Chapter

12

The Design of Optical Systems: General

12.1 Introduction

In the immediately preceding chapters, we have been concerned with the *analysis* of optical systems, in the sense that the constructional parameters of the system were given and our object was the determination of the resultant performance characteristics. In this chapter we take up the *synthesis* of optical systems; here the desired performance is given and the constructional parameters are to be determined. A large part of the synthesis process is, of course, concerned with analysis, since optical design is in great measure a systematic application of the cut-and-try process.

There is no "direct" method of optical design for original systems; that is, there is no sure procedure that will lead (without foreknowledge) from a set of performance specifications to a suitable design. However, when it is known that a certain type of design or configuration is capable of meeting a given performance level, it is a fairly straightforward process for a competent designer to produce a design of the required type. Further, modest improvements to existing designs can almost always be effected by well-established techniques. Thus, it is apparent that a good portion of the ammunition in a lens designer's arsenal consists of an intimate and detailed knowledge of a wide range of designs, their characteristics, limitations, idiosyncrasies, and potentials. Here is one part of the art in optical design; basically it consists of the choice of the point at which the designer begins.

The electronic computer, in the course of little more than a decade, radically modified the techniques used by optical designers. Previously a designer resorted to all manner of ingenious techniques to avoid tracing rays because of the great expenditure of time and effort involved. The computer has reduced raytracing time by many orders of magnitude, and it is now easier to trace rays through a system than it is to speculate, infer, or interpolate from incomplete data. A computer can even be made to carry through the entire design process from start to finish, more or less without human intervention. The results produced by such a process are nonetheless intricately dependent on the starting point elected (as well as the manner in which the computer has been programmed), so that a great deal of art (if perhaps somewhat less personal satisfaction) is still present in even the most automatic technique.

The ordinary design process can be broken down into four stages, as follows: first, the selection of the type of design to be executed, i.e., the number and types of elements and their general configuration. Second, the determination of the powers, materials, thicknesses, and spacings of the elements. These are usually selected to control the chromatic aberrations and the Petzval curvature of the system, as well as the focal length (or magnifying power), working distances, field of view, and aperture. (Choices made at this stage may affect the performance of the final system tremendously, and can mean the difference between success and failure in many cases). In the third stage, the shapes of the elements or components are adjusted to correct the basic aberrations to the desired values. The fourth stage is the reduction of the residual aberrations to an acceptable level. If the choices exercised in the first three stages have been fortuitous, the fourth stage may be totally unnecessary. At the other extreme, the end result of the first three stages may be so hopeless that a fresh start from stage 1 is the only alternative.

In fully automatic computer design procedures, a portion of stage 1 and all of stages 2, 3, and 4 may be accomplished more or less simultaneously (using an approach that might take a human computer a lifetime or two to slog through). Computer design techniques are discussed in Sec. 12.8.

The basic principles of optical design will be illustrated by three detailed examples in the following sections. A simple meniscus (box) camera lens will be used to show the effects of bending and stop shift techniques, as well as the handling of a simplified exercise in satisfying more requirements than there are available degrees of freedom. An achromatic telescope objective will introduce material choice, achromatism, and multiple bending techniques. An air-spaced (Cooke) triplet anastigmat will illustrate the problem of controlling all the first- and third-order aberrations in a system with just a sufficient number of degrees of freedom to accomplish this and will further illustrate the technique of material selection. The design characteristics of several additional types of optical systems are discussed in Chaps. 13 and 14.

At this point it should be emphasized that the design procedures implied by the discussions in Secs. 12.2, 12.4 to 12.6, and to some extent 12.7, while perfectly valid, are presented here as a way of explaining the principles, relationships, limitations, etc., involved in the design. These procedures are rarely used today; the computer, especially the desktop personal computer, or PC, has enough computing power so that every designer can have access to some sort of automatic lens design program. Nonetheless, a knowledge of these procedures and principles is of great utility to a designer, even when using an automatic design program. For example, such knowledge helps in selecting a good starting design for the computer and, among other things, often helps in figuring out what went wrong when the designer has asked the computer to do the optically impossible.

12.2 The Simple Meniscus Camera Lens

There are just two elements to work with in the design of a meniscus camera lens, the lens itself and the aperture stop. If, for the moment, we restrict ourselves to a thin, spherical-surfaced element, the parameters which we may choose or adjust are the material of the lens, its focal length, its shape (or bending), the position of the stop, and the diameter of the stop. With these degrees of freedom we must design a lens which will produce an acceptable image on a given size of film. This implies that all the aberrations of the system must be "sufficiently" small. It is immediately apparent that the spherical aberration will be undercorrected and that the Petzval curvature will be inward-curving (and equal to $-h^2\phi/2n$); these are the immutable characteristics of a simple lens. Thus, the element power, the size of the aperture, and the field of view must be chosen small enough so that the effects of these aberrations are tolerable. The lens material usually chosen is common crown glass or acrylic plastic, on the basis of cost, since a box camera lens must be inexpensive. A high-index crown does not produce enough improvement in the Petzval curvature to warrant its increased cost; a flint glass would introduce increased chromatic aberrations.

We find ourselves with just two uncommitted degrees of freedom, namely the bending of the lens and the position of the stop. Now in a simple undercorrected system it is axiomatic that for a given (i.e., fixed) shape of the lens (or lenses), the position of the stop (the "natural" stop position—see Sec. 3.4) for which the coma is zero is also the position for which the astigmatism is the most overcorrected (i.e., most backward-curving). Since the Petzval surface will be inward (toward the lens) curving, some overcorrected astigmatism is desirable.

Thus the design technique is straightforward: we choose (arbitrarily) a shape for the lens, determine the stop position at which coma is zero, and evaluate the aberrations. By repeating this process for several bendings and graphing the aberrations as a function of the shape, we can then choose the best design.

There are several ways in which this can be accomplished. Since this is a simple lens of moderate aperture and field, the third-order aberrations are quite representative of the system and one would be quite safe in relying on them. The design could also be handled by trigonometric raytracing. For this example we will work out the design using the thin-lens (G-sum) third-order aberration equations and then check the results by raytracing.

Assuming that the glass has an index of 1.50 and a V-value of 62.5, we will set up the G-sum equations for a focal length of 10, an aperture diameter of 1.0, and an image height of 3 (all in arbitrary units and all subject to scaling and adjustment later). Thus, the element power $\phi = 1/10 = 0.1$, and the total curvature $c = c_1 - c_2 = \phi/(n-1) = 0.2$. With the object at infinity, $v_1 = 0$. Using the G-values worked out in Example G of Chap. 10, we find that the spherical and coma (stop at the lens) given by Eqs. 10.8m and 10.8n are

$$\text{TSC} = -0.145833C_1^2 + 0.05C_1 - 0.005625$$

$$\text{CC} = -0.0625C_1 + 0.01125$$

Now the position of the stop can be determined by solving Eq. 10.8g for Q when CC* is zero.

$$\text{CC}^* = 0 = \text{CC} + Q \cdot \text{TSC}$$

$$Q = \frac{-\text{CC}}{\text{TSC}}$$

Equations 10.8o, p, and r give us

$$\text{TAC} = -0.0225$$

$$\text{TPC} = -0.015$$

$$\text{TAchC} = -0.008$$

and by substituting the above into Eqs. 10.8h, j, and l, we get the following expressions for the third-order astigmatism, distortion, and lateral color with the stop as defined by Q above.

$$\text{TAC}^* = -0.0225 + 2Q \cdot \text{CC} + Q^2 \cdot \text{TSC}$$

$$\text{DC}^* = -0.0825Q + 3Q^2\text{CC} + Q^3\text{TSC}$$

$$\text{TchC}^* = -0.008Q$$

Having established the above relationships, we now select several values for C_1 and evaluate the third-order aberrations for each. The results are indicated in the tabulation of Fig. 12.1 and the graph of Fig. 12.2. Note that $X_s = \text{PC}^* + \text{AC}^*$ and $X_t = \text{PC}^* + 3\text{AC}^*$. [Here we revert to the older symbol (X) for field curvature rather than the currently popular Z.]

A study of Fig. 12.2 can be quite rewarding. First, we note that there are two regions which appear most promising, namely the meniscus shapes at either side of the graph. On the left, the lens is concave to the incident light and (since Q is positive) the stop is in front of the lens. To the right the lens is convex to the incident light and the stop is behind the lens. Both forms have more undercorrected spherical aberration than the less strongly bent shapes, but both have their field curvature "artificially" flattened by overcorrected astigmatism. Note that the form with the least spherical aberration (where CC = 0 and the stop is in contact with the lens) has the most strongly inward curving field. *This inward-curving field is characteristic of any thin optical system with the stop in contact,* since by Eqs. 10.8p and 10.8h

$$\text{Stop in contact } X_T = \text{PC}^* + 3\text{AC}^* = \frac{-h^2\phi\,(3n+1)}{2n}$$

C_1	− 0.4	− 0.2	0.0	+ 0.2	+ 0.4	+ 0.6	+ 0.8
ΣSC	− 0.98	− 0.43	− 0.11	− 0.03	− 0.18	− 0.56	− 1.18
ΣCC	+ 0.036	+ 0.024	+ 0.011	− 0.001	− 0.014	− 0.026	− 0.039
Q	+ 0.74	+ 1.11	+ 2.00	− 0.86	− 1.53	− 0.93	− 0.66
l_p	− 1.23	− 1.84	− 3.33	+ 1.43	+ 2.55	+ 1.56	+ 1.26
ΣAC*	+ 0.087	+ 0.077	0.00	− 0.429	− 0.028	+ 0.040	+ 0.059
X_s	− 0.21	− 0.22	− 0.30	− 0.73	− 0.33	− 0.26	− 0.24
X_t	− 0.04	− 0.07	− 0.30	− 1.59	− 0.38	− 0.18	− 0.12
ΣDC	− 0.02	− 0.03	− 0.08	+ 0.07	+ 0.06	+ 0.03	+ 0.02
% Dist.	− 0.7%	− 1.1%	− 2.5%	+ 2.3%	+ 2.1%	+ 1.0%	+ 0.7%
ΣTchC	− 0.006	− 0.009	− 0.016	+ 0.007	+ 0.012	+ 0.007	+ 0.005

Figure 12.1 Tabulation of the third-order aberrations of a thin lens with the stop at the coma-free position, for various values of C_1.

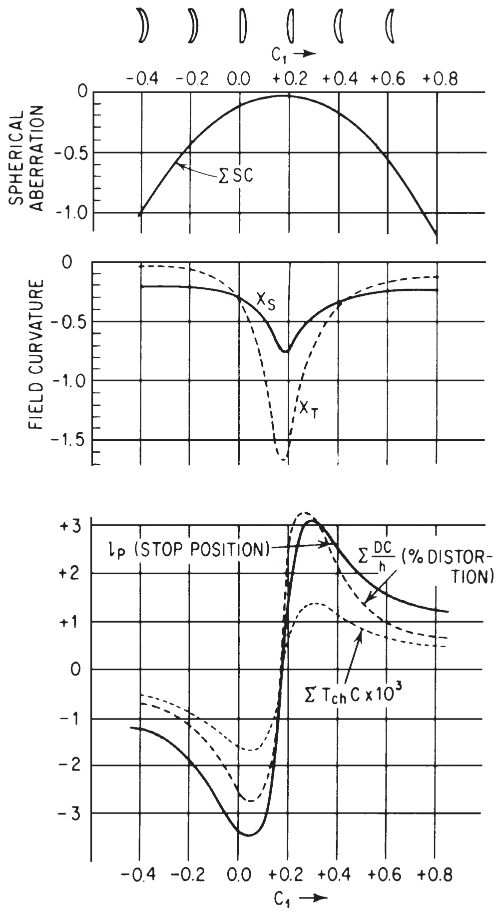


Figure 12.2 The third-order aberrations of a thin lens ($f = 10$, $y = 0.5$, $h = 3$, $n = 1.5$) with the stop at the coma-free position, plotted as a function of the curvature of the first surface (C_1).

Selecting the bending $C_1 = -0.2$ for further investigation, we note that $Q = +1.11$ (from Fig. 12.1). Since $Q = y_p/y$ and $y = 0.5$, we find $y_p = 0.555$. The slope of the principal ray in object space which will yield an image height $h = +3$ with a focal length of $+10$ is $u_p = +0.3$. The stop position is thus

$$l_p = \frac{-y_p}{u_p} = \frac{-0.555}{+0.3} = -1.85$$

or 1.85 units to the left of the lens.

