

PART

1

DESIGN

1

TECHNIQUES OF FIRST-ORDER LAYOUT

Warren J. Smith*

*Kaiser Electro-Optics, Inc.
Carlsbad, California*

1.1 GLOSSARY

A, B	scaling constants
d	distance between components
f	focal length
h	image height
I	invariant
j, k	indices
l	axial intercept distance
M	angular magnification
m	linear, lateral magnification
n	refractive index
P	partial dispersion, projection lens diameter
r	radius
S	source or detector linear dimension
SS	secondary spectrum
s	object distance
s'	image distance
t	temperature
u	ray slope
V	Abbe number
y	height above optical axis
α	radiometer field of view, projector field of view
ϕ	component power ($= 1/f$)

*Deceased.

1.2 FIRST-ORDER LAYOUT

First-order layout is the determination of the arrangement of the components of an optical system in order to satisfy the first-order requirements imposed on the system. The term “first-order” means the paraxial image properties: the size of the image, its orientations, its location, and the illumination or brightness of the image. This also implies apertures, f -numbers, fields of view, physical size limitations, and the like. It does not ordinarily include considerations of aberration correction; these are usually third- and higher-order matters, not first-order. However, ordinary chromatic aberration and secondary spectrum are first-order aberrations. Additionally, the first-order layout can have an effect on the Petzval curvature of field, the cost of the optics, the sensitivity to misalignment, and the defocusing effects of temperature changes.

The primary task of first-order layout is to determine the powers and spacings of the system components so that the image is located in the right place and has the right size and orientation. It is not necessary to deal with surface-by-surface ray-tracing here; the concern is with components. “Components” may mean single elements, cemented doublets, or even complex assemblies of many elements. The first-order properties of a component can be described by its Gauss points: the focal points and principal points. For layout purposes, however, the initial work can be done assuming that each component is of zero thickness; then only the component location and its power (or focal length) need be defined.

1.3 RAY-TRACING

The most general way to determine the characteristics of an image is by ray-tracing. As shown in Fig. 1, if an “axial (marginal)” ray is started at the foot (axial intercept) of the object, then an image is located at each place that this ray crosses the axis. The size of the image can be determined by tracing a second, “principal (chief),” ray from the top of the object and passing through the center of the limiting aperture of the system, the “aperture stop;” the intersection height of this ray at the image plane indicates the image size. This size can also be determined from the ratio of the ray slopes of the axial ray at the object and at the image; this yields the magnification $m = u_0/u'_k$; object height times magnification yields the image height.

The ray-tracing equations are

$$y_1 = -l_1 u_1 \quad (1)$$

$$u'_j = u_j - y_j \phi_j \quad (2)$$

$$y_{j+1} = y_j + d_j u'_j \quad (3)$$

$$l'_k = -y_k / u'_k \quad (4)$$

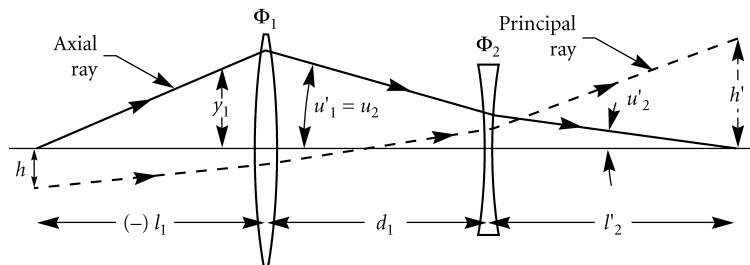


FIGURE 1

where l and l' are the axial intersection distances of the ray before and after refraction by the component, u and u' are the ray slopes before and after refraction, ϕ is the component power ($\phi = 1/f$), y_j is the height at which the ray strikes the j th component, and d_j is the distance from the j th to the $(j+1)$ th component. Equations (2) and (3) are applied sequentially to the components, from object to image.

These equations can be used in two different ways. When the components and spacings are known, the image characteristics can readily be calculated. In the inverse application, the (unknown) powers and spaces can be represented by symbols, and the ray can be traced symbolically through the postulated number of components. The results of this symbolic ray-tracing can be equated to the required characteristics of the system; these equations can then be solved for the unknowns, which are the component powers and spacings.

