



NanoSurf IV: Traceable measurement of surface texture at the nanometre level

by

Richard Kenneth Leach BSC MSC MINSTP CPHYS MION

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Engineering

University of Warwick, Centre for Nanotechnology and
Microengineering, Department of Engineering

February 2000

CHAPTER 1

THE MEASUREMENT OF SURFACE TEXTURE

"[That] man is the measure of all things."

Protagoras (485 BC)

1.1 INTRODUCTION

Whilst dimensions, shapes and other physical structures of engineering components can be unambiguously specified in terms of length or mass, the property of surface texture has an essentially qualitative aspect. A quantitative measurement can be assigned indirectly only by reference to instrumentation operating in accordance with geometric parameters. Surface texture is only just beginning to become a characteristic that can be exactly described and measured. As the fields of precision engineering and nanotechnology become a reality, the need for high-accuracy, truly measurable surface texture becomes more and more important. Today, most surface texture instruments are indirectly traceable to national standards. The use of transfer standards, such as step heights and sinusoidal gratings, whilst playing an important role in characterising an instrument's performance, usually contributes significantly to the uncertainty of a given measurement.

This thesis describes a body of work that makes a major contribution towards developing a new traceable instrument for measuring surface texture: NanoSurf IV. The novel aspects of this research lie in the marriage of many ideas in precision instrument design with sound metrological principles. To begin, I shall try to give the reader an understanding of what the field of surface texture is about and show why it needs to be measured, present a short history of the field and discuss the latest developments. These introductory sections have been kept short and, hopefully, simple. The main part of the thesis then concentrates on describing the development and characterisation of the NanoSurf IV instrument. To this end many important aspects of surface texture measurement, both experimental and theoretical, have been either omitted or over-simplified. Further investigation into any of the subject areas touched upon in this thesis can be carried out by reading some of the, in my opinion, excellent papers referenced at the end of each chapter.

1.2 WHAT IS SURFACE TEXTURE?

Everything in Nature has some form of surface texture. This applies to all naturally occurring entities, such as the grass growing outside my window, and all artificial entities, such as the computer keys that I am using to type this text. It is very difficult to state definitively what surface texture is, but in most cases it can be described as the structure of a surface when the overall shape of the surface has been removed. An example most easily describes this. A beer-can has the overall shape of a cylinder. However, the metal surface that forms the cylinder has its own shape on a much smaller scale (see figure 1.1).

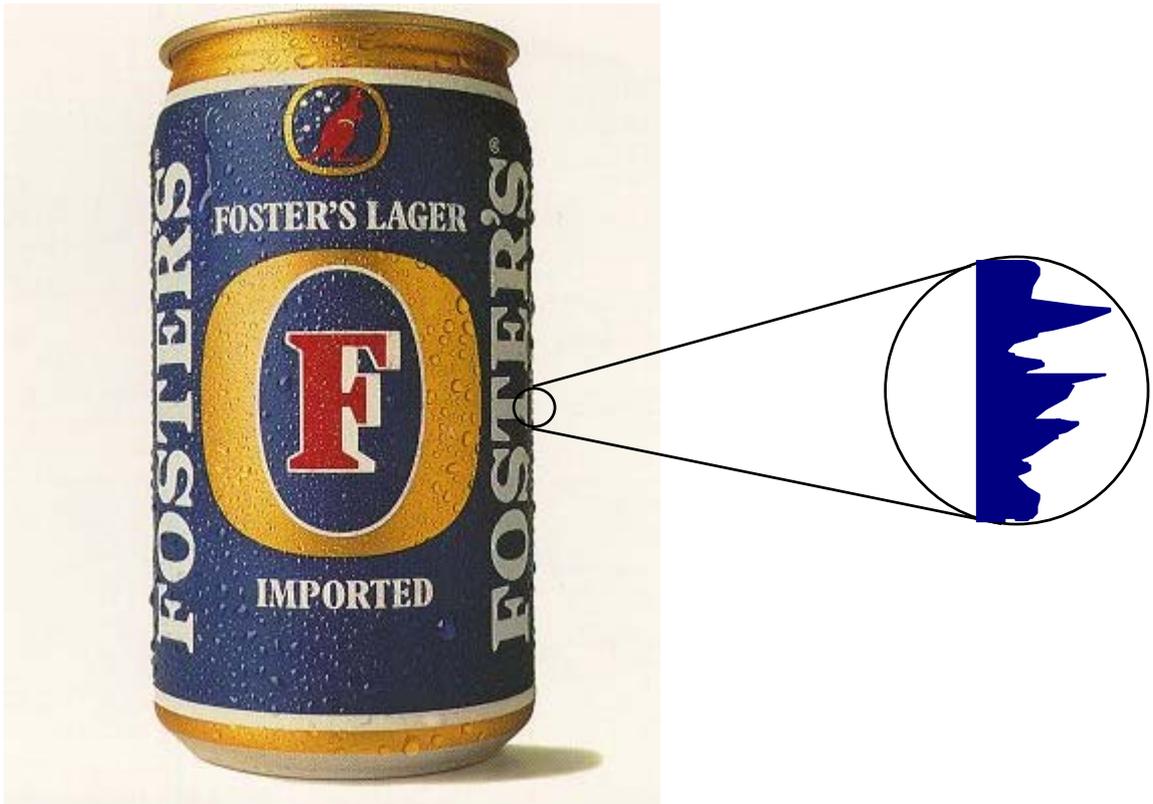


Figure 1.1 The difference between the shape of a surface and its surface texture

The beer-can may feel smooth if one runs a finger across it, but examining this apparently smooth surface under a magnifying device will reveal a complex surface that is a result of the crystal structure of the metal and paint on the surface plus the processes that have been used to manufacture the beer-can. It is the latter structure that is generally referred to as surface texture. Whitehouse (1997) stated that *"Surface geometry is taken to be the departure of real surfaces from their intended shapes. These deviations are caused by the machine process and are called roughness [texture]. Other longer wavelength and shape deviations are caused by errors in the machine tool or sometimes environmental problems. These latter deviations can be avoided, the former cannot!"* This claim could do with the rider "in principle" but is nevertheless illuminating.

The above descriptions of surface texture are not very rigorous and fall down in many examples. This is a factor of the very nature of the subject. When is a given surface structure not part of its texture? Thomas (1999) suggested that the best way to approach the subject is not *"What is the shape of a surface?"*, but rather, *"How did the surface acquire its shape?"* Consider figure 1.2(a) that represents the cross-section of a surface that is theoretically perfectly flat and perfectly smooth (Reason, Hopkins, *et. al.* 1944). Of course, the surface in figure 1.2(a) could never be found in reality as the attainment of perfect flatness is not possible and any real object will be made up of lumpy things called atoms. Now suppose that the surface is slightly deformed as shown in figure 1.2(b), the dip in the middle being 15 μm . If this was a 150 mm diameter optical flat, the un-aided eye will see no imperfections in the flatness of the surface. Next, suppose that instead of being simply curved, the surface is wrinkled as shown in figures 1.2(c) to 1.2(e). The wrinkles, which may or may not be sinusoidal, are to be considered constant in height, but to decrease progressively in wavelength. Again suppose the amplitude is 15 μm . As the wavelength is reduced there will come a time when the eye becomes conscious of a slight ripple in the reflected images of an object containing, for example, straight lines. The ripple will become more and more pronounced as the wavelength is decreased, the magnitude of the effect depending on the slope angle of the undulations. When the wavelength reduces to a few micrometres, the eye will cease to accept the surface as a mirror and regard it as having a sort of shiny roughness. Thus, a smooth surface has been transformed into a rough one simply by changing the wavelength of the undulation, and without changing its height at all. The average heights of the surfaces shown in figures 1.2(c) to 1.2(e) are identical. This begs two questions that will not be answered here: *"Will the surfaces in figures 1.2(c) to 1.2(e) have the same functionality?"*, and *"Would a surface texture measuring instrument give the same reading on these surfaces?"*.

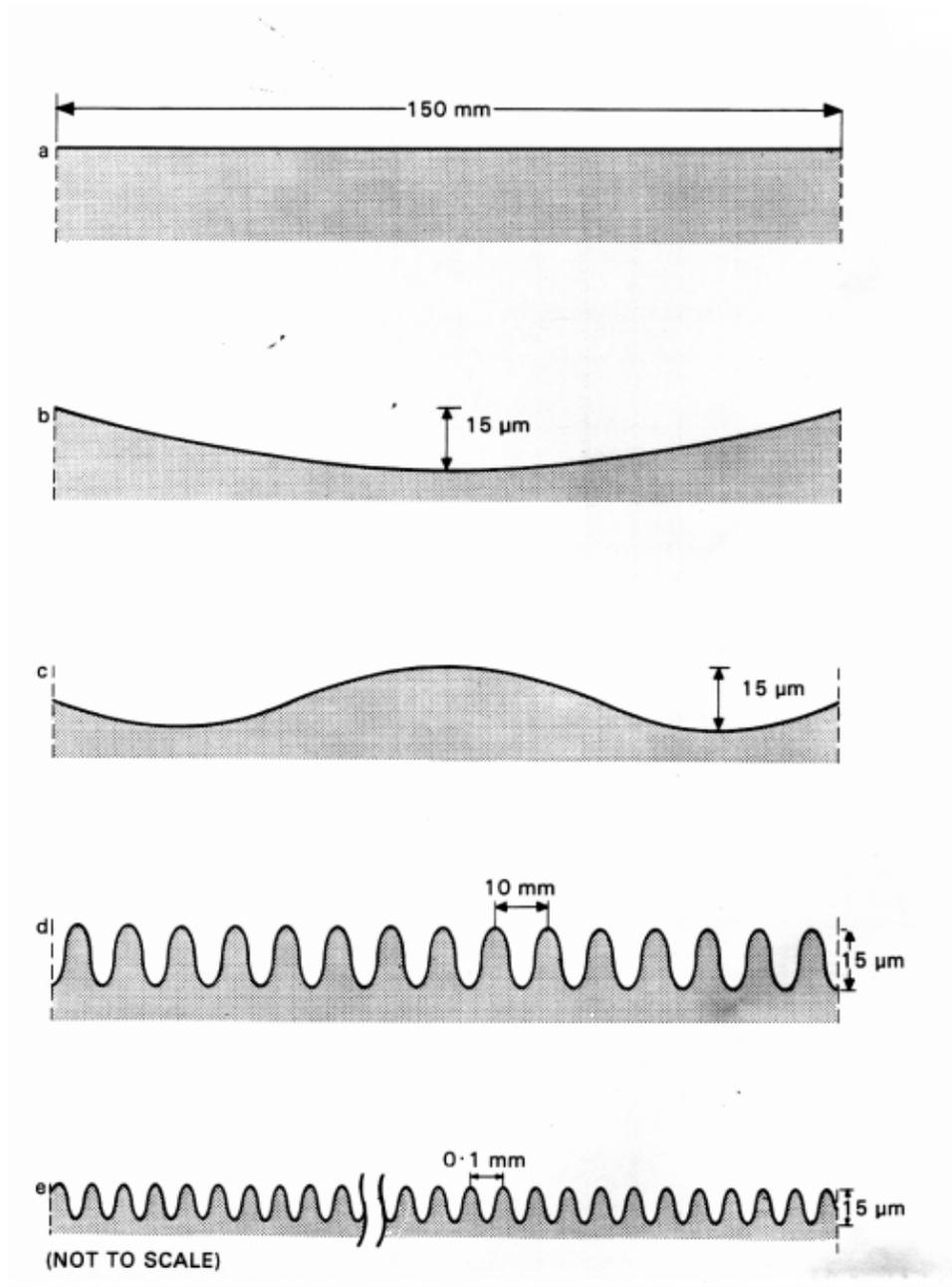


Figure 1.2 Surfaces having the same height structure with different spacings

Thus we see that surface texture can play an important role when considering the functionality of a given object and that surface wavelength undulations are equally as important as surface height variations. For example, it would be extremely difficult to pick the beer-can up if its surface texture were too smooth! When manufacturing any

object various parameters have to be carefully controlled in order to ensure the desired end product meets its given specification. In the field of optics, a mirror must have a surface texture that will give a high enough reflectivity in the wavelength band in which it will be used. Most mirrors require a high degree of specular reflection. A surface feature that is too large can cause unwanted geometrical reflections and a surface feature that is too small can cause light to be diffracted into unwanted angles of reflection. For these reasons it is commonly necessary to control the surface texture of an object during manufacture. To control surface texture means we have to measure it. This leads us to the following question posed in the next section.

1.3 WHY MEASURE SURFACE TEXTURE?

1.3.1 ENGINEERING AND MANUFACTURING

In his keynote speech to the Institute of Mechanical Engineers McKeown (1986) stated that the main wealth creating activity of industrialised nations - and many developing countries, is manufacturing. All control systems require some form of measurement. This applies equally to control in manufacturing. To ensure the quality of a given product, measurements must be applied. Ultra-precision machining with tolerances of the order of a few nanometres is now a reality and only strengthens the argument for manufacturing control. Table 1.1 presents a list of manufacturing tolerances on components for a range of modern products.

Tolerance band	Mechanical	Electronic	Optical
$\pm 5 \mu\text{m}$	mechanical watch parts machine tool bearings gears	electrical relays resistors, capacitors silicon wafer thickness	lenses prisms multimode fibre connectors
$\pm 0.5 \mu\text{m}$	ball and roller bearings hydraulic servo-valves ink-jet nozzles aerostatic bearings	CCD arrays quartz oscillators thin film pressure transducers thermal printer heads	optical scales IC exposure masks monomode optical fibres
$\pm 0.05 \mu\text{m}$	gauge blocks ultra-precision X-Y tables diamond indenter tip radius	IC memories electronic video discs LSI	optical flats diffraction gratings x-ray mirrors
$\pm 0.005 \mu\text{m}$	-	VLSI super-lattice thin films	ultra-precision diffraction gratings

Table 1.1 Manufacturing tolerances of components for a range of modern manufactured products (KcKeown 1986)

Some simple examples of areas where surface texture is controlled and hence measured in traditional manufacturing are:

Machining

The measurement of surface texture is used extensively in manufacturing machining processes because it is very sensitive to changes in either the process or the machine tool. This sensitivity can be used to stabilise production using statistical process control or to simply point out what is going wrong (Whitehouse 1998). In simple cases, the value of the R_a parameter or average roughness (surface texture parameters are summarised in Appendix A) can be used, but where specific information on the process is required the power spectrum of the surface must be measured and analysed. From the power spectrum, process variables such as feed rate, degree of tool chatter, slideway error and machine tool wear can be monitored and ultimately controlled

Tribology

A fluid can interact with a surface via a multitude of physical and chemical means. The effectiveness of a given fluid lubricant will be determined to a large extent by the surface texture over which it flows (Blau 1992). As two surfaces in running contact are at rest they are in surface-to-surface contact or separated by any adsorbed fluid films. As motion begins they become separated by a complete film of lubricant and the coefficient of friction falls up to a point at which increase in speed can cause increased shear in the film. At this point the coefficient of friction begins to rise. So long as a complete film of lubricant exists surface roughness does not greatly influence the wear of the surfaces. However, a complete film is not always maintained for any length of time, as sudden or excessive loads break the film and the surfaces make contact. Contact gives rise to friction that heats and lowers the viscosity of the film, thus making it less able to withstand the

load. Eventually the surfaces will seize. Notwithstanding the importance of the lubricating film and running-in compounds, evidently it is the roughness or texture of the surfaces that ultimately causes a bearing system to fail.

Corrosion

Failure of a bearing or similar engineering material can also be caused by corrosion (Shreir, Jarman, *et. al.* 1995). For example, corrosive products can settle in the bottoms of the valleys, being held there to a great extent by capillary action. The corrosive agent can then begin to undermine the hills, weakening them until they fall away. This process continues and the depth of material removed at each stage of corrosion will depend upon the wavelength of the hills of the original surface texture.

Fracture toughness of materials

Surface texture can also affect the fracture toughness of a machined component. Sharp scratches cause concentration of stress in a material. More material can be used to increase the fracture toughness of a part, but in areas where weight must be a minimum, for example the aerospace industry, it pays to obtain better fracture toughness by giving a better surface finish (Ashby 1999).

The above are just a few examples where surface texture has a profound effect on a finished part and its ultimate function.

1.3.2 MICROELECTRONICS

The current growth of microelectronics has been driven primarily by rapid improvements in price and performance of products ranging from microprocessors and electronic memory to telecommunications devices. These improvements have been made possible by continuous reduction of critical feature dimensions, increases in circuit densities and implementation of robust and reliable processes (Vachet & Young 1995). Circuit features are fast approaching 100 nm width - the limit of classical photolithographic manufacturing techniques (Brittain, Paul, *et. al.* 1998).

Surface	Function	RMS roughness (nm)
Poly Si etch damage	Gate Conductor	> 1.0
Poly Si Rapid Thermal Chemical Vapour Deposition	Gate Conductor	3.0 - 15.7
SiO ₂	Gate Dielectric	0.05 - 0.2
TiN	Adhesion and Barrier	0.5 - 1.0
Al-Cu (1%)	Conductor	8 - 10
Bare Si	Starting material	< 0.1

Table 1.2 Typical RMS roughness values for some critical semiconductor thin films and for the Si substrate (from Sullivan 1995). Note that the cut-off length was not stated.

New techniques are becoming available to combat the 100 nm limit but at these levels of manufacturing accuracy the dimensions of the connecting wires dominate over the

dimensions of the circuit elements and finally, at less than 50 nm quantum mechanical effects will come into play. This continual process of feature miniaturisation requires surface texture to be controlled to an ever higher degree.

Vorburger, Dagate, *et. al.* (1997) discuss using scanning probe microscopy in the microelectronics industry and suggest that there are several types of surface texture measurements that are important for quality control in microelectronic manufacturing. Some of these are shown in table 1.2 (Sullivan 1995). Considerable research in surface texture measurement related to microelectronics is also performed in non-industrial laboratories.

1.3.3 NANOTECHNOLOGY

The above examples where surface texture measurement is required stem mainly from mechanical engineering or manufacturing disciplines. Surface texture measurement does, and will continue, to have a large bearing in the newer fields of atomic manipulation and nanotechnology. It is now common to refer to surface texture measurement in these sub-microscopic regimes as nanometrology. This section, as an example of work in this regime, describes a discipline that is gathering momentum in the research laboratories around the world. For exactly the quality control reasons described in §1.2, surface texture measurement in the nanometre regime will continue to underpin future developments.

Nanotechnology is technology that exploits structures on an atomic and molecular scale. Franks (1987) defines nanotechnology as the “*technology where dimensions and tolerances in*

the range 0.1 to 100 nm play a critical role. This spans the range from the typical radius of an atom to the wavelength of light.”

Everything has physical attributes such as size, shape, mass, strength and stiffness - this includes molecules. Nanotechnology aims to use molecular-scale objects as components of molecular machines in the same way as Nature has developed such molecular machines as enzymes, DNA and proteins. It seems that if we were able to arrange atoms we would gain a better control of the structure of matter and its interaction with the physical universe (Drexler 1990). However, the dreams of the alchemists would still not be realised as the 0.1 to 100 nm scale covers the sciences of chemistry and biology, not that of nuclear physics.

Historically, there have been two approaches to nanotechnology. The first is the top-down approach - miniaturisation. This approach follows the successes of the watchmaker and the integrated circuit manufacturer. In a famous lecture, Richard Feynman (1960) discusses the top-down approach and states *“Why cannot we write the entire 24 volumes of the Encyclopedia Britannica on the head of a pin?”* Microtechnology has evolved largely via a top-down approach to manufacturing such things as computer circuits. But, micro-machining is still in its infancy with micromotors and microgears made from silicon technology only just being exploited in useful devices (see for example Brooks 1999, Knight 1999). The top-down approach to nanotechnology should eventually reach the scale of atomic devices but it only shapes things - it does not give ultimate control of the atomic or molecular world. Control will, probably, be acquired using a bottom-up approach to nanotechnology (Gimzewski 1998).

The bottom-up approach is directly inspired by chemistry and biology. It is the approach that Nature itself seems to have chosen. Proteins and enzymes, *etc.* can be used as structural elements and can be linked to form complex machines with dedicated purposes. Their nature and actions are largely determined by their shape and the way they fold and join. For example, certain protein chains can be made as stiff as steel by arranging the manner in which they fold about each other. Research has also shown that motors and gears, *etc.* can be made using certain artificial polymer chains or artificial proteins.

The advent of the scanning tunnelling microscope (STM) and atomic force microscope (AFM) (Young, Ward, *et. al.* 1972, Binnig, Rohrer 1983, Binnig, Quate, *et. al.* 1986) brought the realisation of atomic manipulation to fruition. In 1992 a group of researchers from IBM carried out the first piece of atomic marketing when they spelt the company's name using xenon atoms (Eiler & Schweizer 1990) and later (Jonathon & Rugar 1992) using laser produced pits and an AFM in the data track of a plastic optical disk (figure 1.3).

However, Drexler (1990) considers atomic manipulation using STM and AFM technology as only a step on the road to true nanotechnology. He states that true control on the nanoworld requires the use of self-replicators. These entities would be produced using much the same technology as that used for crystal growth and chemical synthesis. Drexler's self-replicators would not only be capable of copying themselves, but would also be capable of synthesising other more complex entities. The essential differences between these and other forms of more traditional manufacturing would be the effects of selective-stickiness and thermal motion (Edelstein 1999). For example, in a solution thermal motions shift a typical protein by its own diameter every millionth of a second; fine motions (of an atomic diameter or so in size) happen many billions of times per

second. These constant motions allow molecules to 'explore' their environments quickly, finding any complementary molecules and sticking to them. Manufacturing using Drexler's bottom-up approaches would allow such devices as three-dimensional molecular electronics, effectively grown using self-replicators and proteins (or the like) as scaffolding. The limit here, of course, is not that of the technology but of a law of physics - Heisenberg's uncertainty principle.

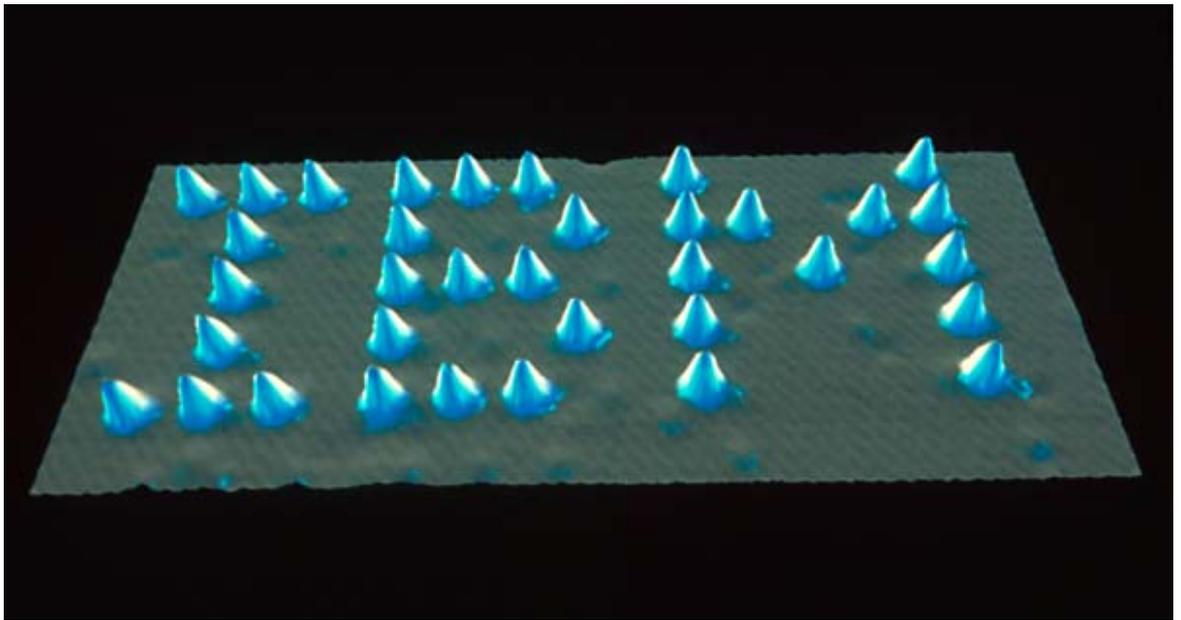


Figure 1.3 IBM's company logo produced using an AFM and a laser (IBM Corporation, Research Division, Almaden Research Centre). The spacing between adjacent pits (which look like peaks) is 200 nm

1.4 A BRIEF HISTORY OF SURFACE TEXTURE MEASUREMENT

Before the turn of the nineteenth century the measurement of surface texture was primarily carried out by making use of our senses of sight and touch. By simply looking at a surface one can easily tell the difference between a freshly machined lump of glass

and one that has been lapped and fine-polished. Touch was utilised by running a finger or finger nail along a surface to be measured and feeling any texture present on the surface. With a few technological modifications, these two methods for measuring surface texture are still the most widely used today.

One of the earliest attempts at controlling surface texture was made in the USA by a company that mounted samples of textures produced by different methods in cases (Shaw 1936) which were given to the machinist, who was expected to obtain a texture on his or her workpiece as near to that specified as possible. This was a suitable method for controlling the appearance of the workpiece but did not in any way indicate the magnitude of the surface texture.

Perhaps the first stylus method was to drag a sapphire needle attached to a pick-up arm across the surface being tested (Harrison 1931). As with a radiogram, the vibration so produced gave rise to sound in a speaker and variation in the electrical current reading on a voltmeter. The method was calibrated by comparing the measured results to those obtained with a sample having a texture that should have been given to the workpiece. This method did not give rise to many benefits over the visual appearance method and it would be expected that the amplitude of the current reading will bear a greater relation to the pitch of the texture rather than its depth. Speed and motion of the sapphire needle plus the frequency response of the pick-up and amplifier influenced the results of this method.

Few metrologists can doubt the influence on the world of surface texture measurement, and indeed the entire field of engineering metrology, played by two brothers named Thomas Smithies Taylor and William Taylor plus their associate William S. Hobson. The

three men went into business in Leicester in 1886 manufacturing optical, electrical and scientific instruments (Hume 1980). In the 1880s, photography was developing rapidly and Taylor, Taylor and Hobson (T,T&H) started making photographic lenses. The company still holds a leading position in the world for cinematograph and television lenses.

The first metrology instrument manufactured by T,T&H was a screw diameter measuring machine (originally designed by Eden at NPL). This instrument was used extensively for armaments manufacture during the First World War. In 1945 J. Arthur Rank, the British flour miller and millionaire film magnate, purchased the shares in the company. Until 1996, Rank Taylor-Hobson was still part of the Rank Organisation.

Richard Reason (1944) who was employed by T,T&H attributed the origin of surface stylus measurements to Gustav Schmaltz of Germany in 1929. Schmaltz (1929) used a pivoted stylus drawn over the surface with a very lightweight mirror being attached to the stylus. A beam of light reflected in the mirror traced a graph on a moving photographic chart, providing a magnified, although distorted, outline of the surface profile. In 1934 William Taylor learned of the work of Abbott and Firestone (1933) in developing methods of measuring surface texture. In their 1933 paper Abbott and Firestone discuss the needs for international standardisation in the surface texture field both in terms of the measuring instruments themselves and in the parameters the instruments measure. They discuss the use of a similar instrument to that of Schmaltz and name it a profilograph. Abbott's instrument was put on the market in 1936. Schmaltz later produced a microscope (known as the light-section microscope) that observed the surface at an angle of incidence of 45° . This gave additional magnification ($\sqrt{2}$) to that of the

microscope but was only suitable for quite coarse surface textures since the optical magnification was necessarily limited.

In the mid-1930s the area where accurate surface measurement was required was mainly in finely finished bearing surfaces, such as those used in aircraft engines. The stylus and mirror arrangement was limited to about 4000× magnification but an order of magnitude more was needed. Therefore, Reason rejected optical magnification and used the principles of a stylus drawn across the surface with a variable inductance pick-up and electronic amplification.

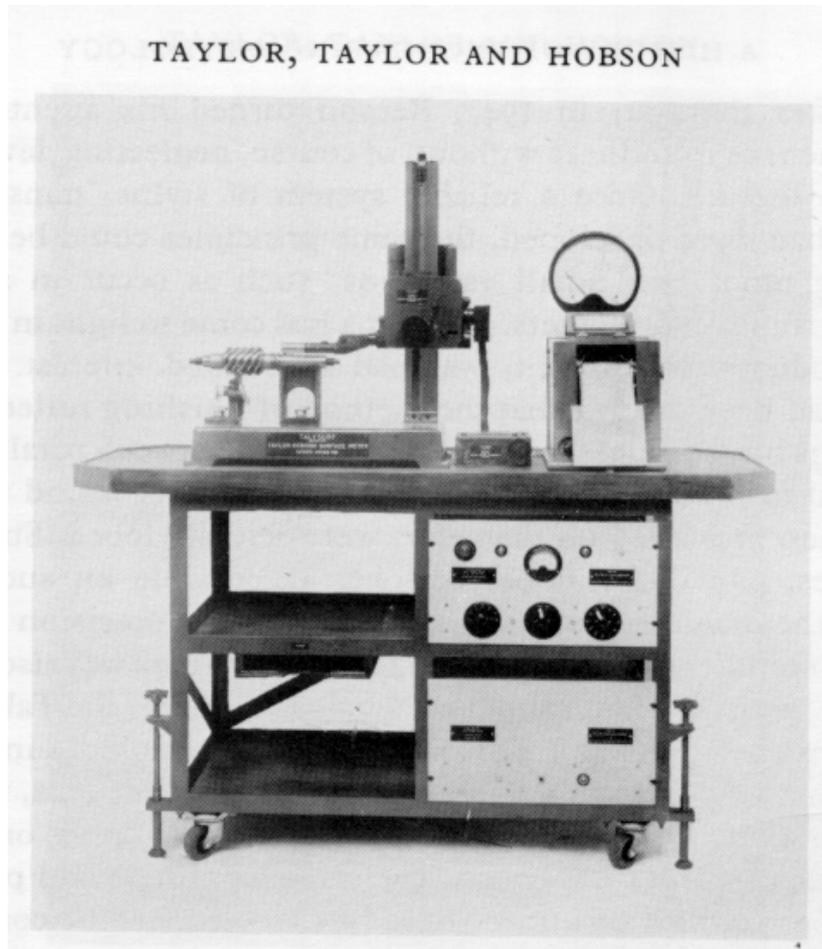


Figure 1.4 The Talysurf from Hume (1980)

Along the lines of Abbott, in 1940 Rolt (at NPL) was pressing for surface texture measurement to produce a single number that would define a surface and enable comparisons to be made. The number most readily obtainable from a profile graph was the average value, obtained using a planimeter. Eventually, T,T&H put the Talysurf onto the market. (Note that the name Talysurf comes from the Latin *talea*, which roughly translates to “measurement”, and not the name Taylor). This instrument provided a graph and the average surface roughness value read directly from a meter. Figure 1.4 is a photograph of the original Talysurf instrument.

Another method of measuring surface texture was due to Linnik of the Mendeleef Institute in Leningrad (1930) and interferometers for this method were made by Hilger and Watts and by Pitter Valve Engineering in Britain. Since the diffraction limit of visible light is half the illuminating wavelength and the interference method depends on the observation of fringe displacement caused by surface roughness, peak-to-peak variations of more than 250 to 500 nm could not easily be distinguished.

In 1947 Reason turned his attention to the measurement of roundness and in 1949 the first roundness testing machine, the Talyrond was produced. The Talyrond used a stylus arm and electrical transducer operating on the same principle as the Talysurf. These two, plus other instruments, paved the way for the Talystep instrument that is essentially the grandfather of the NPL NanoSurf instruments (Reason 1973). The Talystep uses the sensitive electronic transducer technique to measure very small steps or discontinuities in a surface and is thus able to measure thin-film steps of near-molecular thickness. Further developments in surface texture measurement are described in §1.5.

1.5 WHERE ARE WE NOW?

1.5.1 THE PARAMETER RASH

In 1992 Whitehouse coined the term “The parameter rash” to describe the proliferation in recent years of parameters to specify surface texture. Around the 1940s and 1950s it became evident that surface texture could be used to “finger-print” the sequence of operations making up a manufacturing process. The choice of which surface texture parameter should be used varied from country to country. In Great Britain it was the centre-line average, in the USA it was the RMS average and in Europe it was the peak-to-valley parameter. When any variable in a production process is changed, the best way to prevent problems developing is to test the work-piece functionality under the new regime. Unfortunately, this was not always what happened in practice and many variants on the existing surface texture parameters were introduced. This proliferation was increased by the introduction of cheap and powerful microcomputers that could calculate complex parameters with ease.

Industry still suffers from the parameter rash - no universally accepted parameter set is in existence. Also, a measured parameter is very rarely quoted with its associated uncertainty. The need for a comprehensive set of parameters and the background behind the subject of parameters is given by Stout, Sullivan, *et. al.* (1993). A list of two dimensional parameter descriptions is given in Appendix A and the latest proposed three-dimensional parameters can be found in Dong, Sullivan, *et. al.* (1994a, 1994b). Stout, Sullivan, *et. al.* (1993) have attempted to prevent the parameter rash for three-dimensional

surface texture measurement by introducing a set of seventeen internationally recognised parameters covering spatial, amplitude, hybrid and functional parameters.

In 1998 the International Institution for Production Engineering Research (CIRP) carried out a survey of the industrial use of surface texture parameters whereby a questionnaire was circulated to relevant staff in companies asking them which parameters they used (Leach 1999). The data will be used as input to a revision of the relevant ISO standards. The following observations were made, where the ISO parameters are described in Appendix A and the ten-point height is the average height of the five highest peaks and the five lowest valleys within an evaluation length:

- A total of 63 different parameters were considered in the questionnaire.
- 133 industrial companies from 13 countries returned the questionnaire.
- R_a , R_z and the ten-point height R_z are most widely used. Approximately 20% of companies use only R_a . Approximately 40% of companies are covered by R_a , R_z and or the ten point height R_z .
- Approximately 60% of the companies are covered by 10 parameters.
- 51 parameters (81%) are used by less than 5% of the companies. Out of these, 22 parameters (35%) are not used at all by any of the companies.
- 14% of the companies use parameters that are not defined in the ISO standard.

1.5.2 MECHANICAL STYLUS INSTRUMENTS

Stylus instruments are by far the most common instruments for measuring surface texture today. A typical stylus instrument consists of a stylus that physically contacts the surface being measured and a transducer to convert its vertical movement into an electrical signal. Other components can be seen in figure 1.5 and include: a pickup, driven by a motor and gearbox, which draws the stylus over the surface at a constant speed; an electronic amplifier to boost the signal from the stylus transducer to a useful level; and a device, also driven at a constant speed, for recording the amplified signal (Thomas 1999).

The part of the stylus in contact with the surface is usually a diamond tip with a carefully manufactured shape. Owing to their finite shape, some styli on some surfaces will not penetrate into valleys and will give a distorted or filtered measure of the surface texture. Consequently, certain parameters will be more affected by the stylus shape than others. The effect of the stylus shape has been extensively covered elsewhere (see for example McCool 1984, DeVries & Li 1985, O'Donnell 1993, Whitehouse 1994, Dagnall 1998). The effect of the stylus forces can have a significant influence on the measurement results and too high a force can cause damage to the surface being measured. Howard & Smith (1994) and Chetwynd, Liu, *et. al.* (1996) have developed constant force stylus instruments to improve the fidelity between the surface and the stylus tip plus reduce surface damage and dynamic errors.

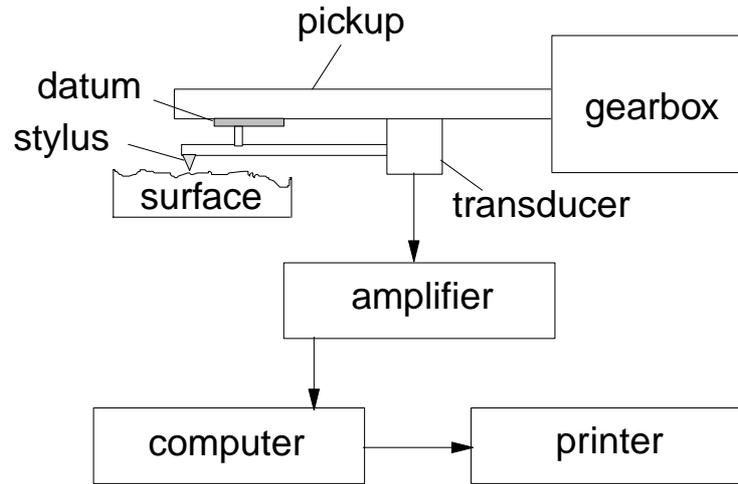


Figure 1.5 Schema of a typical skidless stylus instrument (see text for a description of a skid)

To enable a true cross section of the surface to be measured, the stylus, as it is traversed across the surface, must follow an accurate reference path that has the general profile of, and is parallel to, the nominal surface. Such a datum may be developed by a mechanical slideway, as on the NanoSurf instruments. The need for accurate alignment of the object being measured is eliminated by the surface datum device in which the surface acts as its own datum by supporting a large radius of curvature spherical (or sometimes with different radii of curvature in two orthogonal directions) skid fixed to the end of the hinged pick-up. At the front end of the pick-up body the skid rests on the specimen surface. Alternatively a flat surface or shoe, freely pivoted so that it can align itself to the surface, may be used. Two skids may be used on either side of the stylus.

Whitehouse (1994) describes all the aspects of stylus instruments, including skids, in a great deal detail. The main sources of error associated with a stylus instrument are simply listed below:

- surface deformation;
- amplifier distortion;
- finite stylus dimensions;
- lateral deflection;
- effect of skid or other datum;
- relocation upon repeated measurements;
- effect of filters - electrical or mechanical;
- quantisation and sampling effects;
- dynamic effects;
- environmental effects;
- effect of incorrect data processing algorithms.

Modern stylus instruments regularly obtain measurements of surface texture with sub-nanometre resolutions (Garratt & Mills 1996) but struggle to obtain true traceability of these measurements in each of their axes. A recent comparison of modern industrial stylus instruments in Europe reported alarming spreads in the measurement results (Morrell 1998 and see section 6.6). It is worth pointing out here that many of the pitfalls of mechanical stylus techniques are often highly exaggerated, for example, the wear on the surface caused by a stylus. Whitehouse (1999) rounded off a paper presented at the 1999 LAMBAMAP conference with the statement: *“The stylus technique is not dead or dying. It has virtually just been born!”*

1.5.3 OPTICAL INSTRUMENTS

There are many types of optical instruments for measuring surface texture and only the more popular methods are described in any detail here. Other methods not described below include: heterodyne (Leonhardt, Rippert, *et. al.* 1988), Moiré, holographic, polarisation (Mashimo, Nakamura, *et. al.* 1997), direct Fourier transform (Huynh, Kurada, *et. al.* 1991), ellipsometry (See, Somekh, *et. al.* 1996), fibre optic sensors (Domanski & Wolinski 1992), wavelength scanning (Yamaguchi, Yamamoto, *et. al.* 1998) and surface plasmon resonance (Watts, Sambles, *et. al.* 1997).

Optical methods are generally faster than stylus instruments and offer non-contact profiling. There are many commercially available optical instruments on the market that can give high-resolution three-dimensional scans of a surface. Beautiful images of the surface can be generated in seconds. From the point of view of a metrologist, as with most other surface texture measuring instruments, a question arises: are these images true representations of the surface? Optical instruments can suffer from a number of complications that will not always be apparent when measuring a surface. These include:

- in metals, non-homogeneity (different phases) may give rise to apparent height changes of up to 10 nm due to the effect of the phase change on reflection (Leach 1994);
- in glasses or ceramics local refractive index changes and contaminant films produce nanometre changes (Raine & Franks 1985);
- phase changes at conducting surfaces as a function of incidence angle;
- surface field anomalies caused, for example, by multiple scattering, sharp points acting as secondary sources of light plus edge and wall effects (Whitehouse 1999);

1.5.3.1 Geometrical optics

Light section microscope

In the light section microscope a slit is imaged on to the surface to be measured at an angle of incidence of 45° and viewed by a microscope objective at an angle of reflection of 45° . The reflected image appears as a straight line if the surface is smooth and as an undulating line if the surface is rough. The relative vertical magnification of the profile is the cosecant of the angle of incidence, in this case about 1.4. Resolution is about $0.5\ \mu\text{m}$ and it is quite easy to measure peak-to-valley roughness (Thomas 1999). The origin of the light section microscopes was described in §1.4 and they are still available commercially.

Optical profiler

The most common form of optical profiler is the focus detection instrument, sometimes referred to as an optical follower. As shown in figure 1.6, an illuminating light is reflected by a dichroic mirror, and is focused by an objective to a diffraction limited spot at the focus plane B (Stout 1994). Since the diameter of the spot is very small, the surface is measured by scanning the light spot over the measured surface. If the focused spot can be kept on the top of the surface topography by adjusting the objective or the specimen vertically in the scanning process, or in other words, if the top point of the surface texture being measured is a constant spot in the focus plane in the scanning process, then the vertical dimensional information of the surface topography is determined by the movement of the objective lens or specimen. Since the key problem of the focus detection

technique is to detect focus as sensitively and conveniently as possible, or to make use of the principle of optical focus, many types of focus detection methods have been developed. These methods differ mainly in (i) focus detection mechanisms, (ii) vertical scan method and (iii) horizontal scan method.

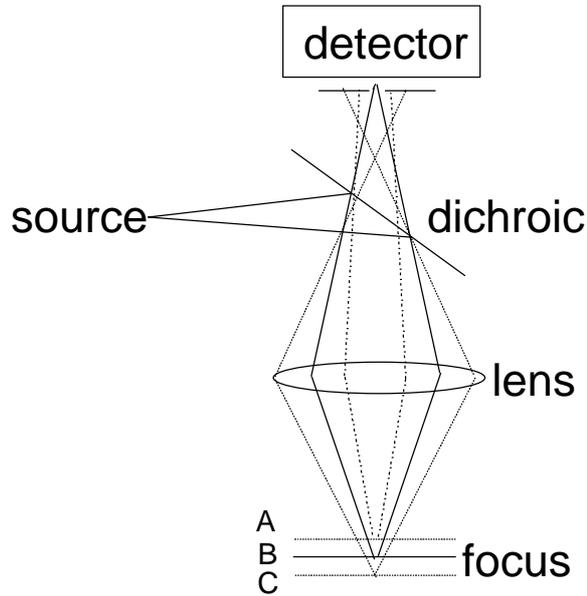


Figure 1.6 Schema of an optical profiler

Confocal microscope

The confocal scanning optical microscope produces images which are formed with light from a limited zone around the focal plane (Gee & McCormick 1992). These images are optical slices or sections through the surface which can be processed to provide non-contact three-dimensional information. For a more detailed description of confocal microscopy the reader is referred to Lichtman (1994)

1.5.3.2 Physical optics

Scattering instruments

Much interest has been shown over the years in the use of scattering for the measurement of surface texture. Several authors have reviewed the previous work (Vorburger & Teague 1981, Bennett 1982, Stover 1990, Olgivy 1991). Three general methodologies have been applied (Griffiths, Bruce, *et. al.* 1995). First, a well-understood area of light scattering theory is used to construct an instrument that will give absolute measurements of surface texture when a number of conditions are met. Second, if a more general theory is applied, a number of assumptions are made that mean no hard and fast connection with a particular parameter can be made, but that the instrument can be used to measure a wide range of surface texture. The third approach is application-specific, *i.e.* an instrument is designed that solves the problem at hand.

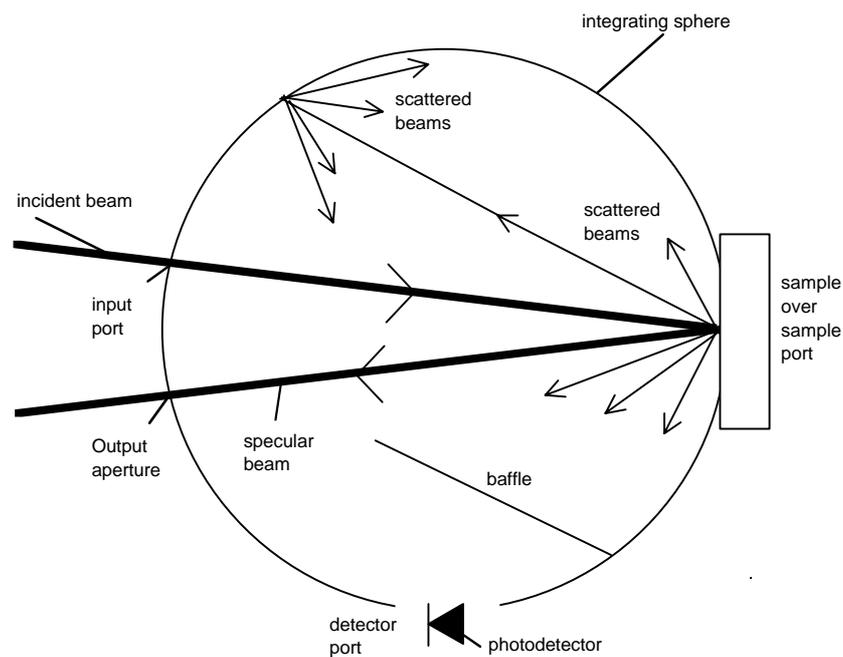


Figure 1.7 Geometry and layout of an integrating sphere

An example of the first type of methodology is the use of an integrating sphere to measure the surface texture of gauge blocks (Leach 1998). The method measures the total intensity of light that is reflected from a gauge surface and compares it with the intensity of light that is diffusely scattered. The ratio of the two intensities is known as the total integrated scatter (TIS) and it is proportional to the square of the RMS surface texture parameter, Rq . The method relies on a number of assumptions that have to be justified for the type of surface being measured. The diffuse component of the reflected intensity is measured by collecting the light in an integrating sphere (Ulbricht 1900). An integrating sphere takes the form of a hollow sphere which has its inside walls coated with a high diffuse reflecting material. Any light that is reflected from the inner surface of the sphere gives rise to a constant flux of electromagnetic power through any surface element of the sphere. The sphere has a number of ports to allow the radiation in and out, the gauge surface to be irradiated and photodetectors to detect the integrated light. Figure 1.7 shows the geometry and layout of a typical integrating sphere.

Interference instruments

Many forms of interferometry are used to measure surface texture, but the most common are interference microscopy, for example the Mirau and Linnik (1930) microscopes, polarisation interferometry (Dyson 1970) plus the full-field methods described by Tolansky (1948). The use of phase-stepping techniques and swept- or multiple-frequency source methods are also becoming popular (Kuwamura & Yamaguchi 1997). There are a number of commercially available phase-shifting interferometric microscopes that give three-dimensional images of surfaces with very short measurement times (Caber 1993, Wyant 1995), but these are subject to the limitations described in the introduction to this

section. Many forms of interference methods are described by Thomas (1999) and Whitehouse (1994).

A sub-nanometre resolution system developed at NPL uses a microscope objective with a numerical aperture of 0.5 in conjunction with a birefringent lens to provide a focused spot from a laser source, with the first minimum in the diffraction pattern 1.54 μm in diameter and a defocused orthogonally polarised beam 10 μm in diameter (Downs, Mason, *et. al.* 1989). Fourier analysis of traces obtained from a number of surfaces indicates that with this objective the system has a surface wavelength range from 0.5 to 15 μm . The trace displayed is an interferometrically obtained path difference of the focused probe beam and the defocused reference beam.

