NASA-CR-194758

FINAL
IN-35-CR
8 CIT
198529
209 P

FINAL TECHNICAL REPORT

NAG3-1300

Design of a Surface Deformation Measuring Instrument for the Surface Tension Driven Convection Experiment (STDCE-2)

for:

Alexander D. Pline
NASA Lewis Research Center
21000 Brookpark Road
Cleveland, OH 44135

by:

H. Philip Stahl, Ph.D.
Rose-Hulman Institute of Technology
5500 East Wabash Avenue
Terre Haute, IN 47803

23 December 1993

(NASA-CR-194758) DESIGN OF A
SURFACE DEFORMATION MEASURING
INSTRUMENT FOR THE SURFACE TENSION
DRIVEN CONVECTION EXPERIMENT
(STDCE-2) Final Technical Report
(Rose-Hulman Inst. of Tech.) 209 p

N94-23285

Unclas

G3/35 0198529

Design of a Surface Deformation Measuring Instrument for the Surface Tension Driven Convection Experiment (STDCE-2)

OUTLINE

1.0 Introduction

2.0 Objectives

3.0 Accomplishments

4.0 Student Involvement

5.0 Master's Thesis

6.0 Publications

7.0 Presentations

Appendix A: Publication Reprints

Appendix B: Presentation Viewgraphs

Design of a Surface Deformation Measuring Instrument for the Surface Tension Driven Convection Experiment (STDCE-2)

1.0 Introduction

This Final Technical Report covers the work accomplished (under NAG3-1300) from 1 October 1991 to 1 October 1993. The grant is a direct result of Dr. H. Philip Stahl's (of Rose-Hulman Institute of Technology) participation in the NASA/ASEE Summer Faculty Fellowship Program at NASA Lewis Research Center sponsored by Case Western Reserve University and the Ohio Aerospace Institute.

The Surface Tension Driven Convection Experiment (STDCE) is a fundamental fluid physics experiment designed to provide quantitative data on the thermocapillary flow of fluid under the influence of an increased localized surface temperature. STDCE flew on the Space Shuttle Columbia in the First United States Microgravity Laboratory (USML-1) in June 1992. The second flight of this experiment (STDCE-2) is scheduled for 1995. The specific science objectives of STDCE-2 are to determine the extent and nature of thermocpillary flows, the effect of heating mode and level, the effect of the liquid free-surface shape, and the onset conditions for and nature of oscillatory flows. In order to satisfy one of these objectives, an instrument for measuring the shape of an air/oil free-surface must be developed.

2.0 Objectives

The purpose of the first year of this project was to conduct a feasibility study to develop one or more non-contact (i.e. optical) techniques for measuring the surface deformation of an air-oil free-surface under the influence of a localized thermal load (i.e. CO2 laser beam or electric heater) which could be integrated into the anticipated reflight of the Surface Tension Driven Convection Experiment (STDCE-2). While initially these measurements were to be performed in a laboratory environment, the instrument had to be capable of operating in a Space Transport System (STS) environment for the United States Microgravity Laboratory flights USML-2.

The primary purpose of the second year of this project was to design, build, calibrate, and fully characterize the instrument(s) prototyped during the first year. A secondary purpose was to actively participate in the design of an engineering model and flight system for STDCE-2/USML-2.

Additionally, both years had the goal of developing algorithms and software to automatically analyze the output data provided by the surface deformation measuring instrument.

3.0 Accomplishments

A prototype instrument was developed, designed, built, calibrated, and fully characterized for measuring the surface deformation of an air-oil free-surface under the influence of a CO2 laser beam. The prototype instrument utilizes a modified two-channel Ronchi test technique where a collimated beam of light is reflected off of the oil free-surface and passed through a sinusoidal amplitude grating. The transmitted irradiance pattern provides quantifyable data about the surface's slope as a function of position. These results are described in great detail in the following publications:

> Meyers, William S., Design and Calibration of a Two Channel Ronchi System to Contour a Dynamic Fluid Surface, Master's Thesis, Rose-Hulman Institute of Technology, 1993.
>
> Meyers, William S., and H. Philip Stahl, "Sensitivity of Two-Channel Ronchi Test to Grating Misalignment," SPIE Vol. 1994, 1993.
>
> Meyers, William S., and H. Philip Stahl, "Contouring of a Free Oil Surface," SPIE Vol. 1755, 1992.
>
> Meyers, William S., Brent C. Bergner, Ronald D. White, David J. Huff, and H. Philip Stahl, "Contouring of a Free Oil Surface," SPIE Vol. 1779, 1992.

Algorithms to analyze the output slope data provided by the surface deformation measuring instrument were developed and reported in:

> Fischer, David J., John T. O'Bryan, Robert Lopes, and H. Philip Stahl, "A Vector Formulation for Interferogram Surface Fitting," Applied Optics, Vol.32, No.25, pp.4738, 1 September 1993
>
> Fischer, David J., and H. Philip Stahl, "Mechanism for Surface Fitting of Interferometric Slope Data," SPIE Vol. 2003, 1993.
>
> Fischer, David J., and H. Philip Stahl, "A Vector Formulation for Ronchi Shear Surface Fitting," SPIE Vol. 1755, 1992.
>
> Fischer, David J., John T. O'Bryan, and H. Philip Stahl, "A Vector Formulation for Interferogram Surface Fitting," SPIE Vol. 1779, 1992.

The developed algorithms were implemented in a software program called THIN which has been delivered to NASA Lewis Research Center (LeRC). This program is being extended by NASA LeRC to automatically analyze Ronchi patterns using image processing techniques.

A modular flight system was designed for the anticipated reflight of the Surface Tension Driven Convection Experiment (STDCE-2) on the United States Microgravity Laboratory (USML-2). It was designed to satisfy the following science requirements:

- observe the entire surface for three different sized chambers (12 mm, 20 mm, and 30 mm).
- measure slopes as small as 5 μm/mm.
- measure slopes as great as 30 μm/mm without vignetting.
- spatially sample the surface at least 10 times per diameter.
- reconstruct the surface shape along a given diameter.
- and, indicate when the oil is filled to its flat surface position.

The design of this instrument is published in:

> Stahl, H. Philip, "Conceptual Design of a Surface Deformation Measuring Instrument for the Surface Tension Driven Convection Experiment (STDCE-2)," SPIE Vol. 2005, 1993.

4.0 Student Involvement

Aside from the technical objectives, a major goal of this project was to provide a number of students with their first exposure to a 'real-world' research project. This was accomplished and the Principle Investigator is exceedingly proud of each student's achievements.

William S. Meyers developed and calibrated the prototype instrument for his Master's of Science in Applied Optics with assistance from Brent Bergner, Leif Sorensen, Ron White, and Dave Huff. Bill is now pursuing a Ph.D. in Bio-Physics at the University of Indiana/Purdue University in Indianapolis. He is performing research on optical tomographic imaging of the human body.

David J. Fischer developed and encoded the Ronchi pattern slope analysis algorithms with assistance from John O'Bryan and Dr. Robert Lopez of the Mathematics Department. Dave was the recipient of a full fellowship and is pursuing a Ph.D. in Optics at the University of Rochester

The software package THIN was developed through the efforts of: Kent Flint, Brad Freriks, Kurt Louis, Tony McAllister, Steven Reid, Brad Rodeffer, John Snider, Jason Snyder, Kevin Stultz, and Dr. Dale Oexmann of the Computer Science Department.

Kevin Stultz is pursuing an M.S. in Computer Science at the University of Alabama in Huntsville. He is performing research on image processing methods to analyze Ronchi patterns. His research is a direct extension of the work begun under this grant.

The participation of each student is summarized in the attached table.

Finally, two students participated in the NASA/ASEE Summer Internship Program at NASA LeRC sponsored by OAI and CWRU - Bill Meyers (Summer 92 and 93) and Kevin Stultz (Summer 93). And, Dave Fischer participated in the NASA/ASEE Summer Internship Program at NASA Langley in the Summer of 1992.

Student	Involvement	Major	Yrs.	Level	Post Grant Plans
Bill Meyers	Hardware Team Leader	Applied Optics	2 yrs.	Senior/Grad	Ph.D. in Bio-Physics
Brent Bergner	Team Member	Applied Optics	2 yrs.	Junior/Senior	Employment
Leif Sorensen	Team Member	Applied Optics	1 yr.	Senior	MS RHIT
Ron White	Team Member	Mech Eng	1 yr.	Junior	Employment
Dave Huff	Team Member	Mech Eng	1 yr.	Junior	Employment
Dave Fischer	Math Team Leader 2nd Yr.	Applied Optics	2 yrs	Junior/Senior	Ph.D. in Optics
John O'Bryan	Math Team Leader 1st Yr.	Math/Physics	1 yr.	Senior	Deceased
John Snider	Software Team Leader 1st Yr.	Computer Sci	1 yr.	Senior	Ph.D. in Comp Sci
Kent Flint	Team Member 1st Yr.	Computer Sci	1 yr.	Senior	Employment
Brad Freriks	Team Member 1st Yr.	Computer Sci	1 yr.	Senior	Employment
Kurt Louis	Team Member 1st Yr.	Computer Sci	1 yr.	Senior	Grad School
Steve Reid	Team Member 1st Yr.	Computer Sci	1 yr.	Senior	US Army
Tony McAllister	Software Team Leader 2nd Yr.	Computer Sci	2 yrs.	Junior/Senior	Employment GM
Brad Rodeffer	Team Member 2nd Yr.	Computer Sci	1 yr.	Senior	Employment Tex Int
Jason Snyder	Team Member 2nd Yr.	Computer Sci	1 yr.	Senior	Seminary
Kevin Stultz	Team Member 2nd Yr.	Applied Optics	2 yrs.	Junior/Senior	MS in Comp Sci

5.0 Master's Thesis

Meyers, William S., Design and Calibration of a Two Channel Ronchi System to Contour a Dynamic Fluid Surface, Master's Thesis, Rose-Hulman Institute of Technology, 1993.

6.0 Publications

Fischer, David J., John T. O'Bryan, Robert Lopes, and H. Philip Stahl, "A Vector Formulation for Interferogram Surface Fitting," Applied Optics, Vol.32, No.25, pp.4738, 1 September 1993

Fischer, David J., and H. Philip Stahl, "Mechanism for Surface Fitting of Interferometric Slope Data," SPIE Vol. 2003, 1993.

Fischer, David J., and H. Philip Stahl, "A Vector Formulation for Ronchi Shear Surface Fitting," SPIE Vol. 1755, 1992.

Fischer, David J., John T. O'Bryan, and H. Philip Stahl, "A Vector Formulation for Interferogram Surface Fitting," SPIE Vol. 1779, 1992.

Meyers, William S., and H. Philip Stahl, "Sensitivity of Two-Channel Ronchi Test to Grating Misalignment," SPIE Vol. 1994, 1993.

Meyers, William S., and H. Philip Stahl, "Contouring of a Free Oil Surface," SPIE Vol. 1755, 1992.

Meyers, William S., Brent C. Bergner, Ronald D. White, David J. Huff, and H. Philip Stahl, "Contouring of a Free Oil Surface," SPIE Vol. 1779, 1992.

Stahl, H. Philip, "Conceptual Design of a Surface Deformation Measuring Instrument for the Surface Tension Driven Convection Experiment (STDCE-2)," SPIE Vol. 2005, 1993.

7.0 Presentations

Stahl, H. Philip, "Measuring Deformations in Space," University of Arizona Optical Sciences Center Colloquium, Tucson, AX 1993.

APPENDIX A

Publication Reprints

Conceptual Design of a Surface Deformation Measuring Instrument for the Surface Tension Driven Convection Experiment (STDCE-2)

H. Philip Stahl
Stahl Optical Systems
5 Hyvue Dr.
Newtown, CT 06470

ABSTRACT

The planned Space Shuttle Experiment STDCE-2 (Surface Tension Driven Convection Experiment) requires a non-contact Surface Deformation measuring instrument to monitor the shape of a dynamic fluid surface. This paper presents a conceptual design for this instrument which best satisfies the various science, engineering, and managerial constraints.

1.0 INTRODUCTION

The ability to grow semi-conductor crystals, biological crystals, or to solidify metal alloys (on earth or in space) is limited by convective fluid flow which introduces imperfections into the lattice structure. Convection arises from two sources: gravity and surface tension. Gravity driven convection (natural convection or buoyancy) occurs when there is a temperature difference within a fluid volume. Colder, heavier fluid is pulled down causing warmer, lighter fluid to rise. Surface-tension driven convection (thermocapillary flow) occurs when there is a temperature difference across the surface of a fluid. Because a warmer fluid has a lower surface tension, it is pulled towards the colder surface liquid (which has a higher surface tension). As the warmer fluid moves across the surface, it draws liquid up to the surface, producing a flow.

On Earth, buoyancy driven convection dominates crystal growth and alloy solidification. Also, gravity causes lattice structures to 'sag' and limits the size of crystals which can form in suspension. But in microgravity, buoyancy flow is reduced to 10^{-6} and crystals of any size remain in suspension. Thus, material processing in space offers great promise. However, thermocapillary flow still exists and must be understood.

2.0 SURFACE TENSION DRIVEN CONVECTION EXPERIMENT

The Surface Tension Driven Convection Experiment (STDCE) is a fundamental fluid physics experiment designed to provide quantitative data on the thermocapillary flow of fluid under the influence of an increased localized surface temperature [1]. STDCE flew on the Space Shuttle Columbia in the First United States Microgravity Laboratory (USML-1) in June 1992. The second flight of this experiment (STDCE-2) is scheduled for the Fall of 1995.

The objective of STDCE was to determine the extent and nature of steady-state thermocapillary flow as a function of heating mode and level, and liquid free-surface shape. It demonstrated the existence of microgravity steady-state thermocapillary flow in a 10 cm diameter chamber for both flat and curved surface shapes when heated by either an electric heater or a CO_2 laser beam (Figure 1).

The objective of STDCE-2 is to determine the onset conditions for and nature of oscillatory thermocapillary flow as a function of heating mode and level, liquid free-surface shape, and container size. For STDCE-2 the 10 cm chamber is replaced by three chambers of 3.0, 2.0, and 1.2 cm in diameter.

To fully accomplish the science objectives, three parameters must be monitored: surface temperature distribution, bulk fluid flow, and surface deformation. Both STDCE and STDCE-2 monitor the surface temperature distribution with a HgCdTe thermal imager. And, bulk fluid flow is monitored by observing the motion of

SPIE Vol. 2005, 1993

tracer particles (which have been uniformly mixed into the fluid) as they move through a sheet or beam of light. But, while STDCE had no means for measuring surface deformation, STDCE-2 will (Figure 2).

A deformation measuring capability has been added to STDCE-2 because current research indicates that both steady-state and oscillatory flow states produce deformations of the air/oil free surface. Additionally, it is hypothesized that there is a coupling between oscillatory flow type and deformation shape, velocity flow, and surface temperature field. Depending upon the experimental conditions, steady state flow produces two characteristic steady state surface deformations: a 'Pimple' (bump), and a 'Dimple' (hole). Oscillatory flow occurs with either pimples or dimples and produces both standing wave and rotary surface deformations.

3.0 EVALUATION OF ALTERNATIVES

The purpose of this project is to design, build, calibrate, and fully characterize one or more non-contact (i.e. optical) prototype instruments for measuring the surface deformation of an air-oil free-surface under the influence of a localized thermal load (i.e. CO2 laser beam or electric heater) which can be integrated into the anticipated re-flight of the Surface Tension Driven Convection Experiment (STDCE-2). While initially these measurements will be performed in a laboratory environment, the instrument must be capable of operating in a Space Transport System (STS) environment for the United States Microgravity Laboratory flights USML-2.

Because the surface is a fluid, it cannot be contoured mechanically. It must be contoured optically. There are two ways to measure a two-dimensional reflective surface: interferometrically or geometrically. For most optical applications, interferometers are generally preferred to geometrical tests. They are more accurate and sensitive. And, because they measure a surface's shape while most geometrical tests measure a surface's slope, they are easier to interpret. Unfortunately, because the anticipated surface deformations are on the order of 100 μm, conventional visible interferometry is not practical. The fringes density will be too great for the camera to resolve. One solution to this problem is to use an infrared interferometer, but such a technique might interfere with the localized heating process. Therefore, a geometrical technique is required.

Because geometrical techniques measure the surface's slope instead of its height, they have specific advantages and disadvantages. They are insensitive to tilting of the free-surface cause by G-Jitter. But, they require two orthogonal data sets in order to reconstruct its shape. These two data sets can be either x-slope and y-slope or r-slope and θ-slope. Whether two data sets are truly required depends upon whether the surface is rotationally symmetric. If the surface is rotationally symmetric, then it may be possible to satisfy the mission's science requirements with just x-slope or r-slope. But, if the surface is non-rotationally symmetric, then it is necessary to acquire two orthogonal data sets. Unfortunately, the surface deformations observed in the laboratory are not rotationally symmetric. But, a unique feature of these deformations which may eventually lead to a simple solution is that their outer edge is at a constant height.

Many classical geometric test techniques (Foucault knife-edge test, wire test, Hartmann screen test, Ronchi test, etc.) were considered. Of these, three which provide data over the entire aperture were selected for further evaluation: Ronchi grating, Hartmann test, and Projected grid.

In the projected grid approach, a linear or crossed grid pattern is placed directly in front of an extended light source (illuminated diffuser screen). This pattern is then imaged onto the reflective sample surface. The reflected wavefront is re-imaged by another optical system onto the video camera. The projected grid approach was eliminated because it lacked sensitivity, dynamic range, and flexible spatial sampling.

In the single-channel Ronchi grating test, a collimated beam of light is reflected off of the sample surface (Figure 3). The reflected wavefront is focused through a grating to form a Ronchigram which is imaged onto a video camera. The Ronchi approach has many advantages. By selection of the grating spatial frequency, it has variable sensitivity and dynamic range. By selection of the grating location, it has adjustable spatial sampling. Unfortunately it does not provide simultaneous orthogonal slope data sets.

The Hartmann screen test can provide both x- and y-slope information simultaneously. But, it has limited spatial resolution. And, for this application, it is not practical because the screen would obscure the upper surface flow visualization function. An alternative is to use crossed gratings in a Ronchi test. However, the complex patterns produced are difficult to analyze.

Another way to obtain simultaneous x- and y-slope data is with a two channel Ronchi approach (Figure 4). With the two-channel Ronchi approach, a collimated beam of light (laser or white) is projected onto the reflective sample surface. The reflected beam is collected, split into separate x- and y-slope measurement channels, passed through vertical and horizontal gratings, and projected onto an observation screen for viewing by a video camera. A two channel Ronchi instrument was demonstrated which simultaneously acquires x- and y-slope data for an input wavefront [2]. Unfortunately, because of packaging, cost, and schedule issues, a two channel system could not be implemented.

Based upon managerial issues (packaging, cost, and schedule) and science issues (sensitivity, dynamic range, and spatial sampling) a single channel Ronchi system was selected. The managerial constraints and science requirements were then iterated, resulting in Table 1.

Table 1: STDCE-2 Surface Deformation Measuring Instrument Science Requirements

Observe the entire surface for three different sized chambers: 12 mm, 20 mm, and 30 mm.

Measure slope as small as 5 μm/mm.

Measure slope as great as 30 μm/mm without vignetting.

Spatially sample the surface at least 10 times per diameter.

Temporally sample the surface at least 30 times per second.

Reconstruct the surface shape at least along a diameter.

Indicate when the oil is filled to its flat surface position.

4.0 THE RONCHI TEST

The Ronchi test is performed by observing the shadow pattern produced when a focused beam of light is passed through a periodic structure, such as a line grating, which is located either at focus or away from focus a distance z (Figure 5). If the beam is ideal, the shadow pattern looks like the grating. But, if the beam is deformed, then the pattern is deformed. A Ronchigram is obtained when the shadow pattern is recorded at an image of the surface under test, such that it is superimposed ('painted') upon the image. Information about the surface is obtained by correlating shadow lines with physical locations on the surface. The distance between adjacent shadows indicates how much the surface slope changes between those two locations. The amount of slope required to go from one shadow to its neighbor is called the equivalent wavelength. Alternatively, the Ronchi test can be though of as a sheared interferogram produced by overlapping diffraction orders.

The performance of the Ronchi system with an amplitude sine grating is completely specified by its equivalent wavelength:

$$\lambda_{eq} = \frac{d}{2(F/\#)}$$

Since the F/# of the measurement beam is fixed by the optical system, everything depends upon grating spacing, d. This dependance is extremely important. It allows for the measurement sensitivity to be varied on orbit by inserting into the beam gratings with different spatial frequencies (Figure 6). If the surface deformation is large, insert a coarse grating. If it is small, use a fine grating.

Accuracy defines the uncertainty of a given measurement. For the current sampling system, the measurement accuracy is approximately $\lambda_{eq}/8$. A measurement system can report numbers to an arbitrary precision, but they may not be accurate or repeatable. The minimum measurable value of a given system is defined to be twice its measurement accuracy. For the current sampling system the minimum measurable slope is approximately $\lambda_{eq}/4$. The maximum amount of full chamber slope which the system can measure is determined by how many shadow lines (fringes) the video camera can clearly resolve in the Ronchigram. For the current sampling system, the Ronchigram will be limited to approximately 16 lines pairs for a maximum slope of 16 λ_{eq}.

Table 2. Performance Summary for F/2.25 System

Grating Frequency [lp/mm]	Grating Line Space [mm/lp]	λ_{eq} [μm/mm]	Accuracy [μm/mm]	Minimum Slope [μm/mm]	Maximum Slope [μm/mm]
0.5	2.0	444.44	55.56	111.11	7111
1.0	1.0	222.22	27.78	55.56	3556
2.0	0.5	111.11	13.89	27.78	1778
2.5	0.4	88.89	11.11	22.22	1422
5.0	0.2	44.44	5.56	11.11	711
10.0	0.1	22.22	2.78	5.56	356

5.0 FLIGHT SYSTEM CONCEPTUAL DESIGN

The flight free-surface deformation measuring instrument was designed in six modules (Figure 7): light source, projection/imaging, polarization, pupil relay, measurement, and camera. Each module overcomes various engineering and/or packaging challenges while allowing the total system to meet the science requirements. Each module is designed to be independently assembled and bench aligned before integration into the final system. And, each module is designed to prevent unnecessary exposure of the crew to light radiation.

5.1 The Oil Surface

The object under test is a free surface of 2 cSt silicon oil. It has a refractive index of 1.39. And, an irradiance reflection coefficient of 2.7%. Because of this low reflection coefficient, the source must have at least 10 mW of power for the video camera to observe the Ronchi pattern. Also, ghost reflections from the optical components can obscure the oil surface reflection.

Because the free-oil surface is a specular reflector, it can be treated as a plane mirror. To implement the Ronchi test, a collimated illumination beam is required. Two illumination geometries were considered: Off-axis, and on-axis. The on-axis geometry was selected for several reasons. First, it gets the lens closer to the fluid surface, thus it can accept higher slope errors. Second, it has no off-axis projection distortion or cosine scaling. And third, it is easy to package. The on-axis geometry requires the beam to travel in two directions through the projection/imaging module. This can cause ghost reflection problems.

5.2 Light Source Module

The source module provides a collimated beam of light which uniformly illuminates the object under test. This is accomplished with a laser diode and a Galilean beam expanding telescope (Figure 8). A laser diode was selected because it is a very-bright source capable of surviving the launch. (White-light and LED sources were also considered.) Coherence is not a factor in selecting a light source. A Galilean beam expander was selected because it is compact and does not have a spatial filter which could become misaligned during launch. The beam diameter is defined by the projection/imaging module's entrance pupil to be 22 mm in diameter. The beam is magnified (30X) and truncated to balance uniformity and power.

5.3 Projection/Imaging Module

The projection/imaging module has two functions: illumination and imaging. Its illumination function magnifies or de-magnifies the light beam to illuminate the oil surface for each of the three chamber sizes (12 mm, 20 mm, 30 mm). Its imaging function has two requirements: to form an image of the different sized surfaces under test at a fixed sized pupil location, and to pass a reflected wavefront with at least 30 μm/mm of slope without vignetting. The illumination and imaging functions determine the magnification properties of this module. The vignetting requirement determines its *F/#*. Packaging issues determines its clear aperture.

The illumination function is accomplished with two AFOCAL lens pairs (Figure 9), an outer pair and an inner pair. The outer pair forms a Keplerian telescope with unit magnification. The inner pair forms a Galilean telescope which is positioned about the outer pair's internal focus and functions as a pseudo field lens. This lens provides magnification/de-magnification when it is flipped. To illuminate the 20 mm chamber, remove the flip lens such that the 22 mm source beam is relayed unchanged. To illuminate the larger or smaller chambers, insert the flip lens either forward or backward to magnify or de-magnify the beam (Table 3). The final illumination beam size depends upon the flip lens magnification factor. The flip lens is not symmetric about the focal point and does not rotate about the focal point.

Table 3 Image Magnification Module Fields of View for Various Flip Lens Magnification Factors

Chamber	Diameter	1.4 X	1.5 X	1.6 X	1.7 X
Smallest	12 mm	15.7 mm	14.7 mm	13.75 mm	12.9 mm
Middle	20 mm	22 mm	22 mm	22 mm	22 mm
Largest	30 mm	30.8 mm	33 mm	35.2 mm	37.4 mm

Imaging of the different sized chambers into a fixed size pupil is accomplished by running the beam backwards through the system. The Keplerian telescope, without the flip lens, relays an image of the surface in the middle sized chamber. Inserting the flip lens, the system relays images of the larger or smaller sized chambers. For proper imaging the surface must be in the front focal plane of the lens closest to the oil. The resultant image will be in the back focal plane of the last lens of the projection/imaging module. This image is transferred to the camera module by the pupil relay and measurement modules - forming a Ronchigram.

The projection/imaging module (as well as the pupil relay and measurement modules) must pass a maximum slope of 30 μm/mm without vignetting. This places an F/# requirement on these components. The faster their F/# the more slope they can pass. Initially, the optical system was designed to fully collect a wavefront reflected from a 250 μm high, 10 mm diameter spherical deformation. However, this was relaxed because of packaging and schedule issues. Packaging factors limit the lens's maximum diameter and minimum focal length. Schedule require them to be commercially available.

The selected components are F/2.25 (90 mm focal length, 40 mm diameter) achromatic doublets with approximately 2 waves of spherical aberration at 633 nm. This amount of aberration is acceptable given the magnitude of the anticipated oil surface deformations. With proper calibration, this error can be removed from the measurement. Table 4 gives the maximum surface feature and slope which can be image with the current system. Please note that the 30 μm/mm specification is not satisfied for the 30 mm diameter chamber.

Table 4 Maximum Observable Feature and Slope without Vignetting for each Sample Chamber as a function of Lens Clear Aperture

40 mm Clear Aperture

Chamber Diameter	δ_{max}	Ω_{max}
12 mm	208 μm	62.4 μm/mm
20 mm	277 μm	41.6 μm/mm
30 mm	208 μm	20.8 μm/mm

36 mm Clear Aperture

Chamber Diameter	δ_{max}	Ω_{max}
12 mm	181 μm	54.3 μm/mm
20 mm	222 μm	33.3 μm/mm
30 mm	125 μm	12.5 μm/mm

5.4 Polarization Module

The polarization module eliminates ghost reflections and efficiently uses the source's available optical power. Since the projection/imaging module is used in both directions and because the oil's reflection coefficient is small, ghost reflections from the illuminated optical surfaces can seriously obscure the oil reflection. To eliminate this problem, the polarization of the beam is manipulated such that the ghost reflections are vertically polarized and the oil reflection is horizontally polarized. This is accomplished with two polarizers and a quarter wave plate (Figure 10). The first polarizer defines the polarization of the ghost reflections. The quarter wave plate rotates the oil reflection polarization by 90 degrees. And, the second polarizer blocks all of the ghost reflections and passes the oil reflection. To maximize the available power, a polarization beam splitter is used. All of the source's vertical light is transmitted into the projection module. And, all of the returning horizontal light is reflected into the relay module. A conventional 50/50 beam splitter could be used, but it would throw away 75% of the available light.

5.5 Pupil Relay Module

The pupil relay module is a unit magnification AFOCAL system. Its primary function is to extend the beam path such that the measurement module is at a location where the mission specialist can insert the gratings. Additionally, it serves as an alternative measurement module allowing for gratings to be inserted into parts of the beam not accessible with the measurement module. This is required to properly measure hole deformations.

5.6 Measurement Module

The measurement module allows gratings of different line spacings to be inserted into different locations of the focused oil reflection forming a Ronchigram. Since the lens F/#'s are known, changing the grating line spacing changes the equivalent wavelength. The equivalent wavelength depends only upon the grating line spacing and is independent of where the grating is placed in the beam. Grating placement in the beam determines the number of spatial sample points across the beam. To insure at least 10 sample points per diameter, a coarse grating must be placed at a greater distance from the measurement module focus than a fine grating. The distance from focus for a given grating depends entirely upon its line spacing:

$$z = 10\, d\, (F/\#) = 22.5\, d$$

Table 5 summarizes the distances needed for each grating to produce 10 lines.

Table 5. Distance from Focus for Spatial Sample Points

Grating Frequency [lp/mm]	Grating Line Space [mm/lp]	Z for 10 Samples per Diameter	Z for 15 Samples per Diameter	Z for 20 Samples per Diameter
0.5	2.0	45.0 mm	67.5 mm	90.0 mm
1.0	1.0	22.5 mm	33.75 mm	45.0 mm
2.0	0.5	11.25 mm	16.88 mm	22.5 mm
2.5	0.4	9.0 mm	13.5 mm	18.0 mm
5.0	0.2	4.5 mm	6.75 mm	9.0 mm
10.0	0.1	2.25 mm	3.375 mm	2.5 mm

To reconstruct a two-dimensional surface contour, separate horizontal and vertical gratings, or crossed gratings can be inserted into the beam. Radial or circular zone gratings may be tried.

5.6.1 Fill Level Indication

Silicon oil has a very low surface tension and spreads easily on any surface. To prevent it from flowing out of the container, the rim of the chamber has a sharp edge that 'pins' the oil in place and is 'barrier' coated. As the container is filled in low gravity, the liquid will form a deep spherical curvature. When completely filled, the surface will be flat. If over-filled, the surface will form a spherical dome. By placing a grating at the focal point of the measurement module, direct visual evidence of the oil's surface state is provided to the Mission Specialist. When the oil surface is not flat, there is a line pattern. As the oil approaches a flat surface, the pattern spreads out (has fewer lines) until all the lines disappear when the surface is flat. If the container is under or overfilled, the lines will reappear and increase in number.

5.7 Camera Module

The Ronchigram produced by the measurement module is imaged onto a diffuser plate where a real image is formed. This image is viewed by a video camera and recorded on video tape for subsequent (post-flight) data analysis [3]. The physical distance from the diffuser screen to the camera is determined by the focal length of the camera lens, the camera sensor format, and the size of the image on the diffuser screen:

$$d = \left[1 + \frac{h_{image}}{r_{sensor}}\right] f_{camera\ lens}$$

For the flight system, the diffuser image radius is 10 mm, the sensor radius is 2.2 mm, and the camera lens focal length is 28 mm. Thus, the stand-off distance is 155 mm. The camera is a 1/2 inch format RS-170 video camera.

In addition to being able to spatially resolve the surface feature without vignetting, the system is required to temporally resolve the oscillation. It is assumed that the period of this oscillation is on the order of 2 to 5 seconds. Thus, a standard video camera with a 60 Hz field rate is adequate. Finally, it is assumed that there may be high-speed temporal gitter in addition to the lower-speed temporal oscillation. This gitter could reduce the contrast of the pattern. To eliminate gitter the camera will need a high speed shutter and the source will need sufficient output power to expose the camera.

6.0 CALIBRATION

If the optical system were aberration free, the Ronchigram of an ideal flat mirror would be a series of perfect straight lines. However, given the total number of positive optical components in the optical system, such a pattern is impossible. By measuring the beam reflected from a 'perfect' reference-flat insert, after assembly, these errors can be characterized and removed from all measured data. Additionally, the system can be calibrated by testing a series of known diamond-turned test plate samples.

Calibration of the prototype instrument was accomplished by measuring a known amount of defocus aberration introduced by translating the collimating lens:

$$W_{020} = -\frac{\epsilon_z}{8\ (F/\#)^2}$$

$$where:\ \epsilon_z = Longitudinal\ Aberration$$

The result of this calibration gives a measurement accuracy of approximately ± 10%. Measurements made on the known test samples confirm this result.

The system is designed for the insertion of interchangeable gratings at various locations in the beam. Thus, grating alignment errors can be a problem. To assess this, the data analysis process has been evaluated for its sensitivity to errors in the position of one grating relative to its orthogonal partner. In general, the Ronchi test is relatively insensitive to small grating placement errors (except for translation along the surface normal[4]) for grating frequencies less than 3 to 5 lp/mm.

7.0 CONCLUSIONS

A prototype Ronchi instrument to monitor steady-state and oscillatory deformations, and fluid fill level of a free-surface of oil has been developed, demonstrated, and characterized for STDCE-2. A single-channel Ronchi configuration has been selected as the flight instrument. The flight instrument is currently being designed.

The Ronchi test is a shearing interferometer. The number of sheared wavefronts and the magnitude of the shear is dependant upon the characteristics of the grating. The fringe spacing (equivalent wavelength) in a Ronchigram is proportional to the spacing of the grating and the F/# of the test beam. The dynamic range and sensitivity of the Ronchi test can be changed by simply changing the grating spacing. The prototype two channel Ronchi system has an accuracy of ± 5%. For a shearing system, the maximum measurable error is determined by the surface curvature.

8.0 ACKNOWLEDGEMENTS

This project was funded by a NASA research grant (NAG3-1300) from the NASA Lewis Research Center under the direction of Alexander Pline, Chief Scientist for STDCE-2, to Rose-Hulman Institute of Technology where Dr. Stahl was an Assistant Professor of Physics and Applied Optics. William S. Meyers developed and calibrated the prototype instrument for his Master's of Science in Applied Optics with assistance from Brent Bergner, Leif Sorensen, and Ron White. David J. Fischer developed and encoded the Ronchi pattern slope analysis algorithms with assistance from John O'Bryan and Dr. Robert Lopez. The software package has been developed through the efforts of: Kent Flint, Brad Freriks, Kurt Louis, Tony McAllister, Steven Reid, Brad Rodeffer, John Snider, Jason Snyder, Kevin Stultz, and Dr. Dale Oexmann. Finally, Patricia Martin of Analex and David Heipel of ADF have been instrumental in converting this conceptual design into flight hardware.

9.0 BIBLIOGRAPHY

1. Surface Tension Driven Convection Experiment, NASA, 1992.
2. Meyers, William S., and H. Philip Stahl, "Contouring of a free oil surface," SPIE Vol.1755, 1992.
3. Fischer, David J., and H. Philip Stahl, "A vector formulation for Ronchi shear surface fitting," SPIE Vol.1755, 1992.
4. Meyers, William S., and H. Philip Stahl, "Sensitivity of two-channel Ronchi test to grating misalignment," SPIE Vol.1994, 1993.

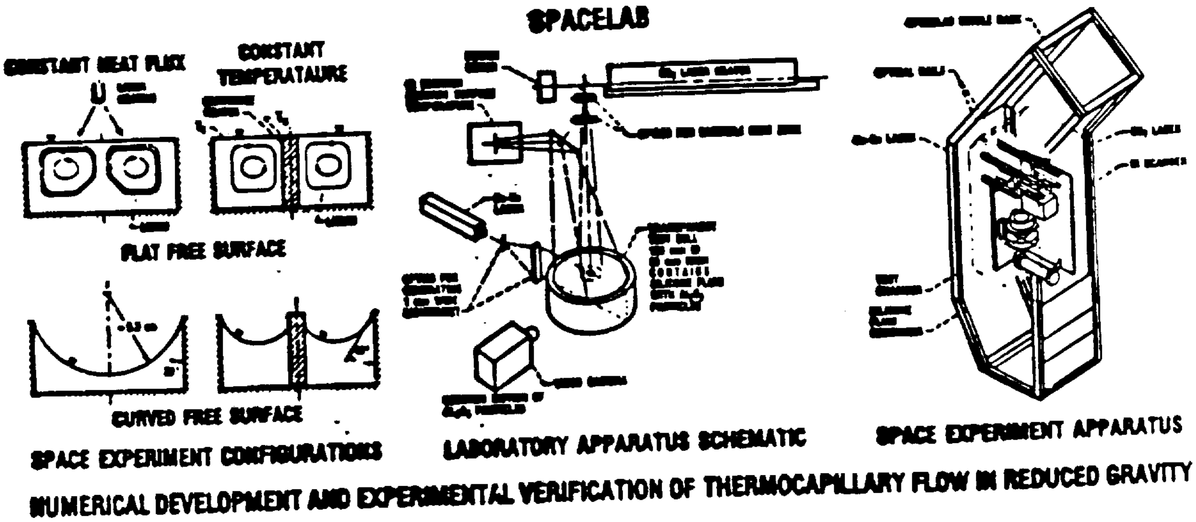


Figure 1. Surface Tension Driven Convection Experiment.

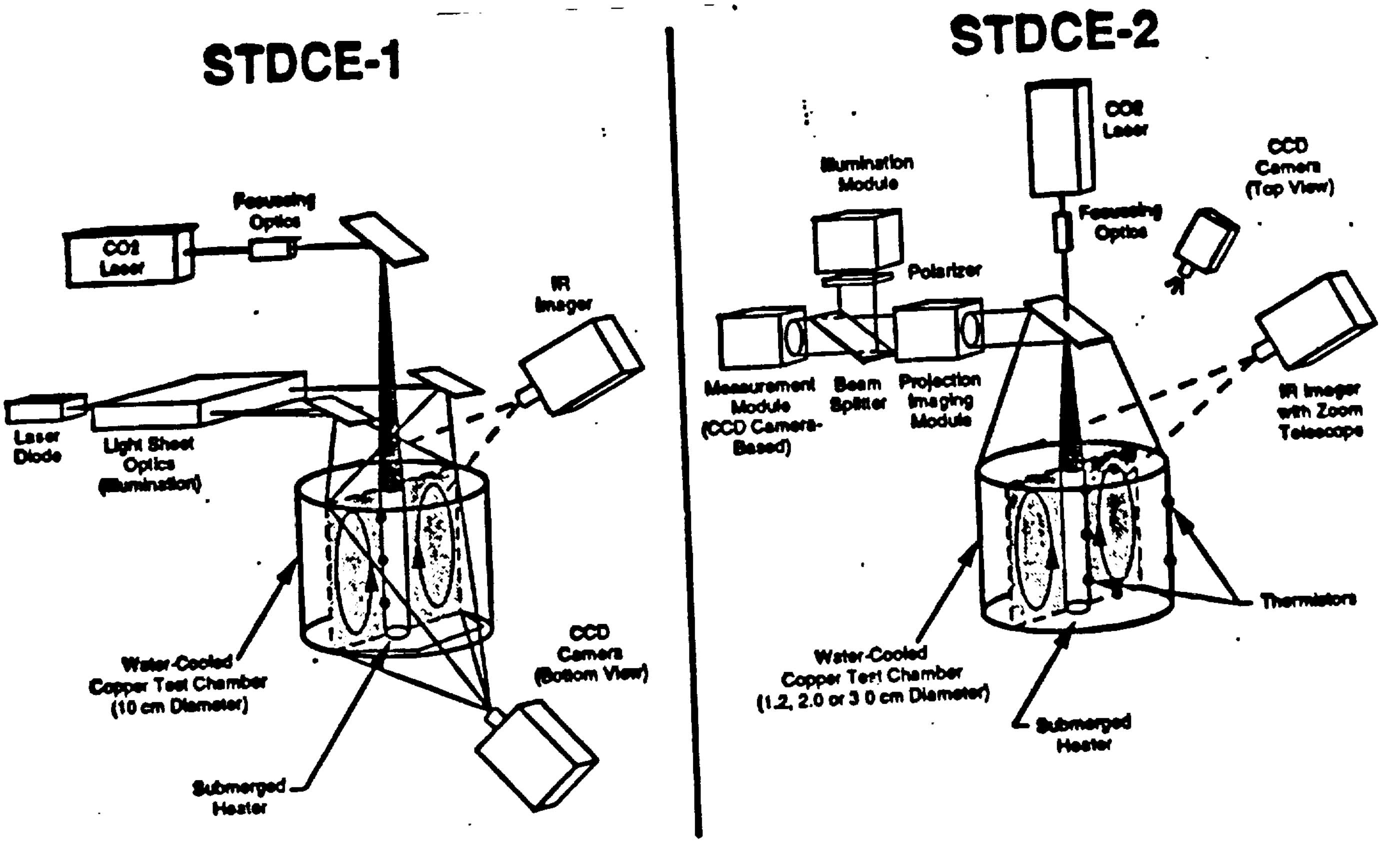


Figure 2. Comparison of STDCE-1 and STDCE-2.

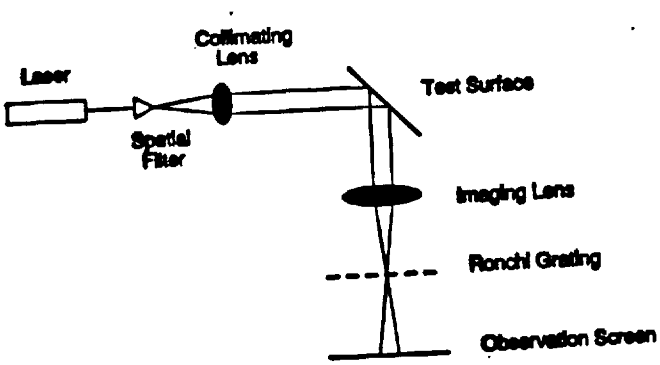


Figure 3. Single Channel Ronchi Test.

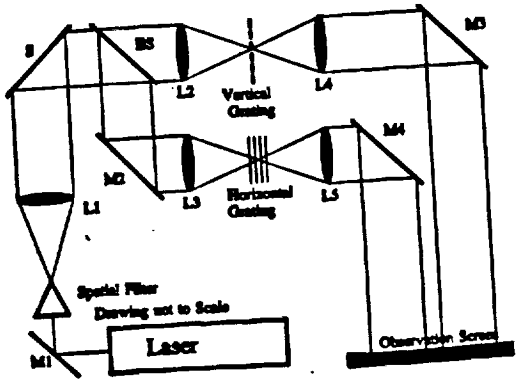


Figure 4. Two Channel Ronchi Test.

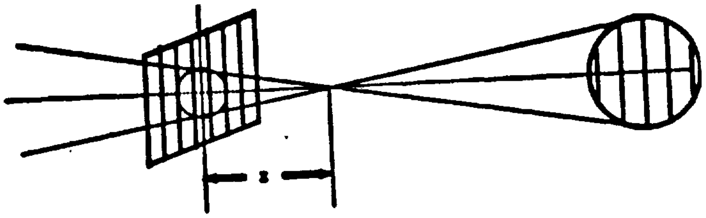
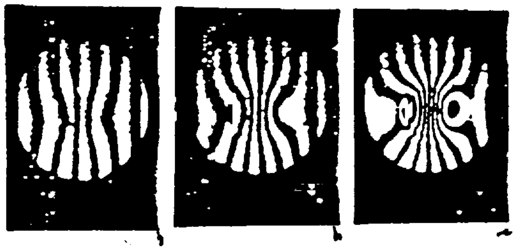


Figure 5. The Ronchi Test.



0.5 cycles/mm 1.0 cycles/mm 2.0 cycles/mm

Figure 6. Measurement of a 25 micrometer bump with three different gratings at a fixed distance z from focus producing three different equivalent wavelengths.

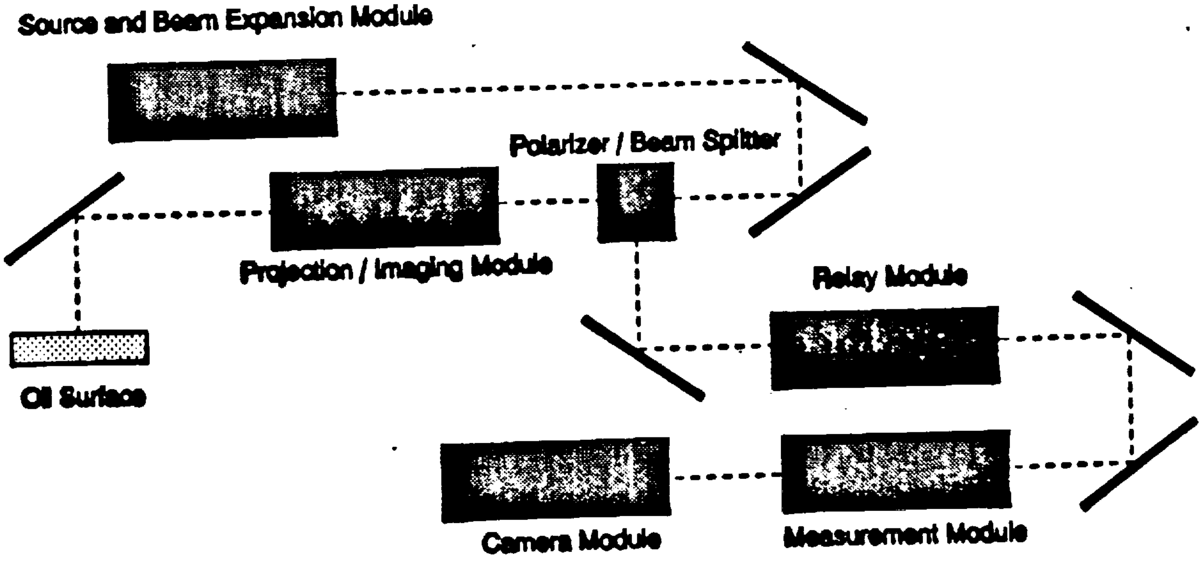


Figure 7. Schematic Diagram of the Flight System

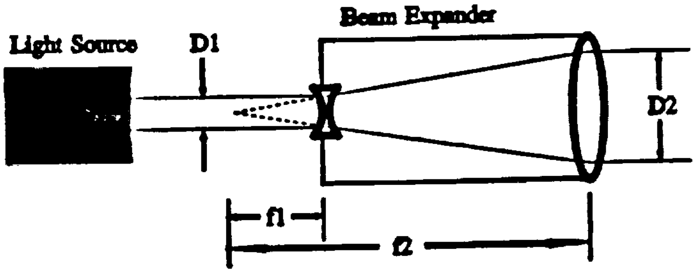


Figure 8. Schematic Diagram of the Illumination Module.

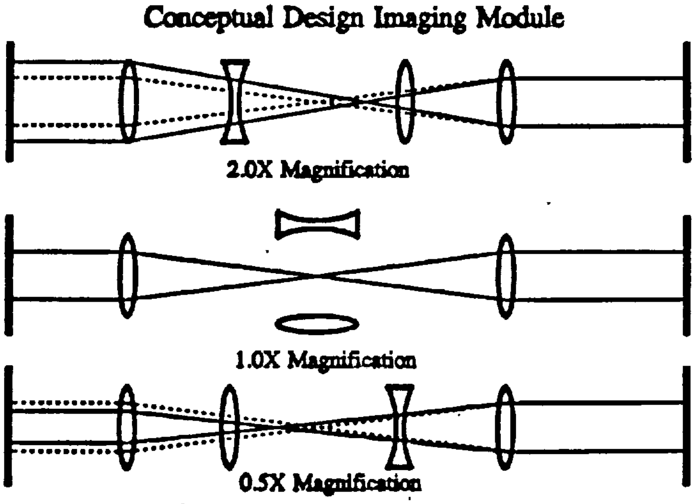


Figure 9. Schematic Diagram of the Projection/Imaging Module Magnification Optics.

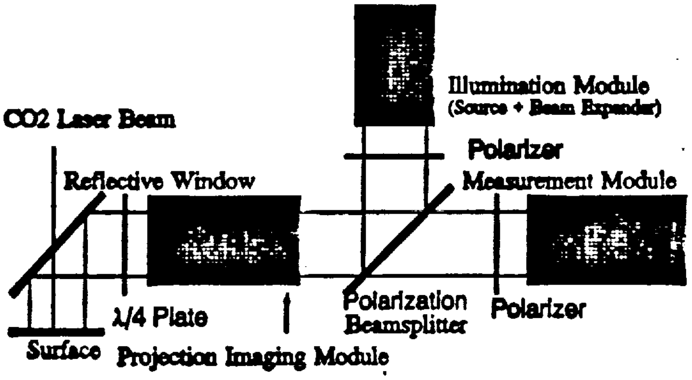


Figure 10. Schematic Diagram of the Polarization Module Component Placement.

Contouring of a free oil surface

William S. Meyers and H. Philip Stahl

Rose-Hulman Institute of Technology, Department of Physics and Applied Optics
5500 Wabash Avenue, Terre Haute, Indiana 47803

ABSTRACT

A non-contact (i.e. optical) technique is required to measure the deformation of a free oil surface under the influence of a localized thermal load. This deformation is caused by surface tension driven thermal convective flow inside the fluid, and can be as large as 250 micrometers. Therefore, conventional interferometry is not possible. Instead, a Ronchi technique is proposed for contour mapping the oil surface. This paper presents a two channel Ronchi instrument, and some preliminary results from the analysis of two aspheric surfaces, a diamond machined metal surface and the free oil surface.

1. INTRODUCTION

Scientists at Case Western Reserve University and NASA Lewis Research Center are conducting research on a variety of fundamental phenomena inherent in low gravity processes. One phenomenon being studied is thermocapillary flow, or fluid flow driven by surface tension. Thermocapillary flows result when temperature differences arise across a liquid/gas interface. These differences cause a change in the surface tension along the interface, which results in thermocapillary flow. Improvements in the understanding of thermocapillary flow may lead to technological breakthroughs in the space processing of semiconductor crystals. These experiments were a part of the 13-day Microgravity Laboratory (USML-1) shuttle mission in the summer of 1992. The experiment described in this paper was not on that mission, but may be included on future shuttle missions.

One of the parameters of thermocapillary flow to be studied is the ability of the free surface to deform during flow. The development and implementation of a non-contact (i.e. optical) method to monitor these surface deformations as a function of experimental parameters such as fluid properties, heating modes, is crucial to the understanding of the physics of thermocapillary flow.

The experimental surface is a three centimeter diameter pool of silicon oil. Deformations as large as 250 micrometers are produced on the surface of the oil by thermocapillary flow. An optical technique instead of a mechanical technique was chosen to contour the free surface because touching the surface with a mechanical probe would deform the surface even more. There are two basic categories of optical techniques: geometric and interferometric. A geometric method measures surface slope, while an interferometric method measures surface height. The geometric approach was chosen over the interferometric technique for two reasons. First, conventional visible wavelength interferometry would produce a fringe pattern too dense to analyze because of the anticipated deformation range of up to 250 micrometers. And, while infrared interferometric techniques could produce a resolvable fringe pattern, the infrared beam might affect the system's thermal loading. Second, geometric techniques are dependant on the slope of the surface and are therefore insensitive to any jitter or tilt experienced by the oil surface while in space. One disadvantage of using geometric techniques is that two orthogonal data sets, either x-slope and y-slope or r-slope and Θ-slope are needed to completely reconstruct the surface shape.

Five classical geometric techniques were considered: Foucault knife-edge test, Hartmann screen test, wire tests, grating interferometry, and Ronchi test. The knife-edge and wire tests were not chosen because they provide data for only a single zone or diameter, not for the entire aperture. The Hartmann screen can provide two orthogonal data sets at one time, but it has a limited spatial resolution when compared with the Ronchi test. The Ronchi test was chosen as the technique to use for three reasons: the small dimensions of the experimental pool are tailored to the Ronchi requirement of a collimated beam, the Ronchi data is less complicated to analyze, and it can be integrated with phase-measuring techniques.

SPIE Vol. 1755, 1992

2. BACKGROUND

2.1 Relevant work

The Ronchi test was discovered by Italian physicist Vasco Ronchi in 1921. Ronchi was attempting to measure the radius of curvature of a concave spherical mirror. Upon looking through the grating at the mirror, the surface defects of the mirror became apparent. The mirror surface appeared lined with twisted and irregular fringes.[1] Since its conception, the Ronchi test has been applied to the measurement of the aberrations or defects in many different applications and using many different test setups.

A single channel Ronchi test is shown in Fig. 1. A collimated wavefront is incident upon a test surface. The reflected wavefront contains information about the surface contour. The collecting lens focuses the wavefront onto a Ronchi grating. The irradiance pattern produced on the observation screen is a sheared interferogram called a Ronchigram, where the irradiance across the pattern is proportional to the slope of the test surface in the direction perpendicular to the grating structure.

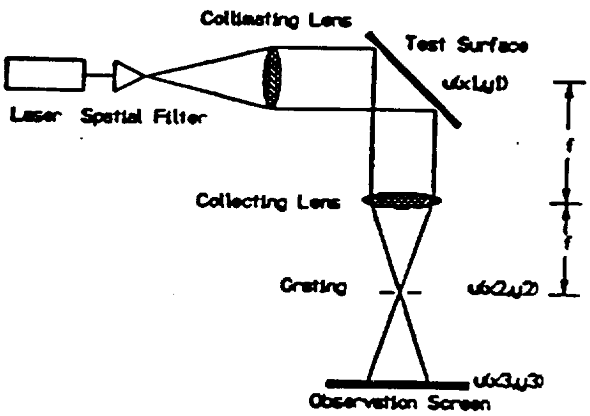


Figure 1 Single channel Ronchi test

3. RONCHI THEORY

3.1 Linear system theory

The Ronchi test can be theoretically explained as either a geometric ray or an interferometric test. From a geometric ray point of view, the fringes in a Ronchigram are the result of the deviation of a ray from its ideal path because of slope errors. In an interferometric test the fringes in the Ronchigram are produced from the interference between overlapping diffraction orders. The overlapping wavefronts are produced by a shearing of the original wavefront. Therefore the Ronchi test can be approached as a sheared wavefront interferometer.

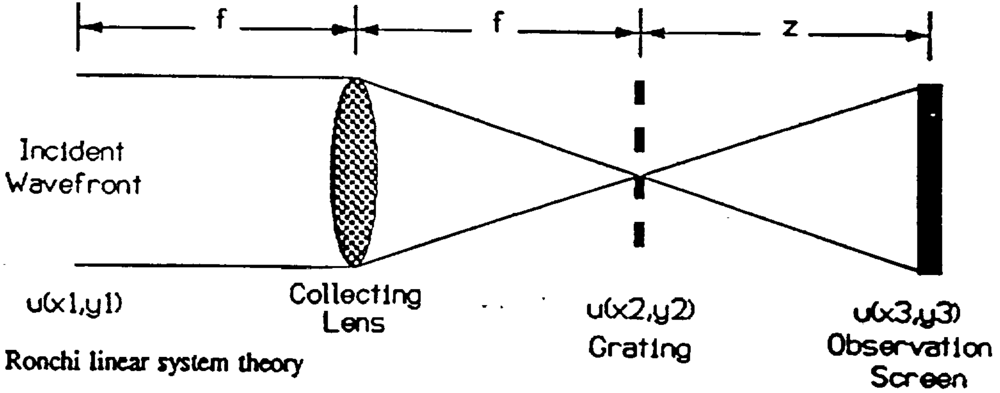


Figure 2 Ronchi linear system theory

The basic function of the Ronchi test can be explained from a linear system theory point of view using the geometry shown in Fig. 2. The wavefront reflected off of the test surface, $u(x_1,y_1)$, contains information about the surface contour. The wavefront observed at the focus of the lens, directly before the grating, is the Fourier transform of the input wavefront.

$$U(\xi_2,\eta_2) = \mathcal{F}\{u(x_1,y_1)\} \tag{1}$$

After the grating, the wavefront is modulated by the grating transmission function, t(x2,y2).

$$u'(x'_2,y'_2) = U(\xi_2,\eta_2)\, t(x'_2,y'_2)$$

$$u'(x'_2,y'_2) = U(\frac{x'_2}{\lambda f},\frac{y'_2}{\lambda f})\, t(x'_2,y'_2) \ . \tag{2}$$

At the observation screen, the wavefront amplitude is the Fourier transform of the wavefront at the focal plane of the lens.

$$u(x_3,y_3) = \mathcal{F}\{u'(x'_2,y'_2)\}$$

$$u(x_3,y_3) = \mathcal{F}\{U(\frac{x'_2}{\lambda f},\frac{y'_2}{\lambda f})\, t(x'_2,y'_2)\}$$

The wavefront at the observation screen is a scaled version of the incident wavefront, u_i , convolved with the Fourier transform of the grating transmission function.

$$u(x_3,y_3) = u(\frac{f x_3}{z}, \frac{f y_3}{z}) * T(\frac{x_3}{\lambda z},\frac{y_3}{\lambda z}) \tag{3}$$

When the distance z (distance from the grating to the observation screen) is equal to the focal length of the collection lens, the wavefront scale factor disappears and equation 3 is the original wavefront convolved with the Fourier transform of the grating transmission function. The grating transmission function is the important mechanism in the Ronchi test.

The irradiance at the observation screen is given by:

$$E(x_3,y_3) = <|u(x_3,y_3)|^2> \tag{4}$$

3.2 Square wave grating

The type of grating used in the Ronchi test determines the interpretation of the fringes in the Ronchigram. A square wave grating was used in this experiment. A square wave grating is a series of transparent and opaque straight line bands. The Fourier transform of a square wave grating is a sinc function convolved with a comb function. The square wave function and its Fourier transform are shown in Fig. 3.

When the width of the bands in the square wave grating is equal to one half of the grating period, the even orders of the sinc function are missing. If the contributions of the higher odd orders (3, 5, 7, ..) are

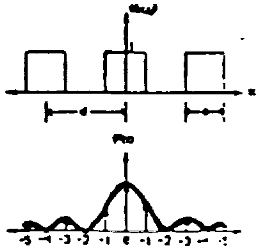


Figure 3 50/50 Duty cycle square wave grating assumed to be negligible, the Fourier transform of a square wave grating can be approximated as three delta functions. Plugging this result into equation 3 gives the wavefront at the observation screen for a square wave grating with a 50/50 duty cycle.

$$u(x_3, y_3) = u_a + u_b + u_c = u(\frac{x_3}{\lambda z}) + u(\frac{1}{\lambda z}[x_3 - d]) + u(\frac{1}{\lambda z}[x_3 + d]) \qquad (5)$$

Equation 5 is of the same form as the general equation for a lateral shearing interferometer.[2]

$$W(x - \frac{\Delta}{2}, y) - W(x + \frac{\Delta}{2}, y) = m\lambda \qquad (6)$$

The Δ in equation 6, referred to as the shear distance, is the distance between the centers of the wavefronts. Figure 4 shows a schematic of the shearing effect of the square wave grating.