As an example, given the starting ray data, y_1 and u_1 , we get

$$\begin{aligned} u'_1 &= u_1 - y_1 \phi_1 \\ y_2 &= y_1 + d_1 u'_1 = y_1 + d_1 (u_1 - y_1 \phi_1) \\ u'_2 &= u'_1 - y_2 \phi_2 \\ &= u_1 - y_1 \phi_1 - [y_1 + d_1 (u_1 - y_1 \phi_1)] \phi_2 \\ y_3 &= y_2 + d_2 u'_2 = \text{etc.} \end{aligned}$$

Obviously the equations can become rather complex in very short order. However, because of the linear characteristics of the paraxial ray equations, they can be simplified by setting either y_1 or u_1 equal to one (1.0) without any loss of generality. But the algebra can still be daunting.

1.4 TWO-COMPONENT SYSTEMS

Many systems are either limited to two components or can be separated into two-component segments. There are relatively simple expressions for solving two-component systems.

Although the figures in this chapter show thick lenses with appropriate principal planes, “thin” lenses (whose thickness is zero and whose principal planes are coincident with the two coincident lens surfaces) may be used.

For systems with infinitely distant objects, as shown in Fig. 2, the following equations for the focal length and focus distance are useful:

$$f_{AB} = f_A f_B / (f_A + f_B - d) \quad (5)$$

$$\phi_{AB} = \phi_A + \phi_B - d \phi_A \phi_B \quad (6)$$

$$B = f_{AB} (f_A - d) / f_A \quad (7)$$

$$F = f_{AB} (f_B - d) / f_B \quad (8)$$

$$h' = f_{AB} \tan u_p \quad (9)$$

where f_{AB} is the focal length of the combination, ϕ_{AB} is its power, f_A and f_B are the focal lengths of the components, ϕ_A and ϕ_B are their powers, d is the spacing between the components, B is the “back

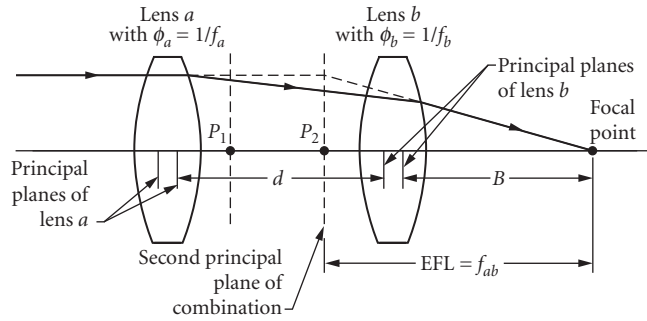


FIGURE 2

focus” distance from the B component, F is the “front focus” distance, u_p is the angle subtended by the object, and h' is the image height.

If f_{AB} , d , and B (or F) are known, the component focal lengths can be found from

$$f_A = df_{AB}/(f_{AB} - B) \quad (10)$$

$$f_B = -dB/(f_{AB} - B - d) \quad (11)$$

These simple expressions are probably the most widely used equations in optical layout work.

If a two-component system operates at *finite* conjugates, as shown in Fig. 3, the following equations can be used to determine the layout. When the required system magnification and the component locations are known, the powers of the components are given by

$$\phi_A = (ms - md - s')/msd \quad (12)$$

$$\phi_B = (d - ms + s')/ds' \quad (13)$$

where $m = h'/h$ is the magnification, s and s' are the object and image distances.

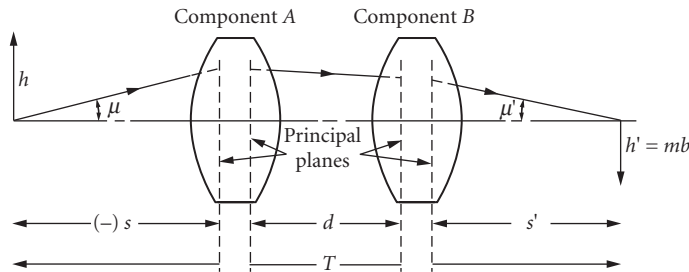


FIGURE 3

In different circumstances, the component powers, the object-to-image distance, and the magnification may be known and the component locations are to be determined. The following quadratic equation [Eq. (14)] in d (the spacing) is solved for d :

