

Performance Factors

After paraxial formulas have been used to select values for component focal length(s) and diameter(s), the final step is to select actual lenses. As in any engineering problem, this selection process involves a number of tradeoffs, including performance, cost, weight, and environmental factors.

The performance of real optical systems is limited by several factors, including lens aberrations and light diffraction. The magnitude of these effects can be calculated with relative ease.

Numerous other factors, such as lens manufacturing tolerances and component alignment, impact the performance of an optical system. Although these are not considered explicitly in the following discussion, it should be kept in mind that if calculations indicate that a lens system only just meets the desired performance criteria, in practice it may fall short of this performance as a result of other factors. In critical applications, it is generally better to select a lens whose calculated performance is significantly better than needed.

DIFFRACTION

Diffraction, a natural property of light arising from its wave nature, poses a fundamental limitation on any optical system. Diffraction is always present, although its effects may be masked if the system has significant aberrations. When an optical system is essentially free from aberrations, its performance is limited solely by diffraction, and it is referred to as diffraction limited.

In calculating diffraction, we simply need to know the focal length(s) and aperture diameter(s); we do not consider other lens-related factors such as shape or index of refraction.

Since diffraction increases with increasing f-number, and aberrations decrease with increasing f-number, determining optimum system performance often involves finding a point where the combination of these factors has a minimum effect.

ABERRATIONS

To determine the precise performance of a lens system, we can trace the path of light rays through it, using Snell's law at each optical interface to determine the subsequent ray direction. This process, called ray tracing, is usually accomplished on a computer. When this process is completed, it is typically found that not all the rays pass through the points or positions predicted by paraxial theory. These deviations from ideal imaging are called lens aberrations.

The direction of a light ray after refraction at the interface between two homogeneous, isotropic media of differing index of refraction is given by Snell's law:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \quad (1.20)$$

where θ_1 is the angle of incidence, θ_2 is the angle of refraction, and both angles are measured from the surface normal as shown in figure 1.14.

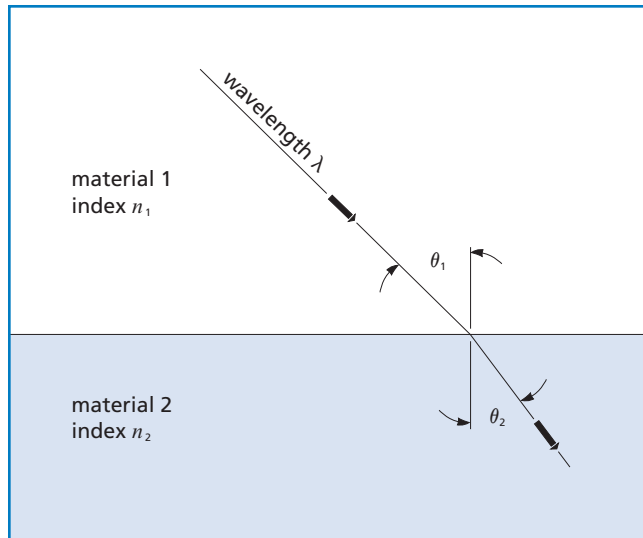


Figure 1.14 Refraction of light at a dielectric boundary

APPLICATION NOTE

Technical Assistance

Detailed performance analysis of an optical system is accomplished by using computerized ray-tracing software. CVI Melles Griot applications engineers are able to provide a ray-tracing analysis of simple catalog-component systems. If you need assistance in determining the performance of your optical system, or in selecting optimum components for your particular application, please contact your nearest CVI Melles Griot office.

Even though tools for the precise analysis of an optical system are becoming easier to use and are readily available, it is still quite useful to have a method for quickly estimating lens performance. This not only saves time in the initial stages of system specification, but can also help achieve a better starting point for any further computer optimization.

The first step in developing these rough guidelines is to realize that the sine functions in Snell's law can be expanded in an infinite Taylor series:

$$\sin \theta_1 = \theta_1 - \theta_1^3/3! + \theta_1^5/5! - \theta_1^7/7! + \theta_1^9/9! - \dots \quad (1.21)$$

The first approximation we can make is to replace all the sine functions with their arguments (i.e., replace $\sin \theta_1$ with θ_1 itself and so on). This is called first-order or paraxial theory because only the first terms of the sine expansions are used. Design of any optical system generally starts with this approximation using the paraxial formulas.

