How to beat the Rayleigh resolution limit: A lecture demonstration

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td>doi:<a href="http://dx.doi.org/10.1119/1.1463736">http://dx.doi.org/10.1119/1.1463736</a></td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:27413766">http://nrs.harvard.edu/urn-3:HUL.InstRepos:27413766</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
How to beat the Rayleigh resolution limit: A lecture demonstration

Wolfgang Rueckner
Harvard Science Center, 1 Oxford Street, Cambridge, Massachusetts 02138

Costas Papaliolios
Department of Physics, Harvard University, 17 Oxford Street, Cambridge, Massachusetts 02138

(Received 6 July 2001; accepted 31 January 2002)

An experiment is described in which the effect of instrument aperture size on resolution is easily demonstrated. Using diffraction as a tool rather than a limit to resolution, one can further demonstrate the possibility of detecting the presence of more than one object, even when the instrument cannot resolve the objects. © 2002 American Association of Physics Teachers.

[DOI: 10.1119/1.1463736]

I. INTRODUCTION

The resolution limit of an optical instrument is given by the Rayleigh criterion. This limit can be easily demonstrated to students in a large lecture hall. The images of two light sources are observed as a function of instrument aperture. Having passed the limit of resolution by using too small an aperture, it nevertheless is possible to detect the presence of two unresolved light sources by invoking an elegant technique known as interferometry. This technique is presently being used to resolve close binary stars and to measure the diameters of single stars, which also can be easily demonstrated.

II. DEMONSTRATING THE RAYLEIGH LIMIT OF RESOLUTION

Most introductory physics texts discuss the Rayleigh limit of resolution: two objects are just resolvable if they are separated by an angle \( \theta \) (in rad) given by

\[
\theta = 1.22 \frac{\lambda}{D},
\]

where \( D \) is the diameter of the objective lens and \( \lambda \) is the wavelength of the light being used. Photographs and/or diagrams usually accompany the discussion. A demonstration of the effect of instrument aperture size on resolution is a marvelous teaching tool that would complement a textbook treatment of the subject. The ability to vary the aperture in a continuous manner and observe its effect is more powerful than static textbook illustrations and, in any case, will clarify even the best of illustrations. The experiment is simple and relatively inexpensive, especially if your department already has a small telescope and video camera. The technical details of the apparatus are discussed in Sec. V.

A telescope equipped with a camera adapter and video camera is set up at the front of the lecture hall. The telescope is aimed and focused on a pair of very closely spaced (0.5-mm separation) point light sources at the rear of the lecture hall—a distance of some 20+ meters away. At full aperture (approximately 10 cm or 4 in. in our case) the two points of light are easily resolved and well separated [Fig. 1(a)].

An adjustable aperture is now placed in front of the telescope. With a telescope magnification of approximately 125, it is necessary that this aperture be mounted on a separate tripod and not touch the telescope—even so, walking around on a cement floor still jiggles the image, as do air currents in the room! As the aperture is slowly closed, the students can see the images of the two points of light grow in size as they smear out [Fig. 1(b)]. Closing the aperture further results in the smeared images melding together [Fig. 1(c)]. The 2.8-cm-diam aperture happens to be the Rayleigh resolution limit aperture for our particular geometry. Not visible in these images are the diffraction rings around the central bright area—one or two of these rings are clearly visible in the live video images, probably because viewing live video integrates 60 frames/s whereas the images in Fig. 1 are single video frames. Closing the aperture even further results in a large blurry spot [Fig. 1(d)], and by no stretch of the imagination can one claim to see two sources of light. At this stage the students can fully appreciate that further telescope magnification (in hopes of resolving the two sources of light) would be absolutely futile. Two points should be noted: all four images in Fig. 1 are the same scale and, in the live video images, one can see the diameter of the diffraction rings increase as the telescope aperture is decreased.

III. BEYOND THE RAYLEIGH LIMIT: INTERFEROMETRY

We begin by replacing the adjustable aperture on the tripod with a simple hand-held aperture—namely, a piece of cardboard with a 1.3-cm (1/2 in.) hole in the center. The image is as before [Fig. 1(d)]—by no stretch of the imagination can one see two sources of light.

