NOTE ON THE BEST FOCUS IN THE PRESENCE OF SPHERICAL ABERRATION

By WANG TA-HANG,
Imperial College of Science and Technology

Communicated by L. C. Martin. Received 14 *November* 1940

ABSTRACT. The location of the best axial focus of an optical system in the presence of spherical aberrations of various orders is considered upon the criterion of maximum axial light intensity. Adhering to Rayleigh's quarter-wave-length limit, it is shown by approximate formulation that for axial aberrations sensible up to the primary, secondary and tertiary order, the allowable tolerance for such parts of the aberrations when stated in terms of optical path-difference can reach λ, 4λ, and 15λ respectively, provided that the foci of various zones are suitably controlled in design.

§ 1. INTRODUCTION

The present paper deals with the location of the best focus of an axial pencil of rays in the image space when residual spherical aberrations of various orders are present. The treatment is based on the criterion that the best location is the point where the axial light intensity is a maximum. Cases are considered when spherical aberrations of various orders are under control in practical designing work. Rayleigh's quarter-wave-length limit is adhered to because an optical instrument would fall very little short of its utmost theoretical perfection if all the light arrived at the focus with differences of optical path not exceeding that limit. This has been proved both mathematically (Conrady, 1919; Buxton, 1921; Martin, 1926) and experimentally (Conrady, 1926). It also simplifies the mathematical treatment given in the present paper. Results obtained are compared with those obtained by other writers and published elsewhere. Methods applied to practical design are also discussed, with special emphasis on the case of tertiary aberration.

§ 2. THEORY

Let B in figure 1 represent a point on the optical axis near the final focus in the image space of an axially symmetrical optical system. AP is a spherical surface concentric with B. We can consider this as a reference surface, with which the wave front (originating from an axial luminous object point and distorted on passing through this optical system) can be compared. In accordance with the notation used by Professor A. E. Conrady, the product of the radial distance from the reference surface to the wave front and the refractive index in the space concerned is termed the optical path difference, written symbolically as OPD. If y is the intersection height of the ray with this reference surface (or a quantity proportional to the numerical aperture of the ray) and y_0 that at the extreme margin, then OPD can be put in the form of a power series:

$$\mathrm{OPD_B} = a_0 + a_1\left(\frac{y}{y_0}\right)^2 + a_2\left(\frac{y}{y_0}\right)^4 + a_3\left(\frac{y}{y_0}\right)^6 + \ldots, \qquad \ldots\ldots(1)$$

where the suffix B means a value taken with respect to a reference surface with B as centre. Professor Conrady's treatment disregarded the constant term a_0 in the above expression. In that case the wave front has its vertex in contact with the reference surface. We will, however, retain this constant term for a reason to be given below, but at present we restrict it to the magnitude of the order of a few wave-lengths of light, so that it becomes comparable in magnitude with the other terms in the same expression.

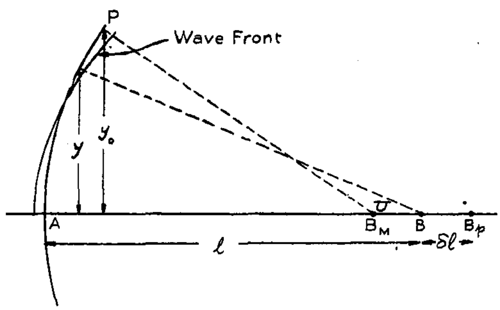

Figure 1.

Imagine the point B to lie in the vicinity of the best focus. It follows that the various terms in expression (1) will have the following meanings :

a_0 *term.* As pointed out above, this term is related to the proper reference surface chosen. The choice of its value is arbitrary and there is no effect on the quality of the image formation, but we shall see in the following sections how a proper choice of this term leads to simplification of the mathematical treatment, where a_0 is set at a value representing the " mean phase " of the wave front.

$a_1(y/y_0)^2$ *term.* This term is related to the position of the point B and is the only sensible term representing the focal shift, provided that B is very near to the focus and the angle of the cone of rays is not too big. To see this point, note that at the paraxial focus, the OPD expression starts from the term in y^4 (disregarding the constant term a_0). It is not difficult to show, using Fermat's principle and simple trigonometrical deductions, that that part of OPD caused by a focal shift δl from the paraxial focus is given by

$$\mathrm{OPD(focal\ shift)} = 2N\delta l \sin^2\frac{U}{2},$$

provided that $(\delta l)^2$ is negligible, where U is the angle subtended by the height of y from B. Expanding the above expression in powers of $\sin U$ and replacing $\sin U$ by y/l, we obtain

$$\text{OPD (focal shift)} = \tfrac{1}{2}N\delta l\,(\sin^2 U + \tfrac{1}{4}\sin^4 U + \tfrac{1}{8}\sin^6 U + \dots)$$
$$= \tfrac{1}{2}N\delta l\left[\left(\frac{y}{l}\right)^2 + \tfrac{1}{4}\left(\frac{y}{l}\right)^4 + \dots\right].$$

