Compact description of substrate-related aberrations in high numerical-aperture optical disk readout

Sjoerd Stallinga

Optical disks are read out by focusing a beam of high numerical aperture (NA) through the substrate. Deviations of the thickness from the nominal value result in spherical aberration; tilting the substrate results in coma. Exact analytical expressions for the rms aberration per micrometer thickness mismatch (for spherical aberration) and per degree tilt (for coma) are derived. The paraxial estimates for these sensitivities proportional to NA^4 (spherical aberration) and NA^3 (coma) underestimate the exact values by a factor of ~ 2 for the value NA = 0.85, corresponding to the new Blu-ray disk format. Expansion of the aberration function in Zernike aberrations shows that the exact aberration functions are well described by the lowest-order Zernike spherical aberration (A_{40}) and coma (A_{31}) term for all but the very highest NA values.

OCIS codes: 080.1010, 210.4590.

1. Introduction

The long-term trend in data storage on optical disks is toward disks with an ever higher capacity. This may be achieved by a decrease in wavelength λ and an increase in the numerical aperture (NA) of the scanning objective lens, as the size of the focal spot scales as λ/NA. The compact disk (CD) operates at a wavelength $\lambda = 0.785$ μm and a numerical aperture NA = 0.45–0.50, the digital versatile disk (DVD) operates at a wavelength $\lambda = 0.660$ μm and a numerical aperture NA = 0.50–0.65, and the recently proposed Blu-ray disk[1–3] (BD) and advanced optical disk[4] (AOD) make use of GaN lasers with wavelength $\lambda = 0.405$ μm, with BD using a numerical aperture NA = 0.85, AOD a numerical aperture NA = 0.65.

The beam is focused onto the data layer of the disk through a substrate layer of thickness d. The scanning objective lens is designed in such a way that the spherical aberration resulting from focusing through this layer is compensated for, so that the scanning spot at the data layer is nominally free from aberrations. As the NA increases, the sensitivity to substrate-related aberrations increases as well. For example, tilting the disk results in coma, and a mismatch of the substrate thickness from the nominal value results in spherical aberration, and both aberrations are known to increase strongly with NA. In order to maintain sufficient tolerance for coma from disk tilt, the increase in NA is accompanied by a decrease in substrate thickness. While CDs have $d = 1.2$ mm, DVDs and AODs have $d = 0.6$ mm, and BDs even have $d = 0.1$ mm.

It is known that coma and spherical aberration scale as NA^3 and NA^4, respectively, in the limit of small NA.[5,6] The current practice of optical disk readout is increasingly further away from this limit, making the accuracy of these scaling laws questionable. In order to model these aberrations adequately, one must resort to numerical ray-tracing methods. It is the object of this paper to provide a compact analytical description of the substrate-related aberrations in optical disk readout that is exact for arbitrary values of NA. The main thesis of this paper is that the paraxial scaling laws greatly underestimate the true aberrations. The differences amount to a factor of ~ 2 for the NA = 0.85 of the BD format. In addition, it will be shown that the expansion of the aberration functions in Zernike aberration terms is more convenient than the series expansion in NA underlying the Seidel aberration terms. The lowest-order Zernike

Sjoerd Stallinga (sjoerd.stallinga@philips.com) is with Philips Research Laboratories, Professor Holstlaan 4, Eindhoven 5656AA, The Netherlands.

Received 9 June 2004; revised manuscript received 14 September 2004; accepted 18 September 2004.

0003-6935/05/060849-10$15.00/0

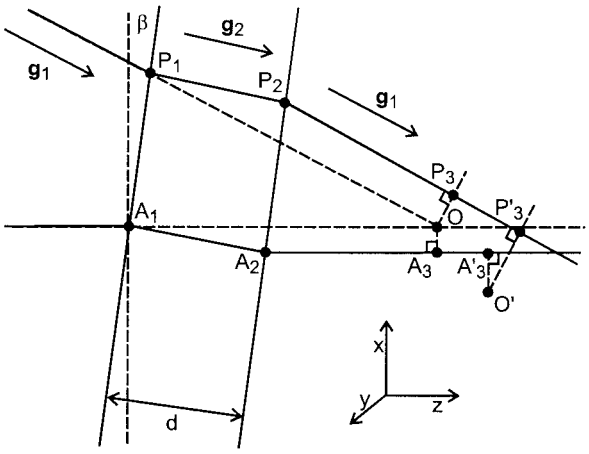

Fig. 1. General ray path along $P_1P_2P_3$ and chief ray path along $A_1A_2A_3$ through a plate of thickness d tilted over an angle β. The reference point is the focus of the incident spherical wave O and may be changed to O' in order to minimize the rms aberration (in that case, O' is the diffraction focus).

terms alone are already sufficient to describe the exact rms aberration within a few percent or less, even up to NA values around NA = 0.90.

Previous work dealt either with the small NA limit[5,6] or with an expansion in Seidel aberrations based on the Gaussian focus as the reference point.[7] Differences with the present paper concern (i) the use of the diffraction focus instead of the Gaussian focus as the reference point for the aberration function, (ii) the analytical calculation of exact rms values of the aberration function (this value is the measure of choice for describing the quality of the scanning spot), (iii) the analytical calculation of the Zernike coefficients, and (iv) an arbitrary refractive index of the incidence medium.

The content of this paper is as follows. In Section 2, a review of a general method to calculate the aberration function is given, in Section 3, this is applied to spherical aberration from thickness mismatch, and, in Section 4, to coma from disk tilt. In Section 5, extensions of the methods used to near-field systems and birefringent media are discussed.

2. Aberration Function

Consider a spherical wave in a medium of refractive index n_1 converging on O that passes a plan-parallel plate of refractive index n_2 and of thickness d tilted over an angle β. The wave-front aberration is defined as the difference in optical path length for a general ray along $OP_1P_2P_3$ and for the chief ray $OA_1A_2A_3$.[7]

$$W = [OP_1P_2P_3] - [OA_1A_2A_3], \tag{1}$$

where P_1 and P_2 are the refraction points of a general ray, A_1 and A_2 are the refraction points of the chief ray, O is the reference point, and the lines OP_3 and OP_3 are perpendicular to P_2P_3 and A_2A_3, which defines P_3 and A_3, respectively (see Fig. 1). In the following, the virtual focus point O is taken to be the center of the coordinate frame; the z axis is taken to be the optical axis. The position vectors of the points P_j are $\mathbf{r}_j$ ($j = 1, 2, 3$).