The important concept to get across to the students at this point is that changing the position of the aperture does not change the position of the image. This idea is readily demonstrated by moving the cardboard around. No matter where the 1.3-cm hole is (as long as it is somewhere within the bounds of the 10.2-cm-diam telescope opening), the image remains in the same spot. Students can be reminded that the image position must be stationary because every part of a lens (or mirror) participates in, and contributes to, the entire image. Moving the 1.3-cm hole around is the same as sampling the contributions to the final image from different parts of the mirror (or lens, if it is a refractor).

The next step is to image a single point light source (by switching over to a second slide that has been prepared with just one pinprick—details of the light sources are discussed in Sec. V). The image appears to be no different than Fig. 1(d); there is no difference between the images of a single point light source and two unresolved point light sources. Having established that, a mask with two 1.3-cm holes is now placed in front of the telescope. The two 1.3-cm aper-
Fig. 1. Four video images of the two point light sources. The loss of resolution is evident as the telescope aperture is reduced.

...tures do not produce two separate images, but rather two images superimposed upon each other. Again, the fact that this is so was just demonstrated previously. However, as in Thomas Young’s famous double slit experiment, an alternating pattern of constructive and destructive interference bands is now present in the image. These are like Young’s fringes produced by two 1.3-cm holes rather than two slits. As in the case of Young’s experiment, the fringe spacing can be shown to vary with the hole separation: the farther the holes are moved apart, the closer the fringe spacing, as is shown in the five images in the left column of Fig. 2. Having demonstrated that to the students, we now go back to the situation of observing two unresolved points of light, but now with the two 1.3-cm apertures in front of the telescope. Note that one could either use the previous slide with the two closely spaced pinpricks or use another slide with even more closely spaced pinpricks. In fact, the images in the right-hand column of Fig. 2 are not only of another such slide, but additionally, the distance between the telescope and double pinhole source of light was also increased by a factor of 2. Having doubled the distance of the pinholes from the telescope, even the full-aperture (10.2-cm) telescope is just below the Rayleigh limit in ability to resolve the double light source. Furthermore, closing down the telescope aperture by almost an order of magnitude (to 1.3 cm) makes it hopelessly impossible to resolve the double light source. Referring back to Fig. 1, imagine what the bottom image would look like if the two sources of light shown in the top image were closer than the Raleigh limit of resolution.

Starting with the two 1.3-cm holes as close together as possible (2.3 cm between centers), the image appears not much different from that produced by one point light source [see Fig. 2(a), right-hand column]. As the holes are moved apart, the fringe spacing again becomes smaller, just as was the case for the single source. However, as the holes are moved still further apart, there comes a point when the fringes become washed out [Fig. 2(c), right-hand column]. The fringes come back again as the holes are moved yet further apart [Figs. 2(d) and 2(e)]. This remarkable behavior is due to the fact that there are two sources of light, notwithstanding that they are unresolved. What is going on?

As a first approximation, let us analyze the situation as follows. Suppose that the two unresolved sources of light have an angular separation, measured from the telescope, equal to \( \theta_{sep} \). The light from each source passes through the two apertures and produces interference fringes. The light (and interference fringes) from one of the two sources is superimposed on that from the other source at the detector (camera). However, the superposition is not exact because the two sources are at slightly different locations. The angular shift between the two interference patterns is \( \theta_{shift} = \theta_{sep} \). This shift by itself is too small to be seen in Fig. 2(a). The shift is illustrated schematically in Fig. 3(a)—the spacing of the interference fringes due to the spacing of the two apertures is about a factor of 4 larger than the shift between the two patterns (due to the separation of the two points of light), and their superposition is a similar pattern with the intensity maxima smeared out every so slightly. The smearing is exaggerated in Fig. 3(a).

