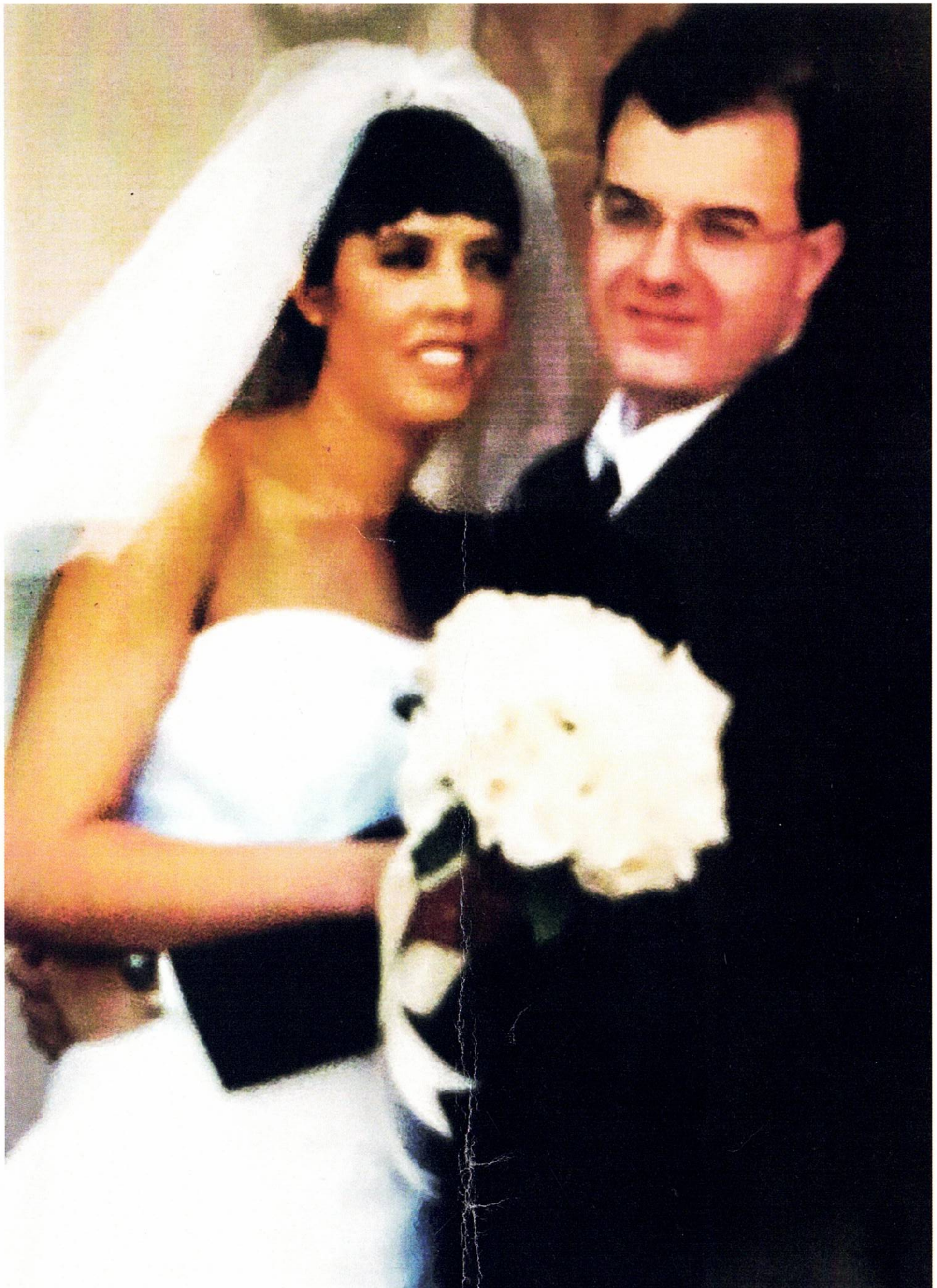
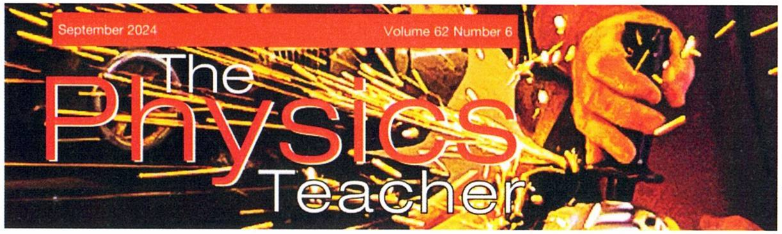


Adrienne Kelly and Alex Small
June 28, 2003

Professor Alex Small

May 17, 1977 - December 7, 2024
“A good man”



A Simple Derivation of the Rayleigh Criterion

Alexander R. Small, California State Polytechnic University Pomona, Pomona, CA

Rayleigh's criterion[1] is important for specifying the minimum separation needed for two objects to produce distinct, nonoverlapping images when observed with diffraction-limited optical instruments. The bound arises from that the phenomenon of waves passing through an aperture spread out due to diffraction, rather than traveling in straight lines like rays. We thus obtain spread-out images of point sources, rather than point images.