We must of course convert our thin lens to a real lens. A ray with a slope of $+0.3$ through the upper edge of the stop (diameter $= 1.0$) will strike the lens at a height of 1.05, and we shall assume a diameter of twice this for the lens. We determine the curvature of the second surface from $C_2 = C_1 - C = -0.2 - 0.2 = -0.4$, and compute the sagittal heights of the surfaces for the diameter of 2.10. Thus for our lens to have an edge thickness of 0.1, it must have a center thickness of $CT = ET + SH_1 - SH_2 = 0.1 - 0.11 + 0.23 = 0.22$. We now trace an oblique fan of four equally spaced meridional rays through the system and calculate two values of coma (by Eq. 10.6d), one from the upper three rays and one from the lower three. By linear interpolation between the two overlapping three ray bundles, we find that a bundle with a chief ray axial intercept of $L_{pr} = -1.664$ will have zero coma. This is the stop position for the *thick* lens (vs. $l_{pr} = -1.85$ for the *thin* lens.)

The results of a raytrace analysis are shown in Fig. 12.3. The field curvature and spherical aberration forecast by the thin-lens third-order computations are shown as circled points, and the agreement with the actual raytrace is quite good. Note that complete TOA plots could be derived from our knowledge of the manner in which the TOA vary with aperture and image height (see the tabulation of Fig. 3.16). For example, knowing that (longitudinal) third-order spherical varies as Y^2 and that SC = −0.429 for $Y = 0.5$, we could determine that SC = −0.107 for $Y = 0.25$ and plot it accordingly. In Fig. 12.3 the dashed lines in the ray intercept plots indicate the portions of the ray fan which are intercepted by the stop.

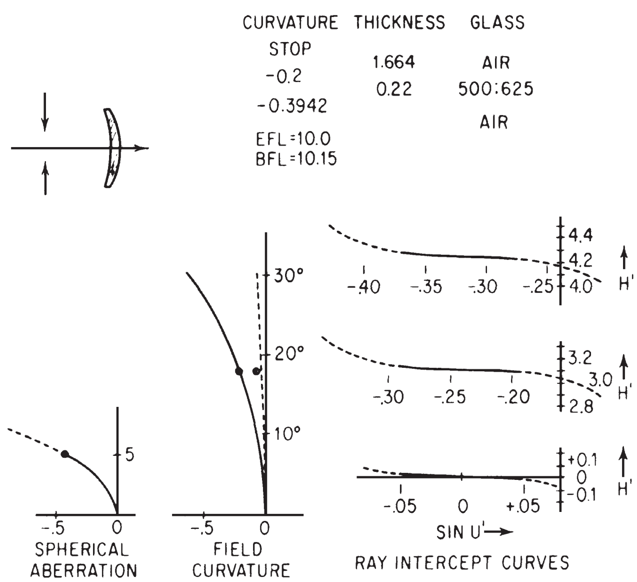


Figure 12.3 The aberrations of a rear meniscus camera lens. The circled points indicate the aberrations predicted by the thin-lens third-order aberration equations (G-sums).

To complete the design we would next scale the entire system to the actual focal length desired. (Note that all the linear dimensions of any system, including the aberrations, may be multiplied by the same constant to effect a change in scale. No additional computation is necessary.) Next an appropriate size for the aperture would be selected, i.e., one which would reduce the aberration blurs to sizes commensurate with the intended application.

The lens form that we have elected to design in this example has the aperture stop in front, i.e., to the left of the lens. This is often referred to as the *rear-meniscus* form. From Fig. 12.2 it is apparent that there is a similar *front-meniscus* form with the stop behind (to the right of) the lens. *Question:* Which is the better design? On the basis of aberration correction, the rear meniscus is slightly better. However, there are several points on which the front meniscus is superior. In a camera, the length of the camera will be approximately equal to the lens focal length for the front meniscus, whereas for the rear meniscus we must add the distance to the stop, resulting in a significantly longer camera. Further, in an inexpensive camera, the shutter is usually a simple spring-driven blade located at the aperture stop. Thus, for the rear meniscus, the shutter mechanism is exposed to the environment; in the front meniscus, the lens acts as a protective window. Finally, and perhaps most important, in the front meniscus, the lens is out in front and quite visible to the customer, whereas in the rear meniscus, all the customer ever sees is the less appealing shutter mechanism. These latter "commercial" reasons are why the front-meniscus form has been universally used for inexpensive cameras since the 1940s. Apparently there is more to optical engineering than aberration correction.

At the start of this section we assumed that the lens would be thin and its surfaces spherical. If we increase the thickness of a meniscus lens and maintain its focal length at a constant value by adjusting one of the radii, it is apparent from the thick-lens focal-length equation (Eq. 2.28) that we must either reduce the power of the convex surface or increase the power of the concave surface to maintain the focal length as the thickness is increased. Either change will have the effect of reducing the inward Petzval curvature of field. This principle (i.e., separation of positive and negative surfaces, elements, or components in order to reduce the Petzval sum) is a powerful one and is the basis of all anastigmat designs.

The value of aspheric surfaces is limited in a design as simple as the box camera lens. However, if the lens is molded from plastic, an aspheric surface is as easy to produce as a spherical one; many simple cameras now have aspheric plastic objectives. The aspheric surface affords the designer additional freedom to modify the system to advantage. A diffractive surface could be used to achromatize the lens (and affect the other aberrations as well).

12.3 The Symmetrical Principle

In an optical system which is *completely* symmetrical, coma, distortion, and lateral color are identically zero. To have complete symmetry a system must operate at unit magnification and the elements behind the stop must be mirror images of those ahead of the stop. This is a principle of great utility, not only for systems working at unit power, but even for systems working at infinite conjugates. This is due to the fact that, although coma, distortion, and lateral color are not completely eliminated under these conditions, they tend to be drastically reduced when the elements of any system are made symmetrical, or even approximately so. For this reason many lenses which cover an appreciable field with low distortion and low coma tend to be generally symmetrical in construction.

If we were to apply this principle to the meniscus camera lens, we would simply use two identical menisci equidistant on either side of the stop. The resulting lens would be practically free of coma, distortion, and lateral color. The periscopic lens, shown in Fig. 12.4, makes use of this principle. Symmetry, plus the thick meniscus principle (to flatten the field) achieves a very remarkable astigmatic field coverage of ±67° for the Hypergon lens, which is also shown in Fig. 12.4. This is accomplished at the expense of a heavily undercorrected spherical aberration which limits its useful speed to about *f*/30 or *f*/20.

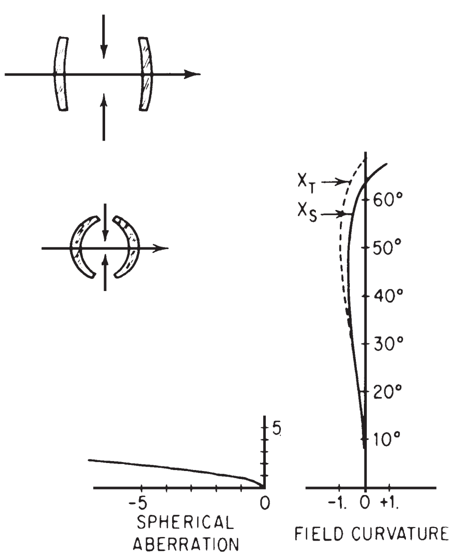


Figure 12.4 Symmetrical (simple) meniscus lenses. The upper sketch shows a periscopic-type lens composed of two identical meniscus lenses. The lower sketch shows the Hypergon (U.S. Patent 706,650-1902), whose nearly concentric construction allows it to cover a total field of 135° at *f*/30. The inner and outer radii of the Hypergon differ by only one-half percent, producing a very flat Petzval curvature. Aberrations shown are for a focal length of 100.

12.4 Achromatic Telescope Objectives (Thin-Lens Theory)

An achromatic doublet is composed of two elements, a positive crown glass element and a negative flint glass element. (Stated more generally, an achromatic doublet consists of a low-relative-dispersion element of the same sign power as the doublet and a high-relative-dispersion element of opposite sign.) As degrees of freedom we have the choice of glass types for the elements, the powers of the two elements, and the shapes of the two elements.

We assume here that we are designing a telescope objective, that the stop or pupil will be located at the lens, and that the lens will be thin. The astigmatism of a thin lens in contact with the stop is fixed, regardless of the number of elements, their index, or their shapes. Equation 10.8o indicates TAC = $(h^2\phi u'_k)/2$ for a single element. Since the power of a doublet is simply the sum of the powers of the elements, this equation applies to a doublet as well as a singlet. Thus we cannot affect the astigmatism (and can do very little about the Petzval curvature). The field will be strongly inward-curving.

With reference to Fig. 12.5, it is apparent that we have only four variable parameters with which to correct the aberrations. Actually, one parameter must always be assigned to control the focal length in any lens design. Thus we have three variables left; we will use them to correct spherical aberration, coma, and axial chromatic aberration.

Since the lens is to be free of chromatic aberration, we must assign the element powers to the determination of focal length and the control of chromatic aberration. Again we begin by using the thin-lens third-order aberration equations; assigning the subscripts a and b to the two elements, Eq. 10.8r gives us

$$\Sigma\text{TAchC} = \text{TAchC}_a + \text{TAchC}_b = \frac{Y_a{}^2\phi_a}{V_a u'_k} + \frac{Y_b{}^2\phi_b}{V_b u'_k}$$

Since the elements are to be cemented together or very nearly in contact, we can substitute $y_a = y_b = y$ and $u'_k = -y/f$ to get

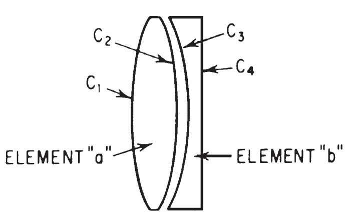


Figure 12.5 Achromatic doublet.

$$\Sigma \text{TAchC} = -fy\left[\frac{\phi_a}{V_a} + \frac{\phi_b}{V_b}\right] \tag{12.1}$$

We now set ΣTAchC = 0 (or some other value, if desired) and make a simultaneous solution of Eq. 12.1 with

$$\frac{1}{f} = \phi_a + \phi_b \tag{12.2}$$

to get the necessary powers for the elements. For zero chromatic, we get

$$\phi_a = \frac{V_a}{f(V_a - V_b)} \tag{12.3}$$

$$\phi_b = \frac{V_b}{f(V_b - V_a)} = \frac{-\phi_a V_b}{V_a} \tag{12.4}$$

Having determined ϕ_a and ϕ_b, we can now write thin-lens equations for the third-order spherical and coma in terms of the shapes of the elements [after tracing a marginal (thin-lens) paraxial ray to determine the values for u'_k of the combination and υ (or υ') for each element]. Since the aperture stop will be at the lens, $Q = 0.0$ and the coma will be given by Eq. 10.8n. After the appropriate substitutions for h, y, $C_a = \phi_a/(n_a - 1)$, $C_b = \phi_b/(n_b - 1)$, and the G-factors, we arrive at an equation of the following general form for coma:

$$\begin{aligned}\Sigma \text{CC} = \text{CC}_a + \text{CC}_b &= K_1C_1 + K_2 + K_3C_3 + K_4 \\ &= K_1C_1 + K_3C_3 + (K_2 + K_4)\end{aligned} \tag{12.5}$$

where C_1 and C_3 are the curvatures of the first surfaces of the elements (Fig. 12.5), and K_1 through K_4 are constants. (Note that by using the alternate form of Eq. 10.8n for element a, the equation could be written in C_2 and C_3, the curvatures of the adjacent inner surfaces). Now for any desired value of ΣCC, we find that

$$C_3 = \frac{\Sigma \text{CC} - K_1C_1 - K_2 - K_4}{K_3}$$

or, combining constants

$$C_3 = K_5C_1 + K_6 \tag{12.6}$$

Thus for any shape of element a, Eq. 12.6 indicates the unique shape for element b which will give the desired amount of coma.

In similar fashion we can write an expression for the thin-lens third-order spherical (using Eq. 10.8m) in the following form:

$$\Sigma \mathrm{TSC} = \mathrm{TSC}_a + \mathrm{TSC}_b = K_7 C_1^2 + K_8 C_1 + K_9 + K_{10} C_3^2 + K_{11} C_3 + K_{12} \quad (12.7)$$

By substituting the value for C_3 from Eq. 12.6 into 12.7, and combining constants, we get a simple quadratic equation in C_1 of the form

$$0 = C_1^2 + K_{13} C_1 + K_{14} \quad (12.8)$$

which can be solved for the value of C_1. When used with the value of C_3 given by Eq. 12.6, this will yield a doublet with spherical and coma of the desired amounts. (Note that because Eq. 12.8 is a quadratic, there may be one, two, or no solutions.)

For a first try, one would use the above procedure with ∑TAchC, ∑TSC, and ∑CC equal to zero (or whatever values are desired). Next, appropriate thicknesses are inserted, and the system tested by raytracing to determine the actual values of spherical, coma (or OSC), and axial color. If these are not within tolerable limits, the thin-lens solution can be repeated using (for the desired ∑TAchC, ∑TSC, and ∑CC) the negatives of the corresponding values determined by raytracing. This process converges to a solution very rapidly.

