similarly to the RW case. Thus the actual SC integration is performed along the intersection path between S and the xz plane. As previously, we assume the Kirchhoff boundary condition which gives the fields on S in terms of rays passing through the optical system. Certainly, imposing these conditions reduces the accuracy of the SC result and requires that the point P is located far enough from the Gaussian cylinder surface.

The electric and magnetic fields, incident on S can be expressed using (9), (17), (18), (19), (20), (21) as follows:

$$\mathbf{E} = \sqrt{\cos \theta} \left[l_{0x} \exp(i\gamma_{0x}) (\mathbf{x}_u \cos \theta + \mathbf{z}_u \sin \theta) + l_{0y} \exp(i\gamma_{0y}) \mathbf{y}_u \right] \quad (33)$$

$$\mathbf{H} = \sqrt{\cos \theta \varepsilon / \mu} \left[l_{0x} \exp(i\gamma_{0x}) \mathbf{y}_u - l_{0y} \exp(i\gamma_{0y}) (\mathbf{x}_u \cos \theta + \mathbf{z}_u \sin \theta) \right] \quad (34)$$

We have dropped the eikonal from the expressions of the fields assuming that any phase differences between the two incident polarizations are adequately represented in the phase deviations γ_{0x} and γ_{0y} . Substituting (33) and (34) into (31) and (32) evaluates the focused field in a slightly different way than it is done by the Debye-Wolf integral. Introducing some approximations it can be shown that the SC integral is reduced to the RW integral [9].

4. Focused field quadratic phase discussion

It is a well known fact that a Fourier transform of an object, as obtained at the back focal plane of a paraxial focusing optical system, is multiplied by some quadratic phase. Mathematically, this quadratic phase factor can be derived by the Fresnel-Kirchhoff integral ([10] Eq. 5-19):

$$U_f(x_f, y_f) = \frac{A \exp \left[\frac{jk}{2f} \left(1 - \frac{d_0}{f} \right) (x_f^2 + y_f^2) \right]}{j\lambda f} \cdot \iint_{-\infty, +\infty} t_0(x_0, y_0) \exp \left[-\frac{jk}{f} \left(1 - \frac{d_0}{f} \right) (x_0 x_f + y_0 y_f) \right] dx_0 dy_0 \quad (35)$$

For the moment, we neglect the aperture of the lens in the above expression. The x_0 and y_0 are the coordinate axes normal to the direction of the axis of optical system z at the position of the object, located at a distance d_0 from the lens, as shown in Fig. 2. The object is represented by the transfer function $t_0(x_0, y_0)$. A coordinate system of x_f and y_f axes is defined at the back focal plane of the lens having focal length of f . Thus, according to (35), when there is no object and t_0 represents the optical system aperture, the above quadratic phase is expected at the focal plane unless the aperture is placed exactly at the lens front focal plane.

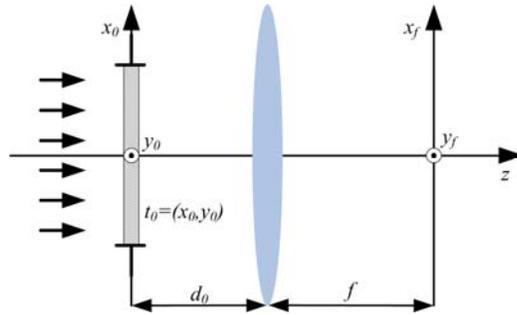


Fig. 2. Focusing optical system.

The corresponding analysis of the focal field by Debye integral, as presented in [6] produces a piecewise quasi constant phase at the focal plane of an aplanatic optical system. So is the case with the Debye-Wolf integral which is a rigorous vector extension of the Debye integral. Consequently when a phase obtained by RW focusing of a plane wave is evaluated, it does not contain the quadratic phase either. The lack of the quadratic phase is not the only discrepancy of the Debye approach. It is known that in optical systems with low Fresnel number the results of Debye integral become less accurate until the approach completely breaks down for systems with Fresnel number below 1 [4]. It must be noted that the above comparison considers optical systems of moderate numerical apertures, such as 0.2, where both approaches have reasonable accuracy. The same effects are observed in a 2D analysis.