In practice, for an optical instrument with Rayleigh-limit definition, we can very nearly always find one space, either the object space or the image space or spaces between the lenses, in which the marginal U-value does not exceed 1/8 radian. For instance, for a microscope objective, the marginal U-value at the image side usually does not exceed 1/30 radian; for a telescope of considerable light-gathering power the relative aperture is of the order of f/5, corresponding to a marginal U of 1/10 radian. Let us refer hereafter to such a space. Consider the case having the biggest possible value of U in the above expression, say $U = 1/8$. We see that the second term is only 1/250 of the first term, and the third term is still smaller. Hence we have good reason to say that the first term is the only sensible part in the OPD expression affected by focal shift, provided the above-mentioned restriction has been adhered to. Thus

$$a_1\left(\frac{y}{y_0}\right)^2 = \frac{N}{2}\delta l\frac{y^2}{l^2} \quad \text{or} \quad a_1 = \delta l\frac{N}{2}\frac{y_0^2}{l^2}. \qquad \dots\dots(2)$$

Note that a_1 is proportional to the focal shift δl.

One may also notice in the later sections, where spherical aberrations of various orders are dealt with, that it is justifiable to neglect the higher-order terms for the focal shift.

$a_2(y/y_0)^4$ *term.* This term represents the primary spherical aberration.

$a_3(y/y_0)^6$, $a_4(y/y_0)^8$, *etc.* These are terms representing secondary, tertiary, and other higher-order aberrations respectively.

Consider the light intensity at the point B resulting from disturbances derived from each zone at the reference surface, extending from the optical axis to the full aperture. For practical purposes we may assume that the amplitude of the disturbances over the reference surface is constant for all zones, but the phase differs from zone to zone by the amount $\dfrac{2\pi \mathrm{OPD_B}}{N\lambda} = \dfrac{2\pi \mathrm{OPD_B}}{\lambda_0}$, λ being the wavelength in the space concerned and λ_0 that in air. Again, confining ourselves to small apertures as before, by Huygens' principle we can express the disturbances at B at any instant t by the complex expression

$$S_B = \frac{2\pi A}{\lambda l}\int_0^{y_0} e^{i2\pi\left[\left(\frac{t}{T}-\frac{l}{\lambda}\right)+\frac{\mathrm{OPD_B}}{\lambda_0}\right]}\,y\,dy$$
$$= \frac{2\pi A}{\lambda l}e^{i2\pi\left(\frac{t}{T}-\frac{l}{\lambda}\right)}\int_0^{y_0} e^{i2\pi\frac{\mathrm{OPD_B}}{\lambda_0}}\,y\,dy.$$

Put $(y/y_0)^2 = z$. Then

$$\left.\begin{aligned} S_B &= \frac{\pi A}{\lambda l} e^{i2\pi\left(\frac{t}{T}-\frac{l}{\lambda}\right)} \int_0^1 e^{i2\pi \frac{\mathrm{OPD_B}}{\lambda_0}} dz \\ &\equiv S_0(\eta + i\xi), \end{aligned}\right\} \qquad \ldots\ldots(3)$$

where $S_0 = \frac{\pi A}{\lambda l} e^{i2\pi\left(\frac{t}{T}-\frac{l}{\lambda}\right)}$ is the disturbance at B, assuming perfect axial image-formation, when the OPD for all zones vanishes over the whole aperture, and

$$\eta = \int_0^1 \cos 2\pi \frac{\mathrm{OPD_B}}{\lambda_0} dz, \quad \xi = \int_0^1 \sin 2\pi \frac{\mathrm{OPD_B}}{\lambda_0} dz. \qquad \ldots\ldots(4)$$

Thus the amplitude given by the present system as compared with the ideal one bears the ratio

$$\frac{S_B}{S_0} = \eta + i\xi,$$

and the corresponding ratio in intensity measure is given by

$$R^2 \equiv \left|\frac{S_B}{S_0}\right|^2 = \eta^2 + \xi^2. \qquad \ldots\ldots(5)$$

Equations (5), (4), and (1) present R as a function of the coefficients a_1, a_2, a_3, etc. In practical designing work, such as microscope-objective design, where the best axial focus is aimed at, once the general arrangement is assigned, the main object to be achieved consists in adjusting the optical system in all admissible ways so that the controllable aberrations of various orders tend to make the best compromise against the residual uncontrollable ones, i.e. R must be a maximum. Thus, regarding a_j, a_k, etc., as parameters, the above criterion requires that

$$\frac{\partial R}{\partial a_j} = 0, \quad \frac{\partial R}{\partial a_k} = 0, \text{ etc.} \qquad \ldots\ldots(6)$$