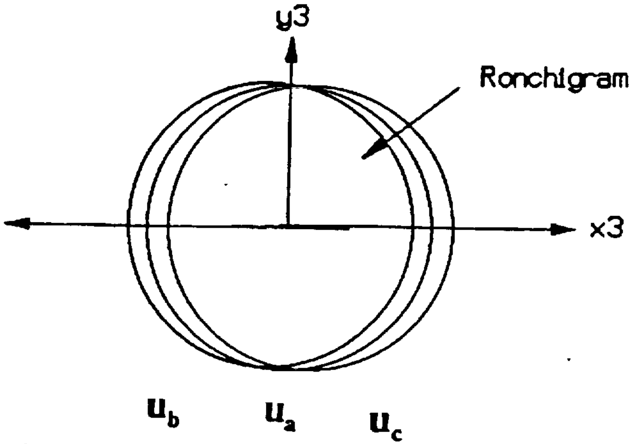


Figure 4 Sheared wavefronts

The irradiance at the observation screen is given in equation 7. Just as in a shearing interferometer, the interference fringes in the Ronchigram are produced in the overlap region of the laterally sheared wavefronts. These overlapping regions are expressed in the irradiance equation as the cross terms in parenthesis.

$$E(x_3, y_3) = E_a + E_b + E_c + (u_a u_b^* + u_a^* u_b) + (u_a u_c^* + u_a^* u_c) + (u_b u_c^* + u_b^* u_c) \qquad (7)$$

There are three fringe patterns produced in the Ronchigram, the interference between wavefronts (a,b), (a,c), and (b,c). Given our setups coarse gratings (50 lp/in), the actual shear distance is very small. The observed irradiance (Ronchigram) is produced from the overlap region of all three wavefronts.

3.3 Equivalent wavelength

In interferometry, the fringe spacing in an interferogram is proportional to the wavelength of light used in the interferometer. The irradiance at any point in an interferogram, E(x,y), can be expressed as a function of measurement wavelength and optical path difference (OPD) between the two interfering wavefronts.

$$E(x,y) = E_0(1 + \cos(\frac{2\pi OPD}{\lambda})) \qquad (8)$$

$$where\ \lambda = measurement\ wavelength$$

The OPD in a shearing interferometer is given by the difference between the wavefront and the sheared wavefront.

$$OPD = W(x,y) - W(x + \Delta x,y) = \Delta W(x,y) \tag{9}$$
$$\text{where } \Delta = \text{shear distance}$$

The relationship between the differential wavefront and the slope of the wavefront is shown in equation 10.

$$\Delta W(x,y) = \frac{\delta W(x,y)}{\delta x}\Delta x = \alpha(x,y)\Delta x \tag{10}$$
$$\text{where } \alpha(x,y) = \text{wavefront slope}$$

Inserting equation 10 into equation 8 gives the irradiance equation for a shear interferometer. This is also the irradiance equation of a Ronchi test, since the Ronchi test is similar to a shearing interferometer.

$$E(x,y) = E_o(1 + \cos(\frac{2\pi\alpha\Delta x}{\lambda})) \tag{11}$$

The equivalent wavelength, λ_{eq}, of a shearing interferometer is defined as the measurement wavelength divided by the shear distance.

$$\lambda_{eq} = \frac{\lambda}{\Delta x} \tag{12}$$

Inserting the equivalent wavelength definition into equation 11, the irradiance equation can be rewritten as equation 13.

$$E(x,y) = E_o(1 + \cos(\frac{2\pi\alpha}{\lambda_{eq}})) \tag{13}$$

Thus, by proper scaling of the interferogram, conventional interferogram analysis techniques will provide a direct measurement of the wavefront slope.

3.3.1 Equivalent wavelength (Square wave grating)

In order to derive an equivalent wavelength for a square wave grating with a 50/50 duty cycle the dominant fringe pattern in the Ronchigram has to be determined. The shear between wavefronts (a,b) and (a,c), in Fig. 4, is dominant over the shear between wavefronts (b,c) for two reasons. First, the intensity of the wavefront u_a, the zero diffraction order, is much higher than the intensity of the other two wavefronts. Second, the fringe pattern produced from (a,b) and (a,c) are identical and they overpower the fringe pattern produced from (b,c).

The shear between wavefronts (a,b) and (a,c) is identical and is equal to

$$\Delta x = \frac{\lambda z}{d} \tag{14}$$

The shear distance is normalized by dividing the shear by the beam radius. Normally, the beam radius is equal to the radius of the collection lens because the lens pupil is over filled and the distance z is equal to the focal length of the collection optics.

$$\Delta' x = \frac{\Delta x}{x_{max}} = \frac{2\lambda f}{D}$$

Substituting the normalized shear into equation 12, the equivalent wavelength for the Ronchigram produced from (a,b) and (a,c) is

$$\lambda_{eq} = \frac{d}{2(f/\#)}. \qquad (15)$$

The fringe spacing in the Ronchigram is independent of the measurement wavelength. It is only dependant upon the spacing of the grating and the f/# of the test beam. This means that the Ronchi test can be performed with either a monochromatic or white light source. In addition, the sensitivity of the Ronchi test can be changed by simply changing the grating frequency.

4. TWO CHANNEL DESIGN

In order to reconstruct the surface contour, two orthogonal slope patterns, either x-slope and y-slope or r-slope and θ-slope, are required. A single channel Ronchi test can be used to characterize a static surface by rotating the grating 90o during the experiment to obtain two orthogonal slope patterns. The oil surface is a dynamic surface and requires that two orthogonal slope patterns be obtained in real time. A two channel Ronchi test has two separate arms, one with a vertical grating and one with a horizontal grating. The two channel test can produce two orthogonal slope patterns in real time.

4.1 Optical layout

A schematic of the two channel Ronchi test is shown in Fig. 5. A collimated wavefront is incident on the oil surface. The wavefront reflected off of the surface contains information about the surface contour. The reflected wavefront is split and directed through two separate channels. One of the wavefronts is focused onto a horizontal grating by collecting lens, L3, producing a Ronchigram with vertical slope information. The other wavefront is focused onto a vertical grating by collecting lens, L2, producing a Ronchigram with horizontal slope information. The two channels are staggered so that the distance from the oil surface to each collection lens is equal. The wavefronts leaving lenses L4 and L5 are positioned side-by-side on an observation screen.

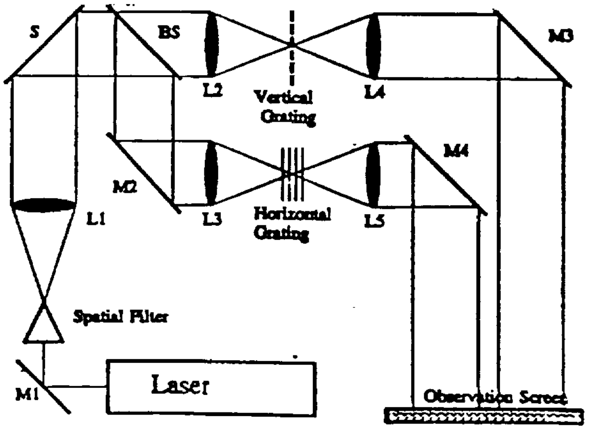


Figure 5 Two channel Ronchi test

4.2 System f/#

The surface under test should be located in the front focal plane of the collection optics. The maximum surface slope that will enter the collection lens of the Ronchi test is dependant upon the f/# of the collection lens. From the geometry shown in Fig. 6, it is clear that the f/# requirement to see a given slope is.

$$\frac{f}{\#} = \frac{1}{2\tan(2\alpha)} \tag{16}$$

A graph of the system f/# required for a given surface slope is provided in Fig. 7. The two channel Ronchi test designed for this experiment is a f/4 system. Therefore a surface slope of up to approximately 60 μm/mm will enter the two channel Ronchi test. The maximum expected slope of the oil surface is 26.7 μm/mm.

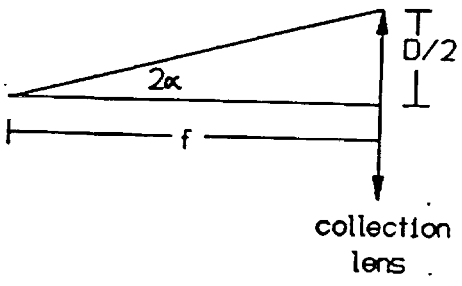


Figure 6 System geometry

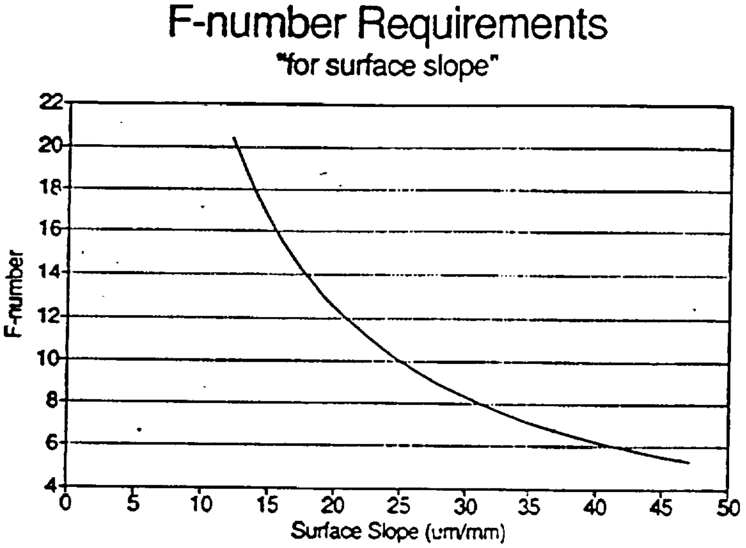


Figure 7 System f/# requirements

4.3 Grating frequency selection

One of the advantages of the Ronchi test is that the dynamic range and resolution of the system are dependant on the characteristics of the gratings and the f/# of the system not the measurement wavelength. Therefore a simple change of the gratings will permit a different range of deformations to be monitored. This gives the Ronchi system great flexibility.

The proper grating spacing, d, for a given maximum surface slope can be found using equation 17. To enhance data evaluation, the optimum number of fringes per pupil radius is 10 fringes. Fewer fringes would reduce the accuracy and more fringes may introduce additional complexity.

$$d = \frac{2 f \alpha_{max}}{m} \tag{17}$$

where f =focal length, α_{max}, = maximum surface slope and m = fringes per radius.

5. TWO CHANNEL CALIBRATION

5.1 Calibration experiment

The two channel Ronchi test was calibrated by testing a known amount of introduced defocus. The defocus was introduced by translating the collimating lens, L1, in Fig. 5. Defocus was chosen to calibrate the system for three reasons: 1) It is easy to introduce. 2) It is easy to analyze. 3) The amount of defocus introduced is directly proportional to the magnitude of the translation. Three different frequency square wave grating were used in the experiment. Each of the different gratings has a different equivalent wavelength and will therefore provide a test of the equivalent wavelength equation for square wave gratings with a 50/50 duty cycle. Table 1 is a listing of all the trials ran in the calibration experiment.

Table 1 Listing of Trials in Two Channel Calibration Experiment

Trial	Grating Frequency [lp/in]	Equivalent Wavelength [micrometers]	Translation [mm]	Defocus [Waves]
1	250	12.10	2.0	1.045
2	250	12.10	3.0	1.623
3	250	12.10	4.0	2.237
4	250	12.10	5.0	2.903
5	150	20.16	3.0	0.950
6	150	20.16	4.5	1.509
7	150	20.16	6.0	2.049
8	150	20.16	7.5	2.588
9	50	60.48	5.5	0.650
10	50	60.48	8.0	0.882
11	50	60.48	9.5	1.076

5.2 Ronchigram analysis

The two channel Ronchi test produced two Ronchigrams for each trial. These Ronchigrams were of orthogonal slope patterns, x-slope and y-slope. Each slope map was digitized and fit to a set of Zernike polynomials. The two slope maps were then combined to produce the original surface contour by linear combinations of the Zernike polynomials.[3]

5.3 Calibration results

Figure 8 is a graph of the data shown in Table 1. The trials for each grating frequency demonstrate a linear relationship between the amount of lens translation and the amount of introduced defocus. Multiplying each defocus value by the appropriate equivalent wavelength value produces the result shown in Fig. 9. All the trials fall upon the same line. The equivalent wavelength equation does relate data taken with different grating frequencies.

Figure 10 shows the percent error between the data and the best fit line shown in Fig. 9. The average percent error of the current two channel system is ±5%. The percent error for small amounts of defocus is higher than five percent. It is speculated that this is because of the inaccuracy involved in placing the gratings at best focus.

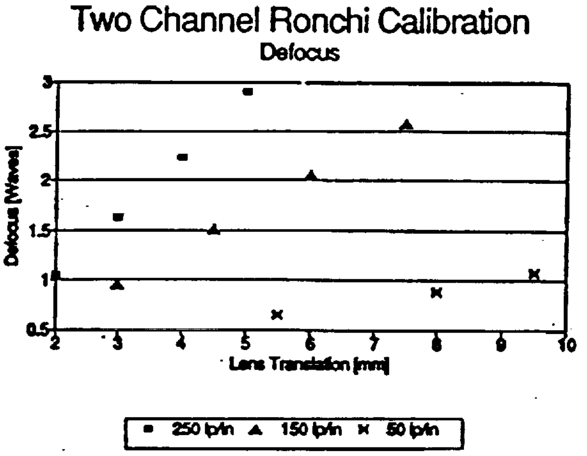


Figure 8 Two channel calibration data

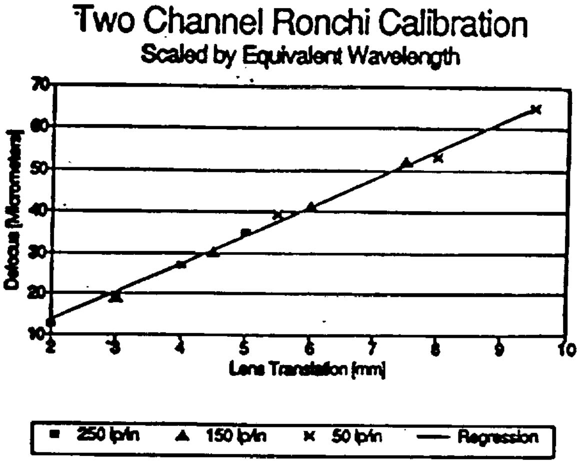


Figure 9 Data scaled by equivalent wavelength

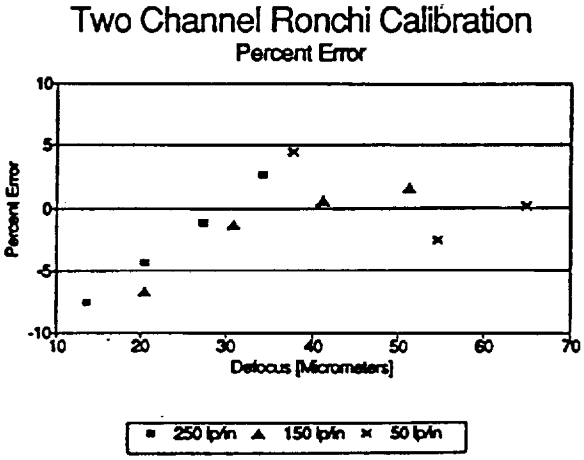


Figure 10 Percent error

6. PRELIMINARY DATA

The equivalent wavelength equation needs to be modified when the test object is a surface instead of a wavefront as in the case of the two channel calibration. The cosine of the reflected angle is multiplied by the denominator of the equivalent wavelength equation. When the reflected beam is normal to the test surface equation 15 is the proper scale factor (cos (0) = 1).

$$\lambda_{eq} = \frac{d}{2(f/\#)\cos(\theta)} \qquad (18)$$

where θ = *the angle between the optic axis of the collection lens and the surface normal.*

6.1 Diamond machined known surface

A diamond machined aluminum surface was examined with the two channel Ronchi test. The surface was machined flat except for a 5 mm diameter area in the center of the surface. This central area contained a 54 micrometer depression. A Ronchigram of the surface and a contour map of the surface generated from the Ronchigram information are shown in Fig. 11. The two channel system measured the depression to be 52.4±2.6 micrometers. The area around the depression is flat.

6.2 Oil surface

The oil surface usually has either a depression or a bump at the location the CO$_2$ beam strikes the surface. The size of the feature can be changed as well as the polarity by adjusting the size of the incident CO$_2$ beam and the laser power level. Fig. 12 shows the Ronchigrams and a contour map generated from the Ronchigrams for a bump on the surface. The CO$_2$ beam size on the surface was 6 mm. The two channel system measured the bump to be 17±1 micrometers.

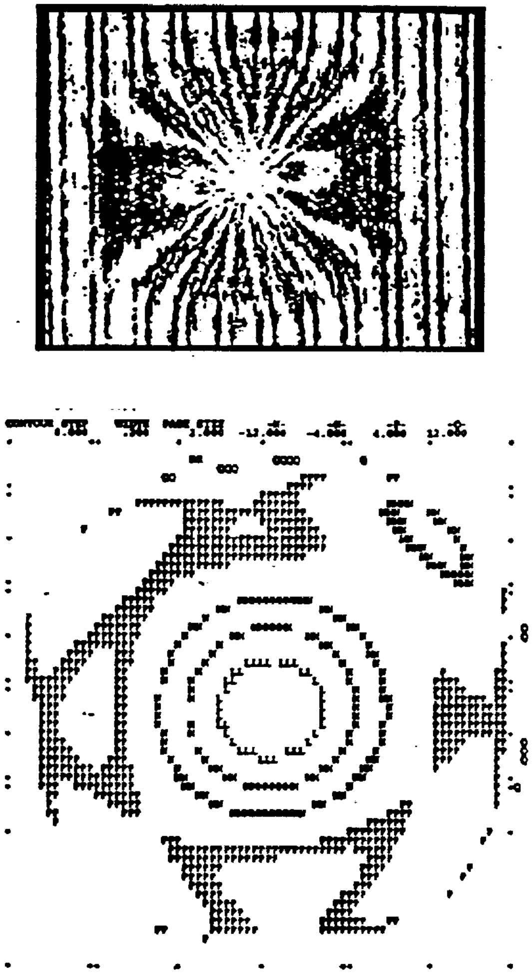


Figure 11 Diamond turned surface

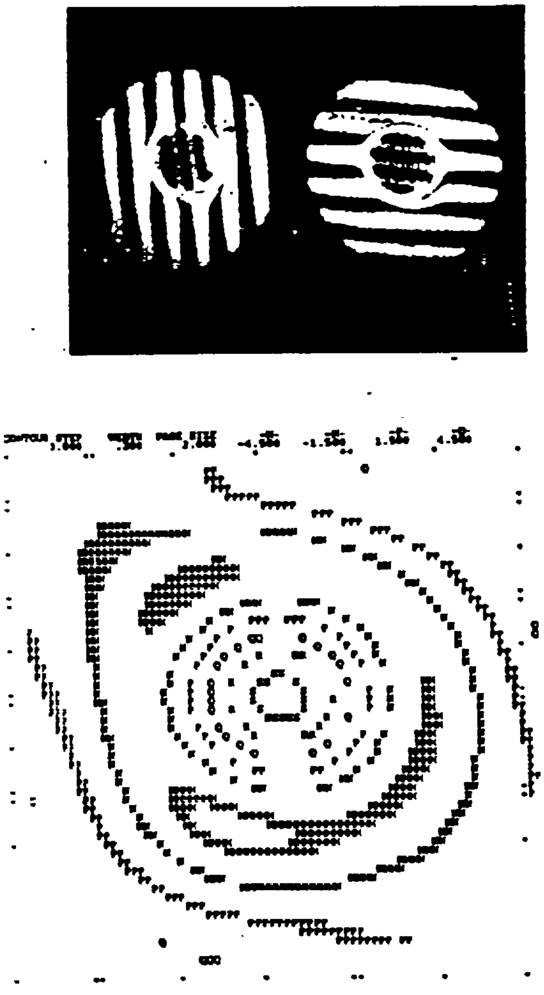


Figure 12 Free oil surface

7. CONCLUSIONS

The Ronchi test can be treated as a shearing interferometer. The number of sheared wavefronts and the magnitude of the shear is dependant upon the characteristics of the grating. The fringe spacing in a Ronchigram is proportional to the equivalent wavelength of the Ronchi system. The equivalent wavelength is dependant upon the spacing of the grating and the f/# of the test beam. The sensitivity of the Ronchi test is both known and variable. It can be changed by simply changing the grating spacing.

The two channel Ronchi system was calibrated and has an accuracy of ±5%. The preliminary data demonstrates the ability of the Ronchi test to contour an aspheric surface.

8. ACKNOWLEDGEMENTS

This project is funded by a NASA research grant (NAG3-1300) from the NASA Lewis Research Center. A special thanks to the people of NASA Lewis and the Ohio Aerospace Institute, especially Alexander Pline, for the support they have provided on this project.

9. REFERENCES

1. Vasco Ronchi, "Forty Years History of a Grating Interferometer," *Applied Optics.* 3, 437-451, 1964.

2. Murty, M. V. R. K., from Ch. 4 Optical Shop Testing. (John Wiley and Sons: New York) 1978.

3. Fisher, David and H. Philip Stahl, "Vector Formulation for Ronchi shear interferogram fitting," *SPIE Conference No. 1755*, July 1992.

Contouring of a free oil surface

William S. Meyers, Brent C. Bergner, Ronald D. White, David T. Huff, and H. Philip Stahl

Rose-Hulman Institute of Technology, Department of Physics and Applied Optics
5500 Wabash Avenue, Terre Haute, Indiana 47803

ABSTRACT

A non-contact (i.e. optical) technique is required to measure the deformation of a free oil surface under the influence of a localized thermal load. This deformation is caused by surface tension driven thermal convective flow inside the fluid, and can be as large as 250 um. Therefore, conventional interferometry is not possible. Instead, a Ronchi technique is proposed for contour mapping the oil surface. This paper presents a design concept for a two channel Ronchi instrument and some preliminary results from a feasibility study using a single channel Ronchi instrument.

1. INTRODUCTION

Scientists and engineers at Case Western Reserve University and NASA Lewis Research Center are conducting research on a variety of fundamental phenomena inherent in low gravity processes. An important phenomenon considered to have a significant impact on processes such as containerless processing and crystal growth is that of thermocapillary flow, or fluid flow driven by surface tension. Thermocapillary flows result when temperature differences arise across a liquid/gas interface. These differences cause a change in the surface tension along the interface, which results in thermocapillary flow. Improvements in the understanding of thermocapillary flow will help lead to technological breakthroughs in space processing of semiconductor materials.

The ability of the free surface to deform during thermocapillary flow is an important aspect of this problem. These deformations can alter the surface temperature distribution, the driving force in the flow, and the resulting fluid flow patterns, causing an unsteady flow which is detrimental to the crystal growth process. Therefore, the development and implementation of a technique to study the surface deformation as a function of experimental parameters such as fluid properties, heating modes, is crucial to the understanding of the underlying physics.

This paper describes the design of a technique for measuring surface deformations of a 3 cm diameter pool of silicon oil. Since the surface is a fluid, touching it would change its shape. Therefore the technique must be non-contact, i.e. optical. Two techniques were considered: geometric and interferometric. A geometrical technique measures surface slope, while an interferometric technique measures surface height. The geometric approach was chosen over the interferometric approach for two reasons. First, conventional visible wavelength interferometry would produce a fringe pattern too dense to analyze because of the anticipated deformation range of up to 250 micrometers. And, while infrared interferometric techniques could produce a resolvable fringe pattern, the infrared beam might affect the system's thermal loading. Second, geometric techniques are dependant on the slope of the surface and are therefore insensitive to any jitter or tilt experienced by the oil surface while in space. One disadvantage of using geometric techniques is that two orthogonal data sets, either x-slope and y-slope or r-slope and Θ-slope are needed to completely reconstruct the surface shape.

Five classical geometric techniques were considered: Foucault knife-edge test, Hartmann screen test, wire tests, grating interferometry, and Ronchi test. The knife-edge and wire tests were not chosen because they provide data for only a single zone or diameter, not for the entire aperture. The Hartmann screen can provide two orthogonal data sets at one time, but it has a limited spatial resolution when compared with the Ronchi test. The Ronchi test was chosen as the technique to use for three reasons: the small dimensions of the experimental pool are tailored to the Ronchi requirement of a collimated beam, the Ronchi data is less complicated to analyze, and it can be integrated with phase-measuring techniques.

SPIE Vol. 1779, 1992.

2. BACKGROUND

The Ronchi test was discovered by the Italian physicist Vasco Ronchi in 1923. He noticed that when a grating was positioned near the center of curvature of a mirror that the image of the grating was superimposed upon the original grating. The combination of the two gratings produced a moire type effect whose pattern depended upon the aberrations of the mirror. Ronchi felt that this method was a good check for mirror quality.[1]

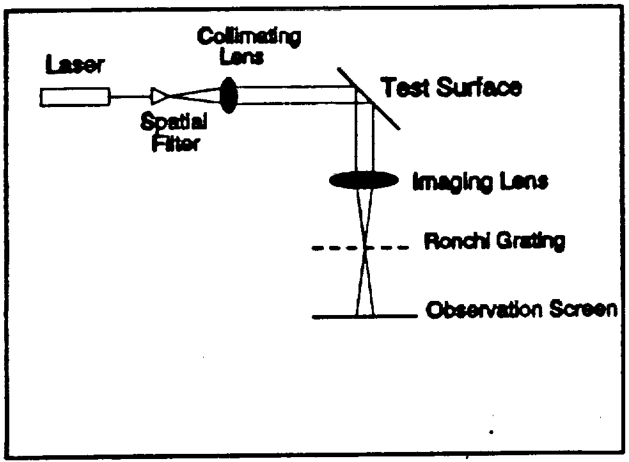


Figure 1 Single channel Ronchi test

The properties of the Ronchi test can be understood by examining the single channel Ronchi setup shown in Fig. 1. The source in Fig. 1 is a "point source" created by spatially filtering a HeNe ($\lambda = 632.8$ nm) laser beam. A lens is positioned one focal length in front of the spatial filter, producing a collimated beam of light. The collimated wavefront is incident upon the test surface and reflected to an achromatic imaging lens. The reflected wavefront contains information about the surface deformations. The imaging lens focuses this wavefront onto a Ronchi grating. The irradiance pattern produced on the observation screen is a sheared interferogram called a Ronchigram, where the irradiance across the pattern is proportional to the slope across the test surface.

The Ronchi test measures wavefront aberrations present in a system. Therefore, if there are no aberrations, the Ronchigram will have no fringes. When the test surface in Fig. 1 is a plane mirror no aberrations are introduced into the collimated wavefront. Therefore, if the grating is at the focus of the imaging lens, either a bright spot or a dark spot is produced on the observation screen. A dark spot indicates that the focused beam is incident upon a dark band in the grating and a bright spot indicates that the focused beam is incident upon an area between two dark bands in the grating. Now, if the grating is moved away from the focal plane of the imaging lens, defocus is introduced. Defocus appears as straight line fringes in a Ronchigram.

When the surface is aberrated, the Ronchigram contains fringes which are proportional to the slope of the surface. Since the Ronchi test measures the wavefront's slope in a direction perpendicular to the grating lines, it is necessary to obtain two orthogonal patterns to fully represent the surface shape.

3. CERTIFICATION OF A SINGLE CHANNEL RONCHI TEST

For any measuring instrument, it is necessary to develop a certification technique. For the Ronchi test, this can be accomplished by introducing a known amount of aberration and then testing to see if that amount is actually measured. In this section, a certification process using defocus is discussed.

3.1 Theory

An ideal spherical wavefront is one which will produce a perfect point image at its focal point. Such a wavefront is described by,

$$w(x,y) = \frac{x^2 + y^2}{2R} \qquad (1)$$

where (x,y) are pupil coordinates, and
R is the wavefront radius of curvature.

If the wavefront is defocused by an amount $\Delta w(x,y)$,

$$w'(x,y) = w(x,y) + \Delta w(x,y) \tag{2}$$

it will focus at a slightly different location, $R+\varepsilon_z$, where ε_z is called the longitudinal ray aberration. The wavefront error introduced by defocus can be described by,

$$\Delta w(x,y) = w_{020}\,(x^2 + y^2) \tag{3}$$

where w_{020} is the Defocus coefficient.

Figure 2 shows a reference wavefront and an aberrated wavefront with a profile function of $W(x,y)$. The longitudinal aberration, ε_z, and the transverse aberration, ε_y, are also shown. The slope error, α, is the angle between the intersection of the reference wavefront and the aberrated wavefront surface normals. The slope error is equal to the partial derivative of the profile function W with respect to y.

$$\alpha = \frac{\partial W}{\partial y} \tag{4}$$

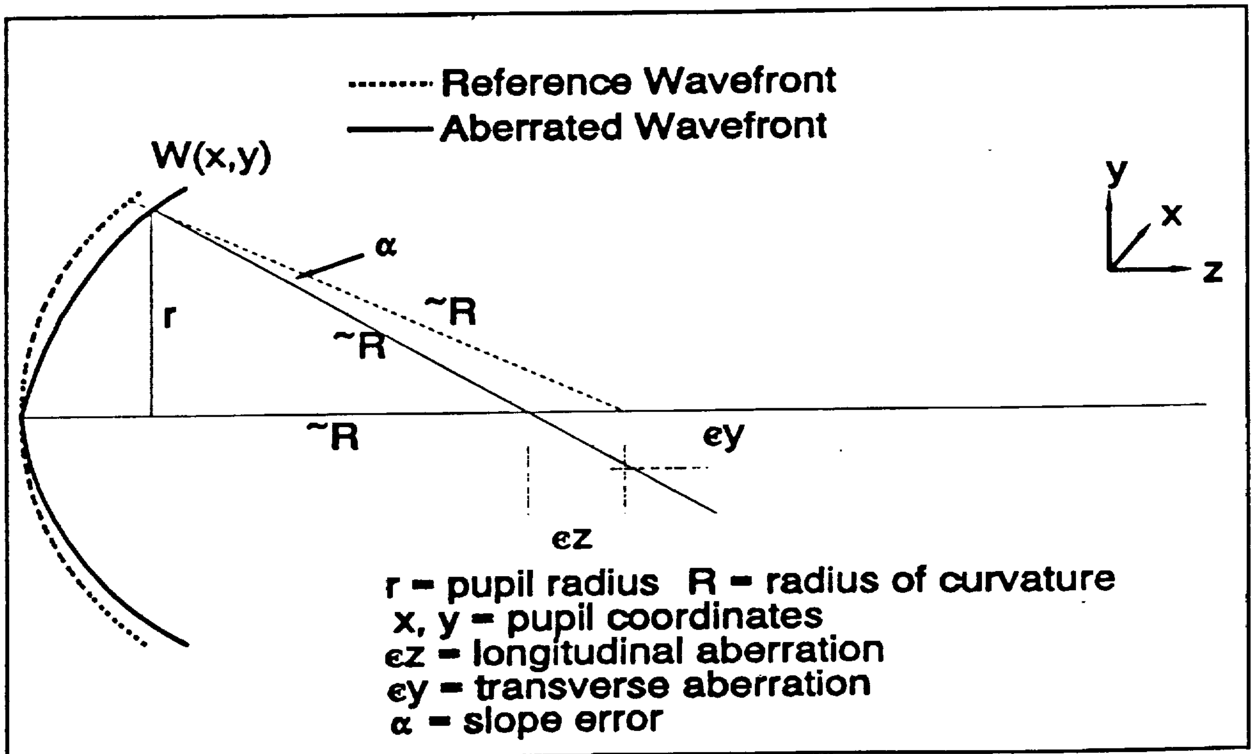


Figure 2 Comparison of reference and aberrated wavefronts

Using similar triangles and the small angle approximation it can be shown from Fig. 2 that:

$$\varepsilon_y = \alpha R \qquad \varepsilon_z = \frac{R}{r}\,\varepsilon_y \tag{5}$$

where r = pupil radius.

Since Ronchigrams are in normalized pupil coordinates, it is necessary to normalize x, y, and α.

$$x' = \frac{x}{r} \qquad y' = \frac{y}{r} \qquad \alpha = -\frac{1}{r}\frac{\partial W}{\partial y'} \tag{6}$$

where r = pupil radius and (x',y') = normalized pupil coordinates.

By substitution of equations 4 and 6 into 5, the longitudinal and transverse aberrations can be rewritten as:

$$\varepsilon_y = -R\,\frac{\partial W}{\partial y} \qquad \varepsilon_z = -\frac{R^2}{r}\,\frac{\partial W}{\partial y} \tag{7}$$

$$\varepsilon_y = -\frac{R}{r}\,\frac{\partial W}{\partial y'} \qquad \varepsilon_z = -\frac{R^2}{r^2}\,\frac{\partial W}{\partial y'} \tag{8}$$

Taking the partial derivative with respect to y of the defocus term in equation 3, writing it in normalized pupil coordinates, and substituting it into equation 8 yields equation 9.

$$\varepsilon_z = -2\,w_{020}\,\frac{R^2}{r^2} \tag{9}$$

The f-number of the system, (f/#), shown in Fig. 2 is the radius of curvature of the reference sphere divided by the pupil diameter (2r).

$$\frac{f}{\#} = \frac{R}{2r} \tag{10}$$

Substituting the f-number into equation 9 provides an equation for the defocus in a wavefront that is dependant only upon the amount of longitudinal aberration and the system f-number.

$$w_{020} = \frac{-\varepsilon_z}{8\,(\frac{f}{\#})^2} \tag{11}$$

Thus, by simply translating the Ronchi grating, a known amount of aberration can be introduced.

3.2 Measuring defocus

The number of fringes in a Ronchigram is related to the partial derivative of the wavefront aberration[2].

$$\frac{\partial W}{\partial y} = -\frac{m\,d}{R} \tag{12}$$

where R = radius of curvature m = number of fringes per radius,
and d = grating spacing.

Writing equation 12 normalized pupil coordinates gives:

$$\frac{\partial W}{\partial y'} = -\frac{m\,d\,r}{R}. \tag{13}$$

Taking the partial derivative of the defocus term in equation 3 and writing it in normalized coordinates yields:

$$\frac{\partial \Delta w(x,y)}{\partial y'} = 2\, w_{020} \tag{14}$$

Setting the right sides of equations 13 and 14 equal and substituting in the definition of the f/# (equation 10) results in the following relation:

$$\omega_{020} = -\frac{m\, d}{4\,(\frac{f}{\#})} \tag{15}$$

Thus, the amount of defocus in the wavefront can be determined by counting the number of fringes per pupil radius in the observed Ronchigram.

3.3 Certification experiment

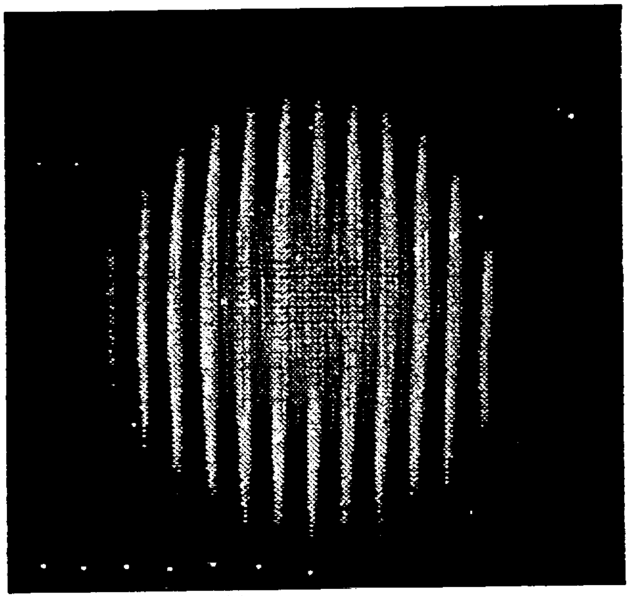


Figure 3 Defocus fringe pattern

The single channel Ronchi test shown in Fig. 1 was certified by testing a known amount of introduced defocus. The grating was initially placed in the focal plane of the lens and its position was recorded for a reference value. The grating was then displaced a known amount towards the lens and the output fringe pattern was photographed. An example defocus fringe pattern is shown in Fig. 3. Translating the grating is equivalent to introducing longitudinal aberration, ε_z, into the system. The number of fringes can be determined directly from the photographs for each translation. The experiment was performed with different frequency gratings and different grating orientations to insure that equation 15 was correct and to get a visual feel for the effects of using different gratings.

Table 1 presents the experimental data for various translations and grating frequencies. Equation 11 was used to calculate the value of defocus from a known amount of translation. Equation 15 was used to calculate the value of defocus from an observed number of fringes. The difference between the two calculations is shown in the defocus difference column in Table I. The reason that there are two different values for the f-number is because a portion of the data was taken with the grating in the vertical direction and a portion was taken with the grating in the horizontal direction. The elliptical beam produced by the test surface, which is oriented at 45°, results in two different f-numbers's for the different grating orientations.

The defocus values from the two different methods compare very well. The largest defocus difference is 38 microns. The orientation of the grating and the grating frequency do not seem to have any effect on the accuracy of calculated defocus values. The results show that the theory is correct and that the amount of defocus can be accurately calculated.

Table 1 Comparison of Calculated Defocus Values

Trial	Grating Spacing (lines/mm)	F-number 2.17±0.05 3.08±0.09	Translation ε (±0.01 mm)	Observed Fringes	Defocus from Translation (mm)	Defocus from Fringes (mm)	Defocus Difference (mm)
1	9.843	2.17	0.71	4.5	0.019	0.026	0.007
2	9.843	2.17	1.62	7.5	0.043	0.044	0.001
3	9.843	2.17	3.60	13.0	0.096	0.076	0.019
4	9.843	2.17	4.92	18.0	0.131	0.105	0.025
5	1.969	2.17	5.16	4.0	0.137	0.117	0.020
6	1.969	2.17	8.40	6.5	0.223	0.190	0.033
7	1.969	2.17	7.42	5.5	0.197	0.161	0.036
8	1.969	2.17	10.81	8.5	0.287	0.249	0.038
9	5.906	3.08	1.14	4.0	0.015	0.027	0.012
10	5.906	3.08	3.78	7.0	0.050	0.048	0.002
11	5.906	3.08	6.23	12.0	0.082	0.082	0.000
12	5.906	3.08	7.83	15.0	0.103	0.103	0.000

4. TWO CHANNEL DESIGN

In order to fully characterize the surface, it is necessary to measure the surface slope in two orthogonal directions. The single channel Ronchi system can be used to obtain two orthogonal slope patterns by simply rotating the grating 90° during the experiment. However this method only works for static surfaces like the plane mirror presently being used to calibrate the system. Since the pool of oil is a dynamic surface two orthogonal fringe patterns must be obtained simultaneously in real time. A two channel Ronchi test can produce two orthogonal fringe patterns in real time.

4.1 Optical layout

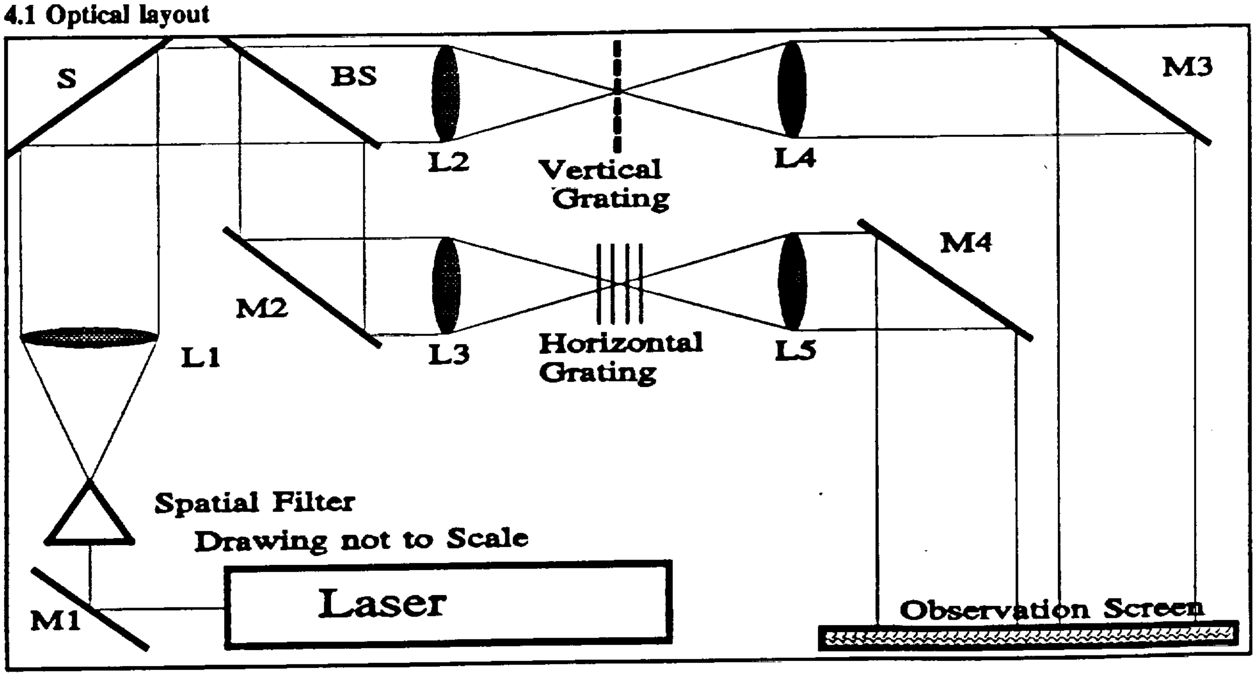


Figure 4 Two channel Ronchi test

A schematic of a two channel Ronchi test is shown in Fig. 4. The two channel test was designed to fit on a standard 2' by 3' optical breadboard. The space requirements of the optical component mounts are not shown in Fig. 4, but they were taken into account when parameters such as focal length of the lenses and the diameters of the elements were chosen.

The source shown in Fig. 4 is a HeNe (λ = 632.8 nm) laser. The light emitted by the laser strikes a fold mirror (M1) and is directed into a spatial filter (SP). The collimating lens, L1, is positioned one focal length away from the spatial filter, thereby producing a collimated beam of light. The collimated beam or plane wavefront is incident upon the test sample at 45°. The wavefront reflected off of the test sample contains information about the contour of the test surface. The reflected wavefront is then divided into two wavefronts by the beamsplitter (BS) and directed through two separate arms. One of the wavefronts is focused onto a vertical grating by L2 producing a Ronchigram with horizontal slope information, while the other wavefront is focused onto a horizontal grating by L3 producing a Ronchigram with vertical slope information. The two slope maps are imaged by lenses L4 and L5 onto the observation screen. The lenses L1-L5 were chosen to be achromats to insure that minimal amounts of spherical aberration are introduced.

The wavefronts exiting L4 and L5 are positioned side-by-side on an observation screen as pictured in Fig. 4 by the fold mirrors, M3 and M4. Additional optics are required in order to form an image of the two Ronchigrams onto a video camera.

4.2 Sample and sample mount

The test sample to be used in the two channel Ronchi setup was designed and manufactured by undergraduate mechanical engineers and applied optics students at Rose-Hulman. It was designed to simulate the free oil surface. Three criteria were used in the design of the test sample: the surface must be 3 cm in diameter, it must be capable of creating both odd and even deformations, and it must be large enough to insure that stress fringes from its mounting do not overlap with experimental deformation fringes. The test surface is a circular piece of plexiglas mounted between two metal rings. The back metal ring is connected to a tip-tilt mount. A black circle of 3 cm diameter is centered on the plexiglas surface to fiducialize the test surface.

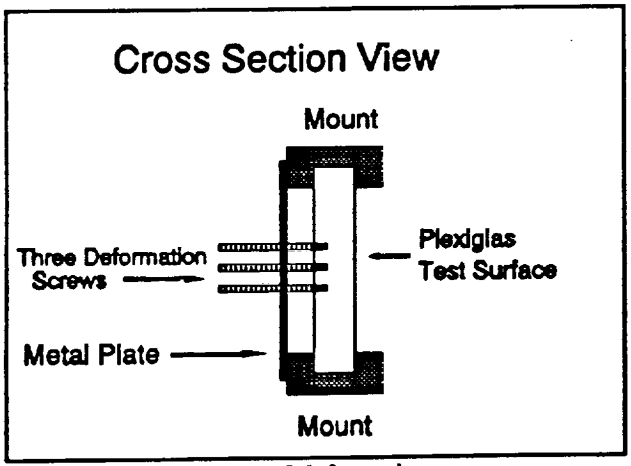


Figure 5 Arrangement of deformation screws

Deformations in the test surface are created by adjusting the position of one of three threaded screws which are glued into the back side of the plexiglas. The three screws are held in place with a threaded nut which is attached to a metal plate across the back of the sample mount. The metal plate is thick enough to insure that the plexiglas surface will deform before the metal plate. The middle screw is located at the center of the 3 cm test area and the other two screws are vertically displaced 1 cm on each side of the center screw. The screws can be translated forward or backwards to create an even deformation in the surface (a bump or a dimple). An odd deformation (a S-shaped deformation) can be created by a combination of screw adjustments, screwing the top screw back and the bottom screw forward or vice versa. The arrangement of the screws in the back of the plexiglas is shown in Fig. 5.

4.3 Vignetting

The next step is to bring the two designs, optical and mechanical, together to insure that no unwanted vignetting is introduced into the system. Vignetting occurs when the outer portion of the beam is partially blocked by the components or component mounts in the system. When designing an optical system, vignetting calculations are important to insure that the sizes of the optical components meet the beam size requirements of the experiment. The actual vignetting calculations for the two channel Ronchi test are not provided, but some of the guidelines followed are provided.

First, when ordering optical components it is a good rule of thumb to allow 20% of the component diameter to be used for the mounting of the component. Second, be aware of any surfaces in the design that are tilted relative to the beam path, like the test surface and the beamsplitter in the two channel Ronchi design. The beam that strikes the test surface needs to be large enough to overfill what appears to be an elliptical test surface due to the tilt of the surface and the clear aperture of the beamsplitter, which is reduced because of tilt, must be large enough to allow the reflected beam to pass through the mounted beamsplitter. Third, be sure that none of the components on the table are blocking a beam that is propagating towards another component. For example, the mounts for mirrors M3 and M4 in Fig. 4 were designed to insure that the mount for M4 did not block the beam from M3 on its way to the observation screen. Fourth, be sure that the output beams can be positioned on the output device. In Fig. 4 the output device is the observation screen.

4.4 Grating frequency selection

One of the major advantages to using the Ronchi test is that the dynamic range and resolution of the system are dependant only on characteristics of the grating used. Therefore if the experiment ever requires that a different range of deformations be monitored then only the grating needs to be changed, not the whole setup. This gives the system great flexibility.

In order to choose the proper grating spacing one needs to know the maximum anticipated surface slope, the relative sphere radius of curvature, and the maximum number of fringes per pupil radius. The proper grating spacing for a given experiment can be calculated using equation 16, where R is the reference sphere radius of curvature, α_{max} is the maximum anticipated surface slope, and m is the desired number fringes per pupil radius.

$$d = \frac{2 R \alpha_{max}}{m} \qquad (16)$$

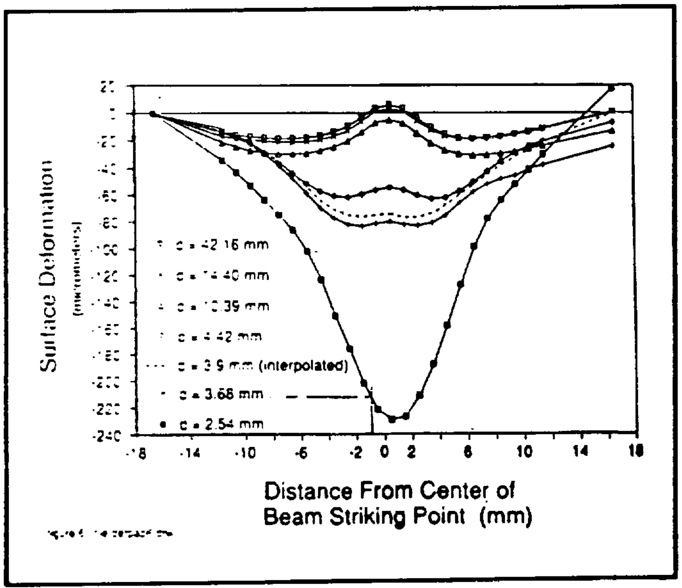


Figure 6 Anticipated deformations

In the Ronchi test, the reference sphere radius of curvature is determined by the imaging lens focal length. For the free oil surface, the anticipated maximum slope was found from Fig. 6 to be 26.7 um/mm. To enhance data evaluation, the optimum number of fringes per pupil radius is 10. Fewer fringes would reduce accuracy and more fringes may introduce additional complexity.

4.5 Diffraction effects

The Ronchi test can be viewed from a physical optics point of view as a diffraction phenomenon. Just as a diffraction pattern has higher diffracted orders so does a Ronchigram. The clarity of the fringe pattern in the output Ronchigram can be increased by removing some of the noise of the higher diffracted orders from the pattern. A spatial filtering experiment was performed to remove some of the noise and it had some positive results. A better approach is to use a sine grating instead of a step grating. The sine grating produces only the diffraction orders 0, and ±1. It will remove higher order noise from the fringe patterns and increase fringe clarity.

5. CONCLUSIONS

A two channel Ronchi instrument has been designed to contour the free oil surface. It can produce two orthogonal fringe patterns at the same time, thus allowing a dynamic surface to be monitored in real time. It is also designed to allow for easy implementation of a phase measuring capability in order to distinguish between bumps and dimples on the surface.

6. ACKNOWLEDGEMENTS

This project is funded by a NASA research grant (NAG3-1300) from the NASA Lewis Research Center. A special thanks to the people of NASA Lewis, especially Alexander Pline, for the support they have provided on this project.

7. REFERENCES

[1,2] Malacara, D., Optical Shop Testing. (John Wiley and Sons: New York), 1978.

Sensitivity of two-channel Ronchi test to grating misalignment

William S. Meyers and H. Philip Stahl

Rose-Hulman Institute of Technology
Center for Applied Optics Studies
5500 East Wabash Avenue
Terre Haute, Indiana 47803

ABSTRACT

The Ronchi test is performed by placing a periodic grating at or near the focus of an optical system. Like most geometric tests, the Ronchi test measures wavefront slope instead of the contour of the wavefront like conventional interferometry. Therefore when a Ronchi test is used to determine the contour of a non-symmetric surface, two orthogonal slope interferograms or Ronchigrams are required to reconstruct the surface. A two-channel Ronchi test with a horizontal and a vertical grating can be used to reconstruct a non-symmetric test surface. This paper investigates the errors introduced by different rotational and translation misalignments of the gratings in the two-channel Ronchi test relative to one another. The Ronchigrams are analyzed with a modal technique based on Zernike polynomials. The severity of each misalignment is dependant upon the frequency of the gratings used and on the f/# of the system.

1. INTRODUCTION

The Ronchi test is performed by placing a periodic structure at or near the focus of an optical system and monitoring the interference pattern or Ronchigram produced. The Ronchi test can be explained as a geometric test or an interferometric test. From a geometric ray point of view, the fringes in a Ronchigram are the result of the deviation of a ray from its ideal path because of slope errors. As an interferometric test, the fringes in the Ronchigram are produced from the interference between overlapping diffraction orders. The overlapping diffraction orders are sheared versions of the original wavefront. Therefore the Ronchi test can be approached as a shearing interferometer.

Shearing interferometers measure the slope of a wavefront instead of the height of the wavefront like conventional interferometers. One of the disadvantages of shearing interferometers is the requirement of two orthogonal slope maps of the surface, either x-slope and y-slope or r-slope and θ-slope, to completely reconstruct a surface without symmetry. If the surface under test is rotationally symmetric or if it has a static contour a single channel Ronchi test, Fig. 1, could be used to reconstruct the surface. The two orthogonal slope maps could be generated by rotating the grating ninety degrees during the experiment or by taking one photograph and analyzing it for both slope maps. However, if the surface under test is not rotationally symmetric and is dynamic, a two-channel Ronchi instrument must be implemented, Fig. 2.

The main differences between the two-channel Ronchi instrument and the single channel instrument are the addition of a second measurement module and a pupil relay module. The relay module creates an intermediate image of the surface under test at the plane of the second beamsplitter. The image created by the relay module is the object for both of the measurement modules. The addition of the second measurement module allows orthogonal slope information to be obtained in real time. One of the measurement modules obtains a vertical, y-direction, slope map of the test surface, S, with a horizontal grating. The second module obtains a horizontal, x-direction, slope map with a vertical grating. The output of each channel is imaged onto an observation screen. It should be noted that the orientation of the image of the surface at the

SPIE Vol. 1994, 1993.

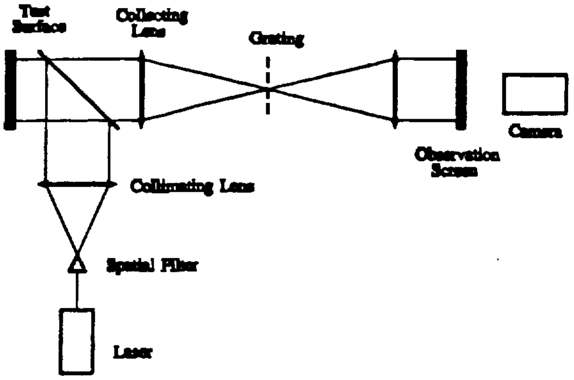


Figure 1 Single channel Ronchi system

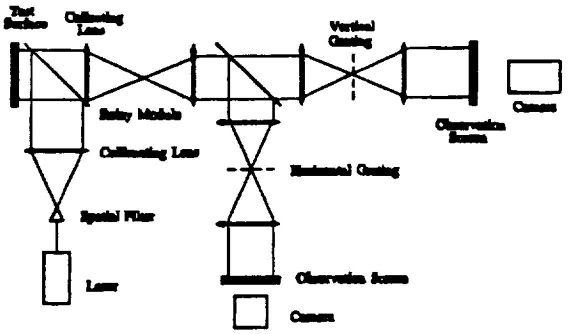


Figure 2 Two-channel Ronchi system

observation screen for the two-channel system is opposite to the orientation in the single channel due to the addition of the pupil relay module.

The wavefront at the observation screen is the wavefront reflected off of the test surface convolved with the Fourier transform of the grating transmission function[1] :

$$u(x_3,y_3) = u(x_3, y_3) * T\left(\frac{x_3}{\lambda z},\frac{y_3}{\lambda z}\right) \tag{1}$$

provided that the surface under test is in the front focal plane of the relay module and that the observation screen is in the back focal plane of the measurement modules. Therefore the grating transmission function is the important mechanism in the Ronchi test. It dictates how the fringes in the Ronchigrams should be interpreted.

2. INTERFEROGRAM ANALYSIS

Whenever two coherent wavefronts are at the same point in space, their amplitudes superimpose to create an interference pattern. The interference pattern contains information about the differences between the two wavefronts or their optical path difference (OPD). The two dimensional irradiance pattern created by the interfering wavefronts is known as an interferogram.

2.1 Interferogram theory

In interferometry, the fringe spacing in an interferogram is proportional to the wavelength of light used in the interferometer. The irradiance, E(x,y), at any point in an interferogram can be expressed as a function of the measurement wavelength and the optical path difference (OPD) between the two interfering wavefronts.

$$E(x,y) = E_o\left(1 + \cos\left(\frac{2\pi OPD}{\lambda}\right)\right) \tag{2}$$
$$\textit{where } \lambda = \textit{measurement wavelength}$$

The OPD in a shearing interferometer is given by the difference between the wavefront and the sheared wavefront.

$$OPD = W(x,y) - W(x+\Delta x,y) = \Delta W(x,y) \tag{3}$$
$$\textit{where } \Delta = \textit{shear distance}$$

The relationship between the differential wavefront and the slope of the wavefront is given by:

$$\Delta W(x,y) = \frac{\partial W(x,y)}{\partial x}\Delta x = \alpha(x,y)\Delta x \tag{4}$$
$$\textit{where } \alpha(x,y) = \textit{wavefront slope.}$$

Substituting equation 4 into equation 2 gives the irradiance equation for a shearing interferometer which is also the irradiance equation for the Ronchi test.

$$E(x,y) = E_o\left(1 + \cos\left(\frac{2\pi\alpha(x,y)\Delta x}{\lambda}\right)\right) = E_o\left(1 + \cos\left(\frac{2\pi\alpha(x,y)}{\lambda_{eq}}\right)\right) \tag{5}$$
$$\textit{where } \lambda_{eq} = \frac{\lambda}{\Delta x} = \textit{equivalent wavelength}$$

Thus, by proper scaling of the shear interferogram via the equivalent wavelength, conventional interferogram analysis techniques can be used to provide a direct measurement of the wavefront slope.

2.2 Ronchi test equivalent wavelength

To properly interpret a Ronchigram it is necessary to derive an equivalent wavelength, λ_{eq}, for the Ronchi test. The type of grating used in the Ronchi test governs how the fringes in the resultant interferogram or Ronchigram should be interpreted. An expression can be derived for the equivalent wavelength of the Ronchi system based on the characteristics of the grating and the geometry of the setup.

Since the lateral placement of the grating in the Ronchi test is arbitrary, the amplitude sinusoidal grating can be modelled as an amplitude cosine grating whose transmission function and Fourier transform are:

$$\mathcal{F}\left\{1+\cos\left(\frac{2\pi x}{d}\right)\right\} = \delta(x) + \delta(x-d) + \delta(x+d) \tag{6}$$
$$\textit{where } d = \textit{the grating spacing.}$$

The Fourier transform of the amplitude cosine grating is a set of three delta functions. Substitution of equation 6 into equation 1 yields the expression for the wavefront at the observation screen.

$$u(x_3, y_3) = u_a + u_b + u_c = u\left(\frac{x_3}{\lambda f}\right) + u\left(\frac{1}{\lambda f}[x_3 - d]\right) + u\left(\frac{1}{\lambda f}[x_3 + d]\right) \tag{7}$$

The wavefront at the observation screen is three sheared versions of the original wavefront. The irradiance at the observation plane is given by:

$$E(x_3 y_3) = <|u(x_3 y_3)|^2>$$

$$E(x_3, y_3) = E_a + E_b + E_c + (u_a u_b^* + u_a^* u_b) + (u_a u_c^* + u_a^* u_c) + (u_b u_c^* + u_b^* u_c). \tag{8}$$

The interference fringes in a Ronchigram are produced by the overlapping wavefronts which are the cross-terms in parenthesis in equation 8. The fringes correspond to changes in wavefront slope in a direction perpendicular to the shear direction. A low spatial frequency grating produces a small shear distance which in turn produces almost complete overlap between the wavefronts. High spatial frequency gratings produce large shear distances which reduce the amount of overlap. Therefore high frequency gratings do not test the full aperture of the wavefront.

There are three sets of cross-terms in equation 8. Therefore there are three sets of sheared wavefronts: (a,b), (a,c), and (b,c). In order to derive an equivalent wavelength for the sine grating, the dominant fringe pattern in the Ronchigram must be determined. The shear between wavefronts (a,b) and (a,c) is dominant for two reasons. First, the intensity of the wavefront u_a, the zero diffraction order, is much larger than the intensities of the other two wavefronts. Second, the fringe pattern produced from (a,b) and (a,c) are identical and they overpower the fringe pattern produced from (b,c). The equivalent wavelength is dominated by the shear between (a,b) and (a,c).

The shear between wavefronts (a,b) and (a,c) is identical and is equal to :

$$\Delta x = \frac{\lambda f}{d} \tag{9}$$

where f = focal length, and d = grating spacing.

Because interferograms are analyzed in normalized coordinates, the shear distance should be normalized by the maximum radius, x_{max}, the radius of the exit pupil of the system, which is the largest radius wavefront which can be tested.

$$\Delta x' = \frac{\Delta x}{x_{max}} = \frac{\lambda f}{d\, x_{max}}$$

Replacing x_{max} by D/2 where D is the diameter of the exit pupil of the system, gives:

Substituting the normalized shear into the definition for λ_{eq} results in:

$$\Delta x' = \frac{2\lambda f}{dD} = \frac{2\lambda(f/\#)}{d}. \tag{10}$$

$$\lambda_{eq} = \frac{\lambda}{\Delta x'} = \frac{d}{2(f/\#)}. \tag{11}$$

The equivalent wavelength of the Ronchi test is independent of the measurement wavelength. It only depends upon the spacing of the grating and the f/# of the system. This means that the Ronchi test can be performed with either a monochromatic or a white light source. In addition, the sensitivity of the Ronchi test can be changed by simply changing the grating frequency.

