Calibration and adjustment of coherence scanning interferometry

This item was submitted to Loughborough University’s Institutional Repository by the/an author.

Additional Information:

• A Doctoral Thesis. Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University.

Metadata Record: [https://dspace.lboro.ac.uk/2134/17357](https://dspace.lboro.ac.uk/2134/17357)

Publisher: © Rahul Mandal

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: [https://creativecommons.org/licenses/by-nc-nd/4.0/](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Please cite the published version.
Calibration and Adjustment of Coherence Scanning Interferometry

by

Rahul Mandal

Doctoral Thesis
Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University

28 November 2014

© by Rahul Mandal 2014
Contents

Abstract .................................................................................................................. 5
Acknowledgement ................................................................................................. 7
Nomenclature ......................................................................................................... 8
Abbreviation ........................................................................................................... 11
Chapter 1: Introduction and Literature Review .............................................. 14
  1.1 Introduction to Surface Metrology .............................................................. 14
  1.2 Overview of Different Surface Measurement Techniques ..................... 15
      1.2.1 Contact Profilers ...................................................................................... 15
      1.2.1.1 Stylus Instruments ................................................................................ 15
      1.2.1.2 Scanning Probe Microscopes ............................................................... 16
      1.2.2 Non-contact Profilers .............................................................................. 19
      1.2.2.1 Confocal Microscopy ............................................................................ 20
      1.2.2.2 Single Point Focus Detection Methods ............................................... 21
      1.2.2.3 Focus Variation Microscopes ............................................................... 22
      1.2.2.4 Phase Shifting Interferometry ............................................................. 23
      1.2.2.5 Interferometric Microscopic Optical Profilers .................................... 23
  1.3 Introduction to Coherence Scanning Interferometry (CSI) .................... 24
      1.3.1 CSI Working Principle ........................................................................... 25
      1.3.2 Interference Objectives .......................................................................... 27
      1.3.3 Formation of White Light Fringes .......................................................... 30
      1.3.4 Fringe Analysis ....................................................................................... 36
      1.3.5 Application of CSI .................................................................................. 43
      1.3.6 Surface Measurement Errors .................................................................. 44
Chapter 2: Traditional Calibration of Coherence Scanning Interferometry

2.1 Introduction ........................................................................................................ 62

2.2 CSI Calibration of Axial Scales .......................................................................... 63
2.2.1 Artefact Details ............................................................................................... 63
2.2.2 Calibration Procedure ..................................................................................... 64
2.2.3 Results ............................................................................................................. 65
2.2.4 Problems in Axial Calibration .......................................................................... 66

2.3 CSI calibration of lateral scales ............................................................................ 67
2.3.1 Artefact Detail .................................................................................................. 67
2.3.2 Results ............................................................................................................. 68
2.3.3 Problems ......................................................................................................... 69

2.4 Cut-off Frequencies in CSI .................................................................................. 70
2.4.1 Optical Cut-off Frequency .............................................................................. 70
2.4.2 Instrument Cut-off Frequency ......................................................................... 71
2.4.3 Calculation of Optical and Instrument Cut-off Frequencies ......................... 71
2.4.4 Calculation of Cut-off Frequencies: Zygo Instrument .................................... 71
2.4.5 Calculation of Cut-off Frequencies: TH Instrument ....................................... 73

2.5 Effects of Aliasing ............................................................................................... 74

2.6 Focusing .............................................................................................................. 76

2.7 Conclusion ........................................................................................................... 80

Chapter 3: Coherence Scanning Interferometry: Linear Theory of Surface Measurement ................................................................................................................. 83
Chapter 6: Measurement and Adjustment of Distortion in CSI

6.1 Introduction ................................................................................................................. 197
6.2 Effects of Distortion in CSI measurement ................................................................. 198
6.3 Measurement of Distortion ......................................................................................... 201
6.4 Quantification of Tilt Related Errors ......................................................................... 204
6.5 Distortion Compensation ............................................................................................ 206
6.6 Effect of Distortion in the Measurement ...................................................................... 208
6.7 Conclusion ..................................................................................................................... 210

Chapter 7: Discussion and Future Work .......................................................................... 212

7.1 Discussion ..................................................................................................................... 212
7.2 Future Work .................................................................................................................. 215
7.3 Publications Arising from the Work in This Thesis ..................................................... 218

Appendix ............................................................................................................................. 220

Effect of adhesion forces on the silica spheres ................................................................. 220

References ......................................................................................................................... 224
Abstract

Coherence scanning interferometry (CSI) is a non-contacting optical technique which is widely used for the measurement of surface topography. CSI combines the lateral resolution of a high power microscope with the axial resolution of an interferometer. As with any other metrology instrument, CSI is calibrated to define measurement uncertainty. The traditional calibration procedure, as recommended by instrument manufacturers, consists of calibration of the axial and lateral scales of the instrument. Although calibration in this way provides uncertainties for the measurement of rectilinear artefacts, it does not give information about tilt-related uncertainty. If an object with varying slope is measured, significant errors are observed as the surface gradient increases.

In this thesis a novel approach of calibration and adjustment for CSI using a spherical object is introduced. This new technique is based on three dimensional linear filtering theory. According to linear theory, smooth surface measurement in CSI can be represented as a linear filtering operation, where the filter is characterised either by point spread function (PSF) in space domain or by transfer function (TF) in spatial frequency domain. The derivation of these characteristics usually involves making the Born approximation, which is strictly only applicable for weakly scattering objects. However, for the case of surface scattering and making use of the Kirchhoff approximation, the system can be considered linear if multiple scattering is assumed to be negligible. In this case, the object is replaced by an infinitely thin
foil-like object, which follows the surface topography and, therefore, is called the “foil model” of the surface.

For an ideal aberration free instrument, the linear characteristics are determined by the numerical aperture of the objective lens and the bandwidth of the source. However, it is found that the PSF and TF of a commercial instrument can depart significantly from theory and result in a significant measurement error. A new method, based on modified inverse filter to compensate the phase and amplitude-related errors in the system PSF/TF, is demonstrated.

Finally, a method based on de-warping to compensate distortion is discussed. The application of the linear theory as well as modified inverse filter is dependent on the assumption of the shift invariance. As distortion introduces a field dependent magnification, the presence of distortion for CSI with relatively large field of view, restricts the applicability of the linear theory. Along with this restriction, distortion also introduces erroneous height measurement for objects with gradients. This new approach, based on de-warping, resolves the problems associated with distortion.
Acknowledgement

Firstly, I would like to thank my supervisor Prof Jeremy Coupland for his encouragement, support and guidance over the last four years. The dissertation would not have been possible without his help and supervision. I am also thankful to my industry supervisor Prof Richard Leach for his help and advice during the course of research.

I am grateful to NPL, Taylor and Hobson and Loughborough University for providing financial support to this project.

My special thanks to Prof Jon Petzing and Jagpal singh for the help they provided in the metrology laboratory during the use of Zygo New View 5000.

My thanks also go to my fellow PhD students and other group members who have given me advice when needed and I acknowledge them for maintaining a friendly research environment. I also want to thank Dr Joshua Vande Hey, Phil Kokoszka and Thomas Bartsch for proof reading my thesis.

Finally I would like to thank my parents and family for their unconditional love, support and inspiration during my pursuit of education.
### Nomenclature

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>\langle A \rangle</td>
<td>Field average of the phase gap</td>
</tr>
<tr>
<td>A(r)</td>
<td>Upper surface of the object for foil model</td>
</tr>
<tr>
<td>A(x,y)</td>
<td>Phase gap at (x, y)</td>
</tr>
<tr>
<td>C</td>
<td>Dewarping matrix</td>
</tr>
<tr>
<td>D</td>
<td>Distortion coefficient</td>
</tr>
<tr>
<td>E(r)</td>
<td>The electric field</td>
</tr>
<tr>
<td>E_r(r)</td>
<td>Illuminating field</td>
</tr>
<tr>
<td>E_s(r)</td>
<td>Scattered field</td>
</tr>
<tr>
<td>E_t(r)</td>
<td>Transmitted field</td>
</tr>
<tr>
<td>E_m(r')</td>
<td>Reconstructed field from boundary</td>
</tr>
<tr>
<td>f_{optical}</td>
<td>Optical cut-off frequency</td>
</tr>
<tr>
<td>f_{inst}</td>
<td>Instrument cut-off frequency</td>
</tr>
<tr>
<td>G(r)</td>
<td>The Green’s function</td>
</tr>
<tr>
<td>\tilde{G}_{ideal}(k)</td>
<td>Transfer function for ideal far field imaging</td>
</tr>
<tr>
<td>\tilde{G}_{NA}(k)</td>
<td>Numerical aperture restricted TF for far field imaging</td>
</tr>
<tr>
<td>h</td>
<td>Surface height</td>
</tr>
<tr>
<td>H(r)</td>
<td>The impulse response/ point spread function (PSF)</td>
</tr>
<tr>
<td>\tilde{H}(k)</td>
<td>Transfer function (TF)</td>
</tr>
<tr>
<td>H_{holo}(r)</td>
<td>PSF for a monochromatic holographic reconstruction</td>
</tr>
<tr>
<td>H_{B^{QM}}(r)</td>
<td>PSF for CSI in quasi-monochromatic mode</td>
</tr>
<tr>
<td>H_{B}(r)</td>
<td>PSF for CSI considering the Born approximation</td>
</tr>
<tr>
<td>H_F(r)</td>
<td>PSF for CSI considering the foil model</td>
</tr>
<tr>
<td>\tilde{H}_{inv}(k)</td>
<td>TF for the inverse filter</td>
</tr>
<tr>
<td>\tilde{H}^{mod}_{foil}(k)</td>
<td>TF for the modified inverse filter</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>$I(k, z, \theta)$</td>
<td>Recorded interference signal</td>
</tr>
<tr>
<td>$I(r)$</td>
<td>The input of a linear system</td>
</tr>
<tr>
<td>$I_{dc}$</td>
<td>The dc irradiance</td>
</tr>
<tr>
<td>$I_R$</td>
<td>Intensity of light reflected from the reference mirror</td>
</tr>
<tr>
<td>$I_O$</td>
<td>Intensity of the light reflected from the object</td>
</tr>
<tr>
<td>$k$</td>
<td>Wave number</td>
</tr>
<tr>
<td>$k_0$</td>
<td>The free space wave number,</td>
</tr>
<tr>
<td>$k_{max}$</td>
<td>Maximum wave number</td>
</tr>
<tr>
<td>$k_{min}$</td>
<td>Minimum wave number</td>
</tr>
<tr>
<td>$k_r$</td>
<td>The reference wave vector</td>
</tr>
<tr>
<td>$n(r)$</td>
<td>The refractive index</td>
</tr>
<tr>
<td>$N_A$</td>
<td>Numerical aperture</td>
</tr>
<tr>
<td>$n_s$</td>
<td>The outward surface normal</td>
</tr>
<tr>
<td>$O(r)$</td>
<td>The output of a linear system</td>
</tr>
<tr>
<td>$R$</td>
<td>Radius of the sphere shown in figure 3.2</td>
</tr>
<tr>
<td>$r_b$</td>
<td>Distance to the boundary</td>
</tr>
<tr>
<td>$\Delta r_{lateral}^x$</td>
<td>Resolution in $x$ direction</td>
</tr>
<tr>
<td>$\Delta r_{lateral}^y$</td>
<td>Resolution in $y$ direction</td>
</tr>
<tr>
<td>$\Delta r_{axial}$</td>
<td>Resolution in $z$ direction</td>
</tr>
<tr>
<td>$S$</td>
<td>Area of the sphere shown in figure 3.2</td>
</tr>
<tr>
<td>$S(k)$</td>
<td>The spectral distribution of the source</td>
</tr>
<tr>
<td>$U(r)$</td>
<td>Source term</td>
</tr>
<tr>
<td>$V$</td>
<td>Volume of the sphere shown in figure 3.2</td>
</tr>
<tr>
<td>$W(k)$</td>
<td>Weighting function used in modified inverse filter</td>
</tr>
<tr>
<td>$z$</td>
<td>The axial position</td>
</tr>
<tr>
<td>$\frac{\partial}{\partial n}$</td>
<td>Partial derivative in the outward normal direction</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>$\phi$</td>
<td>The phase change on reflection</td>
</tr>
<tr>
<td>$\theta_B$</td>
<td>Brewster’s angle</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Fringe visibility</td>
</tr>
<tr>
<td>$\theta (x,y)$</td>
<td>Interference phase at $(x,y)$</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>The free space wavelength of light</td>
</tr>
<tr>
<td>$\bar{\lambda}$</td>
<td>Effective wavelength</td>
</tr>
<tr>
<td>$\Delta(r)$</td>
<td>Object defined as the refractive index contrast</td>
</tr>
<tr>
<td>$\Delta_F(r)$</td>
<td>Object defined as the foil model</td>
</tr>
<tr>
<td>$\delta(r - r')$</td>
<td>Three dimensional Dirac delta function</td>
</tr>
</tbody>
</table>
## Abbreviation

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFM</td>
<td>Atomic force microscope</td>
</tr>
<tr>
<td>CCD</td>
<td>Charged coupled device</td>
</tr>
<tr>
<td>CCI</td>
<td>Coherence correlation interferometry</td>
</tr>
<tr>
<td>CMM</td>
<td>Coordinate measuring machine</td>
</tr>
<tr>
<td>CPM</td>
<td>Coherence probe microscope</td>
</tr>
<tr>
<td>CR</td>
<td>Coherence radar</td>
</tr>
<tr>
<td>CSI</td>
<td>Coherence scanning interferometry</td>
</tr>
<tr>
<td>EM</td>
<td>Electron microscopy</td>
</tr>
<tr>
<td>FED</td>
<td>Frequency domain analysis</td>
</tr>
<tr>
<td>FT</td>
<td>Fourier transform</td>
</tr>
<tr>
<td>FV</td>
<td>Focus variation</td>
</tr>
<tr>
<td>HCF</td>
<td>Helical conjugate function</td>
</tr>
<tr>
<td>IFT</td>
<td>Inverse Fourier transform</td>
</tr>
<tr>
<td>IM</td>
<td>Interference microscope</td>
</tr>
<tr>
<td>LED</td>
<td>Light emitting diode</td>
</tr>
<tr>
<td>LVDT</td>
<td>Linear variable differential transformer</td>
</tr>
<tr>
<td>MCM</td>
<td>Mirau correlation microscope</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>MEMS</td>
<td>Micro-electro-mechanical system</td>
</tr>
<tr>
<td>MOEMS</td>
<td>Micro-opto-electro-mechanical system</td>
</tr>
<tr>
<td>NA</td>
<td>Numerical aperture</td>
</tr>
<tr>
<td>NPL</td>
<td>National physical laboratory</td>
</tr>
<tr>
<td>NIST</td>
<td>National institute of standards and technology</td>
</tr>
<tr>
<td>OPD</td>
<td>Optical path difference</td>
</tr>
<tr>
<td>PSI</td>
<td>Phase shifting interferometry</td>
</tr>
<tr>
<td>PSF</td>
<td>Point spread function</td>
</tr>
<tr>
<td>PUPS</td>
<td>Pupil plane scanning white light interferometry method</td>
</tr>
<tr>
<td>QM</td>
<td>Quasi-monochromatic mode</td>
</tr>
<tr>
<td>RMS</td>
<td>Root mean square</td>
</tr>
<tr>
<td>SEM</td>
<td>Scanning electron microscope</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to noise ratio</td>
</tr>
<tr>
<td>SPM</td>
<td>Scanning probe microscope</td>
</tr>
<tr>
<td>STM</td>
<td>Scanning tunnelling microscope</td>
</tr>
<tr>
<td>SWLI</td>
<td>Scanning white light interferometry</td>
</tr>
<tr>
<td>TEM</td>
<td>Transmission electron microscope</td>
</tr>
<tr>
<td>TF</td>
<td>Transfer function</td>
</tr>
<tr>
<td>TH</td>
<td>Taylor and Hobson</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>VSI</td>
<td>Vertical scanning interferometry</td>
</tr>
<tr>
<td>WLI</td>
<td>White light interferometry</td>
</tr>
<tr>
<td>2D</td>
<td>Two dimensional</td>
</tr>
<tr>
<td>3D</td>
<td>Three dimensional</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction and Literature Review

1.1 Introduction to Surface Metrology

Surface metrology is the branch of science that is concerned with the measurement and characterisation of surface topology [1-4]. The condition of engineered surfaces can have a profound effect, for example, on lubricating flows and has a strong influence on the frictional losses and lifetime of bearings [1]. With the development of micro-electronics and micro electromechanical sensors (such those found in mobile phones) surface metrology is increasingly important as a means to measure the form of structured surfaces and even extends to the measurement of thin films [2, 4]. Traditionally surface measurement has been achieved using stylus based instrumentation and these instruments form the basis of ISO standards [5-9]. In recent years however, the capability of surface measuring instruments has been greatly extended by the introduction of scanning probe techniques and optical instrumentation. In the following section a general overview of different surface measurement instruments is given.
1.2 Overview of Different Surface Measurement Techniques

1.2.1 Contact Profilers

The development of surface measurement instruments started with contacting instruments. There are two main types of instruments that may be classed as contact profilers, a) stylus profilers and b) scanning probe microscopes. Both types utilise a probe which scans across the surface and physically interacts with the surface in order to give the measurement data. The stylus type of instrument can provide traceable measurements with high vertical (nm) and lateral resolution (µm) and is relatively inexpensive to build which makes this approach the method of choice for routine measurement as well as for calibration purposes. On the other hand, scanning probe systems measure (to a greater or lesser extent) the sample probe interaction which along with high resolution can potentially provide information about the material properties of the object [1-4, 10].

1.2.1.1 Stylus Instruments

As the name suggests, these instruments employ a stylus which scans across the surface of the object and gives information about the height profile. The stylus of the instrument is typically made out of a hard material such as diamond with a radius of curvature of 0.5 to 50 µm. The shape and radius of curvature of the stylus, along with the surface shape [11] and sampling interval between the data points [12], determine the lateral resolution of the system [13, 14]. The size, shape and cone angle of the stylus tip also determine the measured aspect ratio of the structures and regulates the maximum measurable slope. A smaller and sharper styli radius allows
the stylus to follow the shape of the surface more easily, which in turn increases the resolution of the system. However, a sharp stylus tip can result in higher pressure which may potentially deform the surface and lead to measurement error. In the worst case, it can damage the surface of the object. Usually a stylus load of 0.1 mg to 50 mg is applied to minimise deformation [3] but there is clearly a trade-off between resolution and measurement error that depends on the object and the stylus dimension.

While scanning the surface the vertical motion of the stylus is detected by a linear variable differential transformer (LVDT) and the corresponding recorded signal is converted into height data. Stylus profilers can measure surface roughness with a RMS of 0.5 angstrom with lateral resolution of 0.1 – 0.2 µm. These profilers can be used to measure profiles of surfaces up to 200 mm in length or further by stitching several scans [1, 2]. The process is relatively slow however [10, 15], and the primary disadvantages of stylus instrumentation as a process tool are both speed and the chance of damage to the object surface.

1.2.1.2 Scanning Probe Microscopes

The other type of profiler is the scanning probe microscope (SPM). Here a fine probe is either moved in close proximity to the test surface, usually within few angstroms, or the probe is in direct contact with the surface while measuring the surface information [16- 18]. Due to the nature of the surface interaction the applied force is generally small compared to general stylus instruments. SPMs can be categorised as a) scanning tunnelling microscope and b) atomic force microscopy, sometimes referred to as scanning force microscope.
a. Scanning Tunnelling Microscope

The scanning tunnelling microscope (STM) was first reported by Binnig and Rohrer in 1982 [19] and they won a Nobel Prize in physics in 1986 for designing the same [20]. In this case a metal probe with a fine tip is brought close to the conducting or semiconducting test surface until a tunnelling current is detected [21]. A voltage is applied between the tip and the object surface to obtain the tunnelling current. The tunnelling starts when the tip is less than 1 nm away from the surface and as the probe is moved closer to the surface the amount of tunnelling current increases exponentially. The probe is connected with a pizeo electric controller, which scans along the surface, in order to get the surface data. It has been successfully utilized to evaluate optical surfaces since the mid-1980s [22-24]. The tunnelling microscope can operate in two modes a) constant height mode and b) constant current mode. In constant height mode, a constant height is maintained during the measurement and the surface information is obtained by identifying the changing current as it scans the surface. In order to provide surface topology the probe/surface interaction characteristic must be accurately known and in addition, there is the potential to damage the surface (particularly while moving the surface or the tip during measurement of rough surfaces). Constant current mode overcomes these problems as either the tip or the object surface is moved closer or further away as the scan goes along in order to maintain the same tunnelling current. The probe position now traces the surfaces topology. Both the operating modes provide excellent axial resolution. Despite the advantages, the main problem is that it can only be used for measurement of conductive surfaces so surfaces such as glass cannot be measured using this technique.
b. Atomic Force Microscopy

Atomic force microscopy (AFM) can be thought of as a modified version of STM that overcomes some of its limitations. In 1986 Binnig invented AFM (the patent was filed in 1986 [25]) and the first experimental implementation was made in the same year by Binnig, Quate and Gerber [26]. It can measure any surface irrespective of its conductivity. In AFM the tip of probe is located at the free end of a flexible cantilever. The tip is made to contact the surface and as it scans through the object the forces between the tip and the sample cause the probe to bend. The deflection data is recorded and analysed in order to obtain the surface information. In 1989 Alexander et al. [27] developed a readout system where the deflection is measured by a laser signal reflected from a mirror mounted on the cantilever. Another deflection measurement technique which incorporated use of a diode laser was reported by Sarid et al [28]. This is the contact mode of AFM. In addition to the contact mode, AFM can be operated in a non-contact mode. In that mode, the attractive force between the surface and the probe tip is measured while the tip is oscillated in high frequency [29-30]. A third operating mode of AFM known as the tapping mode was reported in 1993 by Zhong et al and became the most common operation mode for AFM. In this mode the surface information is obtained by lightly tapping the surface with the probe which is oscillating at the frequency close to the cantilever’s resonance frequency [31, 32].

AFM is widely used for surface measurement purposes and provides lateral resolution of nanometer scale, and sub-nanometer axial resolution [33]. The drawback is its speed and the cost of the measurement. As it scans point by point on the surface, it is a slow speed process. Like any other contact method, contact and tapping mode AFM have the potential to damage the surface, and similar to stylus
instrument measurement the uncertainties for contacting AFM depend on the probe shape for measurement of rough structures and high aspect ratio structures [34].

Though contact profilers provide high resolution, often this comes at the cost of low speed and high expense. In contrast to using a mechanical probe to map surface topography, non-contact methods are gaining popularity, especially optical profilers which determine the surface by sensing the best focus position on the test object over a full field of view in a relatively short time duration.

1.2.2 Non-contact Profilers

Non-contact optical methods hold significant advantages over contacting instrumentation [35]. Firstly, they do not use any probe or stylus for measurement, which eliminates the risk of damaging the surface. Usually contact profilers are slow as they scan the surface point by point. In contrast, for most non-contact techniques the object is illuminated with an electromagnetic wave and the response is obtained over the whole field of view, which allows information to be recorded much faster (except for the confocal microscopy).

Electron microscopy (EM) is one of the non-contact techniques which was reported back in 1933 by Ruska and over the years it has been used for surface topography measurement [36]. There are two primary types of electron microscopes a) scanning electron microscopes (SEM) and b) transmission electron microscope (TEM). For surface measurement purposes SEM is used most and it provides resolution in the nm range. In SEM high energy electrons are accelerated towards the object surface and on impact produce secondary electrons which are detected and processed to form an image of the surface [37]. For a conventional SEM the samples must be
electrically conductive, and electrically grounded to prevent the accumulation of electrostatic charge. As non-conductive objects need to be coated with conducting material (such as gold) before being placed in the vacuum chamber, SEM has limited capability for measurement of non-conducting materials.

The most commonly used non-contact profilers are based on optical techniques. Some of these such as focus detection, phase shifting interferometry, confocal microscopy and interference microscopic optical profilers are discussed in the following section.

1.2.2.1 Confocal Microscopy

Confocal microscopy is a technique that is now used routinely to provide 3D sectioned images in medicine and is beginning to gain acceptance as a surface measuring tool. In 1961 Minsky patented a technique based on modifications to a biological microscope that can reduce the stray light in the system in order to improve the image quality. This later became popular as confocal microscopy [38]. Though confocal microscopy was initially designed to measure biological samples, by the 1980s it was widely used in the measurement of engineering surfaces. In a confocal microscope light from a point illumination is focused on a point of the object surface, rather than illuminating the whole surface. The image of this point is filtered through a pinhole that is placed at the confocal image of the point of focus before being received by the detector. This confocal geometry restricts the out of focus illumination and stray light, and results in high resolution and high signal-to-noise ratio images. It can be used with two types of scanning. In one case a two dimensional (2D) image is formed by x-y lateral scanning while in the other axial scanning is used to obtain a three dimensional (3D) profile [39-41]. Over the years
several methods have been reported to improve the speed of the data acquisition for confocal microscopy. Most of them are based on modification of the confocal aperture [39].

