Advances in Ultra-low Contact Force Nanometric Surface Metrology

Lowell Paine Howard, 24th June, 1993.

The University of Warwick,

in cooperation with the University of North Carolina at Charlotte.

Submitted to the Dept. of Engineering, University of Warwick for the Doctor of Philosophy degree

Abstract

This dissertation describes the theoretical design, practical construction and experimental use of a novel profiler intended to bridge the gap between atomic force microscopes (AFMs) and conventional stylus instruments. More specifically, it may be regarded as a hybrid instrument, combining the long-range of stylus instruments with the low contact force, high-speed operation of the AFM. The heart of the new instrument is a miniature capacitance-based force probe, constructed of glass and ceramic materials chosen primarily for thermal stability. This force probe can sense forces encompassing the range from atomic force levels (10^{-7} N) to stylus instrument levels (10^{-4} N). Probes used in subsequent studies range from ISO standard spherical diamond styli (radii 2, 5 and 10 μm) to 20 nm radius Berkovich diamond tips. A custom designed low excitation voltage, high frequency capacitance gage, used to monitor the sub-nanometer displacements of the force probe is presented. To measure surface profiles, the force probe is mounted on a PZT actuator and, much like an AFM, follows a contour of constant force under servo control.

The specimen traverses underneath the force probe on an ultra-precision kinematic slideway using a flat glass datum surface and polymeric dry bearings. A novel, inexpensive laser interferometer used to monitor specimen position and control data acquisition of the profiler is described. In this manner, profiler repeatability is enhanced to the nanometer level in two axes. Profiler performance is tested for repeatability, noise force servo bandwidth and temperature stability. A force servo response bandwidth of 300 Hz was ascertained. This compares favorably with the sub-ten Hz responses of stylus instruments.

A series of experiments designed to validate the high-speed performance of the profiler are presented. This high speed operation is some 10 to 100 times faster than conventional stylus instruments. Dynamic, non-linear interactions between the stylus tip and specimen are first derived and then examined experimentally. These dynamic interactions may eventually make it possible to measure specimen internal damping and interface stiffness or mechanical properties at the point contact level.

Acknowledgments

The Author would like to take this opportunity to express his sincere gratitude to Professor Bob Hocken at UNC Charlotte and Dr. Stuart Smith at the University of Warwick. Bob provided me with a chance to work in the field of Precision Engineering and has stood by me throughout the past four years. He showed me how exciting it could be to design and build new types of instruments. Almost every time I think I have stumbled upon something new, Bob can show me how this "new idea" has really been around for years. It was through his efforts, along with Professor Keith Bowen at Warwick, that a cooperative agreement was forged, allowing me to study internationally at both Charlotte and Warwick. Thanks Bob, I owe you my career.

It is difficult to find enough good things to say about Stuart Smith. Besides being incredibly bright, Stuart is also a genuinely good and decent human being. For his assistance and efforts, I shall be eternally grateful. After working with him over the past years, I must agree with everyone else in saying that "he is one of the smartest people I have ever met". Keep up the intensity!

Thanks also to the many people with whom I have only had the chance to work briefly with. People like Dr. Derek Chetwynd and Salam Harb at Warwick, and Drs. Jay Raja and David Trumper at UNC Charlotte. Your numerous suggestions have been appreciated. Last but not least, a special thanks to the technicians at the University of Warwick Centre for Nanotechnology: Steve, Dave, Frank and Rhod. Your careful craftsmanship and friendship was instrumental to the success of this project. May our paths continue to cross long into the future.

Table of Contents

1 INTRODUCTION

Chapter summary

This chapter discusses some of the instrumentation used for nanometric surface metrology. In the following discussion, the scanning tunnelling microscope, the scanning electron microscope and x-ray scattering methods will not be discussed for sake of brevity. Most of the instrumentation discussd herein, may, for sake of discussion, be grouped into two categories; optical and mechanical. The optical instruments generally focus their beams onto a small spot or are collimated into a broad-beam, area type measurement. Mechanical methods have evolved into two closely related camps; Stylus profilers and Atomic force microscopes. Although both utilize a sharp probe tip and rely on tip/specimen interaction forces for their operation, the instruments serve different needs and are not at this time considered interchangeable. Instrumentation for each family of instruments will be discussed with emphasis on the broader aspects of each instrument's role in surface measurement.

Nanometric surface metrology

Nanometric surface metrology is an exacting endeavor which currently represents the state of the art of modern instrument technology. Modern manufacturing dictates that surface features be measured to 0.1 nm (1×10^{-10} m) resolution over a dynamic range of many tens of micrometers. The lower limits of instrument sensitivity are often close to the physical limits of both the sensors being used to measure the surface and the traverse mechanisms and datums used to trace the profile. A large number of considerations are important in a nanometric surface instrument: Environment, choice

of materials, precision machining, electronics, optics, tribology, theory of vibration, microelectronics fabrication techniques, control theory and numerical methods. Each of these disciplines will be significant and must be assessed in instrumentation of such a high precision.

Surface topography was largely ignored for many years while efforts were directed toward more direct, first order effects of the manufacturing process. Events of the Second World War and the ensuing manufacturing advances shed some light on the importance of surface finish. Today, surface topography is of great importance in many diverse fields such as contact mechanics, heat transfer, electrical contact, tribology and orthopedics. Surface finish is of central importance in the quality control of many manufactured goods including optical coatings, microelectronics, computer disk substrates, magnetic recording tapes, automotive and aero-engine parts, hydraulics, photographic film and the printing process.

In the lexicon of surface metrology, the height of topography on a surface is referred to as the *amplitude*, probably an outgrowth of the fact that height is represented by instruments as an electrical signal, the amplitude of which is representative of the height. Spatial measurements in the plane of the test specimen are referred to as *wavelength*. All stylus instruments, whether two or three dimensional, record amplitude versus wavelength information. In the era before the advent of inexpensive microprocessors, output was recorded on chart recorder paper or smoked glass and interpretation was done manually with drafting scales and calipers.

Mass produced micro controllers have changed all that, and an ever-increasing plethora of statistical descriptors can be had at the touch of a button or click of a mouse. These parameters, a separate field in their own right, are part of the study of *computational metrology*. Computational metrology for surfaces concerns evaluating statistical parameters used to quantitatively describe the quality of a surface. Since different surface preparation methods produce differing surface topographies, there are

many such parameters available to the metrologist. The arduous task is selecting the best parameter to describe the "goodness of surface" in a specific situation. Selecting the right parameter can be a daunting situation given the "parameter rash" as described by Whitehouse [1]. Two of the most often used parameters are arithmetic or center-line average (R_a) and root-mean-square or RMS roughness (R_q). These parameters are now strictly defined in a number of national and international standards such as ISO 4287, BS 1134, ANSI B46.1 and DIN 4768.

The wavelength, or spatial measure of surface topography is broken down into two domains based on the traditional machining process used to produce surfaces; waviness and roughness. Waviness consists of spatial wavelengths generally greater than 0.8 mm. Processes which affect waviness include chatter from the tool/workpiece interaction, warpage of the workpiece from clamping and fixturing forces, machine slideway errors and residual stresses from the machining process. Processes affecting roughness include spindle speeds, lubricants and removal rates for traditional machining and finishing operations such as bead blasting, lapping and polishing. According to Reason[2], surface metrology was pursued from the short (roughness) end of the wavelength spectrum to satisfy an urgent need for this information during the second world war. He states that, were it not for the impetus of the war, surface topography may have been pursued from the longer (waviness) end of the spectrum which is more difficult to measure.

In practice, the roughness and waviness components of the total profile are separated by filtering. Such filtering is specified under many standards, ISO 4287 being the most universal. Gradually, analog, electronic filtering of the transducer signal (the 2RC filter for example) is being phased-out in favor of more flexible digital filtering. Since almost all modern stylus instruments employ a microprocessor, different filter algorithms are easily implemented.

Optical surface metrology methods

Given the rather close relationship of the study of surface topology to the manufacturing process, it is not surprising that in recent years, surface metrology instrumentation has been developed and marketed with speed of operation in mind. This has meant, traditionally, the use of optical methods for determining surface topography. Optical techniques generally may be classified into two categories: Focused and area techniques. Focused methods scan a small spot of light across the test specimen and record the resulting interactions for display and analysis on a computer system. Since a light beam has no mass, it may be rastered very quickly over a surface by means of a vibrating mirror or galvo-scanner before passing through the focusing optics. The limitation here is that a flat depth of field is extremely difficult to maintain across a large area. Circumventing this problem, the light beam may be left stationary while the test specimen is scanned on a small stage. The limitation in this case being the speed that the more massive stage and sample can be scanned.

All focused methods are subject to the Rayleigh diffraction limit as the limit to their lateral, or spatial resolution. Simply stated, the Rayleigh limit gives the smallest angular separation obtainable (image resolution) for a given lens aperture and light wavelength[3]. The minimum resolvable angular separation is

$$\Delta\theta = \frac{1.22\lambda}{2NA} \qquad (1.1)$$

where λ is the wavelength of the light and $N.A.$ is the numerical aperture of the focusing objective. Note that this equation is only an approximation and in the real world of non-perfect optical components results are always worse. The physical limitations of resolution are the difficulty of building short wavelength UV and X-ray sources and the required focusing optics.

The mere existence of a rule such as the diffraction limit (or any other rule for that matter) means that those who are dedicated and resourceful will seek to find ways

around it. Breaking the law in this case is accomplished by circumventing Rayleigh, operating in the optical near-field where diffraction effects are obviated. The optical near field region extends just some 10 nm or so away from the surface. To illuminate and gather light at this short range, specially tapered optical fibers[4] are positioned to within nanometers of the surface by piezoelectric scanners (more on these later), see figure 1.1.

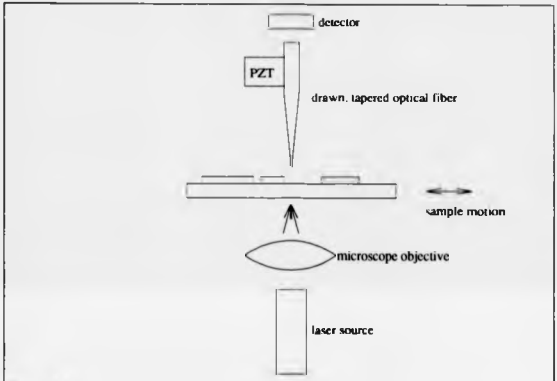


Figure 1.1. Collection mode Near-field Scanning Optical Microscopy (NSOM).

Lateral resolution is limited only by the aperture size and can be on the order of 10 nm[5]. Much more exciting work is expected from this burgeoning field.

Focused methods

While interferometric techniques and optical lever arms (techniques which shall be defined later in this chapter) may both be used with focused light, the majority of commercial instruments fall into either the focus follower or quality- of-focus niches.

The focus follower is generally a derivative of the Focault knife technique[6] used in compact disc audio players and CD ROM drives. This method uses a voice coil motor to profile a contour of constant focus on the specimen.

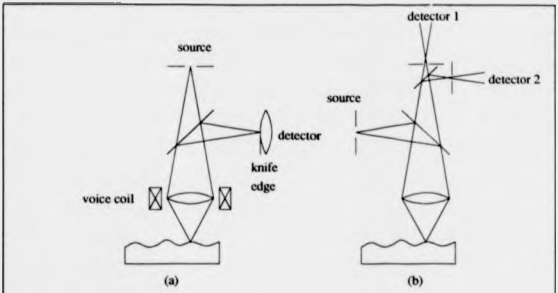


Figure 1.2. Focus follower (a) similar to that used in CD player, and quality of focus probe (b).

The quality-of-focus technique is slightly different in that the objective lens remains stationary throughout the scan. In the so-called Simon probe[7], the amount of light incident on a pair of apertured photodiodes is compared, with one of the photodiodes mounted slightly ahead of, and the other slightly behind the focal plane. As the specimen is moved into and out of focus, the beam waist of the light inside the microscope moves a corresponding distance. By comparing the size of the cones of light on either side of the focal plane, a differential measurement may be ascertained. Simpson[8] pointed out that even a simple version of this probe could be useful at the nanometer level. Further refinements are found in the literature[9] of the Simon technique, but all appear to be too dependent on surface roughness to be of commercial value. An interesting variation of the technique is presented by Kohno[10]

using a critical angle prism to produce a very compact sensor with claimed nanometer resolution and sub-micron spot size.

All of the focused techniques suffer from a number of maladies (other than the diffraction limit) including poor performance on high aspect ratio surfaces and regions of differing reflectivities or indicies of refraction. Both of the former are routinely encountered on integrated circuits. Focus methods can measure both roughness and waviness depending on the travel of the displacement mechanism used.

A variation of the focused methods has been described by Ennos and Virdee[11]. Essentially and optical lever arm, the technique measures the slope of a surface using a beam of focused laser light. The total profile is obtained by integrating the slope information obtained by the instrument.

Area methods

If the optical beam is collimated and expanded for use in an area technique, large areas of the testpiece may be measured with no physical rastering of the either the optical probe or specimen. Area techniques fall into two general categories, and may be used for the determination of roughness and form errors. In general, large aperture interferometers are used for form measurement and microscope interferometers and scattered light techniques are used for roughness estimation. Of the large aperture interferometers, the most commonly used are the simple Fizeau interferometer and the phase measuring interferometer; usually a Fizeau with the addition of a computer and some optical phase shifting apparatus. First, the simplest case, the Fizeau. As shown in figure 1.3 (a), this interferometer consists of only a monochromatic light source and an optical flat.

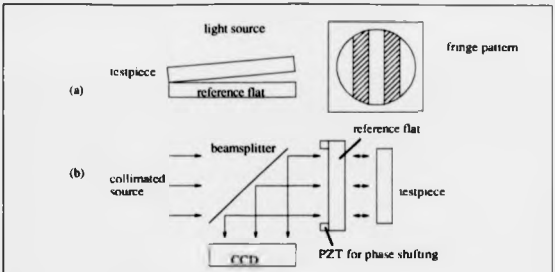


Figure 1.3. Simple Fizeau interferometer (a) and Phase measuring Fizeau (b).

Placing the test-piece on the flat invariably results in some sort of small wedge angle being formed between the two pieces due to surface dust and contamination. With the eye of a skilled artisan, flatness may be determined down to about $\lambda/4$ simply by observing the fringes at an oblique angle and estimating straightness. More sophisticated variations of this technique include capturing the fringes on photographic film for analysis with mechanical drafting instruments or, more recently, capture of the fringe pattern on a television camera for analysis by digital computer. This latter method has largely been superseded by the modern phase measuring interferometer[12].

The phase measuring interferometer, PMI or "phase stepper" as it is sometimes called, uses a CCD television camera to capture a fringe pattern for computer analysis, but adds a phase shifting device to alter the length of the optical beam path, figure 1.3 (b). Shifting the optical path difference a small, pre-determined amount by moving the reference mirror with a piezoelectric element allows one of several different computer algorithms to interpolate the phase of each pixel in the interferogram. The result of these computations is a three-dimensional map of the surface with topography being

related to phase. Speed of the measurement is quite fast with modern computers, on the order of 30 seconds. Setup time, however is considerably longer, requiring the surface to be precisely aligned to the reference flat to within a few milliradians. Vertical resolution can be as high as 0.1 nm with a dynamic range of just a few micrometers however, limiting use of the technique to polished or ground surfaces. Horizontal resolution is calculated by determining the effective aperture size and dividing by the number of pixels in one row of the CCD array. For a good 1024 X 1024 pixel array over a 100 mm aperture, this comes to approximately 100 μm.

There are drawbacks however. Since the phase information is directly translated into surface topography, surfaces with differing indices of refraction will introduce distortion into the computed topography. Polished silicon carbide is a classic example of this effect, but it is prevalent on semiconductor wafers as well, especially after a few dielectric layers have been deposited. Any surface with spatially disjoint thin films is a poor candidate for the phase measuring interferometer. Also, since the phase is ultimately computed from intensity maps of the CCD array, differing reflectivities on the surface as well as nonuniformities of the CCD detectors will cause similar deleterious effects. There has been some recent documentation of non-repeatabilities in these instruments when measuring flat surfaces[13] which the Author has reproduced in the laboratory. These anomalous results are obtained any time the interferometer is operated away from the null fringe condition.

Scattered light techniques use the diffracting properties of a surface with regular features. Such surfaces are commonly produced from most milling and turning processes in the workshop. Scattered light systems can provide good roughness data in only a few milliseconds and are particularly well suited for *in situ* inspection in a production environment. The pattern of a diffracted optical beam is analyzed by a linear photodetector array which is then digitized and analyzed in a computer, figure 1.4.

related to phase. Speed of the measurement is quite fast with modern computers, on the order of 30 seconds. Setup time, however is considerably longer, requiring the surface to be precisely aligned to the reference flat to within a few milliradians. Vertical resolution can be as high as 0.1 nm with a dynamic range of just a few micrometers however, limiting use of the technique to polished or ground surfaces. Horizontal resolution is calculated by determining the effective aperture size and dividing by the number of pixels in one row of the CCD array. For a good 1024 X 1024 pixel array over a 100 mm aperture, this comes to approximately 100 μm.

There are drawbacks however. Since the phase information is directly translated into surface topography, surfaces with differing indices of refraction will introduce distortion into the computed topography. Polished silicon carbide is a classic example of this effect, but it is prevalent on semiconductor wafers as well, especially after a few dielectric layers have been deposited. Any surface with spatially disjoint thin films is a poor candidate for the phase measuring interferometer. Also, since the phase is ultimately computed from intensity maps of the CCD array, differing reflectivities on the surface as well as nonuniformities of the CCD detectors will cause similar deleterious effects. There has been some recent documentation of non-repeatabilities in these instruments when measuring flat surfaces[13] which the Author has reproduced in the laboratory. These anomalous results are obtained any time the interferometer is operated away from the null fringe condition.

Scattered light techniques use the diffracting properties of a surface with regular features. Such surfaces are commonly produced from most milling and turning processes in the workshop. Scattered light systems can provide good roughness data in only a few milliseconds and are particularly well suited for *in situ* inspection in a production environment. The pattern of a diffracted optical beam is analyzed by a linear photodetector array which is then digitized and analyzed in a computer, figure 1.4.

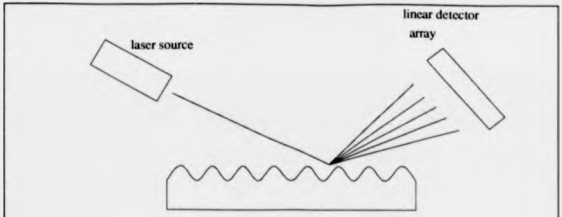


Figure 1.4 Scattered light roughness apparatus.

Problems arise when assessing surfaces with features on the order of the wavelength of the light being used[14].

Random surfaces such as paint finish are evaluated by use of the glossmeter. The glossmeter is often used on smooth, shiny surfaces such as paint, and give an indication as to how a surface *looks*. Glossmeters work by illuminating the surface with a collimated laser beam, trapping the specular reflection in a light trap and measuring the amount of total scattered light inside an integrating sphere. A very high gloss surface will produce little scattered light and a large specular reflection.

Discussion

None of the optical techniques, however can give a representation of the true mechanical surface. The time-tested metrologists' axiom of matching the probe to the test-piece rings true. When testing optical components, try to use an optical method. When testing mechanical parts, observe the mechanical surface. Furthermore, many parts with spatially disjoint dielectric coatings such as integrated circuits and flat panel displays are impossible to measure non-destructively with surface figure interferometers. For reliable readings, such parts must be coated with a rather thick

layer of vacuum deposited metal, such as aluminum or gold. Also, optical techniques work best when complemented by mechanical methods for periodic proving. *Redundant metrology* is always useful to compare one method against another, using mechanical methods as a "reality check". The builders of the Hubble Space Telescope learned this truth the hard way, relying on only one optical test[15]. Had they had the opportunity to check their mirror with a mechanical method, they would have discovered that it was 2 μm out of form[16] [17]. Further references to optical surface metrology techniques may be found in a thorough, but now slightly-dated, review of the subject by Vorburger and Teague [18], and in a more recent review by Sherrington and Smith[19].

Mechanical contact techniques

Mechanical contact techniques are the only methods to determine the nature of the true mechanical surface. They have endeared themselves to metrologists over the years because of their ability to "plow" through dust and surface contaminant layers and provides a readily calibrated quantitative output.

Nomenclature

It is useful, when discussing instrumentation, to define a strict set of terms which comprise all the different functional parts of the instrument. The author is particularly enamored with the nomenclature scheme proposed by Teague[20] [21]. Using this nomenclature, the basic elements of precision instruments may be defined and then grouped together to model the various instrument configurations. The basic tasks of a dimensional metrology instrument are:

From Teague

- "Realizing a metric in the measurement space,

- Establishing the metric in a coordinate reference frame,
- Generating and measuring repeatable motion relative to the coordinate reference frame, and
- Linking the testpiece to the coordinate system by a probe."

Realizing a metric means, of course, building a displacement sensor to measure displacement, a force sensor to measure force, etc. This sensor must then be fixed to a frame of reference with respect to a datum and repeatable motion produced to either the metric or the testpiece within the reference frame. Finally, a probe must be used to transfer the measurand from the testpiece to the metric.

Stylus Instruments.

Using Teague's guidelines, we may establish the four basic components of any stylus instrument:

1. A relative displacement transducer such as a LVDT, LVDI or capacitance gage.
2. A datum surface of known character which may be a separate entity, or defined by,
3. the traverse mechanism. Often traces a linear or arcuate path (2D), but may be three dimensional as well.
4. A diamond stylus probe; 2, 5 or 10 μm spherical tip radii or a chiseled or cleaved diamond point with a tip radii of some tens of nanometers.