2.3 Surface reconstruction from slope information

Wavefront phase estimation from slope data is divided into two categories, modal and zonal. Zonal techniques estimate the phase value over a local region while modal techniques determine the coefficients of a set of aperture polynomials.[2] Rimmer developed a method for zonal estimation.[3] An example of a modal technique is the use of the Zernike polynomials to produce a global fit of the data.[4-5] In both cases, a least squares estimation is used to reconstruct the wavefront phase. Work has also been published on the estimation of slope data with the use of Fourier transform techniques.[6-7] This paper uses a modal algorithm to reconstruct the wavefront shape through a linear combination of Zernike coefficients which are individually fit to orthogonal x- and y-slope data sets.

All of the trials presented in this paper were analyzed with a modified version of the FRINGE software originally from the University of Arizona. Table 1 is a listing of the first eight Zernike polynomials. Table 2 provides the relationships between the Zernike coefficients of the slope data and the Zernike coefficients of the wavefront shape. For example, the wavefront x-astigmatism coefficient, A4, is dependant upon the x-tilt from the x-slope map, B1, and the y-tilt from the y-slope map, C2. Finally, the Seidel coefficients can be calculated from the first eight Zernike coefficients as shown in Table 3.

Table 1 First eight Zernike polynomials

Term	Aberration	F(x,y)
0	Piston	1
1	x-tilt	x
2	y-tilt	y
3	defocus	$2(x^2+y^2)-1$
4	x-astigmatism	x^2-y^2
5	y-astigmatism	$2xy$
6	x-coma	$3(x^3+xy^2)-2x$
7	y-coma	$3(x^2y+y^3)-2y$
8	spherical	$6(x^2+y^2)^2-6(x^2+y^2)+1$

Table 2 Wavefront Zernike coefficients from linear combinations of slope Zernike coefficients

Term	Aberration	Combination
A1	x-tilt	cannot be determined
A2	y-tilt	cannot be determined
A3	defocus	(B1+C2)/8-A8
A4	x-astigmatism	(B1-C2)/4
A5	y-astigmatism	(B2-C1)/4
A6	x-coma	(B4+B3+C5)/9
A7	y-coma	(B5-C4+C3)/9
A8	spherical	(B6+C7)/16

B coefficients are the x-slope.
C coefficients are the y-slope.
A coefficients are the surface shape.

Table 3 Relationship between Seidel and Zernike coefficients

Seidel term	Zernike combination	
Tilt	Magnitude	$((A1-2A6)^2+(A2-2A7)^2)^{1/2}$
	Angle	$\mathrm{Tan}^{-1}((A1-2A6)/(A2-2A7))$
Defocus		$2A3-6A8\pm((A4)^2+(A5)^2)^{1/2}$
The sign of the radical is chosen to minimize the magnitude.		
Astigmatism	Magnitude	$\pm 2((A4)^2+(A5)^2)^{1/2}$
	Angle	$0.5\,\mathrm{Tan}^{-1}(A5/A4)$
The sign of the radical is opposite the sign used for defocus.		
Coma	Magnitude	$3((A6)^2+(A7)^2)^{1/2}$
	Angle	$\mathrm{Tan}^{-1}(A7/A6)$
Spherical		6A8

3. EFFECT OF GRATING MISALIGNMENT

When aligned correctly the gratings in the two-channel Ronchi test should be at best focus of the system, perpendicular to the optic axis, and orthogonal to each other. However, misalignments can occur in the placement of the gratings which can produce measurement errors. The gratings have six degrees of freedom, three rotational and three translational as shown in Fig. 3.

Experiments were performed with an f/4, two-channel system shown in Fig. 4 to investigate the effect of each misalignment on the system performance. In each case, the x-slope grating was used as the reference and the y-slope grating was manipulated relative to the reference grating. The gratings used in the experiments were 50 lp/in, 50/50 duty cycle square wave gratings. The equivalent wavelength for these gratings can be approximated as the equivalent wavelength for the amplitude sinusoidal gratings. The collimating lens, L1, was defocused for all the experiments so that there was a sufficient number of sample points across the aperture. This is analogous to displacing the grating away from focus. Defocusing the collimating lens produces an initial "tilt" fringe pattern to be produced in both channels. The number of tilt

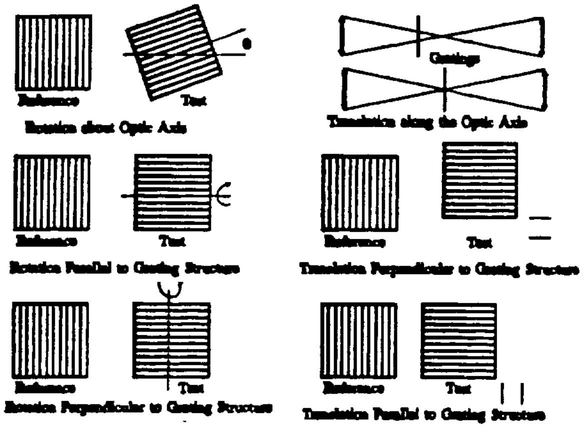


Figure 3 Geometry to determine tilt fringes

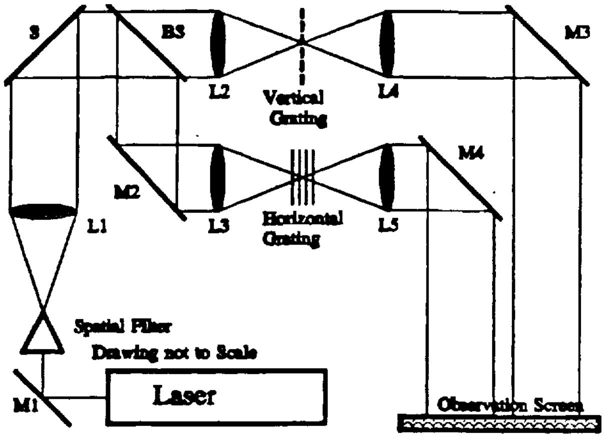


Figure 4 Two channel Ronchi system

fringes can be determined from similar triangles in Fig, 5 and from a knowledge of the grating frequency used in the experiment.

$$W_{fixed} = B1 = \frac{\epsilon_z \nu}{f/\#} \quad and \quad W_{variable} = C2 = \frac{\epsilon_z \nu}{f/\#} \tag{12}$$

$$where\ \epsilon_z = the\ distance\ from\ the\ grating\ to\ focus,$$
$$\nu = grating\ frequency,\ and\ f/\# = the\ system\ f\text{-}number.$$

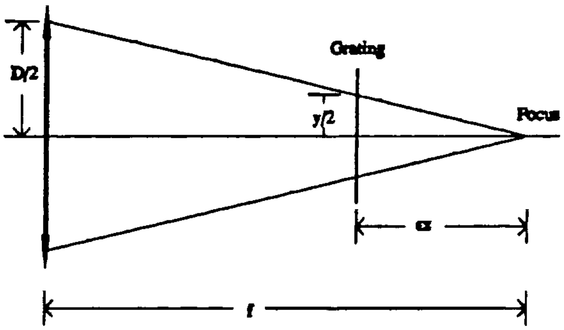


Figure 5 Geometry to determine tilt fringes

The W_{fixed} term is the initial aberration measured in the x-slope or reference channel while $W_{variable}$ is the initial aberration measured in the y-slope or test channel of the two-channel Ronchi system.

3.1 Rotation about the optic axis

The first degree of freedom tested is a rotation about the optic axis. The fixed x-slope grating and the variable y-slope grating are at the focal plane of the system and are perpendicular to the optic axis but they are not orthogonal to each other. The y-slope grating is rotated by an angle θ. This rotation reduces the amount of y-tilt in the y-slope map by cos(θ) and increases the amount of x-tilt in the y-slope map by sin(θ). Therefore a measurement error is introduced into the C1 and C2 terms in Table 2.

$$C1 = \frac{\epsilon_z \nu}{f/\#} \sin(\theta) \qquad C2 = \frac{\epsilon_z \nu}{f/\#} \cos(\theta) \tag{13}$$

To determine the effect of such an error, the relationships in Table 2 and Table 3 were simulated with a spreadsheet program. The B1 term was set equal to the expression given in equation 12 (ideal case), the C1 and C2 terms were set equal to the expressions given in equation 13 (error case), and all the other B and C terms were set to zero. The misalignment causes an error in the amount of Seidel defocus and astigmatism measured. The error in the astigmatism is twice the defocus error. The results of the theoretical model for the Seidel astigmatism is shown in Fig. 6. Figure 6 compares the magnitude of the error for different grating frequencies and system f/#'s.

Generally, a misalignment is not critical until it alters the amount of aberration measured by a quarter of a wave. The sensitivity of the Ronchi test to non-orthogonal gratings is directly proportional to the grating frequency and inversely proportional to the system f/#. Figure 7 compares the experimental results with the theoretical model for non-orthogonal gratings. The experimental data agrees with the theoretical prediction. A rotation of ±4° is acceptable for the f/2.25 system while a 50 lp/in grating while only a rotation of ±2° is acceptable for the same system with a 100 lp/in grating.

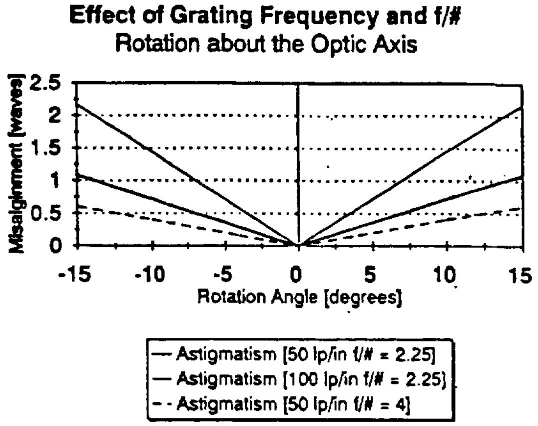


Figure 6 Comparison of different grating frequencies and system f/#'s

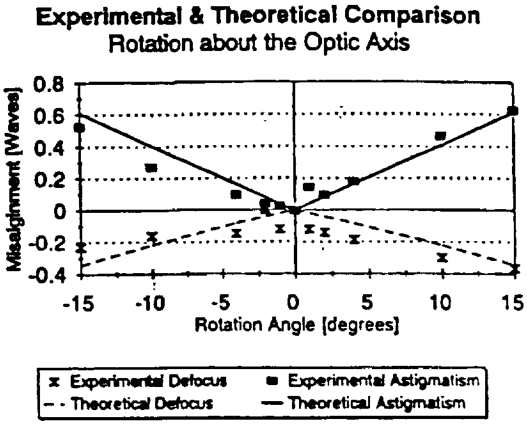


Figure 7 Comparison of experimental and theoretical data

3.2 Rotation parallel to the grating structure

To test the effect of rotation parallel to the grating structure the fixed and variable grating are at the focal plane of the system and are orthogonal to each other but the variable grating is not perpendicular to the optic axis. As the grating is tilted the number of fringes in the test channel Ronchigram increase. This misalignment alters only the C1 term not the C2 term. The theoretical model for this misalignment is derived from the geometry shown in Fig. 8. Figure 8 shows the wavefront leaving the collection lens and coming to a focus. A grating is shown in its aligned position and then in its misaligned position.

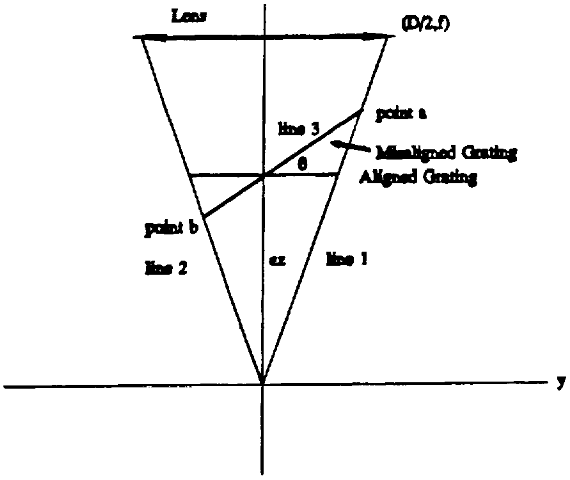


Figure 8 Geometry for rotation parallel to grating structure

In order to determine the number of fringes in the Ronchigram, it is necessary to know the amount of the grating which is illuminated. This is determined by calculating the length of the line segment ab in Fig. 8. The equations for lines 1, 2, and 3 are:

$$\text{Line 1} \qquad z = 2\, f/\#\, y \tag{14}$$

$$\text{Line2} \qquad z = -2\, f/\#\, y \tag{15}$$

$$\text{Line 3} \qquad z = \tan(\theta)\, y + \epsilon_z \tag{16}$$

The coordinates of points a and b can be found by computing the intersection of the appropriate lines. The number of fringes produced in the Ronchigram from the line segment ab is:

$$C2 = \left[\left[\left[\epsilon_z\left[\frac{1}{2f/\# - \tan(\theta)} + \frac{1}{2f/\# + \tan(\theta)}\right]\right]^2 + \left[2f/\#\epsilon_z\left[\frac{1}{2f/\# - \tan(\theta)} - \frac{1}{2f/\# + \tan(\theta)}\right]\right]^2\right]^{\frac{1}{2}}\right]\nu \tag{17}$$

Theoretical models were generated with the C2 term equal to the expression given in equation 17. The only Seidel aberration that was affected by this misalignment was astigmatism. Figure 9 compares the amount of aberration introduced for various grating frequencies and system f/#'s. Once again the amount of aberration introduced is directly proportional to the frequency of the grating and inversely proportional to the system f/#. However, a change in the system f/# will cause a slightly greater change in the amount of aberration introduced than a change in the grating frequency would. Figure 10 shows the comparison of the experimental results and the appropriate theoretical curve for the system shown in Fig. 4. Both curves have the same form although the theoretical curve is slightly higher than the experimental curve. This can be accounted for in the inaccuracy of the measurement of the system f/#. The theoretical defocus curve lies upon the x-axis.

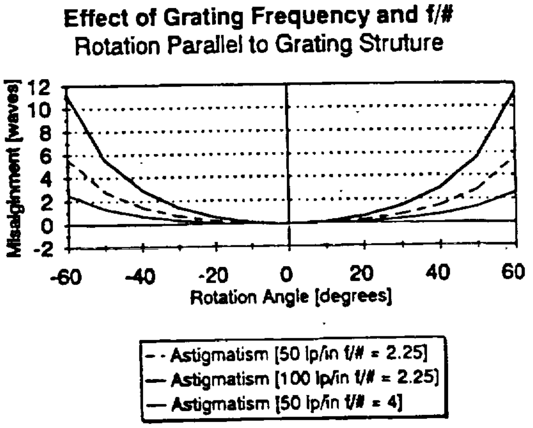


Figure 9 Comparison of different grating frequencies and system f/#'s

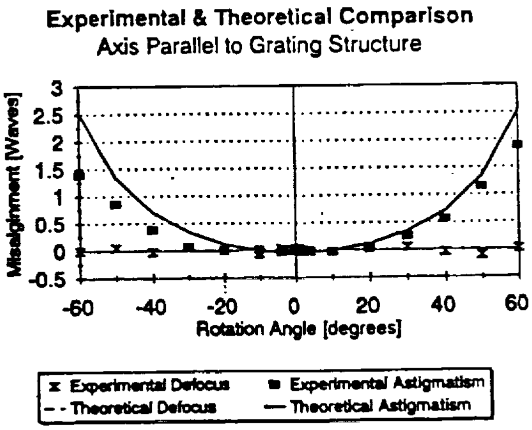


Figure 10 Comparison of experimental and theoretical data

3.3 Rotation perpendicular to the grating structure

A theoretical model for the misalignment of the test grating perpendicular to the grating structure has not yet been developed. The area of the grating that is illuminated is identical to the area of the grating that is illuminated in the case of rotation parallel to the grating structure. The difficulty is that rotating the grating in this manner causes the fringes to "fan out", expanding

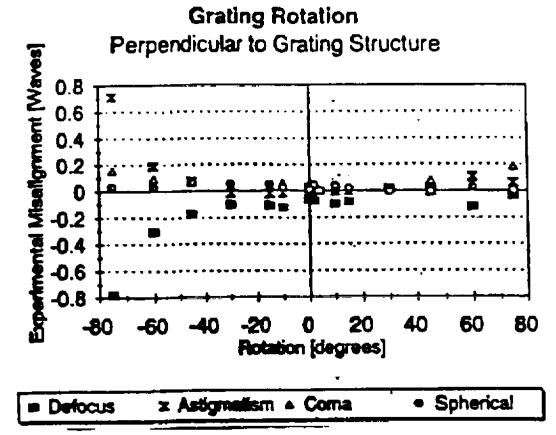


Figure 11 Experimental results for rotation perpendicular to the grating structure

on one side of the Ronchigram and compressing on the other side. Figure 11 shows the experimental results. This misalignment does not have a significant effect on the performance of the system shown in Fig. 4.

3.4 Translation along the optic axis

The fixed and variable gratings are perpendicular to the optic axis and are orthogonal to each other, but the test grating is displaced along the optic axis. Displacing the grating towards focus reduces the number of fringes that are present in the Ronchigram while displacing the grating further away from focus increases the number of fringes. This misalignment introduces an error into the C2 term.

$$C2 = \frac{\epsilon_z \nu}{f/\#} + \frac{\delta\epsilon_z \nu}{f/\#} \qquad (18)$$

where $\delta\epsilon_z$ = *the amount of translation along the optic axis.*

This error should affect both the Seidel defocus and astigmatism coefficients. Translating the grating further from focus will cause an increase in the amount of astigmatism measured while not changing the amount of defocus. This is because the variable channel Ronchigram will have more fringes than the fixed channel Ronchigram. The software characterizes the additional fringes as astigmatism. Translating the grating towards focus will increase the amount of astigmatism and decrease the amount of defocus measured. This is because the test channel Ronchigram will have fewer fringes than the reference Ronchigram.

Figure 12 compares the effect of this misalignment on systems of various f/#'s and grating frequencies. The magnitude of the error is directly proportional to the grating frequency and inversely proportional to the f/# of the optical system. Figure 13 shows the agreement between the experimental results for the system shown in Fig. 4 and the theoretical model for that system. A translation misalignment along the optic axis becomes important for the f/2.25 system at ±0.5 mm for a 50 lp/in grating and ±0.25 mm for a 100 lp/in grating.

3.5 Translation perpendicular to the grating structure

Displacing the test grating perpendicular to its grating structure relative to the reference grating has no effect on the spacing of the fringes in the Ronchigrams. However if this displacement can be performed in a controlled periodic motion, for instance with a piezo-electric transducer, then it can be used to implement phase measuring techniques.[8]

3.6 Translation parallel to the grating structure

Translation parallel to the grating structure does not have any effect on the spacing or orientation of the fringes in the Ronchigrams.

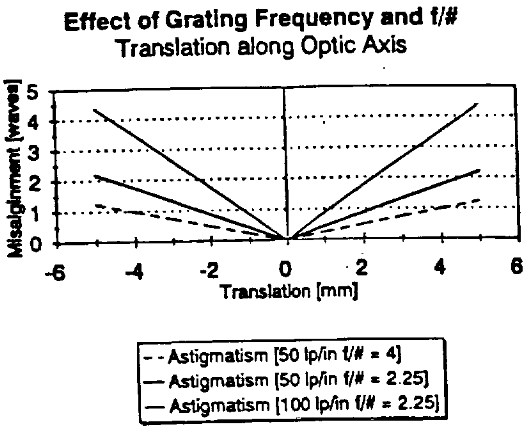


Figure 12 Comparison of different grating frequencies and system f/#'s

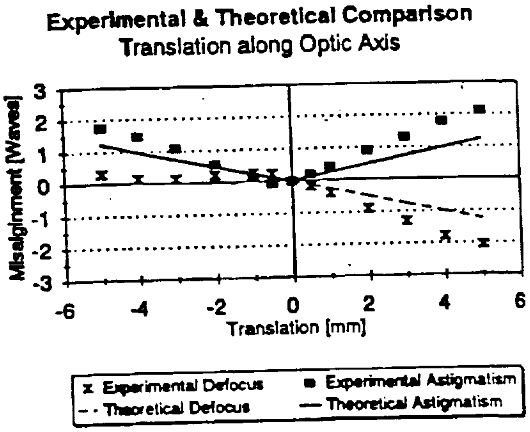


Figure 13 Comparison of experimental and theoretical data

4. CONCLUSIONS

One method to fulfill the requirement of two orthogonal slope maps to reconstruct a surface is to use a two-channel Ronchi instrument. The effect of different grating misalignments were investigated. The results presented are based on the use of Zernike polynomials to fit the slope data and then using linear combinations of those Zernikes to reconstruct the surface shape. Of the six possible misalignments only four had an effect on the performance of the two-channel Ronchi instrument, the three rotational misalignments and the translation along the optic axis.

The severity of these misalignments is dependant upon the frequency of the grating used in the test and on the f/# of the system. Fast optical systems with higher frequency gratings require more care in the alignment of the gratings than slower optical systems with low frequency gratings. The misalignments become critical when they alter the amount of aberration measured by a quarter of a wave, $\lambda_{eq}/4$. The point at which each misalignment becomes critical depends upon the characteristics of the system and grating being used. The two most critical misalignments, independent of system specifications, are translation along the optic axis and rotation about the optic axis.

5. ACKNOWLEDGEMENTS

This work is funded by a NASA research grant (NAG3-1300) from the NASA Lewis Research Center. A special thanks to the people of NASA Lewis and the Ohio Aerospace Institute, especially Alexander Pline, for the support they have provided this project.

6. REFERENCES

1. W. S. Meyers and H. P. Stahl, "Contouring of a free oil surface," *SPIE Conference Proceedings*, **V 1755**, July 1992.

2. W. H. Southwell, "Wave-front estimation from wave-front slope measurements," *Journal of the Optical Society of America*, **V 70** No. 8, pp 998-1006, August 1990.

3. M. P. Rimmer, "Method for Evaluating Lateral Shearing Interferograms," *Applied Optics*, V 13 No. 3, pp 623-629, March 1974.

4. N. H. Davis and T. A. Fritz, "Application of Zernike Polynomials to Reduction of Wavefront-Slope Data," FRINGE User's Manual. University of Arizona, 1982.

5. D. J. Fischer and H. P. Stahl, "A vector formulation for Ronchi shear surface fitting," *SPIE Conference Proceedings*, V **1779**, July 1992.

6. K. R. Freischlad and C. L. Koliopoulos, "Modal estimation of a wave front from difference measurements using the discrete Fourier transform," *Journal of the Optical Society of America*, V **3** No. 11, pp 1852-1861, November 1986

7. F. Roddier and C. Roddier, "Wavefront reconstruction using iterative Fourier transforms," *Applied Optics*, V **30** No. 11, pp 1325-1327, April 10, 1991.

8. T. Yatagai, "Fringe scanning Ronchi test for aspherical surfaces," *Applied Optics*, V **23**, pp 3676-3679, October 15, 1984.

Vector formulation for interferogram surface fitting

David J. Fischer, John T. O'Bryan, Robert Lopez, and H. Philip Stahl

Interferometry is an optical testing technique that quantifies the optical path difference (OPD) between a reference wave front and a test wave front based on the interference of light. Fringes are formed when the OPD is an integral multiple of the illuminating wavelength. The resultant two-dimensional pattern is called an interferogram. The function of any interferogram analysis program is to extract this OPD and to produce a representation of the test wave front (or surface). This is accomplished through a three-step process of sampling, ordering, and fitting. We develop a generalized linear-algebra vector-notation model of the interferogram sampling and fitting process.

Key words: Interferometry, polynomial fitting, optical testing.

1. Introduction

Whenever two coherent wave fronts exist at the same point in space at the same time, they superimpose. The irradiance value at that point is determined by the optical path difference (OPD) between the two wave fronts. Interferometry uses this phenomenon to quantify the shape of a given surface or wave front relative to some reference. The two-dimensional irradiance pattern created by the OPD between the test and reference wave fronts at all points is commonly called an interferogram. The challenge is to extract a representation of the test surface (or wave front) from such an interferogram (assuming that we have perfect knowledge of the reference surface or wave front). This process is accomplished by a three-step process: data sampling, ordering, and fitting.

While the sampling and ordering steps can be accomplished by a number of different methods, fringe digitization, phase-measuring interferometry, or Fourier decomposition, to produce a convenient representation of the surface under test, this sampled and ordered data (regardless of how they were acquired) are generally fit to some polynomial set (such as the ubiquitous Zernikes). Fringe digitization software has been available to accomplish this process since the 1960's[1] and has been discussed at length for Zernike polynomials by using conventional summation notation with some matrix notation.[2–5]

Zernike polynomials are used because they form a basis set for circular apertures. But they are not a basis set for noncircular apertures. In these cases other polynomial sets, such as Zernike–Tatian[6] or Zernike–Mahajan[7,8] polynomials (for annulus apertures) and Legendre[9] polynomials (for rectangular apertures), may provide better results. Ideally an interferogram analysis software package should allow for easy substitution between these and other polynomial sets; however, most programs are developed exclusively for Zernike polynomials. The first step in developing a more general software package capable of using any basis set is to develop a general mathematical model of the fitting process, completely independent of any specific polynomial set.

In this paper we review the interferogram analysis process (sampling, ordering, and fitting) and develop a linear-algebra vector notation that clearly and concisely describes the interferogram sampling and fitting process. The development is completely general (i.e., not limited to Zernike polynomials) and assumes only that the polynomials form a basis set. The approach performs a least-squares fit of the data to the polynomial set, orthogonalizes this result with a Gram–Schmidt algorithm, and solves for the final weighting coefficients of the initial polynomial set with a backsubstitution technique.

2. Interferogram Analysis

Because an interferogram has nearly infinite resolution, the actual number of points that can be measured is limited only by the sampling technique. There are two main methods for acquiring the data: digitization and phase-measuring interferometry. With digitization, data points are acquired by sampling along the center of the light or dark fringes. Usually the interferogram is photographed, and the

---

The authors are with the Rose-Hulman Institute of Technology, 5500 Wabash Avenue, Terre Haute, Indiana 47803.

Received 9 July 1992.
0003-6935/93/254738-06$06.00/0.

image is placed on a digitizing tablet where the data points are digitized by a human operator. Alternatively, the image can be acquired with a video camera and digitized by image-processing techniques. Phase-measuring interferometry by contrast is an electronic process where a video camera samples the wave-front phase at every point within its detector array—thus creating a high-density, uniform grid of data. An illustration of the difference in the two techniques is shown in Figs. 1 and 2. In either case the function of the sampling process is to acquire data in the form m, x, y, where x, y gives the position of the point and m is the order of the OPD at that point.

Once the interferogram has been sampled, the acquired data must be ordered. This is because interferometry measures the wave-front phase ϕ (or OPD) in units of modulo 2π (or modulo λ). Thus, even though a point in the interferogram represents a quantity greater than 2π, no measured quantity is larger than 2π. For example, an x, y point may have a phase of 9π, but the measured m will equal 0.5, which is found by taking the magnitude modulo 2π. Therefore the phase must be artificially reconstructed by ordering the data. This is done by assigning integer order number values to each series of measured data points.

For manual or video digitization, data points along a given fringe all have the same order, and each adjacent fringe has an incrementally larger order number. Typically such data consist of integer m values, regardless of whether one digitizes bright or dark fringes. Assignment of these values is accomplished by the human operator, in the case of manual digitization, or the video digitization software itself. Alternatively, a phase-measuring interferometer, which is not confined to sampling only along a given fringe, measures fractional values of m. However, its data are still limited by the modulo 2π condition and must be ordered by a process called phase unwrapping.

To illustrate the sampling and ordering process, consider a one-dimensional parabolic wave front interfered with a planar wave front to produce an irradiance pattern (Fig. 3). From the irradiance profile it is easy to visualize how an interferogram measures the total phase modulo 2π. Bright fringes occur for even multiples of π (i.e., $\phi = 0, 2\pi, 4\pi, \ldots$), and dark fringes occur for odd multiples of π (i.e., $\phi = \pi, 3\pi, 5\pi, \ldots$). If the pattern is sampled on each dark fringe, all the m values will be 0.5. In the typical ordering process, these fringes will be assigned values of $m = 1, 2, 3, 4, 5$. Clearly the result represents a parabolic wave front (Fig. 4), but there is a constant phase offset between the original wave front and the reconstructed data. To extract the original wave front, it is necessary to fit the ordered data to a parabolic function. Furthermore, if this were a real wave front with aberrations, the only way to determine their contribution is to perform a general polynomial fit.

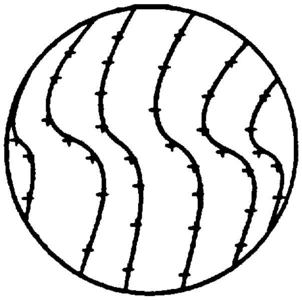


Fig. 1. When an interferogram is sampled by digitization, data points are usually placed along each fringe (either bright or dark).

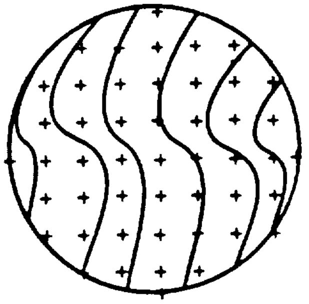


Fig. 2. When an interferogram is sampled by phase-measuring interferometry, data points are placed at each detector location. This usually results in a uniform high-density sampling of the pupil.

Fitting is typically accomplished by a least-squares method with a Gram–Schmidt orthogonalization. Traditionally interferograms are fit to Zernike polynomials for several reasons. Zernikes look like classical Seidel aberrations. They provide good results for the circular aperture interferogram of most optical components. And, since they are orthogonal, they are easy to manipulate. Thus it is possible to add or subtract coefficients without seriously affecting the quality of the fit. In general these coefficients provide information about how flat the surface is, whether

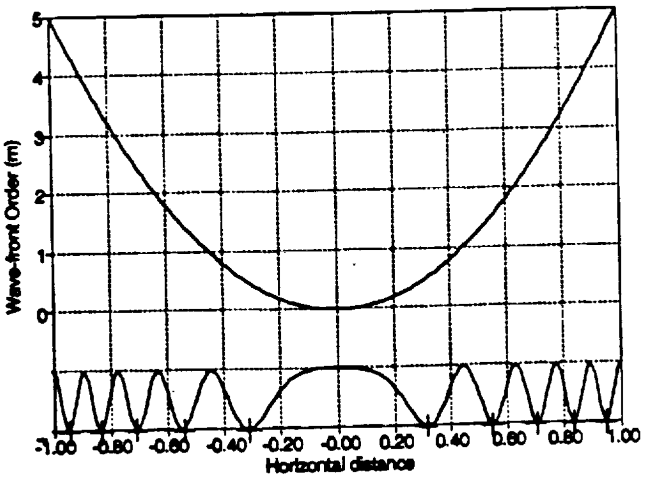


Fig. 3. Top curve, One-dimensional slice of a parabolic wave front. Bottom curve, Resultant irradiance pattern formed when this slice interferes with a plane wave front. The plusses along the bottom of the graph represent data points sampled at irradiance minima.

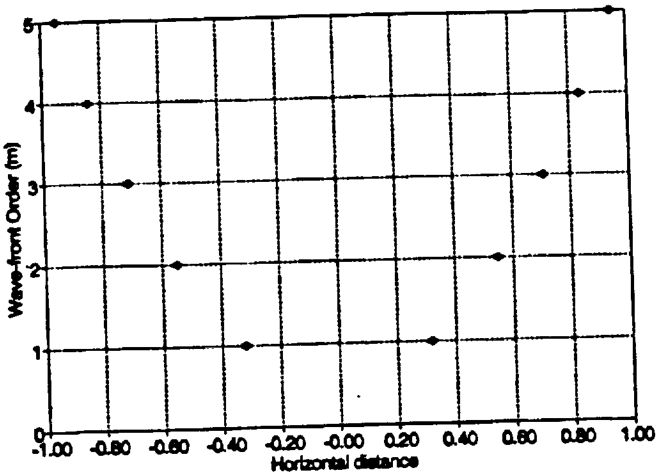


Fig. 4. When plotted, the ordered sample points clearly represent a parabolic shape.

a feature is a bump or hole, whether the part is concave or convex, and what aberrations are present.

3. Vector Formulation

A. Ideal Surface

The surface height function at any point in the exit pupil $Z(x, y)$ can be represented by a linear combination of M polynomials $F(x, y)$ and their weighting coefficients G:

$$Z(x, y) = \sum_{j=1}^{M} F_j(x, y) G_j, \tag{1}$$

where the polynomials are typically a standard set such as the Zernike polynomials. If the coefficients are known, the surface function at any given point is found by summing the product of each polynomial multiplied by its coefficient. However, in optical testing the coefficients are not known. Thus to determine a representation of the surface function one samples and orders the wave front at many discrete points, fits these data to a polynomial set, and solves numerically for the coefficients.

B. Discrete Data

As discussed in Section 2, the first step in interferogram analysis is to sample (and order) the continuous surface function producing a set of N discrete data points. With Eq. (1) each sampled point in the pupil (x_r, y_r) has a measured value Z_r (where r is the sample index) that can be represented by a linear combination of M polynomials and coefficients:

$$Z_r(x_r, y_r) = \sum_{j=1}^{M} F_j(x_r, y_r) G_j, \tag{2}$$

where $N > M$. Alternatively, the value at each sample point can be expressed as a matrix of M polynomials multiplied by a matrix of M coefficients:

$$[Z_r] = [F_1(x_r, y_r), \ldots, F_M(x_r, y_r)] \begin{bmatrix} G_1 \\ \vdots \\ G_M \end{bmatrix}. \tag{3}$$

Now, if the sampled and ordered data are thought of as a sequence of N points,

$$\begin{bmatrix} Z_1 \\ \vdots \\ Z_N \end{bmatrix}, \tag{4}$$

the complete set of all N data points can be represented as a matrix of polynomials multiplied by the coefficient matrix:

$$\begin{bmatrix} Z_1 \\ \vdots \\ Z_N \end{bmatrix} = \begin{bmatrix} F_1(x_1, y_1), & \ldots, & F_M(x_1, y_1) \\ \vdots & & \vdots \\ F_1(x_N, y_N), & \ldots, & F_M(x_N, y_N) \end{bmatrix} \begin{bmatrix} G_1 \\ \vdots \\ G_M \end{bmatrix}. \tag{5}$$

To simplify this notation, we define three vectors. First, the measured values are defined as a vector with N rows. Second, each polynomial F_j evaluated over all N data points is a vector with N rows. Third, the M coefficients are a vector with M rows:

$$\mathbf{Z} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_N \end{bmatrix}, \quad \mathbf{F}_j = \begin{bmatrix} F_j(x_1, y_1) \\ \vdots \\ F_j(x_N, y_N) \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} G_1 \\ \vdots \\ G_M \end{bmatrix}. \tag{6}$$

With this notation, the sampled data can be represented by

$$\mathbf{Z} = [\mathbf{F}_1, \ldots, \mathbf{F}_M]\mathbf{G}. \tag{7}$$

C. Fitting to the Polynomials

The second step in solving for the polynomial coefficients is to fit the N data points to the polynomial set. This is typically accomplished with a least-squares method, where S is defined as the sum of the square of the difference between the data points and the fitted polynomials. With the vector notation, S is given by

$$S \equiv |[\mathbf{Z} - [\mathbf{F}_1, \ldots, \mathbf{F}_M]\mathbf{G}]|^2, \tag{8}$$

or, as more commonly expressed by summation notation,

$$S \equiv \sum_{r=1}^{N} \left[Z_r - \sum_{j=1}^{M} G_j F_j(x_r, y_r) \right]^2. \tag{9}$$

If a perfect fit were possible, S would be equal to zero, because there would be no difference between the measured values and the representing polynomials evaluated at the corresponding points. However, there are always differences between the real surface and its representation. Therefore S is nonzero.

The function of a least-squares fit is to find the coefficients for a given set of polynomials, which minimizes S.

The coefficients are found by taking the derivative of S with respect to each coefficient and setting the result equal to zero. For example, minimizing S for G_k gives

$$\sum_{r=1}^{N} Z_r F_k(x_r, y_r) = \sum_{r=1}^{N} \sum_{j=1}^{M} G_j F_j(x_r, y_r) F_k(x_r, y_r). \quad (10)$$

The result of minimizing S for all coefficients can be expressed by using summations in matrix form as

$$\begin{bmatrix} \sum_{r=1}^{N} Z_r F_1(x_r, y_r) \\ \vdots \\ \sum_{r=1}^{N} Z_r F_M(x_r, y_r) \end{bmatrix} = \begin{bmatrix} \sum_{r=1}^{N} F_1(x_r, y_r) F_1(x_r, y_r) \ldots \sum_{r=1}^{N} F_M(x_r, y_r) F_1(x_r, y_r) \\ \vdots \qquad\qquad \vdots \\ \sum_{r=1}^{N} F_1(x_r, y_r) F_M(x_r, y_r) \ldots \sum_{r=1}^{N} F_M(x_r, y_r) F_M(x_r, y_r) \end{bmatrix} \times \begin{bmatrix} G_1 \\ \vdots \\ G_M \end{bmatrix}. \quad (11)$$

This expression can be greatly simplified with vector notation:

$$\begin{bmatrix} \mathbf{Z} \cdot \mathbf{F}_1 \\ \vdots \\ \mathbf{Z} \cdot \mathbf{F}_M \end{bmatrix} = \begin{bmatrix} \mathbf{F}_1 \cdot \mathbf{F}_1 \ldots \mathbf{F}_1 \cdot \mathbf{F}_M \\ \vdots \qquad \vdots \\ \mathbf{F}_M \cdot \mathbf{F}_1 \ldots \mathbf{F}_M \cdot \mathbf{F}_M \end{bmatrix} \mathbf{G}, \quad (12)$$

where the dot product of the vectors $\mathbf{F}_i$ and $\mathbf{F}_j$ gives the sum of the products of both polynomials over every data point:

$$\mathbf{F}_i \cdot \mathbf{F}_j = \sum_{r=1}^{N} F_i(x_r, y_r) F_j(x_r, y_r). \quad (13)$$

And the dot product of the vectors $\mathbf{Z}$ and $\mathbf{F}_j$ gives the sum of the products evaluated at all data points:

$$\mathbf{Z} \cdot \mathbf{F}_j = \sum_{r=1}^{N} Z_r F_j(x_r, y_r). \quad (14)$$

D. Solving for the Coefficients Assuming Orthogonal Polynomial Vectors

If the $\mathbf{F}$ vectors are orthogonal over the sampled data points, the minimization process produces a diagonal matrix,

$$\begin{bmatrix} \mathbf{Z} \cdot \mathbf{F}_1 \\ \vdots \\ \mathbf{Z} \cdot \mathbf{F}_M \end{bmatrix} = \begin{bmatrix} \mathbf{F}_1 \cdot \mathbf{F}_1 \ldots & 0 \\ \vdots & \vdots \\ 0 & \ldots \mathbf{F}_M \cdot \mathbf{F}_M \end{bmatrix} \mathbf{G}, \quad (15)$$

and the coefficients can be found by multiplying both sides of the equation by the inverse of the diagonal matrix:

$$\mathbf{G} = \begin{bmatrix} \dfrac{\mathbf{Z} \cdot \mathbf{F}_1}{\|\mathbf{F}_1\|^2} \\ \vdots \\ \dfrac{\mathbf{Z} \cdot \mathbf{F}_M}{\|\mathbf{F}_M\|^2} \end{bmatrix}. \quad (16)$$

However, most orthogonal polynomials are only orthogonal over a continuous range, not over discrete points. Thus the solution process is not as easy as it seems at first. To overcome this limitation, there are three possible approaches. The first is to assume that the polynomials are orthogonal. This may work if there is enough sampled data to approximate a continuous range, but it can introduce error. The second is to take the inverse of the nondiagonal matrix, but this can be numerically difficult and inaccurate. Third, the polynomials can be orthogonalized. Typically, this is accomplished with a Gram–Schmidt technique.

4. Gram–Schmidt Orthogonalization

A. Gram–Schmidt Orthogonalization Process

The Gram–Schmidt orthogonalization process takes M arbitrary vectors in an M-dimensional space (M space) and creates M orthogonal vectors that form a basis for that space. A basis for an M space is essentially a coordinate system for that space. Since the basis vectors are linearly independent, all other vectors in that space may be formed from linear combinations of the basis vectors.

To illustrate the Gram–Schmidt process, we consider two arbitrary vectors $\mathbf{F}_1$ and $\mathbf{F}_2$ (Fig. 5), which are to be orthogonalized to produce new orthogonal vectors $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ (Fig. 6). To begin the orthogonalization process, we define the first orthogonal vector $\mathbf{\Phi}_1$ to be the first arbitrary vector $\mathbf{F}_1$. Now, with the first vector orthogonalized, all other vectors can be orthogonalized by projecting the current vector onto each preceding orthogonalized vector (Fig. 7). Each

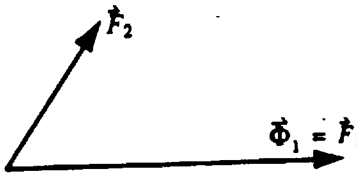


Fig. 5. Two arbitrary vectors in a two-dimensional space. The first vector is defined to be the first orthogonal vector of the space.

Fig. 6. Two orthogonal vectors in a two-dimensional space. The function of the Gram–Schmidt orthogonalization process is to create these two vectors.

projection is subtracted from the current vector, leaving the orthogonal component. This process is continued for the rest of the vectors, each time subtracting off the projections of the current vector onto all previous vectors. This creates a set of **M**, orthogonal vectors, which replaces the original set of vectors.

In general, the Gram–Schmidt orthogonalization process can be expressed for any vector $\mathbf{F}_j$ as

$$\mathbf{\Phi}_j = \mathbf{F}_j - \sum_{s=1}^{j-1} \frac{\mathbf{F}_j \cdot \mathbf{\Phi}_s}{\mathbf{\Phi}_s \cdot \mathbf{\Phi}_s} \mathbf{\Phi}_s. \tag{17}$$

By letting D_{js}, represent the coefficient of $\mathbf{\Phi}_s$,

$$D_{js} = \frac{F_s \cdot \mathbf{\Phi}_s}{\mathbf{\Phi}_s \cdot \mathbf{\Phi}_s}, \tag{18}$$

the orthogonalization of any vector $\mathbf{F}_j$ can be expressed as

$$\mathbf{\Phi}_s = \mathbf{F}_j - \sum_{s=1}^{j-1} D_{js} \mathbf{\Phi}_s. \tag{19}$$

These orthogonalized vectors are now substituted into the fitting process, replacing the original arbitrary polynomial vectors.

B. Applying the Gram–Schmidt Process to the Vector Formulation

From the fitting process, the polynomial vectors, $\mathbf{F}_1$ through $\mathbf{F}_M$ [Eq. (7)], are orthogonalized and replaced by the new orthogonal vectors, Φ_1 through Φ_M, resulting in the calculation of a new and different set of weighting coefficients. These new coefficients are named Γ_j for j in 1–M. Therefore the original coefficient vector **G** is replaced by the vector **Γ**. The original measured surface function values are now fitted in terms of the new vectors:

$$\mathbf{Z} = [\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_M]\mathbf{\Gamma}. \tag{20}$$

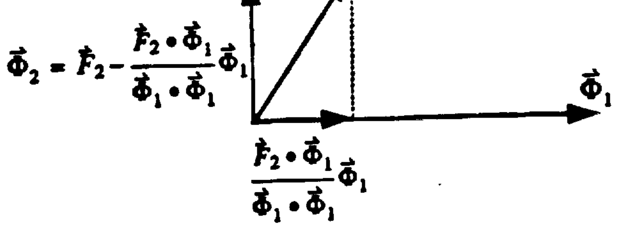


Fig. 7. Gram–Schmidt orthogonalization process, which subtracts the projection of the second vector onto the preceeding vector to create an orthogonal vector.

The system is then solved, as shown above, by multiplying both sides of the minimized least-squares fit equation by the inverse of the resultant diagonal matrix. The Γ coefficients are now known and can be used to represent the original surface:

$$\mathbf{\Gamma} = \begin{bmatrix} \dfrac{\mathbf{Z} \cdot \mathbf{\Phi}_1}{\|\mathbf{\Phi}_1\|^2} \\ \vdots \\ \dfrac{\mathbf{Z} \cdot \mathbf{\Phi}_M}{\|\mathbf{\Phi}_M\|^2} \end{bmatrix}. \tag{21}$$

5. Finding G in Terms of Γ by Backsubstitution

The polynomials fitted to the measured values are typically a standard set, such as the Zernike polynomials. Therefore, since the polynomials do not change, the coefficients found in the fitting process of the original polynomials can be used to describe the test surface and serve as a basis of comparison between other test surfaces. However, when the $\mathbf{F}_1$ through $\mathbf{F}_M$ vectors were orthogonalized, the polynomials F_1 through F_M were modified, creating a new set of polynomials, Φ_1 through Φ_M, that are dependent on the measured values. Thus the Γ coefficients that compose **Γ** cannot serve as a basis of comparison, since every set of Γ coefficients is related to a different set of polynomials. Because of this, it is necessary to find a means to convert **Γ** to **G**.

The first step in finding **G** in terms of **Γ** is to express the orthogonalization process by using matrix notation. From Eq. (19) the orthogonalization of all **M** vectors can be shown as follows:

$$\begin{bmatrix} \mathbf{\Phi}_1{}^T \\ \vdots \\ \mathbf{\Phi}_M{}^T \end{bmatrix} = \begin{bmatrix} \mathbf{F}_1{}^T \\ \vdots \\ \mathbf{F}_M{}^T \end{bmatrix} - \begin{bmatrix} 0 & \cdots & & & 0 \\ D_{21} & 0 & \cdots & & 0 \\ D_{31} & D_{32} & 0 & \cdots & 0 \\ \vdots & \vdots & & & \vdots \\ D_{M1} & D_{M2} & \cdots & D_{MM-1} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{\Phi}_1{}^T \\ \vdots \\ \mathbf{\Phi}_M{}^T \end{bmatrix}. \tag{22}$$

Calling the matrix of the D_{js} coefficients $\mathscr{D}$ for simplicity, we can solve Eq. (22) for the orthogonal vectors, where $\mathscr{I}$ is an identity matrix of size $M \times M$:

$$\begin{bmatrix} \mathbf{\Phi}_1{}^T \\ \vdots \\ \mathbf{\Phi}_M{}^T \end{bmatrix} = (\mathscr{I} + \mathscr{D})^{-1} \begin{bmatrix} \mathbf{F}_1{}^T \\ \vdots \\ \mathbf{F}_M{}^T \end{bmatrix}. \tag{23}$$

Recalling Eqs. (7) and (20), we can clearly see that two equivalent representations of the surface exist, one using the original polynomial vectors and the other using the orthogonalized polynomial vectors. Therefore the following statement can be made:

$$[\mathbf{F}_1, \ldots, \mathbf{F}_M]\mathbf{G} = [\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_M]\mathbf{\Gamma}. \qquad (24)$$

Now, if we rewrite Eq. (23) using transposes

$$[\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_M]^T = (\mathcal{J} + \mathcal{D})^{-1}[\mathbf{F}_1, \ldots, \mathbf{F}_M]^T, \qquad (25)$$

and the transpose of both sides of the resultant equation is taken, an expression for the matrix of $\mathbf{\Phi}_1$ through $\mathbf{\Phi}_M$ in terms of $\mathbf{F}_1$ through $\mathbf{F}_M$ is produced:

$$[\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_M] = [\mathbf{F}_1, \ldots, \mathbf{F}_M][(\mathcal{J} + \mathcal{D})^{-1}]^T. \qquad (26)$$

Substituting Eq. (26) into Eq. (24),

$$[\mathbf{F}_1, \ldots, \mathbf{F}_M]\mathbf{G} = [\mathbf{F}_1, \ldots, \mathbf{F}_M][(\mathcal{J} + \mathcal{D})^{-1}]^T\mathbf{\Gamma}, \qquad (27)$$

and dividing both sides by the matrix of the polynomial vectors provide the desired solution of $\mathbf{G}$ in terms of the D_{js} coefficients and $\mathbf{\Gamma}$:

$$\mathbf{G} = [(\mathcal{J} + \mathcal{D})^{-1}]^T\mathbf{\Gamma}. \qquad (28)$$

6. Conclusion

Interferometry is an optical testing technique in which information about a test surface is extracted from an interferogram. This is accomplished with a three-step process. First, the data are sampled, usually by one of two main methods: digitization or phase-measuring interferometry. Second, the sampled data points are ordered. Third, the ordered data are fit to a polynomial set. Typically, Zernike polynomials are used because they form a basis set for circular apertures. However, they are not a basis set for noncircular apertures. In these cases, other polynomial sets such as Zernike–Mahajan (for annulus apertures) and Legendre (for rectangular apertures) may provide better results. In this paper we developed a completely general linear-algebra vector-notation model to describe the polynomial fitting process independent of the polynomial set. The basic process is to describe the sampled data as a linear vector: Fit that data to an arbitrary polynomial set by using a least-squares method. Orthogonalize the polynomials to the data vector by using a Gram–Schmidt technique. Solve for the original polynomials weighting coefficients by using a backsubstitution technique.

From an implementation point of view, by treating the polynomial set as a vector, it is possible to write a algorithm that has no knowledge of the actual polynomial set being fitted. Thus it is simple to change between polynomial sets for various aperture shapes. Also, it is not necessary to actually solve for the orthogonal polynomial coefficients to arrive at the original polynomial weighting coefficients by using the backsubstitute technique.

This project was supported in part by NASA research grant NA63-1300 from NASA-Lewis Research Center, Cleveland, Ohio, and was a result of H. Philip Stahl's participation in the American Society of Engineering Educators/NASA Summer Faculty Fellowship Program. Special thanks is given to Alex Pline of NASA-Lewis.

References

1. M. P. Rimmer, C. M. King, and D. G. Fox, "Computer program for the analysis of interferometric test data," Appl. Opt. **11**, 2790–2796 (1972).
2. J. Y. Wang and D. E. Silva, "Wave-front interpretation with Zernike polynomials," Appl. Opt. **19**, 1510–1518 (1980).
3. Cheol-Jung Kim, "Polynomial fit of interferograms," Ph.D. dissertation (Optical Sciences Center, University of Arizona, Tucson, Ariz., 1982).
4. D. Malacara, J. M. Caprio-Valadez, and J. J. Sanchez-Mondragon, "Wave-front fitting with discrete orthogonal polynomials in a unit radius circle," Opt. Eng. **24**, 672–675 (1990).
5. J. L. Lewis, W. P. Kuhn, and H. P. Stahl, "The evaluation of a random sampling error on the polynomial fit of subaperture test data," in *Optical Testing and Metrology II*, C. P. Grover, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **954**, 88–94 (1988).
6. W. H. Swantner and W. H. Lowrey, "Zernike–Tatian polynomials for interferogram reduction," Appl. Opt. **19**, 161–163 (1980).
7. V. N. Mahajan, "Zernike annular polynomials for imaging systems with annular pupils," J. Opt. Soc. Am. **71**, 75–85 (1981).
8. V. N. Mahajan, "Zernike annular polynomials for imaging systems with annual pupils: errata," J. Opt. Soc. Am. **71**, 1408 (1981).
9. J. L. Rayces, "Least-squares fitting of orthogonal polynomials to the wave-aberration function," Appl. Opt. **31**, 2223–2228 (1992).

Mechanism for surface fitting of interferometric slope data

David J. Fischer
H. Philip Stahl

Rose-Hulman Institute of Technology
Center for Applied Optics Studies
5500 East Wabash Avenue
Terre Haute, IN 47803
(812) 877-8253

ABSTRACT

One approach for obtaining a surface representation is to fit Zernike polynomials (in a least squares sense) to discrete data points in the full aperture. The mathematics for this have been described using both matrix[1] and vector notation[2]. Additionally, vector notation has been used to describe how to obtain a surface representation from orthogonal (x, y) slope data[3]. The result of that paper was a matrix operator for linearly combining the first eight Zernike polynomial coefficients fit to x- and y-slope data to produce a Zernike polynomial surface representation. This paper extends that process by presenting a systematic approach for obtaining the linear relationship between slope and surface using the first 49 Zernike polynomials.

1. INTRODUCTION

Full-aperture interferometry techniques are used to measure surface characteristics of a test object via non-destructive optical methods.[1] Traditionally, Zernike polynomials are fit (in a least squares sense) to the measured data to produce a standardized representation of the surface. The parameters which describe the surface's characteristics are, the weighting coefficient of the Zernike polynomials. This fitting process has been explained in detail using both matrix and vector analysis[1,2].

While surface interferometry is he simple to analyze, it is not always used; in some cases it is better to measure slope, such as the Ronchi test does. Because, these techniques yield slope data in a particular direction, to create a surface representation, it is necessary to measure slope in two orthogonal directions: in the x- and y-directions. The data is not fit to the derivatives of the Zernike polynomials because the derivatives are not orthogonal. Rather, to reconstruct the surface, these orthogonal slope data sets are fit to Zernike polynomials to create a representation of the test object's slope, and combined according to a linear operator.[3] This paper describes the process of creating a polynomial surface representation using the first 49 Zernike polynomials based on a least squares fit of Zernike polynomials to orthogonal slope data.

2. FITTING ZERNIKE POLYNOMIALS TO THEIR DERIVATIVES

The first step in creating a Zernike polynomial, surface representation based on fits of orthogonal slope data is to determine which linear combinations of Zernike polynomials will recreate the x- and y-derivatives of the Zernike polynomials. To do this, it is first necessary to examine the Zernike polynomials.

2.1 Description of Zernike polynomials in R terms

The Zernike polynomials are a basis set, orthogonal over a unit circle. There is an infinite number of these polynomials, since they are constructed from a generation formula, but in many applications only the first 36 or 49 polynomials are used. They are commonly used for fitting data from interferometric applications because they represent the Seidel aberrations well.[4] In this paper, two different notations will be used to represent the Zernike polynomials. They are the Z_n notation and the $R_r^{\pm m}$ notation. In the case of the Z_n notation, the subscript, n, indicates which polynomial is being represented. The subscript is a positive, integer value starting at zero. Thus, Z_{10} refers to the eleventh Zernike polynomial. For the $R_r^{\pm m}$ notation, r is the radial magnitude of the polynomial and m is the radial frequency. The sign of m also designates whether the trigonometric term involved in the Zernike polynomial is a sine or a cosine: $m \geq 0$ designates a cosine and $m < 0$ designates a sine. Finally, $r \geq 0$ always, and $r \geq |m|$ always. These cases are summarized below:

SPIE. Vol. 2003, 1993.

Z_n = a Zernike polynomial

n indicates polynomial referred to.

$R_r^{\pm m}$ = a Zernike polynomial

r = radial magnitude

The trigonometric term is $\begin{bmatrix} \sin(m\theta), m < 0 \\ \cos(m\theta), m \geq 0 \end{bmatrix}$

Figure 1. Notation of Zernike polynomials

In the remainder of this paper, the Zernikes will be considered using rectangular coordinates (x and y), not polar (r and θ). Examples of this notation follow:

$$R_4^{-2} = Z_{12} = (4r^2 - 3)r^2 \sin(2\theta) = 8x^3y + 8xy^3 - 6xy$$

$$R_3^3 = Z_9 = r^3 \cos(3\theta) = x^3 - 3xy^2$$

Figure 2. Illustration of notation of Zernike polynomials

2.2 Selection of terms to fit

The next step is to create the x- and y-derivatives from the Zernike polynomials. The data is not directly fit to the derivatives of the Zernike polynomials because the derivatives are not orthogonal. For each of the first 49 Zernike polynomials, use linear combinations of Zernike polynomials. Instead of trying a combination of all 49 polynomials to determine which polynomials construct a given derivative, the $R_r^{\pm m}$ notation clearly indicates which polynomials to include in the linear combination.

The first step is to determine the highest order term in the summation. This is found according to the following rules given in Figure 3, below:

To find $\frac{\partial R_r^m}{\partial x}$ use R_{r-1}^{m+1} unless m+1 > r-1, then use R_{r-1}^{m-1}

To find $\frac{\partial R_r^m}{\partial y}$ use $R_{r-1}^{-(m+1)}$ unless m+1 > r-1, then use $R_{r-1}^{-(m-1)}$

(a) $m \geq 0$

To find $\frac{\partial R_r^m}{\partial x}$ use R_{r-1}^{m+1} unless |m-1| > r-1, then use R_{r-1}^{m+1}

To find $\frac{\partial R_r^m}{\partial y}$ use $R_{r-1}^{-(m+1)}$ unless |m-1| > r-1, then use $R_{r-1}^{-(m+1)}$

(b) $m < 0$

Figure 3. The four cases for determining the highest order term of a linear summation which produces the derivative of the Zernike polynomial specified by the initial R term.

The rest of the terms used in the linear summation are found from the highest order term. This is shown below, with an example following.

To find $\frac{\partial R_r^m}{\partial x}$ for m ≥ 0,even, use $R_r^{m-2}, R_r^{m-4}, \ldots, R_r^0, R_{r-2}^m, R_{r-2}^{m-2}, \ldots, R_{r-2}^0, \ldots, R_0^0$

To find $\frac{\partial R_r^m}{\partial y}$ for m > 0, odd, use $R_r^{m-2}, R_r^{m-4}, \ldots, R_r^1, R_{r-2}^m, R_{r-2}^{m-2}, \ldots, R_{r-2}^1, \ldots, R_1^1$

To find $\frac{\partial R_r^m}{\partial x}$ for m < 0,even, use $R_r^{m+2}, R_r^{m+4}, \ldots, R_r^2, R_{r-2}^m, R_{r-2}^{m-2}, \ldots, R_{r-2}^{-2}, \ldots, R_0^{-2}$

To find $\frac{\partial R_r^m}{\partial y}$ for m < 0, odd, use $R_r^{m+2}, R_r^{m+4}, \ldots, R_r^{-1}, R_{r-2}^m, R_{r-2}^{m+2}, \ldots, R_{r-2}^{-1}, \ldots, R_1^{-1}$

Figure 4. Selection of all terms used in a linear summation to produce a given derivative. The initial R value is the one chosen according to Figure 3. It should be recalled that m ≤ r always, so R terms involving a superscript value greater than its subscript will not exist, and are not used in the linear summation.