**1.2.2 Single Point Focus Detection Methods**

Single point focus detection is similar in principle to confocal microscopy. A single point focus detection instrument was first reported by Simon in 1970 [42] and a modified version of it was published by Simpson in the following year [43]. In this case the height information is obtained by finding focus on a single point on the object surface and adjusting the height of the focusing lens until the focus is achieved. Sawatari et al [44] developed this instrument using a normal microscope with vertical laser illumination focused on the object, but the image was split into two parts using two beam splitters. In 1987 Broadman et al incorporated optical wedges in order to deflect the beams instead of using two separate beam splitters [45]. Two photo-detectors are used, one before the image plane, and the other one at an equal distance behind the image plane. If the focusing lens is not at a low or high position, there will be a difference in the signal in the two detectors. The difference signals from the two detectors are monitored to balance the position of the focusing lens to get vertical resolution on a nanometer scale.

The lateral resolution of this type of instrument is dependent on the size of the focus spot at the object surface, which is usually 1 to 1.5 µm in diameter [45]. In this instrument it is assumed that light reflected from the surface goes back to the detectors. For steep slopes or rough surfaces the scattered light may not able to reach the detector. As this system is dependent on the difference signal from the two detectors, these types of surfaces will cause errors.
1.2.2.3 Focus Variation Microscopes

In the 1990s a new concept was developed for use with high resolution electron microscopes [46] known as focus variation (FV). It uses a small depth of focus optical system with vertical scanning to provide topographical and colour information from variation of focus. Small depth of focus causes only small regions of the object to image sharply while the vertical scanning captures data for each regions being focused during the scan. In contrast to many other optical techniques which use coaxial illumination, the maximum measureable slope in this case is not dependent on the numerical aperture of the objective. It is therefore possible to measure slope angles exceeding 80 degrees using a different type of illumination such as a ring light. Another advantage is that the measurement provides optical colour image which enables easy identification and localization of measurement fields of distinctive surface features [47-48].

As the FV technique depends on analysing the variation of focus, it relies on the inherent micro or nano scale roughness of a surface and is therefore only applicable to surfaces where the focus varies sufficiently during the vertical scanning process. Thus, transparent objects and object with small localized roughness are hardly measureable. The lateral resolution of the system is dependent on the numerical aperture of the objective and therefore has a direct effect on quantification of surface roughness parameters. It typically provides a lateral resolution of the order of 2 µm and an axial resolution of the order of 20 nm [48].
1.2.2.4 Phase Shifting Interferometry

Another type of surface measurement technique called phase shifting interferometry (PSI) is the method of choice for measuring the form of spherical lenses. The earliest reference of PSI dates back to 1966, while the development of this technique accelerated in the 1970s [49]. PSI can be used to obtain fast 3D profiles of very smooth surfaces with a nanometer resolution. PSI typically utilises a laser beam to illuminate the object under measurement and mixes the light scattered from by the object with light scattered from a reference object. The surface topography of the object is usually measured by sequentially shifting the phase of the reference beam by known amounts and measuring the resulting interference pattern. The relative surface heights are then calculated from the fringe data by different processing steps including an unwrapping algorithm to remove the phase ambiguities. PSI provides axial resolution in the nanometer to angstrom region with a lateral resolution in micrometers. These instruments are usually limited to the measurement of smooth polished homogeneous surfaces since measurement of rough surfaces with dissimilar optical properties introduces several errors in the measurement [49- 51]. For PSI, the height difference between two adjacent data points must be less than a quarter of a wavelength, otherwise height ambiguity of multiple of half wavelengths exists.

1.2.2.5 Interferometric Microscopic Optical Profilers

The interferometric measurement of surface topography is by far the most efficient surface measurement method [35]. This type of profiler is an extension of the standard white light microscope where the microscope objective is replaced with an interferometric objective. The light reflected from the surface is made to interfere with the reference light inside the objective, and the interference pattern is recorded.
This kind of profiler usually exploits an axial scan where the optical path length is varied by either moving the object or moving the objective. As it scans through the focus a number of interferograms are recorded depending on the number of axial frames. The recorded interference pattern is analysed to obtain the surface information from the brightest fringe position [15]. This technique has been called scanning white light interferometry but is now more often referred to as coherence scanning interferometry CSI and the method is discussed in the following section.

1.3 Introduction to Coherence Scanning Interferometry (CSI)

Coherence scanning interferometry (CSI) is a 3D imaging technique which combines the vertical resolution of an interferometer with the lateral resolution of a high power microscope. The concept of the CSI was first outlined in a seminal work by Davidson et al in 1987 [52]. They demonstrated the basic working principle of a scanning interference microscope and applied it in order to improve the lateral resolution for measurement of smooth surfaces specifically for semiconductor applications. They used a Linnik interferometer and named the technique “coherence probe imaging”.

In 1990, Kino [53] presented a interferometric surface measurement system which is similar to the one discussed by Davidson et al., but used a Mirau objective instead of Linnik interferometer. Interestingly, the application of Mirau interferometry for surface measurement dates back to 1985 when Bhusan et al developed a Mirau optical profilometer based on the Leitz Mirau interferometer [54, 55]. This design has several advantages over the Linnik interferometer. It is compact, light weight and
as a single objective is used, is much more tolerant to vibration and aberrations (as a near-common path interferometer). In the same year as Kino, Lee et al published work on a coherence scanning microscope [56] where height measurement using coherence scanning was demonstrated, but utilizing Michelson interferometer. During the 1990s, similar techniques of surface measurements were reported by different researchers [57-62]. All of these techniques follow the same basic functional principles as CSI, however different terms were found to address it in the patents, technical and commercial literature. The terms coherence radar [58], coherence correlation interferometry, white light interferometry [59], scanning white light interferometry [60, 61], vertical scanning interferometry [62] have all been used to describe what is now known as coherence scanning interferometry [6]. In the following section the basic theory of CSI operation as well as the known error sources are discussed following material taken from review articles [57, 63], patents [64], good practice guides [65, 66] and ISO standard documentation [6].

1.3.1 CSI Working Principle

CSI provides the height dependent variation in fringe visibility related to optical coherence and utilises the principle of interference microscopy to measure 3D surface profiles. Figure 1.1 illustrates a basic CSI instrument.
CSI generally requires a low coherence, white light source with a broadband spectrum, such as tungsten halogen, incandescent or arc lamps or white light LEDs. Usually continuous sources are employed for the measurement, but for moving objects a pulsed light source can be used [67-68].

CSI generally records a number of frames of data in order to calculate surface heights at each detector point. The object wave-front which is reflected from the sample object interferes with the reference wave-front reflected from the reference surface. During measurement the interference signal changes as the optical path difference between the object beam and reference beam changes. This is achieved in practice either by vertical scanning the interference objective relative to the test
surface [69], or by moving the stage on which the test surface is placed using a piezoelectric transducer or motor scanner. Because a broadband and distributed light source is used interference is only observed when the object and reference paths are approximately matched. The interference can be thought of as an optical focus sensor where the position of maximum contrast in the interference signal determines the best focus position. Several types of interference objectives are available depending on the position of beam splitter and reference surface. These objectives are used for different applications depending on the allowable magnification and permissible scan length or working distance [15].

1.3.2 Interference Objectives

The three main types of Interference objectives which are used, are a) Michelson b) Mirau and c) Linnik configurations as shown in figure 1.2. All of these objectives are based on the Michelson interferometer which is essentially an equal path two beam interferometer. The difference between them arises from the position of the reference surface and beam splitter. In order to obtain high contrast fringes and successful identification of the surface, it is important that the position of the reference surface matches the position of best focus position for the objective. This is to ensure that the optical path length from the beam splitter to the focused object surface and from the beam splitter to the reference surface is the same such that the fringes are formed on the focused surface. In most Mirau objectives the position of the reference is fixed, however, as magnification increases (50x and beyond) this is adjustable to account for temperature drift. Although it is often overlooked this adjustment is critical to the performance of the objective as discussed in section 2.6.
Michelson interference objectives consist of an objective, a beam splitter and a separate reference surface. As they incorporate a beam splitter inside, these

Figure 1.2: Interference Objectives (a) Michelson Objective (b) Mirau Objective (c) Linnik Objective
objectives need to have a long working distance which restricts their usage to only low magnification operations.

The Mirau interferometer was designed by Mirau in 1952 [70]. It contains two small glass plates between the objective lens and the test surface. In one of the plates (as shown in figure 1.2 b)) which is also known as the compensating plate, contains a small reflective spot that acts as the reference surface. The position and the size of this spot are extremely important for proper working of this objective. The size of the spot should be bigger than the field of view since it is in the conjugate plane of the best focus plane of the objective. The other plate acts as a beam splitter. These interferometers are used between 10 x to 80 x magnification. This is because at magnification lower than 10x, the reference mirror blocks too much of the aperture. Due to the compact structure, Mirau objectives are the most commonly used objectives in CSI instrument. A 50 X Mirau objective was used for all the experiments described in this thesis.

A Linnik setup is consists of a beam splitter, two matched microscope objectives and a reference mirror, which makes the objective heavier and more expensive compared to the other two. In order to obtain fringes both objectives need to be adjusted properly with the beam splitter which makes the application of it difficult in practice. Potentially it can be used for almost any magnification, though Linnik systems are most often used with a high-magnification objective that has a short working distance.
This discussion on the microscope objectives shows that depending on the magnification, numerical aperture and field of view requirement, different objectives may be selected. In the next section formation of white light fringes in CSI is described.

1.3.3 Formation of White Light Fringes

Due to the wide bandwidth of the light source, individual interference fringes are produced for each of the wavelengths in the source spectrum and the white light interferogram can be considered to be the superposition of these independent fringe patterns as shown in figure 1.3. As the spacing of the fringes for each constituent wavelength of the source is different, different fringes will align only around one point where the optical path difference (OPD) is zero for each wavelength, creating the highest contrast fringe at that point. Moving away from the zero OPD point the fringe contrast drops rapidly as the optical path difference increases and the interference fringes no longer coincide, as shown in figure 1.3 b) which results in the formation of a packet of fringes. It is for this reason the fringes are said to be localized in space, denoting the surface by the highest contrast fringe. The fringe packet is usually referred to as the fringe envelope which is centred around the zero order fringe or the maximum contrast fringe corresponding to the zero OPD position. The shape and the size of the fringe envelope depends on the source bandwidth, the broader the bandwidth of the source spectrum, the narrower the width of the envelope. At low numerical apertures, the width of the fringe envelope is a measure of the coherence length of the source and is given by [15],

\[
\text{Coherence length} = \frac{\lambda^2}{\Delta\lambda}
\]

(1.1)
Figure 1.3: Formation of white light fringes (a) fringes for individual wavelengths of the source is shown in different gray scales ($\lambda_1$, $\lambda_2$, $\lambda_3$) (b) the summation of all the individual fringes create white light fringes.

The fringe formation in a typical CSI using a Mirau objective is shown in Figure 1.4 for the case of a spherical object as the objective is scanned along the $z$ direction. It creates focused fringes when the optical path difference between the light reflected back from the sphere and reference surface is zero. As it scans upward, focused
fringes are first obtained for the edges of the sphere and then for the top portion of the sphere.

Figure 1.4: Fringe formation for a sphere while the objective is scanned progressively upward. Fringes showing which part of the sphere are in focus for a given position of the scan.

A basic model of fringe formation for CSI can be derived from the well known equations of two beam interference [71]. In this case the recorded interference signal $I(k, z, \theta)$ at a particular point $(x, y)$ for a monochromatic point source of wavelength $\lambda$ can be written as,
\[ I(k, z, \theta) = I_R(k) + I_O(k) + 2\sqrt{I_R(k)I_O(k)} \cos(4\pi k (h - z) \cos(\theta) + \phi(k)) \]

where \( k = \frac{1}{\lambda} \) is the wave number, \( I_R \) is the intensity of light reflected from the reference mirror and \( I_O \) is the intensity of light reflected from the object corresponding to the point \((x, y)\). The \((x, y)\) coordinates are omitted from the equation for simpler representation. \( h \) is the local height of the object at the point \((x, y)\) with respect to the corresponding point in the reference mirror and \( \cos(\theta) \) is the direction cosine of the beam’s incident angle \( \theta \) onto the object (in this case considering the source is on axis with the object it \( \cos(\theta) \) will be 1). The final term \( \phi(k) \) represents the phase change on reflection which is introduced by the material of the object. The phase difference \( \varphi \) between the two interfering beams is represented by the term under the cosine form in equation (1.1)

\[ \varphi = 4\pi k (h - z) + \phi(k) \]  

(1.3)

The interference fringes recorded by each detector pixel can be represented in a simpler form,

\[ I(z) = I_{dc}(z)[1 + \gamma(z)\cos(\varphi)] \]

(1.4)

where, \( I_{dc} \) is the dc irradiance and represented as \( I_{dc} = I_R + I_O \). \( \gamma \) is the fringe visibility and is represented as \( \gamma = \frac{2\sqrt{I_R I_O}}{I_R + I_O} \). Fringe visibility is dependent upon the intensity of the object and reference wavefront as well as the degree of coherence and it is at a maximum when the OPD is zero. The effect of the broadband light source is incorporated in equation 1.5 [72] such that,
\[ I(k,z) = \int S(k)I_{dc}(k,z)[1 + \gamma(z)\cos(\varphi)]\,dk \]

(1.5)

where \( S(k) \) is the spectral distribution of the source. If the source has a Gaussian
distribution, the second term in the integral will be a cosinuoidal variation multiplied
by the Gaussian distribution. This is shown in figure 1.5, where the Gaussian part
acts as an envelope of the fringe and called the envelope function or the modulation
envelope.

Figure 1.5: CSI fringe and the modulation envelope
Figure 1.6: (a) A yz section through the fringe pattern for a spherical surface as shown in figure 1.4 (b)(c)(d) fringe packet through different pixel along z direction shows the position of the packet changes depending on the position of the surface.
1.3.4 Fringe Analysis

The shape of the object is determined by finding the surface from the localized fringe pattern created for each spatial point registered during a vertical scan. Since 1980 several methods have been reported for the analysis of white light interferometry [57]. Most of the methods are based on the estimation of interference phase which is used for surface evaluation based on the phase shifting interferometry data. In this section the methods relevant to CSI will be discussed. During surface measurement, it is assumed that the fringe signal is similar at each point in the lateral ($x, y$) plane as shown in figure 1.6, while its axial position ($z$) is different due to changes in the topography of the sample. Each pixel is therefore independent and the data processing techniques are used on each separately. The highest contrast fringes are obtained when the OPD is zero, and consequently the estimation of the peak of the modulation envelope provides information of the object surface (56, 58, 73, 74). Though this technique is one of the popular surface finding methods, there are other ways to determine the surface height, namely phase estimation [54, 75] and frequency domain analysis [76-78]. Many algorithms first compute the position from the modulation envelope and then adjust it using the phase estimation technique [62, 79-82]. In the following section a brief description of different surface finding methods for CSI is provided.

1) Detection of Modulation Envelope

Generation of surface topography using envelope detection is one of the most popular techniques. In most of the cases the surface is realized by calculating the peak of the modulation envelope.
It was reported by Lee et al. in 1990 where they had calculated the modulation envelope by demodulating the fringes [56]. The peak amplitude of the envelope function and corresponding z location is deduced for each pixel in the field of view, thus producing a high lateral resolution surface topography. For their technique it was shown that as in the basic theory above, the lateral and the axial resolution of the system are independent. The low spatial coherence of the illumination source contributes to the increased lateral resolution while the low temporal coherence limits the cross talk between the vertically adjacent points resulting in higher axial resolution.

The next notable publication utilising the envelope detection approach was by Dresel in the year 1992 [58]. This paper is important as it was the first to report the measurement of rough surfaces using CSI with a technique which is a combination of time of flight and interferometry. It showed that a CSI system could provide useful measurements of surfaces with sufficient roughness to generate random speckle pattern. As the phase of the interference fringes statistically varies from speckle to speckle, phase estimation could not be applied. On the other hand, due to the short coherence length interference occurs only within those speckles which correspond to surface elements with the same OPD as the reference mirror as the object is scanned in the axial direction. Thus, the surface profile is determined by measuring the occurrence of the interference for each surface point.

Another approach was described by Caber [74] in 1993, who applied techniques common to conventional communication theory to demodulate the envelope of the fringes and measure the degree of fringe modulation or
coherence instead of the phase of the fringes. In this method the signal from the CCD camera is digitally processed and filtered using high speed digital signal processing hardware which allows demodulation to be done during scanning.

2) Centroid Detection

This technique can be thought of as an extension of the envelope detection method as reported by Ai et al. in 1997 [83]. After obtaining the envelope, the position can be found by determining the peak or for a better approximation fitting a curve (least square fitting [81] or linear fitting [82]) to the envelope. According to this method, the position of the envelope is found by calculating the envelope’s centre of mass using the equation

\[
h = \frac{\sum_{i=1}^{N-1} y_{zi}}{\sum_{i=1}^{N-1} y_{zi}}
\]

(1.6)

where, \( y \) represents the envelope function, \( z \) is the axial position and \( h \) is the height that is output. This algorithm is straightforward to implement and is reported to provide a better estimate of surface height than methods using only the peak value [83].

3) Phase Estimation

Phase estimation is used widely for white light phase shifting interferometry [2, 15, 73]. The phase of the fringes not only depends on the OPD between the two arms of an interferometer but is also dependent on the complex refractive index of the object. Phase estimation was first used in a work
published by Bhusan et all in 1985 [54]. In that communication Mirau interferometry was applied in order to find the surface topography of the magnetic tapes, and the phase was measured using the method known as the integrating bucket technique. According to this technique the reference surface is moved at a constant velocity rather than in steps (as it is done for the phase stepping technique by Wyant et al [84]) which results in less vibration related errors. Different researchers employed similar techniques during the 1990s for fringe analysis in interference microscopy, namely work by Montegomery et al in 1993 where the phase stepping method was used [85]. A year after that Hariharan et al [75] had proposed a method to overcome the error generated due to non-uniform contrast which is produced by moving the reference mirror to vary optical the pathlength. These techniques had achieved high axial resolution on the order of few nanometers. Though phase estimation provides good vertical resolution, it suffers from the phase ambiguity problem. In the next section a method which combines the phase estimation and envelop detection is discussed.

4) A Combination of the Above Two Techniques

Combined envelope detection and phase estimation techniques are effective methods for surface topography measurement, but they have some drawbacks. The phase estimation which is applied for the phase shifting techniques can only be used when the phase difference between two adjacent points is less than a quarter wavelength. On the other hand, the envelope detection is sensitive to errors due to aberration, diffraction, vibration and noise. To address this a method was reported by Larkin where the envelope
detection was used only to resolve fringes and much finer resolution of the surface height is obtained by finding the interference phase [79].

In order to find the surface height in this method, first a height map \( H(x, y) \) is calculated using fringe envelope detection or other coherence based analysis. The second step is to find the interference phase map, \( \theta(x, y) \), which is able to provide the precise position corresponding to the frequency \( k_0 \). A mathematical relation is deduced, where the combined surface topography \( h(x, y) \) is given by [66]

\[
h(x, y) = \frac{\theta(x, y)}{4\pi k_0} + \frac{1}{k_0} \text{round} \left( \frac{A(x, y) - \langle A \rangle}{4\pi} \right),
\]

(1.7)

where \( A(x, y) \) is the phase gap between the two different analysis results for an individual pixel \((x, y)\) and is represented as \( A(x, y) = \theta(x, y) - 2\pi k_0 H(x, y) \). \( \langle A \rangle \) is the field average of the phase gap \( A(x, y) \). and the function \text{round}() denotes the nearest integer value.

Computing the surface profile with the combination of these two methods has been reported for white light interferometry (PSI) by Larkin at 1996 [79]. In that publication the coherence envelope peak was detected first and this was followed by finding the phase using a nonlinear five point algorithm. In the following year a similar surface measurement technique was published by Sandoz [80, 81].

Surface measurement using this technique leads to erroneous measurements known as fringe order errors. This error mostly appears for surfaces having
different optical properties along the field of view and sharp edges which cause incorrect identification of the fringe order [86]. Haraski et al. used a linear fitting algorithm in order to find the peak of the envelope and used a five point algorithm to determine the phase [82]. Another detailed analysis of improving the performance was reported by De Groot based on techniques to determine fringe order by introducing frequency domain analysis [76-78]. This new technique became widely used as a surface profiling technique by providing a better alternative to the envelope detection algorithms.

5) Frequency Domain Analysis (FDA)

In 1994 De Groot and Deck showed that the position of the surface can be determined by processing the fringe in the frequency domain [77]. This process starts with Fourier transform (in this case a digital fast Fourier transform FFT) of the fringe pattern as reported by Kino et al (in 1990 [53]). The magnitude of the transform represents the strength of the spectrum at a given wavelength, and the phase represents the phase of the interference signal for that wavelength [76-78].

In FDA an FFT of the interference function $I(z)$ is calculated and the high spatial frequency content of the signal is analysed, such that

$$I(k) = \int_{-\infty}^{+\infty} I(z) \exp(-i2\pi k z) \, dz$$  \hspace{1cm} (1.8)

where the phase is given by

$$\theta(k) = \tan^{-1}\left(\frac{Im\{I(k)\}}{Re\{I(k)\}}\right)$$  \hspace{1cm} (1.9)
Here \( \theta(k) \) is the phase value calculated and \( \text{Im}\{I(k)\} \) and \( \text{Re}\{I(k)\} \) are the imaginary and real values of the Fourier transform of the interference function.

The phase can be described as a linear function of spatial frequency \( k \). The phase of the Fourier transform of the intensity is plotted against the range of white light frequencies in figure 1.7. The primary estimate of the surface height \( h(x,y) \) corresponding to a particular point \( (x,y) \) can be calculated from the gradient of the line of best fit from the plot. The phase gap \( A(x,y) \), as described in equation (1.10), can be deduced simply from the axis intercept of the best fit line. The phase is, therefore,

\[
\theta(x,y) = A(x,y) + 2\pi kh(x,y)
\]

(1.10)

This approach measures surface topography without relying on fringe contrast. A scan along the axial direction over a range of OPD values results
in a digital representation of the interference function for each image point in the field of view of the instrument. The Fourier transform of each individual interferogram results in a sequence of phase values which is used to determine the local surface height.

After this discussion of different fringe analysis techniques it is worth noting that the real fringe data is likely to deviate from the ideal form for reasons including; aberrations and dispersion introduced by the objective lens, surface tilt, roughness and multiple scattering effects, and the noise introduced during the measurement process, all of which will result in measurement errors [86].

1.3.5 Application of CSI

The primary use of CSI reported has been for testing of smooth surfaces where it was mainly used for semiconductor applications [52] which were later extended to characterization and measurement of different high aspect ratio structures for MEMS and MOEMS devices [53-55, 85, 87-89]. Along with the measurement of smooth surfaces several researchers have applied it to the measurement of rough surfaces when it was recognized that CSI could provide useful measurements for surfaces with sufficient roughness to generate completely random speckle [58, 90].

In recent years it has also been used for thickness measurements. The measurement of films [91, 92] of more than 2 μm thickness results in formation of two localized fringe packets corresponding to the top and bottom surfaces. This application is been extended to measurement of thin films by Mansfield et al where a helical conjugate
function (HCF) was introduced to obtain thickness information for the films between 20 nm to 2 µm [93, 94]. Another technique of thickness evaluation was reported by Lega et al. in 2008 [95] which employs pupil plane scanning white light interferometry method (PUPS).

1.3.6 Surface Measurement Errors

Despite these significant advantages, CSI is also known to be prone to certain measurement errors. In 1990 Hillmann [96] reported that the results obtained using optical methods for measurement of a roughness standard differed from the results obtained from the stylus instrument. Later several other researchers have published documents regarding CSI errors [62, 82, 97, 98] within a decade, namely: fringe order errors, ghost steps, bat wing effects, slope dependent errors, material effects and multiple scattering effects. The majority of the errors reported in literature have been observed when the surface gradient is large compare to the numerical aperture of the objective. However, there are other different reasons including aberrations and dispersion introduced by the objective lens, surface tilt, roughness and multiple scattering effects, and the noise introduced during the measurement process [86, 99]. The following describes the errors observed.

