Strehl ratio for focusing into biaxially birefringent media

Sjoerd Stallinga

Philips Research Laboratories, Professor Holstlaan 4, 5656 AA Eindhoven, The Netherlands

Received February 19, 2004; revised manuscript received June 11, 2004; accepted June 21, 2004

The Strehl ratio for focusing into biaxially birefringent media with small birefringence is investigated. An analytical expression for the Strehl ratio as a function of the polarization and birefringence parameters is derived in the paraxial limit from a generalization of the scalar aberration theory. The expression agrees well with results of a numerical calculation.

OCIS codes: 260.1440, 260.1960, 210.4770, 180.1790.

1. INTRODUCTION

The present paper is an extension of the investigation into the light distribution close to focus in biaxially birefringent media.[1] The theory of Ref. 1 is based on the Debye approach to vectorial diffraction and on the small-birefringence approximation. A discussion of these basic approximations, along with references to related work on the focal light distribution in various circumstances, can also be found in Ref. 1.

A good measure of the effect of aberrations on the imaging quality of a lens is the Strehl ratio, which is defined as the ratio between the peak intensity of the aberrated focal spot and that of the unaberrated focal spot. There is a simple connection between this Strehl ratio and the aberration function W in the scalar theory of diffraction, namely,[2]

$$I = 1 - (\langle W^2 \rangle - \langle W \rangle^2), \tag{1}$$

where the angle brackets indicate an average over the pupil. This relation is valid for small aberrations, i.e., when the resulting Strehl ratio is not much smaller than 1. According to the Maréchal criterion[2] the focal spot is diffraction-limited if the Strehl ratio is above 0.8. When the Strehl ratio drops below this limit the blurring of the focal spot is considered to be too much for reliable imaging of the smallest resolvable details [which are of the order of λ/2NA (numerical aperture)]. For defocus the criterion corresponds to a peak–valley aberration of λ/4. This implies that the plane waves that make up the focal field in the Debye approach (each plane wave corresponding to a different pupil point) can interfere destructively for defocus values larger than the diffraction limit. This result is simply generalized to all other aberrations by the Maréchal criterion.

It is the aim of this paper to use the theory of Ref. 1 to find an expression for the Strehl ratio of the birefringence-induced spot distortion. Such an analytical expression would greatly simplify the analysis of how much birefringence can be tolerated in an imaging system. It is proposed here to generalize the Maréchal criterion of scalar aberrations, i.e., a spot that is distorted by birefringence effects is said to be diffraction-limited if the Strehl ratio is larger than 0.8. Such a tolerance criterion is particularly useful in the field of data storage on optical disks. In the readout of optical disks the beam is focused onto the data layer through a plastic substrate/cover layer. Birefringence effects can occur in the substrate/cover layer of optical disks as a result of mechanical stresses that are introduced into the plastic during manufacturing.

The content of this paper is as follows. The results of the previous study[1] are briefly summarized in Section 2. An analytical expression for the Strehl ratio is derived from a generalization of the scalar aberration theory in Section 3. The paper is concluded with a discussion of the main results in Section 4.

2. LIGHT DISTRIBUTION CLOSE TO FOCUS IN THE DEBYE APPROACH

Consider a collimated beam of light of wavelength λ propagating in the z direction that is focused by a lens into a biaxially birefringent medium with refractive indices n_1, n_2, and n_3 that deviate slightly from the average refractive index $\bar{n}$:

$$n_1 = \frac{\bar{n}}{(1 - \Delta n_\parallel/\bar{n})^{1/2}} \approx \bar{n} + \Delta n_\parallel/2, \tag{2}$$

$$n_2 = \frac{\bar{n}}{(1 + \Delta n_\parallel/\bar{n})^{1/2}} \approx \bar{n} - \Delta n_\parallel/2, \tag{3}$$

$$n_3 = \frac{\bar{n}}{(1 - 2\Delta n_\perp/\bar{n})^{1/2}} \approx \bar{n} + \Delta n_\perp, \tag{4}$$

where $\Delta n_\parallel$ is the lateral birefringence and $\Delta n_\perp$ is the axial birefringence. The third of the three principal axes of the biaxial medium is taken to be parallel to the optical axis (z axis); the first and second axes make an angle γ with the x axis and y axis, respectively.

The cone of light converging to focus has a top angle 2β, making NA $= \bar{n} \sin\beta$. In the Debye approach the electromagnetic field close to the focal point is expressed as a superposition of plane waves. Each of these plane waves

1084-7529/2004/122406-08$15.00

corresponds to a point on the reference sphere of radius R. The propagation direction of the plane wave is the line between the corresponding point on the reference sphere and the (geometrical) focal point of the lens. Associated with each point on the reference sphere are three orthonormal vectors,

$$\hat{p} = (\cos\theta\cos\phi,\ \cos\theta\sin\phi,\ -\sin\theta), \tag{5}$$

$$\hat{s} = (-\sin\phi,\ \cos\phi,\ 0), \tag{6}$$

$$\hat{k} = (\sin\theta\cos\phi,\ \sin\theta\sin\phi,\ \cos\theta), \tag{7}$$

where θ and ϕ are the polar and azimuthal angle, respectively, parameterizing the reference sphere. The wave vector of the plane wave associated with each point is $\mathbf{k} = \bar{n}k\hat{k}$ with $k = 2\pi/\lambda$, and this plane wave contributes to focus if the polar angle $\theta \leqslant \beta$. This sharp cutoff in the angular spectrum of plane waves is the basic assumption of the Debye approach. The validity of this approach is discussed in detail in Refs. 3 and 4. The vectors $\hat{p}$ and $\hat{s}$ are the two basis polarization vectors. In the small-birefringence approximation the component of the electric field along $\mathbf{k}$ may be neglected, implying that the electric field is in the plane spanned by $\hat{p}$ and $\hat{s}$, just as the dielectric displacement is. Instead of the polar angle θ the radial pupil coordinate $\rho = \sin\theta/\sin\beta$ is often used. This maps the part of the reference sphere for $\theta \leqslant \beta$ to the unit circle.