The straightforward computation of equation (4) in connection with (6) is rather elaborate. For practical purposes let us introduce certain approximations in the following manner :—

Expand sin and cos $\frac{\mathrm{OPD_B}}{\lambda_0}$ and put into equations (4):

$$\eta = \int_0^1 \left[1 - \frac{1}{2!}\left(\frac{\mathrm{OPD_B}}{\lambda_0}\right)^2 + \frac{1}{4!}\left(\frac{\mathrm{OPD_B}}{\lambda_0}\right)^4 - \ldots\right] dz,$$

$$\xi = \int_0^1 \left[\frac{\mathrm{OPD_B}}{\lambda_0} - \frac{1}{3!}\left(\frac{\mathrm{OPD_B}}{\lambda_0}\right)^3 + \ldots\right] dz.$$

Remembering that in equation (1) the constant a_0 is under our arbitrary choice, set a_0 such that

$$\int_0^1 \mathrm{OPD_B}\, dz = 0. \qquad \ldots\ldots(7)$$

Let the OPD expression bearing the condition (7) be distinguished by the symbol,* Then we have

$$\left.\begin{aligned}\eta &= \int_0^1 \left[1 - \frac{1}{2!}\left(\frac{\text{OPD}_{\text{B}}^*}{\lambda_0}\right)^2 + \frac{1}{4!}\left(\frac{\text{OPD}_{\text{B}}^*}{\lambda_0}\right)^4 + \ldots\right] dz, \\ \xi &= -\int_0^1 \left[\frac{1}{3!}\left(\frac{\text{OPD}_{\text{B}}^*}{\lambda_0}\right)^3 - \ldots\right] dz.\end{aligned}\right\} \quad \ldots\ldots(8)$$

What is the physical significance of condition (7)? Referring to figure 2, the curve O′ABC represents the OPD at various zones extending from $z=0$ to $z=1$. Condition (7) means a choice of the axis OZ such that the shaded areas between the curve and the axis on both sides of the latter are equal. Thus OZ is at the " mean " phase of the curve. It will not be far from midway between the extreme values of OPD_{B}^*. In this diagram these extreme values are at the points O′ and C.

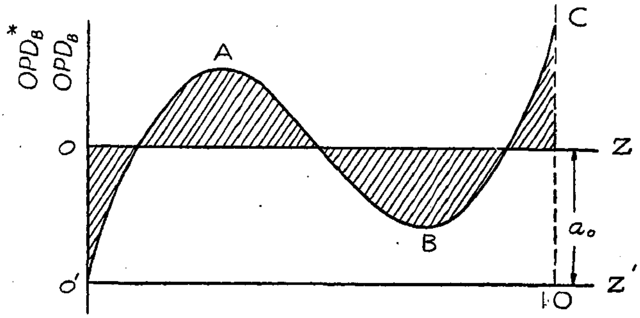

Figure 2.

Having specified the starred OPD_{B}^* value as above, to guard against ambiguity we will hereafter use the plain OPD for those expressions having the constant term $a_0=0$. Thus OPD_{B} refers to the axis O′Z′ where $\text{OPD}_{\text{B}}=0$ at $z=0$.

If we adhere to Lord Rayleigh's quarter-wave-length limit as the tolerance for the extreme difference of OPD_{B}^*, based on the above reasoning, the absolute maximum value of OPD_{B}^* will not be far from $\frac{1}{8}\lambda_0$ when the full tolerance is allowed, whilst a large proportion of the zones will still be covered by values below that amount. When $\text{OPD}_{\text{B}}^*=\frac{1}{8}\lambda_0$, then $\dfrac{2\pi\text{OPD}_{\text{B}}^*}{\lambda_0}=\dfrac{\pi}{4}$, and the magnitude of $\cos\dfrac{2\pi(\text{OPD}_{\text{B}}^*)}{\lambda_0}$ differs from $1-\frac{1}{2}\left(\dfrac{2\pi\text{OPD}_{\text{B}}^*}{\lambda_0}\right)^2$ by not more than 2 per cent. If we retain only the first two terms in the expression for η under the integral in (8), the error introduced into the evaluation of η will seldom be 2 per cent, and the total error will be sensibly less, owing to contributions of zones at smaller OPD_{B}^* under the integral, where the approximation would be more accurate.

Consider next the value of ξ. Take the extreme case by putting $\dfrac{\text{OPD}_{\text{B}}^*}{\lambda_0}=\dfrac{\pi}{4}$ throughout. ξ will have the value 1/10. Hence $|\xi|<\dfrac{1}{10}$, $\xi^2<\dfrac{1}{100}$. The error in the evaluation of $R^2=\eta^2+\xi^2$ by neglecting ξ^2 would hardly exceed 1 per cent.