Speckle instruments

When a rough surface is illuminated with partially coherent light, the reflected beam consists in part of random patterns of bright and dark regions known as speckle. The spatial pattern and contrast of the speckle depend on the optical system used for observation, the coherence condition of the illumination and the surface texture of the scatterer (Vorburger & Teague 1981). Speckle has two very important properties: contrast and number of spots per unit area. The contrast can be shown to have a relationship to the correlation length of the surface whereas the speckle density is more concerned with the resolving capability of the imaging system (Whitehouse 1994). Surface information has been obtained using the contrast of speckle patterns produced in the first instance near to the image plane, and second, near to the diffraction or defocus plane. Polychromatic speckle patterns have also been used, as have the correlation properties of

two speckle patterns. From these methods surface texture height values from 0.01 μm to 25 μm have been estimated.

1.5.4 SCANNING PROBE MICROSCOPY

1.5.4.1 Scanning tunnelling microscope

Since its invention in 1972 (see §1.3.3) the scanning tunnelling microscope (STM) has become the most widely used instrument for probing the nano-world around today. The number of applications where an STM is used in some way or another are enormous and are not confined to the research laboratories (Vorburger, Dagate, *et. al.* 1997). The STM has two modes of operation. In the constant current mode, a thin metallic tip is brought close enough to a surface, at a convenient operating voltage (2 mV to 2 V), so that the quantum tunnelling current is measurable. The tip is scanned over the surface while the current between it and the surface is measured (see figure 1.8).

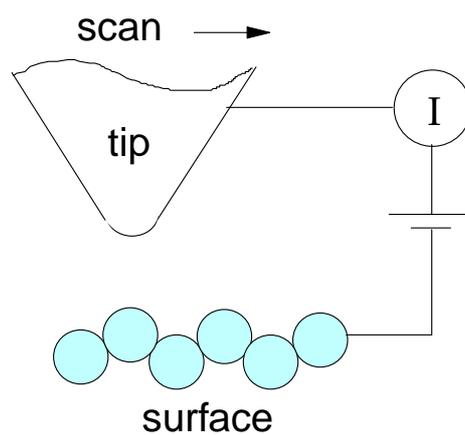


Figure 1.8 Schema of an STM

A feedback network changes the height of the tip to keep the current constant. Since the current varies exponentially with the gap between the tip and the surface, this keeps the gap nearly constant. An image consists of a map of tip height versus lateral position (Hansma & Tersoff 1987). Alternatively, in the constant height mode a tip can be scanned across a surface at nearly constant height and constant voltage while the current is measured. STMs can have resolutions in the vertical and horizontal axes of 0.1 nm or less. Figure 1.9 shows a typical STM image of a quantum corral clearly indicating atomic resolution.

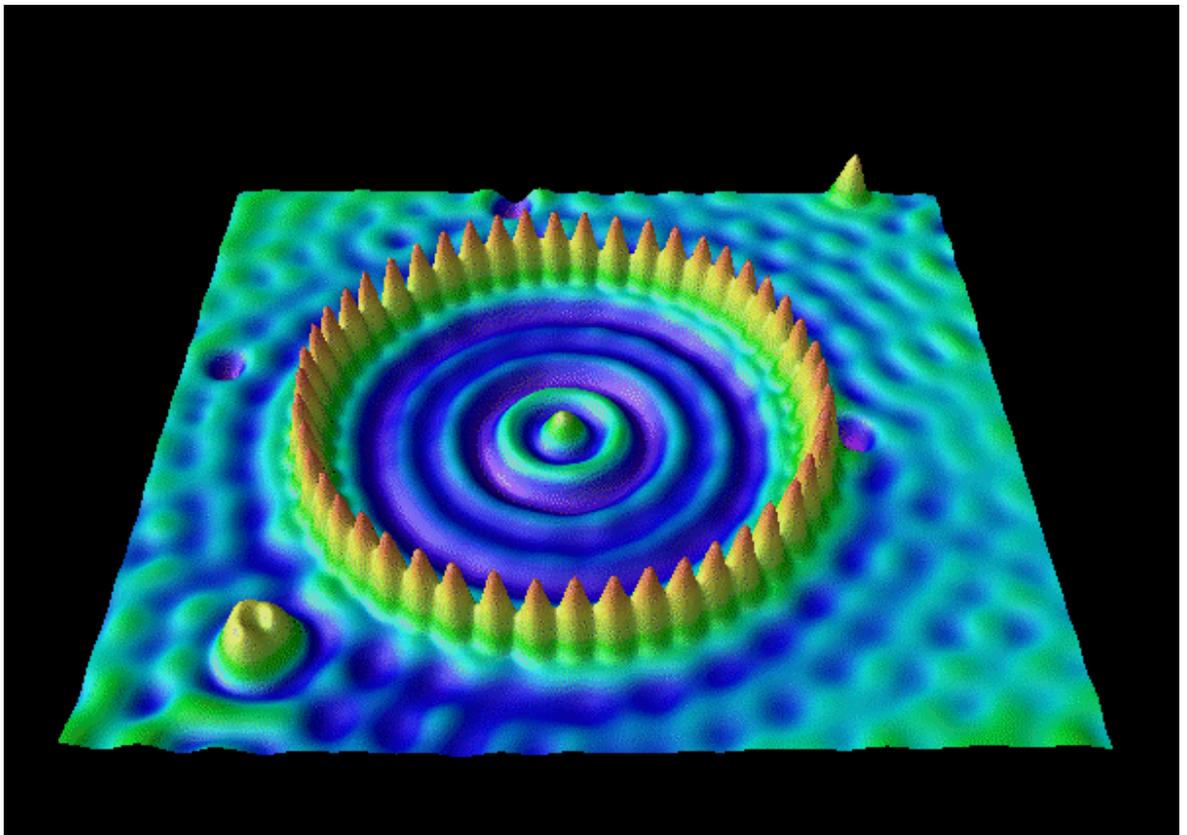


Figure 1.9 This STM image shows surface state electrons on Cu(111) confined to closed structures (corrals) defined by barriers built from Fe adatoms. The radius of the coral is 7.13 nm
(www.almaden.ibm.com/vis/stm/stm.html)

1.5.4.2 Atomic force microscope

One of the drawbacks of the STM is that it can image conducting surfaces only. The atomic force microscope (see §1.3.3) can image all types of surface, albeit with a slight reduction in horizontal resolution. AFM images are obtained by measurement of the force on a sharp tip created by the proximity to the surface of a sample. The tip is usually at the end of a cantilever arm, the deflections of which can be measured using an STM or some optical means. As the tip is raster scanned over the surface the force is kept constant with a feedback network and the tip follows the contours of the surface.

AFMs are widely used for high resolution imaging of many different types of surfaces. The image scales are usually derived from the voltages that drive the piezoelectric actuators that scan either the AFM tip or the sample. These scales can be calibrated from images of calibration transfer standard samples (Garnaes, Kofod, *et. al.* 2000). However, calibration of the scales is not straightforward because the piezoelectric actuators suffer from hysteresis, creep and ageing effects and often are configured to scan in a bending mode that introduces distortion to the image. For metrology applications the accuracy that can be achieved through calibration of the scales derived from the piezoelectric scanners may not be sufficient; in this case it is necessary to add displacement transducers to the scanning axes of the instruments.

At NPL a metrological atomic force microscope (MAFM) has been constructed. Its primary function is to provide calibration of transfer standards for AFMs with

traceability to the national standard of length (Leach, Haycocks, *et. al.* 2000). The instrument is based on a commercial AFM head with displacement transducers added to measure relative displacement of tip and sample in the x , y and z axes. Since direct traceability of measurement was a primary goal for the instrument, laser interferometers were selected as the displacement transducers. The frequencies of the lasers employed in the interferometers are calibrated directly against primary laser standards at NPL, resulting in the shortest possible calibration chain.

1.5.5 AMPLITUDE-WAVELENGTH MAPS

A useful method for the comparison of the capabilities of different measurement techniques is amplitude-wavelength or A-W mapping (Stedman 1987, Franks, Gale, *et. al.* 1988). This map shows how the different measurement devices interact with surfaces by considering their interaction with surfaces that have a sine wave profile. For a given instrument, a region is defined which shows the combination of amplitudes and wavelengths that each different instrument can measure. A-W maps have been plotted for three common surface texture measuring instruments and are shown in figure 1.10. The right hand upper wavelength limit is defined by instrumental parameters such as the range of stage traverse. The left hand wavelength limit is given by the spatial resolution of the stage or data processing system. The high amplitude limit is normally due to some type of physical constraint on the measurement system and the low amplitude limit is given by physical limitations on the vertical resolution of the instrument. The higher

portion of the diagonally sloping upper bound is due to the included angle of the probe (optical or contacting) that is used by the instrument, with the lower portion due to the radius of curvature of the probe tip. A-W maps can be extended in to a third dimension, where the extra dimension is a measure of the strength of the interaction of the probe with the surface.

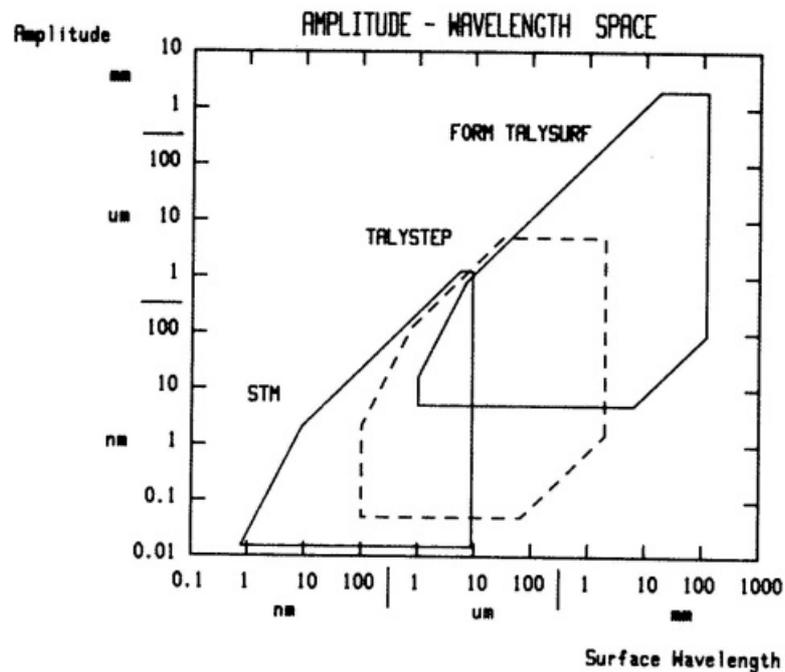


Figure 1.10 Amplitude-wavelength map for three different surface texture measuring instruments

1.6 THE NANOSURF INSTRUMENTS

1.6.1 BACKGROUND

The manufacturing tolerances for X-ray mirrors are extremely demanding, in that departures from the theoretical surface may not exceed a few tenths of a micrometre and

the surface texture should not exceed 1 nm peak to valley (Franks 1977). A central requirement in developing mirror manufacturing techniques and materials, and in assessing and maintaining the quality of the finished product, is the possession of sufficiently sensitive and accurate metrology. In the 1970s and 1980s NPL had a large X-ray Optics Section, headed by Professor Albert Franks, that was concerned with both the manufacture of X-ray components and their metrology (Lindsey 1973, Stedman, Butler, *et al.* 1976, Franks & Gale 1984). At the time the Section was using the Stedman-Stanley profiler (Stedman & Stanley 1979, Wallace, Paul, *et al.* 1981) and a modified Rank Taylor Hobson Talystep (Lindsey 1986).

1.6.2 NANOSURFS I & II

The modified Talystep instrument (the modifications were mainly concerned with environmental isolation of the instrument) did not have the required horizontal scan range so NPL developed NanoSurf I (Lindsey 1986). This instrument incorporated the vertical sensitivity and low noise of the Talystep transducer/stylus system with a precision linear tracking arrangement. Together with an autocollimating laser profilometry system (Ennos & Virdee 1982, Ennos & Virdee 1986), the NanoSurf I provided the necessary measurement range to carry out the metrology on Wolter-type X-ray microscope objectives.

NanoSurf I was followed by NanoSurf II that had a greater horizontal range, improved linear motion accuracy and lower susceptibility to thermal and mechanical fluctuations. Figure 1.11 is a photograph of the NanoSurf II instrument outside its thermal and mechanical isolation systems. Descriptions of the NanoSurf II instrument and its

operation can be found in Lindsey (1986b) and Franks (1992). The NanoSurf II was to become the basis of a technology transfer agreement with Rank Taylor Hobson to develop the Nanostep instrument that is commercially available today (Garratt & Mills 1996).

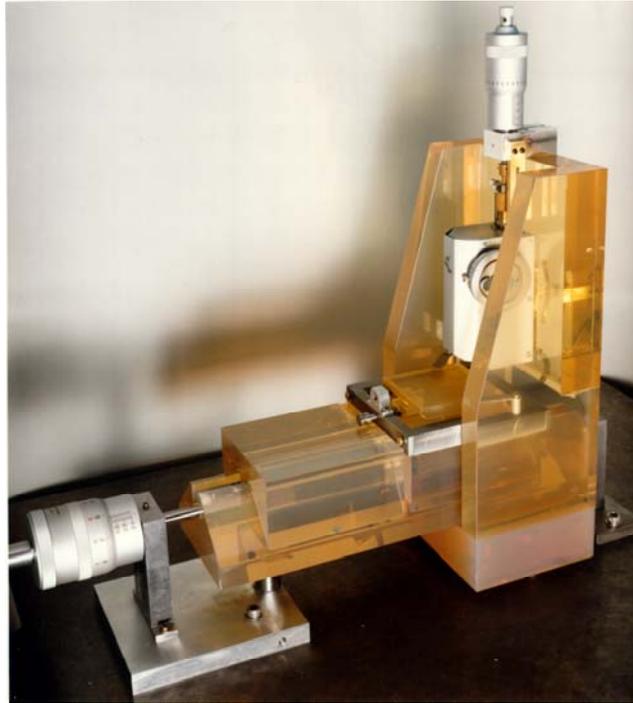


Figure 1.11 The NanoSurf II instrument

1.6.3 NANOSURF III

NanoSurf III was a stylus profilometer having a height range of 5 mm and a scan length of 100 mm that was designed to measure primarily form. The specimen rested on a NanoSurf II-type carriage that was pulled horizontally. The specimen surface was contacted with a 2 mm diameter spherical stylus. The travel of the carriage and the height changes of the stylus were measured using polarisation interferometers. Figure 1.12 is a

photograph of NanoSurf III. Despite excellent design principles and nanometre repeatability it was found that the probing forces that resulted from keeping the stylus arm balanced were too high. This resulted in non-elastic contact between the specimen and stylus and also resulted in the removal of thin films present at the surface. For this reason NanoSurf III is no longer used.

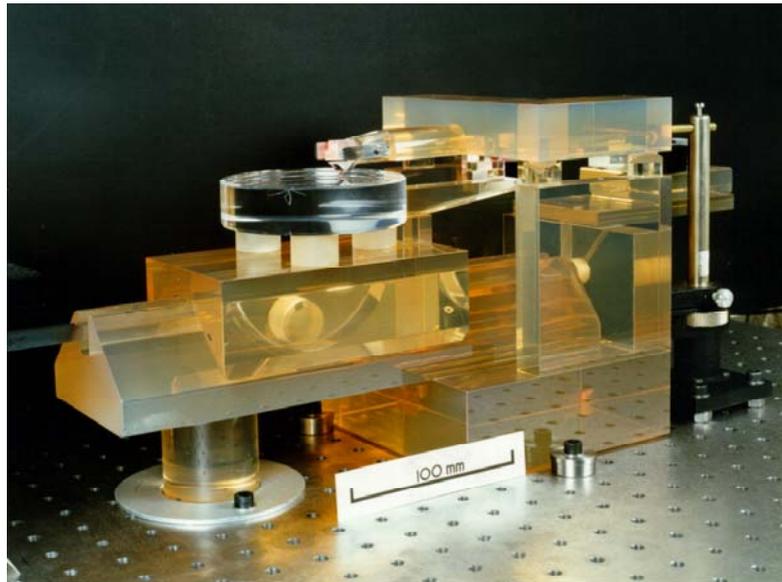


Figure 1.12 The NanoSurf III instrument

1.6.4 NANOSURF IV

The NanoSurfs I and II had one main drawback: lack of traceability and the drawbacks of NanoSurf III have already been discussed. The rest of this thesis describes the design, development and testing of a surface texture measuring instrument, NanoSurf IV, with highly-accurate, fully traceable metrology. Chapter 2 briefly describes the concepts of calibration and traceability in the field of surface texture measurement. Chapter 3 details the design of the mechanical structure, probing system and electronics on NanoSurf IV. Chapter 4 describes the metrology infrastructure of NanoSurf IV. Chapter 5 is an

uncertainty analysis chapter 6 contains the experimental results and chapter 7 presents the discussions and conclusions.

1.7 REFERENCES

- Abbott E J, Firestone F A 1933 Specifying surface quality *Mechanical Engineering* **55** 569-773
- Ashby M F 1999 *Materials selection in mechanical design* 2nd edition (Butterworth Heinemann)
- Bennett J M 1982 Recent advances in surface roughness characterisation *Meas. Sci. Technol.* **3** 1119-1127
- Binnig G, Rohrer H 1983 Scanning tunneling microscopy *Surface Science* **126** 236-244
- Binnig G, Quate C F, Gerner Ch 1986 Atomic force microscopy *Phys. Rev. Lett.* **56** 930-933
- Blau P J 1993 *ASM Handbook 18, Friction, lubrication and wear technology* (ASM International: USA)
- Brittain S, Paul K, Zhao X-M, Whitesides G 1998 Soft lithography and microfabrication *Physics World* **11** 31-36
- Brooks M 1999 Hole in one *New Scientist* **2179** 46-48
- Caber P J 1993 An interferometric profiler for rough surfaces *Appl. Opt.* **32** 3438-3441
- Chetwynd D G, Liu, Smith S T 1996 A controlled-force stylus displacement probe *Prec. Eng.* **19** 105-111
- Dagnall H 1998 *Exploring surface texture* 3rd edition (Taylor Hobson Ltd)
- DeVries W R, Li C-J 1985 Algorithms to deconvolve stylus geometry from surface profile measurements *J. Eng. Ind.* **107** 167-174
- Domanski A W, Wolinski T R 1992 Surface roughness measurement with optical fibers *IEEE Trans. Instrum. Meas.* **41** 1057-1061

- Dong D J, Sullivan P J, Stout K J 1994a Comprehensive study of parameters for characterising three-dimensional surface topography III: Parameters for characterising amplitude and some functional properties *Wear* **178** 29-43
- Dong D J, Sullivan P J, Stout K J 1994b Comprehensive study of parameters for characterising three-dimensional surface topography IV: Parameters for characterising spatial and hybrid properties *Wear* **178** 45-60
- Downs M J, Mason N M, Nelson J C C 1989 Measurement of the profiles of 'super-smooth' surfaces using optical interferometry *Proc. SPIE* **1009** 14-17
- Drexler E K 1990 *Nanotechnology summary* Encyclopaedia Britannica Science and Future of 1990 pp162-177
- Dyson J 1970 *Interferometry as a measuring tool* (The Machinery Publishing Co. Ltd.)
- Edelstein A S 1999 *Nanomaterials: synthesis, properties and their applications* (IOP Publishing)
- Eigler D M, Schweizer E K 1990 Positioning single atoms with a scanning tunnelling microscope *Nature* **344** 524-526
- Ennos A E, Virdee M S 1982 High accuracy profile measurement of quasi-conical mirror surfaces by laser autocollimation *Prec. Eng.* **4** 5-8
- Ennos A E, Virdee M S 1986 Precision measurement of surface form by laser profilometry *Wear* **109** 275 - 286
- Feynman R P 1960 There's plenty of room at the bottom *Engineering and Science* February pp22-36
- Franks A 1977 X-ray optics *Sci. Prog. Oxf.* **64** 371-422
- Franks A, Gale B 1984 Grazing incidence optics for X-ray microscopy *Springer Series in Optical Sciences* **43** 129-138
- Franks A, Gale B, Stedman M 1988 Grazing incidence optics: amplitude-wavelength mapping as a unified approach to specification, theory and metrology *Appl. Opt.* **27** 1508-1517

- Franks A 1987 Nanotechnology *J. Phys. E: Sci. Instrum.* **20** 1142-1451
- Franks A 1991 Nanometric surface metrology at the National Physical Laboratory *Metrologia* **28** 471-482
- Garnaes J, Kofod N, Jørgensen J, Kühle A, Besmens P, Ohlsson O, Rasmussen J B, Lindelof P E, Wilkening G, Koenders L, Mirande W, Hasche K, Haycocks J A, Nunn J, Stedman M 2000 Standards for the calibration of SPMs: design - traceability - calibration - application *Xth Int. Conf. Surfaces* Chemnitz, Germany 277-287
- Garratt J, Mills M 1996 Measurement of the roughness of supersmooth surfaces using a stylus instrument *Nanotechnology* **7** 13-20
- Gee M G, McCormick N J 1990 The application of confocal scanning microscopy to the examination of ceramic wear surfaces *J. Phys. D: Appl. Phys.* **25** A230-A235
- Gimzewski J 1998 Molecules, nanophysics and nanoelectronics *Physics World* **11** 29-33
- Griffiths B J, Wilkie B A, Middleton R H 1995 Surface finish scatters the light *Sensor Review* **15** 31-35
- Hansma P K, Tersoff J 1987 Scanning tunnelling microscopy *J. Appl. Phys.* **61** R1-R23
- Harrison R E W 1931 A survey of surface quality standards and tolerance costs based on 1929-1930 precision-grinding practice *Trans. A.S.M.E.* paper no. MSP-53-12
- Howard L P, Smith S T 1994 A metrological constant force stylus profiler *Rev. Sci. Instrum.* **65** 892-902
- Hume K J 1980 *A history of engineering metrology* (Mechanical Engineering Publications Ltd) ch. 8
- Huynh V M, Kurada S, North W 1991 Texture analysis of rough surfaces using optical Fourier transform *Meas. Sci. Technol.* **2** 831-837
- Knight J 1999 Dust mite's dilemma *New Scientist* **2180** 41-43

- Kuwamura S, Yamaguchi I 1997 Wavelength scanning profilometry for real-time surface shape measurement *Appl. Opt.* **36** 4473-4482
- Leach R K 1994 *Development of an instrument to measure the phase change at reflection* (MSc thesis: Brunel)
- Leach R K 1998 Measurement of a correction for the phase change on reflection due to surface roughness *Proc. SPIE* **3477** 138-151
- Leach R K 1999 Rationalising ISO surface texture parameters *Quality Today* April pp6
- Leach R K, Haycocks J A, Jackson K, Lewis A J, Oldfield S, Yacoot A 2000 Advances in traceable nanometrology at the National Physical Laboratory *euspen Warwick/Tokyo Nanotechnology Symp.* to be published
- Leonhardt K, Rippert K-H, Tiziani H J 1988 Optical methods of measuring rough surfaces *Proc. SPIE* **1009** 22-29
- Lichtman J W 1994 Confocal microscopy *Scientific American* August 30-35
- Lindsey K 1973 The production and topography of X-ray optical components *Proc. X-ray Optics Symp. Mullard Space Science Lab.* 101-121
- Lindsey K 1986a The assessment of ultra smooth substrates and overcoatings *Vacuum* **25** 499-502
- Lindsey K 1986b Sub-nanometre surface texture and profile measurement with NANOSURF 2 *Ann. CIRP* **37** 519-522
- Linnik W 1930 Ein Apparat zur Messung von Verschiebungen in der Sehrichtung *Z. Instrum.* **50** 192-
- Mamin H J, Rugar D 1992 IBM's new ultra dense data storage method *Appl. Phys. Lett.* **8**
- Mashima K, Nakamura T, Tanimura Y 1997 Development of optical noncontact sensor for measurement of three-dimensional profiles using depolarized components of scattered light *Appl. Opt.* **36** 227-234

- McCool J I 1984 Assessing the effect of stylus tip radius and flight on surface topography measurements *Trans. ASME* **106** 202-209
- McKeown P A 1986 High precision manufacturing and the British economy, James Clayton Lecture *Proc. Inst. Mech. Engrs.* **200** 1-19
- McKeown P A 1987 High precision manufacturing technologies in the 21st Century *Ann. CIRP* **36** 495-501
- Morrell R 1998 SMT Project "CERANORM" SMT4-CT96-2078 Mechanical tests for advanced technical ceramics *NPL Report CMMT (D)* **169** 1-64
- O'Donnell K A 1993 Effects of finite stylus width in surface contact profilometry *Appl. Opt.* **32** 4922-4928
- Ogilvy J A *Theory of wave scattering from random rough surfaces* (Adam Hilger: Bristol)
- Raine K W, Franks A 1985 Effects of variations in refractive index on the interferometric measurement of microtopography in glasses **32** 251-254
- Reason R E, Hopkins M R, Garrod R I 1944 *Report on the measurement of surface finish by stylus methods* (Taylor, Taylor & Hobson: Leicester)
- Reason R E 1973 Stylus methods of surface measurement *Bull. Inst. Phys.* Oct. 587-589
- See C W, Somekh M G, Holmes R D 1997 Scanning optical microellipsometer for pure surface profiling *Appl. Opt.* **35** 6663-6668
- Schlesinger G 1942 Surface finish *Inst. Prod. Engrs. Report* Ch 1
- Schmaltz G 1929 Über Glätte und Ebenheit als physikalisches und physiologisches Problem *Zeitschrift des Vereines deutscher Ingenieure* **73** 1461-
- Schreir L L, Jarman R A, Burstein G T 1995 *Corrosion* (Butterworth-Heinemann Ltd)
- Shaw H 1936 Recent developments in the measurement and control of surface roughness *J. Inst. Prod. Engrs.* **15** 369-391
- Stedman M, Butler D W, English T H, Franks A, Gale B, Hutley M C, Stuart P R 1976 Recent developments in X-ray diffraction gratings at NPL *Space Sci. Instrum.* **2** 355-362

- Stedman M, Stanley V W 1979 Machine for the rapid and accurate measurement of profile
Proc. SPIE **163** 99-102
- Stedman M 1987 Mapping the performance of surface-measuring instruments *Proc. SPIE*
803 138-142
- Stout K J, Sullivan P J, Dong W P, Mainsah E, Luo N, Mathia T, Zahouani 1993 *The development of methods for the characterisation of roughness in three dimensions* EUR 15178 EN (Commission of the European Communities)
- Stout K J 1994 *Three dimensional surface topography; measurement, interpretation and applications* (Penton Press: London)
- Stover J C 1990 *Optical scattering* (McGraw-Hill)
- Sullivan N T 1995 Current and future applications of in-line atomic force microscopy for semiconductor wafer processing *Summary report: Second workshop on IASPM*, NISTIR 5752 NIST, USA
- Thomas T R 1999 *Rough surfaces* (Imperial College Press)
- Tolansky S 1948 *Multiple-beam interferometry of surfaces and films* (Oxford University Press)
- Ulbricht R 1900 Die Bestimmung der mittleren räumlichen Lichtintensität durch nur eine Messung *Electrotechnische Zeitschrift* **29** 595-601
- Vachet G, Young M 1995 Critical dimension atomic force microscopy for 0.25 μm process development *Solid State Technology* **12** 57-62
- Vorburger T V, Teague E C 1981 Optical techniques for online measurement of surface topography *Prec. Eng.* **3** 61-83
- Vorburger T V, Dagate J A, Wilkening G, Iizuka K, Thwaite E G, Lonardo P 1997 Industrial uses of STM and AFM *Ann. CIRP* **46** 597-620
- Watts R A, Sambles J R, Hutley M C, Preist T W, Lawrence C R 1996 A new optical technique for characterising reference artefacts for surface profilometry *Nanotechnology* **8** 35-39

- Wallace C A, Paul D F, Lindsey K 1981 Production and evaluation of supersmooth x-ray synchrotron mirror surfaces *Proc. SPIE* **315** 186-192
- Whitehouse D J 1982 The parameter rash - is there a cure? *Wear* **83** 75 - 78
- Whitehouse D J 1993 Horns of metrology *Quality Today* November 22-23
- Whitehouse D J 1994 *Handbook of Surface Metrology* (Institute of Physics Publishing)
- Whitehouse D J 1997a Surface metrology *Meas. Sci. Technol.* **8** 955-972
- Whitehouse D J 1997b Conditioning of the manufacturing process using surface finish *Proc. Laser Metrology and Machine Performance III LAMBDAMAP '97* 3-10
- Whitehouse D J 1998 Surfaces: an essential link in nanotechnology *Nanotechnology* **9** 113-117
- Whitehouse D J 1999 Surface measurement fidelity *LAMBDAMAP* 267-276
- Wyant J C 1995 Computerized interferometric measurement of surface microstructure *Proc SPIE* **2576** 122-130
- Yamaguchi I, Yamamoto A, Yano M 1998 Surface topography by wavelength scanning interferometry *Proc. SPIE* **3479** 24-33
- Young R, Ward J, Scire F 1972 The Topographiner: An instrument for measuring surface microtopography *Rev. Sci. Instrum.* **43** 999-1011

CHAPTER 2

TRACEABILITY AND CALIBRATION

"He gave man speech, and speech created thought, which is the measure of all the universe."

Shelley (1820)

2.1 THE CONCEPT OF TRACEABILITY IN MEASUREMENT

Traceability is defined in BS 5233 (1986) as *The property of a result of a measurement whereby it can be related to appropriate standards, generally international or national standards, through an unbroken chain of comparisons.* Traceability is easier to describe if one considers a well established reference artefact such as the gauge block. A gauge block is a block of rectangular section, made of durable material, with one pair of plane, mutually parallel measuring faces. The length of a gauge block at a particular point on the measuring face is the perpendicular distance between this point and a rigid plane surface of the same material and surface texture upon which the other measuring face has been wrung (BS 4311 1993). 'Wringing' is a technique by which very flat, lapped surfaces can be made to adhere to one another by molecular attraction (Leach, Hart, *et. al.* 1999).

The length of a gauge block is measured in terms of the primary standard of length, *i.e.* the vacuum wavelength emitted by the primary iodine-stabilised He-Ne laser (see chapter 4). Gauge blocks are then used to calibrate the lengths of other standards through

comparison or used to verify the performance of length measuring instruments. This hierarchical system of standards represents the *traceability* of length measurements - in theory any measurement of length can be traced to the definition of the metre. By the nature of the hierarchical structure and loss of accuracy at each comparison stage, it is obvious that the higher in the measurement pyramid, the more accurate must be the standard, with the primary standard being the most accurate. Part of the traceability structure for length is shown in figure 2.1.

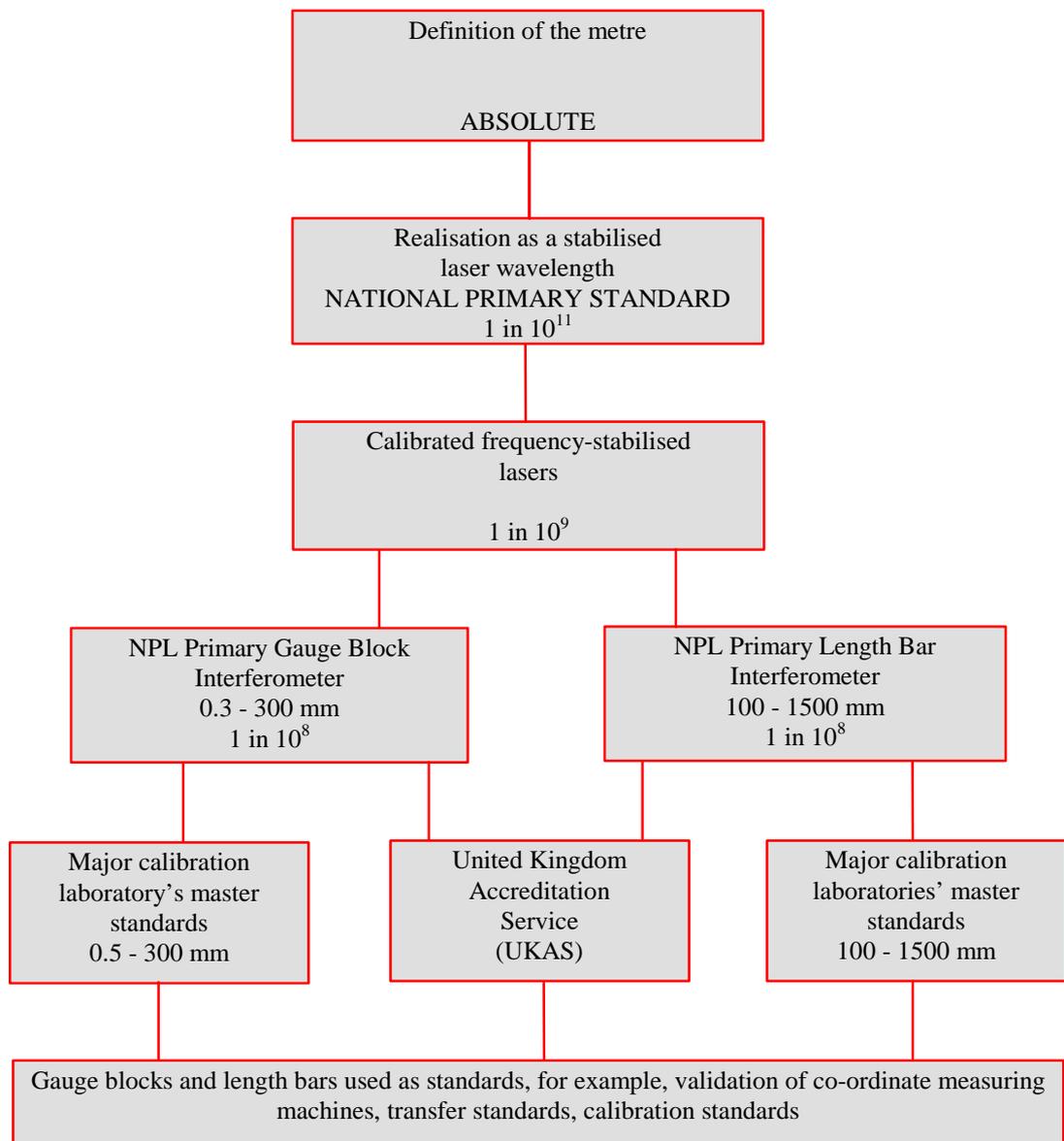


Figure 2.1 Traceability of length measurements.

Note that 1 in 10⁹ implies an accuracy of 1 nm in 1 m

At the head of the traceability chain is the definition of the metre (1983):

*The metre is the length of the path travelled by light in vacuum during a time interval
of 1/299 792 458 of a second.*

This is an absolute standard and is part of the SI system of units. The metre is realised at NPL as a wavelength of a He-Ne laser stabilised to a saturated absorption in neon at 632.991 398 1 nm. Details of the operation of this laser are given in chapter 4. A recent international comparison of iodine-stabilised lasers has shown agreement at the level of 2×10^{-11} and has resulted in a new uncertainty being adopted for the UK realisation of the metre of $\pm 2.5 \times 10^{-11}$.

As the primary laser is the UK's length standard, it follows that it is not used for routine calibration of gauge blocks via interferometry. Instead, commercial stabilised lasers (based on an NPL design) whose wavelengths are calibrated against the primary laser are used.

Annually, NPL measures the length of some 1200 gauge blocks for customers as well as providing a measurement service for laboratories accredited under the United Kingdom Accreditation Service (UKAS). In turn, these calibrated length standards are disseminated throughout UK industry where they are used to demonstrate traceability of length measurement for billions of pounds worth of trade.

Each country, as well as demonstrating national traceability, frequently takes part in international comparisons to ensure that their measurement systems are compatible and demonstrate traceability of length measurement to the internationally agreed definition of the metre.

2.2 CALIBRATION OF SURFACE TEXTURE MEASURING INSTRUMENTS

It is only recently that sound metrological principles have been applied to the measurement of surface texture. Demonstration of the traceability of surface texture measurements is still rare and is often achieved only by the time-consuming task of measuring the characteristics of the separate elements that make up the instrument. There are many different artefacts that can be used to characterise an instrument and a large number of parameters that can be calculated from the measured data. Although there is also a vast library of standards on the subject this thesis will only consider only those published by the International Standards Organisation (ISO). It is also rare to see an uncertainty figure quoted with a measured surface parameter - a taboo in dimensional metrology not seeming to cover surface texture.

2.2.1 HISTORICAL OVERVIEW

The metrological issues described in this section have been explored since the 1930s when the first surface texture instruments were being used. The traditional method of calibrating a surface texture measuring instrument is to traverse a lined calibration artefact. The period and amplitude of the lines were chosen to, ideally, encompass the whole operating bandwidth of the instrument or its transmission characteristics. One of the early standards took the form of acid-etched lines in a substrate and was developed by Timms of NPL (1946). In 1953 Underwood (also of NPL) introduced his Caliblocks which were formed with a diamond tool mounted in a dividing engine and used to rule electro-deposited gold. Schobinger (1956) made specimens by vacuum coating glass surfaces with silica through a wire mask. These early artefacts had limited topography

and could not match the lower spatial bandwidth limits of the instruments they were designed to calibrate.

Sharman of the National Engineering Laboratory (NEL) (1967) was, perhaps, the first to introduce essentially sinusoidal gratings as artefacts. These were formed by vacuum depositing aluminium and chromium on to a lapped steel substrate and for a given peak-to-peak height a single specimen could be used to generate any pitch from 0.25 μm to 13 μm . The pitch variation was achieved by mounting the specimen on a small rotary table and rotating the specimen relative to the direction of traverse of the pick-up head. Much of the literature examined advocates the use of step-height standards, such as wrung gauge blocks for low magnification and lever arm devices or evaporated or etched films for high magnifications (Reason 1967).

Van Hasselt & Bruin (1963) and Bendeli, Duruz, *et. al.* (1974) introduced the idea of using a vibrating platform to simulate the spatial frequencies present in a surface. Calibration of a stylus instrument requires low frequency or static calibration to determine the scale factor and linearity of the low frequency recording equipment. Dynamic calibration is also necessary to determine the transmission characteristics of averaging instruments with meter readouts. They suggest that the problem with using standard artefacts is the difficulty in obtaining satisfactory accuracies for static calibrations at very high magnifications (50 000x and above) without having to rely on the trueness of the attenuators and the effect of the finite shape of the stylus.

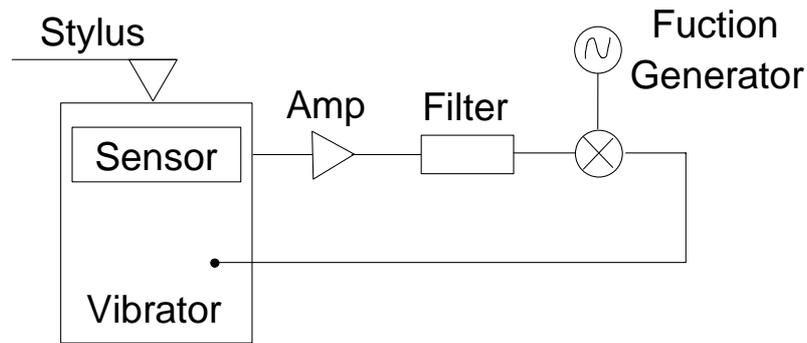


Figure 2.2 Vibrating stylus calibration rig, from Bendeli, Duruz, *et. al.* (1974)

Bendeli's instrumentation is shown in figure 2.2 and its operation is self evident. The method has the following features:

- it provides a way of determining scale factor and linearity of low frequency recorders;
- by applying a random signal to the vibration platform it provides a measurement of transmission characteristics;
- by applying signals that generate desired profile shapes and characteristics it calibrates parameter meters;
- its displacement range is equal to $\pm 2 \mu\text{m}$ with -3 dB cut-off point at 800 Hz;
- calibration is via a spherical-ended capacitance probe with an AC resistance bridge, resolution 2 nm;
- its sensitivity is equal to $0.543 \text{ V } \mu\text{m}^{-1} \pm 0.5\%$ per month;
- its non-linearity is less than 25 nm at extremities, but 10 nm over $\pm 1.5 \mu\text{m}$.

Teague (1978) used interferometrically determined step-heights to calibrate the scale factor of other surface texture measuring instruments. The main sources of error with this technique are due to the poor geometry of the step (microscopic surface texture of the sides, non-flat surfaces and non-parallel planes) and errors in the interferometry. Teague

derived a formula for the normalised uncertainty in a step height measurement, h , which is re-stated here

$$\frac{\delta h}{h} = \frac{\sigma_s}{h(n,s)} \left[1 + \left(\frac{d\delta}{ds} \right)^2 \right]^{\frac{1}{2}} \quad (2.1)$$

where h is given by

$$h = (n + g(s)) \frac{\lambda}{2} \quad (2.2)$$

and n is the integer fringe number, $g(s)$ is a general function of the fringe dispersion, s (the lateral fringe displacement over the fringe spacing), λ is the wavelength of light, σ_s is the variance in the measurement of s and δ is a small error in $g(s)$. Teague then used least-squares methods to calculate the step-height from the measurement data and proved that the uncertainty in assigning a height value to a stylus step-profile is approximately equal to the Ra or Rq of the step's surface texture.

Tsukada & Kanada (1986) describe how it is difficult to specify, for an entire surface, any averaging parameter due to the enormous amount of data required. They also discussed the difficulties of relating the surface texture tolerances stated on engineering drawings to the statistical surface texture parameters. By carrying out repeated measurements at different points on the same surface they reached the following conclusions:

- two-dimensional parameters, such as Ra , Rq and Rz when measured locally at different points on a surface, fluctuate according to a Gaussian distribution;

- each sample standard deviation, S , has a strong correlation with the population mean F and it can be expressed by $S = hF$ (h being a constant);
- three-dimensional surface texture parameters are larger than their two-dimensional brethren.

Whitehouse (1988) discussed the various sources of error contributing to the measurement of surface texture with a stylus instrument. He suggested using ruled or etched standards or crystal lattice spacings as calibration artefacts and discussed the need for a knowledge of the inherent elasticity in the instrument being used. He also advocated using x-ray interferometry to calibrate stylus instruments and described such a system that was developed at Warwick University (Chetwynd, Siddons, *et. al.* 1983). The most critical area, with x-ray techniques, was the mechanical interface between the silicon monolith and the transducer.

Song (1988) discussed the use of D-type random profile specimens for calibration of surface texture instruments (see §2.3 on standards for a description of a D-type specimen).

The disadvantages of the D-type specimens are:

- Ra range limited from 1.5 to 0.2 μm ;
- no smooth datum plane at both sides of the measuring area - this makes comparisons awkward;
- large measurement error resulting from the phase error between the skid and stylus.

Song produced many more random profile specimens to overcome these problems.

Griffith & Grigg (1993) suggested using carbon-60 and other fullerene-like structures as calibration artefacts for surface texture measuring instruments requiring nanometre spacings. Franks (1993) produced Amplitude-Wavelength space diagrams (Stedman 1987) to compare the performance of various instruments and suggested ways of correlating different probing types by assessing the strength of their surface-probe interaction. Jørgensen, Garnoes, *et. al.* (1997) used waffle-plates to measure lateral non-linearity of scanning probe instruments and Fourier techniques to analyse the measurement data.

Haitjema (1997) reported on EUROMET project 301 that compared methods of measuring depth-setting standards (step-heights) between laboratories in Europe. The project involved measuring five artefacts with groove widths of 0.01 mm and 0.1 mm and nominal depths of 32 nm, 64 nm, 160 nm, 1 μm and 3.2 μm . The samples were silicon substrates with a chromium overcoat and the definition of the groove was that stated in ISO 5436. Instruments used by laboratories taking part in the comparison were the Form Talysurf, Talystep, Nanostep and various interference microscopes. Participants were asked to carry out tests on the homogeneity of the specimens and the sensitivity of the depths to their definitions. Results showed that the depths are not sensitive to the definition to within a nanometre. The depth results were in good agreement for the small depths but the 3.2 μm standard gave a sample standard deviation of around 40 nm. Haitjema suggests that this situation must improve and that agreement should be within 1% (*i.e.* 30 nm in 3 μm).

As part of a major European project that is due to finish in December 2000 (see SMT4-CT 97-2176 *Calibration standards for surface topography measuring systems down to nanometric range*), NPL has been producing a number of sinusoidal specimens with varying period and amplitude to be used for calibrating surface texture measuring instruments (NMS

Programme for Length 1996-1999, see §6.5). Replicas of these samples, and even replicas of the replicas can be produced (Daly, Ferguson, *et. al.* 1997). These samples are measured with a stylus instrument and mathematical techniques (similar to reversal methods) can be applied to extract the shape of the stylus. NPL plans to increase the range of sinusoidal samples to include samples with multiple harmonics and varying phases. A number of comparisons of different instruments will then be carried out to assess the samples suitability as calibration artefacts. Watts, Sambles, *et. al.* (1997) have also been involved with the project and have reported on a method using surface plasmons to optically measure the shape of a stylus.

Scheer & Stover (1998) have developed square-wave gratings with depths in the range 1 to 10 nm. The production techniques stem from the silicon wafer technologies. Scheer & Stover measure the amplitude and period of these samples using AFM and angle resolved scatter (ARS). Power spectral analysis is used to correlate instruments with differing bandwidths. Their route to traceability is via the laser source of the ARS instrument.

Haitjema (1998) has recently reported on a thorough investigation of the Taylor-Hobson Form Talysurf stylus instrument. He suggests that deviations in comparisons of various stylus instruments are caused by short wavelength cut-offs due to stylus geometry and probe resonance oscillations. His method of calibration involves splitting the instrument into a number of sub-systems and using various metrological tools to calibrate each sub-system. These separate measurements are described below:

x axis calibration

Calibration of the x axis is required for the definition of the cut-off spatial wavelength, definition of the sampling length and to calculate spacing and hybrid parameters (for example Δq). A graduated rule is used with time-based analysis (this requires a constant traversing speed) to calibrate the x axis internal scales.

z axis calibration

Haitjema uses three methods for calibrating the z axis. Firstly, the stylus traces a standard sphere with known radius and the instrument calculates a polynomial correction for arcuate movement of the probe (this is the standard Taylor-Hobson method). Secondly, the length of two nominally 0.5 mm gauge blocks wrung to a flat surface with a length difference of around 10 μm is measured. The third method uses the set-up shown in figure 2.3. A piezoelectric actuator with capacitive feedback control (DPT) is connected to a function generator and used to move a small gauge block. The gauge block acts as a mirror for a differential plane mirror interferometer and as the contact point for the stylus. This is a very similar system to that of Bendeli, Duruz, *et. al.* (1974) except the traceability path is via the wavelength of the laser source as opposed to a capacitance sensor. The same calibration procedures apply in both instruments.

Straightness datum

An optical flat ($Ra \ll 10 \text{ nm}$) is traced and the known profile minus the measured profile gives the noise. For more confidence in the measurements, the same part of the datum (a steel rod slideway) is used in a measurement of different parts of the flat.

Filter definition and dynamic probe behaviour

The frequency and amplitude of the vibrating gauge are varied and the dynamics of the system are investigated. The filtered and unfiltered data are compared, plus the Rq calculated, with the Talysurf to the standard deviation measured with the interferometer system. From these data probe resonances and the effect of stylus flight can be identified.

