Optical Methods in Experimental Mechanics

Part 45: Measuring Phase Difference—Part I: The Problem

REVIEW AND PURPOSE

The previous article of this series reviewed the fundamental concepts of optical measurements and outlined the ways in which they are utilized in several techniques for measuring strain, velocity, and displacement. We noted that all of the methods that are based on interference require precise measurement of either absolute path length difference or, more commonly, changes of path length difference.

This article begins a short series on precise determination of path length difference by reviewing basic interferometry and by extending this learning to accommodate the less-than-ideal interference usually occurring in practice. These studies lead directly to an understanding of the problems that must be solved in order to obtain useful data from any interference technique.

Until now, we have usually spoken in term of path length difference (PLD) or relative retardation (r) because of their obvious meanings as physical distance. When solving diffraction problems and when measuring PLD by compensation or electronic techniques, common practice is to work in terms of phase difference or phase change. We here gradually slide into this usage as we begin to explore various methods of obtaining precise measurements of phase difference at a single point or over a whole field. The whole-field result from applied experiments is a "change of phase difference" map, but it is often called a "phase map" for short. A photoelasticity fringe pattern, for example, can be thought of as a phase map if the experiment started from zero load and if the relative retardation is converted to phase difference.

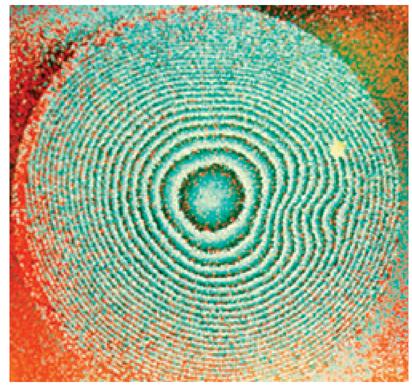
PERFECT COLLINEAR INTERFERENCE

It would be a good idea at this point to go way back and review Parts 2 and 3 of this series, published in 2002, which dealt with interference of light waves, path length, and the generic interferometer. The intensity or irradiance resulting from **ideal** collinear interference of two identical coherent light waves was found to be,

$$I_s = 4A^2 \cos^2\left(\frac{\pi r}{\lambda}\right) \tag{45.1}$$

where I_s is the irradiance created by interference of two waves having the same wavelength λ and the same amplitude A; and r is the relative retardation or path length difference (PLD) between the two waves in spatial units. In light of our learning so far, let us change the form of this equation a bit so it agrees with common practice.

$$I_s = I_{\max} \cos^2\left(\frac{\pi r}{\lambda}\right) \tag{45.2}$$



Change of phase difference map obtained by phase-shifting electronic speckle pattern interferometry showing out-of-plane deformation caused by transient heating of a composite plate that contains a small delamination caused by impact. The anomalies in the fringe pattern show the location of the damage. Photo by X. L. Chen and G. L. Cloud at Michigan State University, ca. 1993.

This article:

- *reviews basic interference concepts,*
- *studies non-ideal interference,*
- *leads to formulation of the two problems that must be solved to obtain precise data from any implementation of interferometry.*

In whole-field experiments, the goal of interference measurement is a map of change of phase difference, relative retardation, or change of path length difference. The result is usually called a phase change map.

The series, Optical Methods - Back to Basics, *is written by University Distinguished Professor Gary Cloud of Michigan State University in East Lansing, Michigan. It began by introducing the nature and description of light and will evolve, with each issue, into topics ranging from diffraction through phase shifting interferometries. The intent is to keep the series educationally focused by coupling text with illustrative photos and diagrams that can be used by practitioners in the classroom, as well as in industry. Unless otherwise noted, the graphics in this series were created by the author.*

The series author, Professor Gary Cloud (SEM Fellow), is internationally known for his work in optical measurement methods and for his book, Optical Methods of Engineering Analysis.

