Subtleties with Young's double-slit experiment: Investigation of spatial coherence and fringe visibility

David P. Jackson,[a] Natalie Ferris, Ruthie Strauss, Hongyi Li, and Brett J. Pearson
Department of Physics and Astronomy, Dickinson College, Carlisle, Pennsylvania 17013

(Received 13 March 2018; accepted 28 June 2018)

We discuss Young's double-slit experiment using a partially coherent light source consisting of a helium-neon laser incident on a rotating piece of white paper. Such an experiment is appropriate for undergraduate students as an independent project or as part of an advanced lab course. As is well known, the resulting interference pattern is observed to disappear and return, depending on the angular size of the source. Interestingly, while the standard theoretical prediction for the light intensity agrees quite well with experimental data when the fringe visibility is high, the prediction is noticeably off when the visibility is low. A first-principles calculation of the light intensity is performed and shown to agree extremely well with the experimental results for all visibilities.
https://doi.org/10.1119/1.5047438

I. INTRODUCTION

Discussions about the nature of light have likely been around as long as humans have existed, and written records of such discussions go back to the philosophers of ancient Greece. By the 1700s, the biggest question being debated was whether light consisted of particles, as put forth by the eminent Sir Isaac Newton, or whether it was comprised of waves, as proposed by Christiaan Huygens.[1] The turning point in this debate is generally credited to be the "double-slit" experiment, carried out by Thomas Young in the early 1800s. In the original version of this experiment, sunlight streams through a tiny hole and a "slip of card" is placed edgewise in the light beam.[2] Light from the two sides of the card then interfere with each other and form an interference pattern on a wall across the room. Young was even able to use his data to calculate the wavelengths of the different colors of light. This version of the "double-slit" experiment is a prime example of beauty, elegance, and simplicity in science, and may very well be the best way to perform the experiment with introductory students.[3]

Today, Young's double-slit experiment is a standard topic in the physics curriculum, and is usually used to demonstrate that light can be described as an electromagnetic wave.[4] The phenomenon is typically explained as follows. As shown in Fig. 1, monochromatic plane waves of wavelength λ are incident on an aperture consisting of two slits, each of width a, separated by a distance d. At the introductory level, the width of the slits is usually neglected, and it is assumed that coherent, circular (in the plane) waves emanate from each slit. By comparing the path lengths from each slit to the observation screen a distance s away, one finds locations of constructive and destructive interference of the two waves. Constructive interference occurs when the two waves are in phase—the path lengths differ by an integer number of wavelengths, $n\lambda$—and destructive interference occurs when the two waves are out of phase—the path lengths differ by a half-integer number of wavelengths, $(n+\frac{1}{2})\lambda$. For $s \gg d$, the angular positions θ_n of the bright spots on the screen are then determined by geometry to obey the relation $d\sin\theta_n = n\lambda$, with a similar expression for the positions of the dark spots.

This simplistic description provides good physical insight as to the main cause of the interference pattern—constructive and destructive interference—and even leads to a reasonably accurate quantitative prediction for the locations and spacing of the fringes. Using the standard approximations $y \ll s$ and $\sin\theta \approx \theta \approx \tan\theta$, the intensity distribution is given by[5]

$$I(y) = \cos^2\left(\frac{kd}{2s}y\right) = \frac{1}{2}\left[1+\cos\left(\frac{kd}{s}y\right)\right], \tag{1}$$

where y measures the off-axis position along the viewing screen, $k = 2\pi/\lambda$ is the magnitude of the wave vector, and we have normalized the intensity to have a (dimensionless) maximum value of 1 at $y = 0$. This equation predicts that the intensity drops to zero at the minima and that all maxima have the same brightness. It is sometimes mentioned that the brightness of the fringes will eventually decrease as you move away from the central maximum because the distance from the double-slits is increasing, but neglecting this minor effect, the brightness of the fringes is predicted to remain constant.

Now, unless the width of the slits is much smaller than the spacing between the slits ($a \ll d$), a quick look at the actual interference pattern will show significant intensity variations, indicating that Eq. (1) cannot be the entire story. Understanding this behavior requires that the slit width be taken into account, and at the introductory level the mathematical derivation is typically bypassed and the result is simply stated: the finite slit width acts to modify the interference pattern so that the fringe maxima are governed by the diffraction pattern of a single slit of width a. Mathematically, the intensity is given by[6]

$$\begin{aligned} I(y) &= \mathrm{sinc}^2\left(\frac{ka}{2s}y\right)\cos^2\left(\frac{kd}{2s}y\right) \\ &= \mathrm{sinc}^2\left(\frac{ka}{2s}y\right)\frac{1}{2}\left[1+\cos\left(\frac{kd}{s}y\right)\right], \end{aligned} \tag{2}$$

where $\mathrm{sinc}(x) \equiv \sin(x)/x$, and the intensity has again been normalized to have the value 1 at $y = 0$. Apart from an overall constant, we see that the (normalized) single-slit intensity distribution, $\mathrm{sinc}^2(kay/2s)$, acts as an envelope for the sinusoidal interference pattern given in Eq. (1). The derivation of Eq. (2), which integrates contributions across the apertures, is usually left for a more advanced course on waves or optics

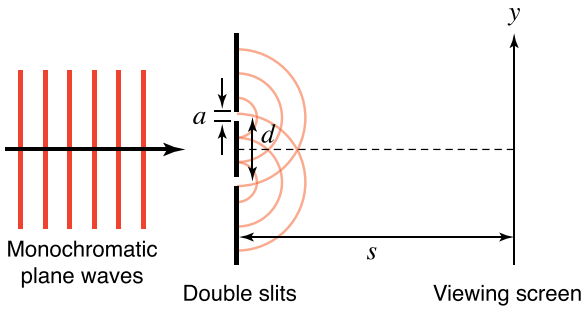

Fig. 1. Basic geometry of the double-slit experiment. Monochromatic plane waves are assumed to be incident on a double-slit aperture with slit widths a and separation d. The interference pattern is viewed on a screen a distance s away.