The assumption that $\sin \theta = \theta$ is reasonably valid for θ close to zero (i.e., high f-number lenses). With more highly curved surfaces (and particularly marginal rays), paraxial theory yields increasingly large deviations from real performance because $\sin \theta \neq \theta$. These deviations are known as aberrations. Because a perfect optical system (one without any aberrations) would form its image at the point and to the size indicated by paraxial theory, aberrations are really a measure of how the image differs from the paraxial prediction.

As already stated, exact ray tracing is the only rigorous way to analyze real lens surfaces. Before the advent of electronic computers, this was excessively tedious and time consuming. Seidel* addressed this issue by developing a method of calculating aberrations resulting from the $\theta_1^3/3!$ term. The resultant third-order lens aberrations are therefore called Seidel aberrations.

To simplify these calculations, Seidel put the aberrations of an optical system into several different classifications. In monochromatic light they are spherical aberration, astigmatism, field curvature, coma, and distortion. In polychromatic light there are also chromatic aberration and lateral color. Seidel developed methods to approximate each of these aberrations without actually tracing large numbers of rays using all the terms in the sine expansions.

In actual practice, aberrations occur in combinations rather than alone. This system of classifying them, which makes analysis much simpler, gives a good description of optical system image quality. In fact, even in the era of powerful ray-tracing software, Seidel's formula for spherical aberration is still widely used.

* Ludwig von Seidel, 1857.

SPHERICAL ABERRATION

Figure 1.15 illustrates how an aberration-free lens focuses incoming collimated light. All rays pass through the focal point F'' . The lower figure shows the situation more typically encountered in single lenses. The farther from the optical axis the ray enters the lens, the nearer to the lens it focuses (crosses the optical axis). The distance along the optical axis between the intercept of the rays that are nearly on the optical axis (paraxial rays) and the rays that go through the edge of the lens (marginal rays) is called longitudinal spherical aberration (LSA). The height at which these rays intercept the paraxial focal plane is called transverse spherical aberration (TSA). These quantities are related by

$$\text{TSA} = \text{LSA} \times \tan(u''). \quad (1.22)$$

Spherical aberration is dependent on lens shape, orientation, and conjugate ratio, as well as on the index of refraction of the materials present. Parameters for choosing the best lens shape and orientation for a given task are presented later in this chapter. However, the third-

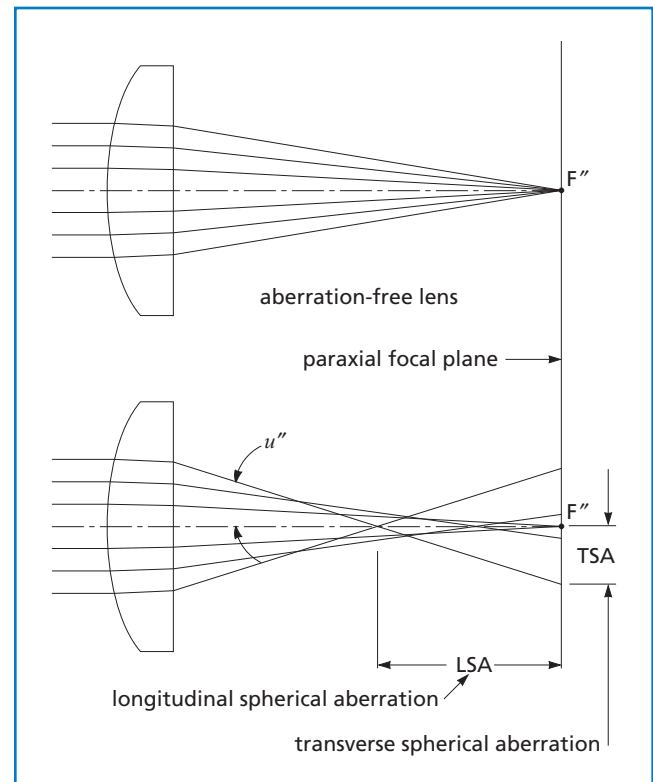


Figure 1.15 Spherical aberration of a plano-convex lens

order, monochromatic, spherical aberration of a plano-convex lens used at infinite conjugate ratio can be estimated by

$$\text{spot size due to spherical aberration} = \frac{0.067 f}{f/\#^3}. \quad (1.23)$$