As the two apertures in front of the telescope are moved further apart, the spacing of the interference fringes shrinks in size, just as when observing a single point light source. Now, here is the crucial difference: when the spacing of the interference fringes has shrunk down to a size equal to twice the shift between the two patterns, then the superposition of the two patterns produces a very washed out intensity pattern. If the two patterns were to have exactly the same intensity (as they would if the two point sources of light were of the same intensity), then their superposition would produce a pattern that is completely washed out, as shown schematically in Fig. 3(c). The corresponding real image [Fig. 2(c)] does not exhibit this completely washed out behavior because the two point sources of light (the two pinpricks in the aluminum foil) were, in fact, not of the same size and therefore not the same intensity. On the positive side, the tech-
nique not only detects the presence of two sources, but also is sensitive to, and capable of, detecting a difference in brightness of the two sources. (As in computer programs, a bug becomes a feature!) Let us now analyze the situation more quantitatively.

For constructive interference, we shall use the same condition as in the case of Young’s double slit experiment, namely the maxima occur when

\[ d \sin \theta = n \lambda, \]  

where \( d \) is the separation of the apertures in front of the telescope, \( \theta \) is the angular separation of the interference pattern fringes, and \( n = 0, \pm 1, \pm 2, \) etc. For \( n = 1 \) and small \( \theta \), the separation of the interference fringes is given by

\[ \theta = \frac{\lambda}{d}. \]  

But if \( \theta = 2 \theta_{\text{sep}} \) (as we argued above), then the angular separation of the two unresolved light sources is

\[ \theta_{\text{sep}} = \frac{\lambda}{2d}. \]  

Because we can easily measure the separation of the apertures, \( d \), we have a method of determining the angular separation of the two unresolved light sources.

As the apertures are moved still further apart, the interference pattern continues to shrink in size and the pattern becomes distinct once more, as seen in the image of Fig. 2(d).

There comes a point when the spacing of the interference pattern is equal to the shift between the two patterns—see the image in Fig. 2(e) and the schematic in Fig. 3(e). At this point the two patterns are in complete registration and their superposition produces a clear and quite distinct intensity pattern. Indeed, the image of the intensity pattern is noticeably more distinct than the corresponding pattern obtained from the single pinhole light source. That difference will be discussed presently. The exact aperture separation at which the two diffraction patterns are in complete registration is a
bit of a judgment call, if done visually. One would expect it to happen at a separation of 8.8 cm but, as can be seen from the image in Fig. 2(e), the superposition of the two patterns produces a clear and distinct pattern even at 9.0 cm. The telescope size of 10 cm prevented exploration of the pattern beyond 9.0 cm.

We continue with the quantitative analysis and take $\lambda$ to be 550 nm (the mean wavelength used to represent the visible spectrum) and $d = 4.40$ cm [Fig. 2(e)], and calculate an angular separation of $6.25 \times 10^{-6}$ rad, ±1%. The distance of the light sources from the telescope was measured to be 44.75 m. Using Eq. (4), the separation of the pinpricks is calculated to be $0.28 \pm 0.003$ mm. Direct measurement of the distance (described in the following paragraph) between the two pinpricks in the aluminum foil gave $0.29 \pm 0.02$ mm.

Although the latter measurement need not be performed in front of the students, especially if the lecturer is pressed for time, it is quite easy to measure the separation of the two pinholes in the slide. The procedure is as follows. Project and focus the pinhole slide onto a screen with a slide projector (this slide projector of course needs a lens). Measure the distance between the pinhole images on the screen with a meter stick. Next, project a blank 35-mm slide onto the screen and measure its dimensions on the screen. The ratio of its actual dimensions (35×24 mm) to the dimensions projected on the screen gives the enlargement factor. The en-
largement factor can now be used to deduce the separation of the pinholes. The 7% uncertainty quoted above is primarily due to the fact that the pinholes were not round in shape and one had to use some judgment in determining the center-to-center distance.