	To find the x-derivative:	To find the y-derivative:
(a)	$Z_{13} = R_5^1$	$Z_{13} = R_5^1$
(b)	$R_5^1 \rightarrow R_4^2$	$R_5^3 \rightarrow R_4^{-2}$
(c)	$R_4^2, R_4^0, R_2^2, R_2^0, R_0^0$	R_4^{-2}, R_2^{-2}
(d)	$Z_{11}, Z_8, Z_4, Z_3, Z_0$	Z_{12}, Z_5
(e)	$\frac{\partial R_5^1}{\partial x} = b_{11}Z_{11} + b_8Z_8 + b_4Z_4 + b_3Z_3 + b_0Z_0$	$\frac{\partial R_5^1}{\partial y} = c_{12}Z_{12} + c_5Z_5$

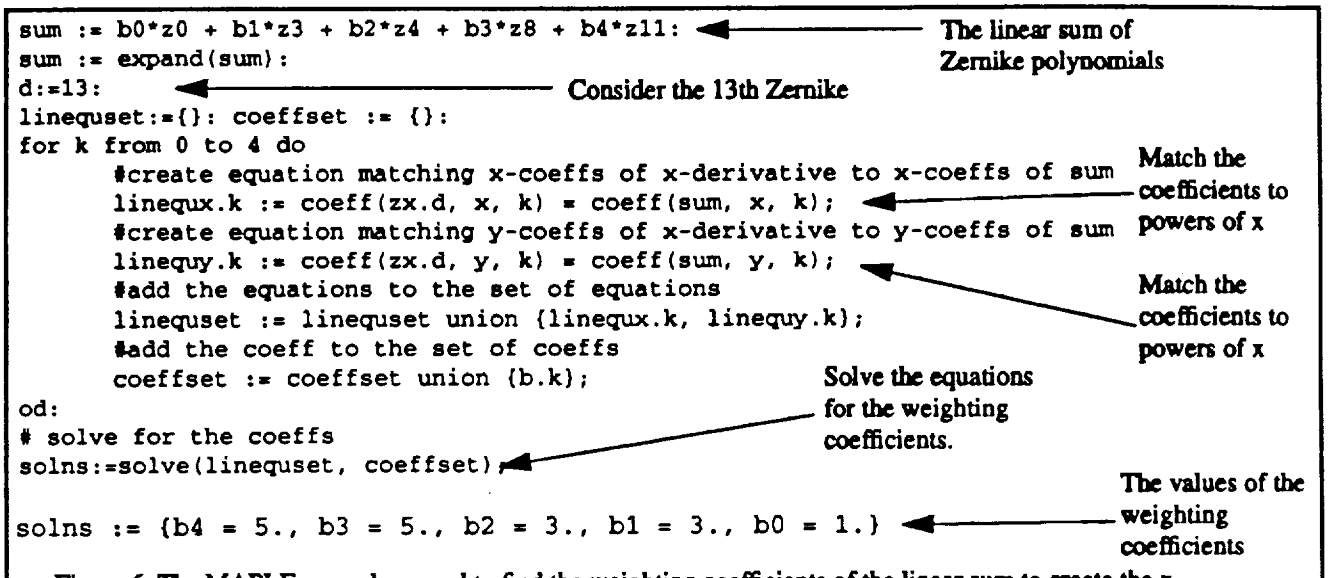
Figure 5. An example of selecting the terms of the linear summation. (a) The Zernike polynomial being considered. (b) Finding the highest order term of the summation for the x- and y-derivatives. (c) All terms involved in the linear summation for the creation of the x- and y-derivatives. (d) The terms expressed in Z_n notation. (e) The linear summation used to create the derivatives; b_{11}, b_8, b_4, b_0 and c_{12} and c_5 are constants.

This process can be repeated for the x- and y-derivatives of the first 49 Zernike polynomials.

2.3 The fitting process

With the polynomials necessary to create each derivative now chosen, it is possible to find the amount of each polynomial needed. Each polynomial in the linear summation is multiplied by an arbitrary constant, as illustrated in Figure 5. To give the exact representation of the derivative in terms of Zernike polynomials, numerical values for these constants were found using a computer algebra application, MAPLE V, on a NeXT computer running NeXTSTEP v3.0. Two approaches to the fitting were used: for simpler summations a simple MAPLE procedure worked well, but for certain higher order terms more effort is involved.

The simple MAPLE procedure begins with a linear sum of the Zernike polynomials that will create the derivative. Each Zernike in the sum is multiplied by a constant. Then, all coefficients are matched between the linear sum and the derivative. (Coefficient matching is the process of finding the coefficients of a given power of x or y on both sides of an equation and setting these terms to be equal.) The set of equations generated by coefficient matching are solved for the constants of the summing polynomials. The MAPLE V codes for doing this for the x- and y-derivatives of the 13th Zernike are shown below.



```
sum := b0*z0 + b1*z3 + b2*z4 + b3*z8 + b4*z11:
sum := expand(sum):
d:=13:
linequset:={}: coeffset := {}:
for k from 0 to 4 do
        #create equation matching x-coeffs of x-derivative to x-coeffs of sum
        linequx.k := coeff(zx.d, x, k) = coeff(sum, x, k);
        #create equation matching y-coeffs of x-derivative to y-coeffs of sum
        linequy.k := coeff(zx.d, y, k) = coeff(sum, y, k);
        #add the equations to the set of equations
        linequset := linequset union {linequx.k, linequy.k};
        #add the coeff to the set of coeffs
        coeffset := coeffset union {b.k};
od:
# solve for the coeffs
solns:=solve(linequset, coeffset);

solns := {b4 = 5., b3 = 5., b2 = 3., b1 = 3., b0 = 1.}
```

The linear sum of Zernike polynomials

Consider the 13th Zernike

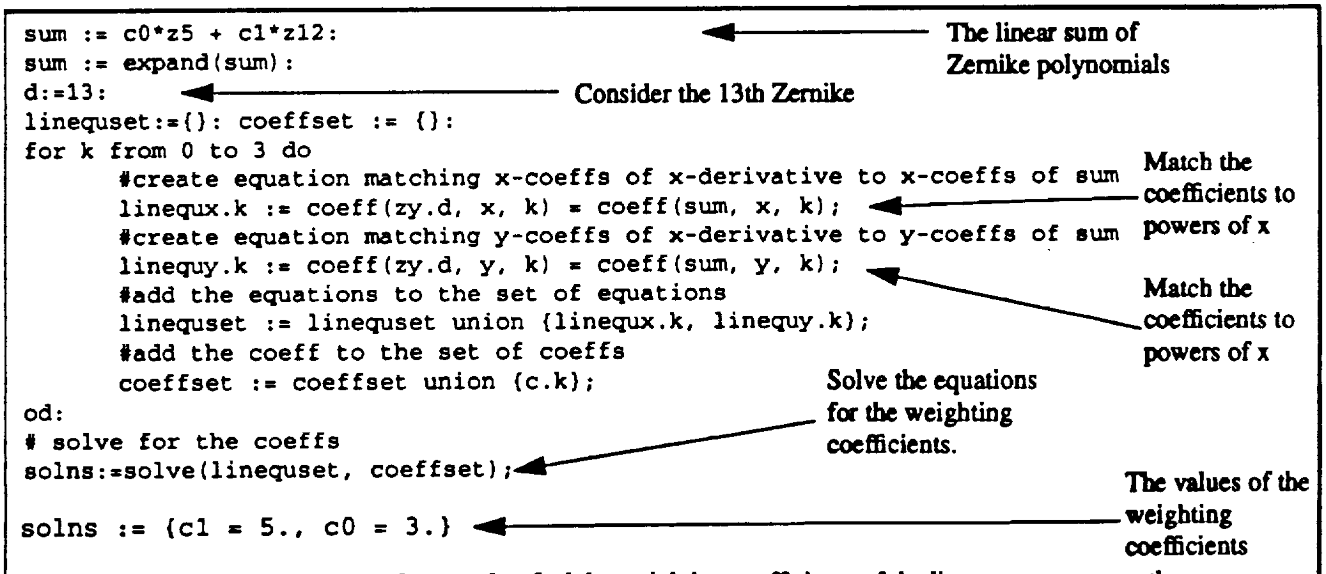
Match the coefficients to powers of x

Match the coefficients to powers of x

Solve the equations for the weighting coefficients.

The values of the weighting coefficients

Figure 6. The MAPLE procedure used to find the weighting coefficients of the linear sum to create the x-derivative of the 13th Zernike polynomial. The coefficients b4 .. b0 correlate to b_{11}, b_8, b_4 and b_0, respectively



```
sum := c0*z5 + c1*z12:
sum := expand(sum):
d:=13:
linequset:={}: coeffset := {}:
for k from 0 to 3 do
        #create equation matching x-coeffs of x-derivative to x-coeffs of sum
        linequx.k := coeff(zy.d, x, k) = coeff(sum, x, k);
        #create equation matching y-coeffs of x-derivative to y-coeffs of sum
        linequy.k := coeff(zy.d, y, k) = coeff(sum, y, k);
        #add the equations to the set of equations
        linequset := linequset union {linequx.k, linequy.k};
        #add the coeff to the set of coeffs
        coeffset := coeffset union {c.k};
od:
# solve for the coeffs
solns:=solve(linequset, coeffset);

solns := {c1 = 5., c0 = 3.}
```

The linear sum of Zernike polynomials

Consider the 13th Zernike

Match the coefficients to powers of x

Match the coefficients to powers of x

Solve the equations for the weighting coefficients.

The values of the weighting coefficients

Figure 7. The MAPLE procedure used to find the weighting coefficients of the linear sum to create the y-derivative of the 13th Zernike polynomial. The coefficients c1 and c0 correlate to c_{12} and c_5, respectively.

Unfortunately, for some cases, this process does not work. The above procedure does not produce integer values for the coefficients. When this happens, it is necessary to use more finesse.

The procedure starts by first identifying the highest order terms in the derivative. The coefficients of these terms are matched to the equivalent terms of the highest order summing polynomials, and the weighting coefficients are solved for. The highest order polynomials are multiplied by the weighting coefficients, and subtracted from the derivative. The process is continued by matching the highest order terms of the remainder to the highest order terms of the remaining summing polynomials and solving for the corresponding coefficients. By iterating this process, all weighting coefficients will be found. This is illustrated for the x-derivative of the 41st Zernike:

```
> zx41;

            6          4  3          5  2          7            4          5
   784.0 y x  + 560.0 x  y  - 336.0 y  x  - 112.0 y  - 840.0 y x  + 168.0 y

                   2         3
        + 180.0 y x  - 60.0 y

> expand(z39);

         6         4  3         5  2        7           4         3  2        5
 35.0 y x  - 35.0 x  y  - 63.0 y  x  + 7.0 y  - 30.0 y x  + 60.0 y  x  - 6.0 y

> expand(z30);

            6          4  3         5  2         7           4         3  2
    63.0 y x  + 105.0 x  y  + 21.0 y  x  - 21.0 y  - 90.0 y x  - 60.0 y  x

                 5           2         3
         + 30.0 y  + 30.0 y x  - 10.0 y

> expand(z23);

           6          4  3          5  2         7           4          3  2
   35.0 y x  + 105.0 x  y  + 105.0 y  x  + 35.0 y  - 60.0 y x  - 120.0 y  x

                5           2         3
        - 60.0 y  + 30.0 y x  + 30.0 y  - 4.0 y

> expand(z26);

                                  4       3  2    5
                             5 y x  - 10 y  x  + y

> expand(z19);

                      4         3  2        5           2        3
              15.0 y x  + 10.0 y  x  - 5.0 y  - 12.0 y x  + 4.0 y

> expand(z14);

                 4         3  2         5           2         3
         10.0 y x  + 20.0 y  x  + 10.0 y  - 12.0 y x  - 12.0 y  + 3.0 y

> expand(z10);

                                       2    3
                                  3 y x  - y

> expand(z10);

                                       2    3
                                  3 y x  - y

> expand(z7);
```

```
                        2          3
              3.0 y x  + 3.0 y  - 2.0 y
```

```
> expand(z2);

                               y
expand(z2);
```

It is seen that the highest order terms in the derivative of Z_{41} are of magnitude seven: yx^6, x^4y^3, y^5x^2, y^7 and there are also three summing polynomials with terms of order seven: Z_{39}, Z_{30}, Z_{23}. Thus, the first three weighting coefficients are found by using the coefficients of the yx^6, x^4y^3, and y^5x^2 terms:

```
> solve({35*b8+63*b7+35*b6 = 784, -35*b8+105*b7+105*b6 = 560, -63*b8+21*b7+105*b6=-
336},
{b8, b7, b6});
>
                         {b7 = 8, b6 = 0, b8 = 8}

> assign(");
> stuff := expand(zx41 - (b8*z39 + b7*z30 + b6*z23));

                              4          5            2          3
          stuff := 120.0 y x  - 24.0 y  - 60.0 y x  + 20.0 y
```

With the first three weighting coefficients found, the process is continued for the yx^4, y^3x^2, and y^5 terms.

```
> solve({5*b5+15*b4+10*b3 = 120, -10*b5+10*b4+20*b3 = 0, 1*b5-5*b4+10*b3=-24}, {b5,
b4, b3});
>
                         {b3 = 0, b4 = 6, b5 = 6}
> assign(");
> stuff2 := expand(stuff - (b5*z26 + b4*z19 + b3*z14));

                                          2         3
                      stuff2 := 12.0 y x  - 4.0 y
```

The last of the weighting coefficients are found by considering the yx^2, and x^3 terms.

```
> solve({3*b2+3*b1=12, -1*b2+3*b1=-4}, {b1, b2});

                             {b1 = 0, b2 = 4}

> assign(");
> stuff3 := expand(stuff2 - (b2*z10+b1*z7));

                               stuff3 := 0
```

The final coefficient, a0, must be equal to zero.

Using the described methods, the constants are found for all polynomials in all linear combinations for the x- and y-derivatives of the first 49 Zernike polynomials.

3. INITIAL APPROXIMATION OF THE SURFACE FROM THE DERIVATIVE FIT

When the orthogonal, slope data is (least squares) fit to Zernike polynomials, Z_0 ... Z_{48}, a set of weighting coefficients for the x-slope and the y-slope are produced. These weighting coefficients will be termed: B_0 ... B_{48} for the x-slope, and C_0 ... C_{48} for the y-slope. The weighting coefficients are now used to generate a surface representation composed of Zernike polynomials.

To find the weighting coefficient for a given Zernike polynomial, first find the largest constant(s) used to recreate its x- and y-derivatives. The coefficients from the least squares fit corresponding to the largest summing constants should then be added together, and divided by the sum of the largest constants. This gives an approximation to the weighting coefficient of the Zernike polynomial to represent the surface.

The 13th Zernike polynomial, Z_{13}, is considered for an example. The largest constants of the x-summation are b_8 and b_{11}, with a value of 5. (See Figure 6) The polynomials corresponding to these constants are Z_{11} and Z_8. Likewise, for the y-derivative, that the largest constant is c_{12}, also with a value of 5, corresponding to Z_{12}. To reconstruct, add the x-slope weighting coefficients (B_{11} and B_8) and the y-slope weighting coefficient (C_{12}). This sum should then be divided by a3+a4+a2 = 15 to normalize. Thus, the initial approximation of the weighting coefficient of Z_{13} based on the slope data is $A_{13} = [B_{11} + B_8 + C_{12}]/15$.

4. IMPROVED APPROXIMATION OF THE SURFACE FROM THE DERIVATIVE FIT

The approximation of the Zernike polynomial representation of the surface found above is improved by subtracting off higher order terms because the higher order terms include lower order terms. The higher order terms which should be removed are those of the same radial frequency of, but with greater radial magnitude than, the Zernike polynomial being considered. These terms are found from the $R_r^{\pm m}$ notation, as shown below:

(a) For a given Zernike polynomial $R_r^{\pm m}$, remove the coefficients of: $R^m_{r+2}, R^m_{r+4}, R^m_{r+4}, \cdots$

(b) For the Zernike polynomial R_5^1, remove the coefficients of: R_7^1, R_9^1, R_{11}^1

For the Zernike polynomial Z_{13}, remove the coefficients: A_{22}, A_{33}, A_{46}

Figure 8. (a) Determining higher order terms to subtract from surface approximation.
(b) An illustration of this procedure using Z_{13}.

The resulting weighting coefficients are thus:

A0 = 0
A1 = [B0] - A6 - A13 - A22 - A46
A2 = [C0] - A7 - A14 - A23 - A34 - A47
A3 = [B1 + B2]/8 - A8 - A14 - A23 - A34
A4 = [B1 - C1]/4 - A11 - A20 - A31 - A44
A5 = [B2 + C1]/4 - A12 - A32 - A45
A6 = [B4 + B3 + C5]/9 - A14 - A23 - A33 - A46
A7 = [B5 - C4 + C3]/9 - A14 - A23 - A34 - A47
A8 = [B6 + C7]/16 - A15 - A24 - A35 - A48
A9 = [B4 - C5]/16 - A18 - A29 - A42
A10 = [B5 + C4]/16 - A19 - A30 - A43
A11 = [B9 + B6 - C7 + C10]/16 - A20 - A31 - A44
A12 = [B10 + B7 + C6 - C9]/16 - A21 - A32 - A45
A13 = [B11 + B8 + C12]/15 - A22 - A33 - A46
A14 = [B12 - C11 + C8]/15 - A23 - A34 - A47
A15 = [B13 + C14]/24 - A24 - A35 - A48
A16 = [B9 - C10]/8 - A27 - A40
A17 = [B10 - C9]/8 - A28 - A41
A18 = [B16 + B11 - C12 + C17]/20 - A29 - A42
A19 = [B12 + B17 - C16 + C11]/20 - A30 - A43
A20 = [B18 +B13 - C14 +C19]/24 - A31 - A44
A21 = [B19 + B14 - C18 + C13]/24 - A32 - A44
A22 = [B20 + B15 + C21]/21 - A34 - A46
A23 = [B21 + C15 - C20]/21 - A35 - A47
A24 = [B22 + C23]/32 - A35 - A48
A25 = [B16 - C17]/10 - A38
A26 = [B17 - C16]/10 - A39
A27 = [B25 + B18 - C19 + C26]/24 - A40
A28 = [B19 + B26 - C25 + C18]/24 - A41
A29 = [B27 + B20 + C28 - C21]/28 - A42
A30 = [B28 + B21 - C27 + C20]/28 - A43
A31 = [B29 + B22 - C23 + C30]/32 - A44
A32 = [B30 + B23 - C29 + C22]/32 - A45
A33 = [B24 + B31 + C32]/27 - A46
A34 = [B32 + C24 - C31]/27 - A47
A35 = [B33 + C34]/40
A36 = [B25 - C26]/12
A37 = [B26 + C25]/12
A38 = [B36 + B25 + B37 - C28]/28
A39 = [B37 + B27 + C27 - C36]/28
A40 = [B38 + B29 + C39 - C30]/32
A41 = [B39 + B30 - C38 + C29]/32
A42 = [B40 + B31 + C41 - C32]/36
A43 = [B41 + B32 - C40 + C31]/36
A44 = [B42 + B33 + C43 - C34]/40
A45 = [B43 + B34 - C42 + C33]/40
A46 = [B44 + B35 - C45]/33
A47 = [B45 - C44 + C35]/33
A48 = [B46 + C47]/48

Figure 9. Weighting coefficients to first 48 Zernike polynomials to reconstruct a surface based on slope measurements. Bn and Cn are the weighting coefficients found by fitting x- and y-derivatives of Zernikes to x- and y-slope, respectively.

A0 = 0
A1 = [B0] - A6 - A13 - A22 - A46 - A33 - A46
A2 = [C0] - A7 - A14 - A23 - A34 - A47
A3 = [B1 + B2]/8 - A8 - A15 - A24 - A35 -48 ← B1+C2 ← -A15-A24-A35 -A48
A4 = [B1 - C1]/4 - A11 - A20 - A31 - A44 ← B1-C2
A5 = [B2 + C1]/4 - A12 - A32 - A45 - A21
A6 = [B4 + B3 + C5]/9 - A13 - A22 - A33 - A46 —— -A13 -A22 ...
A7 = [B5 - C4 + C3]/9 - A14 - A23 - A34 - A47
A8 = [B6 + C7]/16 - A15 - A24 - A35 - A48
A9 = [B4 - C5]/16 - A18 - A29 - A42
A10 = [B5 + C4]/16 - A19 - A30 - A43
A11 = [B9 + B6 - C7 + C10]/16 - A20 - A31 - A44 ← + or -
A12 = [B10 + B7 + C6 - C9]/16 - A21 - A32 - A45
A13 = [B11 + B8 + C12]/15 - A22 - A33 - A46
A14 = [B12 - C11 + C8]/15 - A23 - A34 - A47
A15 = [B13 + C14]/24 - A24 - A35 - A48
A16 = [B9 - C10]/8 - A27 - A40
A17 = [B10 - C9]/8 - A28 - A41 ← + or -
A18 = [B16 + B11 - C12 + C17]/20 - A29 - A42
A19 = [B12 + B17 - C16 + C11]/20 - A30 - A43
A20 = [B18 +B13 - C14 +C19]/24 - A31 - A44
A21 = [B19 + B14 - C18 + C13]/24 - A32 - A44 A45
A22 = [B20 + B15 + C21]/21 - A34 - A46
A23 = [B21 + C15 - C20]/21 - A35 - A47
A24 = [B22 + C23]/32 - A35 - A48
A25 = [B16 - C17]/10 - A38
A26 = [B17 - C16]/10 - A39 ← + or -
A27 = [B25 + B18 - C19 + C26]/24 - A40
A28 = [B19 + B26 - C25 + C18]/24 - A41
A29 = [B27 + B20 + C28 - C21]/28 - A42
A30 = [B28 + B21 - C27 + C20]/28 - A43
A31 = [B29 + B22 - C23 + C30]/32 - A44
A32 = [B30 + B23 - C29 + C22]/32 - A45
A33 = [B24 + B31 + C32]/27 - A46
A34 = [B32 + C24 - C31]/27 - A47
A35 = [B33 + C34]/40 - A48
A36 = [B25 - C26]/12
A37 = [B26 + C25]/12
A38 = [B36 + B25 + B37 - C28]/28 ← C37
A39 = [B37 + B27 + C27 - C36]/28
A40 = [B38 + B29 + C39 - C30]/32
A41 = [B39 + B30 - C38 + C29]/32
A42 = [B40 + B31 + C41 - C32]/36
A43 = [B41 + B32 - C40 + C31]/36
A44 = [B42 + B33 + C43 - C34]/40
A45 = [B43 + B34 - C42 + C33]/40
A46 = [B44 + B35 - C45]/33
A47 = [B45 - C44 + C35]/33
A48 = [B46 + C47]/48

Figure 9. Weighting coefficients to first 48 Zernike polynomials to reconstruct a surface based on slope measurements. Bn and Cn are the weighting coefficients found by fitting x- and y-derivatives of Zernikes to x- and y-slope, respectively.

5. CONCLUSION

The means to obtain a surface representation of by fitting Zernike polynomials (in a least squares sense) to discrete data points in the full aperture has been described in the past via matrix and vector notation. Additionally, vector notation has been used to describe how to obtain a surface representation from orthogonal (x, y) slope data[3]. The result of that paper was a matrix operator for linearly combining the first eight Zernike polynomial coefficients fit to x- and y-slope data to produce a Zernike polynomial surface representation. This paper extended that process by presenting a systematic approach for obtaining the linear relationship between slope and surface using the first 49 Zernike polynomials.

6. ACKNOWLEDGMENTS

This project is supported by a NASA Research Grant (NAG3-1300) from NASA-Lewis Research Center.

7. BIBLIOGRAPHY

1) Fischer, D. J., O'Bryan J. T., and Stahl, H. P., "Vector formulation for interferogram surface fitting", SPIE Proceedings, Vol. 1779, March 19 1992, pp. 80 - 87.

2) Wang J. Y., and D. E. Silva, "Wave-front interpolation with Zernike polynomials", Applied Optics, Vol. 19, No. 9, pp 1510-1518, May 1980.

3) Fischer, D. J., and Stahl, H. P., "A vector formulation for ronchi shear surface fitting", SPIE Proceedings, Vol. 1775, July 17 1993, pp. 80 - 87.

4) Born, Wolf, Principles of Optics, Pergamon Press, New York, 1964, 2nd Ed. p. 464.

A vector formulation for ronchi shear surface fitting

David J. Fischer
H. Philip Stahl

Rose-Hulman Institute of Technology
Center for Applied Optics Studies
5500 East Wabash Avenue
Terre Haute, IN 47803
(812) 877-8253

ABSTRACT

Ronchi interferometry is an optical testing technique similar to shearing interferometry. A coherent wavefront is interfered with a sheared form of itself by placing a periodic grating at, or near, the focus of an optical system. The resultant interference pattern contains information about the wavefront's slope in a direction perpendicular to the grating structure. The wavefront can be reconstructed from two orthogonal slope data sets via the process of sampling, ordering and fitting. This paper develops a linear-algebra vector notation model of the interferogram sampling and fitting process.

1. INTRODUCTION

Whenever two coherent wavefronts exist at the same point in space, they superimpose, to create an interference pattern. The irradiance at any point, E(x,y), as determined by the Optical Path Difference (OPD) between the two wavefronts is described by:

$$E(x,y) = E_0\left[1 + \cos\left(\frac{2\pi}{\lambda}\text{OPD}\right)\right] \tag{1}$$

where

$$\lambda = \text{measurement wavelength} \tag{2}$$

The two dimensional irradiance pattern created by the optical path difference at all points is called an irradiance pattern. For a sheared interferogram, the OPD is given by the difference between the wavefront and its sheared self. This difference is called the differential wavefront.

$$\text{OPD} = W(x,y) - W(x+\Delta x, y) = \Delta W(x,y) \tag{3}$$

By rewriting the differential wavefront as:

$$\Delta W(x,y) = \frac{\delta W(x,y)}{\delta x}\Delta x = \alpha(x,y)\Delta x \tag{4}$$

the wavefront slope can then be written as:

$$\alpha(x,y) = \frac{\delta W(x,y)}{\delta x} = \frac{\Delta W(x,y)}{\Delta x} \tag{5}$$

Therefore, the irradiance pattern can be expressed as:

$$E(x,y) = E_0\left[1 + \cos\left(\frac{2\pi}{\lambda_{eq}}\alpha\right)\right] \tag{6}$$

where:

$$\lambda_{eq} = \frac{\lambda}{\Delta x} = \text{equivalent wavelength} \tag{7}$$

The Ronchi test is similar to shearing interferometry, only the shear is caused by a periodic grating placed at, or near, the focus of the optical system under test. Like shearing interferometry, the resultant irradiance pattern (Ronchigram) contains slope information in a direction perpendicular to the shear direction - the direction of the grating structure. From this similarity, it can be shown that the Ronchigram can be expressed by:

$$E(x,y) = E_0\left[1 + \cos\left(\frac{2\pi}{\lambda_{eq}}\alpha\right)\right] \tag{8}$$

SPIE Vol. 1755, 1992.

where:

$$\alpha = \text{wavefront slope} \tag{9}$$

$$\lambda_{eq} = \frac{d}{2F/\#} \tag{10}$$

$$d = \text{Grating Spacing} \tag{11}$$

$$F/\# = \text{Test Beam } F/\# \tag{12}$$

Thus, by proper scaling of the interferogram, conventional interferogram analysis techniques will provide a direct measurement of the wavefront slope. The challenge is to extract a representation of the test surface from such an interferogram. This will be done using a linear-algebra vector notation model of the interferogram sampling and fitting process

2. INTERFEROGRAM ANALYSIS

Conventional wavefront interferogram analysis is accomplished via a three step process: data sampling, ordering, and fitting. The result of this process is a representation of the wavefront under test. For a shear interferogram, such as a Ronchigram, the analysis provides a representation of the wavefront's slope in a particular direction (such as the x- or y-direction). To obtain a representation of the actual wavefront from a Ronchigram, it is necessary to combine two orthogonal slope representations.

2.1 Sampling

Because an interferogram has nearly infinite resolution, the actual number of points that can be measured is limited only by the sampling technique. There are two main methods for acquiring the data: digitization, and phase-measuring interferometry. In digitization, data points are acquired by sampling along the center of either the light or dark fringes. Usually, the interferogram is photographed and the image is placed onto a digitizing tablet where the data points are digitized by a human operator. Alternatively, the image can be acquired with a video camera and digitized using image-processing techniques. Phase-measuring interferometry, by contrast, is an electronic process where a video camera samples the wavefront phase at every point within its detector array - thus creating a high density, uniform grid of data. An illustration to show the difference in the two techniques is shown in Figure 1 and Figure 2:

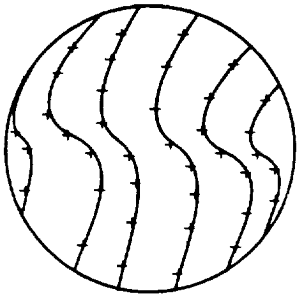


Fig. 1. An illustration of the placement of data points when an interferogram is sampled by digitization

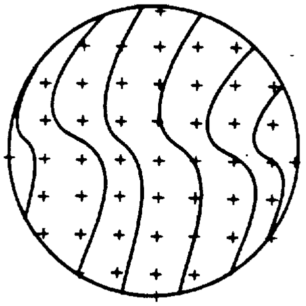


Fig. 2. An illustration of the placement of data points when an interferogram is sampled by Phase-Measuring Interferometry

For this paper, it is assumed that the data is sampled by manual digitization along the dark fringes. The data is recorded in the form (m,x,y) where (x,y) gives the position of the point and (m) is the order of the sampled fringe - which, as will be discussed in the next paragraph, is related to the height of the surface in units of 2π.

2.2 Ordering

Once the data has been sampled, it must be ordered. Because the irradiance pattern is proportional to cos(k*OPD) the data is in units of modulo 2π or modulo λ. Thus, even though a point in the interferogram represents a quantity greater than 2π, no measured quantity will be larger than 2π. For instance, an (x,y) point may have an (m) that is of magnitude 9π. The measured (m) will equal π, which is found by taking the magnitude modulo 2π. Therefore, the magnitude of the data must be artificially reconstructed by ordering the data. This is done by assigning integer order number values to each series of measured data points. For example, all data points along a given fringe will have the same order and each adjacent fringe has an incrementally larger order number.

For manually digitized data, this ordering is accomplished by the trained operator. For video digitization, this is accomplished by image processing techniques. And for phase-measuring, this ordering is accomplished by a process called phase unwrapping.

2.3 Fitting

Once the data has been sampled and ordered, it must be fitted to a polynomial set to represent the shape of the test surface. generally, this is accomplished using a least squares method with a Gram-Schmidt orthogonalization. From this fit several questions can be answered: how flat is the surface, are the features concave or convex, and what aberrations are present.

Traditionally, Zernike polynomials have been used for this representation for several reasons. They look like classical Seidel Aberrations. They provide a good fit to the circular aperture interferogram of most optical components. And, because they are orthogonal, they are easy to manipulate. It is possible to add or subtract coefficients without seriously affecting the quality of the fit.

3. VECTOR FORMULATION

3.1 Purpose of the vector formulation

The purpose of this section is to summarize simply and clearly the analysis process of a Ronchigram using general linear-algebra vector notation, and to develop a generalized mathematical model with which to explore the effects of: sampling parameters, fitting polynomials, and the surface-fitting algorithm on the accuracy of the reconstruction of the test surface. The remainder of this paper presents a general vector based formulation for the surface-fitting process.

3.2 Representing the measured values

The wavefront function at any point in the exit pupil, W(x,y), can be represented by a linear combination of M polynomials, $F_j(x,y)$, and their weighting coefficients, G_j, for j in 1 to M.

$$W(x, y) = \sum_{j=1}^{M} G_j F_j(x, y) \tag{13}$$

Similarly, the wavefront slope at any point in the exit pupil, $\alpha(x,y)$, can be represented by:

$$\alpha(x, y) = \sum_{j=1}^{M} G_j \nabla F_j(x, y) \tag{14}$$

Unfortunately, the Ronchi test does not make general slope measurements. It makes independent measurements of x-slope and y-slope which can be represented as:

$$\alpha_x(x,y) = \sum_{j=1}^{M} G_{xj}(F_{xj}(x,y)) \tag{15}$$

$$\alpha_y(x,y) = \sum_{j=1}^{M} G_{yj}(F_{yj}(x,y)) \tag{16}$$

where:

$$F_{xj}(x,y) = \frac{\delta F_j(x,y)}{\delta x} \tag{17}$$

$$F_{yj}(x,y) = \frac{\delta F_j(x,y)}{\delta y} \tag{18}$$

It is therefore necessary to perform two separate fitting operations to determine G_{xj} and G_{yj}.

A sampled and ordered x-slope interferogram data set be thought of as a series of N measured data points, where N > M. At any point in the interferogram, (x_r, y_r), there is a measured value, α_{xr}, where r is in 1 to N, the number of sampled points. The r^{th} measured x-slope value is then written as:

$$\alpha_{xr}(x_r, y_r) = \sum_{j=1}^{M} G_{xj}(F_{xj}(x,y)) \qquad \text{where } N > M \tag{19}$$

Alternatively, the single measured value, α_{xr}, can also be expressed as a matrix of polynomials multiplied by a matrix of coefficients.

$$\begin{bmatrix}\alpha_{xr}\end{bmatrix} = \begin{bmatrix}F_{x1}(x_r, y_r) & \ldots & F_{xM}(x_r, y_r)\end{bmatrix}\begin{bmatrix}G_{x1}\\ | \\ G_{xM}\end{bmatrix} \tag{20}$$

Thus, the entire series of N, measured x-slope values can be represented as a matrix of polynomials multiplied by the coefficients matrix. (See Equation 21)

$$\begin{bmatrix}\alpha_{x1}\\ | \\ \alpha_{xN}\end{bmatrix} = \begin{bmatrix}F_{x1}(x_1, y_1) & \ldots & F_{xM}(x_1, y_1)\\ | & & | \\ F_{x1}(x_N, y_N) & \ldots & F_{xM}(x_N, y_N)\end{bmatrix}\begin{bmatrix}G_{x1}\\ | \\ G_{xM}\end{bmatrix} \tag{21}$$

In order to simplify this notation, the measured values can be considered to be a vector with N rows. Likewise, the coefficients can be viewed as a vector with M rows, and a given polynomial evaluated at all N data points is a vector with N rows.

$$\vec{\alpha}_x = \begin{bmatrix}\alpha_{x1}\\ | \\ \alpha_{xN}\end{bmatrix} \quad \vec{G}_x = \begin{bmatrix}G_{x1}\\ | \\ G_{xM}\end{bmatrix} \quad \vec{F}_{xj} = \begin{bmatrix}F_{xj}(x_1, y_1)\\ | \\ F_{xj}(x_N, y_N)\end{bmatrix} \tag{22}$$

Using this notation, the fitted values can be represented by:

$$\vec{\alpha}_x = \begin{bmatrix}\vec{F}_{x1} & \ldots & \vec{F}_{xM}\end{bmatrix}\vec{G}_x \tag{23}$$

3.3 Finding the best fit

The polynomials can be fit using a least squares method, where S is defined as the sum of the square of the difference between the data points and the fitted polynomials. If a perfect fit were possible, then S would be equal to zero, because there would be no difference between the measured values and the representing polynomials evaluated at the corresponding points.

However, there are always differences between the real surface and its representation. Therefore, S will be non-zero. The function of a least squares fit is to find the coefficients, for a given set of polynomials, which minimizes S. This process can be described using vector notation (See Equation 24), or as more commonly expressed using summations. (See Equation 25)

$$S \equiv \left\| \left[\vec{\alpha}_x - \left[\vec{F}_{xj} \ \dots \ \vec{F}_{xM} \right] \vec{G}_x \right] \right\|^2 \tag{24}$$

$$S \equiv \sum_{r=1}^{N} \left[\alpha_{xr} - \sum_{j=1}^{M} G_{xj} F_{xj}(x_r, y_r) \right]^2 \tag{25}$$

The coefficients are found by taking the derivative of S with respect to each coefficient and setting the result equal to zero. For example, minimizing S for G_{xk} gives:

$$\sum_{r=1}^{N} \alpha_{xr} F_{xk}(x_r, y_r) = \sum_{r=1}^{N} \sum_{j=1}^{M} G_{xj} F_{xj}(x_r, y_r) F_{xk}(x_r, y_r) \tag{26}$$

The result of minimizing S for all coefficients can be expressed using summations in matrix form as:

$$\begin{bmatrix} \sum\limits_{r=1}^{N} \alpha_{xr} F_{x1}(x_r, y_r) \\ \vdots \\ \sum\limits_{r=1}^{N} \alpha_{xr} F_{xM}(x_r, y_r) \end{bmatrix} = \begin{bmatrix} \sum\limits_{r=1}^{N} F_{x1}(x_r, y_r) F_{x1}(x_r, y_r) & \dots & \sum\limits_{r=1}^{N} F_{xM}(x_r, y_r) F_{x1}(x_r, y_r) \\ \vdots & & \vdots \\ \sum\limits_{r=1}^{N} F_{x1}(x_r, y_r) F_{xM}(x_r, y_r) & \dots & \sum\limits_{r=1}^{N} F_{xM}(x_r, y_r) F_{xM}(x_r, y_r) \end{bmatrix} \begin{bmatrix} G_{x1} \\ \vdots \\ G_{xM} \end{bmatrix} \tag{27}$$

This expression can be greatly simplified with vector notation (See Equation 28), where the dot product of the vectors $\vec{F}_{xi}$ and F_{xj} gives the sum of the products of both polynomials over every data point. (See Equation 29)

$$\begin{bmatrix} \vec{\alpha}_x \bullet \vec{F}_{x1} \\ \vdots \\ \vec{\alpha}_x \bullet \vec{F}_{xM} \end{bmatrix} = \begin{bmatrix} \vec{F}_{x1} \bullet \vec{F}_{x1} & \dots & \vec{F}_{x1} \bullet \vec{F}_{xM} \\ \vdots & & \vdots \\ \vec{F}_{xM} \bullet \vec{F}_{x1} & \dots & \vec{F}_{xM} \bullet \vec{F}_{xM} \end{bmatrix} \vec{G}_x \tag{28}$$

$$\vec{F}_{xi} \bullet \vec{F}_{xj} = \sum_{r=1}^{N} F_{xi}(x_r, y_r) F_{xj}(x_r, y_r) \tag{29}$$

Likewise the dot product of the vectors $\vec{\alpha}_x$ and $\vec{F}_{xj}$ gives the sum of the products evaluated at all data points.

$$\vec{\alpha}_x \bullet \vec{F}_{xj} = \sum_{r=1}^{N} \alpha_{xr} F_{xj}(x_r, y_r) \tag{30}$$

3.4 Solving for the coefficients assuming orthogonal polynomial vectors

If the F vectors are orthogonal over the sampled data points, then the minimization process produces a diagonal matrix (See Equation 31), and the coefficients can be found by multiplying both sides of the equation by the inverse of the diagonal matrix. (See Equation 32) The process is repeated using the y-slope data to find $\vec{G}_y$

$$\begin{bmatrix} \vec{\alpha}_x \bullet \vec{F}_{x1} \\ \vdots \\ \vec{\alpha}_x \bullet \vec{F}_{xM} \end{bmatrix} = \begin{bmatrix} \vec{F}_{x1} \bullet \vec{F}_{x1} & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & \vec{F}_{xM} \bullet \vec{F}_{xM} \end{bmatrix} \vec{G}_x \tag{31}$$

$$\vec{G}_x = \begin{bmatrix} \dfrac{\vec{\alpha}_x \bullet \vec{F}_{x1}}{\|\vec{F}_{x1}\|^2} \\ \vdots \\ \dfrac{\vec{\alpha}_x \bullet \vec{F}_{xM}}{\|\vec{F}_{xM}\|^2} \end{bmatrix} \tag{32}$$

However, most orthogonal polynomials are only orthogonal over a continuous range, not over discrete points. To overcome this limitation, the Gram-Schmidt orthogonalization technique will be used to create two new polynomial sets which is orthogonal over the x-slope data, and over the y-slope data.

4. GRAM-SCHMIDT ORTHOGONALIZATION

4.1 Summary of the Gram-Schmidt Orthogonalization process

The Gram-Schmidt orthogonalization process takes M arbitrary vectors in an M dimensional space and creates M orthogonal vectors that form a basis for that space. The Gram-Schmidt Orthogonalization process can be expressed for any vector, $\vec{F}_{xj}$, as:

$$\vec{\Phi}_{xj} = \vec{F}_{xj} - \sum_{s=1}^{j-1} \frac{\vec{F}_{xj} \bullet \vec{\Phi}_{xs}}{\vec{\Phi}_{xs} \bullet \vec{\Phi}_{xs}} \vec{\Phi}_{xs} \tag{33}$$

By letting D_{js}, represent the coefficient of $\vec{\Phi}_{xs}$ (See Equation 34), then the orthogonalization of any vector, $\vec{F}_{xj}$, can be expressed as shown in Equation 35.

$$D_{xjs} = \frac{\vec{F}_{xj} \bullet \vec{\Phi}_{xs}}{\vec{\Phi}_{xs} \bullet \vec{\Phi}_{xs}} \tag{34}$$

$$\vec{\Phi}_{xj} = \vec{F}_{xj} - \sum_{s=1}^{j-1} D_{jsx} \vec{\Phi}_{xs} \tag{35}$$

4.2 Applying the Gram-Schmidt process to the vector formulation

For the fitting process, the polynomial vectors, $\vec{F}_{x1}$ through $\vec{F}_{xM}$ (See Equation 22), are orthogonalized and replaced by the new orthogonal vectors, Φ_{x1} through Φ_{xM}. Because the original fitting polynomials are replaced by orthogonalized polynomials, a different set of coefficients is found. These new coefficients are named Γ_j, for j in 1 to M. Therefore, the original coefficients vector, $\vec{G}_x$, is replaced by the vector $\vec{\Gamma}_x$. The measured values can now be fitted in terms of the new vectors.

$$\vec{\alpha}_x = \begin{bmatrix} \vec{\Phi}_{x1} & \cdots & \vec{\Phi}_{xM} \end{bmatrix} \vec{\Gamma}_x \tag{36}$$

The system is then solved, as shown previously (See Equations 31 and 32), by multiplying both sides of the minimized least squares fit equation by the inverse of the resultant diagonal matrix. The Γ_x coefficients are now known and can be used to represent the original surface. (See Equation 37) The process is repeated using the y-slope data to find Γ_y.

$$\vec{\Gamma}_x = \begin{bmatrix} \dfrac{\vec{\alpha}_x \bullet \vec{\Phi}_{x1}}{\|\vec{\Phi}_{x1}\|^2} \\ \vdots \\ \dfrac{\vec{\alpha}_x \bullet \vec{\Phi}_{xM}}{\|\vec{\Phi}_{xM}\|^2} \end{bmatrix} \tag{37}$$

5. FINDING $\vec{G}_x$ IN TERMS OF $\vec{\Gamma}_x$

The polynomials fitted to the measured values are typically a standard set, such as the Zernike polynomials. Therefore, since the polynomials do not change, the coefficients found in the fitting process of the original polynomials (See Equation 32) can be used to describe the test surface and serve as a basis of comparison between other test surfaces. However, when the $\vec{F}_{x1}$ through $\vec{F}_{xM}$ were orthogonalized, the polynomials F_{x1} through F_{xM} were modified, creating a new set of polynomials, Φ_{x1} through Φ_{xM}, that are dependent upon the measured values. Thus, the Γ_x coefficients that compose $\vec{\Gamma}_x$ cannot serve as a basis of comparison, since every set of Γ_x coefficients is related to a different set of polynomials. Because of this, it is necessary to find a means to convert $\vec{\Gamma}_x$ to $\vec{G}_x$, and $\vec{\Gamma}_y$ to $\vec{G}_y$.

The first step in finding $\vec{G}_x$ in terms of $\vec{\Gamma}_x$ is to express the orthogonalization process using matrix notation. From Equation 35, the orthogonalization of all M vectors can be shown as follows.

$$\begin{bmatrix} \vec{\Phi}_{x1}^{T} \\ \vdots \\ \vec{\Phi}_{xM}^{T} \end{bmatrix} = \begin{bmatrix} \vec{F}_{x1}^{T} \\ \vdots \\ \vec{F}_{xM}^{T} \end{bmatrix} - \begin{bmatrix} 0 & \cdots & & & 0 \\ D_{x21} & 0 & \cdots & & 0 \\ D_{x31} & D_{x32} & 0 & \cdots & 0 \\ \vdots & \vdots & & & \vdots \\ D_{xM1} & D_{xM2} & \cdots & D_{xMM-1} & 0 \end{bmatrix} \begin{bmatrix} \vec{\Phi}_{x1}^{T} \\ \vdots \\ \vec{\Phi}_{xM}^{T} \end{bmatrix} \tag{38}$$

The matrix of D_{xjs} coefficients is called D_x for simplicity. Solving Equation 38 for the orthogonal vectors, Equation 39 is found, where I is an identity matrix of size M x M.

$$\begin{bmatrix} \vec{\Phi}_{x1}^{T} \\ \vdots \\ \vec{\Phi}_{xM}^{T} \end{bmatrix} = (I + D_x)^{-1} \begin{bmatrix} \vec{F}_{x1}^{T} \\ \vdots \\ \vec{F}_{xM}^{T} \end{bmatrix} \tag{39}$$

Recalling Equation 23 and Equation 36, it is seen that two equivalent representations of the surface exist; one using the original polynomial vectors, and the other using the orthogonalized polynomial vectors. Therefore, this statement can be made:

$$\begin{bmatrix} \vec{F}_{x1} & \cdots & \vec{F}_{xM} \end{bmatrix} \vec{G}_x = \begin{bmatrix} \vec{\Phi}_{x1} & \cdots & \vec{\Phi}_{xM} \end{bmatrix} \vec{\Gamma}_x \tag{40}$$

Equation 40 is then rewritten using transposes (See Equation 41), and the transpose of both sides of the resultant equation is taken. This produces an expression for the matrix of $\vec{\Phi}_{x1}$ through $\vec{\Phi}_{xM}$ in terms of $\vec{F}_{x1}$ through $\vec{F}_{xM}$. (See Equation 42)

$$\begin{bmatrix} \vec{\Phi}_{x1} & \cdots & \vec{\Phi}_{xM} \end{bmatrix}^{T} = (I + D_x)^{-1} \begin{bmatrix} \vec{F}_{x1} & \cdots & \vec{F}_{xM} \end{bmatrix}^{T} \tag{41}$$

$$\begin{bmatrix} \vec{\Phi}_{x1} & \cdots & \vec{\Phi}_{xM} \end{bmatrix} = \begin{bmatrix} \vec{F}_{x1} & \cdots & \vec{F}_{xM} \end{bmatrix} [\,(I + D_x)^{-1}]^{T} \tag{42}$$

Now, Equation 40 is expressed using Equation 42.

$$\begin{bmatrix} \vec{F}_{x1} & \cdots & \vec{F}_{xM} \end{bmatrix} \vec{G}_x = \begin{bmatrix} \vec{F}_{x1} & \cdots & \vec{F}_{xM} \end{bmatrix} [\,(I + D_x)^{-1}]^{T} \vec{\Gamma}_x \tag{43}$$

Both sides of Equation 43 are divided the matrix of polynomial vectors, and $\vec{G}_x$ is found in terms of the D_{xjs} coefficients and Γ_x. Again, this process is repeated using the y-slope data to find $\vec{G}_y$ in terms of Γ_y.

$$\vec{G}_x = [(I+D_x)^{-1}]^T\vec{\Gamma}_x \tag{44}$$

6. WAVEFRONT RECONSTRUCTION FROM WAVEFRONT SLOPE

As was stated in section 2, (INTERFEROGRAM ANALYSIS), a representation of the actual wavefront is obtained by combining two orthogonal slope representations. Therefore, it is necessary to solve for the weighting coefficients for the actual wavefront. This is done by using a linear combination of the orthogonal slope coefficients. (See Equation 45) The x-slope and y-slope coefficients are multiplied by the weighting coefficients, H_x and H_y respectively, which are of size M by M. It should be noted that the scaling matrices are a function of the selected fitting polynomial set.

$$\vec{G} = H_x\vec{G}_x + H_y\vec{G}_y \tag{45}$$

If each element of $\vec{G}$ depends only upon its corresponding slope elements (i.e. the polynomial set is orthogonal under differentiation), then the scaling matrices, H_x and H_y, will be diagonal. However, if the fitting polynomial set is not orthogonal under differentiation, then fitting errors will occur. The Zernike polynomials are not orthogonal under differentiation; they exhibit cross-correlation. This cross-correlation is illustrated for the first nine Zernike polynomials:

Term	$F(x,y)$	$F_x(x,y)$	$F_y(x,y)$
0	1	0	0
1	x	1	0
2	y	0	1
3	$2(x^2+y^2) - 1$	$4x$	$4y$
4	x^2-y^2	$2x$	$-2y$
5	$2xy$	$2y$	$-2x$
6	$3(x^3+xy^2) - 2x$	$3(3x^2+y^2) - 2$	$6xy$
7	$3(x^2y+y^3) - 2y$	$6xy$	$3(x^2+3y^2) - 2$
8	$6(x^2+y^2)^2 - 6(x^2+y^2)^2 + 1$	$24(x^3+xy^2) - 12x$	$24(x^2y+y^3) - 12y$

Table 1: Zernike Polynomials and their derivatives.

There are two solutions to the problem of cross-correlation. The first is to orthogonalize the fitting polynomial sets, F_x and F_y. The second option is to use a different basis set which is orthogonal under differentiation. The standard Zernike polynomials are such a basis set; they can be linearly combined to form their own derivatives. This is illustrated for the first nine Zernike polynomials:

Term	F(x,y)	$F_x(x,y)$	$F_y(x,y)$
Z_0	1	NA	NA
Z_1	x	Z_0	NA
Z_2	y	NA	Z_0
Z_3	$2(x^2+y^2) - 1$	$4Z_1$	$4Z_2$
Z_4	$x^2 - y^2$	$2Z_1$	$-2Z_2$
Z_5	2xy	$2Z_2$	$2Z_1$
Z_6	$3(x^3+xy^2) - 2x$	$3(Z_3+Z_4) + Z_0$	$3Z_5$
Z_7	$3(x^2y+y^3) - 2y$	$3Z_5$	$3(Z_3-Z_4) + Z_0$
Z_8	$6(x^2+y^2)^2 - 6(x^2+y^2)^2 + 1$	$8Z_6 + 4Z_1$	$8Z_7 + 4Z_2$

Table 2: The first nine Zernikes and combinations needed to create the x- and y-derivatives.

In a similar manner, the weighting coefficients for a wavefront can be constructed from the weighting coefficients of the slope representation. This is illustrated for the first nine terms in Table 3.

Term	G	G_x	G_y
G_0	NA	G_{x0}	G_{y0}
G_1	$G_{x0} - (G_{x4}+G_{x3}+G_{y5})/9$	G_{x1}	G_{y1}
G_2	$G_{y0} - (G_{x5}+G_{y4}+G_{y3})/9$	G_2	G_{y2}
G_3	$(G_{x1}+G_{y2})/8 - (G_{x6}+G_{y7})/16$	G_{x3}	G_{y3}
G_4	$(G_{x1}+G_{y2})/4 - (G_{x6}-G_{y7})/16$	G_{x4}	G_{y4}
G_5	$(G_{x2}+G_{y1})/4 - (G_{x7}-G_{y6})/16$	G_{x5}	G_{y5}
G_6	$(G_{x4}+G_{x3}+G_{y5})/9 - G_{x8}/15$	G_{x6}	G_{y6}
G_7	$(G_{x5}-G_{y4}+G_{y3})/9 - G_{y8}/15$	G_{x7}	G_{y7}
G_8	$(G_{x6}+G_{y7})/16$	G_{x8}	G_{y8}

Table 3: Weighting coefficients for the wavefront can be created from the slope coefficients.

Recalling Equation 45, the information in Table 3 can be reduced to the scaling matrices, H_x and H_y as shown:

$$H_x = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -\frac{1}{9} & -\frac{1}{9} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{1}{9} & 0 & 0 & 0 \\ 0 & \frac{1}{8} & 0 & 0 & 0 & 0 & -\frac{1}{16} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & 0 & 0 & 0 & -\frac{1}{16} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & 0 & 0 & 0 & -\frac{1}{16} & 0 \\ 0 & 0 & 0 & \frac{1}{9} & \frac{1}{9} & 0 & 0 & 0 & -\frac{1}{15} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{9} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{16} & 0 & 0 \end{bmatrix} \tag{46}$$

$$H_y = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{1}{9} & 0 & 0 & 0 \\ 1 & 0 & 0 & -\frac{1}{9} & -\frac{1}{9} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{8} & 0 & 0 & 0 & 0 & -\frac{1}{16} & 0 \\ 0 & 0 & -\frac{1}{4} & 0 & 0 & 0 & 0 & \frac{1}{16} & 0 \\ 0 & \frac{1}{4} & 0 & 0 & 0 & 0 & -\frac{1}{16} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{9} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{9} & \frac{1}{9} & 0 & 0 & 0 & -\frac{1}{15} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{16} & 0 \end{bmatrix} \tag{47}$$

7. CONCLUSION

Ronchi shear interferometry is an optical testing technique similar to shearing interferometry. Information about the slope of a test surface is extracted from interferograms. This is done with three main steps. First, the slope data for two orthogonal directions is sampled, usually by one of two main methods: digitization or phase-measuring interferometry. Second, the sampled data points are ordered. And third, the ordered data is fitted to a the derivatives of a polynomial set, via a least squares method and Gram-Schmidt orthogonalization. A linear combination of the slope coefficients is then found to construct the wavefront from the slope representation.

8. ACKNOWLEDGMENTS

This project is supported by a NASA Research Grant (NA63-1300) from NASA-Lewis Research Center.

Some of the work of this project was performed while at NASA Lewis Research Center, Ohio where H. Philip Stahl was a participant in the ASEE/NASA Summer Faculty Fellowship Program.

Special thanks is given to Alex Pline of NASA Lewis.

Special thanks is given to the Ohio Aerospace Institute.

Special thanks is given to Dr. Robert Lopez.

9. BIBLIOGRAPHY

Fischer, D. J., O'Bryan J. T., and Stahl, H. P., "Vector formulation for interferogram surface fitting", SPIE Proceedings, Vol. 1779, March 19 1992, pp. 80 - 87

Lewis, Jeff L., William P Kuhn and H. Philip Stahl, "The Evaluation of a Random Sampling Error on the Polynomial Fit of Subaperature Test Data"

Malacara, Daniel, J. Martin and Caprio-Valadez and J. Javier Sanchez-Mondragon, "Wavefront Fitting with Discrete Orthogonal Polynomials in a Unit Radius Circle", Optical Engineering, Vol. 24, No. 6, pp 672 - 675, June 1990

Nobles, Ben, Applied Linear Algebra, pp 9-18, Prentice-Hall Inc., 1969

Rimmer M.P., "Method for Evaluating Shearing Interferograms", Applied Optics, Vol. 13, No. 3, pp 623 - 629, March 1974

Strang, Gilbert, Linear Algebra and its Applications, pp 166-176, HBJ Inc., 1988

A vector formulation for interferogram surface fitting

David J. Fischer
John T. O'Bryan
H. Philip Stahl

Rose-Hulman Institute of Technology, Department of Physics and Applied Optics
5500 Wabash Avenue
Terre Haute, IN 47803

ABSTRACT

Interferometry is an optical testing technique based on the interference of light. Fringes are formed when the Optical Path Difference (OPD) between a reference beam and an object beam is an integral multiple of the illuminating wavelength. This OPD is extracted through the process of sampling, ordering and interpolating. This paper develops a linear-algebra vector notation model of the interferogram sampling and interpolation process.

1. INTRODUCTION

Whenever two wavefronts exist at the same point in space, they superimpose, to create an interference pattern. The irradiance at any point, E(x,y), as determined by the Optical Path Difference (OPD) between the two wavefronts is described by:

$$E(x, y) = E_0 [1 + \cos(k\text{OPD})]$$

where

$$k = \frac{2\pi}{\lambda}$$

The two dimensional irradiance pattern created by the optical path difference at all points is called an irradiance pattern.

The challenge is to extract a representation of the test surface from such an interferogram. This is accomplished by a three step process: data sampling, ordering, and interpolating.

2. INTERFEROGRAM ANALYSIS

Because an interferogram has nearly infinite resolution, the actual number of points that can be measured is limited only by the sampling technique. There are two main methods for acquiring the data: digitalization, and phase-measuring interferometry. In digitalization, data points are acquired by sampling along the center of either the light or dark fringes. Usually, the interferogram is photographed and the image is placed onto a digitizing tablet where the data points are digitalized by a human operator. Alternatively, the image can be acquired with a video camera and digitalized using image-processing techniques. Phase-measuring interferometry, by contrast, is an electronic process where a video camera samples the wavefront phase at every point within its detector array - thus creating a high density, uniform grid of data. An illustration to show the difference in the two techniques is shown in Figure 1 and Figure 2:

SPIE Vol. 1779, 1992.

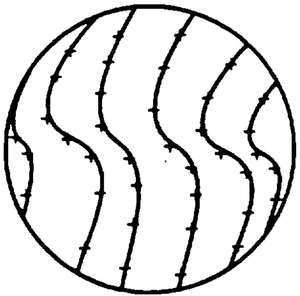


Fig. 1.An illustration of the placement of data points when an interferogram is sampled by digitalization

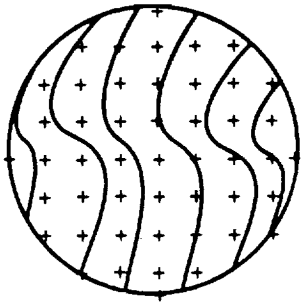


Fig. 2. An illustration of the placement of data points when an interferogram is sampled by Phase-Measuring Interferometry

For this paper, it is assumed that the data is sampled by manual digitalization along the dark fringes. The data is recorded in the form (m,x,y) where (x,y) gives the position of the point and (m) is the order of the sampled fringe - which, as will be discussed in the next paragraph, is related to the height of the surface in units of 2π.

Once the data has been sampled, it must be ordered. Because the irradiance pattern is proportional to cos(k*OPD) the data is in units of modulo 2π or modulo λ. Thus, even though a point in the interferogram represents a quantity greater than 2π, no measured quantity will be larger than 2π. For instance, an (x,y) point may have an (m) that is of magnitude 9π. The measured (m) will equal 0.5π, which is found by taking the magnitude modulo 2π. Therefore, the magnitude of the data must be artificially reconstructed by ordering the data. This is done by assigning integer order number values to each series of measured data points. For example, all data points along a given fringe will all have the same order and each adjacent fringe has an incrementally larger order number.

For manually digitalized data, this ordering is accomplished by the trained operator. For video digitalization, this is accomplished by image processing techniques. And for phase-measuring, this ordering is accomplished by a process called phase unwrapping. Once the data has been sampled and ordered, it must be interpolated to a polynomial set to represent the shape of the test surface. generally, this is accomplished using a least squares method with a Gram-Schmidt orthogonalization. From this fit several questions can be answered: how flat is the surface, are the features bumps or holes, are the features concave or convex, and what aberrations are present.

Traditionally, Zernike polynomials have been used for this representation for several reasons. They look like classical Seidel Aberrations. They provide a good fit to the circular aperture interferogram of most optical components. And, because they are orthogonal, they are easy to manipulate. It is possible to add or subtract coefficients without seriously affecting the quality of the fit.

To help visualize this three step process, two graphs are shown in Figure 3 and Figure 4. Figure 3 shows a one dimensional parabolic wavefront and the irradiance pattern due to the interference of the parabolic wave with a planar reference wave. This represents a slice of what would be a two dimensional interferogram. The pluses on the irradiance pattern represent the sampled points on the dark fringes. The numbers below the sampled points are the integer ordering. Figure 4 is a plot of the ordered points. It is easily seen that a parabolic function could be interpolated to those points to represent the test wave.

