Astronomical Imaging and Photometry: Optics

A very brief overview of Optics

This will serve as an introduction to optics as relevant to issues arising in astronomical imaging. All of optics is subsumed in Fermat's principle, which states, in its simplest form, that light travels along a path between two points which takes the least time; from this one can derive Snell's law, the reflection laws, etc. The connection with waves is simply that the propagation vector of light (a ray) is proportional to the gradient of a wavefront which passes through a point. The optical path along a ray of light at a given wavelength is

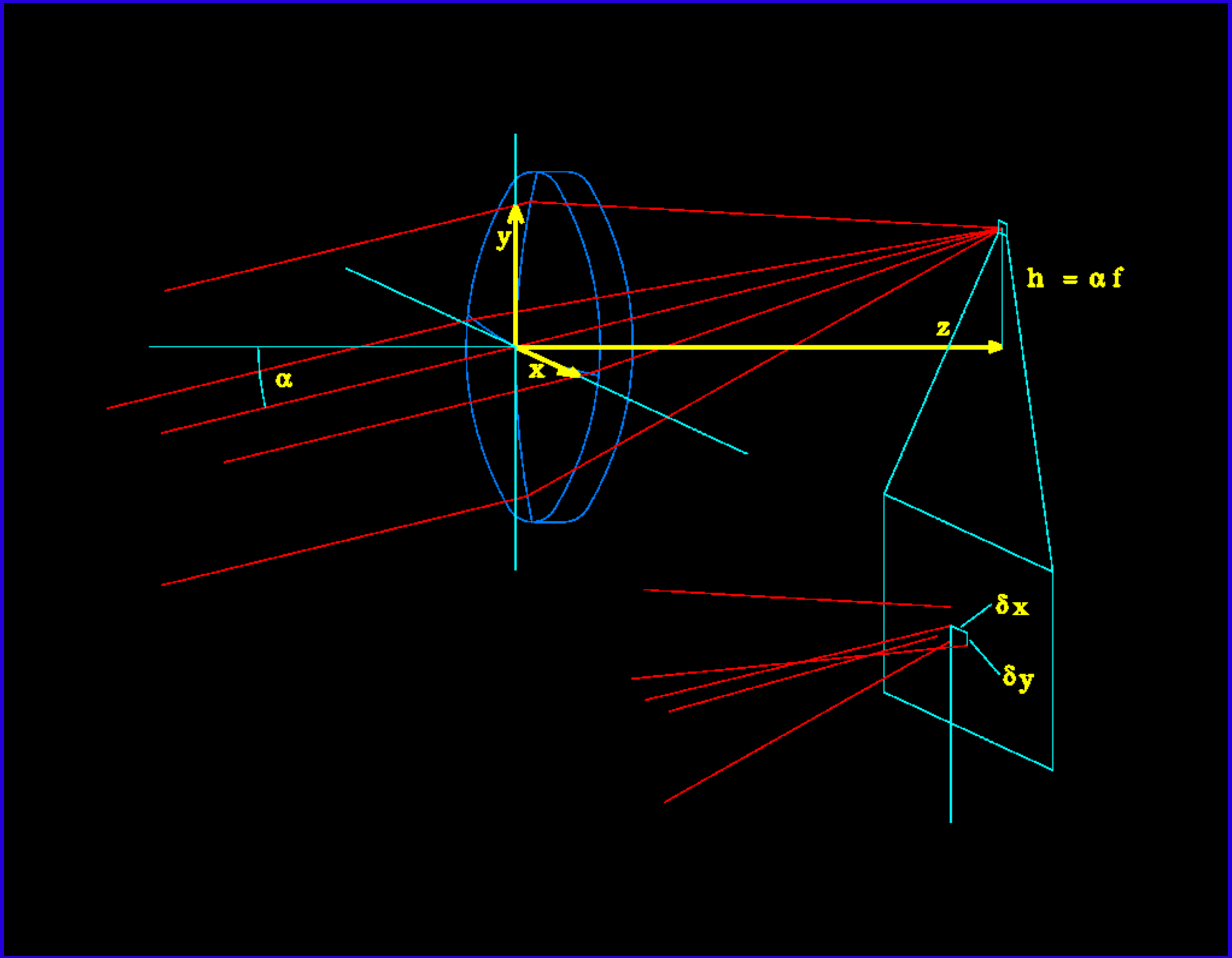
$$OP = \int n(l)\, dl \ \ (\text{cm}) \quad \text{or} \quad OP = \int n(l)\, dl/\lambda \ \ (\text{wavelengths})$$

if $n(l)$ is the refractive index along the path. The propagation vector is just grad OP

Optical systems can be almost arbitrarily complicated, consisting of lenses, mirrors, fibers, index-gradient elements, gratings, holograms, etc, etc. The aspects we are concerned with have to do with the behavior near the focal surface, which determines the properties of the image which falls on the detector and which we are to analyze. It is sufficient here to concentrate on these aspects. We will specialize to circularly symmetric optical systems for simplicity, and at first deal with the geometric optics limit, in which we do not consider interference effects, but just assume that light travels strictly along the optical path and that the intensity is preserved along the path.

The light from a star at very great distances defines a parallel bundle of rays (perpendicular to flat wavefronts) which eventually reach the focal surface. Bundles from different points in the field have central (principal) rays which all intersect to some approximation; this intersection defines the location of the ENTRANCE PUPIL. Likewise, bundles reaching the focal surface have central rays which intersect to some approximation; this intersection defines the EXIT PUPIL. For single simple thin lenses, they are essentially coincident, but are not for any interesting real optical system.

Since the system is assumed circularly symmetric, we can choose the plane containing the entering angle arbitrarily, and we choose it to be the y,z plane, in which z is the optical axis and we erect normal x, y axes in the pupil. The entering angle is α, and the intersection of the ray with the pupil at point x,y completely defines the situation. We show here a simple lens, but the description is completely general.



In first-order optical theory, the approximation is made for all angles γ which rays and the normals to surfaces make with the optical axis that

$$\sin\gamma = \tan\gamma = \gamma .$$

In this case, it is easy to see that optical systems make perfect images; all angle differences upon passing through any surface are just proportional to the radius in the pupil, so those rays will all converge at the same image point.

First-order optics, though not useful at all in examining the quality of images, is useful in defining several global properties of the system.

The height h in the image plane is linear with the field angle α ; the constant of proportionality is the FOCAL LENGTH f:

$$h = \alpha f$$

The ratio of the focal length to the diameter D of the entrance pupil is called the f-ratio F. In the first-order approximation, the angle which a ray coming through the edge of the pupil makes with the principal ray with the same field angle is the RADIUS of the pupil divided by the focal length, 1/(2F). This is called the NUMERICAL APERTURE (NA). Systems with small f-ratios and large NA are called FAST; large f-ratios and small NAs SLOW.

