THE DIFFRACTION OF LIGHT

A PREPRINT

Clayton Williams and Dustin Shipp
Department of Physics
Utah Valley University

July 18, 2019

ABSTRACT

A draft for the Diffraction chapter for PHYS 3115 Introduction to Experimental Physics I. This experiment introduces light diffraction and interference as measurement tools and some basic error analysis techniques. Throughout the experiment, students should consider what else these tools could be used for.

1 Background

1.1 History

The beginning of the 19th century saw a dramatic conclusion to a debate dating from Newton's time on the nature of light when the English physicist Thomas Young established with his double-slit experiment that light can behave in a wave-like manner. Since Christiaan Huygen's work on waves in the 17th century it was well established that waves diffract around obstacles which impede their motion, and that waves undergo interference when they combine. Most interestingly to scientists of the time, it was found that the degree of diffraction depends on the wavelength of the incident light and that light often travels in a seemingly particle-like way due to the very short wavelength of visible light. It was also discovered by Jacques Babinet that complementary apertures produce diffraction patterns which are identical except for their phase.

Until Young's time most optical experiments used prisms, lenses, and mirrors to refract and redirect light; however, his results on diffraction established the use of another optical instrument in the laboratory, that of the diffraction grating. This instrument proved very useful in the hands of Gustave Kirchhoff, who combined diffraction gratings with prisms to separate closely spaced spectral lines and helped establish the science of elemental spectrometry. Fundamental to Kirchhoff's results was the use of a Bunsen burner, invented by Kirchhoff's friend Robert Bunsen [1], which isolated sources of background light. Kirchhoff also attempted to mathematically derive Fesnel's results, with mixed success: while Kirchhoff's methods and approximations are insufficient to apply to the study of electromagnetism in general, they work exceptionally well for visible light and are the basis for most studies in diffraction [2]. Indeed, Babinet's result on complimentary apertures can be derived immediately from Kirchhoff's theory.

Later, in the 20th century, it was discovered by pairing wave-particle duality in quantum theory with the wave function of Erwin Schrödinger that repeating Young's double-slit experiment with single photons or even electrons, fired individually and spaced over long periods, can yield diffraction patterns identical to those of a continuous beam of light provided information about the initial conditions of the system remains sufficiently obscure. Conventional interpretations posit that in these experiments it is the wavefunction of the particle which is diffracted, and indeed treating the system in this way yields results which are mathematically and experimentally consistent. This is particularly interesting because of the collapse of the diffraction pattern when certain information about the system is acquired, and the diffraction pattern can even be restored when this information is destroyed. What these results mean is still mysterious and epistemological debates over determinism and the nature of quantum mechanics ignited by these experiments continue to this day.

1.2 Physical Principles

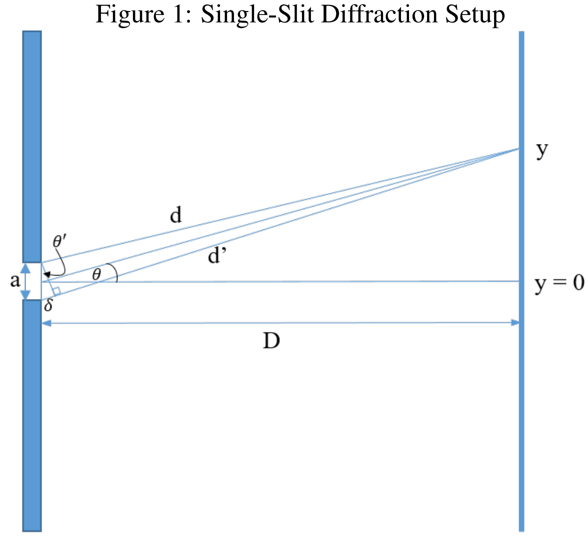
1.2.1 Huygen's Principle

Naturally you are already somewhat acquainted with the principles of interference, that when two waves meet the resultant wave is the sum of the amplitude of the component waves at each point, resulting in both constructive and destructive interference. An understanding of this principle, combined with basic ray-tracing techniques, allows one to accurately predict diffraction patterns without much of a complicated mathematical apparatus. Indeed, as J.D. Jackson notes in his *Classical Electrodynamics,* diffraction is a second order approximation for electromagnetic waves, with ray tracing being the first order. Calculations within the diffraction framework are achieved by the same techniques used in geometric optics save one considers points where 'waves' or rays meet with path lengths differing by odd whole number multiples of $\lambda/2$ (*e.g.* $\lambda/2$, $3\lambda/2$, etc. with λ being the wavelength) as intensity minimums and points with path lengths differing by multiples of λ as intensity maximums. When combined with Huygen's Principle, that all points on a spherical wavefront may be considered to be the source of a spherical wave, this second order approximation becomes a powerful tool. Let's look at this in more detail.