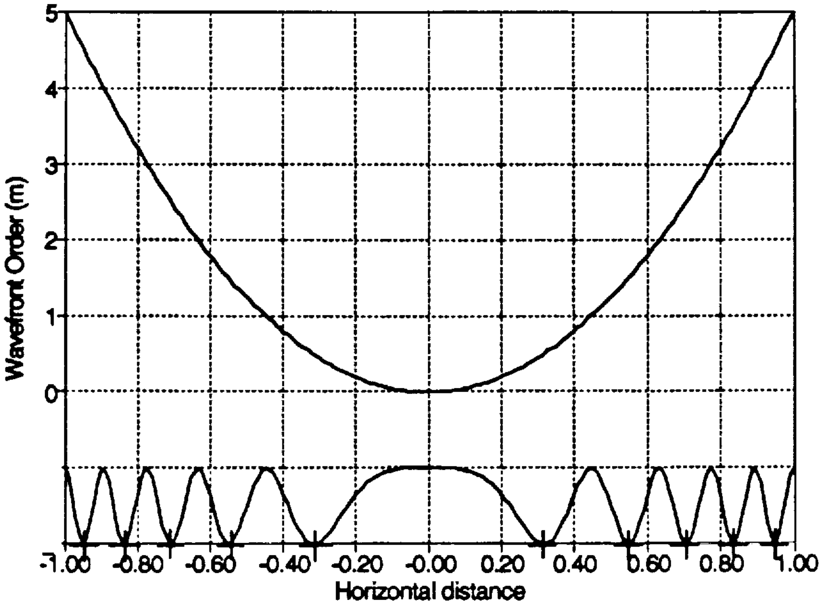


Fig. 3. A1D slice of a parabolic wavefront (top) and the irradiance pattern formed when this slice interferes with a plane wavefront (bottom). The pluses represent data points sampled at irradiance minima.

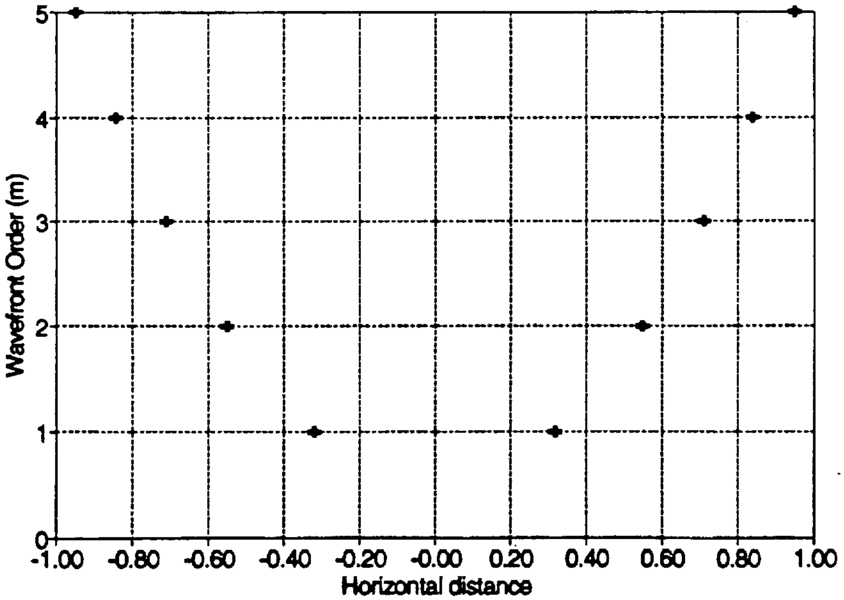


Fig. 4. A plot of correctly ordered data points sampled from the irradiance pattern shown in Fig. 3

3. VECTOR FORMULATION

The purpose of this section is to describe simply and clearly the analysis process of an interferogram using general linear-algebra vector notation, and to develop a generalized mathematical model with which to explore the effects of: sampling parameters, fitting polynomials, and the interpolation algorithm on the accuracy of the reconstruction of the test surface. The remainder of this paper presents a general vector based formulation for the interpolation process.

Any point in the interferogram, Z(x,y), can be represented by a series of M polynomials, F(x,y), and their coefficients, G.

$$Z(x,y) \approx \sum_{j=1}^{M} G_j F_j(x,y)$$

Alternatively, the single point, Z(x,y), can also be expressed as a matrix of polynomials multiplied by a matrix of coefficients.

$$\begin{bmatrix} Z(x,y) \end{bmatrix} = \begin{bmatrix} F_1(x,y) & \dots & F_M(x,y) \end{bmatrix} \begin{bmatrix} G_1 \\ \vdots \\ G_M \end{bmatrix}$$

Thus, if a sampled interferogram can be thought of as a series of N measured data points.

$$\begin{bmatrix} Z(x_1, y_1), Z(x_2, y_2), \dots, Z(x_N, y_N) \end{bmatrix}$$

Then it follows that the complete set of all N data points can be represented as a polynomial matrix multiplied by the coefficients matrix.

$$\begin{bmatrix} Z(x_1, y_1) \\ \vdots \\ Z(x_N, y_N) \end{bmatrix} \approx \begin{bmatrix} F_1(x_1, y_1) & \dots & F_M(x_1, y_1) \\ \vdots & & \vdots \\ F_1(x_N, y_N) & \dots & F_M(x_N, y_N) \end{bmatrix} \begin{bmatrix} G_1 \\ \vdots \\ G_M \end{bmatrix} \qquad (1)$$

In order to simplify this notation, the data points can be considered to be a vector of length N, a given polynomial evaluated at all N data points is a vector of length N, and the coefficients are a vector of length M.

$$\vec{Z} = \begin{bmatrix} Z(x_1, y_1) \\ \vdots \\ Z(x_N, y_N) \end{bmatrix} \quad \vec{F}_i = \begin{bmatrix} F_i(x_1, y_1) \\ \vdots \\ F_i(x_N, y_N) \end{bmatrix} \quad \vec{G} = \begin{bmatrix} G_1 \\ \vdots \\ G_M \end{bmatrix} \qquad (2)$$

Using this notation, the polynomial matrix can be expressed using F vectors.

$$\mathcal{S} = \begin{bmatrix} \vec{F}_1 & \dots & \vec{F}_M \end{bmatrix} \qquad (3)$$

Thus, rewriting equation 1, the surface can be represented as:

$$\vec{Z} \approx \mathcal{S}\vec{G}$$

The polynomials can be fit using a least squares method, where S is defined as the sum of the square of the difference between the data points and the fitted polynomials. If a perfect fit were possible, then S would be equal to zero, because there would be no difference between the data points and the representing polynomials evaluated at those points. However, there are always differences between the real surface and its representation. Therefore, S will be non-zero. The function of a least squares fit is to find the coefficients, for a given set of polynomials, which minimizes S. This process can be described using vector notation

$$\vec{S} \equiv \left| [\vec{Z} - \mathcal{S}\vec{G}] \right|^2 = minimum$$

or as more commonly expressed using summations:

$$S \equiv \sum_{r=1}^{N} \left[Z(x_r, y_r) - \sum_{j=1}^{M} G_j F_j (x_r, y_r) \right]^2 = minimum$$

The polynomial coefficients are found by taking the derivative of S with respect to each coefficient and setting the result equal to zero. For example, minimizing S for G_k gives:

$$\sum_{r=1}^{N} Z(x_r, y_r) F_k (x_r, y_r) = \sum_{r=1}^{N} \sum_{j=1}^{M} G_j F_j (x_r, y_r) F_k (x_r, y_r)$$

The result of minimizing S for all coefficients can be expressed using summations in matrix form as:

$$\begin{bmatrix} \sum_{r=1}^{N} Z(x_r, y_r) F_1 (x_r, y_r) \\ \vdots \\ \sum_{r=1}^{N} Z(x_r, y_r) F_M (x_r, y_r) \end{bmatrix} = \begin{bmatrix} \sum_{r=1}^{N} F_1 (x_r, y_r) F_1 (x_r, y_r) & \dots & \sum_{r=1}^{N} F_M (x_r, y_r) F_1 (x_r, y_r) \\ \vdots & & \vdots \\ \sum_{r=1}^{N} F_1 (x_r, y_r) F_M (x_r, y_r) & \dots & \sum_{r=1}^{N} F_M (x_r, y_r) F_M (x_r, y_r) \end{bmatrix} \begin{bmatrix} G_1 \\ \vdots \\ G_M \end{bmatrix}$$

This expression can be greatly simplified with vector notation to give:

$$\begin{bmatrix} \vec{Z} \bullet \vec{F}_1 \\ \vdots \\ \vec{Z} \bullet \vec{F}_M \end{bmatrix} = \begin{bmatrix} \vec{F}_1 \bullet \vec{F}_1 & \dots & \vec{F}_M \bullet \vec{F}_1 \\ \vdots & & \vdots \\ \vec{F}_1 \bullet \vec{F}_M & \dots & \vec{F}_M \bullet \vec{F}_M \end{bmatrix} \vec{G}$$

where the dot product of the vectors F_i and F_j gives the sum of the products of both polynomials over every data point.

$$\vec{F}_i \bullet \vec{F}_j = \sum_{r=1}^{N} F_i (x_r, y_r) F_j (x_r, y_r)$$

Likewise the dot product of the vectors Z and F_i gives the sum of the products evaluated at all data points.

$$\vec{Z} \bullet \vec{F}_j = \sum_{r=1}^{N} Z(x_r, y_r) F_j (x_r, y_r)$$

Finally, the minimization of S can also be written using the polynomial matrix and its transpose:

$$\mathfrak{J}^T \vec{Z} = \mathfrak{J}^T \mathfrak{J} \vec{G}$$

where the transpose of a matrix is found by rotating its columns so that they become its rows. In the case of the polynomial matrix (See Equ. 3), the columns are composed of polynomial vectors. The polynomials vectors are rotated to become the rows of the polynomial matrix, as shown below:

$$\mathfrak{S}^T = \begin{bmatrix} \vec{F}_1 \\ \vdots \\ \vec{F}_M \end{bmatrix}$$

If the F vectors are orthogonal then the minimization process produces a diagonal matrix

$$\begin{bmatrix} \vec{Z} \bullet \vec{F}_1 \\ \vdots \\ \vec{Z} \bullet \vec{F}_M \end{bmatrix} = \begin{bmatrix} \vec{F}_1 \bullet \vec{F}_1 \ \ldots & 0 \\ \vdots & \vdots \\ 0 & \ldots \ \vec{F}_M \bullet \vec{F}_M \end{bmatrix} \vec{G} \tag{4}$$

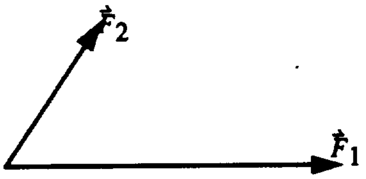
and the coefficients can be found by multiplying both sides of the equation by the transpose of the diagonal matrix.

$$\vec{G} = \begin{bmatrix} \dfrac{\vec{Z} \bullet \vec{F}_1}{\|\vec{F}_1\|} \\ \vdots \\ \dfrac{\vec{Z} \bullet \vec{F}_M}{\|\vec{F}_M\|} \end{bmatrix} \tag{5}$$

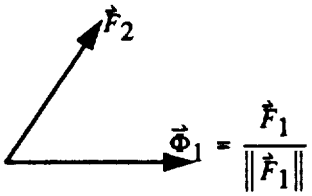
However, most orthogonal polynomials are only orthogonal over a continuous range, not over discrete points. Thus, the solution process is not as easy as it first seems. To overcome this limitation, There are three possible approaches. The first is to assume that the polynomials are orthogonal. This may work if there is enough sampled data to approximate a continuous range, but it can introduce error. The second is to take the transpose of the non-diagonal matrix, but this can be numerically difficult and inaccurate. And third, the polynomials can be orthogonalized. Typically, this is accomplished using a Gram-Schmidt technique. Although this introduces extra work, it should pay for itself in the ease and accuracy it returns in the final solution.

The Gram-Schmidt orthogonalization process takes N arbitrary vectors in an N dimensional space (N-space) and creates N ortho-normal vectors that form a basis for that space. Ortho-normal vectors are vectors which are orthogonal to each other, and are normalized, or have a unit length. A basis for an N-space is essentially a coordinate system for that space, since the basis vectors are linearly independent and all other vectors in that space may be formed from linear combinations of the basis vectors.

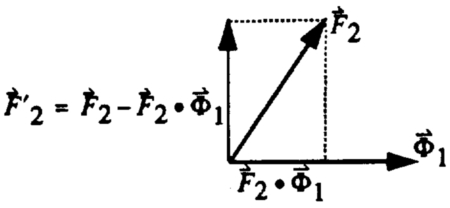
Consider two arbitrary vectors, F_1 and F_2, which are wanted to be replaced by their orthogonal counterparts, Φ_1 and Φ_2.



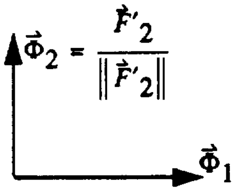
To begin the process, the first vector is normalized by dividing by its magnitude.The result is defined as its orthogonalized counterpart.



After the first vector is orthogonalized, all other vectors are orthogonalized, in a systematic manner, by projecting the current vector onto each preceding orthogonalized vector and subtracting that projection from the vector leaving the orthogonal component.



This component is then normalized by dividing by its magnitude, and renaming it to its orthogonal counterpart. This process is repeated for each polynomial vector in the system.



For the interpolation process, the polynomial vectors, F (See Equ. 2), are orthogonalized and replaced by the new vectors, Φ. Because the original fitting polynomials, F, are replaced by Φ, a different set of coefficients are found. Therefore, the original coefficients, G, are replaced by Γ. The system can now be rewritten using the new vectors, and because the Φ vectors are orthogonal, the matrix is diagonal.

$$\begin{bmatrix} \vec{z} \bullet \vec{\Phi}_1 \\ \vdots \\ \vec{z} \bullet \vec{\Phi}_M \end{bmatrix} = \begin{bmatrix} \vec{\Phi}_1 \bullet \vec{\Phi}_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \vec{\Phi}_M \bullet \vec{\Phi}_M \end{bmatrix} \vec{G}$$

The system is then solved as shown previously by multiplying both sides by the transpose of the matrix. (See Equ. 4, 5) The Γ coefficients are now known and can be used to represent the original surface.

$$\vec{\Gamma} = \begin{bmatrix} \dfrac{\vec{z} \bullet \vec{\Phi}_1}{\|\vec{\Phi}_1\|} \\ \vdots \\ \dfrac{\vec{z} \bullet \vec{\Phi}_M}{\|\vec{\Phi}_M\|} \end{bmatrix}$$

4. CONCLUSION

Interferometry is an optical testing technique in which information about a test surface is extracted from an interferogram. This is done with three main steps. First, the data is sampled, usually by one of two main methods: digitalization or phase-measuring interferometry. Second, the sampled data points are ordered. And third, the ordered data is interpolated to a polynomial set. Typically, Zernike polynomials are used and are fitted with a least squares method and Gram-Schmidt orthogonalization.

In order to describe the interpolating process, a generalized linear-algebra vector notation has been developed in which the data points, fitting polynomials, and coefficients are viewed as vectors. A least squares fit is applied to the vector system representing the test surface. The polynomial vectors are ortho-normalized via the Gram-Schmidt process and then system is solved for the coefficients vector.

5. ACKNOWLEDGMENTS

This project is supported by a NASA Research Grant (NA63-1300) from NASA-Lewis Research Center.
We would like to thank to Dr. Robert Lopez for his help with the mathematics.
We would like to thank John O'Bryan for the time and effort he put into this research.

6. BIBLIOGRAPHY

Koliopoulis, Chris L., "Fringe Analysis of Interferograms", Laser & Optrionics, pp 65 - 75, May 1985
Koliopoulis, Christ L., "Interferometric Optical Measurement Techniques", pp 1 - 67, 1981
Malacara, Daniel, J. Martin and Caprio-Valadez and and J. Javier Sancbez-Mondragon, "Wavefront Fitting with Discrete Orthogonal Polynomials in a Unit Radius Circle", Optical Engineering, Vol. 24, No. 6, pp 672 - 675, June 1990
Nobles, Ben, Applied Linear Algebra, pp 9-18, Prentice-Hall Inc., 1969
Strang, Glbert, Linear Algebra and its Applications, pp 166-176, HBJ Inc., 1988

APPENDIX B

University of Arizona Optical Sciences Center Colloquium

Presentation Viewgraphs

Surface Deformation Measuring Instrument

for the

Surface Tension Driven Convection Experiment (STDCE-2)

by:

H. Philip Stahl, Ph.D.

STDCE

- The Surface Tension Driven Convection Experiment (STDCE) is a fundamental fluid physics experiment designed to provide quantitative data on the thermocapillary flow of fluid under the influence of an increased localized surface temperature.

- STDCE flew on the Space Shuttle Columbia in the First United States Microgravity Laboratory (USML-1) in June 1992.

- The second flight of this experiment (STDCE-2) is scheduled for flight in the Fall of 1995.

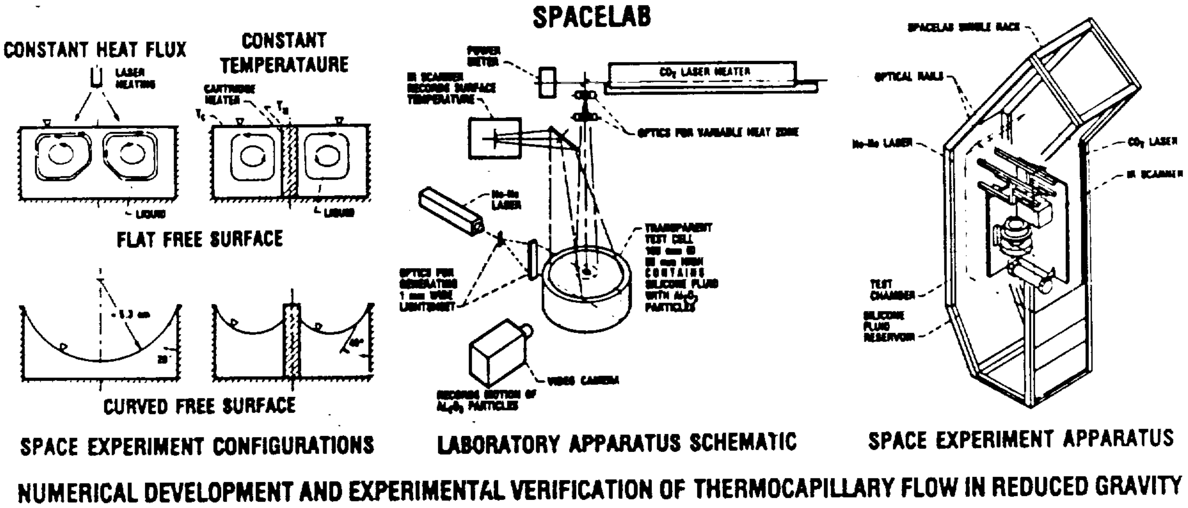


Fig. 1 Surface Tension Driven Convection Experiment.

Motivation

- On Earth, buoyancy flow caused by gravity limits our ability to:

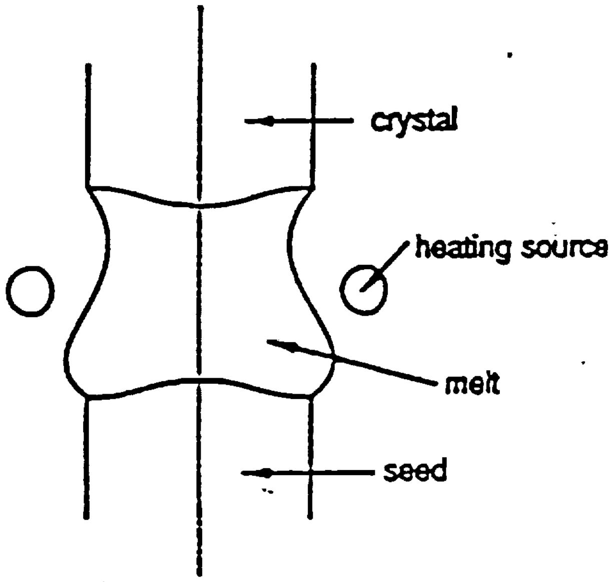
  grow semi-conductor crystals,
  grow biological crystals, and
  solidify metal alloys.

  by introducing imperfections into the lattice structure.

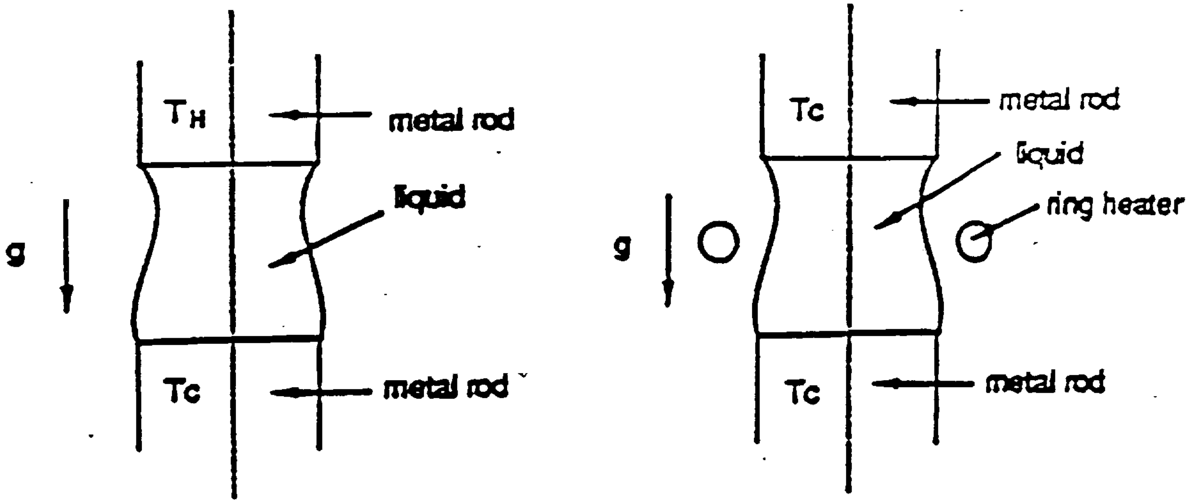
- Also, gravity causes lattice structures to 'sag' and limits the size of crystal which can form in suspension.

- In microgravity, buoyancy flow is reduced to 10^{-6}, offering great promise for material processing in space.

- However, thermocapillary flow still exists and must be understood.



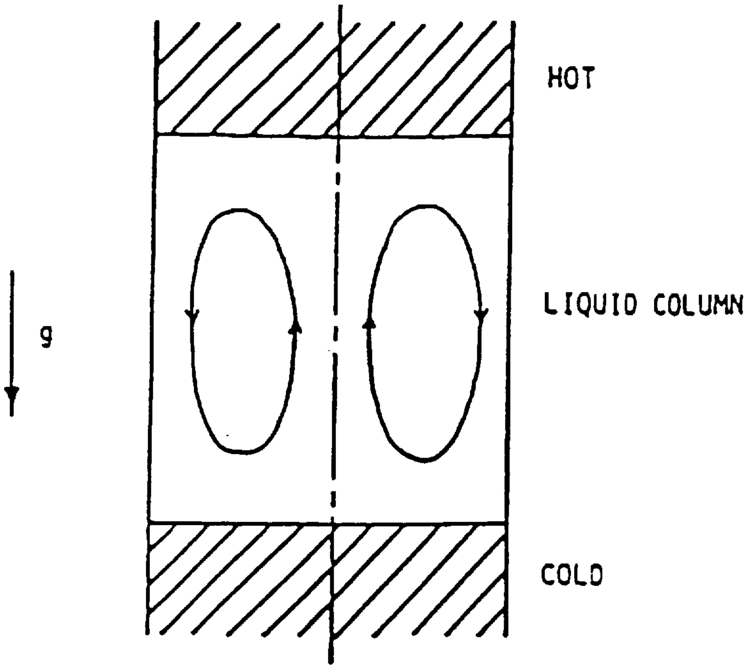
Floating Zone Crystal Growth



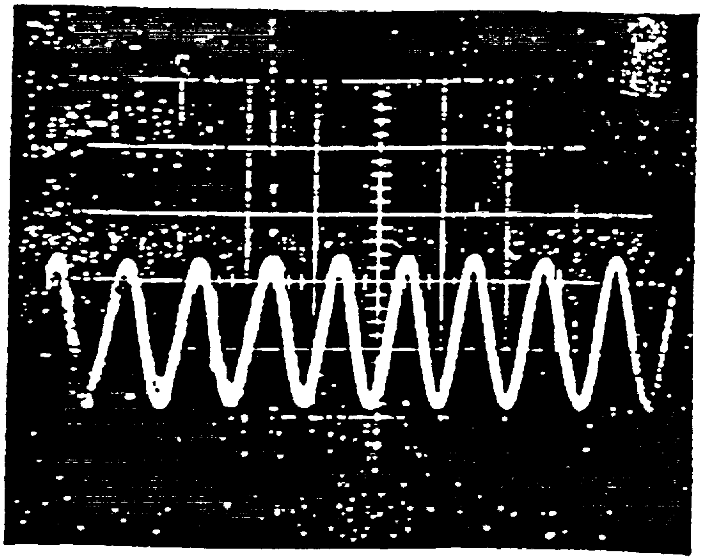
Half-Zone Simulation

Full-Zone Simulation

FLOATING-ZONE SIMULATION



FLOATING-ZONE SIMULATION



TEMPERATURE OSCILLATION PATTERN IN LIQUID

Surface Tension

- Surface tension is a property of a liquid which tends to minimize its surface area. For, example, surface tension causes nearly spherical drops to form when rain falls.

- On Earth, gravity forces usually overpower surface tension forces, and liquids form flat surfaces.

Convection

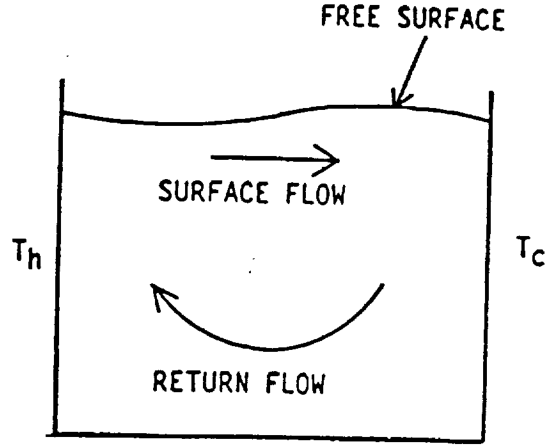
- Convection refers to the flow of a fluid. Natural convection is driven by gravity: colder, heavier fluid is pulled down, causing warmer, lighter fluid to rise.

- On Earth, this buoyancy driven convection occurs whenever there is a temperature difference within a fluid.

- Convection can also be driven by surface tension. As its temperature is increased, a liquid has a lower surface tension. Warmer surface liquid is pulled toward colder surface liquid, which has a higher surface tension. As the warmer liquid moves across the surface, it draws more liquid up to the surface, creating a convection current.

- This surface tension driven convection, also called thermocapillary flow, occurs whenever there is a temperature difference across the surface of a liquid.

BASIC FLOW STRUCTURE



SURFACE FLOW : driven by thermocapillarity and viscosity

RETURN FLOW : driven by pressure gradient

Thermocapillary Flow

- Thermocapillary flow has two states:

  Steady State, and
  Oscillatory.

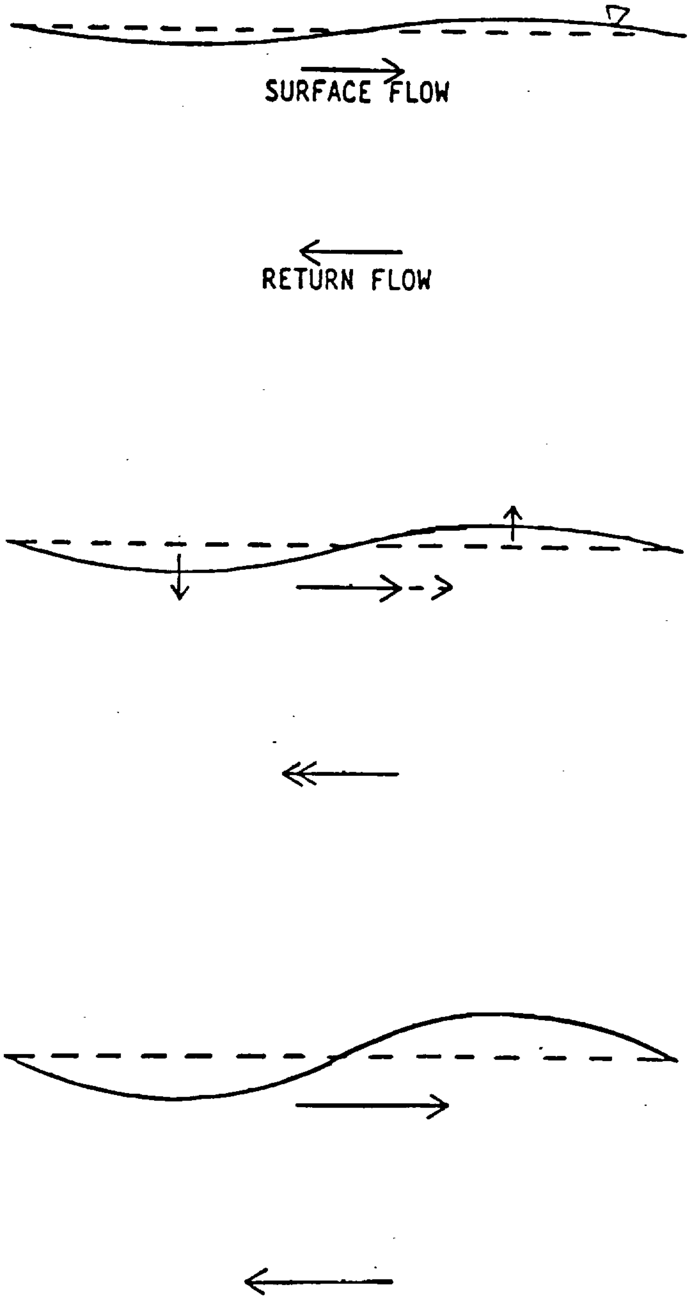
- Steady flows are two-dimensional. They are the same for any symmetrical observer.

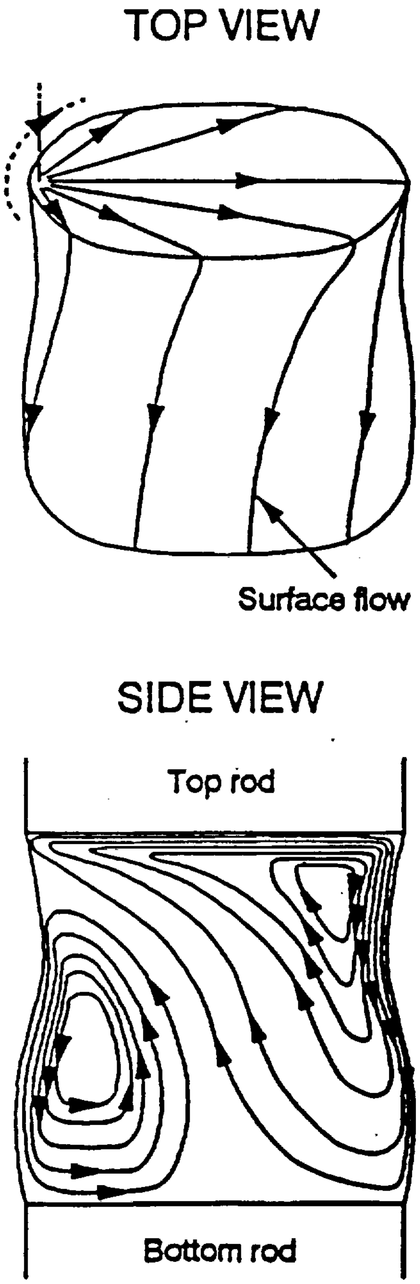
- Oscillatory flows are three-dimensional.

- Surface deformations produced by steady state thermocapillary flow may produce oscillatory flow.

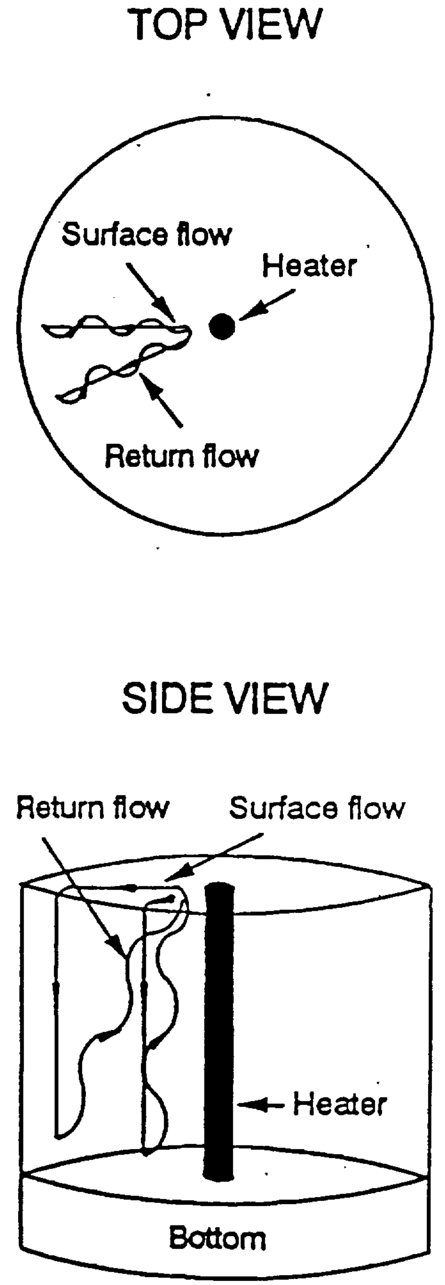
- Oscillatory flow may be detrimental to the crystal growth process and needs to be studied.

TIME LAG IN RETURN FLOW RESPONSE





HALF - ZONE



CIRCULAR CONTAINER

Science Objectives

- The objective of STDCE is to determine the extent and nature of thermocapillary flow as a function of:

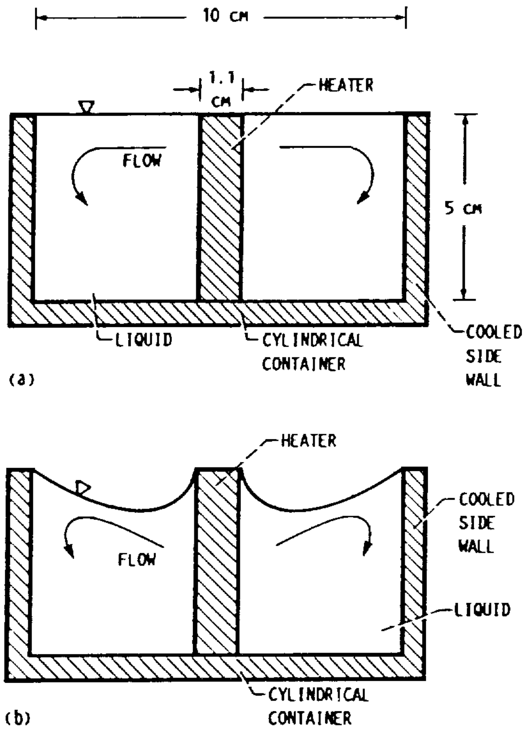
  heating mode and level, and
  liquid free-surface shape.

- The objective of STDCE-2 is to determine the onset conditions for and nature of oscillatory flow as a function of:

  heating mode and level,
  liquid free-surface shape, and
  container size.

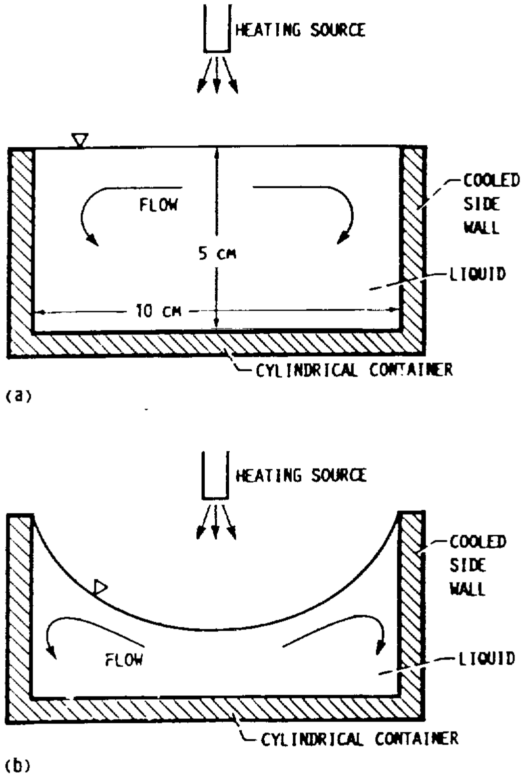
- To accomplish these objectives, three parameters must be monitored:

  surface temperature distribution,
  bulk fluid flow, and
  surface deformation.



(a) Flat free surface.
(b) Curved free surface.

Figure 1.—Constant temperature experiment.



(a) Flat free surface.
(b) Curved free surface.

Figure 2.—Constant flux experiment.

STDCE Configuration

- The STDCE-2 hardware consists of five basic modules:

  Fluid Chamber,
  Heat Source (CO_2 Laser or Cartridge)
  Infrared Imager,
  Flow Visualization, and
  Surface Deformation.

- In STDCE, the fluid was in a single 10 cm diameter container. In STDCE-2, there will be three containers: 1.2, 2, and 3 cm.

- The fluid is heated by illuminating the surface with a CO_2 laser beam or a cartridge heater.

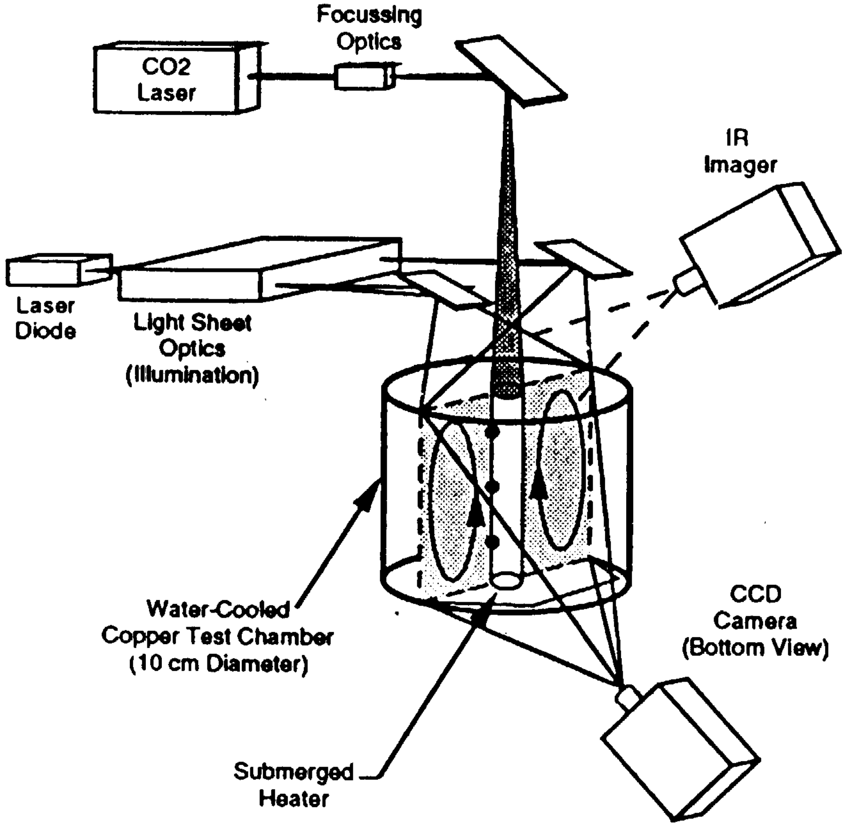
- Surface temperature is measured by a HgCdTe thermal imager.

- Bulk fluid flow is measured by monitoring the motion of microscopic particles which have been uniformly mixing into the fluid. This motion is tracked by shinning a 'sheet' of light through the fluid and observing the light scattered from the particles.

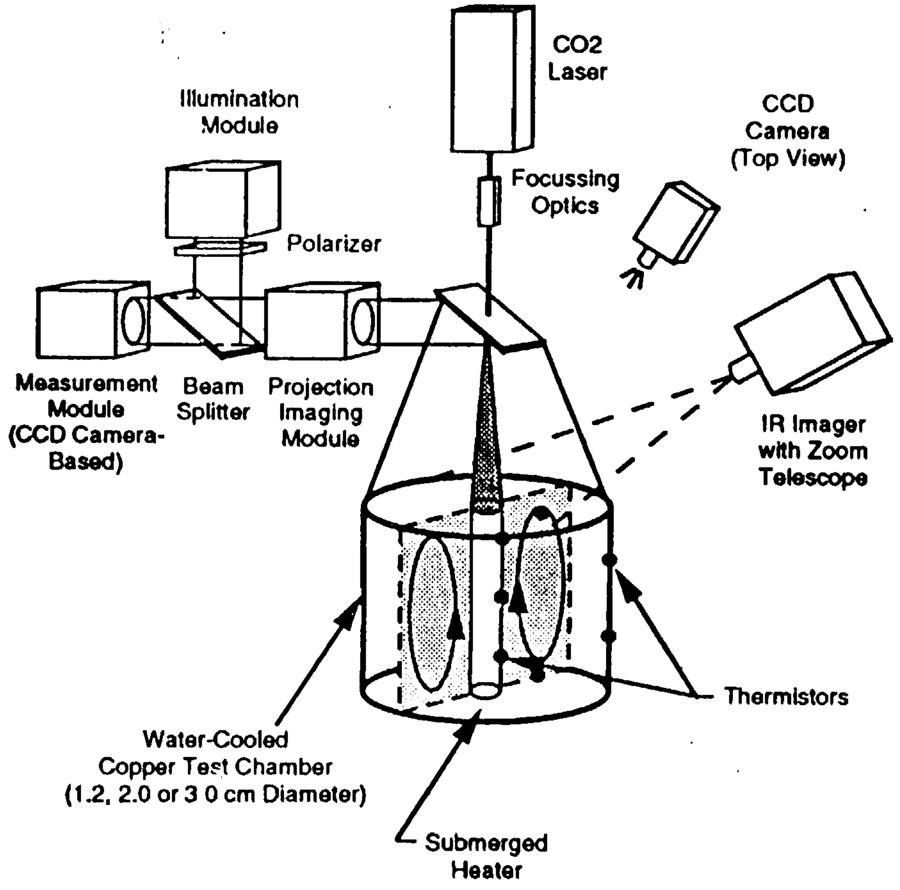
- STDCE had no means for measuring surface deformation. But, STDCE-2 will.

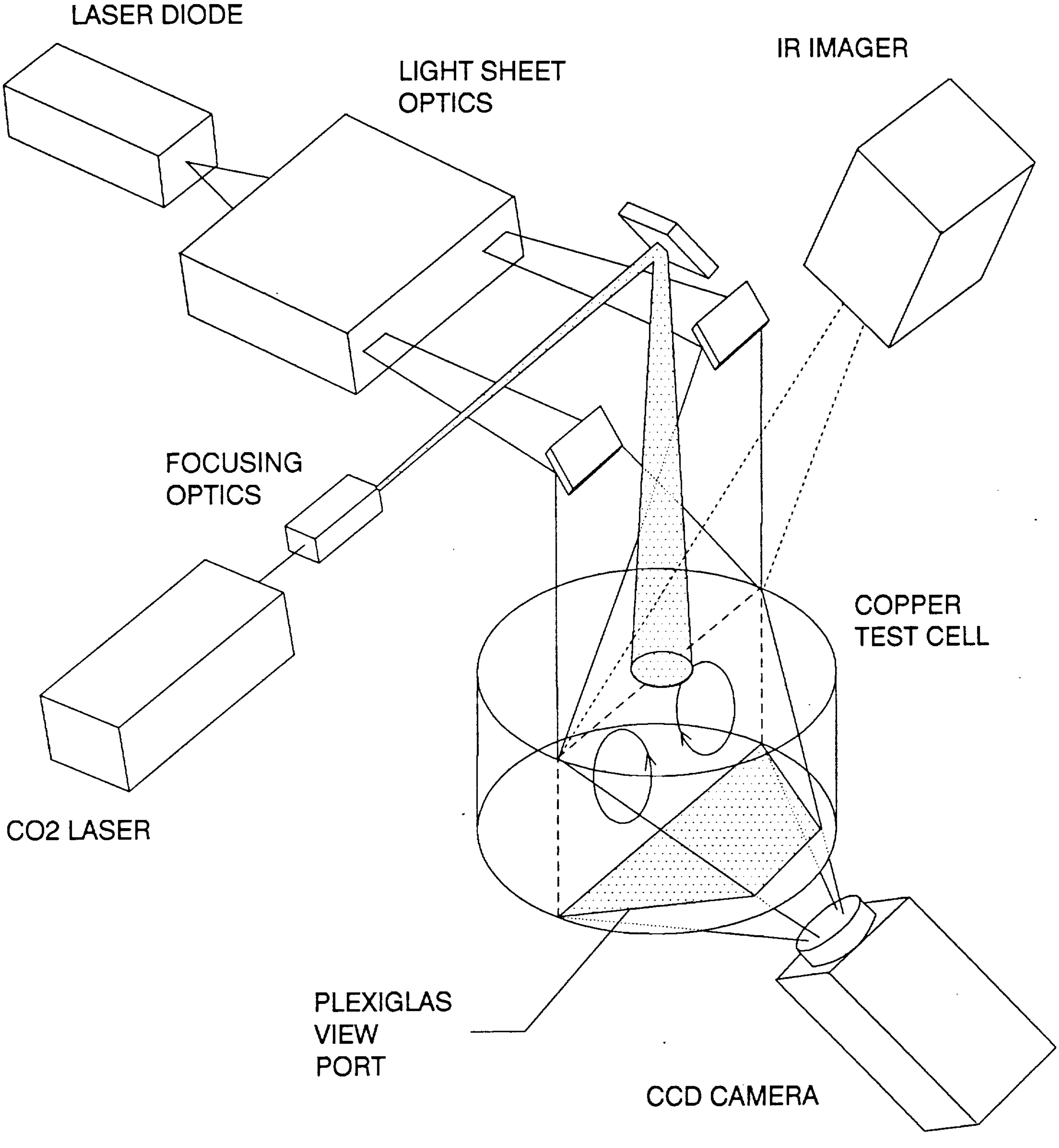
Comparison of STDCE-1 and STDCE-2

STDCE-1



STDCE-2




LASER DIODE
LIGHT SHEET OPTICS
IR IMAGER
FOCUSING OPTICS
COPPER TEST CELL
CO2 LASER
PLEXIGLAS VIEW PORT
CCD CAMERA

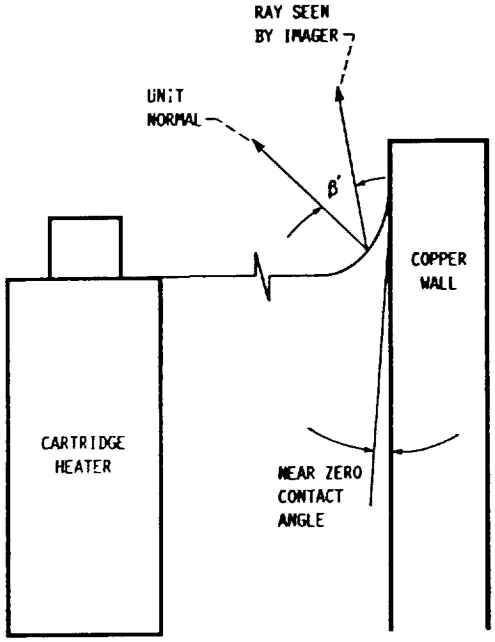


Figure 21.—Meniscus at edge of test cell.

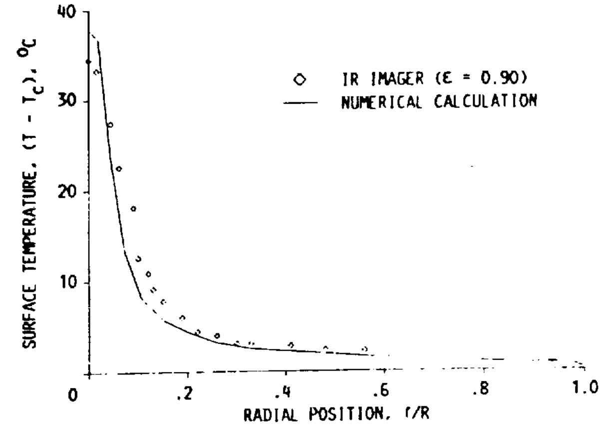


Figure 22.—Comparison of numerical and IR imager surface temperature distribution for constant flux experiment with $\epsilon = 0.90$, $p = 2.7$ W, $T_c = 22.7$ °C, and beam diameter = 5 mm. Numerical calculation used $\mathrm{Ma} = 1.27 \times 10^5$ and $\mathrm{Ra} = 4.87 \times 10^7$.

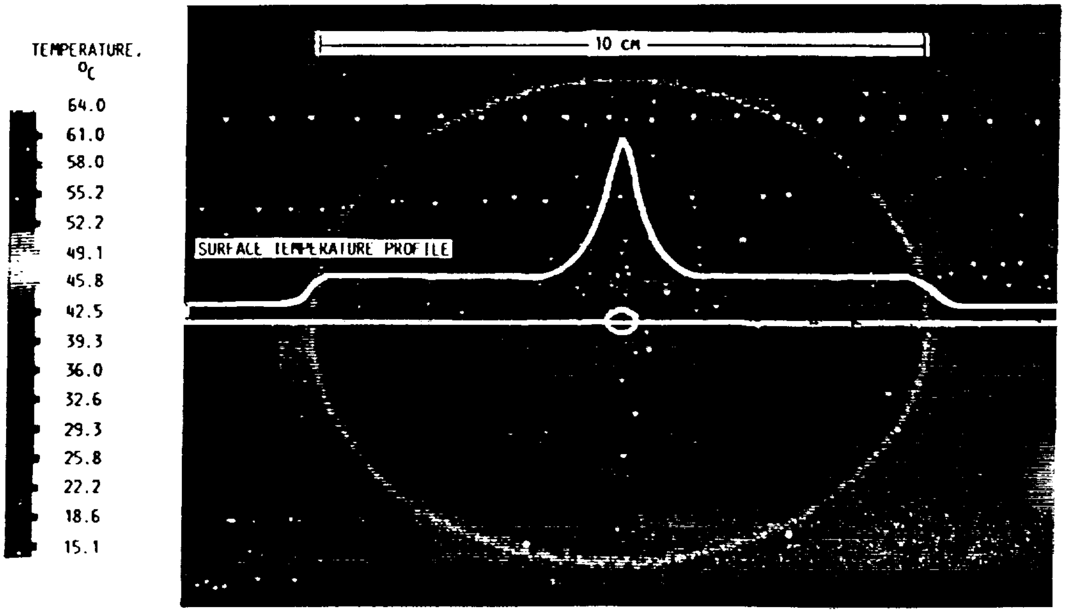


Figure 23.—Constant flux experiment IR image for $T = 38.7$ °C, $R = 50$, $\epsilon = 0.90$, $p = 2.7$ W, and beam diameter = 5 mm.

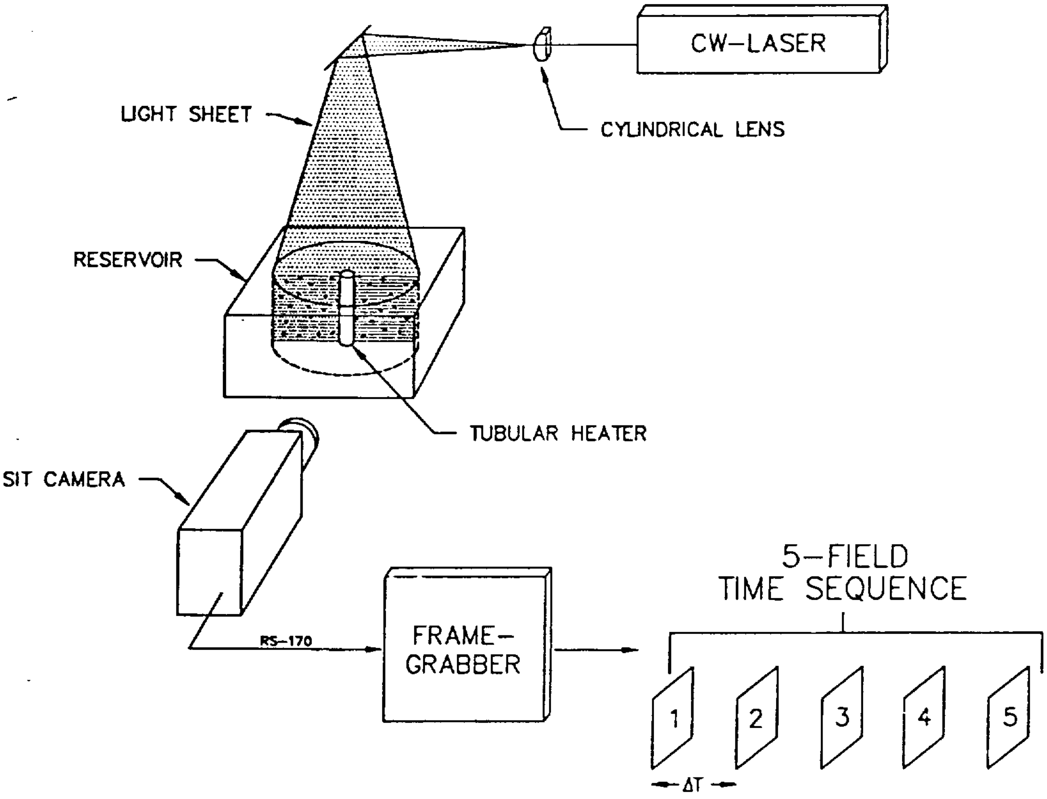


Figure 3 Schematic view of the light sheet illumination, fluid flow reservoir, and data acquisition and processing system. The tubular heater in the reservoir drives the flow. The frame–grabber board digitizes a sequence of video fields separated by ΔT, which is defined in integral multiples of 1/60 second.

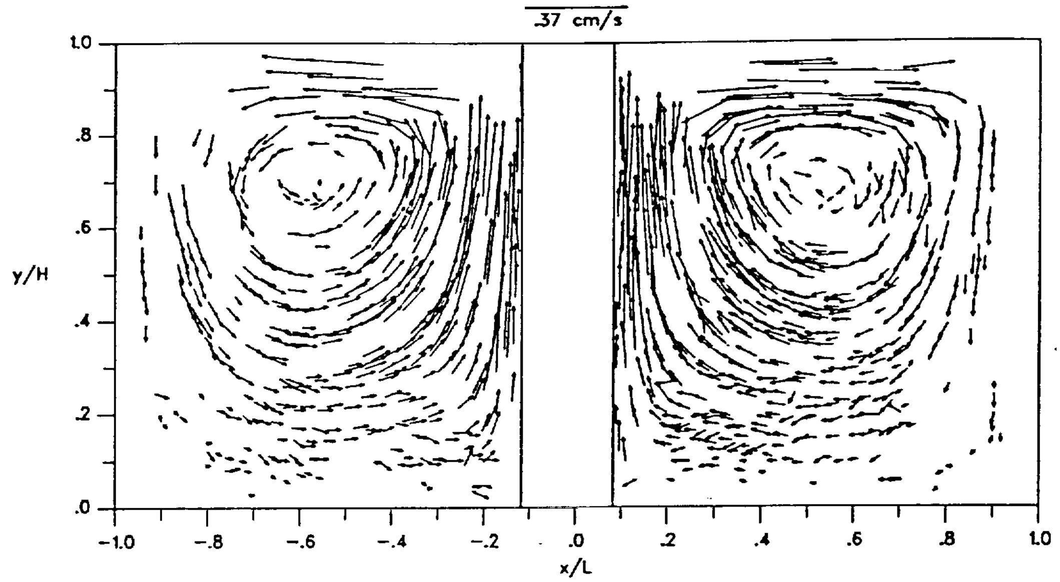


Figure 6 Combined data sequences #1 and #2. A large dynamic range of velocities have been recorded. Many sets of tail to head velocity vectors are observed, indicating particles which were tracked over the entire sampling period.

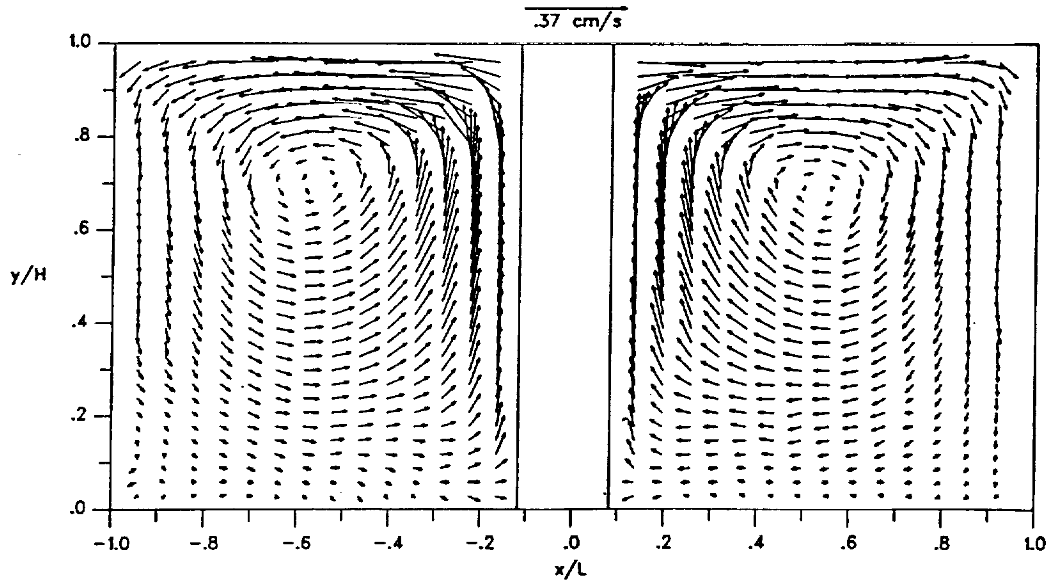


Figure 7 Interpolated 2–D velocity vector plot. The combined data from sequences #1 and #2 have been interpolated over a 32×32 grid. The interpolated data clearly show the counter rotating flow pattern.

Surface Deformation

- On Earth it has been observed that both steady-state and oscillatory thermocapillary flow states produce deformations of the air/oil free surface.
- Steady state flow produces two characteristic steady state surface deformations:

  a 'Pimple' (bump), and
  a 'Dimple' (hole).

- Pimples are produced when the surface is exposed to an external temperature source over a small area.
- Dimples are produced when the same temperature input is spread out over a larger surface area.
- Oscillatory flow produces both standing wave and rotary surface deformations.