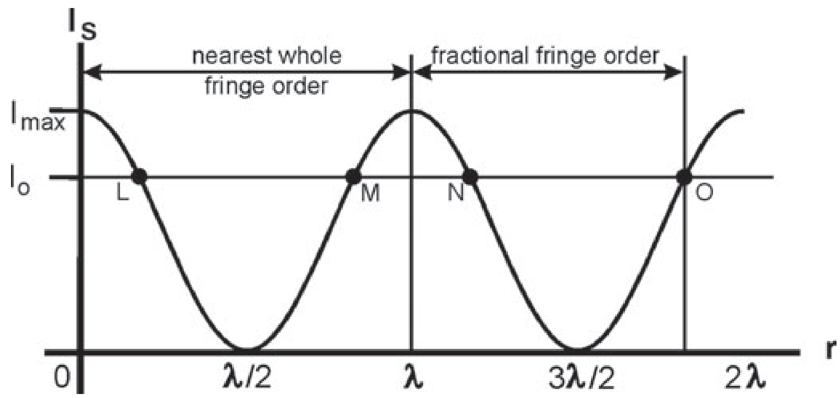
If you have comments or questions about this series, please contact Jen Tingets, journals@sem1.com.

doi: 10.1111/j.1747-1567.2010.00702.x

Where I_{max} is the maximum intensity or irradiance created by the sum of the waves. Later, we will write the PLD in phase angle units, but it seems easier to deal with length units for awhile.

For ideal interference of two identical waves, the observed intensity varies between zero and a maximum as the square of the cosine of the PLD times pi over wavelength.

The sketch below shows the resultant intensity plotted as a function of the retardation.



OBTAINING PATH LENGTH DATA FROM INTENSITY—IDEAL CASE

As we learned, one of the many strengths of all interference techniques lies in the fact that, in principal, simple measurements of intensity allow us to measure PLD's with sensitivity and accuracy equal to a fraction of the wavelength of light. This measurement of PLD can be accomplished either point-by-point or simultaneously over a large field to yield a map of PLD, which, in mechanics applications, is often the desired map of displacement caused by load.

The power of interference techniques is that PLD can, in principal, be obtained from measurements of intensity.

Serious problems immediately arise even in this ideal case. Thinking about these issues is worth some time because it will help us work through the more complex non-ideal situation. The first is the difficulty inherent in all interferometry. Consider the figure above, and suppose that the measured intensity is I_o. This single measurement does not tell us exactly where on the graph we are. The PLD (r) might correspond to that of points L, M, N, O, or an infinity of other values if the graph were extended to the right or left. We have suggested that this difficulty can be solved in some applications by counting interference orders (i.e. fringe orders) as they pass by. In a Michelson interferometer, for example, one starts at a certain PLD, often zero, and counts the alternating intensity maxima and minima as the PLD is changed. The same is true of photoelasticity. For the simple example pictured, suppose the final PLD corresponds to point O. One complete intensity cycle will have passed by as the PLD approaches its final value, so we know that the PLD corresponds to either O or N in the picture. Further, the intensity will have passed through the zero-intensity minimum at a PLD of $3\lambda/2$, so we know that the PLD is between that value and 2λ. In terms of fringe order, we have established the nearest whole- and half-orders, so we know roughly what the actual PLD is for our experiment. We could now interpolate and estimate that the PLD in the example is about 1.8λ, and that approximation might well be sufficient for the application.

Three related problems arise when trying to determine PLD from a measurement of final intensity through use of the retardation-intensity equation, namely:

- *We cannot determine which cycle of the intensity graph is the correct one.*
  - *When possible, this problem can be solved by counting cycles and half-cycles as they pass.*
- *We cannot determine the exact fraction of a cycle without knowing one more datum.*
  - *Possibly an observation of the maximum intensity can be used.*
- *The problem is poorly conditioned near maxima and minima of the graph.*
  - *Large changes of retardation yield only small changes of intensity in these regions.*

Usually, better accuracy is needed. We run into the second difficulty when we follow the plan and substitute the measured intensity I_o into equation 45.2 to solve for r. We have only one measurement and there are two unknowns. The trouble is that we do not know the combined intensities of the interfering beams I_{max}. The obvious solution is to measure it as we pass through a peak value while counting orders. However, in many experiments, the I_{max} varies as cycles of fringe orders pass by, and it almost always varies over the optical field. Also, the problem is poorly conditioned if the I_o lies near a maximum or a minimum, because rather large changes of PLD yield only small changes of intensity in those regions.