1.2.2 Single-Slit Diffraction

Suppose we have a beam of light of wavelength λ incident on a slit of width a, and we are attempting to calculate the location of intensity maxima on a screen. Let the height of the maximum above the center of the slit be y and the distance between the slit and the screen be D. Furthermore, we desire to use the small angle approximation for the angle θ between the perpendicular from the slit to the screen and y be small. See figure 1. By Huygen's Principle we must treat each point on the wavefront as being the source of a spherical wave; in particular, we will focus the two points on the edges of the slit [1]. We can consider only these points because our slit is rather narrow, and by treating these 2 most distant points we can characterize the interference pattern without much loss in accuracy. Then the difference in path length between these two points is $\delta = d - d'$. Note in particular the angle from the central length to the point y, namely θ, is nearly $\theta \approx \tan\theta = y/D$.

Figure 1: Single-Slit Diffraction Setup



Now because of our small angle approximation we see that the rays from our two points on the slit are nearly parallel, at least near the slit. Then the triangle formed by projecting the shorter ray onto the longer one is a right triangle with the interior angle θ' being approximately θ, so $\theta' \approx \theta$ [2]. From our small angle approximation then $\tan\theta \approx \theta$, $\tan\theta' \approx \theta'$, and hence $y/D = \delta/a$ so $y = D\delta/a$. Now since y is a maximum the difference in path lengths is a whole number of

[1] Justify this. Points near the center of the slit combine destructively with wavelets generated on the opposite side of the central line. Let the height of a point above the center of the slit be x. Now use a parallel ray approximation to show that $\mid x \mid \leq d/4$ there is a point opposite the central line a distance x and theese points generate wavefronts which combine destructively, so that the difference in their path lengths to y is $\frac{m\lambda}{2d}$ when the conditions for equation 1 hold.

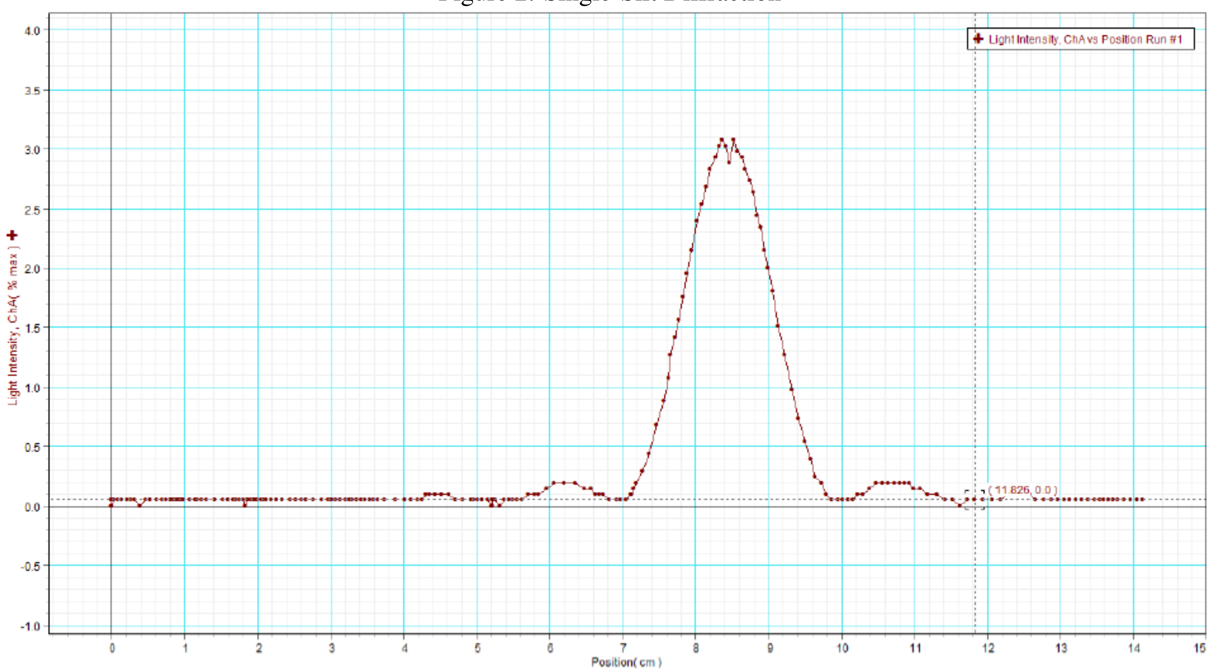
[2] Prove this using the theorems for parallel lines crossed by a transversal (hint: you'll be constructing 2 right triangles).