$$0 = d^2 - dT + T(f_A + f_B) + (m - 1)^2 f_A f_B / m \quad (14)$$

and then

$$s = [(m - 1)d + T] / [(m - 1) - m d \phi_A] \quad (15)$$

$$s' = T + s - d \quad (16)$$

1.5 AFOCAL SYSTEMS

If the system is afocal, then the following relations will apply:

$$MP = -(f_O / f_E) = (u_E / u_O) = (d_O / d_E) \quad (17)$$

and, if the components are “thin,”

$$L = f_O + f_E \quad (18)$$

$$f_O = -L \cdot MP / (1 - MP) \quad (19)$$

$$f_E = L / (1 - MP) \quad (20)$$

where MP is the angular magnification, f_O and f_E are the objective and eyepiece focal lengths, u_E and u_O are the apparent (image) and real (object) angular fields, d_O and d_E are the entrance and exit pupil diameters, and L is the length of the telescope as indicated in Fig. 4.

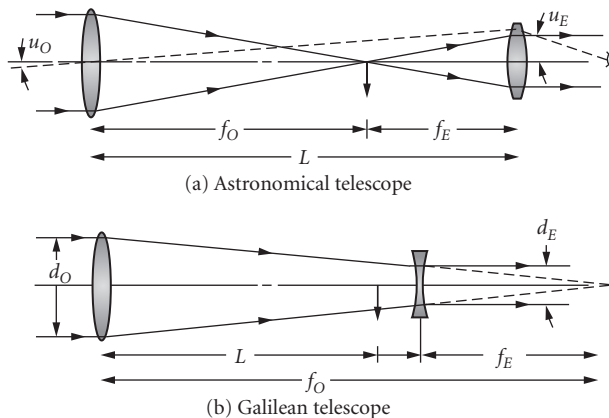


FIGURE 4

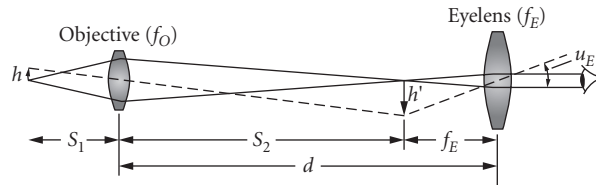


FIGURE 5

1.6 MAGNIFIERS AND MICROSCOPES

The conventional definition of magnifying power for either a magnifier or microscope compares the angular size of the image with the angular size of the object *when the object is viewed from a (conventional) distance of 10 inches*. Thus the magnification can be found from

$$MP = 10''/f \quad (21)$$

for either a simple microscope (i.e., magnifier) or a compound microscope, where f is the focal length of the system. Using the symbols of Fig. 5, we can also write the following for the compound microscope

$$MP = (f_E + f_O - d)10''/f_E f_O \quad (22)$$

$$\begin{aligned} MP &= m_O \times m_E \\ &= (S_2/S_1)(10''/f_E) \end{aligned} \quad (23)$$

1.7 AFOCAL ATTACHMENTS

In addition to functioning as a telescope, beam expander, etc., an afocal system can be used to modify the characteristics of another system. It can change the focal length, power, or field of the “prime” system. Figure 6 shows several examples of an afocal device placed (in these examples) before an imaging system. The combination has a focal length equal to the focal length of the prime system multiplied by the angular magnification of the afocal device. Note that in Fig. 6a and b the same afocal attachment has been reversed to provide two different focal lengths. If the size of the film or detector is kept constant, the angular field is changed by a factor equal to the inverse of the afocal magnification.

1.8 FIELD LENSES

Figure 7 illustrates the function of the field lens in a telescope. It is placed near (but rarely exactly at) an internal image; its power is chosen so that it converges the oblique ray bundle toward the axis sufficiently so that the rays pass through the subsequent component. A field lens is useful to keep the component diameters at reasonable sizes. It acts to relay the pupil image to a more acceptable location.