The electric field $\mathbf{E}$ at a point $\mathbf{r}_p$ in the proximity of focus is given by

$$E_\alpha(\mathbf{r}_p) = \frac{E_0 \pi N}{i} \sum_{j=1,2} F_{\alpha j}(\mathbf{r}_p) A_j, \tag{8}$$

for $\alpha = x, y, z$. Here E_0 is the electric field amplitude in the entrance pupil; $N = R\sin^2\beta/\lambda$ is the Fresnel number; A_1 and A_2 are the x and y components of the polarization vector in the entrance pupil, respectively, normalized such that $|A_1|^2 + |A_2|^2 = 1$; and the functions $F_{\alpha j}$ are defined by the integrals over the unit circle R that maps the part of the reference sphere with $\theta \leqslant \beta$,

$$F_{\alpha j}(\mathbf{r}_p) = \frac{1}{\pi}\int_R \mathrm{d}^2\rho \sum_{l=1,2} \hat{v}_{l\alpha} J_{lj} \exp(i\mathbf{k}\cdot\mathbf{r}_p), \tag{9}$$

where $\hat{v}_l$ are the polarization vectors and where J is a 2×2 matrix, the Jones matrix. All quantities depend on the pupil coordinates ρ and ϕ, but the phase is the only quantity that depends on the image position $\mathbf{r}_p$. The polarization vectors are a linear combination of $\hat{p}$ and $\hat{s}$:

$$\hat{v}_1 = \cos\phi\hat{p} - \sin\phi\hat{s}, \tag{10}$$

$$\hat{v}_2 = \sin\phi\hat{p} + \cos\phi\hat{s}. \tag{11}$$

In the small-birefringence approximation, the Jones matrix is the product of the birefringence Jones matrix J_{bir} and the entrance-interface Jones matrix J_{en}:

$$J = J_{\mathrm{bir}} J_{\mathrm{en}}. \tag{12}$$

There are two polarization eigenmodes in the biaxial medium with aberration functions W_+ and W_-. They correspond to two orthogonal polarization fields that are a linear sum of the p and s polarization fields. The angle between the +mode polarization vector and $\hat{p}$ (which is equal to the angle between the −mode polarization vector and $\hat{s}$) is $\chi - \phi$. The birefringence Jones matrix is then

$$\begin{aligned} J_{\mathrm{bir}} &= R(\chi)\begin{bmatrix}\exp(iW_+) & 0\\ 0 & \exp(iW_-)\end{bmatrix}R(-\chi)\\ &= \begin{bmatrix}\exp(iW_+)\cos^2\chi + \exp(iW_-)\sin^2\chi & [\exp(iW_+) - \exp(iW_-)]\sin\chi\cos\chi\\ [\exp(iW_+) - \exp(iW_-)]\sin\chi\cos\chi & \exp(iW_+)\sin^2\chi + \exp(iW_-)\cos^2\chi\end{bmatrix}. \end{aligned} \tag{13}$$

Here $R(\chi)$ is the rotation matrix:

$$R(\chi) = \begin{bmatrix}\cos\chi & -\sin\chi\\ \sin\chi & \cos\chi\end{bmatrix}. \tag{14}$$

The aberration functions turn out to be

$$W_\pm = \frac{kd}{2\cos\theta}[a \pm (b^2 + c^2)^{1/2}], \tag{15}$$

where d is the distance between the entrance interface and the focal point, and the angle χ follows from

$$\cos(2\chi - 2\phi) = b/(b^2 + c^2)^{1/2}, \tag{16}$$

$$\sin(2\chi - 2\phi) = c/(b^2 + c^2)^{1/2}, \tag{17}$$

where the quantities a, b, and c are defined as

$$a = \Delta n_\perp \sin^2\theta - \tfrac{1}{2}\Delta n_\parallel \sin^2\theta\cos(2\gamma - 2\phi), \tag{18}$$

$$b = \Delta n_\perp \sin^2\theta + \tfrac{1}{2}\Delta n_\parallel (1 + \cos^2\theta)\cos(2\gamma - 2\phi), \tag{19}$$

$$c = \Delta n_\parallel \cos\theta\sin(2\gamma - 2\phi). \tag{20}$$

The interface Jones matrix depends on the Fresnel coefficients t_p and t_s for the interface between the isotropic medium of refractive index n_0 between the lens and the birefringent medium and the birefringent medium itself, where the birefringent medium is approximated as an isotropic medium of refractive index $\bar{n}$,

$$t_p = \frac{2n_0\cos\theta_0}{n_0\cos\theta + \bar{n}\cos\theta_0}, \tag{21}$$

$$t_s = \frac{2n_0\cos\theta_0}{n_0\cos\theta_0 + \bar{n}\cos\theta}, \tag{22}$$

where θ_0 and θ are related by Snell's law $n_0\sin\theta_0 = \bar{n}\sin\theta$. This gives the interface Jones matrix as

$$J_{\text{en}} = R(\phi)\begin{bmatrix} t_p & 0 \\ 0 & t_s \end{bmatrix} R(-\phi) = \begin{bmatrix} t_p \cos^2\phi + t_s \sin^2\phi & (t_p - t_s)\sin\phi\cos\phi \\ (t_p - t_s)\sin\phi\cos\phi & t_s\cos^2\phi + t_p\sin^2\phi \end{bmatrix}. \tag{23}$$

The Jones matrix must be multiplied by an apodization function B that represents the amplitude variation over the reference sphere. The apodization function is given by

$$B = \frac{1}{\sqrt{\cos\theta_0}} = (1 - \rho^2 \text{NA}^2/{n_0}^2)^{-1/4}, \tag{24}$$

assuming an aplanatic lens and a uniform amplitude in the entrance pupil.