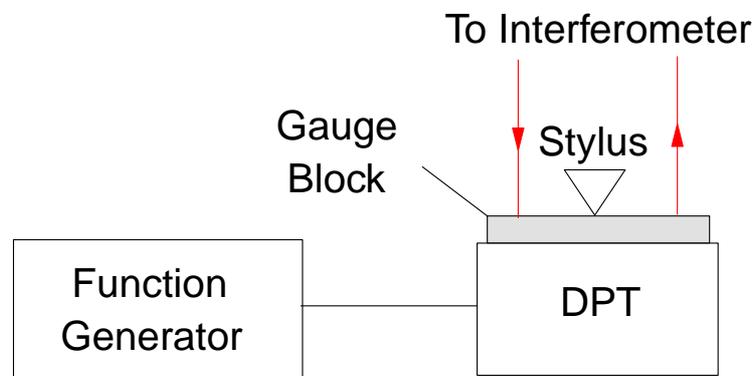


Figure 2.3 Vibrating stylus calibration rig, from Haitjema (1998)

Stylus geometry

The tip of an uncoated razor blade ($< 0.1 \mu\text{m}$ tip radius) is measured and the stylus geometry calculated. Whitehouse (1994) also discussed using a razor blade to measure the stylus geometry along with an etched artefact with decreasing line-widths to measure geometry and wear of the tip.

Measuring force

To determine the static measuring force a mass balance is contacted. Assuming that no stylus flight was apparent from the observations of the dynamic probe behaviour, the dynamic force will be less than the static force and so is not measured. This assumption is not necessarily true (Liu, Chetwynd, *et. al.* 1993). The force is also checked as the direction of probing is reversed.

The above examples illustrate that traceability and calibration in surface texture measurement has involved a multitude of artefacts and measurement strategies. §2.3 will examine the ISO standards that apply in the field.

2.3 SURFACE TEXTURE STANDARDS

In 1994 Whitehouse listed over one hundred standards for surface texture measurement. As more countries adopt International Standards, this number has dropped to around fifty. There are, however, many more standards that relate to surface texture in some way. Appendix B is a list of surface texture standards compiled using the 1998 Standards Infobase.

The current standards only cover the use of stylus instruments for two-dimensional measurements. Also, there are a number of other areas in the standards that are questionable from the point-of-view that they could be misinterpreted or misunderstood. These issues are being addressed by a major European project, headed by Huddersfield

University (SURFSTAND). This project is concentrating on calibration of three dimensional surface texture measuring instruments and defining parameters that can be used to describe aspects of the surface. It is estimated that this project will lead to the publication of specification standards in three to four years.

The most important standard, as far as this thesis is concerned, is ISO 5436 (1985): *Calibration specimens - stylus instruments - types, calibration and use of specimens*. The standard advocates the use of calibration specimens or artefacts to determine the operating characteristics of contact stylus instruments. Four different types of specimen are described:

Type A - These are used to measure the vertical magnification of an instrument (and are an example where the standard lacks rigor in assuming that the instrument has a linear response over the depth of the groove). They come in two sub-groups: *Type A1* - a wide calibrated groove with a flat valley the size of which is dictated by the dimensions of the stylus tip, *Type A2* - same as A1 but with a rounded valley.

Type B - These are used to investigate the geometry of the stylus tip. They also come with two sub-groups: *Type B1* - narrow grooves proportioned to be sensitive to the dimensions of the stylus; *Type B2* - two grids of equal R_a , one sensitive to the tip dimensions the other insensitive.

Type C - These are used to check parameter meters. They consist of a grid of repetitive grooves of similar shape (for example sinusoidal, triangular and arcuate waveforms) with low harmonic amplitudes. Type C specimens have well documented surface texture parameters and can be used to check the horizontal magnification.

Type D - These are used for an overall check of parameter meter calibration. They have irregular profiles in the direction of traverse, but they have the convenience of an approximately constant cross-section along their lengths.

All the specimens can be made of suitably hard materials but glass or quartz is favoured. At the time of writing the realisation of type B1 specimens is still under development.

Calibration of the specimens is carried out using interferometry and the route to traceability is via the calibrated vacuum wavelength of the source. The use of interferometry usually requires the surface to be metal coated and only shallow grooves can be measured without having to resort to specialised fringe analysis techniques. Problems arise due to the differences between the interaction of a contact stylus instrument and the surface and an optical beam and the surface. Examples of problems that arise are quoted in the standard and include the effects of the optical properties of the specimen material and the effects of oblique incidence. However, the magnitude of these effects is assumed negligible by the standard.

The standard goes on to consider the calibration aspects of instruments with a datum skid, the effects of stylus wear, practical use of the artefact and error statements. Very little, if any, real consideration is given to the statement of uncertainty that would either accompany a calibrated instrument or the artefacts used to calibrate an instrument.

This chapter has demonstrated the chaotic state of the field of surface metrology. The rest of the thesis describes the work that has been carried out at NPL in developing a new

traceable surface texture measuring instrument to, hopefully, bring some order to the field.

2.4 REFERENCES

- BS 5233 1986 *British standard glossary of terms used in metrology* (British Standards Institution: London)
- BS 4311 1993 *Gauge blocks and accessories Part 1. New gauge blocks* (British Standards Institution: London)
- Bendeli A, Duruz J, Thwaite E G 1974 A surface simulator for precise calibration of surface roughness measuring equipment *Metrologia* **10** 137-143
- Chetwynd D G, Siddons D P, Bowen D K 1983 X-ray interferometer calibration of microdisplacement transducers *J. Phys. E: Sci. Instrum.* **16** 871-874
- Daly D, Ferguson R, Hutley M C 1997 Replication of optical components using silicone rubber intermediate replica *Proc. SPIE* **3099**
- Franks A 1993 Progress towards traceable nanometric surface metrology *Nanotechnology* **4** 200-205
- Griffith J E 1993 Dimensional metrology with scanning probe microscopes *J. Appl. Phys.* **74** R83-R109
- Haitjema H 1997 International comparison of depth-setting standards *Metrologia* **34** 161-167
- Haitjema H 1998 Uncertainty analysis of roughness standard calibration using stylus instruments *Prec. Eng.* **22** 110-119
- Jørgensen J F, Garnæs J, Thulstrup D, Rasmussen S E 1997 Calibrated surface characterization, problems and solutions *7th Int. Conf. on Metrology and Properties of Engineering surfaces* 358-365
- Leach R K, Hart A, Jackson K 1999 Measurement of gauge blocks by interferometry: an investigation into the variability in wringing film thickness *NPL Report CLM3* 1-17
- Liu X, Chetwynd D G, Wang W 1993 Improvement of the fidelity of surface measurement by active damping control *Meas. Sci. Technol.* **4** 1330-1340

- National Measurement System Programme for Length 1996-1999 MPU 7/28.18
- Reason R E 1967 *ASTME Intern. Conf. On Manufacturing Technology*, Ann Arbor, Michigan
- Schobinger J P 1959 *Investigation into the production of roughness standards* (PhD thesis: Federal Technical College, Zürich)
- Sharman H B 1967 Calibration of surface texture measuring instruments *Proc. Instn. Mech. Engrs.* **182** 319-326
- Sheer B W, Stover J C 1998 Development of a smooth-surface microroughness standard *Proc SPIE* **3141** 78-87
- Song 1988 Random profile precision roughness calibration specimens *Surface Topography* **1** 303-314
- Stedman M 1987 Basis for comparing the performance of surface-measuring machines *Prec. Eng.* **9** 149-152
- Teague E C 1978 Uncertainties in calibrating a stylus type surface texture measuring instrument with an interferometrically measured step *Metrologia* **14** 39-44
- Timms C 1946 Measurement of surface roughness *Metal Treatment* **13** 111
- Tsukada T, Kanada T 1986 Evaluation of two- and three-dimensional surface roughness profiles and their confidence *Wear* **109** 69-78
- Underwood A 1953 Two recent developments for accurate measurement of surface texture *Proc. Symp. on Engineering Dimensional Metrology, NPL* **II** 621
- Van Hasselt R, De Bruin W 1963 Comparative investigation of industrial surface-roughness measuring instruments *Ann. CIRP* **XI** 193
- Watts R A, Samble J R, Hutley M C, Priest T W, Lawrence C R 1997 A new optical technique for characterizing reference artefacts for surface profilometry *Nanotechnology* **8** 35-39
- Whitehouse D J 1988 Nano-calibration of stylus-based surface measurements *J. Phys. E: Sci. Instrum.* **21** 46-51

Whitehouse D J 1994 *Handbook of Surface Metrology* (Institute of Physics Publishing)

CHAPTER 3

THE DESIGN OF NANOSURF IV

"A whole is that which has beginning, middle and end."

Aristotle (384 - 322 BC)

3.1 INTRODUCTION TO PRECISION INSTRUMENT DESIGN

Precision instrument design is a highly specialised science that combines the field of precision engineering with that of metrology. It involves scientific disciplines such as mechanics, materials, electronics, control, thermomechanics, dynamics and software. The needs and drivers for precision instrument design were described in §1.3 as far as surface texture measurement is concerned, but practically all industrial areas where highly accurate measurements are needed require the design of precision instruments. Schellekens, Rosielle, *et. al.* (1998) give a thorough up-to-date review of precision design and Smith & Chetwynd (1992) present an introduction to many of the precision design concepts discussed in this thesis. The trend towards higher precision in industry was highlighted by McKeown (1987) and more specifically for machining accuracy by Taniguchi (1983 and 1996) and Franse (1990). Taniguchi's famous graph predicting future machining accuracy requirements is depicted in figure 3.1.

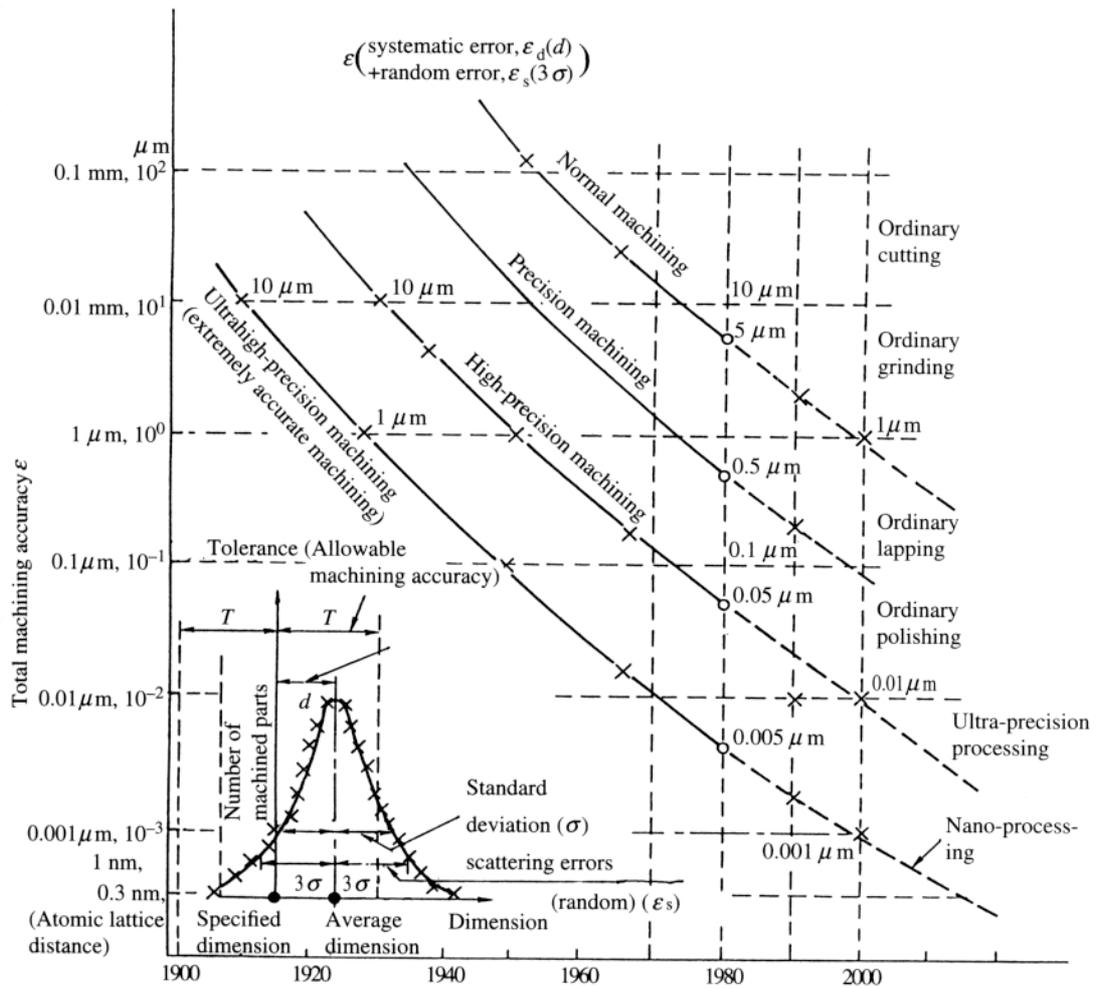


Figure 3.1 Trends in precision machining by Taniguchi (1983)

3.2 ELEMENTS OF PRECISION INSTRUMENT DESIGN

The following sections briefly describe the main elements of precision design with special reference to measuring instruments where necessary. This list follows essentially that of Schellekens, Rosielle, *et. al.* (1998) and is by no means exhaustive.

3.2.1 GEOMETRICAL CONSIDERATIONS

Most precision instrument designs involve parts that are formed from simple geometrical elements such as cylinders, tubes, beams, spheres and boxes to support loads in the system. Surfaces that are used for moving elements are usually formed from flats or cylinders. In practice, however deviations from these ideal shapes and structures occur due to form and surface texture error caused by the machining processes used to manufacture the parts. Geometry is also affected by the environment in which an instrument is housed, for example, vibration, temperature gradients and ageing can cause unwanted dimensional changes. Other factors that can affect the geometry of an instrument include: the effects of the connections between different parts, loading of the structure by the weight of the parts, stiffness and other material characteristics. Also of importance is the effect on the instrument of the part being measured. The above deviations from ideal geometry cause the various parts in an instrument to interact in a way that is very difficult to predict in practice.

3.2.2 KINEMATIC DESIGN

James Clark Maxwell (1890) was one of the first scientists to rigorously consider kinematic design. He stated that:

“The pieces of our instruments are solid, but not rigid. If a solid piece is constrained in more than six ways it will be subject to internal stress, and will become strained or distorted, and this in a

manner which, without the most micromechanical measurements, it would be impossible to specify."

These sentences capture, essentially, the main concepts of kinematic design. Kinematics is a branch of mechanics that deals with relationships between the position, velocity and acceleration on a body. Kinematic design aims to impart the required movements on a body by means of constraints (Nakazawa 1994).

A rigid body possesses six degrees of freedom in motion - three linear and three rotational. In Cartesian co-ordinates the degrees of freedom are in the x , y and z directions plus rotations about each of the axes. A constraint is that which prevents minimally motion in just one of the degrees of freedom. Smith & Chetwynd (1992) state two lemmas of kinematic design:

- (i) any unconstrained rigid body has six degrees of freedom;
- (ii) the number of contact points between any two perfectly rigid bodies is equal to the number of constraints.

This means that

$$\text{Number of constraints} + \text{remaining number of degrees of freedom} = 6.$$

There are often many assumptions applied when carrying out kinematic design. Real bodies are not perfectly rigid and will experience both elastic and possibly plastic deformations under a load. Such deformations will exclude perfect point contacts and cause unwanted motions. For this reason it is often important to choose with care the

materials, shapes and surface texture of a given part. Despite this, kinematic design is an extremely important concept that the designer must master. Two examples of kinematic design will be considered here - the Kelvin clamp and a one degree of freedom motion system. These are, essentially, the only two kinematic designs used on NanoSurf IV.

3.2.2.1 The Kelvin clamps

The Type I and Type II Kelvin clamps are examples of fully constrained systems, *i.e.* ones with six constraints. When designed properly these clamps are very effective where accurate re-positioning is required and are stable to within nanometres (Schouten , Rosielle, *et. al.* 1997).

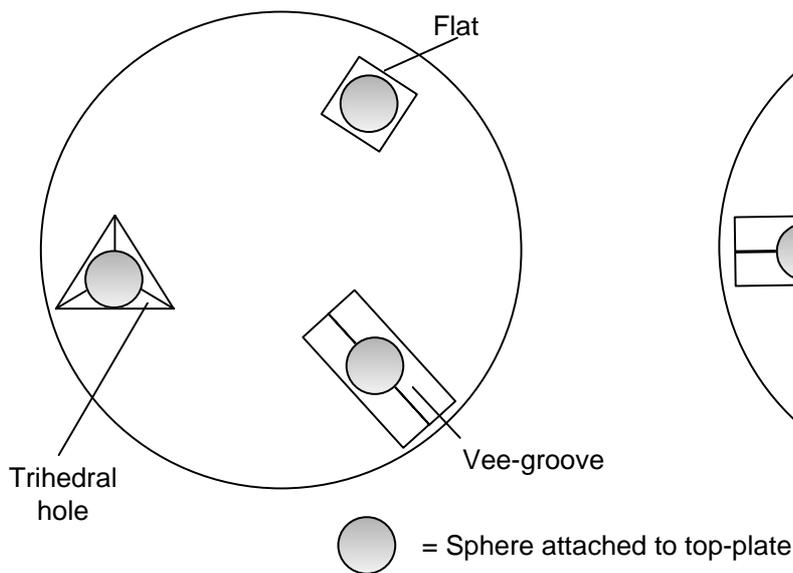


Figure 3.1 (a) Type I Kelvin clamp

Figure 3.2 (b) Type II Kelvin clamp

Both clamps have a top-plate (on which, for example the object to be measured is placed) that has three rigid spheres spaced on a diameter. The three spheres then contact on a flat

and in a vee and a trihedral hole, as in figure 3.2 (a), or in three vee-grooves, as in figure 3.2 (b). In the Type II clamp it is easy to see where the six points of contact, *i.e.* constraints are - two in each vee-groove. In the Type I clamp one contact point is on the flat, two more are in the vee-groove and the final three are in the trihedral hole. The Type I clamp has the advantage of a well defined translational location based on the position of the trihedral hole, but it is much harder to make. A trihedral hole is produced by pressing three spheres together in a flat bottomed hole (the contacting sphere will then touch at a common tangent) or by complex angled machining techniques. For miniature structures an anisotropic etchant can be used on a single crystalline material (Petersen 1982). The Type II is more symmetrical and less influenced by thermal variations. Note that the symmetrical groove pattern confers its own advantages but is not a kinematic requirement: any set of grooves will do provided that they are not all parallel.

3.2.2.2 A single degree of freedom motion device

There are many methods for producing single degree of freedom motion (see for example Smith, Chetwynd, *et. al.* 1987 and Monteiro, Smith, *et. al.* 1996). One method that directly uses the idea of single point contacts is the prismatic slideway (Smith & Chetwynd 1992). The contact points are distributed on two non-parallel flat surfaces as shown in figure 3.3. In practice the spheres would be attached to the carriage. The degrees of freedom in the system can be deduced by considering the loading necessary to keep all five spheres in contact. Firstly, the three point support could be positioned onto the horizontal plane resulting in a linear constraint in the z axis and rotary constraints about the x and y axes. A carriage placed on this plane is free to slide in the x direction until either of the two remaining spheres contacts the vertical face. The x axis linear freedom is then constrained. Further horizontal force would cause the carriage to rotate until the fifth sphere comes

into contact, removing the rotary freedom about the z axis. This gives a single degree of freedom linear motion along the y axis.

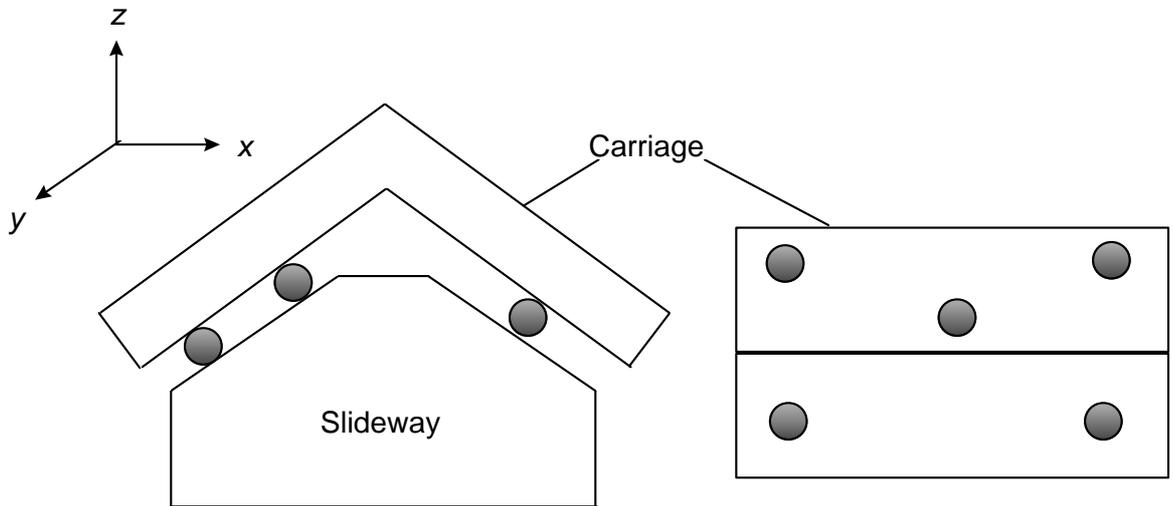


Figure 3.3 (a) Linear slideway and carriage - end view **Figure 3.3 (b)** Carriage - underside view

3.2.3 DYNAMICS

Most precision instruments used for dimensional metrology involve some form of moving part. This is especially true of surface texture measuring instruments. Motion usually requires some form of guideway, this being two or more elements that move relative to each other with fixed degrees of freedom. For accurate positioning, the play and the friction between the parts in the guideway must be reduced (unless the friction characteristics are being used to impart damping on the guideway). To avoid sticking and slipping of the guideway the friction should normally be minimised and kept at a constant value even when there are velocity or acceleration changes. It is also important that a guideway has a smooth motion profile to avoid high accelerations and forces.

The symmetry of a dynamic system plays an important role. With a rotating part the unbalance and mass moment of inertia must be reduced. A linear guideway should be driven through an axis that minimises any angular motion in its travel (its axis of reaction). Stiffness is another important factor; there must be a trade-off between minimising the forces on a guideway and maximising its stiffness. As with the metrology frame the environment in which the instrument is housed affects its dynamic characteristics.

Guideways can be produced using many techniques, but the most popular three are:

- (i) flexures - usually used only over a small range owing to the elastic limit and parasitic motion;
- (ii) linear slideways - like the one used on NanoSurfs I - IV;
- (iii) active feedback control systems (Reay 1993, Hicks & Atherton 1997, Atherton 1998).

3.2.4 THE ABBE PRINCIPLE

The Abbe Principle was first described by Ernst Abbe (1890) of Zeiss and states:

“If errors of parallax are to be avoided, the measuring system must be placed co-axially (in line with) the line in which displacement (giving length) is to be measured on the work-piece.”

Abbe error occurs when the measuring point of interest is displaced laterally from the actual measuring scale location, and when angular errors exist in the positioning system.

Abbe error causes the measured displacement to appear longer or shorter than the true

position, depending on the angular offset. Leach (1995) describes the Abbe error associated with a stage and probing system similar to that on NanoSurf IV and §5.10.7.2 describes the measurement of the errors introduced by Abbe effects on NanoSurf IV.

The Abbe Principle is, perhaps, the most important principle in precision instrument design and is also one that is commonly misunderstood - Bryan (1979) described it as “*the first principle of machine design and dimensional metrology*”. Abbe’s original paper concentrated on one-dimensional measuring instruments. Bryan (1979) re-stated the Abbe Principle for multi-dimensional systems as:

“The displacement measuring system should be in line with the functional point whose displacement is to be measured. If this is not possible, either the slideways that transfer the displacement must be free of angular motion or angular motion data must be used to calculate the consequences of the offset.”

3.2.5 THE METROLOGY FRAME

It is easier to describe the metrology frame of an instrument by splitting it into three types of loop structures: structural loops, thermal loops and metrology loops. These three structures are often interrelated and can sometimes be totally indistinguishable from each other.

3.2.5.1 The structural loop

A structural loop is an assembly of mechanical components which maintain relative position between specified objects. Using NanoSurf IV as an example, looking forward to

figure 3.7 we see the structural loop runs along the base-plate and up the bridge (shown as the L-piece in figure 3.7), through the z slideway and probe, through the object being measured, down through the y and x slideways and back into the base-plate to close the loop. Note that the last example assumes NanoSurf IV is operating without traceability, *i.e.* it is using the inductive probe to measure displacements in the z axis. It is important that the separate components in the structural loop have high stiffness to avoid deformations under loading conditions - deformation in one component will lead to uncompensated dimensional change at the functional or measurement point.

3.2.5.2 The thermal loop

Schellekens, Rosielle, *et. al.* (1998) describe the thermal loop as: “*a path across an assembly of mechanical components, which determines the relative position between specified objects under changing temperatures.*” Much akin to mechanical deformations in the structural loop, temperature gradients across an instrument can cause thermal expansion and resulting dimensional changes. It is possible to compensate for thermal expansion by choosing appropriate component lengths and materials. If well designed and if there are no temperature gradients present, it may just be necessary to make the separate components of an instrument from the same material. Thermal expansion can also be compensated by measuring thermal expansion coefficients and temperatures and applying appropriate corrections to measured lengths. This practice is common in gauge block metrology where the geometry of the blocks being measured is well known (Hughes 1996). Obviously, the effect of a thermal loop can be minimised by controlling the temperature stability of the room in which the instrument is housed.

3.2.5.3 The metrology loop

A metrology loop is a reference frame for displacement measurements, independent of the instrument base. In the case of NanoSurf IV and many surface texture measuring instruments it is very similar to the structural loop. The metrology loop should be made as small as possible to avoid environmental effects. In the case of an optical instrument, relying on the wavelength of its source for length traceability, much of the metrology loop may be the air paths through which the beam travels. Fluctuations in the air temperature, barometric pressure, humidity and chemical composition of these air paths cause changes in the refractive index and corresponding changes to the wavelength of the light (Edlén 1966, Birch & Downs 1994). This can cause substantial dimensional errors. The last example demonstrates that the metrology and structural loops can be quite different.

3.2.6 MATERIALS

Nearly all precision instrument designs involve minimising the influence of mechanical and thermal inputs which vary with time and which cause distortion of the metrology frame. Exceptions to this statement are, of course, sensors and transducers designed to measure mechanical or thermal properties. There are three ways to minimise the effects of disturbing inputs:

- (i) isolate the instrument from the input, for example using thermal enclosures and anti-vibration tables;
- (ii) use design principles and choose materials that minimise the effect of disturbing inputs, for example, thermal compensation design methods, materials with low coefficients of expansion and stiff structures with high natural frequencies;
- (iii) measure the effect of the disturbing influences and correct for them.

The choice of materials for precision instruments is closely linked to the design of the force loops that make up the metrology frame.

3.2.6.1 Minimising thermal inputs

Thermal distortions will usually be a source of inaccuracy. To find a performance index for thermal distortion consider a horizontal beam supported at both ends of length L and thickness h (Cebon and Ashby 1994). One face of the beam is exposed to a heat flux of intensity Q in the y direction that sets up a temperature, T , gradient, dT/dy , across the beam. Assuming the period of the heat flux is greater than the thermal response time of the beam, then a steady state is reached with a temperature gradient given by

$$Q = -\lambda \frac{dT}{dy} \quad (3.1)$$

where λ is the thermal conductivity of the beam. The thermal strain is given by

$$\varepsilon = \alpha(T_0 - T) \quad (3.2)$$

where α is the thermal expansion coefficient and T_0 is the ambient temperature. If the beam is unconstrained, any temperature gradient will create a strain gradient, $d\varepsilon/dy$ in the beam causing it to take up a constant curvature given by

$$K = \frac{d\varepsilon}{dy} = \alpha \frac{dT}{dy} = \frac{\alpha}{\lambda} Q. \quad (3.3)$$

Integrating along the beam gives the central deflection of

$$\delta = C_1 L^2 Q \left(\frac{\alpha}{\lambda} \right) \quad (3.4)$$

where C_1 is a constant that depends on the thermal loads and the boundary conditions. Thus for a given geometry and thermal input, the distortion is minimised by selecting materials with large values of the performance index

$$M_Q = \frac{\lambda}{\alpha}. \quad (3.5)$$

Chetwynd (1987) and Smith & Chetwynd (1992) arrive at the same index by considering other types of thermal load. If the assumption that the period of the heat flux is greater than the thermal response time of the beam is not valid then the thermal mass of the beam has to be taken into account (Chetwynd 1987). In this case thermal conductivity is given by

$$\lambda = \frac{D}{\rho C_p} \quad (3.6)$$

where D is the thermal diffusivity of the beam material, ρ is its density and C_p is its specific heat capacity. In the case of a room with stable temperature and very slow heat cycling equation (3.5) is normally valid.

3.2.6.2 Minimising mechanical inputs

There are many types of mechanical input that will cause unwanted deflections of a metrology frame. These include elastic deflections due to self weight, loading due to the object being measured and external vibration sources. To minimise elastic deflections a high stiffness is desirable. The elastic self-deflection of a beam is described by

$$y = C_2 \frac{Wx^3}{EI} \quad (3.7)$$

where W is the weight of the beam, E is the Young's modulus of the beam material, I is the second moment of area of the cross-section and C_2 is a constant that depends on the geometry of the beam and the boundary conditions. It can be seen from equation (3.7) that, for a fixed design of instrument, the self-loading is proportional to ρ/E - minimising this ratio minimises the deflection.

The natural frequency of a beam structure is given by

$$\omega_n = C_3 \sqrt{\frac{EI}{ml^3}} \quad (3.8)$$

where n is the harmonic number, m is the mass per unit length of the beam, l its length and C_3 is a constant that depends on the boundary conditions. Again, for a fixed design of instrument, ω_n is directly proportional to $\sqrt{E/\rho}$. For a high natural frequency and, hence, insensitivity to external vibrations it is, once again, desirable to have high stiffness. As with the thermal performance index, a mechanical performance index can be given by

$$M_m = \frac{E}{\rho}. \quad (3.9)$$

Insensitivity to vibration will be discussed in more detail in §3.2.8.

3.2.7 SYMMETRY

Symmetry is a very important concept when designing a precision instrument. Any asymmetry in a system normally has to be compensated for. In dynamics it is always better to push or pull a slideway about its axis of reaction otherwise parasitic motions will result due to asymmetry. If a load bearing structure does not have a suitably designed centre of mass, there will be differential distortion upon loading. It would seem that symmetry should be incorporated into a precision instrument design to a maximum extent. The Abbe principle is a good example of symmetry at work. An excellent example of a symmetrical structure (plus many other precision instrument design concepts) is the Tetraform grinding machine developed by Kevin Lindsey at NPL (Lindsey 1992, Hadlington 1999). The symmetrical tetrahedral structure of Tetraform can be seen in figure 3.4. Calculations and preliminary experimental results showed that the Tetraform is extremely well compensated for thermal and mechanical fluctuations. Note that the tetrahedral structure was considered when designing NanoSurf IV, but such a structure would be extremely difficult to manufacture in Zerodur, existing blanks of Zerodur were already in existence (see §3.3.2) plus there were other pragmatic reasons for a more conventional geometry.



Figure 3.4 Kevin Lindsey (on one of the rare occasions that he wore a suit) with the Tetraform grinding machine

3.2.8 VIBRATION ISOLATION

Most precision instruments require some form of isolation from external and internal mechanical excitations. Where sub-nanometre accuracy is required it is essential that seismic and sonic vibrations are suppressed.

3.2.8.1 Sources of vibration

There are two main noise components in a seismic frequency spectrum. Firstly, high-frequency pseudo-random acceleration noise at frequencies equal to and above 1 Hz. The

spectral density for acceleration in a 'quiet place' is typically 10^{-7} to 10^{-6} m s⁻² per $\sqrt{\text{Hz}}$ (Giazotto 1990). Secondly, low-frequency microseismic ground motion at frequencies less than 1 Hz with typical amplitudes in the range 1 to 10 μm (Fritschel, Malvalvala, *et. al.* 1998). The spectrum of seismic vibration is strongly determined by location and time. At less than 50 mHz vibration is caused by atmospheric pressure variations caused by wind turbulence and infrasonic waves (Newell, Richman, *et. al.* 1997).

From 50 to 500 mHz vibrations are caused mainly by ocean waves that generate surface and plate waves in the earth's crust at coastal areas (especially where storm activity is high). The two most dominant frequency components from 50 to 500 mHz are the fundamental ocean wave frequency at 60 to 90 mHz and its second harmonic caused by ocean wave reflections at 120 to 160 mHz (the microseismic peak). Over 500 mHz to a few Hertz the spectrum is dominated by local sources of vibration such as wind-blown vegetation and human activity. Figure 3.5 shows measured vertical amplitude spectral densities for a 'noisy' and a 'quiet' area (Newell, Richman, *et. al.* 1997). Note that the spectrum below 0.1 Hz and above 30 Hz is limited by the seismometer's internal noise.

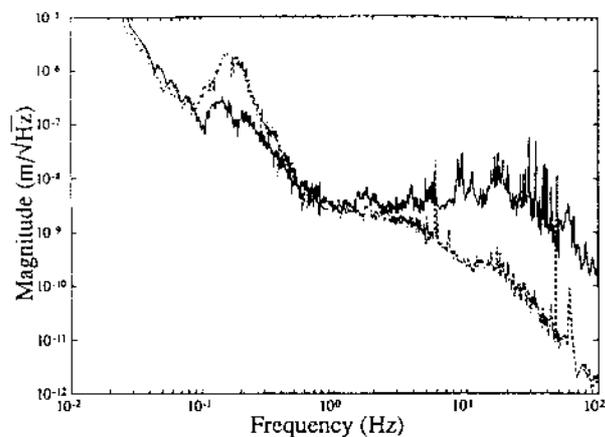


Figure 3.5 Measured vertical amplitude spectral densities. The solid curve is a laboratory on the campus of the University of Colorado, Boulder. The dashed curve is from NIST. The NIST lab is small, remote and separated from the main complex. In addition, all fans and machinery were turned off during the measurements. The spectrum below 0.1 Hz and above 30 Hz is limited by the seismometer's internal noise (Newell, Richman, *et.al.* 1997).

3.2.8.2 Passive vibration isolation

Vibration isolation in both vertical and horizontal directions can be provided by simple springs and pendulums. The transmissibility of an isolator is the proportion of a vibration as a function of frequency that is transmitted from the environment to the structure of the isolator. For a single degree of freedom vibration isolation system the transmissibility, T , is given by (Timoshenko, Young, *et. al.* 1974)

$$T = \frac{\omega_0^2}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\gamma^2\omega_0^2\omega^2}} \quad (3.10)$$

where ω_0 is the resonant frequency of the isolator and γ is the viscous damping factor. Figure 3.6 shows the transmissibility as a function of frequency ratio for various damping factors.

Vibration isolation is provided only above $\sqrt{2}$ times the natural frequency of the system, that is

$$T = \left(\frac{f_0}{f}\right)^2 \quad \text{for } f \ll f_0. \quad (3.11)$$

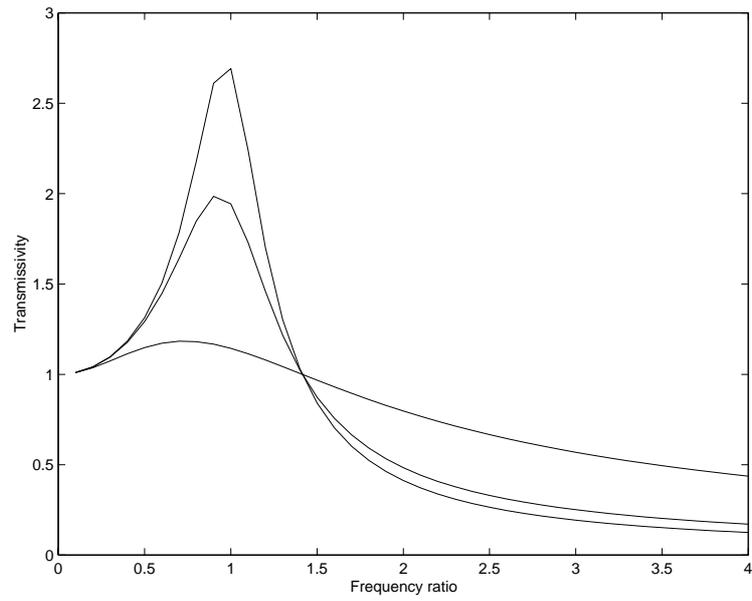


Figure 3.6 Damped transmissibility, T , as a function of frequency ratio (ω/ω_0)

Therefore to provide vibration isolation at low frequencies, the resonant frequency of the isolation system must be as low as possible. The resonant frequency for a pendulum is given by

$$f_0 = \frac{1}{2\pi} \sqrt{\frac{g}{l}} \quad (3.12)$$

and by

$$f_0 = \frac{1}{2\pi} \sqrt{\frac{k}{m}} \quad (3.13)$$

for a spring, where g is the acceleration due to gravity, l is the pendulum length, k is the spring constant and m is the mass.

Re-writing equation (3.13) in terms of the static extension or compression of a spring, δ , gives

$$f_0 = \frac{1}{2\pi} \sqrt{\frac{g}{\delta}} \quad (3.14)$$

since the static restoring force $k\delta = mg$. Thus for a low resonant frequency in a spring system it is necessary to have a large static extension or compression (or use a specialised non-linear spring). For the effect on the resonant frequency of the material properties of the spring see §3.2.7.2.

3.2.8.3 Damping

In vibration isolation systems it is important to have damping, to attenuate excessive vibration near resonance. In equation (3.10) it is assumed that velocity-dependent (viscous) damping is being applied. This is attractive since viscous damping does not degrade the high-frequency performance of the system.

The effects at resonance due to other forms of damping can be represented in terms of an “equivalent viscous damping”, using energy dissipation per cycle as the criterion of equivalence (Beranek 1988). However, in such cases, the value of the equivalent viscous damping is frequency dependent and, therefore, changes the system behaviour. For

hysteresis or structural damping, the damping term depends on displacement instead of velocity.

3.2.8.4 Internal resonances

A limit to high frequency vibration isolation is caused by internal resonances of the isolation structure or the object being isolated (Beranek 1988). At low frequencies the transmissibility is accurately represented by the simple theory given by equation (3.10), but once the first resonance is reached, the isolation does not improve. Typically the fundamental resonance occurs somewhere in the acoustic frequency range. Even with a careful design it is difficult to make a structure of an appreciable size with internal resonant frequencies above a few kilo-hertz.

3.2.8.5 Active vibration isolation

Active vibration isolation is a method for extending the low frequency isolation capabilities of a system, but is very difficult in practice. Single degree of freedom isolation systems are of little practical use because a non-isolated degree of freedom reintroduces the seismic noise even if the other degrees of freedom are isolated. A six degree of freedom isolation system has only recently been demonstrated (Newell, Richman, *et. al.* 1997) for an interferometric gravitational wave detector.

3.2.8.6 Acoustic noise

Acoustic noise appears in the form of vibrations in a system generated by ventilators, music, speech, street noise, *etc.* over a frequency range from about 10 to 1000 Hz in the form of sharp coherent resonances as well as transient excitations (Filinski & Gordon 1994). Sound pressure levels in a typical laboratory environment are greater than 35 dB, usually due to air conditioning systems.

Consider an enclosure that is a simple bottomless rectangular box whose walls are rigidly attached at each edge. When a panel is acoustically excited by a diffuse sound field, forced bending waves govern its sound transmission characteristics and the sound pressure attenuation is determined by the panel mass per unit area (Beranek 1988). The panel sound pressure attenuation (dB) is given by (Brenan, Charette, *et. al.* 1992)

$$\alpha = -10 \log_{10} \left[1 + \left(\frac{\pi \rho_s f}{\rho_0 c} \right)^2 \right] + 5 \text{ dB} \quad (3.15)$$

where ρ_s is its mass per unit area. ρ_0 is the density of air at standard pressure and f is the incident acoustic field frequency. Equation (3.15) suggests the enclosure wall should be constructed from high-density materials to obtain the largest ρ_s possible given the load bearing capacity of any supporting structure. Note that the attenuation decreases for every 20 dB per decade increase in either ρ_s or frequency.

3.3 DESIGN OF NANOSURF IV

3.3.1 THE OVERALL DESIGN

NanoSurf IV is a surface texture measuring instrument that was designed to measure displacements in two nominally orthogonal directions with an accuracy of better than 1 nm. Many instruments can resolve subnanometre features but have no way of demonstrating whether the readout of the co-ordinates from the instrument are correct. They are indicators - not measuring instruments in the true sense of the word (see §2.1 on traceability). NanoSurf IV is simply a stylus instrument, as discussed in §1.5.2, but with traceable metrology inherent in its design. Figure 3.7 shows the subsystems of NanoSurf IV and illustrates how they fit together.

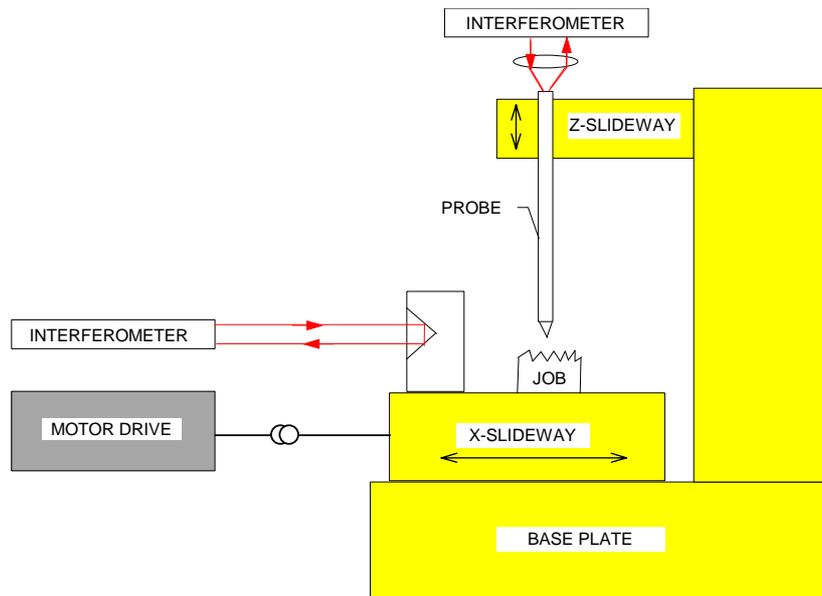


Figure 3.7 Schema of NanoSurf IV

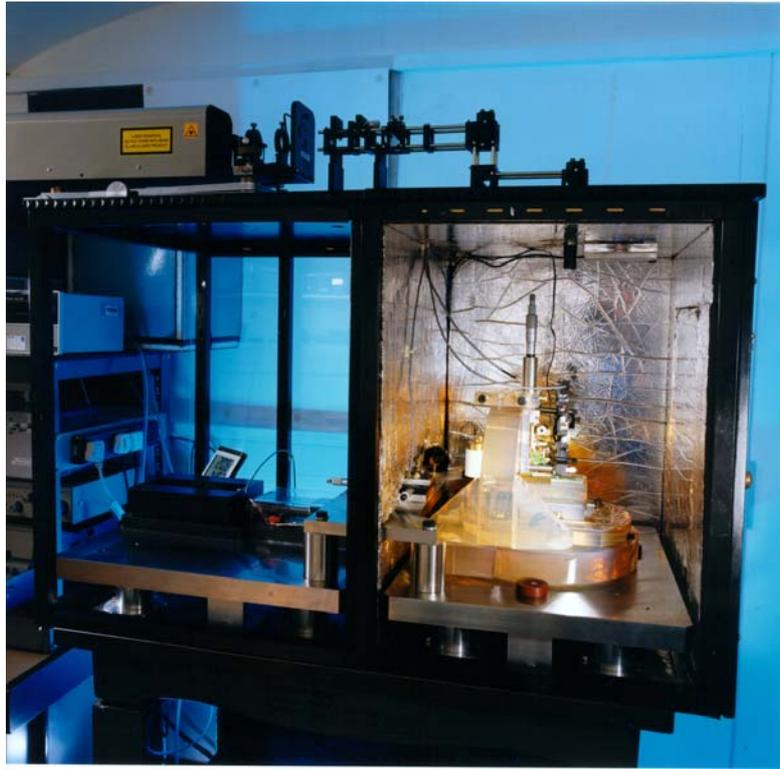


Figure 3.8 The sub-systems of NanoSurf IV

In addition to the core subsystems there are also electronics which receive the signals from the interferometers and control the moving parts in the instrument, a computing system that automates a measurement and carries out any data processing, and systems to measure aspects of the environment in which the instrument is housed. Each subsystem, with the exception of the interferometers, is described in detail in this chapter. The interferometers, essentially the metrology in the system, are described in chapter 4. Figure 3.8 is a photograph of NanoSurf IV with only the electronic and computer subsystems not in view.

A description of the measurement procedure for using NanoSurf IV is needed in order to understand the links between the subsystems. Briefly, with reference to figure 3.7, the object to be measured is placed on a specimen table that in turn sits on the x slideway. The z slideway is used to bring the stylus into contact with the surface. The drive

mechanism pushes or pulls the x slideway via a coupling rod and hence moves the surface during which time the stylus traces out the surface profile. The displacement moved in the x and z axes is measured by two interferometers. Note that NanoSurf IV can operate entirely free from the interferometers. In this case, the x displacement assumes that the scan velocity is constant and that data are collected with uniform time intervals between samples.

3.3.2 THE NANOSURF IV METROLOGY FRAME

The metrology frame is comprised of three basic elements: a base-plate to support the entire structure of the instrument, a bridge that carries a probing system and a vertical stage to position the probe over the specimen being measured. Figure 3.9 is a schema of the instrument where the base-plate can be clearly seen with the x and y slideways upon it. The base-plate is 400 mm in diameter, 70 mm thick and is lapped and polished flat.

The base-plate is located on a vibration isolation table (see §3.3.9) by three pads that are held in place by the mass of the system. The metrology frame is held in place in the horizontal plane by three 40 mm diameter by 20 mm long Tufnol™ cylinders, two of which have a central M6 clearance hole and one of which has an off-centre M6 clearance hole acting as a cam. These are screwed into position on the vibration isolation table at 120° angles around the circumference of the base-plate. As the screws are tightened, the cam ensures the base-plate is held in position.

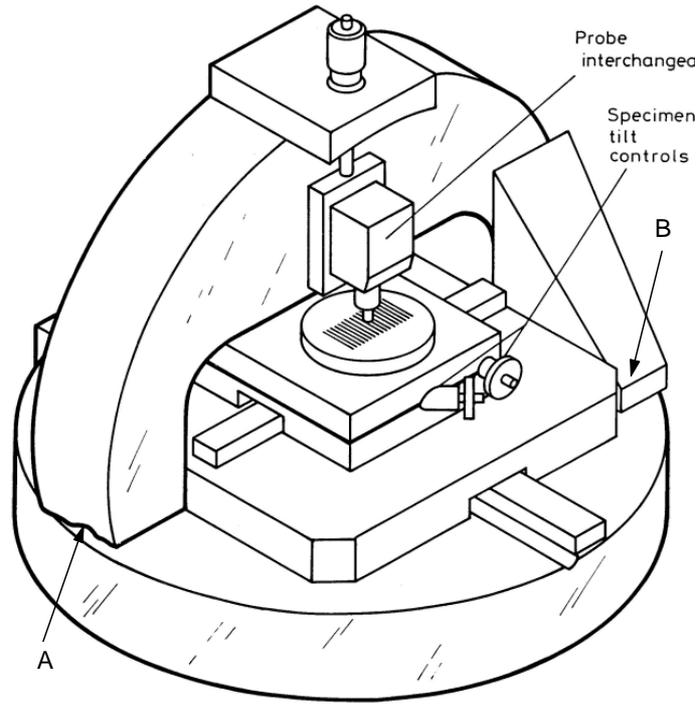


Figure 3.9 The NanoSurf IV metrology frame

The bridge part of the metrology frame is kinematically located on the base-plate via ball contacts and three vee-grooves. The ball contacts are hemispherical silica lenses that have been cemented to the base-plate with a very thin layer of Araldite™ (see end of next paragraph). The vee-grooves are machined into the underside of the bridge. In Figure 3.9 the vee-grooves would be at positions A, B and at a mirror-symmetric position directly opposite to B that cannot be seen in the diagram. For convenience the vee-grooves are not positioned 120° apart. The exact structures of the bridge and the base-plate can be seen in figure 3.10 which shows the NanoSurf IV at an early stage of development. One of the Tufnol cylinders is clearly visible in figure 3.10.