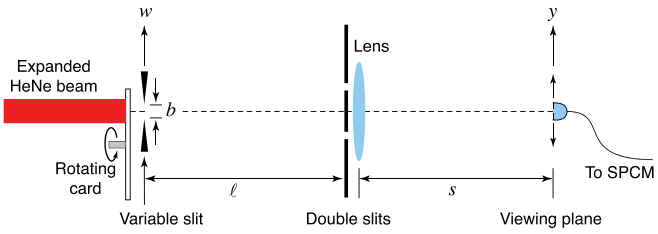

Fig. 2. Experimental setup. A HeNe laser is incident on a rotating white card, and an adjustable slit immediately after the card permits adjustment of the source size. After passing through the double-slit aperture, a (scanned) fiber coupler collects the light and sends it to a single-photon counting module (SPCM) for processing.

in the sophomore or junior years. Experiments to test Eq. (2) are relatively easy to perform, and the agreement with theory can be very good. In terms of the physics curriculum, the derivation and analysis of Eq. (2) provides a reasonably complete description of double-slit interference, so this is usually where the story ends. However, an extension of the traditional treatment of double-slit interference using a partially coherent light source leads to some interesting and unexpected results, making a nice student project or advanced lab experiment.

II. INTERFERENCE WITH PARTIALLY COHERENT LIGHT

The typical double-slit experiment uses a laser as a source of coherent light. The coherence properties of lasers make them ideally suited for standard interference experiments; to a very good approximation a helium-neon (HeNe) laser produces perfectly monochromatic plane waves. Lasers have been in common use since at least the 1980s, so it is unlikely that many of today's students have viewed an interference pattern using anything but a laser. In fact, most students may not even realize that interference effects can be observed using something as simple as a light bulb. Essentially any source of light can be made (spatially) coherent simply by moving it very far away, so that it looks like a point source. Of course, trying to use just any light source in a double-slit experiment leads to two main problems. First, moving the light source far away will decrease the intensity at the viewing screen, making it more difficult to measure the interference pattern. And second, different colors have different wavelengths, so the various interference patterns tend to "wash out" unless care is taken to measure only a small band of wavelengths.

One way of overcoming these problems was pioneered by Thompson and Wolf, who used a light source consisting of a filtered mercury-vapor lamp focused onto a circular aperture.[7] In this experiment, the light is then passed through two circular "slits" and the visibility of the resulting interference pattern is seen to change as the distance between the slits (and hence the spatial coherence) is varied. These researchers demonstrated good qualitative agreement with the general interference law for partially coherent light. Another approach, adopted by Sharpe and Collins, makes use of a sodium-vapor lamp focused onto a variable-width slit; the narrower the slit, the more (spatially) coherent the light source.[8] Not only is this experiment suitable for the undergraduate laboratory, it represents perhaps the most impressive quantitative test of the Thompson-Wolf experiment published to date.

Motivated by simplicity, we utilize a slightly different approach. As diagrammed in Fig. 2, a HeNe laser ($\lambda = 632.8$ nm) incident on a white paper card acts as our source of light; including a beam expander and iris results in an approximately uniform intensity spot. Using a laser makes it extremely easy to produce sufficient photon count rates downstream. Because the laser is spatially coherent, the diffuse scattering that results from the roughness of the paper results in random, position-dependent phase differences of the light, which then interfere to form irregular bright and dark spots ("laser speckle"). Rotating the card by attaching it to a small computer fan with the blades removed ($\sim$3 000 rpm) adds random time-variations to the phases and destroys the spatial coherence of the original laser beam (the detector time-averages over many of the phase variations, creating what is sometimes referred to as a "pseudo-thermal source").[9] Our laser spot can thus be viewed as a collection of monochromatic, incoherent point sources. An adjustable-width slit is placed just after the rotating card, leaving a uniform intensity slit of width b. The light from this source passes through a double-slit aperture that lies a distance ℓ from the source and is followed by a lens with focal length f. A fiber-optic coupler with an adjustable slit in front is then placed where the interference pattern is focused, in our case at a distance s from the lens. The coupler sits on a computer-controlled translation stage that allows us to scan across the interference pattern, with the output of the fiber-optic cable going to a single-photon counting module (SPCM) and finally into a computer for processing.

A few comments are in order. First, the intensity patterns described in Eqs. (1) and (2) are technically valid only when plane waves are incident on the double-slits. Thus, it is common to incorporate an additional (matched) lens in front of the double-slits, with the source located a distance f in front of the lens. In our case, the large distance between the light source and slits made this lens unnecessary. Second, without this second lens, light from the source is expanding (slightly) at the double slits, which causes the interference pattern to be located at a distance that is slightly larger than the focal length f of the lens (hence the fact that $s = 1.06$ m).