The energy density of the electric field is $U = \frac{1}{2}\epsilon_0 \epsilon_{\alpha\beta} E_\alpha E_\beta^*$, where the dielectric tensor $\epsilon_{\alpha\beta}$ is assumed to be real. This may be approximated by the sum over the energy densities of the x, y, and z components of the electric field, where the energy density of the α component of the electric field is $U_\alpha = \frac{1}{2}\epsilon_0 \bar{n}^2 |E_\alpha|^2$ because the difference between the refractive indices is assumed to be much smaller than the average refractive index $\bar{n}$. It is found that

$$U_\alpha = U_0 \sum_{j,l=1,2} F_{\alpha j} F_{\alpha l}^* A_j A_l^* = U_0 \sum_{\mu=0,3} I_{\alpha\mu} M_\mu, \tag{25}$$

with

$$U_0 = \tfrac{1}{2}\epsilon_0 \bar{n}^2 \pi^2 {E_0}^2 N^2, \tag{26}$$

$$I_{\alpha\mu} = \tfrac{1}{2} \sum_{j,l=1,2} (\sigma_\mu)_{jl} F_{\alpha j} F_{\alpha l}^*, \tag{27}$$

and the Stokes parameters

$$M_\mu = \sum_{j,l=1,2} (\sigma_\mu)_{lj} A_j A_l^*. \tag{28}$$

The unit matrix σ_0 and the three Pauli matrices $\sigma_j\,(j = 1, 2, 3)$ have been used in these equations (see Appendix A).

3. STREHL RATIO AND A GENERALIZED ABERRATION THEORY

In this section the scalar aberration theory is generalized and an expression is derived for the Strehl ratio as a function of the birefringence parameters. We will restrict the calculation to the paraxial limit NA $\ll 1$. The use of the Debye approach is then justified provided that the Fresnel number is still much larger than one. The Fresnel number may be expressed as $N = a\text{NA}/\lambda\bar{n}$, with the pupil radius $a = R\sin\beta$. It follows that for a given pupil radius the numerical aperture must satisfy $1 \gg \text{NA} \gg \bar{n}\lambda/a$. For visible light and a pupil radius of the order of 1.5 mm the lower limit is of the order 5×10^{-4}, so the NA must be at least 5×10^{-3}, making this no problem for any practical system. Note that the lower limit for NA is $\bar{n}(\lambda/R)^{1/2}$ if the curvature radius of the reference sphere has a fixed value. Although the analytical approximation strictly holds in the paraxial limit only, we will compare the outcome with numerically calculated results for NA values well beyond the paraxial limit.

The calculation of the Strehl ratio requires finding the position of focus, i.e., the point of highest intensity. The focal point is on the optical axis because of symmetry. It turns out that the z component of the electric field is zero [the Jones matrix is invariant under rotations around the optical axis over π, and the z components of the polarization vectors are proportional to $\cos\phi$ and $\sin\phi$, thus giving a zero pupil average; in other words, the contributions to E_z from the plane waves corresponding to pupil coordinates (ρ, ϕ) and $(\rho, \phi + \pi)$ cancel each other]. It follows that only the terms F_{jk} with $j, k = 1, 2$ determine the field in focus. Moreover, these fields are now given by the pupil averages

$$F_{jk} = \frac{1}{\pi}\int_P \mathrm{d}^2\rho (VJ)_{jk}(\rho), \tag{29}$$

or in matrix notation

$$F = \langle VJ \rangle, \tag{30}$$

where the angle brackets indicate the pupil average. The phase term $\exp(iW)$ of scalar diffraction theory (with W the aberration function) is now replaced by the product of a polarization matrix V and the Jones matrix J. The effect of the focal position on the optical axis z is represented by a defocus term that is absorbed in the Jones matrix. We take the refractive index of the isotropic medium in front of the biaxial medium $n_0 = \bar{n}$ so that the Fresnel coefficients are equal to 1 and the interface Jones matrix is equal to the unit matrix. Numerical calculations show that this has hardly any effect on the Strehl ratio. Apparently the effect of the interface on the focal peak intensity is approximately the same with or without birefringence, so that these effects drop out when the ratio of the two intensities (the Strehl ratio) is considered.

The polarization matrix contains the x and y components of the polarization vectors defined in Eqs. (10) and (11) and the apodization factor defined in Eq. (24), and is given by

$$V = B\begin{bmatrix} v_{1x} & v_{2x} \\ v_{1y} & v_{2y} \end{bmatrix} = R(\phi)\begin{bmatrix} \sqrt{\cos\theta} & 0 \\ 0 & 1/\sqrt{\cos\theta} \end{bmatrix} R(-\phi). \tag{31}$$

In the following, an expansion of the matrix F in the NA is considered. The relevant second-order term of V contains only astigmatic terms, which are proportional to $\rho^2\cos 2\phi$ or $\rho^2\sin 2\phi$. Averaging over the pupil coordinates results in a zero contribution to F. Consequently, to second order in NA the effects of V may be ignored and V may be replaced by the unit matrix. A fourth-order term in NA gives a contribution independent of the presence of birefringence and therefore cancels out in the Strehl ratio.

The matrix F is now simply the average of the Jones matrix J, i.e., $F = \langle J \rangle$, and the Strehl ratio may be expressed as

$$I = \sum_{\mu=0,3} I_\mu M_\mu = \tfrac{1}{2} \sum_{\mu=0,3} \mathrm{Tr}(\sigma_\mu F^\dagger F) M_\mu, \tag{32}$$

where Tr indicates the trace of a matrix, i.e., the sum of its diagonal elements, and where Eqs. (27) and (28) are used.