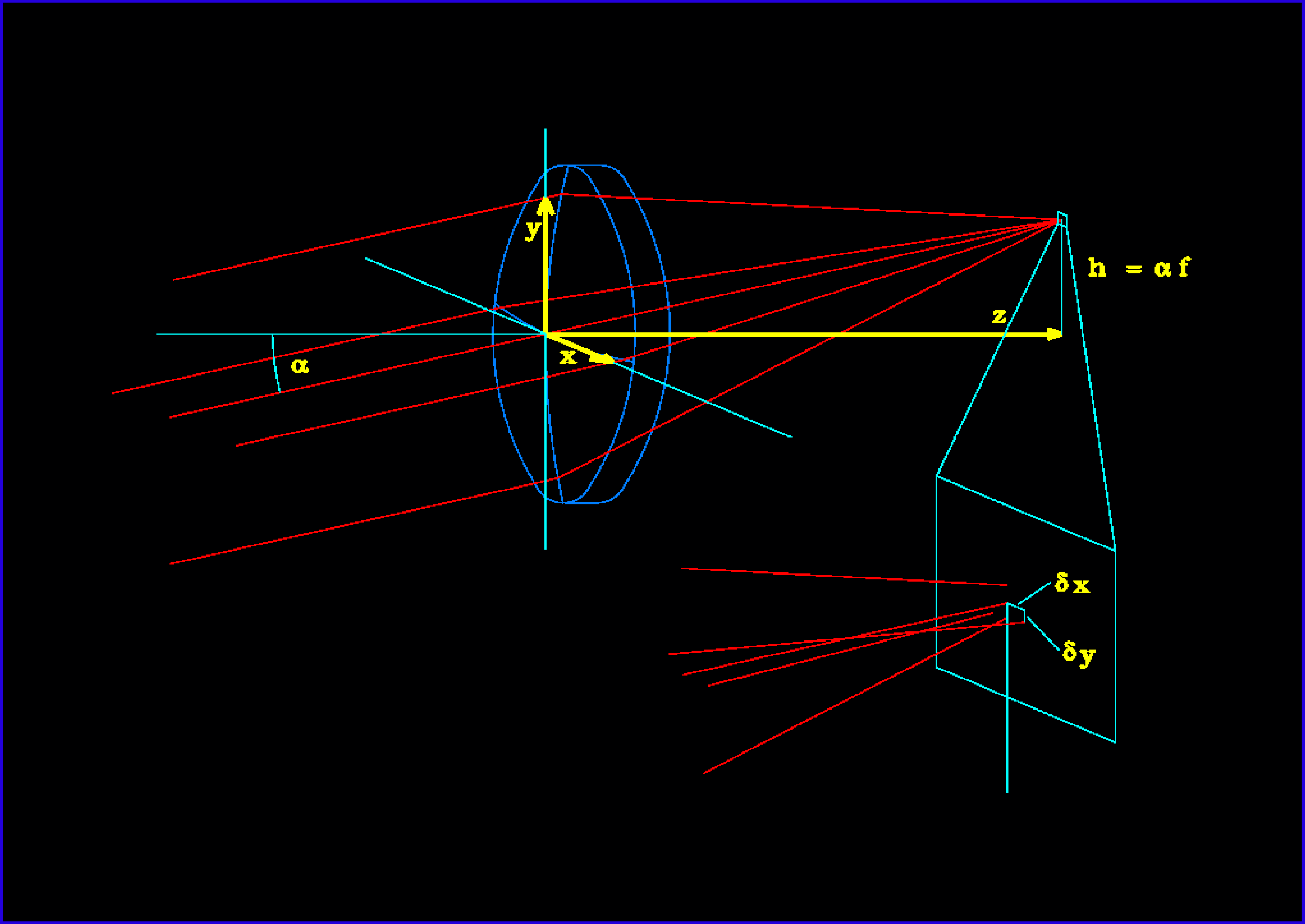
There is now more than two hundred years of experience analyzing optical systems to the next order, in which

$$\sin\gamma = \gamma - \gamma^3/6, \quad \cos\gamma = 1 - \gamma^2/2, \quad \tan\gamma = \gamma + \gamma^3/3$$

This analysis is called, of course, THIRD-ORDER OPTICS, and, as might be expected, is enormously richer than first-order optics. It is adequate for many systems, but fails for very fast or very wide-field systems in which the angles become large enough that fifth and higher-order terms become important. With the advent of fast computing, higher-order optical analysis is essentially obsolete and has been replaced by essentially exact ray-tracing. (ZEMAX, etc)

It is in third order that the imperfections in image formation, called in general ABERRATIONS, first become apparent, and we will here discuss only the classical third-order aberrations.

Let the principal ray have field angle α; it crosses the first-order image plane with height h. In third order, rays entering the pupil with the same field angle but not at the center will NOT, in general, cross the image plane at the same point as the principal ray, but will deviate from that point by the small quantities d_x and d_y in the directions of the x and y axes. These terms will be linear combinations of terms which are third-order in α and in the angles which the rays make to the principal ray coming to the focal plane; those are essentially $\xi = x/R$ and $\eta = y/R$, and we will do the expansion in x and y. Because the angles are gradients of the OP and the obvious symmetries of the geometry, the form of the third-order expansion is not quite arbitrary, and looks like this:



$$\delta x \;=\; (c - a)\alpha^2 x \;+ 2C\alpha xy \;+\; sr^2 x$$

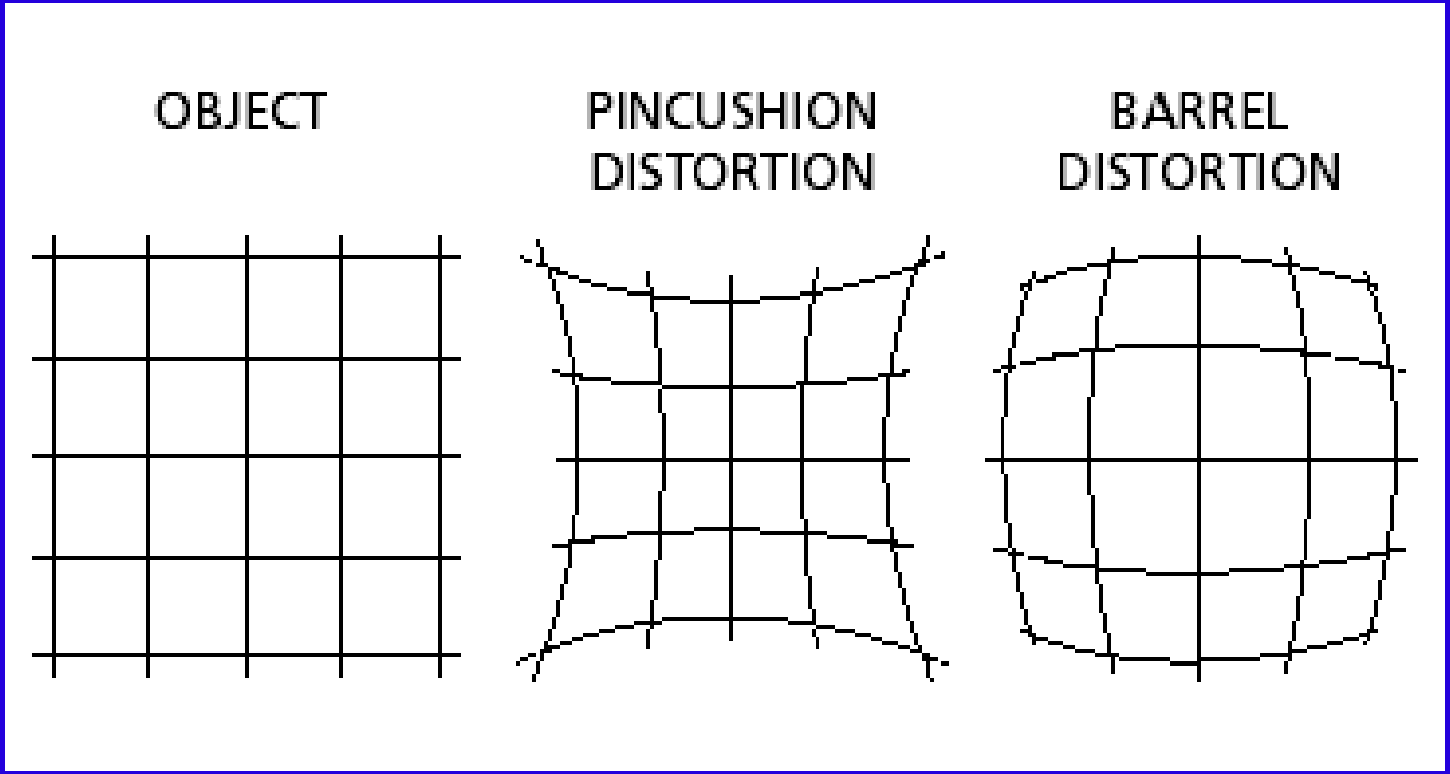
$$\delta y \;= d\alpha^3 \;+ (c + a)\alpha^2 y \;+ C\alpha(x^2 + 3y^2) \;+ sr^2 y$$

$$r = sqrt(x^2 + y^2)$$

The terms have names: The first term is called DISTORTION

$$\delta x = \qquad (c - a)\alpha^2 x + 2C\alpha\, xy + sr^2 x$$
$$\delta y = \underline{d\alpha^3} + (c + a)\alpha^2 y + C\alpha\,(x^2 + 3y^2) + sr^2 y$$

distortion causes NO image degradation, since it depends only on the field angle and is always radial. It DOES cause astrometric errors, since it clearly moves the centroids (indeed, bodily the images) of objects. d>0 is called PINCUSHION DISTORTION; d<0 BARREL DISTORTION.