A. Negligible slit widths

As mentioned, the light source for our experiments consists of a HeNe laser incident on a spinning white paper card, which we treat as a collection of monochromatic *incoherent* point sources that emit in all directions. Thus, to determine the intensity at the viewing plane where we scan the fiber

coupler, we add the *intensities* from these various sources.[10] When the slits have negligible width ($a \to 0$), the calculation is reasonably straightforward.[11] Because the process will be useful to us later, we outline the basic procedure here.

Equation (1) gives the double-slit intensity distribution for a point source of light that resides on axis ($w = 0$ in Fig. 2) when the width of the slits can be neglected. For a point source that is off axis ($w \neq 0$), the interference pattern will be shifted so the central maximum is located at $y \approx -(s/\ell)w$, where we have used the small-angle approximation. Thus, a point source at location w gives rise to an intensity distribution (with the usual normalization) in the viewing plane of

$$I(y) = \frac{1}{2}\left\{1 + \cos\left[\frac{kd}{s}\left(y + \frac{s}{\ell}w\right)\right]\right\}. \tag{3}$$

To obtain the intensity distribution for a finite-sized source, we need to add up the contributions from all point sources within the finite source. However, any point source that lies above or below the plane in Fig. 2 will generate the same fringe system displaced in a direction parallel to the fringes. Therefore, an incoherent line source perpendicular to the plane in Fig. 2 has the effect of simply increasing the amount of light available. The end result is that the aperture can be treated one-dimensionally; we only need to add up point sources along the w-axis in Fig. 2, using intensities that reflect the shape of the source perpendicular to the plane. For example, all points in a rectangular source will have the same intensity, whereas a circular source will have intensities that are larger in the middle and smaller at the edges.

To carry out the calculation for a rectangular source of width b, we note that the intensity due to an infinitesimal element of the source at location w will be proportional to the differential line element dw, giving

$$dI(y) = A\,dw\,\frac{1}{2}\left\{1 + \cos\left[\frac{kd}{s}\left(y + \frac{s}{\ell}w\right)\right]\right\}, \tag{4}$$

where A is an appropriate constant. To find the total intensity, we integrate over the source width ($-b/2 \leq w \leq b/2$) to get

$$I(y) = \frac{1}{2}\left[1 + \operatorname{sinc}\left(\frac{kbd}{2\ell}\right)\cos\left(\frac{kd}{s}y\right)\right], \tag{5}$$

where we have again normalized the result to have the (dimensionless) value 1 at $y = 0$. Because we have assumed the slit widths are negligible, this result can be compared directly with Eq. (1) for coherent illumination. The only difference is that the sinusoidal modulation that governs the fringes now has an amplitude given by a sinc function. Thus, for any $b > 0$, the fringe amplitude will be less than one, implying the intensity minima will *not* drop to zero. In fact, when the sinc function is zero this analysis predicts there will be no fringes at all. Moreover, if the sinc function is negative, then the intensity at $y = 0$ becomes a (local) *minimum* instead of the usual maximum.[7,12]

One measure of fringe contrast in an interferometric system is the visibility V (sometimes referred to as the Michelson visibility), defined by

$$V \equiv \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \tag{6}$$

where $I_{\max}$ and $I_{\min}$ are the intensities of adjacent maxima and minima. Since the cosine function in Eq. (5) is bounded by ± 1, the maximum and minimum of the intensity distribution are given by $\frac{1}{2}[1 \pm |\operatorname{sinc}(kbd/2\ell)|]$, so the visibility for a rectangular source of width b is given by

$$V = \left|\operatorname{sinc}\left(\frac{kbd}{2\ell}\right)\right| \approx \left|\operatorname{sinc}\left(\frac{k\phi d}{2}\right)\right|, \tag{7}$$

where we have defined the angular source size $\phi \equiv 2\arctan(b/2\ell) \approx b/\ell$. Notice that the visibility depends on the wavelength λ (through k), the slit spacing d, and the source size ϕ, but it does *not* depend on the slit width a. This fact should come as no surprise given that the derivation began with Eq. (1) in which the slit width was already neglected.

As shown in Fig. 3, the visibility curve given in Eq. (7) has a maximum of $V = 1$ when the source size is zero and decreases to zero when $\phi = \lambda/d$ (1.05 mrad). After that, the visibility goes through a series of decreasing local maxima separated by zeros that occur when $\phi_n = n\lambda/d$, where n is an integer. In other words, the fringes disappear at regular intervals, only to re-appear with a 180° phase shift resulting from the sign of the sinc function [see Eq. (5)]. While this idealized situation provides good insight into what we should anticipate seeing, it is not expected to agree quantitatively with experimental results since the slit width has been neglected in the theory.