Theoretically, the simplest way to eliminate or reduce spherical aberration is to make the lens surface(s) with a varying radius of curvature (i.e., an aspheric surface) designed to exactly compensate for the fact that $\sin \theta \neq \theta$ at larger angles. In practice, however, most lenses with high surface accuracy are manufactured by grinding and polishing techniques that naturally produce spherical or cylindrical surfaces. The manufacture of aspheric surfaces is more complex, and it is difficult to produce a lens of sufficient surface accuracy to eliminate spherical aberration completely. Fortunately, these aberrations can be virtually eliminated, for a chosen set of conditions, by combining the effects of two or more spherical (or cylindrical) surfaces.

In general, simple positive lenses have undercorrected spherical aberration, and negative lenses usually have overcorrected spherical aberration. By combining a positive lens made from low-index glass with a negative lens made from high-index glass, it is possible to produce a combination in which the spherical aberrations cancel but the focusing powers do not. The simplest examples of this are cemented doublets, such as the LAO series which produce minimal spherical aberration when properly used.

ASTIGMATISM

When an off-axis object is focused by a spherical lens, the natural asymmetry leads to astigmatism. The system appears to have two different focal lengths.

As shown in figure 1.16, the plane containing both optical axis and object point is called the tangential plane. Rays that lie in this plane are called tangential, or meridional, rays. Rays not in this plane are referred to as skew rays. The chief, or principal, ray goes from the object point through the center of the aperture of the lens system. The plane perpendicular to the tangential plane that contains the principal ray is called the sagittal or radial plane.

The figure illustrates that tangential rays from the object come to a focus closer to the lens than do rays in the sagittal plane. When the image is evaluated at the tangential conjugate, we see a line in the sagittal direction. A line in the tangential direction is formed at the sagittal conjugate. Between these conjugates, the image is either an elliptical or a circular blur. Astigmatism is defined as the separation of these conjugates.

The amount of astigmatism in a lens depends on lens shape only when there is an aperture in the system that is not in contact with the lens itself. (In all optical systems there is an aperture or stop, although in many cases it is simply the clear aperture of the lens element itself.) Astigmatism strongly depends on the conjugate ratio.