Shown in figure 1.15 is an archtypical stylus instrument, whose component symbols will be used throughout the following discussion.

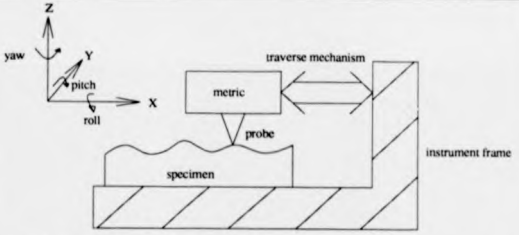


Figure 1.5. Elements describing stylus profiler instrumentation with a reference coordinate system.

Traverse mechanisms and defining the datum.

Using the basic elements described previously and in figure 1.5, most known embodiments of the stylus profiler can be developed from first principles. By defining the measurement loop as the path connecting the probe and metric through the specimen to the instrument frame, each profiler embodiment may be analyzed. A basic profiler configuration which consists of translating the metric and probe across the specimen is illustrated in figure 1.6.

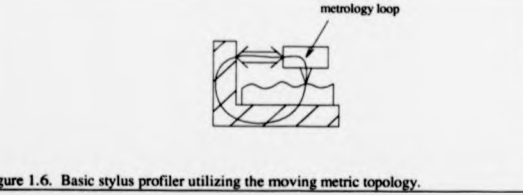


Figure 1.6. Basic stylus profiler utilizing the moving metric topology.

As shown in the figure, the measurement loop passes through the traverse mechanism, forcing it to also define the datum. By using the traverse mechanism as the datum, the measurement is susceptible to noise and errors of straightness in the traverse mechanism. To further elucidate: Consider the case of a ball-slide mechanism typically used to translate optical elements in the laboratory. The ways of such carriage/slide systems are typically precision ground stainless steel rods fixtured into an extruded aluminum base by retaining clips. Any bending of these rods forces the carriage onto a non-rectilinear path, thus causing the datum to be something other than a straight line. This type of error manifests itself as a form error in the final measurement. Noise from the numerous ball bearings will contribute to errors in the roughness portion of the measurement. Our typical ball slide contains many more ball bearings than are needed for strict compliance with kinematic considerations (three). As a result, the bearings will always be competing with one another, constantly redefining the kinematics of constraint, as balls come into, and roll out of, contact. Also, these ball bearings are not perfectly round. As they roll, they will cause the carriage to rise and pitch, undulating several tens of nanometers (if you have very good bearings) with each revolution.

We immediately learn from the forgoing illustration that the traverse mechanism for the moving metric type of profiler must be of uncanny precision. To measure surface form errors and finish in the range below ten nanometers, it is necessary to have an inherent instrumental error of, at most, one nanometer. Elastic flexure mechanisms are one way of providing such a super-precision traverse mechanism. In work pioneered by Jones[22], flexures of several configurations have been found useful for stylus profilers. Historically, flexures have proved very cost effective for stylus instruments for application in the thin-films industry. Thin-film processes are usually calibrated by step height measurements. In the step height measurement, the stylus need not trace a perfectly straight, linear path. In fact, almost any repeatable, nearly-linear path will do. A large radius, arcuate path is simple to produce, requiring only one pivot, figure 1.7.

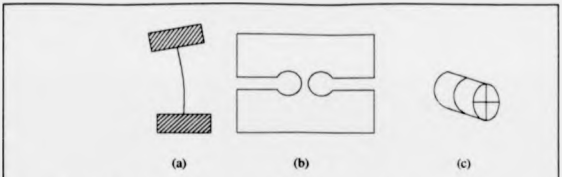


Figure 1.7 Leaf spring (a), monolithic (b) and torsional (c) flexure pivots for arcuate motion of the metric.

The unitized flexure pivot, commonly called a "Bendix" pivot (in the USA) after one manufacturer[23], is simply fitted into round holes drilled into the pivot arm and the instrument base. Though inexpensive to manufacture, the drawback of this type of pivot is that its stiffness in the axial direction is often not as high as that of its cousin, the monolithic notch flexure. Axial stiffness is a key parameter since any sagging or vibration of this axis will manifest itself directly in the measurement, as the errors of the ball-slide did in the earlier example. The notch flexure is realized by drilling two holes in close proximity, as illustrated in figure 1.7 (b). Leaf spring pivots are extremely easy to manufacture using bolted joints and BeCu strips as hinges. Large radius arcs of ten millimeters are possible with flexure pivots. Use of these flexures will figure prominently later on in some case studies of stylus instruments.

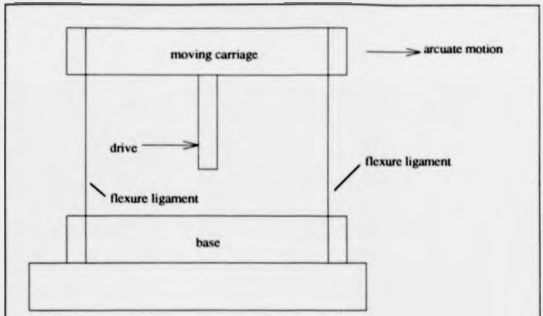


Figure 1.8. Compound spring flexure for tracing a nearly linear path.

For situations where nearly a more linear tracing motion is called for, a compound flexure may be applied, figure 1.8. This variation of the four-bar mechanism produces motion which deviates from the linear by approximately

$$\delta_{vertical} \approx \frac{x^2}{2l} \tag{1.2}$$

where x is the distance moved and l is the length of the flexure ligaments. For more detailed information on flexures, see Smith and Chetwynd[24].

To relax the requirements of the traverse system, a skid may be used near the stylus to establish a local datum with respect to the surface, figure 1.9. With a skid, measurements can even be made from a hand-held instrument with a very simple traverse mechanism.

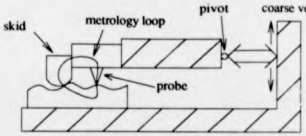


Figure 1.9 Skid type profiler showing localized metrology loop. Note that the skid radius is much greater than the stylus radius to obtain an area average of the datum.

Skids come in many geometries, but the simplest is shown in the figure. By supporting the measurement arm on a pivot (a five degree of freedom coupling), the measurement loop is shortened and no longer includes the traverse mechanism. The advantages are thus threefold:

1. The measurement loop is much shorter, circumventing the instrument frame and reducing external vibration pickup.

2. The traverse mechanism may be much simpler and less expensive to manufacture.

3. Coarse vertical positioning is simplified and specimen leveling is virtually eliminated.

The datum is now defined by the locus of points describing the path of the skid over the surface, form information is lost since the skid/probe system constitutes a mechanical high pass filter. Spatial cutoff of this filter is defined by the distance between the skid and the probe in this example. Different skid geometries exist, each defining a different datum and each being useful for measuring particular types of surface topography. In general, if λ is the spatial wavelength, R_{max} is the peak to valley depth and p is the maximum allowable percentage error for phase considerations, skid radius[25] must not be less than

$$100\lambda^2 / 8pR_{max} \quad (1.3)$$

Keeping this in mind, the disadvantages of skid systems are:

1. Some *a priori* knowledge of the surface topography is useful to select a skid because spatial aliasing and mis-sampling are sometimes problematic.

2. The skid has the potential to scratch and damage many surfaces.

3. Most importantly, form or waviness information is lost.

A skid system may be adapted for external datum operation by removing the skid and resting the moving arm on an external datum plate, figure 1.10.

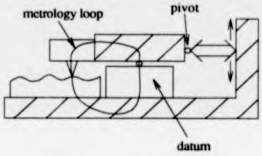


Figure 1.10. Skid type system adapted for external datum use. Moving metric external datum topology.

This single point contact to the datum, together with the hinge connecting the moving arm to the traverse mechanism defines the datum line. While convenient, the system does have several drawbacks. The datum surface must be leveled to so that the metric remains within its operational limits during a measurement cycle. For nanometric measurements, the stiffness of the measurement loop leaves something to be desired. On most systems the moving arm and pivot are of lightweight construction leaving the

metric on the end of a fairly "floppy" cantilever, thus making the instrument susceptible to vibration.

Datum separation from the traverse mechanism may also be accomplished by the configuration shown in figure 1.11.

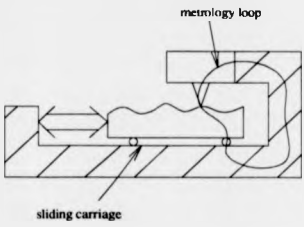


Figure 1.11. Moving specimen external datum topology.

This moving specimen topology eliminates the loop stiffness problems of the moving metric design shown in figure 1.10. The metric may be solidly connected to the instrument frame, through a coarse positioning mechanism. The traverse mechanism drives a kinematically located carriage through a single-degree-of-freedom coupling. More details on the design of these elements will follow. Note that in theory the only factor limiting the horizontal traverse of such a design is the size of the datum plate and the length of travel of the traverse mechanism. This profiler topology is unique in that form errors in the datum plate are often transferred to the measurement as second order effects[24]. This can be illustrated through the 2D model of figure 1.12.

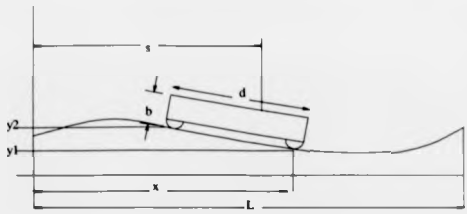


Figure 1.12. Kinematic slideway errors in the vertical (amplitude) axis.

$$\delta_y \approx (x-s)\left[\frac{f(x)-f(x-d)}{d}\right]+b\left(1+\frac{1}{2}\left[\frac{f(x)-f(x-d)}{d}\right]^2\right) \qquad (1.4)$$

Where $f(x)$ is the function yielding the surface of the optical flat, d is bearing pad separation, s is the measurement point, x is the distance between the far bearing pad from the datum and b is the carriage thickness. Since we may select an optical flat with a nominal form error of about 10 nm across 100 mm and bearing pads separated 75 mm apart, datum errors are a *non sequitur* when compared with other potential error sources. Note that this analysis assumes a surface whose undulation wavelength is much greater than the bearing pad separation. If the undulation wavelength of the datum is close to the bearing pad separation distance, an aliasing condition can occur and give false results. True straightness measurements generally require an external datum and several measurement points[27]

Choice of bearings for kinematic support is critical for nanometer applications. Many bearing materials exist and will be discussed in more detail in Chapter 5.

At least two other profiler topologies exist, but are not in common use in commercial instruments. The use of monolithic flexure stages to move the specimen around under

a fixed metric and the use of a differential metric in a metrology frame configuration. A good example of the former is the "Paraflex" stage and profiler by Teague et. al[28]. this 3D profiler used a monolithic 2D flexure stage position the specimen underneath a Talystep head. The latter, metrology frame topology was embodied in a large profiler type measuring machine[29] built by Cranfield Precision Engineering (CUPE), figure 1.13.

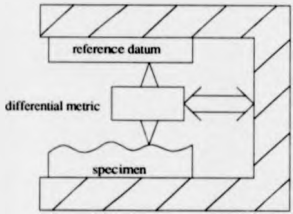


Figure 1.13. Metrology frame concept used in stylus profiler.

The metric in this case being a laser interferometer (more on these later) measuring the differential displacement of the probe with respect to the datum. In a differential measurement such as this one, errors of the traverse mechanism do not effect the profile measurement.

Realizing the metric: Displacement transducers.

It is useful, when comparing transducer systems used in stylus instruments to define a figure of merit. This figure of merit may then be used as a measure of "goodness" when comparing the relative merits and disadvantages of the systems. An ideal transducer would detect stylus displacements smaller than 0.1 nm over a range of many millimeters. Its measurements would be so fast that the speed of traverse could

increase to the point where measurement cycle times become short enough to remove instrumental drift and slowly changing environmental variations. Finally, the mass of such a transducer would be small enough so that the tracking ability of the stylus arm/spring system is not reduced.

With this in mind, the figure of merit is defined as the dynamic range bandwidth product divided by the square root of the added effective mass m_e. So,

$$F.O.M. = \frac{DR \cdot BW^{\frac{1}{2}}}{m_e^{\frac{1}{2}}} \tag{1.5}$$

where dynamic range root bandwidth product is defined by [30]

$$DR \cdot BW^{\frac{1}{2}} = \frac{\Delta x_{max}}{NED} \Delta\omega^{\frac{1}{2}} \tag{1.6}$$

given $\Delta\ x_{max}$ is the maximum sensor displacement, *NED* is the noise equivalent displacement at a bandwidth of $\Delta\ \omega$. NED is defined as the equivalent root-mean-squared voltage or current from the transducer multiplied by whatever scale factor is necessary to convert the voltage or current value to a calibrated displacement value.

The figure of merit provides a mass penalty for heavy transducers which lower the resonant frequency of the stylus tracking system, reducing speed and measurement fidelity (hence the inclusion of the effective mass term). Precise assessment of the figure of merit is often difficult because determining the actual effective mass can be a somewhat nebulous process unless the actual transducer is available for inspection. Nevertheless, the concept of the figure of merit is useful to consider during the initial design process of any nano displacement transducer.

AC bridge transducers

Of the methods used by modern commercial stylus instruments to sense probe displacements, most fall under the category of alternating current bridge methods. AC

bridge methods measure the effects of inductors and capacitors on the magnitude and/or phase of an alternating excitation current. Selecting a detection method for measuring these changes is very important and shall be the subject of this digression into electrical engineering. In the following discussion, detection methods are categorized into two types: Synchronous and asynchronous detection.

In the simplest case, synchronous detection involves the use of an analog multiplier circuit. The two inputs to the multiplier are the bridge signal $e_1(t) = E_1 \sin(2\pi f_1 t + \phi_1)$ and the reference signal $e_2(t) = E_2 \sin(2\pi f_2 t + \phi_2)$ from the excitation oscillator. The product of these two signals is then

$$e_3(t) = \frac{E_1 E_2}{2}\cos(2\pi(f_1 - f_2)t + (\phi_1 - \phi_2)) + \frac{E_1 E_2}{2}\cos(2\pi(f_1 + f_2)t + (\phi_1 + \phi_2)) \tag{1.7}$$

where there are sum and difference terms for frequency and phase. If the frequencies f_1 and f_2 are different, then the operation is known as *heterodyning* and the multiplier is called a *mixer*. If f_1 equals f_2, i.e. they are from derived from the same excitation oscillator, the operation is known as *homodyning* and the multiplier is called a *phase detector*. If the phase detector is followed by a low pass filter to remove the contribution of the summed frequency component, the output reduces to

$$e = \frac{E_1 E_2}{2}\cos(\phi_1 - \phi_2) \tag{1.8}$$

where we are measuring the DC level of a signal which has been frequency shifted down to baseband (DC) by the phase detector. In AC bridge methods, the phase detector is most often encountered as a piece of laboratory equipment called a lock-in amplifier. The "lock-in" contains an excitation source, phase shifter, input amplifiers and filters, one or more phase detectors and low pass filters, figure 1.14.

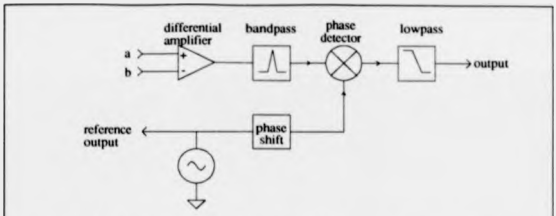


Figure 1.14. A lock-in amplifier used for demodulating stylus transducer signals.

As one might expect, setting the phase of the reference signal ϕ_2 is of critical importance and will be discussed in more detail along with the design of lock-in circuits in a later chapter.

Having thus discussed synchronous detection, we now examine asynchronous detection methods. A full-wave bridge rectifier followed by a low pass-filter is essentially an asynchronous detector, as is an AC voltmeter. Although somewhat cheaper to implement, the asynchronous method is not capable of measuring the phase shift induced by a bridge component, although we shall see that in some applications this may be used to an advantage.

Three transducers are commonly in used by modern stylus instruments: The linear variable differential transformer or LVDT, the linear variable differential inductor or LVDI and the capacitance gage. An LVDT is depicted in figure 1.15.

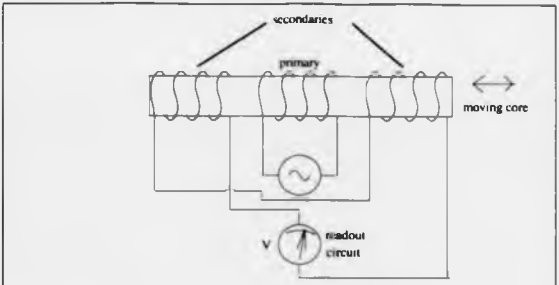


Figure 1.15. A LVDT transducer.

Operating on the principle of variable mutual inductance between the primary and secondary windings, LVDTs are available with maximum stroke lengths of some 200 mm and can have redundant windings in case of failure. Movement of the transformer core changes the magnetic flux linkage ratios of the two secondary windings. The LVDTs used in electronic indicating gages are most often used not with a synchronous demodulator, but with an asynchronous detector. Stray capacitance from cables and winding asymmetries can cause phase shifts in the bridge signal which are difficult to null out (without user intervention) when using interchangeable gage heads. Typical excitation frequencies are on the order of 10 kHz. Because of these high excitation frequencies and the relatively large mass of the moving ferrite core, magnetic forces on the moving core resulting from the excitation signal have a negligable effect on the sensor. To obtain a DC output, the sensor signal may be rectified and filtered to remove the carrier. This asynchronous detection is a convenient and cost effective alternative to phase sensitive demodulators, albeit at the expense of lost resolution. A slightly better detection scheme is depicted in figure 1.16.

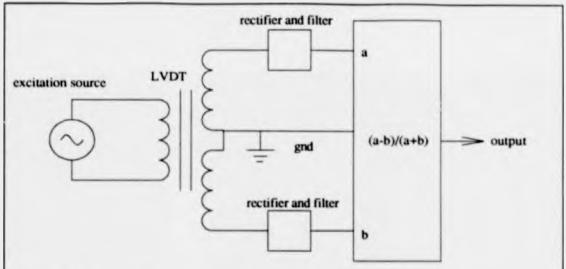


Figure 1.16. Improved asynchronous detection method for LVDTs.

This method, which may be implemented on a single integrated circuit (Analog Devices AD598 is one example) has the advantage of compensating for variations in the amplitude of the excitation oscillator. Each coil signal is rectified and filtered, then input to an analog computation network. No phase adjustments are necessary and accuracy may be around 0.5% full-scale. For maximum performance in instruments, LVDT resolution may be extended by amplifying the drive signal to levels near to core saturation and demodulating the signal with a lock-in amplifier. This method, which must have a carefully adjusted phase shift, can extend the resolution of the LVDT to approximately 0.1 nm with a bandwidth of less than 10 Hz[31].

The LVDI differs from the LVDT in that it is a variable reluctance, self inductance bridge transducer, figure 1.17 (a). Consider the magnetic circuit shown in figure 1.17 (b).

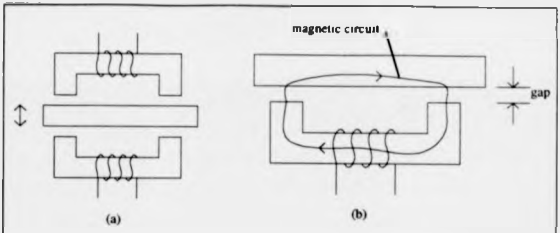


Figure 1.17. LVDI (a) and magnetic circuit for reluctance calculations (b).

The inductance of such a system is defined as

$$L = \frac{\Psi N}{I} \tag{1.9}$$

where Ψ is magnetic flux, N is the number of coil windings and I is the coil current. Recalling that the flux Ψ is the product of N and I divided by the sum of the reluctance of the magnetic circuit,

$$\Psi = \frac{NI}{\Re_1 + \Re_2 + \Re_3 + \Re_4} \tag{1.10}$$

where reluctance

$$\Re = \frac{\ell}{\mu A} \tag{1.11}$$

ℓ is defined as flux path length, μ is the magnetic permeability and A is the cross sectional area of the magnetic circuit. Since the reluctances of the flux paths through the "U" shaped core and the platen are fixed, the total inductance of the coil is dependent on the reluctances of the air gaps. Thus, inductance is inversely proportional to the gap distance. To linearize the output signal, the inductor is used in the differential mode as a matched pair, as depicted in figure 1.17 (a). If the air-gap

separation is allowed to vary a small amount δ, the differential sensor output will be proportional to

$$\frac{1}{\ell+\delta}-\frac{1}{\ell-\delta}=\frac{-2\delta}{\ell^2-\delta^2}\cong -2\delta, \delta << \ell \qquad (1.12)$$

for small changes in gap spacing. The figure-of-merit for the LVDI suffers from its bulk and relatively large mass. Also, since the platen moves between two fixed coils, some type of crash stop mechanism is needed to prevent the platen mechanism from catastrophic failure in the event of motion exceeding normal limits. More will be said about this when the Talystep is reviewed. At least two other possible configurations for LVDIs are possible, shown in figure 1.18.

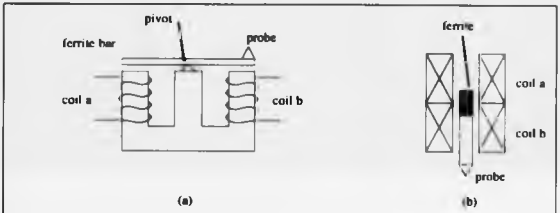


Figure 1.18. E core LVDI (a) and in-line coils (b) designs.