While the above procedure is useful in understanding the nature of the doublet telescope objective, a designer with an optical software computer program could handle this project very easily. The four surface curvatures would be declared as variables, and the merit function would consist of targets for the actual ray-traced values of marginal spherical aberration, coma, and chromatic aberration plus the effective focal length. Given a reasonable starting lens form, the task is trivial, and the nearest solution to the starting form is found immediately.

12.5 Achromatic Telescope Objectives (Design Forms)

Depending on the choice of glass, the relative aperture, the desired values of the aberrations, and also on which solution to the quadratic was selected, the procedure outlined in Sec. 12.4 will result in an objective with one of the forms sketched in Fig. 12.6. In general the edge contact form and, for lenses of modest (up to 3- or 4-in) diameter, the cemented form is preferred, primarily because the relationship between the elements (as regards mutual concentricity about the axis and freedom from tilt) can be more accurately maintained in fabrication. The crown-in-front forms are more commonly used because the

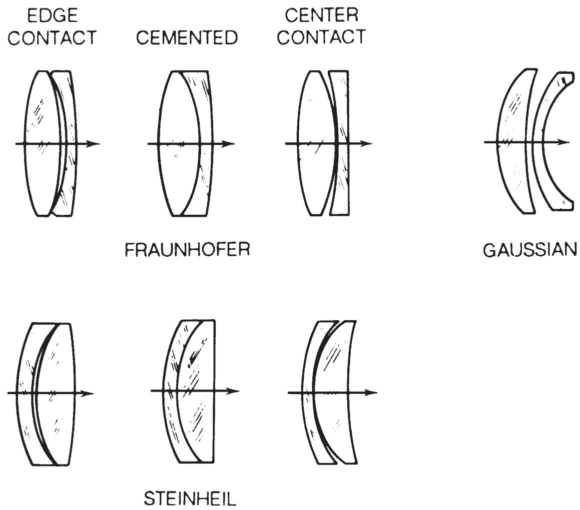


Figure 12.6 Various forms of achromatic doublets. The upper row are crown-in-front doublets and the lower row are flint-in-front. The curvatures are exaggerated for clarity. The center contact form is usually avoided because it is more difficult to manufacture. The shapes indicated are for lenses corrected for a distant object to the left.

front element is more frequently exposed to the rigors of weather; crown glasses are in general more resistant to weathering than flint glasses.

The Fraunhofer and Steinheil forms represent one root of the quadratic of Eq. 12.8, and the Gauss form is the other root. Whether one gets the Fraunhofer or the Steinheil form simply depends on whether the left-hand element is of crown or flint glass. From an image-quality standpoint, there is little difference between them. However the Gauss objective is very different. The Gauss lens has about an order-of-magnitude more zonal spherical aberration residual and slightly (about 20 percent) more secondary spectrum than the Fraunhofer. However, it has only about half the spherochromatism. Another difference is that there is no solution for the Gauss form if the lens elements are too thick; thus the speed is limited to about *f*/5 or *f*/7 to avoid thick elements. The Fraunhofer and Steinheil forms can be corrected at speeds faster than *f*/3 (although the residual aberrations are of course quite large at high speeds).

If one followed the procedure of Section 12.4, a design resulting in a cemented doublet (i.e., $C_2 = C_3$) would be a lucky accident. When a cemented interface is necessary, an alternate procedure is followed. The spherical and coma contribution equations are written in C_2 and C_3 (instead of C_1 and C_3) and C_2 is set equal to C_3, resulting in equations in C_2 (or C_3) which may then be solved for either the desired coma or spherical. If these equations are plotted as a function of the shape of the doublet (i.e., versus C_1 or C_2 or C_4) the resulting graph will look like one of those in Fig. 12.7, in which $\sum$TSC is a parabola and $\sum$CC is a straight line. In the left plot there is no solution for spherical, in the center plot the solutions for spherical and coma occur at the same bending, and on the right there are two possible solutions for spherical with equal and opposite-signed amounts of coma, and often with pronounced meniscus shapes. (These latter solutions are valuable if one desires to utilize the doublets in a symmetrical combination about a central stop, e.g., as an erector or a rapid rectilinear photo lens; the coma can then be used to reduce or overcorrect the astigmatism per Eq. 10.8h.) The exact form obtained is dependent primarily on the types of glass chosen. In general, the spherical aberration parabola can be raised by selecting a new flint glass with a lower index and higher *V*-value, or by selecting a new crown with a higher index and lower *V*. Thus the strongly meniscus solutions of the right-hand plot in Fig. 12.7 result from a glass pair with a small difference in *V*-value. Results approximating those in the middle graph of Fig. 12.7 can be obtained with BK7 (517:642) and SF2 (648:339). The best glass choice depends on the aperture (*f*/#) of the lens.

Figure 12.8 shows the spherical aberration and the spherochromatism of a typical cemented doublet. As previously noted, the field curvature of a thin system with stop in contact is strongly inward and cannot be modified unless the stop is shifted. Thus, systems of this type are limited to applications which require good imagery over relatively small fields (a few degrees from the axis).

It is occasionally desirable to produce a doublet objective with both the zonal and marginal spherical simultaneously corrected. This can be accomplished by using the airspace of a broken contact doublet as an added degree of freedom. The design is begun exactly as in Sec. 12.4, except that two (or more) thick-lens solutions are derived, one

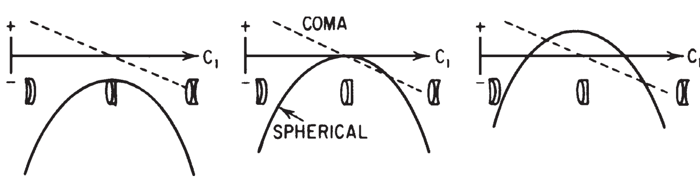


Figure 12.7 The variation of spherical aberration (solid line) and coma (dashed) as a function of the shape of a cemented achromatic doublet. Depending on the materials used there may be two forms with zero spherical (right), one form (center), or no form (left). The center graph is the preferred type since spherical and coma are both corrected.

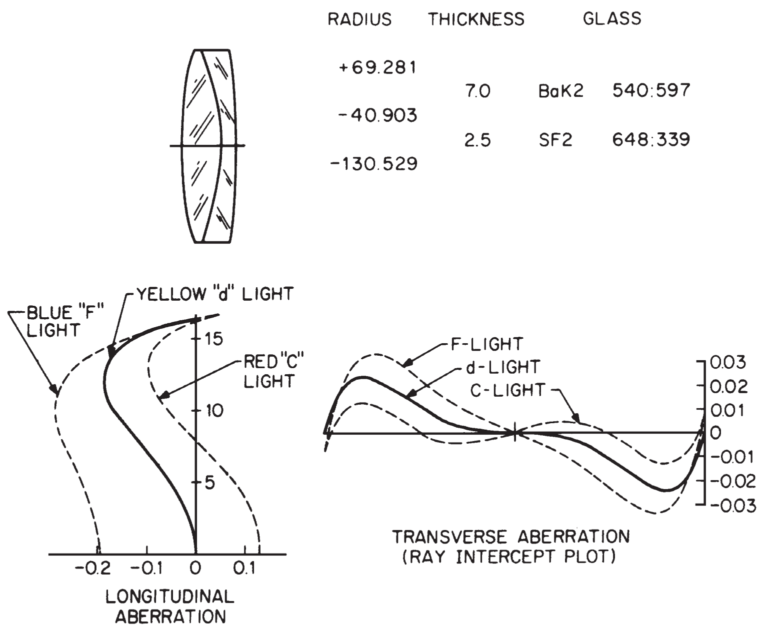


Figure 12.8 The spherical aberration and spherochromatism of a cemented achromatic doublet, efl = 100, *f*/3.0. Note that the chromatic is corrected at the margin. This is good practice if the spherochromatism is large; otherwise the image shows a blue flare. For small amounts, correction at the 0.7 zone is often a better choice.

with a minimum airspace and the other(s) with an increased space. The calculated zonal spherical is then plotted against the size of the airspace, and the airspace with LA_z equal to zero is selected; this form will usually have no zonal OSC. Speeds of *f*/6 or *f*/7 can be attained with practically no spherical or axial coma over the entire aperture. Good glass choices are a light barium crown combined with either a dense flint or an extradense flint; either crown-in-front or flint-in-front forms are possible. In this type of lens the residual axial aberration consists almost solely of secondary spectrum.

Spherochromatism, which is the variation of spherical aberration as a function of wavelength, can be corrected by a change in the spacing between elements (or components) which differ in the sign of their contributions to spherical and chromatic aberration. This general principle may be applied to the doublet achromat in a manner paralleling the use of the airspace to correct zonal spherical; indeed, the basic principle is the same for both aberrations.

The source of spherochromatism can be understood by realizing that (in a cemented doublet) the two outer surfaces contribute undercorrected spherical aberration, while the cemented interface contributes overcorrected spherical. The amount of the contribution varies directly with the size of the index change, or "break," across the surface. The contributions are in balance for the nominal wavelength. At a shorter wavelength all the indices are higher; because of its greater dispersion, the index of the negative (flint) element increases about twice as rapidly as that of the positive (crown) element. The index break at all three surfaces is larger at the shorter wavelength. However, the index break at the outer surfaces is $(n - 1)$, whereas at the cemented surface it is $(n' - n)$; as the wavelength and the indices change, $(n' - n)$ changes proportionately more than does $(n - 1)$. Thus as we go to a shorter wavelength, the overcorrecting contribution of the cemented surface is increased more than the undercorrection from the outer surfaces. The result is that the short-wavelength light is overcorrected compared to the central or longer wavelength. This is spherochromatism.

Now, if the airspace between elements is increased, as indicated in Fig. 12.9, the blue marginal ray, having been refracted more strongly than the red ray by the crown element, will strike the flint element at a lower height than will the red ray. Thus the refraction of the blue ray at the flint will be lessened relative to the red, and its overcorrection reduced accordingly.

A very similar argument can be applied to the reduction of an undercorrected *zonal spherical* (which is caused by an overcorrected fifth-order spherical) by use of an increased airspace. The increased airspace affects the zonal spherical because the undercorrected spherical of the positive element bends the marginal ray toward the axis disproportionately more than the zonal ray. Thus, when the airspace is increased, the ray height at the overcorrecting negative element is reduced proportionately more for the marginal ray than for the zonal ray. The result is that the overcorrection is reduced more at the margin than at the zone, and, when the element shapes are readjusted to correct the marginal aberration, the zonal spherical is reduced. An airspaced doublet with reduced spherochromatism and reduced zonal spherical is shown in Fig.

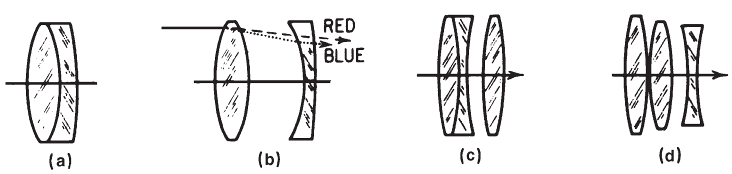
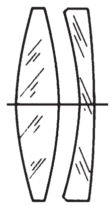


Figure 12.9 The ordinary spherochromatism of a doublet can be corrected by increasing the airspace [shown highly exaggerated in (b)]. This reduces the height at which the blue ray strikes the flint by a greater amount than for the red ray, thus reducing the overcorrection of the marginal blue ray. Sketches (c) and (d) show triplet forms which can be used to correct spherochromatism and spherical zonal residuals simultaneously.



RADIUS	THICKNESS	GLASS
+47.819		
	7.0	BK7 517:642
−56.354		
	3.853	
−46.852		
	2.5	SF1 717:295
−160.303		

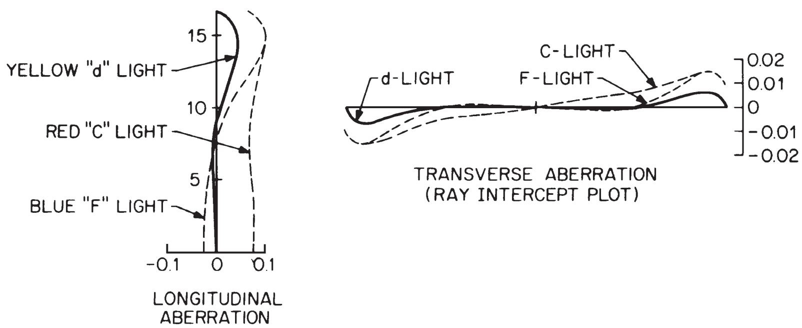


Figure 12.10 The spherical aberration and spherochromatism of an airspaced achromatic doublet, efl = 100, *f*/3.0. The size of the airspace used here is a compromise between the value which would minimize the zonal spherical aberration and that which would minimize the spherochromatism. Compare the residual aberrations with those of the cemented doublet in Fig. 12.8.