Any point O' may be taken as the reference point. Then the points P_3 and A_3 must be replaced by points P_3' and A_3', respectively. However, the optical path lengths must still be calculated starting from point O. Traditionally, the aberration function is defined such that the reference point O' is the Gaussian focus, i.e., the point for which the wave-front tilt in x and y and the defocus are zero. The three degrees of freedom in the form of the three coordinates of O' can always be chosen so that the conditions on zero wave-front tilt and defocus are satisfied. A more appropriate alternative for the high NA systems that are the subject of this paper is to take O' as the diffraction focus, i.e., the point for which the rms value of the aberration function is minimum. Even in the limit of small NA, there is a small difference between the Gaussian and diffraction focus that needs to be taken into account.

The optical path lengths can be expressed as the inner product between ray vectors and position vectors. The ray vector in the surrounding medium (for OP_1 and P_2P_3) is $\mathbf{g}_1$, and the ray vector in the plate is $\mathbf{g}_2$. These ray vectors are perpendicular to the wave fronts and satisfy the eikonal equations[8] $\mathbf{g}_j^2 = n_j^2$ for $j = 1, 2$. It follows that

$$\begin{aligned}[OP_1P_2P_3] &= \mathbf{g}_1 \cdot \mathbf{r}_1 + \mathbf{g}_2 \cdot (\mathbf{r}_2 - \mathbf{r}_1) + \mathbf{g}_1 \cdot (\mathbf{r}_3 - \mathbf{r}_2) \\ &= (\mathbf{g}_2 - \mathbf{g}_1) \cdot (\mathbf{r}_2 - \mathbf{r}_1) + \mathbf{g}_1 \cdot \mathbf{r}_3 \\ &= (g_{2\perp} - g_{1\perp})d, \end{aligned} \tag{2}$$

where $g_{0\perp}$ and $g_{1\perp}$ are the components of the ray vectors perpendicular to the interface, and where it is used that P_2P_3 is perpendicular to OP_3, implying that $\mathbf{g}_1 \cdot \mathbf{r}_3 = 0$, and that the components of $\mathbf{g}_1$ and $\mathbf{g}_2$ parallel to the interface are equal to each other (Snell's law). A similar expression can be found for the optical path length $[OA_1A_2A_3]$ of the chief ray. The wave-front aberration with respect to O then follows as

$$W_O = (g_{2\perp} - g_{1\perp})d - (g_{2\perp}' - g_{1\perp}')d, \tag{3}$$

where the prime indicates the ray vectors for the chief ray path.

When a different reference point O' is chosen, separated from O by the position vector $\Delta\mathbf{r}$, there is an additional contribution to path length, namely,

$$\begin{aligned}[P_3P_3'] &= \mathbf{g}_1 \cdot (\mathbf{r}_3' - \mathbf{r}_3) = \mathbf{g}_1 \cdot \mathbf{r}_3' \\ &= \mathbf{g}_1 \cdot [\Delta\mathbf{r} + (\mathbf{r}_3' - \Delta\mathbf{r})] = \mathbf{g}_1 \cdot \Delta\mathbf{r}. \end{aligned} \tag{4}$$

where it is used that P_3P_3' is perpendicular to $O'P_3'$, i.e. $\mathbf{g}_1 \cdot (\mathbf{r}_3' - \Delta\mathbf{r}) = 0$. This leads to an aberration function with respect to O' given by

$$W_O' = (g_{2\perp} - g_{1\perp})d + \mathbf{g}_1 \cdot \Delta\mathbf{r} - (g_{2\perp}' - g_{1\perp}')d - \mathbf{g}_1' \cdot \Delta\mathbf{r}. \tag{5}$$

This is the general expression for the aberration function of a plane-parallel plate. When the reference point O' is required to be the Gaussian focus, the coordinates of the position shift $\Delta\mathbf{r}$ can be derived from the requirement that the aberration function O' be stationary to ray variations close to the chief ray; when the reference point O' is required to be the diffraction focus, the coordinates of the position shift $\Delta\mathbf{r}$ can be derived from the requirement that the rms value of the aberration function $W_{O'}$ is minimum.

A coordinate system is chosen such that the optical axis of the converging beam of light coincides with the z axis (see Fig. 1). The normal to the plate makes an angle β with the z axis. The incident rays are parameterized by the radial and azimuthal pupil coordinates ρ and θ, respectively. This gives rise to a parameterization of the ray vectors:

$$g_{jx} = -\mathrm{NA}\rho\cos\theta, \tag{6}$$

$$g_{jy} = -\mathrm{NA}\rho\sin\theta, \tag{7}$$

$$g_{jz} = \sqrt{n_j^2 - \mathrm{NA}^2\rho^2}, \tag{8}$$

for $j = 1, 2$, and with NA the numerical aperture. The chief ray is defined by the center of the pupil $\rho = 0$ so that $g_{jx}' = g_{jy}' = 0$ and $g_{jz}' = n_j$, for $j = 1, 2$. The terms related to the shift in reference point are thus

$$\mathbf{g}_1 \cdot \Delta\mathbf{r} - \mathbf{g}_1' \cdot \Delta\mathbf{r} = -\Delta x\mathrm{NA}\rho\cos\theta - \Delta y\mathrm{NA}\rho\sin\theta + \Delta z\left(\sqrt{n_1^2 - \mathrm{NA}^2\rho^2} - n_1\right). \tag{9}$$

It is often convenient to expand the aberration function using Zernike polynomials.[8]

$$W(\rho, \theta) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\left[A_{n,m}Z_n^m(\rho)\cos(m\theta) + A_{n,-m}Z_n^m(\rho)\sin(m\theta)\right], \tag{10}$$

where the integers n and m take values such that $n = m + 2k$ with $k = 0, 1, 2$ and where the Zernike polynomials are defined by

$$Z_{m+2k}^m(\rho) = \rho^m P_k^{0m}(2\rho^2 - 1), \tag{11}$$

where $P_k^{0m}(x)$ is a Jacobi polynomial. The Zernike functions are an orthogonal and complete set of functions on the unit circle (the pupil), so that the associated Zernike coefficients follow from the integrals

$$A_{n,m} = \frac{n+1}{\pi}\int_P \mathrm{d}^2\rho W(\rho, \theta)Z_n^m(\rho)\cos(m\theta), \tag{12}$$

$$A_{n,-m} = \frac{n+1}{\pi}\int_P \mathrm{d}^2\rho W(\rho, \theta)Z_n^m(\rho)\sin(m\theta), \tag{13}$$

where the integration domain P indicates the unit circle, and where the first equation holds for $m = 0$. The square of the rms aberration value is the weighted sum of the square of all Zernike coefficients:

$$W_{\mathrm{rms}}^2 = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{A_{n,m}^2 + A_{n,-m}^2}{2(n+1)}, \tag{14}$$

where the integers n and m take values such that $n = m + 2k$ with $k = 0, 1, 2, \ldots$. This is the property that makes the expansion in Zernike polynomials so useful. It is remarked that a slightly different convention is used compared with Ref. 8; namely, the $m = 0$ terms have no prefactor here.