Considering a focusing problem in 2D enables us to apply the SC integral, as discussed above, with an affordable computational complexity. As it will be shown in the next section, the focal field evaluated by the SC integral does have the quadratic phase, similar to the one produced by the Fresnel-Kirchhoff integral. It must be stressed that the SC and RW approaches integrate the same field incident on the same Gaussian cylinder, but they produce different results. Interestingly, comparison of the amplitudes of the focused fields between the SC and RW results does not reveal significant differences, as opposed to the comparison of the phases.

The quadratic phase of the focused field, as produced by SC approach appears to depend on the numerical aperture of the considered optical system. For lower numerical apertures it approaches the quadratic phase predicted by the Fresnel-Kirchhoff integral, while for higher numerical apertures it approaches zero as the Debye approach suggests. The quantitative analysis of the dependence of the quadratic phase on the numerical aperture is performed in the next section. As the next section considers aplanatic optical system illuminated by a plane wave, the aberration of the wavefront on the Gaussian cylinder Φ is taken to be zero

everywhere. Thus, according to [6], the observed quadratic phase is not connected with primary or higher order optical aberrations.

5. Numerical analysis

5.1 Introductory analysis of a low numerical aperture case

We start by considering a focusing optical system with a relatively low numerical aperture of 0.2. As such, it can be schematically described as in Fig. 1 or in Fig. 2. The system is illuminated from the left by a plane wave propagating in the positive z direction. For the paraxial analysis, the lens aperture is represented by the amplitude transmittance function t_0 . The paraxial representation of a lens as a thin optical element allows to assume that $d_0=0$. Thus the focal field distribution can be calculated based on (35) or by using paraxial optical operators [11]. Both rigorous approaches, RW and SC start with ray representation of the incident illumination at the entrance pupil. This representation neglects the diffraction effects of the entrance pupil as the rays are either stopped by the pupil or continue to the image space. Thus the most natural position for modeling the entrance pupil is at the lens surface. All the rays reaching the image space are described on the surface of the Gaussian cylinder forming an emerging wavefront in the image space. This emerging wavefront is integrated for the focused field evaluation by either Debye-Wolf or Stratton-Chu integrals.

In an optical system, its angular aperture is generally different between the object side and image side [6]. The analysis, presented in this work, is performed having an association of the investigated focusing optical system with microscope objectives in mind. Microscope objectives were originally intended for imaging of small, luminous, species, while in this work we consider a reverse direction of light propagation through the objective as we speak of focusing light. Thus the object space of a microscope objective corresponds to the image space of the focusing optical system presented in the previous sections, and consequently the image space of the microscope objective corresponds to the object space of the investigated optical system. According to the above and to the definition in [6] the relation between the numerical aperture (NA) of the optical system and its aperture angle θ_{max} is given by:

$$NA = n \sin(\theta_{max}) \quad (36)$$

Where n is the refractive index of the optical system image space. In this work we assume $n = 1$. From (36) and the sine condition (5) we can get the half-width of the entrance pupil:

$$h_{max} = f \sin(\theta_{max}) = f \cdot NA \quad (37)$$

For consistency between the rigorous and the paraxial approaches, the amplitude transmittance function t_0 also has the half-width h_{max} . In this work we choose $h_{max} = 1.5\text{mm}$. Thus for $NA = 0.2$ the focal length becomes $f = 7.5\text{mm}$. It must be noted that the exit pupil has a different half-width but it is not relevant to the present work.

The plane wave, incident on the entrance pupil of the optical system can have an arbitrary polarization. In this work we consider two orthogonal polarizations of the incident plane wave x and y . These polarizations span all the other polarizations (linear, circular and elliptic). The paraxial, scalar approach does not make any distinction between the polarizations. Therefore the paraxial results are compared with the focal field component in the direction of the incident illumination polarization.

The focused field amplitudes, as evaluated by different methods, are presented in Fig. 3. The values of the amplitudes are given in percents, relative to the maximum amplitude value, evaluated by (35). The amplitudes, evaluated by different methods are too close to see the difference even when the main lobe is zoomed over the entire graph. Presenting the difference of the amplitudes from the amplitude calculated analytically by (35) gives a better perspective, as shown in Fig. 4. Again the differences are given in percents, relative to the same value as in Fig. 3.