12.10. *Both principles are applicable to more complex lenses as well.* Figure 12.10 shows an example of these principles.

One method of effecting a simultaneous elimination of both spherochromatism and zonal spherical is indicated in Fig. 12.9c. The doublet plus singlet configuration (in any of several arrangements of the elements) introduces still another degree of freedom, namely the balance of positive (crown) power between the two components, which can be used with the airspace to bring about the correction. The airspaced triplet shown in 12.9d is also capable of very good correction, but is more difficult to manufacture. Figure 13.53 illustrates the reduction of spherical by element splitting. Figure 14.3 shows an airspaced triplet telescope objective.

The *secondary spectrum* (SS) contribution of a thin lens is given by Eq. 10.8t; combining this with the requirements for achromatism (Eqs. 12.3 and 12.4), we find that the secondary spectrum of a thin achromatic doublet is given by

$$\mathrm{SS} = \frac{f\,(P_b - P_a)}{(V_a - V_b)} = \frac{-f\,\Delta P}{\Delta V} \tag{12.9}$$

For any of the ordinary glass combinations used in doublets, the ratio $\Delta P/\Delta V$ is essentially constant, and the visual secondary spectrum is about 0.0004 to 0.0005 of the focal length. Similarly, the secondary spectrum of any achromatized combination of two separated components can be shown to be

$$SS = \frac{\Delta P}{D\,\Delta V}\,[f^2 + B\,(L-2f)\,] \tag{12.10}$$

where D is the airspace, B the back focus, and $L = B + D$ is the length from front component to the focal point. Again it is apparent that the ratio $\Delta P/\Delta V$ is the governing factor. Note that in this case the secondary color of two spaced positive lenses is less than that of a thin doublet of the same focal length; conversely, the secondary color of a telephoto lens (positive front component, negative rear component) or reversed telephoto is greater than the corresponding thin doublet.

There are a few glasses which will reduce the secondary spectrum, for example, FK51, 52, or 54 used with a KzFS glass or an LaK glass as the flint element will reduce the visual secondary spectrum to a small fraction of the ordinary value. Note, however, that for most of these pairs $V_a - V_b$ is small, and the powers of the individual elements required for achromatism are higher than with an ordinary pair of glasses. This increase in element power causes a corresponding increase in the other residual aberrations. These glasses, with unusual partial dispersions, generally work poorly in the shop, lack chemical stability, and cannot withstand severe thermal shock.

As mentioned in Chap. 7, calcium fluoride (CaF_2, fluorite) may be combined with an ordinary glass (selected so that $P_a = P_b$) to make an achromat that is essentially free of secondary spectrum. It is also worth noting that there are no ordinary glass pairs which will form a useful achromat in the 1.0- to 1.5-μm spectral band; fluorite can be combined with a suitable glass to make an achromat for this region. Silicon and germanium are useful for achromats at longer wavelengths, as are BaF_2, CaF_2, ZnS, ZnSe, and AMTIR.

A triplet achromat can be used to reduce the secondary spectrum without the necessity of exactly matching the partial dispersions as in the doublet. If one plots the partial dispersion P against the V-value, most glasses fall along a straight line. What is needed to correct secondary spectrum is a pair of glasses with the same partial P, but with a significant difference in V-value. It turns out that in this sort of plot one can synthesize a glass anywhere along a line connecting two glass points by making a doublet of the two glasses. Thus one can arrange a triplet so that two of the elements synthesize a glass with exactly the same partial as the third glass. Some useful Schott glass combinations are (PK51, LaF21, SF15), (FK6, KzFS1, SF15), (PK51, LaSFN18,

SF57); the power arrangement for these combinations is plus, minus, and weak plus, respectively. Other glass manufacturers have equivalent glass combinations. The thin lens element powers for a triplet apochromat can be found from the following equations, which are for a unit power ($f = 1.0$) system.
Define:

$$X = V_a (P_b - P_c) + V_b (P_c - P_a) + V_c (P_a - P_b)$$

Then:

$$\phi_a = V_a (P_b - P_c) / X$$

$$\phi_b = V_b (P_c - P_a) / X$$

$$\phi_c = V_c (P_a - P_b) / X = 1.0 - \phi_a - \phi_b$$

See Fig. 14.2 for an example of an apochromatic triplet telescope objective.

A lens in which three wavelengths are brought to a common focus is called an *apochromat.* Often this term also implies that the spherical aberration is corrected for two wavelengths as well. By properly balancing the glass combinations given above one can achromatize the triplet for four wavelengths; such lenses are called *superachromats.**

Airspaced achromat (dialyte)

A widely airspaced doublet can be made achromatic, but the chromatic correction will vary with the object distance; it will be achromatic only for the design distance. The following equations will yield a separated achromatic doublet which is corrected for an object at infinity.

$$\phi_A = \frac{V_A B}{f(V_A B - V_B f)}$$

$$\phi_B = \frac{-V_B f}{B(V_A B - V_B f)}$$

$$D = \frac{(1 - B/f)}{\phi_A}$$

where f is the focal length, D is the airspace, and B is the back focal length.

*See M. Herzberger and N. McClure, "The Design of Superachromatic Lenses," *Applied Optics,* vol. 2, June 1963, pp. 553–560.

Athermalization

When the temperature of a lens element is changed, two factors affect its focus or focal length. As the temperature rises, all the dimensions of the element are increased; this, by itself, would lengthen the effective and back focal lengths. The index of refraction of the lens also changes with temperature. For many glasses the index rises with temperature; this effect tends to shorten the focal lengths.

The change in the power of a thin element with temperature is given by

$$\frac{d\phi}{dt} = \phi\left[\frac{1}{(n-1)}\frac{dn}{dt} - \alpha\right]$$

where dn/dt is the differential of index with temperature, and α is the thermal expansion coefficient for the lens material. Thus for a thin doublet

$$\frac{d\Phi}{dt} = \phi_A T_A + \phi_B T_B$$

where

$$T = \left[\frac{1}{(n-1)}\frac{dn}{dt} - \alpha\right]$$

and Φ is the doublet power. For an athermalized doublet (or for one with some desired $d\,\Phi/dt$), we can solve for the element powers

$$\phi_A = \frac{(d\Phi/dt) - \Phi T_B}{T_A - T_B}$$

$$\phi_B = \Phi - \phi_A$$

To get an athermalized *achromatic* doublet, we can plot T against $1/V$ for all the glasses under consideration. Then a line drawn between two glass points is extended to intersect the T axis. The value of the $d\Phi/dt$ for the achromatic doublet is equal to the doublet power times the value of T at which the extended line intersects the T axis. Thus one desires a pair of glasses with a large V-value difference and a small or zero T-axis intersection.

An athermal achromatic triplet can be made with three glasses as follows:

$$\phi_A = \frac{\Phi V_A\,(T_B V_B - T_C V_C)}{D}$$

$$\phi_B = \frac{\Phi V_B\,(T_C V_C - T_A V_A)}{D}$$

$$\phi_C = \frac{\Phi V_C\,(T_A V_A - T_B V_B)}{D}$$

where $D = V_A (T_B V_B - T_C V_C) + V_B (T_C V_C - T_A V_A) + V_C (T_A V_A - T_B V_B)$, V_n is the V-value of element n, and T is defined above.

12.6 The Diffractive Surface in Lens Design

A *diffractive surface* as used in lens design is a *fresnel* surface (as shown in Fig. 9.15) "modulo 2π." In other words, it is a fresnel surface where the height of each step is such that the wave front is retarded or stepped by exactly one wavelength. Thus the step height is $\lambda/(n - 1)$, assuming that the surface is bounded by air. For a glass or plastic surface ($n \approx 1.5$), this is a step height of about two wavelengths, as opposed to a step height on the order of tenths of a millimeter or more for an ordinary plastic fresnel. The slope and shape of the fresnel facets can be as defined by a sphere or an aspheric. Note that similar results can be obtained with a local variation of the index of refraction.

Diffraction efficiency

The term *kinoform* indicates a surface with smooth facets. A curved-surface kinoform theoretically can have 100 percent efficiency. A "linear" (cone-shaped) kinoform can be 99 percent efficient. A "binary" surface approximates the smooth fresnel facets with a stair-step contour produced by a high-resolution photolithographic process. The surface relief is created by exposure through a series of masks. The number of levels produced equals 2^n, where n is the number of masks used, hence the name *binary.* The efficiency (i.e., the percentage of light which goes in the desired direction) of a binary surface is limited by the number of levels which are used to approximate the ideal smooth contour of the fresnel facet. A one-mask, 2-level surface is 40.5 percent efficient; a two-mask, 4-level surface is 81.1 percent efficient; a three-mask, 8-level surface is 95.0 percent efficient; a four-mask, 16-level surface is 98.7 percent efficient; and an M-level surface is $[\sin(\pi/M)/(\pi/M)]^2$ efficient. The theoretical efficiency of any diffraction surface, whether kinoform or binary, will be reduced by any fabrication departures from the ideal shape, such as rounding of sharp corners, etc.

Since the wave front is stepped or retarded at each diffractive fresnel step by exactly one wavelength for the nominal wavelength, it is apparent that the coherent behavior of the system is preserved only for the nominal wavelength. At this wavelength, the phase from the top of one zone exactly matches that from the bottom of the preceding zone. The surface is less efficient for other wavelengths, and thus the spectral bandwidth over which a diffractive surface is useful is limited. This limitation may show up as inefficiency or as unwanted diffractive orders, ghosts, stray light, low contrast, etc. The efficiency at other than the nominal wavelength (λ_0) is

$$E = [\sin \pi (1 - \lambda_0/\lambda) / \pi (1 - \lambda_0/\lambda)]^2$$

Over a bandwidth of $(\Delta\lambda)$, the average efficiency is

$$\text{ave } E \approx 1 - [\pi (\Delta\lambda) / 6 \lambda_0]^2$$

Manufacturability

The following expressions allow an estimate of the practicality or manufacturability of a diffractive lens. As indicated above, the step height is $\lambda/(n - 1)$. The radial spacing distance from one fresnel step to the next is approximately

$$\text{Spacing} \approx R\lambda/Y (n - 1) = F\lambda/Y$$

where R is the diffractive surface radius of curvature, F is its focal length, and Y is the radial distance from the axis. The minimum spacing (at the edge of the diffractive lens) is

$$\text{Min spacing} \approx 2\lambda (f/\#) = \lambda/\text{NA}$$

where $f/\# = F/2Y_{max}$ = the relative aperture, and NA = $n \sin u$ = the numerical aperture. The total number of fresnel steps or zones is

$$\text{Number of steps} \approx D^2/8\lambda F$$

where D is the lens diameter. It is apparent that the longer the wavelength and the weaker the power of the diffractive surface, the wider and deeper are the steps, and the easier is the fabrication task. Techniques used for fabrication include single-point diamond turning (especially good for long-wave IR), ion-beam machining, electron-beam writing, laser-beam writing, and photolithography (which is extremely difficult on curved surfaces but effective on plano surfaces). For large commercial quantities, injection-molded plastic elements are an economical choice. Another useful process is epoxy replication. Applications of diffractive optics include hybrid (combined refractive and diffractive) lenses, microlens (size about 50 μm) arrays, anamorphic arrays, prisms, beamsplitters, beam multiplexers, filters, etc.

The Sweatt model

From a lens design standpoint, an easy way to handle and understand the use of a diffractive surface is through the *Sweatt model*. W. C. Sweatt* showed that a raytrace model consisting of a very high index,

---

**J. Opt. Soc. Am.*, vol. 67, 1977, p. 80; vol. 69, 1979, p. 486; *Appl. Opt.*, vol. 17, 1978, p. 1220.

zero-thickness lens could be used to predict the effect of a diffractive surface; the higher the index, the closer the results of the raytrace come to matching the exact diffraction results. An index of about 10,000 is a reasonable value to use. Since the diffractive effect is a direct function of wavelength, the index of the model should vary as the wavelength, and

$$n\ (\lambda) = 1 + (n_0 - 1)\ (\lambda/\lambda_0)$$

where λ_0 and n_0 are the nominal wavelength and index, respectively.

Thus, for the visual region, using *d, F,* and *C* light, we have for

d-light at 0.5875618 μm,

$$n_d = 10{,}001.00$$

F-light at 0.4861327 μm,

$$n_F = 8{,}274.73$$

C-light at 0.6562725 μm,

$$n_C = 11{,}170.42$$

and the Abbe *V*-value,

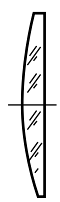
$$V = (n_d - 1) / (n_F - n_C) = -3.45$$

The negative *V*-value results from the fact that the index rises with wavelength instead of dropping as in ordinary refractive materials. The partial dispersion is $P = (n_F - n_d)/(n_F - n_C) = 0.5962$. These extremely unusual values make the diffractive surface a most singular optical material. This low-V-value (i.e., high dispersion) characteristic of a diffractive device indicates that there will be very large amounts of chromatic aberration when a diffractive surface is used over a significant spectral bandwidth.