The expression for the intensity may be rewritten using the Jones matrix for normal incidence (the limiting case where NA = 0) J_0. It can be taken in and out of the pupil average as it does not depend on the pupil coordinates. We may therefore define

$$G = F J_0^\dagger = \langle J \rangle J_0^\dagger = \langle J J_0^\dagger \rangle, \tag{33}$$

giving the intensity as [with $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$]

$$I = \tfrac{1}{2} \sum_{\mu=0,3} \mathrm{Tr}(\sigma'_\mu G^\dagger G) M_\mu, \tag{34}$$

with the transformed σ matrices

$$\sigma'_\mu = J_0 \sigma_\mu J_0^\dagger = \sum_{\mu=0,3} S_{\mu\nu} \sigma_\nu. \tag{35}$$

It is straightforward to check that the transformed σ matrices satisfy the same algebra as the original σ matrices, because J_0 is a unitary matrix. In particular the transformed Stokes vector components

$$\bar{M}_\mu = \sum_{\nu=0,3} S_{\nu\mu} M_\nu \tag{36}$$

satisfy that

$$\sum_{j=1,3} \bar{M}_j{}^2 = \bar{M}_0{}^2 = 1. \tag{37}$$

The intensity may now be written as

$$I = \tfrac{1}{2} \sum_{\mu=0,3} \mathrm{Tr}(\sigma_\mu G^\dagger G) \bar{M}_\mu. \tag{38}$$

The Jones matrix is unitary and can therefore be expressed as

$$J J_0^\dagger = \exp(iK), \tag{39}$$

where K is a 2×2 Hermitian matrix, the aberration matrix, that can be decomposed as

$$K = \sum_{\mu=0,3} W_\mu \sigma_\mu, \tag{40}$$

where the functions W_μ are the aberration functions. Clearly, there are now four relevant aberration functions, the scalar aberration function W_0 and three additional vectorial aberration functions W_j ($j = 1, 2, 3$). These aberration functions are proportional to NA^2 for small NA, as there are no piston terms because of the multiplication with $J_0^\dagger$ (in the limit NA $\rightarrow 0$ the matrix $J J_0^\dagger$ is equal to the unit matrix, so $K = 0$).

For small NA the Jones matrix may be expanded as

$$J J_0^\dagger = 1 + iK - \tfrac{1}{2} K^2, \tag{41}$$

leading to the matrix G

$$G = \sigma_0 + i \sum_{\mu=0,3} \langle W_\mu \rangle \sigma_\mu - \tfrac{1}{2} \sum_{\mu,\nu=0,3} \langle W_\mu W_\nu \rangle \sigma_\mu \sigma_\nu. \tag{42}$$

Using the algebraic properties of the σ matrices it is found that

$$\begin{aligned} G^\dagger G = {} & \sigma_0 - \tfrac{1}{2} \sum_{\mu,\nu=0,3} (\langle W_\mu W_\nu \rangle \\ & - \langle W_\mu \rangle \langle W_\nu \rangle)(\sigma_\mu \sigma_\nu + \sigma_\nu \sigma_\mu) \\ = {} & \sigma_0 - \sum_{\mu=0,3} (\langle W_\mu{}^2 \rangle - \langle W_\mu \rangle^2) \sigma_0 \\ & - 2 \sum_{j=1,3} (\langle W_j W_0 \rangle - \langle W_j \rangle \langle W_0 \rangle) \sigma_j. \end{aligned} \tag{43}$$

Inserting this in the expression for the intensity it follows that

$$\begin{aligned} I = {} & 1 - \sum_{\mu=0,3} (\langle W_\mu{}^2 \rangle - \langle W_\mu \rangle^2) \\ & - 2 \sum_{j=1,3} (\langle W_j W_0 \rangle - \langle W_j \rangle \langle W_0 \rangle) \bar{M}_j \\ = {} & 1 - \sum_{j=1,3} [\langle (W_j + \bar{M}_j W_0)^2 \rangle - \langle (W_j + \bar{M}_j W_0) \rangle^2], \end{aligned} \tag{44}$$

where it is used that $\Sigma_{j=1,3} \bar{M}_j{}^2 = \bar{M}_0{}^2 = 1$. Equation (44) is the sought generalization of the scalar diffraction expression of Eq. (1). Clearly, aberrations reduce the Strehl ratio, just as in the scalar case. In contrast, however, the polarization now influences the Strehl ratio. In fact, part of the scalar aberration function W_0 can be compensated, provided that the additional aberration functions W_j and the polarization in the entrance pupil are such that the terms in angle brackets are zero. Such a polarization effect can exist only if W_0 and the W_j have at least one nonzero type of Zernike aberration in common (e.g., both must have nonzero astigmatism, coma, or spherical aberration). Straightforward variation through use of a Lagrange multiplier for the constraint on the Stokes vector components leads to the values of the transformed Stokes vector for maximum and minimum Strehl ratio as

$$\bar{M}_j = \pm \frac{\langle W_j W_0 \rangle - \langle W_j \rangle \langle W_0 \rangle}{[\Sigma_{j=1,3} (\langle W_j W_0 \rangle - \langle W_j \rangle \langle W_0 \rangle)^2]^{1/2}} \tag{45}$$

for the maximum and minimum being

$$\begin{aligned} I_\pm = {} & 1 - \sum_{\mu=0,3} (\langle W_\mu^2 \rangle - \langle W_\mu \rangle^2) \\ & \pm 2 \left[\sum_{j=1,3} (\langle W_j W_0 \rangle - \langle W_j \rangle \langle W_0 \rangle)^2 \right]^{1/2}. \end{aligned} \tag{46}$$

We now turn to the actual derivation of the analytical expression for the Strehl ratio as a function of birefringence. A useful representation of Jones matrices is

$$J = \exp[i(\tau\sigma_0 + \mu\mathbf{v}\cdot\boldsymbol{\sigma})] = \exp(i\tau)(\cos\mu\sigma_0 + i\sin\mu\mathbf{v}\cdot\boldsymbol{\sigma}), \quad (47)$$

where $\mathbf{v}$ is a three-dimensional vector of unit length. The Jones matrix and the normal-incidence Jones matrix can be represented in this way. The Jones matrix is

$$J = \exp(i\bar{W})R(\chi)\begin{bmatrix}\exp(i\Delta W/2) & 0 \\ 0 & \exp(-i\Delta W/2)\end{bmatrix}R(-\chi) = \exp(i\bar{W})[\cos(\Delta W/2)\sigma_0 + i\sin(\Delta W/2)\mathbf{v}\cdot\boldsymbol{\sigma}], \quad (48)$$

with $\mathbf{v} = [\cos(2\chi), \sin(2\chi), 0]$, $\bar{W} = kda/2\cos\theta$, $\Delta W = kd(b^2 + c^2)^{1/2}/\cos\theta$. The normal-incidence Jones matrix is

