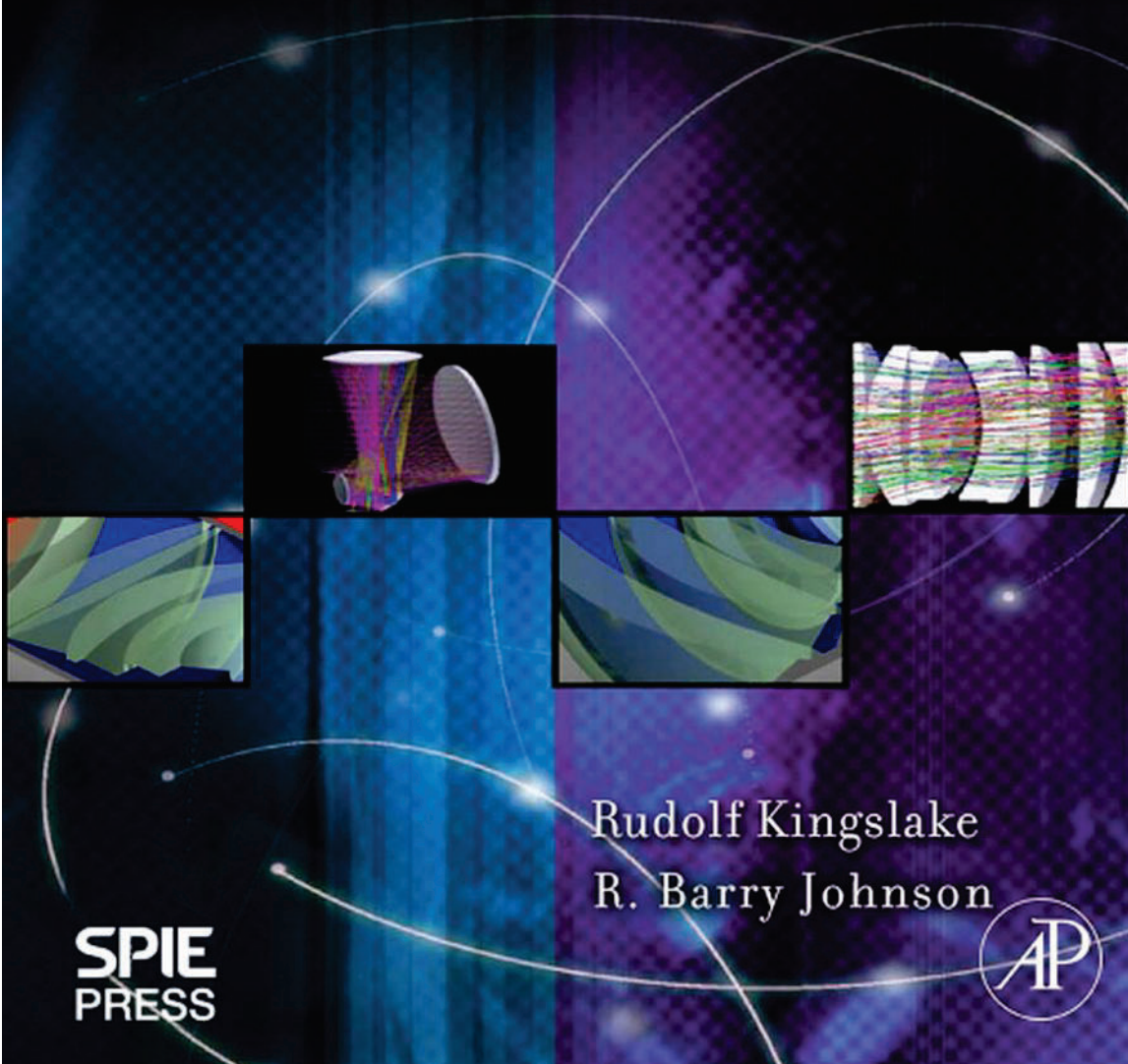


Lens Design Fundamentals

SECOND EDITION



Rudolf Kingslake
R. Barry Johnson

SPIE
PRESS



Lens Design Fundamentals

Second Edition

RUDOLF KINGSLAKE
R. BARRY JOHNSON



AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

SPIE
PRESS

Academic Press is an imprint of Elsevier
30 Corporate Drive, Suite 400
Burlington, MA 01803, USA
The Boulevard, Langford Lane
Kidlington, Oxford, OX5 1 GB, UK

© 2010 Elsevier Inc. All rights reserved.

Co-published by SPIE
P.O. Box 10
Bellingham, Washington 98227-0010 USA
Tel.: +1 360-676-3290 / Fax: +1 360-647-1445
Email: Books@SPIE.org
SPIE ISBN: 9780819479396
SPIE Vol: PM195

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

Application submitted.

ISBN: 978-0-12-374301-5

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

For information on all Academic Press publications
visit our Web site at www.elsevierdirect.com

Printed in the United States

09 10 11 12 13 10 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER

BOOK AID
International

Sabre Foundation

Chapter 8

Oblique Beams

An oblique beam (also called a pencil) of rays from an extraaxial (or off-axis or nonaxial) object point contains meridional rays that can be traced by the ordinary computing procedures already described, and also a large number of skew rays that do not lie in the meridional plane. Each skew ray intersects the meridional plane at the object point and again at a “diapoint” in the image space, and nowhere else. Skew rays require special ray-tracing procedures, which will be discussed in Section 8.3. These are much more complex than for a meridional ray, and it is observed that skew rays were seldom used before the advent of electronic computers; now they are routinely traced by all lens designers since the available computing power of even the most common personal computer is extraordinarily great.

In Chapter 4, we discussed both axial and off-axis/nonaxial aberrations in an analytical sense rather than a causal sense. Axial aberrations have been investigated in some detail in the last several chapters. In this chapter, we will begin a more detailed study of field-dependent astigmatic and comatic aberrations. We begin by looking at the origin of coma and astigmatism, and then the role various types of stops have in lens systems. The remainder of the chapter discusses general ray tracing and graphical representation of skew ray aberrations.

8.1 PASSAGE OF AN OBLIQUE BEAM THROUGH A SPHERICAL SURFACE

8.1.1 Coma and Astigmatism

When a light beam is refracted obliquely through a spherical surface, several new aberrations arise that do not appear on the lens axis. To understand why this is so, we may consider the diagram in Figure 8.1 showing a single refracting surface and an aperture stop that admits a circular cone of rays from an off-axis object point B . We label the rays through the rim of the aperture by their position angles taken counterclockwise from the top as viewed from the image

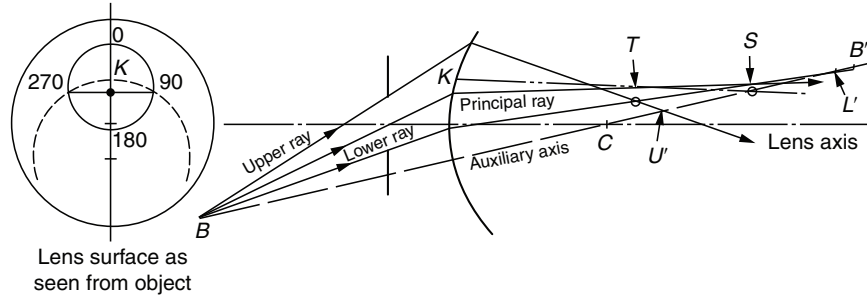


Figure 8.1 Origin of coma and astigmatism.

space, so that the upper ray is called 0° and the lower ray 180° , while the front and rear sagittal rays become 90° and 270° , respectively. The line joining the object point B to the center of curvature of the surface, C , is called the *auxiliary axis*, and obviously there is complete rotational symmetry around this axis just as there is rotational symmetry around the lens axis for an axial object point.

Moreover, because of this symmetry, every ray from the object point B passing through the aperture stop must cross the auxiliary axis somewhere in the image space. If we could trace a paraxial ray from B along the auxiliary axis, it would form an image of B at, say, B' . However, because of the spherical aberration arising at the surface, the intersection point for all other rays will move along the auxiliary axis toward the surface by an amount proportional to the square (approximately) of the height of incidence of the ray above the auxiliary axis. Thus the upper limiting ray might cross the auxiliary axis at, say, U' , and the lower limiting ray at L' . It is at once evident that the upper and lower rays do not intersect each other on the principal ray but in general above or below it; the height of the intersection point above or below the principal ray is called the *tangential coma* (a relic of the old custom of calling meridional rays tangential because they form a tangential focal line).

To find the point at which the two sagittal rays at 90° and 270° intersect the auxiliary axis, we note that these rays are members of a hollow cone of rays centered about the auxiliary axis, all coming to the same focus on that axis. The upper ray of this hollow cone strikes the refracting surface at K , slightly higher than the principal ray, so that the spherical aberration of this ray will be a little greater than that of the principal ray, forming an image at S on the auxiliary axis (shown by the small circle). S lies below the principal ray on our diagram, which indicates the presence of some negative sagittal coma, but not as much as the tangential coma that we found previously. Indeed, it can be shown¹ that for a very small aperture and obliquity, the tangential coma is three times the sagittal coma; the exaggerations in our diagram do not make this relation obvious, but at least both comas do have the same sign.