The achromatic diffractive singlet

If we assume a single element of BK7 ($n_d = 1.5168$, $V = 64.2$, $P = 0.6923$), we can apply Eqs. 12.3 and 12.4 to determine the powers of the singlet and the diffractive element which will produce an achromat. The result is a power of $\phi_a = V_a\Phi/(V_a - V_b) = +0.949\Phi$ for the BK7 element and $\phi_b = +0.051\Phi$ for the diffractive element (where Φ is the desired power of the achromat). The negative *V*-value of the diffractive surface produces an achromat where both elements have positive power. If we allow the diffractive surface to be aspheric (in the actual surface this is done by making the slope and shape of the fresnel facets correspond to those of an aspheric surface), we can produce a singlet of the desired power which is corrected for spherical aberration, chromatic aberration, and coma. The necessary four degrees of freedom are the power and bending of the singlet and the power and fourth-order asphericity (or conic constant) of the diffractive surface.

The resulting design is shown in Fig. 12.11. The residual aberrations (zonal spherical, spherochromatism, and secondary spectrum) can be compared with those of the ordinary achromatic doublet shown in Fig. 12.8. Note that the sign of the secondary spectrum is reversed from that of an ordinary doublet (because of the unusual *P* and *V* of the diffractive surface) and that the spherochromatism is large, more than twice that of the doublet of Fig. 12.8 (and is also of reversed sign). The spherochromatism can be corrected by aspherizing the first surface with a fourth-order deformation term in a manner analogous to adjusting the airspace of the doublet in Fig. 12.10 (i.e., we change the relative heights at which the red and blue rays strike the diffractive surface). The zonal spherical can be removed with a sixth-order deformation term on the first surface. The use of an aspheric surface is an



RADIUS	THICKNESS	MATERIAL		
59.307733	4.50	BK7	1.5168	64.17
−972.261092	0.00	DIFF	10,001.	−3.45
−972.186542*	97.415	AIR		

*Conic constant kappa = −0.186743

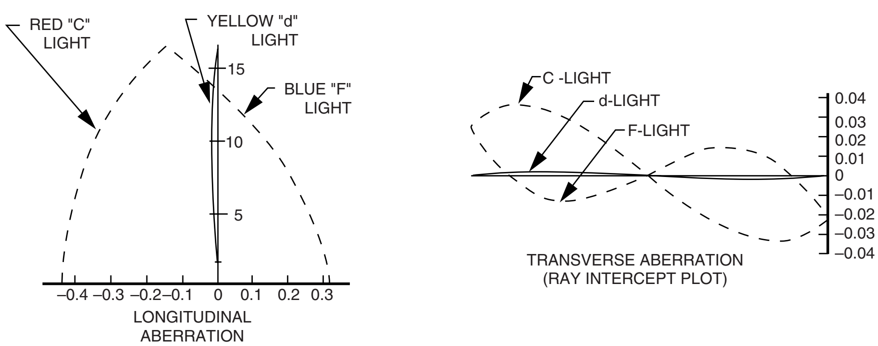


Figure 12.11 The spherical aberration and spherochromatism of a hybrid refractive-diffractive singlet, efl = 100, *f*/3.0. Compare with the doublet of Fig. 12.8 (but note that the scales for LA are different). Both the spherochromatism and secondary spectrum are larger and of the opposite sign from Fig. 12.8. As indicated in the text, the spherochromatism and the zonal spherical can be eliminated easily by aspherizing the first surface (which would be quite a feasible option if the lens were injection-molded from acrylic).

economically practical move, assuming that the lens is to be injection-molded from plastic. The result is a lens whose only axial aberration is about 0.17 mm of secondary spectrum.

Alternately, because photolithographic fabrication is most conveniently done on a flat surface, one might want to limit the lens shape to a plano-convex form and use as degrees of freedom the lens index, its radius, the power of the diffractive surface, and its asphericity. The optimal index is about 1.55 for this lens. If the lens material is acrylic ($n = 1.492$), and if we elect to control focal length, spherical, and chromatic (neglecting coma), the tangential coma at one degree off axis is -0.0156; if the material is polystyrene ($n = 1.590$), it is $+0.0101$.

Achromatic diffractive singlets have been very satisfactorily used in eyepieces, magnifiers, zoom camera lenses, and many other applications where the object field is of relatively uniform brightness. Their compactness and light weight as compared with a glass achromat make them very desirable for many applications such as head-mounted displays. Diffractive surfaces sometimes have proven less satisfactory for systems where there is a high brightness source in (or near) the field of view or a wide spectral bandwidth.

The apochromatic diffractive doublet

Since the unusual V-value and partial dispersion of the diffractive surface are so far from the line of normal glasses in a P versus V plot, we can easily produce an apochromatic lens using two ordinary glasses plus a diffractive surface to eliminate the secondary spectrum.

The element powers for a three-element apochromat can be found using the following equations:

$$X = V_a (P_b - P_c) + V_b (P_c - P_a) + V_c (P_a - P_b)$$

$$\phi_a = \Phi V_a (P_b - P_c) / X$$

$$\phi_b = \Phi V_b (P_c - P_a) / X$$

$$\phi_c = \Phi V_c (P_a - P_b) / X$$

where Φ is the power of the apochromatic triplet, V_i is the V-value, and P_i is the partial dispersion of the ith element.

If we use acrylic ($n = 1.4918$, $V = 57.45$, $P = 0.7014$) and polystyrene ($n = 1.5905$, $V = 30.87$, $P = 0.7108$) for elements a and b, and the diffractive surface ($n = 10{,}001$, $V = -3.45$, $P = 0.5962$) for element c, we get the following starting powers for the elements:

$$\phi_a = +1.9544\Phi \qquad \text{(acrylic)}$$

$$\phi_b = -0.9640\Phi \qquad \text{(polystyrene)}$$

$$\phi_c = +0.0096\Phi \qquad \text{(diffractive)}$$

The lens can be corrected for marginal and zonal spherical aberration, coma, chromatic, spherochromatic, and secondary spectrum using the techniques outlined above. A drawback for this particular lens is that the secondary spectrum varies with aperture and is corrected only at one zone.

12.7 The Cooke Triplet Anastigmat

The Cook triplet is composed of two outer positive crown elements and an inner flint element, with relatively large airspaces separating the elements. This type of lens is especially interesting because there are just enough available degrees of freedom to allow the designer to correct all of the primary aberrations. The basic principle used to flatten the field curvature (i.e., the Petzval sum) is quite simple: the contribution that an element makes to the power of a system is proportional to $y\phi$, and the contribution to the chromatic varies with $y^2\phi$. However, the contribution to the Petzval curvature is a function of ϕ alone and is independent of y. Now in a thin (compact) system, all the elements have essentially the same value of y and the powers of the elements are determined by the requirements of focal length and chromatic correction; consequently, the Petzval radius of a thin doublet is often about $-1.4f$, and rarely exceeds 1.5 or 2 times the focal length. However, when the negative elements of a system are spaced away from the positive elements (so that the ray height y at the negative elements is reduced), the power of the negative elements must be increased to maintain the focal length and chromatic correction of the system. As a result, the overcorrecting contribution of the negative element to the Petzval curvature is increased. Thus by the proper choice of spacing, the Petzval radius can be lengthened to several times the system focal length and the field proportionately "flattened."

From Fig. 12.12, which shows a schematic triplet, we can determine the available degrees of freedom. They are

1. Three powers (ϕ_a, ϕ_b, ϕ_c)
2. Two spaces (S_1, S_2)
3. Three shapes (C_1, C_3, C_5)
4. Glass choice
5. Thicknesses

Of these, items 1, 2, and 3 will be of immediate interest; they total eight variables. Item 4, glass choice, is an extremely important tool, but we will reserve its discussion until later. Item 5, element thickness, is only

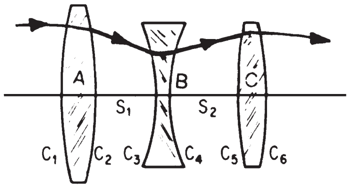


Figure 12.12 The Cooke triplet anastigmat.

marginally effective; in regard to the primary corrections, its effect duplicates that of the spacings.

With these eight degrees of freedom, the designer wishes to correct (or control) the following primary characteristics and aberrations.

1. Focal length
2. Axial (longitudinal) chromatic aberration
3. Lateral chromatic aberration
4. Petzval curvature
5. Spherical aberration
6. Coma
7. Astigmatism
8. Distortion

Thus, there are just the necessary eight degrees of freedom to control the eight primary corrections.

Note that the fact that there are eight variable parameters does not guarantee that there is a solution. The relationships involved are, in several instances, nonlinear, as the thin-lens equations (Eqs. 10.8) indicate. It is entirely possible to choose a set of desired aberration values and/or glass types for which there is no solution. On the other hand, it is also possible that there are as many as eight solutions, as will be seen in the following paragraphs.

Power and spacing solution. The first four items listed immediately above can be seen (by reference to the thin-lens third-order aberration equations) to be functions of element power and ray height (which is a function of spacing); they are independent of element shape. Thus, it is necessary that the powers and spaces be chosen to satisfy these four conditions, which may be expressed as follows:

Power:

$$\text{Desired } \Phi = \frac{1}{f} = \frac{1}{y_a} \sum y\phi \tag{12.11}$$

Axial color:

$$\text{Desired } \Sigma\text{TAchC} = \frac{1}{u'_k} \Sigma \frac{y^2\phi}{V} \tag{12.12}$$

Lateral color:

$$\text{Desired } \Sigma\text{TchC*} = \frac{1}{u'_k} \Sigma \frac{yy_p\phi}{V} \tag{12.13}$$

Petzval sum:

$$\text{Desired } \Sigma\text{PC} = \frac{h^2}{2} \Sigma \frac{\phi}{n} \tag{12.14}$$

where the summation is over the three elements. These expressions are essentially the same as those of Sec. 10.9, and the meanings of the symbols are given there.

The four conditions above must be satisfied by the choice of five variables (three powers plus two spacings). There is one more variable than necessary; this "extra" is utilized in a later step to control one of the remaining aberrations (usually distortion). There are almost as many ways of solving this set of equations as there are designers. Stephens* has worked out the algebraic solution for the triplet, and his paper gives explicit equations for the values of the powers and spaces. An iterative approximation technique (which may be easily modified to apply to systems with more than three components) along the following lines is an alternate method, and its description will help to understand the limits and interrelationships involved in this design.

1. Assume a value for the ratio of the powers of elements c and a. This will be the "extra" degree of freedom mentioned above. ($K = \phi_c/\phi_a = 1.2$ is a typical value.)
2. Choose a value (arbitrary) for ϕ_a. (In the absence of prior experience, $\phi_a = 1.5\Phi$ is suitable.) This determines ϕ_c, since from step 1, $\phi_c = K\phi_a$ and also determines ϕ_b, since Eq. 12.14 can be solved for ϕ_b when ϕ_a, ϕ_c, h, and ΣPC are known or assumed.
3. Choose a value for S_1 (one-fifth to one-tenth of the focal length is suitable).
4. Solve for the value of S_2 which will satisfy Eq. 12.12 (assume that u'_k is equal to Φy_a). This is done by tracing a ray through elements a and b to determine y_a, y_b, and u'_b. Then find S_2 to yield the value of y_c, which satisfies Eq. 12.12. (Note that S_2 may have a negative value on the first try.)

*R. E. Stphens, *J. Opt. Soc. Am.,* vol. 38, 1948, p. 1032.

5. Trace a principal ray (thin-lens paraxial) through the desired stop position, which may be conveniently placed at element b to minimize the labor. Again assume u'_k as in 4 and determine ΣTchC*.
6. Repeat from step 3 with a new choice for S_1 until ΣTchC* is as desired. (As a second guess for S_1, try the average of S_1 and S_2 from the first try.)
7. Determine the system power Φ. If not as desired, scale the value of ϕ_a used in step 2 and repeat from step 2 until a solution is obtained.

Graphs of the relationships between S_1 and ΣTchC* and between ϕ_a and Φ are useful in steps 6 and 7.

Element shape solution. When the element powers and spacings have been determined, there are three uncommitted degrees of freedom, namely the shapes of the three elements (plus the "extra," K, mentioned in step 1 above). These variables must be adjusted so that the spherical, coma, astigmatism, and distortion are corrected to their desired values. Referring to the thin-lens contribution equations of Sec. 10.9, the aberrations can be seen to be quadratic functions of the element shapes; thus, a simultaneous algebraic solution cannot be used and some sort of successive approximation procedure is necessary.