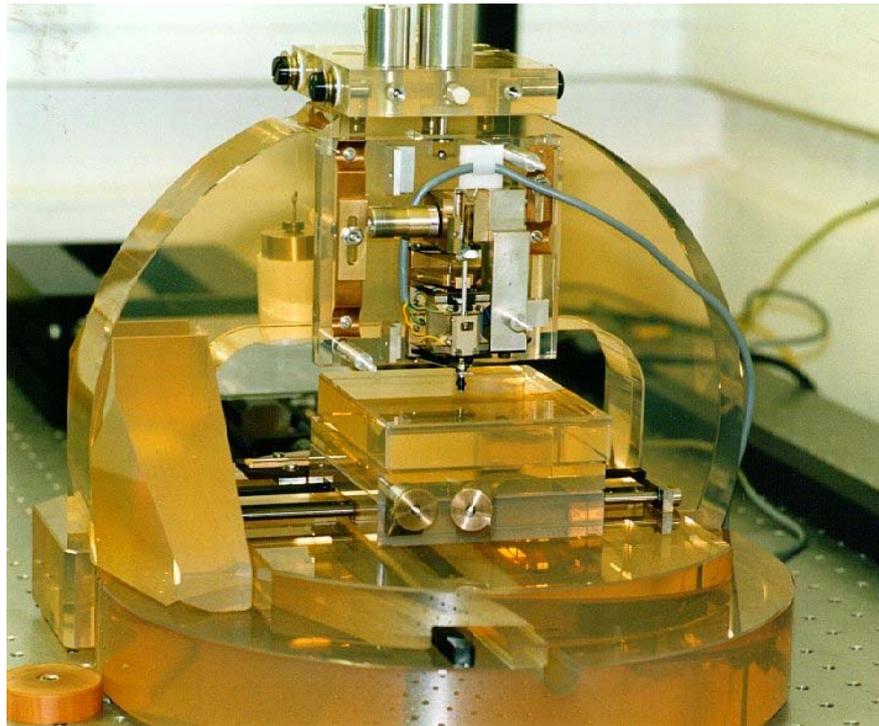


Figure 3.10 The NanoSurf IV metrology frame in its early days of development. Note that the Tufnol cylinder on the bottom left of this photograph is 40 mm in diameter and 20 mm long.

The entire metrology frame is manufactured from a low-expansion material called Zerodur™. Zerodur is a glass ceramic, an inorganic, pore-free material that has both glass-like and crystalline phases. The crystalline phase has negative thermal expansion characteristics (caused by phase change effects), whilst the vitreous phase has positive expansion characteristics. This practically balances out the linear thermal expansion at temperatures between -30 and +70 °C. All the Zerodur parts have been polished and have bevelled edges to prevent micro-cracking in the structure and to alleviate internal stresses caused by the manufacturing processes. Where separate parts are required to be in constant contact, for example the two legs that house two of the vee-grooves in the bridge, they have been cemented together with standard Araldite. Araldite has proven to be a convenient and cheap way to permanently attach Zerodur to Zerodur. So long as

only a thin layer of epoxy is used, the difference in the expansion coefficients between the epoxy and the Zerodur does not cause problems and its slow cure rate ensures that very low mechanical stresses are imparted to the Zerodur structures compared to conventional use of metal screws, *etc.* The main problem with Araldite is that changes in humidity can cause dimensional changes in the structures it is holding together, but, again as long as the interfacing layer is thin this is not a serious problem (Franks 1999).

As discussed above, when considering the material for the metrology frame many factors had to be considered. For this analysis only four materials will be considered: aluminium, steel, Invar™ and Zerodur. There are a great many other materials that could have been chosen to use for NanoSurf IV although experience with the previous NanoSurf designs plus other metrological instruments ruled out almost all other materials (Lindsey, Smith, *et. al.* 1988).

Values of λ , α and the thermal performance index, M_Q , can be seen in table 3.1. On this criterion alone, Zerodur would be the material of choice, *i.e.* its thermal properties give the largest value of M_Q .

Material	Thermal conductivity $\lambda/\text{W m}^{-1} \text{K}^{-1}$	Expansion coefficient $10^{-6} \alpha/\text{K}^{-1}$	Performance index $M_Q/10^6 \text{W m}^{-1}$
Aluminium	236.0	23.1	10.2
Steel	≈ 50.0	10.7	4.7

Invar (64 Fe - 36 Ni)	13.4	0.3	44.7
Zerodur	1.64	≤ 0.02	> 82

Table 3.1 Thermal properties of the design materials. With the exception of Zerodur all the values were obtained from Kaye & Labey (1989). The values for Zerodur were obtained from Schott Glasswerke.

Zerodur was chosen more from a practical point of view than that given purely theoretically above. Design cannot be reduced to calculating one simple equation. The base-plate of NanoSurf IV must be lapped and polished flat and requires a highly accurately machined and finished straight-edge (see next section). It is not possible to achieve the required finish with Invar. Invar is also a very heavy material compared to Zerodur and would make it very difficult to remove the bridge. NanoSurf IV is housed in a laboratory with very stable temperature characteristics and small temperature gradients. The anechoic chamber (see §3.3.8) also helps thermal stabilisation of the instrument. In this environment the thermal conductivity of the metrology frame is of less significance than the expansion coefficient. Provided it has been allowed to settle after being machined, Zerodur also has far better short- and long-term dimensional stability than Invar due to its homogenous molecular structure and pore-free surface (Berthold, Jacobs, *et. al.* 1977). It must also be pointed out that the Invar analysed above is a type known as SuperInvar - normal Invar has a thermal expansion coefficient of the order of 1.3 K^{-1} .

Zerodur is also the best material from a high-stiffness-to-weight ratio point of view (see §3.2.8.2). Table 3.2 gives values of the E , ρ and the mechanical performance index for the four materials considered.

Material	Young's modulus E/GPa	Density $10^3 \rho/\text{Kg m}^{-3}$	Performance index $M_m / \times 10^6 \text{ m}^2 \text{ s}^2$
Aluminium	71.7	27.9	2.57
Steel	≈ 200	78.5	2.55
Invar (64 Fe - 36 Ni)	147	80.5	1.83
Zerodur	90.3	25.3	3.57

Table 3.2 Materials properties of the design materials. The values for steel and aluminium were obtained from Kaye & Laby (1989). The values for Zerodur and Invar were obtained from Schott Glasswerke and Hicks & Atherton (1997).

The drawbacks of Zerodur must also be pointed out - again they all stem from a practical point of view. Zerodur is very brittle and cracks very easily - if the instrument were dropped it would be destroyed. Also, once the metrology frame has been manufactured any subsequent modifications to its mechanical structure would be far easier to make in Invar. Lastly, Zerodur is very expensive to purchase and to machine and can experience long-term dimensional instability (less than with Invar) after being machined (Riehle 1998). Note that the Zerodur on NanoSurf IV was around six years old (from the point at which it was machined) when the first measurements of surface texture were made.

From figures 3.9 and 3.10 it can be seen that the base-plate carries the x and y slideways (§3.3.3) that in turn carry the tilt and level controls and the specimen mounting plate (§3.3.10). The bridge carries the probing system and the z slideway (§3.3.5) plus the z interferometry mount (see chapter 4).

3.3.3 THE SLIDEWAY

The slideways used on NanoSurf IV are of a dry-bearing prismatic design similar to that used on the previous three NanoSurf instruments (Lindsey, Smith, *et. al.* 1988, Garratt & Bottomley 1990). The bearing surfaces for the slideway are the top surface of the base-plate and a straight-edge that is cemented to the base-plate with epoxy resin. Figure 3.11 shows the base-plate viewed from above where the straight-edge can be seen plus a gully for the slideway-to-drive mechanism coupling rod and a ferromagnetic steel rod. Figure 3.12 shows a schema of the base-plate and carriage viewed from the front of the instrument and figure 3.10 is a photograph from a similar angle. Three small polymeric pads (shown in green) can be seen contacting the bearing surfaces. Two more pads make the full complement of five constraints to allow a linear motion in one axis only. The contacting positions of the pads are shown in figure 3.11.

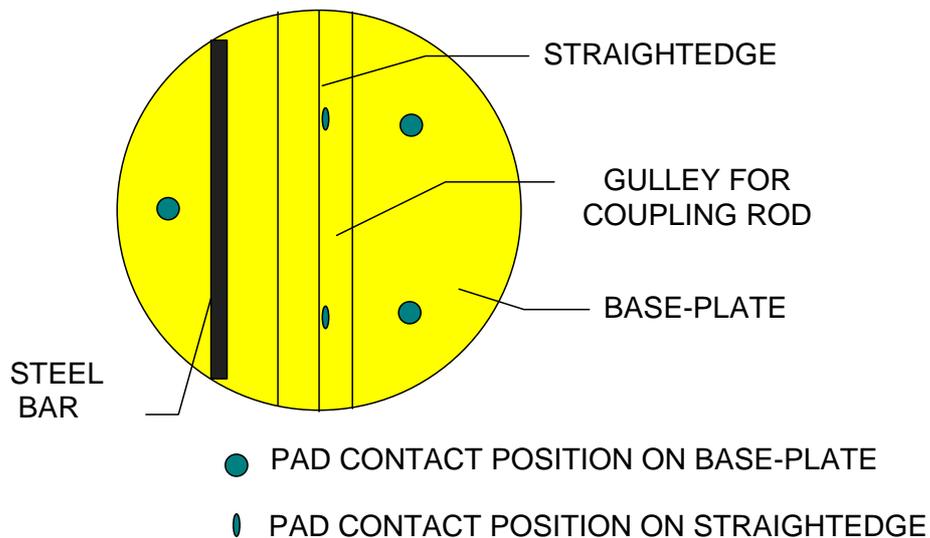


Figure 3.11 Schema of top view of the base-plate showing the position of the bearing pads

Two ceramic magnets are fixed to the underside of the carriage and are attracted to the steel rod to pull two of the pads into contact with the straight-edge. The other three pads are held in contact by the weight of the carriage. The steel bar has deep saw-cuts across it to stop the Zerodur base-plate cracking in the event of a large thermal variation in the room temperature.

This slideway design was used, as opposed to the design used on the previous NanoSurf instruments for purely practical reasons. Firstly, this design is more space conservative when the y axis slideway is mounted on top of the x axis slideway. Also, at the time of manufacture, Zerodur was most easily obtained in the form of polished optical blanks, two of which were used to produce the base-plate and the bridge and one of which was used to produce the x slideway. Note that to save on the cost of buying the materials, most of the other Zerodur components were manufactured from spare material that was available at NPL.

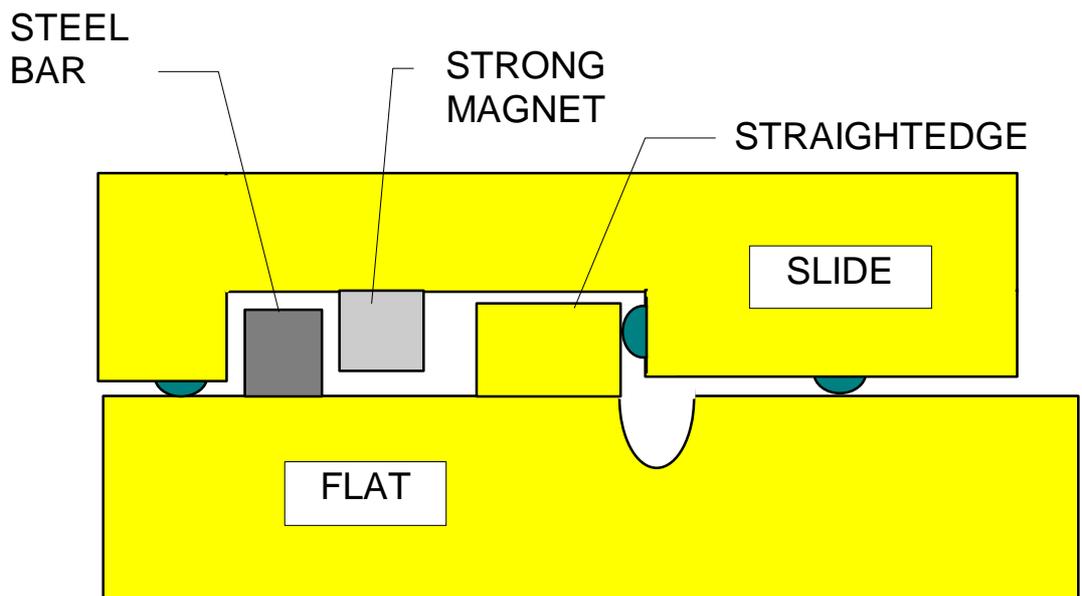


Figure 3.12 Schema of front view of the base-plate and slideway

The bearing pads are made of Glacier DUTM bearing material. This is a composite of poly-tetra-fluoro-ethylene (PTFE) and lead that has been keyed into a porous brass substrate which is then bonded to mild steel (Monteiro & Chetwynd 1996). The straight, long chain, fluorocarbon molecules that make up PTFE produce a high stiffness perpendicular to the direction of sliding and a low shear strength parallel to it. However, the bulk properties of PTFE, such as its low strength, high creep rate and high thermal expansion coefficient, mean that it would not be a good bearing material in bulk form. These detrimental characteristics are overcome by the thin film matrix arrangement of the DU material (Smith & Chetwynd 1992). The overall thermal diffusivity will be roughly the sum of the polymer film value and that for the brass. A very thin film can conduct significant heat to the brass with only a small temperature drop, after which it is rapidly conducted away by the high-diffusivity brass substrate. The stiffness is little affected by the thin layer so that it is almost that of brass (Smith, Harb, *et. al.* 1992). The thickness of the PTFE film is less than 20 μm (Smith & Chetwynd 1992). The pads have been punched into a spherical surface to provide kinematic 'point' contact. The performance of this type of pad-slideway arrangement has been theoretically and experimentally analysed by Monteiro & Chetwynd (1997). Liu, Chetwynd, *et. al.* (1997) have developed a friction test rig to investigate the performance of similar polymeric bearings.

The slideway shown in figures 3.11 and 3.12 is for the x axis. Mounted directly onto this slideway is an orthogonal y slideway of the same design principle. The y slideway is purely for positioning the object to be measured and is driven manually via a simple lead screw. The performance of the x slideway as a positioning device is investigated experimentally in chapter 5 as part of the uncertainty analysis of NanoSurf IV.

3.3.4 THE DRIVE MECHANISM

The drive mechanism for the x slideway is a lead screw driven by a dc servo motor. The drive motor is a McLennan type 3557 CS dc-motor-tacho combination with the advantages of linear current/torque characteristics, low rotor inertia, high efficiency and smooth running. Coupled to the motor is a planetary gearhead again manufactured by McLennan (type 38). The gearbox/motor combination is controlled by a McLennan series PM121 dc servo amplifier. The entire gearbox/motor combination is housed in a heavy metal case and sits on the 66 kg stainless steel plate on rubber vibration isolation pads.

The motor turns a lead screw with a pitch of 0.5 mm and a Merton nut mechanism (Thomas 1974) drives a connecting rod that is attached to the underside of the NanoSurf IV x slideway via a loose screw thread coupling. The lead screw has been lapped to attenuate any asperities.

The Merton nut is made from PTFE and reduces any periodic errors in the manufacture of the lead screw. The loose screw coupling system causes backlash when reversing the direction of the slideway (because the male thread has to make contact with the opposing flank of the female thread) but should theoretically give high friction, point contacts when a constant speed is achieved.

The drive is bi-directional and has nine speeds: 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2 and 5 mm per minute. Figure 3.13 is a photograph of the drive mechanism.



Figure 3.13 The NanoSurf IV
drive mechanism

3.3.5 THE PROBING SYSTEM

The probe on NanoSurf IV is a Taylor-Hobson Nanostep probe that has been described elsewhere (Garratt & Bottomley 1990) and is shown schematically in figure 3.14. Slight modifications have been made to the material structure in the probe and the tip mounting rod has an extra aluminium part attached to the end of the transducer stem to allow mounting of the structure shown in figure 3.15. It was also necessary to replace one of the springs used to balance the probe (see below). NanoSurf IV uses only the flexure movement system, although the inductive displacement sensor can also be used to measure surface texture. The use of interferometry allows for direct traceability of

displacement measurements via the wavelength of the laser source. It has also been shown that the interferometers and corrections (see chapter 4) used are linear to within ± 0.02 nm (Yacoot & Downs 2000). It would be very difficult to achieve this level of accuracy with the Taylor-Hobson probe.

On NanoSurf IV the column shown in figure 3.14 is a Zerodur stage consisting of a base-plate and two cylinders running in two vee grooves, the entire stage being sprung-loaded to the NanoSurf IV bridge. The stage is used for positioning the stylus onto a sample and is counterbalanced against gravity by means of two PTFE weights (the choice of material is arbitrary here). In figure 3.10 the stage can be seen with two phosphor-bronze leaf springs loading it to the bridge and the two PTFE counter weights can be seen through the Zerodur bridge. The weights are held in place with nylon fishing line. The stage is driven by a micrometer that is de-coupled from the measurement and, more importantly, thermal loop during a measurement.

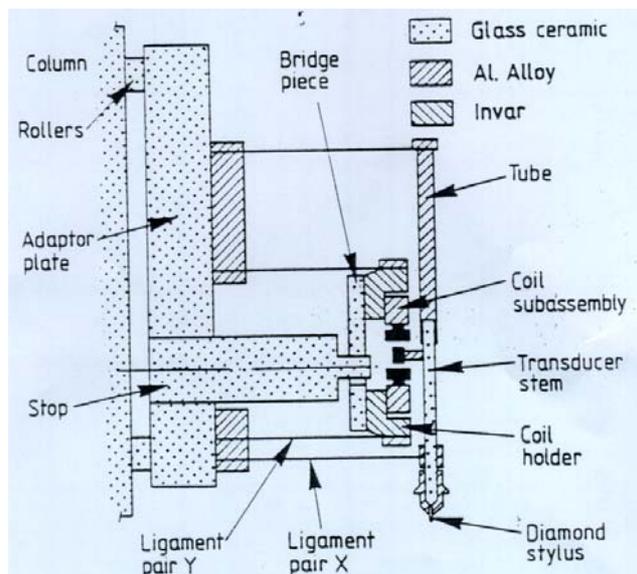


Figure 3.14 Nanostep probe assembly showing the materials used in the measurement loop (courtesy of Taylor-Hobson)

In order to incorporate a laser interferometer as the sensing element, the Nanostep probe had to be modified as shown in figure 3.15. This design complies with the Abbe principle as the polished and aluminised end of the Zerodur rod acts as the plane mirror of the cat's-eye retro-reflector for the z interferometer. The whole structure has a weight of less than 4 g, but is stiff to prevent significant torsional and flexural movement. The light-weight design ensures that the modification does not significantly alter the dynamic response of the existing flexure system. Not shown in figure 3.15 is the lens that focuses the light from the z interferometer onto the polished face of the Zerodur rod. The lens is mounted in a small carbon-fibre housing that is simply cemented to the aluminium structure (labelled aluminium in figure 3.15) using epoxy resin (see chapter 4).

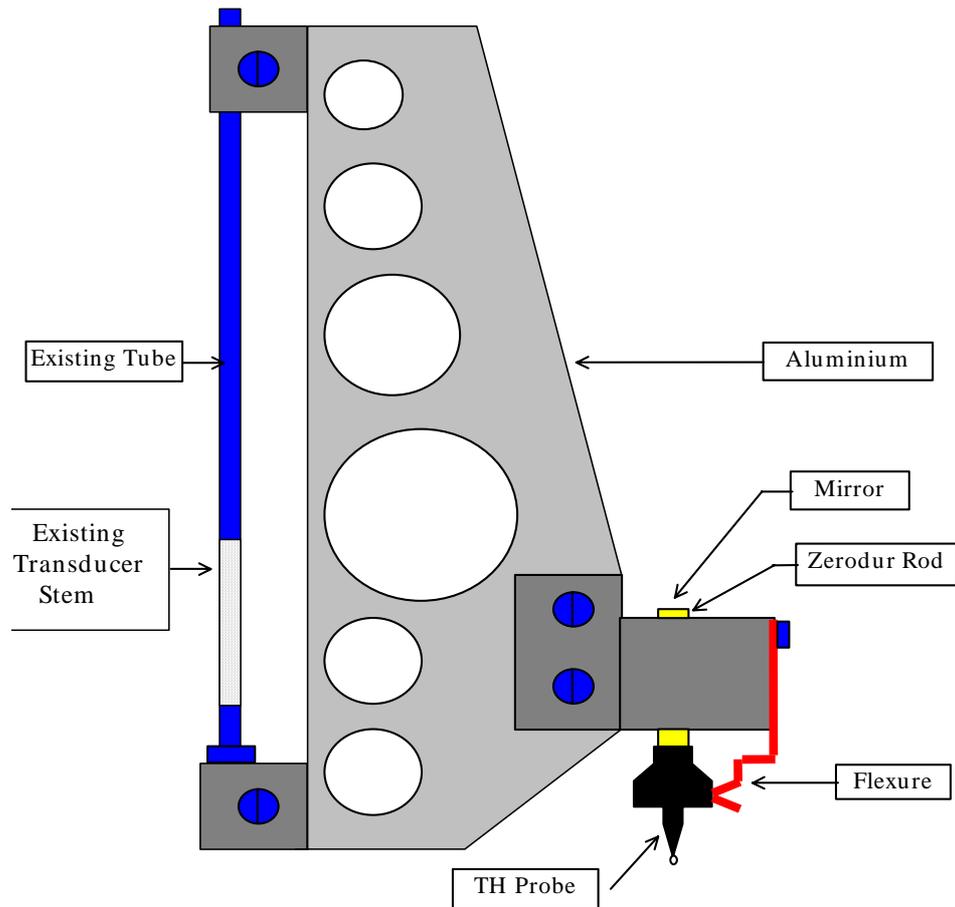


Figure 3.15 Probe modification to allow on-Abbe sensing of z displacements

The probing force on a sample is controlled by the ligament flexure system shown in figure 3.14 and a constant static force is maintained and is adjusted using two springs. The first is a tension spring for coarse adjustment. This spring had to be replaced during re-design to re-balance the system with the extra mass of the aluminium structure shown in figure 3.15. Coarse adjustment of the probing force is carried out by loosening a grub screw and adjusting the tension spring supporting rod (see Taylor-Hobson Talystep manual). A clock spring is used for fine adjustment of the probing force. When the probe is balanced the small armature inside the coil assembly is not contacting either of the magnets (shown as black T-shaped objects in figure 3.14). The Taylor-Hobson electronics

are used to measure the probing force and rely on the assumption that, when balanced, the probing force is linearly dependent upon the deflection of the armature within the coil subassembly. Table 3.3 shows the probing force for the different magnification settings on the Taylor-Hobson electronics.

Magnification	Change in force from zero to full scale on indicator / μN	Recommended central stylus force / μN
x 5000	78	98 - 195
x 10 000	39	98
x 20 000	19.5	39
x 50 000	7.8	39
x 100 000	3.9	19.5
x 200 000	1.9	19.5
x 500 000	0.78	9.8 - 19.5
x 1 000 000	0.39	9.8 - 19.5

Table 3.3 Probing force against magnification setting on the Taylor-Hobson electronics. Note that the probing force is only nominal, it has not been calibrated

The selection of the stylus is dependent on *a priori* knowledge of the approximate texture (frequency and amplitude), hardness of the artefact under test and applied load (stylus force). Stylus force is defined as the force with which the stylus bears on an artefact surface when it is positioned to bring the dial indicator on the Taylor-Hobson electronics to the centre of the dial at any selected magnification range. The force is not constant over the full measuring range but varies linearly with the distance moved by the stylus.

Because there will always be a conflict of interest between selecting a sharp stylus to resolve sharp features and keeping the load per unit area down to minimise artefact damage, stylus selection will always be a compromise.

The stylus force will also vary due to the finite inertia of the stylus/transducer when accelerating up or down asperities on the surface. This variation is, however, very small at the speeds travelled by the NanoSurf IV slideway during a measurement (Whitehouse 1990).

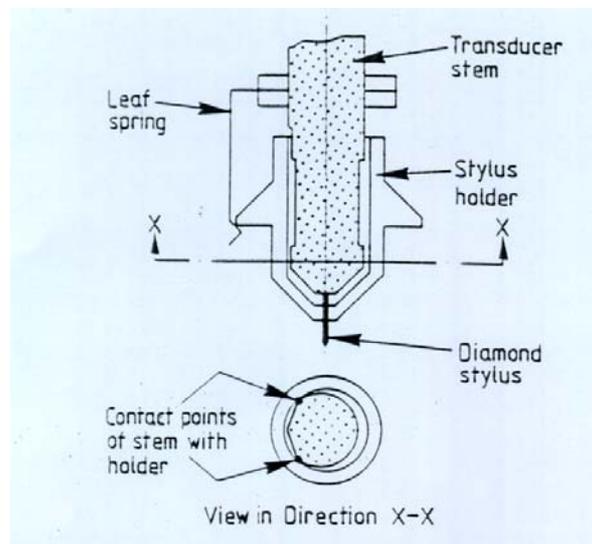


Figure 3.16 Stylus assembly showing the kinematic location (courtesy of Taylor-Hobson)

Six different types of stylus with differing radii of curvature and geometry have been purchased from Taylor-Hobson. These are 0.1, 2 and 10 μm radius truncated pyramidal (chisels), a 1 μm radius conisphere, a 5 μm radius and a 10 μm radius stylii. The Zerodur rod is designed so that a stylus can be mounted kinematically onto its end and re-locate

precisely (see figure 3.16). This ensures that any styli with non-symmetrical geometry are accurately aligned and re-locate. The stylus is held in position with a phosphor-bronze spring flexure.

3.3.6 COMPUTER

The computer that controls the measurements and auxiliary instrumentation is a Dell OptiPlex Gxi IBM compatible with a Intel Pentium 133 MHz processor, 32 Mb RAM and a 2 Gb hard drive unit.

3.3.7 VIBRATION ISOLATION

The entire NanoSurf IV instrument is housed in a basement laboratory at NPL. The instrument sits on a cast iron surface table that is mounted on a 1150 × 1150 × 160 mm concrete plinth. Springs are used to mechanically de-couple the plinth from the foundations of the building. Over the years, however, many small screws, washers, *etc.* have fallen down the gap between the plinth and the floor, and cannot be removed, with the result that the plinth is no longer so well isolated from the foundations. The instrument itself is mounted on a 56 kg stainless steel plate that sits on four anti-vibration balls. Four balls were chosen, as opposed to three, to give the instrument more stability when the top-plate load is varied and to stop the system toppling over if one of the balls burst. The gearbox and drive motor are mounted on a separate 66 kg stainless steel plate which also sits upon four anti-vibration balls. Figure 3.17 shows the two plates mounted on the surface table. The balls are made from squash balls that have been filled with high viscosity silicone oil. A brass valve stops any of the oil from escaping (see figure 3.18).



Figure 3.17 Stainless steel plates sat upon a steel table. The squash ball anti-vibration mounts are positioned at each corner of the plates. Looking at the left -hand plate one can see two of the cups that hold the squash balls

A study of the vibration levels in the laboratory in which NanoSurf IV is housed has been carried out using a piezoelectric accelerometer (Brüel & Kjaer model #4378) and amplifier (model #2526). The sensor and amplifier were calibrated in the frequency range 1 Hz to 20 kHz by comparison with a reference seismometer. The calibration factor was $31.9 \text{ ms}^{-2} \text{ V}^{-1}$. The output from the amplifier is a potential difference that is proportional to the seismic acceleration. The potential difference data was captured by a Stanford Research System FFT spectrum analyser (model SR780).



Figure 3.18 Anti-vibration squash balls

The accelerometer was bolted to a metal plate that was fixed using 'Blu-Tak' to the desired surface ensuring good mechanical coupling and providing a degree of elastic averaging. All cables were fixed using Blu-Tak to the surface to minimise mechanical noise coupled into the system via the cables. Measurements were made on the floor of the laboratory and on the stainless steel plate. The amplifier was electronically grounded during all the measurements by the shortest possible lead. This minimises power line pick up, which can be seen as spikes in the acceleration spectrum at multiples of 50 Hz. The amplifier was set-up for the highest input gain (60 dB) and a low-pass filter of 0.1 Hz was used. The pre-amp output of the amplifier was used, which is the direct output from noise. Blackman-Harris (BMH) windowing was used throughout these experiments as it was found in experiments at NPL to have the least effect on the measurement results over the frequency band being tested (Harris 1978, Roberts, Taylor, *et. al.* 1999).

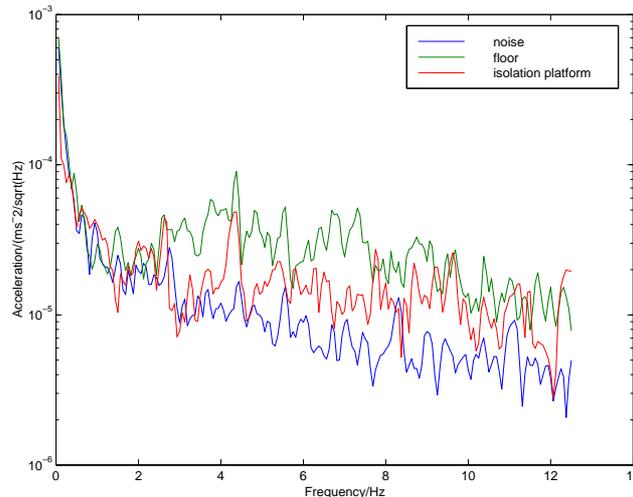


Figure 3.19 Seismic frequency spectrum over a 12.5 Hz span

The noise floor of the amplifier was measured using a dummy accelerometer consisting of a 1200 pF polystyrene capacitor in a shielded metal box connected straight to the input of the amplifier. The value 1200 pF was chosen because it is close to the internal capacitance of the accelerometer (1283 pF).

Figure 3.19 shows the results for a 12.5 Hz frequency span. It is clear that the level of vibration is not much higher than the noise level. The large peaks at frequencies less than 1 Hz are due to $1/f$ or flicker noise present due to material and surface defects in the semiconductors in the amplifier. At frequencies less than 2 Hz the plate is no better an isolator than the floor, but above 2 Hz there is a significant improvement of around 10 dB. Note that the floor in Bushy House is in a basement and this area has shown in the past to be one of the quietest, in a seismic sense, on the whole of the NPL site.

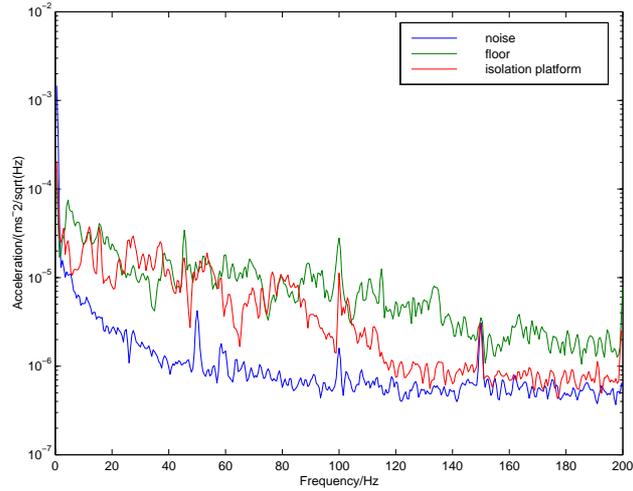


Figure 3.20 Seismic frequency spectrum over a 200 Hz span

Figures 3.20 and 3.21 show the results for the 200 Hz and 1.6 kHz scans respectively. With the exception of a small frequency band from about 25 Hz to 40 Hz the plates are isolating better than the floor. The source of the dominant feature at around 90 Hz was not identified. To calculate the amplitude of the vibration, a , the following equation is applied to the bandwidth limited acceleration data, A ,

$$a = \frac{A}{(2\pi f)^2} \sqrt{f_s k_{BMH}} \quad (3.15)$$

where f is the frequency of vibration, f_s is the sampling frequency and k_{BMH} is a factor required when using BMH windowing. Obvious peaks in the spectrum of the plate occur at 14, 28, 44 and 98 Hz which correspond to RMS amplitudes of around 13.0, 3.2, 2.6 and 0.2 nm.

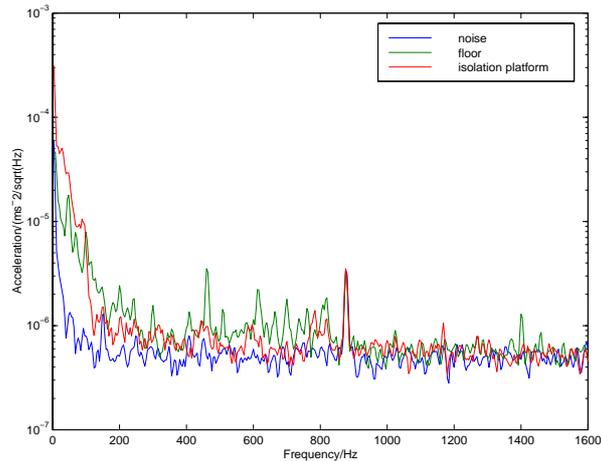


Figure 3.21 Seismic frequency spectrum over a 1.6 kHz span

The isolation plate was also tested with the motor drive on and off. Even at the highest speed, cross correlation analysis showed that there was no significant difference between the frequency spectra when the motor was on or off. The air conditioning system in an adjoining room did not affect the frequency spectrum of the plate.

To isolate the instrument from acoustic sources of noise an anechoic chamber was constructed. This was made from a steel 25 mm square frame structure with 15 mm thick medium-density fibre board (MDF) walls and roof. The walls and roof were lined with a lead-filled foam structure that is coated with aluminium foil. This foam is specifically designed to be a damping agent to acoustic noise.

3.3.8 MISCELLANEOUS

3.3.8.1 Specimen mounting plate

The specimen mounting plate is a 90 mm x 100 mm by 15 mm thick Zerodur plate that has three thin glass pads cemented to its underside using epoxy resin. The pads contact the top-plate of the tilt and level control system. The object to be measured usually has a flat underside and simply sits on the mounting plate under the probe. It is sometimes necessary to use more sophisticated mounting arrangements but it is rarely necessary to use more than the mass of the object being measured to prevent it from moving during a measurement.

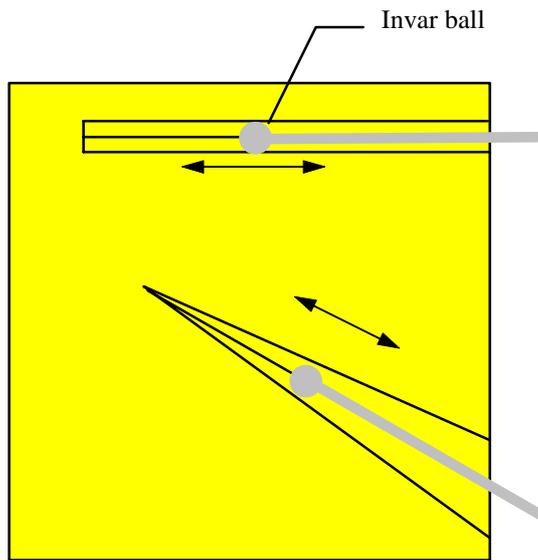


Figure 3.22 Plan view of the tilt and level plate showing the Invar balls attached to rods sat in vee-grooves

3.3.8.2 Tilt and level control

The tilt and level control system consists of a 90 mm x 100 mm by 10 mm thick top-plate and a under-plate of the same dimensions with two grooves machined into it. The position of the grooves is shown schematically in figure 3.22. The depth of the grooves decreases from right to left in figure 3.22 and two Invar balls attached to the end of thin

rods can move along the long axis of each groove. The movement of the balls is controlled by a screw operated manually. Provided the object to be measured is placed centrally on the specimen mounting plate the tilt and level control provides movement in nominally orthogonal directions.

3.4 REFERENCES

- Abbe E 1890 Messapparate für physiker, *Zeits Für Instrumenten-kunde* **10** 446-448
- Atherton P D 1998 Nanometre precision mechanisms *Measurement + Control* **31** 37-42
- Beranek L L 1988 *Noise and vibration control* (Institute of Noise Control Engineering)
- Berthold III J W, Jacobs S F, Norton M A 1977 Dimensional stability of fused silica, Invar and several ultra-low thermal expansion materials *Metrologia* **13** 9-16
- Birch K P, Downs M J 1994 Correction to the updated Edlén equation for the refractive index of air *Metrologia* **31** 315-316
- Brenan C J H, Charette P G, Hunter I W 1992 Environmental isolation platform for microrobot system development *Rev. Sci. Instrum.* **63** 3492-3498
- Bryan J B 1979 The Abbé principle revisited: an updated interpretation *Prec. Eng.* **1** 129-132
- Cebon D, Ashby M F 1994 Materials selection for precision instruments *Meas. Sci. Technol.* **5** 296-306
- Chetwynd D G 1987 Selection of structural materials for precision devices *Prec. Eng.* **9** 3-7
- Edlén B 1966 The refractive index of air *Metrologia* **2** 71-80
- Filinski I, Gordon R A 1974 The minimization of ac phase noise in interferometric systems *Rev. Sci. Instrum.* **65** 575-586
- Franks A 1999 Private communication with the author
- Franse J 1990 Manufacturing techniques for complex shapes with submicron accuracy *Rep. Prog. Phys.* **53** 1049-1094
- Fritschel P, Malvalvala N, Shoemaker D, Sigg D, Zucker M, González G 1998 Alignment of an interferometric gravitational wave detector *Appl. Opt.* **37** 6734-6747

- Garratt J D, Bottomley S C 1990 Technology transfer in the development of a nanotopographic instrument *Nanotechnology* **1** 28-43
- Giazotto A 1990 Interferometric detection of gravitational waves: the Virgo project *Proc. Elizabeth and Frederick White Research Conf.* 110-135
- Hadlington S 1999 The cleanest cut of all *EPSRC Newslines* Winter 20-21
- Harris F J 1978 On the use of windows for the harmonic analysis with discrete Fourier transform *Proc. IEEE* **66** 51-84
- Hicks T R, Atherton P D 1997 *The NanoPositioning Book* (Queensgate Instruments Ltd)
- Hughes E B 1996 Measurement of the linear thermal expansion coefficient of gauge blocks by interferometry *Proc. SPIE* **2088** 179-189
- Kaye G W C, Laby T H 1989 *Tables of physical and chemical constants and some mathematical functions* (Longman Scientific & Technical)
- Leach R K 1995 A novel 3-axis CMM probe calibration facility *NPL Report MOM* **126** 1-22
- Lindsey K 1982 Innovation: Polishing off the Japanese *The Sunday Times* 3 January
- Lindsey K, Smith S T, Robbie C J 1988 Sub-nanometre surface texture and profile measurement with NANOSURF 2 *Ann. CIRP* **37** 519-522
- Lindsey K 1992 Tetraform grinding *Proc. SPIE* **1573** 129-135
- Liu X, Chetwynd D G, Gardner J W 1997 Surface characterisation of electro-active thin polymeric bearings *7th Int. Conf. on Metrology and Properties of Engineering surfaces* 274-280
- McKeown P A 1986 High precision manufacturing in the British economy, James Clayton Lecture *Inst. Mech. E. Soc.* **200**
- Monteiro A F, Smith S T, Chetwynd D G 1996 A super-precision linear slideway with angular correction in three axes *Nanotechnology* **7** 27-36
- Monteiro A F, Chetwynd D G 1997 Load-dependent behaviour in PTFE-glass slideways *Prec. Eng.* **20** 153-160
- Nakazawa H 1994 *Principles of precision engineering* (Oxford Science Publications)

- Newell D B, Richman S J, Nelson P G, Stebbins R T, Bender P L, Faller J E, Mason J 1997 An ultra-low-noise, low-frequency, six degrees of freedom active vibration isolator *Rev. Sci. Instrum.* **68** 3211-3219
- Petersen K E 1982 Silicon as a mechanical material *Proc. IEEE* **70** 420-456
- Reay N K 1993 Sub-nanometre closed-loop positioning for optics and X-Y stage control using capacitance displacement sensors and piezo-electric actuators *Proc. SPIE* **2088** 150-159
- Riehle F 1998 Use of optical frequency standards for measurements of dimensional stability *Meas. Sci. Technol.* **9** 1042-1048
- Roberts M, Taylor P, Gill P 1999 Laser linewidth at the sub-hertz level *NPL Report CLM8* 1-65
- Schellekens P, Rosielle N, Vermeulen J, Vermeulen M, Wetzels S, Pril W 1998 Design for high precision: current status and trends *Ann. CIRP* **2**
- Schott Glaswerke Zerodur - *Precision from glass ceramics*. (Technical sales brochure)
- Schouten C H, Rosielle P C J N, Schellekens P H J 1997 Design of a kinematic coupling for precision applications *Prec. Eng.* **20** 46-52
- Smith S T, Chetwynd D G, Bowen D K 1987 Design and assessment of monolithic high precision translation mechanisms *J. Phys. E: Sci. Instrum.* **20** 977-983
- Smith S T, Chetwynd D G 1992 *Foundations of ultraprecision mechanism design* (Gordon and Breach Science Publishers)
- Smith S T, Harb S, Chetwynd D G 1993 Metrological characterization of polymeric bearings at the nanometre level *J. Phys. D: Appl. Phys.* **25** A240-A248
- Thomas C G 1974 *Engineering Metrology* (Butterworth & Co.)
- Taniguchi N 1983 Current status in and future trends of ultra-precision machining and ultrafine materials processing *Ann. CIRP* **32**.

Taniguchi N 1996 *Nanotechnology: integrated processing systems for ultra precision and ultrafine products* (Oxford University Press)

Timoshenko S P, Young D H, Weaver W 1974 *Vibration problems in engineering* 4th Edition (John Willey & Sons Inc.)

Whitehouse D J 1990 Dynamic aspects of scanning surface instruments and microscopes
Nanotechnology **1** 93-102

Yacoot A, Downs M J 2000 The use of x-ray interferometry to investigate the linearity of the NPL Differential Plane Mirror Interferometer *Meas. Sci. Technol.* **11** 1126-1130

CHAPTER 4

THE METROLOGY

"That an electron has spin, location and so forth even when it is not being measured."

Albert Einstein (1879 - 1955)

4.1 WHICH TYPE OF INTERFEROMETER?

NanoSurf IV requires traceable measurement of displacement in two orthogonal directions with an uncertainty of measurement of less than 1 nm. There are many configurations of optical interferometer that can achieve this goal, a few examples of which are given here. Many more configurations were considered, but they are far more bulky than the simple Michelson interferometer (Michelson 1893) that was adopted, and require very specialised optics and electronics to operate. Other methods of displacement measurement, such as inductive or capacitive transducers were considered but they do not offer a direct route to traceability and would not deliver the full dynamic range and accuracy required by NanoSurf IV (Downs & Nunn 1998).

4.1.1 HETERODYNE INTERFEROMETRY

Perhaps the most widely available interferometer commercially is the heterodyne interferometer (Liu & Klinger 1979, Demarest 1998). The basic form of a heterodyne

interferometer is shown in figure 4.1. Typically, a Zeeman-stabilised laser source emits two closely spaced orthogonal polarisation frequencies, ν_1 and ν_2 , separated by around 1 MHz (Williams 1993). A beam-splitter separates off part of the signal from both polarisation states which are mixed on detector D1 to provide a reference beat $\nu_1 - \nu_2$. The transmitted components are then split at a polarising beam-splitter (PBS). Components ν_1 and ν_2 traverse the fixed reference arm and the measurement arm, respectively, and are recombined at the polarising beam-splitter and detected at D2. In this case the detected beat is $\nu_1 - \nu_2$ when the measurement retroreflector is stationary. During motion of the cube-corner, frequency ν_2 is Doppler-shifted by $\Delta\nu$, the sign being dependent on the direction of motion. The detected beat is thus given by $\nu_1 - \nu_2 \pm \Delta\nu$. The reference and Doppler-shifted beats are counted independently and subtracted to give $\Delta\nu$. This beat difference gives rise to a count rate, dN/dt , which is equal to $\nu(2s/c)$, where s is the velocity of the cube-corner and c is the velocity of light. Integration of the count over time t leads to a fringe count, $N = 2d/\lambda$.

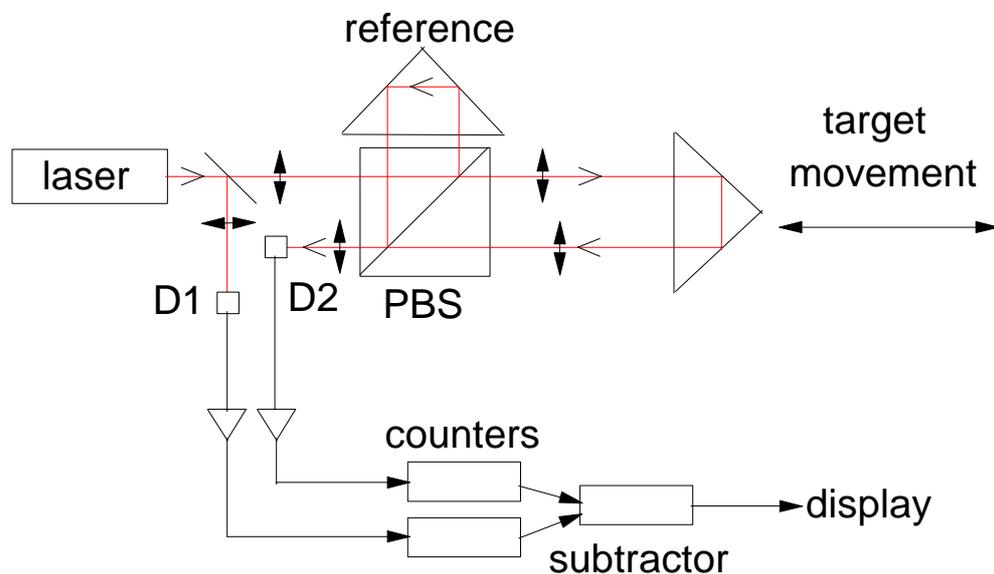


Figure 4.1 A heterodyne interferometer

With a typical reference beat of around 1 MHz, it is possible to monitor $\Delta\nu$ values up to 1 MHz before introducing ambiguities due to the beat crossing through zero. This limits the target speed possible in this case to $< 0.3 \text{ m s}^{-1}$, which could be a constraint in some applications. An alternative method of producing a two-frequency laser beam is to use an acousto-optic frequency shifter. This method has the advantage that the frequency difference can be much higher, so that higher count rates can be handled (Hariharan 1992).

Many variations on the theme in figure 4.1 have been developed which improve both the speed of response, measurement accuracy and resolution. Modern commercial heterodyne interferometers can be configured to measure both displacement and angle (Peggs, Lewis, *et. al.* 1999).

There are, however, fundamental drawbacks of the heterodyne technique that limit their accuracy to a few nanometres. These drawbacks are mainly caused by periodic non-linearities caused by polarisation leakage in the optics (Rosenbluth & Bobroff 1990), but can also be caused by electronic limitations of the phase meters (Oldham, Kramar, *et. al.* 1993), Wu & Su 1996, Stone & Howard 1998), ellipticity of the Zeeman laser beam, stray reflections (Wu & Deslattes 1998) and beam misalignment (De Freitas & Player 1993). Heterodyne laser-based interferometers are, on the whole more complex than a single-frequency Michelson design and could not easily be miniaturised enough to interface with the probing system on NanoSurf IV.