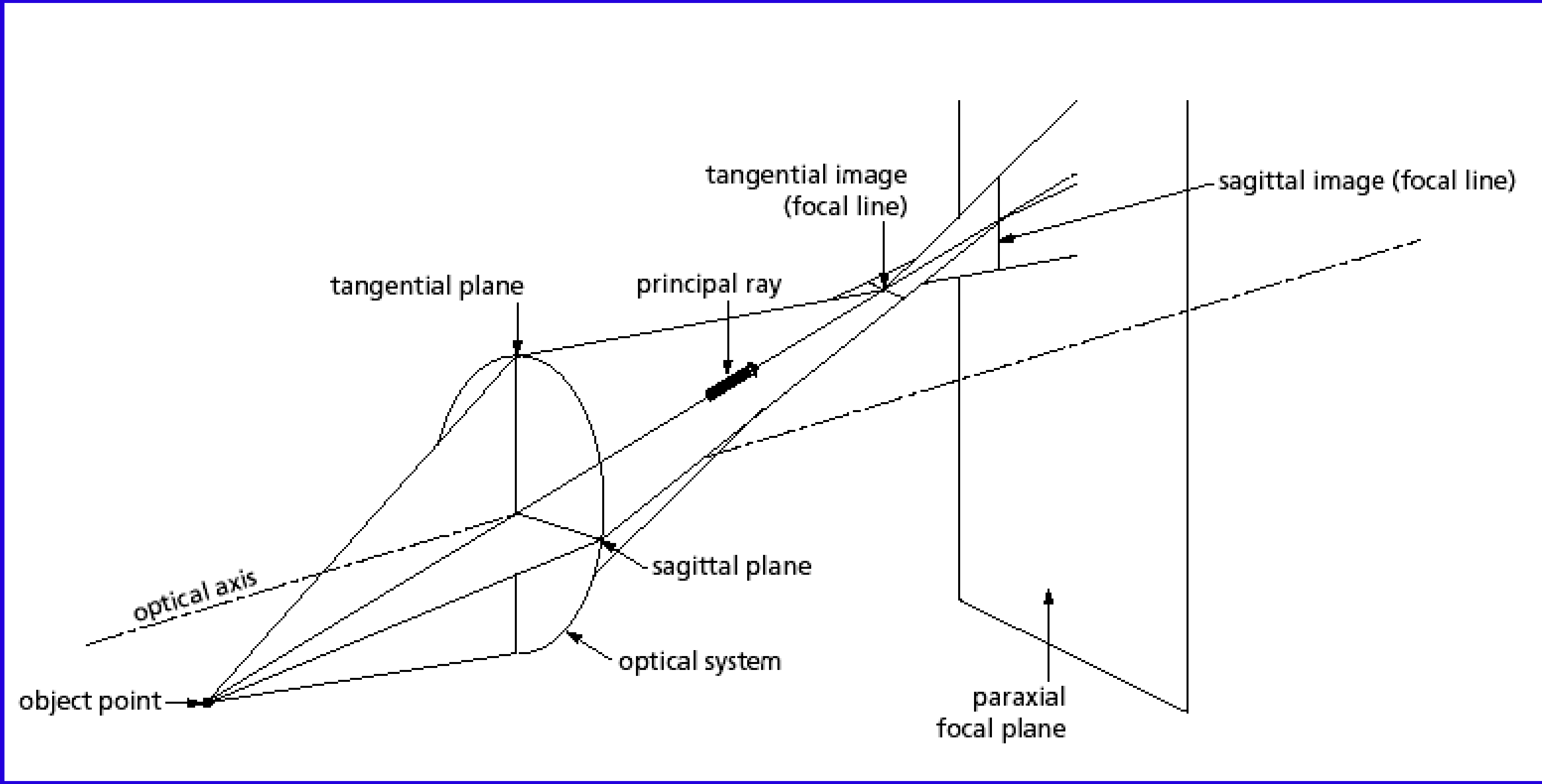
The second term has two parts.

$$\delta x \;=\; \qquad\qquad (c - a)\alpha^2 x \;+ 2C\alpha\, xy \qquad\qquad + \; sr^2 x$$
$$\delta y \;= d\alpha^3 \;+ \underline{(c + a)\alpha^2 y} \;+ C\alpha(x^2 + 3y^2) \;+ sr^2 y$$

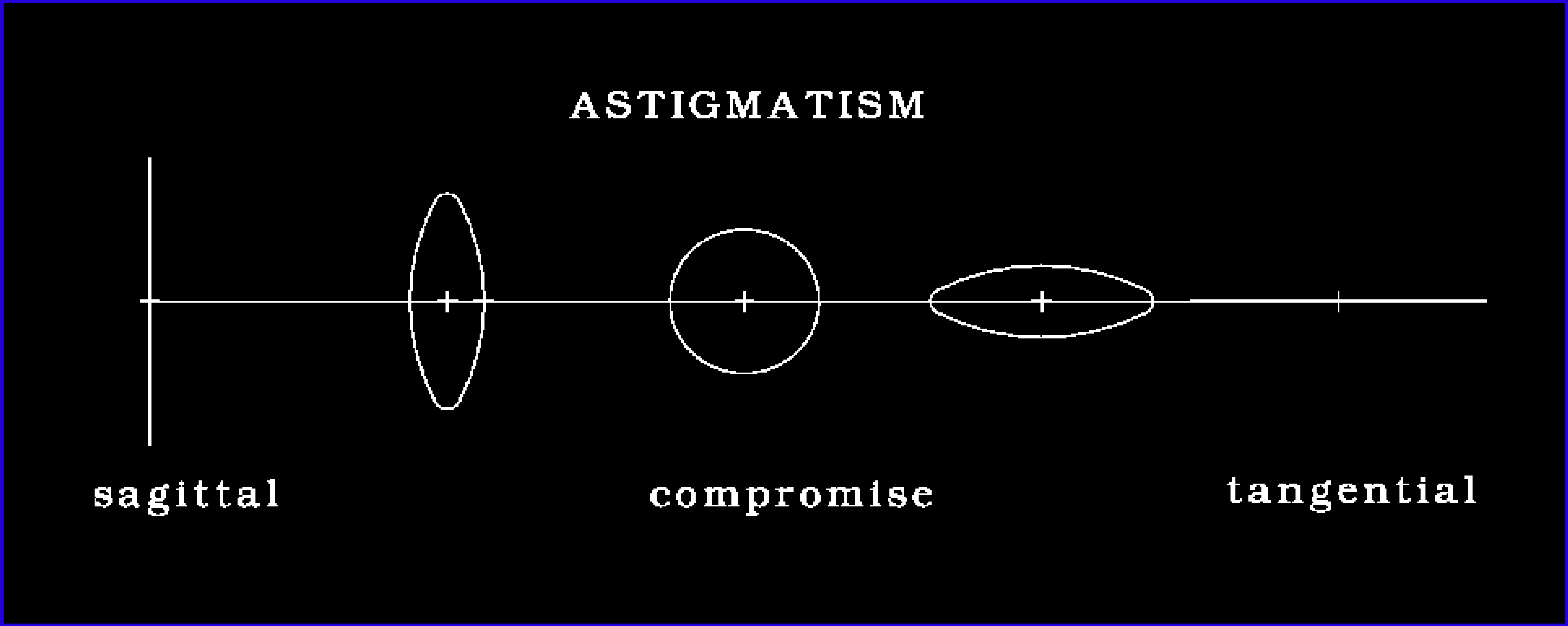
The first (c) part clearly represents a curvature of the focal surface. The deviation is proportional to the coordinate in the pupil, i.e. to the angle the ray makes coming to the focal surface, so they will come to a focus SOMEWHERE ELSE, at a distance from the first-order focal surface which is proportional to the square of the field angle a. This is called <u>FIELD CURVATURE</u> and does not represent image degradation if one can curve the detector.

The second (a) part clearly represent DIFFERENT field curvatures for sets of rays for which x=0, ie which enter in the y,z plane (tangential rays) , and sets of rays which enter in the x,z plane (sagittal rays). If a is nonzero, the system has ASTIGMATISM.