3. Thickness Error and Spherical Aberration

We consider the spherical aberration resulting from focusing through a nontilted substrate ($\beta = 0$) with a thickness error Δd. The aberration function with respect to the reference point O is given by

$$W_O = \Delta d\sqrt{n_2^2 - \mathrm{NA}^2\rho^2} - \Delta d\sqrt{n_1^2 - \mathrm{NA}^2\rho^2} - \Delta d(n_2 - n_1). \tag{15}$$

In order to find the aberration function with respect to the diffraction focus, an axial shift of the reference point Δz must be applied so that the rms value of the resulting aberration function is minimum. This gives the aberration function for the reference point O' as

$$W_{O'} = \Delta d[\sqrt{n_2^2 - \mathrm{NA}^2\rho^2} + \alpha\sqrt{n_1^2 - \mathrm{NA}^2\rho^2} - (n_2 + \alpha n_1)], \tag{16}$$

where

$$\alpha = \frac{\Delta z}{\Delta d} - 1. \tag{17}$$

The rms aberration is

$$W_{\mathrm{rms}} = \sqrt{\langle W_{O'}^2\rangle - \langle W_{O'}\rangle^2}, \tag{18}$$

where the pupil average is defined as

$$\langle A\rangle = \frac{1}{\pi}\int_P \mathrm{d}^2\rho A(\rho), \tag{19}$$

and the integration is over the unit circle. Using the functions

$$f_j = \sqrt{n_j^2 - \rho^2\mathrm{NA}^2} \tag{20}$$

for $j = 1, 2$, we may write that

$$W_{O'} = \Delta d(f_2 + \alpha f_1), \tag{21}$$

where the piston term related to the chief ray contribution to the aberration function $\Delta d(n_2 + \alpha n_1)$ is left out, as this term drops out of the expression for the rms aberration. The relevant pupil averages then follow as

$$\langle W_{O'} \rangle = \Delta d(\langle f_2 \rangle + \alpha \langle f_1 \rangle), \tag{22}$$

$$\langle {W_{O'}}^2 \rangle = (\Delta d)^2 (\langle {f_2}^2 \rangle + 2\alpha \langle f_1 f_2 \rangle + \alpha^2 \langle {f_1}^2 \rangle). \tag{23}$$

It is found that (all integrals are evaluated using Mathematica[9])

$$\langle f_j \rangle = \frac{2}{3\mathrm{NA}^2} \left[n_j^3 - (n_j^2 - \mathrm{NA}^2)^{3/2} \right], \tag{24}$$

$$\langle f_j^2 \rangle = n_j^2 - \frac{1}{2} \mathrm{NA}^2, \tag{25}$$

$$\begin{aligned} \langle f_1 f_2 \rangle = \frac{1}{4\mathrm{NA}^2} \Bigg\{ & n_1 n_2 (n_1^2 + n_2^2) - (n_1^2 + n_2^2 \\ & - 2\mathrm{NA}^2) \sqrt{n_1^2 - \mathrm{NA}^2}\, \sqrt{n_2^2 - \mathrm{NA}^2} + (n_1^2 \\ & - n_2^2)^2 \log\left(\frac{\sqrt{n_1^2 - \mathrm{NA}^2} + \sqrt{n_2^2 - \mathrm{NA}^2}}{n_1 + n_2} \right) \Bigg\}. \end{aligned} \tag{26}$$

The optimum value of α is found by minimization of the quadratic expression for ${W_{\mathrm{rms}}}^2$ in α:

$$\alpha = -\frac{\langle f_1 f_2 \rangle - \langle f_1 \rangle \langle f_2 \rangle}{\langle {f_1}^2 \rangle - \langle f_1 \rangle^2}. \tag{27}$$

The rms wave-front aberration can now be expressed in terms of the calculated pupil averages:

$$\left(\frac{W_{\mathrm{rms}}}{\Delta d} \right)^2 = \langle {f_2}^2 \rangle - \langle f_2 \rangle^2 - \frac{[\langle f_1 f_2 \rangle - \langle f_1 \rangle \langle f_2 \rangle]^2}{\langle {f_1}^2 \rangle - \langle f_1 \rangle^2}. \tag{28}$$

These analytical expressions allow for an easy evaluation of the axial focus shift and the rms aberration in all cases. Figure 2 (solid curve) shows the rms aberration as a function of NA for $n_1 = 1$ (air-incident case) and $n_2 = 1.6$. Apparently, $W_{\mathrm{rms}}/\Delta d$ increases strongly with NA up to a value of $W_{\mathrm{rms}}/\Delta d = 0.017$ for the limiting case NA = 1.

The exact results can be compared with the paraxial limit of low NA. A straight-forward Taylor expansion gives a focal shift to lowest order in NA,

$$\Delta z_{\mathrm{par}} = \frac{n_2 - n_1}{n_2} d, \tag{29}$$

and to second order in NA,

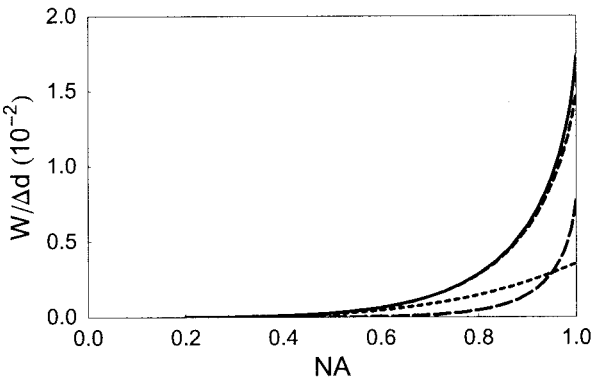

Fig. 2. Exact rms aberration for spherical aberration from thickness mismatch normalized by the thickness mismatch (solid curve), the rms aberration according to the paraxial approximation (small-dashed curve), the lowest-order spherical aberration contribution W_{40} (medium-dashed curve, nearly coinciding with the solid curve), and the higher-order spherical-aberration contributions $\sqrt{{W_{60}}^2 + {W_{80}}^2}$ (long-dashed curve) as a function of NA for $n_1 = 1.0$ and $n_2 = 1.6$.