One other form of inductive transducer is used in stylus instruments. The simple inductor bridge using a single inductive pickup coil. Illustrated in figure 1.19, the inductive pickup is analyzed by first assuming a uniform flux path throughout the pickup coil.

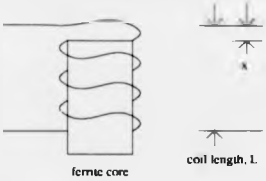


Figure 1.19 Single coil inductive pickup.

Reluctance of the magnetic circuit within this region is

$$\Re = \frac{L-x}{\mu_o A} + \frac{x}{\mu_1 A} \tag{1.13}$$

with the inductance then

$$L = \frac{N^2}{\frac{L-x}{\mu_o A} + \frac{x}{\mu_1 A}} \tag{1.14}$$

If $\mu_1 >> \mu_o$, as in the case of a ferrite core and L>>x, the output is linear for small x. Given a coil length of 10 mm, the output is linear to approximately 1% for displacements up to 100 μm using the previous relation.

All of the transducers described thus far have been relatively large, on the order of one cm^3 volume. The mass of the moving core has been relatively large, adding to the mass of the stylus pickup arm. In an effort to reduce the mass caused by the core material, capacitance gaging circuits have been utilized in recent years. Capacitance gaging will be discussed in greater detail in Chapter 4. On a basic level, capacitance sensors fall into two categories: Variable gap and variable area. From the basic relation for the parallel plate capacitor

$$C = \frac{\varepsilon A}{d} \qquad (1.15)$$

where ε is the permittivity of the dielectric medium (usually air), A is the area of electrode plate overlap and d is the gap between the plates. A typical noise equivalent displacement for a precision capacitance gage is approximately 1 picometer per root Hz (pm $Hz^{-1/2}$).

Stylus probe selection.

Selection of the stylus probe will depend in large part on what type of specimen is being measured and what information is being sought. Probe contact mechanics dictate the size of the contact patch with the specimen. Contact mechanics will be discussed in greater detail in Chapter 2. In general, information about the optical scattering properties of a surface to visible light are contained in spatial wavelengths around to 1μm and shorter. In these cases, a cleaved diamond stylus with the smallest radius possible is often used. (As an aside, these cleaved diamond probes are incredibly durable, maintaining a stable, sharp geometry for decades with careful use[32]). Many tribological and mechanical properties are defined at somewhat longer wavelengths and are more easily measured by the larger, ISO standard probes of micrometer order radii. In the microelectronics industry, probe geometries are driven by somewhat different requirements, primarily line width measurements and step height characterization. As process technology shrinks line widths to one micrometer and below, stylus probes must shrink as well to reduce probe convolution effects. Indeed, the spherical tipped stylus is even giving way to "hammer head" shaped silicon probes, ion milled specifically for measuring sub-micron line widths[33]. The time-tested metrologist's credo of matching the probe to the artifact is continuing to prove true.

Case histories

Instruments will be evaluated on the basis of the following points:

1. Transducer design and sensitivity.

2. Traverse mechanism and datum.

3. Probe choices available and contact force.

4. Ergonomics. Does the implementation of the design have features which make it easy to use?

5. Economics. Does the instrument contain a high percentage of OEM parts or does it contain many custom parts? If it contains many custom parts, can the market price recover a reasonable profit?

Talystep: Mid 1960s.

The first instrument suitable for measurement of film thickness on the nanometer level was the Talystep. Designed by R.E. Reason in the years prior to the instrument's introduction in the mid-sixties, it quickly established itself as the stock-in-trade tool of the engineering research community and the optical coatings and the nascent microelectronics industries. The profiler had several features which were revolutionary for their time: An LVDI transducer head capable of 0.5 nm resolution and an arcuate path flexure traverse mechanism.

The Talystep transducer head was, until the development of scanning probe microscopes in the early 1980s, the favored transducer for nanometric surface science. Using an LDVI, whose operation was outlined earlier, the transducer combined several unique features together in one package. Some mechanism is needed with the LVDI to prevent a tip crash from applying a catastrophic force to the specimen, stylus and measurement head. Crash protection was provided by supporting the LVDI coils on a second spring mounted flexure mechanism. In normal operation this flexure rests against a stop block, effectively removing the spring mechanism from the measurement

loop. But, when overextended, the LVDI platen will lift the coil assembly off its' stop, supported by the spring flexure.

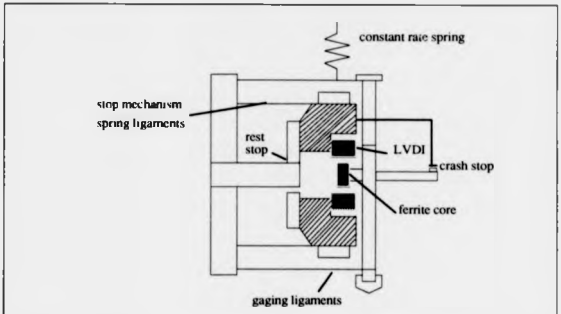


Figure 1.20. Talystep head showing crash stops and force adjustment spring.

To allow adjustment of stylus contact force, the probe flexure is connected to a constant force spring mechanism, relieving the probe of its own weight. In this manner, force can be adjusted from 10 to 300 μN (1 to 30 mgf). Contact force can be further reduced to 0.6 μN (0.06 mgf) by modifying this spring[34]. Low force stylus operation was extremely important in the measurement of surface roughness using the chisel shaped diamond styli. These chisel styli were cleaved diamond with a flat tip of 0.1 by 2.5 μm at the end. Under forces of more than a few tens of μN, these tips could contact the specimen with enough pressure to cause scratches and give the *appearance* of great surface damage. The nature of this damage is currently the subject of some debate and was in part, the impetus for the Author's work in this field.

The traverse mechanism was one of the first to be of the arcuate flexure type. Recalling that with the arcuate traverse mechanism, the profiling datum is the path

traced by the flexure, high stability is needed in the axial direction. Implementation involved hanging the transducer head and coarse approach mechanism by a deep cantilevered flexure hinge, made very stiff in the vertical direction. The other end of the hinge was attached to a slot in the aluminum casting used to house part of the profiler. Horizontal traverse was on the order of ±1 mm, with a small motor-gearbox-micrometer for a drive mechanism.

Ergonomically, the Talystep receives mixed scores for its easily adjusted transducer head but difficult to master leveling system. The transducer head itself, while capable of the highest resolution possible, had a very low resonant frequency (around 10 Hz) which severely limited traverse speed for roughness measurements. The instrument was not capable of automated operation and fell out of favor amongst manufacturers when other, more automated instruments came along.

Economically, very few OEM parts were used, with some assemblies containing a fairly large number of small parts. This arrangement could only work as long as there was little competition in the market and premium prices could be charged.

Alpha-Step: Mid 1970s.

With massive expansion in the microelectronics and thin-films industries, room was opening up for another player in the market to compete with Talystep. The original Talystep design, then over ten years old, was failing to keep pace with the demands of the microelectronics industries[35]. Most notably, the throat clearance of the Talystep was under 50 mm, not enough to profile the 125 mm wafer size then in use. Leveling the Talystep was also perceived as a problem. Microelectronics fabrication houses (or "fabs" in the vernacular) wanted to use unskilled workers for more and more production tasks. Leveling a Talystep was a chore for a patient technician with an understanding of the geometry of the leveling platform. A production oriented instrument with "one button scanning" was conceived and brought to market in some six months time.

The new instrument was able to profile an entire 125 mm wafer and had automatic leveling so that relatively unskilled technicians could make reliable measurements. OEM components would be used whenever possible, as the design and prototype cycle was extremely short for the then start-up company (Tencor Instruments). The design made extensive use of commercially available flexure pivots, commonly referred to in the United States as "Bendix pivots". These pivots were used to form a two-axis gimbal mount for an aluminum tube used as the scan arm.

The scan arm was driven by a string drive mechanism and a small Swiss gear motor. The leveling axis was driven by another small gear motor driving a micrometer head, rocking the gimbal on its axis to provide leveling. Using analog and descrete digital control circuitry (no micro processors), the unit would automatically level a flat or step height sample placed in contact with the stylus arm. Several lead sheet and foam dampers are located along the inside of the aluminum casting (which serves as the housing) to reduce vibrations. Dampers are also located on the small gear motors and inside the tubular scan arm.

The transducer was miniaturized by utilizing a commercially available LVDT some 10 mm in diameter. The stylus arm, with contact force adjustment provided by sliding a small shuttle-mass on the tracking arm had very little mass added by the ferrite slug used for the inductive pickup. Resolution of the standard instrument was specified as 5 nm, probably close to an order of magnitude worse than the Talystep. A variety of probes were available, with most users choosing a 2 or 5 μm stylus for step height measurements. Stylus force was not easily adjustable, with a range of 1 to 20 milligrams force.

Coarse approach is provided by a lifting cam and actuating lever. This cam lifted the entire sample translation stage to a point just out of contact with the specimen. Final approach was provided by a fan shaped gear attached to the end of the lifting lever, engaging a small, toothed wheel. This sine bar type arrangement allowed rapid

swapping of samples, a distinct advantage in production situations. Trace recording was achieved by a small chart recorder located inside the control unit. Combined weight of both the control and profiler units is 14 Kg, with the whole system fitting neatly into a suitcase.

The Alpha-Step was primarily designed for ergonomics. An easy to use instrument, it soon usurped the Talystep as the new stock-in-trade tool of the microelectronics industry. A high percentage of OEM parts made the unit simple and cheap to manufacture for a small, growing company. The Talystep, with its superior resolution remained an instrument of choice in many research situations for some years to come.

Nanostep and P2: Late 1980s.

Driven by demands for measuring form errors for X-ray optics and flat panel display glass, stylus instruments were in need of overhaul by the late 1980s. With scanning tunneling and atomic force microscopes rapidly filling the needs of researchers in the short (roughness) end of the spatial spectrum, stylus instrument makers turned their attention to the more difficult long (waviness) components. Arcuate flexures could only be pressed into service for traces of 10 mm or less in length. What was needed was a super-precision bearing for a optically flat datum surface. These bearings were found as the fruits of decades of tribology work by researchers such as Tabor and Bowden[36]. Polytetrafluoroethane (PTFE) solid bearings of several varieties were developed which address a number of concerns, albeit in different embodiments.

Removal of heat from friction at the bearing surface of primary concern in polymer bearings. Low friction polymers useful for precision bearings are in general poor conductors of heat away from the bearing surface. In the case of PTFE, this heat can be great enough to cause outgassing and the development of micro-bubbles on the bearing surface. One way of removing such heat energy [37]is to use a thin layer of polymer on a metal backing. Heat is conducted safely away from the bearing surface if the film thickness is on the order of 10 μm or less. Alternatively, the PTFE may be

bound to a metal or carbon fiber matrix, capable of sinking away heat. By whichever method, the result is a super-precision carriage and slideway system with noise levels on the order of 50 pm.

Work at NPL on such a slideway system culminated in the Nanosurf2 instrument which eventually led to a technology transfer to Rank Taylor Hobson[38] where it became commercially available as the Nanostep. Based on Talystep transducer head technology, the instrument is a glass-ceramic embodiment of the low thermal effects by materials approach. Much of the measurement loop is constructed of Zerodur glass ceramic, giving the instrument a temperature coefficient of approximately 7 nm/°C. Maximum length of traverse is 50mm, with speed limited by the bandwidth of the modified Talystep transducer head (8 Hz). Lack of automation has made the instrument primarily a tool for research, rather than a production quality control tool.

Development of profiler products at Tencor Instruments produced an automated profiler based on a optical flat datum with a 200 mm travel in two axes, the P2 series of instruments. This two axis slideway drags a specimen carriage across an optical flat by a taut wire coupling. Precision polymer bearings of a proprietary nature are used. The profiler is fully automated for production use. Although a low moment of inertia measuring transducer is available for low contact force profiles, bandwidth is still limited to 4 Hz, making high-speed profiles unlikely.

Atomic Force Microscopes (AFMs)

The atomic force microscope is a more recent offspring [39] of the scanning tunneling microscope. Sensing minute deflections of a probe located at the end of a cantilever beam, the AFM follows a contour of constant force acting between the probe tip and the specimen, across a surface in a raster pattern.

High resolution atomic force microscopy is possible because of the short range nature of the intermolecular forces between atoms in the stylus and those on the surface of the

specimen. The Leonard-Jones potential provides a basis for understanding these forces, even though it is a experimentally derived pair potential between two atoms in a gas. The force between two atoms is

$$F = \frac{A}{X^m} - \frac{B}{X^n} \qquad (1.16)$$

where A and B are experimentally derived constants, as are the power laws m and n[40]. The coefficients *m* and *n* are often expressed as 6 and 12, respectively[41]. The first term in the equation relates to the van der Waals attractive forces, while the second term relates to the shorter-range repulsive forces acting between atoms. Despite the sometimes poor fit of this equation to real physical situations, a plot of the intermolecular forces is enlightening.

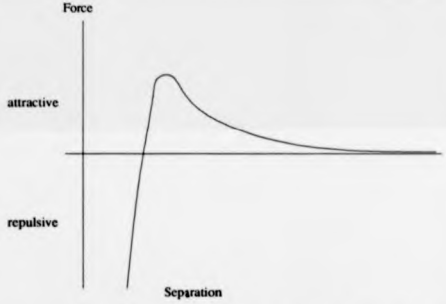


Figure 1.21. Intermolecular forces between two atoms showing attractive and repulsive regions.

The figure has purposely drawn without a scale because all the constants in eq. 1.16 are dependent upon materials, environment and geometry. Typically though, in AFM

work, the the weaker attractive van der Waals forces extend some tens of nanometers from the surface while the much stronger repulsive forces dominate at very close (pm) seperations. Consider the case of a small sphere attached to a spring. As the sphere is brought near the surface, van der Waals attractive forces pull against the spring, drawing it closer to the surface. At some point, the forces of molecular attraction become greater than the restoring force of the spring. At this point, the sphere is "snapped-in" to the surface and will remain in contact with it until the spring is stretched enough to provide a sufficient amount of energy to free the sphere from contact.

AFM Instrumentation

Using once again the nomenclature of Teague, elements of the AFM may be modeled using the following basic elements:

1. A force transducer such as the optical lever or optical interferometer combined with a cantilever spring.

2. A three dimensional raster-scanning mechanism capable of sub-nanometer controlled displacements (usually piezoelectric.)

3. A control system to servo the scanning mechanism to maintain a contour of constant force.

4. A probe with a tip radius on the order of 10 to 20 nm. Most often, these tips are batch processed on a microcantilever using silicon micromechanics technologies.

Two methods of surface mapping are possible with the aforementioned nomenclature. *Scanned specimen* and *scanned probe*. The scanned specimen technique rasters a small specimen with respect to a stationary probe. This allows for simple force transducer construction, but places considerable constraint on the size and mass of the specimen. Typically, thin slices only a few millimeters thick and about 10 mm on a

side. Current trends are toward a scanned probe, stationary specimen topology. No constraints are placed on the sample, but more ingenuity is involved in force transducer design, as it must be compact and light enough to be rastered by the scanning system.

Force Transducers

All of the following force transduction methods make use of the linear cantilever spring assumption. To wit: A uniform cantilever beam of known cross-section and length behaves as a linear spring for small displacements, relating beam deflection to applied force at the tip by Hooke's law. All of these methods have met some success in working instruments.

Tunneling probe

The most sensitive displacement transducer known, electron tunneling from a point probe to a conductive specimen can undergo an order of magnitude change in tunneling current for a 0.1 nm change in tip/sample separation. By applying a small bias voltage of a few tens of millivolts to the tip and measuring the current through the tunneling gap, an indication of gap distance can be ascertained. For the idealized one dimensional case, tunneling current, I is exponential with gap distance[42],

$$I \propto \left(\frac{V}{d} e^{(-A\phi^{\frac{1}{2}}d)} \right). \qquad (1.18)$$

Where V is the gap bias voltage, ϕ is the average barrier height and d is the gap distance. By processing the current signal through a logarithmic amplifier, a nearly linear signal may be obtained, as first measured by Teague[43].

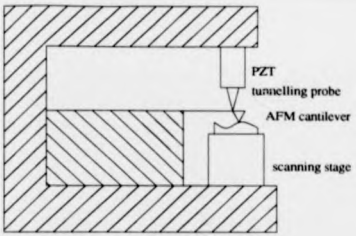


Figure 1.22. Tunneling AFM sensor operation.

Such was the system first used to detect cantilever beam deflection in the first AFM[44]. Tunneling quickly fell from favor though, because of the local nature of the effect. Surface topography and contamination of the force cantilever electrode produce noise in this system. Although electron tunneling is an extremely sensitive distance measuring transducer, surface conditions in air make it extremely difficult to use as an reliable metric.

Optical lever-arm

An adaptation of the time-tested photoelectric galvanometer, this method has long been used to study micro-deflections down to the limitation imposed by their thermal noise limit. Use of lever arms in atomic force microscopes came in around 1989 by Hansma, et. al[45]. The principle is very simple, figure 1.23

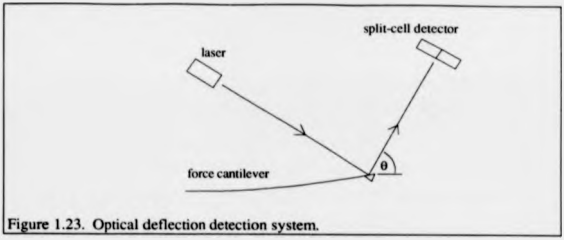


Figure 1.23. Optical deflection detection system.

A collimated or focused optical beam from a semiconductor laser is incident on the reflective surface on the reverse side of the force cantilever. Deflections of the cantilever will result in an angle change $\delta\theta$ of the mirrored surface. The reflected beam undergoes an angle change of $2\delta\theta$, where it strikes a position sensitive detector. This position sensitive detector may be as simple as an opaque slit in front of a photodiode, where the output is a convolution of the optical beam distribution with the slit shape. Normally however, the position sensitive detector is a segmented PIN diode used with a differential transimpedence amplifier to yield a linear position voltage. The segmented PIN diode may be a two element bi-cell for simple beam deflection monitoring, or a four element quad-cell to monitor beam torsion as well as simple deflection. The quad-cell scheme is useful because, in principle, it allows probe friction against the surface to be monitored as well as normal forces. Calibration in the frictional axis remains problematic as of this writing.

A drawback of the lever-arm scheme is that the structure required to firmly fixture the laser diode, force cantilever and detector in place is too bulky to be easily scanned by piezoelectric actuators at high scan rates. An interesting variation is presented by Meyer[46] where the cantilever is scanned only a few microns, and the focused light spot is allowed to "wander" on the cantilever. Such a scenario is subject to errors resulting

from warpage of the cantilever. Nonetheless, it continues to be researched[47] and is used in commercial instruments. In many AFMs, the lever arm components make it somewhat difficult to obtain a simultaneous optical microscope view of the cantilever and surface without more costly long working range objective lenses.

Fiber optic interferometer

Essentially a one-dimensional Fizeau interferometer, this embodiment attempts to make use of fiber optics to construct an interferometric transducer at the force cantilever with very little added mass. First applied to AFM by Rugar[48], the technique soon evolved into a more practical form[49] (Rugar, 1989). Light from a laser diode is coupled into a single mode fiber optic and sent to an optical power splitter. The splitter, or 3 dB coupler in the vernacular equally divides the light into two exiting fibers. One of these fibers is cleaved flat at the end and left open to air. This open end allows a Fresnel reflection of about 4% intensity from the cleaved face, making the reference beam of the interferometer. Placing a mirrored cantilever tip some ten micrometers from this end, the signal reflection from the force cantilever has a very short (20 μm) optical path and is almost immune to air turbulence. Combining in the fiber, the reference and signal beams form interference fringes in the fiber. These fringes may be observed by placing a detector on the free fiber on the launch side of the power splitter, Figure 1.24. Differential measurement with a second detector on the remaining free fiber on the interferometer side of the power splitter reduces laser diode intensity variations in the final measurement. Noise equivalent displacements may be as low as 20 pm RMS.

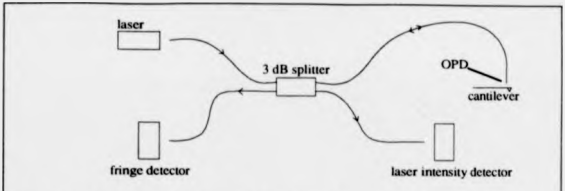


Figure 1.24. Fiber optic interferometer detection scheme.

Several problems are present in this interferometer scheme. First, a problem with all interferometers using only one detection signal. The output of such an interferometer is, of course sinusoidal, not linear. Since only one signal is present, direction can not be discerned. Thus, linear operation is only possible around the *quadrature points*, the nearly linear regions around the zero crossings (see Chapter 4 on distance measuring interferometry for more detail). For a HeNe laser and a Michelson interferometer, this linear region is only about 20 nm. The problems compound if operation shifts from one quadrature point to the next. The slope of such a case is opposite, causing a positive feedback condition and a jump to the next operating point, at a different, indeterminate force.

Alignment of the force cantilever to the fiber end is quite critical, but can be achieved by carefully gluing cantilevers to a fixturing jig[50]. Mechanical and thermal stability of the fiber optics are usually quite poor, with difficulty usually encountered with coupling light into the fiber and getting light out of the waveguide. Polarization sensitivities of the power splitter are also detrimental to operation. Furthermore, the depth of fringe modulation changes with optical path length[51]. As of this writing, the fiber optic interferometer has only been used on experimental, research instruments.