A system with nonzero astigmatism does not make perfect images. At a focal position which is halfway between the sagittal and tangential foci, the image is round but not a point. At the tangential focus, the image is a horizontal line, and at the sagittal focus, a vertical line; at other foci, the image is an ellipse.



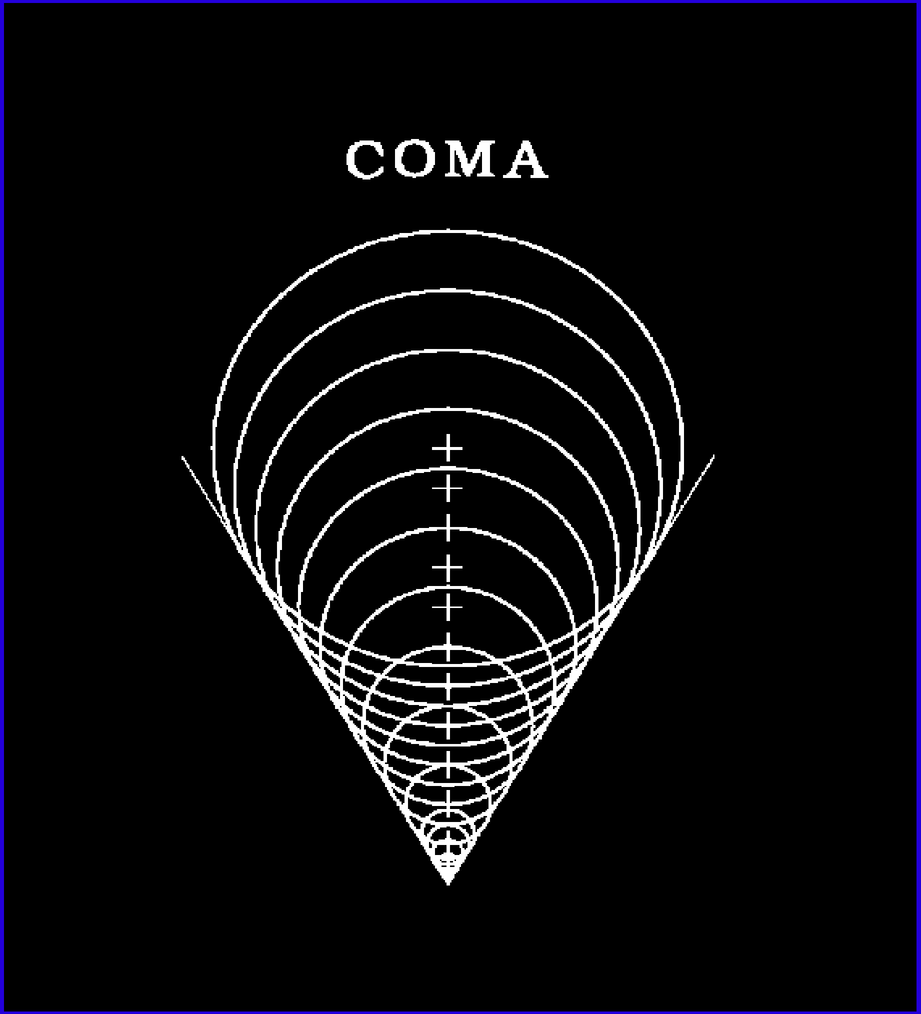
As you go through focus, the images look like this:



The next term, linear in the field angle and quadratic in the pupil coordinates, is called COMA.

$$\delta x = (c - a)\alpha^2 x + 2C\alpha xy + sr^2 x$$
$$\delta y = d\alpha^3 + (c + a)\alpha^2 y + \underline{C\alpha\,(x^2 + 3y^2)} + sr^2 y$$

It can easily be shown that this aberration maps circles in the pupil into circles in the image plane, with radii and deviations in the central position which are quadratic in the radius in the pupil. Thus the circles lie in a cone on the focal plane:

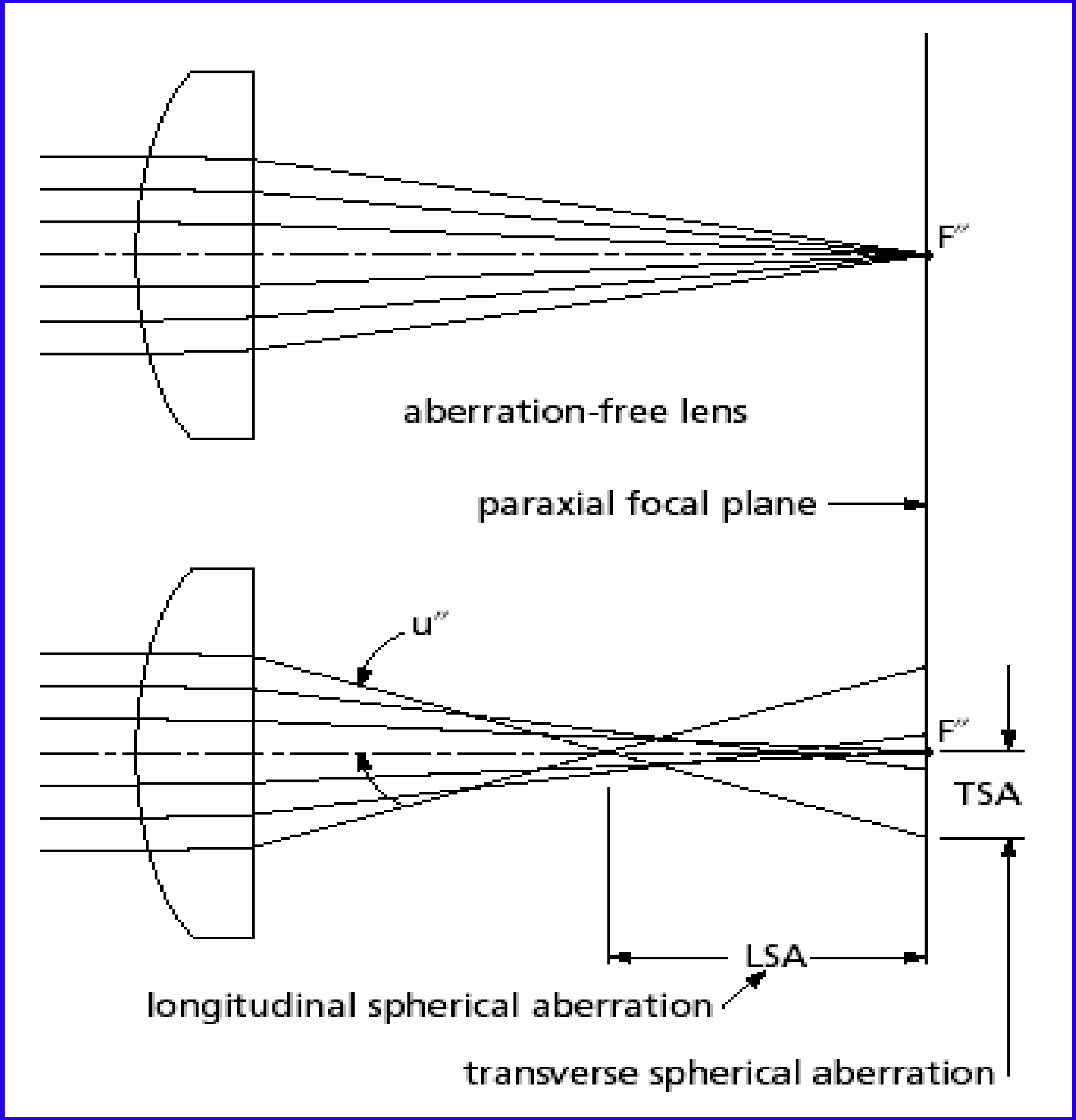


The last term is circularly symmetric in the focal plane, and is called SPHERICAL.

$$\delta x = (c - a)\alpha^2 x + 2C\alpha\, xy + sr^2 x$$
$$\delta y = d\alpha^3 + (c + a)\alpha^2 y + C\alpha\,(x^2 + 3y^2) + \underline{sr^2 y}$$