$$\Delta z = \Delta z_{\mathrm{par}} \left(1 + \frac{n_1 + n_2}{4 n_1 {n_2}^2} \mathrm{NA}^2 \right), \tag{30}$$

where the term in brackets gives the lowest-order correction term of the paraxial result. Similarly, the rms aberration to lowest (fourth) order in NA is

$$W_{\mathrm{rms,\,par}} = \frac{1}{6\sqrt{5}} \frac{{n_2}^2 - {n_1}^2}{8 {n_2}^3 {n_1}^2} \Delta d \mathrm{NA}^4, \tag{31}$$

giving the well-known paraxial NA^4 scaling for spherical aberration, and to the next (sixth) order in NA,

$$W_{\mathrm{rms}} = W_{\mathrm{rms,par}} \left(1 + \frac{3{n_1}^2 + 2{n_2}^2}{4 {n_1}^2 {n_2}^2} \mathrm{NA}^2 \right), \tag{32}$$

where the term in brackets gives the lowest-order correction term of the paraxial result. The accuracy of the paraxial result can be estimated based on the correction term. If the relative error is required to be smaller than ϵ, then the NA must be smaller than $2n_1 n_2 \sqrt{\epsilon} / \sqrt{3{n_1}^2 + 2{n_2}^2}$, which boils down to an upper limit of NA = 0.36 for $n_1 = 1$ and $n_2 = 1.6$ and an error of less than 10%. Clearly, this upper limit is lower than the NA values of current optical disk formats, proving the inadequacy of the paraxial scaling laws.

The paraxial rms aberration can be alternatively derived by starting from the aberration function with respect to the Gaussian focus (this expression is identical to the one derived by Braat[7] for $n_1 = 1$):

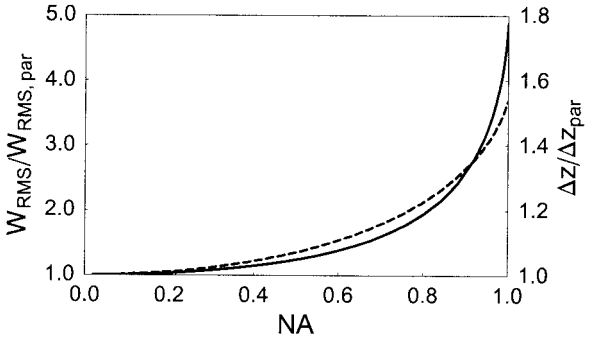

Fig. 3. Ratio of the exact rms aberration and the paraxial NA^4 estimate for spherical aberration from thickness mismatch (solid curve, left axis) and the ratio of the exact focal shift and the paraxial focal shift for spherical aberration from thickness mismatch (dashed curve, right axis) for $n_1 = 1.0$ and $n_2 = 1.6$.

$$W_{O'} = \Delta d\left(\sqrt{{n_2}^2 - \mathrm{NA}^2\rho^2} - \frac{n_1}{n_2}\sqrt{{n_1}^2 - \mathrm{NA}^2\rho^2} - \frac{{n_2}^2 - {n_1}^2}{n_2}\right). \quad (33)$$

A Taylor expansion in $\rho\mathrm{NA}$ gives to lowest order the fourth-order Seidel term:

$$W_{\mathrm{par}} = \frac{{n_2}^2 - {n_1}^2}{8{n_2}^3{n_1}^2}\,\Delta d\mathrm{NA}^4\rho^4. \quad (34)$$

The choice of the Gaussian focus as reference point does not give rise to a minimum rms aberration in the paraxial limit. In order to find the paraxial diffraction focus, a defocus term must be added to W_{par}, and the rms value of the resulting aberration function must be minimized. This is a common practice in optical design, described in, e.g., Ref. 10, and results in Eq. (31).

It appears that the paraxial results are far from accurate for the NA values in practical optical disk readout systems, as was already apparent from the Taylor expansion in NA. Figure 2 (small dashed curve) shows the paraxial result, and Fig. 3 shows the ratio of the focal shift and the paraxial focal shift $\Delta z/\Delta z_{\mathrm{par}}$ and the ratio of the true rms aberration and the paraxial rms aberration $W_{\mathrm{rms}}/W_{\mathrm{rms,par}}$ as a function of NA for $n_1 = 1$ and $n_2 = 1.6$. Clearly, both ratios increase with NA, very steeply when $\mathrm{NA} \to 1$. For NA = 0.50 (CD), the deviation of Δz from Δz_{par} is only 7%, for NA = 0.54 (DVD) it is already 13%, and for NA = 0.85 (BD) it is even 27%. The low NA formula for the rms aberration underestimates the true aberration for NA = 0.50 (CD) with a factor 1.24, for NA = 0.65 (DVD) with a factor 1.48, and, for NA = 0.85 (BD) even with a factor 2.20. Apparently, spherical aberration increases significantly more rapidly with NA than with the well-known NA^4 law.

An alternative to the Seidel expansion of the exact aberration function in powers of $\rho\mathrm{NA}$ is the expansion in Zernike polynomials of azimuthal order $m = 0$, ${Z_{2k}}^0(\rho) = P_k(2\rho^2 - 1)$, where $P_k(x)$ are the Legendre polynomials. The associated Zernike coefficients follow from the integrals

$$\begin{aligned} A_{2k,0} &= \frac{2k+1}{\pi}\int_P \mathrm{d}^2\rho W_{O'}(\rho){Z_{2k}}^0(\rho) \\ &= 2(2k+1)\int_0^1 \mathrm{d}\rho\ \rho W_{O'}(\rho){Z_{2k}}^0(\rho) \\ &= \Delta d n_2 \xi_k\left(\frac{\mathrm{NA}}{n_2}\right) + \alpha\Delta d n_1 \xi_k\left(\frac{\mathrm{NA}}{n_1}\right), \end{aligned} \quad (35)$$

where the functions $\xi_k(\epsilon)$ are given by

$$\xi_k(\epsilon) = 2(2k+1)\int_0^1 \mathrm{d}\rho\ \rho\sqrt{1-\epsilon^2\rho^2}P_k(2\rho^2 - 1). \quad (36)$$

For $k = 0$ (piston), $k = 1$ (defocus), $k = 2$ (lowest-order spherical aberration), and $k = 3$ (higher-order spherical aberration) it may be found (using Mathematica) that

$$\xi_0(\epsilon) = \frac{1 - \sqrt{1-\epsilon^2}\,(1-\epsilon^2)}{3\epsilon^2}, \quad (37)$$

$$\xi_1(\epsilon) = \frac{(8 - 10\epsilon^2) - \sqrt{1-\epsilon^2}\,(8 - 6\epsilon^2 - 2\epsilon^4)}{15\epsilon^4}, \quad (38)$$

$$\xi_2(\epsilon) = \frac{(96 - 168\epsilon^2 + 70\epsilon^4) - \sqrt{1-\epsilon^2}\,(96 - 120\epsilon^2 + 22\epsilon^4 + 2\epsilon^6)}{105\epsilon^6}, \quad (39)$$

$$\xi_3(\epsilon) = \frac{(640 - 1440\epsilon^2 + 1008\epsilon^4 - 210\epsilon^6}{315\epsilon^8} - \frac{\sqrt{1-\epsilon^2}\,(640 - 1120\epsilon^2 + 528\epsilon^4 - 46\epsilon^6 - 2\epsilon^8)}{315\epsilon^8}. \quad (40)$$

