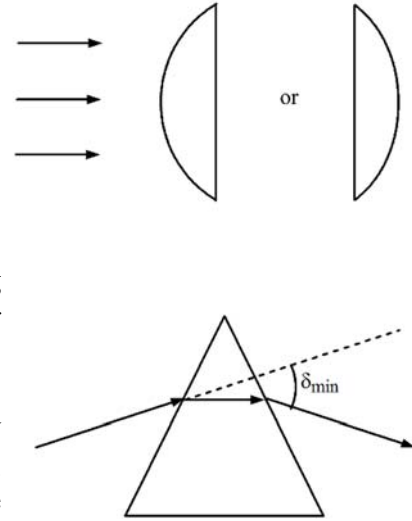


LAST TIME: Reviewed thick lenses and lens systems, examples, chromatic aberration and its reduction, aberration function, spherical aberration

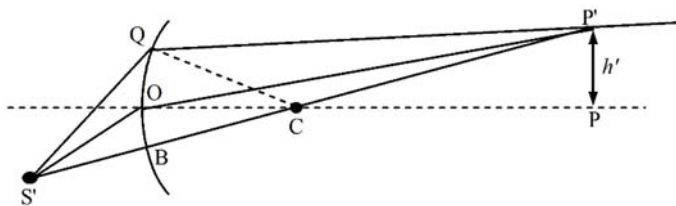
How do we reduce spherical aberration? The most obvious way is to reduce the size of the lens because it is the light that creates the marginal rays that are primarily responsible for spherical aberration. However, what else does this do?

Depending on the type of lens being used, spherical aberration may be reduced by changing the orientation of the lens. Consider what happens with a planar-convex lens that we might use as a collimating lens or a focusing lens. Here are the two alternatives that show how we could orient the lens. Which way do you think it would be best to orient the lens to reduce spherical aberration? Let's get some guidance from the notion we mentioned before concerning dispersing prisms. An equilateral triangular prism shows an interesting feature when a ray of light is incident on the prism. Consider the situation shown in the figure. The minimum deviation seems to refract the light less, and it would cross the optical axis (focus) farther away from the lens. This puts it closer to the paraxial ray focus point and reduces the spherical aberration. Therefore, although it seems counter intuitive, it is best to put the curve side closer to the source to make the lens appear more like a prism.



This method works only if we have a very specific type of lens. The last method is to reshape the lens to be aspherical in shape. As I have mentioned before, having specially items made is usually quite expensive simply because someone has to program an instrument to make the item.

This particular aberration function we derived earlier was done assuming that our point source was on the optical axis. What do we do if it is off the axis as shown?



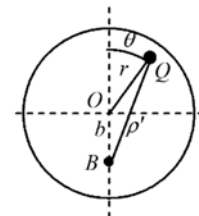
$$a'(Q) = (S'QP' - S'BP') \\ = c(BA)^4 = c\rho'^4$$

$$a'(O) = (S'OP' - S'BP') \\ = c(BO)^4 = cb^4$$

Finally, $a(Q) = a'(Q) - a'(O) = c(\rho'^4 - b^4)$, where the figure labeling the variable is shown. Use $\rho'^2 = r^2 + b^2 + 2rb \cos \theta$ to obtain

$$a(Q) = c(r^4 + 4r^2b^2 \cos^2 \theta + 2r^2b^2 + 4r^3b \cos \theta + 4rb^3 \cos \theta).$$

The triangle OBC is approximately similar to triangle CPP'. Therefore, $b = kh'$. We may then write an aberration function as



$$a(Q) = {}_0C_{40}r^4 + {}_1C_{31}h'r^3 \cos \theta + {}_2C_{22}h'^2r^2 \cos^2 \theta + {}_2C_{20}h'^2r^2 + {}_3C_{11}h'^3r \cos \theta.$$

First term is spherical aberration – no angular dependence

Second term is coma

Third term is astigmatism

Fourth term is curvature of field

Fifth term is distortion

On Monday, I will bring my computer and show some videos of how some of these aberrations look. I also will show how mirrors make excellent collectors of light without regard to imaging.