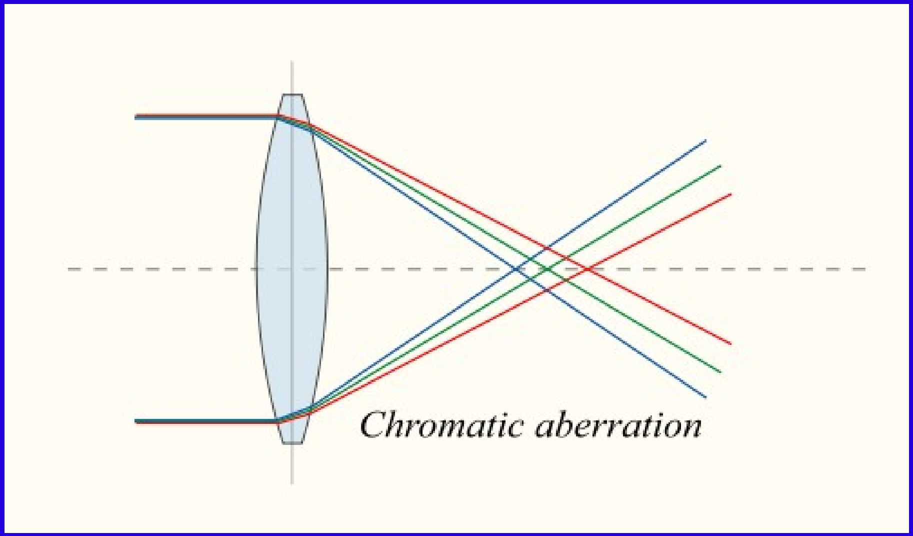
Since the deviation is in the direction of the pupil vector and is just proportional to the cube of the distance of the ray from the center, the rays from a given radius in the pupil fall on a circle in the focal plane with the center at the first-order focal position. The radius of the circle is proportional to the cube of the radius in the pupil, but the aberration is uniform over the focal plane; there is no α -dependence. Spherical aberration is insidious (ask the Hubble folks). There is no really good compromise focus, since most of the area of the system is at large pupil radii, where the focus is changing most rapidly.

SPHERICAL ABERRATION

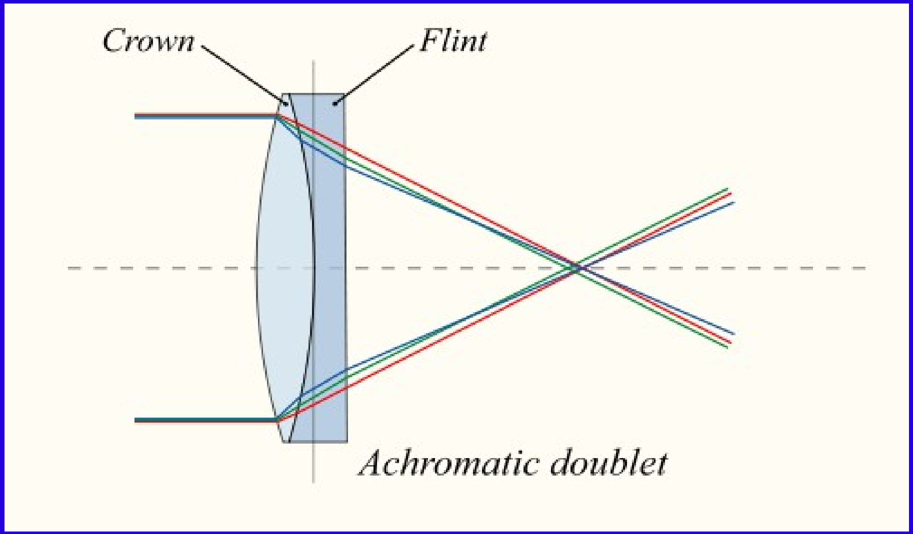


CHROMATIC ABERRATION

All glasses have refractive indices which decrease with wavelength, so simple lenses have focal lengths which depend on wavelength:



This can be `fixed' by using two lenses with different indices and dn/dl Such a lens is called an `achromat'. Many glasses can be used in complex systems.



THE DIFFRACTION LIMIT

In the previous discussion, we have considered light as propagating along rays with no wave interference effects. In this limit, a star image in a system with no aberrations is a point, or in any case, the real geometrical size of geometrical image, a few microarcseconds for stars at typical galactic distances. Consideration of the wavefronts through a system lead one to the conclusion that in a perfect optical system, the OP is the same for all rays at the focus. Real optical systems with real light waves do not make point images. If one considers rays entering at very small angles to the geometric path to a star, the OP for such rays differ by a small fraction of a wavelength from the rays following the exact geometric path. These paths are thus essentially equivalent, and it is only at angles such that the interference of waves along the perturbed path interfere destructively with waves along the geometric path is the intensity substantially reduced.

Consider an telescope with an entrance pupil diameter (typically the primary mirror diameter in a reflecting telescope) D, with a star image from a star at infinity, say, on the optical axis. Consider a point in the focal plane a short distance away from the geometric star image, corresponding geometrically to an angle a away from the star on the sky. There are rays from this point back to the pupil, but they are clearly not quite the rays dictated by Fermat's principle; in particular, they do not have zero optical path difference to the star, as rays at the geometric focus do. The PHASE difference between one of these rays and the rays which define the geometric center is the angle α times the pupil height y (the physical path difference) times $2\pi/\lambda$, $d\phi = 2\pi\alpha y/\lambda$, Thus it is plausible that the electric field in the focal plane, which is just the superposition of the fields propagating along possible rays, is

$$E(\alpha) = \text{Const.} \int \exp(i\, 2\pi\alpha y/\lambda)\, dxdy \qquad \text{over the pupil.}$$

In general, αy becomes $\vec{\alpha} \cdot \vec{x}$, and the integral is just the Fourier transform of the pupil, with argument $\vec{k} = 2\pi\alpha/\lambda$

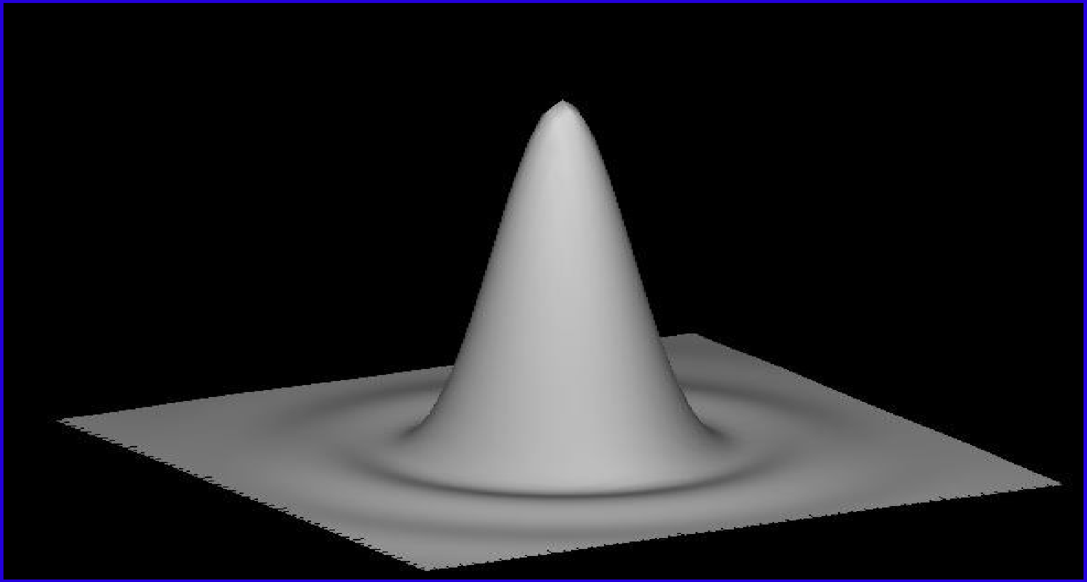
Note that the INTENSITY is proportional to the SQUARE of the electric field, so the intensity is proportional to the square of the Fourier transform of the entrance pupil.

What does this look like? If the angle a is λ/D, then the OP changes from one side of the aperture to the other by 2π, and there is complete destructive interference in the focal plane.

If the pointing is changed by $\lambda/2D$, the effect is roughly half as big, so one expects the half width at half maximum to be at roughly this angle, and the FWHM to be about λ/D.