We thus see that the extreme upper and lower rays of the marginal zone come to a focus at T , while the extreme front and rear rays come to a different focus at S . The longitudinal separation between S and T is called the astigmatism of the image, and evidently both coma and astigmatism arise whenever a light beam is refracted obliquely at a surface. It is essential to note that each surface in a lens has a different auxiliary axis, and that the proportion of coma and astigmatism therefore varies from surface to surface. It is thus possible to correct coma and astigmatism independently in a lens system provided there are sufficient degrees of freedom available.

In Sections 4.3.3 and 4.3.4, additional information about computing astigmatism and coma using exact ray tracing and the relationship to aberration coefficients was presented. In the next chapter, we will discuss coma, the Abbe sine condition, and offense against the sine condition. Astigmatism, Coddington equations, the Petzval theorem, distortion, and lateral color are explored in more depth in Chapter 11. Also in that chapter, the important symmetry principle will be introduced.

8.1.2 Principal Ray, Stops, and Pupils

At this point, it is necessary to define several important terms. The *aperture stop* or *stop* of a lens is the limiting aperture associated with the lens that determines how large an axial beam may pass through the lens. The stop can be an element within the lens system or a mechanical element such as a hole in a disk. A mechanical stop that can vary its opening size is also called an *iris*.

The *marginal ray*, also called the *rim ray*, is the extreme ray from the axial point of the object through the edge of the stop. As discussed in Section 4.2, the *entrance pupil* is the image of the stop formed by all lenses preceding it when viewed from object space. It also is the reference surface used to define ray coordinates, that is, (ρ, θ, H_y) . By convention, the entrance pupil is aberration free. In a similar manner, the *exit pupil* is the image of the stop formed by all lenses following it when viewed from image space. The exit pupil is used as a reference surface for exiting wavefronts from the lens. Very often the rays incident at the exit pupil are not rectilinearly mapped onto the exit pupil due to pupil aberrations.

Consideration of the mapping error is necessary to properly compute image energy distribution, *MTF*, and diffraction. These two pupils and the stop are all geometric images of one another. The entrance and exit pupils can each be real or virtual images of the aperture stop located at finite distances or at infinity dependent on the optical configuration before and after the stop. For example, if an aperture stop is placed between the object and a singlet lens, and closer to the lens than the focal length, then the entrance pupil is clearly real and the exit

pupil is virtual. In general, the spatial location of the pupils with respect to the stop can be in order (for example [exit, entrance, stop], [entrance, exit, stop], and [entrance, stop, exit]).

The *principal ray* is defined as the ray emanating from an off-axis object point that passes through the center of the stop. In the absence of pupil aberrations, the principal ray also passes through the center of the entrance and exit pupils. As the obliquity angle of the principal ray increases, the defining apertures of the components comprising the lens may limit the passage of some of the rays in the entering beam, thereby causing the stop not to be filled with rays. The failure of an off-axis beam to fill the aperture stop is called *vignetting*.² The ray centered between the upper and lower rays defining the oblique beam is called the *chief ray*. When the object moves to large off-axis locations, the entrance pupil often has a highly distorted shape, may be tilted, and/or displaced longitudinally and transversely, and no longer perpendicular to the lens axis.

Indeed, without this tilting of the entrance pupil a fisheye lens covering a full $\pm 90^\circ$ in the object space would not transmit any light at the edge of the field. Due to the vignetting and pupil aberrations, the chief and principal rays may become displaced from one another. In some cases, the principal ray is vignetted while the chief ray is never vignetted as long as light passes through the lens at the considered obliquity angle. The terms principal ray and chief ray are frequently used interchangeably; however, once vignetting occurs, the distinction must be made.

DESIGNER NOTE

It is important that the lens designer understands how the optical design program being used handles the aiming of the chief ray. Typically, the chief ray is aimed toward the center of the (vignetted) entrance pupil, which is generally acceptable in the early stages of design. In the final stages, the chief ray should be aimed at the center of the (vignetted) stop. The reason for this is that additional computational time is required to aim at the (vignetted) stop. Since the stop is a real surface, the entrance pupil may well suffer aberrations. If the entrance pupil is considered unaberrated, then the stop is likely aberrated in theory at least. A design that may appear quite satisfactory using an unaberrated entrance pupil can perform in practice differently since the actual stop is unaberrated, thereby changing what rays actually pass through the lens system! Remember that the vignetted stop is made up of portions of the actual stop and boundaries of various lens elements (see Section 8.1.3).

The *field stop* is an aperture that limits the passage of principal rays beyond a certain field angle. The image of the field stop when viewed from object space is called the *entrance window* and is called the *exit window* when viewed from image space. The field stop effectively controls the field of view of the lens

system. Should the field stop be coincident with an image formed within or by the lens system, the entrance and exit windows will be located at the object and/or image(s).

A *telecentric stop* is an aperture located such that the entrance and/or exit pupils are located at infinity (see Section 12.5.3). This is accomplished by placing the aperture in the focal plane. Consider a stop placed at the front focal plane of a lens. The stop image is located at infinity and the principal ray exits the lens parallel to the optical axis. This feature is often used in metrology since the measurement error is reduced when compared to conventional lens systems because the centroid of the blur remains at the same height from the optical axis even as the focus is varied.

8.1.3 Vignetting

In many lenses, and particularly those having a considerable axial length, an oblique pencil may be unable to traverse the lens without part of the beam being obstructed by the end lens apertures. For instance, in the triplet lens shown in Figure 8.2 the upper rays of the 20° oblique beam are cut off by the rear lens aperture, and the lower rays by the front aperture, so that the beam fails to fill the iris. This process is known as *vignetting*, the oblique beam is projected onto the plane perpendicular to the axis in the object space having the shape shown in the figure. Vignetting is one of the reasons why the illumination on the film/detector-array in a camera falls off at increasing transverse distances from the lens axis. Other reasons are (a) the $\cos^3$ law, (b) distortion of the entrance pupil at high obliquities, and (c) image distortion. The effects of these various factors have been discussed elsewhere.⁴

To plot the vignetting diagram of a lens, the locations of the upper and lower “rim” rays are readily found by trial, but it is then necessary to determine the radii of the upper and lower limiting circular arcs. The lower arc obviously has the same radius as the front lens aperture, but the upper arc is the image of the rear aperture as seen through the lens. Its radius bears the same ratio to the radius of the entering axial beam as the diameter of the rear aperture itself bears to the diameter of the axial beam as it emerges from the rear of the lens.

In addition to the circles corresponding to the front and rear lens apertures, an oblique beam is limited also by the iris, and the image of the iris must therefore be projected into the object space along with the image of the rear aperture. To locate this iris image, we add a ray parallel to the upper and lower rim rays and passing through the center of the iris. This middle ray is projected into the vignetting diagram in Figure 8.2, and we draw a circle about it having the diameter of the entering axial beam because the axial beam necessarily fills the iris completely. The vignetted area of the oblique beam is shown shaded. The

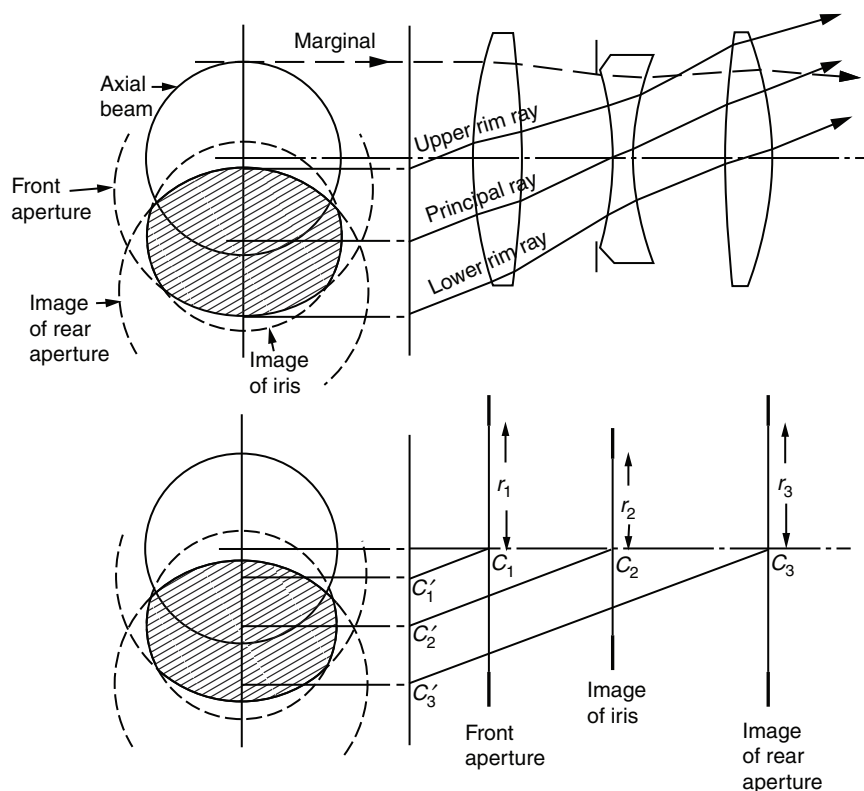


Figure 8.2 Vignetting diagrams.