Thin-lens paraxial marginal and principal rays are traced through the three elements. The principal ray is traced so that the aperture stop is at lens b; both y_p and Q for lens b will be zero.

1. Assume an (arbitrary) value for C_1 and calculate TAC^*_a (the astigmatism contribution) for element a by Eq. 10.8h (a value of $C_1 = 2.5\Phi$ is a reasonable first choice).
2. Since the stop is located at element b, TAC_b will not change with bending (Eq. 10.8o). Now solve Eq. 10.8h for the shape of element c, that is, the value of C_5, which will give TAC^*_c which will yield the desired ΣTAC when combined with AC^*_a and AC_b. Normally there are two solutions for C_5 and the more reasonable one is used.
3. Now CC^*_a and CC^*_c (the coma contributions) are calculated from Eq. 10.8g. Since the equation for CC_b is linear in C_3 (Eq. 10.8n, since $Q_b = 0$), it can be solved for the unique value of C_3 which will yield the desired ΣCC*.
4. The value of ΣTSC (spherical aberration) is now determined from Eq. 10.8m.
5. The procedure is repeated from step 1 with a new value of C_1, and a graph of ΣTSC is plotted against C_1. The shape C_1 for which ΣTSC is equal to the desired value is chosen and the corresponding values

of C_3 and C_5 are determined so that $\sum$TSC, $\sum$TAC*, and $\sum$CC* are simultaneously as desired.

6. If $\sum$DC* (distortion) is within acceptable limits, well and good; if not, a new power and space solution must be made with a different value of $K = \phi_c/\phi_a$. The value of $\sum$DC* can be plotted for several values of K as an aid in effecting a solution.

Note that in step 5, there may be two, one, or no solutions for the desired $\sum$TSC. The best triplets seem to result from cases where the parabolic plot of $\sum$TSC just barely reaches the desired level. Step 6 also may have no, one, or two solutions. Thus, with two possible solutions in each of steps 2, 5, and 6, there are, theoretically at least, eight possible solutions for the thin-lens Cooke triplet. As indicated above, it is also possible that for a given set of conditions, there may be no solution. Usually, however, there is only one "reasonable" solution; occasionally there are two.

The next step is the addition of thickness to the design. Center thicknesses for the crown elements are chosen to give workable edge thicknesses; the second surface curvatures (C_2, C_4, and C_6) are adjusted to hold the thick-element powers exactly to the thin-lens powers. Airspaces are chosen so that the principal points of the elements are spaced apart by the thin-lens spacings. In this way, the thick-lens triplet will have exactly the same focal length as the thin-lens version.

The thick lens is now submitted to a trigonometric raytrace analysis and the values of the seven primary aberrations are determined. If (as is likely) the aberrations are not as desired, a new round of design is initiated, with the new "desired" thin-lens aberration values adjusted to offset the difference between the raytracing results and the desired final values. For example, if the original "desired" $\sum$TSC was -0.2 and the raytracing yielded a marginal spherical, $TA_m = +0.2$, the new "desired" $\sum$TSC would be set at -0.4, assuming that the desired end result was $TA_m = 0.0$

Initial choice of desired aberration values. In general, the initial choice for the "desired" third-order aberration sums should be small, undercorrected amounts, since the higher-order aberrations are usually overcorrecting. Spherical, Petzval, and axial chromatic follow this rule. Since the Cooke triplet is relatively symmetrical, the residuals of distortion, coma, and lateral color are small, and initial "desired" values of zero are appropriate. The desired Petzval sum should be definitely negative. For high-speed lenses, the Petzval radius is frequently as short as two or three times the focal length; moderate-aperture systems (*f*/3.5) usually have $\rho = -3f$ to $-4f$; slow systems may have $\rho = -5f$ or longer. One reason for this relationship is that the flatter (less undercorrected) the Petzval surface is made, the higher the element powers; hence the higher the residual aberrations, especially zonal spherical. The value chosen for the desired ∑PC is also an important factor in determining whether or not there is a solution for step 5 in the curvature determination process. The "desired" astigmatism sum is best set slightly positive, between zero and about one-third the absolute value of the Petzval sum, so that the inward curvature of the Petzval surface is offset by the overcorrected astigmatism.

Glass choice. The choice of the glass to be used in the triplet is one of the most important design factors. From field (Petzval) curvature considerations, it is desirable that the positive elements have a high index of refraction and the negative element a low one to reduce $\sum\phi/N$. As usual, the *V*-value of the positive elements should be high and that of the negative element low in order to effect chromatic correction. For the positive elements, one of the dense barium crowns is the usual choice, although the light barium crowns on one hand and the rare earth (lanthanum) glasses on the other are frequently used. Although triplet designs are possible with ordinary crown glass or even plastics, their performance is relatively poor.

It turns out that the interrelated requirements of Eqs. 12.11 through 12.14 lead to long systems (i.e., S_1 and S_2 are large) when the difference between the *V*-values of the positive and negative elements is large. A lens with a large vertex length will, at any given diameter, vignette at a smaller angle than will a short lens. Further, it turns out that the longer the lens: (1) the smaller the spherical zonal and (2) the smaller the field coverage (i.e., the higher-order astigmatism and coma are greater and limit the angle over which a good image can be obtained when the lens is long). Thus, long systems are appropriate for high-speed, small-angle systems; short systems for small-aperture, wide-angle applications. As a very rough rule of thumb, the vertex length of a triplet is frequently equal to the diameter of the entrance pupil.

The length of the triplet can be controlled by the choice of the glasses used. For example, if a shorter system is desired, the substitution of a flint with a higher *V*-value (or a crown with a lower *V*-value) will produce the necessary change. To get a longer system, use a higher *V*-value crown and/or a lower *V*-value flint. (However, note that a system which is *too* long will have no solution for the element shapes. The ray height on the negative element may be so low that its overcorrecting contribution to the spherical aberration is insufficient to offset the undercorrection of the positive elements simultaneously with the requirements for coma and astigmatism correction as well as chromatic and Petzval.)

Interestingly enough, *this relationship between vertex length and zonal spherical and field coverage is a general one and applies to most anastigmats.** Thus, if an anastigmat design has too much zonal spherical and more than enough angular coverage, one can simply choose new glasses to lengthen the system and strike the desired balance between field and aperture, or vice versa. There are, of course, limits to the effectiveness of this technique.

In general, the higher the index of the crown (positive) elements and the lower the index of the flint, the better the design will be. In other words, with all else equal, a triplet with a more positive index difference (n crown − n flint) will have a smaller zonal spherical and/or a wider field coverage. See also Fig. 13.52 for the effect of index on aberrations.

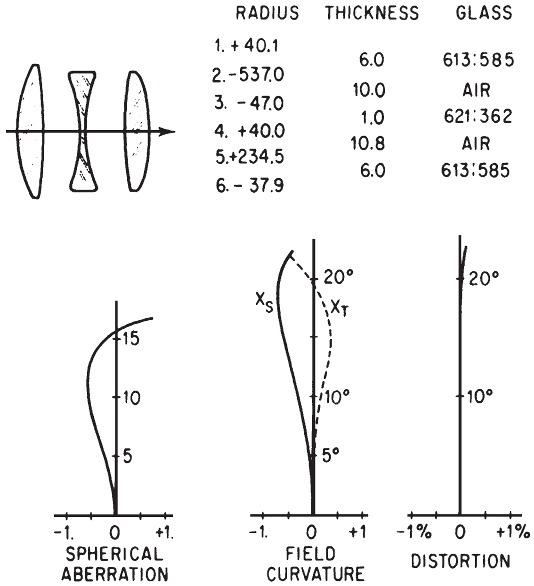
Figure 10.9 showed a triplet of relatively high aperture and modest field coverage. Figures 12.13 and 12.14 illustrate triplets of reduced vertex length and increasingly smaller aperture and wider fields of view. Needless to say, the Cooke triplet is best designed using an automatic computer lens design program of the type described in Sec. 12.9. However, the automatic design program can be better utilized and better results will be obtained if the designer has mastered the information in this section. Figures 14.9 and 14.10 also show Cooke triplet designs. Figure 14.39 shows a triplet with an aspheric field corrector, suitable for use in a point-and-shoot camera, and Fig. 14.41 shows an infrared (8–14 μm) triplet. Figure 14.42 is another IR triplet-based lens of very high speed (*f*/0.55).

12.8 A Generalized (Nonautomatic, Old-Fashioned) Design Technique

The preceding sections have outlined specific design approaches for three particular types of optical systems. This section will describe a generalized approach to optical design. Because of the varied nature of different types of optical systems, this description will be unnecessarily elaborate for many simple cases and must, because of limitations of space and knowledge, fall short of completeness for elaborate and specialized systems. The reader will recognize generalizations of most of the procedures set forth in the preceding sections.

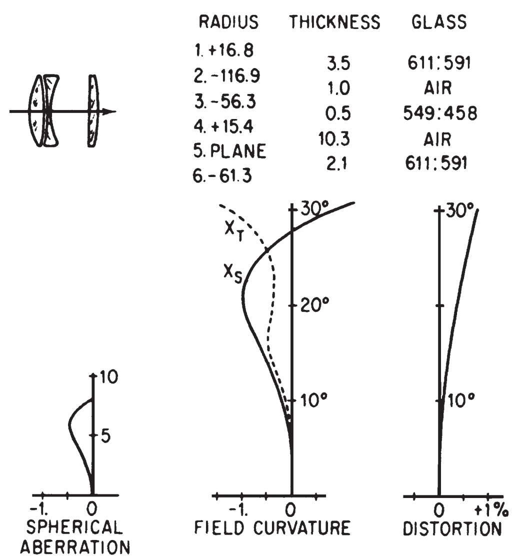
This section describes the design process as if it were to be executed "manually," i.e., without the benefit of a modern "automatic" optical design software program. The aim of this section is not necessarily to prescribe the operations indicated but to outline the structure of the

*See W. Smith, *J. Opt. Soc. Am.,* vol. 48, 1958, pp. 98–105.



RADIUS	THICKNESS	GLASS
1. + 40.1		
	6.0	613:585
2. -537.0		
	10.0	AIR
3. - 47.0		
	1.0	621:362
4. +40.0		
	10.8	AIR
5. +234.5		
	6.0	613:585
6. - 37.9		

Figure 12.13 A Cooke triplet anastigmat of moderate aperture and coverage. Compare with Figs. 10.9 and 12.14. English Patent 155,640-1919. Focal length is 100 units. This design is of the type made for use in slide projectors.



RADIUS	THICKNESS	GLASS
1. +16.8		
	3.5	611:591
2. -116.9		
	1.0	AIR
3. -56.3		
	0.5	549:458
4. +15.4		
	10.3	AIR
5. PLANE		
	2.1	611:591
6. -61.3		

Figure 12.14 A Cooke triplet of small aperture and wide coverage. Compare with Figs. 10.9 and 12.13. German Patent 287,089-1913. Focal length is 100 units.

basic design approach, including some techniques which may in fact be a valuable supplement to understanding and executing the design process.

General considerations. The first step in the design process is the organization of the requirements to be imposed on the optical system into terms of aperture, focal length, field coverage, resolution (or blur spot size), spectral bandwidth, transmission, mechanical or space limitations, and the like. In some systems, e.g., telescopes, the preliminary design may be profitably carried out member by member; if this is to be done, requirements for the individual members are established.

The general configuration of the system is established next. Ordinarily the designer will conduct a survey of known designs (books, technical periodicals, the patent files, and the designer's own experience are the primary sources) to determine whether the performance requirements are within the "state of the art." If so, the designer will select a type of system which is just capable of meeting the requirements (i.e., the most "economical" choice) and will proceed to adjust its parameters to achieve the optimum balance of correction for the particular application at hand. If the performance requirements are beyond the capability of any known design, the designer will select a design form which is felt to be "most likely to succeed." The designer will analyze it thoroughly to determine its shortcomings and then attempt to improve its characteristics. In many instances, there is no directly applicable prior art from which to begin further effort. In such circumstances, a thorough analysis of the first-order (gaussian) requirements is conducted and a system is invented, utilizing the basic design principles exemplified in known designs on a piecemeal basis to accomplish the necessary ends.

In the following paragraphs, we assume that the general design type has been established, either by selection or invention. The next major step in the design process is the correction of the primary aberrations, or at least as many of them as are necessary and feasible. This procedure has been accurately described as the art of solving a number (say n) of second- (or higher-) order equations in m unknowns; one must ascertain initially that m (the number of *effective* variables) exceeds or equals n (the number of aberrations).