The expression for the intensity in the focal plane for a perfect system with a uniform circular pupil is the square of the Fourier transform of a uniform disk of diameter D, the AIRY function:

$$P_0(\vec{\alpha}) = \frac{\pi D^2}{4\lambda^2}\left[\frac{2J_1(\pi D|\vec{\alpha}|/\lambda)}{\pi D|\vec{\alpha}|/\lambda)}\right]^2,$$



This result holds for a perfect circular optical system with no aberrations and no central obstruction. In general the diffraction PSF is the square of the Fourier transform of the pupil, AND if there are phase imperfections in the incoming beam, from the atmosphere or imperfect optics, those phase differences as functions of location $\vec{x}$ in the pupil are simply incorporated into the transform, and the result is correct. Amplitude modulation from coating imperfections or on-purpose apodization across the pupil can likewise be incorporated.

The diffraction resolution limit, λ/D, is, numerically,

$\gamma_d = 0.21\ \lambda(\mu)/D(m)$ ***arcseconds***

A 4-meter telescope in the visible has a diffraction limit of about 25 milliarcseconds, but seeing, which we will discuss in the next lecture, limits resolution in even excellent conditions at excellent sites to an order of magnitude worse than this, so for many if not most, purposes we do not need to worry about diffraction very much in seeing-limited observing from the ground, EXCEPT for one thing:

Big telescopes are all reflecting systems, and either have a secondary mirror in the beam or have a focal surface structure (correctors, detectors, etc) in the beam. This means that the pupil has a hole in it, but this is generally not very significant, given the tiny scale of the diffraction image compared to seeing. What IS nearly always significant is that the central structure is supported by some kind of spider made of thin rods or vanes. Since the scale of these structures is very small in their thin dimension compared to the size of the pupil, there is little interference between the energy they diffract and the diffraction of the pupil. Under these circum-stances, it is easy to see that since the Fourier transform of the real pupil is that of the pupil without the spider MINUS the FT of the spider, and the cross term in the square, which represents the interference between them, is negligible, the fraction of the energy *scattered* by the spider is the same as that blocked by the spider (this is a manifestation of the same OPTICAL THEOREM you have probably met in quantum mechanics). The angles involved are then the diffraction limit associated with the WIDTH of the spider vanes, which is typically of the order of one or a few centimeters, in the direction perpendicular to the spider vane,

and a size similar to the diffraction limit of the telescope in the direction along the vane. This small dimension is enlarged by seeing, and the typical result is a pair of spikes emerging from a star image for each supporting vane, perpendicular to the long direction of the vane, with a width determined by the seeing and a length of rougly the diffraction limit associated with the centimeter or so width of the vane, i. e. of order ten arcseconds. The intensity distribution along the diffraction spike is just the square of the Fourier transform of a tophat the width w of the vane,

$$I = Const\ (\lambda/2\pi\alpha)^2 \sin^2(\pi\alpha w/\lambda)$$

so has a core of width of order $2\lambda/w$, and falls like α^{-2} outside this. The wavelength dependence can be seen graphically on HST color images, where one sees the rainbow along the spikes. Remember that the total fraction of energy in the spikes is the fraction of the area obscured by them.

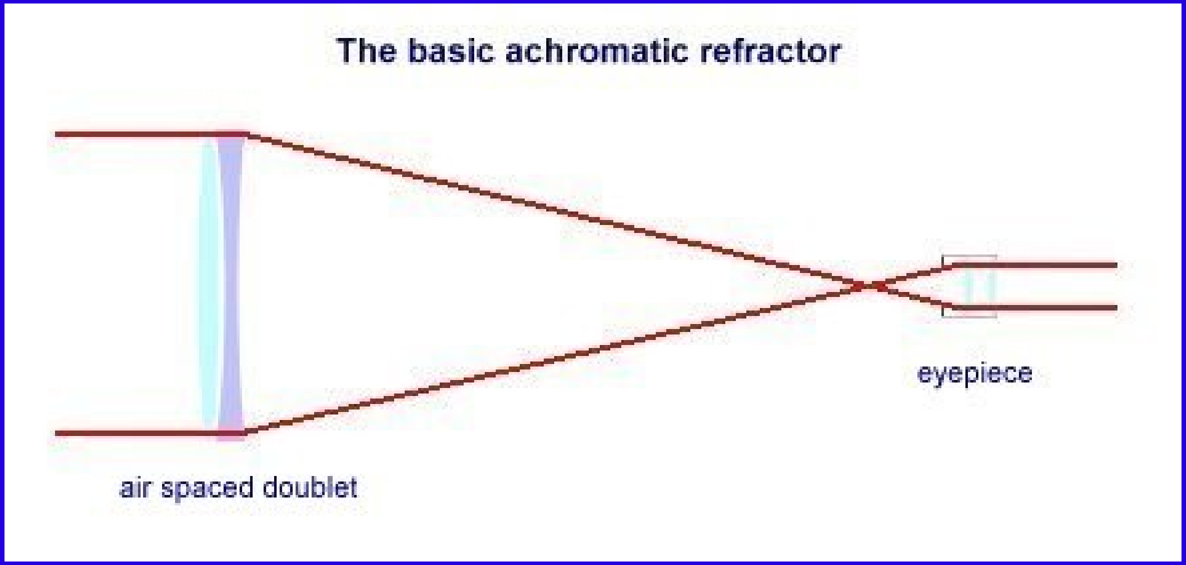
These spikes are a real nuisance in image processing, and can give rise to many spurious object detections close to bright stars if one is not very, very careful.

Image of center of dust nebula around V838 Mon near maximum, showing modulation of diffraction spikes.

Telescopes and Optical Systems

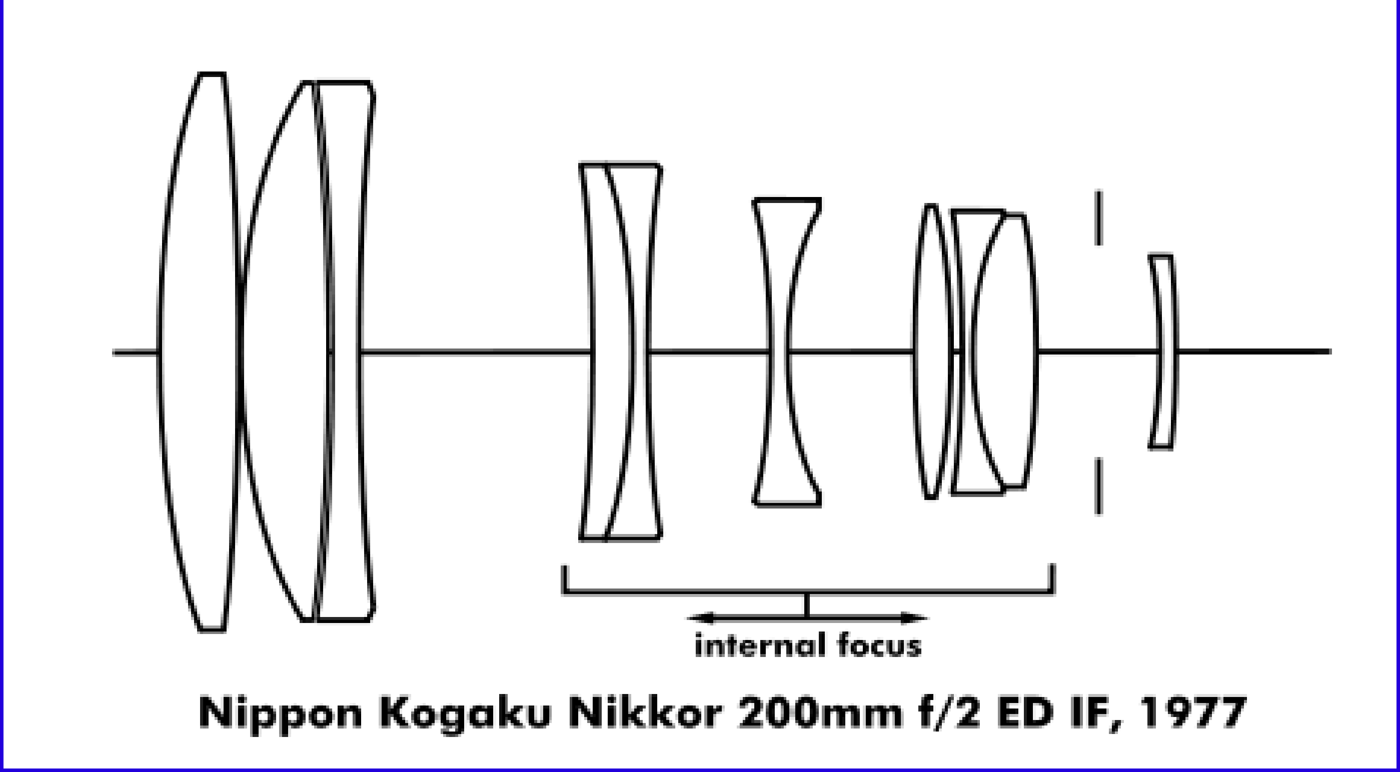
A telescope takes light from very distant objects entering the entrance pupil as parallel rays, and brings the light to a focus with some well-defined focal length and f/ratio (NA).