In summary, we find that it is possible but often not satisfactory to obtain PLD by counting interference orders and then obtaining somehow two measurements of intensity to use in the equation that relates intensity to PLD. The problems are compounded in real-world experiments, because they rarely approximate the ideal case. Monochromatic photoelasticity is probably the only technique of experimental mechanics that closely approaches ideal interferometry. Improved knowledge and techniques are required for precise measurements in almost all other implementations.

The problems found in the ideal case are compounded in real-world interferometry. More analysis is needed.

NON-IDEAL INTERFERENCE

In practice, interferometry is complicated by a number of factors, including:

1. The interfering waves likely do not have identical amplitudes, and they might differ by a large amount.
2. The waves might not have exactly the same polarizations.
3. The waves might not be perfectly coherent in temporal or spatial coordinates.
4. The starting path length difference probably is not zero or a multiple of the wavelength, meaning the initial intensity is not a maximum.
5. The intensity versus PLD relationship might be contaminated by vibrations in the setup, air currents, or other noise that causes it to deviate from its ideal shape.
6. The interfering waves probably do not travel along the same axis, meaning they undergo oblique interference, a problem that was discussed in Part 4 of these articles.

Consider for now only the first of the problems enumerated above. Dealing with that one shows how to attend to the most of the rest of the list, as will become clear eventually.

Complications experienced in non-ideal interference include:

- *The interfering waves do not have identical amplitudes.*
- *They might not have the same polarizations.*
- *They might not be perfectly coherent.*
- *The starting PLD is probably not zero or at a maximum or minimum.*
- *The intensity versus PLD relationship is probably contaminated by noise.*
- *The interfering waves might not be brought together along the same axis.*

COLLINEAR INTERFERENCE OF TWO WAVES HAVING DIFFERENT AMPLITUDES

Intuition and knowledge of the ideal case would probably lead you to the correct form of the relationship between intensity and PLD for the case where the interfering waves have different amplitudes. To be certain, let us outline the derivation, the result of which dominates interferometry calculations. Start, as we did in Part 2 of this series, with two simple coherent harmonic waves that are brought together along the z-axis. For variety we will say that one leads the other by a PLD equal to r. This time, however, allow the waves to have different amplitudes. Rather than use complex variables, which would be quicker, we regress to trigonometric forms that are easy to visualize.

Analysis of the collinear interference of two waves that have differing amplitudes shows how to attend to the complications mentioned above and suggests what measurements are required for precise determination of phase difference.

$$\begin{aligned} E_1 &= A\cos\left[\frac{2\pi}{\lambda}(z - vt + r)\right] \\ E_2 &= B\cos\left[\frac{2\pi}{\lambda}(z - vt)\right] \end{aligned} \tag{45.3}$$

Add these scalar amplitudes of the wave vectors together as usual. While at it, simplify matters by writing the equations in terms of initial phase difference $\phi = \frac{2\pi}{\lambda}(z - vt)$ and a change of phase difference $\delta\phi = 2\pi r/\lambda$. We have for the scalar amplitude of the sum of the waves,

$$E_s = A\cos(\phi + \delta\phi) + B\cos\phi \tag{45.4}$$

The goal is to see how the interference modulates the total intensity. One good approach is to divide up the scalar wave amplitude as follows,

$$E_s = \frac{1}{2}(A + B)[\cos(\phi + \delta\phi) + \cos\phi] + \frac{1}{2}(A - B)[\cos(\phi + \delta\phi) - \cos\phi] \tag{45.5}$$

