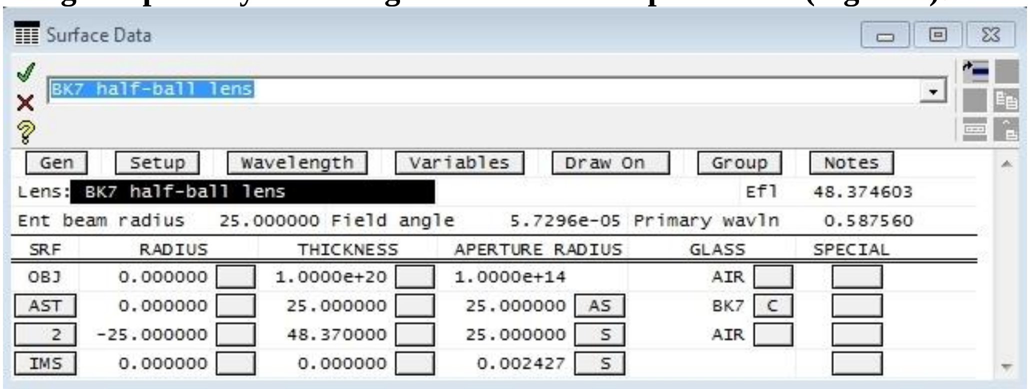
DESIGN OF OPTICAL SYSTEMS

A (very) short[1] course in optics design:

Optical systems are designed using ray tracing software like OSLO, Zemax, CodeV (to name the best known). We already used Oslo to illustrate some geometrical aberrations and thus learned the basics that are summarized here:

1) **Describing an optical system using the surface data spreadsheet (Figure 1):**



Surface Data

BK7 half-ball lens

Gen | Setup | Wavelength | Variables | Draw On | Group | Notes

Lens: BK7 half-ball lens — Efl 48.374603

Ent beam radius 25.000000 Field angle 5.7296e-05 Primary wavln 0.587560

SRF	RADIUS	THICKNESS	APERTURE RADIUS		GLASS		SPECIAL
OBJ	0.000000	1.0000e+20	1.0000e+14		AIR		
AST	0.000000	25.000000	25.000000	AS	BK7	C	
2	-25.000000	48.370000	25.000000	S	AIR		
IMS	0.000000	0.000000	0.002427	S			

Figure 1: Example of a surface data spreadsheet describing an optical system.

Except the first line and the last line, each line of the spreadsheet describes an optical interface:

- "Radius" is the radius of curvature on the interface. RD > 0 means that the center of the sphere is to the right of the interface position. (In Figure 2, RD[1] is positive). RD = 0 is a flat surface.
- "Thickness" is the distance to the next interface (measured on the optical axis of the system as described in Figure 2). Negative values are needed if a mirror is used.

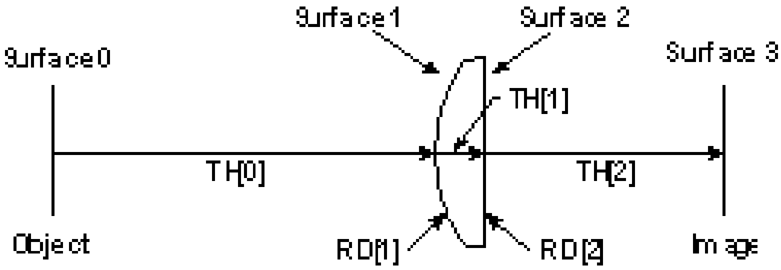


Figure 2: Meaning of some column values in the surface data spreadsheet.

- "Glass" describes the material in which the light propagates after the interface (over a distance "thickness").
- "Aperture radius" is the diameter of the open aperture of the lens divided by 2. This value is usually calculated by the software as visible by the letter S on the corresponding button. In order to act on the width of the light-beam, modify the value of the "Ent(trance) beam radius" above the spreadsheet. The "aperture radius" can however also be set manually to implement an iris diaphragm.

The first line describes the object and its distance to the first interface (value in the thickness column).

[1] Please consider this lab work just as a glance into the world of optics design. Typically an assistant optics designer should have at least 5 000 h of experience in ray-tracing. At more than 10 000 h one may be able to make a full complex design independently. (For comparison: An employee usually works 1 600 h per year.)

The last line describes the screen and contains the (manual) autofocus-function in the thickness column. The position where an observer would place the screen in order to obtain the sharpest image of a homogeneously illuminated object, is best described by the "best RMS OPD" (RMS = root mean square, OPD = optical path difference).

