Spherical aberration, one of the five Seidel Aberrations, concerns the incorrect mapping of a point source on one side of a spherical lens onto an image on the opposite side that is smeared along the optical axis. Through the use of an apparatus that allows the comparison of aperture width to the location of the critical focus (the focus closest to the lens), we have examined this phenomenon. We find that our results are in excellent agreement with the behavior expected by Seidel’s formula for third-order optics, and that the observed correlation between aperture width and critical focus location is commensurate with that of the theoretical prediction. However, we do note a discrepancy between the absolute location of the critical foci and the expected values, largely due to a slight inaccuracy regarding the measurement of one of the base distances.

I. INTRODUCTION AND HISTORY OF THE PROBLEM

Optical systems utilizing the geometry of refractive and reflective media can imperfectly transmit the light passing through them, producing a variety of aberrations. In 1857, Philipp Ludwig von Seidel identified five such aberrations that corresponded to the simplest treatment of the full optical system, the so-called third order or Seidel optics. These five aberrations are coma, astigmatism, curvature of field, distortion, and spherical aberration, each of which deals with a specific type of imperfect transmission arising from a different orientation of the source.

The subject of this lab is spherical aberration, which involves a specific phenomenon arising from a point source impinging on a spherically shaped lens. Rays passing through the lens will be bent unevenly, producing a distributed region of focused light rather than a single point. This results in an unclear image, which can only be alleviated through the use of corrective optics.

II. BASIC PHYSICS

The location at which light rays emitted from a point source impinge on the surface of a spherical lens is labeled A in FIG. 1. The value $\phi$ is the angle between the line connecting the point of intersection and the center of the lens and the optical axis (the line connecting the center of the lens and the source).

With this in mind, the equation relating the path length of a given ray inside and outside the lens is given by:

$$\frac{n_1}{l_0} + \frac{n_2}{l_i} = \frac{1}{R} \left( \frac{n_2s_i}{l_i} - \frac{n_1s_0}{l_0} \right)$$  \[1\]

The values of $l_i$ and $l_0$ are dependent on $\phi$ and therefore the substitution of the values of $l$ with a formula involving $\phi$ quickly produces an equation that can only be solved by approximating the value of $\sin(\phi)$ and $\cos(\phi)$. This approximation takes the form:

$$\cos(\phi) = 1 - \frac{\phi^2}{2!} + \frac{\phi^4}{4!} - \frac{\phi^6}{6!} + \ldots$$

$$\sin(\phi) = \phi - \frac{\phi^3}{3!} + \frac{\phi^5}{5!} - \frac{\phi^7}{7!} + \ldots$$  \[2\]

For small values of $\phi$ (rays nearly parallel to the optical axis, or paraxial), all but the first terms can be ignored. This is the so-called first order approximation, where $\cos(\phi) \approx 1$ and $\sin(\phi) \approx \phi$. In this narrow regime surrounding the optical axis, most of the incident rays aren’t too terribly distorted, and so they converge on a common focal point.

However, in order to examine larger values of $\phi$, and therefore images whose visual information might impinge on a large section of the lens, one must include higher order terms. The terms that still concern monochromatic light (i.e. the terms that are large enough that prism-like splitting of the light into different parts based on
color is not yet an issue) are contained within the next value in the series. Ludwig von Seidel first studied the implications of substituting \( \cos(\phi) \approx 1 - \frac{\phi^2}{2!} \) and \( \sin(\phi) \approx \phi - \frac{\phi^3}{3!} \) into the above equation and produced the five different aberrations that bear his name. Substituting the third order approximation into the above equation produces the formula:

\[
\frac{n_1}{s_0} + \frac{n_2}{s_i} = \frac{n_2 - n_1}{R} + h^2 \left( \frac{n_2}{s_0} \left( \frac{1}{s_0} + \frac{1}{R} \right)^2 + \frac{n_2}{2s_i} \left( \frac{1}{R} - \frac{1}{s_i} \right)^2 \right)
\]

(3)

This equation contains the initial first-order result, but includes an additional \( h^2 \) term that corresponds to the higher-order correction. This higher order correction produces the characteristics of the spherical aberration. Rays that impinge on the lens further away from the optical axis are bent more dramatically and therefore come to focus closer to the lens than rays that impinge on points closer to the axis. This extra bending produces the characteristic smearing of spherical aberration.