4.1.2 MICHELSON/JAMIN INTERFEROMETRY

Figure 4.2 is a schema of a differential plane mirror interferometer developed at NPL (Downs & Nunn 1998). The beam from the laser is separated by a Jamin beam-splitter, creating two beams which are displaced laterally and parallel to each other. Figure 4.2 shows how polarisation optics can be used to convert the Michelson part of the interferometer into a plane mirror configuration, but a retro-reflecting configuration could just as easily be employed. After a double passage through the waveplate, the beams are transmitted back to the Jamin beam-splitter where they recombine and interfere. The design of the Jamin beam-splitter coating is such that the two signals captured by the photo-detectors are in phase quadrature and so give the optimum signal-to-noise conditions for fringe counting and sub-dividing (see §4.2.7). In this configuration only the differential motion of the mirrors is detected.

With the design shown in figure 4.2 sub-nanometre accuracy can be achieved (Downs & Nunn 1998). The only drawback with this design is the need for a much more bulky system than that described later in §4.1.4. These interferometers, which are available commercially through NPL, are typically 120 x 100 x 40 mm in size and weigh around 250 g. This would have been far too bulky for the *z* axis interferometer on NanoSurf IV and would have required considerable resource to miniaturise.

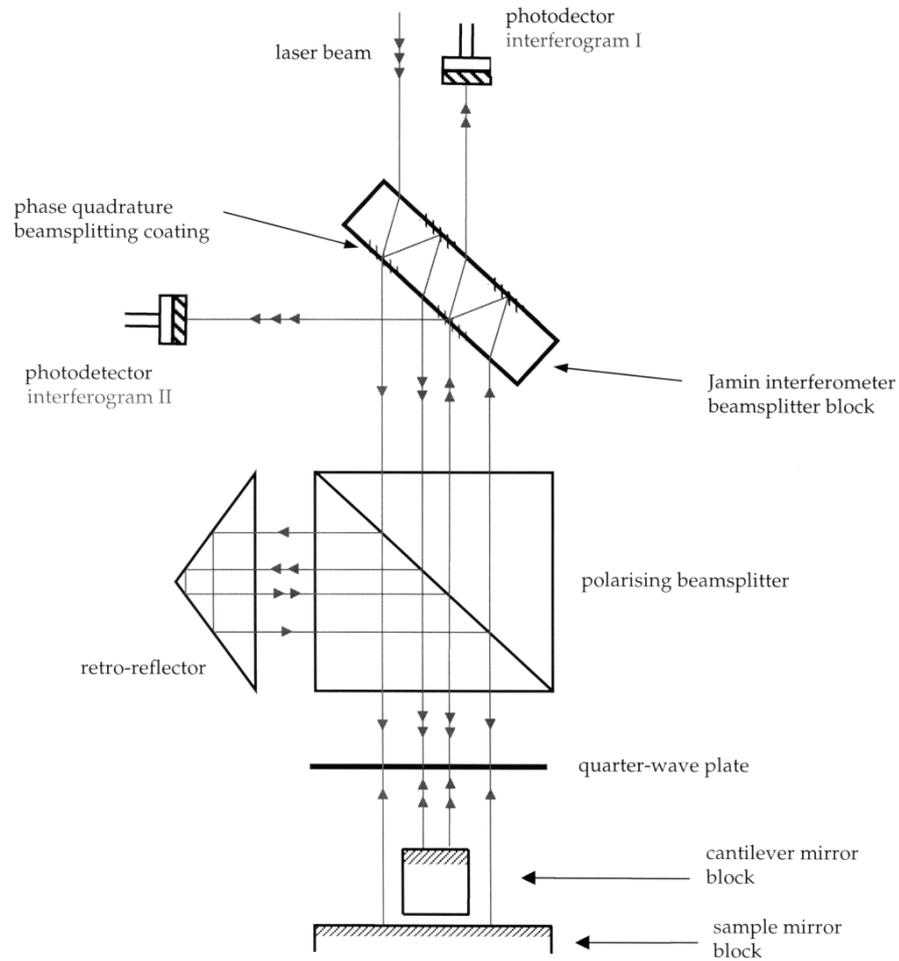


Figure 4.2 A differential plane mirror interferometer. Note the arrow convention on the beams: at the Jamin the three-arrow beam is split into a single and a double-arrow beam. This makes it easy to follow the paths through the interferometer

4.1.3 SWEPT-FREQUENCY INTERFEROMETRY

Swept-frequency interferometry using laser diodes or other solid-state lasers is becoming popular due to the versatility of its sources and its ability to measure length absolutely (Barwood, Gill, *et. al.* 1993). At the moment these interferometers are achieving high resolutions but relatively low accuracies. Their present state-of-the-art would not make

them suitable for the interferometers on NanoSurf IV, but a short description of their operating principles is given for completeness.

Consider the case of a laser aligned to an interferometer of free spectral range ν_{FSR} . If the output of the laser is scanned through a frequency range ν_s , N fringes are generated at the output of the interferometer (Barwood, Gill, *et. al.* 1998). Provided the frequency scan range is accurately known, the free spectral range and hence the optical path length, L , may be determined from counting the number of fringes. For a Michelson or Fabry-Pérot interferometer in vacuum, the optical path length is given by

$$L = \frac{c}{2\nu_{FSR}} = \frac{Nc}{2\nu_s}. \quad (4.1)$$

It is generally convenient to use feedback control techniques to lock the laser to particular fringes at the start and finish of the scan and so make N integral. For scans of up to several gigahertz, two lasers are typically used, which are initially tuned to the same frequency. One laser is then scanned by ν_s , and the difference frequency counted directly as a beat by means of a fast detector with around 25 GHz frequency response. This, together with the number of fringes scanned, enables the optical path length to be determined. The number and size of the sweeps can be used to improve the accuracy and range of the interferometer.

4.1.4 MICHELSON INTERFEROMETRY

The two interferometers on NanoSurf IV are retro-reflecting Michelson configurations and are described in §4.2. Owing to the fact that the light from the laser is collimated

before entering the interferometers, the actual configuration is Twyman-Green. This configuration was chosen as it has been shown to give sub-nanometre uncertainty in the past (Downs, Birch, *et. al.* 1995), it lacks polarisation optics and therefore avoids the errors due to polarisation leakage described in §4.1.1, it is available at NPL and is inherently simple.

4.2 DESIGN OF THE INTERFEROMETERS ON NANOSURF IV

The next section describes the design of the interferometers on NanoSurf IV. Considerable depth has been given to the description of the laser source as this provides the primary route of traceability to the definition of the metre.

4.2.1 THE LASER SOURCE

A frequency-stabilised laser is used for the source of the interferometers on NanoSurf IV to ensure a high enough accuracy (see chapter 6) and because the type of laser used has proved to be very reliable in the past for the interferometers used on the well-established gauge block interferometers used at NPL (Lewis 1994). The calibration of the two interferometric displacement transducers is carried out every three years by calibrating the frequency of the laser source (see §5.10.2.1).

4.2.1.1 Theory of the helium-neon laser

The tube of a continuous wave helium-neon (He-Ne) gas laser contains a mixture of approximately 8 parts of helium to 1 part of neon at a total pressure of a few millibars. The laser consists of an optical cavity, similar to that of a Fabry-Pérot etalon, formed by a

plasma tube with optical quality mirrors at both ends. The gas in the tube is excited by a high voltage discharge of approximately 1.5 to 2.5 kV, at a current of approximately 5 to 6 mA. The discharge creates a plasma in the tube that emits radiation at various wavelengths corresponding to the multitude of allowed transitions in the helium and neon atoms.

The coherent radiation emitted by the He-Ne laser at approximately 632.8 nm wavelength corresponds to the $3s_2 - 2p_4$ transition in neon. The excited $3s_2$ level is pumped by energetic $2s_0$ helium atoms colliding with the neon atoms; the $2s_0$ helium energy level is similar in energy to the $3s_2$ level of neon and the lighter helium atoms are easily excited into the $2s_0$ level by the plasma discharge (see figure 4.3). The excess energy of the collision is approximately equal to kT , *i.e.* it is easily removed by the atoms in the plasma as kinetic (thermal) energy.

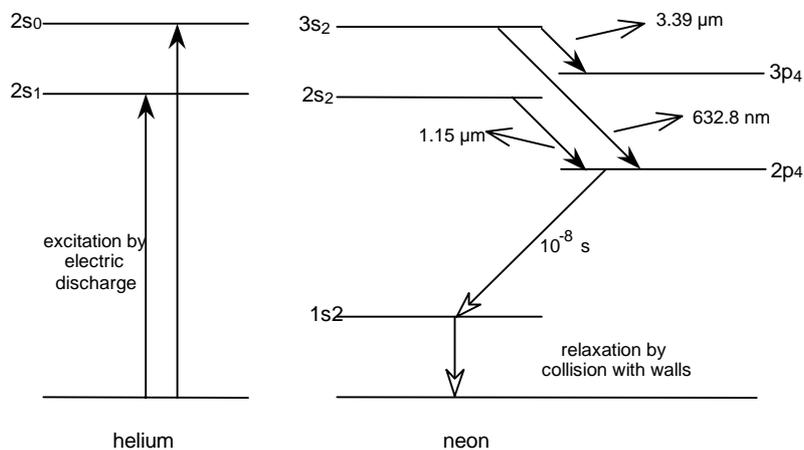


Figure 4.3 Energy levels in the He-Ne gas laser for 632.8 nm radiation (Milonni & Eberly 1988)

The collisional pumping of the $3s_2$ level in neon produces the selective excitation or population inversion that is required for lasing action. The $2p$ neon state decays in 10^{-8} seconds to the $1s$ state, maintaining the population inversion. This state relaxes to the

ground state by collision with the walls of the plasma tube. The laser gain is relatively small and so losses at the end of the mirrors must be minimised by using a high reflectance coating, typically 99.9%. The output power is limited by the fact that the upper lasing state reaches saturation at quite low discharge powers, whereas the lower state increases its population more slowly. After a certain discharge power is reached, further increase in the power leads to a decrease in the population inversion, and hence lower light power output.

The 632.8 nm operating wavelength is selected by the spacing of the end mirrors, *i.e.* by the total length of the optical cavity, l_c . The length of the cavity must be such that the waves reflected by the two end mirrors are in phase. The wavelength of successive axial modes are then given by

$$2l_c = m\lambda . \quad (4.2)$$

These modes are separated in wavenumber by

$$\Delta\sigma = \frac{1}{2l_c} \quad (4.3)$$

or in terms of frequency

$$\Delta\nu = \frac{c}{2l_c} \quad (4.4)$$

where c is the speed of light in a vacuum. This would lead to a series of narrow lines of similar intensity in the spectrum, if it were not for the effects of Doppler broadening and

the Gaussian distribution of atoms available for stimulated emission. The availability of atoms for stimulated emission is given by

$$A(\sigma) = \frac{1}{a\sqrt{\pi}} e^{-\frac{\sigma^2}{a^2}} \quad (4.5)$$

where

$$a^2 = \frac{2kT\sigma_0^2}{Mc^2} \quad (4.6)$$

k is Boltzman's constant, T is the absolute temperature of the gas, M is the molecular mass of the gas and σ_0 is the line centre. This distribution has a half width at

$$w_\sigma = 2a\sqrt{\ln 2} . \quad (4.7)$$

When a particular mode is oscillating, there is a selective depopulation of atoms with specific velocities (laser cooling) which leads to a dip in the gain profile. For modes oscillating away from the centre of the gain curve the atomic population for the two opposite directions of propagation are different due to the equal but opposite velocities. For modes oscillating at the centre of the gain curve, the two populations become a single population of effectively stationary atoms. Thus a dip in the gain profile occurs at the centre of the gain curve - the so called Lamb dip (CCDM 1973). The position of the Lamb dip is dependent on other parameters of the laser such as the position of the gain curve and can be unstable.

For early lasers with typical cavity lengths of 1 m the mode spacing was 0.5 m^{-1} , with a gain profile width of approximately 5.5 m^{-1} . Thus several axial modes were present in the gain profile with gains sufficient for laser action, and so two or more modes would operate simultaneously, making the laser unsuitable for coherent interferometry. By using a shorter tube and then carefully lowering the power of the discharge and hence lowering the gain curve, it is possible to achieve single mode operation.

Having selected a single mode, the fundamental limitation to the linewidth (and thus the temporal coherence) is spontaneous noise fluctuations which broaden the line into a Lorentzian function given by (Jaseja, Javan, *et. al.* 1963)

$$w_{\sigma} = \frac{8\pi hc^2 \sigma w_c^2}{P} \quad (4.8)$$

or

$$\Delta \nu = \frac{\pi h \nu}{P} (\Delta \nu_c)^2 \quad (4.9)$$

where P is the operating power of the laser, h is Plank's constant and w_c is the linewidth due to the cavity. However, a more common-place limitation is the thermal vibration of the tube size which directly affects the wavelength of the oscillating mode. For the lowest frequency stretching mode of the laser tube

$$w_{\sigma} = \sigma \sqrt{\frac{2kTV}{E}} \quad (4.10)$$

where V is the volume of the tube and E is the Young's modulus of the tube material.

4.2.1.2 Single mode laser wavelength stabilisation schemes

To allow a laser to be used in interferometry with coherence lengths above a few millimetres it must operate in a single mode and there have been many proposed schemes for laser stabilisation (White 1965).

The Lamb dip, mentioned above, was used in an early stabilisation scheme. Here the intensity of the output beam was monitored as the length of the cavity was modulated, for example by piezo-electric translators (PZTs). Alternatively, mirrors external to the laser cavity are used that could be modulated - the output intensity being monitored and the laser locked to the centre of the Lamb dip (Rowley & Wilson 1972). The reproducibility of lasers locked to the Lamb dip is limited by shift of the Lamb dip centre as the pressure of the gas inside the laser tube varies and also by a discharge current dependent shift. The large width of the Lamb dip itself (about 5×10^{-7} of the laser frequency) also limits the frequency stability obtainable from this technique.

Use has also been made of tuneable Fabry-Pérot etalons in a similar system. Other groups have locked the output of one laser to the frequency of a second stabilised laser. Others have used neon discharge absorption cells (White, Gordon, *et. al.* 1964) where the laser was locked to the absorption spectrum of neon in an external tube, the theory being that the unexcited neon would have a narrower linewidth than the neon in the laser discharge.

4.2.1.3 Laser frequency-stabilisation using saturated absorption

The technique with the greatest stability is used in the Primary Reference lasers which realise the UK's Primary Standard of Length and involves controlling the length of the laser cavity to alter the wavelength and locking the wavelength to an absorption line in saturated iodine vapour (Brillett & Cérez 1981, Cérez & Bennett 1979, Cérez, Brillett, *et. al.* 1981). This is a very stable technique since the absorption takes place from a thermally populated energy level that is free from the perturbing effects of the electric discharge in the laser tube.

If the output beam from a laser is passed straight through an absorption cell, then absorption takes place over a Doppler broadened transition. However, if the cell is placed in a standing wave optical field the high intensity laser field saturates the absorption and a narrow dip appears at the centre of the absorption line corresponding to molecules which are stationary or moving perpendicular to the direction of beam propagation. This dip produces an increase in the laser power in the region of the absorption line. The absorption line is reproducible and insensitive to perturbations. The line width is dependent on the absorber pressure, laser power and energy level lifetime. Saturated absorption linewidths are typically less than 1×10^{-8} of the laser wavelength.

In a practical application an evacuated quartz cell containing a small iodine crystal is placed in the laser cavity and temperature controlled to 23 °C. The laser mirrors are mounted on PZTs and the end plates are separated by Invar bars to ensure a thermally stable cavity. A small frequency modulation is then applied to one of the PZTs. This leads to an amplitude modulation in the output power which is detected using a phase sensitive detector and fed back to the other PZT as a correction signal. The frequency control system employs a photodiode, low noise amplifier, coherent filter and phase

sensitive detector followed by an integrating filter. Figure 4.4 is a schema of the iodine-stabilised He-Ne instrumentation.

Detection of the absorption signal at the laser modulation frequency results in a first derivative scan that shows the hyperfine components superimposed on the sloping background of the neon gain curve. The laser may be servo-locked to any of these lines, the frequency of which has been fixed internationally at the time of the re-definition of the metre in 1983 in terms of the speed of light (Editor's note in Metrologia 1983).

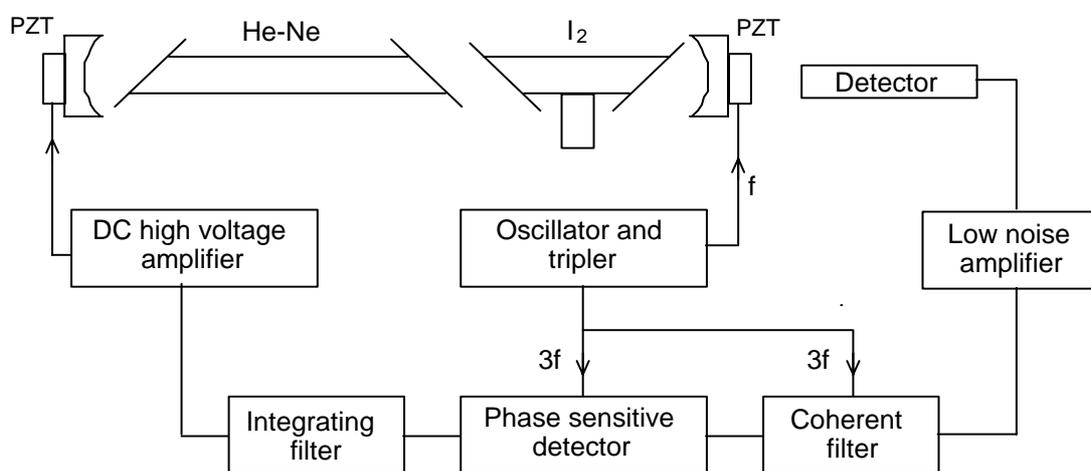


Figure 4.4 Schema of an iodine-stabilised He-Ne laser

Iodine-stabilised He-Ne lasers can achieve frequency stability (Allan 1966) of a few parts in 10^{13} over a period of a few minutes with long term reproducibility of a few parts in 10^{11} .

4.2.1.4 Zeeman-stabilised 633 nm lasers

An alternative technique to saturated absorption is used in the commercial laser used for the NanoSurf IV interferometers. The method of stabilisation is based on the Zeeman effect (Morris, Ferguson, *et. al.* 1975, Umeda, Tsukiji, *et. al.* 1980, Fellman, Junger, *et. al.* 1987). A longitudinal magnetic field is applied to a single mode He-Ne laser tube, splitting the normally linearly polarised mode into two counter-rotating circular polarisations. A field strength of 0.02 T is sufficient to split the modes, which remain locked together at low magnetic field, to produce the linear polarisation. These two modes differ in frequency by 300 kHz, around a mean frequency corresponding to the original linear mode (Baer, Kowalski, *et. al.* 1980).

The wavelength difference between the two modes is due to each of the two modes experiencing a different refractive index and therefore, different optical path length, in the He-Ne mixture. This arises due to magnetic splitting of an atomic state of neon, shown in figure 4.5.

The $\Delta m = +1$ mode couples with the left polarised mode and the $\Delta m = -1$ mode couples with the right polarised mode. The relative frequencies of the polarisation modes are given by

$$\omega_{\pm} = \frac{cN}{2Ln_{\pm}} \quad (4.11)$$

where L is the cavity length, n_{\pm} the refractive index for the mode and N the axial quantum number (Rowley 1990).

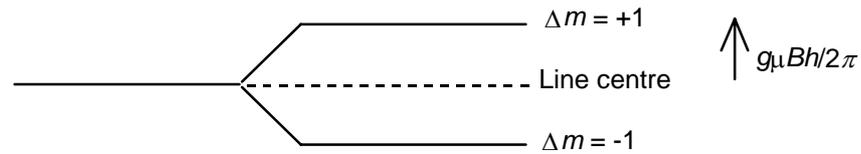


Figure 4.5 Magnetic splitting of neon - g is the Landé g factor, μ the Bohr magneton, B the field strength, and h is Planck's constant

The important feature of the Zeeman split gain curve is that the position of ω_0 does not vary with magnetic field strength - it remains locked at the original (un-split) line centre, and thus a very stable lock point. If one combines the two oppositely polarised components, one observes a heterodyne beat frequency between them given by

$$\Delta\omega = \omega_+ - \omega_- = \frac{cN}{2L} \left(\frac{1}{n_+} - \frac{1}{n_-} \right) \quad (4.12)$$

which is proportional to $\omega_0[\chi_+(\nu) - \chi_-(\nu)]$, where $\chi_+(n)$ and $\chi_-(n)$ are dispersion functions for the left and right polarised modes respectively. For a more complete derivation see Tomlinson & Fork (1968). As the laser is tuned by altering the cavity length, L , the beat frequency will pass through a peak that corresponds to the laser frequency being tuned to ω_0 .

This tuning curve can be used as an error signal for controlling the laser frequency. The particular method used to modulate the laser cavity is usually thermal expansion. A thin foil heater is attached to the laser tube and connected to a square-root power amplifier. Two magnets are fixed onto the tube to provide the axial magnetic field. A polarising is

used, together with a photodetector and amplifier to detect the 300 kHz beat frequency. This error signal is fed to various stages of counters and amplifiers and then to the heater.

The laser tube requires a period of approximately 10 minutes to reach the correct temperature corresponding to the required tube length for operation at frequency ω_0 . A phase-locked loop circuit then fine-tunes the temperature and length of the cavity to stabilise the laser at the correct frequency. This last process takes only a few seconds to achieve lock. The frequency stability of the laser is 5×10^{-10} for 1 s averages and is white-noise limited for averaging times between 100 ms and 10 minutes. The day-to-day reproducibility of the laser frequency is typically $\pm 5 \times 10^{-10}$. There is also a linear drift of frequency with the total amount of time for which the laser has been in operation. This is due to clean up of the helium-neon mixture whilst undergoing discharge. The rate of drift is unique to each laser, but is stable with respect to time, and can be ascertained after a few calibrations of the laser frequency. As an example, Sasagawa & Zumberge (1989) showed drift rates of 0.3 to 5.7 ± 0.5 MHz per calendar year, though these were for frequency against date, rather than against operational time. Rowley (1990) reported a drift rate of -1×10^{-11} per hour of operation.

4.2.1.5 Frequency calibration of the Zeeman-stabilised 633 nm laser

The calibration of the laser frequency is achieved by launching the light from the Zeeman-stabilised laser into a multi-mode optical fibre, which terminates near to the primary iodine stabilised laser. The beam combines with the Primary laser via a and the beat signal between the two frequencies is measured with a photodetector (see figure 4.6). This beat frequency is easy to detect, and monitor via a computer, over a number of

hours. A typical value of the beat signal is 260 MHz, with the iodine standard (g component) at approximately 473 612 345 MHz.

The Zeeman-stabilised laser emits two polarisations which are separated by 0.5 MHz. The two components are present in the beam at the interferometer which thus measures with a wavelength that is the mean of the two components. During the laser calibrations, beats between each of these frequencies and the iodine frequency are measured. The mean of these is deemed to be the calibrated wavelength of the Zeeman-stabilised laser under test.

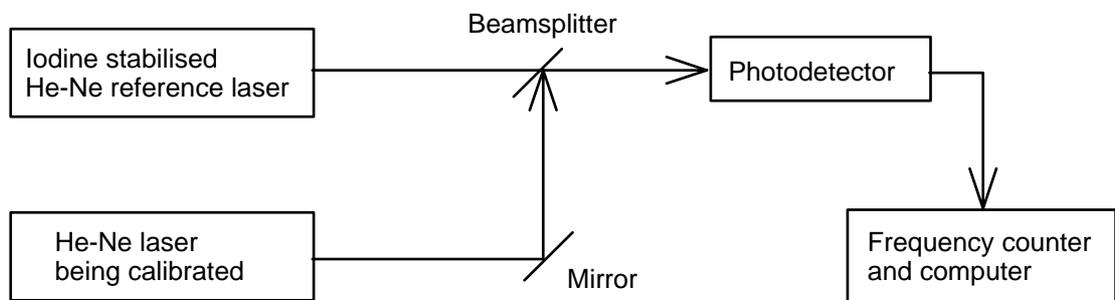


Figure 4.6 Calibration scheme for Zeeman-stabilised laser

The Zeeman-stabilised laser incorporates a modulation signal, which is applied to the heaters to achieve the lock point. This signal is a square wave of frequency 3 to 5 Hz. This causes a periodic contraction and expansion of the laser tube, and thus, the frequency of the laser undergoes a sinusoidal modulation. The depth of this modulation is 7 MHz peak-to-peak and approximately 70 ° out of phase with the square wave modulation signal.

The heater signal is inverted and can be used as a synchronisation signal for the timing of phase stepping digitisation. Thus it is important to calibrate the laser at the

synchronisation point, that occurs at the rising edge of the inverted signal (see figure 4.7). By using a frequency to voltage converter, it is possible to synchronise the frequency determination to the correct edge in the modulation signal.

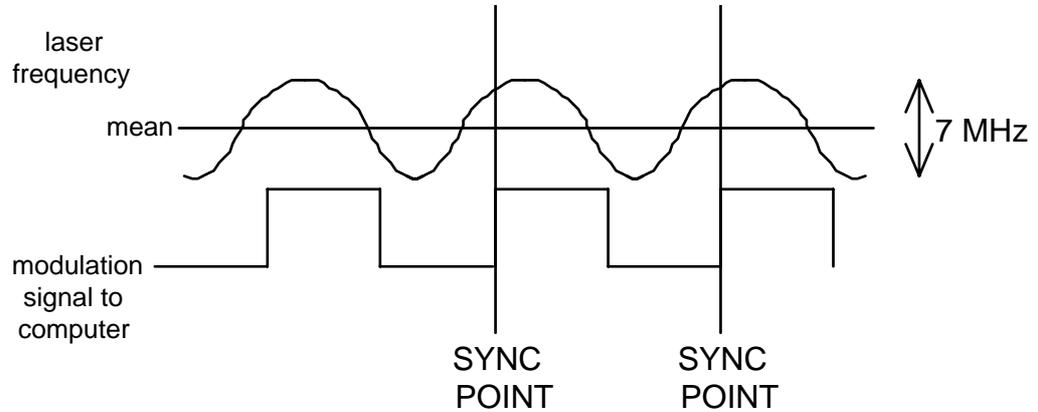


Figure 4.7 Synchronisation diagram for red Zeeman-stabilised laser

4.2.2 THE LAUNCH OPTICS

The launch optics for the interferometers are shown schematically in figure 4.8. Lenses L1 and L2, the BS and mirrors M2, M3 and M4 are mounted on a Spindler & Hoyer Microbench system (see the top of figure 3.8) and with the exception of mirror M1 and the interferometers, the entire launch optics are mounted on hardened steel plates attached to the roof of the instrument chamber. The actual laser is a 633 nm nominal wavelength Zeeman-stabilised laser with an output power of around 5 mW (Browne & Sharpe Spectrum model SRL). The two counter-rotating circularly polarised beams from the laser enter an acousto-optic modulator (AOM) where they are shifted in frequency by about 20 MHz. For this interferometer design the main purpose of the AOM is to reduce the amount of optical feedback into the laser cavity that can cause instabilities in the frequency-stabilisation scheme. Any light that travels back through the AOM receives a further 20 MHz frequency shift and is out of the resonant range of the laser cavity. The

laser and AOM are kinematically mounted on a two-plate arrangement that allows adjustment in a direction orthogonal to the beam propagation and also allows tilting and rotational adjustments about a horizontal axis.

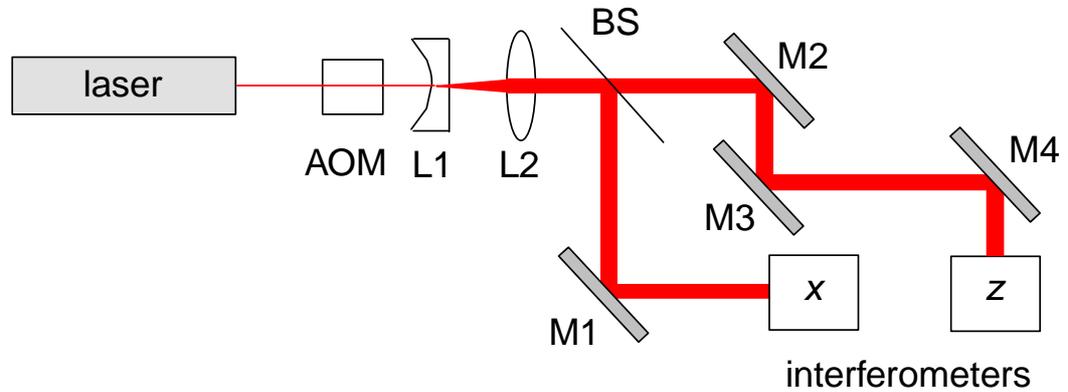


Figure 4.8 The launch optics. Annotations are described in the text.

Lenses L1 and L2 form a Galilean telescope that collimates and expands the beam by a factor of four. Lens L2 can be moved along the axis of propagation using a sprung-loaded lens mount and a micrometer (see below). To determine the values of the focal lengths of L1 and L2, the equations for their total length along the axis of propagation of the beam, L_T , and the magnification of the telescope, M , are solved simultaneously thus

$$L_T = f_2 - f_1 \text{ and } M = -\frac{f_2}{f_1} \quad (4.13)$$

where f_1 and f_2 are the focal lengths of L1 and L2 respectively, the magnification is set at 4 and the L1 to L2 distance can only be a certain size owing to practical reasons. It was also important to calculate whether plano-concave or plano-convex lenses could be used as these will minimise spherical aberration of the beam (Welford 1986). To do this the laser

spot size due to spherical aberration was compared to the spot size due to diffraction and the following inequality should apply (Melles-Griot 1998)

$$\frac{0.067 f}{(f/d)^3} \leq \frac{\lambda f}{w\pi} \quad (4.14)$$

where w is the beam waist and d is the lens diameter. Only at considerable cost were plano-convex lenses available off-the-shelf that could satisfy inequality (4.14) and mount in the Microbench system. Therefore, a high-quality achromatic doublet lens was chosen for L2. The degree of collimation was simply checked by allowing the beam to fall on a wall some distance from the telescope and using the adjustment micrometer to move L2 into the position that ensures a constant beam diameter as it propagates to the wall. This method is suitably accurate as shown in the calculation in §5.10.3.

The beam exits the collimating telescope and is split into two components by a beam-splitter, BS in an adjustable mount. One component of the beam travels vertically downwards and is directed into the x axis interferometer by mirror M1. M1 is mounted on a 500 × 200 × 25 mm stainless steel plate that is in turn mounted on the anti-vibration plate of the motor drive and gearbox by two stainless steel cylindrical legs. Figure 3.8 shows the mirror mounting plate that in this photograph appears to be connected to the plate on which the x interferometer is mounted. This, however, is not the case - the only physical connection between NanoSurf IV and the motor drive plate is through the rod that couples the drive to the slideway.

The component of the beam that is transmitted through BS is reflected at mirrors M2, M3 and M4 into the anechoic chamber housing the z axis interferometer. This arrangement of

mirrors, two of which are adjustable, allows independent adjustment of the beam's propagation direction into the z axis interferometer. This extended mirror arrangement does not cause problems due to thermal drift or mechanical vibration because it is not part of common-path interferometry.

4.2.3 THE X AXIS INTERFEROMETER

The x axis interferometer is mounted on a stainless steel plate similar to the one that M1 is attached to, with some of the material being relieved to accommodate the main body of NanoSurf IV. Figure 4.9 is a schema of the x axis interferometer. The beam from M1 is incident on a beam-splitter; the reflected part of the beam is directed to a reference cube-corner retro-reflector and the transmitted part is directed to the moving cube-corner retro-reflector that is fixed to the top surface of the x slideway with epoxy resin. The beam-splitter and the small mirror combination are mounted on an aluminium plate that is in turn attached to a linear slideway that allows the plate to be translated in a direction orthogonal to the propagation direction of the incident beam and in the horizontal plane. This motion is necessary when first aligning the optics on the two stainless steel plates. The measurement and reference beams interfere at the beam-splitter. The phase of the beams that have been reflected at the beam-splitter surface is shifted by $\pi/2$ radians by means of a chromium/gold/chromium thin-film coating (Raine & Downs 1978). The beam-splitter has a small wedge angle to lower the effects of stray reflections (see chapter 5).

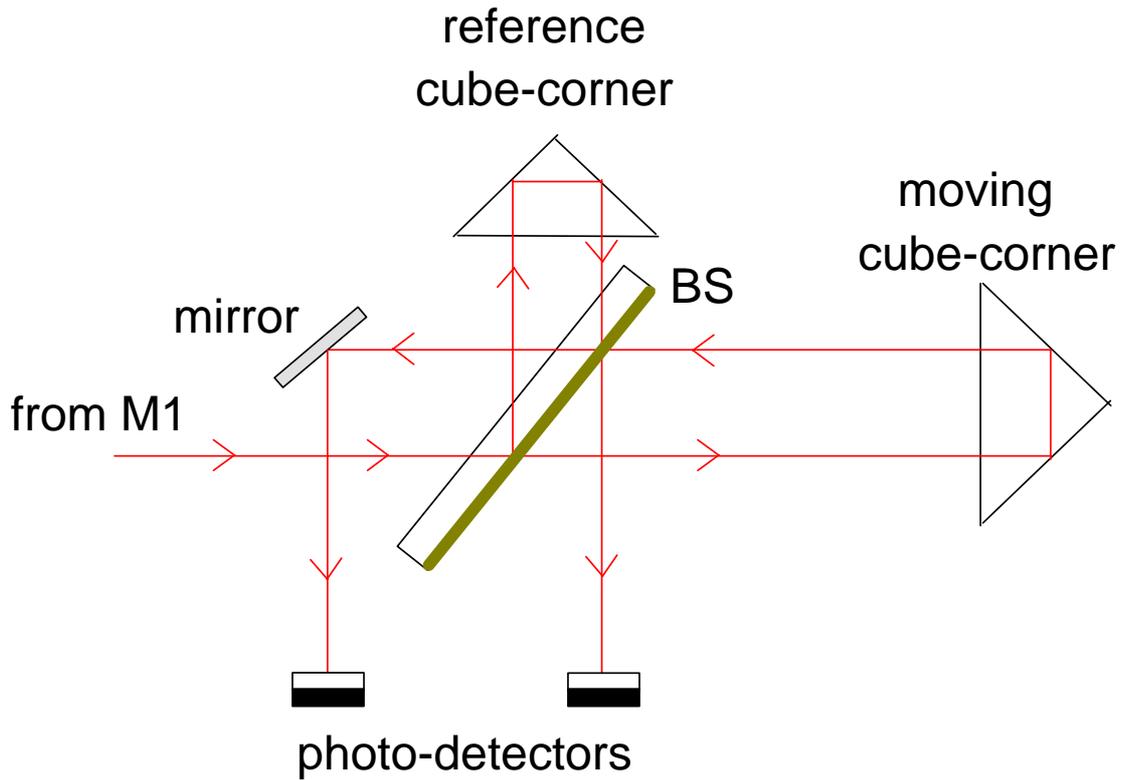


Figure 4.9 Schema of the x axis interferometer. Note that the effects of refraction within the beam-splitter have been ignored

As shown in figure 4.9, the two beams in phase quadrature are directed onto two photo-detectors that are mounted in a small Tufnol housing. The photo-detectors each have a bug-eyed lens directly in front of the 5 mm² silicon photo-diode array (L.O.T. - Oriel model OSD 5-5TB). This allows for less stringent alignment of the incident beams onto the arrays.

4.2.4 THE Z AXIS INTERFEROMETER

The configuration of the z axis interferometer is similar to that of the x axis as shown in figure 4.9. The only differences between the two interferometers are the position of the

small mirror, the position of the photo-detectors and the addition of a compensating plate, with identical wedge angle and thickness to the beam-splitter, that corrects the beam for refraction within the beam-splitter (this compensating plate was purely for convenience of alignment and was not required on the much larger x axis interferometer). The beam-splitter, compensating plate and the small mirror are housed in a 40 x 30 x 15 mm aluminium housing and the photo-detectors are housed in a small Tufnol box which has apertures to allow the incident beam in and out, and the interference signals in.

The beam enters the main chamber of NanoSurf IV via the mirror arrangement shown in figure 4.8 and a small aperture in the chamber's roof. The interferometer, photo-detector and the reference cube-corner retro-reflector are mounted on an aluminium plate in a vertical orientation shown in figures 4.10 and 3.8. The aluminium plate has a linear slideway, that moves the interferometer in a direction orthogonal to the incident beam propagation, and is sprung-loaded via a spring-steel flexure hinge onto a micrometer to facilitate angular adjustments. The entire plate is kinematically mounted on the three Invar pillars that are in turn attached to the z stage. Three springs play the part normally played by gravity in kinematically mounting the plate and three micrometers allow angular adjustment of the plate. The photo-detector housing is fixed to a carbon-fibre plate that is in turn fixed to the main aluminium plate with epoxy resin.

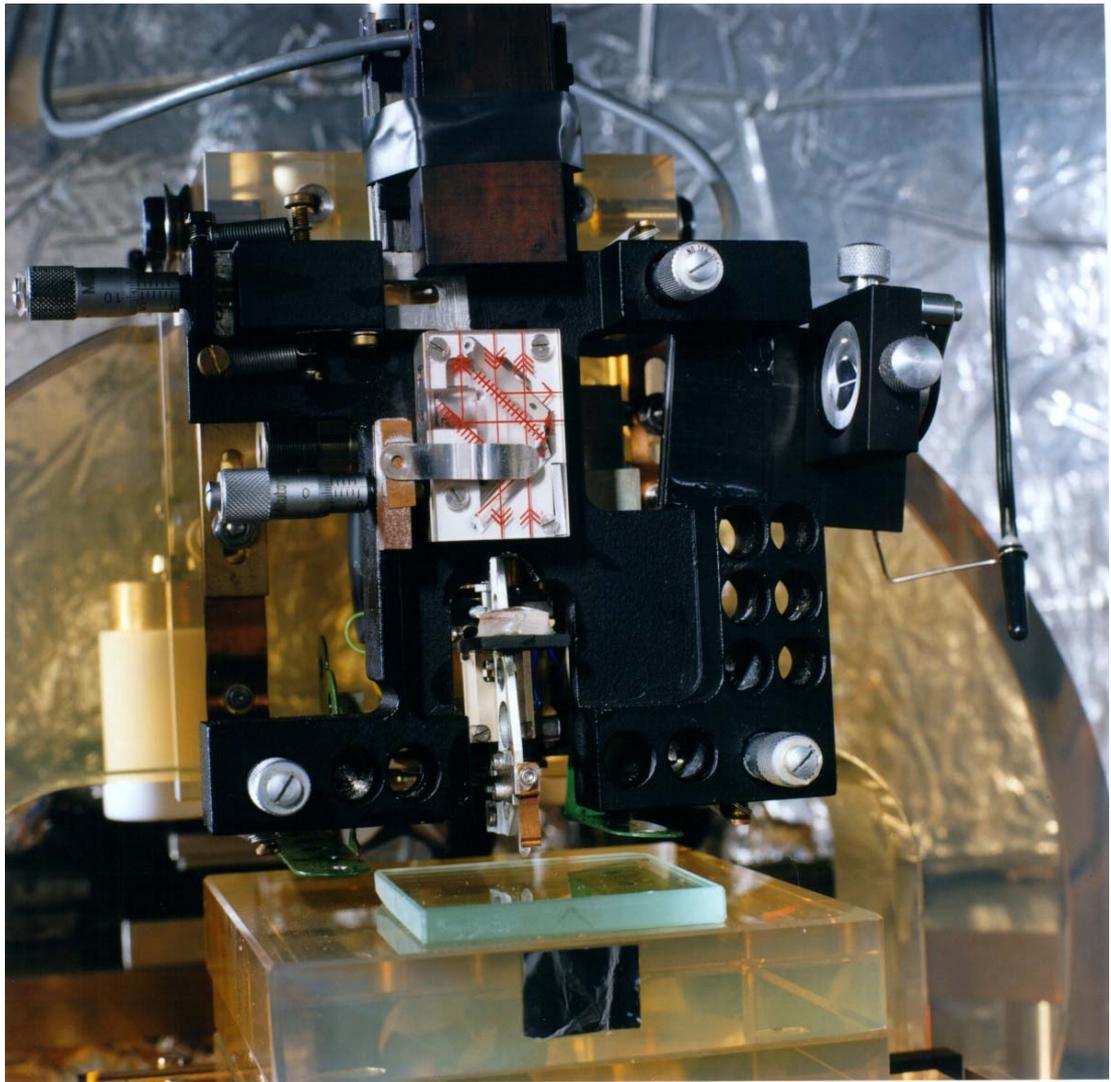


Figure 4.10 The z axis interferometer

The measurement cat's eye retro-reflector comprises the mirrored end of the Zerodur rod described in §3.3.5 and a 10 mm diameter, 25 mm focal length achromatic doublet lens that is fixed to a small carbon-fibre plate with a hole in it with epoxy resin. The plate and lens system is glued to the aluminium part of the probe shown in figure 3.15 with the probe mounted in a bench Twyman-Green interferometer that had been set up as part of another project. This assured the focal position and angular alignment of the cat's-eye.

4.2.5 THE PHOTO-DETECTION SYSTEM

The signals from the photo-detectors are processed in the same way for each interferometer so will not be considered separately. The photo-detectors operate in photovoltaic, or zero-bias, mode. Silicon photo-detectors consist of p-type and n-type semiconductor film layers one above the other. These layers cause a depletion region that is free from any mobile electric charge carriers, the width of which depends on the applied voltage (even with no applied voltage there will be a depletion region due to charge carrier diffusion). When a photon is absorbed in a semiconductor, an electron-hole pair is formed. Photo-current results when photon-generated electron-hole pairs are separated, electrons passing to the n-type region and holes to the p-type region. Alternatively, holes and electrons may recombine, thereby causing no charge displacement and no contribution to photo-current. The highest probability of the electron-hole pairs separating exists in the depletion region where the strongest electric field exists.

The photovoltaic mode of operation means that zero bias voltage is applied across the photo-diode. The external circuitry has a load resistance that causes the resultant voltage across the photo-diode to be linearly dependent on the incident radiation level. This mode has low noise due to the almost complete elimination of dark leakage current but is slow because it is the diffusion current that dominates, rather than the much faster transit of the charge carriers through the depletion region under the influence of an electric field. However, at the frequencies employed in the NanoSurf IV interferometers the response time is perfectly adequate.

In operation the two signals in phase quadrature are incident upon the photo-detectors where a proportional photo-current is generated and fed via a shielded cable to an electronic circuit known as a sine/cosine optimiser.

4.2.6 ELECTRONIC SINE AND COSINE OPTIMISATION

In a perfect interferometer the two signals incident on the two photo-detectors would be equal in amplitude, have zero dc bias and have a phase difference of exactly $\pi/2$. In practice these criteria are not realised and in order to obtain nanometric path length resolutions, corrections have to be applied to the phase quadrature signals (Birch 1990).

The NanoSurf IV system carries out two stages of optimisation: electronic signal conditioning using the sine and cosine optimisers and mathematical corrections applied after the data has been collected by the computer (see §4.2.10).

The sine and cosine optimiser electronics, originally designed by Keith Birch of NPL, are not considered in detail here. The circuitry consists of two zero-bias amplifiers that convert the current from the photo-detectors to a potential difference and allow the operator to match the amplitudes of the two signals using potentiometers. A further stage of amplifiers allows the dc offsets to be adjusted to zero. The two outputs from the circuit are then input to two channels of an oscilloscope and viewed in XY mode to produce a Lissajous figure. In the perfect case described above the oscilloscope spot would sweep out a perfect circle as the measurement retro-reflector is moved through one fringe, *i.e.* the sine signal as the abscissa and the cosine signal as the ordinate. The optimiser potentiometers are adjusted to produce a circular Lissajous figure with an origin at zero (as judged by eye). Provided the interferometer optics are not adjusted, this operation

should only be required once, however in practice a check is carried out before each measurement and any necessary adjustments are made.

Some changes had to be made to the optimiser electronics to allow low-noise operation. The electronic circuit was mounted into a thick aluminium box, designed to shield the circuit from electromagnetic interference and power was supplied via a low-noise external power supply (originally power to the circuit was supplied by an encapsulated supply mounted on the circuit board). The input stage amplifier was also modified to act in photo-voltaic as opposed to reverse bias mode. These changes significantly improved the noise output from the optimisers, especially decreasing the amount of electrical pick-up from the main supply.

4.2.7 FRINGE COUNTING ELECTRONICS

It is important that the method used to count the fringes is precisely synchronised with the samples of the signals otherwise approximately half-wavelength transient errors can appear in the estimates of position - the system described here uses digital processing to eliminate such errors (designed by Ian Robinson of NPL). The electrical signals from the photo-detectors are digitised using two 12-bit analogue to digital converters. The signals are sampled every 400 ns producing two streams of digital data. Numbers representing the average intensity of the light are subtracted from each stream to produce two new streams of data. The position counter, which is a 24-bit bi-directional counter, counts the number of half-wavelengths traversed by the interferometer. The direction of the count is determined by the direction of the zero crossing on one stream and the counter is enabled if the sign of the other stream is positive. On receipt of an external trigger pulse the

instrument samples the data stream and stores the two offset ADC readings and the corresponding values of the position and time counters. The time counter is a 24-bit counter which is incremented every 50 ns. These four values are input into the computer via an optical-fibre data bus (see §4.2.8). The relative position of the moving retro-reflector of the interferometer can be determined by combining the first three values as described above. The value from the time counter is used to calculate the time between samples to allow the computation of average velocities and accelerations from successive samples of the position of the retro-reflector.

4.2.8 INTERFACING WITH THE COMPUTER

The fringe counting electronics are interfaced to the computer via an NPL-developed optical fibre ring interface bus (Robinson 1991). This allows the two fringe counting circuits to be electrically separated from the noise-generating processing section. This eliminates currents flowing through the apparatus via the data connections and allows the reduction of magnetically induced interference.

4.2.9 THE NANOSURF IV SOFTWARE

The program structure for the NanoSurf IV is split into two sections. Firstly, a *set-up* section that allows the user to input data such as user name, scanning speed (see §3.3.4 for the range of speeds), sampling frequency, stylus type and additional information relating to the specimen being measured or the procedure used. During the *set-up* routine the output from the *z* axis interferometer is continually displayed as a graph on the screen and the user can change the scale and offset. This section of the program is

extremely useful when trying to locate a particular feature on the specimen, for example, a step height.

The measurement section of the program uses the data entered in the set-up section and starts taking data from the x and z axis interferometers when the user clicks on a start button. Similarly data collection is stopped by clicking a stop button. As with the set-up section the output from the z axis interferometer is displayed on the screen. Once the measurement is complete the user has the option of saving the data. The save routine stores the values of displacement, fringe count, fringe fraction, the x and y values from the Lissajous figure and the time (from the counter cards). Another file is also saved with the same filename, but different file extension, with the set-up data.

The routine to read the data from the counter cards is written in Borland C++ version 5.1 and has been compiled as a dynamic link library (DLL). The rest of the software is written in Microsoft Visual Basic version 5.0. The entire system runs under the Windows 95 operating system.