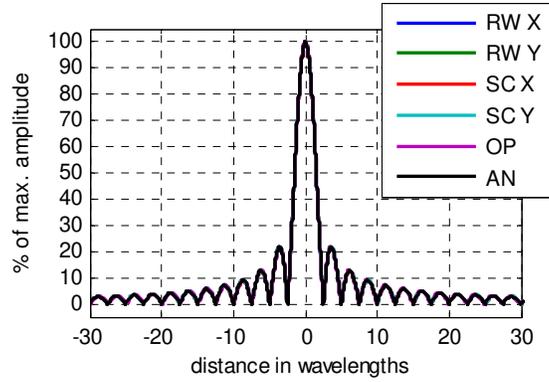


Fig. 3. Focused field amplitudes. Prefix meaning: RW – evaluation by RW method, SC – evaluation by SC method, OP – evaluation by paraxial operators, AN – evaluation by analytic paraxial expression. Suffix meaning: X – incident x polarization, Y – incident y polarization.

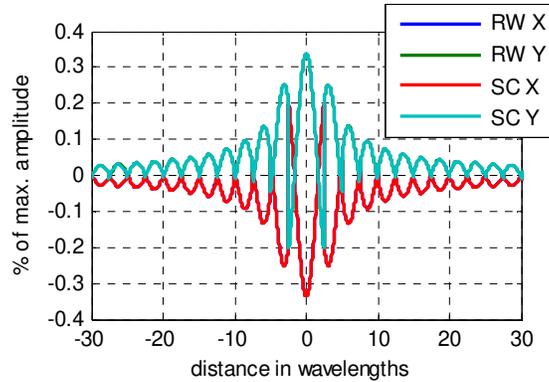


Fig. 4. Rigorous focused field amplitude differences from the analytical paraxial result. Prefix meaning: RW – evaluation by RW method, SC – evaluation by SC method. Suffix meaning: X – incident x polarization, Y – incident y polarization.

In Fig. 4, the line corresponding to SC X completely overlaps (and hides) the line corresponding to RW X; similarly, SC Y overlaps RW Y. The differences between the overlapping lines are of an order of 0.0005%.

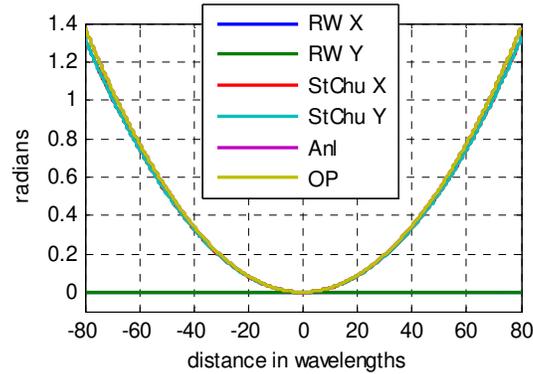


Fig. 5. Wavefront phase after the removal of “sync” jumps. Prefix meaning: RW – evaluation by RW method, SC – evaluation by SC method. Suffix meaning: X – incident x polarization, Y – incident y polarization.

We now turn to the investigation of the phase of the focused field components corresponding to the incident illumination polarization and the phase of the complex focused wavefront in the paraxial case. It is implied by Fig. 3 that there are multiple zero crossings of the amplitude along the x axis, which are due to the “sinc”-like behavior of the complex function representing the wavefront. The zero crossings incur phase jumps of π . These jumps are unwanted as they complicate the analysis of the quadratic phase of the wavefront, which is the investigation subject in this work. Clearing those jumps can be accomplished by multiplying the wavefront by the sign of its real or imaginary part. The phase remaining after the multiplication is shown in Fig. 5. There are three groups of overlapping graphs in Fig. 5. The first group comprises paraxial results, where the analytically calculated quadratic phase overlaps the phase of the wavefront evaluated by the paraxial operators. The second group comprises the phases of the results obtained by SC approach, where the phase corresponding to the incident y polarized plane wave overlaps and hides the phase corresponding to x polarized incident plane wave. Their phase also looks like quadratic, but it appears to have less steep slope, which can be a result of a smaller coefficient. The third group comprises the phase of RW approach results. Again the phases corresponding to different incident illuminations overlap and they both are zero.