1) Fringe Order errors

Fringe order errors cause sudden jumps of about half the mean wavelength in the measured surface topography and are known as fringe order or $2\pi$ errors. In the previous discussion it was observed that the measurement of surface topography is dependent on the phase measurements. However, the inherent cyclic nature of the phase measurement often causes a fringe order ambiguity or a phase error of $2\pi$. 
During surface extraction from the fringe data, it is not possible to determine if an optical phase jump in the data is due to a surface feature (such as, step, sharp edge, high surface roughness, different surface materials), or to the wrong identification of the brightest fringe [100]. It is reported that fringe order errors are often field dependent and the presence of them generally increases towards the edge of the field of view [101].

2) Ghost Steps

A ghost step error is a type of fringe order error which introduces a step corresponding to a phase jump of $2\pi$ while measuring a perfectly flat object for some of the CSI instruments [102]. It results in a surface error of half the mean wavelength. Field dependent dispersion, which is inherent in Mirau and Linnik type interferometers, can be a possible reason for it [98, 100].

3) Batwing Effect

The batwing effect is an error that often appears when measuring a step discontinuity which is less than coherence length of the light source [62, 78]. As the name suggests the error looks like a bat wing or ringing at the surface discontinuity. It is usually caused by the interference between the reflections of the waves normally incident on the top and bottom surfaces following diffraction from an edge [86].

4) Multiple Scattering

Multiple scattering is caused during the measurement of rough surfaces in a CSI which results in overestimation of the surface roughness. The effect of multiple scattering is discussed by Gao et al [99] for measurement of a silicon V Groove (with smooth walls and an internal angle of 70.52 degree) where the incident light is
reflected and scattered multiple times. As shown later, in CSI weak scattering is assumed such that multiple scattering is neglected. However for the V groove it was shown that due to multiple scattering, a clear peak with its apex at the bottom of the profile is observed. A detailed theoretical analysis of multiple scattering from a V groove and other high-aspect-ratio structures is reported by Lobera et al [103]. This theory can be used to explain the over-estimated surface roughness measurements reported by Hillmann [96].

5) Dispersive Effects in Dissimilar Materials.

The optical properties of the materials that make up the object surface are integral to the CSI measurement process because different materials exhibit different phase changes on reflection which will affect the surface height measurements [98, 104]. For an example, metallic surfaces will introduce errors due to their complex refractive index which influences the phase of the fringes within the modulation envelope. If the surface is made out of a single material, the entire surface will be shifted in the vertical direction, keeping the relative surface topography same. However, the problem will appear when dissimilar materials are measured as each material will cause a different phase shift. A comparative measurement of an object with a chrome line deposited on a glass surface has been reported in order to demonstrate the effect [66]. In this case it was shown that the tactile instrument gives a step height of 60.8 nm while the CSI measurement measured 37.3 nm.

1.3.7 A Demonstration of CSI Problems

It was mentioned while describing the formation of white light fringes in CSI that, each pixel is independent. In practice neighboring pixels cannot be viewed as
independent interferometers. During fringe formation their responses are coupled, which results in a significant change in measurement when the surface gradient changes quickly or in the presence of a discontinuity such as a step.

Examples of errors in CSI measurement are shown in figure 1.7. These errors were observed while measuring a sinusoidal grating of 8 µm pitch and peak to peak amplitude of 466 nm [86]. In the picture the dotted line represents the original profile and the continuous line indicates the measured profile. It is clear that the measured profile only follows the actual profile at the top and bottom portion of the sinusoid. As the gradient increases, $2\pi$ errors are observed, which results in a measurement of peak to peak amplitude of 1 µm.

Figure 1.7: CSI measurement errors (reproduced with permission from [99])

1.4 Research Novelty and Chapter Summary

The main objective of the project described in this thesis was to implement a calibration and adjustment protocol for CSI. The problems associated with CSI when measuring tilted samples are well documented in the literature [87], however currently there is no calibration procedure available to provide tilt dependent

47
measurement uncertainty. In this thesis, this problem is considered from first principles and a method of calibration and adjustment is proposed. To begin with the CSI measurement process is described in terms of linear systems theory where the filter characteristic is expressed as a 3D point spread function (PSF). In contrast with the 2D PSF which is frequently used to characterise the lateral resolution of optical systems this new approach defines for the first time both the axial and lateral resolution and explains the response of the system to tilted surfaces.

This chapter has given a general overview of different surface measurement techniques which was followed by a detailed discussion of CSI. In the next chapter the calibration process that is currently used for CSI is discussed in practice, and some important parameters such as the effect of different operating modes and focusing of the reference mirror in the objective (Mirau) will be examined.

The third chapter discusses the linear theory of CSI and defines the point spread function (PSF) and transfer function (TF) assuming a weakly scattering object volume. In this case the assumption of weak scattering is observed by the application of the Born approximation. The theory is then extended to consider surface scattering. In this case the 3D object is replaced mathematically by an infinitely thin foil-like layer placed at the object boundary and the PSF and TF of the CSI are changed accordingly. We call this model the “foil model” of the surface.

In the fourth chapter the PSF of a CSI is measured in practice according to the foil model of a spherical artefact. Comparison of the measured PSF with a calculated PSF for an ideal CSI reveals that there are some slow but significant variations in the imaginary part of the PSF that will lead to measurement errors. It is important to mention that there are two CSI instruments which are used in this project, i. e. the
Zygo New View 5000 and the Taylor Hobson CCI. Due to ease of accessibility most of the experiments were performed first using the Zygo instrument and then later with the Taylor Hobson CCI. As both the instruments operate on the same principle, results corresponding to one (in this case Zygo) are displayed primarily in different chapters. However, for situations where the instruments produced different responses, results from Taylor and Hobson CCI are shown and analysed as well.

Having measured the PSF and TF responses that characterise the measurement error, an adjustment or compensation method is proposed in the fifth chapter. It is shown that a filter based on modified inverse filtering can compensate for the systematic errors. Application of the modified inverse filter also increases the contrast of the fringes and makes the fringe envelope more compact which helps the identification of the surface.

Another problem that is addressed in the final chapter of the thesis is distortion. It was found that the application of the modified inverse filter compensates the errors for the Zygo instrument which has a field of view of around 140 µm by 105 µm and the filter is found to provide shift invariant compensation. In contrast the Taylor Hobson CCI has a field of view which is three times the field of view of Zygo and exhibits some distortion In this case the compensation is not shift invariant. Although some methods have been proposed to compensate for distortion they do not completely remove the effect and cannot be used with the inverse filtering method applied here. Finally a method based on a dewarping algorithm is proposed to remove the distortion at source. Using this approach shift invariant compensation can be used to reduce systematic measurement errors to the order of a few nanometres.
References


95. De Lega, X. Colonna, and Peter J. de Groot. “Characterization of materials and film stacks for accurate surface topography measurement using a white-
light optical profiler.” Photonics Europe. International Society for Optics and Photonics, 2008.


Chapter 2: Traditional Calibration of Coherence Scanning Interferometry

2.1 Introduction

The purpose of calibration for any measuring instrument is to define measurement uncertainty, traceability, repeatability and reproducibility [1]. This allows the user to make a judgment of the significance of their measured data. The most common method of calibration is to measure an artefact that is geometrically similar to the object under study and has previously been measured by another calibrated instrument. A traceable calibration provides a statement of uncertainty in terms of the primary standard; in this case, it is the standard unit of length – the meter [1, 2].

At present the calibration procedure recommended by the manufacturers of CSI instrumentation can be used to verify the lateral and vertical scales of the instrument [1- 5]. This is achieved by measuring pre-calibrated rectilinear artefacts and subsequently by comparing the measurement results to those specified in the calibration certificate. Before describing this procedure, however, it is noted briefly that this traditional approach to calibration is limited as it is restricted to rectilinear artefacts such as step and grid plates. Measurements taken from these artefacts result from surfaces which have a normal that is aligned with or close to the optical axis of the instrument and consequently there is no information concerning the response of the instrument to surface tilt. As will be described later, one variable that has a
significant effect on the instrument performance when applied to tilted surfaces is the focus of the reference flat within the Mirau objective. The traditional calibration process does not reveal focussing error.

A detailed calibration procedure of different surface measurement instruments according to the ISO specification standards is described in the Good Practice Guides published by National Physical Laboratory (NPL) [1, 2]. This chapter illustrates the recommended calibration procedure applied to a commercial CSI instrument – the Zygo New View 5000. Calibration of the axial scale is described, followed by the lateral scales. The effect of magnification, camera resolution and aliasing are then discussed.

### 2.2 CSI Calibration of Axial Scales

The calibration of the axial scale of a CSI instrument is generally performed by measuring a calibrated step height. The step height comes with a certificate of calibration which specifies the measurement temperature and humidity (e.g. temperature $20\pm1^\circ\text{C}$ and humidity $43\pm2\%$). It is important that these conditions are maintained throughout the measurement procedure to achieve specified uncertainty.

#### 2.2.1 Artefact Details

Step height artefacts which are used for calibration are normally manufactured from quartz or silicon wafers with a positive step (etched step) or negative step (etched trench) [1]. It is preferable to have the artefact surface made from single material to
avoid a differential phase change on reflection which will corrupt the calibration process [6]. A wide range of different step heights should be used to calibrate the instrument to ensure the performance of the system for different height objects. From a user’s point of view, a calibrated step height comes with the instrument, and it must be measured before starting any experiments. The results are then compared with the results provided in the calibration certificate to ensure the performance quality of the instrument.

The artefact shipped with the Zygo New View 5000 is a chrome coated step etched on a quartz wafer. The step has been calibrated and certified by VLSI standards incorporated. In this case the certified value of the step height is 1.844 µm ± 0.011 µm at the temperature 20 ±1°C and humidity 43 ± 2%.

2.2.2 Calibration Procedure

Typically the step height is measured before and after each measurement cycle. To illustrate the process the step height was measured here using the 50X objective with 1X zoom condition. This objective has an adjustment collar that can be used to set the focus of the reference flat within the objective. In this case the step was placed on the measurement stage and the Mirau objective was focused on the upper surface of the step edge. The collar was adjusted until fringes appeared at the step edge. The x and y tilt adjustment screws were adjusted to minimise the tilt, to ensure only one fringe was on the step. The step height was measured according to ISO 5436-1 [7] using the CSI measurement software. The results are shown in figure 2.1.
2.2.3 Results

Figure 2.1 shows the 3D plot of the measured step surface. In figure 2.2, the measurement reveals the value of the step height to be 1.85 µm, which is within the tolerance range mentioned in the calibration certificate. This process was repeated five times to ensure reproducibility of the value. The measured step height was found to be 1.847 µm ± 0.013 µm. This is well within the specified tolerance limit with an average error of 0.16%.

Figure 2.1: 3D view of the axial calibration artefact
2.2.4 Problems in Axial Calibration

Although the above method can be used to verify the vertical scale it does not provide a proper estimate of measurement uncertainty. This is because it is merely a single measurement that might have been taken in the sweet spot of the instrument. A well-known problem with CSI is nonlinearity of z scale due to the piezo controlled movement of the z stage [8]. When the instrument is built the piezo element is calibrated using interferometric techniques [9] and a look-up table is produced. For the user, calibration of the full axial scale requires a large range of calibrated step heights which in this case is not provided by the instrument manufacturer. Alternatively, Giusca et al have suggested a way of calibrating the total scan range with the available step height by measuring it multiple times inside the scan range such that the calibrated scan range overlaps with each other. [5] This process was not undertaken however, the effects of z scale nonlinearity with this instrument were apparent in the experiments which are discussed later in chapter 5.
For the remainder of the experiments conducted using this instrument, it was ensured that all the measurements are done in same range of the z axis (i.e. performed in the sweet spot), which will make all the results comparable.

2.3 CSI calibration of lateral scales

The purpose of lateral calibration is to determine the uncertainty of measurements for the lateral scales of CSI and measure the pitch of a pre-calibrated square grid pattern along x and y direction. Similar to the step height artefacts these artefacts also come with calibration certificate specifying the calibrated pitch values.

2.3.1 Artefact Detail

Applications of two different types of artefacts are reported for the lateral calibration of CSI [1, 2]. They are typically (a) square grid patterns and (b) concentric circular discs etched on silicon wafer. In this experiment an artefact with a square grid pattern was chosen. The artefact consists of square grid patterns of different pitches made out of a platinum coated silicon dioxide layer which has been etched on a silicon wafer. This is a type of positive step (etched step) [1, 2]; while the application of a negative step (etched trench) is also reported in literature [5]. The artefact has square grids of pitch starting from 3 µm to 500 µm. In this context, to calibrate the CSI with a field of view of 140 µm by 105 µm, the square grid with a pitch of 10 µm (9.996 µm ± 0.019 µm at temperature 20 ± 1°C and humidity 56 ± 2 %) was selected.
2.3.2 Results

Figure 2.3: 3D view of the later calibration artefact
Figure 2.3 shows the 3D measurement of the square grid patterns of 10 micron pitch. Figure 2.4 shows the measurement of the pitch. The pitch measured here is 10 micron. This process was repeated five times to ensure the reproducibility of the results giving the pitch as 9.995±0.018 µm. The results show that the measurement is within the defined calibrated uncertainty range of the artefact.

2.3.3 Problems

Although, the results show that the measurements are within the uncertainty limit provided in the calibration certificate, the calibration of lateral scales is also dependent on the size of the field of view of the instrument. As the field of view
increases, the effect of field dependent aberration becomes dominant. In chapter 6 the effect of field dependent aberration is discussed for the Taylor and Hobson CCI which has a field of view almost three times larger than the Zygo instrument. It is shown in that chapter that field dependent aberration: distortion introduces a field dependent magnification in the measurement, which will result in the wrong measurement of the object features, around the edges of the field [10]. Along with that, it will also introduce, tilt dependent measurement errors. The effects of these errors and their resolution are also discussed in that chapter.

In addition to the traditional calibration it is also useful at this time to discuss some other characteristics of CSI. In particular the lateral resolution or cut-off frequency is often specified as follows.

2.4 Cut-off Frequencies in CSI

The cut-off frequency of CSI is an important factor in defining the performance of the system as it determines the instruments ability to detect the higher spatial frequency present in the object. There are two types of cut-off frequency that can be used to specify the system performance i.e. optical cut-off frequency and instrument cut-off frequency.

2.4.1 Optical Cut-off Frequency

Optical cut-off frequency defines the maximum spatial frequency that can be resolved by the optical system. It is usually limited by the numerical aperture of the objective lens which together with the properties of the source determines the lateral extent of the transfer function in frequency domain or the lateral bandwidth of the
instrument [11] as reported by Coupland et al. (A detailed 3D mathematical analysis for CSI transfer function is discussed later in chapter 3.)

2.4.2 Instrument Cut-off Frequency

Instrument cut-off frequency is the maximum limit of the spatial frequency that the camera can resolve. It is directly related to the field of view of the instrument and determined by pixel size and pixel number of the camera. It is important that the instrument cut-off frequency is the same or more than the optical cut-off frequency in order to resolve the highest frequency present in the system response or all the optical frequencies present in the system.

2.4.3 Calculation of Optical and Instrument Cut-off Frequencies

In this thesis, the experiments were performed using two different types of CSI instrument. The following section describes the calculation of the optical and instrument cut-off frequency of both the instruments.

2.4.4 Calculation of Cut-off Frequencies: Zygo Instrument

The first instrument is the Zygo New View 5000 series CSI. The calculation of the optical and instrument cut-off frequency for that system is shown below

To calculate the optical cut off, the necessary information is wave number and numerical aperture of the system [11]. It can be mathematically defined as,

\[
f_{optical} = 2k_{max} NA
\]

(2.1)
where $k_{\text{max}} = 1/\lambda_{\text{min}}$. Considering the mean operating wavelength as 0.6 µm and bandwidth of 0.12 µm, the optical frequency limit calculated for an objective with NA 0.55 is 2.03 cycles/µm.

On the other hand, the instrument cut-off frequency ($f_{\text{inst}}$) is independent of the NA and operating wavelength of the instrument; instead it depends on the pixel magnification and pixel number of the CCD camera. Due to this reason it is possible to have a different instrument cut-off frequency even when two instruments use the same NA objective. It is defined as

$$f_{\text{inst}} = \frac{\text{total number of pixels}}{\text{total field size} \times 2}$$

(2.2)

The factor of 2 in the denominator arises because according to Nyquist theory, at least 2 pixels are required in order to record a fringe successfully.

For the Zygo New View 5000, the total number of pixels in $x$ direction is 640 and the size of one pixel is 10.948 µm. While imaging with an objective with 50x magnification the length of the field is 140.13 µm in $x$ direction. This calculates the instrument cut-off as 2.29 cycles/µm. As the pixels are square, it is the same in the $y$ direction as well. It shows that in this case, the instrument is able to resolve all the spatial frequencies present in the optical signal.

There is another Mirau objective of magnification 10x which is supplied with the Zygo New View 5000. In a similar way, the cut-off frequencies are calculated for this objective of NA = 0.3 using equation (2.1) and (2.2). The optical cut-off frequency is found out to be 1.11 cycles/µm and instrument cut-off frequency is
0.454 cycles/µm. The result shows that with the 10x objective, the instrument is not able to resolve all the optical frequencies which will result in aliasing (as discussed in section 2.5).

2.4.5 Calculation of Cut-off Frequencies: TH Instrument

The second instrument used in this work is the Taylor and Hobson CCI which has a field of view of 330 µm by 330 µm. It can be operated in different working modes. The most commonly used modes for measurements are 4M and XY mode. The measurements for 4M mode with, 50x Mirau objective and 1 X zoom condition will be discussed first.

The optical cut-off is the same as the Zygo New View 5000, as it only depends on the effective wavelength of the source and the NA of the objective. However, the instrument cut-off frequency will be different due to different camera dimensions. The pixel size in this case is 0.164 µm, which results in the instrument cut-off frequency to be 3.05 cycles/µm. The results indicate that TH instrument operating in 4M mode with 50x objective can resolve all the optical spatial frequencies present.

In contrast if the operating mode is changed from 4M to XY mode, the number of pixels in the field is reduced by half, from 2048 to 1024. This will effectively reduce the instrument cut off frequency by a factor of 2. As a result the fringes will not be resolvable.

If the objective is changed to a 20x Mirau objective (NA 0.4), the optical cut-off frequency becomes 1.48 cycles/µm. The instrument cut-off frequency is calculated as 1.22 cycles/µm, which means that the system is not able to resolve higher optical frequencies.
2.5 Effects of Aliasing

So far the optical and instrument cut-off frequencies were calculated for different objectives and operating modes for both Zygo New View 5000 and Taylor Hobson CCI instruments. It is important to discuss the actual physical significance of the cut-off frequencies and why it is important in measurement using CSI and in calibration.

When the instrument cut-off frequency is less than the optical cut-off a CSI will not able to resolve the higher spatial frequency present in the optical field. The most noticeable error will be for an object with a variable grating frequency such as the resolution artefact proposed by NPL [12]. As the frequency of the object feature increases and goes over the cut-off frequency limit, corresponding features will be unresolvable.

The problem gives rise to aliasing which is common in communication theory [13]. In this case it can be shown that at least two samples (i.e. pixels) are necessary to resolve each cycle of a periodic component. This is the Nyquist frequency [14]. For the measurement of a flat surface using a CSI, any mode can potentially be used, as in frequency domain; the information will be concentrated around the centre. The problem will arise during the measurement of a tilted object (or object having all different tilts like a sphere). When the spatial frequency of the fringes corresponding to a particular tilt is more than the instrument cut-off frequency, those frequencies will be wrapped around the centre and in space domain it will change the direction of the fringe pattern as shown in figure 2.5.

In figure 2.5 the problem of aliasing is shown from measurement of a sphere of around 300 µm using Zygo instrument with the 10x objective. A xz section of the
recorded fringe pattern is shown in figure 2.5 (a). Figure 2.5 (b), a zoomed section of figure 2.5(a) is shown where the un-aliased fringes are present. As the tilt angle increases the direction of the fringe changes due to aliasing as shown in figure 2.5(c). Interestingly, the pixel-by-pixel method to deduce the position of the sample surface is unaffected by the aliasing. It is only during advanced phase unwrapping methods or the calibration methods discussed later in chapter 4, that aliasing becomes a problem.
2.6 Focusing

It is important to focus the microscope objective in order to see a nice and crisp image that is achieved when the focal plane of the objective and the object plane
coincide. In case of the interferometric objective the situation is different, as along with an image of the object, the fringe pattern on the object is also recorded. This requires the fringe pattern to be formed on the object surface which is in focus with the objective. On the other hand the optical path length of the light reflected from the object surface and from the reference mirror should be the same to obtain the interference fringes. In order to obtain fringes at best focus, the position of the reference mirror needs to be set at the best focus position relative to the objective. In figure 2.6 it is described for a Mirau objective where the virtual mirror position is shown by a dotted line. Figure 2.6 (a) shows that the objective is focused on the object and the reference surface is placed at the focus of the objective in either direction which makes both the object and reference optical path lengths the same. The virtual mirror position matches with the object surface which results in the fringes being formed on the focused image. If the reference mirror is not in focus with the objective, the fringes will not be formed on the focused object plane as shown in figure 2.6 (b). Generally interferometric objectives are designed and adjusted by the manufacturer to provide fringes on the focused surfaces. However, it is different in some higher magnification objectives where a collar is included with the objective to focus the reference surface.
The effect of defocus on the CSI measurement is reported by Petzing et al in NPL Good Practice Guide [1] and is reproduced in figure 2.7. Here the measurement of 0.3 µm ruby using 50x NA=0.55 objective for five different stages of focus shows that defocusing progressively gives rise to $2\pi$ errors at greater slope angles.

Figure 2.6: Focusing in Mirau objectives (a) focused (b) defocused
Figure 2.7: In focus and out of focus surface data and profile (reproduced with permission from [1])
2.7 Conclusion

In this chapter the basic calibration of CSI, which is performed by measuring pre-calibrated step height artefact for axial scale and a square grid pattern for the lateral scales, was demonstrated. The effect of some other parameters such as instrument cut-off frequency and defocusing, were also highlighted. The cut-off frequency defines the instrument’s ability to detect higher spatial frequencies. Optical cut-off frequency determines the maximum spatial frequency that could be generated by the instrument, which is dependent on the numerical aperture of the objective and the operating wavelength. In contrast instrument cut-off frequency depends on the pixel dimension and field of view of the camera. A limited instrument spatial frequency will restrict the ability to detect higher spatial frequencies as well as limit the maximum slope which can be measured using the system. As the CSI records fringes along with the image of the object, for successful detection of the surface, it is important for the fringes to coincide with the focused image. Although most of the interferometric objectives are designed by the manufacturer to satisfy this condition, for high magnification objectives, users need to adjust it.

Calibration of the CSI is one of the objective of this thesis which is achieved by the calibration of machine axis or xy and z axis. The calibration process described in the section 2.2 and 2.3 can provide uncertainty information regarding axial and lateral scales, however, as the system has not been calibrated using tilted artefacts, the tilt related uncertainty was not available. In order to better characterise the CSI, a linear theory of 3D imaging has been developed [10]. In work described in the next chapter, the linear theory has been extended to cover a comprehensive model of surface scattering. For reasons that will become apparent we call this the “foil model
of surface scattering”. In subsequent chapters, the 3D characterisation of CSI instrumentation is explained and a calibration and adjustment procedure is proposed.

References


Chapter 3: Coherence Scanning

Interferometry: Linear Theory of Surface Measurement

3.1 Introduction

In the last chapter the current calibration process of a commercial CSI instrument was described. It was discussed that the current calibration process is not able to estimate the tilt related problems, as the calibration is done only for axial and lateral scales. In this chapter the calibration and adjustment procedures, that are central to this thesis and rest on a linear theory of surface measurement, are developed.

The application of linear systems theory was first reported in the field of electrical engineering [1, 2] where the output of the system can be characterised as a weighted sum of the input variables. This theory became widely popular in the field of communication theory, due to the simple relationship between the input and output parameters that is described in terms of impulse response in time domain or equivalently the frequency response in the frequency domain [2, 3].

Mathematically the output $O(r)$ of a linear system can be represented by a convolution of the input $I(r)$ with the impulse response $H(r)$ such that

$$O(r) = \int I(r')H(r - r')dr'$$

(3.1)
Taking the Fourier transform the same operation can be written in the frequency domain as

\[ O(\mathbf{k}) = I(\mathbf{k}) \cdot H(\mathbf{k}) \]  

(3.2)

where, \( O(\mathbf{k}) \) is the Fourier transform of the output function, \( I(\mathbf{k}) \) is the Fourier transform of the input function and \( H(\mathbf{k}) \) is the Fourier transform of the impulse response and represents the frequency response of the system.