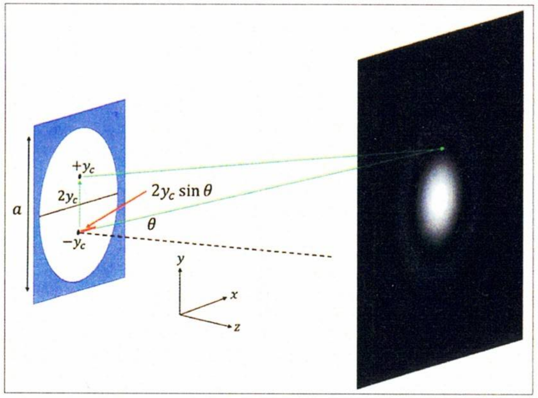


Fig. 1. Schematic of the circular diffraction problem, with the aperture approximated as two sources located $\pm y_c$ from the center. Rayleigh diffraction pattern courtesy of Wikimedia Commons. (https://commons.wikimedia.org/wiki/File:Airy-pattern.svg)

Quantifying the extent of the spread involves the diffraction pattern of a circular aperture, and introductory textbooks usually just assert the key formula relating angular spread θ (see Fig. 1) to wavelength λ and aperture diameter a: $\sin\theta = 1.22\lambda/a$.[2] While the presence of the ratio of wavelength to aperture diameter can be explained via analogies to Fraunhofer diffraction problems that are solvable by beginning students, the factor of 1.22 (more precisely, 1.2196) requires Bessel functions, which are well beyond introductory classes. Here we show an approximate method for finding the dark rings in the diffraction pattern, via a pedagogically accessible two-point interference problem. The resulting prediction is $\sin\theta = 1.1781\lambda/a$, accurate to 3.4%.

We treat the aperture as a collection of infinitesimal point sources; dark rings in the diffraction pattern occur when the point sources interfere destructively. An exact calculation would require integrating over an infinite number of points, but approximate calculations only require finite numbers of points. In the simplest approximation, we divide the aperture into just two portions, approximating each with a point source, as shown in Fig. 1. The top portion is somewhat closer to the detector, the bottom portion somewhat farther, and when the distances differ by a half-wavelength, we get destructive interference.

The key question is where to place the point sources. A plausible approximation is to represent the half-circles with point sources at their centers of mass, or centroids. We can motivate this assumption by noting that in the Fraunhofer theory of diffraction, all paths from the aperture to a given point on the detector are roughly parallel, so path differences depend on $y \sin\theta$, as shown in Fig. 1.[3,4] The average path difference hence depends on the average y-coordinate, i.e., the centroid coordinate. When the sources are at $\pm y_c$, as shown in Fig. 1, the path difference is $2y_c \sin\theta$, and setting this to $\lambda/2$ gives

$$y_c \sin\theta = \lambda/4. \qquad (1)$$

We still need to relate y_c to the half-circle diameter a, via explicit calculation of the centroid coordinate. As shown in Fig. 2, we break the half-circle into strips with height y, width dx, and centroid $y/2$. The centroid of the entire half-circle is the weighted average of the centroids of the strips, weighting by strip area $y\,dx$. The resulting integral is

$$y_c = \frac{\int_{-\frac{a}{2}}^{\frac{a}{2}} \frac{1}{2} y^2 dx}{\frac{1}{2}\pi\left(\frac{a}{2}\right)^2} = \frac{4}{\pi a^2}\int_{-\frac{a}{2}}^{\frac{a}{2}}\left[\left(\frac{a}{2}\right)^2 - x^2\right]dx = \frac{2a}{3\pi}. \qquad (2)$$

The denominator in the first step is the area of a half-circle; the integrand in the second step follows from writing the equation of a circle as $y^2 = (a/2)^2 - x^2$.

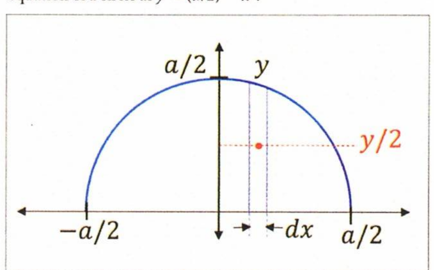


Fig. 2. Geometry and notation for finding the centroid of a half-circle.

 DOI: 10.1119/5.0138716

This is Professor Small's last published research article

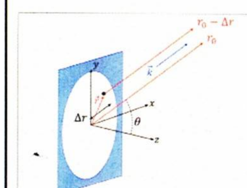


Fig. 3. Geometry of interference. The path length from the origin of the aperture plane to a distant detector is r_0, while the distance from a point at position r in the aperture plane is $r_0 - \Delta r$. Both paths are parallel to the wave vector k.