Laser diode interferometer

Laser diode interferometers are a new variant of the fiber optic interferometer. A truly clever innovation, the laser diode interferometer consists of the laser diode package and the force cantilever. No other optics are necessary. Developed at the University of Arizona by Sarid et. al[52] the laser diode interferometer can have a sensitivity as high as 3 pm per root Hertz[53]. The sensor has recently been patented[54] and brought to market in a stand-alone AFM application[55].

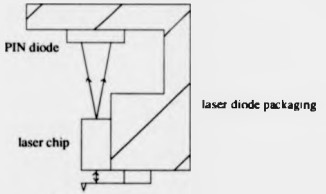


Figure 1.25. Self-contained laser diode interferometer.

The price of this convenience comes at the expense of a few modifications of the laser diode package. First, the theory of operation: If a mirrored surface is placed some 10 μm from the front facet of a single-mode laser diode, a Fabry Perot cavity is formed, externally coupled to the front facet of the laser. Most laser diodes are packaged with an integral PIN diode located close to the back facet of the diode chip, for monitoring optical intensity of the laser. By making use of this back-facet PIN diode, a complete, integrated detector is possible. Energizing the laser diode with a constant current just beyond the threshold required for laser operation, the back-facet PIN diode is monitored for fringe information.

The main modification to the diode laser package involves removing the window from the front of the case to allow easy access to the laser diode chip itself. A cantilever must then be glued to the copper heatsink block which mounts the laser diode. It should be noted that the laser diode represents a constant heat source near the cantilever. Also, the same range problems that beleaguer the fiber optic interferometer, plague the laser diode interferometer. That is, undetected shifting of the operating point from one quadrature point to another.

Strain gage

Embedding a strain gage in a into a microfabricated cantilever makes for a tiny force metric. Using the piezoresistive effect of p-type silicon, Tortonese[54] reports minimum detectable deflections of around 0.2 nm with a bandwidth of 1 kHz. Cantilever deflections are monitored by connecting the force probe as one element of a Wheatstone bridge. In operation in a commercial instrument, this bridge is DC biased, with each bridge element being approximately 2.4 KΩ. It should be noted that the microcantilever dissipates nearly 1 mW of power in this configuration. Placing it in close proximity with the specimen causes the cantilever to couple radiatively into the specimen surface. In this case, the probe becomes an excellent emmisivity monitor. The device is also extremely temperture unstable and closely resembles a hot-wire anenmometer used for measuring air-flow. Figure 1.26 shows a microscope interferogram of a commercial resistive force probe.

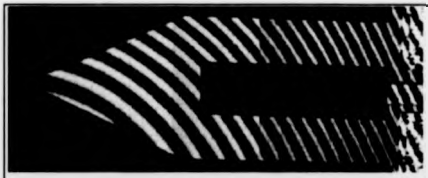


Figure 1.26. Microscope interferogram of a commercial piezoresistive strain gage atomic force microscope probe (Zygo Maxim, UNC Charlotte Precision Engineering Lab.).

Capacitance gage

Alternating current bridge methods, like those used in stylus instruments, can also be adapted for use on atomic force microscopes. Inductive transducer methods are simply too bulky to be used on miniature AFM cantilevers. The capacitance bridge can be of use in AFM detection by virtue of the almost negligible added mass of thin-film capacitance plate electrodes. If a small 10 μm or so air gap can be formed between the reverse side of the force cantilever and another stationary electrode, (figure 1.27) the subsequent capacitor can be monitored for changes by a sensitive capacitance bridge. The detailed design and construction of such bridges will be the subject of Chapter 4.

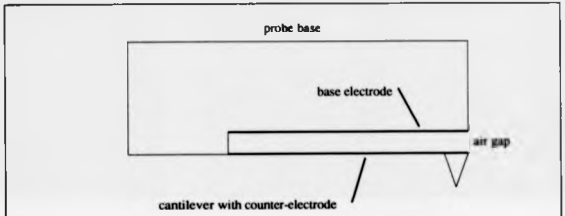


Figure 1.27. Principle of operation for capacitive AFM sensor.

Earlier work by Göddenhenrich [57] with a wire flexure supported foil electrode achieved sensitivities on the order of 10 pm. Work by Neubauer [58] et. al. with a high frequency bridge was able to detect frictional as well as normal forces on a wire cantilever with a noise level of 30 pm.

Scanning mechanisms

Although *all* of the different scanning mechanisms developed for probe microscope applications cannot be mentioned here, the vast majority at the time of writing can be

lumped into one of three categories. Not all choices of actuator can be considered here either. Piezoelectrics, despite their well-documented problems with hysteresis and creep[59], remain the choice of actuator for scanned probe microscopes. Relatively cheap, piezoelectrics produce motion on the order of 1 μm/100V, depending on materials and geometry. Strain in the material brings a change in the net dipole moment, resulting in a net charge across the material. This mode of operation is useful for transduction of dynamic signals. Reverse piezoelectricity used for actuators converts applied charge to a strain in the material.

Piezoelectric tripods

One of the earliest, and possibly the first application of piezoelectrics to nanometer scale scanning was Russel Young's Topografiner[60] in 1972. This groundbreaking instrument was the direct predecessor of modern scanning tunneling microscopes. In it, three piezoelectric elements were used to support a flexible metal post. Motions of the piezoelectric elements bend the post slightly and can produce a raster scanning pattern. The scan pattern is slightly arcuate, but introduces error on the order of about one part per million deviation from a rectilinear path with the geometry used in this design.

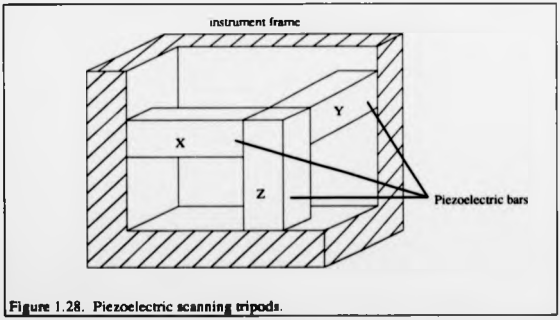


Figure 1.28. Piezoelectric scanning tripods.

Piezoelectric tripod performance is limited by the joints in the connecting piezoelectric bars. These bars can be over constrained if the rods are solidly fixed to both the actuator and the tripod. Conversely, lost motion can result from low stiffness in the glued joints used to join the bars. In recent years, the scanning tripod has largely given way to other actuators such as the tube scanner.

Piezoelectric tube scanners

Seeking a simpler, more elegant solution to the scanner problem, Binnig and Smith[61] developed a device resembling a ceramic phonograph pickup using reverse piezoelectricity. A thin walled tube of radially polarized piezoceramic is coated with a solid electrode on its inside diameter and a quartered electrode pattern on the outside , figure 1.29.

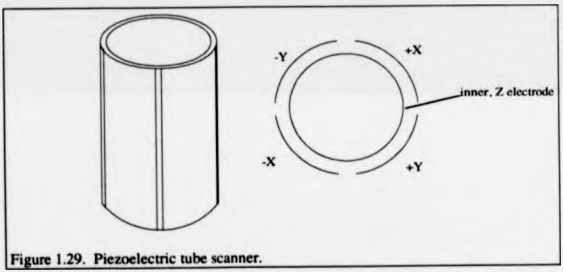


Figure 1.29. Piezoelectric tube scanner.

By driving opposing quarters in a push-pull configuration, the tube may be made to bend along its axis. Z axis motion is achieved by raising the potential of the inside electrode, causing all quarters to contract or expand equally. For three dimensional scanning, five high voltage amplifiers are required, one for each electrode. For atomic

level scanning, a tube length of approximately 12.5 mm is commonly used. These tubes often have maximum travels of a few micrometers and resonant frequencies on the order of 20 kHz. Longer range scans require longer scanning tubes, up to 100 mm for a 100 μm scan. Resonant frequencies for these tubes are at best only a few kHz. As scan lengths for tube scanners get longer, problems arising from creep, hysteresis and arcuate motion become more prevalent.

Piezo-driven monolithic flexure stages

With recent trends toward longer range, metrological microscopes, monolithic flexure stages have enjoyed a recent resurgence. Flexure stages have the advantages of smooth motion, relatively free from hysteresis and stick-slip friction. They may be arranged to act as lever arms to amplify to motions of piezoelectric actuators to scan lengths of 1mm square. The first commercially available stage, the "Piezo-Flex" was described by Scire and Teague[62] in 1978, Figure 1.30.

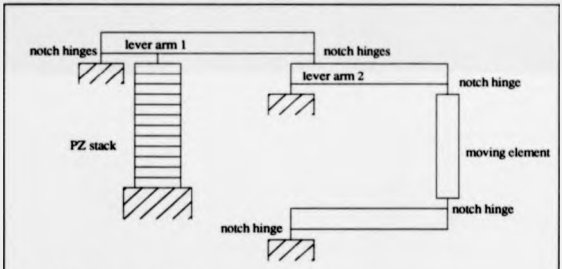


Figure 1.30. Original "Piezo-flex" stage design.

Many monolithic flexure designs make use of the notch hinge, produced by jig boring two circular holes close together. The lever shapes are often cut by wire EDM, which

has become increasingly popular and is the only way to machine thin walls. Aluminum alloy is a common material, with brittle materials such as silicon and Zerodur glass-ceramic occasionally used. The advantages of brittle materials in metrology tools is such that any damage to the flexure mechanism great enough to change the character of the device will also render it useless. In other words: It works perfectly or it breaks!. One disadvantage of the notch hinge is that it concentrates stress and severely limits maximum displacements. One remedy for this is to wire erode bladed flexures as described by Xu and Smith[63], distributing the stress over a larger volume.

Miscellaneous:

Many other varieties of actuators exist, but are not in common use for microscopes at this time. Electromagnetic voice coil actuators[64] are often used to move large loads with displacements up to 1 mm. Requiring much less electronic resources than piezoelectrics, they can be successfully used in any application where their dissipated heat is not a problem. Although power dissipation is not a problem for very small displacements, the devices tend to be rather bulky. They are typically employed on larger precision machines such as step and repeat cameras[65] where size is not a priority.

Much smaller micromachines often use the forces of electrostatic attraction as a motive element. Easy to implement wherever there are two electrode plates separated by a gap, the method scales extremely well into the micro-miniature. While dielectric breakdown was often cited as a limiting factor, it appears as though diminutive electrode gaps, on the order of the mean free path of air, begin to approximate a vacuum environment and its higher breakdown threshold [66]. Although large amounts of energy may be transferred by this method, it is often relatively difficult to design effective mechanisms on planar microelectronics substrates. So-called electrostatic comb drives are sometimes used to generate motion across the plane of the substrate[67].

A relatively recent addition to the commercial arena, electrostrictive drives are becoming quite common in precision instruments because of their low creep, hysteresis and driving voltage when compared to piezoelectrics. Application of an electric field acts to rotate the polar domains with a behavior characteristic of electrostatic actuators. Sensitivities are somewhat better than piezoelectrics, around 30 V/μm for a typical drive. They are not as stiff as piezoelectrics and have poor low temperature performance. The devices are fairly non-linear at low voltages, but may be considered to have linear regions at about one half of their maximum working voltage. Linearization is thus accomplished by either operating in this region with a DC bias or using computational techniques such as look-up tables or a closed loop servo with an external sensor. As of this writing, electrostrictive actuators are more costly then piezoelectrics, but rapidly falling in price.

Magnetostrictive drives employ an change of strain energy for changes in the uniform magnetic field surrounding it. Best used as very stiff actuators for precise control of large, heavy machine elements, magnetostriction has seen little use in precision instrumentation. A magnetostrictive drive would be attractive because of its low cost of associated electronics and ability to move large loads. New materials such as Terfenol-D are currently being explored[68].

AFM Control systems

The function of the control system in the AFM is to maintain a constant applied force between a probe and specimen if the two are rastered relative to each other, a contour of constant force is followed. Since the topographic signal that is recorded is usually derived from the control system output, the servo system must follow a contour that accurately describes the specimen surface. Errors of the controller become errors of the measurement. Given the importance of the servo system to commercial instruments, very little public-domain research exists on the subject.

Whitehouse [69] addresses measurement fidelity by first attacking the issue of suitable mathematical models. Once defined, these models are used to determine the optimum damping characteristics of a second-order stylus tracking system. From this analysis, an optimum value of ζ (damping cefficient)is determined to be 0.59. This figure is of importance when we realize that many AFM following systems are either of the proportional plus integral (PI) type, or simply a high gain proportional controller with very little damping. Rate dependent damping, the "D" term of a PID controller, is not often used due to the fear that the differentiator will be unstable and add excess noise. A low-pass filter used a rate-limiter does much to alleviate these problems.

Typically, force probe cantilevers are high Q, lightly damped devices with resonant frequencies in the tens of kHz. This is to ensure not only high speed operation, but also so that any external excitations (thus causing the system to vibrate at the resonant frequency) cause oscillations that are far beyond the sensor bandwidth. The addition of some damping could allow much larger cantilevers with no loss of bandwidth or performance. More will be said on controllers in Chapter 2.

Force probes

Microfabricated AFM cantilevers with integral probe tips are the result of a decade of silicon micromechanism development ushered in by the publication of Petersons's landmark review paper[70] in the early 1980s. Throughout the decade, anisotrophic etching and bulk micromachining techniques were perfected, as were surface micromachining techniques to produce low-stress thin films. Early AFM force cantilevers were made of gold foil or platinum wires, laboriously hand-assembled with diamond or etched wire tips. Today, virtually all AFM work is done with microfabricated cantilevers. Introduced in 1989 by Quate and Albrecht[71], many improvements have followed. Originally fabricated from SiO_2 or Si_3N_4 thin films, the cantilevers have since been fabricated using single crystal silicon[72]. Tip materials range

from silicon nitride to single crystal silicon pyramidal styli with approximately 10 nm radii for the single crystal Si embodiment.

Fabrication of a commercially available cantilever product usually involves the etching of silicon to form pit-shaped molds for the Si_3N_4 layer, figure 1.31, after Albrecht.

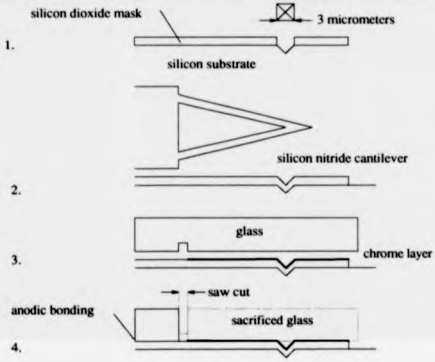


Figure 1.31. Microcantilever fabrication for silicon nitride cantilevers.

A layer of chromium is then deposited onto the cantilever, not extending past the base to the substrate and a Pyrex glass plate is anodically bonded[73] to the substrate, adhering to the Si_3N_4 but not to the chrome on the cantilever. The glass above the cantilever is then removed with a saw cut and all the silicon is sacrificially etched by either KOH or EDP etching. The finished cantilever is neither uniform in geometry or free from resisual stresses.

Discussion

Upon first glance, a lay person could only conclude that there is little if any similarity between stylus instruments and atomic force microscopes. As briefly outlined, they have very little in common in their metrics, or transducer systems. The traverse or scanning mechanisms are completely different, and the entire *modus operandi* is different: A contour of displacement versus a servoed contour of constant force. Yet both techniques provide topographical information to the surface metrologist. Repulsive mode AFM operation is thus physically very similar to that of the stylus profiler.

AFM instrumentation has, until recently, been concerned with the very short end of the spatial spectrum or small features. The easy end from an instrumentation standpoint as Reason pointed out. Tracing a datum has been relatively easy at micron traverse ranges using arcuate motion scanners. Metrology in any axis has been largely ignored, with linear relationships assumed for piezoelectric extensions. As instrument makers are now finding out, the shift from the short traverse to the long is arduous; the datum must be smoothly defined and metrology must be performed on the scanning mechanisms. As the new atomic force instruments grow to larger sizes, the influences of moving parts on the structural loops will need to be assessed. Environmental control grows more stringent as instrument loops grow longer and measurement times increase for large samples. In other words, all the elements of precision instrument design that have dogged makers of stylus instruments for years must now be applied to long-range atomic force microscopes as well. Also, metrologists must study the nature of the tip/specimen interaction forces at a levels three to six orders of magnitude lower than were encountered with stylus instruments.

---

[1] D. Whitehouse, "The parameter rash-is there a cure?", *Wear*, 83, pp. 75-78, (1982).

[2] R.E. Reason, "Progress in the appraisal of surface topography during the first half-century of instrument development", *Wear*, **57**, p 1-16, (1979).

[3] E. Hecht, Optics 2nd Ed., .Addison-Wesley, Reading, Mass., 1988, p 422.

[4] E. Betzig, P.L. Finn and J.S. Weiner, "Combined shear force and near-field scanning optical microscopy", *Appl. Phys. Lett.*, **60**(20), p. 2484, (1992).

[5] E. Betzig, M Isaacson and A. Lewis, "Collection mode near-field scanning optical microscopy", *Appl. Phys. Lett.*, **51**(25), p 2088, (1987).

[6] O. Dupuy, "High precision optical profilometer for the study of microgeometrical surface defects", *Proc. Inst. Mech. Engrs.*, **182**(3K), p 255-259, (1967/68).

[7] J. Simon, *Appl. Opt.*, **10**, p 2337, (1970).

[8] J.A. Simpson, "Use of a microscope as a noncontacting microdisplacement measurement device", *Rev. Sci. Instrum.*, **42**, p 1378-1380, (1971).

[9] F.S. Jing, A.W. Hartman and R.J. Hocken, "Noncontacting optical probe", *Rec. Sci. Instrum.*, **58**(5), p 864-868, (1987).

[10] T. Kohno, N. Ozawa, K. Miyamoto and T. Musha, "High precision optical surface sensor", *Appl. Opt.*, **27**(1), p 103-108, (1988).

[11] A.E. Ennos and M.S. Virdee, "Precision measurement of surface form by laser autocollimation", *Proc. SPIE*, **398**, pp. 252-257, (1983).

[12] K. Creath, Progress in Optics, Vol. XXVI, E. Wolf, ed., Elsevier Sci. Pub. B.V., 1988, pp. 351-393.

[13] C.J. Evans and W.T. Estler, "Some observations on the performance of commercial phase measuring (Fizeau) interferometers used in surface figure metrology", *Proceedings of the ASPE 1991 annual conf., Santa Fe, NM*, pp. 54-57, (1991).

[14] J.H. Rakels, "Nanometric surface finish measurement by optical diffraction", *Measurement + Control*, pp. **24**(2), pp. 56-59, (1991).

[15] National Aeronautics and Space Administration, "The Hubble space telescope optical systems failure report", *NASA TM-103443*, (NASA, Washington, DC, 1990).

[16] L. Furey, T. Dubos, D. Hansen and J. Samuels-Schwartz, "Hubble space telescope primary -mirror characterization by measurement of the reflective null corrector", *Applied Optics*, **32**(10), pp. 1703-1714, (1993).

[17] J.B. Breckenridge and H.J. Wood, "Space Optics: An introduction by the Editors", *Appl. Optics*, **32**(10), p. 1677, (1993).

[18] T.V. Vorburger and E.C. Teague, "Optical techniques for on-line measurement of surface metrology", *Precision Eng.*, p 61-83, (1981).

[19] I. Sherrington and E.H. Smith, "Modern measurement techniques in surface metrology: Part II; Optical instruments", *Wear*, **125**, p 289-308, (1988).

[20] E.C. Teague, "Nanometrology", *AIP Conference Proceedings 241, Scanned Probe Microscopy, Santa Barbara, CA*, HK Wickramasinghe Ed., p 371-407, (1992).

[21] E.C. Teague and C. Evans, "Patterns for precision instrument design (mechanical aspects)", *American Society of Precision Engineering Tutorial, Santa Fe, NM*, (1991).

[22] R.V. Jones, Instruments and Experiences, Papers on Measurment and Instrument Design, John Wiley and Sons, NY, pp. 122-144 , (1988).

[23] Now manufactured by Lucas Aerospace, 211 Seward Ave, PO Box 457, Utica, NY 13503, USA.

[24] S.T. Smith and D.G. Chetwynd, Foundations of Ultraprecision Mechanism Design, Gordon and Breach Scientific Publishers, Reading, U.K., see chapter 4 for many flexure designs, (1992).

[25] R.E. Reason, Ibid., p 9, (1979).

[26] S.T. Smith and D.G. Chetwynd, Ultraprecision Mechanism Design, Gordon and Breach Science Publishers, Reading, U.K., p. 262, (1992).

[27] J.B. Bryan, "The Abbé principle revisited: An updated interpretation", *Precision Engineering*, 2, pp. 129-132, (1979).

[28] F.E. Scire and E.C. Teague, "Piezodriven 50 μm range stage with subnanometer resolution", *Rev Sci. Instrum.*, **49**(12), pp. 1735-1740, (1978).

[29] P.B. Leadbeater, W.J. Wills-Moren and P.A. McKeown, "Ultra Precision Stylus Profilometry for Large Optical Components, *Proceedings ASPE 1990 annual conference, Rochester, NY.*,, P. 114-117.

[30] C.J. Brenan, T.D. Doukoglou, I.W. Hunter and S. Lafontaine, "Characterization and use of a novel optical position sensor for microposition control of a linear motor", *Rev. Sci. Instrum.*, **64**(2), p 349-356, (Feb., 1993).

[31] D.J. Whitehouse, "Some ultimate limits on the measurement of surfaces using stylus techniques", *Meas. Control*, **8**, pp. 147-51, (1975).

[32] Conversation with E.C. Teague, NIST Precision Engineering Division, December, 1992.