The first representation of linear theory in optical systems was initiated with the work of Abbe [4, 5] and Rayleigh [6] during the period between 1870 and 1900. Abbe demonstrated the concept of spatial frequencies to describe image formation in microscopic systems in the early 1870s. Though he did not use the term spatial frequency at that time, his work built the foundations for the application of linear theory in optics.

The first fully formulated theory of the optical transfer function was reported by French scientist Duffieux in 1935 [7]. In his publications he represented the transfer function as a “transmission factor” representing the image formation in a 2D incoherent imaging system. He also showed that the Abbe theory of image formation could be represented in terms of Fourier analysis. Later, in 1946 he published a book [8, English version 9] in which he applied Fourier methods to optics to describe imaging using coherently and incoherently illuminated objects. By applying the convolution theorem he showed that the Fourier transform of the intensity distribution of an image can be closely approximated by the product of the Fourier transform of the object distribution and Fourier transform of a point image. It is
noted that for an optical system the impulse response is represented as the point spread function (PSF) which is the image of the point object by that system. The Fourier transform of it is called the transfer function (TF).

Following from the work of Duffieux, linear systems theory has been applied extensively to the characterization of optical imaging systems. Perhaps most notably was the work of Otto Schade in the 1950s [10] who utilised this linear theory to characterise and improve the performance of TV camera lenses [11, 12]. During the same time H.H. Hopkins used linear theory to represent the quality of optical systems and derived the relationship between the linear theory with the common aberrations present in the optical system, which he published in the book *Wave Theory of Aberrations* [13]. During the 1950s Hopkins published several papers where he discussed the effects of different design parameters on the transfer characteristics (PSF and TF) [14 - 17]. It was shown that the image of a point source made by a diffraction limited coherent imaging system takes the form of the Fourier transform of the pupil function [18]. The theory discussed so far is the early work on two dimensional (2D) analysis of the PSF and the TF which can be found in detail in the books by Goodman [19] and Williams and Becklund [7]. For the work discussed in this thesis a 3D representation of the CSI in terms of 3D linear theory is necessary.

In 1969 Wolf [20] was the first to show the existence of a linear relationship between the physical characteristics of a 3D object and the measured response of a holographic system. He demonstrated that the plane wave component of the illumination and scattered field are connected to 3D spatial frequencies with object function which is defined using the refractive index contrast. Shortly after Wolf in 1970, Dandliker and Weiss [21] computed the 3D distribution of refractive index contrast (which defines the object), by determining the complex amplitude
distribution of the scattered field from holographic reconstruction. It was shown that in holography under certain circumstances, 3D image formation can be represented as a 3D linear shift invariant filtering operation. Both the approaches demonstrated by Wolf and Dandlíker, are for a holographic system which is generally a coherent imaging system. The 3D analysis of an incoherent microscopic system was reported by Striebl in 1985 who demonstrated the relation between input and output of a microscope by means of a 3D transfer function [22]. His work can be thought of as an extended version of the work on transfer theory by Frieden [23]. Frieden had developed a transfer function theory for 3D image formation by applying the concepts of the 2D TF derived by Duffieux [9]. In a similar way Fercher [24] extended this approach to describe coherence tomography, where recording has been done in multiple wavelengths while Sheppard applied similar theory to describe the image formation for confocal microscopes [25-30]. Shortly afterwards, during the 1990s he built on this work by deriving transfer characteristics to represent the performance of systems with high numerical aperture [31-33], and extended this technique to different aperture configurations [34-37]. Later a holistic overview of the 3D transfer characteristics of different 3D optical systems was reported by Coupland and Lobera [38]. This work highlighted for the first time the differences and similarities offered by a range of optical techniques including holography, tomography, confocal microscopy and CSI.

As will be described later, for the case of surface measurement instruments linear theory takes a slightly different form in which the surface is described by an infinitely thin membrane that covers the surface [39]. It will be referred as a “foil model” of the surface [39-43].
By way of comparison with tactile instruments the PSF can be thought of as a virtual “probe tip” whose physical dimensions define the resolution of the system in a similar manner to the radius of the ball end of a tactile probe [44]. It is noted however that the geometric centre of a spherical tactile probe follows a line that is always displaced from the true surface by a distance equal to its radius as shown in figure 3.1a). The continuous line in figure 3.1 a) is the actual surface and the dashed line represent the measured surface. In contrast the PSF of an optical measuring instrument is essentially the “blur” function that degrades the 3D image of the foil model [39] that follows the exact contours of the surface as shown in figure 3.1b).

![Figure 3.2: Representation of surface measurement in terms of (a) tactile measurement (b) PSF](image-url)
In the following chapter the linear theory of surface measurement is explained. Starting from the Helmholtz equation it is shown that light propagation can be described as a 3D filtering operation applied to the source distribution; the filter used for this operation is defined by the free space Green’s function. Furthermore an instrument that records and reconstructs an image of the source distribution effectively applies further filtering with characteristics defined by the numerical aperture of the system. It is important to note however, that the source distribution is a linear function of the parameter that defines the object only for the case of weak scattering. Hence the system is only linear in terms of the object parameter (in this case the refractive index distribution) if the object causes a small perturbation to the illuminating field or in other words the illumination field is essentially the same in the presence or absence of the object. This is the well-known Born approximation and is rather restrictive for volume objects as it can only be applied for small changes in refractive index or small objects such as particles suspended in fluid [44]. It is not generally applicable to the comparatively large changes in refractive index that are typical of 3D scattering objects. For the case of surface scattering, however, it is far less restrictive and large changes in refractive index can be accommodated provided the effects of multiple scattering are negligible.

3.2 Theory

The analysis of CSI presented in this section starts from scalar diffraction theory. When a wave is propagating in a linear, isotropic, homogeneous medium, a scalar approximation of rigorous vector diffraction theory can be made. Scalar diffraction is
simpler yet retains important features so that an approximation to three dimensional (3D) imaging techniques can be made. In order to apply scalar diffraction theory the medium should be linear. This implies that if the wave propagating in the medium can be decomposed into several components, after propagation the wave will be the same as the superposition of the effect of propagation for all individual wave components. In addition the medium should be isotropic implying that its properties are independent of the polarization and homogeneous implying that the permittivity or refractive index is constant throughout the region of propagation. [19].

According to scalar diffraction theory a monochromatic electric field propagating in a medium of complex refractive index obeys the Helmholtz equation.

$$\nabla^2 E(r) + 4\pi^2 k_0^2 n^2(r)E(r) = 0,$$

(3.3)

where $k_0$ is the free space wave number, $k_0 = \frac{1}{\lambda}$ where $\lambda$ is the free space wavelength and $n(r)$ is the refractive index.

The electric field $E(r)$ can be written as a superposition of the illuminating field $E_r(r)$ (that will be present in the absence of the object) and the scattered field $E_s(r)$ (the additional field that is observed when the object is present) such that

$$E(r) = E_r(r) + E_s(r)$$

(3.4)

For the case of monochromatic systems it is usual to represent the object by its scattering potential [45, 46]. However, since polychromatic optical instruments are
of interest here the object is defined as the refractive index contrast \( \Delta(\mathbf{r}) \) given by [38, 39].

\[
\Delta(\mathbf{r}) = 4\pi^2 \left( 1 - n^2(\mathbf{r}) \right).
\]  

(3.5)

Using these substitutions the Helmholtz equation can be written,

\[
(\nabla^2 + 4\pi^2 k_0^2)E_s(\mathbf{r}) = k_0^2 \Delta(\mathbf{r})(E_r(\mathbf{r}) + E_s(\mathbf{r}))
\]

\[= U(\mathbf{r}).\]

(3.6)

This is the free space scalar wave equation with source terms \( U(\mathbf{r}) \). This differential equation can be represented in integral form using the Green’s function of the Helmholtz’s operator \( G(\mathbf{r} - \mathbf{r'}) \), defined by [22]

\[
(\nabla^2 + 4\pi^2 k_0^2)G(\mathbf{r} - \mathbf{r'}) = \delta(\mathbf{r} - \mathbf{r'}),
\]

(3.7)

where \( \delta(\mathbf{r} - \mathbf{r'}) \) is a 3D Dirac delta function. It can be shown that the Green’s function that here represents an outgoing spherical wave propagating in free space is given by

\[
G(\mathbf{r} - \mathbf{r'}) = \frac{e^{2\pi j k_0 |\mathbf{r} - \mathbf{r'}|}}{4\pi |\mathbf{r} - \mathbf{r'}|}.
\]

(3.8)
Multiplying equation (3.6) by \( G(\mathbf{r} - \mathbf{r}') \) and equation (3.7) by \( E_s(\mathbf{r}) \) and subtracting the resulting equations from each other gives

\[
G(\mathbf{r} - \mathbf{r}')\nabla^2 E_s(\mathbf{r}) + 4\pi^2 k_0^2 E_s(\mathbf{r})G(\mathbf{r} - \mathbf{r}') - E_s(\mathbf{r})\nabla^2 G(\mathbf{r} - \mathbf{r}')
- 4\pi^2 k_0^2 G(\mathbf{r} - \mathbf{r}')E_s(\mathbf{r}) = U(\mathbf{r})G(\mathbf{r} - \mathbf{r}') - E_s(\mathbf{r})\delta(\mathbf{r} - \mathbf{r}').
\]

(3.9)

Let us interchange \( \mathbf{r} \) and \( \mathbf{r}' \) since \( G(\mathbf{r} - \mathbf{r}') = G(\mathbf{r}' - \mathbf{r}) \). After integrating both sides with respect to \( \mathbf{r}' \) on a volume \( V \), bounded by a large sphere of radius \( R \) with area \( S \), centered around the origin \( O \) in the region of the scatterer as shown in figure 3.2, equation (3.9) can be represented as

Figure 3.2: Schematic for derivation of equation (3.10)
\[
\int_{V} G(r' - r) \nabla^2 E_s(r') - E_s(r') \nabla^2 G(r' - r) \, d^3 r' = \int_{V} U(r') G(r' - r) \, d^3 r' - \int_{V} E_s(r') \delta(r' - r) \, d^3 r'.
\]

(3.10)

Converting the volume integral on the left to a surface integral by applying Green’s theorem [19], and rearranging terms, equation (3.10) can be written,

\[
E_s(r) = \int_{V} U(r') G(r' - r) \, d^3 r' - \int_{S} \left[ G(r' - r) \frac{\partial E_s(r')}{\partial n} - E_s(r') \frac{\partial G(r' - r)}{\partial n} \right] \, dS,
\]

(3.11)

where \( \frac{\partial}{\partial n} \) signifies a partial derivative in the outward normal direction at each point on the surface \( S \). The surface integral on the right hand side does not contribute to the scattered field according to the Sommerfeld radiation condition [19]. The scattered field can be represented as a convolution of the source term with the Green’s function

\[
E_s(r) = \int_{V} U(r') G(r - r') \, d^3 r' = G(r) \otimes U(r)
\]

(3.12)

This equation suggests that the scattered field \( E_s(r) \) can be thought of as linear shift invariant filtering operation applied to the source distribution \( U(r) \). The PSF or the linear filter is simply the free space Green’s function.
3.3 Far Field Imaging

It will now be shown that reconstruction of a monochromatic scattered field from the information present in the measurement from a distant boundary surface can be represented as a linear filtering operation. The transfer characteristics are derived for both the ideal case of full recording of the boundary field and an incomplete recording over a surface of restricted numerical aperture.

According to the argument presented in the previous section, the field at the spherical boundary, $E_s(r_b)$, due to an object defined by the source distribution $U(r)$ as shown in figure 3.3, can be written as

$$E_s(r_b) = \int U(r)G(r_b - r) \, d^3r$$

(3.13)

If the spherical boundary is placed in the far field such that $|r_b| \gg |r|$ as shown in figure 3.3, the free space Green’s function can be re-written such that,

$$G(r_b - r) = \frac{e^{2\pi jk_0|r_b - r|}}{4\pi|r_b - r|} \approx \frac{e^{2\pi jk_0|r_b|}}{4\pi|r_b|} e^{-2\pi jk_0r \frac{r_b}{|r_b|}},$$

(3.14)

Using equation (3.14), equation (3.13) can be expressed as

$$E_s(r_b) = \frac{e^{2\pi jk_0|r_b|}}{4\pi|r_b|} \int U(r) e^{-2\pi jk_0r \frac{r_b}{|r_b|}} \, d^3r$$

(3.15)
Let us now consider the reconstruction of the source distribution from the information contained in the boundary field. This process could be realised in practice by an optical reconstruction of the real image from the holographic recording media placed at the boundary or a numerical reconstruction from a digital recording of the boundary field. The reconstruction in all cases is essentially done by back propagation of the boundary field and can be achieved by using the complex conjugate of the Green’s function $G^*(r' - r_b)$ in the Kirchhoff integral [19, 44]

$$E_m(r') = \int_\Sigma \left[ G^*(r' - r_b) \frac{\partial E_S(r_b)}{\partial n} - E_S(r_b) \frac{\partial G^*(r' - r_b)}{\partial n} \right] ds,$$

(3.16)

where $E_m(r')$ is the reconstructed field from the boundary, $\frac{\partial}{\partial n}$ is the field derivative in the direction of the outward surface normal, and the surface integral is performed
over the boundary surface Σ. Without loss of generality, assuming the boundary is spherical of radius $r_b$, and using the far field approximation once again we have

$$\frac{\partial G^*(r' - r_b)}{\partial n} = -2\pi j k_0 G^*(r' - r_b)$$

(3.17)

and

$$\frac{\partial E_S(r_b)}{\partial n} = 2\pi j k_0 E_S(r_b).$$

(3.18)

Substituting values from equation (3.17) and (3.18) into equation (3.16) the reconstructed field is given by

$$E_m(r') = 4\pi j k_0 \int_{\Sigma} G^*(r' - r_b) E_S(r_b) \, ds$$

$$= \frac{j k_0}{4\pi} \int_{\Sigma} \left[ \int U(r) e^{-2\pi j k_0 r \cdot r_b / |r_b|^2} \, d^3 r \right] \frac{1}{|r_b|^2} e^{2\pi j k_0 r' \cdot r_b / |r_b|^2} \, ds.$$

(3.19)

Using the sifting properties of the Dirac Delta function [45], equation (3.19) can be written as the indefinite integral

$$E_m(r') = \frac{j k_0}{4\pi} \int \left[ \int U(r) e^{-2\pi j k_0 r \cdot r_b / |r_b|^2} \, d^3 r \right] \frac{1}{|r_b|^2} \delta(|r_b| - r_0) e^{2\pi j k_0 r' \cdot r_b / |r_b|} \, d^3 r_b.$$

(3.20)

Finally, making the substitution, $k/k_0 = r_b/r_0$, it is found that
\[
E_m(r') = \frac{j}{4\pi k_0} \int \int U(r) \ e^{-2\pi j k r} \ d^3r \ \delta(|k| - k_0) e^{2\pi j k r'} \ d^3k.
\]

(3.21)

In this expression the bracketed term can be recognised as the source spectrum, \( \tilde{U}(k) \). Taking the Fourier transform of equation (3.19), the spectrum of the reconstructed field \( \tilde{E}_m(k) \) can be written

\[
\tilde{E}_m(k) = \tilde{U}(k) \tilde{G}_{\text{ideal}}(k),
\]

(3.22)

where \( \tilde{G}_{\text{ideal}}(k) \) is the transfer function given by

\[
\tilde{G}_{\text{ideal}}(k) = \frac{j}{4\pi k_0} \delta(|k| - k_0)
\]

(3.23)

Consequently the reconstruction process can be considered as a linear filtering operation that modifies the source spectrum by selecting only the spatial frequency components that lie on the sphere of radius \( k_0 \). In analogy with the terminology used in X-ray diffraction the sphere of radius \( k = k_0 \) in the frequency domain it is referred to as Ewald’s sphere [45]. Thus for ideal case of full field recording the transfer function is Ewald’s sphere itself. In practice, however, most optical instruments collect light over a finite part of the sphere defined by the numerical aperture. In this case the transfer function, \( \tilde{G}_{NA}(k, k_0) \), can be represented by a portion of the sphere such that [38, 39]

\[
\tilde{G}_{NA}(k, k_0) = \frac{j}{4\pi k_0} \delta(|k| - k_0) \ step \left( \frac{k \cdot \hat{\sigma}}{k_0} - \sqrt{1 - N_A^2} \right),
\]

(3.24)
where \( step(x) \) is the Heaviside step function and \( \hat{\mathbf{e}} \) is a unit vector in the direction of observation (i.e. along the optical axis of the instrument) and \( N_A \) is numerical aperture of the system, which is defined as the sine of the maximum allowable angle over which the system can accept or emit light. In figure 3.4 a) the half of the maximum allowable angle is shown as \( \theta \) and considering the surrounding medium is air, the numerical aperture is \( N_A = \sin \theta \).

Equation (3.24) defines the 3D TF of a far field imaging system of finite aperture and is shown in figure 3.4 b). The 3D PSF is the inverse Fourier transform of this expression and is given by

\[
G_{NA}(\mathbf{r}, k_0) = \int \frac{j}{4\pi k_0} \delta(|\mathbf{k}| - k_0) \text{step} \left( \frac{k \cdot \hat{\mathbf{e}}}{k_0} - \sqrt{1 - N_A^2} \right) e^{j2\pi \mathbf{k} \cdot \mathbf{r}} d^3k,
\]

(3.25)

Equations (3.24) and (3.25) define the response of a far field imaging system of restricted numerical aperture. Correspondingly the resolution of the far field imaging system can be represented in terms of Nyquist distance. The Nyquist distance is defined as the shortest spatial period that can be faithfully reproduced by a system of given bandwidth [47]. For a frequency representation that is symmetric about the origin, the Nyquist distance, in a given direction is reciprocal to the total extent of the frequency representation in that direction. The axial (\( \Delta r_{\text{axial}} \)) and lateral Nyquist distance (\( \Delta r_{\text{lateral}}^{x}, \Delta r_{\text{lateral}}^{y} \)) can therefore be found out by,

\[
\Delta r_{\text{lateral}}^{x} = \frac{1}{\Delta k_{\text{lateral}}^{x}} \approx \frac{\lambda}{2N_A}
\]

\[
\Delta r_{\text{lateral}}^{y} = \frac{1}{\Delta k_{\text{lateral}}^{y}} \approx \frac{\lambda}{2N_A}
\]
\[ \Delta r_{\text{axial}} = \frac{1}{\Delta k_{\text{axial}}} = \frac{\lambda}{(1 - \sqrt{1 - N_A^2})} \]

(3.26)

It can be seen from the above expressions that for the system with restricted NA, NA plays an important role in defining the resolution. The maximum lateral resolution that can be achieved considering \( N_A = 1 \), is half the operating wavelength while the maximum axial resolution is about a wavelength.

Figure 3.4: a) Numerical aperture representation for a microscopic objective b) numerical aperture restriction in the spectrum plane on Ewald’s sphere
It is noted once again that the reconstructed field can be thought of as a 3D filter applied to the source distribution $U(\mathbf{r})$. For the case of a measuring instrument it is necessary to consider how the source distribution is related to the function that defines the object – in this case the refractive index contrast defined by equation (3.5). This is discussed in the following section.

3.4 Application of Linear Theory to CSI

So far it has been shown that the scattered field from an object can be represented as a linear filtering operation applied to the source distribution that is characterised by a PSF that is the free space Green’s function. It was also shown that for a far field instrument, the reconstruction of the scattered field from the information present at the boundary surface has undergone further filtering since only plane wave components that pass through the instruments numerical aperture are retained. It is important to realise that in both cases the filter is applied to the source distribution which is a function of the input illumination and the function that characterises the object (in this case the refractive index contrast). In this section, it is shown that in certain circumstances the source distribution is linearly related to the object function and consequently the optical instrument can be characterised as a linear filtering operation applied to the object function. In the following section the necessary linear theory is developed with reference to volume scattering and later surface scattering, however before this, it is worth considering the interferogram that is output by CSI.

3.4.1 CSI Output

The basic working principle of CSI was discussed in the first chapter. The purpose of this section is to develop expressions that relate the output of a CSI instrument to the
illuminating field $E_r(\mathbf{r})$ and scattered field $E_s(\mathbf{r})$ defined in the previous section. As described in chapter 1 a CSI instrument exploits an interference objective to record the interference between light scattered from the object surface $E_s(\mathbf{r})$ and the light reflected from the reference surface $E_r(\mathbf{r})$ as shown in figure 3.5. However, the field recorded by the instrument is a filtered version of this interference. For the case of a flat reference surface due to reflection the reference light is simply a negative copy of the reference field $-E_r(\mathbf{r})$. Consequently the recorded interference pattern is between the measured scattered field $E_m(\mathbf{r})$ and $-E_r(\mathbf{r})$. The recorded intensity can be written as

$$I(\mathbf{r}) = |E_m(\mathbf{r}) - E_r(\mathbf{r})|^2$$

$$I(\mathbf{r}) = |E_r(\mathbf{r})|^2 + |E_m(\mathbf{r})|^2 - E_m(\mathbf{r})^*E_r(\mathbf{r}) - E_m(\mathbf{r})E_r(\mathbf{r})^*.$$

(3.27a)

It is noted that these terms are similar to those obtained for the case of holography and can be separated in the frequency domain in a similar manner. In this case, however a 3D interferogram is obtained and it is necessary to consider the bandwidth occupied by each term in the 3D frequency space ($\mathbf{k}$-space).
The first term $I_1(\mathbf{r})$ in equation (3.27a) is the intensity of the reference or the illumination field $E_r(\mathbf{r}) = e^{2\pi j \mathbf{k} \cdot \mathbf{r}}$ which is uniform. Taking Fourier transform,

$$I_1(\mathbf{k}) = \int |E_r(\mathbf{r})E^*_r(\mathbf{r})| e^{-j2\pi \mathbf{k} \cdot \mathbf{r}} d^3\mathbf{r} = \delta(\mathbf{k}),$$

(3.27b)

Hence this represents the ‘DC term’ in holography.

The second term as shown in equation (3.27a) is the intensity of the measured scattered field $E_m(\mathbf{r})$. Taking the Fourier transform

$$\tilde{I}_2(\mathbf{k}) = \int |E_m(\mathbf{r})E^*_m(\mathbf{r})| e^{-j2\pi \mathbf{k} \cdot \mathbf{r}} d^3\mathbf{r}$$
\[
= \int \tilde{E}_m(k')\tilde{E}_m(k + k')^* \, d^3k'.
\]

(3.27c)

This can be recognized as the autocorrelation of the measured field. As described in the previous section for a far-field instrument \( E_m(k) \) is band limited by the numerical aperture such that the spectrum lies in \( \tilde{G}_{NA}(k) \) as shown in figure 3.4. In the frequency domain therefore this term can be thought of as the auto correlation of \( \tilde{G}_{NA}(k) \) as shown as in figure 3.6.

The third term and fourth terms are functions of the reference field and the scattered field and hence correspond to the ‘signal terms’ in holography. Taking the Fourier transform of the third term,

\[
\tilde{I}_3(k) = \int E_m(\mathbf{r})^*E_r(\mathbf{r}) \, e^{-j2\pi k \mathbf{r}} \, d^3\mathbf{r}
\]

\[
= \int E_m(\mathbf{r})^*e^{j2\pi k \mathbf{r}} \, e^{-j2\pi k \mathbf{r}} \, d^3\mathbf{r}
\]

\[
= \int E_m(\mathbf{r})^* e^{-j2\pi(k-k_r) \mathbf{r}} \, d^3\mathbf{r}
\]

\[
= \tilde{E}_m(k - k_r)^*
\]

(3.27d)

Note once again that \( \tilde{E}_m(k) \) is band limited and must occupy a space described by \( \tilde{G}_{NA}(k)^* \). The complex conjugate of \( \tilde{G}_{NA}(k) \), will be on the other side of \( \tilde{G}_{NA}(k) \). This term must be found in the region where \( \tilde{G}_{NA}(k - k_r) \) is non zero.
Similarly, the fourth term in equation (3.27a) must occupy the space where \( \tilde{G}_{NA}(\mathbf{k} + \mathbf{k}_r) \) is non zero which will be on the opposite side of the term described in equation (3.27d) in frequency domain, such that

\[
\tilde{I}_4(\mathbf{k}) = \int E_m(\mathbf{r})E_r(\mathbf{r})^* e^{-j2\pi\mathbf{k}\cdot\mathbf{r}} d^3r
\]

\[
= \int E_m(\mathbf{r}) e^{-j2\pi\mathbf{k}\cdot\mathbf{r}} e^{-j2\pi\mathbf{k}_r\cdot\mathbf{r}} d^3r
\]

\[
= \int E_m(\mathbf{r}) e^{-j2\pi(\mathbf{k}+\mathbf{k}_r)\cdot\mathbf{r}} d^3r
\]

\[
= \tilde{E}_m(\mathbf{k} + \mathbf{k}_r).
\]

(3.27e)

Furthermore it is noted that the above has been derived assuming a single monochromatic illumination. In CSI we have many incoherent sources and these terms therefore occupy slightly different regions in \( \mathbf{k} \)-space. We can represent this as

\[
I(\mathbf{k}_r, \mathbf{r}) = \sum_{\mathbf{k}_r} [ |E_r(\mathbf{k}_r, \mathbf{r})|^2 + |E_m(\mathbf{k}_r, \mathbf{r})|^2 - E_m(\mathbf{k}_r, \mathbf{r})^* E_r(\mathbf{k}_r, \mathbf{r})
\]

\[
- E_m(\mathbf{k}_r, \mathbf{r}) E_r(\mathbf{k}_r, \mathbf{r})^* ]
\]

(3.27f)

Taking this into account figure 3.6 shows various terms
Neglecting the first three terms, the reconstructed field from the CSI can be represented by the fourth term as shown in equation (3.27e). The discussion will now continue to the linear theory of optical instruments applied to volume scattering objects.