wavelengths, so rays originating from the top and bottom of the slit combine constructively [3]. Hence $\delta = d - d' = m\lambda$ for some m a natural number, and

$$y = \frac{m\lambda D}{a}, \quad m \in \mathbb{N}. \tag{1}$$

If we plot light intensity as a function of position on the screen, then positions of minima and maxima become quite apparent, as shown in the experimentally obtained data in figure 2. Note that the greatest intensity is at the $m = 0$ order maxima, in the center of the diffraction pattern.

Figure 2: Single-Slit Diffraction



Now if it seems that this calculation required an excessive amount of approximation, recall that the key assumption (and essentially our only assumption) was the small angle approximation from which all our calculations followed. The approximation $\sin\theta \approx \theta$ is accurate to less than 5% error up to $30^\circ = 0.53\,\mathrm{rad}$, and hence we can expect that for low order maxima the main source of error in our experiment will not be that due to the small angle approximation [4]. If you still aren't convinced, complete the diffraction experiment to determine exactly how far our theoretical calculation is from reality! [5].

1.2.3 Double-Slit Diffraction

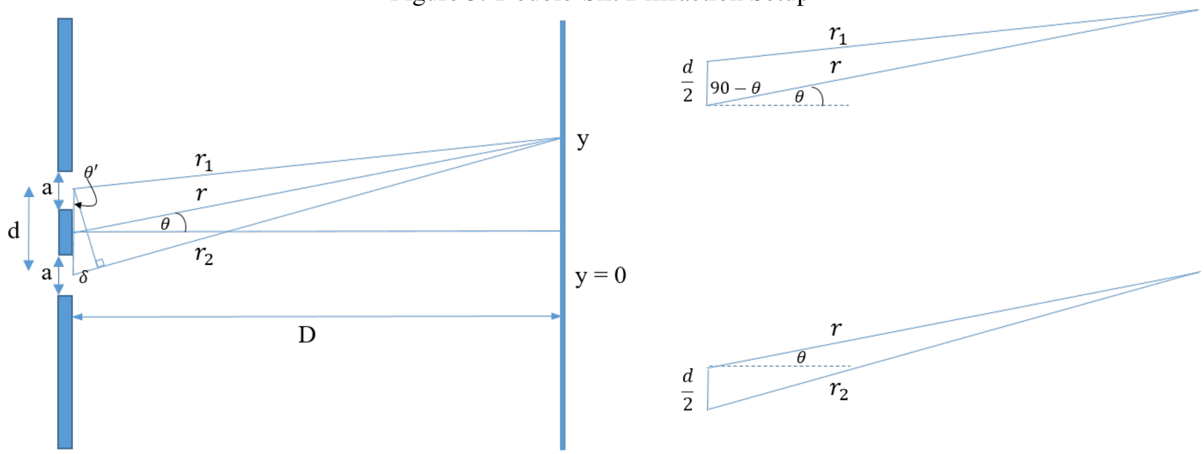
The setup for double-slit diffraction is quite similar to that for single-slit diffraction. The main difference in our presentation is the use of the variable d to refer to the displacement between slits and variables r, r_1, and r_2 to refer to lengths between the slit and screen.

[3] Draw a diagram of wave interference convincing yourself that if the path length differs by a multiple of a whole wavelength the result is constructive interference, while if the path length differs by a multiple of a half wavelength the result is destructive interference.