The simplest telescope uses a lens (simple or achromatic) for the *OBJECTIVE*, ie the optic which is primarily responsible for creating the image.



Fast Refracting Lens Systems for Wide-Field Imaging

Lens system with very small f/ratios (fast) for imaging can be quite complex. They are used in commercial photography (your SLR has one) and large ones are used in astronomical spectrographs. They often have many exotic glasses and aspheric surfaces. This example is a fast Nikon lens for 35mm photography, with 10 elements.



Reflecting Optical Systems

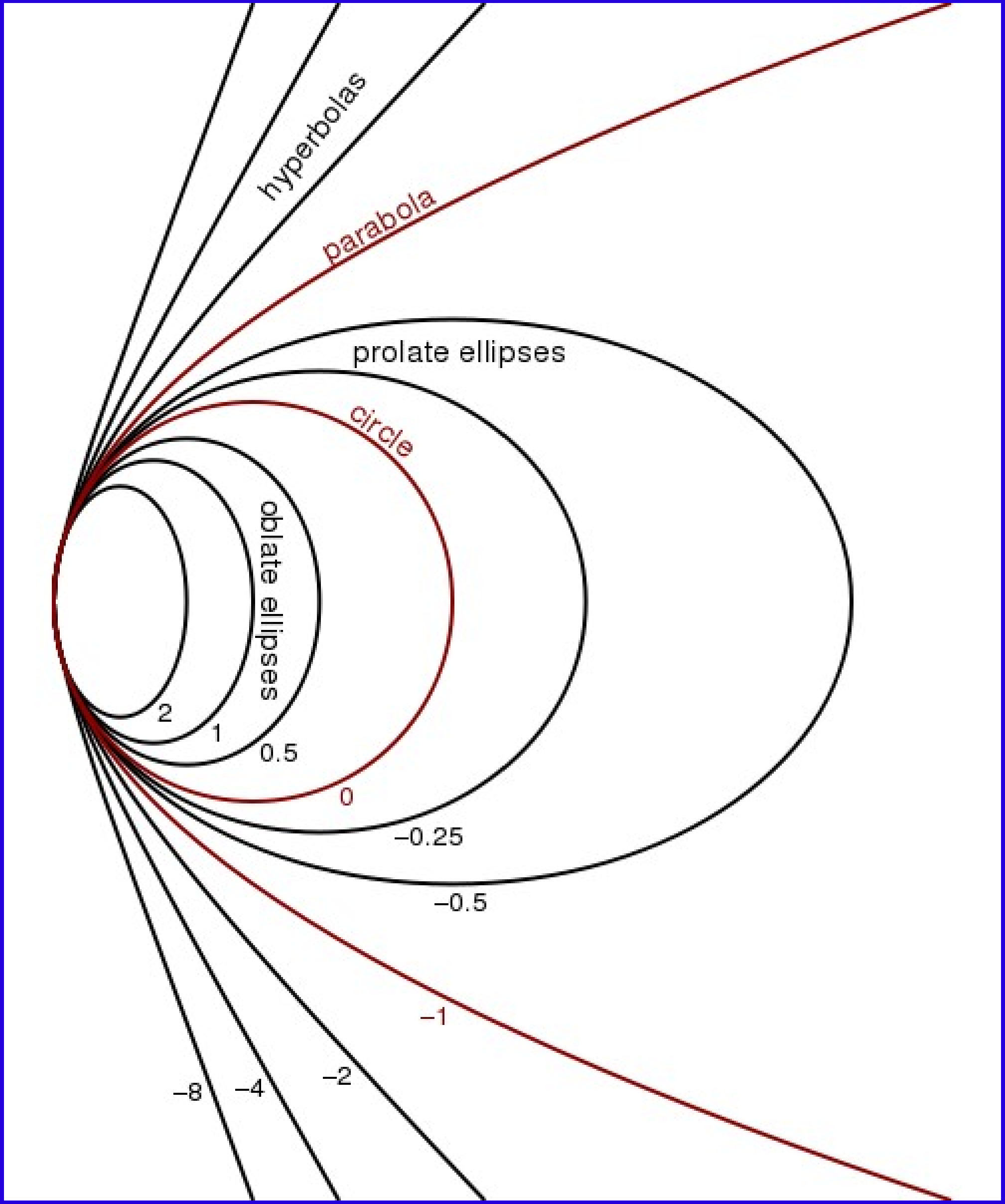
But all large telescopes these days are reflecting—that is, the objective is a mirror, not a lens. A concave mirror can produce images just as a convex lens does. Most mirrors are shaped (approximately) like conic sections of revolution—paraboloids, ellipsoids, hyperboloids, spheres. The properties of these surfaces are:

1. Paraboloid: All rays parallel to the axis of the paraboloid converge exactly to one point, the focus, after reflection. (This is obviously the prototypical objective (primary) mirror.

2. (prolate) Ellipsoid: All rays from one focus of the ellipsoid converge exactly to the other focus after reflection.

3. Hyperboloid: (one sheet) All rays which would have converged at the focus behind the sheet converge exactly to the focus in front of the sheet, (whichever sheet you use) after reflection.

4. Sphere: Obviously, all rays from the center of curvature coverge back to the center of curvature.



Prescription:

Define the CONIC CONSTANT $K=-e^2$. Positive K corresponds to OBLATE ellipsoids; the foci define a ring, not points on the axis.

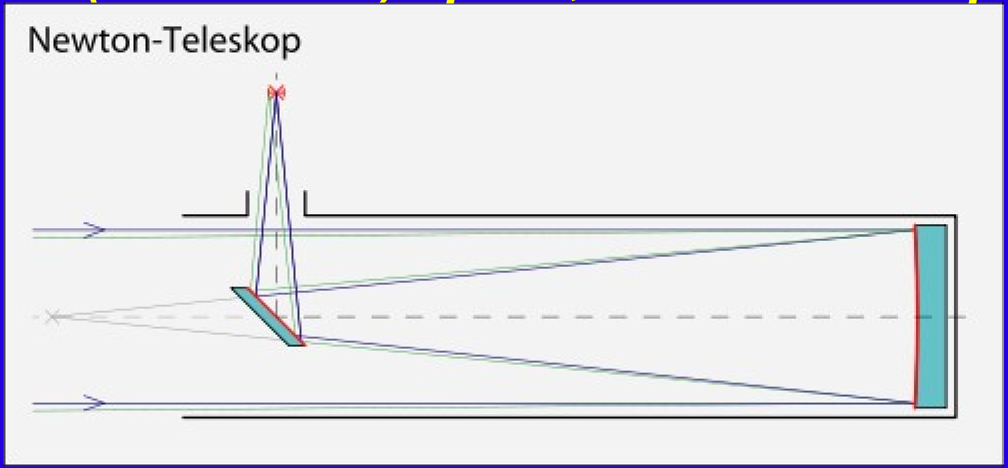
K>0 are oblate ellipsoids
K=0 is a sphere
0>K>-1 are prolate ellipsoids
K=-1 is a paraboloid
K< -1 are hyperboloids

All these curves have the same osculating radius R near the origin. The surface is