### 3.4.2 Volume Scattering

In section 3.3 it was shown that the source distribution is given by

\[ U(\mathbf{r}) = k_0^2 \Delta(\mathbf{r})(E_r(\mathbf{r}) + E_s(\mathbf{r})) \]

(3.28)

It is noted that it depends on both the illuminating field and scattered field. Since the latter also depends on the object function, it is in general a non-linear function of the refractive index contrast.
If the medium is weakly scattering however, such that the scattered field is small compared to the illumination field, and the source distribution can be written as

\[
U(\mathbf{r}) = k_0^2 \Delta(\mathbf{r})(E_r(\mathbf{r}) + E_s(\mathbf{r})) \approx k_0^2 \Delta(\mathbf{r})E_r(\mathbf{r})
\]

(3.29)

This is the Born Approximation [45] and is valid when the refractive index differs only slightly from unity such that the refractive index contrast is small or in words the object is small such as particles suspended on fluids. Considering this approximation, the measured field can be written as

\[
E_m(\mathbf{r}) = k_0^2 \int V \cdot \mathbf{G}_{NA}(\mathbf{r} - \mathbf{r}') \Delta(\mathbf{r}')E_r(\mathbf{r}') \, d^3 r'
\]

(3.30)

Replacing \( E_m(\mathbf{r}) \) in the 4th term of equation (3.27a) and representing as the output for CSI, \( O_B(\mathbf{r}) \), (where the subscript \( B \) is for the Born approximation)

\[
O_B(\mathbf{r}) = k_0^2 \int V \cdot \mathbf{G}_{NA}(\mathbf{r} - \mathbf{r}') \Delta(\mathbf{r}')E_r(\mathbf{r}')E_r^*(\mathbf{r}) \, d^3 r'
\]

(3.31)

Considering \( E_r(\mathbf{r}) \) is a monochromatic plane wave illumination, the above expression can be written as a 3D convolution with the object function \( \Delta(\mathbf{r}) \) [38] such that

\[
O_B(\mathbf{r}) = k_0^2 \int \mathbf{G}_{NA}(\mathbf{r} - \mathbf{r}')e^{-2\pi jk_r(\mathbf{r} - \mathbf{r}')} \Delta(\mathbf{r}') \, d^3 r',
\]

(3.32)
where $\mathbf{k}_r$ is the reference wave vector. The above expression is identical to that of the reconstruction of a hologram by a single monochromatic plane wave. This can be considered as a linear filtering operation on the object function $\Delta(\mathbf{r})$ (or the refractive index contrast) rather than on source distribution $U(\mathbf{r})$. The corresponding PSF can be represented as

$$H_{holo}(\mathbf{r}) = k_0^2 G_{NA}(\mathbf{r}) e^{-2\pi j \mathbf{k}_r \cdot \mathbf{r}}, \quad (3.33)$$

and the transfer function is presented as the Fourier transform of equation (3.33)

$$\tilde{H}_{holo}(\mathbf{k}) = k_0^2 G_{NA}(\mathbf{k} + \mathbf{k}_r) \quad (3.34)$$

The schematic of the transfer function is shown in figure 3.7. From equation (3.34) it is clear that the TF of the single illumination holographic reconstruction will be similar to the TF for the far field imaging system as shown in figure 3.4 except it will be shifted by the incident wave vector $\mathbf{k}_r$. 

106
It is straightforward to extend the analysis from the single monochromatic illumination as shown in equation (3.34) to demonstrate the TF corresponding to the quasi-monochromatic (QM) illumination. If CSI is operated in a QM illumination, the interferogram will be the sum of the each plane wave component illuminating the field over the numerical aperture $N_A$. The TF ($\tilde{H}^\text{QM}_B(\mathbf{k})$) can be written as,

$$\tilde{H}^\text{QM}_B(\mathbf{k}) = k_0^2 \int \tilde{G}_{NA}(\mathbf{k}) \tilde{G}_{NA}(\mathbf{k} - \mathbf{k}_r) d^3k_r$$

$$= k_0^2 \tilde{\alpha}_{NA}(\mathbf{k}) \otimes \tilde{G}_{NA}(\mathbf{k}),$$

(3.35)

and the PSF is
\[ H_B^{QM}(\mathbf{r}) = \int G_{NA}^2(\mathbf{r}, k_0) k_0^2 \, dk_0. \]  

(3.36)

A schematic diagram showing the TF is shown in figure 3.8. A \( k_y k_z \) and \( k_x k_y \) plot of the 3D TF of QM mode CSI is shown in figure 3.8 a). The TF in this case is represented as the convolution between two of the TFs for the far field imaging system to incorporate all the illumination directions within the numerical aperture. Corresponding axial and lateral resolutions can be found by the reciprocal of the bandwidth as shown in figure 3.8 b) and is represented as

\[
\Delta r_{\text{lateral}}^x = \frac{1}{\Delta k_{\text{lateral}}^x} \approx \frac{\lambda}{4N_A}
\]

\[
\Delta r_{\text{lateral}}^y = \frac{1}{\Delta k_{\text{lateral}}^y} \approx \frac{\lambda}{4N_A}
\]

\[
\Delta r_{\text{axial}} = \frac{1}{\Delta k_{\text{axial}}} = \frac{\lambda}{2 \left(1 - \sqrt{1 - N_A^2}\right)}
\]

(3.37)

The resolution expressions in equation (3.37) show that both the lateral and axial resolution is increased by a factor of 2. The corresponding TF and PSF are calculated for numerical aperture of 0.55 and mean wavelength of 600 nm as shown in figure 3.9.
Figure 3.8: (a) $k_y k_z$ and $k_x k_y$ section of 3D TF of QM mode CSI (b) calculating the bandwidth
Figure 3.9: Sections through a) TF (absolute value) and b) PSF (real part) using Born approximation for CSI operating in a quasi-monochromatic mode.
The mode used most commonly for CSI is the one utilising white light or broad band illumination. For a broad band light source, the effects of all different wavelengths present in the illumination need to be considered while determining the transfer characteristics. This can be incorporated by integrating the TF of the QM mode over all spectral components present in the source spectrum. Thus for the case of CSI with broad band white light illumination the TF can be written as,

$$\tilde{H}_B(\kappa) = \int k_0^2 S(k_0) [\tilde{G}_{NA}(\kappa) \otimes \tilde{G}_{NA}(\kappa)] dk_0,$$

(3.38)

where $S(k_0)$ is the spectral density expressed as a function of the wavenumber $k_0$.

The PSF is

$$H_B(r) = \int G_{NA}^2(r, k_0) k_0^2 S(k_0) dk_0,$$

(3.39)

The corresponding TF and PSF is calculated considering a numerical aperture of 0.55, mean wavelength 600 nm with a bandwidth of 120 nm, as shown in figure 3.10 (a) and 3.10 (b)
Figure 3.10: Sections through a) TF (absolute value) and b) PSF (real part) using Born approximation.
The resolution of wideband CSI can be estimated as follows. Calculation of the lateral resolution is straightforward, which is the inverse of the lateral extent of the transfer function and in this case it is dependent on the $k_{\text{max}}$ or the highest spatial frequency content in the lateral direction. Here, the effects of all different spectral components are needed to be considered. It is important to mention that, as
wavelength ($\lambda$) and wave number ($k$) are inversely proportional, $k_{\text{max}}$ will be $1/\lambda_{\text{min}}$ and $k_{\text{min}} = 1/\lambda_{\text{max}}$.

$$\Delta r_{\text{lateral}}^x = \frac{1}{\Delta k_{\text{lateral}}} = \frac{1}{4k_{\text{max}}N_A} \approx \frac{\lambda_{\text{min}}}{4N_A}$$

$$\Delta r_{\text{lateral}}^y = \frac{1}{\Delta k_{\text{lateral}}} = \frac{1}{4k_{\text{max}}N_A} \approx \frac{\lambda_{\text{min}}}{4N_A}$$

However for axial resolution; the situation is a bit more complicated. The axial situation can be described such that

$$\Delta r_{\text{axial}} = \frac{1}{\Delta k_{\text{axial}}} = \frac{1}{2(k_{\text{max}} - k_{\text{min}}) + 2k_{\text{min}} \left(1 - \sqrt{1 - N_A^2}\right)}$$

$$\approx \frac{1}{2 \left(\frac{1}{\lambda_{\text{min}}} - \frac{1}{\lambda_{\text{max}}}\right) + \frac{2 \left(1 - \sqrt{1 - N_A^2}\right)}{\lambda_{\text{max}}}}.$$  

(3.40)

Different spectral components in the TF will interfere resulting in the fringe pattern in the axial direction of the PSF as shown in figure 3.10 b). Generally, for interference patterns the distance between fringes is half the wavelength of the signal. However, in this case due to the presence of different spectral components, the distance between the fringes is half the effective wavelength ($\bar{\lambda}$). It can be calculated as the inverse of effective wave number ($\bar{k}$) derived from the centroid of the TF shown in equation (3.41) [48] such that

$$\bar{k} = \frac{1}{2} \left( \frac{1}{\lambda_{\text{min}}} + \frac{1}{\lambda_{\text{max}}} \right) \left( \frac{1}{\lambda_{\text{max}}} - \frac{1}{\lambda_{\text{min}}} \right)$$

(3.41)
3.4.3 Surface Scattering

In the last section it was demonstrated that the recorded interferogram in CSI is similar to holography where the interference between the light reflected from the reference surface and light scattered from the object surface is recorded. Here CSI is represented as a linear filtering operation where the filter is characterized by the PSF in the space domain and the TF in the frequency domain. It is also shown that the output fringe pattern can be obtained by convolving the object function with the PSF, and the corresponding equation is derived considering the Born approximation. Though the Born approximation provides a good way to estimate the theoretical performance of the system, the approximation rests on the assumption of weak scattering, which implies that the incident field is weakly perturbed by the object. This is reasonable for objects that are characterized by small variations in refractive index, such as cellular tissue or if the object consists of sparse point like objects, but is rarely justified for general 3D objects. So the attention is moved to a surface scattering model [49-53] to incorporate the effects of light scattered from the object. For the case of strong surface scattering from the interface between two homogenous media, however, the process can be represented as linear provided that there is no multiple scattering [49- 51]. In order to relate these two apparently disparate processes, scattering by an object characterized by the function $\Delta(\mathbf{r}) = 4\pi^2 \left(1 - n^2(\mathbf{r})\right)$ (equation 3.5) is considered as shown in figure 3.12.
If the object is illuminated by the reference field $E_r(r)$, then the scattered field $E_s(r)$ is given by the integral form of the Helmholtz equation as shown in equation (3.12),

$$E_s(r) = \int_{\mathcal{V}} U(r') G(r - r') d^3r',$$

(3.42)

where, $G(r) = e^{i 2 \pi k_0 |r|}/(4\pi |r|)$ is the free-space Green’s function that defines a point source and $U(r') = k_0^2 \Delta(r')(E_r(r') + E_s(r'))$. Replacing the value of $U(r')$,

$$E_s(r) = k_0^2 \int G(r - r') \Delta(r')(E_s(r') + E_r(r')) d^3r'.$$

(3.43)

It is clear, however, that the only contribution to the integral is from regions where $\Delta(r') = 4\pi^2 (1 - n^2)$ is non-zero (i.e. from the volume occupied by the object itself) and the scattered field can, therefore, be written as the volume integral,

$$E_s(r) = k_0^2 \int G(r - r') \Delta(r') E_t(r') d^3r'.$$

(3.44)
where $E_t(\mathbf{r})$ is the transmitted field inside the object boundary. According to the surface scattering model, the fields present on the surface are the incident field $E_r(\mathbf{r})$, the scattered field $E_s(\mathbf{r})$ and the transmitted field $E_t(\mathbf{r})$ inside the object boundary. Since inside the object [39]

$$
(\nabla^2 + 4\pi^2n^2k_0^2)E_t(\mathbf{r}) = 0
$$

(3.45a)

and

$$
(\nabla^2 + 4\pi^2k_0^2)G(\mathbf{r} - \mathbf{r}') = 0.
$$

(3.45b)

Multiplying equation (3.45a) with $G(\mathbf{r} - \mathbf{r}')$ and equation (3.45b) with $E_t(\mathbf{r})$ and subtracting the resultant equations,

$$
G(\mathbf{r} - \mathbf{r}')E_t(\mathbf{r}) = \frac{1}{4\pi^2k_0^2(n^2-1)}(E_t(\mathbf{r})\nabla^2 G(\mathbf{r} - \mathbf{r}') - G(\mathbf{r} - \mathbf{r}')\nabla^2 E_t(\mathbf{r})).
$$

(3.46)

Substituting the value of $G(\mathbf{r} - \mathbf{r}')E_t(\mathbf{r})$ from equation (3.46) to equation (3.44)

$$
E_s(\mathbf{r}) = \int \left( (G(\mathbf{r} - \mathbf{r}')\nabla^2 E_t(\mathbf{r}) - E_t(\mathbf{r})\nabla^2 G(\mathbf{r} - \mathbf{r}'))d^3r.
$$

(3.47)

Applying Green’s theorem to the right hand side of the equation (3.47),
\[ E_s(r) = \int \left( G(r - r') \frac{\partial E_t(r)}{\partial n} - E_t(r) \frac{\partial G(r - r')}{\partial n} \right) ds \]

(3.48)

Equation (3.48) is Kirchhoff’s integral. This equation shows that for the case of surface scattering from homogenous medium the scattered field can be represented as the field present in the boundary. The process now can be linearized by assuming the appropriate surface boundary conditions. Following Beckman [49], if the surface is illuminated by a plane wave, \( E_r(r) \), the boundary field and its normal derivative are given by

\[ E_t(r) = (1 + R)E_r(r) \]

(3.49)

\[ \frac{\partial E_t(r)}{\partial n} = 2\pi j k_r \hat{n}_s (1 - R)E_r(r), \]

(3.50)

where \( \hat{n}_s \) is the outward surface normal, and \( R \) is the Fresnel amplitude reflection coefficient. Further it is assumed that \( R \) is a constant over the range of scattering angles of interest. Beckmann has discussed the validity of these boundary conditions in detail but for application this context, it is outlined below:

i) The surface must be slowly varying on the optical scale such that the local radius of curvature is more than the wavelength (\( \lambda \)). This is the Kirchhoff or physical optics approximation. In this approximation the field at any point on the surface is represented by the sum of the incident and reflected fields with the reflection coefficient (\( R \)) of the plane which
is tangent at the particular point. So the field at the surface is represented as the field that would be present on the tangent plane at that point as shown in figure 3.13, where the surface is shown by the continuous line and the tangent is by the dashed line. As equation (3.49) is fulfilled exactly in the case of an infinite plane, it will make a good approximation for locally flat surfaces where the radius of curvature of the surface is large compared to the wavelength as shown in figure 3.13 a). However, if the radius of curvature is less than the wavelength, i.e. with the existence of sharp edges or sharp points the approximation breaks down and can’t be applied [49].

![Figure 3.13: The tangent at a general point on the surface. The radius of curvature is a) large (n times) and b) small in comparison with the wavelength (λ).](image)

ii) For a perfect conductor the reflection coefficient is constant \( R = 1 \)

iii) More generally the reflection coefficient \( R \) depends on polarisation (p or s). For Fresnel’s reflection coefficient, it is observed that reflectance of both of the orthogonal polarisations increases if the light is incident at an
angle greater than Brewster’s angle, which would require incorporating the effect of polarization in the equations. So, for a simpler calculation it is considered that the light is incident at less than Brewster’s angle. For light incident from a medium of refractive index \( n_1 \) to a medium of refractive index \( n_2 \), Brewster’s angle \( (\theta_B) \) is represented as \( \theta_B = \tan^{-1}\left(\frac{n_2}{n_1}\right) \). For light travelling from a lighter medium to a denser medium Brewster’s angle will always be more than 45 degrees. Therefore, providing the angles of incidence are less than 45 degrees, the sum of the state normal to the plane of incidence is considered to be approximately constant.

iv) For a dielectric, the field at the lower boundary and its gradient may depart markedly from those given in equations (3.49) and (3.50) due to propagation through the object. However, this component of the field will generally be separable from that scattered from the top boundary due to the extra path length travelled.

Figure 3.14: Surface scattering to a distant boundary
It has been shown previously that the process of reconstructing the field using the boundary fields is a linear filtering operation and a similar approach will be implemented in this case. First the field at a distant boundary \( \mathbf{r}_b \) from the object is computed on the distant boundary \( \Sigma \) as shown in figure 3.14 by applying the boundary values in equation (3.49) and (3.50). Later it will be shown that the field can be reconstructed using the boundary field only.

Replacing the value of \( E_r(\mathbf{r}) = e^{2\pi j k_e \mathbf{r}} \), equation (3.49) and (3.50) the transmitted field and normal derivative, respectively, can be written as

\[
E_t(\mathbf{r}) = (1 + R) e^{2\pi j k_e \mathbf{r}}
\]

(3.51)

\[
\frac{\partial E_t(\mathbf{r})}{\partial n} = 2\pi j (1 - R)e^{2\pi j k_e \mathbf{r}} \mathbf{k}_e \cdot \hat{n}_s.
\]

(3.52)

Since the boundary is at a large distance \( \mathbf{r}_b \gg \mathbf{r} \), the far field Green’s function can be written as,

\[
G(\mathbf{r}_b - \mathbf{r}) \approx \frac{e^{2\pi j k_0 |\mathbf{r}_b|}}{4\pi |\mathbf{r}_b|} e^{-2\pi j k_0 |\mathbf{r}_b|}.
\]

(3.53)

The normal derivative of Green’s function is

\[
\frac{\partial G(\mathbf{r}_b - \mathbf{r})}{\partial n} = -2\pi j G(\mathbf{r}_b - \mathbf{r})_0 \frac{\mathbf{r}_b}{|\mathbf{r}_b|} \cdot \hat{n}_s.
\]

(3.54)
To calculate the field at the distant boundary, the Kirchhoff integral in equation (3.48) can be written by incorporating the distance to the boundary $\mathbf{r}_b$ such that

$$E_s(\mathbf{r}) = \int \left( G(\mathbf{r}_b - \mathbf{r}) \frac{\partial E_t(\mathbf{r})}{\partial n} - E_t(\mathbf{r}) \frac{\partial G(\mathbf{r}_b - \mathbf{r})}{\partial n} \right) d\mathbf{s}$$

(3.55)

By replacing the values of the field and Green’s function from equations (3.51), (3.52), (3.53) and (3.54) in the Kirchhoff integral in equation (3.55) the following expression is obtained

$$E_s(\mathbf{r}) = 2\pi j \frac{e^{2\pi j k_0 |\mathbf{r}_b|}}{4\pi |\mathbf{r}_b|} \int e^{-2\pi j \left(k_0 |\mathbf{r}_b| - k_0 \mathbf{r} \cdot \mathbf{k}\right)} \left[ R \left(k_0 |\mathbf{r}_b| - k_0 \mathbf{r} \cdot \mathbf{k}\right) + \left(k_0 |\mathbf{r}_b| + k_0 \mathbf{r} \cdot \mathbf{k}\right)\right] \hat{n}_s d\mathbf{s}$$

(3.56)

Considering the boundary conditions, in particularly condition (iv), a region of interest on the upper surface of the object can be defined as function $A(\mathbf{r})$ such that

$$A(\mathbf{r}) = W(r_x, r_y) \delta \left(r_z - s(r_x, r_y)\right),$$

(3.57)

where $W(r_x, r_y)$ is a window function that defines the illuminated area and $s(r_x, r_y)$ is the 2D function that defines the height of the surface. Equation (3.56) represents the field at the boundary as a surface integral. It can be represented as a volume integral and shown in equation (3.58)

$$E_s(\mathbf{r}_b) \approx j \frac{e^{2\pi j k_0 |\mathbf{r}_b|}}{2 |\mathbf{r}_b|} \int e^{-2\pi j \left(k_0 |\mathbf{r}_b| - k_0 \mathbf{r} \cdot \mathbf{k}\right)} \left[ R \left(k_0 |\mathbf{r}_b| - k_0 \mathbf{r} \cdot \mathbf{k}\right) + \left(k_0 |\mathbf{r}_b| + k_0 \mathbf{r} \cdot \mathbf{k}\right)\right] \hat{n}_s A(\mathbf{r}) d^3\mathbf{r}$$

(3.58)
$E_s(r_b)$ is the scattered field calculated at the boundary. Now, it will be shown that the field at some other position on the surface can be reconstructed from the boundary field only as shown in section 3.4. The measured field at a point with a distance $r'$ from the boundary can be calculated by the back propagation of the boundary field. Accordingly the Green’s function this will be a point sink defined by the conjugate of the Green’s function and the measured field can be written as the Kirchhoff integral

$$E_m(r') = \int \left[ G^*(r' - r_b) \frac{\partial E_s(r_b)}{\partial n} - E_s(r_b) \frac{\partial G^*(r' - r_b)}{\partial n} \right] ds$$

(3.59)

Substituting the values of the field at the boundary and Greens function and assuming without loss of generality, that the boundary surface is spherical (similar to figure 3.3) such that

$$E_m'(r') = \frac{-k_0}{2} \int_\Sigma \frac{1}{|r_b|^2} e^{-2\pi j \left( k_0 \frac{r_b}{|r_b|} \cdot r \right)} R \left( k_0 \frac{r_b}{|r_b|} - k_r \right) + \left( k_0 \frac{r_b}{|r_b|} + k_r \right) \cdot \hat{n}_S \times \frac{A(r)}{|n_S \cdot z|} d^3r e^{2\pi j k_0 r' \frac{r_b}{|r_b|}} ds.$$  

(3.60)

Equation (3.60) describes the field at a point $r'$, but it is in the surface integral form. It can expressed in the volume integral form using the sifting properties of the Dirac delta function.
\[ E_m'(r') = \frac{- k_0}{2} \int \frac{1}{|r_b|^2} e^{-2\pi j (k_0 r_b - k r)} \left[ R \left( k_0 \frac{r_b}{|r_b|} - k \right) \right] \]

\[ + \left( k_0 \frac{r_b}{|r_b|} + k_r \right) \cdot \hat{n}_A \cdot \hat{z} \int e^{2\pi j k_0 r' \frac{r_b}{|r_b|} \delta(|r_b| - r_s)} d^3 r_b. \]

(3.61)

Considering \( k'/k_0 = r_b/r_0 \)

\[ E_m'(r') = - \frac{1}{2k_0} \int \left[ e^{-2\pi j (k' - k_r) \cdot r} \right] \left[ R(k' - k_r) + (k' + k_r) \right] \cdot \hat{n}_A \cdot \hat{z} \int e^{2\pi j k_0 r' \frac{r_b}{|r_b|} \delta(|r_b| - r_s)} d^3 r_b. \]

(3.62)

A further simplification can be made by considering the phase within the bracketed integral in equation (3.62). Since the phase of the complex exponential changes in the direction defined by \( (k' - k_r) \), only regions of the surface where the surface normal is in this direction will contribute to the integral. This is the principle of stationary phase and is illustrated in figure 3.14. Noting that case in these regions the term \( (k' + k_r) \cdot \hat{n}_s \) is negligible and \( \hat{n}_s = (k' - k_r)/|k' - k_r| \), equation (3.62) becomes

\[ E_m'(r') = - \frac{R}{2k_0} \int \int e^{-2\pi j (k' - k_r) \cdot r} \left( \frac{|k' - k_r|^2}{(k' - k_r) \cdot z} \right) A(r) \cdot d^3 r \cdot \delta(|k'|) \]

\[ - k_0) e^{2\pi j k' \cdot r'} d^3 k'. \]