Interestingly, the condition that the visibility goes to zero when $\phi_n = n\lambda/d$ can be understood qualitatively using an argument similar to that used when determining where the minima occur in a single-slit diffraction pattern. Although we are dealing with incoherent sources, we are still adding up a series of sinusoidal intensity distributions. For a point source at $w = 0$, the central maximum of the intensity distribution will occur at $y = 0$. However, a point source at the edge of the source (at $w = -b/2$) will have its central maximum shifted to the point $y \approx -(s/\ell)w = sb/2\ell$. Now, these two intensity distributions will be completely out of phase when the maximum of one is located at the minimum of the other. For the point source at $w = 0$, the first minimum occurs when $d \sin\theta = \lambda/2$, which corresponds to $y \approx s\lambda/2d$. Equating this to $y = sb/2\ell$ then leads to $b/\ell = \phi = \lambda/d$, precisely the condition for the first zero of the visibility function. The argument then proceeds exactly as with the single-slit diffraction case; namely, the intensity distribution from a point source "just above" $w = 0$

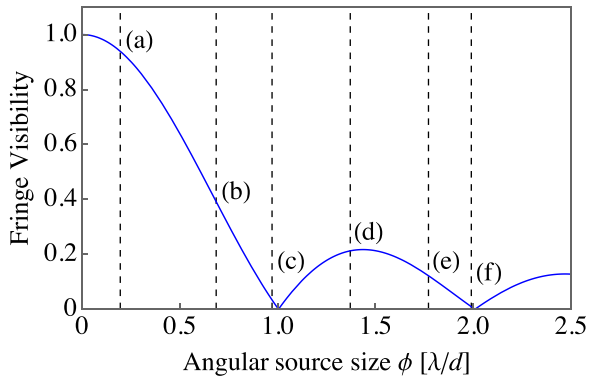

Fig. 3. The fringe visibility as described by Eq. (7) in units of λ/d. The fringes are predicted to disappear and reappear periodically as the source size increases. The dashed lines represent values of ϕ used in our experiment, with the letters corresponding to the panels of Figs. 5 and 6.

will be out of phase with a point source "just above" $w = -b/2$, and so on, so that all across the source, intensity pairs will effectively cancel out.

Two comments are worth emphasizing here. First, because we are adding intensity distributions as opposed to fields, they will not sum to zero even if they are perfectly out of phase. But the "humps" of the intensity distributions will cancel out, leaving a new intensity distribution without any humps at all (in the present case, the resulting intensity distribution will be constant). Second, it may have dawned on the reader that there is a strong similarity between the single-slit diffraction pattern and the visibility function in Eq. (7). This is not a coincidence and is precisely the reason we are able to use the single-slit diffraction argument to understand the zeros of the visibility function. The formal connection between these two functions is governed by the van Cittert-Zernike theorem, which is briefly discussed in Sec. II B.

B. Finite slit width

To extend the results above to account for the width of the slits, a more formal derivation of the intensity distribution is normally undertaken. The details can be found elsewhere;[7,13] here we simply present the results. Assuming the source is centered at $w = 0$ and that the intensity of the light across the individuals slits is the same, the double-slit intensity distribution is found to be

$$I(y) = 2I^{(0)}(y)\left[1 + |\gamma_{12}|\cos\left(\frac{kd}{s}y + \beta_{12}\right)\right], \tag{8}$$

where $I^{(0)}(y)$ is the intensity distribution due to one of the slits by itself (either one, since they are assumed equal), γ_{12} is the "complex degree of spatial coherence" between two points (1 and 2), and β_{12} is the phase that arises due to the sign of γ_{12} (*not* due to any path-length difference). To calculate γ_{12} one typically makes use of the van Cittert-Zernike theorem, which states that when d, $b \ll \ell$, the complex degree of coherence is equal to the normalized Fourier transform of the irradiance distribution across the source.[14,15]

Subject to the usual small angle and far-field approximations, Eq. (8) is a very general result that applies to situations having different source shapes or various types of "double-slits." For example, this equation can be applied to a rectangular source and a traditional (rectangular) double-slit aperture, but it also applies if the source has a circular shape and/or if the "double-slit" consists of two circular apertures. It is important to note, however, that the function γ_{12} characterizes the coherence of the light source at two specific *points* in space—the "locations" of the slits—and therefore the visibility component of Eq. (8) still represents an approximation in which the slit widths are assumed to be infinitesimally small.

In order to make use of Eq. (8), we need to apply it to our specific situation. Determining $I^{(0)}(y)$ is relatively straightforward, as this is simply the intensity distribution due to either one of the slits, i.e., it is the single-slit intensity distribution, which we have already seen is proportional to $\mathrm{sinc}^2(kay/2s)$. Next, we use the Van Cittert-Zernike theorem to find γ_{12}. If the irradiance distribution is uniform across a rectangular (slit) source of width b, we find[7,14] $\gamma_{12} = \mathrm{sinc}(k\phi d/2)$, where $\phi \approx b/\ell$. Taking into account the fact that $\beta_{12} = 0$ when $\mathrm{sinc}(k\phi d/2) > 0$ and $\beta_{12} = \pi$ when $\mathrm{sinc}(k\phi d/2) < 0$, Eq. (8) becomes (with our usual normalization)

$$I(y) = \mathrm{sinc}^2\left(\frac{kay}{2s}\right)\frac{1}{2}\left[1 + \mathrm{sinc}\left(\frac{k\phi d}{2}\right)\cos\left(\frac{kd}{s}y\right)\right]. \tag{9}$$

Equation (9) is a beautiful result that combines both the single-slit intensity envelope from Eq. (2) that arises due to the finite width of the slits, as well as the sinc visibility function from Eq. (5) due to a finite source size. Figure 4 shows a typical plot of Eq. (9), showing that both the maxima and the minima are bounded by envelope functions. Because we are now considering both a finite source size *and* a finite slit width, it may not be a complete surprise that Eq. (9) takes this form [we might even have *guessed* this result simply by referring to Eqs. (2) and (5)]. That being said, we note that the visibility function that appears here is precisely the same visibility function that appears in Eq. (5), which was derived when neglecting the width of the slits. Thus, while Eq. (9) is obviously an improvement over Eq. (5), we should not expect perfect agreement with experimental results, particularly as the width of the slits becomes more important.