4.2.10 POST PROCESSING

As described in §4.2.6, the sine and cosine signals from the interferometers need to be corrected for dc offsets and differential gains. The method chosen to correct the signals has been used successfully at NPL (Birch 1990) for nearly twenty years and is a modified version of that developed by Heydemann (1981). This method only requires a single-frequency laser source (homodyne) and does not require polarisation optics. Birch (1990) used computer simulations of the correction method to predict a fringe-fractioning

accuracy of 0.1 nm. Other methods, that also claim to obtain sub-nanometre uncertainties, use heterodyne techniques (Link, von Martens 1998) and polarisation optics (Usada, Dobonsz, *et. al.* 1998), but are more complicated than the design used on NanoSurf IV.

Heydemann used two equations that describe an ellipse

$$U_{1d} = U_1 + p \quad (4.15)$$

and

$$U_{2d} = \frac{U_2 \cos a - U_1 \sin a}{G} + q \quad (4.16)$$

where U_{1d} and U_{2d} represent the noisy signals from the interferometer containing the correction terms p , q and a as defined by equations (4.21), (4.23) and (4.24) respectively, G is the ratio of the gains of the two detector systems and U_1 and U_2 are given by

$$U_1 = R_D \cos \delta \quad (4.17)$$

$$U_2 = R_D \sin \delta \quad (4.18)$$

where δ is the instantaneous phase of the interferograms. If equations (4.15) to (4.18) are combined they describe an ellipse given by

$$R_D^2 = (U_{1d} - p)^2 + \frac{[(U_{2d} - q)G + (U_{1d} - p)\sin a]^2}{\cos a}. \quad (4.19)$$

If equation (4.19) is now expanded out and the terms are collected together, an equation of the following form is obtained

$$AU_{1d}^2 + BU_{2d}^2 + CU_{1d}U_{2d} + DU_{1d} + EU_{2d} = 1 \quad (4.20)$$

with

$$\begin{aligned} A &= [R_D^2 \cos^2 a - p^2 - G^2 q^2 - 2Gpq \sin a]^{-1} \\ B &= AG^2 \\ C &= 2AG \sin a \\ D &= -2A[p + Gq \sin a] \\ E &= -2AG[Gq + p \sin a] \end{aligned}$$

Equation (4.20) is in a form suitable for using a linearised least-squares fitting routine (Forbes 1987) to derive the values of A through E , from which the correction terms can be derived from the following set of transforms

$$a = \sin^{-1} \left[\frac{C}{(4AB)^{1/2}} \right] \quad (4.21)$$

$$G = \left[\frac{B}{A} \right]^{1/2} \quad (4.22)$$

$$p = \frac{2BD - EC}{C^2 - 4AB} \quad (4.23)$$

$$q = \frac{2AE - DC}{C^2 - 4AB} \quad (4.24)$$

$$R_D = \frac{[4B(1 + Ap^2 + Bq^2 + Cpq)]^{1/2}}{5AB - C^2}. \quad (4.25)$$

Consequently, the interferometer signals are corrected by using the two inversions

$$U_1' = U_{1d} - p \quad (4.26)$$

and

$$U_2' = \frac{(U_{1d} - p)\sin a + G(U_{2d} - q)}{\cos a} \quad (4.27)$$

where U_1' and U_2' are now the corrected phase quadrature signals and therefore the phase of the interferometer signal is derived from the arctangent of (U_2'/U_1') .

The arctangent function varies from $-\pi/2$ to $+\pi/2$, whereas for ease of fringe-fractioning a phase, θ , range of 0 to 2π is preferable. This is satisfied by using the following equation

$$\theta = \tan^{-1}(U_2'/U_1') + \pi/2 + \Lambda \quad (4.28)$$

where $\Lambda = 0$ when $U_{1d} > p$ and $\Lambda = \pi$ when $U_{1d} < p$.

The Heydemann correction described above is carried out on both the x and z interferometer data after the measurements. The data processing routines are written in MATLAB version 5.0 (The Maths Works Inc.). Routines have also been written to remove a linear fit from the data, either using a least-squares method or by selecting two points on a profile by eye.

4.3 REFERENCES

- Allan D W 1966 Statistics of atomic frequency standards *Proc. IEEE* **54** 221-230
- Baer T, Kowalski F V, Hall J L 1980 Frequency stabilisation of a 0.633 μm He-Ne longitudinal Zeeman laser *Appl. Opt.* **19** 3173-3177
- Barwood G P, Gill P, Rowley W R C 1993 Laser diodes for length determination using swept-frequency interferometry *Meas. Sci. Technol.* **4** 988-994
- Barwood G P, Gill P, Rowley W R C 1998 High-accuracy length metrology using multiple-stage swept-frequency interferometry with laser diodes *Meas. Sci. Technol.* **9** 1036-1041
- Birch K P 1990 Optical fringe subdivision with nanometric accuracy *Prec. Eng.* **12** 195-198
- Brillet A, Cézé P 1981 Laser frequency stabilisation by saturated absorption *J. de Phys. (France)* **42 (C-8)** 73-82
- CCDM 1973 5th Session (*BIPM Sevres, France* 92310) M58
- Cézé P, Brillet A, Hartman F 1974 Metrological properties of the R(127) line of iodine studied laser saturated absorption *IEEE Trans. Inst. Meas.* **IM-23** 526-528
- Cézé P, Bennett S J 1979 Helium-neon laser stabilisation by saturated absorption in iodine at 612 nm *Appl. Opt.* **18** 1079-1083
- De Freitas J M, Player M A 1993 Importance of rotational beam alignment in the generation of second harmonic errors in laser heterodyne interferometry *Meas. Sci. Technol.* **4** 1173-1176
- Demarest F C 1998 High-resolution, high-speed, low data age uncertainty, heterodyne displacement measuring interferometer electronics *Meas. Sci. Technol.* **9** 1024-1030
- Downs M J, Birch K P, Cox M G, Nunn J W 1995 Verification of a polarization-insensitive optical interferometer system with subnanometric capability *Prec. Eng.* **17** 1-6

Downs M J, Nunn J W 1998 Verification of the sub-nanometric capability of an NPL differential plane mirror interferometer with a capacitance probe *Meas. Sci. Technol.* **9** 1437-1440

(Editor's note) Documents concerning the new definition of the metre 1983 *Metrologia* **19** 163-177

Fellman T, Junger P, Stahlberg B 1987 Stabilisation of a green He-Ne laser *Appl. Opt.* **26** 2705-2706

Forbes A B 1987 Fitting an ellipse to data *NPL Report DITC 95/87* 1-21

Hariharan P 1992 *Basics of interferometry* (Academic Press Inc.)

Heydemann P L M 1981 Determination and correction of quadrature fringe measurement errors in interferometers *Appl. Opt.* **20** 3382-3384

Jaseja T S, Javan A, Townes C H 1963 Frequency stability of He-Ne masers and measurements of length *Phys. Rev. Lett.* **10** 165-167

Lewis A J 1994 Measurement of length, surface form and thermal expansion coefficient of length bars up to 1.5 m using multiple-wavelength phase-stepping interferometry *Meas. Sci. Technol.* **5** 694-703

Link A, von Martens H-J 1998 Amplitude and phase measurement of the sinusoidal vibration in the nanometer range using laser interferometry *Measurement* **24** 55-67

Liu L S, Klinger J H 1979 Theory and application of laser interferometer systems *Proc. SPIE* **192** 17-26

Melles-Griot Catalogue 1997-1998

Michelson A A 1893 Comparison of the international metre with the wave-length of the light of cadmium *Astronomy and Astro-Physics* **12** 556-560

Milonni P W, Eberly J H 1988 *Lasers* (Wiley Interscience: Canada)

Morris R H, Ferguson J B, Warmiac J S 1975 Frequency stabilisation of internal mirror He-Ne lasers in a transverse magnetic field *Appl. Opt.* **14** 2808-

- Oldham N M, Kramar J A, Hetrick P S, Teague E C 1993 Electronic limitations in phase meters for heterodyne interferometry *Prec. Eng.* **15** 173-179
- Peggs G N, Lewis A J, Oldfield S 1999 Design for a compact high-accuracy CMM *Ann. CIRP* **48** 417-420
- Raine K W, Downs M J 1978 Beam-splitter coatings for producing phase quadrature interferometer outputs *Optica Acta* **25** 549-558
- Robinson I A 1991 An optical-fibre ring interface bus for precise electrical measurements *Meas. Sci. Technol.* **2** 949-956
- Rosenbluth A E, Bobroff N 1990 Optical sources of non-linearity in heterodyne interferometers *Prec. Eng.* **12** 7-11
- Rowley W R C, Wilson D C 1972 Optical coupling effects in frequency stabilisation of single mode gas lasers *Appl. Phys. Lett.* **11** 475-476
- Rowley W R C 1990 The performance of a longitudinal Zeeman-stabilised He-Ne laser (633 nm) with thermal modulation and control *Meas. Sci. Technol.* **1** 348-351
- Sasagawa G S, Zumberge M A 1989 Five year frequency stability of a Zeeman-stabilised laser *Appl. Opt.* **28** 824-825
- Stone J A, Howard L P 1998 A simple technique for observing periodic nonlinearities in Michelson interferometers *Prec. Eng.* **22** 220-232
- Tomlinson W J, Fork R L 1968 Properties of gaseous optical masers in weak axial magnetic fields *Phys. Rev.* **164** 480-483
- Umeda N, Tsujiki M, Takasaki H 1980 Stabilised ^3He - ^{20}Ne transverse Zeeman laser *Appl. Opt.* **19** 442-450
- Usada T, Dobonsz M, Kurosawa T 1998 Evaluation method for frequency characteristics of linear actuators in the sub- μm stroke range using a modified Michelson-type interferometer *Nanotechnology* **9** 77-84
- Welford W T 1986 *Aberrations of optical systems* (Adam Hilger)

- White A D, Gordon E I, Labuda E F 1964 Frequency stabilisation of single mode gas lasers
Appl. Phys. Lett. **5** 97-98
- White A D 1965 A two-channel laser frequency control system *IEEE J. Quant. Elect.* **1** 322-
- Williams D C 1993 *Optical methods in engineering metrology* (Chapman & Hall)
- Wu C-M, Su C-S 1996 Nonlinearity in measurements of length by optical interferometry
Meas. Sci. Technol. **7** 62-68
- Wu C-M, Deslattes R D 1998 Analytical modelling of the periodic nonlinearity in heterodyne interferometry *Appl. Opt.* **37** 6696-6700

CHAPTER 5

UNCERTAINTY ANALYSIS

“Measure the measurable, make the immeasurable measurable.”

Galeleo Galilei (1564 - 1642)

5.1 INTRODUCTION

The uncertainty analysis of NanoSurf IV has been split into two sections. Firstly, the uncertainty of a length measurement made using the interferometers and secondly the uncertainty in a measurement of a surface roughness parameter. At all times the uncertainty calculation follows internationally accepted guidelines for the evaluation and expression of uncertainties as laid out in the *Guide to the Expression of Uncertainty in Measurement* (1993), herein simply referred to as the *GUM*. Most of this chapter has also been published as part of an NPL Report (Leach 1999).

5.2 COMBINED STANDARD UNCERTAINTY

The combined standard uncertainty, $u_c(L)$ is an estimate of the standard deviation of the distribution of possible values (or probability distribution) of a length measurement made using an interferometer, L . The combined standard uncertainty, as its name implies, is a

quadrature sum of the uncertainties $u(x_i)$ of all the influence factors x_i , weighted by a sensitivity coefficient [§5.1.2 *GUM*]

$$u_c^2(L) = \sum_{i=1}^N c_i^2 u^2(x_i) + \sum_{i=1}^N \sum_{j=1}^N \left[\frac{1}{2} c_{ij}^2 + c_i c_{ij} \right] u^2(x_i) u^2(x_j) \quad (5.1)$$

where $u(x_i)$ are the standard uncertainties attributed to the influence quantities x_i , and where the sensitivity coefficients

$$c_i = \frac{\partial L}{\partial x_i}, \quad c_{ij} = \frac{\partial^2 L}{\partial x_i \partial x_j}, \quad c_{ijj} = \frac{\partial^3 L}{\partial x_i \partial^2 x_j} \quad (5.2)$$

are the partial derivatives of the model equation. Note that equation (5.1) assumes the input quantities are uncorrelated and is only an approximation of the uncertainty.

It is convenient to think of equation (5.1) as consisting of two parts: terms containing $u^2(x_i)$, and cross terms containing $u^2(x_i)u^2(x_j)$ that take into account the degree of non-linearity in the model. The cross terms may be referred to as second order terms by some authors in the literature, including the *GUM*.

5.3 TYPE A AND TYPE B UNCERTAINTIES COMPONENTS

In the *GUM* there is considerable concern about distinguishing between Type A and Type B uncertainty components. The distinction between Type A and Type B uncertainties relates to the manner in which they are established and not how they are subsequently

treated when they are combined. Note that there is not always a simple correspondence between the classification into categories A and B and the previously used classification into random and systematic uncertainties (Colclough 1987). Type A uncertainties are those for which repeated measurements are made and the standard deviation is evaluated from the data, and used as the standard uncertainty. Type B uncertainties are those for which repeated measurements cannot simply isolate the influence, and the uncertainty must be referred to by some other means [§4.2, §4.3 *GUM*].

The rectangular distribution occurs in Type B uncertainty evaluation and is used frequently in this chapter. As explained in §4.4.5 of the *GUM*, if data to estimate the uncertainty distribution of an influence parameter are limited, often an adequate and useful approximation is to assume an upper $+a$ and a lower $-a$ bound for a range of equally probable values. The standard uncertainty (for example, §4.3.7 *GUM*) is then found by integrating over the rectangular distribution to give $a/\sqrt{3}$. Similarly, a reading from a meter read-out of resolution b has a standard uncertainty of $b/\sqrt{12}$ [§F.2.2.1 *GUM*]; the reading is considered to be a rectangular distribution bounded by $\pm a = +b/2$ and $-b/2$.

5.4 DEGREES OF FREEDOM

The degrees of freedom, ν_{x_i} are estimated for each of the $u(x_i)$. For Type A uncertainty evaluation, the degrees of freedom $\nu = n - 1$, where n is the number of observations taken to determine the arithmetic mean [§G.3.3 *GUM*]. In many cases, the uncertainty evaluation is Type B, and the degrees of freedom are estimated according to the relative

uncertainty in the uncertainty $\Delta u/u$, or to put it another way - the judgement of reliability or confidence in the estimation of the uncertainty is given by [§G.4.2 GUM]

$$v_i \approx \frac{1}{2} \left[\frac{\Delta u(x_i)}{u(x_i)} \right]^{-2}. \quad (5.3)$$

Degrees of freedom for influence parameter uncertainties are combined in one step for the overall uncertainty budget effective degrees of freedom, by applying the Welch-Satterthwaite formula [§G.4.1 GUM]

$$v_{eff} = \frac{u_c^4(x)}{\sum_{i=1}^N \frac{u_i^4(x)}{v_i}}. \quad (5.4)$$

Degrees of freedom are always rounded-up to the next integer [§G.4.1 GUM].

5.5 EXPANDED UNCERTAINTY

It is desirable to express uncertainties so that for most of the measurements the measured value is within the uncertainty range of the 'true value'. The expanded uncertainty [§6.2 GUM]

$$U = k u_c(L) \quad (5.5)$$

is defined as the combined uncertainty multiplied by a coverage factor k . The value of the coverage factor is chosen depending on the level of confidence that would facilitate the interpretation of the uncertainty and the number of degrees of freedom. Most measurements are expressed with a value of k between two and three. $k = 2$ corresponds to a confidence level of approximately 95% assuming a high number of degrees of freedom [§6.2.2 GUM].

The steps involved in this uncertainty analysis are based upon those presented by Decker, Ulrich, *et. al.* (1998):

- Step 1:** Analyse the measurement process and identify the influence quantities.
- Step 2:** List any simplifying assumptions and their impact on the measurement.
- Step 3:** Form a mathematical model of the measurement in terms of the influence quantities (expressed in an optimal form).
- Step 4:** Evaluate the sensitivity coefficients of the influence quantities.
- Step 5:** Evaluate the standard uncertainties and degrees of freedom of the influence quantities.
- Step 6:** Determine correlated components.
- Step 7:** Calculate the combined and expanded uncertainties and degrees of freedom for the overall process.

5.6 MEASUREMENT PROCESS

At this stage only the displacement of either the cat's-eye in the z axis or the corner-cube in the x axis is analysed. The following influence quantities are considered:

Laser source:

- short and long-term calibration;
- frequency stability;
- intensity stability;
- intensity distribution;
- wavefront characteristics;
- polarisation.

Interferometer characteristics:

- collimation/obliquity effects;
- quality of the optical components;
- air refractive index ;
- stray beams;
- diffraction effects;
- dead path error;
- air turbulence effects.

Metrology frame:

- thermal expansion;
- mechanical expansion and rigidity;
- Abbe offset;
- cosine error;
- co-ordinate orthogonality.

Detection system:

- linearity of photo-detectors;
- detector geometry;
- electronic noise;
- frequency response;
- resolution.

Computing/software:

- quality of mathematical fits and models.

By identifying the specific characteristics of the measurement procedure and considering the conditions of the laboratory, some of these influence quantities can be combined or be deemed negligible. If an influence quantity were found to have an effect that is less than 1 pm, then it is considered negligible. When an influence quantity is deemed negligible it is because sound physical laws have been applied to model its magnitude and effect, or because it has been measured to be negligible. It is important to appreciate that the final uncertainty in length measurement is for a strict range of conditions and techniques.

5.7 ASSUMPTIONS

Evaluation of measurement uncertainty must be for a specific measurement scenario. The specifics of the measurement and the influence factors should be well defined before trying to consider their uncertainties. At this stage there are no simplifying assumptions

apparent to the NanoSurf IV measurement technique. Of course, the assumption is made that operators of NanoSurf IV are applying best laboratory practice that is free of blunders. The ISO 9001 Quality Management System in operation at NPL should ensure this.

5.8 MATHEMATICAL MODEL OF THE MEASUREMENT

The primary influence factors recognised above can be expressed algebraically and combined to yield a mathematical model representing the measurement. Starting from first principles the measured displacement of one of the retro-reflectors in an ideal Michelson interferometer is given by

$$L_{\phi} = \frac{\lambda}{2\pi} \Delta\phi \quad (5.6)$$

where λ is the wavelength of the laser source and $\Delta\phi$ is the difference in the phase from the reference and measurement arms of the interferometer. A non-ideal interferometer will measure a length given by

$$L = L_{\phi} + L_{\Omega} + L_n + L_t + L_m + L_A + L_d + L_T + L_j \quad (5.7)$$

where L_{Ω} is a correction to the measured length for the effects of diffraction, L_n is a correction due to the change in the refractive index of the air in which the laser operates, L_t is that for thermal effects on the metrology frame, L_m is that for mechanical effects on the metrology frame, L_A is that for the imperfect optics and stray beams, L_d is that for the

dead path length, L_T is that for the effects of air turbulence and L_j is that for the measurement set-up, *i.e.* Abbe and cosine errors. Once again, it assumed that the input quantities are uncorrelated. Note that some of the terms listed above are gathered together in later sections (see §5.10.4)

5.9 EVALUATION OF THE SENSITIVITY COEFFICIENTS

The equation for the combined standard uncertainty is applied to the mathematical model describing the measurement. Simply making the substitution of the influence variables in place of the x_i in equation (5.1), the combined standard uncertainty can be written as

$$u_c^2(L) = c_{L_\phi}^2 u^2(L_\phi) + c_{L_\Omega}^2 u^2(L_\Omega) + c_{L_n}^2 u^2(L_n) + c_{L_t}^2 u^2(L_t) + c_{L_m}^2 u^2(L_m) + c_{L_A}^2 u^2(L_A) + c_{L_d}^2 u^2(L_d) + c_{L_T}^2 u^2(L_T) + c_{L_j}^2 u^2(L_j) + \text{higher order terms} \quad (5.8)$$

where the sensitivity coefficients, c_i , for the first order terms are

$$c_{L_\phi} = \frac{\partial L}{\partial L_\phi}, c_{L_\Omega} = \frac{\partial L}{\partial L_\Omega}, c_{L_n} = \frac{\partial L}{\partial L_n}, c_{L_t} = \frac{\partial L}{\partial L_t}, c_{L_m} = \frac{\partial L}{\partial L_m}, c_{L_A} = \frac{\partial L}{\partial L_A}, c_{L_d} = \frac{\partial L}{\partial L_d}, c_{L_T} = \frac{\partial L}{\partial L_T}, c_{L_j} = \frac{\partial L}{\partial L_j}. \quad (5.9)$$

Calculating the partial derivatives c_i , c_{ij} and c_{ijj} in equation (5.8) determines the sensitivity coefficients for the uncertainty in displacement measurement. In the case of equation (5.8)

all of the first order coefficients are equal to unity and the higher order terms are equal to zero. Each term in equation (5.8) is now examined in detail.

5.10 STANDARD UNCERTAINTIES OF INFLUENCE QUANTITIES

5.10.1 OPTICAL PATH DIFFERENCE

The combined uncertainty in the measurement of the optical path difference is given by

$$u_c^2(L_\phi) = c_{\Delta\phi}^2 u^2(\Delta\phi) + c_\lambda^2 u^2(\lambda) + \text{higher order terms.} \quad (5.10)$$

The first order sensitivity coefficients are given by $(\lambda/2\pi)$ and $(\Delta\phi/2\pi)$ respectively. For convenience, c_λ is expressed as L/λ . The second-order nature of the all higher order coefficients means they are negligible.

5.10.1.1 Measured parameters influencing the measurement of the optical path difference

Vacuum wavelength

The laser wavelength is calibrated by beat frequency measurement against one of the standard NPL iodine-stabilised lasers. The wavelength stability is quoted as $\pm 1 \times 10^{-9}$

over 24 hours with a drift of $\pm 1 \times 10^{-8}$ between calibrations (every 2500 hours). Table 5.1 shows the uncertainty contributions.

Source	Size	k	Standard uncertainty	Sensitivity coefficient	Contribution
Primary laser accuracy	$2.5 \times 10^{-11} \lambda$	1	$2.5 \times 10^{-11} \lambda$	L/λ	$2.5 \times 10^{-11} L$
Stability of laser	$1 \times 10^{-9} \lambda$	1	$1 \times 10^{-9} \lambda$	L/λ	$1 \times 10^{-9} L$
Yearly drift range	$1 \times 10^{-8} \lambda$	1	$1 \times 10^{-8} \lambda$	L/λ	$1 \times 10^{-8} L$
					$1 \times 10^{-8} L$

Table 5.1 Summation of uncertainties due to the laser source

Phase difference

The uncertainty in the phase difference measurement cannot easily be differentiated from other sources of uncertainty. Its effect is measured in §5.10.4 and summarised in table 5.2. Notice that the effect of this uncertainty is considered twice in the uncertainty analysis -

once in combination with the calibration parameters of the laser wavelength and again in §5.12 as a purely random and experimentally determined parameter.

5.10.1.2 Overall uncertainty due to the measured optical path difference

Combining the values for the sensitivity coefficient and the measured parameters gives a total uncertainty in the measurement of the optical path difference of

$$u_c(L_\phi) = \sqrt{(1 \times 10^{-2} L)^2 + 0.37^2} \text{ nm} \quad (5.11)$$

where L is in millimetres.

Source	Size	k	Standard uncertainty	Sensitivity coefficient	Contribution
Random phase fluctuation	0.37 nm	1	0.37 nm	1	0.37 nm
					0.37 nm

Table 5.2 Summation of uncertainties due to fluctuations in the phase measurement

5.10.2 DIFFRACTION EFFECTS AND MISALIGNMENT

As the interferometers are illuminated by a laser source, the shift of the phase and changes in the curvature of the wavefronts lead to systematic errors, to which there must

be added the error caused by misalignments. Consider an ideal interferometer illuminated by a monochromatic Gaussian beam that produces two interfering beams. No diffraction is assumed to occur in the optical system. The error in the fringe spacing when the interference pattern is focussed onto the detector is given by (Mana 1989)

$$\frac{2\Delta\lambda_0}{\lambda} = \left(\frac{1}{k^2 w_0^2} \right) \left(\frac{[1 + 2u^2 / w_f^2] e^{-2u^2/w_f^2} - 1}{e^{-2u^2/w_f^2} - 1} \right) + \left(\frac{2}{k^2 w_0^2} \right) \left(\frac{b}{2D} \right)^2 \quad (5.12)$$

where λ is the wavelength of the laser source, λ_0 is the fringe displacement, k is the wavenumber, w_0 is the waist of the laser, u is the aperture radius, D is the absolute distance of the lens from the waist of the beam, f is the focal length of the lens and w_f and b are given by

$$w_f = 2 \left(\frac{f}{b} \right) w_0 \quad (5.13)$$

and

$$b = k w_0^2. \quad (5.14)$$

Before calculating the error term given by equation (5.12) it is necessary to calculate the effective beam waist due to the collimating telescope. To do this a new waist is found for each of the optical elements by using the following (Self 1983)

$$\frac{1}{f} = \frac{1}{s''} + \frac{1}{s + z_R^2 / (s - f)} \quad (5.15)$$

and

$$z_R'' = m^2 z_R \quad (5.16)$$

where s is the object to lens distance, s'' is the image to lens distance, m is the magnification and z_R is the Rayleigh length or the distance over which the beam radius spreads by a factor of $\sqrt{2}$. The Rayleigh length is related to the beam waist by

$$z_R = \frac{\pi w_0^2}{\lambda}. \quad (5.17)$$

From the numerical values of the interferometer systems, $\Delta\lambda_0$ turns out to be less than 30 fm and is considered negligible.

Assuming the size of the detector is a lot larger than the spot size and that

$$\frac{\alpha}{\theta} \leq \frac{z}{|z - l|} \quad (5.18)$$

where α is the misalignment angle, l is the distance between the waist and the retro-reflector, z is the distance between the waist and the detector, $\theta = \lambda / (\pi w_0)$, the normalised error can be reduced to

$$\frac{4}{\theta^2} \frac{2\Delta\lambda_0}{\lambda} = 1 - \left(\frac{v^2 - 2}{8v^2} \right) \sigma^2. \quad (5.19)$$

In equation (5.19) $v = w/w_0$, $\sigma = 2\alpha/(v\theta)$ and

$$w^2 = w_0^2 \left[1 + 4 \left(\frac{z}{b} \right)^2 \right]. \quad (5.20)$$

Inserting values for the parameters in equation (5.19), even if the interferometer were misaligned by as much as $\alpha = 5^\circ$, $\Delta\lambda_0$ would only be equal to 0.1 pm. This source of uncertainty is, therefore, taken to be negligible.

According to the above the correction for the effects of diffraction and misalignment of the interferometer, $L_\Omega = 0$, as are its associated uncertainties.

5.10.3 AIR REFRACTIVE INDEX

When performing optical interferometry in air, it is important to correct the laser wavelength for the refractivity of the air through which it passes. The correction factor, the refractive index, is applied to the vacuum wavelength of the light emitted by the laser

$$\lambda = \frac{\lambda_{vac}}{n} \quad (5.21)$$

where λ_{vac} is the wavelength in vacuum and n is the refractive index of air, for ambient conditions.

5.10.3.1 Equations for the refractivity of air

In 1965 Edlén reviewed the most recent work, collated findings and issued new formulae for the dispersion of air. The formulae derived in that paper have since been widely used to correct for the refractivity of air, with a minor correction to the humidity term suggested by Birch & Downs (1988) and a further correction suggested by Bönsch and Potulski (1998) for the latest standard conditions. The calculation starts with the dispersion of dry air for the new standard conditions, temperature $t = 20$ °C (ITS-90), pressure $p = 100\,000$ Pa and 0.04% carbon dioxide content, describing the refractivity of standard air dependent on the wavenumber $\sigma = 1/\lambda$.

$$(n-1)_N \cdot 10^8 = 8091.37 + \frac{233\,3983}{130 - \sigma^2} + \frac{15\,518}{38.9 - \sigma^2}. \quad (5.22)$$

A CO₂ content x , differing from 0.0004, changes the refractivity to

$$(n-1)_x = (n-1)_N [1 + 0.5327(x - 0.0004)]. \quad (5.23)$$

The deviation of temperature and pressure from the reference conditions is taken into account by

$$(n-1)_{tp} = \frac{(n-1)_x p}{93214.6} \left[\frac{(1+10^{-8})(0.5953-0.009876t)p}{1+(0.003661t)} \right]. \quad (5.24)$$

The influence of water vapour with partial pressure f is calculated, which results in the refractive index for moist air

$$n_{tpf} - n_{tp} = -f(3.802 - 0.0384\sigma^2) \cdot 10^{-10}. \quad (5.25)$$

The uncertainty attributed to the empirical determination of the numerical coefficients in this equation is $\pm 1 \times 10^{-8}$ at the one standard deviation level of confidence (Birch & Downs 1994).

The length correction that accounts for the refractive index of air is given by

$$L_n = -nL \quad (5.26)$$

and its combined standard uncertainty

$$u_c^2(L_n) = \left(\frac{\partial L_n}{\partial p}\right)^2 u_c^2(p) + \left(\frac{\partial L_n}{\partial x}\right)^2 u_c^2(x) + \left(\frac{\partial L_n}{\partial t}\right)^2 u_c^2(t) + \left(\frac{\partial L_n}{\partial f}\right)^2 u_c^2(f) + \left(\frac{\partial L_n}{\partial \lambda}\right)^2 u_c^2(\lambda) + \text{higher order terms} \quad (5.27)$$

Calculating the partial derivatives and putting in the values of the reference conditions the sensitivity coefficients are given by (see §5.11.2 for the reference values of humidity)

$$\begin{aligned}
 \left(\frac{\partial L_n}{\partial p} \right) &= 2.68 \times 10^{-9} \text{ L / Pa} \\
 \left(\frac{\partial L_n}{\partial x} \right) &= 1.4 \times 10^{-10} \text{ L / ppm} \\
 \left(\frac{\partial L_n}{\partial t} \right) &= -9.30 \times 10^{-7} \text{ L/}^\circ\text{C} . \\
 \left(\frac{\partial L_n}{\partial f} \right) &= -3.8 \times 10^{-10} \text{ L / Pa} \\
 \left(\frac{\partial L_n}{\partial \lambda} \right) &= 1.22 \times 10^{-5} \text{ L / } \mu\text{m}
 \end{aligned}
 \tag{5.28}$$

The largest cross term is given by

$$\frac{1}{2} c_u^2 + c_t c_m = 8 \times 10^{-9} \text{ L}^2 / ^\circ\text{C}^2,
 \tag{5.29}$$

but from equation (5.1) this would multiply $u^2(x_i)$ and it is therefore considered that all cross terms are negligible.

5.10.3.2 Conversion of humidity units

The humidity is measured using a dewpoint hygrometer, which displays results in the form of $^\circ\text{C}$ dewpoint temperature. Magnus' relation (BS 1339: 1965) is used to convert $^\circ\text{C}$

dewpoint into partial pressure of water vapour for use by the Edlén equations and is given by

$$f = 10^{\left(0.7857 + \frac{7.5t_{dp}}{237.3+t_{dp}}\right)} \quad (5.30)$$

where t_{dp} is the dewpoint temperature in °C. Partially differentiating gives

$$\left(\frac{\partial f}{\partial t_{dp}}\right) = \ln(10) \times \left[\frac{-7.5t_{dp}}{(237.3+t_{dp})^2} + \frac{7.5}{237.3+t_{dp}} \right] \times 10^{\left(0.7857 + \frac{7.5t_{dp}}{237.3+t_{dp}}\right)}. \quad (5.31)$$

In most dimensional metrology laboratories the humidity is controlled at around 10 °C dewpoint. Substituting $t_{dp} = 10$ °C gives

$$\left(\frac{\partial f}{\partial t_{dp}}\right) = 82.2 \text{ Pa } ^\circ\text{C}^{-1}. \quad (5.32)$$

Hence a variation in humidity of 1 °C dewpoint alters the partial pressure by 82.2 Pa and hence the refractive index by 3.12×10^{-8} .

5.10.3.3 Measured parameters influencing the refractive index of air

Air temperature

The air temperature is measured by platinum resistance thermometers (PRTs) inside the anechoic chamber and close to the object being measured. The calibration certificate of the

PRTs states a calibration uncertainty of ± 0.005 °C at a 95% confidence level. The self-heating of the PRT (Downs, Ferris, *et. al.* 1990) was assessed by observing a typical PRT resistance reading as the current was increased by a factor of $\sqrt{2}$ from its initial value of 1 mA. Table 5.3 summarises the contributions due to the air temperature measurement.

Air pressure

The pressure transducer is located inside the equipment rack and is connected to the anechoic chamber via PVC tubing. The calibration certificate quotes an uncertainty of ± 5 Pa at a confidence level of 95% for the primary standard, and shows a variation of ± 20 Pa in calibrated results. Table 5.4 summarises the contributions due to the measurement of air pressure.

Source	Size	k	Standard uncertainty	Sensitivity coefficient	Contribution
Resistance bridge accuracy	3 mK	$\sqrt{3}$	1.7 mK	$9.30 \times 10^{-7} L/K$	$1.6 \times 10^{-9} L$
PRT calibration	5 mK	1.96	2.6 mK	$9.30 \times 10^{-7} L/K$	$2.4 \times 10^{-9} L$
ITS90 equations	0.13 mK	1	0.13 mK	$9.30 \times 10^{-7} L/K$	$1.2 \times 10^{-10} L$
PRT inter-calibration drift	2 mK	1	2 mK	$9.30 \times 10^{-7} L/K$	$1.9 \times 10^{-9} L$

Uncertainty Analysis

PRT - air lag	10 mK	$\sqrt{3}$	5.8 mK	$9.30 \times 10^{-7} L/K$	$5.4 \times 10^{-9} L$
Self heating of PRT	12 mK	$\sqrt{3}$	6.9 mK	$9.30 \times 10^{-7} L/K$	$6.4 \times 10^{-9} L$
					$9.1 \times 10^{-9} L$

Table 5.3 Summation of uncertainties due to air temperature measurement

Humidity

The humidity transducer is located inside the equipment rack and is connected to the anechoic chamber via PVC tubing. The calibration certificate quotes an uncertainty of ± 0.2 °C at a confidence level of 95%. Table 5.5 summarises the contributions due to the measurement of humidity.

Source	Size	k	Standard uncertainty	Sensitivity coefficient	Contribution
Primary standard uncertainty	5 Pa	1.96	2.6 Pa	$2.68 \times 10^{-9} L/Pa$	$7.0 \times 10^{-9} L$
Sensor variability	20 Pa	$\sqrt{3}$	11.6 Pa	$2.68 \times 10^{-9} L/Pa$	$3.1 \times 10^{-8} L$
Sensor resolution	1 Pa	$\sqrt{12}$	0.3 Pa	$2.68 \times 10^{-9} L/Pa$	$8.0 \times 10^{-10} L$

$3.2 \times 10^{-8} L$

Table 5.4 Summation of uncertainties due to air pressure measurement

Source	Size	<i>k</i>	Standard uncertainty	Sensitivity coefficient	Contribution
Dewpoint meter calibration	0.2 °C DP	1.96	0.10 °C DP	$3.0 \times 10^{-8} L/^\circ\text{C DP}$	$3.0 \times 10^{-9} L$
Magnus' equation	0.2 °C DP	$\sqrt{3}$	0.12 °C DP	$3.0 \times 10^{-8} L/^\circ\text{C DP}$	$3.6 \times 10^{-9} L$
Interface resolution	0.5 °C DP	$\sqrt{12}$	0.14 °C DP	$3.0 \times 10^{-8} L/^\circ\text{C DP}$	$4.2 \times 10^{-9} L$
					$6.3 \times 10^{-9} L$

Table 5.5 Summation of uncertainties due to air humidity measurement where DP = Dewpoint temperature

Carbon dioxide content

Typical changes in the carbon dioxide content in a laboratory, due to such things as human respiration, can be up to 100 ppm (Downs 1998). Table 5.6 shows the effect of this departure on the uncertainty due to the refractive index.

Source	Size	k	Standard uncertainty	Sensitivity coefficient	Contribution

Departure from standard conditions (<i>i.e.</i> 400 ppm)	100 ppm	1	100 ppm	1.47 × 10 ⁻¹⁰ L/ppm	1.47 × 10 ⁻⁸ L
					1.47 × 10 ⁻⁸ L

Table 5.6 Summation of uncertainties due to air carbon dioxide content

Vacuum wavelength

The combined uncertainty attributed to the vacuum wavelength through its contribution to the refractive index of air is very small. Using values from table 5.1 and equation (5.28) gives

$$u_c(\lambda) = 8 \times 10^{-14} L \tag{5.33}$$

which is considered negligible.

Uncertainty of the Edlén equations

Table 5.7 presents the contribution due to the uncertainty in the Edlén equations.

Source	Size	k	Standard uncertainty	Sensitivity coefficient	Contribution

Accuracy of Edlén equations	3×10^{-8}	3	1×10^{-8}	L	$1.0 \times 10^{-8} L$
					$1.0 \times 10^{-8} L$

Table 5.7 Summation of uncertainties due to the Edlén equations

5.10.3.4 Overall uncertainty due to the refractive index of air measurement

Combining the values for the sensitivity coefficients and the measured parameters taken from tables 5.3 to 5.7 gives a total uncertainty in the refractive index measurement of

$$u_c(L_n) = 3.8 \times 10^{-8} L \tag{5.34}$$

where L is the measured length.

5.10.4 EFFECTS ON THE METROLOGY FRAME AND DETECTOR ELECTRONICS

It is very difficult to separate the effects of thermal and mechanical changes on the metrology frame plus the effects of the detector electronics and data processing. It is also very difficult to rigorously model these effects. For these reasons, the stability of the metrology frame, detection system and mathematical algorithms has been experimentally measured with no regard as to the source of any fluctuations, *i.e.* a correction for the stability of the metrology frame and detection system is equal to $(L_t + L_m + L_T + L_\phi)$. It

must also be remembered that other random uncertainty contributions, such as refractive index fluctuations, will also be present in these measurements. The measured uncertainty term is, therefore, expected to be a pessimistic value.

To measure the stability of the metrology frame and detection system, the output of the interferometers was monitored as described in §6.2 with the retro-reflectors stationary. The results are given in table 5.8. Only the highest (z axis) value of the standard deviations for the two axes has been used here. During the experiments the humidity and barometric pressure were monitored for stability - any noticeable drift or large fluctuations voided the experiment.

Source	Size	k	Standard uncertainty	Sensitivity coefficient	Contribution
x axis dimensional change	0.37 nm	1	0.37 nm	1	0.37 nm
					0.37 nm

Table 5.8 Summation of uncertainties due to dimensional fluctuations of the metrology frame

5.10.5 EFFECTS OF IMPERFECT OPTICS AND STRAY BEAMS

Due to the very high spatial and temporal coherence of the laser source, stray light can interfere with beams reflected from the surfaces present in the reference and measurement arms of the interferometers. The dominant effects are usually due to unwanted reflections and isolated strong point scatterers, both leading to random and non-random spatial variations in the scattered phase and amplitude (Hariharan 1997). This analysis does not attempt to isolate sources of stray reflection.

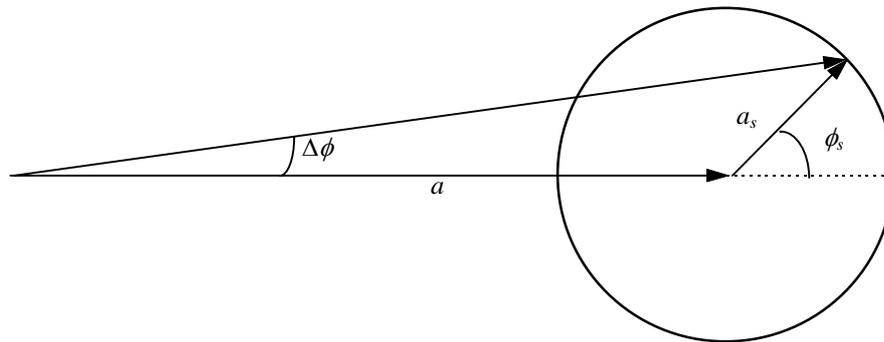


Figure 5.1 Phase shift of a beam produced by coherent stray light that has travelled an additional path

Assuming the stray light affects one beam only, its amplitude, a_s , adds vectorily to the amplitude of the main beam, a as in figure 5.1. In the case of a Michelson interferometer, the complex amplitude at any point in the interference pattern is obtained by summing the complex amplitude of the stray light and the complex amplitudes of the beams reflected from the reference and test surfaces. If we assume that the beams reflected from the reference and test surfaces have unity intensity, the resultant complex amplitude is given by the relation

$$a(x, y) = \left| 1 + e^{-i\phi(x,y)} + a_s(x, y)e^{-i\phi_s(x,y)} \right| \quad (5.35)$$

where $\phi_s = (2\pi/\lambda)p$, and p is the difference in the lengths of the optical paths traversed by the stray light and by the test beams relative to the beam reflected from the reference surface. The intensity in the interference pattern is then

$$I(x, y) = a^2(x, y) = 2 + |a_s|^2 + 2a_s \cos \phi_s + 2[1 + a_s \cos \phi_s] \cos \phi + 2a_s \sin \phi_s \sin \phi. \quad (5.36)$$

Assuming $a_s \sim 1$, $\Delta\phi$ can be found by equating the following

$$[1 + a_s \cos \phi_s] \cos \phi + a_s \sin \phi_s \sin \phi \equiv A \cos(\phi - \Delta\phi). \quad (5.37)$$

After some simple trigonometry and assuming $\Delta\phi$ is a small angle the magnitude of the phase error is given by

$$\Delta\phi \approx a_s \sin \phi_s. \quad (5.38)$$

To minimise the effects of stray reflections all the optical components have been thoroughly cleaned, the retro-reflectors are mounted at a non-orthogonal angle to the beam propagation direction (to avoid reflections off the front surfaces) and all the non-critical optical surfaces are anti-reflection (AR) coated. It is extremely difficult, if not impossible, to measure the amplitude of the stray light, simply because it propagates in the same direction as the main beams. The value of a_s is taken to be equal to the value of

the reflection coefficient of the AR coat, $a_s = 0.004$. This gives a value for $\Delta\phi$ of ± 0.4 nm.

Table 5.9 summarises this uncertainty contribution.

Source	Size	k	Standard uncertainty	Sensitivity coefficient	Contribution
Uncertainty due to stray light	0.4 nm	$\sqrt{3}$	0.23 nm	1	0.23 nm
					0.23 nm

Table 5.9 Summation of uncertainties due to stray light

5.10.6 DEAD PATH LENGTH UNCERTAINTY

Dead path length, d , is defined as the difference in distance in air between the reference and measurement retroreflectors and the beamsplitter when the interferometer measurement is initiated (Zanoni 1988). Dead path error occurs when there is a non-zero dead path and environmental conditions change during a measurement. The equation below (Zanoni 1988) yields the displacement, D , for a single pass interferometer such as those used on NanoSurf IV

$$D = \frac{N\lambda_{vac}}{n_2} - \frac{\Delta nd}{n_2} \quad (5.39)$$

where N is half the number of fringes counted during the displacement, n_2 is the refractive index at the end of the measurement, Δn is the change in refractive index over the measurement time: that is $n_2 = n_1 + \Delta n$, and n_1 is the refractive index at the start of the

measurement. The second term on the right hand side of equation (5.39) is the dead path error, L_{dp} .

Dead path error is corrected for in the software but there is still an uncertainty in the correction given by

$$u_c^2(L_{dp}) = \left(\frac{\partial L_{dp}}{\partial n_1}\right)^2 u_c^2(n_1) + \left(\frac{\partial L_{dp}}{\partial n_2}\right)^2 u_c^2(n_2) + \left(\frac{\partial L_{dp}}{\partial d}\right)^2 u_c^2(d) + \text{higher order terms} \quad (5.40)$$

The dead path imposes several critical measurement conditions on NanoSurf IV. Firstly, a sample must always be mounted such that the initial position is at the position of zero dead path. This can only be achieved to within 1 mm and $u_c^2(d)$ is ± 1 mm using a steel rule (plus the optical elements will be affected by thermal expansion and refractive index variations). Also, the environmental conditions have to be monitored. Any drifts in temperature of greater than ± 0.1 °C and pressure greater than ± 10 Pa void a measurement. Note that higher order terms in equation (5.40) are all negligible. Table 5.10 summarises the uncertainty in the measurement and correction for the dead path length.

Source	Size	k	Standard uncertainty	Sensitivity coefficient	Contribution
Uncertainty due dead path	0.20 nm	$\sqrt{3}$	0.12 nm	1	0.12 nm
					0.12 nm

Table 5.10 Summation of uncertainties due to dead path

5.10.7 UNCERTAINTIES DUE THE MEASUREMENT SET-UP

5.10.7.1 Cosine error

There will always be some misalignment of the measurement axis to the axis of motion of the stage. The measurement axis is the central line parallel to the in and out beams to and from the measurement retroreflector. The misalignment manifests itself as an error in the measured length that is directly proportional to the cosine of the angle between the measurement axis and the axis of motion of the stage. The cosine error always causes the interferometer to measure shorter than the actual distance travelled by the stage.

The cosine error in the x axis can be measured (and minimised) by removing the beamsplitter, attaching a mirror to the front of the measurement retroreflector and auto-reflecting the laser beam onto the output aperture of the laser. The alignment can then be checked as the stage is moved through its full range. The worst case measured was a displacement from the aperture of less than 1 mm for a path length of around 1 m. This corresponds to a cosine error of $1.5 \times 10^{-10} L$ and is negligible.

Provided the z interferometer block can be set up so as to get interference, the small length over which it operates ensures a negligible cosine error in this axis.

5.10.7.2 Abbe error

The Abbe offset has been made negligible in the z axis by virtue of the interferometer and probe design, *i.e.* the displacement measuring axis is co-axial with the displacement to be measured (this was checked by tracing a $\lambda/40$ flat at different areas of the probe's range - no significant differences above the instrumental noise were measured). In the x axis there can be an Abbe offset of up to 5 mm.

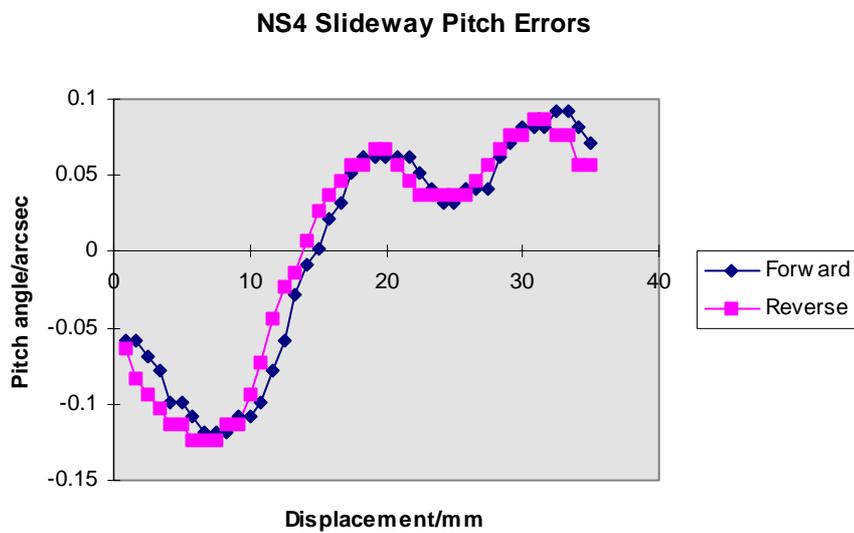


Figure 5.2 x axis slideway pitch errors

To determine the Abbe error in the x axis, the straightness of the slideway must be determined. This was carried out by setting up an autocollimator with a small aluminised microscope cover-slip glued to the front of the specimen mounting plate with epoxy resin. To ensure that the slideway and autocollimator had the same mechanical earth (the stainless steel vibration isolation plate) a small, lightweight autocollimator was constructed from a compact disc reading head (Sony Model D50 Mk II, Armstrong &

Fitzgerald 1992). The autocollimator has a resolution of 0.01 second of arc and was calibrated using an NPL indexing table. Figures 5.2 and 5.3 show the straightness results for the x axis slideway. Note that there is no Abbe sensitivity to roll. From both the graphs it is apparent that the repeatability of these measurements is excellent when the direction of travel is reversed.