We can summarize that as far as the paraxial approximation holds, both paraxial and rigorous methods, discussed herein, are in a good agreement regarding the amplitude of the focused field in the direction of the incident illumination polarization. When considering the focused field phase, the RW approach, based on the Debye integral, does not produce the quadratic phase, as opposed to the paraxial approach, based on the Fresnel-Kirchhoff integral. The approach, based on the SC integral, yields a quadratic phase which is slightly different from the paraxial result. As it was discussed in the earlier sections, the SC approach involves fewer approximations than the RW approach. The results presented above suggest that the field evaluated by the SC approach has a better correspondence to the paraxial results. Thus it can be suggested that SC approach is more accurate than RW approach also when optical systems of higher numerical aperture are concerned. It must be noted that these conclusions are independent of the polarization of the incident illumination.

5.2 Amplitude evaluation for higher numerical apertures

To the best of our knowledge, this work is the first one that uses SC integral for tight focusing. Therefore it is important to verify that the focused field amplitude on the back focal plane, as calculated by the SC approach corresponds to that evaluated by the widely accepted RW approach. For this purpose we evaluate the amplitude of the focal field by the two approaches for different numerical apertures of an optical system. We choose to keep the $h_{max} = 1.5\text{mm}$, thus different numerical aperture values are achieved by different focal length values, according to (37). Practically, this also means that the number of samples taken on the Gaussian cylinder changes with different arc lengths defined by the angle θ_{max} and different focal distances f , as the number of samples, per wavelength distance along the arc, is kept constant.

Presenting focused field amplitudes similarly to Fig. 3 would give no visible indication about the differences between them. Thus we present the absolute differences between the amplitude results calculated by RW and SC, which are normalized by the maximum amplitude value of the SC result. Correspondingly, these absolute differences are presented in percents of the maximum amplitude for E_x field component in Fig. 6, E_y field component in Fig. 7 and E_z field component in Fig. 8.

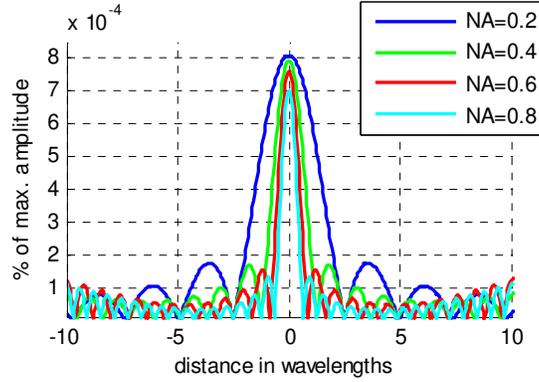


Fig. 6. Absolute amplitude difference of the E_x field component normalized by the maximum amplitude value. Different lines correspond to different numerical apertures.

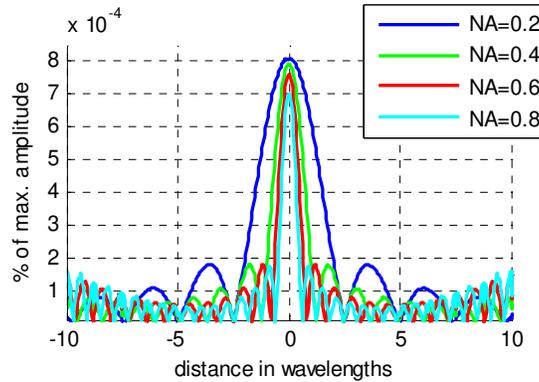


Fig. 7. Absolute amplitude difference of the E_y field component normalized by the maximum amplitude value. Different lines correspond to different numerical apertures.

The differences are presented for a selected subset of numerical apertures. Generally, in the vicinity of the geometrical focal point, the difference is limited by 10^{-5} of the maximum amplitude values for all the field components. It is interesting to note that the difference in the center is smaller for higher numerical apertures. This comes in agreement with [4], stating that the Debye approach is more accurate for optical systems with higher Fresnel numbers. It is also evident that the result for E_z field, in Fig. 8, contains a significant number of point-wise numerical errors, which overlay a difference pattern that is otherwise similar to that of the other fields.

The Debye approximation represents each ray that is directed to the geometrical focus from the Gaussian cylinder as a plane wave in the vicinity of the geometrical focus. This approximation is expected to become less accurate when higher distances from the geometrical focal point are considered. Thus we expect that the differences between the RW approach that uses Debye approximation and the SC approach that does not make this approximation, become greater with a greater distance from the geometrical focal point. In order to make a qualitative analysis of the dependence of the differences on the distance from the geometrical focal point we need to perform some averaging of the curves presented in Figs. 6-8. The averaging is required for better visualization of the differences that behave in an oscillatory way with amplitude increasing with the distance from the geometrical focus.