[4] To motivate the small-angle approximations, note the Taylor series for $f(x) = \sin x = x - \frac{1}{6}x^3 + ... \approx x$, for small values of x. Indeed, $x^3 \to 0$ more quickly than $x \to 0$. A similar case is true for $h(x) = \tan x$. Plot $x, \sin x$, and $\tan x$ on the same graph near $x = 0$ to visualize this. It is worth noting that the small angle approximation $\tan\theta \approx \theta$ is less accurate than the approximation $\sin\theta \approx \theta$. Make a table demonstrating the relative errors of these approximations for small angles to convince yourself of their accuracy and usefulness, and to gain an understanding of when these approximations can reliably be used.

[5] Perhaps add a note on the difficulty of evaluating Kirchoff's Integral.

Figure 3: Double-Slit Diffraction Setup



The derivation of the location of maxima is somewhat more involved than for the single-slit case and left as an exercise [6]. Without deriving the relation, and simply knowing that for double-slit diffraction

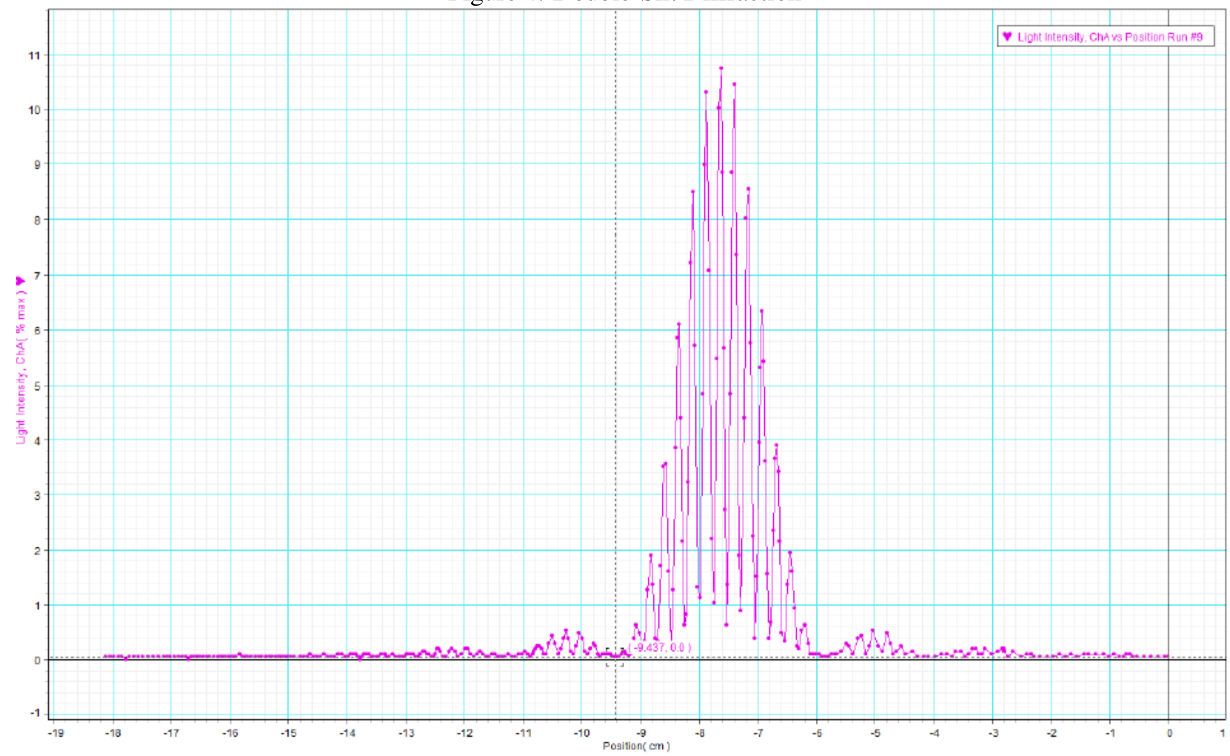
$$d \sin\theta = m\lambda, \quad m \in \mathbb{N}, \tag{2}$$

we can analyze double-slit diffraction in general. First, note that information about the location y of maxima can be obtained from $\sin\theta$ when D is known. Now recognize that two closely spaced slits each generate their own single-slit patterns superimposed on the double-slit pattern. This superposition becomes apparent when viewing the double-slit diffraction pattern on a screen (see figure 4). In the figure the large-scale structure of the pattern is a single-slit diffraction pattern, with the finer details resulting from the double-slit interference. Indeed, double-slit diffraction patterns are contained within single-slit envelopes, which you can confirm during the course of your experiment [7]. This means you can use equation 2 to determine the separation of a double slit and equation 1 to determine the width of the slits, once you determine the location of a single-slit maxima.