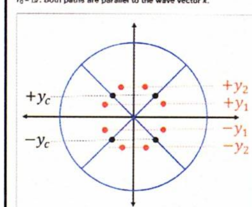


Fig. 4. Circular aperture divided into eight equal segments. The y-coordinates of the centers of mass of the eight wedges are denoted by $\pm y_1$ and $\pm y_2$. The y-coordinates of the centers of mass of the quadrants are denoted by $\pm y_c$.

We can combine Eqs. (1) and (2) to write the resolution criterion as

$$\sin\theta = 3\pi\lambda/8a = 1.1781\lambda/a. \qquad (3)$$

The factor of 1.1781 differs from Rayleigh's prediction (1.2196) by only 3.4%. Given the crudeness of our two-point approximation, this is impressive agreement.

To determine whether this accuracy is physically meaningful or merely coincidental, we can check whether our approximation's predictions remain consistent in a finer-grained model. We show next that dividing the aperture into eight segments gives the same result as Eq. (3). This demonstration requires a slightly more formal treatment of interference and diffraction, which we develop now. Interested readers can consult standard optics textbooks for more details.[5]

Adding contributions from multiple segments means adding sinusoidal waves of the form $\cos(kr - \omega t)$, where r is the path length from the segment in the aperture plane to a distant detector, $k = 2\pi/\lambda$ is the wave number, and ω and t are frequency and time. We measure r relative to a reference path from the center of the aperture, whose length we label r_0 in Fig. 3. In the limit of Fraunhofer diffraction, a path from an off-center segment of the aperture differs from r_0 by a segment of length Δr in the vicinity of the aperture, and in Fig. 3 Δr is the projection of $\mathbf{r}$ (position in the aperture plane) onto the central path. Since the central path is parallel to the wave vector $\mathbf{k}$, we can write $k\Delta r$ as the dot product $\mathbf{k}\cdot\mathbf{r}$.

Due to the circular symmetry of the diffraction pattern, we can simplify our task by limiting our attention to a diffracted wave traveling up and out of the aperture plane, i.e., traveling in the y–z plane. The wave vector $\mathbf{k}$ thus has only y and z components, and $z = 0$ in the aperture plane, so we can write $\mathbf{k}\cdot\mathbf{r}$ as $ky \sin\theta$. (This dot product has the sine rather than the cosine because the angle between $\mathbf{k}$ and $\mathbf{r}$ is $90° - \theta$.) The contribution of a segment i to the diffracted field thus obeys the following proportionality:

$$E_i \propto \cos(kr_0 - ky_i \sin\theta - \omega t). \qquad (4)$$

Suppose now that we divide the aperture into quadrants via the usual x- and y-axes. Due to the left–right symmetry of the half-circle, each quadrant's centroid has the same y-coordinate as the half-circle, so y_c is still given by Eq. (2). Equation (3) again follows, as the path difference between the top and bottom sources is half a wavelength.

Subdividing the quadrants into octants, as shown in Fig. 4, is more interesting. We have drawn the octants symmetrically for convenience, but the crucial assumption is simply that the octants have equal areas. It follows that a quadrant's centroid coordinate y_c is the average of the octants' centroids y_1 and y_2:

$$\frac{y_1 + y_2}{2} = y_c. \qquad (5)$$

We can rewrite that condition as $y_2 = 2y_c - y_1$, giving the following:

$$ky_2 \sin\theta = 2ky_c \sin\theta - ky_1 \sin\theta = \pi - ky_1 \sin\theta. \qquad (6)$$

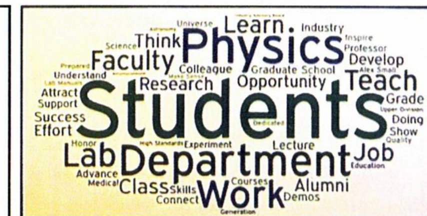
The π in the last step follows from the fact that $y_c \sin\theta = \lambda/4$ [Eq. (1)] and $k = 2\pi/\lambda$.

The diffracted field is the sum of eight terms of the form in Eq. (4), but left–right symmetry makes half of the terms redundant. Of the four remaining terms, two have $\pm ky_1 \sin\theta$ inside the cosine, and the others have $\pm ky_2 \sin\theta$, which is given by Eq. (6). We can thus write

$$E \propto \cos(kr_0 - ky_1 \sin\theta - \omega t) + \cos(kr_0 - \pi + ky_1 \sin\theta - \omega t) + \cos(kr_0 + ky_1 \sin\theta - \omega t) + \cos(kr_0 + \pi + ky_1 \sin\theta - \omega t). \qquad (7)$$

The four cosines in Eq. (7) occur in pairs that are identical except for the inclusion of $\pm\pi$ inside the cosine. Adding $\pm\pi$ to the argument of a sinusoid multiplies the cosine by -1, and so each pair of terms cancels. We thus find that the diffracted field is zero when we divide into octants and the angle is given by Eq. (3).

Our approach has significant limitations, of course.



Presented by the Optical Society of Southern California