$$J_0 = R(\gamma)\begin{bmatrix}\exp(ia_\parallel/2) & 0 \\ 0 & \exp(-ia_\parallel/2)\end{bmatrix}R(-\gamma) = \cos(a_\parallel/2)\sigma_0 + i\sin(a_\parallel/2)\mathbf{v}_0\cdot\boldsymbol{\sigma}, \quad (49)$$

with $\mathbf{v}_0 = [\cos(2\gamma), \sin(2\gamma), 0]$, and where

$$a_\parallel = \frac{2\pi d\Delta n_\parallel}{\lambda}, \quad (50)$$

$$a_\perp = \frac{2\pi d\Delta n_\perp}{\lambda}. \quad (51)$$

The transformation matrix S then follows as

$$S = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 - 2\sin^2(a_\parallel/2)\sin^2(2\gamma) & \sin^2(a_\parallel/2)\sin(4\gamma) & \sin a_\parallel \sin(2\gamma) \\ 0 & \sin^2(a_\parallel/2)\sin(4\gamma) & 1 - 2\sin^2(a_\parallel/2)\cos^2(2\gamma) & -\sin a_\parallel\cos(2\gamma) \\ 0 & -\sin a_\parallel \sin(2\gamma) & \sin a_\parallel\cos(2\gamma) & \cos a_\parallel \end{bmatrix}. \quad (52)$$

The product matrix $JJ_0^\dagger$ is given by

$$JJ_0^\dagger = \exp(i\bar{W})\{[\cos(\Delta W/2)\cos(a_\parallel/2) + \sin(\Delta W/2)\sin(a_\parallel/2)\mathbf{v}\cdot\mathbf{v}_0]\sigma_0 + i[\sin(\Delta W/2)\cos(a_\parallel/2)\mathbf{v} - \cos(\Delta W/2)\sin(a_\parallel/2)\mathbf{v}_0 + \sin(\Delta W/2)\sin(a_\parallel/2)\mathbf{v}\times\mathbf{v}_0]\cdot\boldsymbol{\sigma}\}. \quad (53)$$

The aberration matrix K can be determined from this expression in all cases. To find the Strehl ratio, only the paraxial limit of small NA needs to be analyzed. To that end we substitute $\sin\theta = \rho\mathrm{NA}/\bar{n}$ into the expressions for $\bar{W}$, ΔW, and χ and expand $JJ_0^\dagger$ up to second order in $\rho\mathrm{NA}/\bar{n}$, thus arriving at the aberration matrix K up to second order in $\rho\mathrm{NA}/\bar{n}$. This procedure is followed by use of the computer algebra package MATHEMATICA, and the resulting four aberration functions are found as

$$W_0 = \tfrac{1}{2}[a_\perp - \tfrac{1}{2}a_\parallel\cos(2\gamma - 2\phi)]\left(\frac{\rho\mathrm{NA}}{\bar{n}}\right)^2, \quad (54)$$

$$W_1 = \frac{1}{2}a_\perp[\cos(2\gamma)\cos(2\gamma - 2\phi) + \cos(a_\parallel/2)\mathrm{sinc}(a_\parallel/2)\sin(2\gamma)\sin(2\gamma - 2\phi)] \times \left(\frac{\rho\mathrm{NA}}{\bar{n}}\right)^2, \quad (55)$$

$$W_2 = \frac{1}{2}a_\perp[\sin(2\gamma)\cos(2\gamma - 2\phi) + \cos(a_\parallel/2)\mathrm{sinc}(a_\parallel/2)\cos(2\gamma)\sin(2\gamma - 2\phi)] \times \left(\frac{\rho\mathrm{NA}}{\bar{n}}\right)^2, \quad (56)$$

$$W_3 = \frac{1}{2}a_\perp\sin(a_\parallel/2)\mathrm{sinc}(a_\parallel/2)\sin(2\gamma - 2\phi)\left(\frac{\rho\mathrm{NA}}{\bar{n}}\right)^2, \quad (57)$$

with $\mathrm{sinc}\,x \equiv (\sin x)/x$. Clearly, W_0 contains defocus due to axial birefringence and astigmatism due to lateral birefringence, whereas the W_j contain only astigmatism, which is present only for nonzero axial birefringence. By shifting the image plane the defocus contribution to W_0 can be eliminated. This term will therefore be left out in the following. The relevant pupil averages are

$$\langle(W_1 + \bar{M}_1W_0)^2\rangle = \frac{1}{96}(2a_\perp\cos(2\gamma) - \bar{M}_1a_\parallel)^2\left(\frac{\mathrm{NA}}{\bar{n}}\right)^4 + \frac{1}{24}{a_\perp}^2\cos^2(a_\parallel/2)\mathrm{sinc}^2(a_\parallel/2) \times \sin^2(2\gamma)\left(\frac{\mathrm{NA}}{\bar{n}}\right)^4, \quad (58)$$

$$\langle(W_2 + \bar{M}_1W_0)^2\rangle = \frac{1}{96}(2a_\perp\sin(2\gamma) - \bar{M}_2a_\parallel)^2\left(\frac{\mathrm{NA}}{\bar{n}}\right)^4 + \frac{1}{24}a_\perp^2\cos^2(a_\parallel/2)\mathrm{sinc}^2(a_\parallel/2) \times \cos^2(2\gamma)\left(\frac{\mathrm{NA}}{\bar{n}}\right)^4, \quad (59)$$

$$\langle (W_3 + \bar{M}_3 W_0)^2 \rangle = \frac{1}{96} \bar{M}_3{}^2 a_\parallel{}^2 \left(\frac{\text{NA}}{\bar{n}}\right)^4 + \frac{1}{24} \times a_\perp{}^2 \sin^2(a_\parallel/2) \text{sinc}^2(a_\parallel/2) \left(\frac{\text{NA}}{\bar{n}}\right)^4 . \quad (60)$$