Use the identities for the sum and difference of the cosines of two different angles and sort the parts out to obtain,

$$E_s = \left[(A+B)\cos\frac{\delta\phi}{2}\cos\left(\phi+\frac{\delta\phi}{2}\right)\right] - \left[(A-B)\sin\frac{\delta\phi}{2}\sin\left(\phi+\frac{\delta\phi}{2}\right)\right] \tag{45.6}$$

The quantity in each square bracket is in the usual form of an amplitude times a wave function. Clearly, there are two waves that have different amplitudes, both of which contain the change of phase difference. One wave is out of phase with the other by $\pi/2$. It is convenient and in this case valid to claim that the total intensity I_s will be the sum of the intensities of the individual waves, which is the sum of the squares of the amplitudes,

$$I_s = (A+B)^2\cos^2\frac{\delta\phi}{2} + (A-B)^2\sin^2\frac{\delta\phi}{2} \tag{45.7}$$

Expand the squared expressions and also use the identity $\sin^2 x = 1 - \cos^2 x$ to obtain after gathering up terms,

$$I_s = (A-B)^2 + 4AB\cos^2\frac{\delta\phi}{2} \tag{45.8}$$

Note that if the amplitudes of the interfering waves are equal, this result matches that given above for ideal interference. Although the result is already in useful form, it makes more sense if we use the identity $\cos^2 x = \frac{1}{2}(\cos 2x + 1)$ to obtain,

$$I_s = A^2 + B^2 + 2AB\cos\delta\phi \tag{45.9}$$

The first two terms are the intensities of the individual waves, so we can write,

$$I_s = I_A + I_B + 2\sqrt{I_A I_B}\cos\delta\phi \tag{45.10}$$

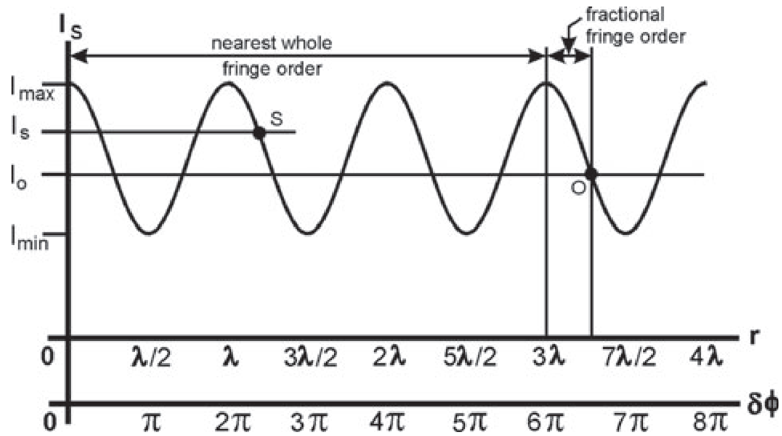
It is interesting to note that a plot of equation 45.8 containing $\cos^2\frac{\delta\phi}{2}$ is identical to a graph of equation 45.10, which contains $\cos\delta\phi$. This fact has surprised generations of students.

In practice, we usually do not care about the actual intensities of the interfering waves beyond making them as near equal as possible to maximize modulation or swing from dark to light. Now that we understand the relationship between observed intensity and the change of phase difference, we can convert equation 45.9 to an alternate useful form by noticing that the maximum intensity will be at $\delta\phi = 0$, which gives $I_{s\,\max} = A^2 + B^2 + 2AB$; and the minimum is where $\delta\phi = \pi$, which means $I_{s\,\min} = A^2 + B^2 - 2AB$. The result is,

$$I_s = \frac{I_{\max}+I_{\min}}{2} + \frac{I_{\max}-I_{\min}}{2}\cos\delta\phi \tag{45.11}$$

The observed intensity can be seen as the average of the intensities of the two waves plus half their difference times the cosine of the phase difference.