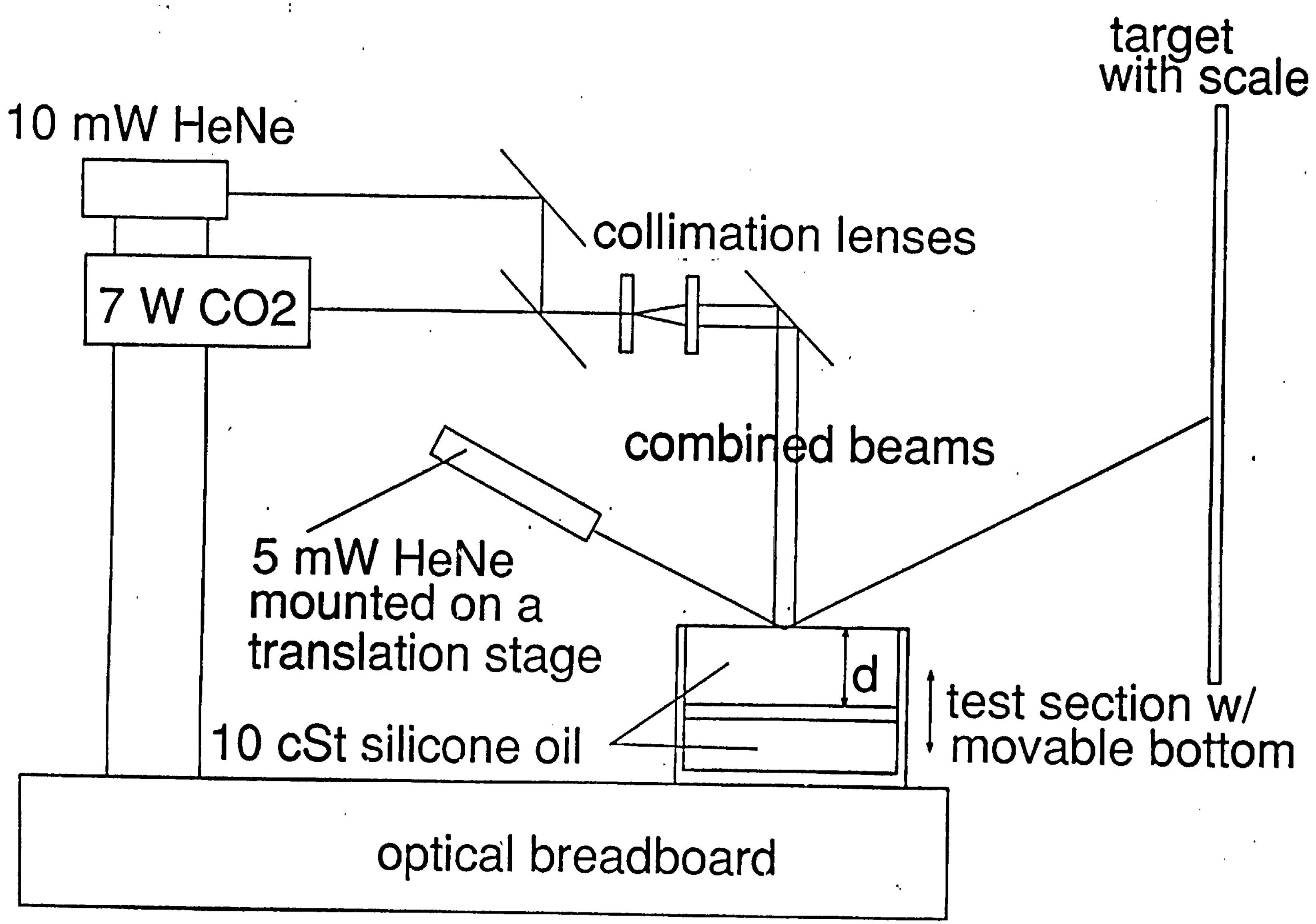


figure 1 file:defpap1.drw

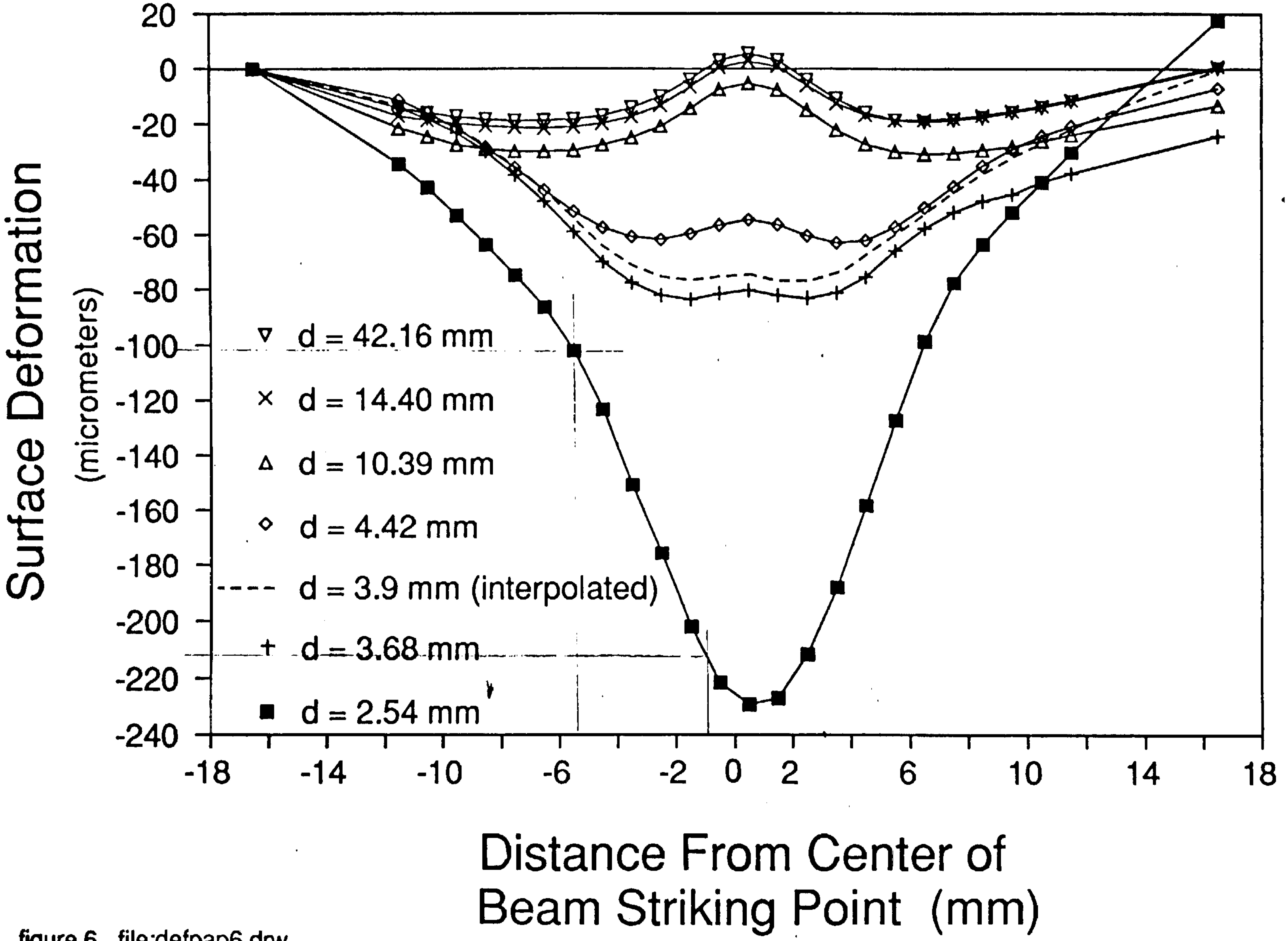
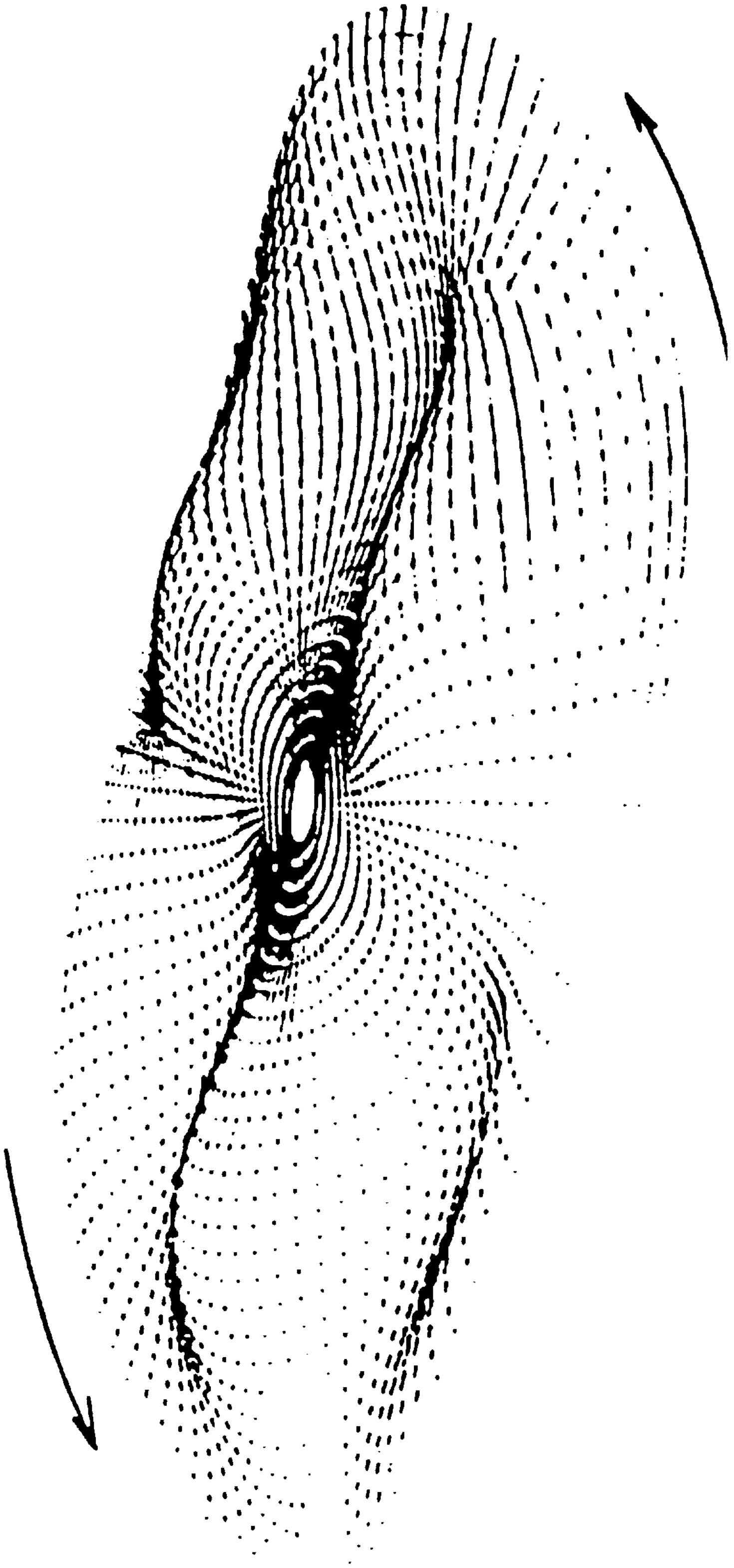


figure 6 file:defpap6.drw



Purpose

- The purpose of this project is to design, build, calibrate, and fully characterize one or more non-contact (i.e. optical) prototype instruments for measuring the surface deformation of an air-oil free-surface under the influence of a localized thermal load (i.e. CO2 laser beam or electric heater) which can be integrated into the anticipated re-flight of the Surface Tension Driven Convection Experiment (STDCE-2)
- While initially these measurements will be performed in a laboratory environment, the instrument must be capable of operating in a Space Transport System (STS) environment for the United States Microgravity Laboratory flights USML-2 and/or USML-3.

Measurement Approach

- Because the surface is a fluid, it cannot be contoured mechanically. It must be contoured optically.

- There are two ways to optically contour a large two-dimensional reflective surface:

  interferometrically or
  geometrically.

- Generally, it is easier and more accurate to use an interferometer than to perform a geometrical test because interferometers measure a surface's shape while most geometrical tests measure a surface's slope.

- Unfortunately, because of the large surface deformations, conventional visible interferometry is not practical. The fringes will be too dense for the camera to resolve.

- One solution to this problem is to use an infrared interferometer, but such a technique might interfere with the localized heating process.

- Therefore, classical geometrical techniques were evaluated: Foucault knife-edge test, wire test, Hartmann screen test, Ronchi test, grating interferometry, or reflective moire.

Geometrical Techniques

- Geometrical techniques measure the surface's slope instead of its height. Thus, there are specific advantages and disadvantages:
- Advantage: they are insensitive to tilting of the free-surface cause by G-Jitter.
- Disadvantage: two orthogonal data sets are required to reconstruct its shape. These two data sets can be either x-slope and y-slope or r-slope and θ-slope.

Surface Symmetry

- A fundamental question which has a profound impact upon the selection of the final approach is whether or not the surface is rotationally symmetric.

- If the surface is rotationally symmetric, then it may be possible to satisfy the mission's science requirements with just x-slope or r-slope.

- But, if the surface is non-rotationally symmetric, then it is necessary to acquire two orthogonal data sets.

- The surface deformations are not rotationally symmetric.

- But, their outer edge is at a constant height.

Candidate Techniques

- Three techniques which provide data over the entire aperture were evaluated:

  Ronchi grating,
  Hartmann test, and
  Projected grid.

Projected Grid

- In the projected grid approach, a linear or crossed grid pattern is placed directly in front of an extended light source (illuminated diffuser screen).

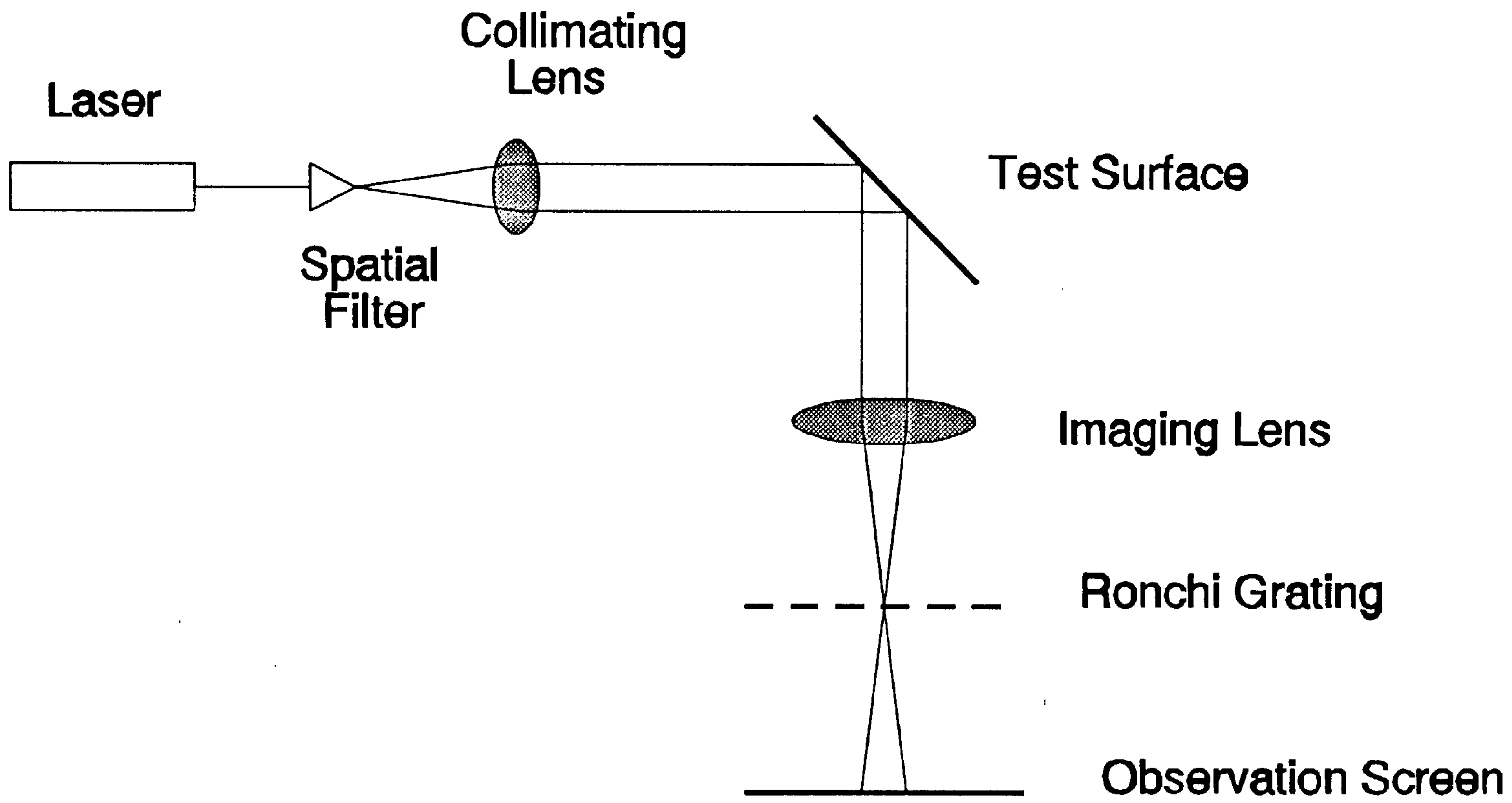
- This pattern is then imaged onto the reflective sample surface.

- The reflected wavefront is re-imaged by another optical system onto the video camera.

- The projected grid approach was base lined early but eliminated because it lacked sensitivity, dynamic range, and flexible spatial sampling.

Ronchi Grating

- In the Ronchi grating test, a collimated beam of light is reflected off of the sample surface.
- The reflected wavefront is focused through a grating to form a Ronchigram which is imaged onto a video camera.
- The Ronchi approach has many advantages. By selection of grating it has variable sensitivity and dynamic range. By location of grating it has flexible spatial sampling.
- Unfortunately, it does not provide simultaneous orthogonal slope data sets.


Laser
Spatial Filter
Collimating Lens
Test Surface
Imaging Lens
Ronchi Grating
Observation Screen

Hartmann Screen Test

- The Hartmann screen test can provide both x- and y-slope information simultaneously.

- Unfortunately, it has limited spatial resolution.

- And, for this application, it is not practical because the screen would obscure the upper surface flow visualization function.

- A practical implementation of the Hartmann test is to use crossed gratings in a Ronchi test. However, the complex patterns produced are difficult to analyze.

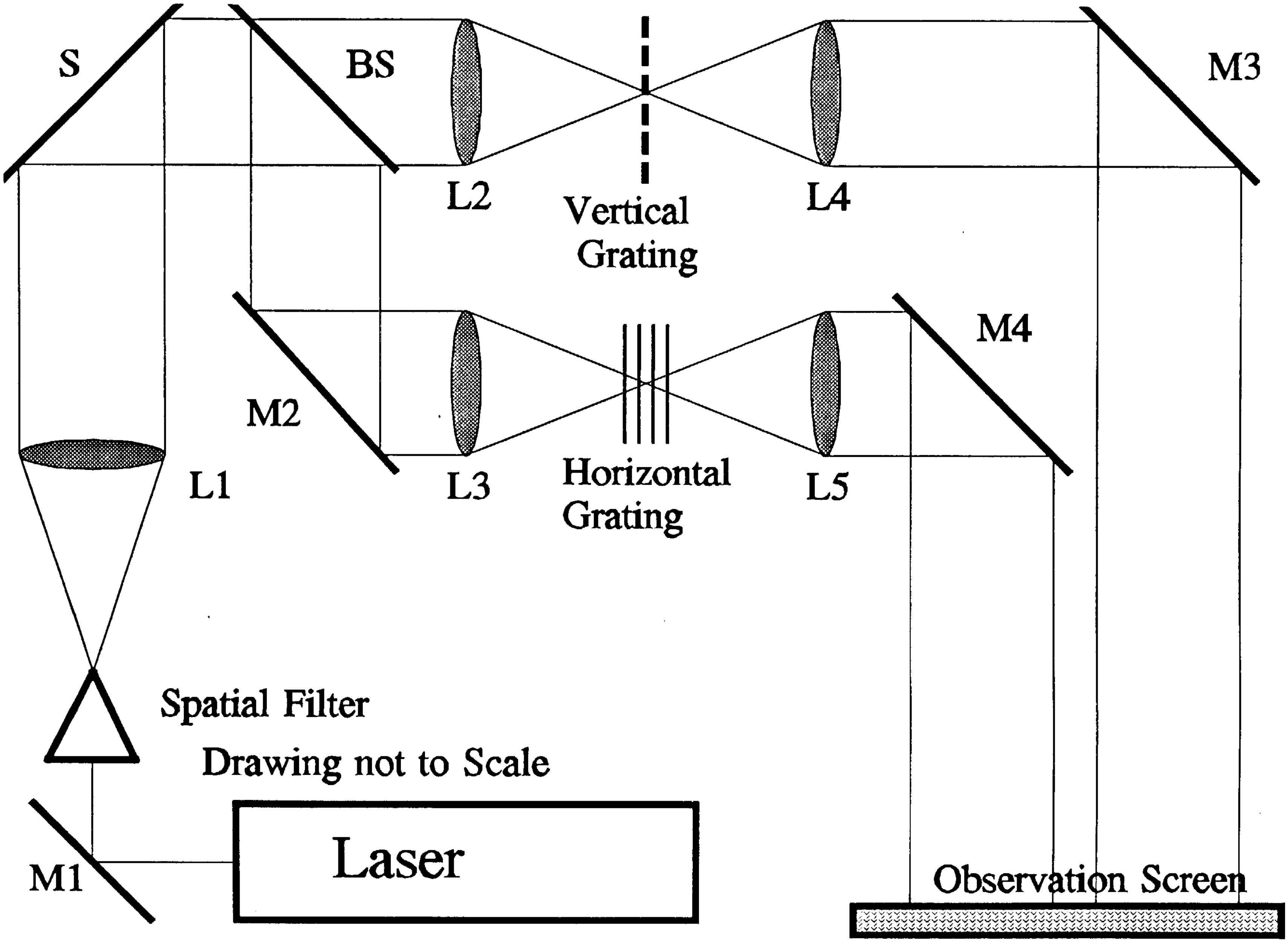
Two Channel Ronchi

- Another way to obtain simultaneous x- and y-slope data is with a two channel Ronchi approach.

- For the two-channel Ronchi approach, a collimated beam of light (laser or white) is projected onto the reflective sample surface.

- The reflected beam is collected, split into separate x- and y-slope measurement channels, passed through vertical and horizontal gratings, and projected onto an observation screen for viewing by a video camera.

- A two channel Ronchi instrument was demonstrated which simultaneously acquires x- and y-slope data for an input wavefront. Unfortunately, because of packaging, cost, and schedule issues, a two channel system could not be implemented.


S
BS
M3
L2
Vertical
Grating
L4
M4
M2
L1
L3
Horizontal
Grating
L5
Spatial Filter
Drawing not to Scale
Laser
M1
Observation Screen

The Ronchi Test

- The Ronchi test is performed by observing the shadow pattern produced when a focused beam of light is passed through a periodic structure, such as a line grating.

- If the beam is ideal, the shadow pattern looks like the grating. But, if the beam is deformed, then the pattern is deformed.

- A Ronchigram is obtained when the shadow pattern is recorded at an image of the surface under test, such that it is superimposed ('painted') upon the image.

- Information about the surface is obtained by correlating shadow lines with physical locations on the surface. The distance between adjacent shadows indicates how much the surface slope changes between those two locations.

- The amount of slope required to go from one shadow to its neighbor is called the equivalent wavelength.

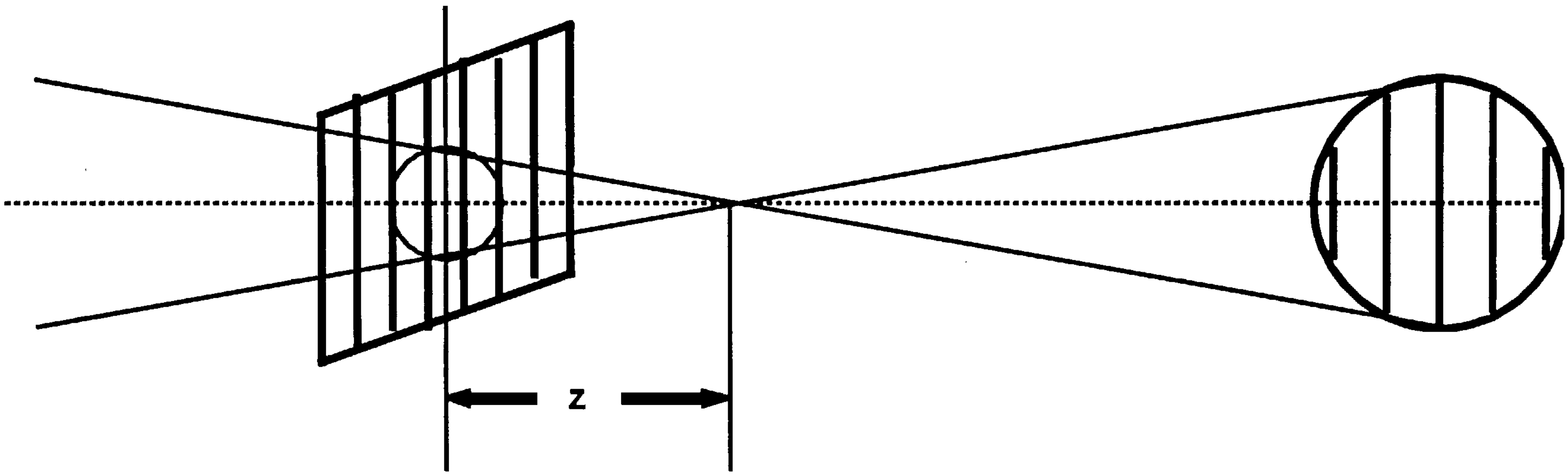


Figure 1 The Ronchi Test

The Ronchi Test

- The Ronchi test can be explained as either a geometric ray or an interferometric test.

- From a geometric ray point of view, the fringes in a Ronchigram are the result of the deviation of a ray from its ideal path because of slope errors.

- From the interferometric point of view, the fringes in the Ronchigram are produced from the interference between overlapping diffraction orders. The overlapping wavefronts produce a shearing of the original wavefront.

- Therefore the Ronchi test can be modeled as a sheared wavefront interferometer.

Advantages of the Ronchi Test

- Measures Slope: It is insensitive to G-jitter induced tilting of the oil surface.
- Self-Referencing: It is proportional to the difference between the test wavefront and itself displaced laterally.
- Common Path: The wavefront and its sheared self travel the same optical path, therefore, the test is insensitive to atmospheric turbulence and mechanical vibration.
- Known and Variable Sensitivity (Equivalent Wavelength): Measurement sensitivity is determined wholly by geometric factors (grating period and beam F/#) which can be varied in a controlled manner.
- Known and Variable Spatial Sampling: The number of spatial sample points across the aperture can be controlled by displacing the grating a known distance from focus.

Disadvantages of the Ronchi Test

- Slope patterns are difficult to interpret visually.

- Two orthogonal data sets (x-, y-slope; or r-, θ-slope) are required to completely characterize a given wavefront.

- Orthogonal slope information can be obtained with a single-channel system using a crossed-grating, but the patterns are even more difficult to interpret and analyze.

Shear Interferometry

- Whenever two coherent wavefronts are at the same point in space at the same time, they superimpose to form an interference pattern called an interferogram.

- The irradiance at any point, E(x,y), cause by this superposition can be expressed as a function of measurement wavelength and optical path difference (OPD) between the two interfering wavefronts.

$$E(x,y) = E_0 \left(1 + \cos\left(\frac{2\pi\ OPD}{\lambda}\right)\right)$$

where: λ = *measurement wavelength*

- For a shear interferogram, the OPD is given by the difference between the wavefront and its sheared self.

$$OPD = W(x,y) - W(x + \Delta x, y) = \Delta W(x,y)$$

where: Δx = *shear distance*

$\Delta W(x,y)$ = *differential wavefront*

Shear Interferogram

- The irradiance for a sheared interferogram is given by:

$$E(x,y) = E_0 \left(1 + \cos\left(\frac{2\pi \; \Delta W(x,y)}{\lambda}\right)\right)$$

- Relationship between differential wavefront and slope is:

$$\Delta W(x,y) = \frac{\delta W(x,y)}{\delta x} \Delta x = \alpha(x,y) \; \Delta x$$

where: $\alpha(x,y)$ = *wavefront slope*

- Inserting this into the irradiance equation gives:

$$E(x,y) = E_o \left(1 + \cos\left(\frac{2\pi\alpha}{\lambda_{eq}}\right)\right)$$

where: $\lambda_{eq} = \frac{\lambda}{\Delta x}$ = *equivalent measurement wavelength*

Ronchigram

- Similarly, a Ronchigram is described by:

$$E(x,y) = E_o \left(1 + \cos\left(\frac{2\pi\alpha'}{\lambda_{eq}}\right)\right)$$

where: $\alpha' = \dfrac{\delta W}{\delta x'}$ = *Normalized Slope*

$x' = \dfrac{x}{r}$ = *Normalized Pupil Coordinate*

r = *Pupil Radius*

$\lambda_{eq} = \dfrac{d}{2\ F/\#}$

d = *Grating Spacing*

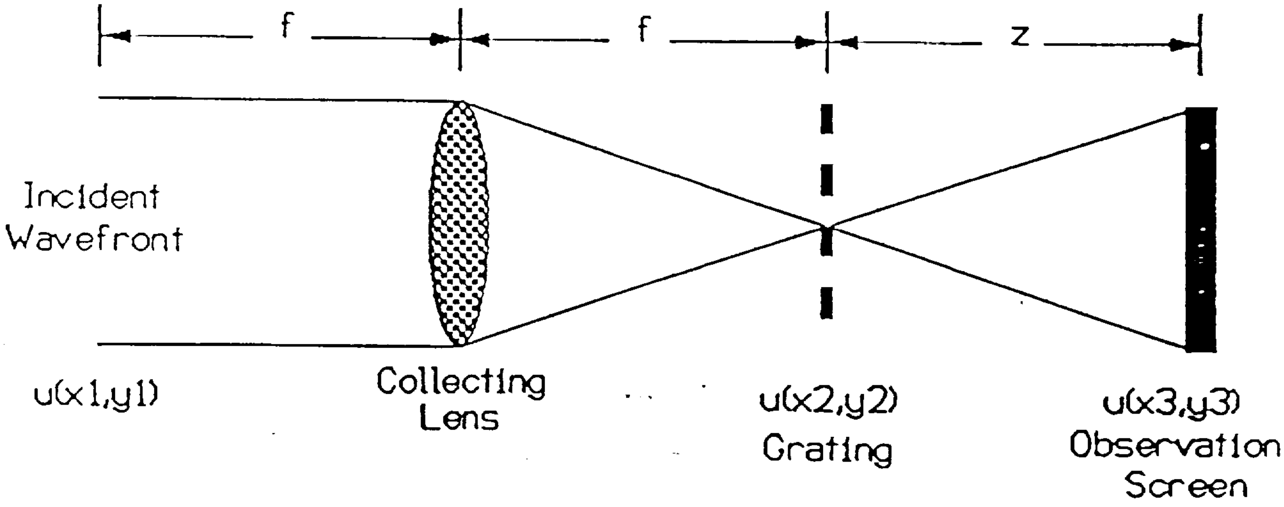
$F/\#$ = *Test Beam F/#*

- Thus, with proper scaling, conventional interferogram analysis provides a direct measurement of the slope.

- Finally, if the grating is shifted forward of focus, the irradiance pattern becomes:

$$E = E_o \left[1 + \cos\left(\frac{2\pi}{\lambda_{eq}}\left[\left(\frac{\delta W}{\delta x'}\right) - \left(\frac{z}{4\ (F/\#)^2}\right) x'\right]\right)\right]$$

Ronchi Theory



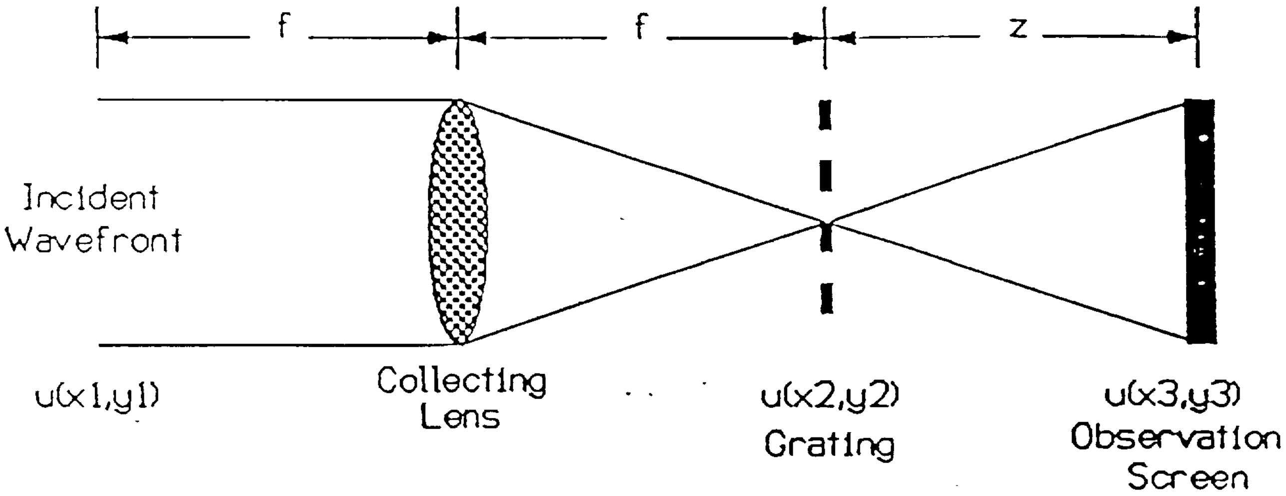
- The wavefront in the front focal plane of the collecting lens is

$$u(x_1,y_1)$$

- The wavefront in the back focal plane of the collecting lens (without the grating) is the Fourier transform of the wavefront in the front focal plane.

$$u(x_2,y_2) = \mathscr{F}\{u(x_1,y_1)\} = U(\frac{x_2}{\lambda f},\frac{y_2}{\lambda f})$$

Ronchi Theory Continued



- The wavefront is modulated by the transmission function of the grating.

$$u(x'_2,y'_2) = U(\frac{x'_2}{\lambda f},\frac{y'_2}{\lambda f})\ t(x'_2,y'_2)$$

- At the observation screen, the wavefront is given by

$$u(x_3,y_3) = \mathscr{F}\{u(x'_2,y'_2)\}$$

$$u(x_3,y_3) = u(\frac{f\,x_3}{z},\frac{f\,y_3}{z}) * T(\frac{x_3}{\lambda z},\frac{y_3}{\lambda z})$$

- When z = f the wavefront at the observation screen is the original wavefront convolved with the Fourier transform of the grating transmission function.

Sine Wave Grating

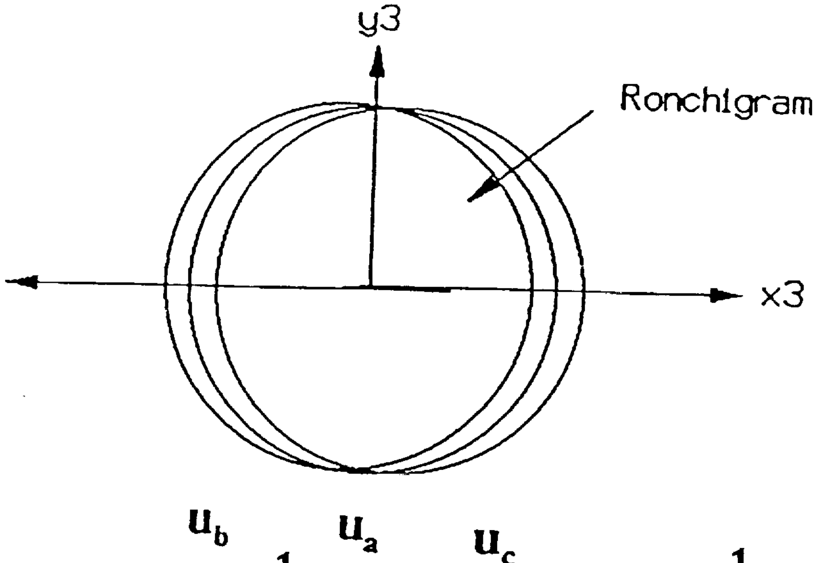
- **A sine wave grating is a series of transparent and opaque straight line bands of the following form.**

$$t(x_2,y_2) = 1 + \cos(\frac{x}{d})$$

- **The Fourier transform of the sine wave function is three delta functions.**

$$\mathscr{F}\{1 + \cos(\frac{x}{d})\} = \delta(x) + \delta(x - d) + \delta(x + d)$$

Sine Wave Grating



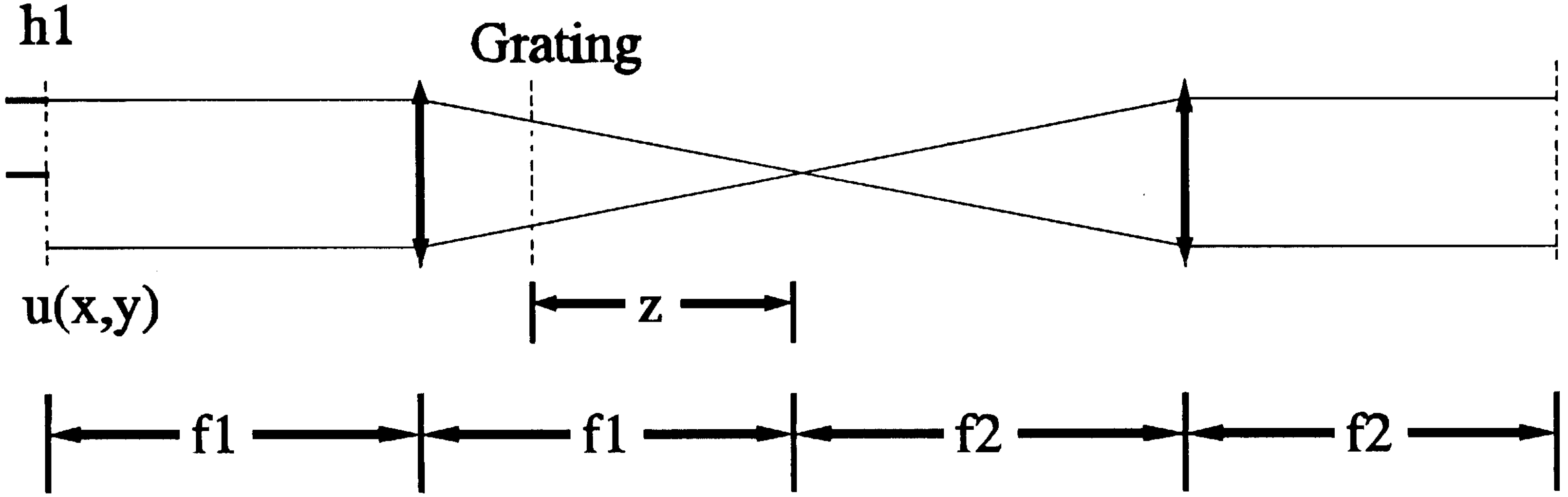
$$u(x_3,y_3) = u_a + u_b\left(\frac{1}{\lambda z}[x - d]\right) + u_c\left(\frac{1}{\lambda z}[x+d]\right)$$

- **The irradiance at the observation screen is**

$$E = E_a + E_b + E_c + (u_a u_b^* + u_b u_a^*) + (u_a u_c^* + u_c u_a^*) + (u_b u_c^* + u_c u_b^*) \ .$$

- **The cross terms (overlapping regions) are the terms that interfere to produce the Ronchigram.**

Conceptual Design
Measurement Module
Fresnel Theory



$$E = Eo\ [\ 1 + \cos\ [\ \frac{2\pi}{\lambda eq}\ [\ \frac{\delta W}{\delta x'} - (\ \frac{z}{4(f/\#)\hat{}2}\ )x'\ ]]]$$

$\lambda eq = d\ /\ [2(f/\#)]$ $\qquad$ $x' = x\ /\ h1$

Flight System

- Based upon managerial issues (packaging, cost, and schedule) and science issues (sensitivity, dynamic range, and spatial sampling) a single channel Ronchi system was selected.
- The managerial constraints and science requirements were then appropriately modified.

Surface Deformation Science Requirements

- Observe the entire surface for three different sized chambers: 12 mm, 20 mm, and 30 mm.
- Measure slope as small as 5 μm/mm.
- Measure slope as great as 30 μm/mm without vignetting.
- Spatially sample the surface at least 10 times per diameter.
- Temporally sample the surface at least 30 times per second.
- Reconstruct the surface shape at least along a diameter.
- Indicate when the oil is filled to its flat surface position.

Science Requirement

Field of View

- The surface deformation measuring apparatus must be able to view the entire surface area of each fluid chamber at an optimum camera spatial resolution.

- The chamber sizes are:
  - 30 mm
  - 20 mm
  - 12 mm

- The apparatus's imaging system should have at least three different fields of view which can be used to view any of the three chambers.

Science Requirement

Measurable Surface Features

The apparatus must be able to measure three different types of surface features:

- A central bump or 'pimple',
- A central hole or 'dimple', and
- A large-area temporal oscillatory motion.

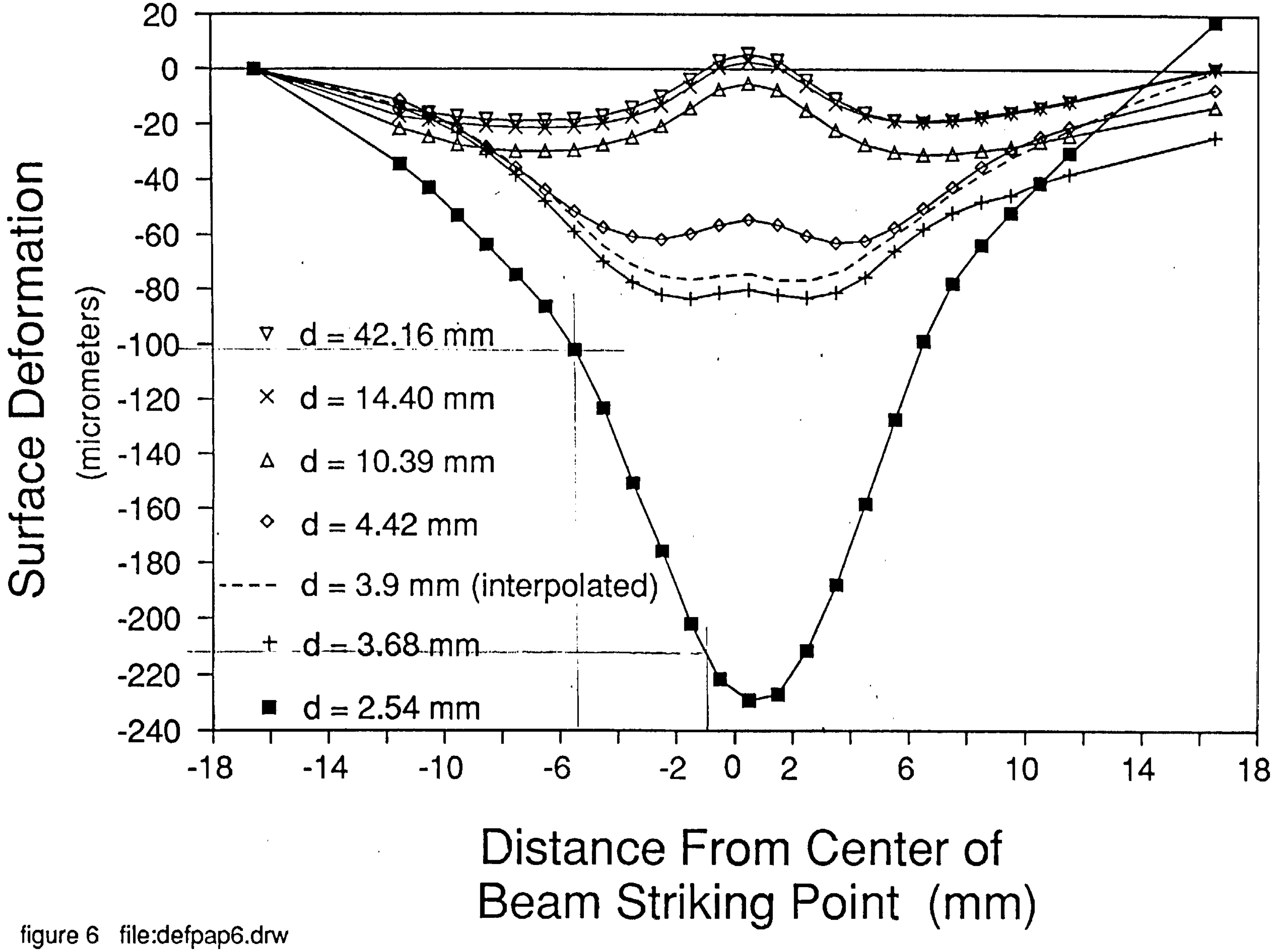


figure 6 file:defpap6.drw

Science Requirements

Central Bump

- The central bump is assumed to be spherical in shape with a constant radius of curvature, given by:

$$W(y) = A\left[1 - \frac{y^2}{r^2}\right]$$

where: A = Maximum Vertical Height, and
r = Maximum Lateral Radius.

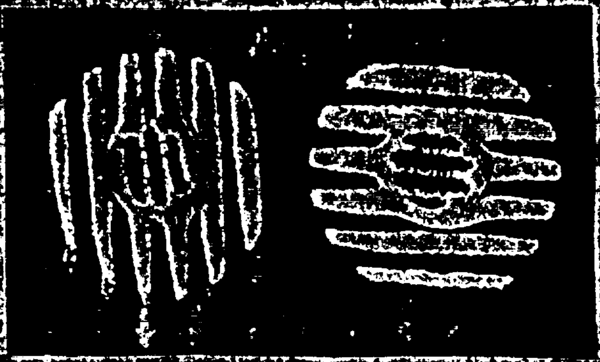
- The maximum measurable surface slope of the central bump occurs at y = r and is given by:

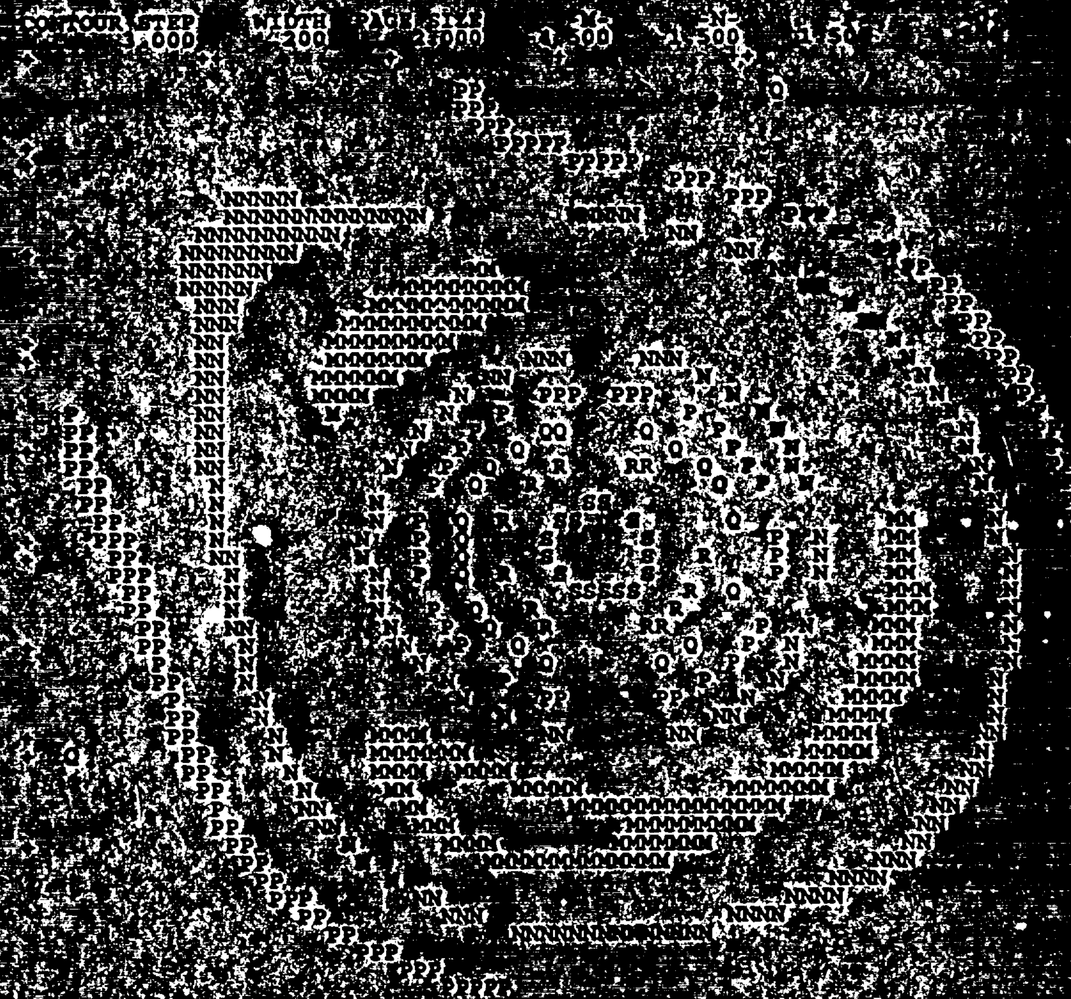
$$\left.\frac{\delta W(y)}{\delta y}\right|_{\max} = -\frac{2A}{r}$$

- The maximum measurable surface curvature of the central bump is given by:

$$\left.\frac{\delta^2 W(y)}{\delta y^2}\right|_{\max} = -\frac{2A}{r^2}$$

Preliminary Data from Oil Surface (Bump)
[illegible] V; Beam Diameter at Oil Surface = 6 mm
Size of [illegible] Micrometers





```
CONTOUR STEP     WIDTH   PAGE SIZE     X-       N-
      3.000       .200      2.000    1.300    1.500    1.50
```

Science Requirements

Central Hole

- The central hole is assumed to be Gaussian in shape and is given by:

$$W(y) = -A\, e^{\left(-\pi \frac{y^2}{r^2}\right)}$$

where: A = Maximum Vertical Height, and
r = Maximum Lateral Radius.

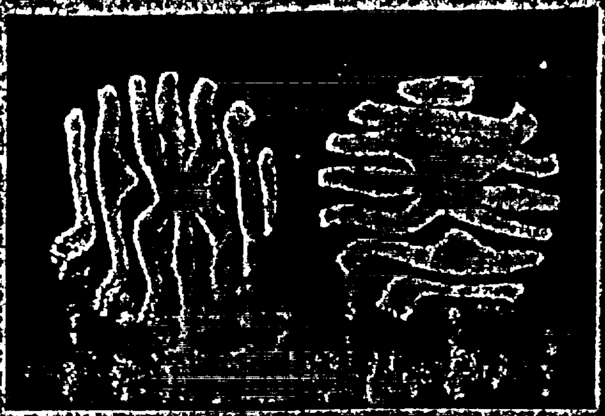
- The maximum measurable surface slope of the central hole occurs at $y^2 = r^2 / 2\pi$ and is given by:

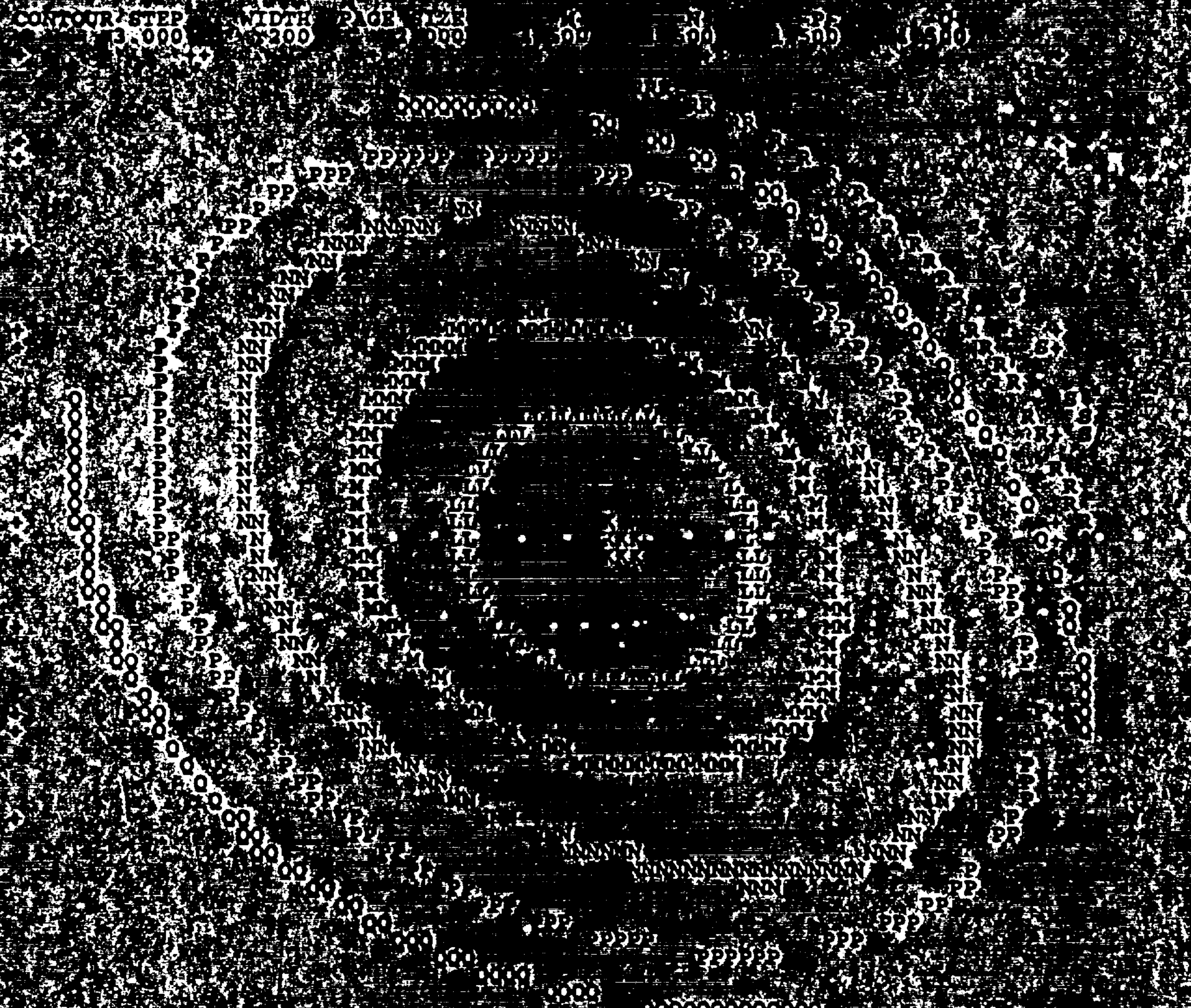
$$\left.\frac{\delta W(y)}{\delta y}\right|_{\max} = \frac{\sqrt{2\pi}\, A}{r}\, e^{\left(-\frac{1}{2}\right)} = \frac{1.5\, A}{r}$$

- The maximum measurable surface curvature of the central hole occurs at y = 0 and is given by:

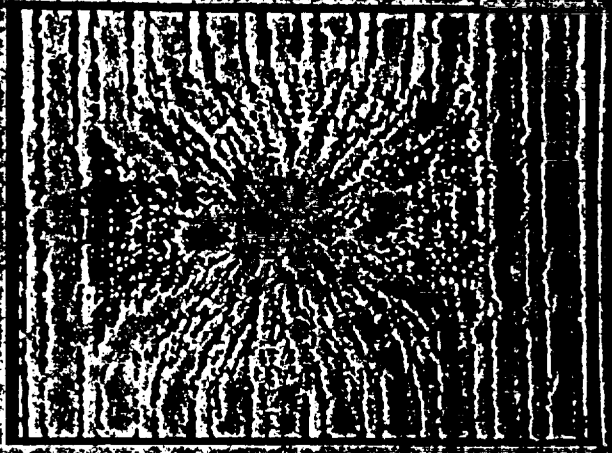
$$\left.\frac{\delta^2 W(y)}{\delta y^2}\right|_{\max} = \frac{2\pi\, A}{r^2} = \frac{6.3\, A}{r^2}$$

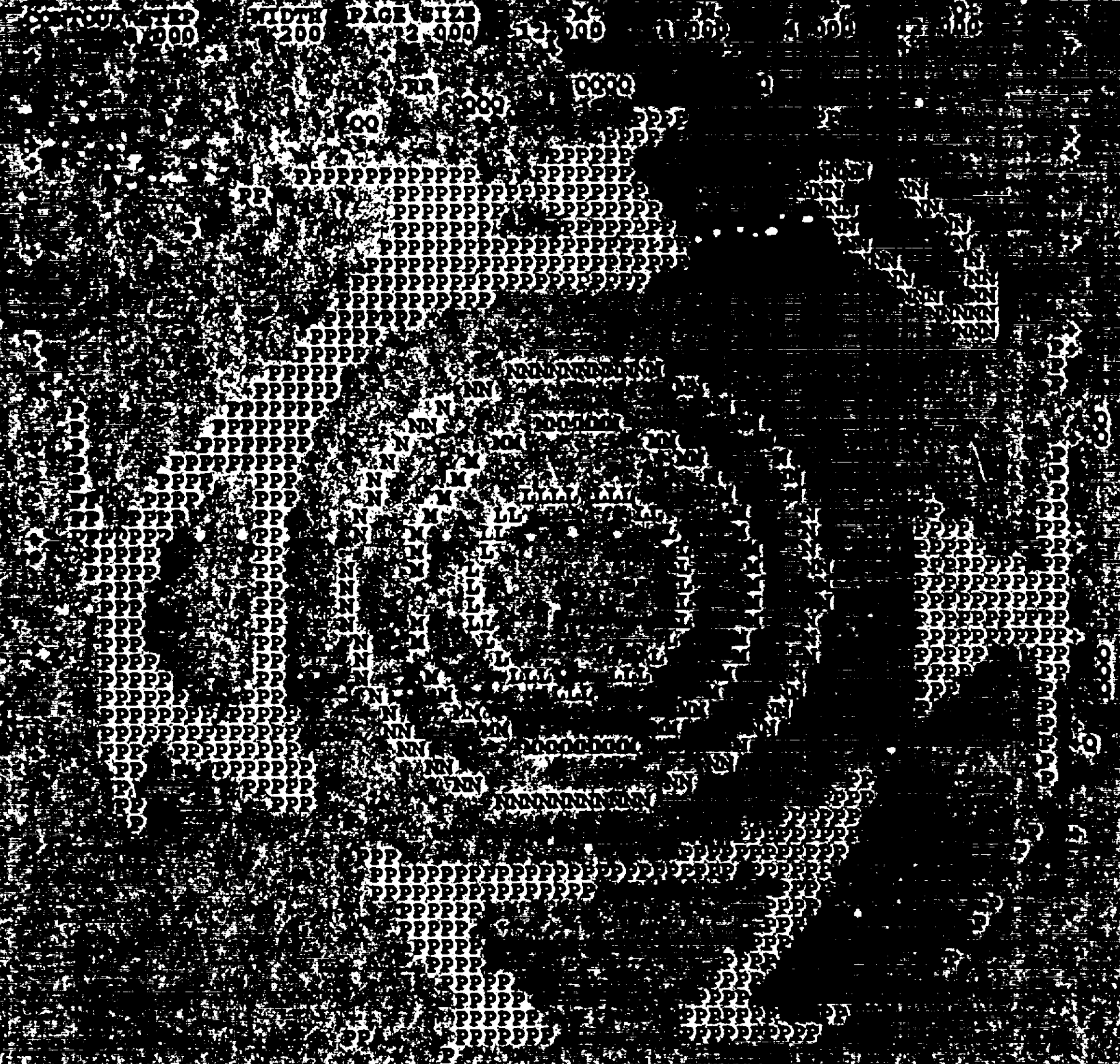
Preliminary Data from Oil Surface (Depression)
[illegible] Voltage = 23 V, Beam Diameter at Oil Surface = 19 mm
Depression Size = 22 Micrometers





Preliminary Data Known Surface (54 Micrometer Depression)
Measured Depression Size 52 Micrometers





Science Requirements

Maximum Resolvable Features

- The largest feature that the apparatus must resolve, whether it is a bump or a hole, is assumed to have the following properties:

Vertical Height	$A = 250\ \mu m$
Lateral Radius	$r = 5\ mm$

- The maximum assumed measurable surface slope and curvature for a bump are:

Slope	$100\ \mu m/mm$
Curvature	$20\ \mu m/mm^2$

- The maximum assumed measurable surface slope and curvature for a hole are:

Slope	$75\ \mu m/mm$
Curvature	$63\ \mu m/mm^2$

Interferogram Analysis

- **A minimum of 3 pixels per fringe are required to extract the correct information about the test signal.**

Rise, Run, and Feature Sign

- **A good rule of thumb is to have 4 pixels per fringe.**
- **The curvature corresponding to one fringe per 4 pixels can be calculated by:**

$$\beta = \frac{\lambda_{eq}}{4ar}$$

where β = *curvature,* λ_{eq} = *equivalentwavelength,*
a = *pixel width, and* r = *radius of test beam.*

- **For $\lambda_{eq} = 60.48$ um, r = 10 mm, and a = 19.5 um the curvature is calculated to be 77.5 um/mm^2**
- **Since the calculated curvature is larger than the required curvature of 63 um/mm^2, the required curvature can be resolved and each fringe covers more than four pixels.**
- **Since each cycle of the test function is larger than four pixel widths the Nyquist rate is met ($\xi_s > 2\xi_0$).**

Science Requirement

Minimum Resolvable Feature

- The smallest feature that the apparatus must resolve, whether it is a bump or a hole, is assumed to have the following properties:

Vertical Height	$A = 10\ \mu m$
Lateral Radius	$r = 5$ mm

- The minimum assumed measurable surface slope and curvature for a bump are:

Slope	4.0 $\mu m/mm$
Curvature	0.8 $\mu m/mm^2$

- The minimum assumed measurable surface slope and curvature for a hole are:

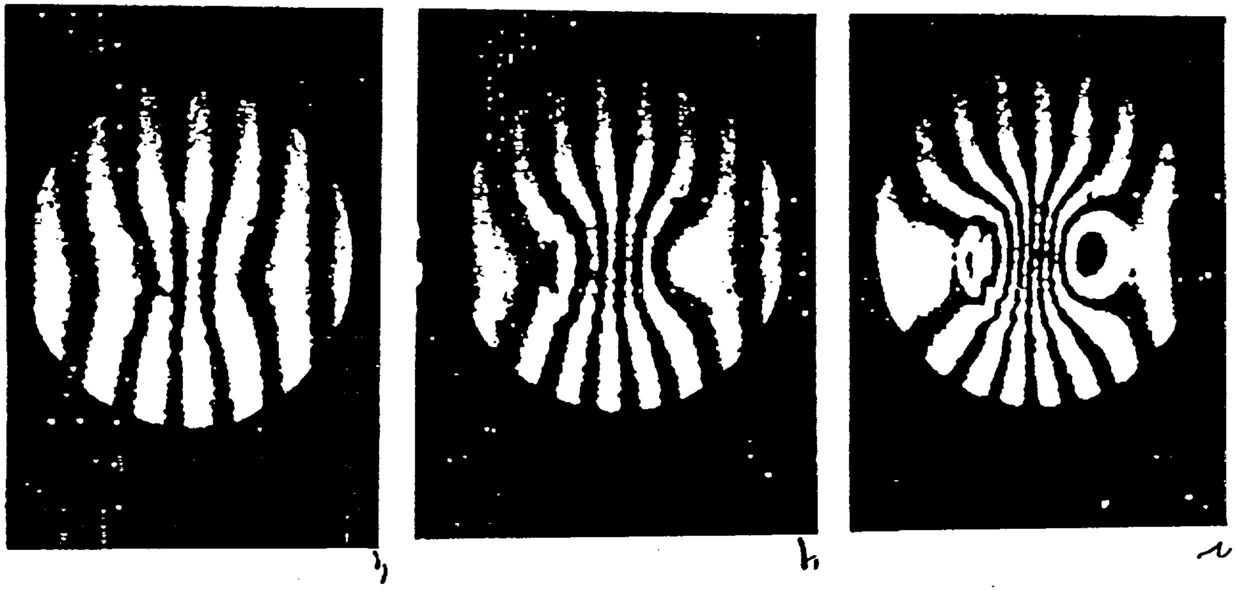
Slope	3.0 $\mu m/mm$
Curvature	2.5 $\mu m/mm^2$

Ronchi Measurement Performance

- The performance of the Ronchi system is completely specified by its equivalent wavelength:

$$\lambda_{eq} = \frac{d}{2(F/\#)}$$

- Since the F/# of the measurement beam is fixed by the optical system, everything depends upon grating spacing.
- This dependance is extremely important. It allows for the measurement capability of the system to be varied on orbit by simply inserting different gratings into the beam.
- If the surface deformation is large, insert a coarse grating.
- If it is small, use a fine grating.