Manual correction of the primary aberrations. The powers and spacings of the elements which are to comprise the system can usually be determined on a highly rational basis. First, the elements must be so arranged as to provide the desired focal length, aperture, field, and so on for the system. Throughout this entire design stage, the value of a scale drawing cannot be overemphasized; such a drawing will prevent attempts to design impossible elements, such as those with negative edge thickness, or with hyperhemispheric concave surfaces. If the ray paths are roughed in (on a first-order basis), one is also less apt to require magnifications and apertures which lead to slope or incidence angles which exceed 90°.

The usual method for correction of the aberrations is bending the elements, i.e., changing the shape of the elements while maintaining a constant power and position. However, certain aberrations are unaffected (or affected only slightly) by bending. These are axial (longitudinal) chromatic aberration, lateral color, Petzval curvature, and, to a certain extent, distortion. The chromatic aberrations and the Petzval curvature *must* be corrected in the power and space layout if they are ever to be corrected. The third-order contribution equations, especially the thin-lens versions, are most useful at this stage, and it is ordinarily a relatively straightforward procedure to adjust the system so that the ∑TAchC, ∑TchC*, and ∑PC are equal to values which have been selected as desirable (or at least acceptable). See Eqs. 12.11 through 12.14.

It then remains to correct the spherical, coma, astigmatism, and distortion to their desired values. A number of alternate procedures are often available at this step. Unless the designer has prior experience with the type of system under construction, or unless the design effort is a minor modification of an existing design, it is probably best at this stage to make a graph of the aberration contributions from each element (or component) as a function of the element shape. Then, from a set of such graphs, a region (or regions) in which a solution is possible can be selected. These graphs can be plotted from data obtained by the use of the thin-lens contribution equations, the surface contribution equations, or in some cases by direct raytracing. The latter two methods are appropriate for electronic computer work; one element at a time is bent and the changes in the final aberrations are plotted. The thin-lens expressions have the advantage that the "n equations in m unknowns" are explicitly that and can be handled analytically.

When a "region of solution" is selected (by whatever means), a method of differential correction is usually applied. The partial differentials of the aberrations against shape, $\delta A/\delta C$ (or $\Delta A/\Delta C$), are determined along with the values of the aberrations for a trial prescription. The desired amount of change of each aberration (ΔA) is determined from the analysis of a trial prescription and the necessary number (n) of simultaneous equations of the general form

$$\Delta A_n = \sum_{i=1}^{i=K} \left(\frac{\delta A_n}{\delta C}\right)_i \Delta C_i \qquad (12.15)$$

are set up and solved to yield the required values of ΔC_i. Because of the nonlinearity of the equations (i.e., the partials vary as the shape is changed), the first solution is seldom precise. However, the preselection of the region of solution limits the size of the ΔC's so that the linear simultaneous solution of Eqs. 12.15 is a good approximation; a series of such solutions converges rapidly on the required design form.

It is sometimes advisable to limit the number of parameters used in the technique described above. Because a limited number of aberrations are to be controlled, the problem is simplified if only an equal number of variables are used, *provided that these variables are effective and admit of a solution.* The preliminary graphs of the aberrations (versus element shapes) and the subsequent selection of a region of solution are strongly recommended as insurance against ineffective parameters and insoluble sets of simultaneous equations.

Certain systems lend themselves to an iterative technique which can be a powerful design tool. For example, assume that three aberrations, *A, B,* and *C* are to be corrected by the adjustment of three parameters, *x, y,* and *z*. An initial trial prescription is modified by changing one of the parameters, say *z,* until one of the aberrations, say *C,* is "corrected." Then parameter *y* is arbitrarily changed and a new value of *z* is determined to maintain the correction of *C*. Parameter *y* is varied in this manner until the aberrations *B* and *C* are simultaneously corrected. Then parameter *x* is changed; with each change of *x,* parameters *y* and *z* are adjusted as above to hold the aberrations *B* and *C* at the desired values. Parameter *x* is varied in this manner until aberration *A* is brought to correction simultaneously with *B* and *C*. In such a process, graphs of *C* means *z, B* versus *y,* and *A* versus *x* are quite useful.

If the thin-lens aberration expressions have been used in any of the preceding steps, it is necessary to add thickness to the elements. This is generally done by adjusting the secondary curvature of each thick element to hold the thick-element power equal to the thin-lens-element power. The spacing between elements is then adjusted so that the spacing of the thick-element principal points is equal to the thin-lens spacing. This method serves to retain the overall system power and working distances at the same values as the thin-lens systems. Some designers prefer to adjust the secondary curvatures to maintain the Petzval curvature precisely. The exact procedure used to go from thin to thick is not critical; what may be important is that the procedure of introducing thickness be rigorously consistent (in order that the differential trigonometric correction method will be accurate).

Trigonometric correction. When the third-order aberrations have been brought to desired values, it is necessary to trace rays trigonometrically to determine the actual state of correction of the system. It will usually differ by a small amount from that predicted by the third-order aberrations. However, a step or two of differential correction as outlined five paragraphs above will usually bring the trigonometric correction home; in most systems, the *change* in the trigonometrically determined aberrations is quite close to the change predicted by third-order aberration calculations.

Reduction of residual aberrations. After the primary aberrations have been brought to correction, the design is tested for residual aberrations. The primary aberrations are generally corrected for only a single zone of the aperture or field and can be expected to depart from correction in all other zones, as previously discussed in Chapter 3. Several general principles can be given for the reduction of residuals; their variety and extent make a catalog of specific remedies too extensive for inclusion.

If there are any "leftover" parameters that were not used in the correction of the primary aberrations, these may be systematically varied and their effects on the residuals noted and used. In addition to the obvious and continuously variable parameters of bendings, powers, and spacings, the choice of glass types is often an effective leftover. Also the possibility that more than one region of solution exists should not be overlooked, since this is, in effect, an extra parameter.

An analysis of the source of the third-order surface contributions will often pinpoint one or two surfaces or elements which are especially heavy contributors. The elimination or reduction of a single large contribution will often reduce residual aberrations. This can be accomplished by introducing a correcting element near the offender (for example, convert a single element into a compound component, perhaps an achromat), by splitting the offending element into two elements whose total power equals that of the original, by raising the index, or (infrequently) by shifting the offender to a location where the incidence angles of the rays on its surfaces are reduced. Compounding or splitting an element introduces two new variable parameters: the ratio of the powers of the two elements (although the best split ratio is often close to 50-50) and the shape of the added element. An additional possibility is that a drastically different shape for the troublesome element may reduce its contribution to an acceptable level.

The specific remedies for spherochromatism, zonal spherical, and field coverage set forth in Secs. 12.5 and 12.7 have fairly general applicability. Another specific is the introduction of a zero-power meniscus element or a concentric meniscus element into the system. Depending on how and where it is used, a meniscus can be effective in modifying zonal spherical, Petzval curvature, or astigmatism.

An aspheric surface can be a powerful design tool for the reduction of residuals or the elimination of primary aberrations (especially distortion, astigmatism, and spherical) which will yield to no other design techniques. One should, if at all possible, temper one's enthusiasm for the easy way out which the aspheric surface represents with the knowledge that several spherical elements may usually be added to a design for less than the cost of producing a single precise aspheric surface. As a consequence of this fact, aspherics are seldom used except where absolutely necessary for space or weight considerations, or where cost is no object (as in one-of-a-kind instruments), or where the required precision of the surface is very low (as in molded condenser elements). Although injection-molded plastic element and diamond-turned surfaces are often aspheric, and glass aspherics can be molded, tooling cost is the limitation here.

In general, where residuals are a problem, it is wise to reconsider the initial power and space layout for the entire system. It is sometimes possible to revise the layout in such a way that the powers of the elements or the "work" ($y\phi$ or $y_p\phi$) done by the elements can be reduced. This is an extremely rapid and effective way of reducing residuals. An initial choice of too small a value for the Petzval sum will result in elements of high power and large residuals. A change to allow a more inward-curving field is the obvious remedy for this situation for ordinary lenses.

Aberration balancing. The final stage in the optical design process consists of balancing the aberrations, or "touching up" the design. Here the experienced designer frequently departs from what may seem to be the best state of correction in order to minimize the overall effects of the residual aberrations. In the presence of zonal spherical, spherochromatism, and astigmatism, the interrelationships of the aberrations with each other, and with the position selected for the focal plane, often allow an improvement to be made by selecting a deliberately uncorrected state. We have previously (Sec. 11.3) seen that the best spherical correction as regards OPD occurs when the marginal spherical is zero and the reference plane is shifted toward the zonal focus; the minimum *geometrical* blur spot size (Sec. 11.7) requires that the marginal spherical be undercorrected. Thus, if the application of the system is such that a resolution significantly less than the diffraction-limited resolution is of prime importance, and if the zonal spherical is large in terms of OPD, then an undercorrected marginal spherical is in order. Except in a camera lens, an overcorrected marginal spherical is rarely desirable; it does permit a higher resolution and reduces focus shift when the system is stopped down, but it reduces the image contrast.

Another reason for preferring a slightly undercorrected spherical is that the oblique spherical aberration (y^3h^2) is almost always overcorrected and the axial undercorrection will counterbalance this tendency. The overcorrected oblique spherical also causes the *effective* field curvature to be more backward-curving than indicated by the x_s and x_t curves given by Coddington's equations (Eq. 10.5). This is especially true for the tangential field curvature. For this reason the astigmatism is seldom made overcorrected enough to cause a backward-curving tangential field; ordinarily one desires a correction somewhere between $x_t = 0$ and $x_t = x_s = x_p$. Note that the focus position is usually chosen inside the paraxial focal plane and that the field curvature should be judged with this in mind.

We have previously noted that the Petzval curvature in most anastigmats is preferably left somewhat inward-curving in order to minimize element powers and aberration contributions.

The obvious choice of the 0.707 zone of the aperture as the zone at which to correct the longitudinal chromatic is rarely the best choice unless the spherochromatism is small. In the presence of spherochromatism and an undercorrected zonal spherical, the inward shift of the best focus from the paraxial focus allows the overcorrected spherical of the blue light to produce a halo or blue haze in the image. This can be eliminated, or reduced, by correcting the chromatic at a larger zone of the aperture.

The reader should bear in mind that the preceding comments are intended to apply to normal types of lenses in which (as is usually the case) the higher-order residuals are somewhat larger than desirable.

12.9 Automatic Design by Electronic Computer

The fantastically high computation speed of the electronic computer makes it possible to perform a major portion of the optical design task on an "automatic" basis. One possible approach is essentially a duplication of the process that a designer goes through in correcting the primary aberrations of a system. The computer is presented with an initial prescription and a set of desired values for a limited set of aberrations. The machine then computes the partial differentials of the aberrations with respect to each parameter (curvature, spacing, etc.) which is to be adjusted and establishes a set of simultaneous equations (Eqs. 12.15), which it then solves for the necessary changes in the parameters. Since this solution is an approximate one, the computer then applies these changes to the prescription (assuming that the solution is an improvement) and continues to repeat the process until the aberrations are at the desired values. When there are more variable parameters than system characteristics to be controlled, there is no unique solution to the simultaneous equations; in this case, the computer will add another requirement, namely that the sum of the squares of the (suitably weighted) parameter changes be a minimum. This allows a solution to be found and has the added advantage that it holds the system close to the original prescription. Since the solution of simultaneous equations may call for excessively large changes to be applied, the computer is usually instructed to scale down the changes if they exceed a certain predetermined value.

This "simultaneous" technique is a useful one. Even modest-sized computers are capable of handling this problem without difficulty and several inexpensive computer programs of this type are available, often based on third-order aberration contributions. Since the designer is in rather close control of the situation, this technique is, in effect, simply an automation of conventional methods as described in the preceding section. Thus, the designer should have a fairly good knowledge of the system, and the system must have a solution reasonably close to the initial prescription. This type of approach is very efficient for making modest changes in designs or for touching-up a design. It also makes easy work of systems with exceedingly complex interrelationships of the variables, such as the older meniscus anastigmats of the Dagor or Protar type.

Fully automatic lens design optimization

There are many other approaches to automatic design; almost all of them are characterized by the use of a "merit function." The merit function is a single numerical value which indicates to the computer whether any given change has improved the lens or not. Obviously, representing the total performance of a lens system by a single number is a rather tricky business and considerable care must be taken in the choice of the merit function; at times it seems that the "design" of the merit function is more demanding than the design of the lens which the merit function is intended to represent. Some approaches use a merit function of the following sort: A large number of rays are traced from each of several points in the field of view. For each image point, the distance of each ray intersection (with the image plane) from the "ideal" location for that ray is computed and the sum of the squares of these distances is taken. Then the sum of the sums for the several image points is the merit function. Since the merit function will be large if the image blur spot is large, it is apparent that a small value of the merit function is desirable.