III. EXPERIMENTAL RESULTS AND DISCUSSION

A schematic of the experimental setup is shown in Fig. 2. With the exception of the SPCM, most of the equipment used is fairly inexpensive. Because we had previously purchased a set of SPCMs (PerkinElmer SPCM-AQ4C) for performing single-photon experiments, such a detector was an obvious choice for us, as they essentially remove any concerns about low-light detection ability.[16,17] The experiments are carried out in a darkened room, though small amounts of stray light can be seen under doors, around windows, and from computer monitors. We chose a distance $\ell = 0.8\,\mathrm{m}$ and used a lens of focal length $f = 1\,\mathrm{m}$ with the detector placed where the interference pattern is focused, in our case a distance $s = 1.06\,\mathrm{m}$ from the lens. Our double-slit has spacing $d = 600\,\mu\mathrm{m}$ and slit width $a = 150\,\mu\mathrm{m}$, and a computer running LabVIEW is used to record the data from a programmable coincidence-counting unit (Altera DE2).[18] An experimental scan consists of ~90 three-second measurements, with the detector being moved between each measurement. We use a computer-controlled motorized translation stage (Thorlabs Z825B) to move the detector, an entire scan taking approximately five minutes to complete.[19] The source size is altered after each scan by adjusting the variable slit in Fig. 2. For our setup, count rates at the central maximum ranged from around 40,000–200,000 per second, depending on the width of the adjustable slit. Background rates varied from ~7,000–14,000 per second,

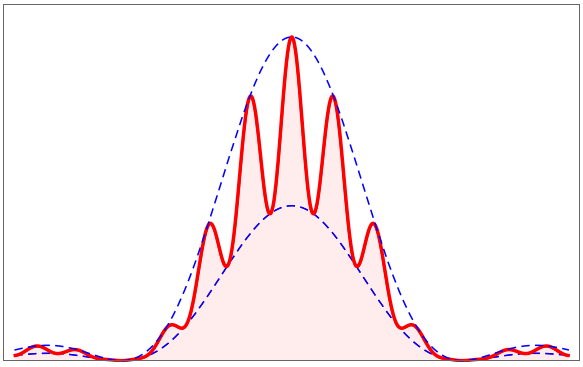

Fig. 4. A sample plot from Eq. (9) showing that both the fringe maxima and the fringe minima are governed by envelope functions (dashed).

with stray light accounting for the majority (∼60–80%) of background.

A. Comparison with the standard model

Figure 5 shows a sample of our experimental results (circles), along with the predictions (curves) from Eq. (9), for a variety of source sizes after subtracting the background, defining $y = 0$ at the central fringe (max or min), and normalizing the data. It is worth mentioning that there are no adjustable parameters in these predictions. There are several notable features of this figure. First, as expected from Fig. 3, the fringe visibility is seen to decrease, then increase with an overall phase shift, and then decrease again as the source size gets larger. It is interesting for students to observe that one can obtain an interference pattern in which the central fringe is a (local) minimum instead of a maximum, and to see "interference" patterns with essentially no fringes. Second, while Eq. (9) does an admirable job of predicting these results, it is clear that the predictions fall short at times, particularly when the visibility is low. Specifically, when the prediction shows essentially no modulation [panels (c) and (f)], the experimental results continue to show fringes. In other words, the observed fringe visibility never disappears completely.

In addition to the visibility issues, Fig. 5 shows that the standard model does not always get the fringe locations precisely correct; this is most apparent in panels (b) and (e). It is clear from Eq. (9) that the functions primarily responsible for the fringe spacing and visibility—those inside the square brackets—are independent of the slit width a. Evidently, the actual situation is not so simple.

B. A more complete model

While the standard model agrees with the data fairly well overall, there are some significant shortcomings that can be somewhat disconcerting to students. This discrepancy between theory and experiment can be explained by the complex correlation function being a function of two points, while in actuality the two slits have finite widths. Recall that the van Cittert-Zernike theorem led to a complex degree of spatial coherence of $\mathrm{sinc}(k\phi d/2)$, precisely what was obtained in Eq. (5) when the slit width was neglected. Indeed, the fact that the parameter a does not appear in the visibility function at all is a sign that Eq. (9) is most likely incomplete. The issue of finite-sized apertures in measurements of fringe visibility has been considered before.[20,21] In particular, finite-width apertures lead to non-zero visibility even at locations where the standard model shows no visibility. Surprisingly, there do not appear to be any subsequent reports of quantitative experimental measurements documenting this effect. While there have been a limited number of studies comparing data with the standard model,[8,22–26]