Consequently, the error due to the replacement of R in the full expression by the approximate value

$$R=\int_0^1\left[1-\tfrac{1}{2}\left(\frac{2\pi \mathrm{OPD}_{\mathrm{B}}^*}{\lambda_0}\right)^2\right]dz \qquad \ldots\ldots(9)$$

will be sensibly less than 5 per cent, probably below 1 per cent. To illustrate this by an example, the writer has evaluated R for a proper $\mathrm{OPD_B}$ (expression (11) in the next section), allowing full tolerance, first, by the simplified expression (9), and then by the expression (8), taking into account one more term in the integrals η and ξ than in our simplified expression. The difference between these two values of R is only 0·3 per cent. Thus there is good reason to adopt the simplified expression (9) provided Rayleigh's quarter-wave-length limit is allowed as tolerance.

We have seen that to find a maximum for R requires the solution of equation (6). Putting R in the expression (9), we get

$$\frac{\partial R}{\partial a_j}=\int_0^1\frac{\partial}{\partial a_j}\left[1-\tfrac{1}{2}\left(\frac{2\pi \mathrm{OPD}_{\mathrm{B}}^*}{\lambda_0}\right)^2\right]dz$$
$$=\int_0^1-\frac{2\pi^2}{\lambda_0^2}\mathrm{OPD}_{\mathrm{B}}^*\frac{\partial}{\partial a_j}(\mathrm{OPD}_{\mathrm{B}}^*)\,dz.$$

Writing $\mathrm{OPD}_{\mathrm{B}}^*$ in ascending powers of z, namely,

$$\mathrm{OPD}_{\mathrm{B}}^*=a_0+a_1z+a_2z^2+a_3z^3+\ldots,$$

the equation becomes

$$\int_0^1 z^j(a_0+a_1z+a_2z^2+\ldots)\,dz=0, \qquad \ldots\ldots(10)$$

which involves only simple integration. Remember here that condition (7) has to be involved in the calculation,

$$\int_0^1\mathrm{OPD}_{\mathrm{B}}^*\,dz=\int_0^1(a_0+a_1z+a_2z^2+\ldots)\,dz=0. \qquad \ldots\ldots(7')$$

This condition corresponds to $\partial R/\partial a_0=0$ by (9).

In the following sections we will consider some cases separately.

§ 3. PRIMARY ABERRATION

For systems where the OPD values are well represented by expressions in powers of y up to the fourth order only, while the higher-order terms are insensible throughout the aperture, we can express

$$\mathrm{OPD}_{\mathrm{B}}^*=a_0+a_1z+a_2z^2.$$

This is the case when only the primary aberration is within the scope of consideration. Assume that the primary aberration cannot be controlled. Here the only controllable parameter is the a_1, corresponding to a shift of focus from the paraxial. We require, according to equations (10) and (7′),

$$\int_0^1 (a_0 + a_1 z + a_2 z^2)\, dz = 0,$$
$$\int_0^1 z(a_0 + a_1 z + a_2 z^2)\, dz = 0.$$

Integrate, and solve for a_0, a_1 in terms of a_2:

$$\left.\begin{aligned} a_1 = -a_2, \quad a_0 = +\tfrac{1}{6}a_2, \\ \mathrm{OPD}_{\mathrm{B}}^{*} = \tfrac{1}{6}a_2 - a_2 z + a_2 z^2. \end{aligned}\right\} \qquad \ldots\ldots(11)$$

The shape of $\mathrm{OPD}_{\mathrm{B}}^{*}$ is illustrated in figure 3. The extreme $\mathrm{OPD}_{\mathrm{B}}^{*}$ difference is $a_2/4$. Using $\lambda_0/4$ as tolerance, we should have $|a_2/4| \leqslant \lambda_0/4$ or $|a_2| \leqslant \lambda_0$, indicating

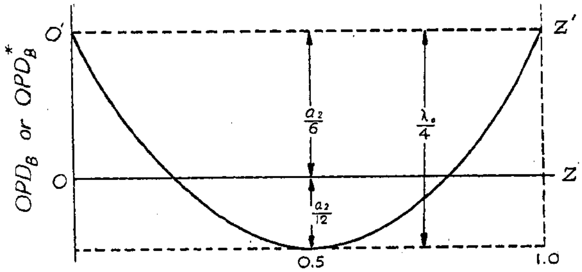

Figure 3.

that the maximum allowable spherical aberration in terms of OPD is a whole wave-length.

In fact, apart from the constant term, expression (11) has just the same form as that obtained by other authors based on the criterion of minimum phase difference (Conrady, 1923; Martin, 1930). Hence the same conclusions can be drawn, namely:

(1) The best physical focus lies midway between the extreme marginal and paraxial foci.

(2) The residual difference of optical paths is only one-quarter of the magnitude attained by them at the extreme marginal or paraxial focus.