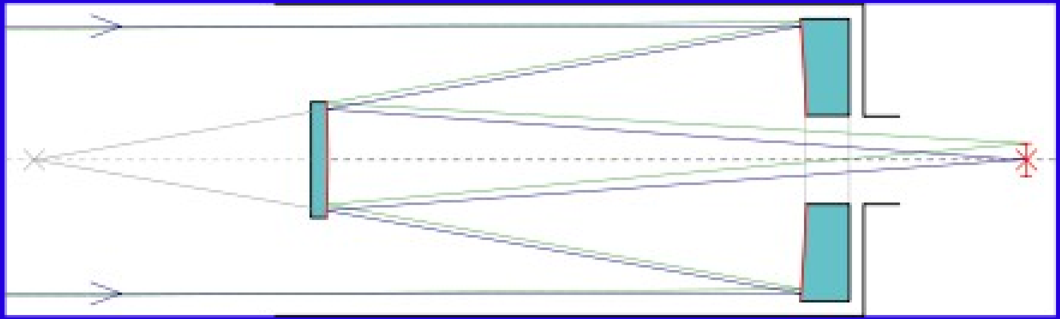
$$y^2 + 2Rx + (K+1)x^2 = 0$$

so $x \sim y^2/(2R)$ for small y

The simplest refecting telescope is the Newtonian. It has a paraboloidal prmary (objective) mirror, and a flat mirror to make the focus accessible; for sufficiently large telescopes, the flat diagonal is not necessary and one can work at PRIME FOCUS. The primary aberration is coma, and is quite terrible without other (`corrector') optics, which can be quite complex. (cf HSC)

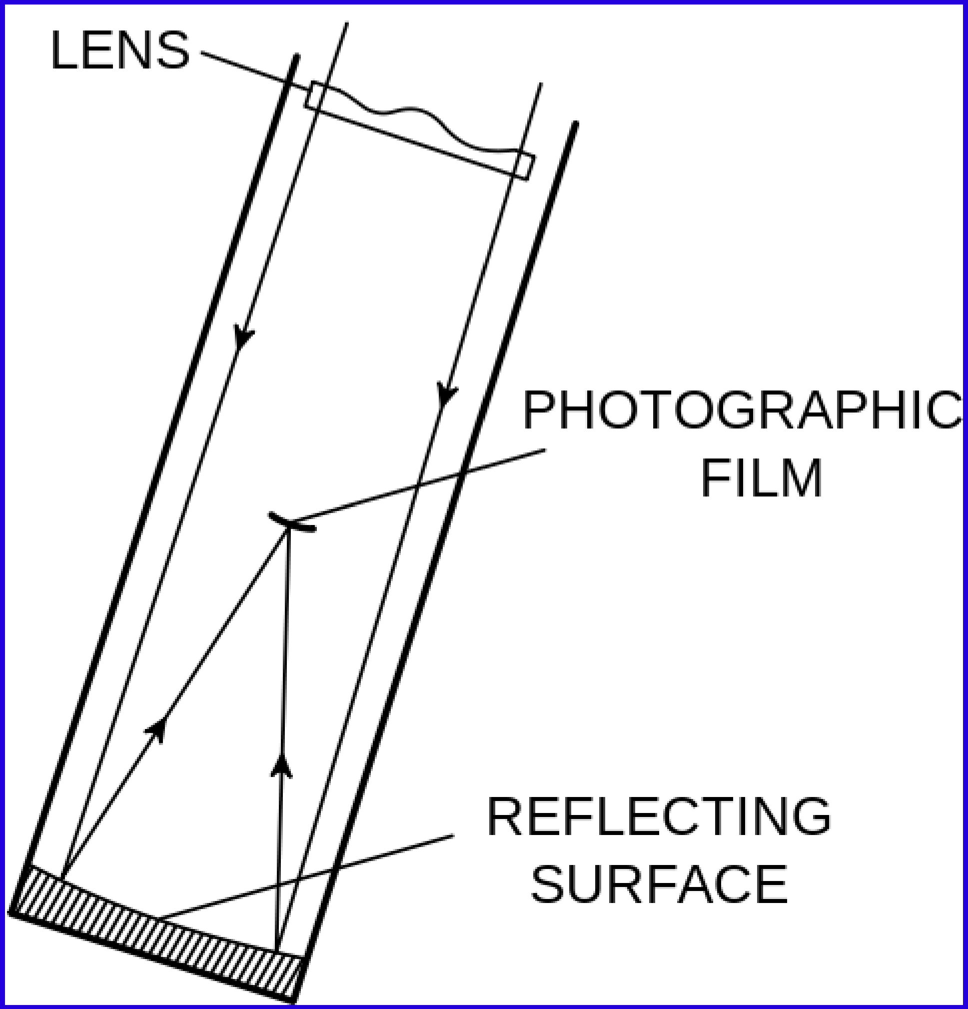


if one introduces a convex hyperboloid with one focus at the focus of the paraboloid and the other well behind the primary mirror, one gets the Cassegrain. All large telescopes are prime focus/cassegrain systems. The coma can be corrected by making both mirrors hyperboloids, which results in the Ritchey-Chretien design, also widely used. A simple auxiliary aspheric lens also gets rid of astigmatism, resulting in quite wide fields with good images.



Catadioptric systems; the Schmidt camera

It is possible to combine reflecting and refracting systems; usually the idea is to let the mirror do most of the `work' forming the image, and to use lenses only to correct the aberrations of the mirror. The simplest and most beautiful of these is the Schmidt camera, which uses a spherical mirror (which by itself has ghastly spherical aberration-- but remember that spherical is field-independent) with a thin fourth-order aspheric corrector plate at the center of curvature of the mirror whose sole role is to correct the spherical aberration of the spherical primary. Since it is at the center of curvature and its power does not change much with field angle, the result is excellent imaging over huge (many degree) fields with very fast optics.



There is a theorem that all opticians are insane. This was certainly true of Bernhard Schmidt, an Estonian optician who invented the Schmidt camera. But it turns out that he did NOT invent the Schmidt camera; someone else did it, 400 milliion years ago.

The eye of the giant sea scallop:

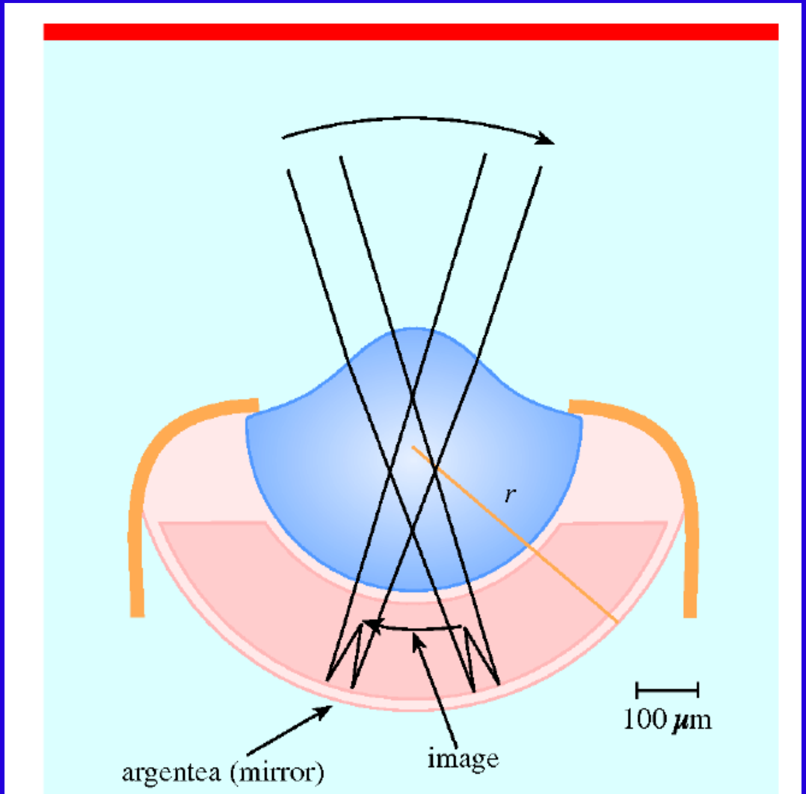


Figure 3. The image on the retina is mainly formed by reflection.