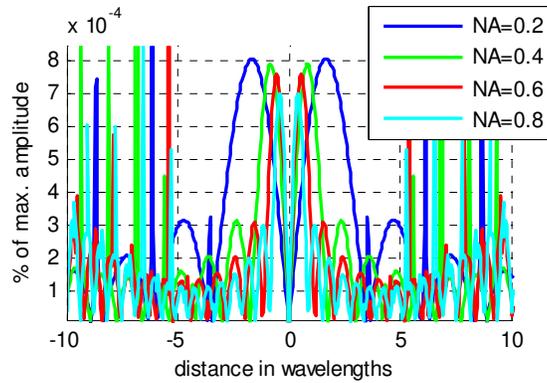


Fig. 8. Absolute amplitude difference of the E_z field component normalized by the maximum amplitude value. Different lines correspond to different numerical apertures.

The averaged difference is, of course, only an approximation that shows a general trend or contour, rather than a result having an exact quantitative meaning. The averaged differences are shown in Figs. 9, 10 and 11. One can notice that with a small exception for the E_x field, the difference increase with a distance from the geometrical focus becomes steeper with higher numerical aperture of the system. A possible explanation can be a comparison between the plane waves and the spherical (cylindrical in 2D) waves. Each point on the Gaussian cylinder is both a representative of a ray directed to the geometrical focus and a secondary source of cylindrical waves, according to Huygens principle. In the former case it is evaluated as a plane wave in the vicinity of the focal point and in the latter case it is a cylindrical wave with radius equal to the focal distance. With increasing numerical aperture this radius becomes smaller (as the entrance pupil radius is kept constant) and therefore the difference between the plane wave and the cylindrical wave becomes larger when the same distance from the focal point is considered.

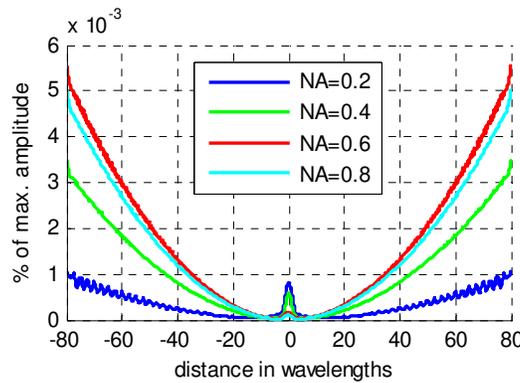


Fig. 9. Approximate contours of absolute amplitude difference of the E_x field component normalized by the maximum amplitude value. Different lines correspond to different numerical apertures.

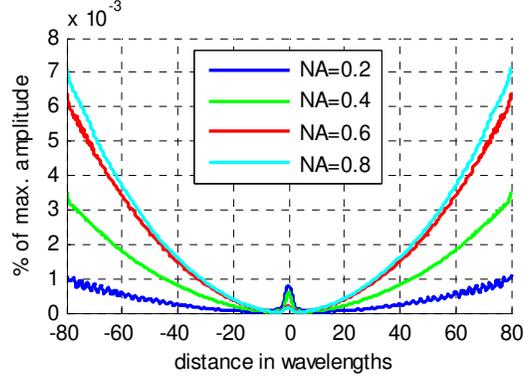


Fig. 10. Approximate contours of absolute amplitude difference of the E_y field component normalized by the maximum amplitude value. Different lines correspond to different numerical apertures.

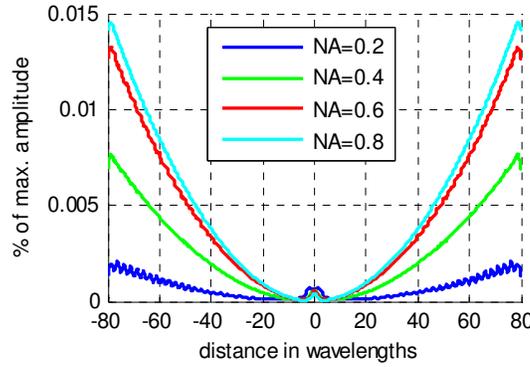


Fig. 11. Approximate contours of absolute amplitude difference of the E_z field component normalized by the maximum amplitude value. Different lines correspond to different numerical apertures.