Finally, referring back to Fig. 2(e), we address the question, why the image of the intensity pattern of the double pinhole source is noticeably more distinct than the corresponding pattern obtained from the single pinhole light source. In short, the answer has to do with the fact that the diameter of the single pinhole is larger than either of the two pinholes in the double source. (As an aside, that is why the overall intensity in all the images of the single pinhole source is greater than in the corresponding images of the double pinhole source.) To make the effect even more pronounced, the diameter of the single pinhole was enlarged from roughly 0.4 mm to 0.57 ± 0.03 mm and the experiment repeated. The images in Fig. 4 are dramatically different from those in Fig. 2. The pinhole behaves more like an extended light source than a point source. Imagine the two halves of the single pinhole as being two sources of light (very close together). We can then apply the same argument used in the analysis of the double pinhole source experiment to explain the “washing out” behavior—the spacing of the interference fringes is equal to twice the shift between the two patterns. However, in this case the two patterns are from the two halves of the pinhole and the shift is due to the separation distance between the two halves. Again we use the condition that

\[ \theta_{\text{sep}} = \frac{\lambda}{2d} \]

Now the question is, what is the separation between the two halves. It is not simply equal to the radius of the pinhole. The diffraction pattern is a result of adding the intensities of light from the two halves. In this analysis each of the halves is a semicircle and the intensity of light is proportional to the area from which it comes. Let us assume that the effective center of a luminous semicircle of light is located at the position that bisects the area, that is, having equal areas on both sides of this effective center would provide equal intensities of light radiating from either side of this center. The distance between the two effective centers will then be defined as the separation between the two halves of the pinhole, as is illustrated schematically in Fig. 5. The effective centers of each of the two halves are indicated by dotted lines and are a distance \( b \) from the center of the pinhole. An estimate of the distance \( b \) can be obtained by approximating the segment of the circle between the dotted line and the center (of the circle) as a rectangle of width \( b \) and height \( 2r \). If we set its area equal to 1/4 the total area, then \( b = 0.39r \). This value is clearly a slight underestimate. For a better approximation, we apply plane geometry to the circle segments and calculate the areas in terms of the parameters \( b \) and \( r \). As expected, this calculation yields a slightly higher value for \( b \), equal to 0.404\( r \); this calculation could be given as a homework problem to the class. In this model we shall consequently use a value of 0.81\( r \) for the effective separation, \( x \), between the two pinhole halves. The angular separation is \( \theta_{\text{sep}} = x/L \), where \( L \) ( = 44.75 m) is the distance of the pinhole source from the telescope. Thus we have that

\[ r = 1.23 \frac{\lambda L}{2d} \]

and the diameter of the pinhole can be written in terms of these parameters as \( x \)
If we refer back to Fig. 4, we see that the diffraction pattern washes out somewhere between \( d = 4.3 \) and 6.3 cm—the washout is not a sudden phenomenon and some judgment must be invoked if the separation at which it happens is determined simply by visual means. An aperture separation of \( d = 5.3 \) cm is a reasonable assumption in this case. By using this value we obtain a diameter of 0.57 mm for the pinhole, in excellent agreement with the actual size.

In conclusion, not only can this technique be used to measure the separation between two unresolved sources of light, it also can be used to determine the size of a single source of light. In terms of astronomical observations, not only can the separation of unresolved stars be measured, but the sizes of stars as well!\(^7\)

IV. LONG BASE-LINE INTERFEROMETRY (REF. 8)

The experiment described in Sec. III is a model of the so-called long baseline interferometry technique. The technique requires the superposition of two images from two telescopes separated by some appreciable distance. Superposition is accomplished electronically when using radio telescopes, but for optical telescopes the light from the two telescopes must be brought together to interfere. In the demonstration described in Sec. III, the telescope performs the task of superposing two images onto the video camera that records their interference. The two apertures in front of the telescope play the part of two separate instruments (telescopes) aimed at the same unresolved object. In practice, it is necessary to vary the distance between the two telescopes just as we varied the distance between the apertures. When the distance between the telescopes is about half the Rayleigh limit diameter of resolution, we have the situation modeled in Fig. 2(c). By using two (or more) telescopes separated by large distances, it is possible to attain angular resolutions that depend on the separation between the telescopes, rather than on their individual sizes. This technique has been exploited in the field of radio astronomy, providing resolutions far greater than optical telescopes have been able to provide, despite the fact that the wavelength of radio radiation is much larger than that of optical radiation.

This technique is more difficult to apply in the optical part of the spectrum because of the very short wavelengths of light and because of the much higher frequency of oscilla-

\[
\text{diameter} = \frac{1.23 \lambda L}{d}.
\]  

Fig. 5. Geometry of the single pinhole source of light to determine the effective separation between the two halves of the source.