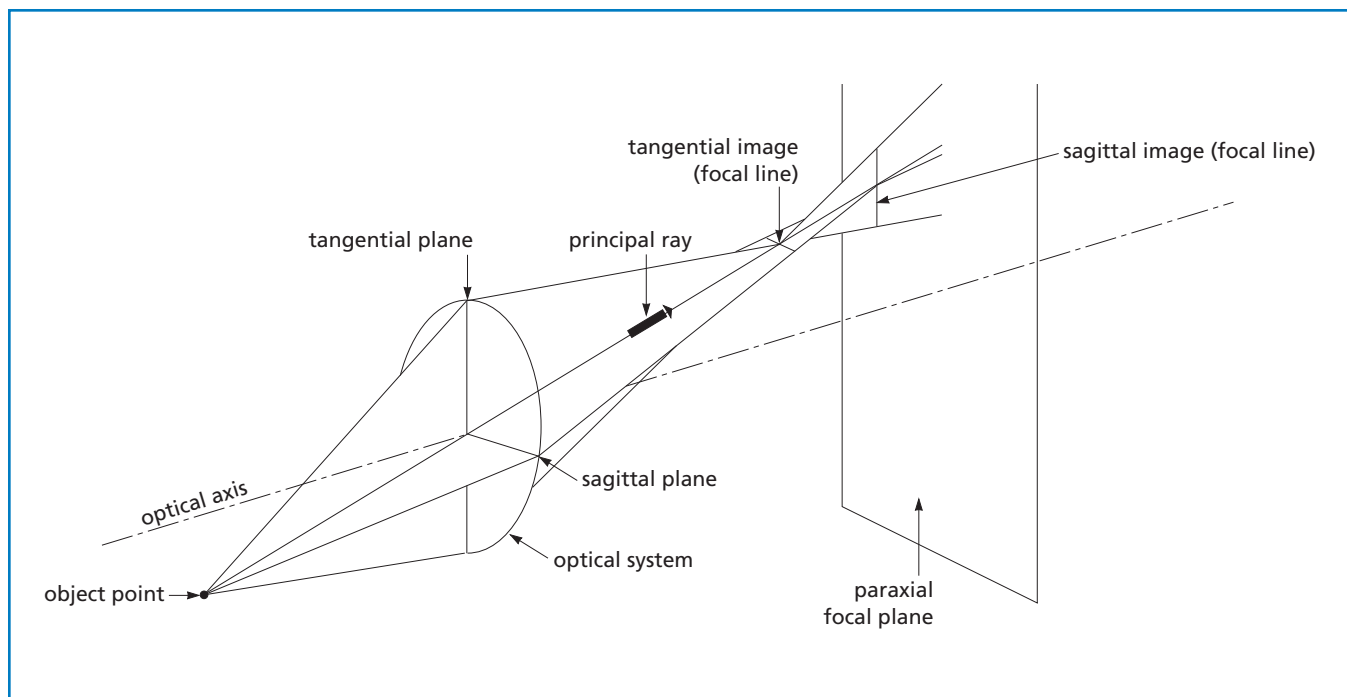


Figure 1.16 Astigmatism represented by sectional views

COMA

In spherical lenses, different parts of the lens surface exhibit different degrees of magnification. This gives rise to an aberration known as coma. As shown in figure 1.17, each concentric zone of a lens forms a ring-shaped image called a comatic circle. This causes blurring in the image plane (surface) of off-axis object points. An off-axis object point is not a sharp image point, but it appears as a characteristic comet-like flare. Even if spherical aberration is corrected and the lens brings all rays to a sharp focus on axis, a lens may still exhibit coma off axis. See figure 1.18.

As with spherical aberration, correction can be achieved by using multiple surfaces. Alternatively, a sharper image may be produced by judiciously placing an aperture, or stop, in an optical system to eliminate the more marginal rays.

FIELD CURVATURE

Even in the absence of astigmatism, there is a tendency of optical systems to image better on curved surfaces than on flat planes. This effect is called field curvature (see figure 1.19). In the presence of astigmatism, this problem is compounded because two separate astigmatic focal surfaces correspond to the tangential and sagittal conjugates.

Field curvature varies with the square of field angle or the square of image height. Therefore, by reducing the field angle by one-half, it is possible to reduce the blur from field curvature to a value of 0.25 of its original size.

Positive lens elements usually have inward curving fields, and negative lenses have outward curving fields. Field curvature can thus be corrected to some extent by combining positive and negative lens elements.