Thus, the equation for spherical aberration included above provides a relationship between aperture and focal length for the rays that are not paraxial. By varying the extent of the most widely spread ray impinging on the lens, a keen observer can control the horizontal location of where the foci start to appear.

III. DESCRIPTION OF THE EXPERIMENT

The procedure described above loosely summarizes the intended procedure being performed in this experiment. The apparatus in question consisted of three groups of components, listed in FIG. 2. The first group consisted of a light source, a lamp, and a diffuser that removed the coherent image of this lamp. The second group consisted of a 50µm pinhole, a variable aperture, and a spherical lens oriented to refocus the light coming from the pinhole. This group was mounted on a removable stand to allow the aperture to be removed from the device and measured without compromising the relative distance between the pinhole, lens, and aperture. The third group consisted of a CCD camera mounted on a stand whose horizontal location could be changed via the use of a micrometer screw.

Initial calibration first involved ascertaining the distance from the case holding the CCD camera and the lip of the aperture when the micrometer screw on the CCD stand was set to its zero value. This provided a baseline for all subsequent measurements, and allowed for the value of the micrometer screw to be read with only a constant correction. This calibration involved use of a digital caliper whose relative inexactness necessitated the taking of multiple measurements. Thus, three measurements in total were taken for this distance, and the average was assigned as the measured value for this stance. In addition, two other values factored into the calibration distance; the depth between the aperture lip and the actual aperture and the depth between the outer casing of the CCD camera and the CCD chip itself. Again, due to the inaccuracy of the calipers, multiple measurements contributed to an average value rather than a single measurement of the requested values.

Once the experimenters successfully performed the initial calibration, the next step involved finding the correspondence between the size of the aperture and the distance from the lens at which the first focus appeared. In this procedure, the distance was the independent variable and the size of the aperture was the dependent, due to the fact that the micrometer screw was a far more reliable value to set and which could then be roughly approximated multiple times with the aperture.

Thus, the distance from the CCD to the lens was set, and the experimenters varied the width of the aperture until the bright pattern first appeared on the CCD chip. The experimenters then removed the stand containing the aperture and measured the width of the aperture using the digital caliper. Because of the inaccuracy of the caliper, as well as the fact that measuring the aperture had a chance of increasing its size slightly, multiple measurements of the critical aperture width for a given focal distance were necessary to get an accurate result.

The distance between the CCD to the lens was first set to the base value and increased by 3mm using the micrometer screw. At each increment, the mean width of the critical aperture was ascertained by recording four aperture widths and averaging.

Midway through the process of collecting the data the maximum possible distance possible for the micrometer stand was achieved, but the smallest possible width of the aperture was not. Thus, the CCD on the micrometer stand had to be moved back on the stand in order to allow a larger maximum distance (and therefore a smaller aperture). The experimenters then recalibrated the apparatus and continued the measurements, this time starting with a value for the micrometer screw that was commiserate with the maximum distance of the furthest location recorded by the previous configuration. Once the values of this initial measurement were verified to be in the same regime as the previous, the CCD was moved...
back using the procedure listed above until the experimenters reached a maximum useful distance.