(3.63)
Rearranging the terms in equation (3.63)

\[ E_m'(r') = \frac{-R}{2k_0} \int \left( \left( \frac{|k' - k_r|^2}{(k' - k_r) \cdot z} \right) \delta(|k'| - k_0) e^{-2\pi j(kr - kr') \cdot x} \int A(r) e^{2\pi jkr'} d^3 r \right) d^3 k'. \]

(3.64)

The interference that is recorded in the CCD is practically the intensity of the interference between the measured scattered field \( E_m(r) \) and the reference field \(-E_r(r)\) as shown in figure (3.5) and described in equation (3.27a) and (3.27f). The output term can be written as,

\[ O(r) = E_m(r) e^{-2\pi jk_r r} \]

(3.65)

Substituting the values of \( E_m(r') \) from equation (3.64) to equation (3.65), the output field is written as

Figure 3.15: Principle of stationary phase
\[
O(r') = \frac{-R}{2k_0} \int \int e^{-2\pi j (k' - k_r) \cdot r} \left( \frac{|k' - k_r|^2}{(k' - k_r) \cdot z} \right) A(r) d^3r \delta(|k'| - k_0) \] 
\[
- k_0) e^{2\pi j k r' r'} e^{-2\pi j k_r r'} d^3k',
\]

(3.66)

Rearranging the terms in equation (3.66)

\[
O(r') = \frac{-R}{2k_0} \int \left[ \left( \frac{|k' - k_r|^2}{(k' - k_r) \cdot z} \right) \delta(|k'| - k_0) \right] e^{2\pi j (k' - k_r) r'} \int e^{-2\pi j (k' - k_r) \cdot r} A(r) d^3r] d^3k'.
\]

(3.67)

Defining the object as the foil model of the original object, \( \Delta_F(r) \) as

\[
\Delta_F(r) = 4\pi j RA(r) = 4\pi j RW(r_x, r_y) \delta(r_z - s(r_x, r_y)).
\]

(3.68)

For an ideal case, substituting the value of \( \tilde{G}_{\text{ideal}}(k', k_0) \) from equation (3.23),

\[
\tilde{G}_{\text{ideal}}(k', k_0) = \frac{j}{4\pi k_0} \delta(|k'| - k_0),
\]

the output can be written as

\[
O(r') = \int \left[ \left( \frac{|k' - k_r|^2}{2(z \cdot (k' - k_r))} \right) \tilde{G}_{\text{ideal}}(k', k_0) e^{2\pi j (k' - k_r) r'} \right] \int e^{-2\pi j (k' - k_r) \cdot r} \Delta_F(r) d^3r] d^3k'.
\]

(3.69)

However in practice most optical instruments collect light over a finite numerical aperture, \( N_A \), so in this case \( \tilde{G}_{\text{ideal}}(k', k_0) \) will be replaced by \( \tilde{G}_{NA}(k', k_0) \) as shown in equation (3.24)

Replacing \( k = k' - k_r \) in equation (3.69)
\[
O(r') = \int \int \Delta_F(r) e^{-2\pi jkr'} d^3r \left[ \left( \frac{|k|^2}{2(k_\cdot z)} \right) \tilde{G}_{NA}(k + k_r, k_0) e^{2\pi jkr'} \right] d^3k'
\]

(3.70)

or

\[
\tilde{O}_F(k) = \tilde{\Delta}_F(k) \left( \frac{|k|^2}{2k_\cdot z} \right) \tilde{G}_{NA}(k + k_r, k_0)
\]

(3.71)

Finally, incorporating the spectral density \(S(k_0)\) of the source which is a function of wavenumber and integrating over all illumination wave vectors, \(k_r\), within the numerical aperture and all wave numbers, the transfer function \(\tilde{H}(k)\) is given by

\[
\tilde{H}_F(k) = \left( \frac{|k|^2}{2k_\cdot z} \right) \int \int \tilde{G}_{NA}(k_r, k_0) \tilde{G}_{NA}(k - k_r, k_0) d^3k_r S(k_0) dk_0
\]

(3.72)

The output of the CSI corresponding to this surface scattering model will be,

\[
\tilde{O}_F(k) = \tilde{\Delta}_F(k) \tilde{H}_F(k)
\]

(3.73)

Equation (3.72) defines the response of a CSI of restricted numerical aperture when it is applied to the foil model of an object surface as defined by equation (3.68).

Figures 3.15 a) and 3.15 b) illustrate the TF and PSF of an ideal instrument, respectively, having \(NA = 0.55\), a mean wavelength of \(\lambda = 600\) nm and a Gaussian spectral density, \(S(k_0)\), with a bandwidth of 120 nm (FWHM at 1/e²).
Figure 3.16: Sections through a) TF (absolute value) and b) PSF (real part) using the ‘foil’ approximation.
Figure 3.10 shows the transfer characteristic calculated using the Born approximation, and Figure 3.16 shows it for the ‘foil’ model. In the next section the characteristics of these two will be compared.

### 3.6 Discussion

The TF defined in equation (3.72) is similar to the “effective transfer function” as described by Sheppard in the context of confocal microscopy [50, 51]. Comparing the TF calculated using the ‘foil’ model with the TF defined using the Born approximation

$$\tilde{H}_F(k) = \left(\frac{|k|^2}{2k \cdot z}\right) \int \int \tilde{g}_{NA}(k_r, k_0) \tilde{g}_{NA}(k - k_r, k_0) d^3 k_r S(k_0)dk_0$$

(3.74)

$$\tilde{H}_B(k) = \int \int \tilde{g}_{NA}(k', k_0) \tilde{g}_{NA}(k - k', k_0) d^3 k' k_0^2 S(k_0)dk_0$$

(3.75)

Although the forms of equations (3.74) and (3.75) are similar, they differ in the weighting term, $\left(\frac{|k|^2}{2k \cdot z}\right)$ in equation (3.74) and the factor $k_0^2$ that weights the spectrum of the illumination source, $S(k_0)$, in equation (3.75). This difference in the weighting is observed in the absolute values of the TF as shown in figure 3.10(a) and 3.15 (a).

It should also be noted that in each derivation the object function to which the filtering operation is applied is defined in a slightly different way. In the case of the Born approximation this is
\[ \Delta_\theta(r) = 4\pi^2(1 - n^2(r)), \]

(3.76)

whereas the foil model of the object is defined by

\[ \Delta_F(r) = 4\pi jR \delta \left( r_z - s(r_x, r_y) \right), \]

(3.77)

It should be remembered that the foil model is strictly only valid for perfect conductors when the reflection coefficient is independent of incidence angle and is equal to unity. However, for instruments with low numerical aperture, the angle of incidence is restricted and it is reasonable to replace the reflection coefficient with its value at normal incidence, \( R = (1 - n)/(1 + n) \) such that

\[ \Delta_F(r) \approx 4\pi j \left( \frac{1 - n}{1 + n} \right) \delta \left( r_z - s(r_x, r_y) \right). \]

(3.78)

Finally, returning to equation (3.71) it is noted that for low numerical aperture, and consequently small angles of incidence, the factor \( |k' - k_r|^2/2(k' - k_r).z \approx k_0 \), and following a similar derivation

\[ \tilde{H}_F(k) \approx \int \int \tilde{G}_{NA}(k_r, k_0) \tilde{G}_{NA}(k - k_r, k_0) d^3k_r k_0 S(k_0) dk_0, \]

(3.79)

Equivalently, the PSF is given by

\[ H_F(r) \approx \int \tilde{G}_{NA}^2(r, k_0) S(k_0) dk_0. \]

(3.80)
Equations (3.76) and (3.78) define the properties of the foil model of the surface that is used in practice and it is straightforward to show that when refractive index is close to unity \( (n \to 1) \) and for perpendicular illumination for which numerical aperture tends to zero \( (N_A \to 0) \) the output interferograms of a CSI that are predicted using the Born approximation and foil model are in fact identical.

3.7 Conclusion

In the previous chapter CSI calibration mechanisms were discussed. As the current calibration procedures are based on calibrating the axial and lateral scales, the tilt related problems often get overlooked. In this thesis the calibration and adjustment of commercial CSI system are described in terms of linear theory where the CSI surface measurement is represented as a linear filtering operation and the filter is characterised by the PSF in the space domain and the TF in the frequency domain. In this chapter the linear theory was derived from the Helmholtz equation and it was shown that for far field imaging, the system can be represented as a linear filtering operation applied to the source term while for full field recording the transfer function is represented by Ewald sphere. As, in practical cases, most of the optical instruments receive light for the solid angle defined by the numerical aperture of the system, the numerical aperture restriction is incorporated in the expression of the transfer function. In a similar way CSI system is described in terms of linear theory and corresponding expression for transfer characteristics is described. Although this linear theory provides a reasonable approximation to practical results, it was derived using the Born approximation which is appropriate only when the object is small or weakly scattering such that there is a small changes in refractive index. However, for the case of surface scattering without considering the Born approximation, the
process is found to be linear if appropriate boundary conditions are assumed. In this case, the Kirchhoff or physical optics approximation together with the assumption of no multiple scattering, the object can be replaced by an infinitely thin foil-like membrane which has been called the “foil model” of the surface. The PSF and TF corresponding to the Born approximation and the foil approximation are calculated and compared. It is shown that the results from both approaches converge to the same result when both the numerical aperture and refractive index change tend to zero.

This chapter also describes the importance of the PSF with respect to system performance. The PSF defines the resolution of the system which is represented by the Nyquist distance. The axial and lateral dimensions of the PSF determine minimum distance the system can resolve in those directions, respectively, and are calculated by the reciprocal of the bandwidth in the respective directions in the frequency domain. The measurement of the PSF of a system will provide overall information about system performance, and comparing the measured PSF with its ideal counterpart will indicate the systematic errors in the system. Therefore, the measurement of the PSF is important in order to calibrate the system in terms of linear theory which will be discussed in the next chapter.

References


8. Duffieux P M, [L’intégrale de Fourier et ses applications a l’Optique] (Rennes) (1946)


10. Schade, O H. “Electro-optical characteristics of television systems”, RCA review, IX:5 (Part I), 245 (Part II), 490 (Part III), 653 (Part IV), (1948)

11. Schade, O H. “Image gradation, graininess and sharpness in television and motion picture systems.” Journal of the Society of Motion Picture and Television Engineers 56.2, 137-177. (1951)


Chapter 4: Measurement of the Point Spread Function

4.1 Introduction

In the previous chapter coherence scanning interferometry (CSI) was represented as a linear filtering operation where the filtering is characterised by the point spread function (PSF) in space domain and transfer function (TF) in frequency domain. A numerical expression for the interferogram output by CSI was derived as a convolution between the object function and the PSF of the system. Although this linear theory provides a good approximation of practical observations, it was derived using the Born approximation [1-5]. This approximation is only valid for weakly scattering objects, where either the object size or the refractive index contrast is small which is rarely true in practice. However for the case of surface scattering where the radius of curvature of the object surface is large compared to the wavelength and the surface is smooth such that multiple scattering can be neglected, the process can also be represented as linear [2]. In this case the object can be represented as an infinitely thin foil like membrane and thus named as ‘foil model’ or ‘foil approximation’. In the last chapter expressions for the PSF, TF as well as output fringe pattern were derived corresponding to the foil model of the object
which will be applied in this chapter to measure the linear characterises of a CSI system.

Though the PSF is by definition the system response to a point object, practically it is not possible to obtain or implement an ideal point object. In this chapter the PSF is derived from the measurement of a known object - a spherical surface. Due to the unavailability of a suitable reference spheres with a surface roughness in the optical scale, measurements were made using mercury droplets [3]. These droplets are easy to make and due to surface tension the upper surface is spherical. The shape and form of the mercury droplets is discussed and a method to calculate the radius of the droplets is also demonstrated. Using the radius information, a foil surface was designed which was utilised to calculate PSF from the recorded fringes. Finally, the measured PSF was compared with the ideal PSF as discussed in the last chapter.

4.2 Theory

In the last chapter it was demonstrated that the PSF defines the resolution of the system and an expression corresponding to the axial and lateral dimensions of the PSF were derived for an ideal system. The measurement of the PSF not only provides the information regarding the resolution, but comparison of the measured PSF with the ideal PSF also indicates the presence of systematic errors in the system. In this section the basic theory of PSF measurement is outlined.

In practice it is very hard to find a point like object or at least one that is significantly smaller than the PSF. For the case of a point like object (e.g. fine powder in suspension), the radius of curvature would be substantially less than one wavelength which will restrict the application of Kirchhoff or physical optics approximation and
thus would violate the validity of the scattering theory discussed in the last chapter (section 3.4.3 condition i)). It is also noted that small particles behave as dipole sources which will make the calculation strongly dependent on the polarisation (which can be otherwise avoided according to section 3.4.3 condition iii)) for the all of the different light components (incident, scattered and transmitted). It is therefore more appropriate to use an extended object that is more representative of those measured.

The calculation of PSF is demonstrated here using the system response of a known object. The calculation is best performed in the frequency domain. In this case the measured TF is given by

$$\hat{H}(\mathbf{k}) = \frac{\hat{O}(\mathbf{k})}{\hat{A}_{foil}(\mathbf{k})}$$  \hspace{1cm} (4.1)

where, $\hat{O}(\mathbf{k})$ is the Fourier transform of the recorded interferogram. $\hat{A}_{foil}(\mathbf{k})$ is the Fourier transform of the designed foil surface of the object while $\hat{H}(\mathbf{k})$ is the computed TF.

### 4.3 Suitable Calibration Artefacts

As described in equation (4.1) the TF can be obtained by dividing the Fourier transform of the foil surface by the Fourier transform of the output response. In order to apply the foil theory in practice, an object, the Fourier transform of which uniformly covers the TF of an ideal system (of the same numerical aperture and bandwidth as calculated in the last chapter) is needed. In this section possible object
options for this measurement will be explored. First a plane surface will be considered.

A plane surface perpendicular to the $z$ direction can be represented as consisting of delta functions in the $xy$ plane. The Fourier transform will be a line ($k_z$ direction) in the spectrum space because,

$$\iint \delta(z)e^{-2\pi j(k_x x + k_y y + k_z z)} \, dx \, dy \, dz$$

$$= \iint e^{-2\pi j(k_x x + k_y y)} \, dx \, dy = \delta(k_x)\delta(k_y)$$

(4.2)

In other words, the Fourier transform of the plane along $xy$ direction is the cross section of the planes corresponding to the $k_x$ and $k_y$ directions which will result in an infinitely thin line along their cross section along $k_z$ as shown in figure 4.1. The plane described in equation (4.2) and figure 4.1 is flat, however if there is a tilt in the
plane in the space domain, a corresponding tilt (in the opposite direction) will be observed in the frequency domain.

The description above shows that the information obtained from the measurement of a single plane is not sufficient to measure the PSF. An alternative option could be a combination of the measurement of several tilted planes that require the plane to be tilted in all possible directions with respect to a particular point, preferably the centroid of the plane. Although this is possible in principle, the centre of rotation must be stable and constant between the measurements. An easier option is to use a spherical reference artefact. A sphere can be thought of as a summation of tilted planes of all different angles and its Fourier transform adequately cover the frequency domain such that \( |\tilde{\Delta}_{foil}(k)| > 0 \) for all \( |\tilde{\phi}(k)| > 0 \).

It is in the primary consideration of the foil theory described in chapter 3 that, the surface have to be optically smooth and for this reason the spherical artefact needed for this purpose must have surface roughness in the order of nm. The form of the sphere is also important, as the form error will directly affect the phase of the measured TF. Thus application and design of the foil model requires a smooth spherical surface with a diameter that enables the sphere to fit in the field of view of the CSI instruments; a) Zygo New View 5000 and b) Taylor and Hobson (TH) CCI. The field of views were 140 \( \mu m \times 105 \mu m \) and 330 \( \mu m \times 330 \mu m \) respectively using a 50x objective (with numerical aperture 0.55). Therefore, a spherical surface of radius 10 to 150 micron will be suitable for both the instruments.

Several spherical artefacts were considered for this measurement. Early candidates were tungsten probes and ruby spheres (similar to what is used in CMM probes [8]) supplied by NPL and TH respectively as shown in figures 4.2a) and 4.2b). The
accuracy of the spherical form of these was considerably less than the nm form required.

4.4 Mercury Spheres

In order to progress the project, it was decided to use mercury droplets to demonstrate the calibration and adjustment procedure.

A detailed analysis on the shape of the mercury droplets is reported by Smithwick III [9]. In that paper microscopic mercury droplets are created by condensation of mercury vapour in partial vacuum. It is described that the shape of the mercury droplet is spherical and gravitational distortion effects are negligible for droplets that have a diameter lower than 500 µm. However, it will be demonstrated in the next section that the top portion of a mercury droplet of 50 µm diameter can be considered as spherical and can be used for calibration purposes.
4.4.1 Calculation of Sphericity of the Mercury Droplets

According to Laplace’s law for liquid spherical droplet the pressure difference $\Delta p$ between the liquid and surrounding atmosphere is given by,

$$\Delta p = \frac{2\gamma}{R}$$

where $\gamma$ is the surface tension of the mercury air interface and $R$ is the radius of curvature. Replacing the values of $\gamma = 486\text{mN/m}$ and $R = 25\mu\text{m}$ the pressure comes out to be $\sim 0.384$ atm. This pressure will cause a mercury droplet to take a spherical shape. However, due to gravity the pressure at the top and pressure at the bottom are slightly different which can distort the shape from the spherical form as shown in figure 4.3.

![Figure 4.3: Mercury droplets (a) ideal and (b) the effect of gravitation distorts the shape](image)

In this experiment the droplet is supported by a glass slide. In this case the nominally spherical droplet has a flat underside. However, the pressure increase at any point due to gravity must be balanced by the pressure induced by surface tension and
consequently the maximum change in radius curvature at any point must be less than the hypothetical case of the “floating droplet” above.

In order to calculate the deviation from the spherical form, the radius of curvature of the mercury droplet at the top and bottom surfaces are calculated as follows.

The pressure at the top of the droplet can be written as

$$P_{\text{top}} = P_0 + \frac{2\gamma}{R_{\text{top}}}$$

(4.4)

where $P_{\text{top}}$ and $R_{\text{top}}$ are pressure and radius of curvature of the top surface of the droplet and $P_0$ is ambient pressure.

The pressure at the bottom $P_{\text{bottom}}$ is

$$P_{\text{bottom}} = P_0 + \frac{2\gamma}{R_{\text{bottom}}} = P_{\text{top}} + \rho g$$

(4.5)

where $R_{\text{bottom}}$ is the radius of curvature at the bottom of the droplet. $h$, $\rho$, $g$ are the height of the droplet, the density of mercury and the gravitational acceleration respectively.

From equation (4.4) and (4.5) it follows that

$$P_0 + \frac{2\gamma}{R_{\text{bottom}}} = P_0 + \frac{2\gamma}{R_{\text{top}}} + \rho g$$

$$\frac{2\gamma}{R_{\text{bottom}}} - \frac{2\gamma}{R_{\text{top}}} = \rho g$$
\[ 2\gamma \left( \frac{1}{R_{\text{bottom}}} - \frac{1}{R_{\text{top}}} \right) = h\rho g \] 

(4.6)

Considering the height \( h \) of the droplet is 50 \( \mu \)m and the radius at the top of the droplet is assumed to be 25 \( \mu \)m and replacing the values of \( \gamma = 486.5 \) mN/m, \( g = 9.8 \) m/s\(^2\), \( \rho = 1.36 \times 10^4 \) Kg/m\(^3\), gives

\[ R_{\text{bottom}} = 24.995 \mu \text{m} \]

For a mercury droplet of about 25 \( \mu \)m radius, the maximum deviation in the top and bottom of the radius is therefore around 5 nm. However, when CSI is used to measure the droplet, due to the finite numerical aperture only a small region of the sphere is measured as shown in figure 4.4. For a lens with NA of 0.55 the top 30 degree of the droplet will be considered.
The change in the height measurement due to the difference in top and bottom of the sphere radius is demonstrated in figure 4.4. The difference in the measured height $\Delta h$ (at the numerical aperture limit) is represented as,

$$\Delta h = R_{\text{top}} \cos \theta - \Delta R - R_{\text{bottom}} \cos \phi$$

where $\phi = \sin^{-1} \left( \frac{R_{\text{top}} \sin \theta}{R_{\text{bottom}}} \right)$. Considering $\theta = 30^\circ$, the height difference will be 0.6 nm.

It should be noted that this is an exaggeration of the form deviation as it stems from the assumption of a sphere of minimum radius $R_{\text{bottom}}$. Thus, it can be concluded that the mercury droplet is spherical with a tolerance of less than a nm within the allowable numerical aperture. In the next section the radius measurement of the droplet is discussed.
4.4.2 Measuring the Radius of the Droplets

Though it is possible to predict the shape of a droplet, its radius is unknown. The wrong radius will lead to a phase error in the measured PSF, which will later be propagated to the adjustment procedure by means of inverse filtering (described in the next chapter). As a result, the wrong radius will introduce an error as described in figure 4.4. For a sphere of radius 25 \( \mu \text{m} \), if the measurement is out by 0.3 \( \mu \text{m} \), then the height error will be 0.0475 \( \mu \text{m} \).

It is easy to measure the diameter of a spherical artifact on a plane by measuring the distance between the top and the plane it is resting on, however for mercury droplets the situation is different. As the droplet makes an angle with glass surface [9], this concept cannot be applied.

An indirect way of measuring the radius for a spherical object was mentioned by Weise et all [10, 11]. It was shown that during the measurement of a spherical surface using confocal microscope a strong signal is obtained for the center position of the spherical surface, along with the one for the top of the surface, as it scanned through. This is because the light incident perpendicular to the spherical surface retraces the same path after reflection, which appears as if the signal is originated from the center itself. Figure 4.5 describes the situation.
As the transfer characteristics (PSF and TF) of CSI operating in a quasi-monochromatic mode are similar to confocal microscopy [1], this method of measuring the radius was applied here to calculate the radius of the mercury droplets.

Figure 4.6 shows the absolute value of the \(yz\) cross section of the fringe pattern recorded by operating CSI in quasi-monochromatic mode. This figure shows the signal due to a glass slide, mercury droplet, and the center of the mercury droplet. A \(z\) section through the middle of the field of figure 4.6 is shown in figure 4.7 which includes packets of fringes at the top and the center position of the spherical surface.
The fringe packet on the left hand side of figure 4.7 corresponds to the centre of the sphere, while the fringe packet on the right hand side is from the top surface. The difference between the two provides the measurement of sphere radius. In this case the radius was found to be 27.45 μm.

Figure 4.6: A yz section through the absolute value of the recorded fringe pattern

Figure 4.7: Plot through the centre of the fringe pattern along the z axis.
There are however several problems with the measurement of the radius of the mercury droplet using this technique:

a. As there is no surface present in the central position of the droplet, the phase of fringes cannot be used to define its position.

b. This is a calibration method, and the information about the radius is crucial in the whole calibration process. The measurement of the radius using the same instrument that is going to be calibrated is not ideal. To prove the validity of the theory, however this method was used initially.

4.5 Measurement of Point Spread Function

Before measuring the spherical artefact the instrument was calibrated using lateral and axial calibration artifacts as discussed in chapter 2. Due to easy access to the Zygo New View 5000 CSI, this instrument was considered first.

The measurement of the PSF is described in a step by step manner as follows,

i) At first, the mercury droplet was placed in the middle of the field of view (140 µm by 105 µm). The corresponding interferogram was recorded. A $yz$ section through the recorded interferogram is shown in figure 4.8. The recorded 3D interferogram is Fourier transformed and a $k_yk_z$ section is depicted in figure 4.9. The figure clearly shows the different terms corresponding to the CSI operation, as described in equation (3.27a) in chapter 3 and figure 3.6.
The Fourier transform of the fringes in figure 4.9 shows that the recorded fringe pattern is dominated with background noise, specially the big white line through the middle of the frequency domain. This is mainly due to the

Figure 4.8: $yz$ section of the real part of the recorded fringe pattern.

Figure 4.9: $k_yk_z$ section of the real part of the FT of the recorded fringe pattern.

ii) The Fourier transform of the fringes in figure 4.9 shows that the recorded fringe pattern is dominated with background noise, specially the big white line through the middle of the frequency domain. This is mainly due to the
sharp transition from light to dark at the edges of the recorded data. In order to remove the noise, the interferogram was multiplied by a Hanning window (figure 4.10) to smooth the sharp transition. The effect of the Hanning window is readily observable in the frequency domain and shown in figure 4.11.