(3) The paraxial part of $\mathrm{OPD}_{\mathrm{B}}^{*}$, namely, $a_2 z^2$, may be allowed to reach a whole wave-length without infringement of the Rayleigh limit, as has been proved above.

Whilst the present author reaches the same conclusion as the others did, the physical meanings are different. Based on the principle of minimum phase-difference, the " best " focus derived can only be admitted with reserve. It is the above formulation that gives the proof.

Let us now calculate the reduction of intensity when the full tolerance is allowed. Using expression (9),

$$R = \int_0^1 \left[1 - \tfrac{1}{2}\left(\frac{2\pi \mathrm{OPD}_{\mathrm{B}}^{*}}{\lambda_0}\right)^2\right] dz$$
$$= 1 - \tfrac{1}{2}\left(\frac{2\pi}{\lambda_0}\right)^2 a_2^2 \int_0^1 (\tfrac{1}{6} - z + z^2)^2\, dz = 1 - \tfrac{1}{2}\left(\frac{2\pi}{\lambda_0}\right)^2 \frac{a_2^2}{180}.$$

Put $a_2 = \lambda_0$. Hence

$$R = 1 - 0{\cdot}11. \qquad \ldots\ldots(12)$$

Thus the central maximum amplitude at the best focus has been reduced from the ideal value by 11 per cent. This accords well with the results obtained from strict numerical work (Conrady, 1919; Buxton, 1921).

§ 4. ZONAL ABERRATION

In an optical system of higher aperture, where the secondary aberration—the $a_3(y/y_0)^6$ term—comes out with sensible magnitude, but higher-order aberrations are still inconsiderable, and the primary aberration is under our control by the designer's usual means, we can well represent the aberration by

$$\mathrm{OPD}_{\mathrm{B}}^{*} = a_0 + a_1 z + a_2 z^2 + a_3 z^3.$$

Here, a_1 and a_2 are the controllable parameters. Using a treatment similar to that of § 3, we arrive at a solution giving a_0, a_1, a_2 in terms of a_3. The results are:

$$\left.\begin{aligned} a_0 &= -\frac{1}{20}a_3, \quad a_1 = \frac{3}{5}a_3, \quad a_2 = -\frac{3}{2}a_3, \\ \mathrm{OPD}_{\mathrm{B}}^{*} &= a_3\left(-\frac{1}{20} + \frac{3}{5}z - \frac{3}{2}z^2 + z^3\right). \end{aligned}\right\} \qquad \ldots\ldots(13)$$

The shape of the $\mathrm{OPD}_{\mathrm{B}}^{*}$ is given in figure 4.

The value of R can be calculated in a similar manner:

$$\left.\begin{aligned} R &= 1 - \tfrac{1}{2}a_3^2\left(\frac{2\pi}{\lambda_0}\right)^2 \int_0^1 \left(-\frac{1}{20} + \frac{3}{5}z - \frac{3}{2}z^2 + z^3\right)^2 dz \\ &= 1 - \tfrac{1}{2}a_3^2\left(\frac{2\pi}{\lambda_0}\right)^2 \frac{1}{2800}. \end{aligned}\right\} \qquad \ldots\ldots(14)$$

We noticed in the previous section that the reduction of amplitude is 11 per cent

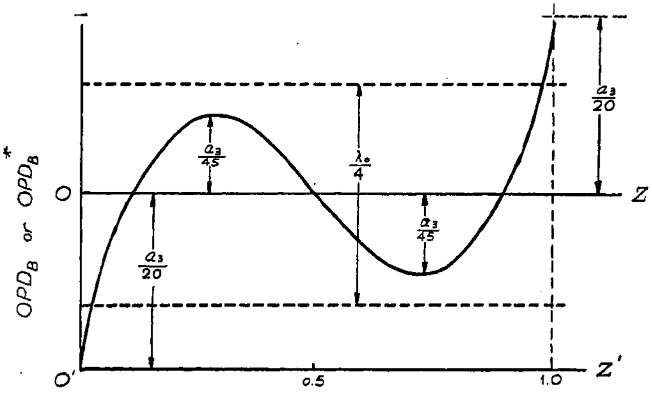

Figure 4.

when the full Rayleigh limit is allowed. In the present case, we may treat the problem in the reverse way. Set the value 11 per cent as the margin for amplitude reduction, below which the image would no longer be considered as perfect.

We can find the allowable values of $\text{OPD}_{\text{B}}^{*}$. From expression (14), to have $R \geqslant 1\text{–}0{\cdot}11$ requires $|a_3| \leqslant 4\lambda_0$. In figure 4 the dotted horizontal lines indicate the Rayleigh limit when $|a_3| = 4\lambda_0$, which is the maximum allowable value of $|a_3|$. Note that there are zones near the axis and the extreme margin whose $\text{OPD}_{\text{B}}^{*}$ values exceed $\lambda_0/8$. But as these are only very narrow zones, we need not fear that the use of the simplified expression (9) will cause serious error.