Over the full 35 mm travel range the pitch and yaw errors correspond to Abbe errors of 3.2 nm and 3.6 nm respectively. However, a typical measurement of surface texture would take place over less than 1 mm and over this displacement the Abbe error for both pitch and yaw is 0.50 nm. Table 5.11 shows the contribution of the Abbe error to the uncertainty. The value of 0.02 arc seconds is the standard deviation of the pitch and yaw error experiments and the angle resolution of the autocollimator is 0.01 arc seconds.

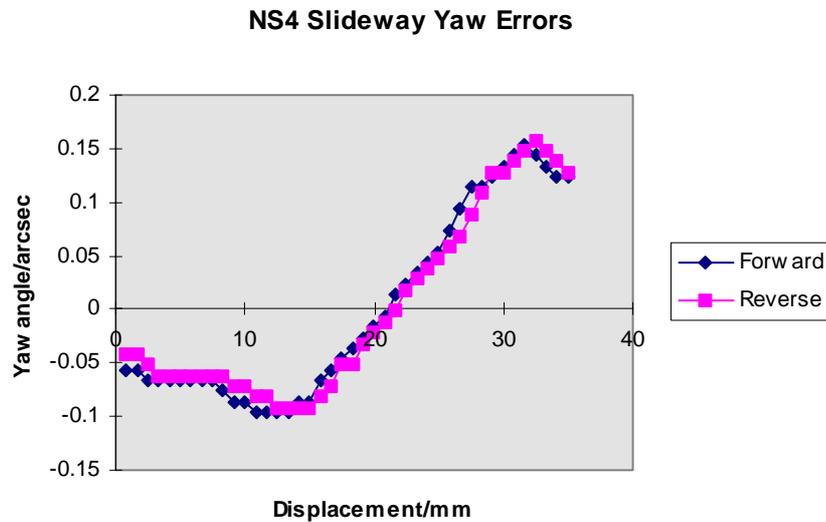


Figure 5.3 x axis slideway yaw errors

Source	Size	k	Standard uncertainty	Sensitivity coefficient	Contribution

Uncertainty due to x straightness	0.02 arc sec	$\sqrt{3}$	0.28	1	0.28 nm
Resolution of autocollimator	0.01 arc sec	$\sqrt{12}$	0.07	1	0.07 nm
					0.29 nm

Table 5.11 Summation of uncertainties due to Abbe error

5.11 DEGREES OF FREEDOM OF THE INFLUENCE QUANTITIES

Table 5.12 lists the degrees of freedom associated with each source of measurement uncertainty plus the magnitude of the uncertainty. For the optical path difference and metrology frame uncertainty contributions, the degrees of freedom are 19 because the noise measurements were repeated twenty times, *i.e.* $(20 - 1) = 19$ degrees of freedom. For the Abbe error contribution ten repeat measurements were performed, hence nine degrees of freedom. For the remaining uncertainty contributions, the standard uncertainty is obtained from rectangular distributions. According to *GUM* [§G.4.3] this assumes the number of degrees of freedom is infinite, or equally that the probability of a measured value lying outside the bounds of the distribution is extremely small.

Applying equation (5.4) gives an effective number of degrees of freedom of 45 .

5.12 EXPANDED UNCERTAINTY FOR A LENGTH MEASUREMENT

Adding the uncertainty contributions in table 5.12 in quadrature and multiplying by a coverage factor of $k = 2$ (justified with 45 degrees of freedom), gives the following combined standard uncertainty

$$u_c^2(L_{x,z}) = 2 \times \sqrt{0.66^2 + (1.5 \times 10^{-3} L)^2} \quad (5.41)$$

Source of uncertainty	Uncertainty / nm	Degrees of freedom
Optical path difference	$\sqrt{(1 \times 10^{-2} L)^2 + 0.37^2}$	19
Air refractive index	$3.8 \times 10^{-2} L$	∞
Metrology frame & detection system	0.37	19
Imperfect optics & stray beams	0.23	∞
Deadpath length	0.12	∞
Abbe error	0.29	9

Table 5.12 Sources of uncertainty and their associated degrees of freedom. L is in millimetres.

As an example, if L_z were equal to 1 mm, then $u_c^2(L_z) = 1.322$ nm and if $L_z = 1$ nm, $u_c^2(L_z) = 1.320$ nm. It is clear that in most cases the length dependent uncertainty contribution is negligible. The uncertainty in NanoSurf IV is quoted as 1.32 nm at $k = 2$ with 45 degrees of freedom.

5.13 UNCERTAINTIES IN MEASURING SURFACE TEXTURE PARAMETERS

In order to put a meaningful number to measured surface texture data and aid in understanding the functionality of a surface, various parameters, such as the average or peak-to-valley height, can be calculated. It is extremely rare to see rigorous uncertainty analysis applied to the calculation of uncertainty in a parameter. This section applies the guidelines laid down in the *GUM* to calculate the uncertainty in a given surface texture parameter. Only the parameters advocated by the ISO standards are considered. Note that the effects of any mechanical, electrical or computational filtering of the measured data are not considered here (with the exception of the calculation of the average reference line).

5.14 REFERENCE DATA

To obtain a reference data set, a 0.2 mm trace of an optical flat (NPL specimen D19) was taken at a speed of 0.1 mm per minute. The x and z displacements were measured at 2048 sampling points. A first-order polynomial least-squares fit was removed from the data - this filtering is necessary due to the lack of levelling of the specimen.

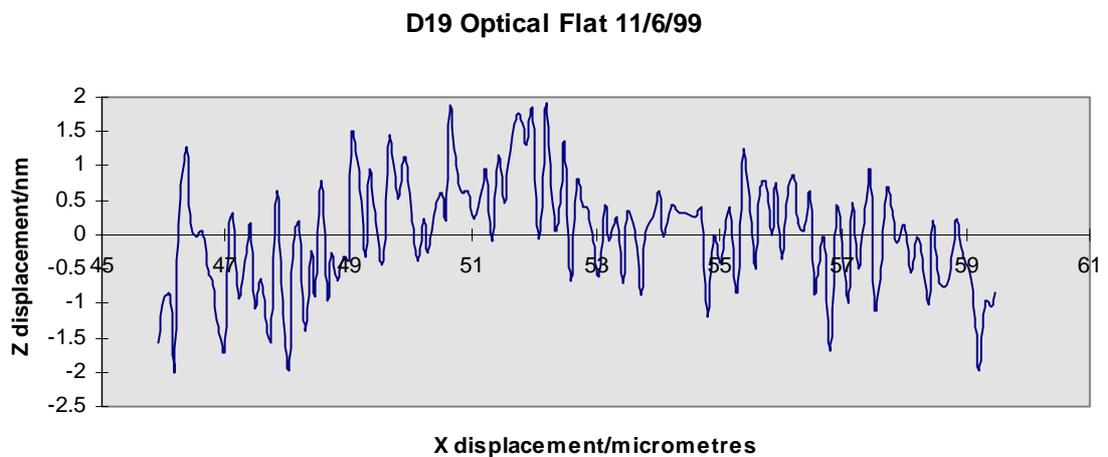


Figure 5.4 The reference data

The dominant frequency component that is clearly evident in the trace is the resonant frequency of the probing system. As this is only a reference set of data, no post measurement filtering or de-trending has been carried out (Rothe, Duparré, *et. al.* 1994). All calculations were carried out using MATLAB (The Math Works Inc.) to ensure the reliability of any mathematical algorithms. Figure 5.4 presents the first 150 data points.

5.15 UNCERTAINTY IN THE CALCULATION OF THE AVERAGE REFERENCE LINE

Before most parameters can be calculated an average reference line has to be fitted to the x and z data. The fitting equation is given by

$$z = a + bx \tag{5.43}$$

where a is the intercept on the z axis and b is the gradient. Uncertainties in a and b must be propagated through to the parameters.

To find $u(a)$ and $u(b)$ we must carry out a least-squares fit to the measured data assuming there are errors in both x and z . In the general case, this problem has no analytical solution, but when $u(x)$ can be assumed proportional to $u(z)$, as is the case with NanoSurf IV, the problem can be circumvented (Bruzzone & Moreno 1998). When only uncertainties in z are considered, the least-squares fitting requires the minimisation of

$$\chi^2 = \sum_{i=1}^n \frac{(z_i - a - bx_i)^2}{u^2(z_i)}. \quad (5.44)$$

The problem has the solutions

$$b^* = \frac{SS_{xz} - S_x S_z}{SS_{xx} - S_x^2} \text{ and } a^* = \frac{S_{xx} S_z - S_x S_{xz}}{SS_{xx} - S_x^2} \quad (5.45)$$

where

$$\begin{aligned} S &= \sum_{i=1}^n \frac{1}{u^2(z_i)} & S_x &= \sum_{i=1}^n \frac{x_i}{u^2(z_i)} & S_z &= \sum_{i=1}^n \frac{z_i}{u^2(z_i)} \\ S_{xx} &= \sum_{i=1}^n \frac{x_i^2}{u^2(z_i)} & S_{zz} &= \sum_{i=1}^n \frac{z_i^2}{u^2(z_i)} & S_{xz} &= \sum_{i=1}^n \frac{x_i z_i}{u^2(z_i)} \end{aligned} \quad (5.46)$$

The fit parameters' uncertainties are then given by

$$u(a^*) = \sqrt{\frac{S_{xx}}{SS_{xx} - S_x^2}} \text{ and } u(b^*) = \sqrt{\frac{S}{SS_{xx} - S_x^2}}. \quad (5.47)$$

If errors in both co-ordinates are considered, the effective variance method states that the expression (Barker & Diana 1974)

$$\chi^2 = \sum_{i=1}^n \frac{(z_i - a - bx_i)^2}{u^2(z_i) + u^2(x_i)} \quad (5.48)$$

must be minimised. Equation (5.48) has no analytical solution in the general case. However, provided $u(x) = du(z)$, where d is a constant (unity in the case of NanoSurf IV), equation (5.48) reduces to

$$\chi^2 = \frac{1}{1+b^2d^2} \sum_{i=1}^n \frac{(z_i - a - bx_i)}{u^2(z_i)}. \quad (5.49)$$

Minimisation of equation (5.49) leads to a quadratic expression for b , but only the root given here is to be considered (Bruzzone & Moreno 1999)

$$db = \frac{1}{2} \left\{ -\frac{1}{db^*} + \frac{db^*}{r^2} + \frac{1}{db^*} \left[\left(1 - \frac{(db^*)^2}{r^2} \right)^2 + 4(db^*)^2 \right]^{1/2} \right\} \quad (5.50)$$

where r is the linear correlation coefficient, given by

$$r^2 = \frac{(SS_{xz} - S_x S_z)^2}{(SS_{xx} - S_x^2)(SS_{zz} - S_z^2)}. \quad (5.51)$$

Once the slope b has been obtained, it is possible to evaluate a by means of

$$a = \frac{S_z - S_x b}{S}. \quad (5.52)$$

For the case of proportional errors in x and z , the uncertainties in a and b can be found from the following

$$u^2(b) = u^2(b^*)f(db^*, r^2) \quad (5.53)$$

and

$$u^2(a) = \frac{1+b^2d^2}{S} \{1 + [u^2(a^*)S - 1]f(db^*, r^2)\} \quad (5.54)$$

where

$$f(db^*, r^2) = (1+b^2d^2)^2 \frac{1 + (db^*)^2/r^2}{[1 + d^2(2bb^* - b^{*2}/r^2)]^2}. \quad (5.55)$$

For the reference data set the uncertainties are $u(a) = 0.46$ nm and $u(b) = 8.7 \times 10^{-6}$. Assuming these uncertainties are rectangularly distributed, they must be divided by $\sqrt{3}$. Once the calculated least-squares curve has been removed from the data, a given value of z_i is given by

$$z'_i = z_i - (a + bx_i). \quad (5.56)$$

An uncertainty in z' is now given by the following equation

$$u^2(z') = u^2(z) + u^2(a) + x^2u^2(b) + b^2u^2(x). \quad (5.57)$$

5.16 UNCERTAINTIES IN THE PARAMETERS

Where a parameter only applies to one length measurement, for example R_p , equation 5.41 should be applied appropriately. Where a parameter is the sum of parameters calculated for their corresponding sampling lengths over the entire evaluation length, a simple quadrature sum should be applied.

5.16.1 UNCERTAINTY IN R_a

The equation given in Appendix B for R_a is for continuous data sampling within the sampling length. For the case of a real surface texture measuring instrument, with a finite sampling length, R_a is given by

$$Ra = \frac{1}{N} \sum_{i=1}^N |Z_i|. \quad (5.58)$$

To calculate the uncertainty, $u_c(Ra)$, equation (5.1) is applied to equation (5.58), *i.e.*

$$u_c^2(Ra) = c_{z_i}^2 u^2(z_i) = \frac{1}{N} \sum_{i=1}^N u^2(z_i) \quad (5.59)$$

where all the higher order terms are equal to zero. The combined standard uncertainty of the R_a parameter is simply the sum of the individual uncertainties divided by the number of samples, or the average uncertainty. There is, however, a subtlety here. By definition, it is not physically possible to have a negative R_a . But, if it is assumed that the measurement

uncertainties are normally distributed, it is possible to have a confidence interval that allows negative Ra values. For example, $Ra = 0.1 \text{ nm} \pm 0.5 \text{ nm}$ could have values of Ra that vary from -0.4 nm to 0.6 nm . The problem has arisen due to the arbitrary assumption that the uncertainties in Ra are normally distributed about the “true value”, whereas the mathematical definition of Ra rules this assumption out - any measurement noise will positively bias the value of Ra . As each surface will have a different distribution of surface heights, a value for Ra can only be stated with a standard uncertainty - to state an expanded uncertainty has implied that we have prior knowledge of the height distribution of the surface. Of course, the distribution could be calculated for each surface, but this would not be practical without rigorous software backup (Cox & Harris 1999).

For the reference data set, $Ra = 0.65 \text{ nm}$ and its standard uncertainty is 0.6 nm .

5.16.2 UNCERTAINTY IN Rq

The equation for a sampled evaluation of Rq is given by

$$Rq = \sqrt{\frac{1}{N} \sum_{i=1}^N Z_i^2} \quad (5.60)$$

The associated uncertainty in Rq is given by

$$u_c^2(Rq) = c_{z_i}^2 u^2(z_i) = \frac{1}{N^2 Rq^2} \sum_{i=1}^N z_i^2 u^2(z_i). \quad (5.61)$$

The higher order terms are not equal to zero but will be negligibly small. For the reference data set $Rq = 1.4$ nm with a standard uncertainty of 0.40 nm. Again, because the value for Rq is always positive, the same points raised above for Ra apply for Rq .

5.16.3 UNCERTAINTY IN Rsk AND Rku

The uncertainties in the skewness and kurtosis of a surface cannot be calculated using the guidelines laid down in the *GUM*. This is because they are dimensionless nonlinear parameters and their definitions lead to complexities with the higher order terms that give nonsensible results.

5.17 REFERENCES

British Standard BS 1339 (1965) (London: British Standards Institution)

Armstrong T R, Fitzgerald M P 1992 An autocollimator based on the laser head of a compact disc player *Meas. Sci. Technol.* **3** 1072-1076

Barrel H, Sears J E 1939 The refraction and dispersion of air for the visible spectrum *Phil. Trans. Roy. Soc. London* **A238** 1-64

Birch K P, Downs M J 1988 The results of a comparison between calculated and measured values of the refractive index of air *J. Phys. E: Sci. Instrum.* **21** 694-695

Birch K P, Downs M J 1994 Correction to the updated Edlén equation for the refractive index of air *Metrologia* **31** 315-316

Bönsch G, Potulski E 1998 Measurement of the refractive index of air and comparison with the Edlén formulae *Metrologia* **35** 133-139

Bruzzone H, Moreno C 1998 When errors in both coordinates make a difference in the fitting of straight lines by least squares *Meas. Sci. Technol.* **9** 2007-2011

Colclough A R 1987 Two theories of experimental error *J. Res. NBS* **92** 167-185

Cox M J, Harris P 1999 Up a GUM tree? *The Standard* **99** 36-38

Decker J E, Ulrich A, Pekelsky J R 1998 Uncertainty of gauge block calibration by mechanical comparison: a worked example for gauge blocks of dissimilar materials *Proc. SPIE* **3477** 225-246

Downs M J, Ferris D H, Ward R E 1990 Improving the accuracy of the temperature measurement for gases by correction for the response delays in the thermal sensors *Meas. Sci. Technol.* **1** 717-719

Downs M J 1998 Realisation of the optimum performance from length-measuring interferometers by atmospheric refractometry *Proc. SPIE* **3477** 46-53

Edlén B 1953 The dispersion of standard air *J. Opt. Soc. Am.* **43** 339-344

Edlén B 1966 The refractivity of air *Metrologia* **2** 71-80

GUM Guide to the expression of uncertainty in measurement, 1st Edition, International Organization for Standardization (ISO), Switzerland 1993

Hariharan P 1997 Interferometric measurements of small-scale surface irregularities: sources of errors *Opt. Eng.* **36** 2330-2334

Leach R K 1999 Calibration, traceability and uncertainty issues in surface texture metrology *NPL Report CLM7* 1-56

Mana G 1989 Diffraction effects in optical interferometers illuminated by laser sources *Metrologia* **26** 87-93

Rothe H, Duparré A, Jacobs S 1994 Generic detrending of surface profiles *Opt. Eng.* **33** 3023-3030

Self S A 1983 Focussing of spherical Gaussian beams *Appl. Opt.* **22** 658-661

Zanoni C A 1988 *High precision interferometric linear and angular displacement interferometry*
Zygo Corporation document

CHAPTER 6

EXPERIMENTAL RESULTS

"In the real world, absolute precision is an ideal that can never be reached."

Richard P Feynman

6.1 INTRODUCTION

The aim of this chapter is to present a sufficient amount of experimental results obtained from NanoSurf IV to demonstrate that the instrument is operating correctly. This is by no means an easy task. When precision instruments are designed it is common to assume that, because the instrument uses laser interferometry as its metrology, it is automatically correct. This is a poor assumption in practice - the degree of correctness of the results from an instrument must be proven. Often proof can only be obtained by comparison with other instruments that have themselves been rigorously tested for accuracy, uncertainty and traceability. Chapter 5 demonstrated that interferometry in non-ideal conditions does not always give the desired result - instead only a probability that the result lies within certain bounds, *i.e.* the expanded uncertainty.

A sound method for testing the performance of an instrument is to use what are known as natural or quantum standards (see, for example, Cook 1972, Thywissen, Johnson, *et.al.* 1998). For scanning probe microscopes a good quantum standard is the lattice spacing of

atoms in a crystal or similar natural structure (see §2.2). This sort of standard is suitable for instruments that operate in the lower left hand corner of amplitude-wavelength space (*i.e.* sub-nanometre accuracy in both amplitude and wavelength), but is no use for an instrument like NanoSurf IV. The physical size of the stylus used on NanoSurf IV rules out the use of quantum standards - to date no stable quantum structures exist in the size region of 100 nm.

For the reasons described above a series of experiments have been designed to characterise NanoSurf IV that aim to show that it is operating within its specification and that it compares sufficiently well with other traceable instruments.

All equipment that forms NanoSurf IV was given at least one hour to stabilise before any measurements were carried out and all the surface traces were always checked by running the instrument in both forward and reverse directions. The experiments described below did not highlight any significant effects associated with the direction of traverse.

6.2 DETERMINATION OF NOISE LEVELS

A number of experiments were carried out to assess the noise levels on NanoSurf IV. Some of results of these experiments have been used in chapter 5 as inputs to the uncertainty analysis.

6.2.1 STATIC NOISE LEVELS

The static noise level of an interferometer is defined here as the standard deviation of 500 measurements of displacement at 100 Hz sampling frequency (duration 5 seconds). In the case of the z axis interferometer, the probe was brought into contact with a flat specimen with a stylus force of 0.1 mN and in the case of the x axis interferometer the path lengths from the beamsplitter to the retro-reflectors were made nominally equal. Figures 6.1 and 6.2 show the static noise in the x and z axes respectively. The standard deviations for these results are 0.20 nm and 0.37 nm for the x and z axes respectively. The same test was also carried out over 5 seconds at 50, 200 and 300 Hz sampling frequencies but no significant change in the standard deviation of the results was measured. Figures 6.1 and 6.2 show unfiltered results.

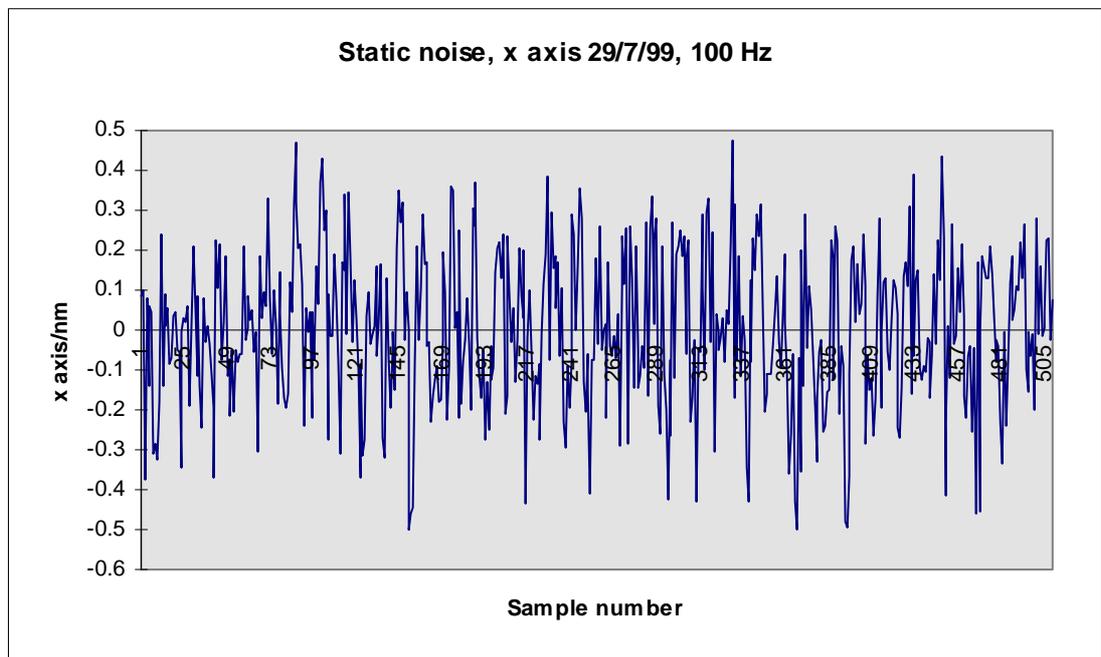


Figure 6.1 Static noise in the x axis interferometer over 5 seconds

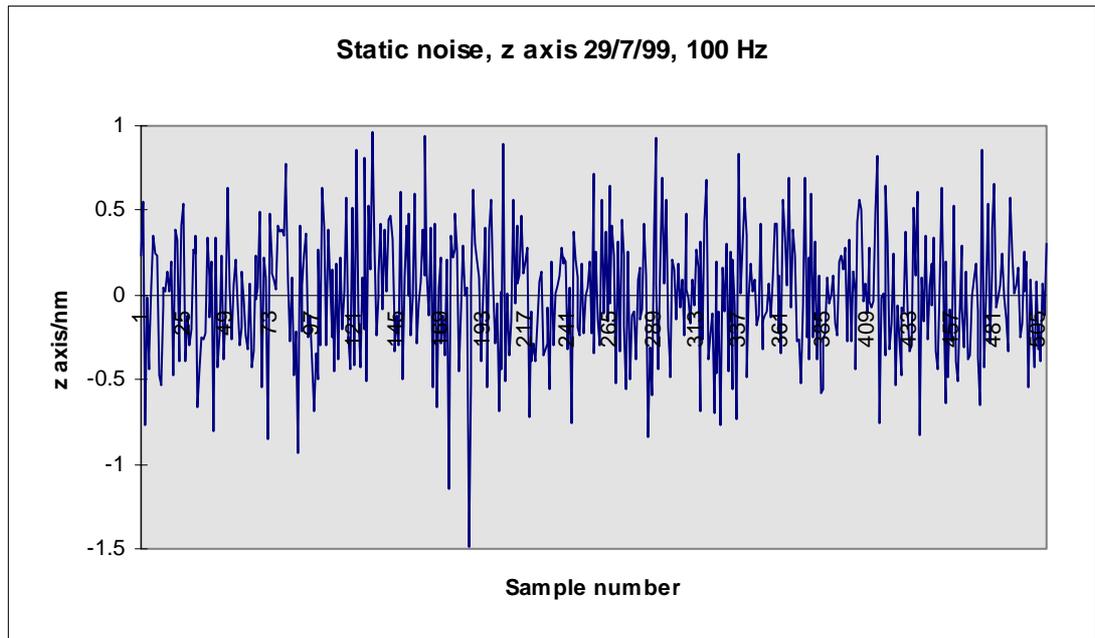


Figure 6.2 Static noise in the z axis interferometer over 5 seconds

Figure 6.3 and 6.4 show the static noise results in frequency space following a fast Fourier transform. It is clear that, with the exception of some minor mechanical resonance peaks the frequency response in both x and z axes is effectively flat.

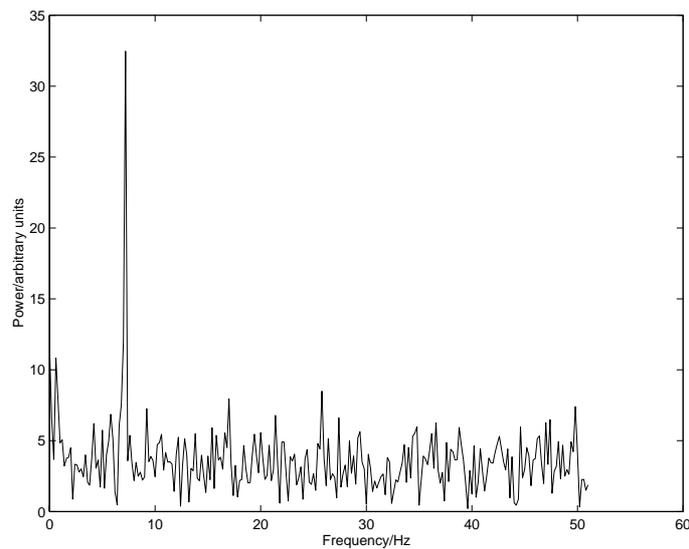


Figure 6.3 Frequency response of the static noise in the x axis interferometer

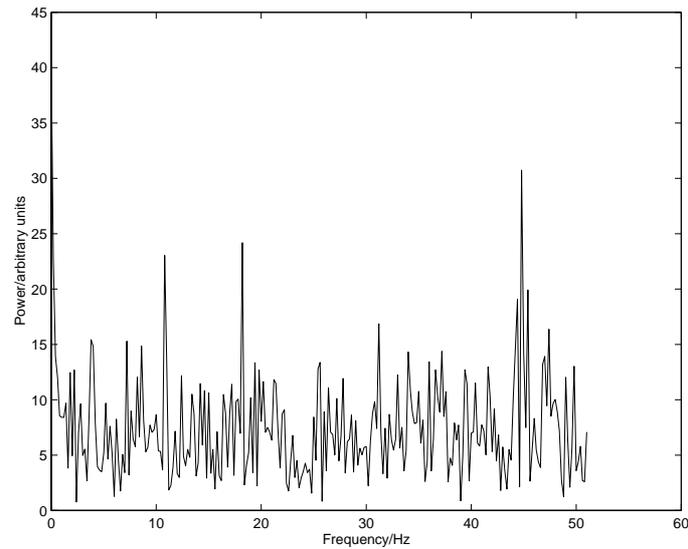


Figure 6.4 Frequency response of the static noise in the z axis interferometer

6.2.2 DYNAMIC NOISE LEVELS

The dynamic noise levels on NanoSurf IV are very difficult to measure as a measurement must be in progress while the noise levels are being assessed and it is impossible to completely separate the effects of the static and dynamic noise from the actual surface being measured. Dynamic noise (not necessarily dynamic effects) will only affect the very high accuracy measurements of for example, an optical flat. Where low uncertainty is required only the slower of the nine speeds is used. To investigate the effect of dynamic noise at the lower scanning speeds a 10 μm trace of an optical flat was made at speeds of 0.01, 0.02, 0.05 and 0.1 mm per minute. A 2 μm radius chisel stylus, a 100 Hz sampling rate and a nominal stylus force of 4 μN were used for each scan.

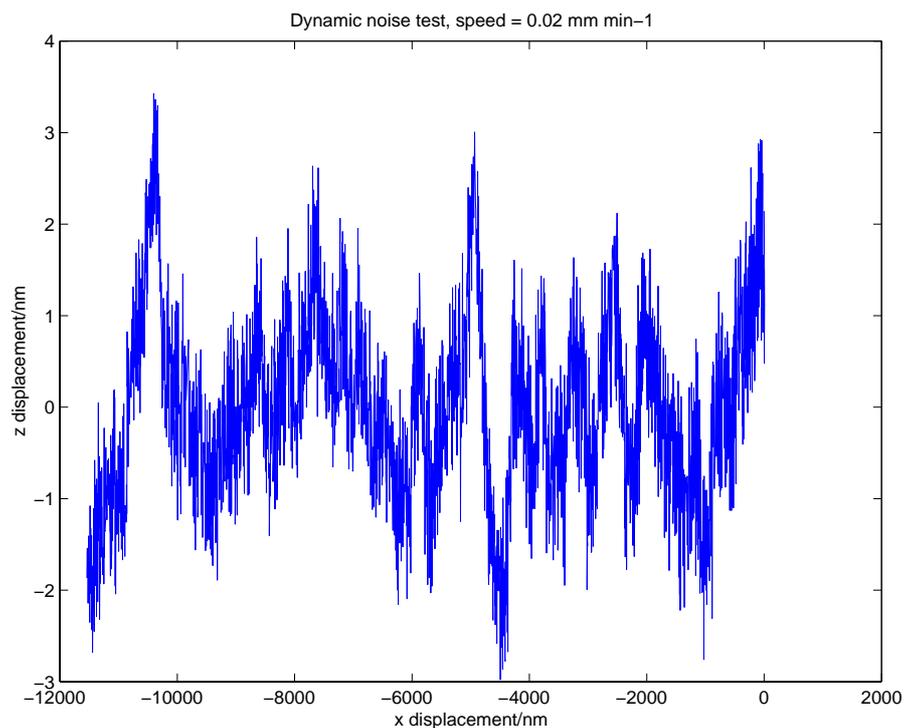


Figure 6.5 Dynamic noise test on specimen D19

Figure 6.5 shows a typical scan of the flat at a speed of 0.02 mm per minute. A second order least squares polynomial fit has been removed from the data to minimise the effect of the topography of the flat. The standard deviations at the four speeds are shown in table 6.1.

Scan speed / mm per minute	Standard deviation of the z displacement / nm
0.01	1.07
0.02	0.98
0.03	1.08
0.10	0.89

Table 6.1 Dynamic noise results

These results show that there is no significant difference between the noise in the system at the different speeds and that the noise level is not increasing with speed. These tests also point out the difficulty of measuring dynamic noise at the nanometre level when the expanded uncertainty of the instrument is around one nanometre.

6.3 MAGNIFICATION STANDARDS

Many surface texture measuring instruments are calibrated using magnification standards that take the form of rectangular grooves in a substrate. These are useful for checking the vertical magnification factor of an instrument but do not give metrological information regarding horizontal magnification or the transmission characteristics (frequency response) of an instrument. BS 6364: 1987 and BS 1134: 1988 refer to these standards as Type A1 and describe the measurement process as follows:

A continuous straight mean line equal in length to three times the width of the groove is drawn over the groove to represent the upper level of the surface and another to represent the lower level, both lines extending symmetrically about the centre of the groove (see figure 6.5). To avoid the influence of any rounding of the corners, the upper surface on each side of the groove is to be ignored for a length equal to one-third of the width of the groove. The surface at the bottom of the groove is assessed only over the central third of its width. The portions to be used for assessment purposes are therefore those at A, B, and C in figure [6.5]. The depth d of the groove shall be assessed perpendicularly from the upper mean line to the mid-point of the lower mean line. A significant number, not less than five, of evenly distributed traces shall be taken.

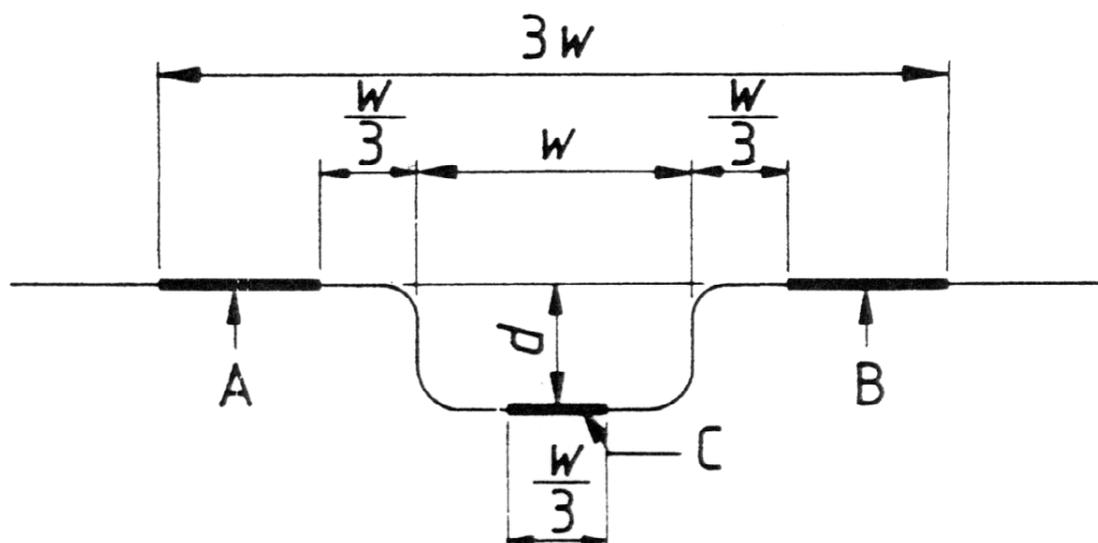


Figure 6.5 Procedure for the assessment of Type A1 standards

A computer program, named `depth.m` was written using MATLAB (MathWorks, Inc.) to analyse Type A1 artefacts according to BS 6365. Artefacts were obtained from Taylor-Hobson and take the form of three parallel grooves that have been etched in a glass substrate (the presence of three grooves is merely historical - only the central groove is used in practice). There are two sets of grooves on each artefact. Artefact one has a nominally $3.0\ \mu\text{m}$ and $0.3\ \mu\text{m}$ set of grooves and artefact two has a $30\ \text{nm}$ set of grooves plus one other set that has not been used. The depth of the grooves has been measured by Taylor-Hobson's UKAS laboratory using a Form Talysurf (3.0 and $0.3\ \mu\text{m}$ grooves) and a Talystep ($30\ \text{nm}$ groove) measuring instruments. A nominally $2\ \mu\text{m}$ radius stylus tip was used on the Form Talysurf and a $12.5\ \mu\text{m}$ tip on the Talystep. The traceability of the Taylor-Hobson instruments is obtained by using a mechanical magnification lever and traceable length standards. The calibration certificate (serial number: 24836, May 1997) did not give the scan speed or sampling rate of the instruments (I later found out that the scan speed of the Form Talysurf was $2\ \text{mm}$ per second). The mean value of the depth has been calculated from five measurements taken $0.8\ \text{mm}$ apart.

The values quoted on the Taylor-Hobson certificate are shown in table 6.1. These measurements were repeated using NanoSurf IV. As far as possible the same measurement procedure as Taylor-Hobson was employed, although a slower scanning speed of 2 mm per minute was used, a 10 μm radius stylus tip was used for the nominally 30 nm groove and a 100 Hz sampling rate was used consistently. The NanoSurf IV results are shown in table 6.2.

The uncertainty of NanoSurf IV is much smaller than that quoted in table 6.2, but when measuring this type of artefact it is the surface texture of the nominally flat areas that can dominate the uncertainty (Teague 1978). The highest Ra value of the three areas used to calculate the height, d , has been added in quadrature to the uncertainty in height measurement with a $k = 1.96$ value used to bring the uncertainty in line with the 95% quoted on the Taylor-Hobson certificate.

Figures 6.6, 6.7 and 6.8 show typical traces of the three standards measured using NanoSurf IV and indicate the functions employed by `depth.m`. It can be seen from figure 6.8 that, when the magnitude of the surface texture of the nominally flat areas of Type A1 standards approaches the magnitude of the depth height, the method advocated in the standard begins to give what looks like dubious results. However, when one calculates the angle between the red and yellow lines in figure 6.8, it can be seen that such errors in the height of the groove do not exceed the magnitude of the surface roughness, in this case Ra is around 3 nm.

The only significant difference between the Taylor-Hobson and the NanoSurf IV results is for the nominally 3 μm groove - the difference is 33 nm. However, the difference does fall within the uncertainties of both instruments and is therefore not significant.

It must be pointed out at this stage of the development of NanoSurf IV that comparisons like the one described above, between NanoSurf IV and instruments that are expected to have far greater uncertainties of measurement, are limited in their use. In this case the results agreed well, but if this level of agreement had not resulted, it would not be NanoSurf IV that would be suspected of causing large disagreements.

Mean height of the central line	Standard deviation of measurement	Maximum uncertainty of calibration
2.98 μm	0.01 μm	$\pm 0.05 \mu\text{m}$
0.3 μm	$< 0.01 \mu\text{m}$	$\pm 0.05 \mu\text{m}$
0.031 μm	$< 0.001 \mu\text{m}$	$\pm 0.005 \mu\text{m}$

Table 6.1 Results for depth standards quoted in Taylor-Hobson certificate

Mean height of the central line	Standard deviation of measurement	Maximum uncertainty of calibration
2.947 μm	0.005 μm	$\pm 0.004 \mu\text{m}$
0.296 μm	0.005 μm	$\pm 0.003 \mu\text{m}$
0.033 μm	0.001 2 μm	$\pm 0.002 \mu\text{m}$

Table 6.2 Results for depth standards using NanoSurf IV

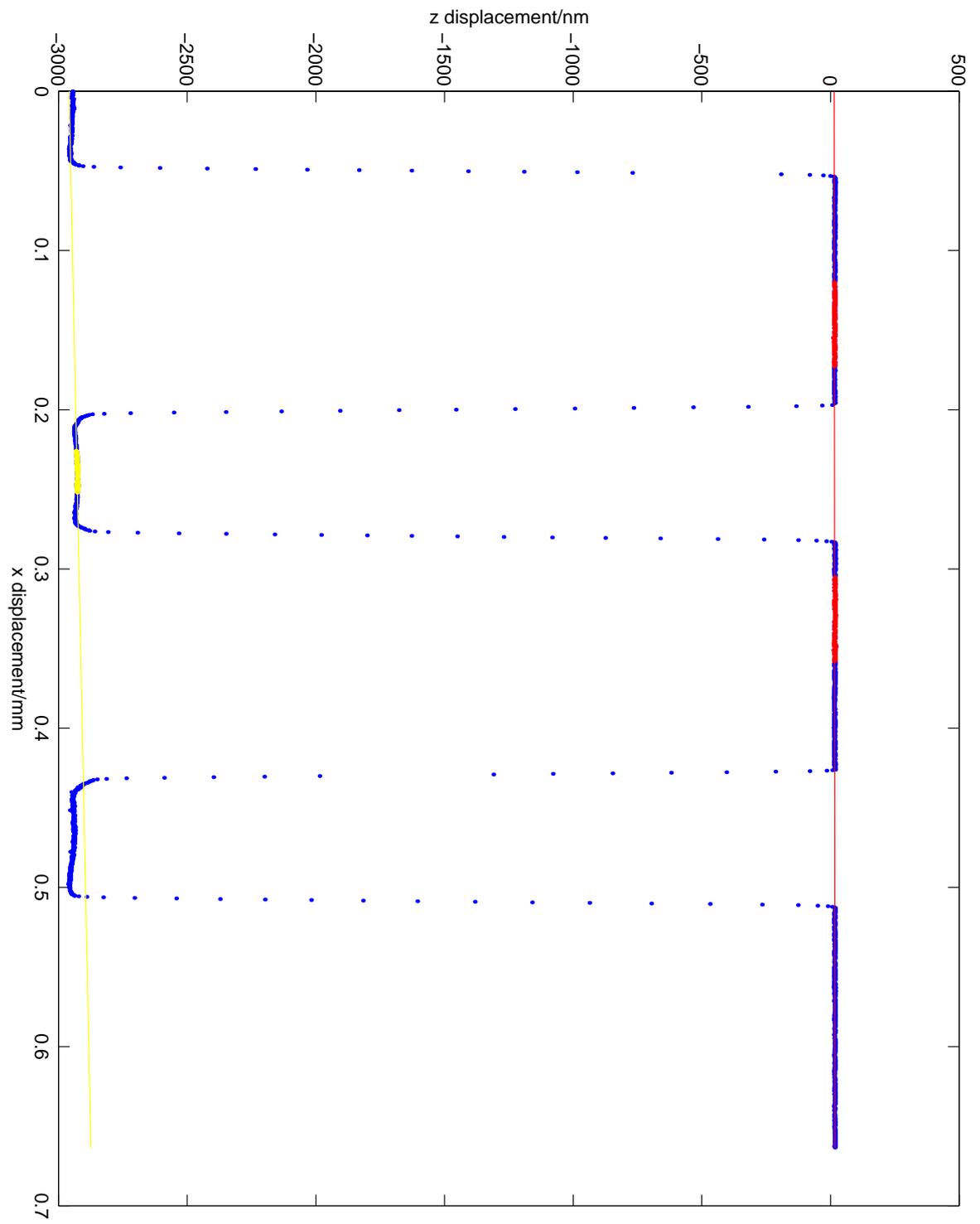


Figure 6.6 Example printout from `depth.m` for a nominally $3\ \mu\text{m}$ magnification artefact

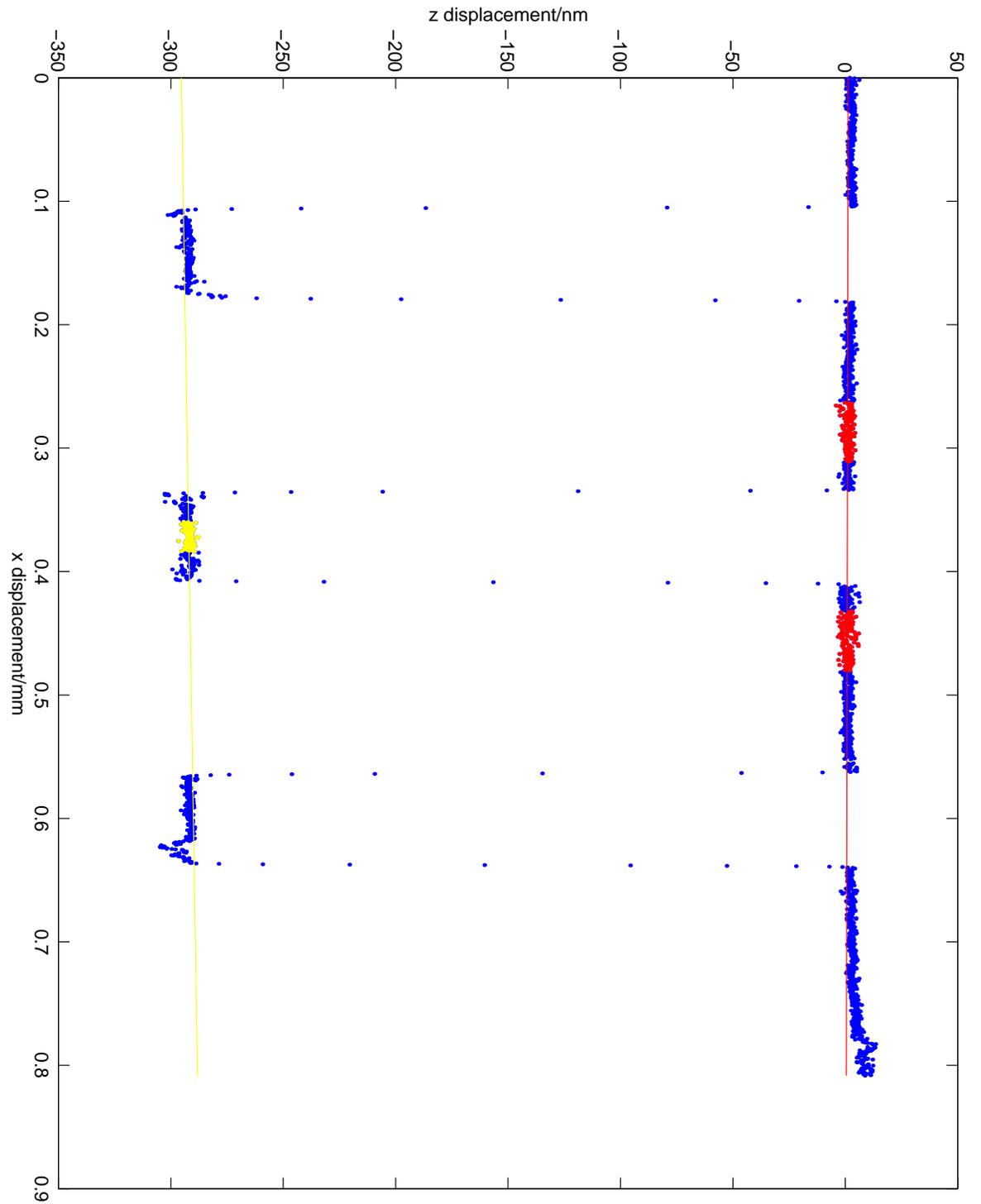


Figure 6.7 Example printout from `depth.m` for a nominally $0.3\ \mu\text{m}$ magnification artefact

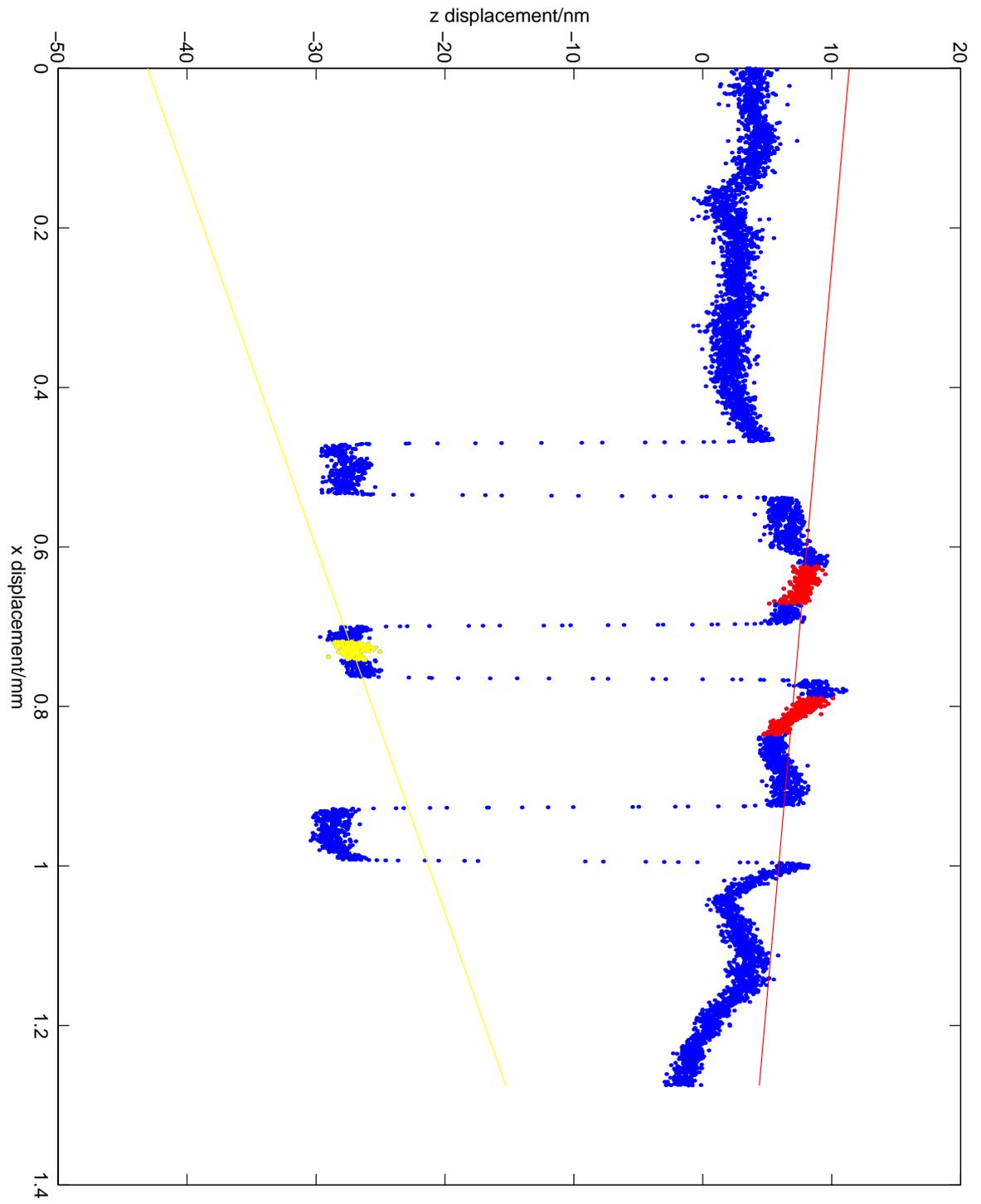


Figure 6.8 Example printout from `depth.m` for a nominally $0.03\ \mu\text{m}$ magnification artefact

6.4 COMPARISON USING PTB ROUGHNESS STANDARDS

The Physikalisch-Technische Bundesanstalt (PTB) in Germany supply Type D (see §2.3) roughness artefacts as part of their calibration service to internal and external customers (Hillman 1993). The surfaces they use have irregularly ground profiles that recur in the direction of measurement after a set length and are constructed from hardened stainless steel. The specimens are measured under set conditions and a number of surface texture parameters are calculated from the measurement data. PTB kindly lent me two of their specimens, hereafter known as the ‘fine roughness standard’ and the ‘superfine roughness standard’. The PTB measurements in general conform to ISO 3274 and ISO 4288 with the exception of the choice of cut-off length for the fine roughness standard which conforms to DIN 4768 (see Appendix A for a list of specification standards).