2) Some definitions in geometrical optics

For an off-axis object point (for example (0, PY, 0)) one defines two planes: The *tangential plane* (or meridional plane) that comprises the optical axis and the object point (abbreviated T in some plots) and the *sagittal plane* that is perpendicular to the tangential plane and comprises the object point and the center of the aperture stop.

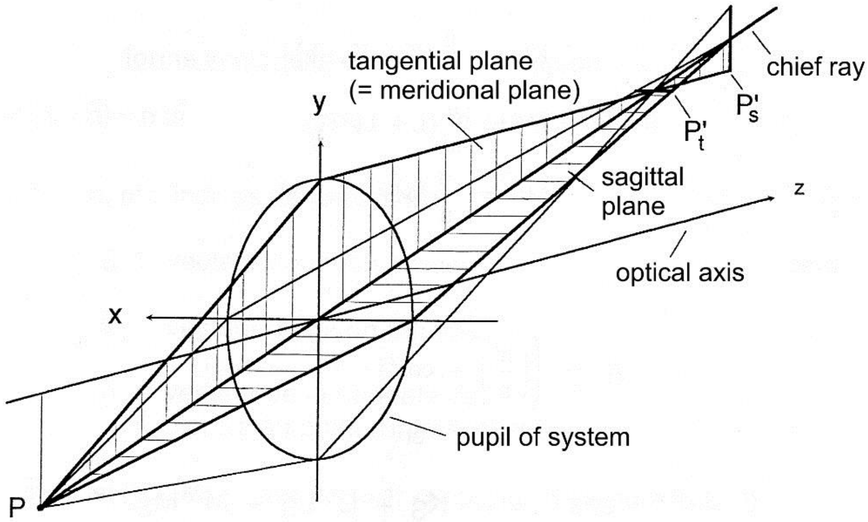


Figure 3: Some definitions in ray optics. P is vertically displaced from the optical axis ($y_P \neq 0$). Thus the tangential plane is the vertical plane (y-z-plane) and the sagittal plane contains the x-axis. P'_s is the image formed by sagittal rays and P'_t is the image formed by tangential rays.

The ray from the object point through the center of the aperture stop is called the *chief ray*. The distance between the chief ray and the optical axis at an image location defines the size of the image.

Another important ray is the *marginal ray*. It is the tangential (or meridional) ray that starts at the point where the object crosses the optical axis, and touches the edge of the aperture stop of the system. This ray is useful, because it crosses the optical axis again at the locations where an image will be formed.

3) Configuring the surface data spread sheet for design tasks

Designing an optical system involves variations of some of its parameters in order to obtain the wished performances. However many parameters exist which may be dependent on one another. In order to avoid manual adjustment of the whole chain of dependent parameters after each variation of an independent parameter, these dependencies can be implemented in the spreadsheet. These functions are accessed by clicking on the buttons to the right of the parameters. The most frequent dependencies are:

- *"Make the lens as small as possible, but large enough to let pass all rays":* Aperture radius / solved (switched on by default).
- *"Make the lens as thin as possible, but thick enough to hold it without risking breakage":* Thickness / solves / edge thickness. Remark: It is not possible to use both features simultaneously as the edge thickness solve will ask for a fixed aperture radius.
- *"Move the screen always to the paraxial focus":* This can be achieved automatically by using the thickness of the last material (the before-last line) and making it dependent on a 'solve' of a ray similar to the marginal ray, but staying close to the optical axis, which is called "axial ray" in OSLO.

- *“Move an element between two other elements that keep their position”:* Go to the second distance (the one that should be adjusted automatically) and choose: Thickness / Pickups / minus thickness / constant (sum of both thicknesses) / multiplier (1.0).

The influence of the remaining parameters or at least some of them can then be observed using sliders under Optimize / Slider-Wheel design. The defined parameters can then conveniently be modified using the sliders and the dependent parameters can be observed in the surface data spreadsheet, the autodraw window and another graphics window, the content of which is chosen at slider-wheel design time.

4) Evaluating an optical system

Load the following sample from the Lens database / public / demo triplet (f50mm f/4 20deg).

Oslo will display the “report graphics” plot which summarizes the performances of the optical system. The implemented ‘nearly symmetrical’ design is one way to minimize several types of aberrations in an optical system.

Ray intercept curves

To the left you find the ‘ray intercept curves’ that represent the image space displacements as a function of object-space fractional coordinates. Two examples are given in Figure 4.