Thus we come to the conclusion:—

(1) With a proper choice of primary spherical aberration and focal shift, the maximum allowable zonal term in $\text{OPD}_{\text{B}}^{*}$ can be as high as 4 wave-lengths.

Take the OPD_{p} value with respect to the paraxial focus:

$$\text{OPD}_{\text{p}} = a_3\left[-\frac{3}{2}\left(\frac{y}{y_0}\right)^4 + \left(\frac{y}{y_0}\right)^6\right],$$

$$\frac{d\text{OPD}}{dy} = -\frac{a_3}{y_0}\left[6\left(\frac{y}{y_0}\right)^3 - 6\left(\frac{y}{y_0}\right)^5\right].$$

This derivative vanishes at $y=0$ and $y=y_0$. Since $\dfrac{1}{N}\dfrac{d\text{OPD}_{\text{p}}}{dy}$ indicates the angular aberration with respect to the paraxial focus, this means that the marginal ray would pass through the paraxial focus. Hence we come to the second conclusion:

(2) In order to get the best axial focus, the extreme marginal ray should be brought to the paraxial focus.

In the expression (13) the existence of the y^2 term means that the best focus is not at the marginal-paraxial focus, but at a focal shift given very nearly by

$$\delta l = -\frac{6}{5}a_3\frac{l^2}{Ny_0^{\,2}},$$

by equation (2). Furthermore, put the aberration at various zones in longitudinal measure:

$$\text{B}_{\text{p}}\text{B} = \frac{l^2}{N}\frac{d\text{OPD}_{\text{p}}}{y\,dy} = \frac{2l^2}{Ny_0^{\,2}}\frac{d\text{OPD}_{\text{p}}}{dz}$$

$$= -a_3\frac{l^2}{Ny_0^{\,2}}(6z - 6z^2).$$

$\text{B}_{\text{p}}\text{B}$ is maximum at $z=0{\cdot}5$ or $y=0{\cdot}707$. Call the most remote zonal focus B_{z} and we find

$$\text{BB}_{\text{z}} = -\frac{3}{2}a_3\frac{l^2}{Ny_0^{\,2}} = -\frac{3}{2}\frac{a_3}{N\sin^2 U_0},$$

$$\frac{\delta l}{\text{B}_{\text{p}}\text{B}_{\text{z}}} = \frac{4}{5},$$

where U_0 is the intersecting angle at full aperture. Thus we come to the third conclusion:

(3) The maximum longitudinal zonal aberration is contributed by the 0·707 zonal ray. The best focus is at 4/5 of the distance from the marginal-paraxial focus to the most remote zonal focus.

Comparing equation (13) with those derived by the same kind of treatment, but based on the principle of minimum phase difference, we find, after a slight transposition to give better accordance with the notations used here, that the expression for $\mathrm{OPD_B}$ is given elsewhere (Conrady, 1923; Martin, 1926) by

$$\mathrm{OPD_B}=\frac{9}{16}a_3\left(\frac{y}{y_0}\right)^2-\frac{3}{2}a_3\left(\frac{y}{y_0}\right)^4+a_3\left(\frac{y}{y_0}\right)^6.$$

Apart from the constant term, there is a slight difference in the $a_1(y/y_0)^2$ term, which corresponds to a slight shift to the so-called best focus. The other conclusions reached are the same.

§ 5. TERTIARY ABERRATION

In systems like high-power microscope objectives, the apertures are so big that aberrations of y^8 or even higher orders have to be taken into consideration. The zonal aberration is still controllable in a good design. Thus our controllable

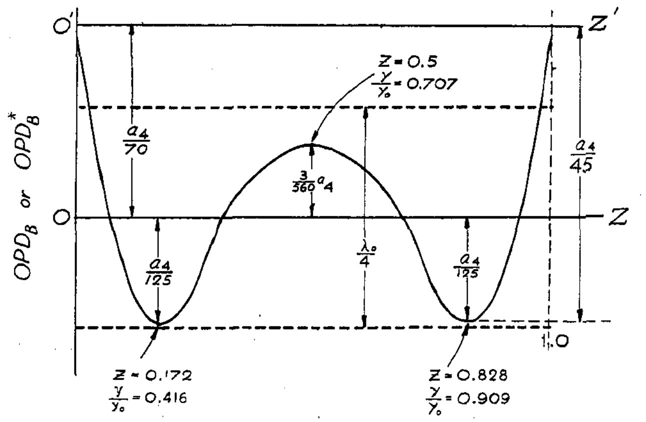

Figure 5.