Analytical expressions for $k > 3$ can also be found but are not listed here. It appears that the largest contribution to the overall rms aberration is from the (lowest-order) spherical aberration coefficient A_{40}. Figure 2 shows the contribution to the total rms aberration from A_{40} (medium-dashed curve) and from the higher-order spherical aberrations A_{60} and A_{80} (long-dashed curve). The A_{40} line nearly coincides with the total rms line, apart from the region very close to A_{40}. The higher-order spherical aberrations contribute significantly only in this region. For $n_1 = 1$, $n_2 = 1.6$, the approximation of the total rms aberration with the lowest-order Zernike contribution A_{40} is accurate within 0.2% for NA = 0.50 (CD),

within 0.5% for NA = 0.65 (DVD), and within 2.3% for NA = 0.85 (BD). It may be concluded that the expansion in Zernike terms converges much faster than the expansion in Seidel terms for the relevant NA values.

The sensitivity for defocus can be analyzed in much the same way. An axial displacement Δz of the lens that converges the beam to O results in an aberration function:

$$W = \Delta z(\sqrt{n_1^2 - \mathrm{NA}^2\rho^2} - n_1). \tag{41}$$

It follows that the rms aberration may be expressed in terms of the averages of the function f_1 as

$$\left(\frac{W_{\mathrm{rms}}}{\Delta z}\right)^2 = \langle f_1^2\rangle - \langle f_1\rangle^2. \tag{42}$$

In the paraxial limit this evaluates to

$$\left(\frac{W_{\mathrm{rms}}}{\Delta z}\right)^2 = \frac{\mathrm{NA}^4}{48 n_1^2}. \tag{43}$$

According to the Maréchal criterion,[8] the spot is a so-called diffraction-limited spot if the Strehl ratio is above $1 - \pi^2/48 \approx 0.8$, or, equivalently, if the rms aberration is below $\lambda/8\sqrt{3} = 0.072\,\lambda$. This corresponds to a peak-valley value of the defocus aberration function equal to $\lambda/4$ and gives the paraxial depth of focus (the maximum defocus value for which the spot is diffraction limited) as $\Delta z = n_1\lambda/2\mathrm{NA}^2$. Equation (42) gives the nonparaxial corrections to the depth of focus. Clearly, the defocus sensitivity increases with NA. This implies that the actuator that controls the position of the objective lens along the optical axis, which is needed to keep the scanning spot in focus, needs to satisfy more stringent requirements as the NA increases.[11]

4. Disk Tilt and Coma

We consider the coma resulting from tilting the plate of thickness d over an angle β in the xz plane. The plate normal is thus $\hat{n} = (-\sin\beta,\ 0,\ \cos\beta)$. It is convenient to introduce a new coordinate frame, the $x'y'z'$ frame. It is obtained from the xyz frame by rotating around the y axis over an angle β. The components of the ray vectors parallel to the plate are here the x' and y' components:

$$\begin{aligned} g_{1x'} &= g_{2x'} \\ &= \cos\beta\, g_{1x} + \sin\beta\, g_{1z} \\ &= -\cos\beta\ \mathrm{NA}\rho\cos\theta + \sin\beta\sqrt{n_1^2 - \mathrm{NA}^2\rho^2}, \end{aligned} \tag{44}$$

$$g_{1y'} = g_{2y'} = -\mathrm{NA}\rho\sin\theta. \tag{45}$$

The perpendicular components are just the z' components:

$$\begin{aligned} g_{1z'} &= -\sin\beta g_{1x} + \cos\beta g_{1z} \\ &= \sin\beta\mathrm{NA}\rho\cos\theta + \cos\beta\sqrt{n_1^2 - \mathrm{NA}^2\rho^2}, \end{aligned} \tag{46}$$

$$\begin{aligned} g_{2z'} &= \sqrt{n_2^2 - g_{1x'}^2 - g_{1y'}^2} \\ &= [n_2^2 - (\cos\beta\ \mathrm{NA}\ \rho\cos\theta - \sin\beta\sqrt{n_1^2 - \mathrm{NA}^2\rho^2}) \\ &\quad + \mathrm{NA}^2\rho^2\sin^2\theta]^{1/2}. \end{aligned} \tag{47}$$

The aberration function for the reference point O is then (leaving out the piston contribution from the chief ray for brevity)

$$\begin{aligned} W_O &= d[n_2^2 - (\cos\beta\ \mathrm{NA}\rho\cos\theta - \sin\beta\sqrt{n_1^2 - \mathrm{NA}^2\rho^2})^2 \\ &\quad + \mathrm{NA}^2\rho^2\sin^2\theta]^{1/2} - d(\sin\beta\mathrm{NA}\rho\cos\theta \\ &\quad + \cos\beta\sqrt{n_1^2 - \mathrm{NA}^2\rho^2}) \\ &= d\sqrt{n_2^2 - n_1^2\sin^2\beta}\ \sqrt{1-t} - d(\sin\beta\mathrm{NA}\rho\cos\theta \\ &\quad + \cos\beta\sqrt{n_1^2 - \mathrm{NA}^2\rho^2}), \end{aligned} \tag{48}$$

with

$$\begin{aligned} t &= [-\sin(2\beta)\mathrm{NA}\rho\cos\theta\sqrt{n_1^2 - \mathrm{NA}^2\rho^2} \\ &\quad + \cos(2\beta)\mathrm{NA}^2\rho^2\cos^2\theta + \cos^2\beta\mathrm{NA}^2\rho^2\sin^2\theta]/(n_2^2 \\ &\quad - n_1^2\sin^2\beta). \end{aligned} \tag{49}$$

This expression is identical to Braat's[7] for $n_1 = 1$ (apart from the irrelevant contribution from the chief ray). The aberration introduced by a small disk tilt β follows from a Taylor expansion in β as

$$W_O = \left(\sqrt{\frac{n_1^2 - \mathrm{NA}^2\rho^2}{n_2^2 - \mathrm{NA}^2\rho^2}} - 1\right) d\beta\mathrm{NA}\rho\cos\theta. \tag{50}$$