With the following relations for the Stokes vector components,

$$\bar{M}_1{}^2 + \bar{M}_2{}^2 + \bar{M}_3{}^2 = 1, \quad (61)$$

$$\bar{M}_1 \cos(2\gamma) + \bar{M}_2 \sin(2\gamma) = M_1 \cos(2\gamma) + M_2 \sin(2\gamma) = \cos(2\epsilon) \cos(2\xi - 2\gamma), \quad (62)$$

it follows that the Strehl ratio may be expressed as

$$I = 1 - \frac{1}{96} \{4[1 + \text{sinc}^2(a_\parallel/2)] a_\perp{}^2 + a_\parallel{}^2 - 4 a_\perp a_\parallel \cos(2\epsilon) \cos(2\xi - 2\gamma)\} \left(\frac{\text{NA}}{\bar{n}}\right)^4 . \quad (63)$$

This equation is the main result of this paper. It shows that the Strehl ratio depends on the retardation parameters $a_\perp$ and $a_\parallel$, the polarization state, and the effective NA in the medium NA/$\bar{n}$. The dependence on these parameters and a comparison with numerical results is discussed in the next section.

4. DISCUSSION AND CONCLUSION

As to the dependence on polarization, the maximum Strehl ratio is obtained for the linear polarization parallel or perpendicular to one of the principal in-plane axes of the birefringent layer (depending on the sign of $\Delta n_\perp \Delta n_\parallel$). For the orthogonal linear polarization the cross term changes sign, so that this polarization state gives the minimum Strehl ratio. For a circular polarization the cross term is absent, giving a value that is the average of the minimum and maximum Strehl ratio.

For the special case of uniaxial birefringence the analytical expression Eq. (63) can be simplified. When the principal axis is parallel to the optical axis (axial birefringence) the refractive indices satisfy $n_1 = n_2$, leading to $\Delta n_\parallel = 0$. In this special case the Strehl ratio is given by

$$I_{\text{ax}} = 1 - \frac{1}{12} a_\perp{}^2 \left(\frac{\text{NA}}{\bar{n}}\right)^4 , \quad (64)$$

in agreement with previous results.[5,6] When the birefringence is uniaxial with the principal axis perpendicular to the optical axis (lateral birefringence), the refractive indices satisfy $n_3 = n_1$ or $n_3 = n_2$. In terms of the birefringence parameters this corresponds to $2\Delta n_\perp = \Delta n_\parallel$ or $2\Delta n_\perp = -\Delta n_\parallel$, respectively. The Strehl ratio then follows as

$$I_{\text{lat}} = 1 - \frac{1}{96} [2 \mp 2 \cos(2\epsilon) \cos(2\xi - 2\gamma) + \text{sinc}^2(a_\parallel/2)] a_\parallel{}^2 \left(\frac{\text{NA}}{\bar{n}}\right)^4 , \quad (65)$$

where the lower sign corresponds to $n_3 = n_1$ and the upper sign to $n_3 = n_2$. The Strehl ratio is maximum for the linear polarization perpendicular to the principal axis ($\xi = \gamma$ and $\xi = \gamma + \pi/2$). This means that the ordinary polarization mode is predominantly excited. The Strehl ratio is then

$$I_{\text{lat,max}} = 1 - \frac{1}{24} \sin^2(a_\parallel/2) \left(\frac{\text{NA}}{\bar{n}}\right)^4 , \quad (66)$$

and is equal to one if $d\Delta n_\parallel = k\lambda$ with $k = 1, 2, 3, \ldots$. For other birefringence values the deviation from one will be very small as well, because the prefactor $(\text{NA}/\bar{n})^4/24$ will generally be less than 1%, even for high NA values. The light distribution close to focus for the (predominantly) ordinary mode will be distorted only slightly, namely, by the effects described by terms of higher order in the NA, e.g., spherical aberration and higher-order astigmatism. It may be concluded that axial birefringence is more damaging than lateral birefringence to spot quality. Even for a circular polarization, the lateral birefringence giving the same Strehl ratio as the Strehl ratio for a given value of the axial birefringence is twice as high as that axial birefringence value.

The analytical expression for the Strehl ratio obtained in Section 3 is strictly valid if the NA is sufficiently small compared with one and if the birefringence is sufficiently small that the resulting Strehl ratio is not much smaller than one. The effect of NA on the accuracy of the analytical approximation can be investigated by comparing its results with numerically calculated Strehl ratios. Figure 1 shows the numerically calculated Strehl ratio as a function of NA for the linear polarization with minimum Strehl ratio and for $\Delta n_\perp = 1.5 \times 10^{-3}$, $\Delta n_\parallel = 4.0 \times 10^{-3}$, $\bar{n} = 1.58$, $\lambda = 405\,\text{nm}$, and thickness d that scales such that $d\text{NA}^2 = 72.25\,\mu\text{m}$. The evaluation of the pupil average is done by straightforward trapezoidal

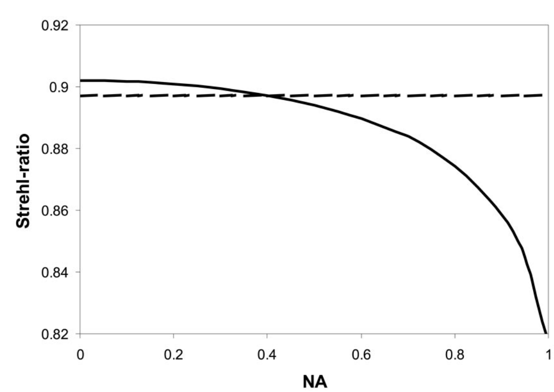

Fig. 1. Numerically calculated Strehl ratio (solid curve) and the analytical approximation of Eq. (63) (dashed line) as a function of NA for the birefringence parameters $\Delta n_\perp = 1.5 \times 10^{-3}$ and $\Delta n_\parallel = 4.0 \times 10^{-3}$; for the linear polarization with minimum Strehl ratio; and for $\bar{n} = 1.58$, $\lambda = 405\,\text{nm}$, and a thickness d that scales such that $d\text{NA}^2 = 72.25\,\mu\text{m}$.