5.3 Quadratic phase evaluation for higher numerical apertures

We turn now to the investigation of the focused field quadratic phase, as produced by the SC approach for different numerical apertures of an optical system. The extraction of the quadratic phase associated with a specific field component was described in subsection 5.1. The quadratic phase is evaluated for optical systems with different numerical apertures, as discussed in section 5.2. As previously, we keep the $h_{max} = 1.5\text{mm}$, thus different numerical aperture values are achieved by different focal length values, according to (37). After phase jumps removal and a certain amount of smoothing, the quadratic phase of the E_x field looks like the one shown in Fig. 12. It is interesting to note that at first the quadratic phase becomes steeper with increasing numerical aperture, but after a certain point it begins to decrease with increasing numerical aperture.

In the 2D system the E_x and E_z fields correspond solely to the incident x polarized illumination and the E_y field corresponds solely to the incident y illumination. Thus it is important to investigate the quadratic phase of all field components for the corresponding incident illumination polarizations. A quantitative measure of a quadratic phase coefficient can be obtained by polynomial fitting of the phase curves as depicted in Fig. 12. The absolute value of the quadratic phase coefficient (reflecting such optical system parameters as focal length of wavelength) is given in Fig. 13, as a function of the numerical aperture. From Fig.

13 it is evident that the numerical aperture after which the quadratic phase factor starts decreasing is about 0.6.

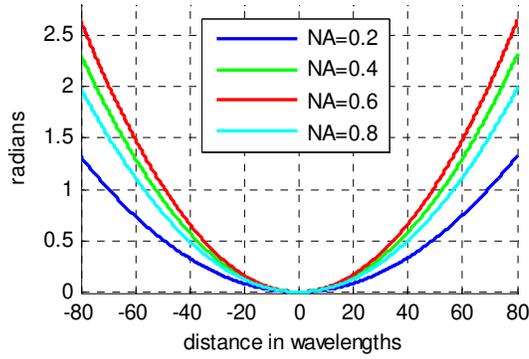


Fig. 12. Quadratic phase of the E_x field component. Different lines correspond to different numerical apertures.

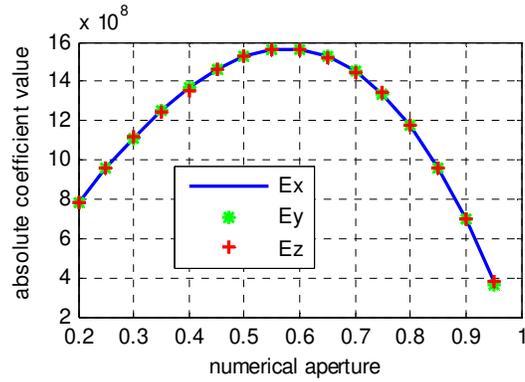


Fig. 13. Absolute value of quadratic phase coefficients of different field components as a function of numerical aperture.

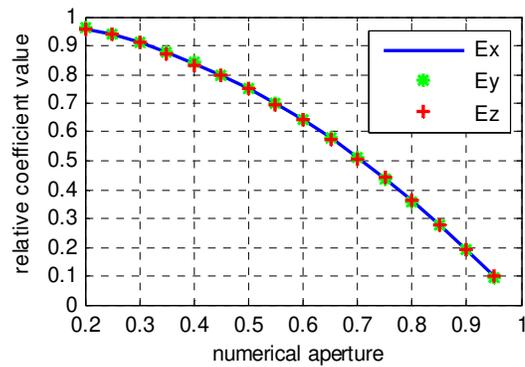


Fig. 14. Relative value of quadratic phase coefficients of different field components as a function of numerical aperture.

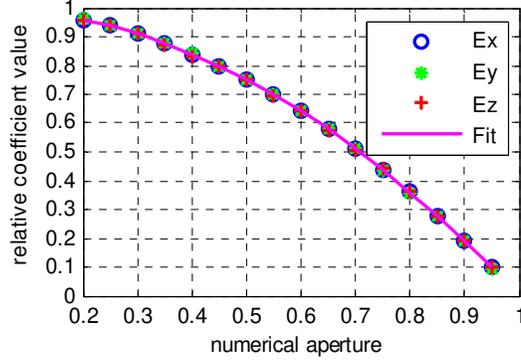


Fig. 15. Relative value of quadratic phase coefficients of different field components compared with a corresponding analytical expression, as a function of numerical aperture. The result corresponding to analytical expression is denoted Fit.