[33] J.E. Griffith, D.A. Grigg, M.J. Vasile, P.E. Russel and E.A. Gerald, "Scanning probe metrology", *J. Vac. Sci. Technol. A*, pp. 674-679, (Jul/Aug, 1992).

[34] J.F. Song and T.V. Vorburger, "Stylus profiling at high resolution and low force", *Applied Optics*, **30**(1), p 42-50, (1991).

[35] The Author is greatly indebted to Mr. Bill Wheeler, the mechanical designer of the Alpha-Step for his extremely useful recollections.

[36] F.P. Bowden and D. Tabor, The Friction and Lubrication of Solids, Part II, Clarenden Press, Oxford, 1964, Chapters XI and XII.

[37] K. Lindsey, S.T. Smith and C.J. Robbie, "Sub-nanometre surface texture and profile measurement with Nanosurf2", *Annals of the CIRP*, **37**(1), p 519-522, (1988).

[38] J.D. Garratt and S.C. Bottomley, "Technology transfer in the development of a nanotopographic instrument", *Nanotechnology*, **1**(1), p 38-43, (1990).

[39] G. Bennig C.F. Quate and C.H. Gerber, "Atomic Force Microscope", *Phys. Rev. Letts.*, **56**(9), p. 930-933, (1986).

[40] D. Tabor, Gases, liquids and solids and other states of matter, 3rd ed., Cambridge University Press, Cambridge, 1991.

[41] J. Israelachvili, Intermolecular and Surface Forces, 2nd ed., Academic Press, NY, 1992, p. 113.

[42] Y. Kuk and P.J. Silverman, "Scanning tunneling microscope instrumentation", *Rev. Sci. Instrum.*, **60**(2), p 165-179, (1989).

[43] E.C. Teague, "Room Temperature Gold-Vacuum-Gold Tunneling Experiments", *J. Res. of the Natl. Bur. of Stds.*, **91**(4), (1986).

[44] G. Bennig, C.F. Quate and C.H. Gerber, "Atomic force microscope", *Phys. Rev. Letts.*, **56**(9), p. 930-933, (1986).

[45] S. Alexander, L. Hellemans, O. Marti, J. Schneir, V. Ealings, P.K. Hansma, M. Longmire and J. Gurley, "An atomic resolution atomic force microscope implemented using an optic lever", *J. Appl. Phys.*, **65**(1), p 164-167, (1989).

[46] G. Meyer and N.M. Amer, "Optical beam deflection atomic force microscopy: the NaCl (001) surface", *Appl. Phys. Let.*, **56**(21), p 2100-2101, (1990).

[47] C.A.J. Putnum, B.G. DeGrooth, N.F. Van Hulst and J. Greve, "A detailed analysis of the optical beam deflection technique for use in atomic force microscopy", *J. Appl. Phys.*, **72**(1), p. 6-12, (1992).

[48] D. Rugar, H.J. Mamin, R. Erlandsson, J.E. Stern and B.D. Terris, "Force microscope using a fiber-optic displacement sensor", *Rev. Sci. Instrum.*, **59**(11), p 2337-2340, (1988).

[49] D. Rugar, H.J. Mamin and P. Guethner, "Improved fiber-optic interferometer for atomic force microscopy", *Appl. Phys. Lett.*, **55**(25), p 2588-2590, (1989).

[50] P.J. Mulhern, T. Hubbard, C.S. Arnold, B.L. Blackford and M.H. Jericho, "A scanning force microscope with a fiber-optic-interferometer displacement sensor", *Rev. Sci. Instrum.*, **62**(5), p. 1280-1284, (1991).

[51] Unpublished research, L.P. Howard, 1990.

[52] D. Sarid, V. Weissenberger, D.A. Iams and J.T. Ingle, "Theory of the Laser Diode Interaction in Scanning Force Microscopy", *IEEE J. Q. Elec.*, **25**(8), p. 1968-1972, (1989).

[53] D. Sarid, D.A. Iams, J.T. Ingle, V. Weissenberger and J. Ploetz, "Performance of a scanning force microscope using a laser diode", *J. Vac. Sci. Technol. A*, **8**(1), p 378-382, (1990).

[54] U.S. patent number 5,025, 658; 25 June, 1991.

[55] D. Sarid, P. Pax, L. Yi, S. Howells, M. Gallagher, T. Chen, V. Elings and D. Bocek, "Improved atomic force microscope using a laser diode interferometer", *Rec. Sci. Instrum.* **63**(8), p 3905-3908, (1992).

[56] M. Tortonese, H. Yamada, R.C. Barrett and C.F. Quate, "Atomic Force Microscopy using a Piezoresistive Cantilever", unpublished document, 1992.

[57] T. Göddenhenrich, H. Lemke, U. Hartmann and C Heiden, "Force microscope with capacitive displacement detection", *J. Vac. Sci. Technol. A*, **8**(1), p 383-387, (1991).

[58] G. Neubauer, SR Cohen, GM McClelland, D Horne and CM Mate, "Force microscopy with a bidirectional capacitance sensor", *Rev. Sci. Instrum.*, **61**(9), p 2296-2308, (1990).

[59] H. Kaizuka and B. Siu, "A simple way to reduce hysteresis and creep when using piezoelectric actuators", *JJAP*, **27**(5), pp. L773-L776, (1988).

[60] R. Young, J. Ward and F. Scire, "The Topographiner: An Instrument for Measuring Surface Topography", *Rev. Sci. Instrum.*, **43**(7), p 999-1011, (1972).

[61] G. Bennig and D. Smith, "Single-tube three-dimensional scanner for scanning tunelling microscopy", *Rev. Sci. Instrum.*, **57**, p 1688-1689, (1986).

[62] F.E Scire and E.C Teague, "Piezodriven 50 μm range stage with subnanometer resolution", *Rev. Sci. Instrum.*, **49**(12), p 1735-1740, (1978).

[63] Y. Xu and S.T. Smith, private communication, (1993).

[64] S.T. Smith and D.G. Chetwynd, "Optimisation of a magnet/coil force actuator and its application to linear spring mechanisms", *Proc. Inst. Mech. Engrs*, **204**(C4), pp. 243-253, (1990).

[65] A.H. Slocum, Precision Machine Design, Prentice Hall, Englewood Cliffs, NJ, 1992, pp. 530-532.

[66] Arthur R. Von Hippel, Dielectrics and Waves, The M.I.T. Press, Cambridge, Mass, 1954, p. 236.

[67] W. Tang, T. Chuong, H. Nguyen, W. Judy and R. Howe, "Electrostatic comb drive of lateral polysilicon resonators", *Sensors and Actuators A*, **A21-A23**, p 328, 1990.

[68] A.E. Clark, J.P. Teter and O.D. McMasters, "Magnetostriction 'jumps' in twinned Tb0.3Dy0.7Fe1.9", *J. Appl. Phys.*, **63**(8), p. 3910-3912, (1988).

[69] D.J. Whitehouse, "A revised philosophy of surface measuring systems", *Proc. Inst. Mech. Engrs*, **202**(C3), p 169-185, (1988).

[70] K.E. Peterson, "Silicon as a mechanical material", *Proc. IEEE*, **70**(5), p 420-456, (1982).

[71] T.R. Albrecht, S. Akamine, T.E. Carver and C.F. Quate, "Microfabrication of cantilever styli for the atomic force microscope", *J. Vac. Sci. Technol. A*, **8**(4), p 3386-3396, (1990).

[72] Ultra Levers™, Park Scientific Instruments, 1171 Borregas Ave., Sunnyvale, CA, 94089.

[73] W.P. Maszara, "Silicon-on-insulator by wafer bonding: A review", *J. Electrochem. Soc.*, **138**(1), p 341-347, (1991).

2 Profiler design considerations and force probe system

Chapter summary

In this chapter the justifications for, and the conceptual design features of the constant force profiler are introduced. Metrology systems for such an instrument are identified as the force probe, the capacitance gage and the laser interferometer. One of these systems, the force probe and servo control system, will be examined in detail. Probe design parameters are first considered and then the production and assembly of two generations of force probe are presented. Servo controller construction and tuning are next presented. Effects of the probe/specimen interaction on measurement dynamics are introduced. These interactions will be revisited in Chapter 7 with further theoretical background and experimental results.

Force profiler justifications

Having examined the instrumental aspects of mechanical contact surface metrology in the introductory chapter, it is concluded that stylus instruments and AFMs differ in four main aspects: Contact force, spatial wavelength capability, bandwidth (or speed of operation), and surface amplitude. A comparison between the two instruments is summarized in table 2.1.

	Stylus Instruments	AFMs (repulsive mode)
Force	10 μN to 1 mN	1 nN to 1 μN
Spatial Wavelength	10 μm < λ < 100 mm	1 nm < λ < 100 μm
Bandwidth	4 Hz < f < 30 Hz	≈300 Hz
Amplitude	0.1 nm < y < 100 μm	1 pm < y < 1 μm

Table 2.1. Summary of stylus instrument and atomic force microscope characteristics.

The force profiler was conceived to bridge the gaps that exist between these operational parameters, giving stylus users a faster, lower force instrument and AFM users a longer range.

The crux of the problem with contact surface metrology instruments is that AFMs simply can't yet measure the long (>0.8 mm) spatial wavelengths needed for engineering materials. On the other hand, stylus profilers can make these form measurements, but often are very slow or are capable of damaging the sample. With stylus instruments, the dynamic forces at the tip/specimen interface are always pushing against the restoring forces of the stylus tracking arm. If the traverse speed is increased (thus increasing the dynamic forces at the tip), the stylus may bounce if the dynamic forces at the tip exceed the restoring forces of the tracking arm. On the other hand, the restoring force can be increased (to prevent bouncing) at the risk of damaging the specimen. Neither option is an especially desirable one. The force profiler was designed to eliminate this "lose-lose situation".

Force profiler concept

The force profiler is a hybrid blend of both AFM and stylus profiler instrument technology. A generalized diagram of the complete instrument is illustrated in figure 2.1. The force sensor, holding an ISO standard stylus tip is located on a PZT actuator.

The output from the sensor is fed back through a PID servo controller to maintain a constant applied force on the force probe by moving the PZT actuator in response to changes in surface topography. This mode of operation is hereby referred to as the repulsive mode. All subsequent profiles have been measured with a net repulsive force between 10^{-5} and 10^{-7} Newtons (1 to 0.01 mgf). No attempt is made to operate the instrument in the attractive mode due to problems associated with surface contaminant layers. The repulsive mode is very similar to low-force stylus operation in that contaminant layers may be pushed out of the way (if sufficient pressures are generated at the point of contact) and the true mechanical surface is followed. A variety of metrological sensors are united in the force profiler to provide an instrument with a high precision and reproducability, figure 2.1.

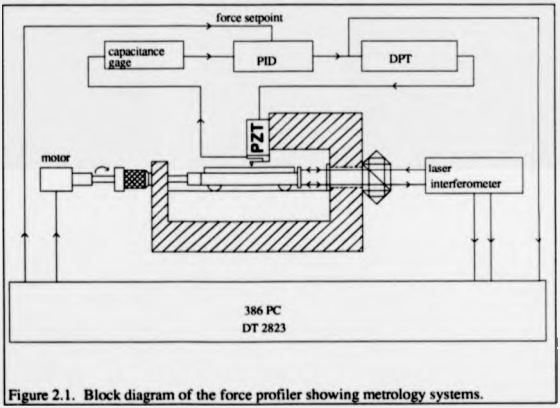


Figure 2.1. Block diagram of the force profiler showing metrology systems.

As shown in the figure, the force sensor utilizes a capacitance-based transducer to measure deflections of an AFM-type cantilever spring. A control system is used to maintain a constant deflection, and therefore applied force at the probe tip. Cantilever

deflections are monitored by a high resolution capacitance gage. The design and operation of this gage will be presented in Chapter 3. Specimen translations on a high-precision carriage and slideway system are monitored by a distance measuring interferometer, described in Chapter 4. Integrating all these parts together is a data acquisition system and control software, which will be examined in Chapter 5.

Constant force servo system

From figure 2.1, it is shown that the complete force metric consists of four main components: The force probe, a capacitance sensor, a servo-controller and a piezoelectric actuator.

Actuator

For both practical and metrological purposes, it is desired that the actuator system meet the following requirements: Resonant frequency at least several hundred Hz, stroke length of at least 10 micrometers and stable, low- drift operation, free of hysteresis and/or creep. Hysteretic effects and creep are very significant in piezoelectric actuators capable of stroke lengths longer than a few micrometers. Generally, the low voltage, high expansion stacked piezoelectric materials used for this application are more susceptible to such effects than high voltage, low expansion ceramics. One solution to these problems is to provide an integral displacement sensor in the piezoelectric actuator to monitor it's extension, and then use this sensor in a closed loop servo control system to remove hysteretic and creep effects from thesystem. This method provides virtual separation of the force and sensing, or metrology, loops and increases the virtual stiffness of the actuator. Increase the load on such an actuator and the control system compensates to maintain a constant piezo extension.

Queensgate Instruments offers such an actuator system (known as the Digital Piezoelectric Actuator, or DPT) in an easy to use package, shown schematically in figure 2.2.

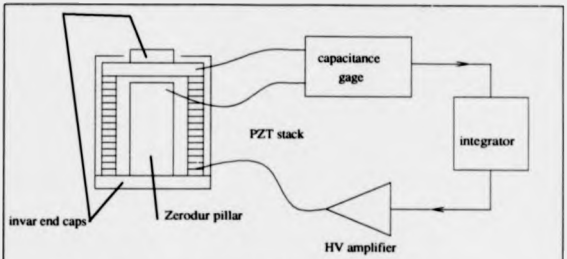


Figure 2.2. Queensgate Instruments DPT actuator system.

An internal capacitance gage located coaxially with the piezoelectric stack measures extension between the end pieces. The Zerodur pillar supporting one electrode transfers the datum from the back end of the DPT where the piezo stack is anchored. This use of low thermal expansion materials results in a low thermal expansion coefficient for the unit , about 15 nm/°C. An integrator in the controller unit closes the loop between the capacitance gage and the piezoelectric stack, with the demand input to the servo controller being the DPTs extension input. All the user need do is supply DC power to the unit and input a demand signal. Specifications for the DPT-C actuator used in the profiler are as follows: 15 μm maximum extension with 0.5 nm RMS noise. Assuming a Gaussian noise distribution, the peak to peak noise excursions are probably 2 to 3 nm. Slew rate is specified as 1 μm/μs, limited by amplifier current. Since the piezoelectric stack in the DPT behaves electrically as a large capacitor, fast excursions place large demands on amplifier current, recalling

$I = C\frac{dV}{dt}$. Also, the integrator in the DPT servo system acts as a low-pass filter, adding phase lag increasing with frequency. The specified -3 dB (half power) cutoff for the unit occurs at approximately 300 Hz. With the added mass of the force metric, this bandwidth can be expected to be slightly lower than the specified, unweighted value.

Probe requirements

There were five main requirements for the force transducer and probe head. They were, in no particular order:

1. The complete sensor must be small enough for mounting on a piezoelectric actuator to be servoed up and down to follow the contours of the specimen surface. A large probe would add mass to the moving system and reduce it's response, slowing the allowable traverse speed and increasing measurement time. The resonant frequency of the system should be at least several hundred Hertz.

2. No alignment or adjustments should be necessary to the metric once installed. No adjustment screws or other alignment methods are permissible on a moving metric system.

3. The probe should be large enough to hold a standard profiler stylus as well as AFM type 'atomic tips'. A typical stylus shank is 0.5 mm diameter by 2.0 mm in length. The stylus can be bonded to the force sensor.

4. The range of forces sensed must span the forces typically encountered in both stylus instruments and AFMs. Thus, force range must vary from 10^{-7} N to 10^{-4} N (one milligram force is approximately 10^{-5} N), bridging the gap between AFMs and stylus profilers.

5. The probe should be quickly interchangeable for scanning using different styli. A collection of probes ranging from ISO standard 2, 5 and 10 μm spherical diamond to 20 nm cleaved tetrahedral, diamond "Berkovich" tips.

6. Obviously, (and all too often overlooked) the probes must be manufacturable from available resources.

Choice of metric and materials

The first criteria, that the force sensor must be small and light enough to mount on the end of a piezoelectric actuator, eliminates all inductive AC sensors. Inductive coils and cores simply involve too much mass for rapid piezo scanning. Interferometry, although capable of high-resolution and bandwidth does not provide a linear output signal, (unless operating around the quadrature point) and would complicate the control system, making a simple analog servo impossible. The capacitance bridge method was one of the few options available which required only a negligible increase in mass while allowing sub-nanometer resolution. Furthermore, both the Author and the researchers at the University of Warwick had gained confidence in the capacitance gaging technique through experimentation in the laboratory. A commercial capacitance gaging circuit could be used for initial trials and later, custom circuitry could be designed and implemented. The capacitance bridge used in the initial experiments[1] will be discussed in much greater detail along with the custom-built electronics supplied by the Author, in Chapter 3.

The probe system is constructed mostly of brittle materials. Zerodur[2], an ultra low expansivity glass-ceramic was chosen for the probe base because of its low thermal expansion, relative ease of machining and good dielectric properties. Aluminum or chromium electrodes are evaporated onto the base and cantilever to provide parallel plate capacitors and contacts for connection to external electronics. The cantilever is machined from a thin (100 μm) piece of microscope cover glass which is subsequently covered with a vacuum deposited electrode. The reason for choosing a brittle material

such as glass instead of a more conventional material such as BeCu is that for metrology purposes, a brittle material performs nearly perfectly, or not at all[3]. Damage to the cantilever will usually be catastrophic, forcing it to be discarded. To a metrologist, the only thing worse than no measurement is a bad measurement.

Probe design and modeling

There are four critical dimensions of length, width, beam thickness and gap distance which effect the following parameters: Spring constant, resonant frequency, damping and nominal capacitance. For the prototype probe, discussed here, the cantilever dimensions were 10 by 2 by 0.1 mm with a 10 µm electrode gap distance.

Dynamic deflection of the cantilever beam can be modeled as a second-order system with the governing equation given by

$$m_e\ddot{x} + d_c\dot{x} + k_c x = 0 \qquad (2.1)$$

where m_e is the effective mass of the cantilever/tip system, d_C is the damping due to the thin squeeze film (air) between the cantilever electrodes, and k_C is the spring constant of the cantilever. The effective mass of a uniform cantilever beam with a concentrated mass at the end is[4]

$$m_e = m_{tip} + 0.24 m_{cantilever} \qquad (2.2)$$

Stylus mass was determined to be approximately 1 mg. The mass of the cantilever is determined by volume

$$m_{cantilever} = \rho BHL \qquad (2.3)$$

where B, H, L, and ρ are the cantilever width, thickness, length between tip and clamped end of the beam, and density, respectively. The spring constant for a uniform cantilever beam is

$$k_c = \frac{EBH^3}{4L^3}, \tag{2.4}$$

where E is the Young's modulus of the material (in our case borosilicate glass). The cantilever spring is then assumed to follow Hooke's law in that applied forces may be measured by monitoring beam deflection, δ . Fractional errors of this assumption are estimated by using the chain rule

$$\frac{dF}{F} = \frac{dE}{E} + \frac{dB}{B} + 3\frac{dH}{H} + \frac{d\delta}{\delta} - 3\frac{dL}{L}. \tag{2.5}$$

It may be seen that errors in estimating cantilever thickness and position of the stylus are of great importance. Note that this model assumes a uniform stress-free beam with zero strain at the boundaries. Damping may be approximated using the theory of unsteady, self-acting, rectangular, squeeze film bearings. Ignoring the small bending of the beam, solving the Reynold's equation[5]

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial z^2} = \frac{12\mu}{h^3} V \tag{2.6}$$

for reaction force, and then dividing by velocity V, we obtain the cantilever damping constant d_c

$$d_c = -\left(\frac{\mu B^3 L}{h_o^3}\right)\left[1 - \frac{192B}{\pi^5 L}\sum_{n=1,3,5...}^{\infty} n^{-5}\tanh\left(\frac{n\pi L}{2B}\right)\right], \tag{2.7}$$

where h_o is the air gap which is taken as the step height on the Zerodur blank and μ is the absolute viscosity of air. The minus sign indicates that the damping force is opposed to the relative motion of the cantilever. The reader should note that this model is not especially accurate, as it assumes the motion of parallel plates. A more accurate model with much more experimental data is currently being developed by another worker at the University of Warwick.

Choosing cantilever dimensions involves other considerations as well. In the initial prototype, a capacitance of about 20 pF was required for use with the bridge

electronics. More recent versions have reduced this value to about 10 pF. Given a static condition with no force acting on the sensor tip, the capacitance C_o is given by

$$C_o = \frac{\varepsilon BL}{h_o} \cong 20pF. \tag{2.8}$$

Finally, the natural, undamped frequency of the cantilever is

$$\omega_n = \sqrt{\frac{k_c}{m_e}}. \tag{2.9}$$

Equations 2.4, 2.7, 2.8 and 2.9 are used to optimize probe response for the desired goals of high capacitance, near critical damping, low spring constant and high resonant frequency.