parameters are a_1, a_2, a_3. At present, as an approximation only, we will assume that only the y^8 term is of sensible magnitude among the higher-order terms. With this restriction, we can obtain a_0, a_1, a_2, a_3 by the same type of treatment as before:

$$a_0=\frac{a_4}{70},\quad a_1=-\frac{2}{7}a_4,\quad a_2=\frac{9}{7}a_4,\quad a_3=-2a_4,$$

from which we come to the conclusion:

(1) For best correction, we should arrange the optical system so that the final $\mathrm{OPD_B^*}$ comes out as

$$\mathrm{OPD_B^*}=a_4\left[\frac{1}{70}-\frac{2}{7}\left(\frac{y}{y_0}\right)^2+\frac{9}{7}\left(\frac{y}{y_0}\right)^4-2\left(\frac{y}{y_0}\right)^6+\left(\frac{y}{y_0}\right)^8\right],\quad \ldots\ldots(15)$$

or at the paraxial focus,

$$\mathrm{OPD_P}=a_4\left[\frac{9}{7}\left(\frac{y}{y_0}\right)^4-2\left(\frac{y}{y_0}\right)^6+\left(\frac{y}{y_0}\right)^8\right]. \qquad \ldots\ldots(16)$$

The shape of the $\mathrm{OPD_B^*}$ curve is shown in figure 5. In a similar way we find

$$R=1-\tfrac{1}{2}a_4{}^2\left(\frac{2\pi}{\lambda_0}\right)^2\left[\frac{1}{44100}\right], \qquad \ldots\ldots(17)$$

assuming that the OPD residues do not greatly exceed the Rayleigh limit. To find the maximum allowable tolerance, if we still use the 11 per cent amplitude reduction limit, for $R\geqslant 1-0{\cdot}11$, $|a_4|\leqslant 15\lambda_0$. The corresponding Rayleigh limit is shown by the two horizontal dotted lines for a_4 at full tolerance. Again we note there are two very narrow zones near the axis and the extreme margin whose $\mathrm{OPD_B^*}$ values exceed $\frac{1}{8}\lambda_0$, but they would not cause serious error on using the simplified expression (9). Hence we come to the second conclusion:

(2) In a well-designed optical system, where the axial spherical aberrations are controllable up to the second order, the tertiary aberration, when put in terms of $\mathrm{OPD_B}$, should not exceed $15\lambda_0$.

In a well-known treatment by Prof. A. E. Conrady (1923), the tolerance for tertiary aberration is given a limiting value of $8\frac{1}{2}\lambda_0$ as the maximum which

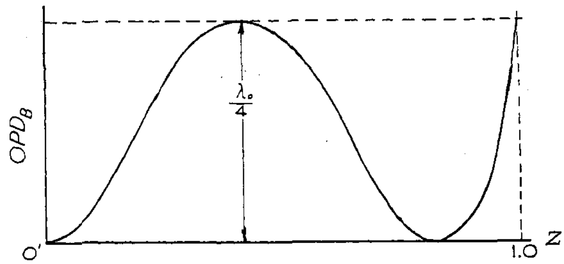

Figure 6.

can be allowed (under stated conditions) without transgression of the Rayleigh quarter-wave-length limit. In this case the OPD curve appears as shown in figure 6. His treatment is based on the idea of minimum optical path-difference. The present treatment shows that the tolerance may be practically doubled. Although we must acknowledge that the assumption of the presence of tertiary aberration with negligible higher-order terms is only a rough approximation, which, from the practical point of view, we must not regard as absolutely trustworthy, a realization of the true maximum permissible tolerance would nevertheless be a good guide.

In the scheme of design suggested by Professor Conrady the control of the primary and secondary aberration is obtained by bringing the paraxial $\frac{10}{11}$ zonal, and $\frac{10}{11}\frac{1}{\sqrt{2}}$ zonal foci to the same point. In the present scheme, in order to allow a greater tolerance, the earlier method has to be modified to some extent.

In equation (16), to find the geometrical ray intersections of different zones, we differentiate it:

$$\frac{d\mathrm{OPD}_{\mathrm{B}}^{*}}{dy}=4\frac{a_4}{y_0}\left[-\frac{1}{7}\left(\frac{y}{y_0}\right)+\frac{9}{7}\left(\frac{y}{y_0}\right)^3-3\left(\frac{y}{y_0}\right)^5+2\left(\frac{y}{y_0}\right)^7\right].$$