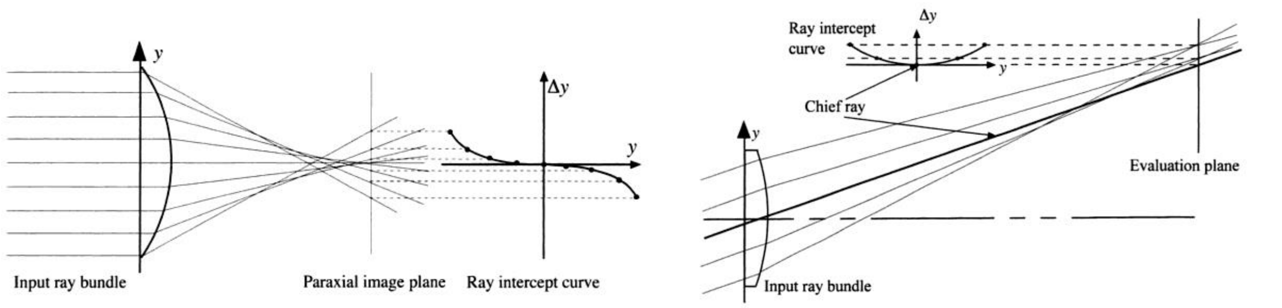


Figure 4.11 Ray intercept curve (right) of a lens with spherical aberration. The errors in the rays in the paraxial image plane are plotted as a function of the location of the input ray on the lens.

Figure 4.12 Lens with coma and ray intercept curve (top). The errors in the rays in the paraxial image plane are plotted as a function of the location of the input ray on the lens.

Figure 4: Examples for ray intercept curves.

For a more detailed explanation of these curves, read the attached chapter of the ‘handbook of optics’. In order to see the axis labels you may generate intercept curves for one field point by using Evaluate / Ray fans / Single field point.

- ➔ What is the optimum curve in this graph? (displayed by an ideal system)
- ➔ How to estimate the diameter of the blur circle from these plots?

Spherical aberration

Typically the longitudinal spherical aberration is displayed (distance in mm between the point where the beam crosses the optical axis and the paraxial focus) as a function of object point position (or angle). Lateral spherical aberration can be calculated by multiplying with tan(direction of chief ray in image space).

- ➔ What is the optimum curve in this graph? (displayed by an ideal system)
- ➔ Describe the changes occurring when you put the screen at different locations by using the manual autofocus function.
- ➔ When looking at the aberration curve, can you tell if the paraxial focus was chosen as reference? Explain.

Distortion
The distortion value is negative for barrel type distortion and positive for pincushion type distortion (Figure 5).

Figure 5 Image of a regular grid made by a system with barrel type distortion (left) and a system with pincushion type distortion (right).

Point spread function and encircled energy
If an optical system is diffraction limited, geometrical ray-tracing no longer provides realistic values. The size of an image point is then determined by the point spread function (PSF). OSLO can also calculate point spread functions and plots them as well as the part of energy of the incoming beam that goes through a circular diaphragm placed at the screen location.

5) Using automatic optimization

Ray-tracing programs include optimization algorithms that can help you to find the best solution to your design-problems. Mathematically speaking optimization is a minimization problem and the critical questions are which are the parameters to modify and which function should be minimized.
The parameters that should be modified are simply indicated in the ‘surface data spreadsheet’ by defining them as ‘variable’.
The function to minimize is usually called ‘merit function’ or in OSLO ‘error function’ and is defined in Optimize / Generate Error Function. A complex black-box error function that is available and that may work for a large range of applications is the GENII Ray aberration error function. However you may define simpler error functions that you can fully understand by choosing the Aberration Operands you want. A description of the available options is given on page 346 of the OSLOProgramRef6X file and an example how to use it is given in the OSLOOpticsReference file (page 40). In short you choose the operands to keep (for example PY, SA3 and EFL) and the weight (importance) each of these operands should have in your optimization. By default, OSLO optimizes all values to zero but this is not reasonable for all operands. If an operand should be optimized to a value different from zero (for example the effective focal length) you need to modify the entry in the ‘definition’ column. For example, in order to optimize a system to an effective focal length of 200 mm change ‘OCM21’ (which is the internal reference of the variable containing the effective focal length) to ‘OCM21 - 200’.
Once variables and error function are defined, the optimization can be started by Optimize / Iterate and the resulting (optimum) values will replace the initial values in the surface data spreadsheet. Pay attention to the fact that different starting values may lead to different results after optimization, even if the problem is always the same. This may occur because the merit function may have several local minima.

Exercises:

1) Simple optimization test

Make a single BK7 lens that has the same effective focal length and the same NA as the triplet from the lens data base. In order to achieve this, start with a plano-convex shape (in the right orientation) having a radius of the curved shape of 25 mm and use automatic optimization using the operands PY, SA3 and EFL.

- ➔ What are the significations of the operands PY and SA3? To which optimum values should they be optimized and why?
- ➔ Briefly compare the performances of the triplet and the singlet lens.