0.5 cycles/mm 1.0 cycles/mm 2.0 cycles/mm

Figure 14 Comparison of Slope Detectable with Different Gratings

Performance Parameters

- Accuracy defines the uncertainty of a given measurement.
- For the current sampling system, the measurement accuracy is approximately $\lambda_{eq}/8$.
- A measurement system can report numbers to an arbitrary precision, but they may not be accurate or repeatable.
- The minimum measurable slope of a given system is defined to be twice its measurement accuracy.
- For the current sampling system the minimum measurable slope is approximately $\lambda_{eq}/4$.
- The maximum amount of slope which the system can measure is determined by how many shadow lines (fringes) the video camera can clearly resolve in the Ronchigram.
- For the current sampling system, the Ronchigram will be limited to approximately 16 lines pairs for a maximum slope of 16 λ_{eq}.

Table 1. Performance Summary for F/2.25 System

Grating Frequency [lp/mm]	Grating Line Space [mm/lp]	λ_{eq} [μm/mm]	Accuracy [μm/mm]	Minimum Slope [μm/mm]	Maximum Slope [μm/mm]
0.5	2.0	444.44	55.56	111.11	7111
1.0	1.0	222.22	27.78	55.56	3556
2.0	0.5	111.11	13.89	27.78	1778
2.5	0.4	88.89	11.11	22.22	1422
5.0	0.2	44.44	5.56	11.11	711
10.0	0.1	22.22	2.78	5.56	356

Science Requirements

Temporal Resolution

- The third required surface feature which must be measured is a large-area temporal oscillation.

- It is assumed that the period of this oscillation is on the order of 2 to 5 seconds. Thus, a standard video camera with a 60 Hz field rate is wholly adequate as the data acquisition device.

- However, if the temporal phenomena were to have a period of less than about 0.2 sec, then a non-standard high-speed video camera may be necessary.

- It is assumed that there may be high-speed temporal gitter in addition to the lower-speed temporal oscillation. The effect of this gitter is to reduce the contrast of the pattern.

- To eliminate gitter the camera will need a high speed shutter and the source will need sufficient output power to expose the camera.

Science Requirement

Fill Level Indication

- Silicon oil has a very low surface tension and spreads easily on any surface. To prevent it from flowing out of the container, the rim of the chamber has a sharp edge that 'pins' the oil in place and is 'barrier' coated.

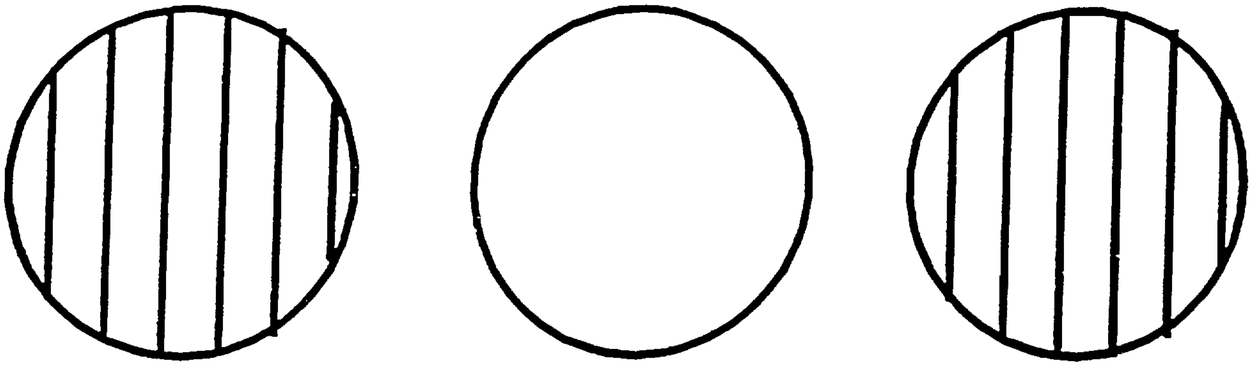
- As the container is filled in low gravity, the liquid will form a deep spherical curvature.

- When completely filled, the surface will be flat.

- If over-filled, the surface will form a spherical dome.

- By placing a grating at a focal point, direct visual evidence of the oil's surface state is provided to the Mission Specialist.

- When the oil surface is not flat, there is a line pattern. As the oil approaches a flat surface, the pattern spreads out (has fewer lines) until all the lines disappear when the surface is flat. If the container is under or overfilled, the lines will reappear and increase in number.



Concave Surface

Flat Surface

Convex Surface

Figure 9. Fill Level Indicator

Flight System Design

- The free surface deformation measuring instrument was designed in modules.

- The flight system consists of six modules:
  light source,
  projection/imaging,
  polarization,
  pupil relay,
  measurement, and
  camera.

- Each module overcomes various engineering and/or packaging challenges while allowing the total system to meet the science requirements.

- Each module is independently assembled and bench aligned before integration into the final system.

- And, each module prevents unnecessary exposure of the crew to light radiation.

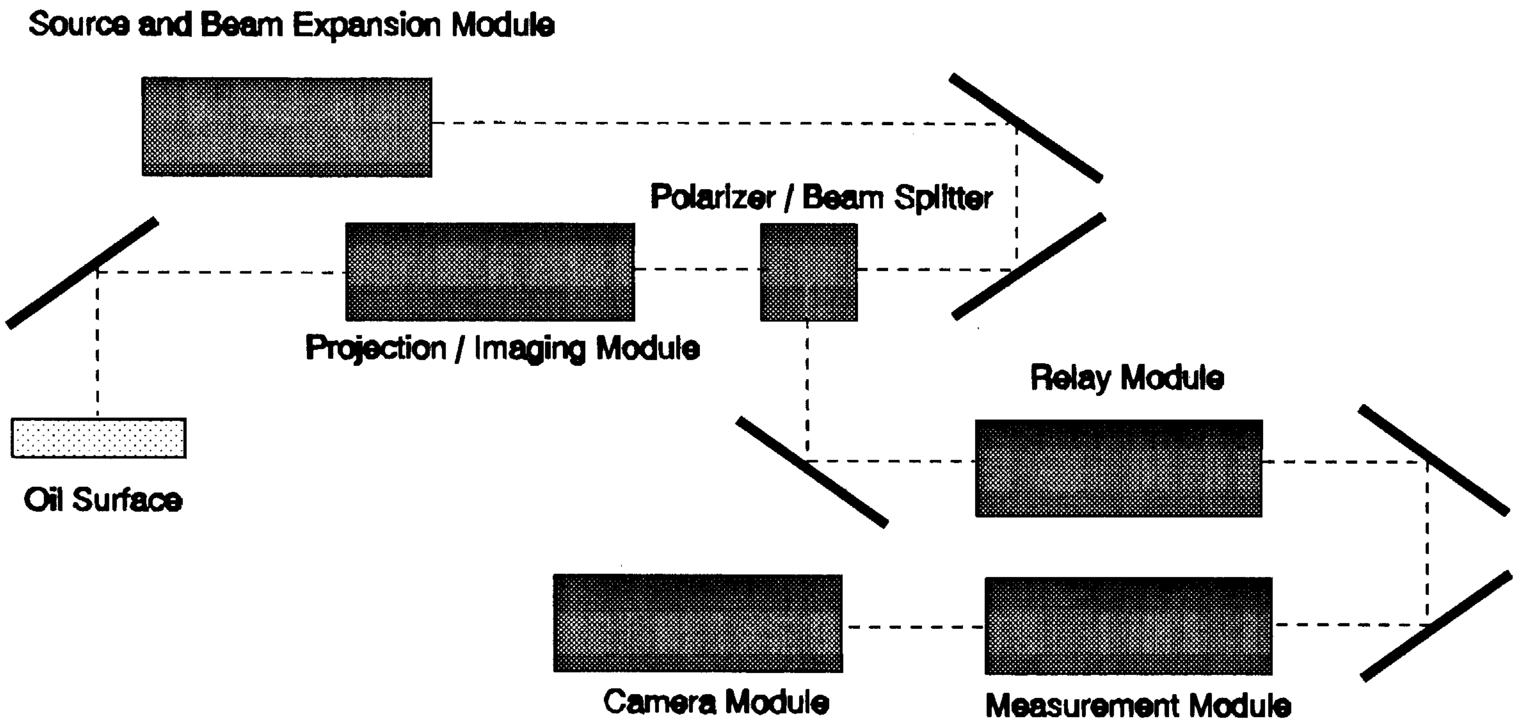
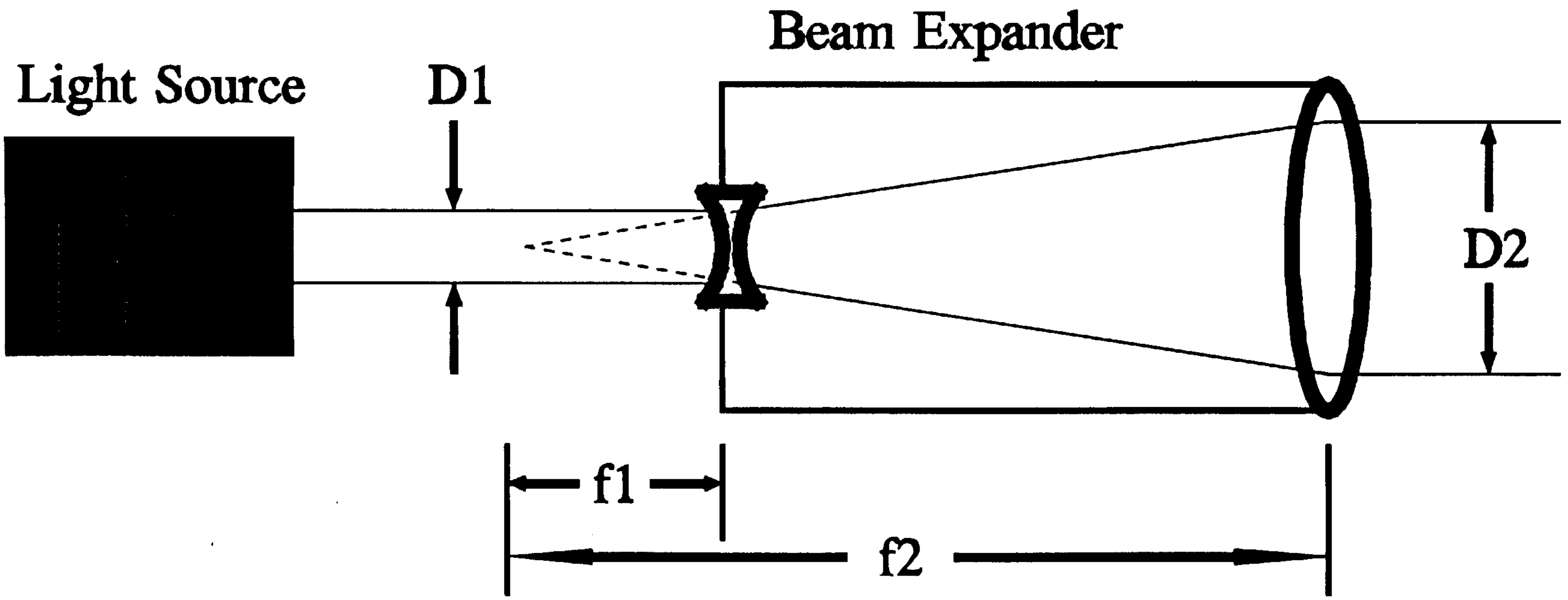


Figure 3. Schematic of Flight System

Source Module

- The source module provides a collimated beam of light which uniformly illuminates the object under test.
- This is accomplished with a laser diode and a Galilean beam expanding telescope.
- A laser diode was selected because it is a very-bright collimated source capable of surviving the launch. (White-light and LED sources were also considered.)
- A commercially available Galilean beam expander was selected because it is compact and does not have a spatial filter which could become misaligned during launch.
- The beam diameter is defined by the projection/imaging module's entrance pupil to be 22 mm in diameter.
- The beam is magnified (30X) and truncated to balance uniformity and power.

Conceptual Design
Illumination Module



Fluid Properties

- The object under test is a free surface of 2 cSt silicon oil.
- It has a refractive index of 1.39
- And, an irradiance reflection coefficient of 2.7%.
- Because of this low reflection coefficient, the source must have at least 10 mW of power for the video camera to observe the Ronchi pattern.
- Also, ghost reflections from the optical components can obscure the oil surface reflection.

Illumination Geometry

- Because the free-oil surface is a specular reflector, we treat it as a plane mirror.

- A collimated illumination beam is required to implement the Ronchi test.

- Two illumination geometries were considered:

  Off-Axis
  On-Axis

- The on-axis geometry was selected for several reasons:

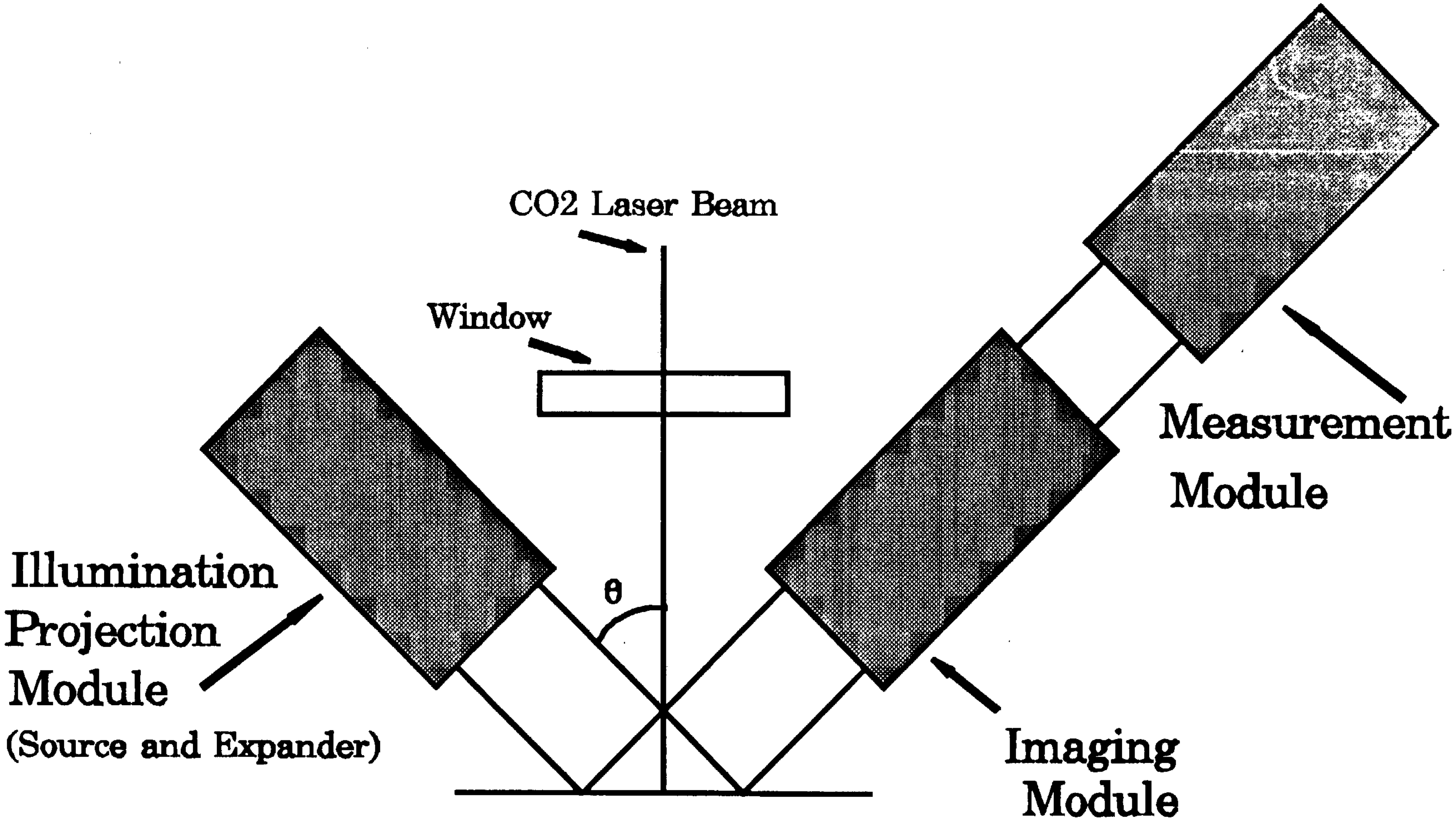
  Able to get lens closer to fluid surface, thus able to observe the highest slope errors.

  No off-axis projection distortion or cosine scaling.

  Easy to package.

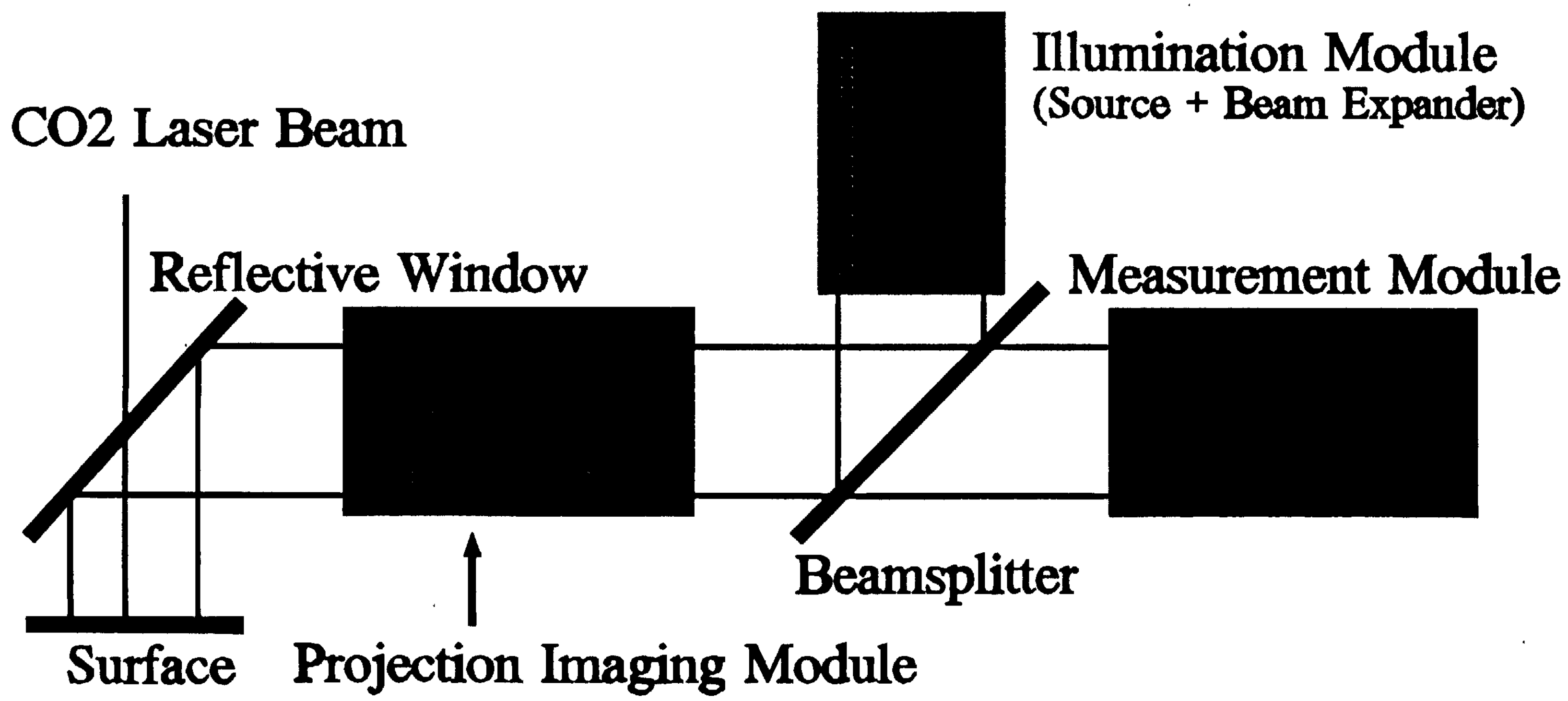
- The on-axis geometry requires the beam to travel in two directions through the projection/imaging module. This can cause ghost reflection problems.

Conceptual Designs
Off-Axis Illumination Geometry



Conceptual Design

On - Axis Illumination Geometry



Illumination Module
(Source + Beam Expander)

CO2 Laser Beam

Reflective Window

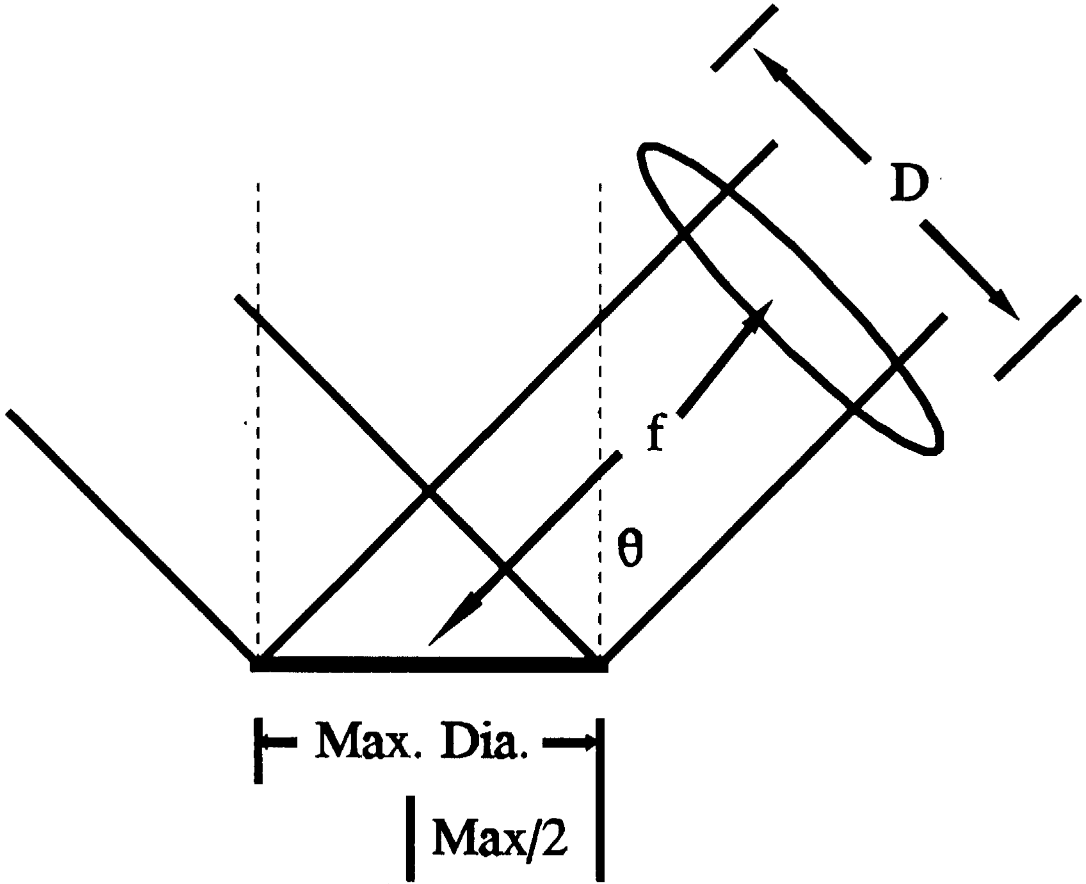
Measurement Module

Beamsplitter

Surface

Projection Imaging Module

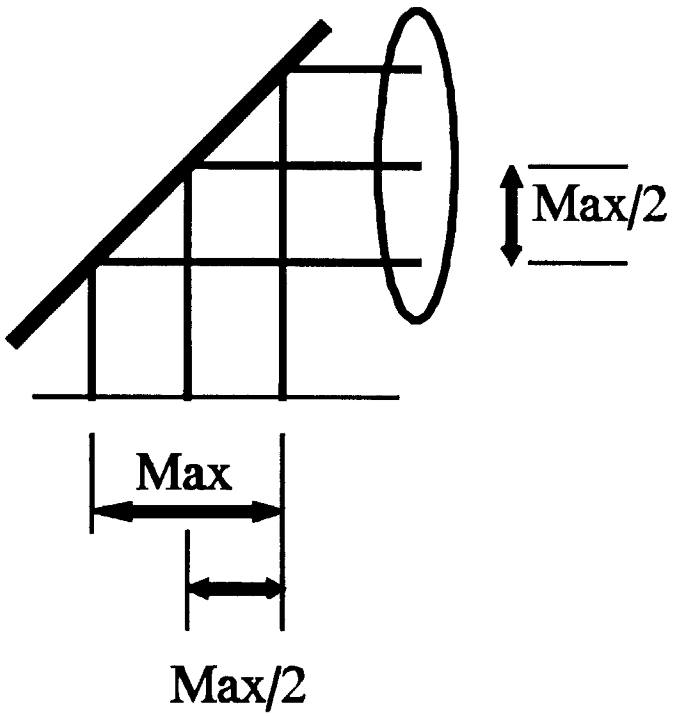
Conceptual Design Off Axis Illumination (Distance to First Lens)



$f >= 1/2\sin(\theta)[D\cos(\theta) - \text{Max.}]$
if $D = 40$ mm   Max. = 30 mm $\theta = 45°$

$f >= 41$ mm

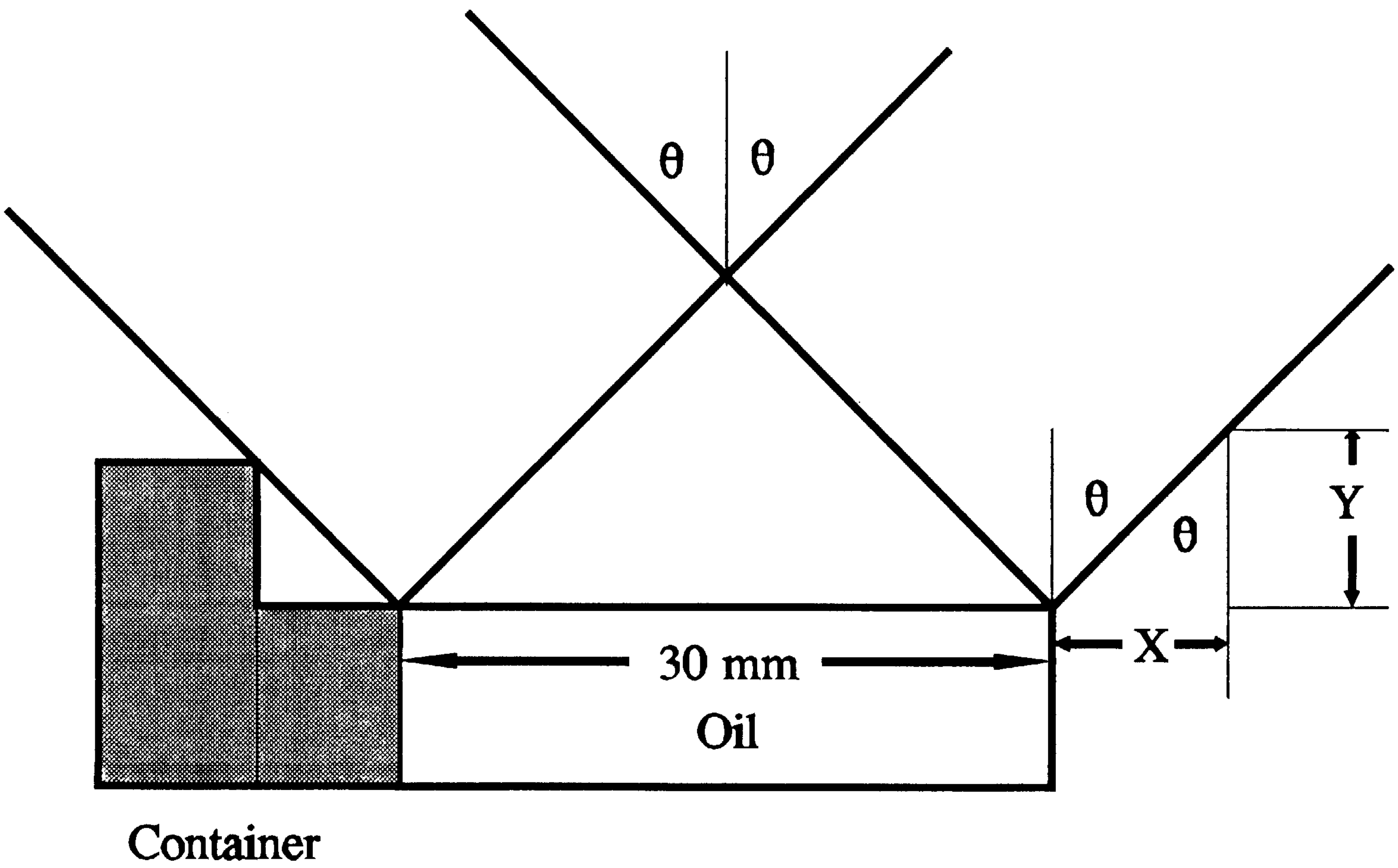
Conceptual Design On Axis Illumination
Distance to First Lens



f >= Max. Chamber Diameter

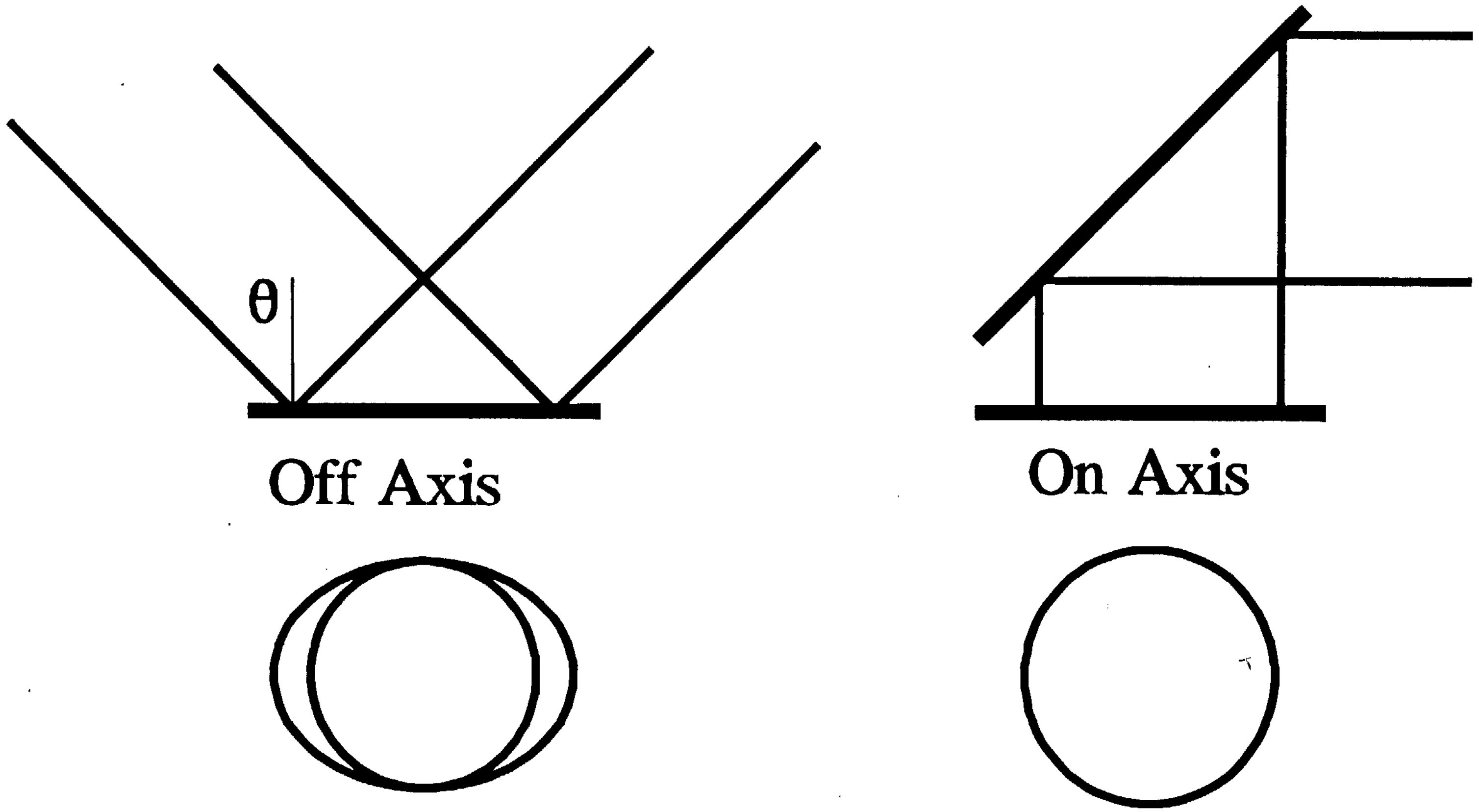
f >= 30 mm

Conceptual Design Off Axis Illumination (Chamber Size vs. Vignetting)



$X/Y >= \tan(\theta)$

Conceptual Design Illumination Geometry Pupil Mapping



Off Axis

On Axis

Minor Axis = Max. Chamber Size

Major Axis = (Max. Size) / cos (θ)

When projected onto the camera the axis will switch

Polarization Module

- The polarization module eliminates ghost reflections and efficiently uses the source's available optical power.

- Since the projection/imaging module is used in both directions and because the oil's reflection coefficient is small, ghost reflections from the illuminated optical surfaces can seriously obscure the oil reflection.

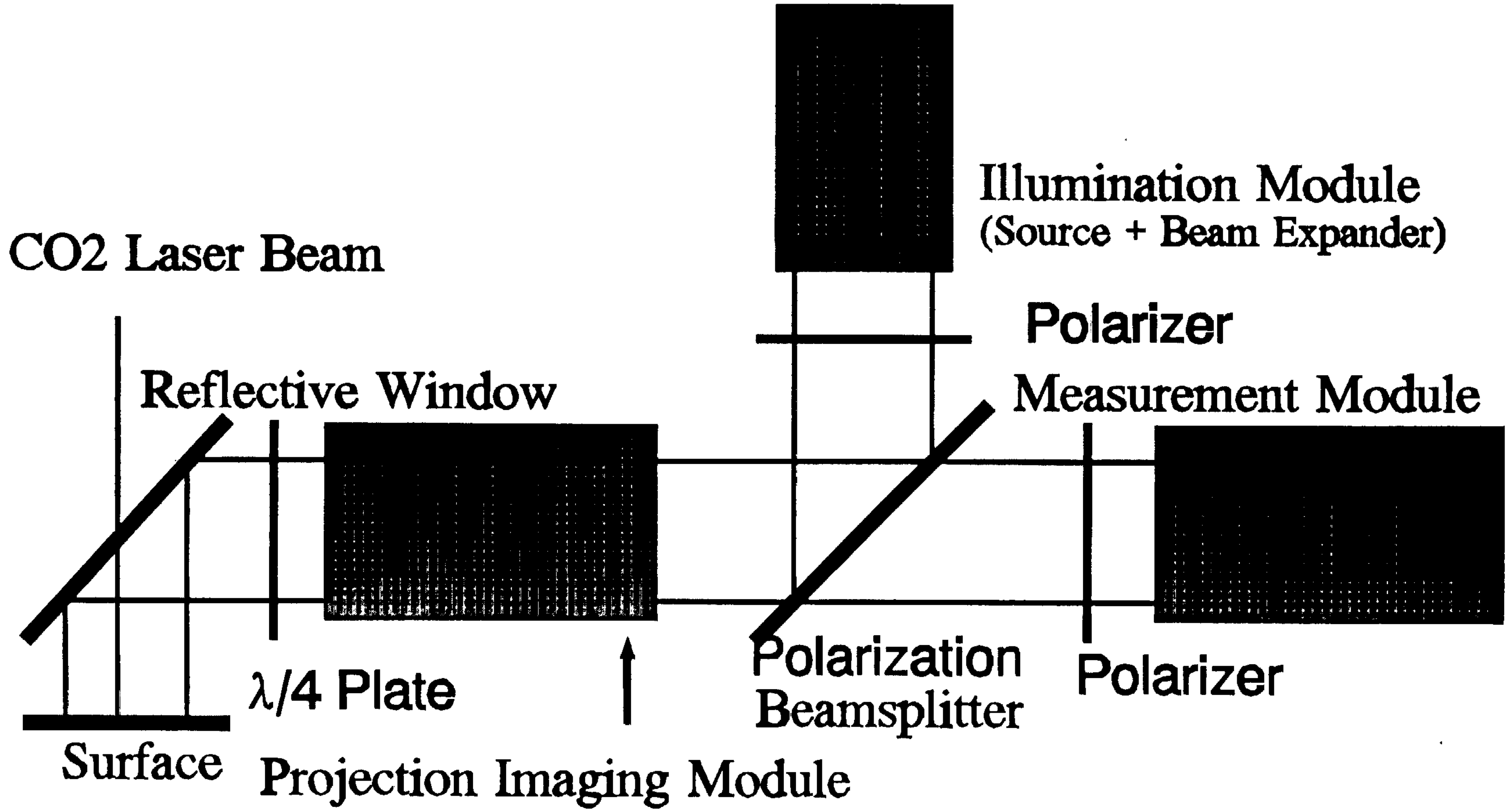
- To eliminate this problem, the polarization of the beam is manipulated such that the ghost reflections are vertically polarized and the oil reflection is horizontally polarized.

- This is accomplished with two polarizers and a quarter wave plate. The first polarizer defines the polarization of the ghost reflections. The quarter wave plate rotates the oil reflection polarization by 90 degrees. And, the second polarizer blocks all of the ghost reflections and passes the oil reflection.

- To maximize the available power, a polarization beam splitter is used. All of the source's vertical light is transmitted into the projection module. And, all of the returning horizontal light is reflected into the relay module. A conventional 50/50 beam splitter could be used, but it would throw away 75% of the available light.

Conceptual Design

On - Axis Illumination Geometry



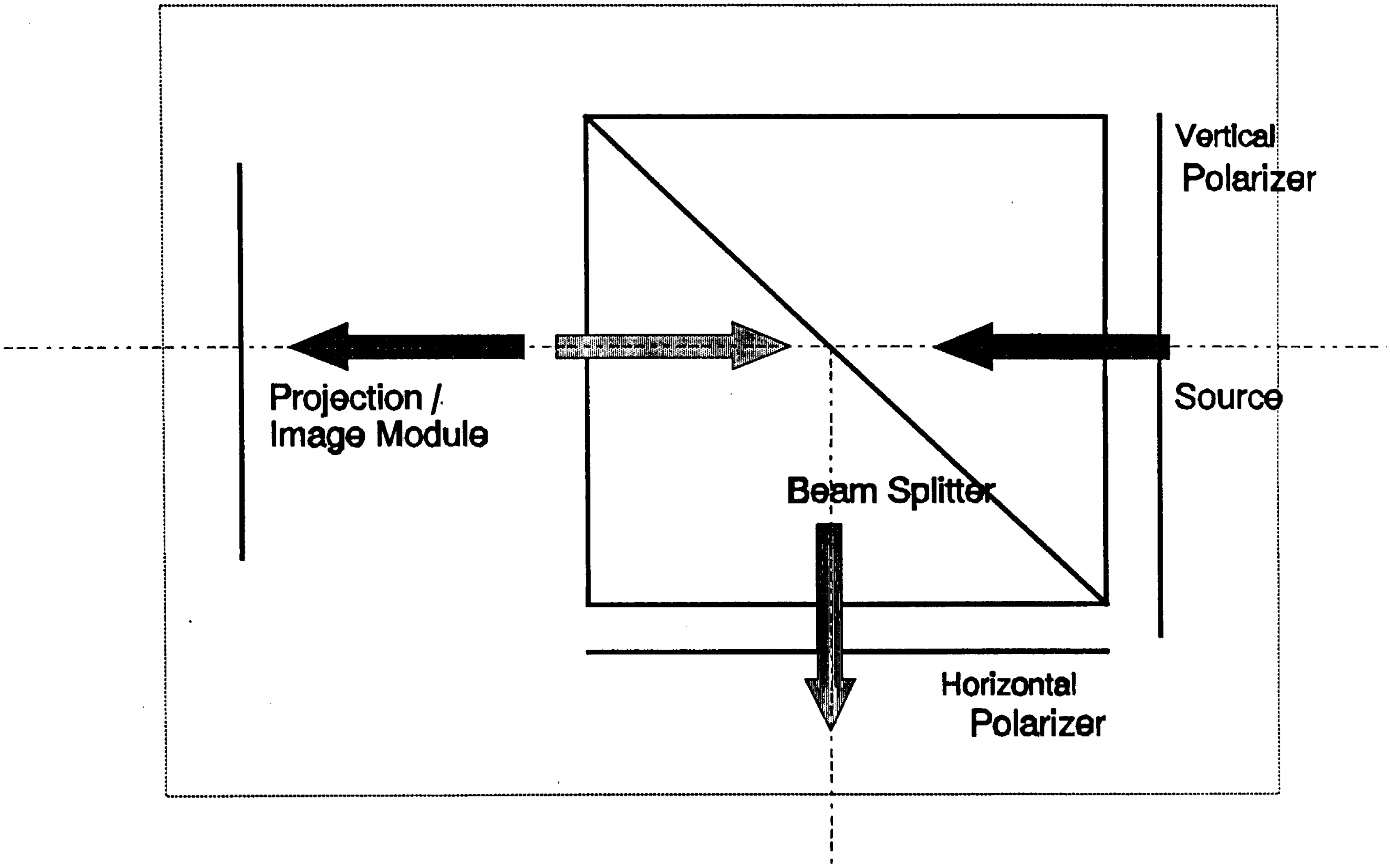


Figure 6. Polarization Module

Projection/Imaging Module

- The projection/imaging module has two functions:

  illumination
  imaging.

- Its illumination function magnifies or de-magnifies the light beam to illuminate the oil surface for each of the three chamber sizes (12 mm, 20 mm, 30 mm).

- Its imaging function has two requirements:

  form an image of the different sized surfaces under test at a fixed sized pupil location, and

  pass a reflected wavefront with at least 30 μm/mm of slope without vignetting.

- The illumination and imaging functions determine the magnification properties of this module.

- The vignetting requirement determines its F/#.

- Packaging issues determines its clear aperture.

Projection/Imaging Module

Illumination Function

- The illumination function is accomplished with two AFOCAL lens pairs, an outer pair and an inner pair.

- The outer pair forms a Keplerian telescope with unit magnification.

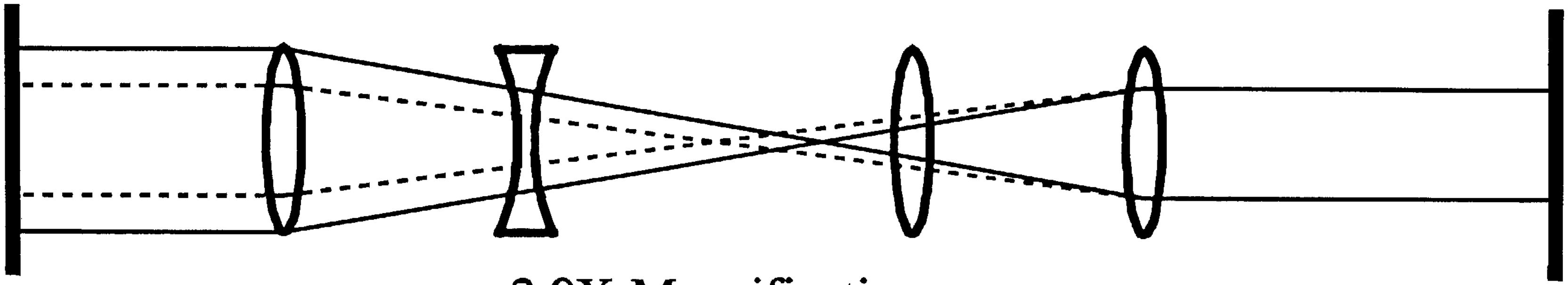
- The inner pair forms a Galilean telescope which is positioned about the outer pair's internal focus and functions as a pseudo field lens. This lens provides magnification/de-magnification when it is flipped.

- To illuminate the 20 mm chamber, remove the flip lens such that the 22 mm source beam is relayed unchanged.

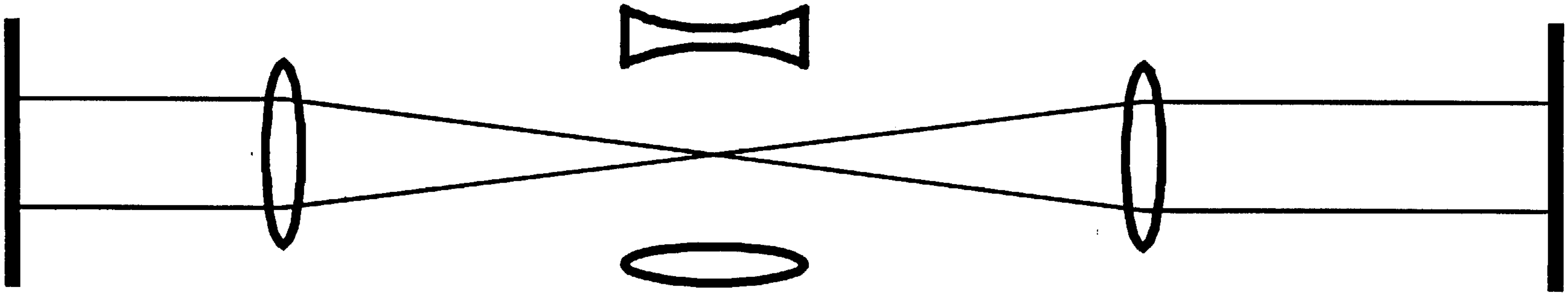
- To illuminate the larger or smaller chambers, insert the flip lens either forward or backward to magnify or de-magnify the beam. The final illumination beam size depends upon the flip lens magnification factor.

- The flip lens is not symmetric about the focal point and does not rotate about the focal point.

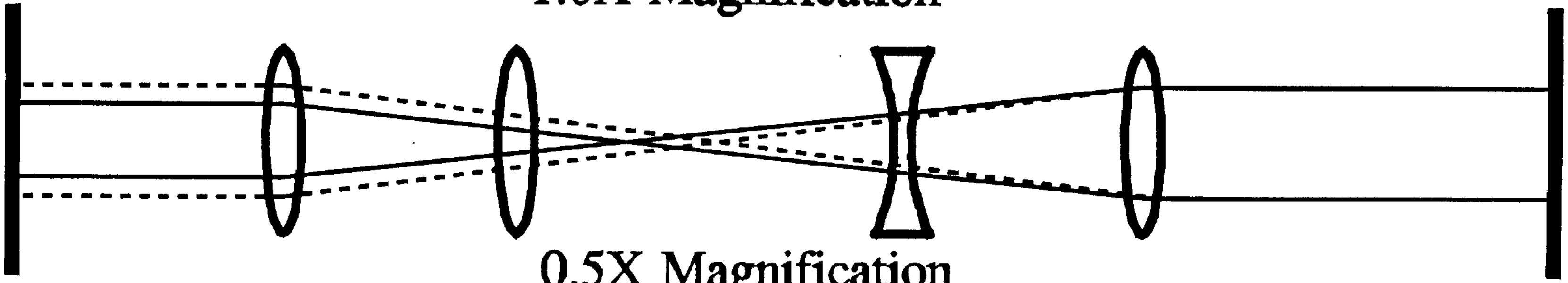
Conceptual Design Imaging Module



2.0X Magnification



1.0X Magnification

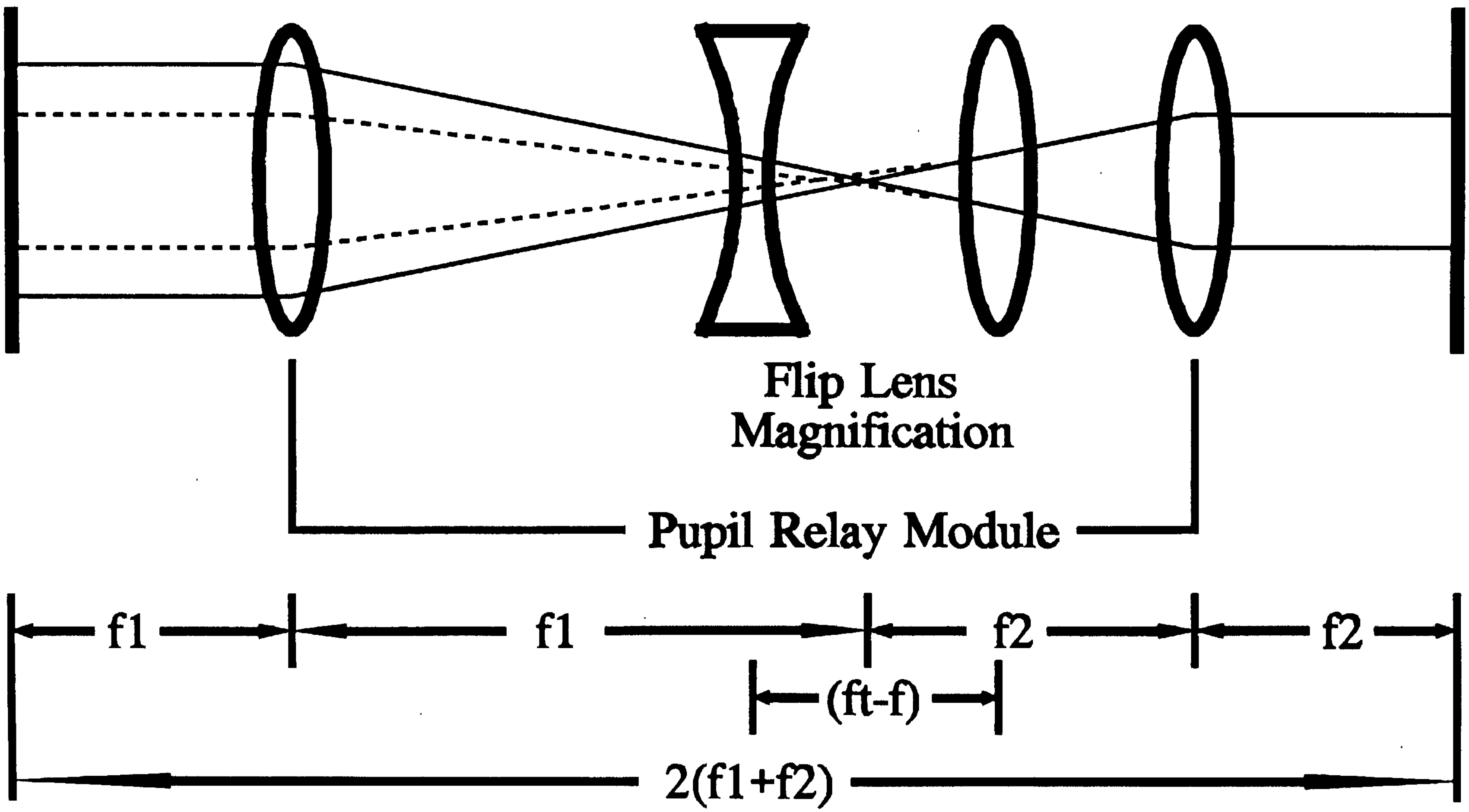


0.5X Magnification

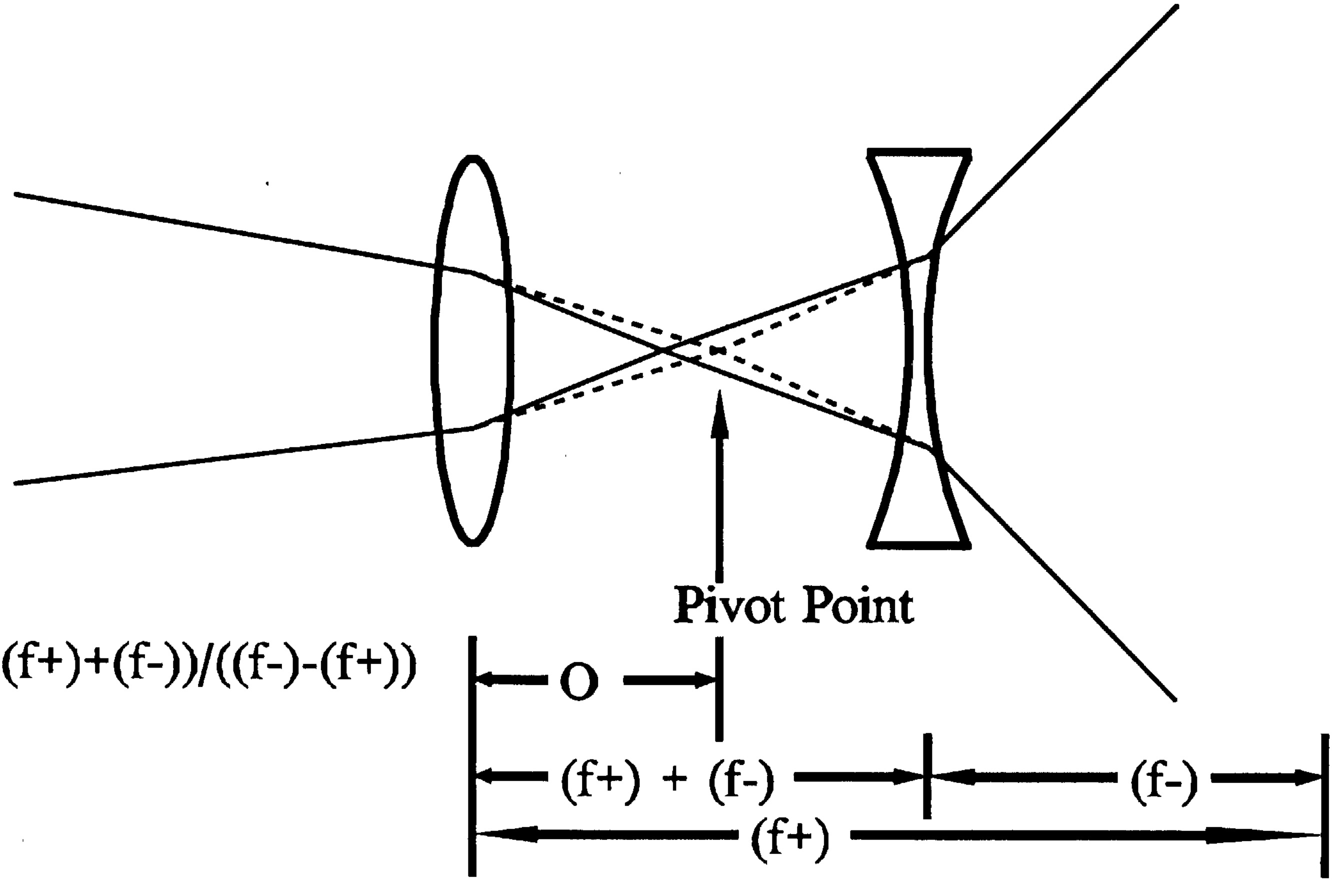
Table 3 Image Magnification Module Fields of View for Various Flip Lens Magnification Factors

Chamber	Diameter	1.4 X	1.5 X	1.6 X	1.7 X
Smallest	12 mm	15.7 mm	14.7 mm	13.75 mm	12.9 mm
Middle	20 mm	22 mm	22 mm	22 mm	22 mm
Largest	30 mm	30.8 mm	33 mm	35.2 mm	37.4 mm

Conceptual Design Imaging Module (2X, 3X)



Conceptual Design AFOCAL Flip Lens



$$O = ((f+)((f+)+(f-))/((f-)-(f+))$$

Projection/Imaging Module

Imaging Function

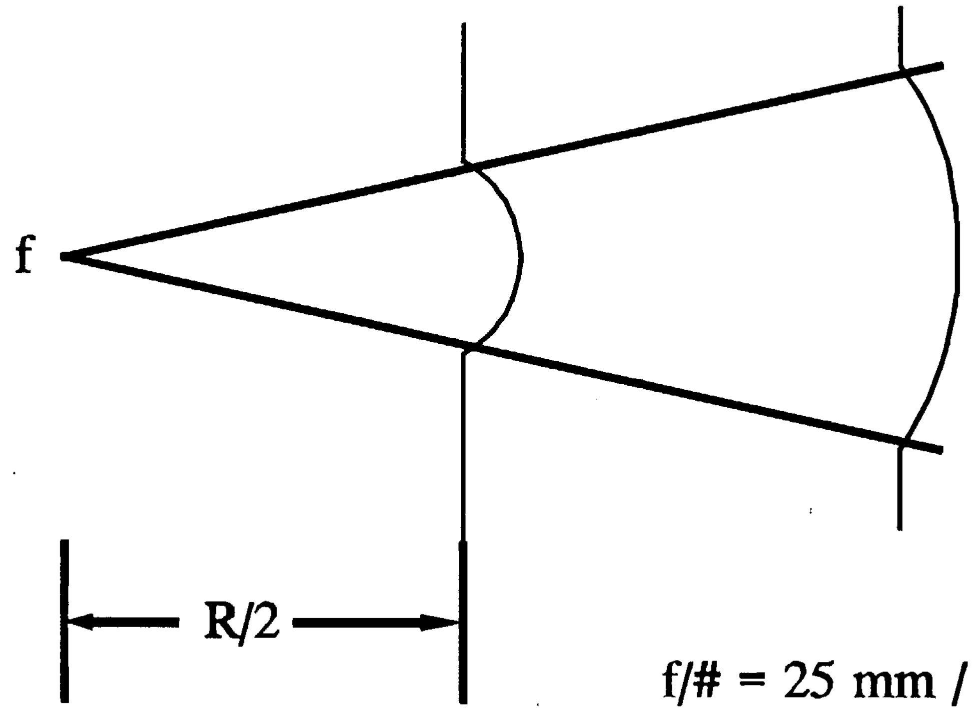
- Imaging of the different sized chambers into a fixed size pupil is accomplished by running the beam backwards through the system.
- The Keplerian telescope, without the flip lens, relays an image of the surface in the middle sized chamber.
- Inserting the flip lens relays images of the larger or smaller sized chambers.
- For proper imaging the surface must be in the front focal plane of the lens closest to the oil.
- The resultant image will be in the back focal plane of the last lens of the projection/imaging module. This image is transferred to the camera module by the pupil relay and measurement modules - forming a Ronchigram.