“vignetting factor” is the ratio of the area of the oblique beam to the area of the axial beam, both measured in a plane perpendicular to the lens axis. It is, of course, an assumption that the images of the iris and of the rear lens aperture are circles; indeed, they are much more likely to be arcs of ellipses, but we make very little error by plotting them as circles.

An alternative method of plotting the vignetting diagram is shown in the lower diagram of Figure 8.2. We begin by determining the location and size of the images of the rear aperture and of the iris, projected into the object space, by use of paraxial rays traced right-to-left from the centers of those two apertures. The front aperture and the two images are shown at C_1 , C_2 , and C_3 , their computed radii being, respectively, r_1 , r_2 , and r_3 . We can now replace the lens by these three circles, and project their centers at any required obliquity onto a vertical reference plane as shown. Knowing the centers of the circles and their radii, it is a simple matter to draw the vignetting diagram directly. Of course, this procedure cannot be as

accurate as the first method, but it is much simpler and generally sufficiently accurate for most purposes. This simple procedure cannot be used for wide-angle or fisheye lenses where the pupil is seriously distorted or tilted.

Another method that can be used relies on the linear nature of paraxial ray tracing. It is easy to show that if two paraxial rays have been traced through a lens system and at each surface y and $\bar{y}$ are known, then the intercept height $\bar{y}$ for any other ray at any surface can be computed without ray tracing.⁴ The general equation is

$$\bar{y}_j - \bar{y}_i = (\bar{y}_i - \bar{y}_1) \frac{y_j}{y_i}. \quad (8-1)$$

Calculation of the coordinate of any ray on the entrance pupil ($j = 1$) having the coordinate $\bar{y}_i$ on the i th surface, with $\bar{y}_1 = 0$, then Eq. (8-1) becomes

$$\bar{y}_1 = (\bar{y}_i - \bar{y}_1) \frac{y_1}{y_i}. \quad (8-2)$$

Table 8.1 contains ray trace data for a simple two-lens system having an internal stop. Lens A diameter is 3.0, the diameter of Lens B is 2.0, and the diameter of the stop is 2.0. From Table 8.1, the entrance pupil is located a distance of 2.5 behind Lens A (or 0.5 to the right of the stop). The marginal ray and two principal rays ($u = 0.1$ and 0.2) were traced. The size of the entrance pupil can be determined in a couple of ways from the data in the table. First, the marginal ray has a height of 1.25 for a stop radius of 1.0. Remember that the linear nature of paraxial ray tracing implies that a stop diameter of 0.80, or the magnification of the stop to form the entrance pupil, is 1.25. A second way is to observe that the principal angle at the entrance pupil is 0.1 and is 0.125 at the stop. Hence, the magnification of the stop to form the entrance

Table 8.1
Ray Trace Data for Lens Systems Demonstrating Vignetting

	Entrance pupil	Lens A		Stop		Lens B	
Surf #	1	2		3		4	
$-\phi$	0	-0.1		0		-0.1	
t	-2.5		2		1		
y } marginal	1.25	1.25		1		0.875	
u } principal	0	-0.2500	-0.125	0	-0.125	0.125	-0.2125
$\bar{y}$ } principal	0	0.1	0.125	0	0.125		0.1125
$\bar{u}$ } principal	0	-0.5000		0		0.2500	
	0.2	0.25		0.25			0.2250

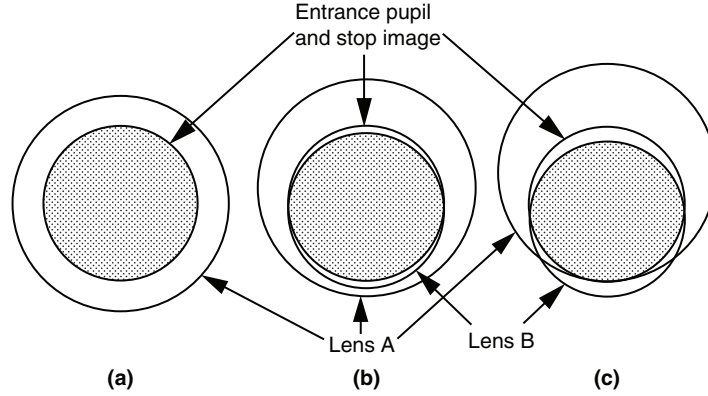


Figure 8.3 Vignetting diagram for the lens system shown in Table 8.1 where (a) is for an axial object, (b) is for a distant object having a field angle of 0.1, and (c) is for a distant object having a field angle of 0.2.

pupil is again found to be $1.25 (u_3/u_1)$. To compute the shift, often called shear, of the i th element's projected center onto the entrance pupil, set $\bar{y}_i = 0$ and use Eq. (8-2) with the data in the table. Determination of the projected size of the lens onto the entrance pupil once again uses Eq. (8-2).

Figure 8.3 depicts the circular apertures for the stop and both lenses contained in Table 8.1. where Figure 8.3a is for an axial object, Figure 8.3b is for a distant object having field angle of 0.1, and Figure 8.3c is for a distant object having a field angle of 0.2. The shaded area illustrates the portion of the lens that can pass light at these three angles. Notice in Figure 8.3b that Lens B is starting to vignette while Lens A is relatively far from vignetting. In Figure 8.3c, Lens B is vignetting more and Lens A is just starting to vignette. The vignetted entrance pupil appears to shift down and become elliptically shaped.

8.2 TRACING OBLIQUE MERIDIONAL RAYS

For any given object point, or for any given obliquity angle if the object is at infinity, a specific meridional ray must be defined by some convenient ray parameter. This may be the height A at which the ray intersects the tangent plane at the first lens vertex, or it may be the intersection length L of the ray relative to the front lens surface. For a ray proceeding uphill from left to right and entering the lens above the axis, A will be positive and L negative.

Whatever ray parameter is chosen, it is necessary to use appropriate “opening equations” to convert the given ray data into the familiar (Q, U) values to trace the ray.

1. *A Finite Object*

If the object point is defined by its H and d_0 (Figure 8.4), then

$$\tan U = -(A - H)/d_0 \quad \text{and} \quad Q = A \cos U$$

If the ray is defined by its L value, then

$$\tan U = -H/(L - d_0) \quad \text{and} \quad Q = -L \sin U$$

2. *A Very Distant Object*

The slope angle of all entering rays is now the same, being equal to the principal-ray slope U_{pr} ; we use only the second of the opening equations to find Q .

3. *Closing Equations*

Having traced an oblique ray through a lens, we generally wish to know the height at which it crosses the paraxial image plane. This is given by (Figure 8.5)

$$H' = (Q' + l' \sin U') / \cos U'$$

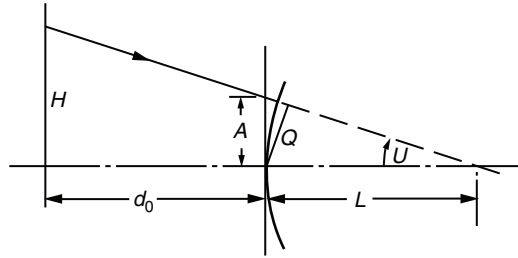


Figure 8.4 Opening equations.

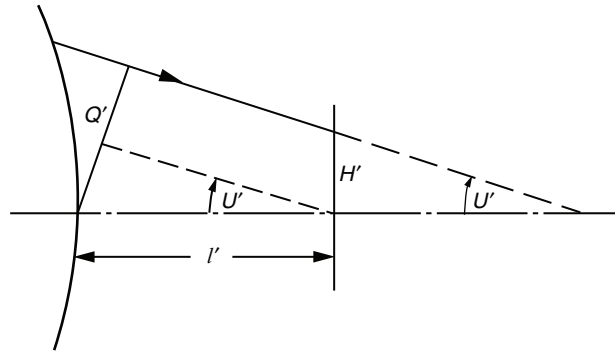


Figure 8.5 Closing equations.