Fig. 6. Telescope with video camera and optics adapter. Counterweights on the side and rear are necessary for balance.

Fig. 7. Double aperture. (a) The hand-held version is made of black paper, folded accordion-style. The aperture separation is adjusted by pulling or pushing the edges of the paper. (b) A more sophisticated version is held in a frame and supported by a camera tripod.
tions of the electromagnetic waves of this light. New technology, however, is now making it possible to do optical interferometry more easily. By using optical telescopes, each about 1 m in diameter, but separated by about 100 m, it is possible to determine stellar diameters to milliarcsecond precision. Space interferometers are now being designed to measure the separations of stars to microarcsecond precision. These measurements will provide a means of detecting earth-sized planets about other stars by making the small wobble of stars evident when these planets orbit their sun.

V. TECHNICAL DETAILS

A. Double light source

It is important that the two light sources be point-like, close together, bright, and adjustable in brightness. How close together will depend on the size of the lecture hall—the smaller the hall, the closer they will have to be so that they cannot be resolved. The sensitivity of the video camera will dictate the minimum brightness.

The simplest method of making a pair of closely spaced point sources is to make two small pinpricks (with a sewing needle) in a piece of aluminum foil. When done by hand (and the aid of a magnifying glass), 1/5-mm-diam holes spaced 1/2 mm apart are easily achieved. Measurement of the actual size and separation is described in Sec. III. The aluminum foil is then mounted in a 35-mm slide mount and placed in a slide projector without a lens. That’s all there is to it! The condenser and light bulb of the projector bathe the pinholes with 300 W of uniform illumination—bright enough for even the least sensitive of video cameras. Alternatively, a 60-W frosted light bulb behind the aluminum foil will provide enough brightness for more light sensitive cameras.

Closing down the aperture of the telescope from 10 cm (4 in.) to 1.3 cm (0.5 in.) diameter obviously compromises its light collecting ability by a factor of about 60. For this reason it is important that the brightness of the light source be made adjustable, even when a camera with automatic gain control is employed. This brightness adjustment is accomplished by connecting a standard light dimmer (available in most hardware stores) to the appropriate pins of the external lamp control connector of the slide projector. The light dimmer must be at the front of the lecture hall (by the telescope) so that the lecturer can simultaneously adjust the telescope aperture with one hand and the appropriate brightness of the light source with the other. Thus, a long connecting cord is required.

B. Telescope

Whether it is a reflector or refractor is not important—ours is an old 4-in. reflector (Celestar™ 4 by Fecker) with a 90-cm focal length. With an 8-mm eyepiece, the telescope magnification is about 112. A finder-scope is essential in such a setup. The choice of f/2.10 A T-to-C camera adapter mount and universal combined camera adapter couple the video camera to the telescope. The telescope’s 8-mm eyepiece is used with the adapter for “eyepiece projection” of the image onto the CCD. The output of the camera is fed to a video projector or large monitor.

Unless the camera is ultralight, it is better to attach an auxiliary camera mount on the telescope so that not all the weight is borne by the universal camera adapter. A laboratory clamp designed for clamping onto large cylindrical vessels serves this purpose well and a simple coupling can be made that secures the camera to the clamp (see Fig. 6). Note that a large counterweight needs to be added to the rear of the telescope to keep everything in balance about its equatorial mount. If the weight of the camera and accessories tends to twist the telescope in its mount, an additional lab clamp can be added to provide for the possibility of adding counterweight to the telescope on the side opposite to the camera. All these additional weights need not be anything fancy, but should be adjustable in some way so that the telescope can be well balanced.

D. Double aperture

The double aperture consists of an opaque mask with two 1.3-cm holes in it. The separation of the holes must be adjustable in the range from almost touching up to the diameter of the telescope aperture. The simplest (and least expensive) means of achieving this is shown in Fig. 7(a)—it is a piece of black construction paper that has two holes cut out; their separation is 10 cm. The paper between the holes is folded “accordion-style” and pulling them apart varies the hole separation. This hand-held scheme works quite well if one wishes to simply demonstrate the effect of aperture separation. However, for quantitative demonstrations, some kind of frame is recommended to hold the aperture separation fixed while measurements are performed. Figure 7(b) shows one possibility that allows for continuous adjustment of the separation from 2.35 to 10 cm. The frame mounts onto a camera tripod.