Firstly, $\frac{d\mathrm{OPD}_{\mathrm{B}}}{dy}$ vanishes at $y=0{\cdot}416$, $y=0{\cdot}707$, $y=0{\cdot}909$ zones. Therefore, if we make ray tracings through these zones, these three rays will meet at the same focus, and this is the best focus. The corresponding $\mathrm{OPD}_{\mathrm{B}}$ values are $\mathrm{OPD}_{\mathrm{B}_{0\cdot416}}=\mathrm{OPD}_{\mathrm{B}_{0\cdot909}}=-\frac{a_4}{45}$, $\mathrm{OPD}_{\mathrm{B}_{0\cdot707}}=-\frac{a_4}{112}$. When $|a_4|=15\lambda_0$, $\mathrm{OPD}_{\mathrm{B}_{0\cdot416}}=\mathrm{OPD}_{\mathrm{B}_{0\cdot909}}=\lambda_0/3$. Secondly, at the axis and at the extreme margin,

$$\left(\frac{d\mathrm{OPD}_{\mathrm{B}}^{*}}{y\,dy}\right)_{y=0}=-\frac{4}{7}\frac{a_4}{y_0^{2}},\quad \left(\frac{d\mathrm{OPD}_{\mathrm{B}}^{*}}{y\,dy}\right)_{y=y_0}=\frac{4}{7}\frac{a_4}{y_0^{2}}.$$

Denote the marginal focus by $\mathrm{B_M}$ and the paraxial by $\mathrm{B_p}$. The distances $\mathrm{B_M B}$ and B B are given by the expressions

$$\mathrm{B_M B}=\frac{l^2}{N}\left(\frac{d\mathrm{OPD}_{\mathrm{B}}^{*}}{y\,dy}\right)_{y=y_0}=+\frac{4}{7}\frac{a_4}{N}\frac{l^2}{y_0^{2}},$$

$$\mathrm{B_p B}=\frac{l^2}{N}\left(\frac{d\mathrm{OPD}^{*}}{y\,dy}\right)_{y=y_0}=-\frac{4}{7}\frac{a_4}{N}\frac{l^2}{y_0^{2}}.$$

Thus the marginal and the paraxial foci lie at equal and opposite distances from the best focus. Denote the longitudinal spherical aberration by LA. LA $=$ B $\mathrm{B_M}$. By allowing $|a_4|$ as $15\lambda_0$, the maximum permissible axial spherical-aberration LA becomes

$$\text{Maximum allowable }|\mathrm{LA}|=\frac{8}{7}\frac{15\lambda_0}{N\sin^2 U_0}\simeq\frac{16\lambda_0}{N\sin^2 U_0},\quad \ldots\ldots(18)$$

which is 16 times the axial focal range of the system. Hence we have two more conditions:

(3) The best focus lies midway between the paraxial and the marginal foci. It is the point where the 0·416, 0·707, and 0·909 zonal rays meet.

(4) The maximum allowable longitudinal spherical-aberration is 16 times the focal range of the system.

In practical designing work, for the final achievement to be in accordance with expression (15), we may either:

(*a*) bring the 0·416, 0·707, and 0·909 zonal rays to the same focus, the tolerance being guarded against by comparing the value of LA obtained with expression (18), or

(*b*) bring the 0·416 and 0·909 zonal rays to the same focus, and in addition to that, $\mathrm{OPD}_{\mathrm{B}_{0\cdot416}}=\mathrm{OPD}_{\mathrm{B}_{0\cdot909}}$, the performance of the system then being checked by observing $|\mathrm{OPD}_{\mathrm{B}_{0\cdot416}}|=|\mathrm{OPD}_{\mathrm{B}_{0\cdot902}}|\leqslant\lambda/3$.

It should be remembered that in a system where tertiary aberration plays a sensible part, aberrations of still higher orders are likely to be present and may even become very big at the extreme margin. Referring to figure 5, the OPD_B^* curve usually tends to become steeper at the extreme marginal zone than we have here assumed. This will cause the marginal focus to shift farther away from the best focus, thus giving a greater LA. Therefore the restriction of LA below a certain limit is naturally a safeguard against higher-order aberrations. In the case of method (*b*) being used, as just suggested, it is advisable still to trace the marginal and paraxial rays to see what the value of LA looks like.

It may also be noted that the so-called focal shift as a controllable factor occurs only in the theory, but not in the practical design. Once the optical system is properly designed, the judgement of the best focus is left to the observer.

§ 6. ACKNOWLEDGEMENTS

In conclusion, the author wishes to express his thanks to Professor L. C. Martin, who carefully read the manuscript and pointed out several corrections which should be made. He has also the pleasure of thanking the authorities of the Imperial College of Science and Technology for allowing him to publish this note.

REFERENCES

BUXTON, A., 1921. *Mon. Not. R. Astr. Soc.* **81**, 547.
CONRADY, A. E., 1919. *Mon. Not. R. Astr. Soc.* **79**, 575.
CONRADY, A. E., 1923. Glazebrook's *Dictionary of Applied Physics*, **4**, 215 (under the title *Microscope, Optics of the*).
CONRADY, A. E., 1926. *Proc. Opt. Conv.* Part 2, 830.
MARTIN, L. C., 1926. *Trans. Opt. Soc.* **27**, 249.
MARTIN, L. C., 1930. *An Introduction to Applied Optics*, **1**, 121–6 (Pitman and Co.).