[6] To derive the location of the maxima, use the small angle approximation for θ as before, noting this makes r_1 and r_2 approximately parallel and equal to r near the slits. Reference the setup in figure 3 and use the law of cosines on the triangles there, along with the relation $r_2 - r_1 = \delta$, to determine δ and note that for constructive interference $\delta = m\lambda$.

[7] In general $(n + 1)$-slit diffraction patterns are contained in n-slit diffraction patterns.

Figure 4: Double-Slit Diffraction



1.3 Why do we care?

Interference and diffraction are most apparent when light propagates through apertures or around objects (Babinet's Principle says these are equivalent cases) that are very small (usually measured in micrometers). It becomes much easier to measure the fringe patterns (millimeters in size) than to measure the aperture or object itself. This becomes a powerful tool in many applications of microscopy and nanoscience. What applications have you heard of?

1.4 Error Propagation

Naturally an essential skill when obtaining, analyzing, and publishing data is the proper handling of uncertainty. Despite our conventionally negative attitude towards the term 'error', it is inevitable in experimentation and, when sufficiently minimized, is essentially synonymous with uncertainty. Error or uncertainty in one measurement may have an effect on the overall experiment which is greater than the uncertainty in just that measurement. This is *error propagation*. To demonstrate error propagation, consider the uncertainty in the measurement of the circumference of a circle, say of circumference C and radius r. By taking a differential (implicit differentiation) on the relation $C = 2\pi r$, we have $\Delta C = 2\pi \Delta r$, and hence the error in r is approximately $\Delta r = \Delta C / 2\pi$. Now for the area of the circle, however, $A = \pi r^2$ results in an error of $\Delta A = 2\pi r \Delta r = 2\pi C \delta C$, and scales with the value of r (and hence your measurement of C).

This process is rather simple when measuring simple properties of circles, but not as intuitive in other experiments. For example, what happens when the differential yields a negative value? Or when there are multiple sources of error in a single measurement? To resolve this conundrum it is standard to add errors *in quadrature.* If an experimenter is determining a value y which is a function of variables $x_1, x_2, ..., x_n$ the uncertainty in $y(x_1, x_2, ..., x_n)$, called σ_y, can be approximated by

$$\sigma_y^2 \leq \sigma_{x_1}^2 \left(\frac{\partial y}{\partial x_1} \right)^2 + \sigma_{x_2}^2 \left(\frac{\partial y}{\partial x_2} \right)^2 + ... + \sigma_{x_n}^2 \left(\frac{\partial y}{\partial x_n} \right)^2, \qquad (3)$$

where σ_{x_i} is the uncertainty in the measurement x_i and the function is evaluated at the mean for each of x_i.

In practice it can be somewhat tricky to determine what uncertainty to use for each σ_x. Ideally each piece of equipment will have an attached statement with a sensitivity rating; however, this is often not the case. It will often be up to your personal discretion to determine what value to use for σ_x, however, common sense and good practice dictates that oftentimes 1/2 the value of the most accurate decimal place on your measurement is a wise choice [8]. For example, with a ruler that measures up to millimeters the uncertainty would be $\pm 0.5\,\mathrm{mm}$. And, it is important to note that when confronted with two reasonably valid choices for the uncertainty it is best practice to pick the larger value.

Now, when designing and conducting experiments it is important to be aware of two kinds of errors: systemic errors and random errors. A systemic error is one which is due to the experimenter or system performing the experiment, and is consistent through each trial. Being aware of these kinds of error and minimizing them is a critical part of experimentation. An example of systemic error may be a difference in delays between sensors receiving a signal and the recording of those signals by a computer. Oftentimes systemic errors can be dealt with by careful examination of your apparatus; however, systemic errors will always persist in any experiment and it is important to note possible sources of systemic error and their magnitude when reporting your results.

The other kind of error mentioned, random error, can be controlled for by simply increasing the number of trials run during your experiment. These errors result from inconsistencies in setup and execution which cannot be controlled. For example, while Hooke's law for springs $\vec{F} = -k\vec{x}$ is incredibly accurate and predictable across an appropriate ranges for x, multiple trials of the same experiment measuring the spring constant k will result in a predictably randomly distributed distribution even for the same spring, which distribution is centered on the mean value for k.