The construction of the merit function as described above is far from the most desirable scheme of things, and in practice many refinements are used. Since the outer portions of the field are frequently less critical than the center, the individual sums may be weighted to take this into account. A modest amount of computation will indicate that, in the presence of a constant fifth-order spherical aberration, the smallest value of the sum of the squares of the ray displacements does *not* represent the best solution from an OPD standpoint. One scheme uses a reduced weighting of large ray displacements in an attempt to take this into account. The choice of the "ideal" intersection point for the rays (for off-axis points) is a complex matter; the use of the gaussian image point is quite misleading if any amount of distortion is present. Similarly, the use of the image-plane intersection of the principal ray as the ideal point can yield a distorted evaluation in the presence of coma. Frequently the separately computed values of distortion and lateral chromatic aberration are added (suitably weighted) into the merit function, and the computer selects the centroid of the blur spot as the "ideal" point.

Other types of merit function are also widely used to characterize the quality of a lens system. A few use the OPD, or wave-front aberration, as the merit function, taking the variance of the wave front for several field points, after selecting the reference point (i.e., image plane) so as to minimize the variance over the field. Another very widely used approach allows the designer to tailor a merit function to suit the application. The merit function entries may be ray displacements, OPD, defocusing, field curvature, chromatic aberrations, the slope, or the curvature of the ray intercept plot, the constructional data of the lens, the ray heights or slopes, or the classical aberrations, plus almost any mathematically possible combination of these.

The merit function, being a collection of aberrations and departures from desired conditions, is obviously misnamed; it properly should be called a *defect function* or *error function.* However, common usage has established "merit function" as a well-understood term, and we will use it here with the understanding that the smaller the merit function, the better the image.

Almost all automatic-lens-design programs allow at least some adjustment to the merit function, even if they do not allow the sort of flexibility described above. Typically, even in a program of limited flexibility, different parts of the aperture, field, or spectrum can be weighted to suit the application and the design form. The general procedure is to have the program optimize a design, for the designer to examine the results, and then to adjust or alter the merit function in such a way as to achieve the desired balance of aberrations and characteristics.

Automatic-lens-design programs operate this way: Each of the construction parameters to be varied is changed (one at a time) by a small amount. The corresponding change in each entry or aberration in the merit function is calculated in order to obtain its partial derivative with respect to the parameter. Then equations of the form of Eqs. 12.15 are set up, one for each aberration or merit function entry. Typically there are many more aberrations in the merit function than there are effective variable parameters in the lens, so a "solution" is made in the least-squares sense, i.e., the variable set is changed in such a way as to minimize the sum of the squares of the differences between the desired value of each aberration and the value predicted by Eqs. 12.15. But Eqs. 12.15 are based on an approximation; the assumption that the relationship between aberration and variable is a linear one. We have seen in Chapters 3 and 10 that, even for third-order aberrations, this is not so, and it is much less linear for the higher-order aberrations. At best then, the solution is an approximate one, but probably significantly improved over the original lens form. At worst, the nonlinearity of the relationships can cause the least-squares process to come up with such an extreme change that the design is not just worse, it may be a totally impossible form with near-zero radii that the rays miss, or near-infinite spacings that cause similarly disastrous results. This problem can be handled by adding to the merit function the sum of the weighted squares of all the parameter changes. This penalizes any large parameter changes and tends to stabilize the process. The weighting can be adjusted to be large where the nonlinearity is a problem, and small where it is not. This is called *damped least squares,* and with a few significant exceptions, is the basis of current automatic lens design programs.

By repeating the approximate solution process until it converges, these programs are capable of driving a rough preliminary design form to the nearest local minimum of the merit function. Depending upon the structure of the merit function, most lens designs have more than one local minimum. Consider the "front" and "rear" meniscus camera lens discussed in Sec. 12.2, or the Fraunhofer and Gauss forms of telescope objectives (Secs. 12.4 and 12.5); these are simple design forms where the merit function has two obvious local minima. An automatic design program will find the minimum nearest to the starting design form which it is given. There is no way that the user of such a program can be certain that a minimum is the best one (i.e., a "global optimum"). The solution space is n-dimensional, where n is the number of variable parameters. In the simple designs discussed in this chapter it was not impractical for us to do a limited, simplified exploration of the solution space. In a design with 20 or 30 variable parameters it is a quite different matter.

In any case, it is apparent that since the design program will seek out the nearest minimum, the selection of the starting point for the process is vitally important. In fact, once the merit function is defined and weighted, the starting design form uniquely defines a single minimum. Obviously the choice of the starting form is a critical factor. Fortunately, it seems that with most merit functions, most nonsimple design types have relatively broad, flat minima, and one can choose a starting point over a fairly large volume in solution space and expect a reasonably good result. An experienced lens designer uses knowledge of successful design types and features to direct the computer to good starting points. The novice designer should study the standard, classical design forms as an aid in selecting appropriate starting points.

The mathematics of this process are written up in many places. Two which explain the basic operations are G. Spencer, "A Flexible Automatic Lens Correction Procedure," *Applied Optics,* vol. 2, 1963, pp. 1257–1264, and W. Smith, in W. Driscoll (ed.), *Handbook of Optics,* New York, McGraw-Hill, 1978.

12.10 Practical Considerations

The following is a partial list of certain design characteristics which, although they may be quite beneficial to the performance of a design, tend to have an undesirable effect on the difficulty and cost of fabrication. Thus, unless you enjoy being unpopular with the opticians who must execute your designs, this list represents things which you should assiduously avoid if at all possible.

1. Materials which are soft and easily abraded.
2. Materials which are thermally fragile and which may split from a mild thermal shock, such as that encountered in blocking or washing under a hot or cold water tap.
3. Materials with low acid resistance or high stain characteristics.
4. Expensive materials. (Often you can find a similar, cheaper glass which is nearly as good.)
5. Thin elements, i.e., those with a large ratio of diameter to the average thickness. Such elements can deform under the stress of blocking or polishing, making an accurate surface geometry almost impossible to produce. Note that a negative element with a substantial edge thickness often can tolerate a center thickness which would be too thin for a weaker element.
6. Thin-edged elements chip easily and, if processed at a diameter larger than the finished one, may become sharp-edged during fabrication. Also a thin-edged element is difficult to mount satisfactorily.
7. A very thick element obviously requires more material and may require an awkward arrangement when blocked. Visualize Fig.

15.2 if the elements are as thick as the diameter. A thin lens with the same radius can have more lenses blocked on a tool because they can be placed closer together at the surface; with the thick lens, there are large gaps between the elements at the surface which make polishing difficult.

8. Very "strong" curves (i.e., with a large diameter-to-radius ratio) lead to fewer elements blocked per tool and the correspondingly increased processing costs, difficulty in polishing surfaces accurately, and difficulty in testing the surface accuracy with a test plate or interferometer.
9. Meniscus elements whose surfaces are concentric or nearly concentric with each other. A monocentric element must be ground and polished so that the two surfaces are properly aligned during these operations; it cannot be "centered" after polishing as an ordinary element can.
10. Nearly equiconvex or equiconcave elements can cause trouble in assembly because it is difficult to tell one side from the other, and the element is liable to be mounted backward.
11. Weakly curved, nearly plane surfaces are more expensive to tool and fabricate than a plane surface. It is almost always possible to force such a design to a plane surface with little or no sacrifice in image quality.
12. Precision bevels. If possible, avoid mounting from a beveled surface. Use a loosely toleranced 0.5 mm by 45° chamfer to eliminate sharp edges; this kind of edge break is almost free.
13. Avoid odd-angle precision bevels. Many shops are tooled for 45°, 30°, or 60°; other angles may require new tooling.
14. Cemented triplets and quadruplets are unpopular in some shops.
15. Tight scratch and dig specifications on surfaces which are not visible to the ultimate customer are usually a waste of money. With a few exceptions (such as surfaces near an image plane or the optics of a high-powered laser system), scratch and dig considerations are purely cosmetic and have no functional effect (unless the lens aperture is so small that a dig can actually obstruct a significant fraction of the beam area).
16. Tight tolerances in general. See Chap. 15 for a discussion of efficient tolerance budgeting.

Bibliography

Note: Titles preceded by an asterisk (*) are out of print.

*Conrady, A., *Applied Optics and Optical Design,* Oxford, 1929. (This and vol. 2 were also published by Dover, New York.)
*Cox, A., *A System of Optical Design,* Focal, 1965 (lens construction data).
Dictionary of Applied Physics, vol. 4, London, Macmillan, 1923.
Farn, M. W., and W. B. Veldkamp, "Binary Optics," in *Handbook of Optics,* vol. 2, New York, McGraw-Hill, 1995, Chap. 8.
Fischer, R. (ed.), *Proc. International Lens Design Conf.,* S.P.I.E., vol. 237, 1980.
*Greenleaf, A., *Photographic Optics,* New York, Macmillan, 1950.
Herzberger, M., *Modern Geometrical Optics,* New York, Interscience, 1958.
*Jacobs, D., *Fundamentals of Optical Engineering,* New York, McGraw-Hill, 1943.
*Kingslake, R. (ed.), *Applied Optics and Optical Engineering,* vol. 3, New York, Academic, 1965 (lens design).
*Kingslake, R., *Lens Design Fundamentals,* New York, Academic, 1978.
*Kingslake, R., *Lenses in Photography,* Garden City, 1952.
*Linfoot, E., *Recent Advances in Optics,* London, Clarendon, 1955.
*Martin, L., *Technical Optics,* New York, Pitman, 1950.
Merte, W., *Das Photographische Objektiv,* Parts 1 and 2, translation, CADO, Wright-Patterson AFB, Dayton, 1949.
Merte, Richter, and von Rohr. *Handbuch der Wissenschaftlichen und Angewandten Photographie,* vol. 1, 1932; *Erganzungswerke,* 1943, Vienna, Springer. Reprinted by Edwards Brothers, 1944 and 1946 (lens construction data).
Merte, *The Zeiss Index of Photographic Lenses,* vols. 1 and 2, CADO, Wright-Patterson AFB, Dayton, 1950 (lens construction data).
MIL-HDBK-141, *Handbook of Optical Design,* 1962.
Peck, W., in Shannon and Wyant (eds.), *Applied Optics and Optical Engineering,* vol. 8, New York, Academic, 1980 (automatic lens design).
Rodgers, P., and M. Roberts, "Thermal Compensation Techniques," in *Handbook of Optics,* vol. 1, New York, McGraw-Hill, 1995, Chap. 39.
Rosin, S., "A New Thin Lens Form," *J. Opt. Soc. Am.,* vol. 42, 1952, pp. 451–455.
Sinclair, D. C., "Optical Design Software," in *Handbook of Optics,* vol. 1, New York, McGraw-Hill, 1995, Chap. 34.
Smith, W. J. (ed.), *Lens Design,* S.P.I.E., vol. CR41, 1992.
Smith, W. J., *Modern Lens Design,* New York, McGraw-Hill, 1992.
Smith, W., in W. Driscoll (ed.), *Handbook of Optics,* New York, McGraw-Hill, 1978.
Smith, W., in Wolfe and Zissis (eds.), *The Infrared Handbook,* Washington, Office of Naval Research, 1985.

Taylor, W., and D. Moore (eds.), *Proc. International Lens Design Conf.,* S.P.I.E., vol. 554, 1985.

Exercises

The exercises for this chapter take the form of suggestions for individual design projects; as such, there can be no "right" answers, and none are given. The effort involved in each exercise is considerable, and it is likely that only those interested in obtaining first-hand experience in optical design will wish to undertake these exercises. The casual reader will, however, be amply rewarded by mentally reviewing the steps he or she would follow in attempting the exercises.

1 Design a symmetrical double-meniscus objective of the periscopic type. Select a bending (a ratio of 3:2 for the curvatures is appropriate), determine the proper spacing for a flattened field, and calculate the thin-lens third-order aberrations for the combination. Analyze the final design by raytracing and compare the results with the third-order calculations. The student may wish to repeat the process for several additional bendings, perhaps including the Hypergon (Fig. 12.4), and to compare the results of each, noting the variations of aperture and coverage.

2 Design an achromatic doublet objective using BK7 (517:642) and SF2 (648:339). Correct the spherical aberration for an aperture of *f*/3.5. Raytrace marginal and zonal rays in *C, D,* and *F* light to evaluate the axial image. Compare the coma obtained by raytracing an oblique fan with the OSC calculation.

3 Design a telescope objective lens consisting of a BK7 singlet and a doublet of BK7 and SF2. Vary the distribution of powers and the spacing to optimize the correction of zonal spherical and spherochromatic.

4 Design a Cooke triplet anastigmat. For a minimal exercise, duplicate the design of Fig. 12.13 using the same glasses and the same power and space layout as a starting point. For a more ambitious project, design the same lens, but derive the power and space layout without recourse to the data of the figure.