With a lateral focus shift Δx, the aberration function for the reference point O' can be written as

$$W_{O'} = d\beta(h_2 + \mu h_1), \tag{51}$$

where

$$h_1 = \mathrm{NA}\rho\cos\theta, \tag{52}$$

$$h_2 = \mathrm{NA}\rho\cos\theta\sqrt{\frac{n_1^2 - \mathrm{NA}^2\rho^2}{n_2^2 - \mathrm{NA}^2\rho^2}}, \tag{53}$$

and

$$\mu = \frac{\Delta x}{d\beta} - 1. \tag{54}$$

In order to find the diffraction focus, we proceed similar to the case of spherical aberration. The relevant pupil averages are

$$\langle h_1\rangle = \langle h_2\rangle = 0, \tag{55}$$

$$\langle h_1{}^2\rangle = \frac{\mathrm{NA}^2}{4}, \tag{56}$$

$$\langle h_2{}^2\rangle = \frac{1}{4\mathrm{NA}^2}\left[(2n_2{}^2 - 2n_1{}^2 + \mathrm{NA}^2)\mathrm{NA}^2 + 2n_2{}^2(n_2{}^2 - n_1{}^2)\log\left(\frac{n_2{}^2 - \mathrm{NA}^2}{n_2{}^2}\right)\right], \tag{57}$$

$$\langle h_1h_2\rangle = \frac{1}{8\mathrm{NA}^2}\left[n_1n_2(3n_2{}^2 - n_1{}^2) - (3n_2{}^2 - n_1{}^2 + 2\mathrm{NA}^2)\sqrt{n_1{}^2 - \mathrm{NA}^2}\,\sqrt{n_2{}^2 - \mathrm{NA}^2} + (3n_2{}^2 + n_1{}^2)(n_2{}^2 - n_1{}^2)\log\left(\frac{\sqrt{n_1{}^2 - \mathrm{NA}^2} + \sqrt{n_2{}^2 - \mathrm{NA}^2}}{n_1 + n_2}\right)\right]. \tag{58}$$

The optimum value of μ is given by

$$\mu = -\frac{\langle h_1h_2\rangle - \langle h_1\rangle\langle h_2\rangle}{\langle h_1{}^2\rangle - \langle h_1\rangle^2} = -\frac{\langle h_1h_2\rangle}{\langle h_1{}^2\rangle}. \tag{59}$$

The rms wave-front aberration can be expressed in terms of the calculated pupil averages as

$$\left(\frac{W_{\mathrm{rms}}}{d\beta}\right)^2 = \langle h_2{}^2\rangle - \langle h_2\rangle^2 - \frac{[\langle h_1h_2\rangle - \langle h_1\rangle\langle h_2\rangle]^2}{\langle h_1{}^2\rangle - \langle h_1\rangle^2} = \langle h_2{}^2\rangle - \frac{\langle h_1h_2\rangle^2}{\langle h_1{}^2\rangle}. \tag{60}$$

Figure 4 (solid curve) shows the rms aberration as a function of NA for $n_1 = 1$ (air-incident case) and $n_2 = 1.6$. Apparently, $W_{\mathrm{rms}}/d/\beta$ increases strongly with NA up to a value of $W_{\mathrm{rms}}/d/\beta = 0.069$ for the limiting case NA = 1.

The exact results can be compared with the paraxial limit of low NA by a Taylor expansion in NA. The focal shift to lowest order in NA is

$$\Delta x_{\mathrm{par}} = \frac{n_2 - n_1}{n_2}\,d\beta, \tag{61}$$

and to second order in NA

$$\Delta x = \Delta x_{\mathrm{par}}\left(1 + \frac{n_1 + n_2}{3n_1n_2{}^2}\,\mathrm{NA}^2\right), \tag{62}$$

where the term in brackets gives the lowest-order correction term of the paraxial result. The rms aberration to lowest (third) order in NA is

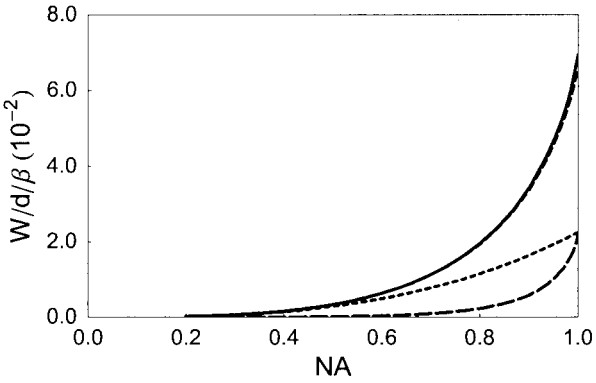

Fig. 4. Exact rms aberration for coma from disk tilt normalized by the product of substrate thickness and disk tilt angle (solid curve), the rms aberration according to the paraxial approximation (small-dashed curve), the lowest-order coma contribution W_{31} (medium-dashed curve, nearly coinciding with the solid curve) and the higher-order coma contributions $\sqrt{W_{51}{}^2 + W_{71}{}^2}$ (long-dashed curve) as a function of NA for $n_1 = 1.0$ and $n_2 = 1.6$.

$$W_{\mathrm{rms,par}} = \frac{1}{3\sqrt{8}}\,\frac{n_2{}^2 - n_1{}^2}{2n_2{}^3n_1}\,d\beta\mathrm{NA}^3. \tag{63}$$

giving the well-known paraxial NA^3 scaling for coma, and to the next (fifth) order in NA

$$W_{\mathrm{rms}} = W_{\mathrm{rms,par}}\left[1 + \frac{3(3n_1{}^2 + n_2{}^2)}{10n_1{}^2n_2{}^2}\,\mathrm{NA}^2\right], \tag{64}$$

where the term in brackets gives the lowest-order correction term of the paraxial result. If the relative error is required to be smaller than ϵ, then the NA must be smaller than $n_1n_2\sqrt{10\epsilon}/\sqrt{3(3n_1{}^2 + n_2{}^2)}$, which boils down to an upper limit of NA = 0.39 for $n_1 = 1$ and $n_2 = 1.6$ and an error of less than 10%.

The paraxial rms aberration can also be derived by starting from the aberration function with respect to the Gaussian focus:

$$W_{O'} = \left(\sqrt{\frac{n_1{}^2 - \mathrm{NA}^2\rho^2}{n_2{}^2 - \mathrm{NA}^2\rho^2}} - \frac{n_1}{n_2}\right)d\beta\mathrm{NA}\rho\cos\theta. \tag{65}$$

The leading lowest-order term is Seidel coma:

$$W_{\mathrm{par}} = -\frac{n_2{}^2 - n_1{}^2}{2n_2{}^3n_1}\,d\beta\mathrm{NA}^3\rho^3\cos\theta. \tag{66}$$

The paraxial diffraction focus and rms aberration are obtained by adding a wave-front tilt term and minimizing the resulting rms, similar to the case of spherical aberration.