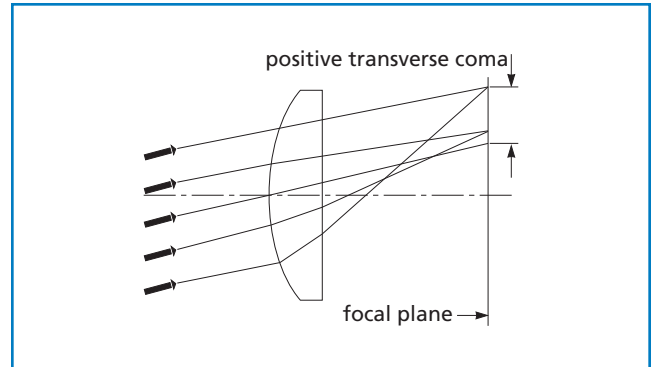


Figure 1.18 Positive transverse coma

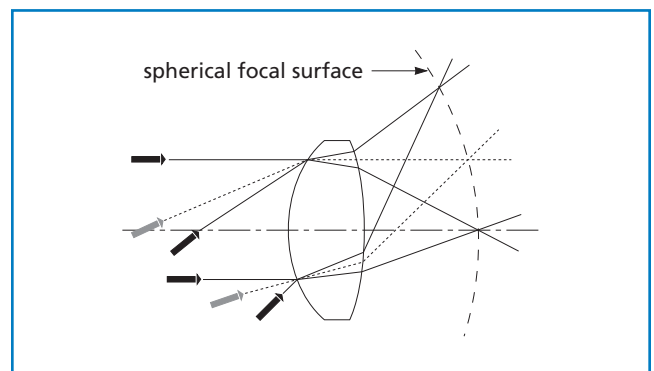


Figure 1.19 Field curvature

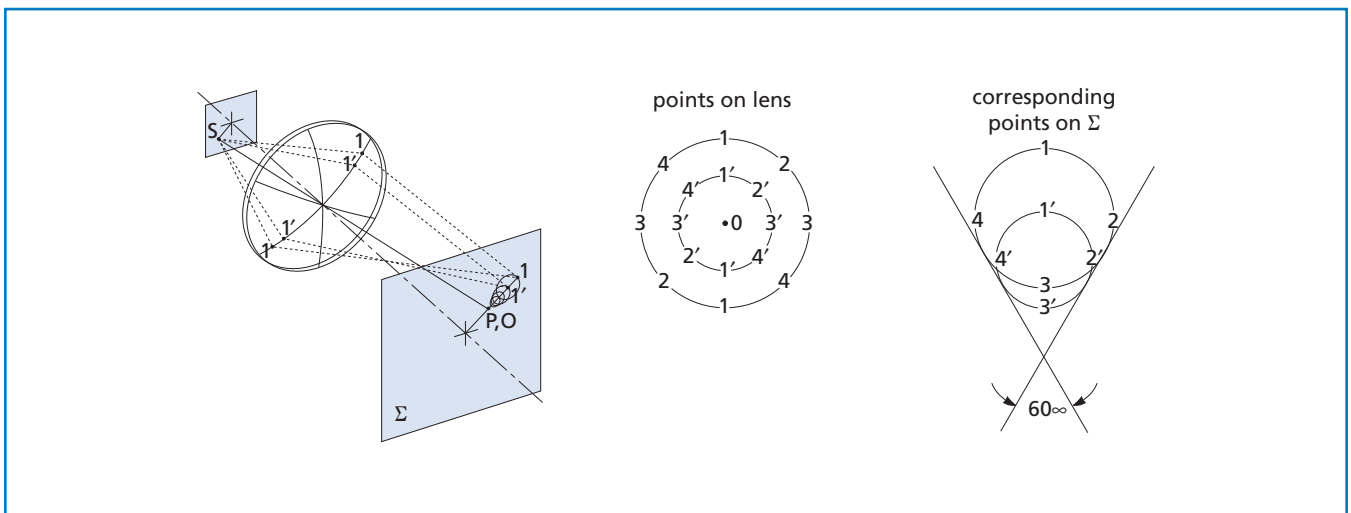


Figure 1.17 Imaging an off-axis point source by a lens with positive transverse coma

DISTORTION

The image field not only may have curvature but may also be distorted. The image of an off-axis point may be formed at a location on this surface other than that predicted by the simple paraxial equations. This distortion is different from coma (where rays from an off-axis point fail to meet perfectly in the image plane). Distortion means that even if a perfect off-axis point image is formed, its location on the image plane is not correct. Furthermore, the amount of distortion usually increases with increasing image height. The effect of this can be seen as two different kinds of distortion: pincushion and barrel (see figure 1.20). Distortion does not lower system resolution; it simply means that the image shape does not correspond exactly to the shape of the object. Distortion is a separation of the actual image point from the paraxially predicted location on the image plane and can be expressed either as an absolute value or as a percentage of the paraxial image height.

It should be apparent that a lens or lens system has opposite types of distortion depending on whether it is used forward or backward. This means that if a lens were used to make a photograph, and then used in reverse to project it, there would be no distortion in the final screen image. Also, perfectly symmetrical optical systems at 1:1 magnification have no distortion or coma.