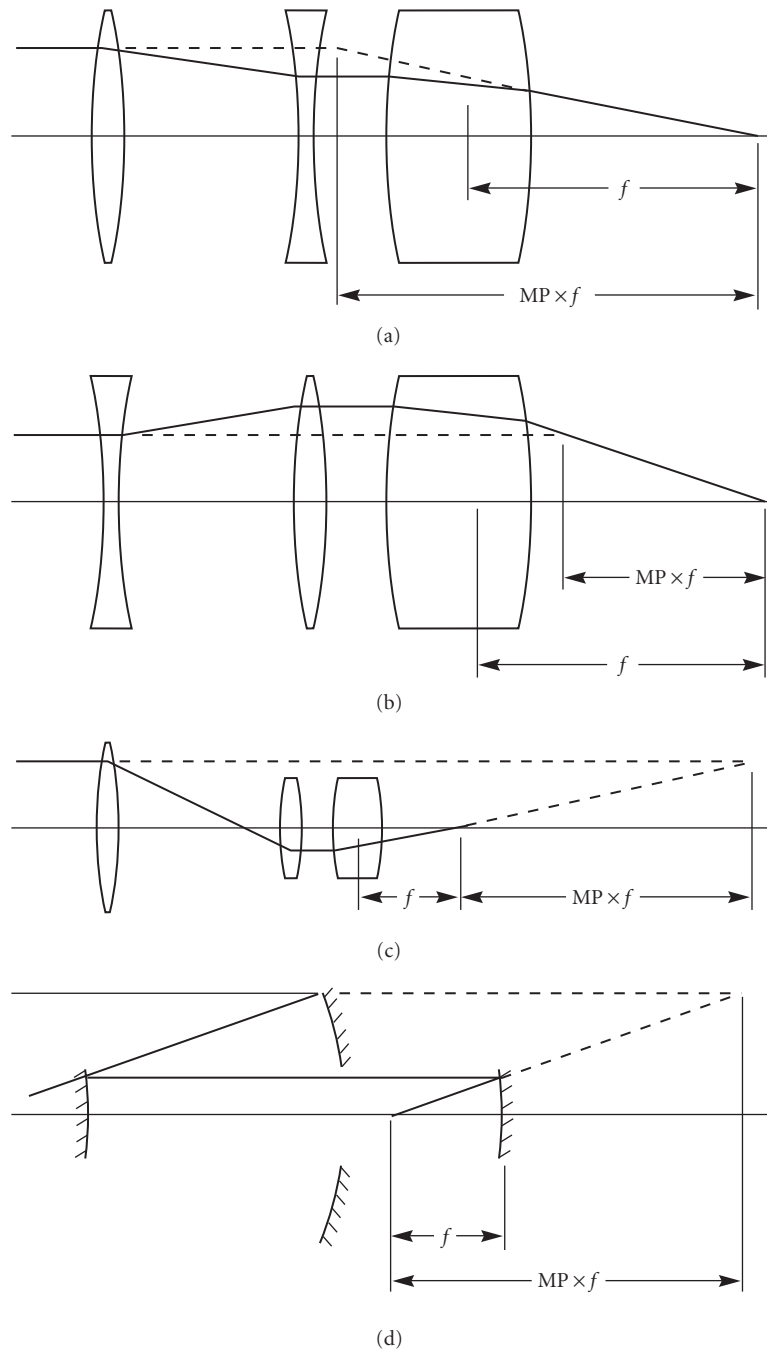


FIGURE 6

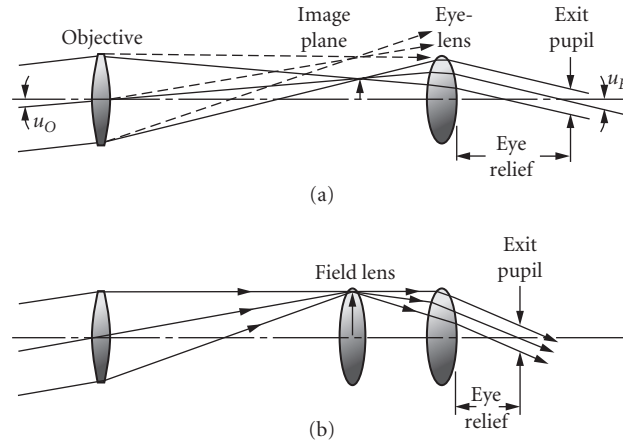


FIGURE 7

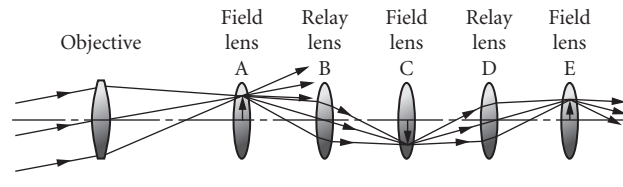


FIGURE 8

The required field lens power is easily determined. In Fig. 7 the most troublesome ray is that from the bottom of the objective aperture; its slope (u) is simply the height that it climbs divided by the distance that it travels. The required slope (u') for the ray after refraction by the field lens is defined by the image height (y), the “eyelens” semidiameter, and the spacing between them. Then Eq. (2) can be solved for the field lens power,