Linearity

To examine the linearity of the cantilever capacitance force probe, consider a uniform beam of length L with a point load of force F on the end. At a distance x away from the fixed base of the beam we have a deflection

$$\delta(x) = \frac{Fx^3}{6EI}(3L - x). \tag{2.10}$$

By using a square law approximation for beam curvature[6], which is mostly valid at the end,

$$\delta(x) \cong \left(\frac{x}{L}\right)^2 \frac{F}{k}. \tag{2.11}$$

The gap between the electrodes as a function of x is

$$h(x) = h_o - \left(\frac{x}{L}\right)^2 \frac{F}{k}. \tag{2.12}$$

By using Eq. (2.12) and integrating over the length of the beam we find the capacitance as a function of force

$$C(F) = \int_0^L \frac{\varepsilon B}{h_o - (x/L)^2 (F/k)} dx \tag{2.13}$$

we have

$$C(F) = \frac{\varepsilon B}{a} \tanh^{-1}\left(\frac{La}{h_o}\right), \text{ where } a = \sqrt{\frac{h_o F}{L^2 k}} \tag{2.14}$$

Since Tanh^{-1} is linear for small values relevant to the probe designs, the sensor is linear for applied deflection up to about 1 micrometer and forces up to over 10^{-5} N, figure 2.3.

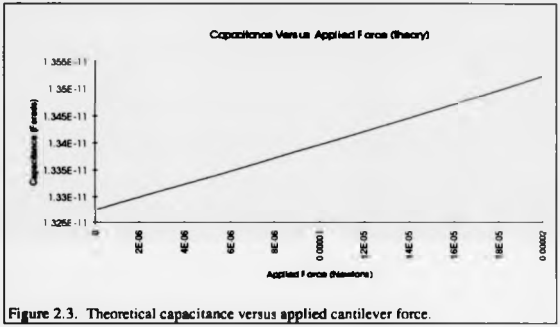


Figure 2.3. Theoretical capacitance versus applied cantilever force.

Experience with these sensors has proven this to be a good estimate. From the derivative of the force/capacitance relation, a typical sensitivity of 2.7×10^{-7} Farads per Newton is found.

Sensitivity may be improved (at the expense of a smaller total capacitance) by reducing the length of the base electrode to one quarter of it's former length and placing it under the end of the cantilever. Computing the integral, for small displacements the change in capacitance for an applied force is reduced by only 25% while total capacitance is

reduced 75%. The resulting increase in sensitivity is a factor of three improvement, at the cost of a capacitance as small as 2 or 3 pF.

One possible modification of the probe system worth discussing is depicted in figure 2.4.

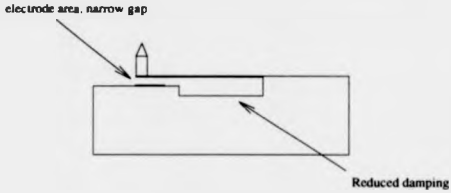


Figure 2.4. Possible modification of probe base to provide greater sensitivity and reduced damping.

As illustrated in the figure, the probe base has been etched twice to produce a narrow air gap with capacitance electrodes at the tip of the cantilever, while increasing the gap seperation along the remainder of the cantilever. The effects of such a modification are thus threefold.

1. The electrodes at the tip of the cantilever more closely approximate parallel plates, making damping calculations more accurate.

2. The overall squeeze film damping is reduced by increasing the gap distance for much of the cantilever. Reduced damping allows the use of thinner gaps in the capacitance electrode region at the cantilever tip, increasing sensitivity.

3. Sensitivity may be further increased due to the greater fractional change in sensor capacitance provided by the enhanced geometry.

Electrostatic effects

The effects of the capacitance gage excitation voltage on the motions of the cantilever are considered by simple electrostatics. For a parallel-plate capacitor with motion perpendicular to the plane of the electrodes, the potential energy of the system is differentiated with respect to electrode spacing to obtain the relation for force

$$U = \frac{1}{2}CV^2 \;,\; F_{drive} = \frac{\partial U}{\partial h} = -\frac{eBL}{2}\frac{V^2}{h^2}. \tag{2.15}$$

Summing this force with the equation for cantilever motion

$$m_c\ddot{x} + d_c\dot{x} + k_c x = F_{drive}\sin(2\omega_{drive}t). \tag{2.16}$$

We solve for the displacement due to the Queensgate Instruments capacitance bridge's 10 $V_{p\text{-}p}$ drive signal at 20 KHz. The electrostatically induced peak displacement is approximately 3 pm at 40 KHz, well above our sensor bandwidth. The RMS value of this excitation will induce a small static displacement of the cantilever. Since this displacement can not be monitored without the aid of the capacitance gage, no motion can be observed.

Snapping is the description of an event which occurs when the electrostatic attractive forces overcome the restoring spring force of the cantilever. Thus, snapping limits the maximum excitation voltage which may be utilized by the capacitance bridge. Summing the electrostatic and spring forces

$$\frac{eBL}{2}\frac{V^2}{h^2} - k_c(h_o - h) = 0 \tag{2.17}$$

and solving for voltage

$$V = \left[\frac{2k_c}{eBL}h^2(h_o - h)\right]^{1/2}. \tag{2.18}$$

Snapping occurs when the first derivative of this function equals zero

$$h = \frac{2}{3} h_0. \tag{2.19}$$

Substituting the latter into the former, the snapping voltage is then

$$V_s = \left[\frac{8}{27} \frac{k_c'}{\varepsilon B L} h_0^3 \right]^{1/2} \tag{2.20}$$

where k_c' is computed for a cantilever beam under uniform, not point loading. Thus,

$$k_c' = \frac{2EBH^3}{3L^4}. \tag{2.21}$$

Thermal noise

The thermal noise, or Brownian motion of system is only applicable in the case of a freely supported cantilever, not in contact with a surface. Since sensor noise levels are often estimated in this fashion however, a determination of the minimum noise levels achievable is useful. Brownian motion, or thermal noise is estimated by equating the potential energy of the cantilever spring with it's thermal energy via the equipartition theorem

$$\frac{1}{2} k_c \delta^2 = \frac{1}{2} k_{Boltzman} T \tag{2.22}$$

so that the RMS deflection of the cantilever is given by

$$\delta = \sqrt{\frac{k_{boltzman} T}{k_c}}. \tag{2.23}$$

About 11 pm of thermal noise is computed for a 37 N/m cantilever. A recent summary of thermal uncertainty is given by Drexler [7] in the context of miniature systems.

Mechanics of contact

Since during repulsive mode AFM operation the stylus probe is in contact with the surface, the stiffness of the contact must be accounted for in any modeling.

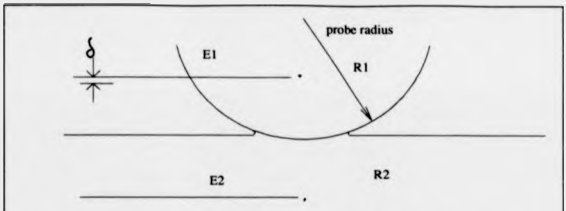


Figure 2.5. Schematic of a spherical stylus in contact with a flat surface.

The mutual approach of two distant points in the probe and specimen is given by δ, from Johnson[8]

$$\delta = \left(\frac{9}{16} \frac{P^2}{RE^{*2}} \right)^{\frac{1}{3}} \tag{2.24}$$

Where P is the applied force and radius of mutual contact, the effective radius R and effective Young's modulus E^* are

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} \text{ and} \tag{2.25}$$

$$\frac{1}{E^*} = \left(\frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2} \right) \tag{2.26}$$

where ν_1 and ν_2 are Poissons ratios and E_1 and E_2 the Young's moduli of the two materials. For the case of a diamond spherical tip (1) on a silicon flat (2), the radius R_1 is considered infinite and it's Young's modulus much larger than silicon. Solving Eq. 2.24for force

$$P = \frac{4}{3} \delta^{\frac{3}{2}} R^{\frac{1}{2}} E^* . \tag{2.27}$$

Assuming that for small displacements, the contact stiffness k_s is given by the derivative of force with respect to displacement δ (Hooke's law), we have

$$k_s = \frac{dP}{d\delta} = 2\delta^{\frac{1}{2}} R^{\frac{1}{2}} E^* . \tag{2.28}$$

Substituting Eq. 2.24 in for δ, we obtain a solution for contact stiffness as a function of force, geometry and Young's modulus

$$k_s = 2\left(\frac{9}{16}\right)^{\frac{1}{6}} P^{\frac{1}{3}} R^{\frac{1}{3}} E^{*\frac{2}{3}} . \tag{2.29}$$

Substituting x_1 for δ in eq. 2.28 and summing spring potential energies, the equation of motion (neglecting damping) for the cantilever beam force probe in contact with the specimen becomes

$$m_e \ddot{x} + k_c x_1 + \alpha x_1^{3/2} = 0 \tag{2.30}$$

where $$\alpha = \frac{4}{3} R^{1/2} E^* , \tag{2.31}$$

m_e is the equivalent mass of the cantilever/probe system and k_C is the spring rate of the cantilever. This non-linear equation may be solved for small displacements δ about a nominal indentation distance x_1. Rewriting the coordinates as

$$x = x_1 + \delta , \tag{2.32}$$

$$m_e \ddot{x} + k_c x + \alpha (x + \delta)^{3/2} = -k_c \delta \tag{2.33}$$

Expanding the term within parenthesis with a binomial series yields

$$\delta^{3/2} + \frac{3}{2}\delta^{1/2} x_1 + \ldots \tag{2.34}$$

which when substituted into (2.33) gives the equation of motion

$$m_e \ddot{x} + k_c x + (3/2)\alpha\delta^{1/2} x = -\left(k_c \delta + \alpha\delta^{3/2}\right). \tag{2.35}$$

Since the right hand side of this equation is a constant, the undamped natural frequency of the system is

$$\omega_n = \sqrt{\frac{k_c + (3/2)\alpha\delta^{1/2}}{m_e}} \tag{2.36}$$

which, when rearranged gives

$$\omega_n = \sqrt{\frac{k_c}{m_e}\left(1 + \frac{9P^{1/3}\left(E^*R\right)^{2/3}}{4k_c}\right)^{1/2}} \tag{2.37}$$

The implications of this relation are many. First, the contact stiffness will tend to increase the resonant frequency of the force metric. Second, if the contact stiffness is commensurate with the force metric stiffness, a large effect may be ascertained during measurements. Examining the term in parenthesis, contact radius and cantilever stiffness are fixed constants. Contact force may be set by servo control to be constant, leaving the variable for Young's modulus affecting the resonant frequency. Thus the prospects for a *scanning Young's modulus microscope* are predicted. If the force probe cantilever stiffness is commensurate with (or dominated by) the contact stiffness, a Bode response may be determined, showing magnitude and phase shifts of the system for contact with different materials. Results of such experiments will be presented in the Chapter 7.

Control system

The control system uses the capacitance gage signal as an input to extend or contract the DPT, maintaining a constant deflection of the probe cantilever. In this manner, the probe follows a contour of constant, net repulsive force. The advantages of such a servo controlled scheme are many. First, since the force metric operates with only small deflections about a fixed setpoint, the requirements for linearity are relaxed. The force metric must only be linear for small displacements about a fixed operating point. Second, the range of operation is limited by the actuator, not the force metric. The

dynamic range of the force probe need only be large enough to handle small swings in the error signal. Third, the system's dynamic character is primarily determined by the type of, and settings used on the control system. Parameters such as rate dependent damping may be altered electronically, rather than mechanically.

Since the system must be fairly immune to steady-state errors, an integrator is used with a proportional term in the servo loop. Rate dependent damping is provided by a differentiator followed by a rate limiter designed not to exceed the slew-rate of the DPT. Given an error signal, e derived from the difference between the actual force derived from the sensor output voltage and the demand force setpoint,

$$e = V_{demand} - V_{actual} \tag{2.38}$$

In the time-domain the resulting PID controller has the form

$$V_{out} = G_P e + G_I \int e\,dt + G_D \frac{de}{dt} \tag{2.39}$$

where G_P, G_I and G_D are gain coefficients for the P, I and D stages respectively. Generally, another parameter, the time constant, T_C, is defined as the reciprocal of the integrator gain. For initial setup, the integrator time constant was chosen to be equal to, or slower than the maximum slew-rate of the DPT actuator, 1 μm/μs.

Electronically, the PID controller is a very simple design, implemented using three parallel stages. One for each control action, figure 2.6.

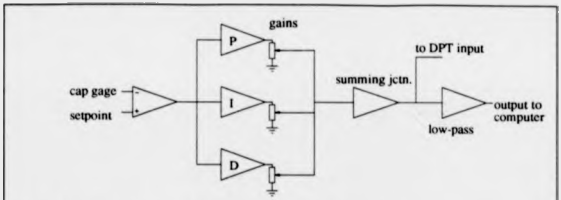


Figure 2.6. Block diagram of three-term controller electronics.

For precision purposes, the error signal is developed by a Analog Devices AD620 single-chip instrumentation amplifier. Using an amplifier such as this with high common-mode rejection ensures a highly linear subtraction of the setpoint voltage from the force probe signal. From the instrument amp, the error signal branches into three, parallel stages. One each for proportional, integral and derivative actions. The gains and time constants are fixed at their maximum values, allowing the signals to be attenuated and summed independently without introducing additional phase shifts. Proportional gain is set to a maximum of 100, integrator time constant is set to the slew-rate of the DPT and the differentiator is rate-limited to 500 Hz. Ten-turn potentiometers allow for independent adjustment of each term, summing to an output stage driving the DPT. Additional output for the data acquisition system is passed through a built-in adjustable, single-pole, low-pass filter with cutoff frequencies from 50 Hz to 800 Hz. This filter cutoff is nominally set to equal the closed-loop servo bandwidth. A complete schematic diagram of the servo controller appears in the Appendix.

Although there are many theoretical discourses that can be made on the subject of tuning a PID controller, the time-tested method of Ziegler and Nichols gives acceptable results upon first application to a stable system such as this. Briefly, the system is first set so that the force metric gains and actuator gains nullify each other. In other words, if the controller were removed and replaced by a straight wire, the

system gain would be unity. The integral and derivative term gains are minimized and the proportional gain is then increased from zero until oscillation occurs. When oscillation is achieved, the gain is noted as G_{pc} and the period of the oscillations as T_c. The gains are then set as follows[9]:

G_P	G_I	G_D
$0.6G_{pc}$	$2.0/T_c$	$T_c/8$

These gains are only a first approximation, however. Further adjustments are best made by applying a disturbance signal to the controller and measuring the response of the system, taking into account all the phase lags of each component , as shown in figure 2.7.

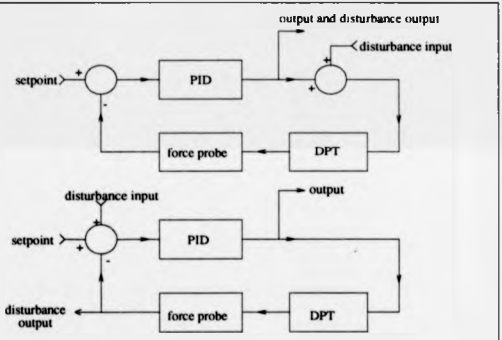


Figure 2.7. Control systems with disturbance inputs for fine-tuning.

Using the disturbance input, step inputs may be tuned for maximum fidelity. A still further adjustment is usually required of the differentiator to minimize system noise

while retaining the desired value of damping. The above procedures were usually found to produce a damping coefficient of 0.6-0.7, with no *a priori* knowledge. This is, in effect the reverse of the process espoused by Whitehouse[10], setting the controller for maximum fidelity and then measuring the damping coefficient. Also, a swept sine waveform from a spectrum analyzer may be applied to the disturbance input and a closed-loop transfer function generated for the entire force metrology servo system. Results of such plots will be presented later in the performance evaluation, Chapter 6.

Prototype force probe

The first probe design was based on the requirement that the probe's nominal capacitance be approximately 20 pF, as dictated by the use of a Queensgate Instruments capacitance bridge. Rapid-replacement of the probe was not of any concern in the prototypes. A 20 by 10 by 6 mm Zerodur flat was prepared for the prototype probes. The cantilevers used were 10 mm long by 2 mm wide with a large 4 by 4 mm square base for contact to the Zerodur base. The probes were machined out, two at a time (see construction section) from a 100 μm thick borosilica glass plate 25 mm square, figure 2.8.

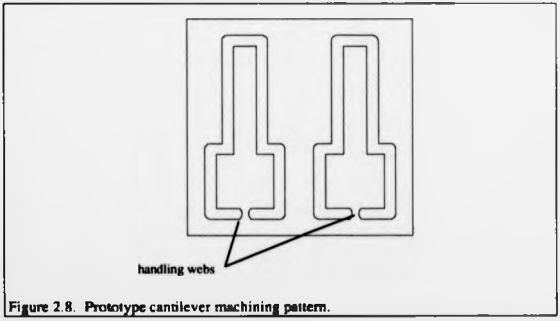


Figure 2.8. Prototype cantilever machining pattern.

The complete prototype force sensor is shown in figure 2.9.

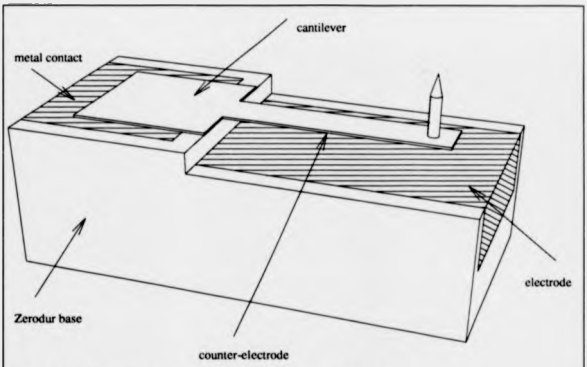


Figure 2.9. Prototype probe showing arrangement of electrodes.

The capacitor is formed by the gap between the electrode on the Zerodur base and the counter-electrode on the facing side of the cantilever. Wires were attached with silver epoxy to the electrodes overlapping the sides of the base. The entire apparatus was then attached to the DPT with two-part epoxy.

Manufacture and construction of an improved force probe

The second-generation probe system consists of four main mechanical components.

- The rapid replacement probe holder,
- a Zerodur probe base,
- a glass cantilever electrode and

- a stylus probe.

These components are depicted depicted in figure 2.10 showing the force probe loaded into the rapid replacement housing.

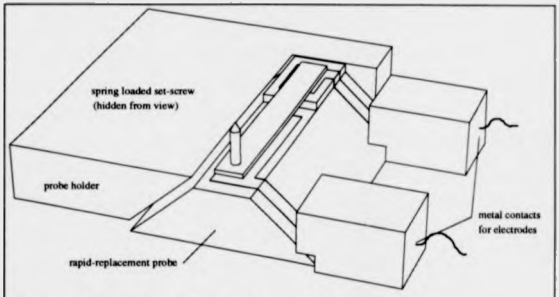


Figure 2.10. Improved force probe and probe holder.

The rapid replacement probe holder performs the dual functions of psuedo kinematically locating the probe base and making electrical contact to the capacitance electrodes. Made of Zerodur for stability, the holder base attaches to the DPT actuator by a single, centrally located M3 screw. The probe holder and contacts are, in turn, bonded to the holder base. The holder base attachment screw is recessed to allow the rear face of the probe base to contact the holder base on a flat, lapped datum. Two rounded contacts serve as stops and electrode wire attachment points for the capacitance gage. The probe base is pushed against these stops by a M3 screw with a spring-loaded ball-ended-plunger pushing against the opposing side. To change probes, the spring plunger is backed off a few turns and the probe base simply pulled out. Installation is simply the reverse, with the whole process taking only about 30 seconds. The Zerodur holder base is first sawed from rod stock and carefully lapped

until a flatness of about one micron is achieved. High speed diamond machining with a NSK Planet 500 air-motor-spindle grinds the center hole and screw recess. Spindle speeds are on the order of 50,000 to 100,000 RPM. For this operation the Zerodur workpiece is fixtured to a piece of float glass with optical blocking wax. Standard water soluble oil coolant is used for all glass grinding operations. The aluminum probe holder parts are attached to the Zerodur holder base by cyanoacrylate adhesive.

The probe base is batch processed so that some 20 to 30 probe bases can be fashioned from a single 75 by 100 mm, 3.5 mm thick slice of Zerodur. The Zerodur is first lapped on both sides and then optically polished to $\lambda/4$ on the electrode face. The plateaus are then formed by chemical etching. Strips of etch tape (vinyl electrician's tape works very well) are placed across the Zerodur plate and the whole piece etched in a solution of 400 ml HF(48%) and 200 ml HCl (35.4%) to a depth of 10 μm (about a 30 second etch time). Tape is then removed and the Zerodur affixed to a piece of float glass by optical wax. Using a slicing saw, the probe bases are then cut into rows. Columns are formed by cutting with a 90 degree diamond vee saw so that the bases are just separated from one another, figure 2.11.

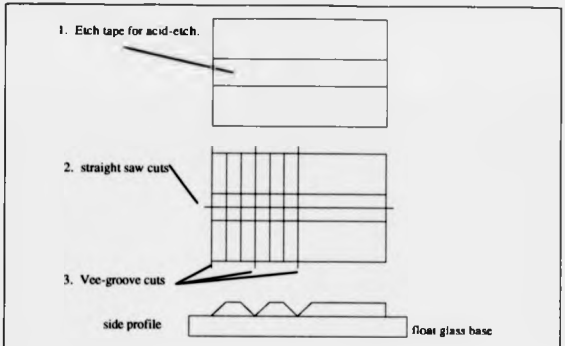


Figure 2.11. Probe base manufacture showing a small section of a larger process.

Heating the float glass on a hot plate frees the probe bases.