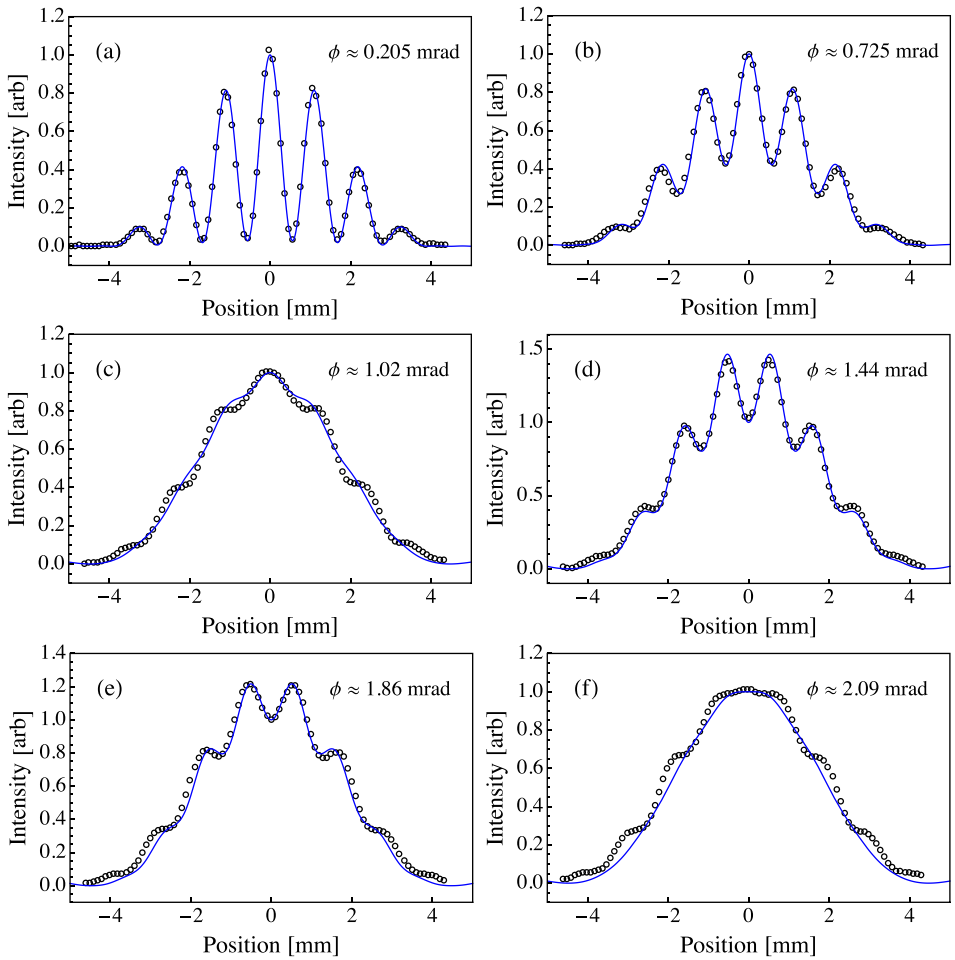

Fig. 5. Experimental data (circles) and theoretical predictions (curves) from Eq. (9) for six difference source sizes. While the standard model does a reasonably good job of predicting the data, there are some obvious shortcomings, particularly when the visibility is low [panels (c) and (f)]. The source sizes used here are represented by the vertical dashed lines in Fig. 3.

none of them address the discrepancies between theory and experiment at low visibilities.

Perhaps the most straightforward way to incorporate the finite width of the slits more fully is to return to the derivation of Eq. (5). This equation arises, in essence, by integrating Eq. (1), the intensity formula when the slit width is negligible, over the size of the source. Our approach is to make use of Eq. (2) instead of Eq. (1), so that we integrate the finite-slit-width intensity formula over the size of the source,

$$I(y) = C\int_{-b/2}^{b/2} \operatorname{sinc}^2\left[\frac{ka}{2s}\left(y+\frac{s}{\ell}w\right)\right] \times \frac{1}{2}\left\{1+\cos\left[\frac{kd}{s}\left(y+\frac{s}{\ell}w\right)\right]\right\} dw, \tag{10}$$

where C is an appropriate constant.

In this form, Eq. (10) is equivalent to Eq. (7) of Ref. 20 (and essentially Eq. (5) of Ref. 21). Here, we take things one step further; while not trivial (and with a little help from MATHEMATICA), this integral can be computed in terms of well-known functions. The (unnormalized) result is somewhat complicated and not particularly enlightening, but is included here for completeness:

$$I(y) = \frac{s^2}{ka^2}[I_c(ky_-) + I_s(ky_-) - I_c(ky_+) - I_s(ky_+)], \tag{11}$$

where $y_\pm = y \pm bs/2\ell \approx y \pm s\phi/2$ are the locations of the images of the source edges (in the viewing plane), and the functions I_c and I_s are given by

$$I_c(x) \equiv \frac{1}{x}\left\{1+\cos(\tilde{d}x)-\cos(\tilde{a}x) - \frac{1}{2}\cos[(\tilde{d}-\tilde{a})x] - \frac{1}{2}\cos[(\tilde{d}+\tilde{a})x]\right\} \tag{12}$$

and

$$I_s(x) \equiv \tilde{d}\,\mathrm{Si}(\tilde{d}x) - \tilde{a}\,\mathrm{Si}(\tilde{a}x) - (\tilde{d}-\tilde{a})\,\mathrm{Si}\big[(\tilde{d}-\tilde{a})x\big] - (\tilde{d}+\tilde{a})\,\mathrm{Si}\big[(\tilde{d}+\tilde{a})x\big], \tag{13}$$

with $\tilde{d} = d/s$, $\tilde{a} = a/s$, and the sine integral function given by[27]

$$\mathrm{Si}(x) \equiv \int_0^x \frac{\sin t}{t}\,dt. \tag{14}$$

Unfortunately, with this full analytic result there is no simple interpretation to the various terms and there is no simple "visibility" function that can be defined. In this sense, the original result shown in Eq. (9) has some obvious advantages: the equation is compact and simple, and it is easy to