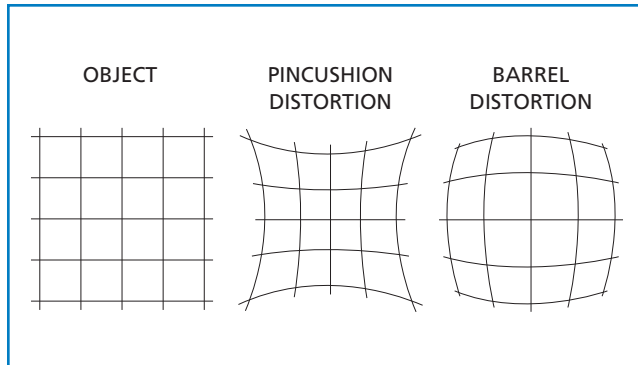


Figure 1.20 Pincushion and barrel distortion

CHROMATIC ABERRATION

The aberrations previously described are purely a function of the shape of the lens surfaces, and they can be observed with monochromatic light. Other aberrations, however, arise when these optics are used to transform light containing multiple wavelengths. The index of refraction of a material is a function of wavelength. Known as dispersion, this is discussed in *Material Properties*. From Snell's law (see equation 1.20), it can be seen that light rays of different wavelengths or colors will be refracted at different angles since the index is not a constant. Figure 1.21 shows the result when polychromatic collimated light is incident on a positive lens element. Because the index of refraction is higher for shorter wavelengths, these are focused closer to the lens than the longer wavelengths. Longitudinal chromatic aberration is defined as the axial distance from the nearest to the farthest focal point. As in the case of spherical aberration, positive and negative elements have opposite signs of chromatic aberration. Once again, by combining elements of nearly opposite aberration to form a doublet, chromatic aberration can be partially corrected. It is necessary to use two glasses with different dispersion characteristics, so that the weaker negative element can balance the aberration of the stronger, positive element.

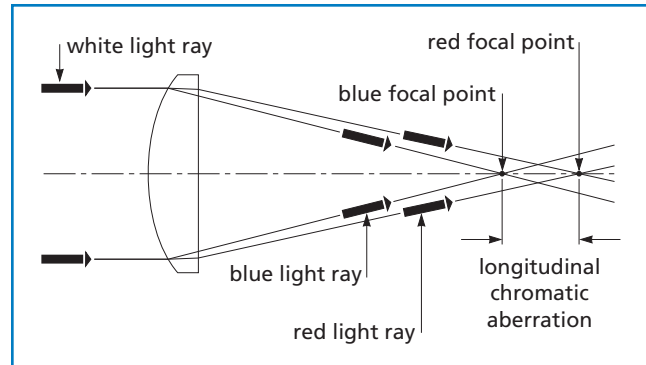


Figure 1.21 Longitudinal chromatic aberration

Variations of Aberrations with Aperture, Field Angle, and Image Height

Aberration	Aperture (ϕ)	Field Angle (θ)	Image Height (y)
Lateral Spherical	ϕ^3	—	—
Longitudinal Spherical	ϕ^2	—	—
Coma	ϕ^2	θ	y
Astigmatism	ϕ	θ^2	y^2
Field Curvature	ϕ	θ^2	y^2
Distortion	—	θ^3	y^3
Chromatic	—	—	—

LATERAL COLOR

Lateral color is the difference in image height between blue and red rays. Figure 1.22 shows the chief ray of an optical system consisting of a simple positive lens and a separate aperture. Because of the change in index with wavelength, blue light is refracted more strongly than red light, which is why rays intercept the image plane at different heights. Stated simply, magnification depends on color. Lateral color is very dependent on system stop location.

For many optical systems, the third-order term is all that may be needed to quantify aberrations. However, in highly corrected systems or in those having large apertures or a large angular field of view, third-order theory is inadequate. In these cases, exact ray tracing is absolutely essential.

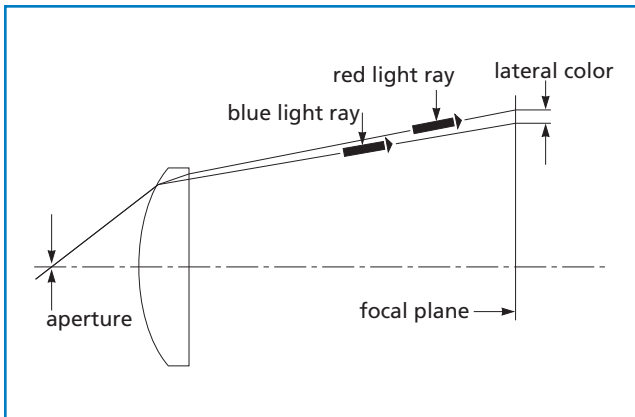


Figure 1.22 **Lateral Color**

APPLICATION NOTE**Achromatic Doublets Are Superior to Simple Lenses**

Because achromatic doublets correct for spherical as well as chromatic aberration, they are often superior to simple lenses for focusing collimated light or collimating point sources, even in purely monochromatic light.

Although there is no simple formula that can be used to estimate the spot size of a doublet, the tables in *Spot Size* give sample values that can be used to estimate the performance of catalog achromatic doublets.