Figure 4.10: \(yz\) section of the real part of the designed Hanning window

Figure 4.11: \(k_yk_z\) section of the real part of the FT of the fringe pattern (after it passed through the Hanning window)
A band pass filter was also applied in the frequency domain, to remove the frequency components that are outside of the defined bandwidth. It was designed so that it follows the ideal TF of the CSI system of same NA and bandwidth and calculated as described in the previous chapter. A $k_yk_z$ section of the bandpass filter and the filtered spectrum are shown in figure 4.12 and 4.13 respectively. After inverse Fourier transforming the filtered spectrum, an overall noise free fringe pattern was obtained as shown in figure 4.14.

Figure 4.12: $k_yk_z$ section of the real part of the band-pass filter.
Using the diameter calculated (as discussed in the section 4.4), the foil model of the surface was designed by considering a series of 1D delta function.

Figure 4.13: $k_y, k_z$ section of the real part of the bandpass filtered spectrum

Figure 4.14: $yz$ section of the real part of the band-pass filtered fringe pattern.

iv) Using the diameter calculated (as discussed in the section 4.4), the foil model of the surface was designed by considering a series of 1D delta function.
following the spherical surface. However, the 1D delta function that defines the foil model of the surface has infinite bandwidth. In order to avoid aliasing in the numerical analysis a 1D Gaussian function was multiplied to define the profile of the surface in the $z$-direction. The position of the modified foil surface was aligned with the interferogram to avoid introducing any phase errors in the TF. The modified foil surface is shown in figure 4.15.

![Figure 4.15: yz section of the absolute part of the foil surface](image_url)

v) The TF was then calculated by dividing the Fourier transform of the interferogram by that of the modified foil. A $k_y k_z$ section of the measured TF is shown in figure 4.16.

vi) Finally, the PSF was calculated by inverse Fourier transformation of the TF. A $yz$ and $xz$ section of PSF is shown in figure 4.17 and figure 4.18.
Figure 4.16: $k_y k_z$ section of the absolute part of the TF

Figure 4.17: $yz$ section of the real part of the PSF
4.6 Observation and Discussion

If the measured transfer characteristics of CSI are compared with their calculated counterparts from last chapter (figure 3.16) the following observations can be made.

a. In the measured transfer function, the spread (bandwidth) in the $k_x$ and $k_y$ directions is shorter than in the calculated TF. This results in a widening of the PSF lateral width and a decrease in lateral resolution of the system (equation 3.40). As described in chapter 3, the fringe formation can be represented as a convolution between the foil surface and the PSF. If the lateral dimension of PSF decreases a more detailed representation of the surface can be found. This situation is comparable to the stylus dimension for stylus profilers. A smaller stylus will follow the shape of the surface more easily which increases the resolution of the system. However, a stylus with
larger dimensions is less able to follow the detailed features on the surface and thus the resolution decreases.

b. It should be noted that there is a maximum surface gradient that can be measured using a CSI instrument. Due to convolutional relationship between the foil surface and the PSF, the output of CSI at a given point can be viewed as the PSF integrated over the surface defined by the foil. If the foil surface is normal to the optical axis it will pass through a single fringe inside the PSF and the absolute value of the surface integral will be a maximum. During the measurement of a surface with a gradient, because of tilt, the foil will pass through several fringes and the surface integral will be substantially reduced. The increase in lateral extent of PSF will reduce it even more, limiting the maximum tilt the system could measure. The situation is described in figure 4.19

![Figure 4.19: Restriction in the measurement of tilted surface due to widening of PSF](image)

c. Measurement of tilt can also be represented in the frequency domain. The Fourier transform of a tilted plane is a tilted line through the frequency space depending on the angle of tilt (as shown in figure 4.1). In the following figure (figure 4.20), the dotted line indicates the ideal TF whereas the continuous line is for the measured TF. As the extent of the measured TF is smaller...
compared to the ideal one, the Fourier transform of some of the tilted planes will lie outside the TF extent, which limits the gradient measured by the system.

![Fourier transform of tilted plane showing it is out of the measured TF, thus it can’t be measured](image)

Figure 4.20: Restriction in the measurement of tilted surface due to narrowing of TF

d. Although the extent of the TF can be used to provide an estimate of resolution and measurable gradient limitations, its phase has an even greater importance if it is used to determine the position of the surface. It is clear from the discussion in the last chapter that the TF of an ideal instrument is a real valued function. In figures 4.21 and 4.22 the real and imaginary parts of the calculated TF are shown. Figure 4.22 shows that the imaginary part of the measured TF is non zero and has a small but significant variation. This is indicative of phase variation and in general will result in a change in both fringe contrast and fringe phase in the interferogram. It will affect the height measurement specifically for an object with different gradients in it. If TF does not have a uniform phase, it will introduce a phase difference in the fringe pattern for an object with varying gradient. It will result in height difference in the surface extracted from the fringes as shown in figure 4.23.
Figure 4.21: $k_y k_z$ section of the real part of the TF

Figure 4.22: $k_y k_z$ section of the imaginary part of the TF
In the figure 4.23, height variation introduced due to the change in the phase of TF is demonstrated. As the Fourier transform of a plane surface (the dotted line) was concentrated to the center of the TF, the phase variation in the TF will not have an effect while measuring it. However for a spherical object the phase difference in the TF will be transferred in the measurement, resulting in erroneous height measurement. This is demonstrated in figure 4.23 b), where the ideal sphere surface is represented using a dashed line while the measured surface from CSI measurement is denoted by a continuous line. In the case of the measurement of a sphere, a $\lambda/2$ height difference will be introduced as a consequence of a phase difference of $\pi$ in the TF.

Figure 4.23: Height variation due to change in phase in TF (a) in frequency domain (b) in space domain.
It is possible to compensate for this error and this process can be thought of as an inverse filtering operation which will be discussed in the next chapter.

4.7 Conclusion

In the previous chapter, CSI was considered as a linear filtering operation. The filtering was characterised by the PSF in the space domain or equivalently the TF in the frequency domain. In this chapter the practical measurement of these functions has been illustrated using a reference artefact - in this case a spherical surface. Several different spherical objects were studied, and for verification of the theory, a mercury droplet was used. It was shown that due to surface tension the top surface of a droplet of mercury of about 50 µm diameter was extremely spherical. The TF of the CSI is measured by dividing the Fourier transform of the recorded fringe pattern with the designed foil surface. Although the measurements look similar to the ideal PSF/TF, the limited lateral extent of TF indicate the restriction on the maximum surface gradient that can be measured using the instrument. Detailed study of TF reveals that there is a slow but significant variation of phase that will change fringe contrast and fringe phase, resulting erroneous surface measurement. In the next chapter the process of compensating the CSI errors using the measured TF will be discussed.

References


Chapter 5: Compensation of Systematic Errors

5.1. Introduction

In the previous chapter the measurement of the point spread function (PSF) from the system response of a mercury droplet was discussed. In this chapter it is shown how this information can be used to improve system response of the CSI system, by means of an inverse filtering technique [1, 2]. Though the conventional inverse filter is able to compensate for the phase errors in the system TF, the amplitude part can potentially decrease the signal to noise ratio of the output. Therefore, a modified inverse filter is designed to produce a flat amplitude and phase response across a well-defined region in the spectrum plane. The filter is applied to the Fourier transform of the recorded interferogram which is inverse Fourier transformed to obtain the modified interferogram. An error surface is calculated by subtracting a reference spherical surface of the measured radius from the measured surface from the modified interferograms. It is later compared with the error surface calculated from the surface obtained from the CSI results. The results show that the error is reduced from about 35 nm to around 3 nm [1, 2].

Shift invariance is the primary assumption for application of linear theory in any system [3]. In order to prove shift invariance of the filter the object (mercury droplet
which was used to calculate PSF in the last chapter) was moved, to the four corners of the field and corresponding fringes were recorded. The modified inverse filter created for the droplet in the center of the field was applied to the interferograms for four different corners. The error surfaces showed similar reduction in the measurement error as observed for the center of the field.

As mercury is not an ideal material to be used as a calibration artefact, it was replaced in this study by silica micro-spheres. The sizes of the microspheres are specified by the manufacturer, but there is not much information available about the shape. The sphericity of the microsphere was verified in this chapter by rotating it through different angles. The modified inverse filter was applied to each measurement and the results establish the repeatability of the whole process.

### 5.2 Inverse Filter

Inverse filtering is a well-known technique that is widely used for image restoration in digital image processing [3]. It can be applied to restore an image which has been blurred by a known linear invariant PSF. It can be interpreted as if the blurred image has been subjected to a linear, space invariant filtering transformation, described as $\tilde{H}_{\text{sys}}(k)$. The image is restored by passing the blurred image through another system whose system response was characterised as $\tilde{H}_{\text{inv}}(k)$

$$
\tilde{H}_{\text{inv}}(k) = \frac{1}{\tilde{H}_{\text{sys}}(k)}
$$

(5.1)
Though the application of the inverse filter is straightforward, it has some well-known problems, especially for the amplitude part of the transfer function (TF) of the inverse filter.

(1) Existence of isolated zeros in the TF will make its inverse ill-defined at those places: As the TF of the inverse filter is reciprocal of the transfer function of the system, the presence of isolated zeros in the transfer function will make corresponding points of the inverse filter undefined.

(2) The inverse filter will boost the noise level of the signal, especially at the places where amplitude of the system TF is low.

For these reasons the use of the conventional inverse filter is limited to correct phase errors in the transfer function. However in this chapter, a modified inverse filter is designed to overcome the problems of amplitude of the inverse filter and obtain a flatter amplitude response. This modified inverse filter is a combination of the phase correction based on the conventional inverse filtering technique, with an additional weighting function associated to overcome the problems in the amplitude part.

5.3. The Modified Inverse Filter

The measured TF of the CSI can be written in terms of phase \( \exp(j\phi(k)) \) and amplitude part \( A(k) \), \( H_{foil}^{meas}(k) = A(k)\exp(j\phi(k)) \). The inverse filter, \( H_{foil}^{inv}(k) \), takes the form,

\[
H_{foil}^{inv}(k) = \frac{\exp(-j\phi(k))}{A_n(k)}, \tag{5.2}
\]

where \( A_n(k) = A(k) / \max[A(k)] \) is the normalised amplitude. In effect, the inverse filter compensates for variations in the phase and amplifies the attenuated parts of
the TF with a gain, \(1/A_n(k)\), in an attempt to flatten the frequency response. For CSI, this has the effect of reducing the lateral dimensions of the PSF and consequently increasing the resolution. In practice; however it is important, to limit the gain as measurement noise (which falls outside the instrument’s response) will otherwise increase the noise level and hence corrupt the signal. In order to overcome the problems of the inverse filter, a modified inverse filter was designed. This filter is a combination of a sigmoid curve and the inverse filter characteristics. It is designed so that it boosts the amplitude of the frequency response of the TF by a factor of 10 where the amplitudes drop to one tenth of its maximum value and then it rolls off to unity. In this case the maximum gain is limited such that the modified inverse filter, \(H_{foil}^{mod}(k)\), takes the form

\[
H_{foil}^{mod}(k) = \frac{\exp(-j\varphi(k))}{W(k)},
\]

(5.3)

where \(W(k)\) is a weighting function [2] defined by

\[
W(k) = (A_n(k) + 0.027)(1 + \exp(-A_n(k) \times 58 + 3.9))/1.5.
\]

(5.4)

The values of the various constants in equation (5.4) are chosen to follow the curve \(1/A_n(k)\) in the region \(0.1 < A_n(k) < 1\) and provide a roll-off to unit gain at \(A_n(k) = 0\). Figure 5.1 shows a plot of the gain of the modified inverse filter, \(1/W(k)\), as a function of normalized amplitude. The values of the coefficients can be varied, which will change the gain curve of the modified inverse filter. Although this profile was not optimized it was found that a maximum gain of 10 was sufficient to flatten
the response over a substantial region of $k$-space (and thereby increasing resolution) without increasing the noise levels significantly.

The modified inverse filter is applied in the frequency domain by multiplying it with the Fourier transform of the interferogram. A modified interferogram was obtained by inverse Fourier transformation of the modified spectrum. Figure 5.2(a) shows the effect of inverse filtering on the spectrum while figure 5.2(b) shows the modified interferogram. Although changes in phase cannot be seen, comparing figure 5.2 (b) with figure 4.14 it can be seen that the amplitude of the fringes is greater at larger tilt angles and the envelope is also reduced. This is mainly due to the gain of the

Figure 5.1: Plot of gain of the weighting function as a function of normalized amplitude.
modified inverse filter boosting the frequencies at the edge of the pass band and, to a limited extent, by the phase correcting properties of the filter.
5.4. Filter Performance

In order to illustrate the benefits of the modified inverse filter, it was first applied to the same droplet that was used to measure the transfer characteristics in the previous chapter (section 4.5). The interferogram was processed as discussed in section 4.5 and the Fourier transform of the fringe pattern was multiplied with the modified inverse filter. The surface height was calculated using frequency domain analysis [4] and an error surface was computed by subtracting the form of an ideal spherical surface of the measured radius. Another error surface was calculated by subtracting the same ideal spherical surface from the surface measurement obtained from the CSI instrument and compared. Figure 5.3(a) shows the error surface obtained from CSI surface measurement which was compared with the error surface calculated from the fringes which was filtered using the modified inverse filter as shown in figure 5.3(b). A height variation in the range of -10 nm to 25 nm and fringe order or $2\pi$ errors at the top and bottom of the droplet (outside of the scale) can be seen in the error surface from the instrument measurement. Using the modified inverse filter it was clear that errors are reduced to the order of ±3 nm and no fringe order errors are apparent.

(b)

Figure 5.2. Modified (a) spectrum ($a k_y, k_z$ section of the real part shown) and (b) $yz$ section fringe pattern.
Figure 5.3: Error surfaces from (a) instrument measurement (b) modified fringes (self-corrected).

So far, the modified inverse filter has been applied to the droplet placed in the centre of the field. A more challenging task is to correct interferograms of known objects placed at different places within the field of view thereby establishing that the calibration and adjustment process is shift invariant. Theoretically, if the system is shift invariant the PSF of the system will be the same throughout the field of view and so is the modified inverse filter. In order to examine the shift invariance of the filter the mercury droplet has been moved and the interferogram was recorded for four corners of the field as shown in figure 5.4. The same filter was applied to all of the four interferograms and corresponding error surfaces were compared.
The error surface from the instrument measurements and the modified fringes for B, C, D and E positions are shown in figure 5.5 and 5.6 respectively. The results are similar to what is obtained in the middle of the field. Another important observation is that the $2\pi$ errors in the polar region (top and bottom) of the error surface of the CSI surface data are significantly reduced for the surface from the modified fringes.
Figure 5.5: Error surfaces from instrument measurements for CBED (clockwise from top left) position as shown in figure 5.4.
Until now, the modified inverse filter has been applied to correct the errors for the droplet which was used to measure the TF of the system. The shift invariance of the designed modified filter has also been verified. The error surface reveals that the filter has reduced the error range from 35 nm to ±3 nm. It is perhaps not surprising that the modified inverse filter is able to correct the measurements of the sphere used to create it, however to prove the general applicability of the filter, it was applied to different sized droplet. Figure 5.7 shows the application of the same modified inverse filter to a droplet of radius 45.4 µm which was about one and half times

Figure 5.6: Error surfaces using the modified inverse filter for CBED (clockwise from top left) position as shown in figure 5.4
bigger than the size of the droplet used to calculate the filter. It can be seen that the error surface (figure 5.7(b)) was similar to the self-corrected surface with an error of about ±3 nm over much of the sphere.

![Figure 5.7: Error surfaces calculated from (a) instrument measurements (b) modified fringes (applying the modified inverse filter used in figure 5.3).](image)

### 5.5. Silica Spheres

Though preliminary experiments using mercury droplets have produced some promising results, there are several problems associated with using mercury. It is not a very stable material and it evaporates over time [5]. For routine calibration, a more stable transferable artefact is required and for this reason a calibration and adjustment protocol for CSI instrumentation based on measurements of silica micro-spheres was investigated. These micro-spheres were ordered from the company Corpuscular inc. and delivered in a vial of de-ionized water. The micro-spheres are specified according to the NIST size standards with diameter 53 ± 1 µm. In case of mercury droplets, it was not possible to determine radius by measuring the distance between the top surface of the droplet and the glass slide. However, for a silica
sphere assuming the sphere is in direct contact with the glass slide, the diameter is measured by calculating the distance between the top of the sphere and the plane of best fit to the substrate surface. Measurement of the sphere dimension in this way has an advantage over the radius measurement described in the last chapter. As it is a differential measurement, the errors due to the CSI instrument will not have an effect in this type of measurement.

In figure 5.8 the schematic representations of the diameter measurement process for a silica sphere. The process is repeated to check the repeatability as well as to determine the mean value of diameter. The diameter of the microsphere used for this work is found to be 53.626 μm ± 0.005 μm.

![Silica microsphere and glass slide](image)

**Figure 5.8: Schematic of radius measurement for silica microsphere**

It was pointed out in the PhD viva that for the case of silica spheres the adhesion force between the sphere and the glass slide could potentially create a deformation in the shape of the sphere at the contact. There are different theories reported to address the contact problem in the presence of adhesion and correspondingly to calculate the deformation [6-9]. A short discussion of the effects of adhesive forces is provided in the appendix. The maximum deformation of the sphere due to adhesive forces is
about 18 nm at the contact region of sphere with the glass surface. Though the contact mechanics provides estimation of the deformation at the contact region the propagation of the contact deformation to the top of the sphere can be explored for silica spheres in future using finite element method. For the calibration purposes, it can be shown (using equation 4.7) that the adhesion forces, may create a maximum height difference of 2 nm at the allowable numerical aperture limit.

After the diameter of the microsphere has been determined, the concentration was focused on verifying the shape of them. The sphericity of an object can be determined by rotating it to various angles. When a spherical object is rotated to different angles, due to symmetry it should always look circular in a two dimensional plane. Sometime if a nearly spherical object having two four or eight fold symmetry is rotated according to a symmetrical angular period, the object can create an impression that it is spherical. In order to avoid that situation, the microsphere was rotated by unequal angular intervals, such as 30, 60, 90, 180, 225 and 315 degrees. A schematic of the rotational procedure is shown in figure 5.9. The fringe pattern corresponding to all different angular positions was recorded. PSF and correction filter was calculated from the sphere at 0 degree and applied to the rest of the interferograms and corresponding error surfaces were calculated by subtracting the modified surfaces from an ideal designed surface.
Figure 5.9: Scheme of rotation (about the z axis) to check the sphericity of the Silica spheres.

Figure 5.10: Error surface for sphere at 0 degree
Figure 5.11: Error surface for sphere at 30 degree

Figure 5.12: Error surface for sphere at 60 degree
Figure 5.13: Error surface for sphere at 90 degree

Figure 5.14: Error surface for sphere at 180 degree
Figure 5.15: Error surface for sphere at 225 degree

Figure 5.16: Error surface for sphere at 315 degree
Figures 5.10 to 5.17 depict the error surfaces for different rotational positions of the sphere. The results show that for all different rotation angles the error surface is circular, which ensures that the shape of the micro-spheres is spherical. However, the other important observation from these results is that the error surface varies for every position. It gives a false impression that the radius being increased until 180 degrees of the rotation and slowly decreases after that. A close observation of the relative position of the micro-spheres in the scan range suggests that the micro-spheres are measured in different parts of the total scan range during different rotations. The corresponding positions of the glass slide and the sphere were noted and shown in Table 5.1. It was found that the sphere was placed at a higher position when it was positioned at 180 degrees, which results in maximum height deviation in
the error surface for that position. This evidence suggests that the piezo electric control of the axial motion is non-linear. A diagram describing the situation is shown in figure 5.18.

Table 5.1: Position of the silica droplets in the scan range

<table>
<thead>
<tr>
<th>Rotation angle (°)</th>
<th>Position of glass slide (µm)</th>
<th>Position of the top of the sphere (µm)</th>
<th>Measured radius in different parts of the z scan range (µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>62.96</td>
<td>5.565</td>
<td>57.395</td>
</tr>
<tr>
<td>30</td>
<td>63.46</td>
<td>5.98</td>
<td>57.48</td>
</tr>
<tr>
<td>60</td>
<td>64.37</td>
<td>6.894</td>
<td>57.476</td>
</tr>
<tr>
<td>90</td>
<td>64.45</td>
<td>6.894</td>
<td>57.556</td>
</tr>
<tr>
<td>180</td>
<td>65.53</td>
<td>7.807</td>
<td>57.723</td>
</tr>
<tr>
<td>225</td>
<td>63.29</td>
<td>5.814</td>
<td>57.476</td>
</tr>
<tr>
<td>315</td>
<td>64.78</td>
<td>7.309</td>
<td>57.471</td>
</tr>
<tr>
<td>360</td>
<td>62.54</td>
<td>4.983</td>
<td>57.557</td>
</tr>
</tbody>
</table>
The information of z non-linearity leads to a new situation, in which the relative position of the micro-spheres becomes important. The non-linearity in the z scale is due to the piezo control of the z stage of this particular CSI instrument: Zygo New View 5000. It was not observable during the calibration of the axial scale as the available calibrated step height is 1.844 µm ± 0.011 µm, which only covers one twentieth of the scan range. A step height of at least half the scan range is required to understand the effect of non-linearity in it. In this case, the sphere acts like a step height and reveals the nonlinearity problem of the instrument.

This problem can be avoided by measuring the object in the same region in z axis. It was verified by measuring a microsphere in the same position for four consecutive
times. The modified inverse filter was applied and corresponding error surfaces are calculated for each measurement and shown in figure 5.19.

Figure 5.19: (a)(b)(c)(d) Error surfaces for repeatability test

The results as shown in figure 5.19 are repeatable. This shows that if the measurements are made in the same position of the z range, the nonlinearity of the piezo scale can be avoided. For the rest of the experiments using the Zygo, it was ensured that all the measurements were done in the same range of z axis.
The other CSI instrument which is available in the University is a Taylor and Hobson (TH) CCI. Zygo New View 5000 and Taylor and Hobson CCI work with same operating principle. Though the software characteristics of the two instruments are different, their optical transfer characteristics are similar. As all of the methods in this thesis are applied to the fringe information, individual results of TH instrument are not shown in each chapter. However, in this case, to show that the non-linearity in the axial scan range occurs only for the Zygo instrument, an experiment was performed and corresponding results are shown. In that experiment a silica microsphere was measured in four different regions of the scan range as shown in figure 5.20. The PSF was measured and correspondingly a modified inverse filter was designed for the microsphere at the lowest position. The modified inverse filter was then applied to the fringe data of the microsphere for all the positions. The corresponding error surface was calculated and shown in figure 5.22. The results show that, though the nonlinearity of the z scale is not generic to every CSI instrument, rather it is problem only for the Zygo instrument.
Figure 5.20: Axial scan range experiment for TH instrument
The TH instrument results shown in figure 5.2 are important because they not only prove the linearity of the axial scale, but also verify the applicability of this theory to calibrate and adjust errors for TH instrument.

Though for the Zygo instrument shift invariance of the modified inverse filter was verified with the mercury droplet, some measurements of the silica micro-spheres are taken to ensure the repeatability of the calibration and adjustment process. The
interferograms are recorded for the B C D and E positions as shown in figure 5.4. The modified inverse filter is applied to fringes for the middle as well as for all the four corners. Error surfaces from the position A are shown and compared with the error surface calculated from the instrument measurements in figure 5.22.

The error surface from the CSI measurement shows a variation of about 35 nm while for the corrected fringe it is about ±3 nm. The results for the corners of the field showed similar trends and are compared in figure 5.23 and 5.24.
Figure 5.23: Error surface from the instrument measurements for CBED (clockwise from top left) position respectively
5.6. Conclusion

In this chapter a method to compensate measurement errors, based on an inverse filter, is proposed. An inverse filter compensates for the phase variation and boosts the amplitude of TF with a gain. However, in practice, due to the reciprocal

Figure 5.24: Error surface using the modified inverse filter for CBED (clockwise from top left) position
relationship of the system TF and inverse filter TF, it can significantly increase the noise level and correspondingly decrease the SNR in the places where the amplitude of system TF is low. This problem was addressed by designing a modified inverse filter which combines the phase compensation of the inverse filter with an additional weighting function to obtain a flatter amplitude response. The modified inverse filter was applied to the fringe spectrum. Comparison of the error surface from both filtered fringes and CSI output showed that application of the inverse filter significantly improved the system performance minimizing the measurement errors from 35 nm to 3 nm.

This method was first applied to mercury droplets that were used to measure the TF as discussed in the previous chapter. Later the procedures described in chapter 4 and in this chapter were repeated to calculate error surface from the silica micro-spheres. In both the cases the similar improvements were observed. Consequently the shift invariance of the correction filter was verified for the Zygo instrument by applying it for the interferogram of the spheres at the four corners of the field. The results of shift invariance are promising, showing similar improvement in the error surfaces.