2) Achromatic doublet

Design a cemented achromatic doublet similar to the one in Figure 6 with the same effective focal length and the same NA as the triplet from the lens data base. The glass map in the appendix will help you to choose the glass types. Compare its chromaticity with the singlet lens and the triplet lens.

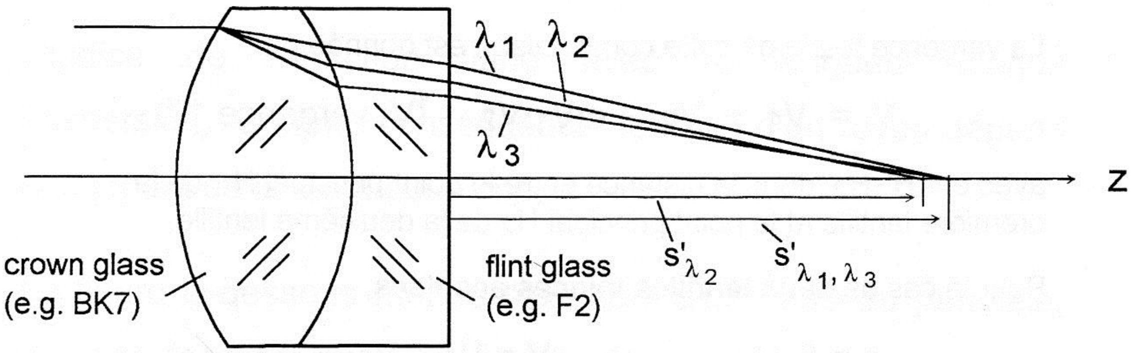


Figure 6 Sketch of a positive cemented doublet lens.

- ➔ Optimize this doublet to reach a spot size lower than the diffraction limit.
- ➔ If you want to use this lens to do a fiber coupling, how much light will you be able to couple to a fiber with a core diameter of 20 µm? (Calculate the Point Spread Function).

3) Zoom lens

Understand and design a very simple zoom lens according to the schematics in Figure 7.

- ➔ Using a geometrical construction, show that the effective focal length of the zoom system in Figure 7 is varying when the two positive lenses are moved.
- ➔ Why does a (far away) object appear larger on the photograph if the lens has a longer focal length? How to quantify the image size in a ray-tracing software?

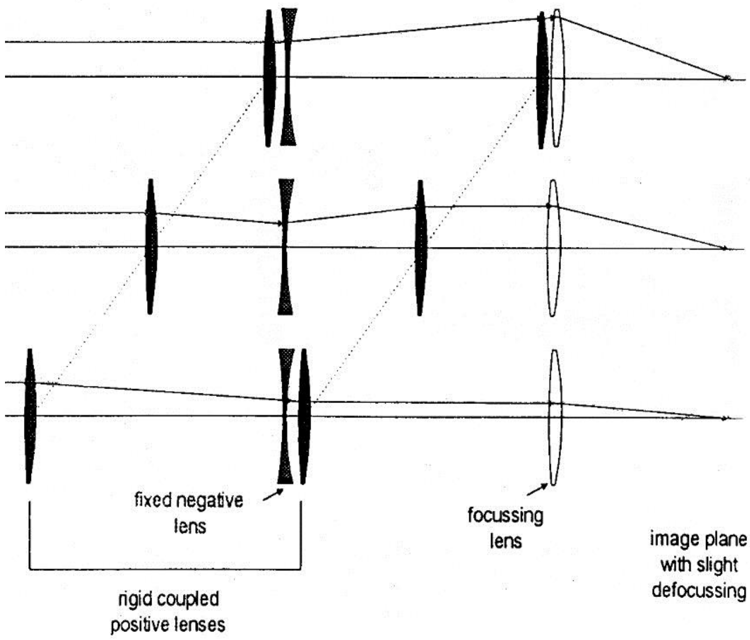


Figure 7 Sketch of a simple zoom lens. The two moving lenses are made out of glass with a high refractive index. The refractive index of the fixed negative lens is lower than this. The image plane of this kind of system is not exactly fixed.