Again, the paraxial results are inaccurate for the NA values in present optical disk readout systems. Figure 4 (small-dashed curve) shows the rms aberration according to the paraxial approximation; Fig. 5 shows the increases with NA of the ratio of the lateral focal shift and the paraxial focal shift $\Delta x/\Delta x_{\mathrm{par}}$ and of the ratio of the true rms aberration and the paraxial rms aberration $W_{\mathrm{rms}}/W_{\mathrm{rms,par}}$ as a function of NA for $n_1 = 1$ and $n_2 = 1.6$. For NA = 0.50 (CD), the deviation of the focal shift from the paraxial result is only 9%, for NA = 0.65 (DVD) it is already 17%, and for NA = 0.85 (BD) it is even 36%. The ratio $W_{\mathrm{rms}}/W_{\mathrm{rms,par}}$ is 1.19 for NA = 0.50 (CD), 1.37 for NA = 0.65 (DVD), and 1.86 for NA = 0.85 (BD). Clearly, coma increases significantly more rapidly with NA than the well-known NA^3 law.

The resulting aberration function may be expanded in Zernike polynomials of azimuthal order $m = 1$. The associated Zernike coefficients follow from the integrals

$$A_{2k+1,1} = \frac{4k+4}{\pi}\int_P \mathrm{d}^2\rho W_{O'}(\rho)Z_{2n+1}{}^1(\rho)\cos\theta = d\beta\mathrm{NA}\,\frac{n_1}{n_2}\,\zeta_k\left(\frac{\mathrm{NA}}{n_1},\ \frac{\mathrm{NA}}{n_2}\right) + \mu d\beta\mathrm{NA}\delta_{k0}, \tag{67}$$

where the function $\zeta_k(\epsilon_1,\ \epsilon_2)$ is defined by

$$\zeta_k(\epsilon_1,\ \epsilon_2) = 4(k+1)\int_0^1 \mathrm{d}\rho\ \rho^3\sqrt{\frac{1-\epsilon_1{}^2\rho^2}{1-\epsilon_2{}^2\rho^2}}\,P_k{}^{01}(2\rho^2-1). \tag{68}$$

For $k = 0$ (wave-front tilt) and $k = 1$ (lowest-order coma), these functions are

$$\zeta_0(\epsilon_1,\ \epsilon_2) = \frac{1}{8\epsilon_1{}^3\epsilon_2{}^5}\left[\epsilon_1\epsilon_2(3\epsilon_1{}^2-\epsilon_2{}^2) - \epsilon_1\epsilon_2(3\epsilon_1{}^2 + 2\epsilon_1{}^2\epsilon_2{}^2+\epsilon_2{}^2)\sqrt{(1-\epsilon_1{}^2)(1-\epsilon_2{}^2)} + (\epsilon_1{}^2-\epsilon_2{}^2)(3\epsilon_1{}^2+\epsilon_2{}^2)\log\left(\frac{\epsilon_1\sqrt{1-\epsilon_1{}^2}+\epsilon_2\sqrt{1-\epsilon_1{}^2}}{\epsilon_1+\epsilon_2}\right)\right], \tag{69}$$

$$\zeta_1(\epsilon_1,\ \epsilon_2) = \frac{1}{16\epsilon_1{}^5\epsilon_2{}^7}\left[\epsilon_1\epsilon_2(15\epsilon_1{}^4-3\epsilon_2{}^2+4\epsilon_1{}^2\epsilon_2{}^2 - 12\epsilon_1{}^4\epsilon_2{}^2+4\epsilon_1{}^2\epsilon_4{}^2) - \epsilon_1\epsilon_2(15\epsilon_1{}^4-3\epsilon_2{}^2 - 4\epsilon_1{}^2\epsilon_2{}^2-2\epsilon_1{}^4\epsilon_2{}^2 + 2\epsilon_1{}^2\epsilon_4{}^2)\sqrt{(1-\epsilon_1{}^2)(1-\epsilon_2{}^2)} + (\epsilon_1{}^2-\epsilon_2{}^2)(15\epsilon_1{}^4+3\epsilon_2{}^2+6\epsilon_1{}^2\epsilon_2{}^2-12\epsilon_1{}^4\epsilon_2{}^2 - 4\epsilon_1{}^2\epsilon_4{}^2)\log\left(\frac{\epsilon_1\sqrt{1-\epsilon_2{}^2}+\epsilon_2\sqrt{1-\epsilon_1{}^2}}{\epsilon_1+\epsilon_2}\right)\right]. \tag{70}$$

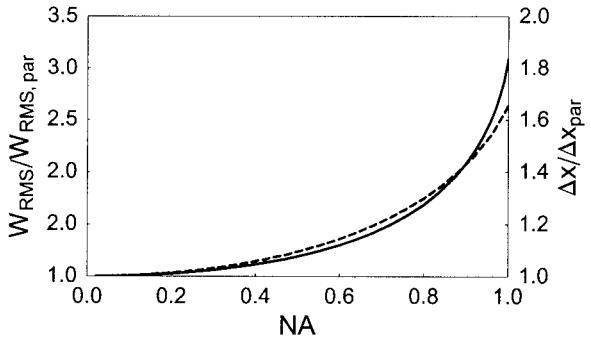

Fig. 5. Ratio of the exact rms aberration and the paraxial NA^3 estimate for coma from disk tilt (solid curve, left axis) and the ratio of the exact focal shift and the paraxial focal shift for coma from disk tilt (dashed curve, right axis) as a function of NA for $n_1 = 1.0$ and $n_2 = 1.6$.

Analytical expressions for higher-order coma coefficients can be obtained easily using Mathematica but are too lengthy to fit to the paper. It turns out that the coma of lowest order has the largest contribution to the total rms aberration, as can be seen in Fig. 4. The contribution to the total rms aberration from A_{31} (medium-dashed curve) nearly coincides with the total rms curve. A small deviation can be seen in the region very close to NA = 1, the only region in which the higher-order coma contributions from A_{51} and A_{71} (long-dashed curve) differ appreciably from zero. For $n_1 = 1$, $n_2 = 1.6$, the approximation of the total rms aberration with the lowest-order Zernike contribution A_{31} is accurate within 0.07% for NA = 0.50 (CD), within 0.3% for NA = 0.65 (DVD), and within 1.0% for NA = 0.85 (BD). In conclusion, the expansion in Zernike terms converges much faster than the expansion in Seidel terms for the relevant NA values, just as for the case of spherical aberration.