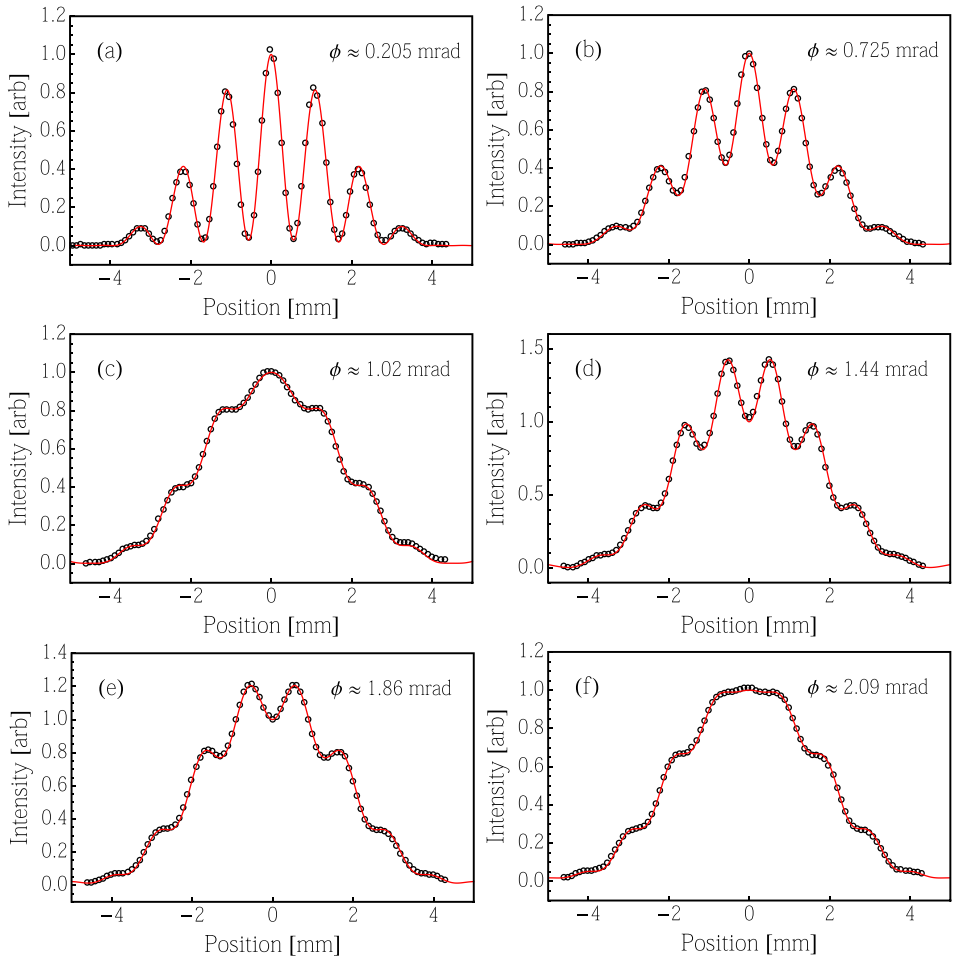

Fig. 6. Experimental data (circles) and theoretical predictions (curves) using the full model of Eq. (11) (once normalized) for six difference source sizes. Compared to Fig. 5, the agreement is significantly better.

understand on physical grounds both the overall envelope function and the visibility function. However, as already noted, the quantitative behavior of the original model is less than ideal in some regimes.

Figure 6 shows the full (normalized) predictions over the same experimental data shown in Fig. 5. As is clear from this figure, the new predictions are significantly better than the original model. In particular, the full model correctly predicts that the fringes never fully disappear, and it also predicts the subtle changes in the fringe spacing [compare panels (b) and (e) of Figs. 5 and 6]. It is worth emphasizing that these predictions are not fits to the data; they are plotted directly on the data after subtracting background, defining $y = 0$ at the central fringe, and normalizing the data. The downside of the full result is that the spectacular quantitative agreement comes at the cost of simplicity. Whereas Eq. (9) is simple and compact and each function has physical meaning, Eq. (11) is very complicated, with 16 separate terms that preclude any sort of physical insight.

IV. CONCLUSION

The work presented here grew out of a summer research experience and senior project with our undergraduate students. There are significant experimental, theoretical, and computational aspects to this project, and we believe it would work well as either a senior project or as part of an advanced lab course. It might also be incorporated into a lab for an upper-level optics course. Although the new theoretical predictions can be written in terms of well-known functions [Eq. (11)], it would probably be simpler for students to do the integral in Eq. (10) numerically using MATHEMATICA or another software package. It is also worth mentioning that while we used scattered HeNe light from a (spinning) card for the extended source and single-photon counting modules for the detectors, other light source-detector combinations could also be employed.