First, however, I want to show you a bit about how the different shaped lenses we discuss earlier can play a role in reducing spherical aberration. Recall that we had at least six different lens shapes we mentioned. You might wonder why they are needed if they all might have the same focal length. Note that our aberration function $a(A) = (n_i l_o + n_t l_i) - (n_i s_o + n_t s_i)$ involves both the object and image distances as well as the distances l_o and l_i . Because the shape of the lens determines l_o and l_i for a given focal length, it should not be surprising that aberrations depend on the lens shape and the object and image distances. To account for these dependencies, it is customary to define two additional parameters, the Coddington shape factor and the position factor. The Coddington shape factor σ is given by

$$\sigma = \frac{R_2 + R_1}{R_2 - R_1}$$

and the inverse of the position factor is given by

$$p = \frac{s_i - s_o}{s_i + s_o}.$$

The longitudinal spherical aberration of a lens may be written in terms of σ and p as

$$\frac{1}{s_{ih}} - \frac{1}{s_{ip}} = \frac{h^2}{8f^3} \frac{1}{n(n-1)} \left[\frac{n+2}{n-1} \sigma^2 + 4(n+1)p\sigma + (3n+2)(n-1)p^2 + \frac{n^3}{n-1} \right]$$

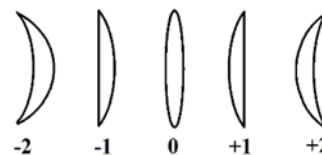
s_{ih} is the location of a ray at height h and s_{ip} is the location of a paraxial ray. The longitudinal spherical aberration is minimized when

$$\sigma = \frac{-2(n^2 - 1)}{n + 2} p.$$

To use this result, we must specify the location of s_o . To make our life easy and to get some results, we take s_o to be infinite. In essence, we are bringing in parallel rays. Then $p = -1$, and if $n = 3/2$, σ is given by

$$\sigma = \frac{2(2.25 - 1)}{3.5} \cong 0.7.$$

The shape factor for a few lenses is shown in the figure. You can see from this that the best way to minimize spherical aberration from an infinite object is to use a planar-convex lens with the convex side toward the object. This will be demonstrated Monday when I show some of these effects.



Let's now begin our study of waves. The reading material for this section is covered in your textbook in sections 2.1 through 2.11 – essentially all of chapter 2. A wave may be defined as a self-sustaining disturbance that carries momentum, energy, and even angular momentum. For the case of mechanical waves, waves that require a medium for their propagation, waves may be transverse or longitudinal. For transverse waves, the direction of the motion of the particles in the medium is perpendicular to the direction of propagation of the wave. For longitudinal waves, the motion of the particles in the medium coincides with the direction of propagation. Solid materials may support either type of wave, transverse or longitudinal, but fluids (gases or liquids) generally support only longitudinal waves. Electromagnetic waves, the ones we will be most concerned with, are transverse and require no medium to travel. They travel in a vacuum and rely on changing electric and magnetic fields to be self-sustaining. Some examples include

Sound waves in air – longitudinal wave that relies on pressure differences to propagate

Waves on a string – transverse waves whose speed depends on the tension in the string and the linear mass density of the string.

Sound waves in a solid – either transverse or longitudinal

We start with the idea that the shape of the disturbance propagates without loss of energy. The shape may be arbitrary, but we expect it to have a functional form given by

$$\psi(x, t) = f(x \pm vt).$$

We may understand why this is the case by recalling the form of functions whose centers are displaced by an amount a from the origin. Remember the equation for a circle is given by

$$x^2 + y^2 = R^2,$$

but if we want the center of the circle to be displaced by an amount a along the x-axis, the equations above becomes