The figure below shows a graph of the observed intensity as a function both of phase difference and of PLD = r.



OBTAINING PATH LENGTH DATA FROM INTENSITY—GENERAL CASE

Refer to the figure above and suppose that we have measured the intensity I_o. The task is to determine the $\delta\phi$ corresponding to this intensity. The problems we met in the ideal case discussed above are compounded. There are three unknowns in either of Equations 45.10 or 45.11. An infinite number of possible cosine waves having the same periods pass through point A, especially if we could not set up the experiment so the starting intensity was at $I_{\max}$. As before, O might lie on any one of the cycles of the wave, and an infinity of others.

Part of the solution might be to somehow count the number of cycles between maximum and minimum intensity in order to determine which cycle is the correct one. After that, it seems that three measurements must be acquired, e.g. $I_{\max}$ and $I_{\min}$, in addition to I_o. Again, this approach is poorly conditioned for the reasons mentioned above, although it can be made to work.

Finally, recall that in a typical experiment, the **change** of retardation or phase difference must be determined. How can we account for the probability that the experiment did not start conveniently at the maximum intensity? Imagine, for example, that we cleverly measured the starting intensity, identified as I_s on the graph shown above. The goal is to obtain the change of $\delta\phi$ between the points S and O. Clearly, the initial $\delta\phi_s$ as well as the final $\delta\phi_o$ must be established in order to compute the change. So, the problem of obtaining PLD from intensity measurements must be faced twice, once for the initial state of the specimen and once for the final state.

To determine phase difference from intensity measurements in the non-ideal case, the following problems must be addressed:

- *Again, we do not know which cycle of the curve is correct.*
  - *Counting cycles and half-cycles of maximum and minimum between start and finish might solve this problem if the experiment allows such a procedure.*
- *There are three unknowns in the equation.*
  - *Measurements of maximum and minimum intensity in addition to the final intensity would allow a determination of the fraction of a cycle.*
- *The phase difference at the start of the experiment must also be measured because it is probably not known a priori.*
- *The procedure is poorly conditioned.*
  - *Small changes of intensity correspond to large changes of phase difference near the maxima and minima of the cosine relationship.*

SUMMARY OF THE PROBLEM AND A LOOK AHEAD

We conclude that a valid measurement of the change of PLD or phase difference between the interfering waves requires precise determinations of the initial and final values of phase difference. The initial value is then subtracted from the final value, a process that tends to magnify uncertainties. For each determination, a minimum of three intensity observations are required, and some method of counting the whole interference cycles between the initial and final states must be incorporated. Finally, the analysis method must minimize uncertainties such as those caused by the shape of the cosine function and by noise.

The problem of establishing phase difference from intensity measurements is, we hope, now well understood. What is the solution? Several well-conditioned approaches to perform the observations and calculate the required change of phase difference have been developed. Some of these, such as various compensation techniques used in photoelasticity before the age of computers, are quite old. Others, including the so-called phase-shifting or phase-stepping methods developed for digital speckle pattern interferometry, are relatively new. No matter the application and the implementation, all these approaches are fundamentally similar in concept. They involve changing the phase difference between the interfering waves by known amounts and observing the resultant intensities.

The next article in this series will examine a simple example to illustrate how to solve the problems outlined above with the goal of cementing the concepts. ■

In summary, to obtain a valid measurement of the change of phase difference:

- *The phase differences at the start and the end of the experiment must be established and the results subtracted.*
- *For each determination of phase difference, at least three intensity measurements are required.*
- *Some method of counting whole interference cycles through the course of the experiment must be incorporated.*
- *The analysis method should be well conditioned to minimize uncertainties.*

Several methods of measuring change of phase difference, some old, some new, have been developed.

- *The basic unifying concept is that they involve creating known phase changes between the optical paths and recording intensity at each change.*
- *A simple illustrative example will be the topic of the next article.*