IV. MEASUREMENTS AND DISCUSSION

In all, the procedure described above involved the examination of the critical aperture width for eleven different CCD camera positions. The first step in comparing the collected data to the theoretical expectation involved the generation of an expectation function from the ray tracing values included in the lab writeup. The initial shape of the expectation curve, as given by a plot of the data with focal point position as the independent variable and aperture width as the dependent, produced a graph that involved a square root and was singular at a specific point. Because it was particularly difficult to fit a function to these data while they were this form, the experimenters converted the set to instead use the aperture width as the independent variable and the focal point position as the dependent. This allowed for the data to be effectively fit to a function of the form \( f(x) = ax^2 + bx + c \), where \( f(x) \) was the focal point position, \( x \) was the critical aperture width, and the values \( a, b, \) and \( c \) were allowed to vary.

Note that the ray tracing algorithm output the focal point position as an absolute distance between focus and emitter, not as a relative distance between aperture and focus (which is obvious, since there is no aperture in the simulation!). The result of this discrepancy between the theoretical and experimental results required that the distance between the aperture and the pinhole be added to the experimental data in order for the results to be comparable. The value for the distance between the pinhole and the aperture was estimated to be 45mm, and so this value was included in the overall experimentally measured distance between focal point and pinhole.

There were four sources of errors that contributed to the total uncertainty: the measurement of the aperture width, the CCD chip inset, the aperture ring inset, and the distance between the aperture and CCD housings. In the case of the aperture width, the errors were actually in the independent variable, and so a conversion as described in allowed these uncertainties to be transferred into errors in the measured focal distance\(^2(p.65)\). The other three errors directly corresponded to the focal length distance, and since these measured values were summed together their errors summed as well\(^2(p.50)\).

With both the necessary correction and the parameters of the expectation function ascertained, the next step involved fitting the expectation function to the assembled data. Figure 3 contains an illustration of this attempt, plotting the theoretical and experimental results side by side. Because the absolute separation between the focal point and pinhole was just an estimation, the fitting algorithm allowed the offset \( c \) of the expectation function to vary, while keeping the \( a \) and \( b \) values- the values maintaining the shape of the curve- constant. Since the expectation function contained no dependence on the y-displacement of the variables, a set of values that had the same relative dependence. That is, while the absolute distances would be preferred, the fundamental behavior of the system was purely relative, and therefore this behavior did not depend on the distance between the pinhole and aperture.

With the value \( c \) allowed to vary, the fit algorithm calculated a net offset value of 108.373mm, which lies within the accepted error of the expectation function’s \( c \) of 108.783±0.2865mm. However, the estimated variance for this attempted best fit was 10.2864. Careful examination of the graphed function revealed that the first experimental measurement (the one colored red in the experiment) lay far outside the range of values expected for the given aperture width. From this inspection, it was decided that this value represented an outlier, and its removal would lead to a better fit.

When this outlier was removed, the estimated variance for the expectation function was much more acceptable; a value of 1.0607. However, a consequence of this omission was that the value of the y-offset became 107.726mm, which lay outside the uncertainty for the expected y-offset from the theoretical prediction.

V. SUMMARY AND CONCLUSION

The excellent fit between the edited data and the expectation value leads inexorably to the conclusion that the expectation function accurately predicts the dependence of aperture width on critical focus location. Since the simulation from which the expectation function was derived is a direct implementation of equation (3), it can therefore be concluded that this equation is an accurate approximation of the behavior of light as it refracts through a spherical lens. While omission of the final value did result in an inaccurate reading of the absolute distance between focus and pinhole, the verification that the vast majority of the data fit the theoretically predicted behavior was far more valuable to the experiment. If this experiment were to be continued, the experimenters would definitely measure the distance between the aperture and the pinhole, removing the necessity to estimate this value and producing a more accurate portrayal of the system.

---

FIG. 3: A comparison of the calculated data and the experimentally observed data. The blue data points are the theoretically predicted values and the blue line is the best quadratic fit for these data. The orange data points are the experimentally measured values, as well as their x- and y- error bars. The red dotted line is the best fit for the observed data when the purple outlier is included, and the orange solid line is the best fit when it is omitted.