$$(x \pm a)^2 + y^2 = R^2.$$

Now, if we want the center of the circle to move along the positive or negative x-axis, we may just set $a = vt$, and the location of the center of the circle will move to the right if we choose the negative sign and to the left if we choose the positive sign. In this case, the location of the y coordinate of the center of the circle remains zero. We could say that the circle is propagating along the positive x – axis with a speed v . Your textbook shows an arbitrary shape propagating along the positive x-axis, but the idea is the same. Any disturbance of the form $\psi(x, t) = f(x \pm vt)$ represents such a wave. Suppose we have a Gaussian function at $t = 0$ given by $f(x) = e^{-ax^2}$. Then $\psi(x, t) = e^{-a(x-vt)^2}$ represents that Gaussian shape as it travels along the positive x-axis with speed v . Sometimes, it is convenient to represent the wave as

$$F\left(-\frac{x \pm vt}{v}\right) = F\left(t - \frac{x}{v}\right).$$

Because many of the fundamental equations of physics are partial differential equations, it is worthwhile checking to see if our function $f(x \pm vt)$ represents a solution to a partial differential equation. Here is one way we may do this. Let

$$\psi(x, t) = f(\xi) \text{ with } \xi = x - vt \text{ and } \psi(x, t) = f(\eta) \text{ with } \eta = x + vt.$$

Because time and space are both involved, we take the partial derivative of $\psi(x, t)$ with respect to x and t to obtain

$$\frac{\partial \psi}{\partial x} = \frac{\partial f}{\partial \xi} \frac{\partial \xi}{\partial x} = \frac{\partial f}{\partial \xi} \text{ and } \frac{\partial \psi}{\partial x} = \frac{\partial f}{\partial \eta} \frac{\partial \eta}{\partial x} = \frac{\partial f}{\partial \eta}.$$

Likewise, we calculate the time derivatives to obtain

$$\frac{\partial \psi}{\partial t} = \frac{\partial f}{\partial \xi} \frac{\partial \xi}{\partial t} = -v \frac{\partial f}{\partial \xi} \text{ and } \frac{\partial \psi}{\partial t} = \frac{\partial f}{\partial \eta} \frac{\partial \eta}{\partial t} = v \frac{\partial f}{\partial \eta}.$$

If we use only the first derivative, we obtain two equations that ψ obeys. The difficulty is that the positive x and negative x propagating waves have different equations. We should try the next higher partial derivative to see if anything changes. I will ask you to do this calculation in the homework and just state the result for one-dimension. What happens is that we obtain the same equation for both directions of propagation given by

$$\frac{\partial^2 \psi}{\partial x^2} - \frac{1}{v^2} \frac{\partial^2 \psi}{\partial t^2} = 0.$$

This equation is known as the one-dimensional wave equation without energy loss. This equation is a second order, linear partial differential equation. Second order refers to the highest order derivative occurring in the equation, and linear means that no cross terms of the function with itself or its derivatives occurs. For equations that are linear, the sum of any solutions is also a solution to the equation. Note that this is a homogeneous partial differential equation; i.e., there is no source term on the right hand side. For this equation, the wave travels unchanged in shape until something interferes with its propagation. Partial differential equations (PDEs) are generally much harder to deal with than ordinary differential equations (ODEs) because of the nature of a partial derivative.

Recall that when you calculate a partial derivative, all variables except the one occurring in the denominator is treated as a constant. This means that when you try to solve even the simplest PDE such as $\frac{\partial f(x,y,z)}{\partial x} = 0$, the solution can be only given by $f(x,y,z) = F(y,z)$. You will learn how to deal with PDEs in other courses. The most useful form of a wave to consider for our purposes in this course is the harmonic wave given by

$$\psi(x, 0) = A \sin kx = f(x).$$

This means that our traveling wave has the form $\psi(x, t) = A \sin k(x \pm vt) = f(x \pm vt)$. Because sine functions are repetitive, we know that there is some distance over which the wave repeats itself. Let's see what that distance is.

$$\psi(x + \lambda, t) = A \sin[k(x + \lambda) \pm kvt] = A \sin(k(x \pm vt) + k\lambda).$$

We know that the sine function is periodic 2π , so $k\lambda = 2\pi$ and $k = \frac{2\pi}{\lambda}$. k is called the propagation number and λ is the spatial period or wavelength.

NEXT TIME: Aberration demonstrations and waves continued