$$\phi = (u - u')/y \quad (24)$$

A periscope is used to carry an image through a long, small-diameter space. As shown in Fig. 8, the elements of a periscope are alternating field lenses and relay lenses. An optimum arrangement occurs when the images at the field lenses and the apertures of the relay lenses are as large as the available space allows. This arrangement has the fewest number of relay stages and the lowest power components. For a space of uniform diameter, both the field lenses and the relay lenses operate at unit magnification.

1.9 CONDENSERS

The projection/illumination condenser and the field lens of a radiation measuring system operate in exactly the same way. The condenser (Fig. 9) forms an image of the light source in the aperture of the projection lens, thereby producing even illumination from a nonuniform source. If the source

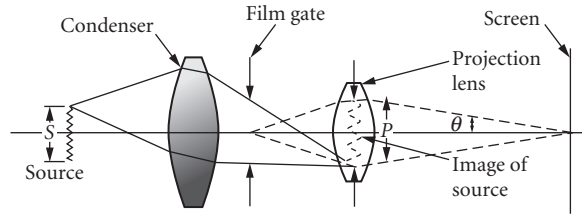


FIGURE 9

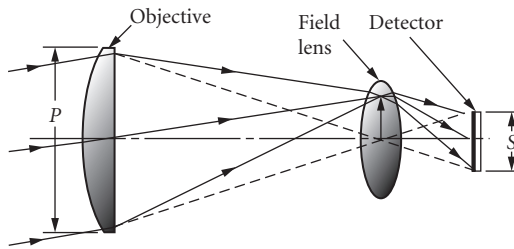


FIGURE 10

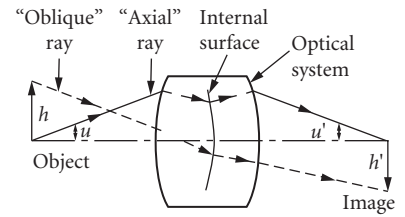


FIGURE 11

image fills the projection lens aperture, this will produce the maximum illumination that the source brightness and the projection lens aperture diameter will allow. This is often called Köhler illumination. In a radiometer type of application (Fig. 10), the field lens images the objective lens aperture on the detector, uniformly illuminating its surface and permitting the use of a smaller detector. Often, the smallest possible source or detector is desired in order to minimize power or maximize signal-to-noise. The smallest possible size is given by

$$S = P\alpha/2n \quad (25)$$

where S is the source or detector size, P is the projection lens or objective aperture diameter, α is the field angle of projection or the radiometer field of view, and n is the index in which the source or detector is immersed. This value for S corresponds to an (impractical) system speed of $F/0.5$. A source or detector size twice as large is a far more realistic limit, corresponding to a speed of $F/1.0$.

The invariant, $I = n(y_2u_1 - y_1u_2)$, where y_1 , u_1 , y_2 , and u_2 are the ray heights and slopes of two different rays, is an expression which has the same value everywhere in an optical system. If the two rays used are an axial ray and a principal (or chief) ray as shown in Fig. 11, and if the invariant is evaluated at the object and image surfaces, the result is

$$hnu = h'n'u' \quad (26)$$

1.10 ZOOM OR VARIFOCAI SYSTEMS

If the spacing between two components is changed, the effective focal length and the back focus are changed in accord with Eqs. (5) through (9). If the motions of the two components are arranged so that the image location is constant, this is a mechanically compensated zoom lens, so called because

1.12 DESIGN

the component motions are usually effected with a mechanical cam. A zoom system may consist of just the two basic components or it may include one or more additional members. Usually the two basic components have opposite-signed powers.