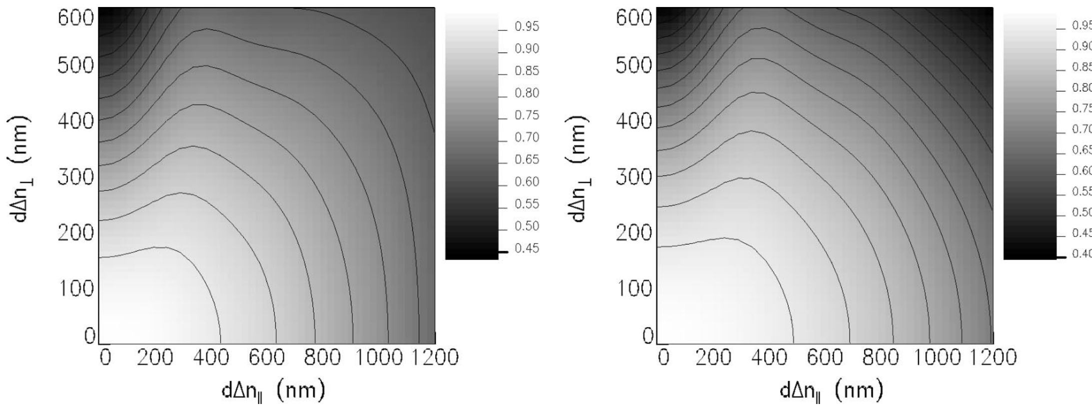

Fig. 2. Numerically calculated Strehl ratio (left) and the analytical approximation of Eq. (63) (right) as a function of the retardation parameters $d\Delta n_{\perp}$ and $d\Delta n_{\parallel}$ for a circular polarization and for $\bar{n} = 1.58$, NA = 0.85, and $\lambda = 405$ nm.

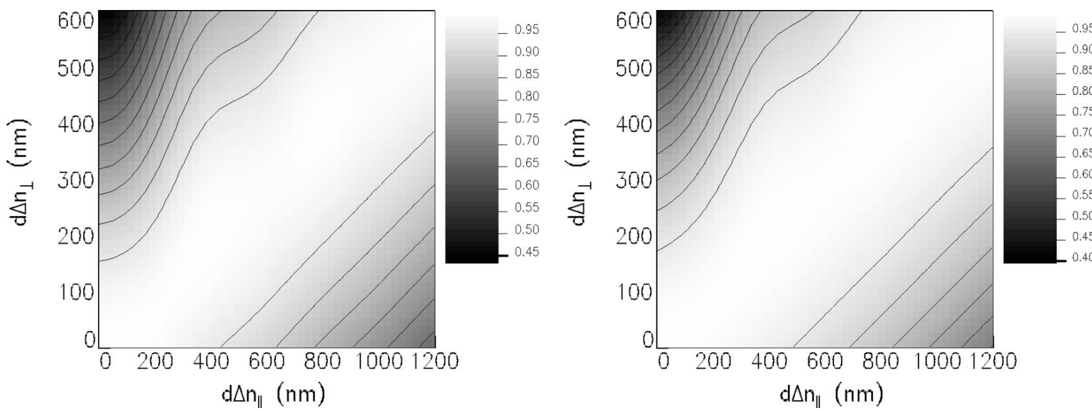

Fig. 3. Numerically calculated Strehl-ratio (left) and the analytical approximation of Eq. (63) (right) as a function of the retardation parameters $d\Delta n_{\perp}$ and $d\Delta n_{\parallel}$ for the linear polarization giving rise to a maximum Strehl ratio and for $\bar{n} = 1.58$, NA = 0.85, and $\lambda = 405$ nm.

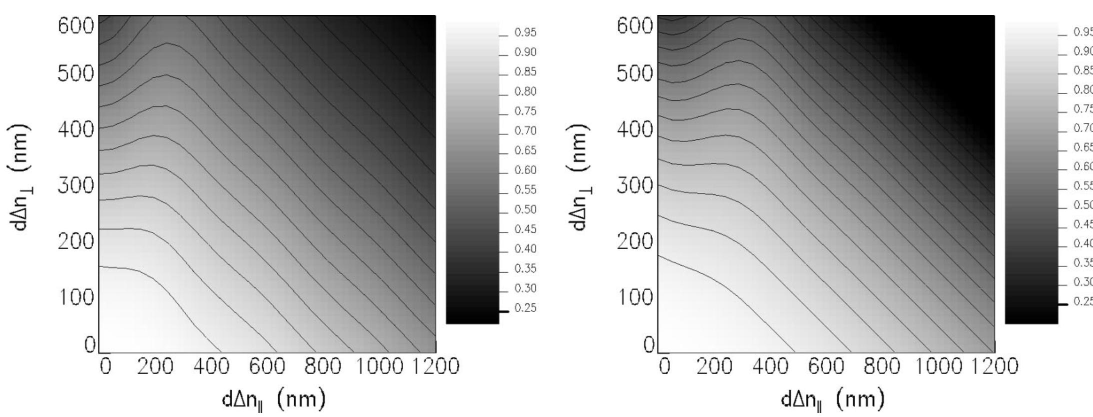

Fig. 4. Numerically calculated Strehl ratio (left) and the analytical approximation of Eq. (63) (right) as a function of the retardation parameters $d\Delta n_{\perp}$ and $d\Delta n_{\parallel}$ for the linear polarization giving rise to a minimum Strehl ratio and for $\bar{n} = 1.58$, NA = 0.85, and $\lambda = 405$ nm. Values of the analytical Strehl ratio below 0.2 have been set equal to 0.2 (upper right corner of right-hand graph) in order to have roughly the same gray scales in both plots. This is already in the regime where the analytical approximation cannot be expected to give reasonable results.

integration[7] over the radial and azimuthal coordinates, resulting in an estimated accuracy of ~0.5%. According to Eq. (63) the Strehl ratio is a constant 0.8971 independent of NA (because of the scaling of the thickness). The numerical Strehl ratio is 0.9019 in the limit NA $\rightarrow$ 0, which is ~0.5% too high and is probably due to numerical errors. For increasing NA, deviations from the constant low-NA value occur. These deviations are roughly parabolic in NA, implying that they are related to aberrations that roughly scale as NA^4. It follows that these deviations are due to higher-order astigmatism and to the difference in spherical aberration between the two eigenmodes of the biaxial medium, which are not taken into account in the analytical approximation.