Projection/Imaging Module

Vignetting

- The projection/imaging module (as well as the pupil relay and measurement modules) must pass a maximum slope of 30 μm/mm without vignetting.
- This places an F/# requirement on these components. The faster their F/# the more slope they can pass.
- Initially, the optical system was designed to fully collect a wavefront reflected from a 250 μm high, 10 mm diameter spherical deformation.
- The actual design was determined by packaging and schedule issues.
- Packaging factors limit their maximum diameter and minimum focal length.
- Schedule require them to be commercially available.

Conceptual Design Reflected Wavefront from Bump



f/# = 25 mm / 10 mm = 2.5

R = (r)^2 / (2 δ) = (5 mm)^2 / (2(0.250 mm)) = 50 mm

Vignetting Requirement

- The selected components are F/2.25 (90 mm focal length, 40 mm diameter) achromatic doublets with approximately 2 waves of spherical aberration at 633 nm.
- This amount of aberration is acceptable given the magnitude of the anticipated oil surface deformations.
- With proper calibration, this error can be removed from the measurement.
- Please note that the 30 μm/mm specification is not satisfied for the 30 mm diameter chamber.

Table Maximum Observable Feature and Slope without Vignetting for each Sample Chamber as a function of Lens Clear Aperture

40 mm Clear Aperture

Chamber Diameter	δ_{max}	α_{max}
12 mm	208 μm	62.4 μm/mm
20 mm	277 μm	41.6 μm/mm
30 mm	208 μm	20.8 μm/mm

36 mm Clear Aperture

Chamber Diameter	δ_{max}	α_{max}
12 mm	181 μm	54.3 μm/mm
20 mm	222 μm	33.3 μm/mm
30 mm	125 μm	12.5 μm/mm

Pupil Relay Module

- The pupil relay module is a unit magnification AFOCAL system.
- Its original function was simply to extend the beam path such that the measurement module is at a location where the mission specialist can insert the gratings.
- Additionally, it serves as an alternative measurement module allowing for gratings to be inserted into parts of the beam not accessible with the measurement module.
- This is required to properly measure hole deformations.

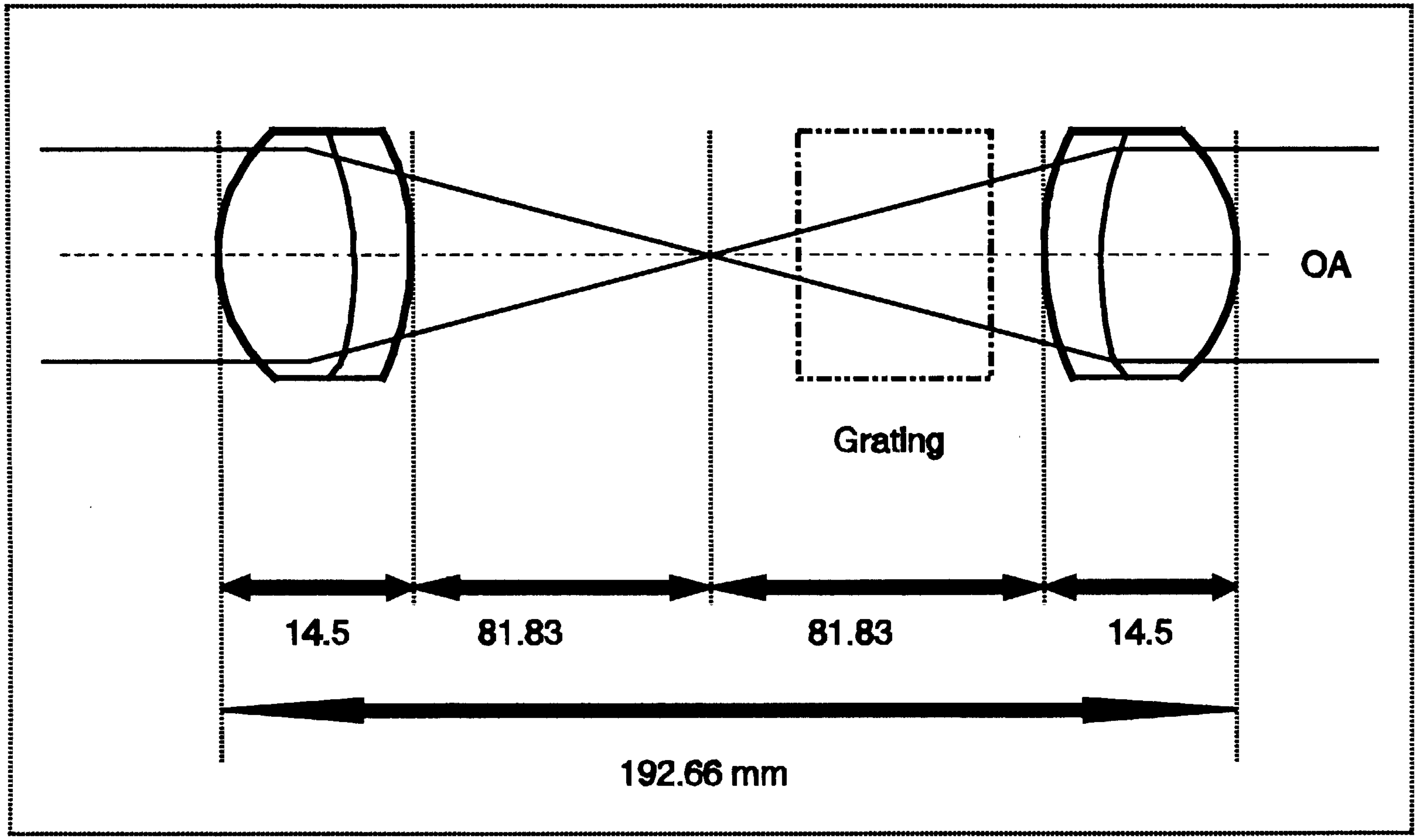


Figure 7. Relay Module
Expansion of Relay
Module in Figure 3.

Measurement Module

- The measurement module allows gratings of different line spacings to be inserted into different locations of the focused oil reflection forming a Ronchigram.

- Since the lens F/#'s are known, changing the grating line spacing changes the equivalent wavelength.

- The equivalent wavelength depends only upon the grating line spacing and is independent of where the grating is placed in the beam.

- Grating placement in the beam determines the number of spatial sample points across the beam.

- To insure at least 10 sample points per diameter, a coarse grating must be placed at a greater distance from the measurement module focus than a fine grating.

- The distance from focus for a given grating depends entirely upon its line spacing:

$$z = 10\,d\,(F/\#) = 22.5\,d$$

- To reconstruct a two-dimensional surface contour, separate horizontal and vertical gratings, or crossed gratings can be inserted into the beam.

- Radial or circular zone gratings may be tried.

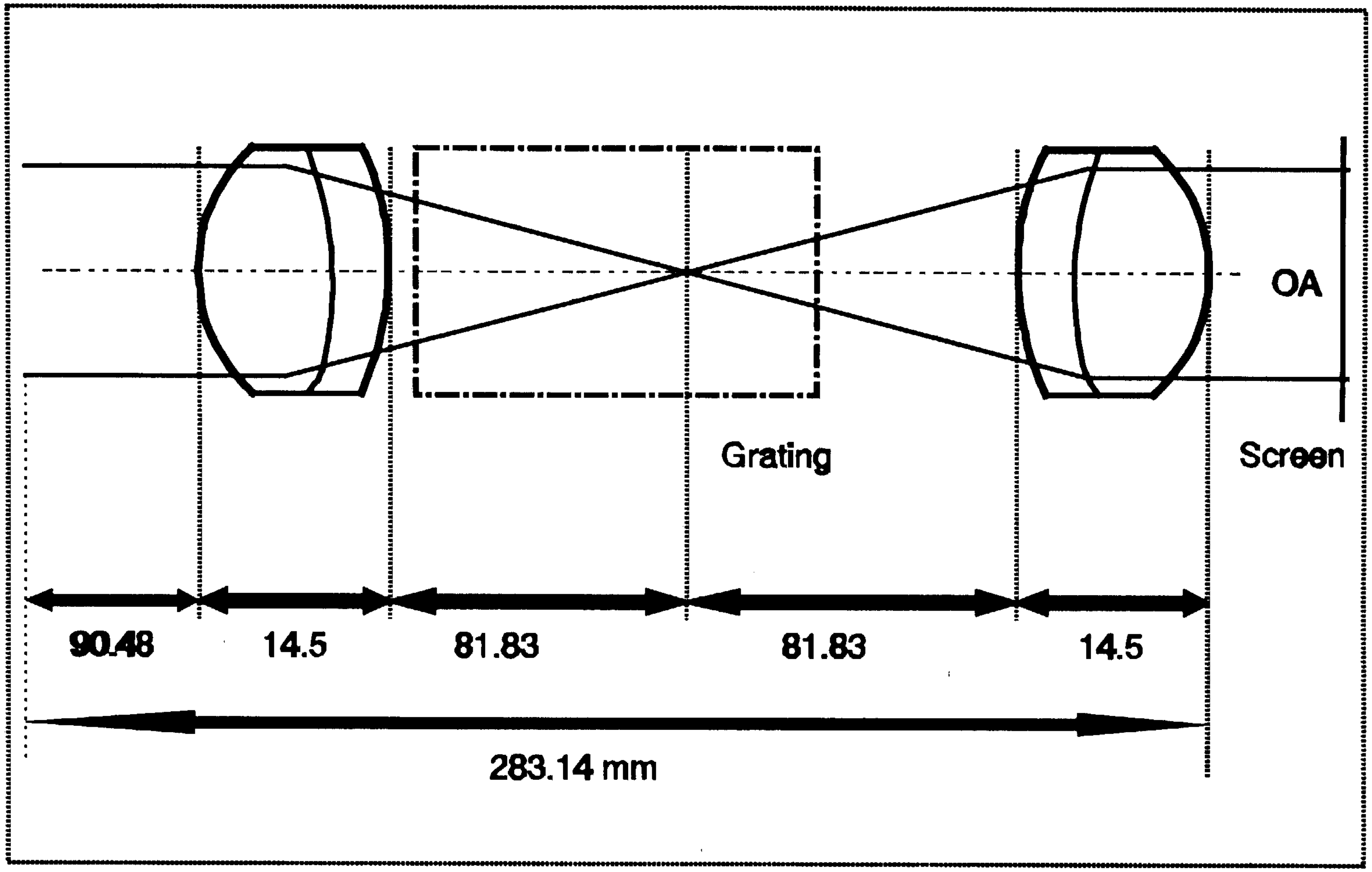
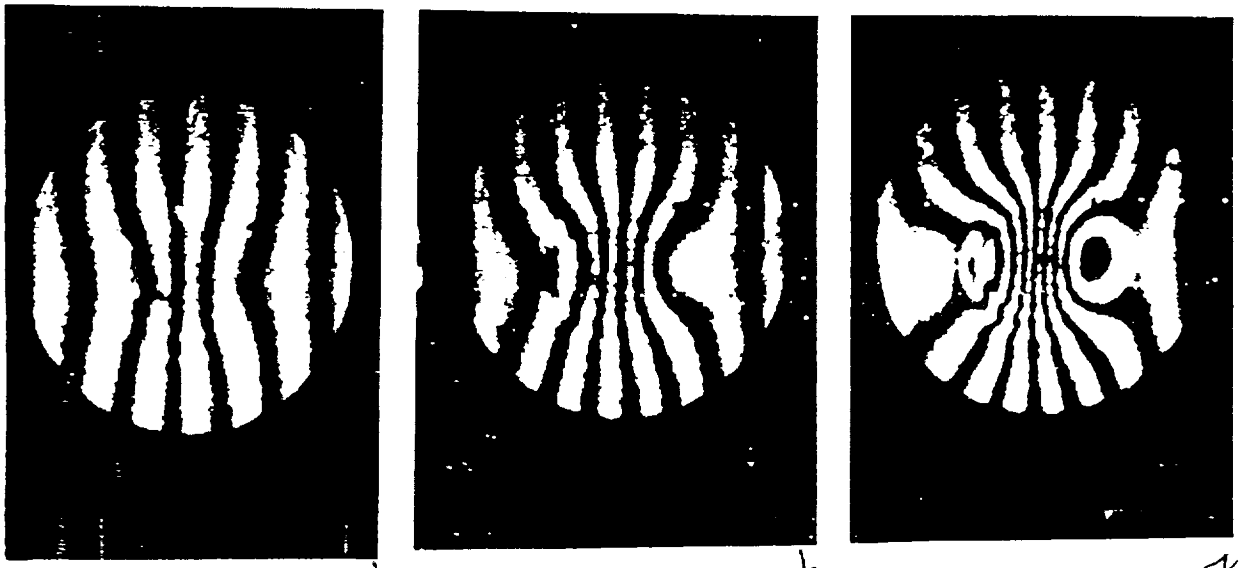


Figure 8. Measurement Module
Expansion of Measurement
Module in Figure 3.



0.5 cycles/mm 1.0 cycles/mm 2.0 cycles/mm

Figure 14 Comparison of Slope Detectable with Different Gratings

Table 2. Distance from Focus for Spatial Sample Points

Grating Frequency [lp/mm]	Grating Line Space [mm/lp]	Z for 10 Samples per Diameter	Z for 15 Samples per Diameter	Z for 20 Samples per Diameter
0.5	2.0	45.0 mm	67.5 mm	90.0 mm
1.0	1.0	22.5 mm	33.75 mm	45.0 mm
2.0	0.5	11.25 mm	16.88 mm	22.5 mm
2.5	0.4	9.0 mm	13.5 mm	18.0 mm
5.0	0.2	4.5 mm	6.75 mm	9.0 mm
10.0	0.1	2.25 mm	3.375 mm	2.5 mm

Camera Module

- The Ronchigram produced by the measurement module is imaged onto a diffuser plate where a real image is formed.

- This image is viewed by a video camera and recorded on video tape for subsequent data analysis.

- The physical distance from the diffuser screen to the camera is determined by the focal length of the camera lens, the camera sensor format, and the size of the image on the diffuser screen:

$$d = \left[1 + \frac{h_{image}}{r_{sensor}} \right] f_{camera\ lens}$$

- For the flight system, the diffuser image radius is 10 mm, the sensor radius is 2.2 mm, and the camera lens focal length is 28 mm. Thus, the stand-off distance is 155 mm.

- The camera is a 1/2 inch format RS-170 video camera.

Calibration

- If the optical system were aberration free, the Ronchigram would be a series of perfect straight lines.

- However, given the total number of positive optical components in the optical system, such a pattern is impossible

- By measuring the beam reflected from a 'perfect' reference-flat insert, after assembly, these errors can be characterized and removed from all measured data.

Certification

- The prototype was certified by measuring a known amount of defocus aberration introduced by translating the collimating lens.

$$W_{020} = -\frac{\epsilon_z}{8\,(F/\#)^2}$$

where: ϵ_z = *Longitudinal Aberration*

- Defocus was used because it is:
  - easy to introduce,
  - linear with translation, and
  - easy to analyze.

- Additionally, the prototype measured diamond-turned samples provided by NASA.

- Measurement accuracy is approximately 5%.

- The flight instrument will be certified by measuring known test plates.

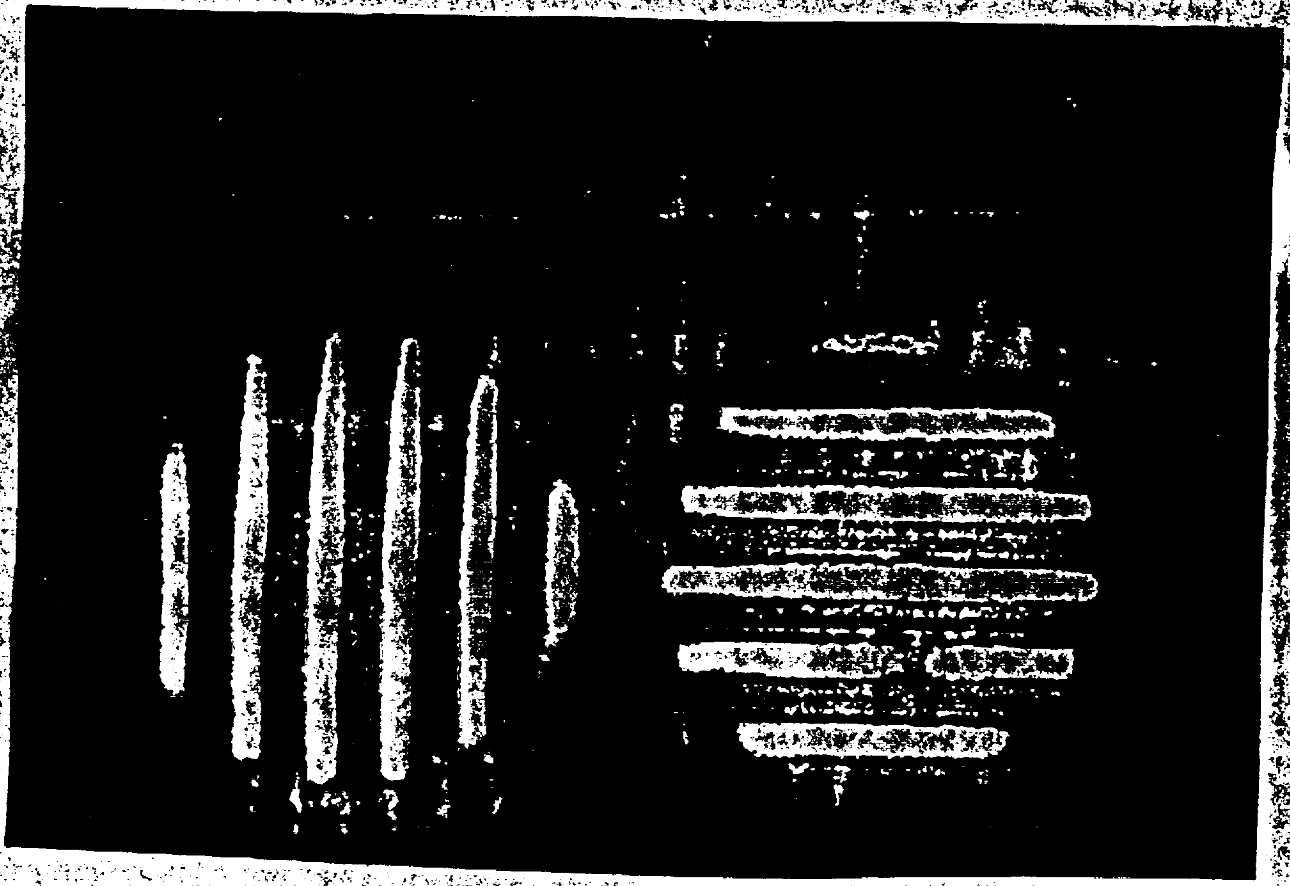
Calibration results

- For each grating, a linear relationship was observed between the amount of lens translation and the amount of measured defocus.

- Multiplying each defocus value by the appropriate equivalent wavelength produces a single line of data.

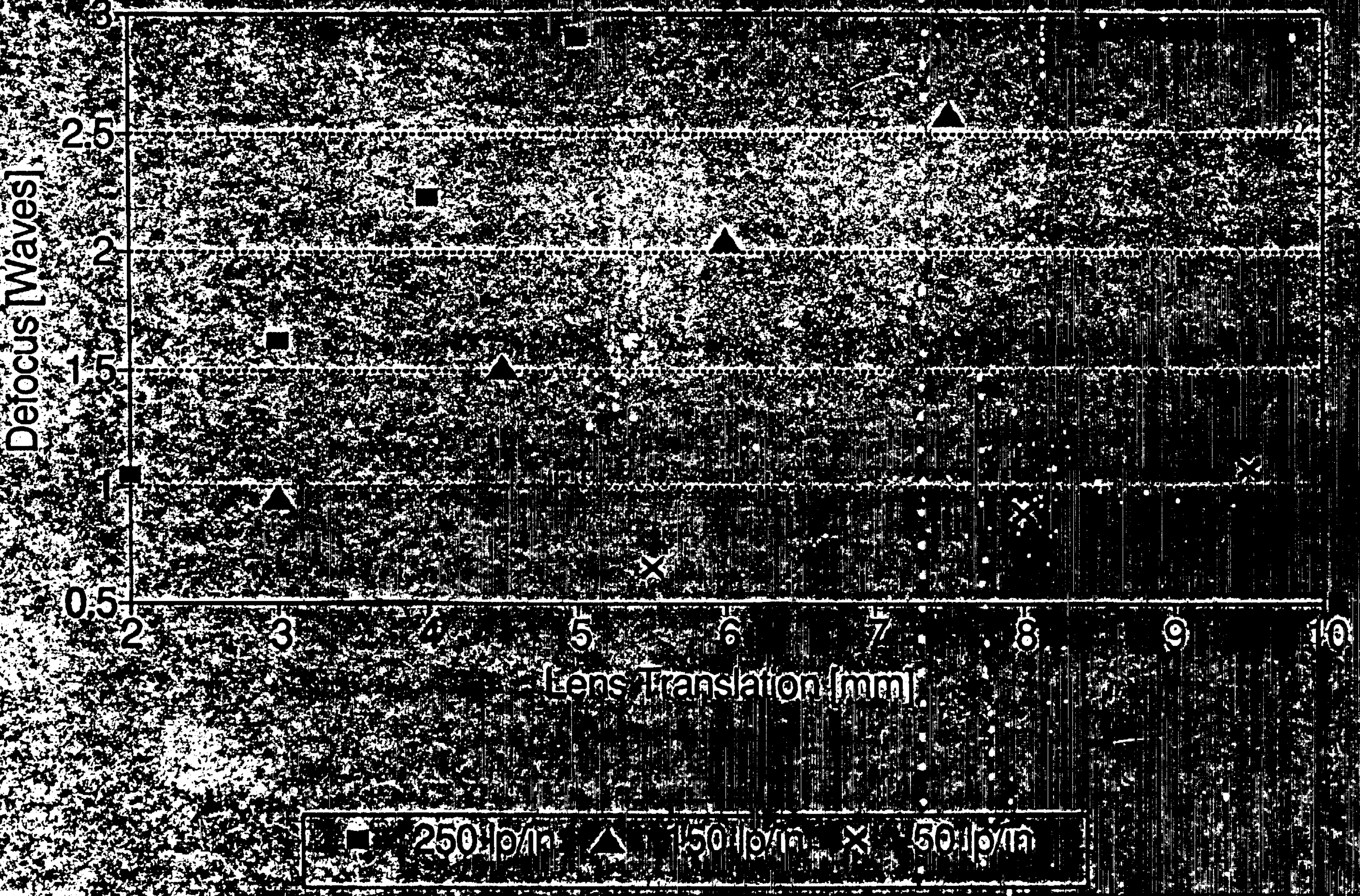
- The average percent error between the data and the best fit line is ± 5%.

Defocus Ronchigram



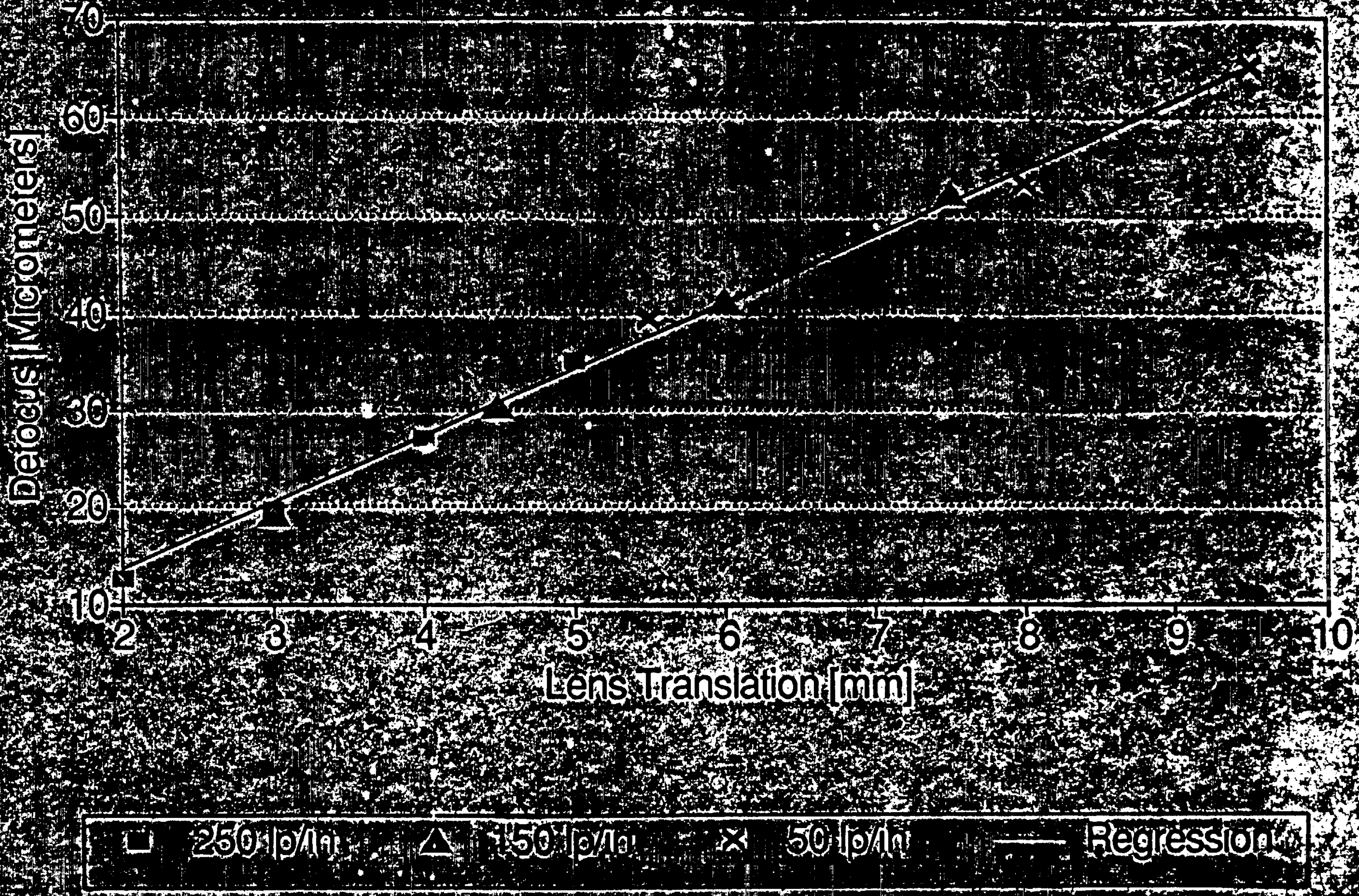
Two Channel Ronchi Calibration

Defocus



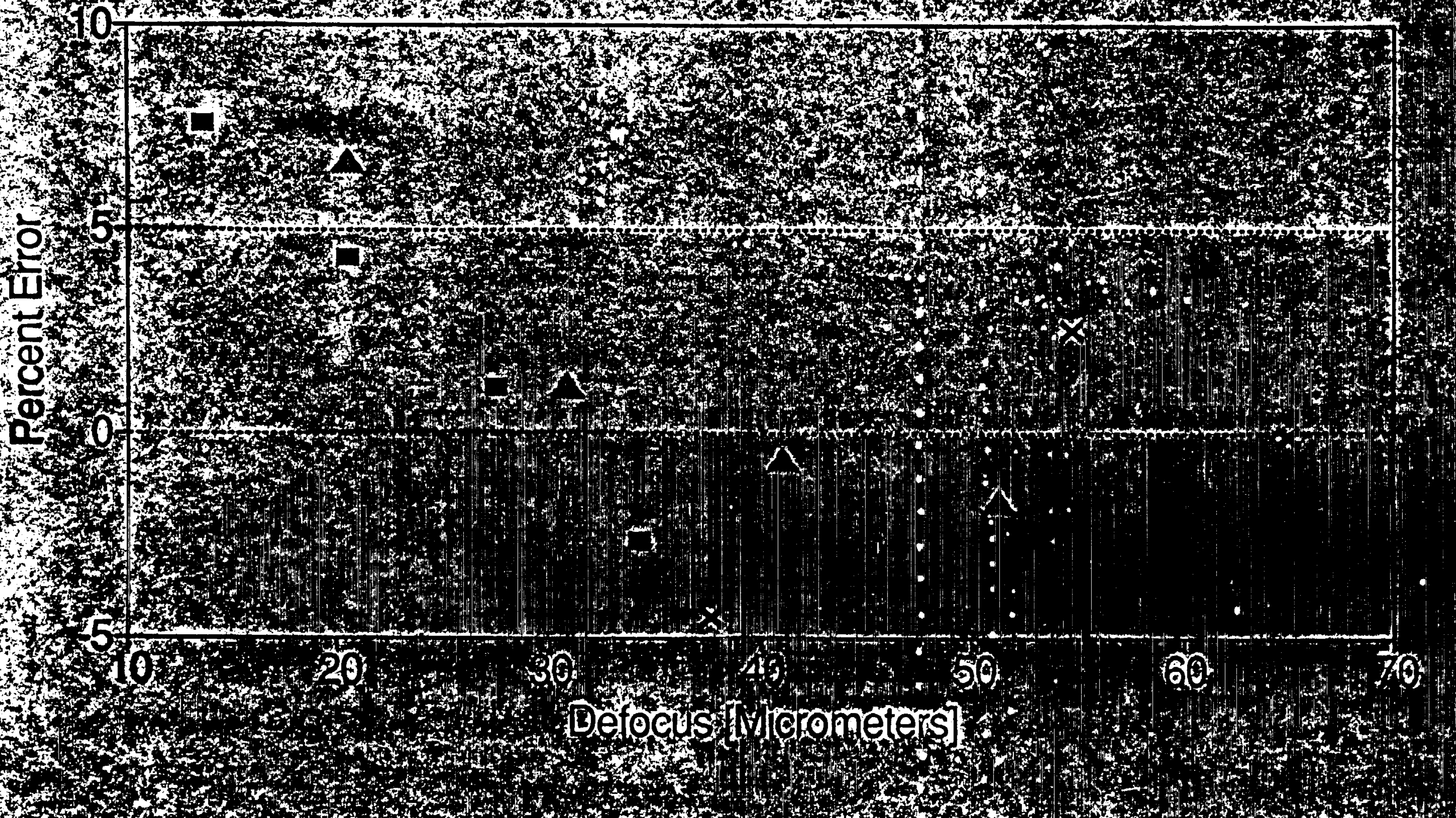
Two Channel Ronchi Calibration

Scaled With Equivalent Wavelength



Two Channel Ronchi Calibration

Percent Error



C-3

Grating Alignment

- The system is designed for the insertion of interchangeable gratings at various locations in the beam.

- Thus, grating alignment errors can be a problem.

- The data analysis was evaluated for its sensitivity to error in the position of one grating relative to its orthogonal partner:
  - Rotation about the Surface Normal
  - Rotation parallel to the Line Structure
  - Rotation perpendicular to the Line Structure
  - Translation along Surface Normal
  - Translation parallel to Line Structure
  - Translation perpendicular to Line Structure

- In general the Ronchi test is relatively insensitive to small grating placement errors.

- Except for translation along the surface normal.

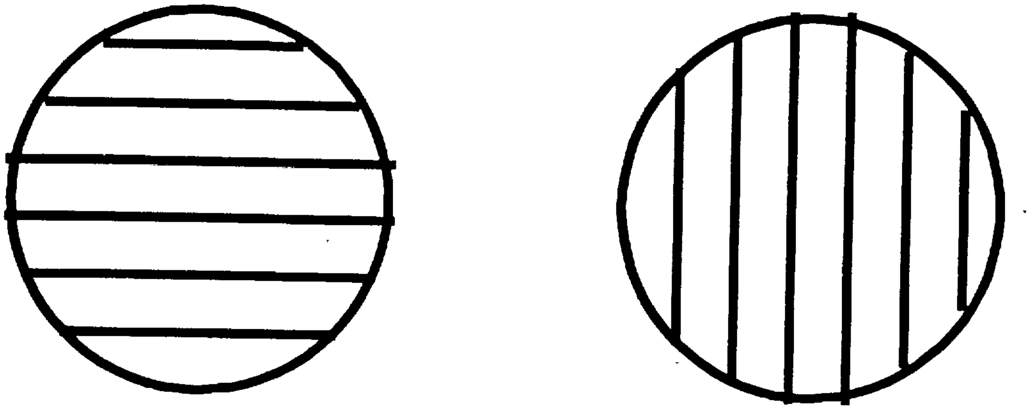


Figure 15 Ronchigrams with Gratings Aligned

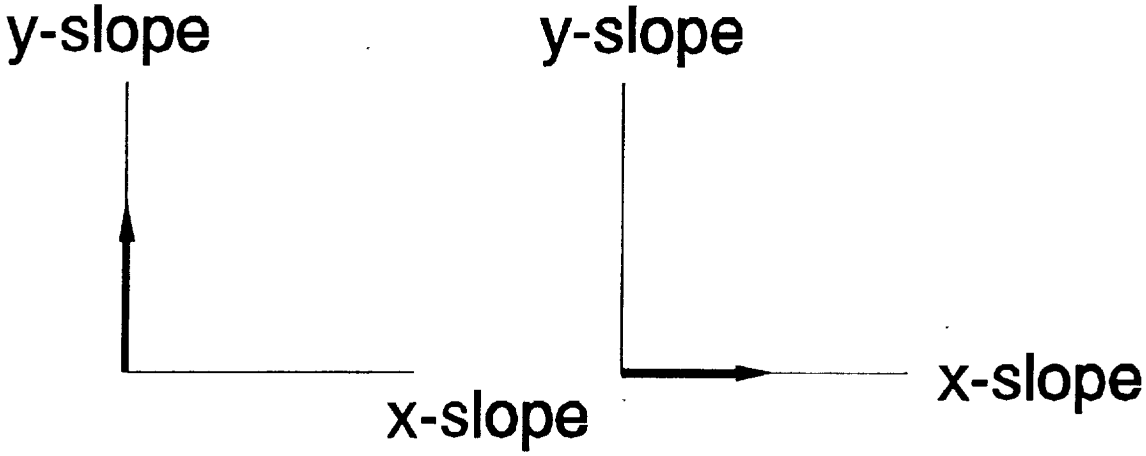


Figure 16 Vectors Corresponding to the Slope in the Ronchigrams in Figure 15.

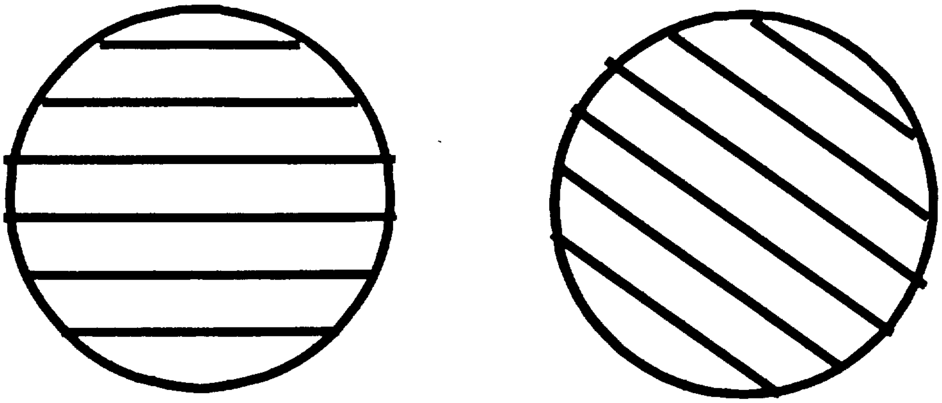


Figure 17 Ronchigrams with Grating Misalignment

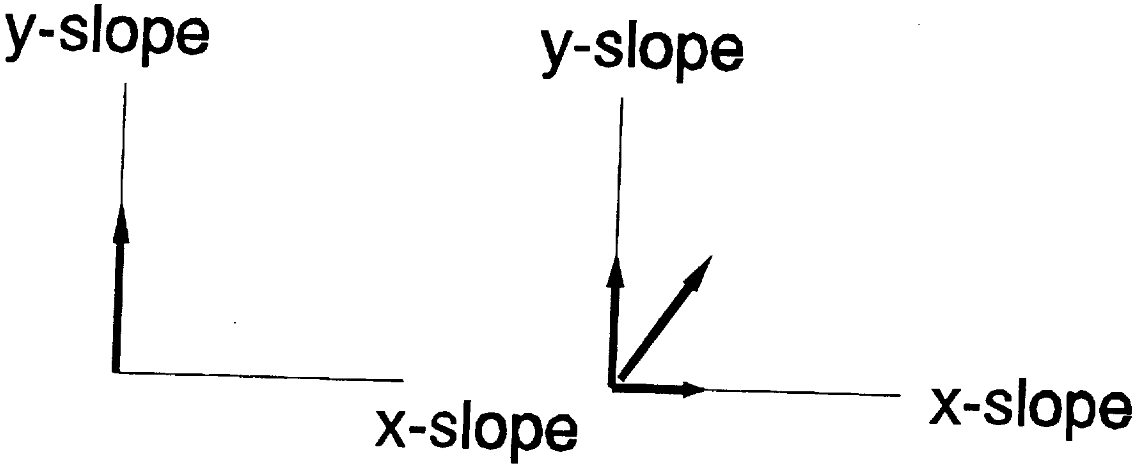
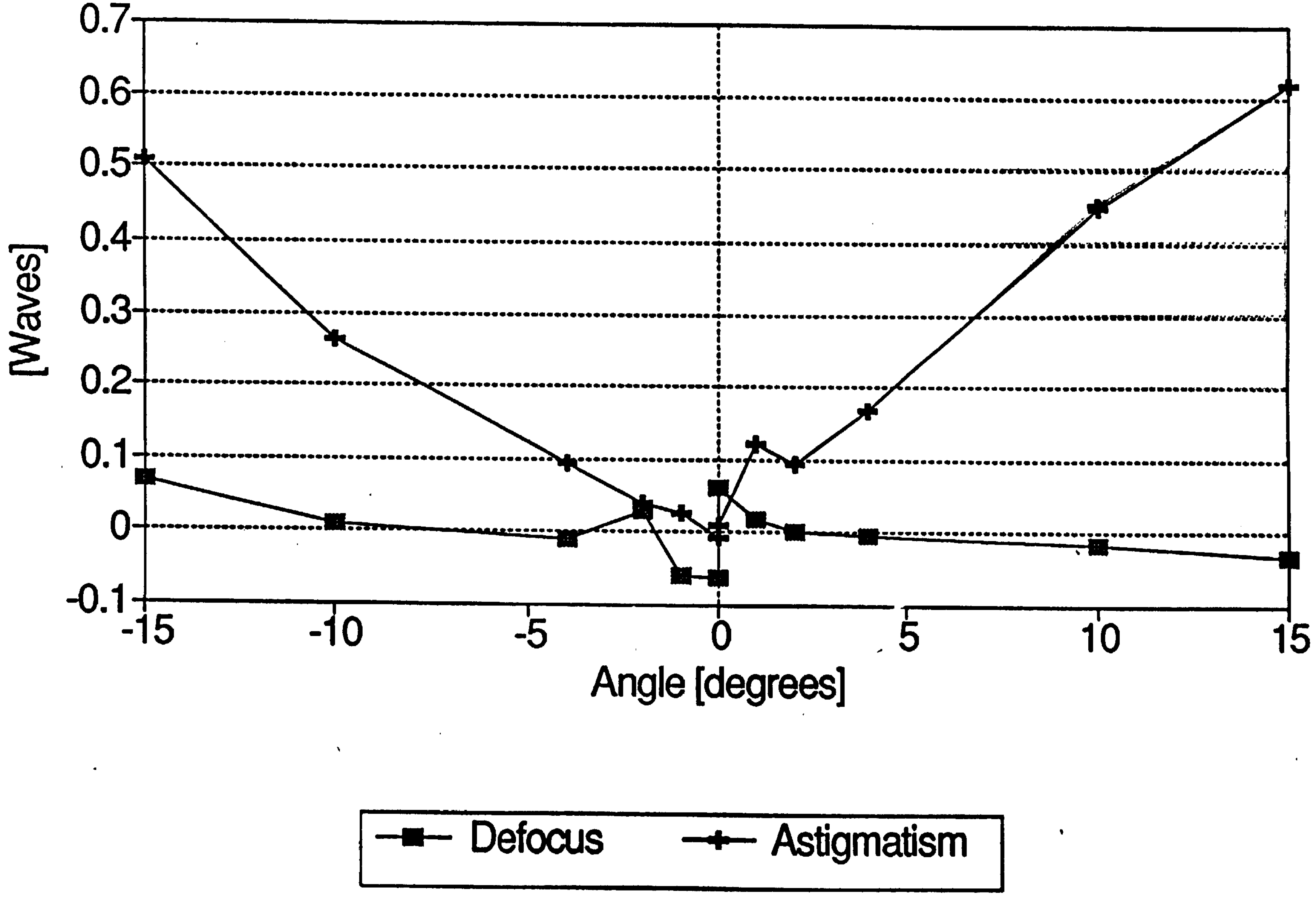


Figure 18 Vectors Corresponding to the Slope in the Ronchigrams in Figure 17.

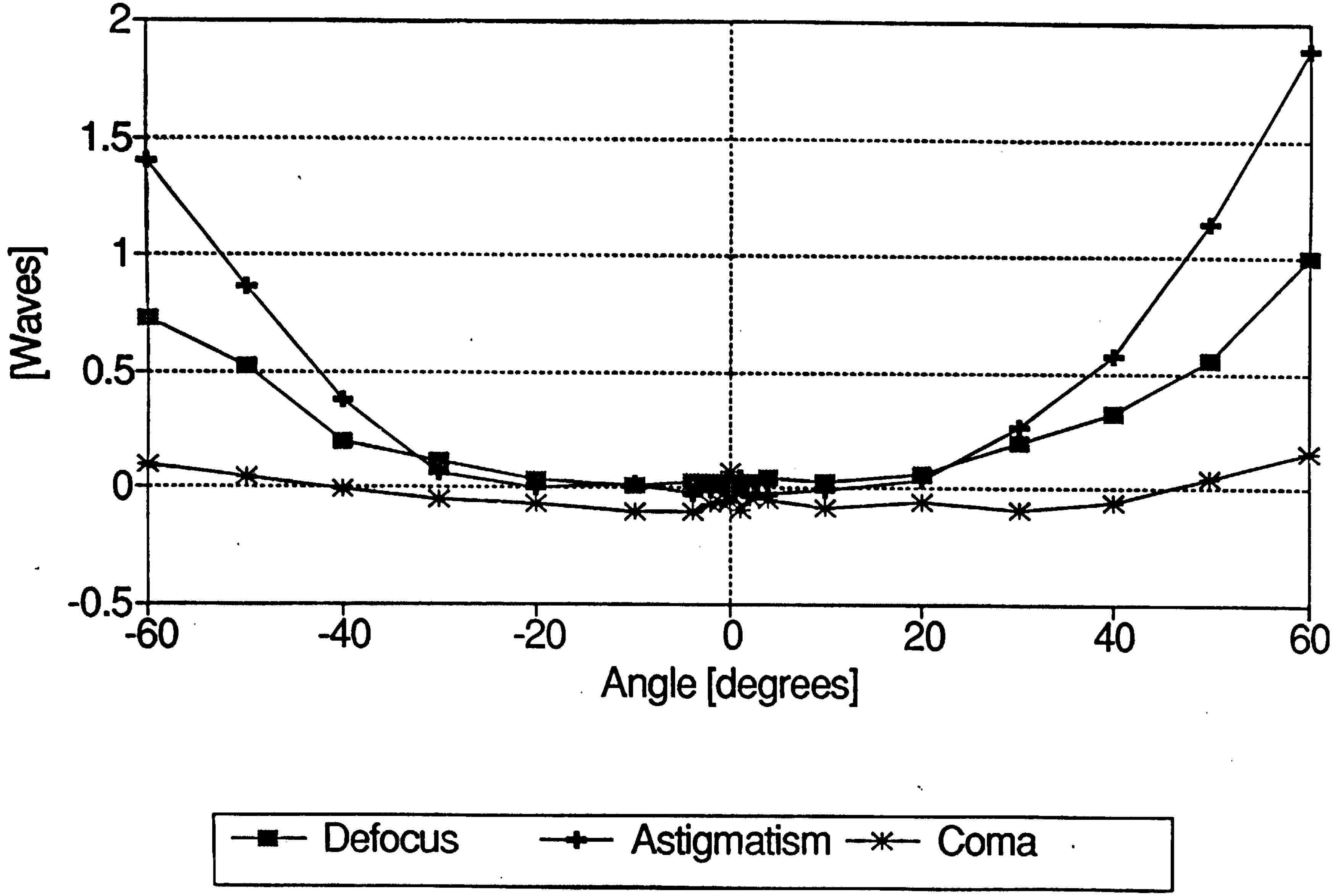
Grating Rotation

Axis Parallel to Surface Normal



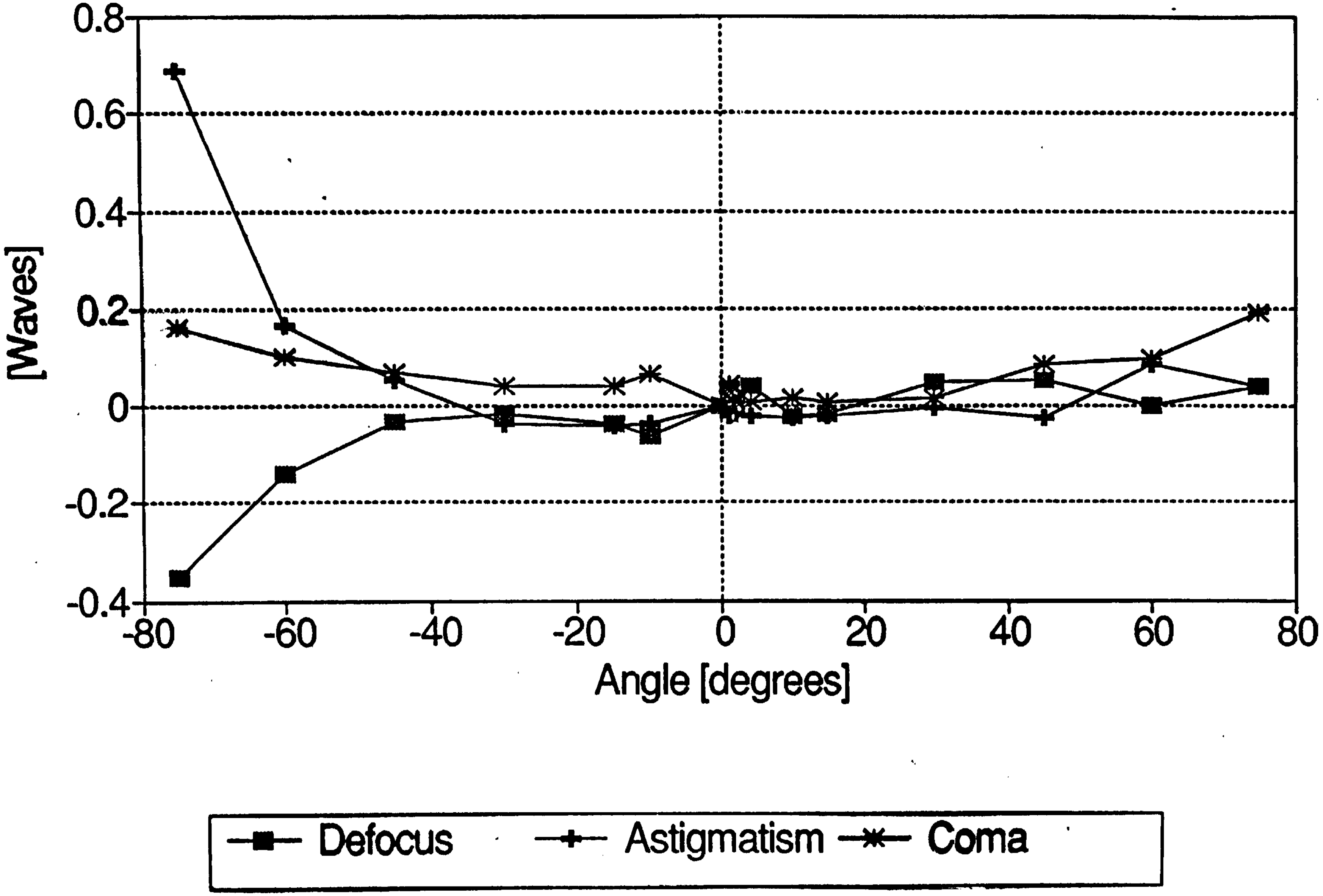
Grating Rotation

Axis Parallel to Grating Structure



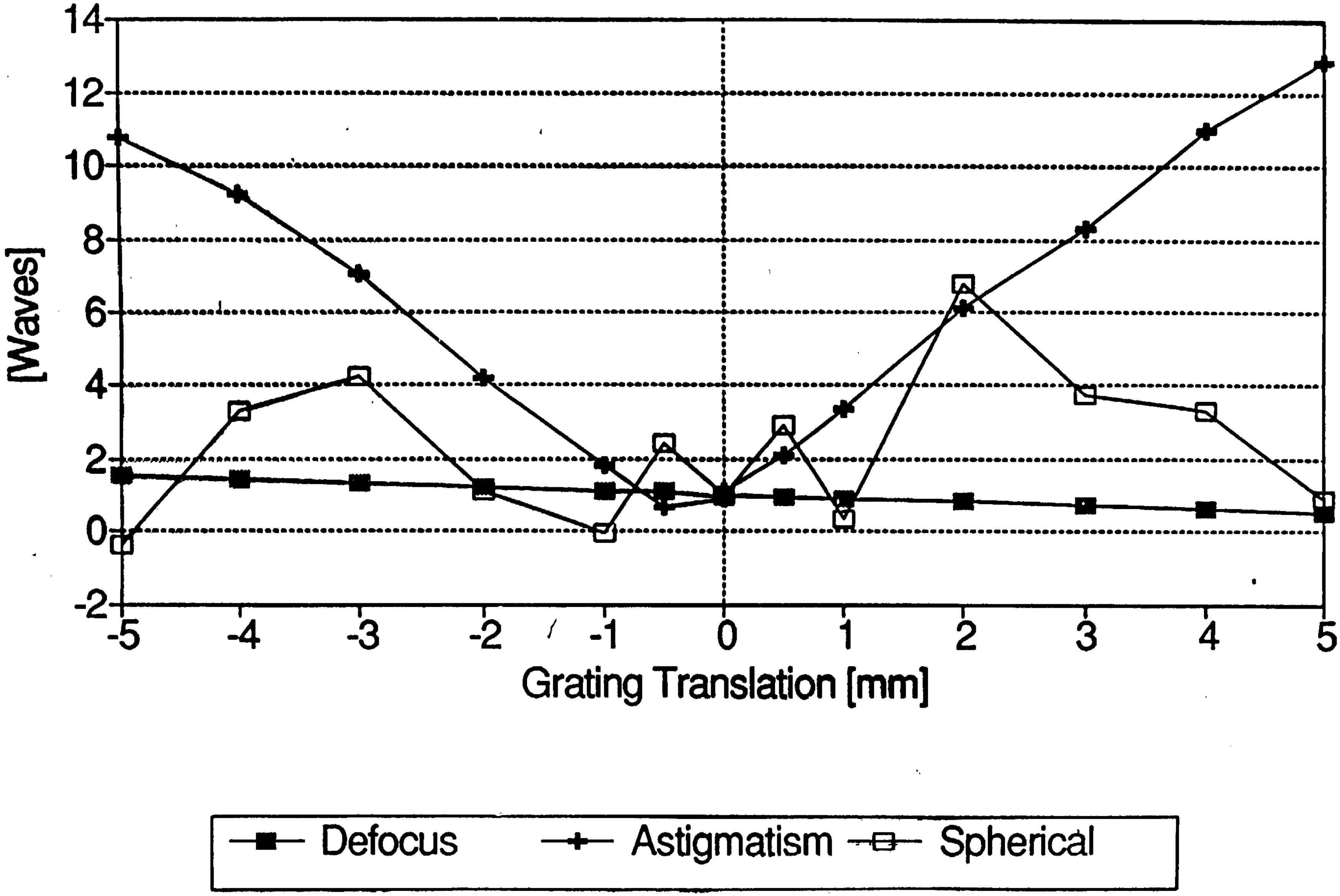
Grating Rotation

Axis Perpendicular to Grating Structure



Grating Translation

Along Surface Normal



Post-Flight Data Analysis

- The recorded Ronchigrams are analyzed after the flight to reconstruct the surface shape along a given diameter - and if possible over the entire surface.
- This is accomplished using standard interferogram analysis techniques to obtain a polynomial representation of each slope pattern.
- Surface shape along a diameter is obtained by integrating across the slope data.
- With orthogonal data sets, surface shape is obtained by a linear combination of Zernike polynomials.

Conclusions

- A prototype Ronchi instrument to monitor:

  Steady State Deformation,
  Oscillatory Deformation, and
  Fill Level

  of a free-surface of oil has been developed, demonstrated, and characterized for STDCE-2.

- A single-channel Ronchi configuration has been selected as the flight instrument.

- The flight instrument is currently being designed

Conclusions

- The Ronchi test is a shearing interferometer.
- The number of sheared wavefronts and the magnitude of the shear is dependant upon the characteristics of the grating.
- The fringe spacing (equivalent wavelength) in a Ronchigram is proportional to the spacing of the grating and the F/# of the test beam.
- The dynamic range and sensitivity of the Ronchi test can be changed by simply changing the grating spacing.
- The two channel Ronchi system has an accuracy of ± 5%.
- For a shearing system, the maximum measurable error is determined by the surface curvature.

Acknowledgement

- This project is funded by a NASA research grant (NAG3-1300) from the NASA Lewis Research Center under the direction of Alexander Pline.

- William S. Meyers developed and calibrated the prototype instrument for his Master's of Science in Applied Optics with assistance from Brent Bergner, Leif Sorensen, and Ron White.

- David J. Fischer developed and encoded the Ronchi pattern slope analysis algorithms with assistance from John O'Bryan and Dr. Robert Lopez.

- The software package has beed developed through the efforts of: Kent Flint, Brad Freriks, Kurt Louis, Tony McAllister, Steven Reid, Brad Rodeffer, John Snider, Jason Snyder, Kevin Stultz, and Dr. Dale Oexmann

Bibliography

Ceperley, Peter H. "Rotating Waves," Am. J. Phys., Vol.60, No.10, pp.938-942, October 1992.

Fischer, David J., and H. Philip Stahl, "A vector formulation for Ronchi shear surface fitting," SPIE Vol.1755, 1992.

Gaskill, J. D., Linear Systems, Fourier Transforms, & Optics, John Wiley & Sons, New York, 1978.

Mantravadi, Murty V. "Lateral Shearing Interferometers," Optical Shop Testing, John Wiley and Sons, New York, 1992.

Meyers, William S., and H. Philip Stahl, "Contouring of a free oil surface," SPIE Vol.1755, 1992.

Ronchi, Vasco, "Forty Years History of a Grating Interferometery," Applied Optics, Vol.3, pp.437-451, 1964.

Surface Tension Driven Convection Experiment, NASA, 1992.