5. Discussion

The aberration sensitivities of the optical disk formats CD, DVD, AOD, and BD may be compared with each other using the exact values given in Table 1. The rms values are expressed in milliwaves (mλ), as that is the relevant measure for the quality of the scanning focal spot. The Maréchal criterion (rms aberration below $\lambda/8\sqrt{3} = 72$ mλ) gives a rule of thumb to calculate tolerances for thickness mismatch in micrometers and tolerances for disk tilt in degrees. From the values in Table 1, it is clear that the decrease in λ and the increase in NA greatly increases the sensitivities, and hence greatly decreases the tolerances for substrate-related imperfections. Especially the recently proposed AOD and BD formats have a small tolerance for disk tilt (AOD) and thickness mismatch (BD). Assuming that a third of the aberration budget of 72 mλ may be spent on substrate-related aberrations, it follows that the disk tilt margin for AOD is only 0.1°, whereas the thickness mismatch margin for BD is only 2.4 μm. These margins are difficult to attain in practice, meaning that an apparatus to write/read AODs must have a

Table 1. Aberration Sensitivities of Optical Disk Formats

Format	λ (μm)	NA	d (mm)	DF[a] (mλ/μm)	CO[a] (mλ/deg)	SA[a] (mλ/μm)
CD[b]	0.785	0.45–0.50	1.2	39–49	63–89	0.22–0.35
DVD[b]	0.655	0.60–0.65	0.6	88–105	101–134	0.97–1.42
AOD	0.405	0.65	0.6	171	218	2.31
BD	0.405	0.85	0.1	332	110	10.07

[a]Sensitivities for defocus (DF), coma from disk tilt (CO), and spherical aberration from thickness mismatch (SA).
[b]The sensitivities are given for the NA values of a read-only drive (lower value) and for a writing drive (higher value).

correction mechanism for disk tilt,[12,13] whereas an apparatus to write/read BDs must have a correction mechanism for thickness mismatch.[14] The sensitivity for thickness mismatch is a problem for both AOD and BD when dual-layer disks are considered. The thickness of the spacer layer between the two data layers is taken to be 40 μm (AOD) and 25 μm (BD), and the objective lens is assumed to be corrected for the depth halfway between the two layers. The spherical aberration is then ± 46 mλ for AOD, and ± 126 mλ for BD, implying that both systems need means for spherical-aberration correction when dual-layer disks are read/written.

The equations derived in this paper are valid for arbitrary NA, and arbitrary refractive indices n_1 and n_2, provided that all waves are propagating, i.e., provided that NA $\leq n_1$ and NA $\leq n_2$. In particular, near-field optical disks may be considered. Here, a solid immersion lens of refractive index n_1 is brought into near contact with the disk, so that the evanescent waves can tunnel through the subwavelength gap into the disk, so that the NA can be larger than the air-incident limit of one. For thermal control reasons, it is advisable to use a thin protective cover layer of refractive index n_2. Deviations in thickness of this cover layer from the nominal thickness will result in spherical aberration. It turns out that this spherical aberration is not excessively high in the region NA > 1, as naive expectations would have it. Instead, it is the difference in the refractive indices n_1 and n_2 and the deviation of NA from the smallest of the two refractive indices that control the aberration sensitivity. For example, for $n_1 = 2.2$, $n_2 = 1.7$, NA $= 1.4$, and $\lambda = 0.405$ μm, the sensitivity for spherical aberration is 19.4 mλ/μm, only a factor of 2 larger than for the case of BD, which is quite acceptable. It is mentioned that, although the gap between the solid immersion lens and the disk is small, the effects of it on the amplitude, phase, and polarization of the transmitted light cannot be neglected, and must be added to the spherical-aberration effect of thickness mismatch discussed here.

The formalism presented in this paper can be extended to (uniaxial) birefringent media as well. For the ordinary mode, there is no difference with isotropic media, for the extraordinary mode the ray vectors, which by definition are normal to the wave fronts, are not parallel anymore to the actual rays. However, this is of no importance for the optical path lengths, which are still given by the inner product of ray vectors and position vectors, similar to the formalism described in Section 2. The birefringent nature of the medium is expressed by the eikonal equation for the extraordinary ray vector $\mathbf{g}_e$, which may be shown to satisfy an equation of the same form as the dispersion relation for the extraordinary-mode plane waves in a uniaxially birefringent medium.[15] The study of the effect of birefringence on aberrations within the framework of geometrical optics is left for future work.

Teus Tukker and Benno Hendriks are thanked for suggestions to improve the manuscript.

References

1. 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).
2. 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).
3. 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).
4. T. Sugaya, "A basic concept for next generation DVD: 0.6 mm substrate disk technology using blue-violet laser," in *Optical Data Storage 2003*, M. O'Neill and N. Miyagawa, eds., Proc. SPIE **5069,** 278–280 (2003).
5. W. T. Welford, *Aberrations of Optical Systems* (Hilger, London, 1986).
6. G. Bouwhuis, J. Braat, A. Huijser, J. Pasman, G. van Rosmalen, and K. Schouhamer Immink, *Principles of Optical Disc Systems* (Hilger, London, 1985).
7. J. Braat, "Analytical expressions for the wave-front aberration coefficients of a tilted plane-parallel plate," Appl. Opt. **36,** 8459–8466 (1997).
8. M. Born and E. Wolf, *Principles of Optics, 6th ed.* (Cambridge University, Cambridge, UK, 1980).
9. S. Wolfram, *Mathematica. A System for Doing Mathematics by Computer, 2nd ed.* (Addison-Wesley, Reading, Mass., 1991).
10. W. J. Smith, *Modern Optical Engineering. The Design of Optical Systems, 2nd ed.* (McGraw-Hill, Boston, Mass., 1990).
11. P. Asthana, B. I. Finkelstein, and A. A. Fennema, "Rewritable optical disk drive technology," IBM J. Res. Dev. **40,** 543–558 (1996).
12. S. Stallinga, J. J. Vrehen, J. Wals, E. Stapert, and E. Verstegen, "Liquid crystal aberration compensation devices," in *Optical Storage and Information Processing*, H.-P. D. Shieh and T. D. Milster, eds., Proc. SPIE **4081,** 50–56 (2000).

13. S. Stallinga and J. J. Vrehen, "Advances in liquid crystal tilt compensation," in *Optical Data Storage 2001*, T. Hurst and S. Kobayashi, eds., Proc. SPIE **4342,** 465–473 (2001).
14. H. R. Stapert, J. Lub, E. J. K. Verstegen, B. M. I. van der Zande, and S. Stallinga, "Photo replicated anisotropic liquid crystalline lenses for aberration control and dual layer read out of optical disks," Adv. Functional Mater. **13,** 732–738 (2003).
15. M. P. van der Meulen, Institute for Theoretical Physics, Valckenierstraat 65 1018 XE Amsterdam, The Netherlands (personal communication, 2001).