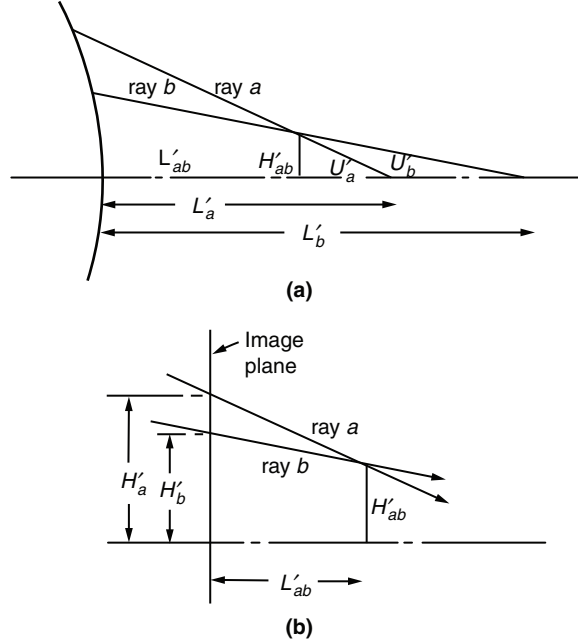


Figure 8.6 Coordinates of the intersection of two rays.

Sometimes we want to know the coordinates of the intersection point of two traced rays, knowing their L' or their Q' value and also their slope angles U' . The formulas to be used are (Figure 8.6a)

$$L'_{ab} = \frac{L'_a \tan U'_a - L'_b \tan U'_b}{\tan U'_a - \tan U'_b}, \quad \text{where } L' = -Q' / \sin U' \quad (8-3a)$$

$$H'_{ab} = -(L'_a - L'_{ab}) \tan U'_a = -(L'_b - L'_{ab}) \tan U'_b$$

8.2.1 The Meridional Ray Plot

Having traced a number of oblique rays through a lens from a given object point, we need some way to plot the results and interpret the mixture of aberrations that exists in the image. This mixture will contain spherical aberration, of course, and also the oblique aberrations coma and meridional field curvature. Astigmatism as such will not appear because it involves sagittal rays, which are not traced in a meridional beam. The two chromatic aberrations will not appear unless colored oblique rays are being traced.

The usual procedure is to plot the intercept height H' of the ray at the paraxial image plane as the ordinate, with some reasonable ray parameter as the abscissa.

For the latter we may use the Q value of the ray at the front lens surface, or the incidence height A of the ray at the tangent plane to the front vertex, or the intersection length L of the ray at the first lens surface. Sometimes we use the height of the ray at the paraxial entrance pupil plane, or its height in the stop. However, there is good reason to use as abscissa the tangent of the ray slope angle U' in the image space. When this is done a perfect image point plots as a straight line, whose slope is a measure of the distance from the paraxial image plane to the oblique image point. The reason for this can be seen in Figure 8.6b, which shows two rays in an oblique pencil having heights H'_a and H'_b at the image plane and emerging slope angles U'_a and U'_b , respectively. The longitudinal distance L'_{ab} from the image plane to the intersection of these rays with one another is given by

$$H'_{ab} = H'_a + L'_{ab} \tan U'_a \text{ and } H'_{ab} = H'_b + L'_{ab} \tan U'_b$$

Eliminating H'_{ab} gives

$$L'_{ab} = \frac{H'_a - H'_b}{\tan U'_b - \tan U'_a} \quad (8-3b)$$

If the data of the two rays are plotted on a graph connecting H' with $\tan U'$, the slope of the line joining the two ray points will be a direct measure of L'_{ab} . Consequently if all the rays in the beam have the same L'_{ab} , their ray points will all lie on a straight line, with the lower rim ray at the left and the upper at the right. The principal ray will fall about midway between the two rim rays. A perfect lens with a flat field will plot as a horizontal straight line (Figure 8.7a). A perfect lens with an inward-curving field plots as a straight line sloping down from left to right (Figure 8.7b). Primary coma is represented by a parabolic graph, the ends being up in the case of positive coma (Figure 8.7c), and down for negative coma (Figure 8.7d). Primary spherical aberration is represented by a cubic

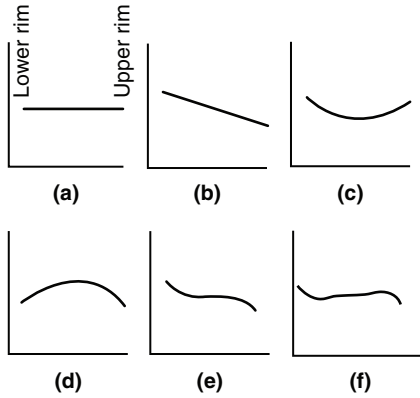


Figure 8.7 Some typical $H - \tan U$ curves: (a) a perfect lens, (b) inward-curving field, (c) positive coma, (d) negative coma, (e) spherical undercorrection, and (f) zonal spherical aberration.

curve, and if the image along the principal ray lies in the paraxial image plane, the middle of the cubic curve will be horizontal (Figure 8.7e). Zonal spherical aberration is revealed by a curve with a double bend, which is a combination of a cubic curve for the primary aberration component and a fifth-order curve for the secondary aberration (Figure 8.7f). Of course, any imaginable mixture of these aberrations can occur, and the experienced designer soon gets to recognize the presence of the different aberrations by the shape of the curve.

8.3 TRACING A SKEW RAY

A skew ray^{5,6} is one that starts out from an extraaxial object point and enters a lens in front of or behind the meridional plane. It should be noted that for every skew ray there is another skew ray that is an image of the first, formed as if the meridional plane were a plane mirror. Thus, having traced one skew ray we have really traced two, the ray in front of the meridional plane and the corresponding ray behind it. These two skew rays intersect each other at the same diapiant (see Figure 2.1).

In tracing a skew ray, we denote a known point on the ray as X_0, Y_0, Z_0 , and the direction cosines of the ray as K, L, M . Of course, in the object space the point X_0, Y_0, Z_0 can be the original object point, and we must somehow specify the direction cosines of the particular entering ray that we wish to trace. This is often done by specifying the point at which the entering ray pierces the tangent plane at the first lens vertex. Then, knowing X_0, Y_0, Z_0 , and K, L, M , we can determine the point X, Y, Z at which the ray strikes the following lens surface, and after refraction it will have a new set of direction cosines K', L', M' and proceed on its way. The ray-tracing problem thus reduces to two steps: the transfer of the ray from some known point to the next surface, and the refraction of the ray at the next surface.

8.3.1 Transfer Formulas

Since the direction cosines of a line are defined as the differences between the X, Y, Z coordinates of two points lying on the line divided by the distance between these points, it is clear from Figure 8.8 that

$$K = \frac{X - X_0}{D}, \quad L = \frac{Y - Y_0}{D}, \quad M = \frac{Z - Z_0 + d}{D}$$

where D is the distance along the ray from the point X_0, Y_0, Z_0 to the point of incidence X, Y, Z , and d is the axial separation of the surfaces. By means of these relationships we see that

$$X = KD + X_0, \quad Y = LD + Y_0, \quad Z = MD + (Z_0 - d) \quad (8-4a)$$

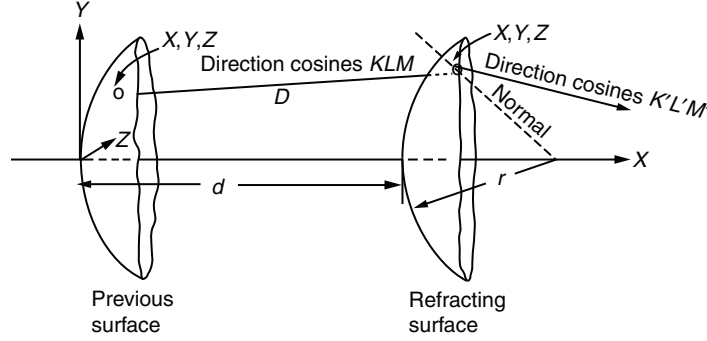


Figure 8.8 Transfer of a skew ray from one surface to the next.

The equation of the next refracting surface is, of course, known. For a sphere of radius r it is

$$X^2 + Y^2 + Z^2 - 2rX = 0 \quad (8-4b)$$

and substituting Eqs. (8-3a) in (8-3b) gives the equation to be solved for D as

$$D^2 - 2rF \cdot D + rG = 0$$

where

$$\begin{aligned} F &= M - \frac{KX_0 + LY_0 + M(Z_0 - d)}{r} \\ G &= \frac{X_0^2 + Y_0^2 + (Z_0 - d)^2}{r} - 2(Z_0 - d) \end{aligned} \quad (8-5)$$

The solution is, of course,

$$D = r \left[F \pm \left(F^2 - \frac{G}{r} \right)^{1/2} \right]$$

The ambiguous sign of the root indicates the two possible points of intersection of the ray with a complete sphere of radius r . Only one of these is useful, and the appropriate sign must be chosen. Remember that D must always be positive. Knowing D we return to Eq. (8-4a) and calculate X , Y , and Z , the coordinates of the point of incidence. For a plane surface,

$$D = G/2F = -(Z_0 - d)/M$$

8.3.2 The Angles of Incidence

It is a well-known property of direction cosines that the angle between two intersecting lines is given by

$$\cos I = Kk + Ll + Mm$$

Here K, L, M are the direction cosines of the ray, and k, l, m the direction cosines of the normal at the point of incidence. For a spherical surface,

$$k = -\frac{X}{r}, \quad l = -\frac{Y}{r}, \quad m = 1 - \frac{Z}{r} \quad (8-6)$$

Hence,

$$\begin{aligned} \cos I &= F - \frac{D}{r} \\ \cos I' &= [1 - (n/n')^2 (1 - \cos^2 I)]^{1/2} \end{aligned} \quad (8-7)$$

For a plane, $\cos I = K$.