Cantilevers are machined from thin, glass cover slips. First, No. 0, 100 μm thick cover slips are inspected with a horizontal profile projector to determine the axis of warp. Most glass cover slips are hot-rolled and have a warp along one axis. Aligning the warp axis so that it is coincident with the minor (width) axis of the cantilevers, not the length, is a very important first step. The cover slips are then fixtured to a piece of float glass by soft, red dental modeling wax. This wax has a very low melting temperature and may be removed in hot kerosene or petrol. A new, albeit untried fixturing method could be Loctite® Cold Block™ 3791[11], a UV cured, water soluble blocking compound. Use of this substance would eliminate the need for solvent cleaning. A 1 mm diameter diamond inside grinding pin is turned by an air motor spindle (which is, in turn, mounted on a CNC milling machine) with a rotary speed of 100,000 RPM. Feed rates used for cantilever cutting are approximately 2 to 3

mm/min. 0.5 mm webs are left on the cantilever ends for later removal. The cover slips are heated and removed from the float glass and cleaned in petrol to remove wax residue. Standard acetone, methanol and DI water cleaning then precedes the vacuum deposition of 100 nm aluminum on both sides.

Mating the cantilever to the probe base is the most critical step in the assembly process. The roughly 10 μm gap between the cantilever and probe base must be free of dust and foreign matter. Also the cantilever must lie parallel to the facing side of the etched step on the base, not touching the opposing electrode. For these reasons, the probe base is fixtured in a horizontal profile projector and the cantilever maneuvered into place while observing the condition of the gap. When in place, the cantilever is held in place by a small weight, then permanently bonded with a small fillet of epoxy along the plateau. As a final check, the probe's capacitance is measured on a HP4284A LCR bridge. With a little practice, high yields may be achieved.

A final step, the attachment of a stylus, involves carefully bonding the stylus onto the tip of the cantilever with a small drop of epoxy.

---

[1] Nanosensor is manufactured by Queensgate Instruments Ltd., Silwood Park, Ascot, Berkshire, SL5 7PW, England.

[2] Zerodur is a registered tradename of Schott Glass Technologies Inc., 400 York Ave., Durea, PA, 18642, USA.

[3] D.K. Bowen, D.G. Chetwynd and DR Schwarzenburger, 'Sub-nanometre displacements calibration using X-ray interferometry', *Meas. Sci. Technol.*, 1, pp. 107-119, (1990).

[4] Cyril M. Harris and Charles E. Crede, eds., Shock and Vibration Handbook, 3rd ed., McGraw-Hill, NY, 1988.

[5] W.A. Gross, Gas Film Lubrication, Wiley and Sons, NY, 1962, p. 318.

[6] K.E. Peterson, *IEEE Trans Electron Dev.*, ED-25, p. 1241, (1992).

[7] K.E. Drexler, Nanosystems. John Wiley and Sons, NY, 1992.

[8] K.L. Johnson, Contact Mechanics. Cambridge University Press, NY, 1985.

[9] W. Bolton, Industrial Control and Instrumentation. Longman Group UK Ltd., 1991, p. 117.

[10] D.J. Whitehouse, "A revised philosophy of surface measuring systems", *Proc Instn Mech. Engrs.*, 202(C3), pp. 169-185, (1988).

[11] Distributed by Optical Manufacturers International Ltd., Romeoville, IL., USA, 1-800-323-7157.

3 DISTANCE MEASURING CAPACITANCE METROLOGY

Chapter summary

Since capacitance gaging is the metric of choice for the force profiler, a thorough review of the fundamentals of capacitance measurement is useful. Also, an examination of the prior art will be made before describing the development of a capacitance measurement circuit capable of fulfilling the needs of the probe's force metric.

Introduction

The use of the capacitor as a displacement measuring transducer has a long history, dating back at least to Villey[1] (1910). In the years since, it has become a widely used transducer not only for displacement measurement, but for fluid mixture measurement and industrial process control. Capacitance gaging is generally recognized as an excellent transduction method readily providing range over resolution on the order of 50,000 to 1. The resolution of the capacitance gage can easily exceed that of optical interferometers, albeit for short ranges on the order of tens of micrometers.

Basic relations

A surprisingly few relations are needed to understand the operation of the capacitance transducer. The archtypical capacitor stores energy in the dielectric between two charged conductors. First, from the definition of capacitance, C expressed in Farads

$$C \equiv \frac{q}{V}, \tag{3.1}$$

where q is charge in Coulombs and V is the voltage potential. The total energy stored in the system is

$$U(t) = \frac{1}{2} CV^2. \tag{3.2}$$

The amount of RMS alternating current able to pass through a given capacitor is given by

$I = \frac{V}{X_c}$ where the reactance, X_C is defined as $X_c = \frac{1}{2\pi fC}$ where f is the excitation frequency. Thus

$$I_{capacitor} = 2\pi fVC. \tag{3.3}$$

So, for a given capacitance, to increase the amount of recoverable signal, we have only two choices. Increase the excitation frequency and/or increase the drive voltage. Excitation frequency is limited mostly by electronics and stray capacitance (more on this later). Drive voltage is limited by the dielectric breakdown of air to around 3000 V/mm. Thus, for a 10 micron gap, we are limited to 30 volts, at most. For increasingly small gaps, less than a few microns, there is evidence that the breakdown voltage actually *increases*. This effect, known as the Paschen curve[2], is due to the fact that as electrode gaps shrink to the same order as the mean free path of the air molecules, the air gap begins to approximate a vacuum. One or two micron gaps have been operated at voltages as high as 300 volts without breakdown[3].

The factor limiting resolution of the capacitance gage is the Johnson (or thermal) noise of the excitation source oscilator and the associated signal recovery electronics. The capacitor itself is unique because (in theory) it produces no noise of its own. General estimations of noise in capacitance gages are useful only as an academic exercise

because each individual circuit topology requires its own unique, extensive noise analysis.

Electrode geometries

The simplest case is that of two, equal sized parallel conductive plates, separated by a dielectric medium. To derive the expression for the system's capacitance, we must solve for charge and potential. In general terms, this is best done by solving the surface integral for charge and solving for the potential between the two plates, then applying equation 3.1. Thus, capacitance

$$C = \frac{\oint_S \varepsilon \vec{E} \cdot dS}{\int_a^b \vec{E} \cdot d\ell} \tag{3.4}$$

Depending on geometries, determining charge densities involves applying Gauss's law and then solving for electric field using Laplace's equation. For the simple case of two equal sized parallel plates with uniform sheet-charge density ρ ,

$$Q = \rho A, \tag{3.5}$$

where A is the plate area. Potential V is then

$$V = \frac{\rho}{\varepsilon} d. \tag{3.6}$$

Capacitance of the two plate system is then

$$C = \frac{Q}{V} = \frac{\varepsilon A}{d}. \tag{3.7}$$

The sensitivities for the various parameters are given by the chain rule

$$dC = \frac{\partial C}{\partial \varepsilon} d\varepsilon + \frac{\partial C}{\partial A} dA + \frac{\partial C}{\partial d} dd, \tag{3.8}$$

$$dC = \frac{A}{d} d\varepsilon + \frac{\varepsilon}{d} dA + \frac{\varepsilon A}{d^2} dd. \tag{3.9}$$

Dividing by C to obtain the relative sensitivities

$$\frac{dC}{C} = \frac{d\varepsilon}{\varepsilon} + \frac{dA}{A} - \frac{dd}{d} \qquad (3.10)$$

it is seen that changes in electrode gap spacing and overlapping area have a direct effect on the measured capacitance. Environmental effects, which effect ε will be examined in a later section in this chapter.

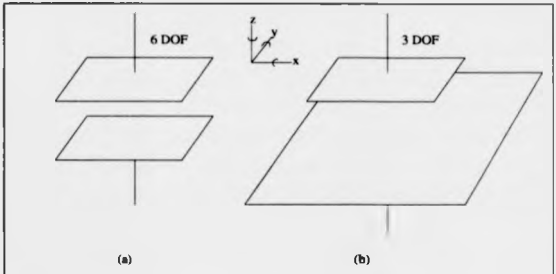


Figure 3.1. Parallel plate capacitor showing 6 degrees of freedom (a) and 3 degrees of freedom (b).

Shown in the previous figure is the ideal, parallel plate capacitor. It is, however, far from ideal in real life. The configuration shown in (a) is sensitive to motions in all six degrees of freedom. For measuring gap separation, this is clearly an undesirable situation. Making one electrode much larger than the other reduces this sensitivity to three degrees of freedom: Pitch, roll and electrode separation. The pitch and roll axis can usually be constrained to second order effects. If need be, these effects can be compensated using the following models. For square electrode plates of width W and length L undergoing a tilt of angle θ, the problem is solved by a simple integration of a infinitely thin strip dC. This integration[4] results in

$$C = \frac{W\varepsilon}{\theta}\ln\left[\left(1+\frac{L\tan(\theta)}{2d}\right)\Big/\left(1-\frac{L\tan(\theta)}{2d}\right)\right]. \qquad (3.11)$$

valid for the condition $d > \theta(L/2)$. For small θ this reduces to

$$C \cong \frac{\varepsilon A}{d}\left[1+\frac{1}{12}\left(\frac{L\theta}{d}\right)^2\right]. \qquad (3.12)$$

The problem is mathematically more difficult for circular plates, but has been solved by Harb, *et. al.*[5]. For circular plates of radius R, the apparent change in displacement δ d for a small change in tilt δ θ may be approximated by

$$\delta d \cong -R^2\theta\delta\theta/d. \qquad (3.13)$$

Returning to equation 3.9, it is clear that area also effects overall capacitance. Four basic types of area transducers are depicted in figure 3.2.

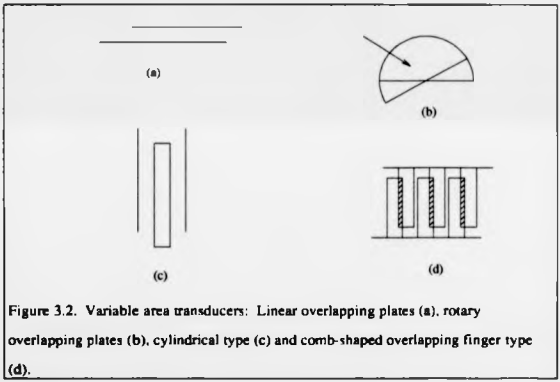


Figure 3.2. Variable area transducers: Linear overlapping plates (a), rotary overlapping plates (b), cylindrical type (c) and comb-shaped overlapping finger type (d).

Due to the fact that all the overlapping electrode, variable area schemes are very sensitive to gap distance and angular motion errors, they are not often used in

nanometer precision applications. The rotary type, familiar to many as a radio tuning "condenser" has been applied to torsional balances[6]. The overlapping finger method has been used as a machine-scale[7], but the problem of sensitivity to electrode spacing persists. The overlapping electrodes do make attractive candidates for sensors, however because the grating period may be much finer than that of an Inductosyn™, which is limited by minimum wire diameter.

Fringing fields at the edges of electrodes are another source of non-linearity in capacitance transducers. If the electrode plate is thick, or the ratio of electrode dimension to separation is small, stray fields from its sides can introduce non-linearities in the measurement. These fields may be reduced by making the electrodes vanishingly thin, reverse tapering them away from the other plate or compensating for them with a correction factor[8]. Failing this, the stray effects may be nullified by using a guarding technique.

Guarding and shielding

Fringing electric fields can never be completely eliminated from the electrode plates at the transducer, but their effects can be removed from the measurement. By using an *active guard ring* electrode running around the circumference of the transducer electrode, the fringing field can be effectively moved to a non-sensitive area[9] (figure 3.3). The active guard is driven at same voltage, in phase with the main electrode.

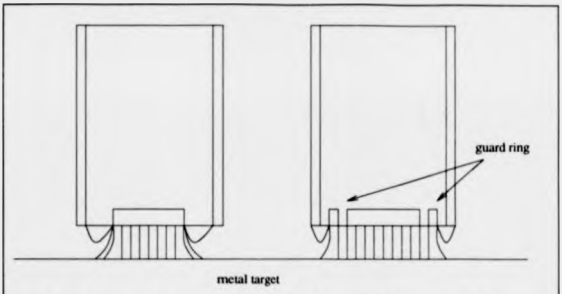


Figure 3.3. Active guard ring showing reduction of fringing field effects on main electrode.

This scheme is especially effective if one of the electrodes is spherically shaped, as are the tooling balls used for machine axis-of-rotation metrology[10]. The positive effects of guarding are not limited to reducing fringing fields at the electrode. The effects of cable capacitance may be greatly reduced, decreasing the measurement's sensitivity to cable length and bending. This is accomplished by "bootstrapping" the inner shield of a triaxial cable with a high speed buffer amplifier (figure 3.4).

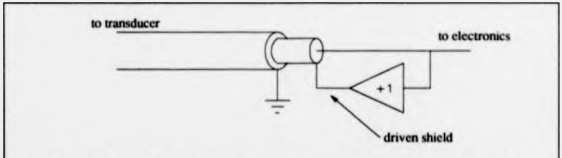


Figure 3.4. Cable guarding showing "bootstrapping" technique with buffer amplifier.

Since the inner conductor and the shield are always at nearly the same voltage, there exists no current flow between them by definition. The buffer amplifier must have a

high input impedance so as not to load the incoming signal and must introduce minimal phase shift between the two signals.

Stray capacitance effects must also be accounted for in any capacitance design. "Strays", are usually in parallel with the sensor and represent the capacitance between the sensor or cable and the earth, instrument frame or any nearby object. Wrapping a third conductor, a grounded shield around the cable keeps this capacitance constant. All other objects in the vicinity of the sensor should be solidly grounded to prevent variations of the sensor's stray capacitance. It is interesting to note that at least one measurement technique, *scanning capacitance microscopy*[11] measures the small, stray capacitance between a sharp tip and a specimen and servos the tip height to maintain a constant capacitance and possibly a constant gap.

Environmental effects

From the analysis of sensitivities in equation 4, it is clear that the dielectric constant of the medium between the plates is of great importance. Precision capacitance standards kept by government standards organizations are made of dimensionally stable, low-loss dielectric materials, such as fused silica[12]. In almost all gaging cases, however, the dielectric medium is air. Not dry air, but moist, humid air of ever-changing temperature and pressure. Herein lies the crux of the problem: How to measure and compensate for these environmental changes?

One possible solution is to compensate indirectly by measuring temperature, pressure and humidity, then applying a correction factor to the measurement. Expressions have been developed for moist air which relate permittivity to these three variables[13]

$$\varepsilon_{moist\ air} = 1 + \frac{211}{T}\left[P + \frac{48P_s}{T}H\right]10^{-6}, \qquad (3.14)$$

where T is absolute temperature (K), P is air pressure (mmHg), P_s is pressure of saturated water vapor at temperature T (mmHg) and H is relative humidity expressed

in percent (%). The previous expression is of limited utility, however. When the relative humidity rises above some 30%, adsorbed layers of water form on the electrode plates[14]. With high levels of humidity, these layers can become many tens of nanometers thick. Water has both a high dielectric constant (ε=80) and a high polar moment. As soon as these adsorbed layers form, the equation is no longer valid. One possible avenue of approach is to coat the electrodes with a substance which resists the formation of water layers, such a rhodium. Further studies of these effects are warranted.

A pragmatic solution to this environmental problem is to make a differential measurement between the probe transducer and a dummy *reference capacitor*, open to ambient air. In theory, environmental changes will affect both capacitors equally, and be nulled if identical geometries are used.

Reference capacitors

As mentioned in the previous section, most standards organizations are interested in maintaining capacitance standards which are stable in changing environmental conditions. For these reasons, capacitance standards are made of stable, solid materials such as fused silica. Since air is not a stable dielectric, it has not found use in standards applications, and high stability air gap capacitors are often difficult to find. A study of catalog specifications showed that some of the *best* air-gapped units have stabilities that approach only 50 ppm. For our purposes of using the reference capacitor as a "dummy" sensor in a differential measurement scheme, care should be taken to match both the electrode area and gap distance. Dimensional stability of the electrode areas is relatively easy to obtain but a stable gap separation requires construction with high stability materials. Zerodur glass ceramic has a thermal expansion coefficient of some part in 10^{-8}. If a reference capacitor is etched from Zerodur, with channels leading to the electrode gap open to ambient, a high stability air gap should result. One such reference is shown in figure 3.5.

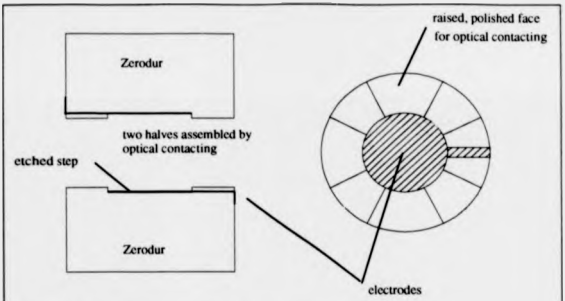


Figure 3.5. Optically contacted Zerodur ambient air reference capacitor.

The two polished faces, with etched steps and thin-film metal electrodes, are permanently assembled by optical contacting. Optical contacting is essentially a permanent wringing process that is used to assemble highly polished, flat pieces of glass. Since the gap is only on the order of ten microns, tolerances are extremely tight, and the capacitance of the finished product is difficult to predict. Highly desirable would be an adjustable capacitor, with areas similar to those encountered on the cantilever beam probes, but with an adjustable gap separation. Such a variable reference could be adjusted for a null, representing the force probe cantilever condition with no external forces applied. A variable reference is shown in figure 3.6.

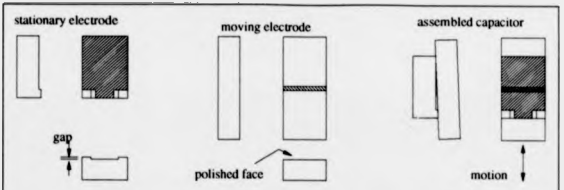


Figure 3.6. High stability variable reference capacitor made from Zerodur glass ceramic.

The basic design uses a very small wedge formed from the etched step along the width of the base. This step is shaped to minimize contact area and to approximate a point contact, preserving kinematic constraint. The upper plate is polished optically flat and has a vacuum deposited electrode with similar dimensions to those of the cantilever. The sides of the upper plate are constrained by two fixed gibbs, 3mm steel ball bearings glued to the end of grub screws. The other side is preloaded into the fixed gibbs by 3mm spring-loaded ball ended plungers. A leaf spring with a point contact forces the two halves together, an aluminum frame both holding the gibbs and shielding the capacitor. The upper plate is driven by a ball-ended screw through a hysteretic drive coupling (figure 3.7).

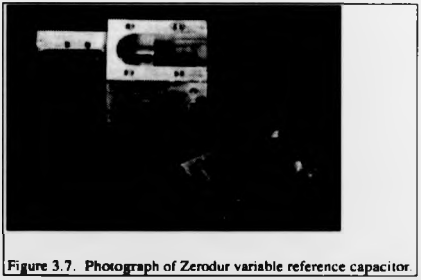


Figure 3.7. Photograph of Zerodur variable reference capacitor

Since many millimeters of screw translation are required to change the capacitance only some ten picoFarads, a high resolution is obtained. Shown in figure 3.8 is a calibration of the capacitor adjustment mechanism.

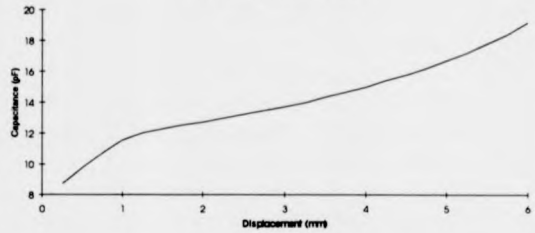


Figure 3.8. Graph of capacitance versus screw translation for the variable reference capacitor.

The data were recorded on a Hewlett Packard 4284A precision LCR bridge. Problems with the design include wear of the point contacts on the Zerodur base and stick-slip of the side gibbs. The former problem is not that great because the capacitor is normally adjusted and then left unattended for long periods of time. Adjustments are infrequent in normal force profiler operation. The latter problem affects the dimensional stability of the gap separation and is much more serious. The aluminum frame expands with temperature much more readily than the Zerodur. We wish the aluminum frame's expansion to have a second order effect on capacitance. If the stiffness of the vertical preloading contact is much greater than those of the side gibbs and the side contacts are free to slip, expansions of the frame will be of little consequence because the basic gap geometry is defined by the Zerodur wedge. Prototypes of this device had relatively stiff steel balls contacting rough surfaces on the sides, preventing easy slip at

the bearing interface. Future versions will need to be slightly larger, having polished sides, with relatively loose BeCu leaf springs for preloading and polymer bearing pads for easy slip along the bearing faces.

A reasonable temporary solution has been found by modifying a commercial stacked mica/metal plate capacitor. Removing most of the sandwiched metal and mica plates from a 100 pF variable capacitor resulted in a 10 pF adjustable capacitor of acceptable stability over short periods representative of the force profiler's measurement cycle time (several minutes at worst).

Capacitance measurement techniques

Many techniques may be used to measure the capacitance of a transducer, each with its own merits and limitations. These techniques may be broadly grouped into three categories for purposes of comparison[15]: Resonance shifting, charge transfer and alternating current bridge methods. Each category has many subsequent sub-groupings which will next be examined with the purpose of selecting the best methods for use in the force profiler instrumentation.

Resonance

The resonance method functions over a wide frequency range, from a few kHz to hundreds of MHz. Two basic principles are used: Either a stable carrier oscillator is passed through a filter whose characteristics are variable with the device under test, or an oscillator whose frequency is dependent on the device under test is measured.

Filter shift networks

By connecting a fixed oscillator to a LC filter containing the device under test, a measure of sensor capacitance may be made by measuring the oscillator amplitude, attenuated by the filter passband (figure 3.9).