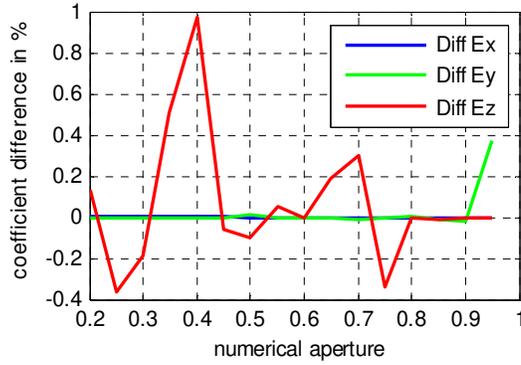


Fig. 16. Difference shown as a function of numerical aperture, between the analytically calculated relative quadratic phase coefficient and corresponding results, derived from different field components. Except for the E_z field and E_y field at $NA=0.95$, other difference results are below 0.02%.

A deeper insight on the behavior of the rigorous quadratic phase coefficient c_v as a function of numerical aperture can be obtained by normalizing it with the quadratic phase coefficient of the paraxial case c_p , appearing in (35):

$$c_{v,norm} = \frac{c_v}{c_p} = c_v \frac{2f}{k} = c_v \frac{f\lambda}{\pi} \quad (38)$$

The rigorous quadratic phase coefficient $c_{v,norm}$, relative to the paraxial quadratic phase coefficient c_p , is shown in Fig. 14, for all the field components, as a function of numerical aperture. The form of the curve in Fig. 14 hints that it can have a polynomial dependence on the numerical aperture. Evaluating polynomial coefficients for all the field components suggests that $c_{v,norm}$ is fitted by the following polynomial:

$$c_{v,norm} = 1 - NA^2 \quad (39)$$

The comparison of the relative coefficients obtained by the numerical field evaluation with the coefficient given by (39) is shown in Fig. 15. For better perception of the actual differences between the compared values, they are shown separately in Fig. 16. The differences corresponding to the E_x and E_y fields are mainly limited by 2×10^{-4} , while the difference

corresponding to the E_z field is higher. This may be due to a higher level of numerical errors in E_z , as evident from Fig. 8.

The investigation above leads to our main result, the expression of the quadratic phase coefficient at the back focal plane of the focusing system:

$$c_v = \frac{k}{2f}(1 - NA^2) = \frac{k}{2f} \left[1 - \left(\frac{h_{\max}}{f} \right)^2 \right] \quad (40)$$

This result shows that the quadratic phase coefficient, derived from the results of the rigorous, vector methods, approaches the value obtained by paraxial analysis for systems with decreasing numerical aperture. At the other extreme, with NA approaching 1, the coefficient approaches zero and the phase function, at the back focal plane, approaches piecewise planar profile.

The result in (40) was obtained for a plane wave illumination at the entrance pupil of the focusing system. Making an analogy with a paraxial case, it can be expected that this result is a property of the optical system rather than of the incident illumination. The verification of this conclusion has yet to be performed.

The interesting result, as obtained above, for a 2D system, hints that some similar quadratic phase behavior can be expected in a 3D system. Given a very high computational complexity, required to perform such an analysis of a 3D system, we can not present any definite conclusions so far. Results of some preliminary 3D system simulations suggest that the quadratic phase of all the three field components of the focused field is close to the one described by (40) for some specific numerical aperture values. Yet, much more work has to be done in this direction.

5. Conclusion

We have used the Stratton-Chu diffraction integral to investigate the phase of the focused field of a 2D optical system. We have found that this phase has a quadratic component, similar to that of the Fresnel-Kirchhoff diffraction integral, as opposed to the result of Debye-Wolf integral which does not produce any quadratic phase component. The coefficient of the investigated quadratic phase was found to be dependent on the square of the numerical aperture as $1-NA^2$. This result was consistent for all field components. A closed analytical form of the quadratic phase coefficient will allow an easy addition of the quadratic phase to focal field values, obtained by Richards-Wolf approach so as to improve the accuracy of the subsequent analysis. If, as we believe, the discussed quadratic phase is independent of the incident illumination, it can be added to the results of any method based on RW approach. Yet, a care should be taken when considering 3D systems, as their proper investigation is still ahead.

Acknowledgments

This work was performed within the NANOPRIM project supported by the European Community under the FP6 program. The authors are pleased to thank P. Einziger for providing constructive comments.