8.3.3 Refraction Equations

To derive the refraction equations, we refer back to Figure 2.3, used in connection with the process of graphical ray tracing. It is reproduced and enhanced in Figure 8.9. In the vector triangle OAB , OA is a vector of magnitude n in the direction of the incident ray, OB is a vector of magnitude n' in the direction of the refracted ray, while AB is a vector of magnitude $n' \cos I' - n \cos I$ in the direction of the normal. Hence, we may construct the vector equation

$$n'\mathbf{R}' = n\mathbf{R} + (n' \cos I' - n \cos I)\mathbf{N}$$

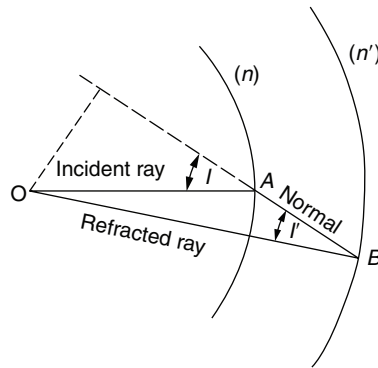


Figure 8.9 Refraction of a skew ray.

where $\mathbf{R}'$, $\mathbf{R}$, and $\mathbf{N}$ are unit vectors. Since the components of a unit vector are simply the direction cosines of the vector, we can resolve the vector equation into its three component equations:

$$\begin{aligned} n'K' &= nK + (n' \cos I' - n \cos I)k \\ n'L' &= nL + (n' \cos I' - n \cos I)l \\ n'M' &= nM + (n' \cos I' - n \cos I)m \end{aligned} \quad (8-8)$$

The direction cosines of the normal, k , l , and m , are given in Eq. (8-6). Hence Eq. (8-8) becomes

$$\begin{aligned} n'K' &= nK - JX \\ n'L' &= nL - JY \\ n'M' &= nM - J(Z - r) \end{aligned} \quad (8-9)$$

where $J = (n' \cos I' - n \cos I)/r$. As a check on our work, we can verify that $(K'^2 + L'^2 + M'^2) = 1$. For refraction at a plane surface these relations become

$$\begin{aligned} M &= \cos I, \quad M' = \cos I' \\ n'K' &= nK, \quad n'L' = nL, \quad J = 0 \end{aligned}$$

8.3.4 Transfer to the Next Surface

This has already been described. The direction cosines K' , L' , M' become the new K , L , M , and we calculate the new point of incidence by Eqs. (8-4), (8-5), (8-7), and (8-9), in order.

8.3.5 Opening Equations

1. Distant Object

Here we have a parallel beam incident on the lens inclined at an angle U_{pr} to the lens axis. Then

$$K = 0, \quad L = \sin U_{\text{pr}}, \quad M = \cos U_{\text{pr}}$$

The point of incidence of the particular skew ray must be determined in some way so that the X , Y , Z can be found. It is common to define the ray by its point of incidence with the tangent plane at the vertex of the first surface. If this is done, it is convenient to regard this tangent plane as the first lens surface with air on both sides of it, and use the general transfer equations to go from the tangent plane to the first refracting surface in the ordinary way.

2. Near Object

Here again we assume a tangent plane at the first lens surface, and we specify the point Y, Z at which the skew ray is to pierce that plane. The X_0, Y_0, Z_0 of the object point are, of course, known and also the distance d between the object and the front vertex. Then

$$K = \frac{X - X_0}{D}, \quad L = \frac{Y - Y_0}{D}, \quad M = \frac{d}{D}$$

where

$$D^2 = d^2 + (X - X_0)^2 + (Y - Y_0)^2.$$

8.3.6 Closing Equations

The closing equations for a skew ray are trivial, since the ray can be transferred to the final image plane by the ordinary transfer equations. This process gives the X', Y' coordinates of the intersection of the ray with the image plane directly. The d is, of course, nothing but the back focal distance from the rear vertex of the lens to the image plane.

8.3.7 Diapoint Location

For some purposes we may desire to determine the diapoint location of the skew ray. As has been stated, this is the point where the ray pierces the meridional plane. The X coordinate of the diapoint is therefore zero, but the other coordinates must be found. By means of a diagram such as that in Figure 8.10 it is easy to show that

$$L'_d = -X'M'/K' \quad \text{and} \quad H'_d = Y' - (X'L'/K')$$

where K', L', M' are the direction cosines of the ray as it emerges from the lens, and Y', Z' are the coordinates of the point where the ray pierces the image plane. The L'_d, H'_d are the required coordinates of the diapoint relative to the midpoint of the image plane and the optical axis of the lens.

8.3.8 Example of a Skew-Ray Trace

To illustrate the kind of record required in the manual tracing of a skew ray by these formulas, we will trace a ray through our old familiar cemented doublet objective, entering at an upward slope of 3° through a point at unit distance

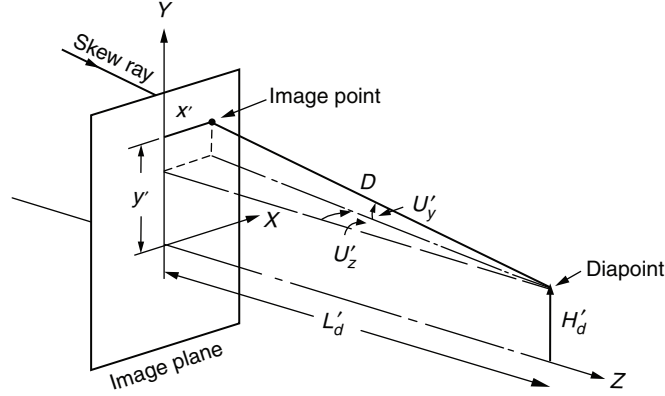


Figure 8.10 Diapoint calculation.

behind the meridional plane and on the same level as the principal ray. Regarding the tangent plane at the first vertex as a refracting surface, the starting data at that surface are

$$\begin{aligned} Z &= 0, & M &= \cos(-3^\circ) = 0.9986295 \\ Y &= 0, & L &= \sin(-3^\circ) = -0.0523360 \\ X &= 1, & K &= 0 \end{aligned}$$

We now transfer the ray from the tangent plane to the first spherical refracting surface in the usual way. The results of the trace are shown in Table 8.2.

8.4 GRAPHICAL REPRESENTATION OF SKEW-RAY ABERRATIONS

8.4.1 The Sagittal Ray Plot

The name *sagittal* is generally given to the 90° and 270° skew rays that lie in a plane perpendicular to the meridional plane, containing the principal ray. This is not one single plane throughout a lens but it changes its tilt after each surface refraction (Figure 8.1). The point of intersection of a sagittal ray with the paraxial image plane may have both a vertical error and a horizontal error relative to the point of intersection of the principal ray, and both these errors can be plotted separately against some suitable ray parameter. This parameter is often the entrance pupil coordinate for $\theta = 90^\circ$ or the horizontal distance x from the meridional plane to the point where the entering ray pierces the tangent plane at

Table 8.2
Manual Tracing of a Skew Ray

	<i>Tangent plane</i>					<i>Image plane</i>
r	∞	7.3895	-5.1784	-16.2225	∞	
d	0	1.05	0.4	11.28584		
n	1	1.517	1.649	1		
$(n/n')^2$		0.4345390	0.8463106	2.719201		
F	0.9986295	0.8000638	0.9673926	0.9952001		
G	0.1353271	1.584704	0.9077069	22.627015		
D	0.0680706	0.8939223	0.4623396	11.368061		
X	1	1.0	0.9584830	0.9457557	-0.0033456	
Y	0	0.0035625	0.0342546	0.0491091	0.6289086	
Z	0	0.0679773	-0.0895928	-0.0276675	0	
$\cos I$		0.9894178	0.9726889	0.9958924		
$\cos I'$		0.9954155	0.9769360	0.9887907		
J		0.0704550	-0.0261467	0.0402796		
K	0	-0.0464436	-0.0275280	-0.0834884		
L	0.0523360	0.0343342	0.0321289	0.0510025		
M	0.9986295	0.9983305	0.9991040	0.9952011		

the first lens vertex. Figure 8.11 shows the layout for an $f/2.8$ triplet photographic objective⁷ having a focal length of 10, and Figure 8.12 shows the set of meridional and sagittal ray plots for this lens having the following prescription:

Radius	t	n	V
4.7350	0.6372	1.7440	44.9 (LaF2)
148.835 (stop)	1.0015	(air)	
-5.8459	0.2705	1.7400	28.2 (SF3)
5.1414	0.9253	(air)	
33.1041	0.6979	1.7440	44.9 (LaF2)
-4.4969	8.4894	(air)	

The meridional plot, of course, has no symmetry, but the two sagittal ray plots do have symmetry. The vertical errors are identical for rays entering at equal distances in front of and behind the meridional plane, these errors being forms of sagittal coma. The horizontal errors are antisymmetrical, so that the error of the 90° ray is equal and opposite to the horizontal error of the 270° ray; these errors represent sagittal field curvature and sagittal oblique spherical aberration, strictly analogous to the effects of tangential field curvature and tangential oblique spherical aberration in the ordinary meridional ray plot.