The formula for the standard error of the mean (standard deviation) is presented in equation 4

$$\sigma = \sqrt{\sum_{i=1}^{N} \frac{x_i - \bar{x}}{N-1}}, \tag{4}$$

which describes the spread of data from the mean $\bar{x}$ due to random error and is optimized when σ is small, tells us much about general ways to improve experiments [9]. Most importantly, it tells us that the influence of random errors decreases with a larger number of samples N, and that the standard deviation is undefined for a sample size of 1. Indeed, this indicates to us that when we run an experiment a single time we have no useful information about the size of the random error and we may conclude that any result obtained from such a shallow experiment is invalid. But, by increasing the number of trials the standard deviation decreases, and with the correspondingly smaller influence of random error on our our data the results are more reliable.

1.4.1 Reporting Your Results

When CERN reported the landmark discovery of the Higgs boson in 2012 they did so with the exemplary caution that characterizes the scientific community when publishing results of any kind. This can be seen in their report, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC* [3]. It is remarkable that nowhere in the article do they state that they have definitively discovered the Higgs boson, and even in the title the most they claim is that they had discovered a new kind of particle. While the particle they discovered seemed to have many of the characteristics of the Higgs boson, CERN researchers waited for consensus of the scientific community that the particle observed is indeed almost certainly the Higgs before declaring their discovery in stronger terms.

Diving more deeply into the CERN article we can learn much about the proper reporting of scientific results, applicable both in landmark cases and in situations more modest. The report includes multiple *Systemic Uncertainties* sections, one for each type of particle event described, in which we find an in-depth analysis of their experiment including background events and possible complications confounding their results. This analysis is not just qualitative and includes a numeric uncertainty associated with every claim, as well as a possible source of that uncertainty. For example, the article states that:

"The dominant experimental uncertainty on the signal yield ($\pm 8\%$ and $\pm 11\%$ [for the $7\,\mathrm{TeV}$ and $8\,\mathrm{TeV}$ data respectively]) comes from the photon reconstruction and identification efficiency, which is estimated with data using electrons from Z decays and photons from $Z \to l^+ + l^- + \gamma$ events."

---

[8] In some cases this method will result in an underreporting of error, especially if you're using the default setting on your measuring apparatus. For example, in some experiments the main source of error may be in human reaction time, which may find it difficult to match the accuracy of a stopwatch with $0.1\,\mathrm{s}$ accuracy. In this case it may be best to state the uncertainty in your measurement as $0.5\,\mathrm{s}$ instead of $0.05\,\mathrm{s}$.

[9] Note this σ, the standard deviation, is dimensionless and related to but not the same as the theoretical uncertainty in your measurement σ_y in equation 3. It is unfortunate that these two are so closely related and use similar symbols, but this is the standard usage.

Without going into the science behind the CERN experiment, we can still be inspired by their methods for accurately reporting data. From this statement not only is their result clear (the 7 TeV and 8 TeV data) but also the uncertainty of that data ($\pm 8\%$ and $\pm 11\%$ respectively), potential sources of that uncertainty (photon reconstruction and identification efficiency), and the method used to calculate that uncertainty (data from Z decays). This is a statement not only of their results but also of their confidence in those results.

When reporting your own results it is important to include as much relevant detail as necessary [10]. Oftentimes you will be reporting the mean result from your data, and stating with that mean the standard deviation, number of trials, and theoretical uncertainty from equation 3. If you do so you will be reasonably covered in case of controversy, so long as you do not claim any discoveries which don't follow from your data. Also, it is common for a calculator to return many more decimal points than are appropriate for your experiment. Report only up to the accuracy given by your theoretical uncertainty [11]. In general, it is best to simply restate what it is you actually measured as well as the method of that measurement, in which case no one can take the results of your experiment away from you. As an example, the measurement of a spring constant in the laboratory may be stated in the following manner:

"We measured the stiffness k of a spring by measuring the displacement of the spring from equilibrium when the spring was hung vertically with one end fixed and various masses hung from the free end. The spring constant k was found to be $k = 0.50 \pm .04\,\mathrm{N\,m^{-1}}$ (or $k = 0.50(4)\,\mathrm{N\,m^{-1}}$) with a standard deviation of $\sigma = 0.043$ using 10 differing masses ranging from $0.1\,\mathrm{kg}$ to $5\,\mathrm{kg}$ Possible sources of error include inaccuracies in the masses used, which were measured by a digital scale with $0.05\,\mathrm{kg}$ accuracy; or inaccuracies in measuring the displacement perhaps caused by imperceptible gentle bobbing of the spring or by measuring along an angle not directly in line with the vertical of the spring-mass system, which displacement was measured by ruler sticks with $0.5\,\mathrm{mm}$ accuracy.".