If a component is working at unit magnification, it can be moved in one direction or the other to increase or decrease the magnification. There are pairs of positions where the magnifications are m and $1/m$ and for which the object-to-image distance is the same. This is the basis of what is called a “bang-bang” zoom; this is a simple way to provide two different focal lengths (or powers, or fields of view, or magnifications) for a system.

1.11 ADDITIONAL RAYS

When the system layout has been determined, an “axial” ray at full aperture and a “principal” ray at full field can be traced through the system. Because of the linearity of the paraxial equations, we can determine the ray-trace data (i.e., y and u) of *any* third ray from the data of these two traced rays by

$$y_3 = Ay_1 + By_2 \quad (27)$$

$$u_3 = Au_1 + Bu_2 \quad (28)$$

where A and B are scaling constants which can be determined from

$$A = (y_3u_1 - u_3y_1)/(u_1y_2 - y_1u_2) \quad (29)$$

$$B = (u_3y_2 - y_3u_2)/(u_1y_2 - y_1u_2) \quad (30)$$

where y_1 , u_1 , y_2 , and u_2 are the ray heights and slopes of the axial and principal rays and y_3 and u_3 are the data of the third ray; these data are determined at any component of the system where the specifications for all three rays are known. These equations can, for example, be used to determine the necessary component diameters to pass a bundle of rays which are A times the diameter of the axial bundle at a field angle B times the full-field angle. In Fig. 12, for the dashed rays $A = +0.5$ and -0.5 and $B = 1.0$. Another application of Eqs. (27) through (30) is to locate either a pupil or an aperture stop when the other is known.

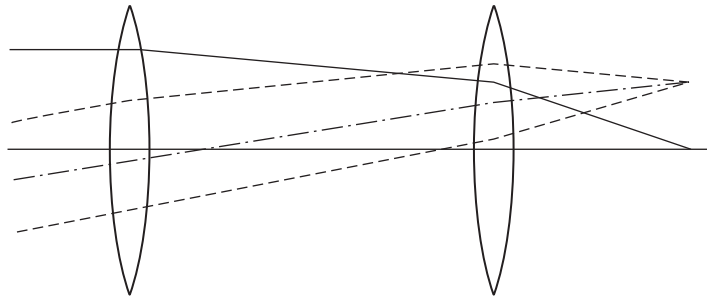


FIGURE 12

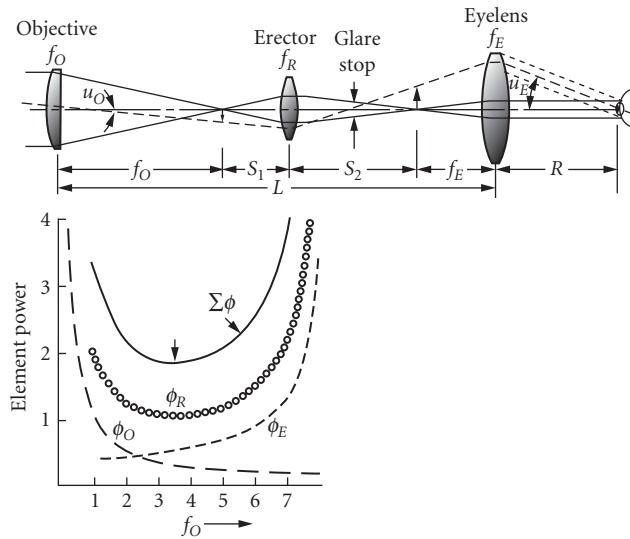


FIGURE 13

1.12 MINIMIZING COMPONENT POWER

The first-order layout may in fact determine the ultimate quality, cost, and manufacturability of the system. The residual aberrations in a system are a function of the component powers, relative apertures, and angular fields. The relationships are complex, but a good choice for a system layout is one which minimizes the sum of the (absolute) component powers, or possibly the sum of the (absolute) $y\phi$ product for all the components.

For example, in Fig. 13 the length, magnification, and the eye relief of the rifle scope are specified. There are five variables: three component powers and two spaces. This is one more variable than is necessary to achieve the specified characteristics. If we take the focal length of the objective component as the free variable, the component powers which satisfy the specifications can be plotted against the objective focal length, as in Fig. 13, and the minimum power arrangement is easily determined.