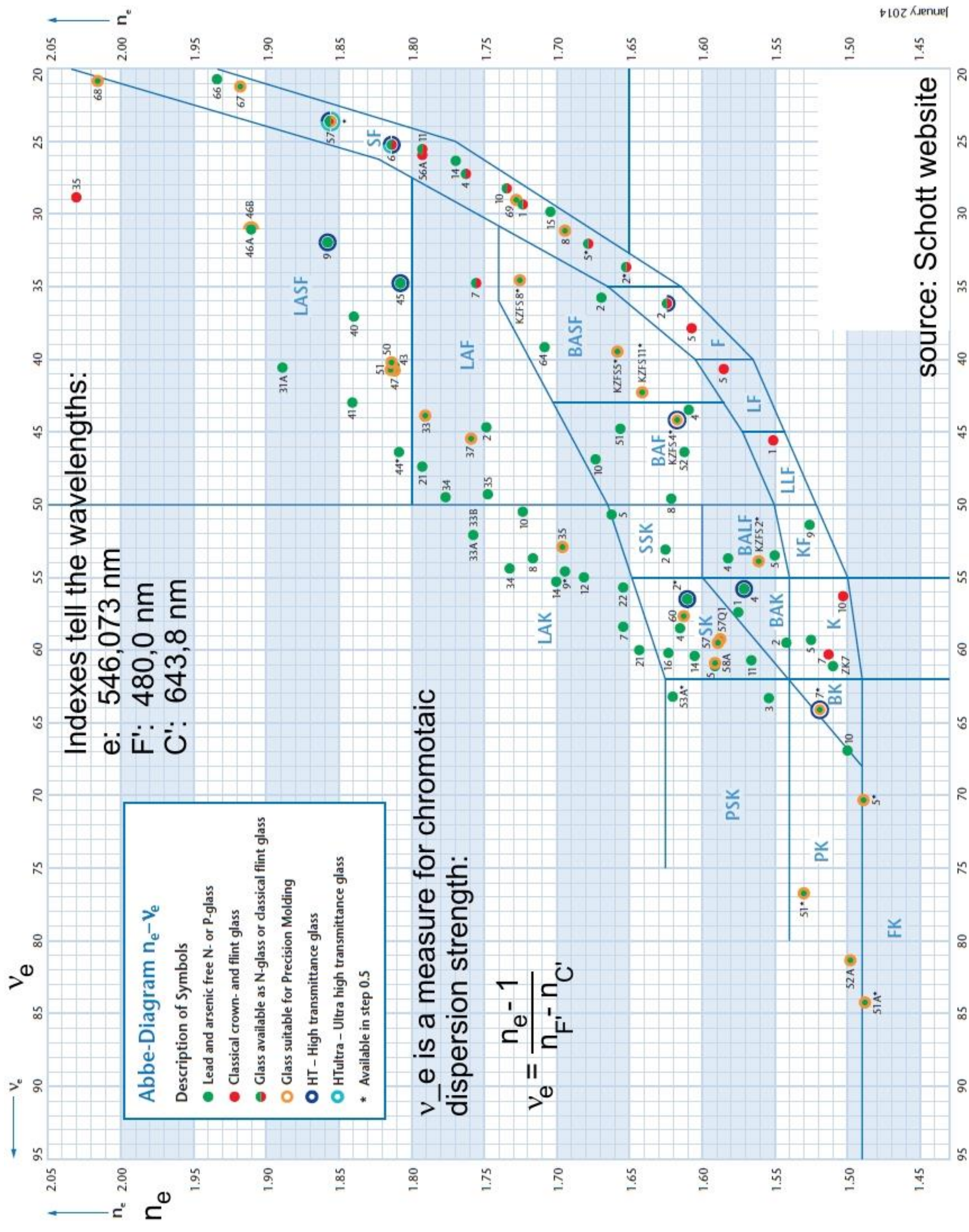
Start with a configuration using symmetric lenses and monochromatic light (to simplify things). Prepare your spreadsheet such that by changing only one thickness parameter the rigidly coupled lenses 'slide' as in reality and the screen is always automatically positioned at the location of the paraxial image.

➔ For your starting configuration:
  - Print three system drawings like the ones in Figure 7.
  - Print the surface data spreadsheet for one of the positions
  - Characterize the relative variation of the image plane location when moving the mobile lenses.

Now minimize the relative variation of the image plane location.
I propose to do this using the Optimize / Slider-wheel design feature of Oslo. Define two slider wheels, one for zooming and one for changing the first radius of curvature of the system. You may also change other parameters but the optimization of the first radius of curvature is essential to achieve a nearly stable image plane location. The asymmetry of the lenses may be observed in Figure 7.

➔ For your final configuration:
  - Print the surface data spreadsheet for one of the positions
  - Characterize the relative variation of the image plane location when moving the mobile lenses.
  - Plot the image size as function of the zoom position
  - Plot the spot diagram size (sharpness) as function of the zoom position
  - If your system is diffraction limited in any zoom position, print the encircled energy plot.

Appendix I: Abbé diagram



CHAPTER 33
ABERRATION CURVES IN LENS DESIGN

Donald C. O'Shea
Georgia Institute of Technology
Center for Optical Science and Engineering
and School of Physics
Atlanta, Georgia

Michael E. Harrigan
Eastman Kodak Company
Electronic Imaging Research Laboratory
Rochester, N.Y.