4 See Ref. 2.
5 Alternatively, a section of a clear plastic ruler can be cut to the size of a B&K 330, 1/3 in. B/W charge-coupled device (CCD) camera. It is one of the more light-sensitive cameras with a specifica-

- Electronic mail: rueckner@fas.harvard.edu
- For example, see D. Halliday, R. Resnick, and J. Walker, Fundamentals of Physics (Wiley, New York, 2001), 6th ed., p. 898. The authors make the comment that “the analysis of such patterns is complex,” but state where the first minimum is located [given by Eq. (1) in this paper].
- Young’s experiment is also discussed in most, if not all introductory physics texts. For example, again see Ref. 1, pp. 905–911 for a full discussion.
- One hole measured 0.26 mm in diameter and the other measured 0.16 mm, which would suggest that the light from the smaller hole is about 2.3 times the intensity from the larger hole.
- See Ref. 2.
- Alternatively, a section of a clear plastic ruler can be cut to the size of a 35-mm slide mount (2 in. × 2 in.) and used as a slide. The numbers and rulings will be projected and their image easily measured on the screen.
- Note the remarkable closeness between the factor of 1.23 in this expression and the correct factor of 1.22 found in the literature. Calculating the factor of 1.22 is a problem of considerable difficulty because it requires a double integration over the surface of the apertures. It was first solved by Airy in 1835 in terms of Bessel functions. For details of this solution, see T. Preston, The Theory of Light (Macmillan, London, 1928), 5th ed., pp. 324–328. Most optics texts simply quote the factor of 1.22 and note the Airy reference. For a more modern reference, see K. D. Møller, Optics (University Science Books, Mill Valley, California, 1988). On p. 347 the author simply states where the first minimum is located, but refers the reader to Problem 7 of Chap. 3. This problem concerns the actual calculation for a circular aperture and the author leads the reader through the problem (pp. 178–181). S. A. Akhmanov and S. Y. Nikitin, Physical Optics
Optics (Clarendon, Oxford, 1997), give a full treatment of diffraction by a round aperture on pp. 289–292 and derive the factor of 1.22.

See, for example, A. R. Thompson, J. M. Moran, and G. W. Swenson, Jr., Interferometry and Synthesis in Radio Astronomy (Wiley, New York, 2001), 2nd ed. These measurements are discussed in Sec. 1.3 and Chap. 12.

Reference 7 is an excellent general reference on this topic as well as for detailed information.

The light dimmer should be rated for at least 600 W (we use a Lutron D-600RH, available for about $5). Another solution is to put the entire slide projector on a Variac™ (autotransformer). However, the projector fan will slow down as the lamp is dimmed—not a good thing if left on too long. A very long extension cord connects the projector to the Variac, which must be at the front of the lecture hall.

Sensitivity is typically expressed in terms of minimum illumination required to produce a usable picture at a particular f-stop. This definition of sensitivity is not very useful in this application but can be used to compare one camera to another. Sometimes manufacturers express the sensitivity in foot-candles. For reference, lux (lx) = lumen/m², foot-candle (fc) = lumen/ft², and thus 1 fc = 10.76 lx.

Orion part # 07127.

Orion part # 5264. For part of the experiment it was not possible to focus the image using these adapter parts. The inexpensive solution worked well—simply put an empty cardboard toilet paper roll between the front of the camera and the telescope eyepiece. Its only purpose is to shield the CCD from light other than that projected onto it by the eyepiece.

Inclined Plane. The inclined plane is a basic building block of introductory physics. Perhaps present-day students consider problems involving it dull and boring because they do not have the opportunity to do experiments with it. This inclined plane at Amherst College is by Ducretet of Paris and was almost certainly bought about 1875 as a replacement for the apparatus lost in the destruction by fire of the original Walker Hall on the Amherst College campus. Its cost was about $20 in 1875 currency. The block is not original; when new, the body on the inclined plane was a roller. (Photograph and notes by Thomas B. Greenslade, Jr., Kenyon College)