Minimizing the component powers will strongly tend to minimize the aberrations and also the sensitivity of the system to fabrication errors and misalignments. The *cost* of an optical element will vary with its diameter (or perhaps the square of the diameter) and also with the product of the diameter and the power. Thus, while first-order layout deals only with components, these relationships still apply reasonably well even when applied to components rather than elements. Minimizing the component powers does tend to reduce the cost on these grounds (and also because it tends to reduce the complexity of the components).

1.13 IS IT A REASONABLE LAYOUT?

A simple way to get a feel for the reasonableness of a layout is to make a rough scale drawing showing each component as single element. An element can be drawn as an equiconvex lens with radii which are approximately $r = 2(n - 1)f$; for an element with an index of 1.5 the radii equal the focal length. The elements should be drawn to the diameter necessary to pass the (suitably vignetted) off-axis

bundle of rays as well as the axial bundle. The on-axis and off-axis ray bundles should be sketched in. This will very quickly indicate which elements or components are the difficult ones. If the design is being started from scratch (as opposed to simply combining existing components), each component can be drawn as an achromat. The following section describes achromat layout, but for visual-spectrum systems it is often sufficient to assume that the positive (crown) element has twice the power of the achromat and the (negative) flint element has a power equal to that of the achromat. Thus an achromat may be sketched to the simplified, approximate prescription: $r_1 = -r_2 = f/2$ and $r_3 = \text{plano}$.

Any elements which are too fat must then be divided or “split” until they look “reasonable.” This yields a reasonable estimate of the required complexity of the system, even before the lens design process is begun.

If more or less standard design types are to be utilized for the components, it is useful to tabulate the focal lengths and diameters to get the (infinity) f -number of each component, and also its angular field coverage. The field coverage should be expressed both in terms of the angle that the object and image subtend from the component, and also the angle that the smaller of these two heights subtends as a function of the focal length (rather than as a function of that conjugate distance). This latter angle is useful because the coverage capability of a given design form is usually known in these terms, that is, h/f , rather than in finite conjugate terms. With this information at hand, a reasonable decision can be made as to the design type necessary to perform the function required of the component.

1.14 ACHROMATISM

The powers of the elements of an achromat can be determined from

$$\phi_A = \phi_{AB} V_A / (V_A - V_B) \quad (31)$$

$$\phi_B = \phi_{AB} V_B / (V_B - V_A) \quad (32)$$

$$= \phi_{AB} - \phi_A$$

where ϕ_{AB} is the power of the achromatic doublet and V_A is the Abbe V -value for the element whose power is ϕ_A , etc. For the visible spectral region $V = (n_d - 1)/(n_F - n_C)$; this can be extended to any spectral region by substituting the indices at middle, short, and long wavelengths for n_d , n_F , and n_C .

If the elements are to be spaced apart, and the back focus is B , then the powers and the spacing are given by

$$\phi_A = \phi_{AB} B V_A / (V_A B - V_B / \phi_{AB}) \quad (33)$$

$$\phi_B = -\phi_{AB} V_B / B (V_A B - V_B / \phi_{AB}) \quad (34)$$

$$D = (1 - B \phi_{AB}) / \phi_A \quad (35)$$

For a complete system, the transverse *axial chromatic* aberration is the sum of $y^2 \phi / V u'_k$ for all the elements, where y is the height of the axial ray at the element and u'_k is the ray slope at the image. The *lateral color* is the sum of $y y_p \phi / V u'_k$, where y_p is the principal ray height.

The *secondary spectrum* is the sum of $y^2 \phi P / V u'_k$, where P is the partial dispersion, $P = (n_d - n_c) / (n_F - n_c)$. Summed over two elements, this leads to an expression for the longitudinal secondary spectrum of an achromatic doublet

$$\begin{aligned} \text{SS} &= f(P_B - P_A) / (V_A - V_B) \\ &= -f(\Delta P / \Delta V) \end{aligned} \quad (36)$$