33.1 GLOSSARY

H	ray height
NA	numerical aperture
OPD	optical path difference
P	petzval
S	sagittal
T	tangential
tan U	slope

33.2 INTRODUCTION

Many optical designers use aberration curves to summarize the state of correction of an optical system, primarily because these curves give a designer important details about the relative contributions of individual aberrations to lens performance. Because a certain design technique may affect only one particular aberration type, these curves are more helpful to the lens designer than a single-value merit function. When a design is finished, the aberration curves serve as a summary of the lens performance and a record for future efforts. For applications such as photography, they are most useful because they provide a quick estimate of the effective blur circle diameter.

The aberration curves can be divided into two types: those that are expressed in terms of ray errors and those in terms of the optical path difference (OPD). OPD plots are usually plotted against the relative ray height in the entrance pupil. Ray errors can be displayed in a number of ways. Either the transverse or longitudinal error of a particular ray relative to the chief ray can be plotted as a function of the ray height in the entrance pupil. Depending upon the amount and type of aberration present, it is sometimes more appropriate to plot the longitudinal aberration as a function of field angle. For example, astigmatism or field curvature is more easily estimated from field plots, described below. Frequently, the curves are also plotted for several wavelengths to characterize chromatic performance. Because ray error plots are the most commonly used format, this entry will concentrate on them.

33.3 TRANSVERSE RAY PLOTS

These curves can take several different forms, depending on the particular application of the optical system. The most common form is the transverse ray aberration curve. It is also called lateral aberration, or ray intercept curve (also referred to by the misleading term ''rim ray plots''). These plots are generated by tracing fans of rays from a specific object point for finite object distances (or a specific field angle for an object at infinity) to a linear array of points across the entrance pupil of the lens. The curves are plots of the ray error at an evaluation plane measured from the chief ray as a function of the relative ray height in the entrance pupil (Fig. 1). For afocal systems, one generally plots angular aberrations, the differences between the tangents of exiting rays and their chief ray in image space.

If the evaluation plane is in the image of a perfect image, there would be no ray error and the curve would be a straight line coincident with the abscissa of the plot. If the curve were plotted for a different evaluation plane parallel to the image plane, the curve would remain a straight line but it would be rotated about the origin. Usually the aberration is plotted along the vertical axis, although some designers plot it along the horizontal axis.

The curves in Fig. 1 indicate a lens with substantial underconnected spherical aberration as evidenced by the characteristic S-shaped curve. Since a change of the evaluation plane serves only to rotate the curve about the origin, a quick estimate of the aberrations of a lens can be made by reading the scale of the ray error axis (y axis) and mentally rotating the plot. For example, the blur spot can be estimated from the extent of a band that would enclose the curve a in Fig. 1, but a similar estimate could be made from the curves b or c, also.

The simplest form of chromatic aberration is axial color. It is shown in Fig. 2 in the presence of spherical aberration. Axial color is the variation of paraxial focus with wavelength and is seen as a difference in slope of the aberration curves at the origin as a

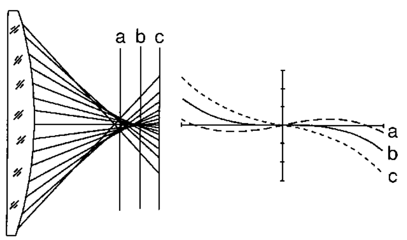


FIGURE 1 (*Left*) Rays exiting a lens are intercepted at three evaluation planes. (*Right*) Ray intercept curves plotted for the evaluation planes: (a) at the point of minimum ray error (circle of least confusion), (b) at the paraxial image plane, (c) outside the paraxial image plane.

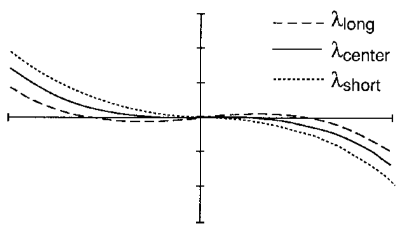


FIGURE 2 Meridional ray intercept curves of a lens with spherical aberration plotted for three colors.

function of wavelength. If the slopes of the curves at the origin for the end wavelengths are different, primary axial color is present. If primary axial color is corrected, then the curves for the end wavelengths will have the same slope at the origin. But if that slope differs from the slope of the curve for the center wavelength, then secondary axial color is present.

A more complex chromatic aberration occurs when the aberrations themselves vary with wavelength. Spherochromatism, the change of spherical aberration with wavelength, manifests itself as a difference in the shapes of the curves for different colors. Another curve that provides a measure of lateral color, an off-axis chromatic aberration, is described below.