Figure 2 shows contour plots of the Strehl ratio as a function of the birefringence parameters according to a numerical calculation and according to the analytical approximation for a circular polarization and for $\lambda = 0.405\,\mu\text{m}$, NA = 0.85, and $\bar{n} = 1.58$ (effective NA in the medium $\text{NA}/\bar{n} = 0.54$). These numbers pertain to the readout of so-called Blu-ray disks.[8–10] This new type of optical disk has a significantly higher storage capacity (~25 GByte) than existing compact disks (CDs, 650 MByte) and digital versatile disks (DVDs, 4.7 GByte) owing to the use of a shorter wavelength (785 nm for CD and 655 nm for DVD) and a larger NA (0.45–0.50 for CD and 0.60–0.65 for DVD). Figures 3 and 4 show the Strehl ratio for the linear polarizations giving rise to the maximum and minimum Strehl ratio, respectively. The numerical results match quite well with those of the analytical expression, although the NA is rather high. For small retardation parameters the analytical expression overestimates the Strehl ratio because of the effects of the high NA, i.e., the difference in spherical aberration between the two modes and higher-order astigmatism mentioned in the previous paragraph. For large retardation parameters the analytical expression underestimates the Strehl ratio because the analytical expression eventually drops below zero, whereas the actual Strehl ratio is always positive. For the case of lateral birefringence and the ordinary polarization mode the Strehl ratio is close to one, leading to a band of values in Fig. 3 around the line $d\Delta n_\perp = d\Delta n_\parallel/2$ for which the focal spot is diffraction limited. For the orthogonal, worst-case polarization (Fig. 4) the focal spot is no longer diffraction-limited for relatively small retardation values. The contours for the circular polarization shown in Fig. 2 are between the two extreme cases.

In conclusion, we have derived a simple analytical equation for the Strehl ratio for focusing into biaxially birefringent media. The expression is quantitatively correct for small birefringence parameters, small retardation ("small" being defined as giving rise to small deviations in the Strehl ratio from one) and for small NA. However, even for large NA the correspondence with numerical results, which are exact in the sense that all effects of high NA are taken into account, is quite reasonable. The analytical approximation can be of use in estimating tolerances for birefringence effects in optical systems, such as in scanning microscopy and in optical disk readout.

APPENDIX A

The unit matrix σ_0 and the three Pauli matrices $\sigma_j\,(j = 1, 2, 3)$ are defined by

$$\sigma_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

$$\sigma_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \sigma_3 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}. \tag{A1}$$

These matrices satisfy

$$\sigma_\mu{}^2 = \sigma_0 , \tag{A2}$$

$$\sigma_0\sigma_j = \sigma_j\sigma_0 = \sigma_j , \tag{A3}$$

$$\sigma_j\sigma_k = -\sigma_k\sigma_j = i\,\epsilon_{jkl}\sigma_l , \tag{A4}$$

$$\sum_{\mu=0,3} (\sigma_\mu)_{jl}(\sigma_\mu)_{l'j'} = 2\,\delta_{jj'}\delta_{ll'} , \tag{A5}$$

with ϵ_{jkl} the (fully antisymmetric) Levi–Civita tensor. The Pauli matrices are often denoted as σ_x, σ_y, and σ_z, where conventionally $\sigma_x = \sigma_2$, $\sigma_y = \sigma_3$ and $\sigma_z = \sigma_1$.

ACKNOWLEDGMENTS

Jean Schleipen and Paul Urbach are thanked for a critical reading of the manuscript.

Sjoerd Stallinga can be reached by e-mail at sjoerd.stallinga@philips.com.

REFERENCES

1. S. Stallinga, "Light distribution close to focus in biaxially birefringent media," J. Opt. Soc. Am. A **21**, 1785–1798 (2004).
2. M. Born and E. Wolf, *Principles of Optics*, 6th ed. (Cambridge U. Press, Cambridge, UK, 1980).
3. J. J. Stamnes, *Waves in Focal Regions* (Hilger, Bristol, UK, 1986).
4. *Selected Papers on Electromagnetic Fields in the Focal Region*, J. J. Stamnes, ed., SPIE Milestone Series Vol. 168 (SPIE Optical Engineering Press, Bellingham, Wash., 2001).
5. S. Stallinga, "Axial birefringence in high NA optical systems and the light distribution close to focus," J. Opt. Soc. Am. A **18**, 2846–2858 (2001).
6. A. B. Marchant, "Cover sheet aberrations in optical recording," in *Optical Disk Systems and Applications*, E. V. LaBudde, ed., Proc. SPIE **421**, 43–49 (1983).
7. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 77: The Art of Scientific Computing*, 2nd ed. (Cambridge U. Press, Cambridge, UK, 1992).
8. M. K. Dekker, N. Pfeffer, M. Kuijper, W. M. Coene, E. R. Meinders, and H. J. Borg, "Blue phase-change recording at high data densities and data rates," in *Optical Data Storage 2000*, D. G. Stinson and R. Katayama, eds., Proc. SPIE **4090**, 28–35 (2000).
9. I. Ichimura, S. Masuhara, J. Nakano, Y. Kasami, K. Yasuda, O. Kawakubo, and K. Osato, "On-groove phase-change optical recording for a capacity of 25 GB," in *Optical Data Storage 2001*, T. Hurst and S. Kobayashi, eds., Proc. SPIE **4342**, 168–177 (2001).
10. M. Kuijper, I. P. Ubbens, L. Spruijt, J. M. ter Meulen, and K. Schep, "Groove-only recording under DVR conditions," in *Optical Data Storage 2001*, T. Hurst and S. Kobayashi, eds., Proc. SPIE **4342**, 178–185 (2001).