8.4.2 A Spot Diagram

The meridional and sagittal ray plots already discussed take account of only the rays passing through a cross-shaped aperture over the lens. To include every possible skew ray, it is necessary to divide the lens aperture into a checkerboard of squares, or a rectangular grid, and to trace a ray through every intersection of the lines. Assuming that each ray carries the same amount of light energy, the assembly of the intersection points of all such rays with the image plane will be a fair representation of the type of image that may be expected when the lens has been fabricated, assembled, and tested.

Actually it requires a large number of rays, say hundreds, to provide a fair approximation to the actual image. Of course, it is unnecessary to trace rays

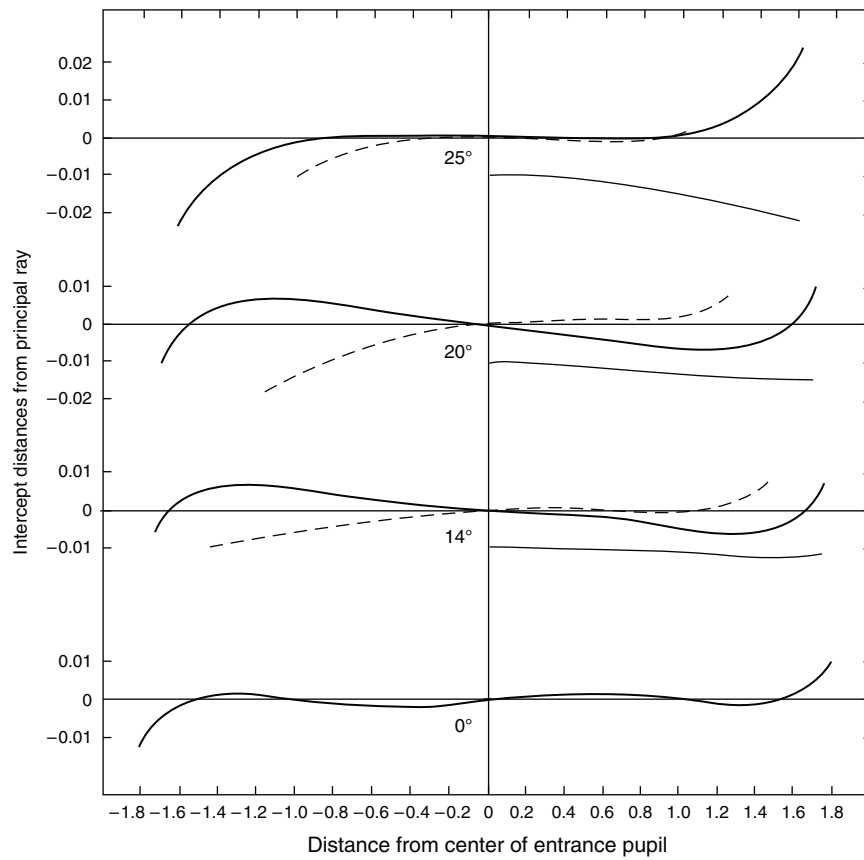


Figure 8.11 Typical ray plots for a triplet lens.

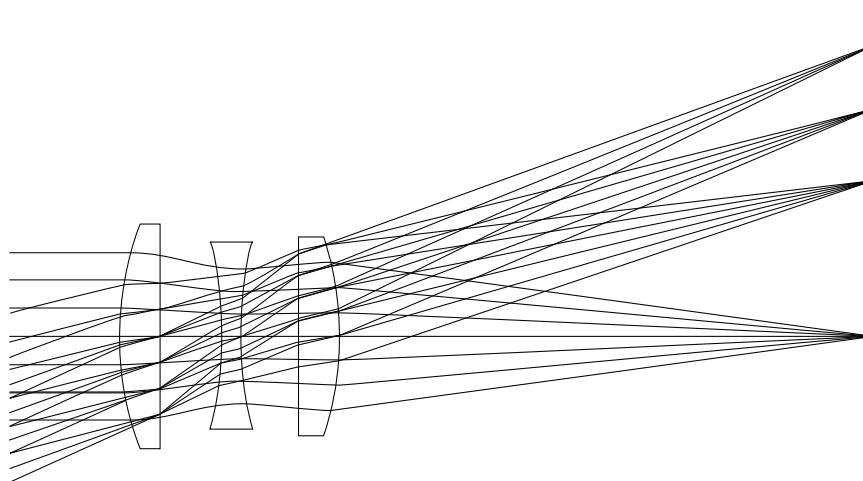


Figure 8.12 Layout of an $f/2.8$ triplet objective lens, U.S. Patent 2,966,825.

both behind and in front of the meridional plane since they are identical, but it is necessary to plot both rays in the image plane.⁸ Such dot patterns are called spot diagrams, and they were obviously never plotted before the advent of high-speed computers to do the ray tracing.⁹

A typical spot diagram for the aforementioned $f/2.8$ triplet photographic objective is shown in Figure 8.13a, where a rectangular pattern was used; for this pattern over a thousand skew rays were traced through each side of the lens aperture. Figure 8.13b shows the same spot diagram except that a hexapolar pattern was used, and a dithered pattern was used in Figure 8.13c. The dithered pattern traces a pseudorandom distribution of rays through the entrance pupil. The intent of using a dithered pattern is to mitigate the symmetry artifacts caused by using either the rectangular or the hexapolar pattern.

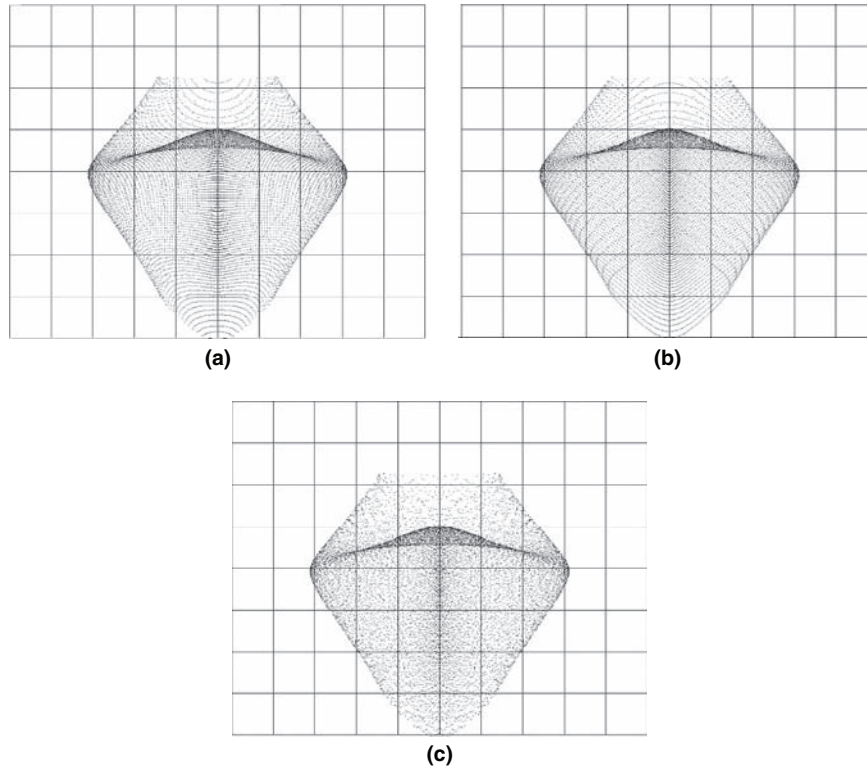


Figure 8.13 A typical spot diagram for an $f/2.8$ triplet lens at 14° off axis ($f' = 10$) for d light. The ray pattern used at the entrance pupil was (a) rectangular, (b) hexapolar, and (c) dithered.