For a point on the axis of the optical system, all ray fans lie in the meridional plane and only one plot is needed to evaluate the system. For off-axis object points, a second plot is added to evaluate a fan of skew rays traced in a sagittal plane. Because a skew ray fan is symmetrical across the meridional plane, only one side of the curve is usually plotted. For all curves the plots are departures from the chief ray location in the evaluation plane (Fig. 3). (In the case of the on-axis point, the chief ray is coincident with the optical axis.) For systems of small-field coverage only two or three object points need to be analyzed, but for wide-angle systems, four or more field points may be necessary.

What can be determined most easily from a comparison between the meridional and sagittal fans is the amount of astigmatism in the image for that field point. When astigmatism is present, the image planes for the tangential and sagittal fans are located at different distances along the chief ray. This is manifested in the ray intercept curves by different slopes at the origin for the tangential and sagittal curves. In Fig. 3 the slopes at the origins of the two curves are different at both 70 percent and full field, indicating astigmatism at both field points. The fact that the difference in the slopes of these two curves has changed sign between the two field points indicates that at some field angle between 70 percent and full field, the slopes are equal and there is no astigmatism there. In addition, the variation of slopes for each curve as a function of field angle is evidence of field curvature.

The off-axis aberration of pure primary coma would be evident on these plots as a

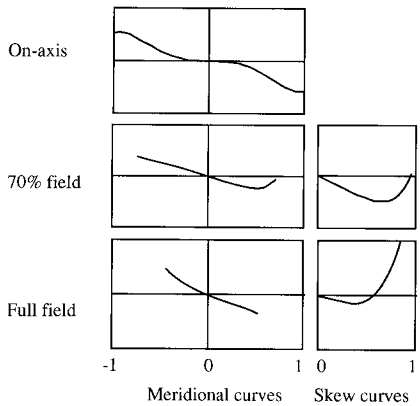


FIGURE 3 Evaluation of a lens on-axis and at two off-axis points. The reduction of the length of the curve with higher field indicates that the lens is vignetting these angles. The differences in slopes at the origin between the meridional and skew curves indicate that the lens has astigmatism at these field angles.

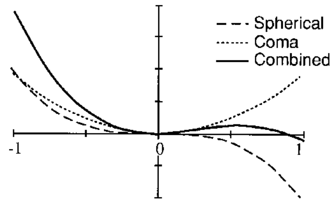


FIGURE 4 Ray intercept curve showing coma combined with spherical aberration.

U-shaped curves for the meridional fan and sagittal fans, the tangential curve being three times larger than the sagittal curve. The "U" will be either upright or upside down depending on the sign of the coma. In almost all cases coma is combined with spherical to produce an S-shaped curve that elongates one of the arms of the "S" and shortens the other (Fig. 4).

The amount of vignetting can be determined from the ray intercept curves also. When it is present, the meridional curves get progressively shorter as the field angle is increased (Fig. 3), since rays at the edges of the entrance pupil are not transmitted. Taken from another perspective, ray intercept curves can also provide the designer with an estimate of how far a system must be stopped down to provide a required degree of correction.

33.4 FIELD PLOTS

The ray intercept curves provide evaluation for a limited number of object points—usually a point on the optical axis and several field points. The field plots present information on certain aberrations across the entire field. In these plots, the independent variable is usually the field angle and is plotted vertically and the aberration is plotted horizontally. The three field plots most often used are: distortion, field curvature, and lateral color. The first of these shows percentage distortion as a function of field angle (Fig. 5).

The second type of plot, field curvature, displays the tangential and sagittal foci as a function of object point or field angle (Fig. 6*a*). In some plots the Petzval surface, the surface to which the image would collapse if there were no astigmatism, is also plotted. This plot shows the amount of curvature in the image plane and amount of astigmatism

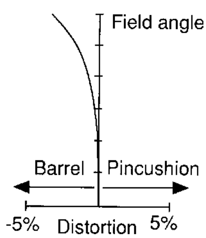


FIGURE 5 Field curve: distortion plot. The percentage distortion is plotted as a function of field angle. Note that the axis of the dependent variable is the horizontal axis.

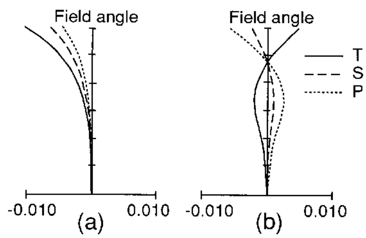


FIGURE 6 Field curve: field curvature plot. The locations of the tangential T and sagittal S foci are plotted for a full range of field angles. The Petzval surface P is also plotted. The tangential surface is always three times farther from the Petzval surface than from the sagittal surface. (*a*) An uncorrected system. (*b*) A corrected system.