The results presented here make use of a variable slit to control the source size, but we have also investigated the use of pinholes to create circular sources with different diameters.[28] In this situation, the standard prediction is similar to Eq. (9), but with visibility function $\mathrm{sinc}(k\phi d/2)$ replaced by $\mathrm{Bessinc}(k\phi d/2)$, where $\mathrm{Bessinc}(x) \equiv 2J_1(x)/x$, with $J_1(x)$ the first-order Bessel function of the first kind. Qualitatively, the results are similar to those presented here; namely, for high visibilities the standard model does a respectable job of predicting the experimental data, but for low visibilities the predictions are noticeably off. For the full-model calculation, it does not appear that the integral corresponding to Eq. (10) can be written in terms of well-known functions and therefore must be computed numerically. Nevertheless, the agreement with the experimental data is, once again, very impressive.[28]

ACKNOWLEDGMENT

The authors gratefully acknowledge financial support from Dickinson College.

[a]Electronic mail: jacksond@dickinson.edu

[1]We note that Al-Hasan Ibn al-Haytham ("Alhazen") is generally credited with putting forth the first theory of light as particles during the beginning of the 11th century; see John Gribbin, *Q is for Quantum: An Encyclopedia of Particle Physics* (Touchstone, New York, 1998), p. 15.

[2]T. Young, "The Bakerian lecture. Experiments and calculations relative to physical optics," Philos. Trans. R. Soc. London **94**, 1–16 (1804).

[3]W. Schneider, "Bringing one of the great moments of science to the classroom," Phys. Teach. **24**, 217–219 (1986).

[4]At a more advanced level, the double-slit experiment can also be used to demonstrate that single photons interfere with themselves when passing through a double-slit aperture; see R. S. Aspden, M. J. Padgett, and G. C. Spalding, "Video recording true single-photon double-slit interference," Am. J. Phys. **84**, 671–677 (2016).

[5]E. Hecht, *Optics*, 3rd ed. (Addison Wesley, New York, 1998), p. 396.

[6]Reference 5, p. 459.

[7]B. J. Thompson and E. Wolf, "Two-beam interference with partially coherent light," J. Opt. Soc. Am. **47**, 895–902 (1957).

[8]J. P. Sharpe and D. P. Collins, "Demonstration of optical spatial coherence using a variable width source," Am. J. Phys. **79**, 554–557 (2011).

[9]A useful discussion of rotating diffusers can be found in W. Martienssen and E. Spiller, "Coherence and fluctuations in light beams," Am. J. Phys. **32**, 919–926 (1964).

[10]For coherent sources one adds the fields before squaring to get the intensity. For incoherent sources the random phase relationship means the "interference term" time-averages to zero. The result is a superposition of intensities for incoherent sources.

[11]Reference 5, pp. 562–566.

[12]B. J. Thompson, "Illustration of the phase change in two-beam interference with partially coherent light," J. Opt. Soc. Am. **48**, 95–97 (1958).

[13]Reference 5, pp. 566–571.

[14]Reference 5, pp. 572–573.

[15]Max Born and Emil Wolf, *Principles of Optics*, 7th ed. (Cambridge U.P., Cambridge, 1999), pp. 572–577.

[16]B. J. Pearson and D. P. Jackson, "A hands-on introduction to single photons and quantum mechanics for undergraduates," Am. J. Phys. **78**, 471–484 (2010).

[17]Educational single-photon counting modules manufactured by Exelitas technologies can be purchased at a reduced cost through the Advanced Laboratory Physics Association (ALPhA); see <https://advlab.org/spqm> for additional information. The use of such detectors could be used in other "low-light" experiments as well.

[18]The Altera DE2 is a commercially produced development and education board. Users must load a program onto the board in order to interface with LabVIEW. We used a coincidence-counting program developed by Mark Beck for single-photon quantum mechanics experiments. For additional information, see Ref. 16 and/or Mark Beck's webpage <http://people.whitman.edu/~beckmk/QM/>.

[19]Our HeNe laser power was reasonably stable; measurements over a 10-minute time period showed a power drift of less than 1%.

[20]B. J. Thompson and R. Sudol, "Finite-aperture effects in the measurement of the degree of coherence," J. Opt. Soc. Am. A **1**, 598–604 (1984).

[21]A. S. Marathay and D. B. Pollock, "Young's interference fringes with finite-sized sampling apertures," J. Opt. Soc. Am. A **1**, 1057–1059 (1984).

[22]D. Bloor, "Coherence and correlation–Two advanced experiments in optics," Am. J. Phys. **32**, 936–941 (1964).

[23]S. Mallick, "Degree of coherence in the image of a quasi-monochromatic source," Appl. Opt. **6**, 1403–1405 (1967).

[24]D. N. Grimes, "Measurement of the second-order degree of coherence by means of a wavefront shearing interferometer," Appl. Opt. **10**, 1567–1570 (1971).

[25]B. Tom King and W. Tobin, "Charge-coupled device detection of two-beam interference with partially coherent light," Am. J. Phys. **62**, 133–137 (1994).

[26]D. Ambrosini, G. S. Spagnolo, D. Paoletti, and S. Vicalvi, "High-precision digital automated measurement of degree of coherence in the Thompson and Wolf experiment," Pure Appl. Opt. **7**, 933–939 (1998).

[27]The sine integral function is one of the so-called trigonometric integrals, all similarly defined; see <https://en.wikipedia.org/wiki/Trigonometric_integral> for additional information.

[28]B. J. Pearson, N. Ferris, R. Strauss, H. Li, and D. P. Jackson, "Measurements of slit-width effects in Young's double-slit experiment for a partially-coherent source," (to be published).