The comparison of Figures 8.13a, b, and c clearly demonstrates the presence of such artifacts and their mitigation. It should be recognized that there is no perfect or best ray pattern. Also, the lens designer should be cautious about making inferences about aberrations in the lens as a consequence of observing such induced artifacts. And finally, spot diagrams are strictly geometric; however, an image of the point source accounting for both aberrations and diffraction can be computed and displayed. Figure 8.14 presents the *point spread function* (PSF) for the same image shown in Figure 8.13. Notice the similarities and differences between them, but realize that the PSF is closer to what will actually be observed.

DESIGNER NOTE

A rough rule of thumb, often called the “three-to-one rule,” can be used to decide if diffraction or geometric aberration dominates. If the rms geometric blur diameter is less than one-third of the diffraction blur diameter, then diffraction is dominant. The converse is also true. In the “in-between” region, both must be considered. The 3:1 ratio can be 5:1 or whatever the designer desires, but not less than 3:1.

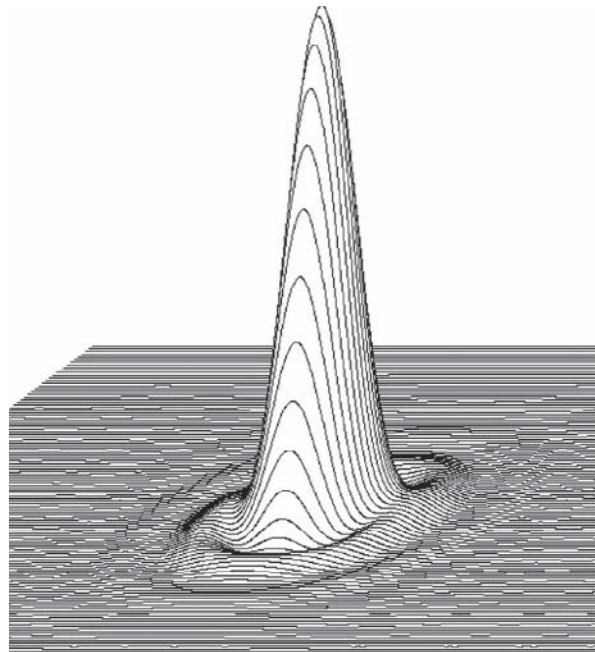


Figure 8.14 Point spread function for the same lens and object as in Figure 8.13.

8.4.3 Encircled Energy Plot

By counting the rays enclosed by a succession of circles of increasing size laid over the spot diagram, it is possible, by a suitable computer program, to plot a graph of the *encircled energy* of a given lens at several obliquities. This assumes that each ray carries the same amount of light energy, a justifiable assumption if the rays are incident in a checkerboard pattern at the entrance pupil of the lens. To include chromatic effects it is necessary to trace many rays in other colors, the size of the checkerboard squares for each wavelength being dependent on the spectral response of the detector intended for that particular lens.

The encircled energy graphs of the $f/2.8$ triplet used above are shown in Figure 8.15. As can be seen, this one plot shows the performance of the lens at several obliquities. Graphs of this type provide the designer with a great deal of useful information, particularly in comparing one design with another. It is noted that Figure 8.15 is based on only geometric ray trace data. Encircled energy plots can be made where the geometric data are multiplied by the diffraction-limited values to produce perhaps a better expectation of what may be observed. Diffraction-encircled energy plots can also be produced which properly account for both diffraction and aberrations; however, computation is more time-consuming than for the other versions.

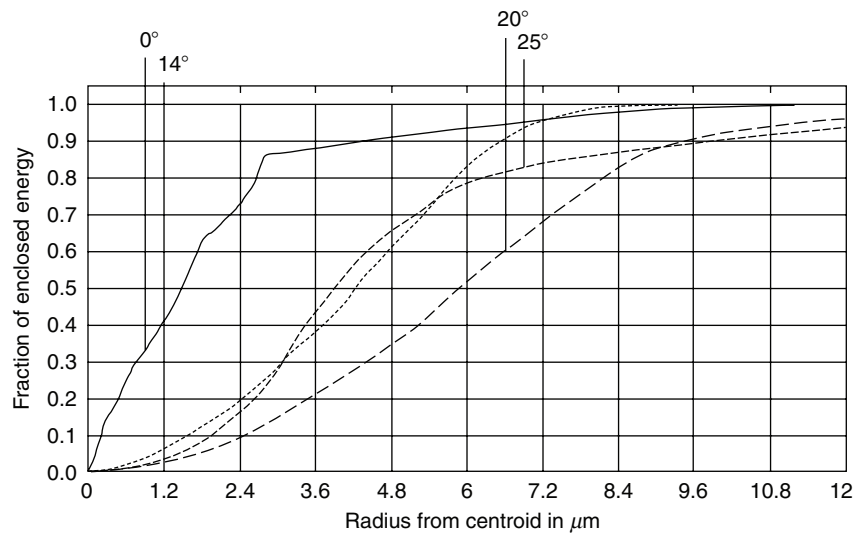


Figure 8.15 Encircled energy plot.

8.4.4 Modulation Transfer Function

An important addition to the tools of the lens designer was the development of the modulation transfer function (*MTF*) for optical systems; serious interest in *MTF* began in the 1950s although it was not actually accepted by most practitioners until the 1970s. Today, the *MTF* is arguably the dominant method for describing the performance of lenses. Willams and Becklund have presented a comprehensive history and study of the optical transfer function (*OTF*).¹⁰ The $MTF = |OTF|$ where the *OTF* is a complex value. Analogous to electronic communication systems, optical systems can also be considered as linear systems and utilize similar theory.¹¹ In an electrical system, the *MTF* is essentially the ratio of the output to the input of a linear system as a function of frequency (i.e., $MTF(f) = \text{output}(f) / \text{input}(f)$).

In a like manner, the *MTF* for a lens is $MTF(v) = \text{output}(v) / \text{input}(v)$, where v is spatial frequency and may be multidimensional, unlike temporal frequency used in electronics. In a simple sense, the *MTF* is a measure of the fidelity the lens-formed image has to the object. Although the measurement should be, and has been, accomplished using sinusoidal spatial targets, most often *MTF* is measured using alternating black and white bars. The resulting *MTF* is known as square-wave *MTF* and the plots are not the same as the sine-wave *MTF*. Figure 8.16 shows the geometrical *MTF* for the triplet lens multiplied by

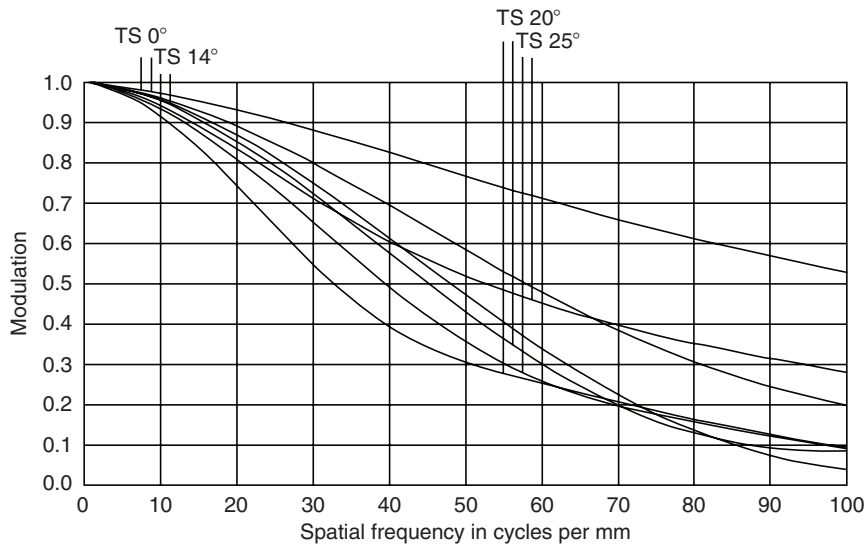


Figure 8.16 Geometrical *MTF* multiplied by the diffraction-limited *MTF*.