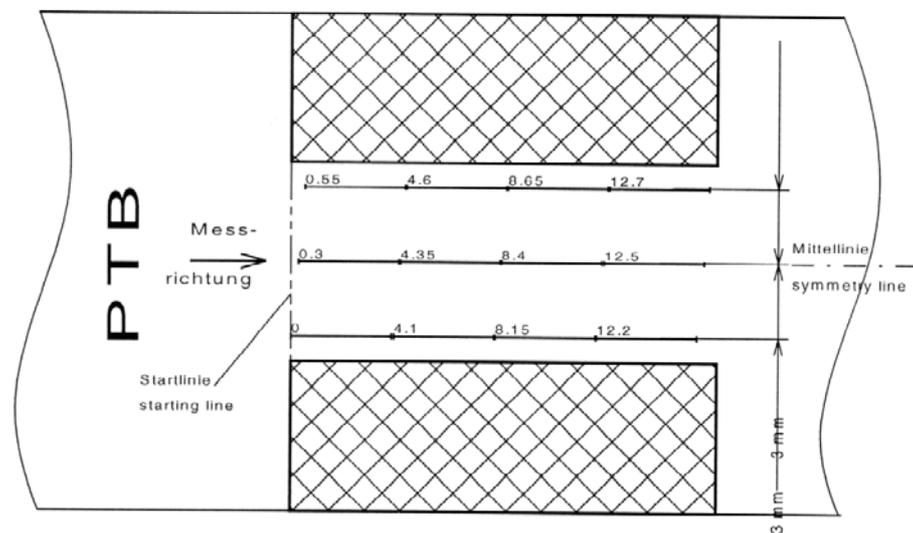


Figure 6.9 Measurement procedure for the fine roughness specimen

The measurement procedures for the fine and superfine roughness standard respectively are shown in figures 6.9 and 6.10 and table 6.3 lists the measurement conditions. As far as possible these conditions were matched when using NanoSurf IV, but the columns titled

'NanoSurf IV' in table 6.3 show that there were some deviations. PTB used a Perthen surface stylus instrument.

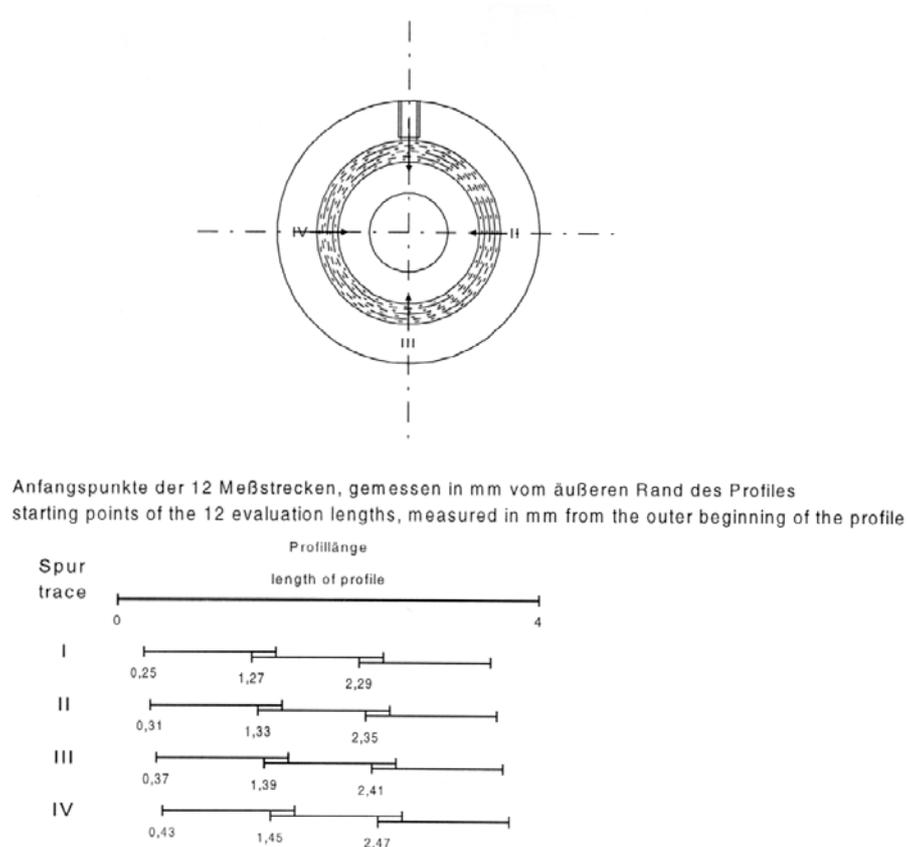


Figure 6.10 Measurement procedure for the superfine roughness specimen

In order to carry out the data processing required to calculate the surface texture parameters that are to be compared with the PTB results, it was necessary to develop a filter to take into account the long wavelength cut-offs, λ_c (Whitehouse 1994). ISO 11562: 1996 (E) stipulates that for surface texture analysis a filter should be phase correct, *i.e.* a filter that does not cause phase shifts which lead to asymmetrical distortions of the measured profile, and the weighting function should correspond to the equation of the Gaussian density function.

	Fine roughness standard		Superfine roughness standard	
	PTB	NanoSurf IV	PTB	NanoSurf IV
Tip radius/ μm	5	5	2	2
Stylus force/mN	0.6	0.2	0.025	0.039
Speed/ mm s^{-1}	0.5	0.083	0.05	0.033
Total length traced/mm	5.6	5.6	1.75	1.75
Total evaluation length/mm	4.0	4.0	1.25	1.25
Cut-off wavelength/mm	0.8	0.8	0.25	0.25
Length between samples/ μm	2.5	2.5	2.5	2.5

Table 6.3 Measurement conditions for the PTB specimens

This leads to a transmission characteristic for a short wave profile component (or high-pass) filter given by

$$\frac{a_2}{a_0} = 1 - e^{-\pi \left(\frac{\alpha \lambda_c}{\lambda}\right)^2} \quad (6.1)$$

where a_0 is the amplitude of the profile before filtering, a_2 is the amplitude of the profile after filtering, λ is the wavelength of the profile and α is defined as

$$\alpha = \sqrt{\frac{\ln 2}{\pi}} . \quad (6.2)$$

A program was written using MATLAB to filter the measured data (`kjfilter.m`). The input parameters are the x and z axis data plus the cut-off wavelength. The program first defines the frequency response of the filter using equation (6.1) and then designs an n^{th} order finite impulse response (FIR) filter using a Hamming window (Stout, Sullivan, *et. al.* 1993). To do this in MATLAB the function `FIR2` is used. `FIR2` produces a real filter, b , with linear phase and coefficients obeying

$$b(k) = b(n + 2 - k) \quad (6.3)$$

where $k = 1, 2, \dots, n + 1$. The z axis data are then digitally filtered using the MATLAB function `FILTER` with input parameters b and the z axis data. `FILTER` is a direct form II transposed implementation of the standard difference equation given by

$$z_f(n) = b(1)z(n) + b(2)z(n - 1) + \dots + b(nb + 1)z(n - nb) \quad (6.4)$$

where z_f is the filtered data. Figure 6.11 is a Bode plot (Baher 1990) of the filter with $n = 255$ and $\lambda_c = 0.25$ mm. Figure 6.12 shows the profile measured on NanoSurf IV of the superfine roughness standard after filtering.

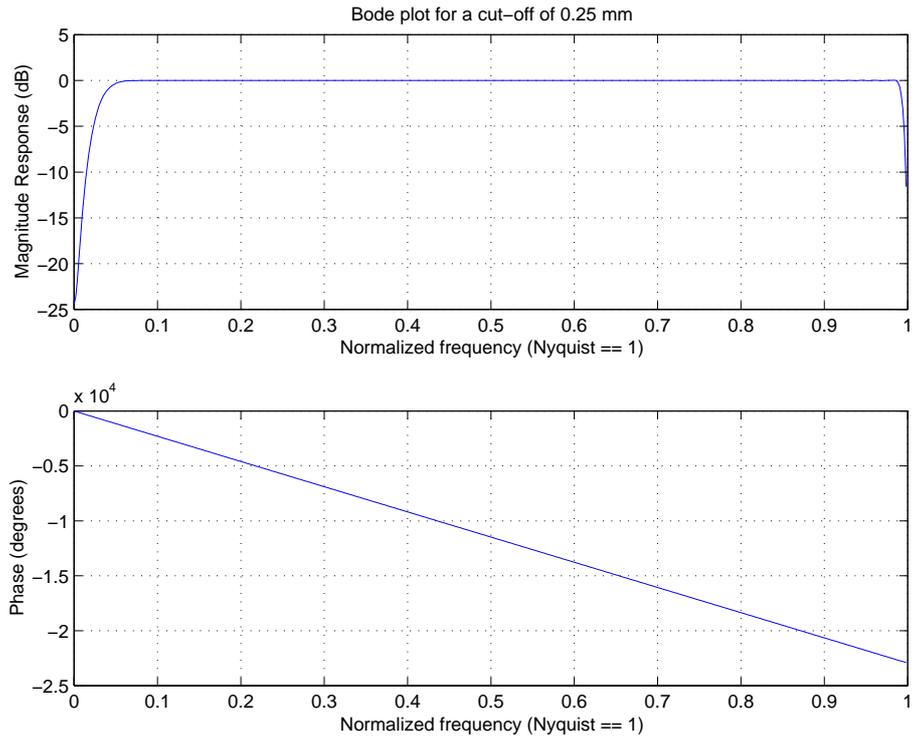


Figure 6.11 Frequency and phase response of kfilter

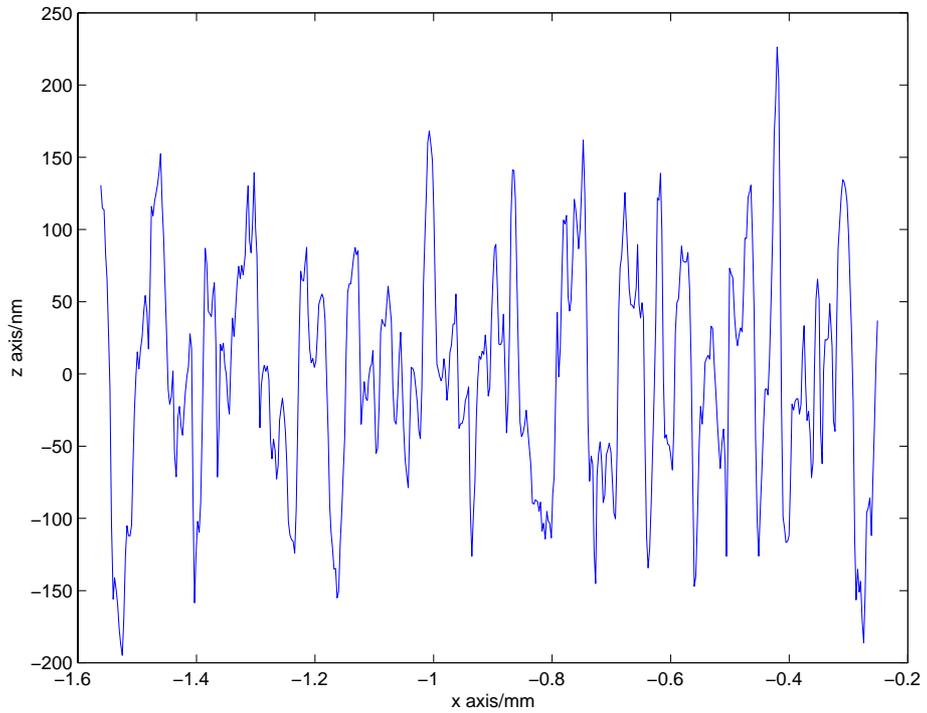


Figure 6.12 PTB superfine roughness standard - filtered data

The documentation that accompanied the roughness standards contained many surface texture parameters that are not stipulated in ISO 4287 (1997); in fact all of the parameters, with the exception of R_a , are defined in DIN 4776 (1990). For this reason, only the measured values of R_a are compared here. Table 6.4 compares the PTB results with those using NanoSurf IV.

Sample (lab)	R_a/nm	Standard deviation /nm	Spread/nm	Difference in R_a/nm	Percentage difference
Fine (NPL)	192.5	1.3	190.6 to 194.8		
Fine (PTB)	191.0	12.0	188.0 to 194.0	1.5	0.8
Superfine (NPL)	59.5	1.8	58.3 to 61.6		
Superfine (PTB)	58.7	1.6	57.2 to 61.2	0.8	1.3

Table 6.4 Comparison of NanoSurf IV results with those of PTB

There is extremely good agreement between the results obtained with NanoSurf IV and those obtained with the PTB instrument. In both cases the PTB value for R_a is slightly smaller than the NanoSurf IV value but these discrepancies are within the uncertainties of the two instruments.

6.5 SINUSOIDAL SAMPLES

From the point of view of “pure” metrology, the ideal calibration artefact for a surface texture measuring instrument is a perfect sine wave. It has a single period, a single amplitude, is represented by a single point in Amplitude-Wavelength space and the measurement is not affected by the instrumental response function. Unfortunately, this type of artefact is not typical of real surfaces. However, if the profile of a near sinusoidal grating is known accurately, the distortion due to the stylus can be determined and used to refine measurements on real surfaces. If standards are available with a range of amplitudes and periods it is possible to verify the calibration of an instrument and determine the instrumental response function.

To date, NPL has developed gratings different amplitudes and periods and is intending to increase the range. To produce the gratings two crossed, expanded and collimated laser beams expose a layer of photoresist on a glass substrate. This results in a near-sinusoidal grating. Electro-formed nickel is then coated onto the grating and the glass material is removed. This forms a master grating that can be copied by further electro-forming. The copy is an inverse of the master grating and if the two are measured a mathematical process can be used to extract the offset curve that results from the use of a stylus with non-zero radius of curvature (Cox & Harris 1997).

NanoSurf IV is used to calibrate these gratings that in turn are used to calibrate other instruments. Only the results for a nominally 8 μm grating are presented here mainly to give confidence in the performance of the x interferometer. Recently, a service for the calibration of gratings and periodic structures based on optical diffraction has been introduced at NPL (Nunn 2000). The technique involves directing a laser beam onto the

surface of the standard and angling the surface of the standard so that the diffracted orders are directed back towards the laser. The method is relatively low cost and although in its simplest form it cannot measure any variation of the period of the standard over areas that are smaller than the beam diameter, it can nevertheless probe large areas or even the entire surface of the standard. The measurement obtained will be the mean period of the standard covered by the beam diameter weighted by the Gaussian intensity distribution of the laser beam. Traceability of the measurements is gained through the frequency of the laser source.

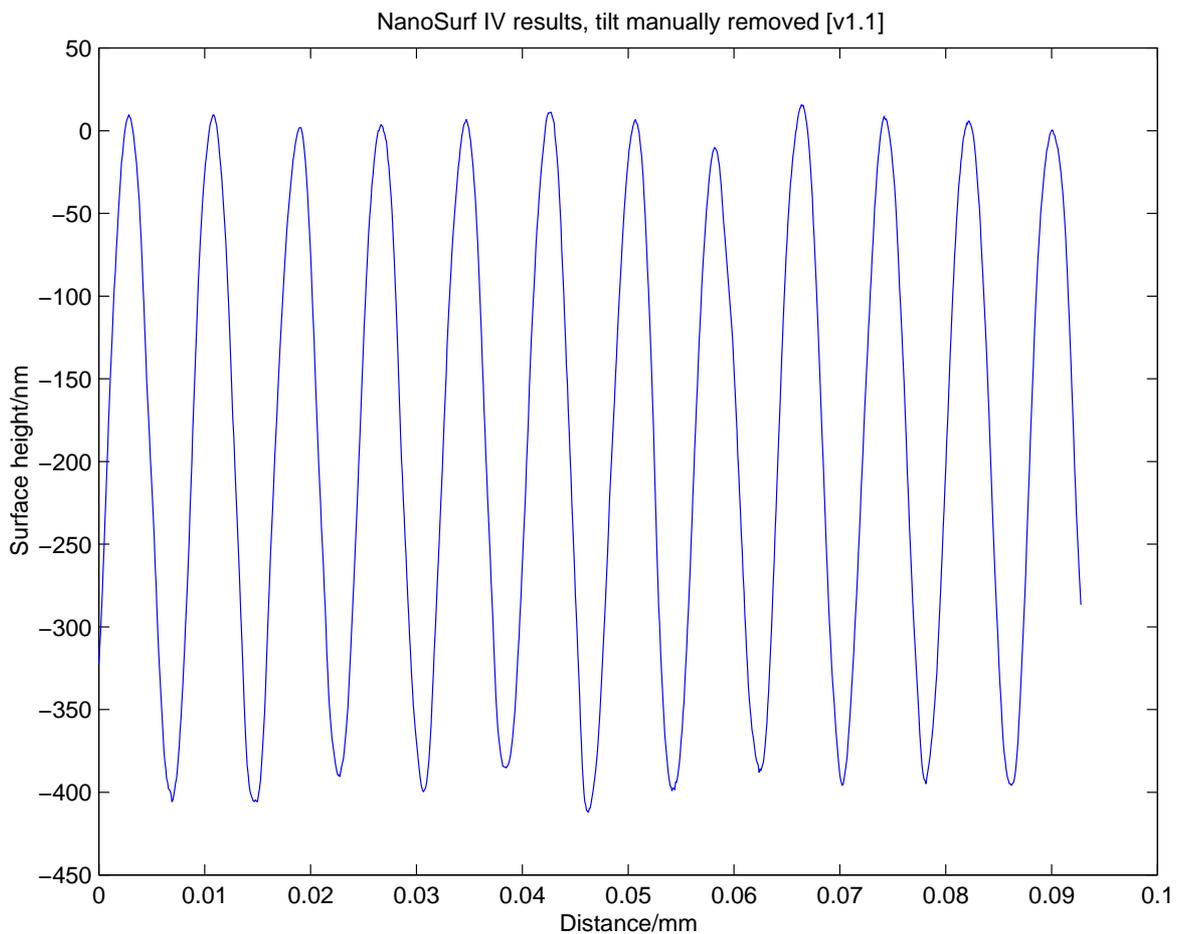


Figure 6.13 Nominally 8 μm period sinusoidal standard grating

Results of measurements obtained both with the above described diffractometer and NanoSurf IV are compared below. Figure 6.13 shows a trace using NanoSurf IV of the master nominally 8 μm period grating, known as α_1^+ .

Measurements of α_1^+ were made along five lines from different areas of the grating. The average amplitude measured by NanoSurf IV was 212.6 nm with a standard deviation of 10.9 nm. The expanded uncertainty in the measurement of a 212.6 nm amplitude is 1.3 nm, so the relatively high standard deviation is due to variations of the amplitude over the surface of the grating. It is clear from figure 6.13 that the amplitude is varying over a single trace. The average standard deviation of the amplitude, *i.e.* the variation of the amplitude for a single trace, for the five areas sampled was 13 nm.

The average period measured by NanoSurf IV was 7928.8 nm with a standard deviation of 0.5 nm, *i.e.* the period is extremely constant over the surface of the grating. The average measurement of the period using the diffractometer was 7926.6 nm with a standard deviation of 0.9 nm and an expanded uncertainty of 3 nm. The difference between the NanoSurf IV result and that of the diffractometer is 2.2 nm, well within the expanded uncertainties of the two instruments. This agreement lends great confidence in the results from the x interferometer.

6.6 CERAMIC ROUGHNESS SPECIMEN

From 1997 to 1998 an EC project, with the acronym 'CERANORM', was designed to address four areas of concern in the formulation of new standards for advanced technical ceramics within a CEN technical committee (TC184, Working Group 3) (Morrell 1998).

This thesis only considers one of the four ceramic specimens that were used in the project and only the results pertaining to surface texture.

Eleven European companies (including the Centre for Materials Measurement and Technology at NPL) were involved in the project, but are not named here for reasons of confidentiality. The materials used were samples of commercial ceramics developed during manufacture of pump seal components. Instruments used were commercial stylus instruments and laser profilometers. NPL used a Taylor-Hobson Form Talysurf which was not considered to be a traceable instrument at the time of the EC project.

Only one of the samples could be measured using NanoSurf IV because the others had peak-to-valley heights that were larger than NanoSurf IV's vertical range (10 μm). The round robin instructions specified a total trace length of 5 mm, from which a evaluation length of 4 mm should be selected, *i.e.* five times the cut-off length (0.8 mm) in accordance with the procedure recommended in ISO 4288: 1996.

Figure 6.14 shows the Ra results for the twelve laboratories involved - Laboratories 1 and 2 are the NanoSurf IV result and the NPL result previously measured for the round robin using a Taylor-Hobson Form Talysurf. The average Ra value for all the laboratories is 0.162 μm with a standard deviation of 0.06 μm ; the NanoSurf IV result was 0.220 μm and the previous NPL result was 0.205 μm . Note that the result from laboratory 13 has not been used to calculate the average as they were the only laboratory to use a laser profilometer and their results were consistently in strong disagreement with those of the other laboratories. It is clear that, whilst the two NPL results are in good agreement with each other, they do not agree well with the average result of all the laboratories.

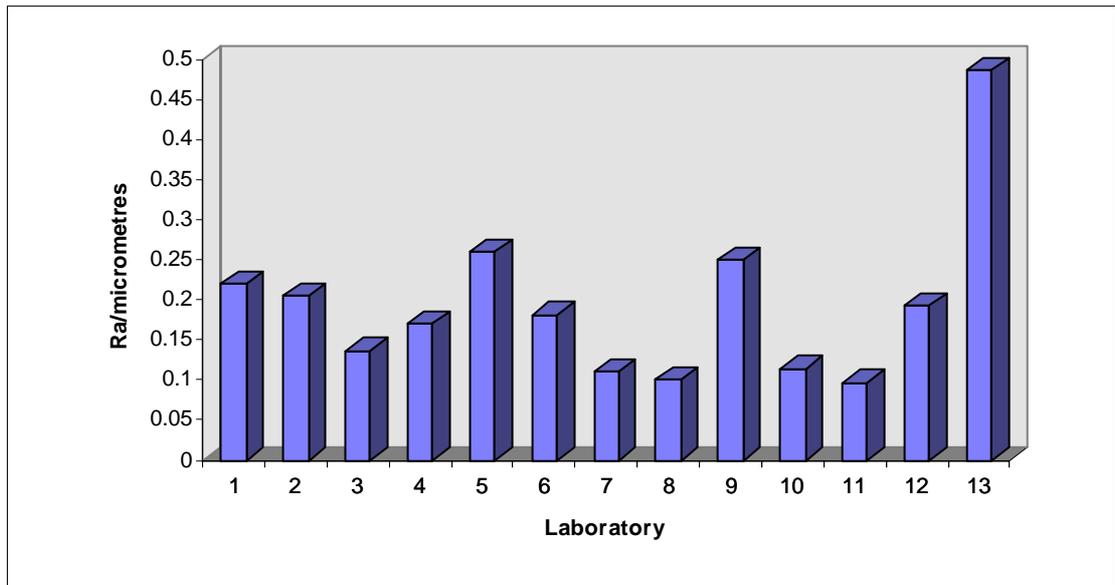


Figure 6.14 *Ra* results for all laboratories, Lab 1 is the NanoSurf IV result and Lab 2 is the previous NPL result

Figure 6.15 shows the results for the *Rt* values. The average *Rt* value for all the laboratories is 2.273 μm with a standard deviation of 0.643 μm . The difference between the two NPL results is 0.095 μm . Considering the magnitude of the *Rt* values this is very good agreement. Note also that the standard deviation of the *Rt* results for the five repeat measurements carried out at NPL was 0.593 μm .

These results show that the two NPL results agree very well, but that there is considerable disagreement from laboratory to laboratory. Feedback following the CERANORM project suggested that, as all the laboratories with the exception of NPL, were using commercial equipment in industrial environments, the NPL result should be taken as the reference result. This is backed up by the NanoSurf IV results.

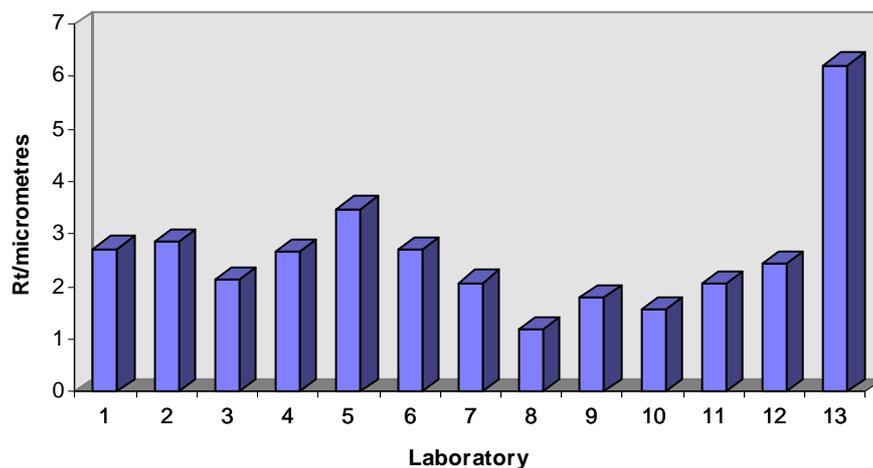


Figure 6.15 *Rt* results for all laboratories, Lab 1 is the NanoSurf IV result and Lab 2 is the previous NPL result

6.7 COMPARISON WITH THE NPL METROLOGICAL AFM

A described in §1.5.4.2, NPL has developed a Metrological AFM (Leach, Haycocks, *et. al.* 2000). One of the 8 μm period gratings, described in §6.5, was measured using the Metrological AFM (MAFM) and NanoSurf IV (in fact one of the gratings had to be cut down to a size that could be mounted on the MAFM). The amplitude results for the two instruments were 159.0 nm and 160.5 nm for the NanoSurf IV and the MAFM respectively. The differences are well within the uncertainties of the two instruments and are extremely encouraging. The periods measured were 7930.0 nm on the MAFM and 7928.8 nm on NanoSurf IV. Once again, these results show the NanoSurf IV instrument is operating correctly.

The following chapter presents the conclusions of this work and discusses future research.

6.7 REFERENCES

Note that any standards mentioned in this chapter are referenced in Appendix A.

Baher H 1990 *Analogue and digital signal processing* (John Wiley & Sons)

Cook A H 1972 Quantum metrology - standards of measurement based on atomic and quantum phenomena *Rep. Prog. Phys.* **35** 463-528

Cox M, Harris P 1997 Length 6: Surface texture data analysis *Internal NPL report ISE/C03/31A/1997-01-24/01* 1-34

Hillman I W 1993 Calibration of contact stylus instruments within the Deutscher Kalibrierdienst (DKD = German Calibration Service) *EUROMET Workshop Traceable Measurement of Surface Texture* **289** 23-30

Leach R K, Haycocks J A, Jackson K, Lewis A J, Oldfield S, Yacoot A 2000 Advances in traceable nanometrology at the National Physical Laboratory *euspen Warwick/Tokyo Nanotechnology Symp.* to be published

Morrell R 1998 SMT Project "CERANORM" SMT4-CT96-2078 Mechanical tests for advanced technical ceramics *NPL Report CMMT(D)169* 1-64

Nunn J W 2000 Calibration of 2 dimensional magnification standards for SPMs and SEMs through optical diffraction: Method, traceability and uncertainties *to be published*

Stout K J, Sullivan P J, Dong W P, Mainsah E, Luo N, Mathia T, Zahouani 1993 *The development of methods for the characterisation of roughness in three dimensions* EUR 15178 EN (Commission of the European Communities)

Teague E C 1978 Uncertainties in calibrating a stylus type surface texture measuring instruments with an interferometrically measured step *Metrologia* **14** 38-44

Thywissen J H, Johnson K S, Dekker N H, Chu A P, Prentiss M 1998 Using neutral atoms and standing light waves to form a calibration artefact for length metrology *J. Vac. Sci. Technol.* **B16** 3841-3844

Whitehouse D J 1994 *Handbook of Surface Metrology* (Institute of Physics Publishing)

CHAPTER 7

CONCLUSIONS

“Seen it all, done it all, can't remember most of it.”

Car bumper sticker

7.1 CONCLUSIONS

The body of work described in this thesis has contributed to the achievement of traceability to national standards of surface texture measuring instruments. NanoSurf IV determines surface texture by measuring the displacement of the tip of a diamond stylus as it is traced across the surface. Traceability is obtained by laser interferometer based displacement measuring systems that have been calibrated directly against a primary iodine-stabilised laser. NanoSurf IV is used to measure transfer standards that can, in turn, be used to verify other surface texture measuring instruments. NanoSurf IV has a vertical range of 10 μm and a horizontal range of 100 mm.

The instrument was specified such that it should be capable of measuring displacement in two axes with an uncertainty (at 95% confidence) of less than a nanometre. This has almost been achieved - the uncertainty in displacement is around 1.3 nm. However, the instrument was designed to operate in a thermally controlled environment despite the rigorous design of the thermal loops and consideration of materials on the instrument. The support for z axis interferometer is only a prototype - a new support structure is

being designed and manufacture is expected to start soon. The new structure is made from Zerodur and will be either machined monolithically or simply joined together with epoxy resin and will have tighter spring-loading to the Invar pillars. The instrument will soon be housed in a temperature-controlled environment ($20\text{ }^{\circ}\text{C} \pm 0.1\text{ }^{\circ}\text{C}$) in the new NPL building (the current date for the move is October 2000). These modifications to the instrument and its environment are expected to push the uncertainty below the nanometre level.

The new design of the x axis slideway proved to operate as well as the more conventional design used on the previous three NanoSurf instruments. The new slideway has the benefits of being more compact and, therefore, reducing the size of the thermal loops in a system.

The use of passive vibration isolation using simple parts such as squash balls proved to work sufficiently well for uncertainties of the order of a nanometre. If this level of uncertainty is to be reduced the instrument should be mounted on a plinth that is isolated from ground-born mechanical vibrations in the foundations. Once again, the new NPL building has been designed to provide the instrument with an isolated plinth.

It has proved almost impossible to fully characterise NanoSurf IV over its entire dynamic range. Comparisons have been made with instruments that do not have equal uncertainties with NanoSurf IV and which are not directly traceable. At the risk of sounding a bit big-headed, how do you calibrate the most accurate calibration instrument? Hopefully, this question will be answered over the next couple of years.

NanoSurf IV provides the scientific community with a platform to carry out research into the next generation of surface texture measuring instruments. It will also allow surface/probe interaction theories to be put to the test. NanoSurf IV is also a fully operational part of the NPL Measurement Services and can provide certified measurements of surface texture.

APPENDIX A

SURFACE TEXTURE STANDARDS

AGMA 118.01 (1995)	Information sheet - gear tooth surface texture for aerospace gearing (surface roughness, waviness, form and lay)
AS 2382 (1981)	Surface roughness comparison specimens
ASME B46.1 (1978)	Surface texture, (surface roughness, waviness and lay)
ASTM F 1048- (1992)	Test method for measuring the effective surface roughness of optical components by total integrated scatter
ASTM F 1438- (1997)	Test method for determining surface roughness by scanning tunneling microscopy for gas distribution system components
ASTM F 1811- (1997)	Practice for estimating the power spectral density function and related finish parameters from surface profile data
BS 1134: PT1 (1988)	Assessment of surface texture - methods and instrumentation
BS 1134: PT2 (1990)	Assessment of surface texture - guidance and general information
BS 2634: PT1 (1987)	Specification for roughness comparison specimens - specification for turned, ground, bored, milled, shaped and planed specimens
BS 2634: PT2 (1987)	Specification for roughness comparison specimens - specification for spark-eroded, shot blasted, grit-blasted and polished specimens
BS 6393 (1987)	As ISO 5436-

BS 6741: PT1 (1987)	Glossary of surface roughness terms - surface and its parameters
BS 6741: PT2 (1987)	Glossary of surface roughness terms - measurement of surface roughness parameters
BS 7900 (1998)	Specification for examination of surface texture of precision steel castings
BS ISO 3274 (1998)	As ISO 3274-
DD ENV 623 PT4 (1994)	Advanced technical ceramics - monolithic ceramics - general and textural properties - determination of surface roughness
DIN 31670 PT8 (1986)	Plain bearings: quality assurance of plain bearings: checking the geometrical tolerances and surface roughness of shafts, collars and thrust collars
DIN 31699 (1986)	Plain bearings: shafts, collars, thrust collars; geometrical tolerances and surface roughness
DIN 3969 PT1 (1991)	Surface roughness of tooth flanks; roughness parameters, surface grades
DIN 40686 SUPP2 (1983)	Surfaces of dense ceramic components for electrical engineering; determination of surface roughness
DIN 4762 (1981)	Surface roughness; terminology; surface & its parameters
DIN 4762 PT1 (1981)	Progressive ratio number values of surface roughness parameters
DIN 4766 PT1 (1981)	Surface roughness associated with types of manufacturing methods; attainable arithmetical mean value of peak-to-valley height Rz
DIN 4766 PT2 (1981)	Surface roughness associated with types of manufacturing methods; attainable arithmetical mean value Ra

DIN 4768 (1974)	Determination of values of surface roughness parameters Ra, Rz, Rmax using electrical contact (stylus) instruments; concepts and measuring conditions
DIN 4768 PT1 (1978)	Determination of surface roughness Ra, Rz, Rmax with electric stylus instruments; basic data
DIN 4772 (1979)	Electrical contact (stylus) instruments for measurement of surface roughness by profile method
DIN 4775 (1982)	Measuring of the surface roughness of workpieces; visual and tactile comparison, methods by means of contact stylus instruments
DIN 4776 (1990)	Determination of surface roughness parameters Rk, Rpk, Rvk, Mr1, Mr2 serving to decrease the material component of roughness profile
DIN 4776 SUPP1 (1990)	Measurement of surface roughness; parameters Rk, Rpk, Rvk, Mr1, Mr2 for description of material portion (profile bearing length ratio) in roughness profile; measuring conditions and evaluation procedures
DIN 54530 PT10 (1993)	Testing of paper and board cores; surface quality; determination of surface roughness and waviness
DIN V ENV 623 PT4	As DD ENV 623 PT4
ISO 1879-	Classification of instruments and devices for measurement and evaluation of the geometrical parameters of surface finish
ISO 1879- (1981)	Instruments for the measurement of surface roughness by the profile method - vocabulary
ISO 1880- (1979)	Instruments for the measurement of surface roughness by the profile method - contact (stylus) instruments of progressive profile transformation - profile recording instruments

ISO 3274- (1975)	Geometrical product specifications (GPS) - surface texture: profile method - nominal characteristics of contact (stylus) instruments
ISO 4287/1-(1984)	Surface roughness - terminology - surface and its parameters
ISO 4287/2- (1984)	Surface roughness - terminology - measurement of surface roughness parameters
ISO 4287 (1997)	Geometrical product specifications (GPS) - surface texture: profile method - terms, definitions and surface texture parameters
ISO 4288- (1983)	Geometrical product specifications (GPS) - surface texture: profile method - rules and procedures for the assessment of surface texture
ISO 468- (1972)	Surface roughness - parameters, their values and general rules for specifying requirements
ISO 5436- (1985)	Calibration specimens - stylus instruments - types, calibration and use of specimens
ISO 11562 (1996)	Geometrical product specifications (GPS) - surface texture: profile method - metrological characteristics of phase correct filters
ISO 11565 (1998)	Geometrical product specifications (GPS) - surface texture: profile method; surface having stratified functional properties
ISO 12085 (1996)	Geometrical product specifications (GPS) - surface texture: profile method - Motif parameters
JIS-B0601 (1982)	Surface roughness - definitions and designation
JIS-B0652 (1973)	Instruments for the measurement of surface roughness by the interferometric method
JIS-K7104 (1976)	Methods for comparison of surface roughness of plastics
MIL-I-45177 (1996)	Instrument, tracer, surface roughness

NAS 30 (1956)	Surface roughness designation
NAS 31 (1958)	Conversion table, surface roughness designations
SIS 81 20 05 (1973)	Concrete surfaces. Determination of surface roughness
SS 674 (1989)	Surface roughness - guidance for the choice of surface roughness
SS 675 (1989)	Surface roughness - measurement of surface roughness by means of electrical profile recording instruments
SS ENV 623-4	As DD ENV 623 PT4
SS ISO 4287-1	As ISO 4287/1-
SS ISO 4287-2	As ISO 4287/2-
SS ISO 4288 (1988)	Rules and procedures for the measurement of surface roughness using stylus instruments
SS ISO 468	As ISO 468-
UNI ISO 1879	As ISO 1879-
UNI ISO 1880	As ISO 1880-
UNI ISO 4287/1	As ISO 4287/1-
UNI ISO 4287/2	As ISO 4287/2-
UNI ISO 4288	As SS ISO 4288
UNI ISO 468	As ISO 468-
VD/VDE 2615 (1988)	Surface roughness measurement of cylindrical gears and bevel gears by means of electrical stylus-type instruments

APPENDIX B

SURFACE TEXTURE PARAMETERS

The following surface texture parameters are taken from ISO 4287: 1997 *Geometrical product specifications (GPS) - Surface texture: Profile method - Terms, definitions and surface texture parameters*.

Name & symbol	Short description
profile peak height, Z_p	distance between the x axis and the highest point of the profile peak
profile valley height, Z_v	distance between the x axis and the lowest point of the profile valley
profile element height, Z_t	sum of the height of the peak and depth of the valley of a profile element
profile element width, X_s	length of the x axis segment intersecting with the profile element
material length of profile at level c , $M(c)$	sum of the section lengths obtained, intersecting with the profile element by a line parallel to the x axis at a given level c
maximum profile peak height, R_p	largest profile peak height within a sampling length
maximum profile valley depth, R_v	largest profile valley depth within a sampling length

maximum height of profile, Rz sum of height of the largest profile peak height and the largest profile valley within a sampling length

mean height of profile elements, Rc mean value of the profile element heights Zt within a sampling length

$$Rc = \frac{1}{m \sum_{i=1}^m Zt_i}$$

total height of profile, Rt sum of the height of the largest profile peak height Zp and the largest profile valley depth Zv within the evaluation length

arithmetical mean deviation of the assessed profile, Ra arithmetic mean of the absolute ordinate values $Z(x)$ within a sampling length

$$Ra = \frac{1}{l} \int_0^l |Z(x)| dx$$

root mean square deviation of the assessed profile, Rq root mean square value of the ordinate values $Z(x)$ with the sampling length

$$Rq = \sqrt{\frac{1}{l} \int_0^l Z^2(x) dx}$$

skewness of the assessed profile, Rsk quotient of the mean cube value of the ordinate values $Z(x)$ and the cube of Rq respectively within the sampling length

$$Rsk = \frac{1}{Rq^3} \left[\frac{1}{l} \int_0^l Z^3(x) dx \right]$$

kurtosis of the assessed profile, Rku quotient of the mean quartic value of the ordinate values $Z(x)$ and the fourth power of Rq respectively within the sampling length

$$Rku = \frac{1}{Rq^4} \left[\frac{1}{l} \int_0^l Z^4(x) dx \right]$$

mean width of the profile elements, RSm mean value of the profile element widths Xs within a sampling length

$$RSm = \frac{1}{m} \sum_{i=1}^m Xs_i$$

root mean square slope of the assesses profile, $R\Delta q$ root mean square value of the ordinate slopes dZ/dX , within the sampling length

material ratio of the profile, $Rmr(c)$ ratio of the material length of the profile elements $MI(c)$ at a given level c to the evaluation length

$$Rmr(c) = \frac{MI(c)}{l}$$

profile section height difference, $R\delta c$ vertical distance between two section levels of given material ratio

$$R\delta c = C(Rmr1) - C(Rmr2); (Rmr1 < Rmr2)$$

relative material ratio, Rmr material ratio determined at a profile section $R\delta c$, related to a reference $C0$

$$Rmr = Rmr(C1)$$

where

$$C1 = C0 - R\delta c$$

$$C0 = C(Rmr0)$$

APPENDIX C

CLEANING OF OPTICS AND SAMPLES

It is important to note that some cleaning systems can physically and/ or chemically modify the specimen, for example etching, oxidation and reduction. In the Dimensional Metrology Section at NPL the usual glass cleaning liquid we use is, Decon™, which is highly alkaline and does dissolve glass very slowly (around 0.1 nm per hour). In the case of metal or metal coated surfaces we use Kodak Photo-Flow™. All water is filtered and de-ionised.

C.1 THE CLEANING PROCEDURE FOR GLASS

The cleaning sequence for the optics that make up the interferometers and for samples to be measured by NanoSurf IV is given below.

- 1) Remove all visible contamination with an appropriate means, for example, grease is removed with an organic solvent. The working surface is never physically scrubbed - the scrubbing pad may contain abrasive particles. Any visible dirt should be blown free with a filtered air line.
- 2) Soak the pre-cleaned sample, completely immersed in a 5% Decon and water mix for 20 minutes. Scrub non-working surfaces with cotton wool soaked in Decon and water mix. Keep surfaces wet at all times. Wash off with water. Place the sample in a second

Decon and water bath and zero-pressure wipe the working surface with freshly broken cotton wool, *i.e.* pull out a lump of cotton wool from the bundle, break it open and use the fresh surface keeping the cotton wool and sample very wet at all times.

- 3) From now on, the process is basically one of getting rid of the Decon. Decon is a powerful cleaner, but, almost by definition, it is a powerful surface active agent, and thus well-attached to the sample surface. Detaching it just takes time with lots of changes of water.
- 4) After the water stage, dry the sample by blowing with filtered air. The purer the water, the more successful this will be.
- 5) The dried surface will be electrostatically charged and will attract and retain dust very efficiently. Therefore, the sample should be transferred to a statically charged box, for example polypropylene.

C.2 CROSS CONTAMINATION

The cleaning sequence can readily be compromised by cross-contamination, for example, a dirty cleaning bath, dirty or dusty gloves, dust from clothes, *etc.*, falling on or transferring to the sample.

Dirt, in this context, includes any material outside the specification or the dissolution spectrum of the cleaning system. Cleaning systems and handling ware (including gloves) should be dedicated to the process and not used for other applications.

C.3 THE CLEANING PROCEDURE FOR METALS

The glass cleaning sequence is basically used for metals except the Photo-Flow wetting agent is used instead of Decon and the subsequent water washing times are much shorter since Photo-Flow does not adhere so well to metals. A total water washing time of 20 to 30 minutes is perfectly adequate.

APPENDIX D

LIST OF PUBLICATIONS

At the time of submission of this thesis the author had published the following papers and reports:

Leach R K 1995 A novel 3-axis CMM probe calibration facility *NPL Report MOM 126* 1-12

Leach R K, Hart A, Jackson K 1997 Design of two systems to measure the phase change at reflection and one system to measure the variability in contact error due to wringing *NPL Report MOT 10* 1-36

Leach R K, Hart A, Jackson K 1997 Measurement of gauge blocks by interferometry: measurement of the phase change at reflection *NPL Report MOT 11* 1-38

Leach R K 1998 Measurement of a correction for the phase change on reflection due to surface roughness *Proc. SPIE 3477* 138-151

Leach R K, Hart A, Jackson K 1999 Measurement of gauge block by interferometry: an investigation into the variability in wringing film thickness *NPL Report CLM 3* 1-17

Leach R K 1999 Measurement of the phase change on reflection in gauge block metrology *Length Bulletin 6*

Leach R K 2000 EUROMET Project 413: Inter-laboratory comparison of measurements of the phase correction in the field of gauge block interferometry *Metrologia* to be published

As a direct result of this thesis the following publications have been written:

Leach R K 1999 Calibration, traceability and uncertainty issues in surface texture metrology *NPL Report CLM 7* 1-52

Leach R K 2000 NanoSurf IV: Traceable measurement of surface texture *Xth Int. Colloquium on Surfaces* Chemnitz, Germany, Jan-Feb, 252-264

Leach R K 2000 Telling the rough from the smooth *Material World* Feb 18-19

Leach R K 2000 NanoSurf IV: Traceable measurement of surface texture at the National Physical Laboratory, UK *8th Int. Conf. Metrology and Properties of Engineering Surfaces* Huddersfield, April, to be published

Leach R K 2000 Traceable measurement of surface texture at the National Physical Laboratory using NanoSurf IV *Meas. Sci. Technol.* **11** 1162-1173

Leach R K, Haycocks J A, Jackson K, Lewis A J, Oldfield S, Yacoot A 2000 Advances in traceable nanometrology at the National Physical Laboratory *euspen Warwick/Tokyo Nanotechnology Symp.* September p112-128