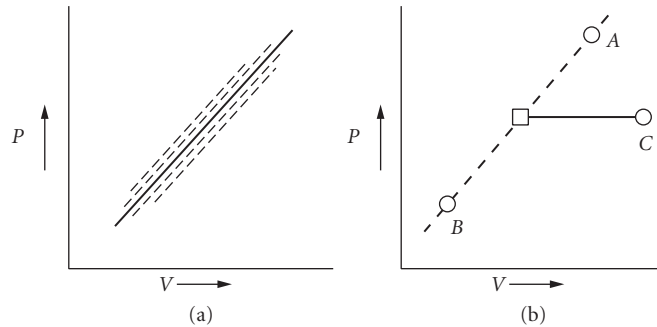


FIGURE 14

This indicates that in order to eliminate secondary spectrum for a doublet, two glasses with identical partial dispersions [so that $(P_A - P_B)$ is zero] are required. A large difference in V -value is desired so that $(V_A - V_B)$ in the denominator of Eqs. (31) and (32) will produce reasonably low element powers. As indicated in the schematic and simplified plot of P versus V in Fig. 14a, most glasses fall into a nearly linear array, and $(\Delta P/\Delta V)$ is nearly a constant for the vast majority of glasses. The few glasses which are away from the "normal" line can be used for apochromats, but the ΔV for glass pairs with a small ΔP tends to be quite small. In order to get an exact match for the partial dispersions so that ΔP is equal to zero, two glasses can be combined to simulate a third, as indicated in Fig. 14b. For a unit power ($\phi = 1$) apochromatic triplet, the element powers can be found from

$$X = [V_A(P_B - P_C) + V_B(P_C - P_A)] / (P_B - P_A) \quad (37)$$

$$\phi_C = V_C / (V_C - X) \quad (38)$$

$$\phi_B = (1 - \phi_C)(P_C - P_A)V_B / [V_B(P_C - P_A) + V_A(P_B - P_C)] \quad (39)$$

$$\phi_A = 1 - \phi_B - \phi_C \quad (40)$$

1.15 ATHERMALIZATION

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

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

$$d\phi/dt = -\phi[a - (dn/dt)/(n-1)] \quad (41)$$

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

$$d\phi/dt = \phi_A T_A + \phi_B T_B \quad (42)$$

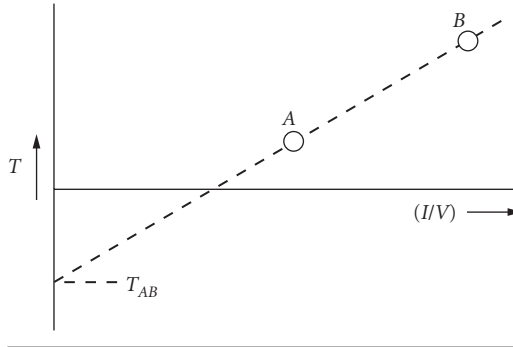


FIGURE 15

where

$$T = [-a + (dn/dt)/(n-1)] \quad (43)$$

and ϕ is the doublet power.

For an athermalized doublet (or one with some desired $d\phi/dt$) the element powers are given by

$$\phi_A = [(d\phi/dt) - \phi T_B]/(T_A - T_B) \quad (44)$$

$$\phi_B = \phi - \phi_A \quad (45)$$

To get an athermalized *achromatic* doublet, a plot of T against $(1/V)$ for all the glasses/materials under consideration is made. A line drawn between two glass points is extended to intersect the T axis as indicated in Fig. 15. Then 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 line intersects the T axis. A pair of glasses with a large V -value difference and a small or zero T axis intersection is desirable.

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

$$\phi_A = \phi V_A (T_B V_B - T_C V_C)/D \quad (46)$$

$$\phi_B = \phi V_B (T_C V_C - T_A V_A)/D \quad (47)$$

$$\phi_C = \phi V_C (T_A V_A - T_B V_B)/D \quad (48)$$

$$D = V_A (T_B V_B - T_C V_C) + V_B (T_C V_C - T_A V_A) + V_C (T_A V_A - T_B V_B) \quad (49)$$

See also Chap. 8, "Thermal Compensation Techniques," by Philip J. Rogers and Michael Roberts.

NOTE: Figures 2, 3, 4, 5, 7, 8, 9, 10, 11, and 13 are adapted from W. Smith, *Modern Optical Engineering*, 2nd ed., McGraw-Hill, New York, 1990. The remaining figures are adapted from *Critical Reviews of Lens Design*, W. Smith (Ed.), SPIE, vol. CR41, 1992.