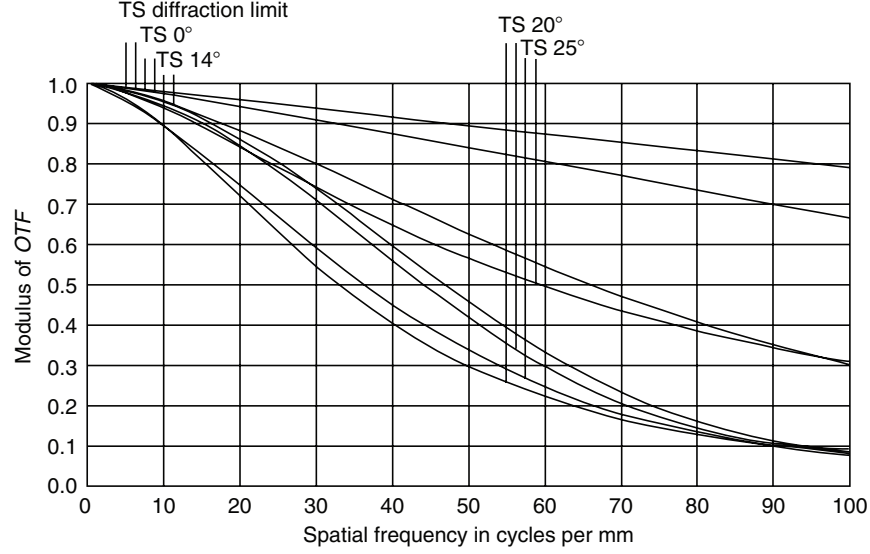


Figure 8.17 Diffraction-based *MTF* and curve showing the diffraction-limited *MTF* for comparison.

the *MTF* for a diffraction-limited lens. Compare this *MTF* to that in Figure 8.17, which shows the diffraction-based *MTF*. As can be observed, the axial geometric *MTF* is underestimated while being overestimated for 14° and is about the same for the greater off-axis object points. Multiplication of the geometric *MTF* by the *MTF* for a diffraction-limited lens should be used only when the geometric blur is not greater than about one-third that of the diffraction blur.

When the lens is to be used over some finite spectral bandwidth, a method has been determined for estimating the blur size and shape for a polychromatic source and an aberration-free lens system.¹² The perfect-image irradiance distribution of a polychromatic point source can be written as

$$E(r) = C_1 \int_0^\infty \tilde{\mathfrak{R}}(\lambda) \left[\frac{2J_1(kD_{ep}r/2)}{kD_{ep}r} \right]^2 d\lambda$$

where $\tilde{\mathfrak{R}}(\lambda)$ is the peak normalized spectral weighting factor and C_1 is a scaling factor. By invoking the central limit theorem to approximate this distribution by a Gaussian function, we obtain

$$E(r) \approx C_2 e^{-(r^2/2\sigma^2)}$$

where C_2 is a scaling constant and σ^2 is the estimated variance of the irradiance distribution. When $\tilde{\mathfrak{R}}(\lambda) = 1$ in the spectral interval λ_{short} to λ_{long} and zero

otherwise with $\lambda_{\text{short}} < \lambda_{\text{long}}$, an estimate of sigma can be written as

$$\sigma = \frac{M\lambda_{\text{long}}}{\pi D_{ep}}$$

where $M = 1.335 - 0.625b - 0.25b^2 - 0.0465b^3$ with $b = (\lambda_{\text{long}}/\lambda_{\text{short}}) - 1$. Should $\tilde{\mathcal{R}}(\lambda) = \lambda/\lambda_{\text{long}}$ in the spectral interval λ_{short} to λ_{long} and zero otherwise, which approximates the behavior of a quantum detector, $M = 1.335 - 0.65b + 0.385b^2 - 0.099b^3$.

The Gaussian estimate of residual error is less than a few percent for $b = 0.5$ and remains useful even as $b \rightarrow 0$. A useful estimation of the modulation transfer function for this diffraction-limited *polychromatic* lens system is given by

$$MTF(v) \approx e^{-2(\pi\sigma v)^2}$$

where v is the spatial frequency. This approximation can provide a useful insight into expected performance limits.

8.5 RAY DISTRIBUTION FROM A SINGLE ZONE OF A LENS

The nature of the various oblique aberrations of a lens may be better understood if we trace a family of rays passing through a single zone of a lens, both on and off axis. We take the cemented telescope doublet used many times before and isolate a single zone of radius one unit. On axis, all the rays from this zone will, of course, intersect at a single point, forming a perfect focus. At an obliquity of only one degree, however, the rays from the zone form a succession of complicated loop figures as shown in Figure 8.18. As before, the upper and lower rim rays are labeled 0° and 180° , while the sagittal rays are 90° and 270° .

Referring to this diagram, the tangential focus is at the intersection of rays 0° and 180° , giving 0.000084 for the amount of tangential coma. The sagittal focus is where the 90° and 270° rays intersect, forming a sagittal coma of magnitude 0.000035, about one-third of the tangential coma. It can be proved that in the absence of higher-order aberrations this ratio should be exactly 3:1.¹³ The presence of field curvature is indicated by the tangential and sagittal foci not being in the same plane as the axial image of the zone. This series of patterns arises at each zone of the lens, and it is clear that when all zones are open together, the resulting image is very complicated indeed.

It should be noted that the diapoint locus of the zone is also indicated in Figure 8.18. It is bisected by the sagittal image, and it ends at upper and lower tangential rays. Conrady refers to the diapoint locus as the *characteristic focal line* of a zone.

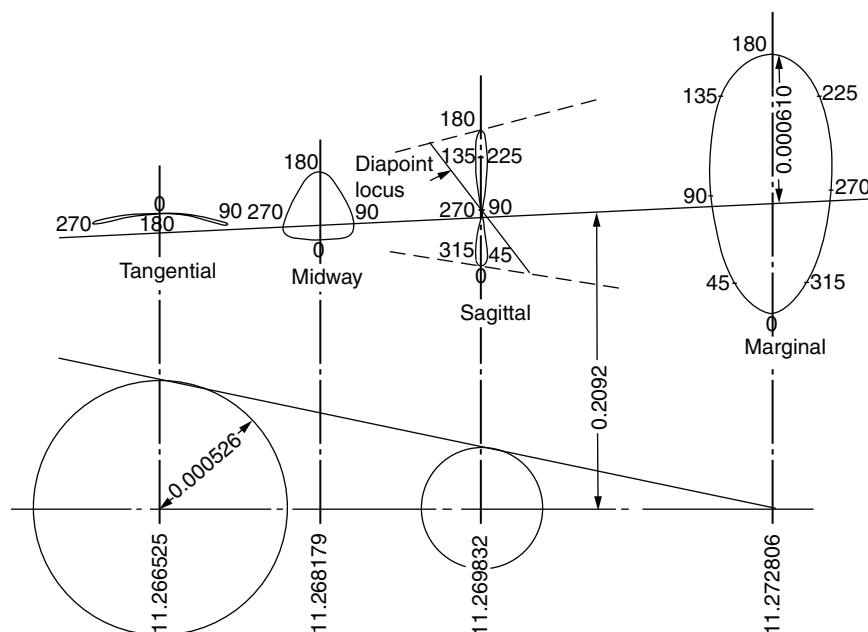


Figure 8.18 Ray distribution from a single zone of a lens.

ENDNOTES

- ¹ A. E. Conrady, pp. 284, 742.
- ² The word *vignetting* is pronounced as “vĭn yĕt’ ting”. It is used in the context that the image formed shades off gradually toward the edges as a consequence of more and more rays in the beam being vignettted as the obliquity angle increases.
- ³ MIL-HDBK-141, *Optical Design*, Section 6.11.8, Defense Supply Agency, Washington, DC (1962).
- ⁴ R. Kingslake, “Illumination in optical images,” in *Applied Optics and Optical Engineering*, II:195, Academic Press, New York (1965).
- ⁵ MIL-HDBK 141, *Optical Design*, Chapter 5, Defense Supply Agency, Washington, DC (1962).
- ⁶ Daniel Malacara and Zacarias Malacara, *Handbook of Lens Design*, Chapter 2, Marcel Dekker, New York (1994).
- ⁷ C. Baur and C. Otzen, U.S. Patent 2,966,825, filed in February (1957).
- ⁸ The context of “behind and in front of the meridional plane” is that one is considered to be viewing the lens from the side with light propagation going from left to right. Front is the space between the observer and the meridional plane and contains entrance pupil coordinates containing θ values from 0° to 180° .

- ⁹ As a historical note, once high-speed computers became available in the late 1950s and into the 1970s, plotters were often not readily available and were rather slow as well. A clever method was devised to use the text printer to generate a useful spot diagram. The technique was to consider each print character position as a ray bin and a character was assigned to each bin to represent the number of rays “hitting” that bin. Of course, a blank meant no rays, but then the count went as 1, 2, 3, ..., 9, A, B, C, ..., Z. Although such a spot diagram is inferior to the spot diagrams of today, a lens designer quickly learned to “read” such character-based spot diagrams. Printers even in the 1950s were quite fast because a primary utilization of computers was for business and government accounting, inventory, and payroll.
- ¹⁰ Charles S. Williams and Orville A. Becklund, *Introduction to the Optical Transfer Function*, Wiley, New York (1989).
- ¹¹ R. Barry Johnson, “Radar analogies for optics,” *Proc. SPIE*, 128:75–83 (1977).
- ¹² R. Barry Johnson, “Lenses,” in *Handbook of Optics, Second Edition*, Vol. II, Chapter 1, pp. 1.39–1.41, Michael Bass (Ed.), McGraw-Hill, New York (1995).
- ¹³ *Hint*: Consider the ray aberration expansion equations in Section 4.2, assume all aberrations are zero except for third-order coma (σ_2), and consider the ratio of the values of ε_y for $\theta = 0^\circ$ and 90° .