Another important discovery in this context is the nonlinearity of the z scale in the Zygo instrument. The measurement of the same sphere in different parts of the scan range leads to different results due to z scale non linearity. This might potentially lead to erroneous measurement if measurements of an artefact are performed in different parts of the scan range. For the rest of the measurements using the Zygo instrument it was ensured that, all measurements of any particular artefact are done in the same scan range to avoid ambiguities created due to the nonlinear axial scale. However for the TH instrument, the axial scale is linear. This was verified by measuring a micro-sphere in four different positions in the total scan range which
gives similar results. These results not only prove the linearity of the z scale for TH instrument but also prove the applicability of the modified inverse filter for TH instrument.

References


Chapter 6: Measurement and Adjustment of Distortion in CSI

6.1 Introduction

In the previous chapter the concept of inverse filtering was applied to compensate systematic errors in two CSI instruments a) Zygo New View 5000 and b) Taylor Hobson CCI. Application of this filter is based on linear theory which relies on the assumption of shift invariance. By measuring the error surface of the spheres at the corners of the field of view, it is shown that for the Zygo instrument the filter is shift invariant. The size of the field is important in this respect as the field dependent aberrations become dominant. As the TH instrument has a field of view which is about 3 times the field of view of the Zygo instrument (330 µm X 330 µm with the 50x objective), it is more susceptible to the field dependent errors. Distortion is the primary field dependent aberration, which can be represented as a field dependent magnification. Apart from displacement errors distortion also introduces height errors while measuring tilted artefacts [1, 2].

In this chapter the effect of distortion in the TH instrument is discussed. First a mathematical interpretation of distortion induced height error is demonstrated for a tilted sample and measurement of the shifts introduced in the lateral coordinates due to distortion is described. A shift invariance test (as discussed in section 5.4) is performed for the TH instrument which shows that due to distortion the radius of the sphere is changed in the lateral direction making them elliptic. Finally a method based on de-warping is proposed to compensate these errors.
6.2 Effects of Distortion in CSI measurement

In 1993, Evans [1] showed that measurement of a plane surface with a tilt produced erroneous measurement where the error was similar to comatic aberration. In 2005 similar field and orientation dependent errors were reported by Deck in a patent which appeared during measurement of a plane with a gradient in low coherence interferometry. In that same publication a solution was proposed, in terms of measuring the gradient of each pixel and compensating accordingly with the use of a look up table. In this section it will be shown that this type of error occurs due to distortion.

Distortion is a third order Seidel aberration which is proportional to the cube of the radial component of the field [3]. For the case of a tilted object, it not only shifts the position of the object, but also introduces a change in height as shown in the following section.

Here, an expression is developed to show the distortion induced tilt dependent height variations for a plane having a tilt in only x direction, such as \( z = mx + c \). In polar coordinates \((\rho, \theta)\), defined by \( x = \rho \cos \theta \) and \( y = \rho \sin \theta \), the change in the radial component \( \rho \) due to distortion, \( \Delta \rho \) can be written as,

\[
\Delta \rho \sim D \rho^3 = D(x^2 + y^2)^{\frac{3}{2}} \tag{6.1}
\]

where \( D \) is the distortion coefficient.

The change in \( x \) coordinate \( \Delta x \) due to distortion can be written as

\[
\Delta x = D \rho^3 \cos \theta = \cos \theta D(x^2 + y^2)^{\frac{3}{2}} \tag{6.2}
\]
The height change $\Delta z$ can be written as

$$\Delta z = \frac{\partial z}{\partial x} \Delta x + \frac{\partial z}{\partial y} \Delta y$$

If $z = mx + c$,

$$\Delta z = m \Delta x$$  \hspace{1cm} (6.3)$$

Replacing the values of $\Delta x$ and $\cos \theta = \frac{x}{\sqrt{x^2+y^2}}$, the change in height is,

$$\Delta z = m D x (x^2 + y^2)$$  \hspace{1cm} (6.4)$$

This shows that, for the case of a tilted surface, the change in height due to distortion creates an effect similar to comatic aberration [3] and is illustrated in figure 6.1. In figure 6.1 a) the dotted line represents the original surface and the continuous line signifies the measured surface. Figure 6.1 b) takes a section through the figure 6.1 a) and shows the variations in the measured height due to distortion.
Figure 6.1 (a) Schematic showing tilted plane (dotted line) and measured surface (continuous line), (b) a section through 1(a)
6.3 Measurement of Distortion

In order to measure the displacement due to distortion, a pre calibrated square grid structure which is generally used to calibrate the lateral scales of CSI (as described in chapter 2), was used. The position of the square grids, are measured and the difference in the position provides a direct measure of distortion. Considering the size of the total field of view and dimension of the square grids, the grid with about 30 µm (29.99 ± 0.18) µm pitch was selected for this experiment. The measurement procedure is described below

1. The grid was placed such that 11 x 11 arrays of square grids cover the entire field of view of 330 µm x 330 µm and corresponding fringe pattern was recorded. In order to reduce the electronic noise in the recorded fringe pattern, pre-processing as described in chapter 4, was applied in the frequency domain.

2. The positions of the square grids were found by cross correlating [4] the fringe pattern for the whole field with the fringe for a single square grid. The cross correlation \( C(x, y) \) is defined such that,

\[
C(x, y) = \iiint f(x', y')g^*(x' - x, y' - y) \, dx' \, dy'
\]

(6.5)

where \( f(x, y) \) is a section of the 3D fringe pattern, and \( g(x, y) \) is of same size as \( f(x, y) \) but containing only one square grid as shown in figure 6.2a) and 6.2b) respectively. To avoid edge effects, a single grid was selected and a Gaussian window applied as shown in figure 6.2 b). The correlation result is shown in figure 6.2 c). The cross correlation indicates the position of the
square grids as a prominent bright peak in the respective positions. The lateral \((x, y)\) positions of the grids were calculated by finding the positions of the peak and the results were stored in a matrix for comparison with the ideal position of the square grids.

3. The ideal positions of the square grids were calculated based on the following assumptions. a) The grids are equally spaced throughout and b) the effect of distortion is negligible at the centre of the field of view. The ideal positions were then compared with the measured one, and the difference is plotted in a vector plot with respect to the ideal positions in figure 6.2d).

4. In figure 6.2 d) the arrows point towards the measured positions of the grid from the ideal positions. This vector plot shows the presence of distortion. The effects of distortion on the total field can be observed by plotting the difference between ideal and measured positions of individual grids with respect to their ideal positions as shown in figure 6.3. These are the plots taken through the middle of the field of view along x and y axis. In the middle of the field for around 70 µm radius the distortion is negligible, but at the edges the effects are maximum around 0.5 µm.

As the specification of the grid pattern for the 30 µm grid pattern is 29.99 µm ± 0.18 µm, it can be claimed that the departure from the ideal positions is within the tolerance of the grid pattern. So, this experiment was repeated by translating the grid pattern in \(x\) and \(y\) direction and every time it gives similar results confirming the presence of distortion.
Figure 6.2 (a) a section through the fringes of the square grid pattern, (b) a small square selected for cross correlation (c) results of the cross correlation (d) vector plot showing the difference in the actual position with the measured position (the scale factor for the arrows is 25).
6.4 Quantification of Tilt Related Errors

The tilt dependent height variation as discussed in section 6.2 was measured as described in the following section. A plane surface was tilted to an angle of 2 degree (in $xz$ plane) and corresponding fringe data was recorded. In order to find the effect of distortion on a tilted plane, the surface from the fringe pattern was extracted and a $xz$ section of the same was compared with a straight line of same slope.

A potential source of error in this measurement is the tilt present in the instrument stage which can cause over or under estimation of the tilt dependent effects. In practice it is very hard to achieve the no fringe (or no tilt) condition on the instrument stage to ensure the flatness of the stage. Even if a single fringe is observed on the object surface due to stage’s position, a 5% deviation is introduced in the tilt angle. The tilt angle was measured from the extracted surface by fitting a straight line to the middle of the measured surface (as the distortion effects dominate at the edges of the field). This showed that the tilt angle of the surface was about
1.88 degrees instead of 2 degrees which lies within the predicted 5% tolerance in the tilt angle. The corresponding theoretical value of the tilt dependent height error due to distortion was therefore calculated using the measured tilt angle as discussed in section 6.2 and shown by the blue line in figure 6.4a).

The measured surface was subtracted from the straight line of the measured angle and the difference was compared with the theoretical predicted value. In figure 6.4a) the calculated difference is shown by the red line. The results show direct correlation of the theoretical and experimental values. The experiment was repeated for a tilt angle of 5 degree which is shown in figure 6.4b). The result for 5 degrees also supports the proposed theory in section 6.2. The error plot shows that the curve is similar to the comatic curve as reported by Evans [1]. A height error of ±20 nm is observed for the tilted surface of 2 degrees (measured 1.88 degree). It increases as the tilt angle is increased and for the tilted surface of 5 degree (measured 4.88 degree) the height error is about ±50 nm.

Figure 6.4 Errors from the measurement of tilted plane (a) 2 degrees (b) 5 degrees
6.5 Distortion Compensation

It was mentioned before that in a previous communication a possible solution for tilt dependent erroneous height measurement [1] was addressed by means of a look up table. That method finds the gradient of the surface at each pixel and compensates the heights corresponding to the gradient from data in a look up table. Although, application of this technique may have solved the symptoms of the height errors, the cause of the problem remains unresolved. The correction of the tilt dependent height errors will only change the heights but the positional errors will be unchanged which will result in wrong positional measurement of the surface features. In this section a solution to this problem is proposed using de-warping.

De-warping is a technique used in digital image processing [4] to correct distorted images. It interpolates the coordinates of the distorted data with its ideal equivalent such that,

\[ u = a_1 + a_2 x + a_3 y + a_4 x^2 + a_5 y^2 + a_6 x y + a_7 x^2 y + a_8 y^2 x + a_9 x^3 + a_{10} y^3 \]  
(6.6)

\[ v = b_1 + b_2 x + b_3 y + b_4 x^2 + b_5 y^2 + b_6 x y + b_7 x^2 y + b_8 y^2 x + b_9 x^3 + b_{10} y^3 \]  
(6.7)

where \( u \) and \( v \) are the coordinates of the distorted field and \( x, y \) are the coordinates of the ideal field, and \( a_1 \) to \( a_{10} \) and \( b_1 \) to \( b_{10} \) are the polynomial coefficients.

\[
\begin{bmatrix}
  u \\
  v
\end{bmatrix} = C [1 \ x \ y \ x^2 \ y^2 \ x y \ x^2 y \ y^2 x \ x^3 \ y^3]^T
\]  
(6.8)
where, \( C = \begin{bmatrix} a_1 & a_2 & \ldots & a_{10} \\ b_1 & b_2 & \ldots & b_{10} \end{bmatrix} \)

The distorted and ideal coordinates in equation (6.8), are replaced by the measured positions and ideal positions of the grids to calculate the de-warping matrix ‘\( C \)’ which is later used to interpolate each pixel of the distorted image to obtain the corrected image.

In this way de-warping was applied to the tilted plane measurements and corresponding height error was calculated. Using the 30 \( \mu \text{m} \) grid (described in section 6.3) the ideal and distorted coordinates were measured as described in section 6.3. Using the resulting de-warping matrix the distortion was removed from the measured fringes and the surface height was measured. Figure 6.5 shows that in this way the height errors for 2 degree is reduced from ±20 nm to ±3 nm and for the tilted plane of 5 degrees it is reduced from ±50 nm to ±5 nm.

![Figure 6.5: Error curve after dewarping a tilted plane of (a) 2 degrees (b) 5 degrees](image)
6.6 Effect of Distortion in the Measurement

As mentioned before, distortion will limit the application of the modified inverse filter which was used to adjust CSI systematic errors in the last chapter. Here it is demonstrated using the similar arrangements that were used to check the shift invariance (in figure 5.4) of the modified inverse filter in the last chapter. The measurement of spheres will be affected by the presence of distortion especially at the edges of the field. Distortion will make the spheres slightly elliptical which will result in changing the radius of the sphere in lateral dimension as shown in figure 6.6. The silica sphere was moved to the four corners of the field of view and corresponding interference fringes were recorded. The modified inverse filter was applied on the Fourier transform of the interferograms, and the error surface was calculated. The results are dominated with $2\pi$ errors, so plot through $x$ and $y$ directions are taken for the position ‘B’ and the unwrapped results of the error surface are shown in figure 6.7a) and b).

![Figure 6.6: Effect of distortion on the field of view](image-url)
In order to compensate the effect of distortion, de-warping was applied on the fringe data as described in section 6.5 and the modified inverse filter was applied to the de-warped fringes. The error surfaces from the de-warped fringes are shown in figure 6.7. The concave shape of the uncorrected error surface in figure 6.7 a) and 6.7 b) indicates the radius change due to distortion, while in figure 6.7 c) and 6.7 d) the effect of radius change is minimised after application of de-warping.
6.7 Conclusion

In this chapter the effect of distortion in CSI measurements has been discussed. As distortion related errors are position dependent, a CSI instrument with a smaller field of view, such as the Zygo New View 5000, has negligible distortion effects. On the other hand, the measurement from a CSI with larger field of view, such as Taylor and Hobson CCI, shows a translation of 0.5 µm of the surface features at the edges of the field. Along with that, distortion also introduces a height error while measuring tilted objects. It is shown here that when a tilted plane is measured in CSI with distortion, the results show similar characteristics as that of comatic aberration. Though the tilt related height measurement problems and possible solutions have been addressed by other researchers by means of look up tables, none of them remove the distortion at source. In this chapter the solution is achieved using a digital image processing technique called de-warping. It is shown that using this technique the tilt dependent height errors are reduced from ±20 nm to ±3 nm for a tilted plane of 2 degrees and for 5 degrees from ±50 nm to ±5 nm.

The modified inverse filter which is used to adjust the CSI systematic errors in this thesis is based on the assumption of shift invariance. However, as distortion causes a position dependent magnification it is not appropriate to an instrument with distortion. The measurement of the spheres at four corners of the field shows that distortion changes the radius along x and y direction resulting in an elliptic profile. In order to remove this error, de-warping must be applied to interferogram before using the modified inverse filter. The results show that the error surface obtained from the de-warped fringes have improved and the change in radius due to distortion is typically 5 nm or better.
References

Chapter 7: Discussion and Future Work

7.1 Discussion

Traditional CSI calibration is based on calibration of the machine axis or $x$, $y$ and $z$ axis. The measurement uncertainty is found by measuring a precalibrated step height artefact for axial ($z$) scale and a square grid pattern for the lateral ($x$ and $y$) scales. As the system is not calibrated using tilted artefacts, the tilt related uncertainty is not available. This thesis addresses this problem for the first time.

In order to provide tilt related uncertainty as well as for better understanding of CSI operation, in the third chapter a linear theory of 3D imaging is described where the CSI surface measurement is represented as a linear filtering operation. The filter characteristic is represented by the PSF in space domain and/or the TF in frequency domain. This linear theory representation is based on the Born approximation. Strictly, the Born approximation can only be applied when the incident field is weakly perturbed by the object. In other words, it can be only implemented when the object is small such as when particles suspended in a medium or there is small change in refractive index between the object and surroundings. For the case of surface measurements it is possible to represent the measurement process as a linear operation with the application of surface scattering model where the 3D object is replaced by an infinitely thin foil-like membrane which has been called the “foil model” of the surface. This modification rests on the following assumptions,

1. The Kirchhoff or physical optics approximation
2. No multiple scattering.
Interestingly, the PSF and TF were calculated using both the approaches and the results are compared. It was found that results from both the approaches converge to the same result when both numerical aperture and refractive index change tend to zero.

From a user perspective, the PSF is defined as the system response corresponding to a point object. The lateral and axial dimensions of PSF determine the minimum resolvable distance in corresponding directions. This is inversely proportional to the bandwidth of the TF which plays an important role in determining the maximum measurable slope of the system.

In the fourth chapter the PSF and TF of a typical CSI instrument was measured. This was done by dividing the system response of a mercury droplet with its designed foil surface. In order to apply the foil model to calculate PSF, the object was selected carefully, so that Fourier transform of the object uniformly covers the region corresponding to the ideal TF of the system. Primarily the measurements were done with the mercury droplets as the top surface of a droplet of about 50 µm diameter or smaller is to a high degree of accuracy spherical. Later on silica spheres were used for the experiments. The measured TF was compared with the ideal TF and it was noticed that there is a slow but significant variation in the imaginary part of the TF which will result in an erroneous surface measurement.

Chapter five discussed the compensation of the systematic errors in CSI measurement using a modified inverse filter to overcome this problem. Application of this filter shows that the systematic errors are reduced to around ±3 nm.

During the course of this work experiments were performed in two CSI instruments, Zygo New View 5000, and Taylor Hobson CCI. Though the two instruments operate
on the same operating principles, they have some differences. The Zygo vertical scanning is driven by a piezo electric transducer, which was found to be non-linear. Initially the nonlinearity was not detected during axial scale calibration the height of the step height artefact used was only 1.844 µm (about 1/20th of the total scan range). The non-linearity only became apparent when the spheres were measured in different positions in the scan range. The Taylor Hobson instrument does not have the axial scale problem, however and with this instrument the calibration process was constant over the axial measurement range.

The application of linear theory is based on the assumption of shift invariance. The shift invariance of the filter was verified for the Zygo instrument which has a field of view of 140 µm x 105 µm; by measuring the sphere in four different corners of the field of view. However, for the case of Taylor Hobson instrument the field of view is 330 µm x 330 µm and the field dependent aberration, distortion is found to be dominant. Distortion introduces a position dependent magnification in the measurement. Its effect is maximum at the edges of the field of view and here, it introduced an error of about 0.5 µm in both x and y directions. Distortion was found to have a significant effect on the measured radius of the sphere which was identified when the shift invariance check was applied to the Taylor Hobson instrument. At the edges of the field, the measured radius of the sphere changed, making it elliptical in shape. Along with the translational errors, distortion also causes height error during measurement of a tilted surface. A theoretical model was shown which described that distortion introduces height error similar to the comatic aberration. A method based on de-warping was proposed in order to compensate the distortion related errors. The results showed that application of de-warping had successfully
compensated the position error due to distortion in the lateral scale as well as reduced the height errors for the tilted surfaces.

7.2 Future Work

The work discussed in this thesis forms the basis of a calibration and adjustment procedure for CSI. Though this procedure was tested and verified on an interferogram recorded from two CSI instruments, it is not incorporated directly in either of them. The next step is to integrate calibration using PSF and adjustment with the modified inverse filter in the current system software.

A flowchart that outlines the calibration and adjustment process as discussed in this thesis is shown in figure 7.1. At first calibration should be done in a conventional way that incorporates calibrating step height artefacts and grid pattern artefacts for axial and lateral scales. This should then be followed by the checking the presence of distortion in terms of comparing the actual and ideal position of the square grids. If distortion is present, the de-warping coefficient is calculated and the fringes are de-warped. These steps are sufficient for measurement of plane surfaces. However, the second part of calibration and adjustment, based on PSF/TF measurement, is necessary for the samples with gradients. As the processing needs considerable amount of memory space and time, the user can perform specific calibration procedure depending on the type of surface to be measured.
Another area which can be studied in the future is phase unwrapping which is used to compensate the $2\pi$ errors present in the measurement. For CSI this occurs due to wrong identification of fringe and creates a sudden surface transition equivalent to...
half the mean operating wavelength. It is usually associated with the gradient present in the measured object. In this work, no separate phase unwrapping algorithm was developed, instead the unwrap function in Matlab is used for the results shown in chapter 6. However, the unwrap function can only be applied to one dimensional data. The next step can be application of an unwrapping algorithm that can be applied to the 3D surface profile.

The main challenge in implementation of this approach is to find out a suitable calibration sphere. None of the artefacts which are used to demonstrate the calibration purposes is perfect. For the mercury spheres, measurement of the radius was a problem along with the fact that mercury is an inherently unstable material. Whereas the silica spheres have the adhesion force acting on the contact region which can potentially change the shape of the sphere. It is shown in the appendix that for a silica sphere the deformation in the contact region is 18 nm. However further work is necessary to determine the effect of the contact deformation on the shape of the top of the sphere. In order to apply the spheres for calibration, quick calculations as shown in chapter 4 (equation 4.7) were used to estimate the maximum possible height difference at the numerical aperture limit (approximately 2nm). Although small, this error remains significant and for this reason a more detailed analysis of the propagation of the deformation using finite element method is proposed for future work.
7.3 Publications Arising from the Work in This Thesis

Journal Papers


Conference papers


Appendix

Effect of adhesion forces on the silica spheres

For silica spheres the adhesion force between the sphere and the glass slide could potentially create a deformation in the shape of the sphere at the contact. There are different theories reported to address the contact problem in the presence of adhesion and correspondingly to calculate the deformation of the surface [1-3].

The problem of normal contact without adhesion was first solved by Hertz [2, 4]. For Hertzian point contact the radius of the contact area can be calculated as,

\[ a = \left( \frac{3FR}{4E^*} \right)^{\frac{1}{3}} \]

(A.1)

where, \( R \) is the radius of the silica sphere, \( E^* \) is the combined elastic modulus of the two surfaces and can be written as,

\[ E^* = \left( \frac{1 - \nu_1^2}{E_1^*} + \frac{1 - \nu_2^2}{E_2^*} \right)^{-1} \]

(A.2)

\( E_1, E_2 \) are Young’s moduli and \( \nu_1, \nu_2 \) are the Poisson’s ratios for the two materials in contact respectively. In this case, considering both materials to be silica, it can be written as
Applying the Young’s modulus value for silica \( E = 70 \text{ GPa} \) and Poisson’s ratio \( \nu = 0.2 \), the combined elastic modulus is calculated as, \( E^* = 36.45 \text{ GPa} \)

Here, as there is no external load is applied, the only force \( F \) is the force acting due to gravity,

\[
F = mg = \rho vg = \frac{4}{3} \pi R^3 \rho g
\]

(A.4)

For a sphere of 25 micron radius, the Force \( F \) is calculated as 1.54 nN, and the radius of contact \( a = 9.25 \text{ nm} \). The corresponding deformation is [4]

\[
\delta = \left( \frac{9F^2}{16E^*^2R} \right)^{\frac{1}{3}}
\]

(A.5)

Substituting the values of \( F, E^* \) and \( R \), the deformation is calculated as 3.4 pm.
According to Hertzian contact theory, the deformation of the sphere in the contact area due to its own weight is negligible. However in practice, when two solid surfaces are brought into close proximity, the van der Waals attraction forces start to have significant effect. In 1932 Bradley proposed a model considering Van der Waals forces between two rigid spheres with perfectly smooth surfaces \[5\]. His work can be modified to represent the adhesive forces acting between a spherical surface of radius \(R\) with a rigid plane as [1],

\[
F_a = 4\pi\gamma R
\]

(A.6)

where \(\gamma\) is the surface energy and for uncontaminated glass this value is 2000 mJ/m\(^2\). Substituting the value of surface energy and radius, the adhesion force is found to be 0.6283 mN.

More specific solutions of adhesive contact were reported by Johnson, Kendall and Roberts (JKR theory) in 1971 [6]. According to JKR theory the adhesive force for

![Figure A.1: Deformation of silica spheres due to adhesion](image)
the aforementioned spheres is \( F_{a}^{JKR} = 3\pi \gamma R = 0.471 \, mN \) and the deformation from JKR theory can be calculated as \( \delta^{JKR} = 15.5 \, nm \).

A few years later 1975 Derjaguin, Muller and Toporov (DMT theory) published an alternative adhesive theory [7]. DMT theory assumes that the contact profile remains the same as in Hertzian contact but with additional attractive interactions outside the area of contact. In this case the adhesive force is represented as \( F_{a}^{DMT} = 4\pi \gamma R \) and corresponding deformation for the aforementioned spheres is \( \delta^{DMT} = 18.7 \, nm \).

In 1976 Tabor did a comparative analysis of these two techniques which showed that both the JKR theory and DMT theory are correct and are spatial cases of the general problem [8]. For absolutely rigid bodies Bradley’s theory is valid, whereas for small rigid spheres DMT theory is used and the JKR theory is applied for large flexible spheres.

In essence, it is shown here that, for silica sphere of 25 micron radius, there is a deformation of nearly 18 nm at the contact region. It should be noted that the deformation is maximum at the contact and its effect at the top surface of the sphere is not known to the best of author’s knowledge. However, in the worst case, the deformation is propagated to the top of the sphere (as described it for the mercury sphere in the previous chapter, section 4.4.1) and the height difference at the numerical aperture limit is calculated (using equation 4.7) to be 2.1 nm.

A more detailed analysis of the effect of contact deformation to the top surface of the sphere can be achieved using finite element analysis which can be an area to explore further as future work.
References