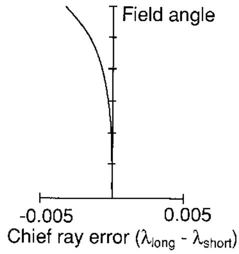


FIGURE 7 Field curve: lateral color plot. A plot of the transverse ray error between red and blue chief ray heights in the image plane for a full range of field angles. Here the distance along the horizontal axis is the color error in the image plane.

over the entire field. In cases of corrected field curvature (Fig. 6*b*), this plot provides an estimate of the residual astigmatism between the axis and the corrected zone and an estimate of the maximum field angle at which the image possesses reasonable correction.

The last of the field curves provides information on color error as a function of field angle (Fig. 7). Lateral color, the variation of magnification with wavelength, is plotted as the difference between the chief ray heights at the red and blue wavelengths as a function of field angle. This provides the designer with an estimate of the amount of color separation in the image at various points in the field. In the transverse ray error curves, lateral color is seen as a vertical displacement of the end wavelength curves from the central wavelength curve at the origin.

Although there are other plots that can describe aberrations of optical systems (e.g., plot of longitudinal error as a function of entrance pupil height), the ones described here represent the ensemble that is used in most ray evaluation presentations.

33.5 ADDITIONAL CONSIDERATIONS

In many ray intercept curves the independent variable is the relative entrance pupil coordinate of the ray. However, for systems with high NA or large field of view, where the principal surface cannot be approximated by a plane, it is better to plot the difference between the tangent of the convergence angle of the chosen ray and the tangent of the convergence angle of the chief ray. This is because the curve for a corrected image will remain a straight line in any evaluation plane.[1] When plotted this way, the curves are called H-tan U curves.

Shifting the stop of an optical system has no effect on the on-axis curves. However, it causes the origin of the meridional curves of off-axis points to be shifted along the curve. In Fig. 8, the off-axis meridional curves are plotted for three stop positions of a double Gauss lens. The center curve (Fig. 8*b*) is plotted for a symmetrically located stop; the outer curves are plots when the stop is located at lens surfaces before and after the central stop.

It is usually sufficient to make a plot of the aberrations in the meridional and sagittal sections of the beam. The meridional section, defined for an optical system with rotational symmetry, is any plane containing the optical axis. It is sometimes called the tangential section. The sagittal section is a plane perpendicular to the meridional plane containing the chief ray. There are some forms of higher-order coma that do not show in these sections.[2] In those cases where this aberration is suspected to be a problem, it may be helpful to look at a spot diagram generated from rays in all sections of the bundle.

For a rotationally symmetric system, only objects in a meridional plane need to be

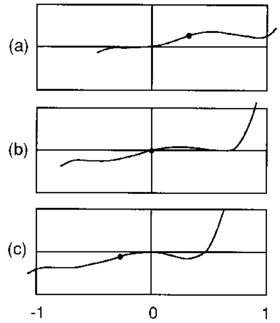


FIGURE 8 The effect of stop shifting on the meridional ray intercept curves of a double Gauss lens. (*a*) Stop located in front of the normal centrally located stop. (*b*) Stop at the normal stop position. (*c*) Stop behind the normal stop position. The dot locates the point on the curve where the origin is located for case (*b*).

analyzed. Also for such systems, only meridional ray errors are possible for purely meridional rays. To observe certain coma types, it is a good idea to plot both the meridional and sagittal ray errors for sagittal rays. It is possible for the meridional section to show no coma and have it show only in the meridional error component of the sagittal fan,[2] but this aberration is normally small.

In addition to plots of the ray error in an evaluation plane, another aberration plot is one that expresses wavefront aberrations as an optical path difference (OPD) from a spherical wavefront centered about the image point. These OPD plots are particularly useful for applications where the lens must be close to diffraction-limited.

33.6 SUMMARY

Aberration curves provide experienced designers with the information needed to enable them to correct different types of aberrations. Chromatic effects are much more easily classified from aberration curves also. In comparison to spot diagrams and modulation transfer function curves, the types of aberrations can be more easily seen and quantified. In the case of diffraction-limited systems, modulation transfer functions may provide better estimates of system performance.

33.7 REFERENCES

1. R. Kingslake, *Lens Design Fundamentals,* Academic Press, San Diego, 1978, p. 144.
2. F. D. Cruickshank and G. A. Hills, "Use of Optical Aberration Coefficients in Optical Design," *J. Opt. Soc. Am.* **50:**379 (1960).