2 The Experiment

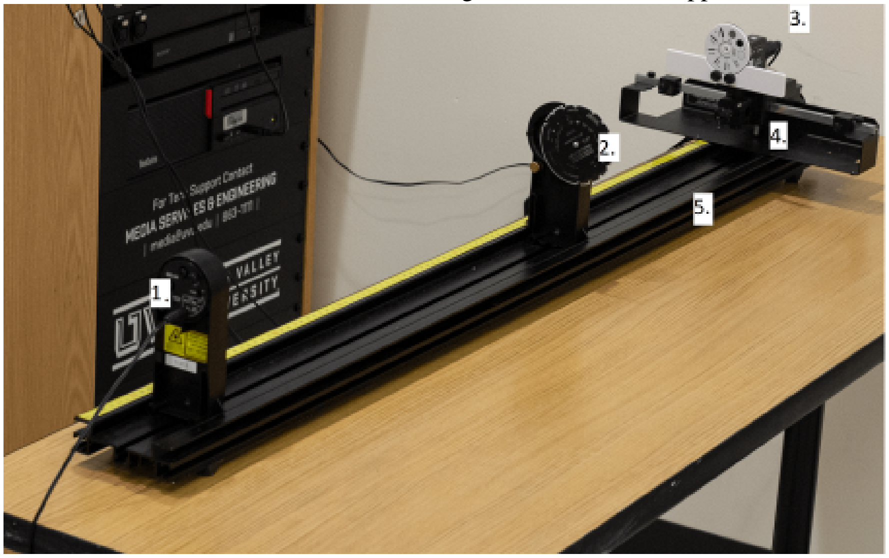
2.1 The Apparatus

The main part of the experiment consists of a track on which various optical instruments are placed. In particular, these instruments include a red $650\,\mathrm{nm}$ laser, which shines on apertures of various dimensions. The resulting diffraction patterns are recorded by a light sensor, which is placed on a track of its own (the linear translator) which is perpendicular to the path of the laser. Measurements along the linear translator are correlated with measurements of light intensity as read by the light sensor, which data is interpreted by a Pasco Scientific 550 Universal Interface and recorded by the software Pasco Capstone. Pasco Capstone allows the student to locate maxima and minima, and in doing so the student will be able to corroborate or refute the results of equations 1 and 2.

[10] The key word is *relevant*. Some reports, even professional ones, will include superfluous and irrelevant information such as the temperature in the room, levels of ambient light, elevation above sea level, etc. Depending on the experiment this may be pertinent information and it will be up to your judgement to determine what to include. In general include all detail which would be practically useful to some attempting to replicate your experiment.

[11] So, for example, instead of reporting $13.54325 \pm .05$ eV report $13.54 \pm .05\mathrm{eV}$.

Figure 5: Diffraction Apparatus



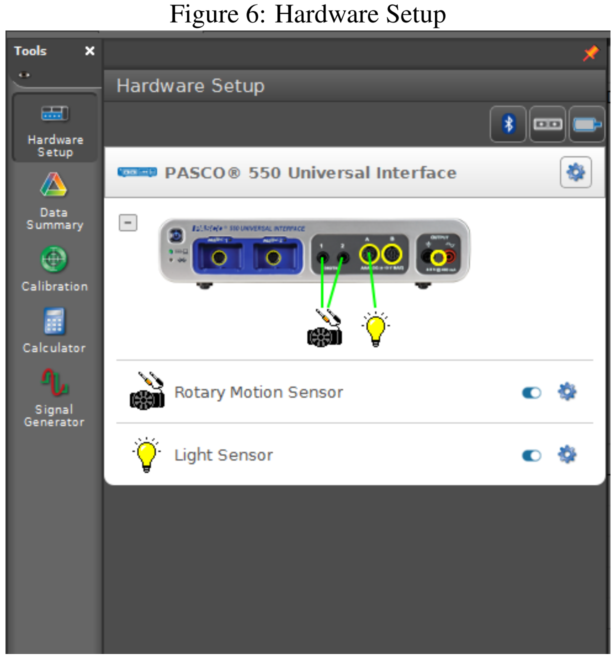
1. 650 nm Laser
2. Aperture Selector
3. Light Sensor
4. Linear Translator
5. Optical Track

2.2 Performing the Experiment

We begin with assembling the apparatus.

1. Insert the nut under the frame of the linear translator into the slot on the track, slide the linear translator into a desirable position, and tighten.
2. Insert the laser and aperture selector into the optical track.
3. Connect the laser and the Pasco 550 Universal Interface to power.
4. Connect the Pasco 550 Universal Interface to the computer via USB and the Rotary Motion Sensor (attached to the linear translator) to the 1 and 2 ports on the Pasco 550 Universal Interface. Connect the light sensor to the A port on the Pasco 550 Universal Interface.

Figure 6: Hardware Setup



Now to set up Pasco Capstone:

1. Open Pasco Capstone on the computer.
2. Click "Hardware Setup".
3. Click the 1 and 2 ports on the image of the Pasco 550 Universal Interface. Scroll down to select the rotary motion sensor.
4. Click on the settings for the Rotary Motion sensor and make sure that the "Medium Pulley" is selected to ensure correct calibration of rotation to distance.
5. Click the A port on the image of the Pasco 550 Universal Interface and scroll down to select the light sensor.
6. Exit hardware setup and open the graph. On the graph's vertical axis select measurements by clicking on the axis and selecting "Light Intensity".
7. Similarly, select "Position" for the horizontal axis.

To make measurements select the desired aperture on the aperture selector and press the "Record" button. Then slowly drag the sensor from one side of the linear translator to the other. It may take several trial runs to learn how to smoothly drag the sensor from one side to the other. Zoom and scale your graph as appropriate.

2.3 Analysis

Now that we have empirical data on our diffraction setup we can compare our results to those predicted by equations 1 and 2. Knowing the wavelength of the laser light is $650\,\mathrm{nm}$ and by measuring the distance from the aperture to the light sensor, we can the locations of maxima y to measure the slit width a and distance between slits d.

1. For a single-slit aperture, use your plot of light intensity to determine the distance from the central maximum to the first order maximum. What is the distance between these points?
2. Use equation 1 to calculate the slit width a. Does this seem reasonable?
3. Perform an error analysis for the single-slit experiment to calculate your predicted uncertainty in the slit width. What contributes to this uncertainty and how can you combine these effects in your calculations?
4. See if your theoretical prediction falls within the error bounds of your experiment. How does this affect your confidence in your measurement?
5. Think of ways that you could reduce the uncertainty in your measurements. Try them out now if you can!
6. Repeat this procedure for various aperture widths and repeat the entire experiment for the double slit experiment. Can you confirm the double-slit diffraction pattern is bounded by the single-slit diffraction pattern of the same aperture width? Are there any variables unaccounted for which affect your experiment?

Carefully collect your findings and record them in a dedicated laboratory notebook. Use this notebook to write a report on diffraction summarizing your findings, including proper error analysis. Answer the following questions:

1. What do your data mean? Why?
2. How can this experiment or the analysis be improved?
3. What predictions can you make and verify in future experiments?

3 The Next Step

What else could you measure using light diffraction? Are there things that this procedure could measure that would be difficult or impossible using other methods?

With these types of measurements in mind, describe another experiment you could perform. What questions could you ask about the things you are measuring? What other equipment would you need? What measurements would you need to make? Could any of the other experiments you have done also be useful?

If this plan seems feasible, it can be the beginnings of the proposal you write for this class. Your designed experiment (measurements and analysis) should take you about 10 class periods. Do some background research to predict what results you would get. Adapt your plan as necessary. Contact your instructor if you need equipment that is currently unavailable. Funding has been obtained for student-designed experiments in the class in past semesters (prior results do not guarantee a similar outcome).

References

[1] Isaac Asimov. *Asimov's Biographical Encyclopedia of Science and Technology*. Doubleday and Company, Inc., Garden City, New York, 1982.

[2] John David Jackson. *Classical electrodynamics*. AAPT, 2 edition, 1999.

[3] Georges Aad, Tatevik Abajyan, B Abbott, J Abdallah, S Abdel Khalek, Ahmed Ali Abdelalim, O Abdinov, R Aben, B Abi, M Abolins, et al. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(1):1–29, 2012.