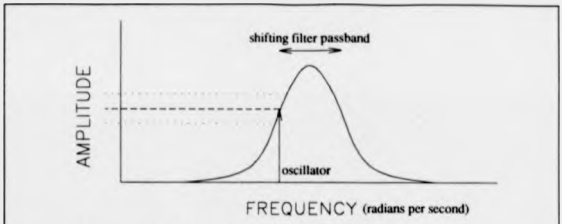


Figure 3.9. Resonance method of shifting filter passband.

Obviously, the "sharpness" of the filter is crucial to the success of such a transduction method. We may describe filter efficiency with the parameter Q, which provides a measure of a filter's efficiency. Q is defined by

$$Q = \frac{\omega_n}{B} \tag{3.15}$$

where ω_n is the center frequency, in radians and B is the bandwidth. In real circuits, Q values of over 100 are difficult to realize due to resistive losses in wires and eddy current losses in inductor cores. Thus, the given bandwidth of a good RLC network is roughly the center frequency (radians) divided by 100. As frequencies increase, the bandwidth of most filters also increased. Something to be considered when designing a filter shifting network.

A filter-shifting technique was developed in the early 1980s at RCA for a capacitive video disc storage technology. Commercially, the venture was a failure, but the hardware and methodologies have since been applied to scanning capacitance microscopes[16]. The systems with the highest spatial resolution (some 10 nm)[17] have been based on the RCA Video Disc system, developed by Matey[18] and shown in figure 3.10.

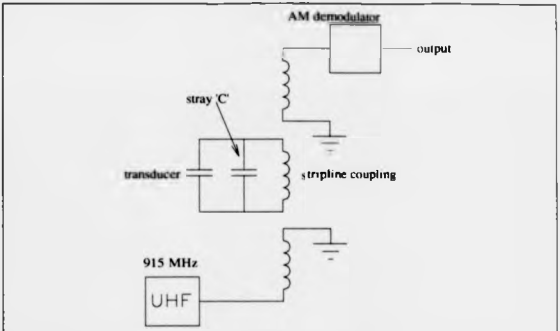


Figure 3.10. RCA Video Disc system of Matey showing measurement of stray capacitance between a sharp tip and a specimen.

In this circuit, a 915 MHz UHF oscillator feeds a stripline filter circuit (a distributed RLC network of printed circuit board material) whose passband is changed by the stray capacitance between the tip and sample. The amplitude modulated output is then rectified and low-pass filtered. Sensitivity in a 1 kHz bandwidth is on the order of 10^{-19} F, which can be improved to 10^{-22} F at higher oscillation frequencies. Oscillator amplitude noise dominates, with a strong 1/f component. Dithering techniques are often used to circumvent the flicker noise. Since this technique measures mostly stray capacitances and requires an extensive knowledge of practical microwave engineering, it was not used in the force profiler.

Oscillator shifting networks

In these methods, the frequency of an oscillator is shifted by changes in the device under test. The oscillator may be of the RC or RLC type. The RC oscillator is easily implemented with digital multivibrator or voltage controlled oscillator circuits at frequencies up to several MHz. Stray capacitances are a common problem, but the

oscillator may be miniaturized and located physically near the capacitance electrodes with minimal connecting cabling. Such RC circuits are usually only sensitive to about 10^{-15} F, with oscillator noise and temperature stability limiting performance. RLC oscillators are usable to frequencies of many hundreds of MHz, generally sharing the limitations of the previous scheme. Two different methods are commonly used to detect oscillator frequency shifts: FM slope detection and frequency counting.

Slope detection requires only a low pass filter element. FM slope detection is given its name because the oscillator's frequency shifts result in signal amplitude changes as the source frequency moves down the slope of the filter's passband, figure 3.11.

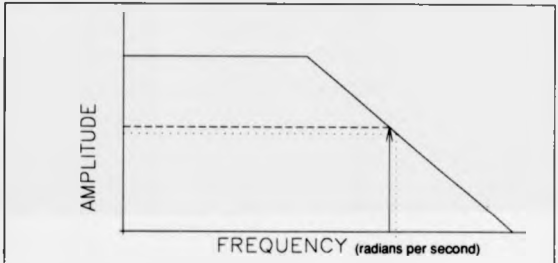


Figure 3.11. FM slope detection method of frequency detection.

Sensitivity of the method is determined by the filter Q. Higher Qs yield sharper cutoffs and more sensitivity. Since it is an analog method, it can be used with a variety of inexpensive filters. Generally, stability is poor (about 10^{-5} at best) because of the sensitive nature of the analog components. The resonance techniques require much manual operator adjustment and are generally very sensitive to stray capacitances. Resonance methods will not be used on the force profiler.

Charge

All charge techniques make use of the basic relation

$$Q = CV. \tag{3.16}$$

Since V can be held constant or to known AC values, the capacitance can be inferred by measuring the charge stored on the transducer. If the electrodes are known to be moving and DC response is not needed in the transducer, a simple charge amplifier may be used to measure electrode vibrations (figure 3.12)

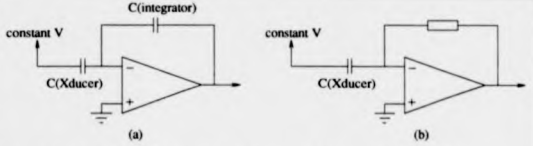


Figure 3.12. Charge amplifier used to measure vibrating electrode displacement (a) and transimpedence amplifier used to measure velocity (b).

In these schemes the voltage across the plates is held constant, leaving the capacitance directly proportional to charge. If the plates are vibrating, the change in charge may be measured directly as a current as in (b). Since by definition $I = dQ/dt$, this circuits output is the velocity of the electrode plates. Integrating the transducer current with a charge amplifier yields position directly These methods are well proven in the field of ultrasonics research where DC response not important. Since the force profiler requires a DC response, these methods were not considered.

Charge pump/switched capacitor

Charge methods capable of DC response may be divided into two basic groupings: Charge pumps/switched capacitors and slope conversion. Switched capacitor methods use integrated circuit analog switches to store charges on the transducer capacitor and then operate on that charge in a variety of ways.

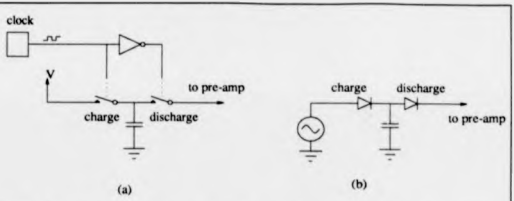


Figure 3.13. Charge pump circuits using analog switches (a) and diodes (b).

Charges stored on the transducer capacitor can be subtracted from charges on a reference capacitor[19] to make differential measurements. Successive approximation A/D techniques may be used to null the amount of charge on the transducer capacitor, providing a direct, digital output[20]. Switched capacitor techniques may be combined with RC oscillators to provide an output in variable duty-cycle form[21]. Building upon the usage of MOSFET switches for charge transfer, complete capacitive readouts can be fabricated using integrated circuit techniques. If the readouts are combined with on-chip micromachined sensors, such as shuttle-mass systems, complete integrated circuit accelerometers can be produced. Analysis of these miniature systems is beyond the scope of this discussion, but system resolutions generally are on the order of tens of femto Farads (10^{-15}), not good enough for application in the force profiler.

Charge pump systems use diodes[22] or analog switches[23] to charge the transducer along one circuit path and then discharge it along another path to the detector. Clock cycles are often in the MHz range. The charge is then integrated over many clock cycles to provide a DC output.

Slope conversion

Slope conversion techniques are a useful variation of the charge pump circuits. Unlike charge pumps, which are integrated over many rapid clock cycles, the slope converter integrates over only one, relatively slow cycle. A typical switching sequence goes like

this: the probe capacitor, discharged to ground is charged with a constant current source. The voltage across such a capacitor is given by

$$V(t) = 1/C \int i(t)dt, \tag{3.17}$$

which for the constant current source is a ramp waveform. Since the coefficient of the integral is the reciprocal of capacitance, the ramp is then *linear* with electrode gap separation. At the end of a fixed time period, the ramp is sampled and held for digitization. The capacitor is then discharged and the whole process starts over again. Bandwidth can be as high as 10 kHz, or more. The circuit is noise limited by the noise of the current source charging the capacitor, usually dominated by resistive Johnson noise. A version of this circuit was presented by Chapman of CUPE[24] and had a noise level of some 10 nm with 12 kHz bandwidth. Cable and probe guarding were also fully implemented in this elegant design. The capacitance electrodes used in this design had a nominal value of some 100 pF, however. Too large for the force profiler probe system.

Alternatively, the capacitor can be permitted to charge until the ramp waveform reaches some predetermined value. Measuring the time interval between the beginning of the ramp and its end value can give a direct digital output if the time is clocked by a counter register. This is essentially single-slope A/D conversion . Schemes for dual-slope ratiometric conversion also exist[25]. A variation of the dual-slope conversion could be applied to differential capacitor measurement if the counter is allowed to count in the *up* direction while charging the transducer capacitor and then to count *down* while charging the dummy reference capacitor. The direct digital output would be free of common mode rejection problems associated with analog electronics.

The charge methods can provide a large range to resolution ratio, but are limited by the 1/f current noise of the charging circuits and are generally used on fairly large capacitors, in the range of 100 pF and up. Since a fairly typical resolution of charge circuits is about 10 nm, they were not considered for the force profiler.

AC bridge methods

Last but certainly not the least, AC bridge reactance measuring methods will next be considered. AC bridge methods are generally recognized as the most sensitive measurements[26], being used in laboratory LCR bridges and dielectric measurement equipment.

Lock-in amplifiers

The lock-in amplifier[27], discussed in chapter 1, is the reactance metrologist's "best friend" in the laboratory. With a lock-in amplifier, measurements can routinely be made down to an amplifiers noise limit. This feat is possible because the lock-in allows high frequency, AC measurements to be made well above the 1/f noise existing in the first few hundred Hertz of the spectrum. Referring back to figure 1.14, it is seen that the lock-in's preamplifiers are AC coupled, negating the effects of DC drift. Their 1/f "flicker" noise (or that of the source) is filtered out of the measurement by the bandpass filter. After processing by the phase detector, the amplifier's bandwidth may be lowered to the sub-Hertz level, integrating the noise and further reducing its effects.

The simplest measurements are those made on a single transducer, by either impedance or admittance. Impedance measurement involves ascertaining the voltage drop across the reactance in question, while admittance measurement is concerned with divining the current through the reactance. These approaches have several drawbacks, however. First, the measurement is not differential. Recalling the large changes induced on air permittivity by temperature changes, the import of the "dummy" transducer is firmly established. Second, a differential measurement allows the use of nulling, where the dummy transducer is tuned to match the nominal value of the active transducer. Since the resulting difference signal approaches zero, the amplifier gain may be increased until either the dynamic range is exceeded, or the signal is dominated by noise. Nulling has another distinct advantage worth considering, oscillator noise cancellation. Since the oscillator noise affects both transducers equally near null, its

noise can be largely differenced out of the measurement. Also, oscillator amplitude drifts tend to cancel each other as well.

Excitation sources

Frequency and amplitude stable sinusoidal sources should be used in precision reactance bridges. Sinusoidal sources should be used rather than square wave sources, for example because of potential distortion from upper harmonics present in the recovered waveform. Even though much of this harmonic distortion can be cured in the preamplifier stage by filtering out all but the fundamental frequency, it is wasteful because useful excitation energy is lost in the higher frequency harmonics. Amplitude stabilized crystal controlled sources can easily have long term frequency stabilities on the order of one part in 10^7 with short term stabilities even higher.

Transformer bridges

One of the most direct ways to construct a suitable sensor for capacitance micrometry is to use a ratio transformer bridge[28]. Some possible bridges, shown in figure 3.14, can use an AC transformer ratio standard (a), or an audio frequency transformer (b) and

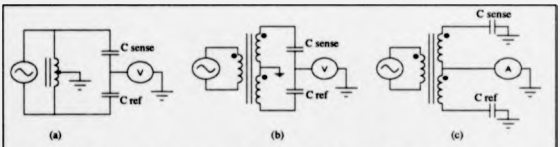


Figure 3.14. Three transformer ratio bridges: ratio (a), voltage (b) and current (c).

(c).

When operating with the fixed ratio transformer (b), the reference capacitor must be variable in order to achieve a null condition with the transducer. The variable ratio transformer[29] (a) allows the use of a fixed reference capacitor and nulling to a part in

10^{-7} via the transformer ratio settings. Guarding the signal is difficult, the transformer is usually placed physically near the transducer to reduce cable capacitance.

The complement of this circuit is shown in figure 3.14 (c). This bridge, described by Patterson and Magrab[30] is wound on a toroidal core with screened cable used as a secondary. This screening serves as a guard and can be enhanced with a bootstrapping amplifier. An added advantage of this circuit is that it allows one of the capacitor electrodes to be grounded.

Electronic bridges

Since transformers tend to be rather bulky, sensitive to temperature variations and microphonics from accelerations, solid state electronics have been utilized in some recent designs. Replacing the transformer with two gain-matched amplifiers, one inverting and the other non-inverting, an equivalent to the circuit shown in figure 3.15 may be formed[31].

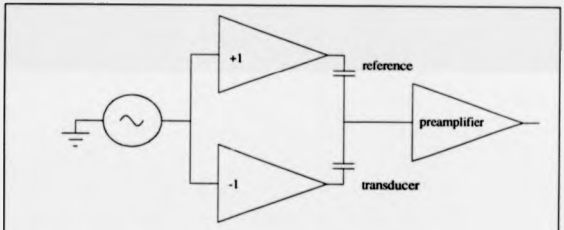


Figure 3.15. Electronic voltage bridge for capacitance measurement.

The active electronics complement to the circuit of figure 3.15 is an electronic current bridge, where current mirrors are used to provide a summing junction for the transducer currents. This circuit remains conceptual. Finally, if a differential drive

voltage is not available, a differential detector may be used instead. Peters, *et. al.*[32] adapts the electronic differential detector to a rotary plate capacitance transducer.

Linearization

The linearity of the charge based slope conversion methodology has been examined, but what of the linearity of AC bridge sensors where capacitance is inversely proportional to electrode gap distance? Inherently linear measurement schemes such as the differential capacitance displacement transducer come to mind. Used in geophysical instruments for many years[33], this transducer is similar in operation to the LVDI described in Chapter 1. Figure 3.16 shows the electrode arrangements.

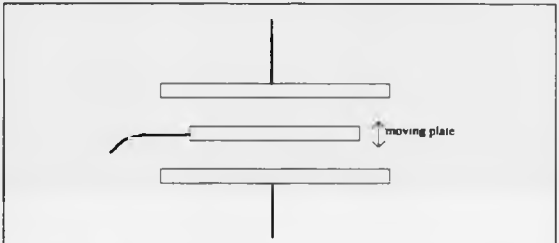


Figure 3.16. Differential Capacitive Displacement Transducer (DCDT) operation.

The advantages of this setup are threefold: First, the transducer is linear for small displacements of the total gap distance. Second, the sensitivity of the system is doubled over those with only one set of electrodes. Lastly, the reference capacitor is built in, making the system very insensitive to changes in the dielectric constant of the air atmosphere. When used with the self-guarding toroidal transformer bridge, the system allows differential measurements of grounded objects to be made.

Providing a linear output with only one set of electrode plates is a challenge most often met with electronic feedback schemes. Of these, op-amp bridges and amplitude or

frequency feedback linearization are often used. Of the op-amp bridges, either the Wayne-Kerr system[34] or the Queensgate Instruments[35] circuit topology may be used. Both systems are illustrated in figure 3.17.

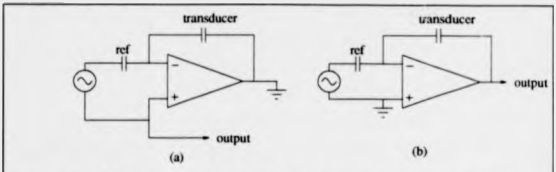


Figure 3.17. Wayne Kerr (a) and Queensgate Instruments (b) linear op-amp bridges.

Both systems take advantage of the fact that a simple inverting operational amplifier has a transfer function

$$V_{out} = -\frac{Z_f}{Z_{in}} V_{in} \tag{3.18}$$

where Z_f is the impedance of the feedback element and Z_1 is the impedance of the series element connected to the summing junction on the inverting node of the amplifier. Given an impedance Z_f that is a capacitive transducer and Z_1 as a reference capacitor,

$$V_{out} = -\frac{d\, C_{ref}}{\varepsilon A} \tag{3.19}$$

if the impedances are treated as pure capacitive reactances. Note the similarity of this circuit to the charge amplifier, except that the DC source has been replaced by an AC excitation signal. Low source impedances negate many cable effects but the system does require an active target with a cable attached to it, preventing it from being grounded. The Wayne Kerr system[36] is essentially the electrical dual of the Queensgate bridge, allowing guarded probe operation with a grounded, passive target. Since the output of the op-amp is grounded, the two input nodes will remain at

essentially the same potential. The signal is recovered off the non-inverting node, which mirrors the inverting input connected to the capacitors. For high gain op-amps with small unbalances between the two capacitors, the output is linear also. Note that as shown, the circuit is unrealizable unless the power supplies and oscillator are allowed to float with respect to ground potential.

Another method of linearizing capacitance gage output is with the aid of a linearizing circuit. While general purpose analog function blocks are well suited for just such a purpose, their linearity would do well to reach the 0.1% level. A slightly better method works well when capacitance sensors are interfaced to a digital computer. The sensor can be calibrated with the aid of high precision length scales and a table of errors computed and stored in the computer's memory for use in linearization. The process is slightly suspect however, due to the frequent calibrations that would almost certainly be necessary keep the system functioning within specification.

Two other feedback linearization schemes are possible. If the capacitive transducer is driven with a sinusoidal source, the series current is

$$I_{cap} = 2\pi\epsilon A \frac{\Delta V_{drive}}{d}. \tag{3.20}$$

A feedback control system can be used to compensate for the reciprocal relationship with gap distance. Either drive voltage or excitation frequency can be adjusted by the servo controller to maintain a constant current through the transducer. Amplitude control[37] uses a voltage controlled amplifier to vary the amplitude of the drive signal to the transducer. Yielding the same result, the drive frequency may be placed under servo control[38] with a voltage controlled oscillator. Dynamic response of both techniques depends on the control system. The systems work well at fairly low bandwidths, below a few hundred Hertz, but can run into control system dependencies at higher frequencies.

Finally, since the force profiler uses capacitive displacement techniques to measure the deflection of a cantilever beam spring, constant force operation results in operation about a constant transducer capacitance, relaxing the requirements for system linearity.

Capacitance gaging for force profiling

The requirements for the capacitance gaging circuitry used in the constant force profiler are:

- Grounded cantilever operation to reduce cantilever and tip charging effects.
- A excitation frequency far above any anticipated cantilever mechanical resonance frequency.
- Low voltage drive for reduced electrostatic attraction between the plates and reduced airborne particle accumulation.
- Design must tolerate a transducer cable length of 1 m or more.
- Noise equivalent displacement must be equal to or better than that of the DPT actuator, approximately 0.5 nm RMS.
- Electronics drift should be an order of magnitude better than the mechanical effects which cause measurement drift, such as thermal expansion of the instrument frame.
- Sensitivity must be adjustable to allow individual force probes to be calibrated.

The Queensgate Instruments capacitance bridges used in the initial prototypes met some, but not all of the aforementioned requirements. Most importantly, the QI circuit did not allow operation with a grounded cantilever and required two cables at the probe. Although sensitivity and drift were excellent, it was felt that a lower excitation voltage and a higher frequency might improve performance. With this in mind, a custom capacitance bridge was designed and constructed, specifically for use on the force profiler.

Oscillator and phase shifter

The basic function of the oscillator is to provide a sinusoidal output to drive the capacitance bridge circuitry. Since the bridge measurement is differential in nature, the effects of frequency drifts on the two capacitors should, in theory, null each other out. The same may be said for amplitude variations. However, owing to the prototypical nature of the circuit in question, it was felt that these effects would be best minimized by a crystal controlled oscillator. The other requirements for the oscillator are that it provide stable, phase-shifted outputs in phase quadrature for possible use in a two-phase lock-in detection scheme. Lastly, the oscillator must provide complementary digital outputs for use by the phase detector. These requirements are met by the circuit shown in figure 3.18.

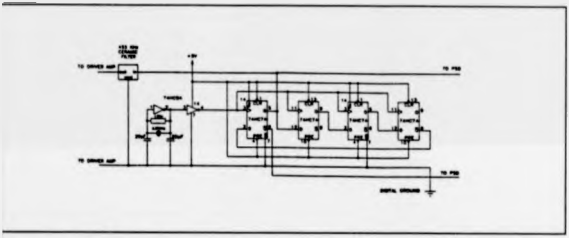


Figure 3.18. Oscillator and quadrature phase-shifter with complementary outputs.

The heart of the system is a modulo-8, walking-ring counter. This flip-flop network divides the clock signal by 8 and provides complementary outputs that are 0, 90, 180 and 270 degrees out of phase with each other. A non-temperature compensated oscillator such as this is stable to a few parts per million[19]. In generating the sinusoidal waveform, some trickery is involved: The square waveform generated by the flip-flops is transformed into a sine wave by removing its harmonics. Doing so involves considerable filtering of the original waveform. A fortunate coincidence is taken

advantage of to achieve this filtering. Ceramic IF filters are used in almost every consumer radio on the planet. These ceramic filters are very stable, very sharp and very inexpensive. Their center frequency is commonly 455 kHz, with input and output impedances around 1000 Ohms. A further coincidence is taken advantage of. Every color television set in North America contains a 3.579 MHz crystal. These crystals are very cheap and readily available. 3.579 MHz divided by 8 falls within the passband of most 455 kHz IF filters. A very stable and inexpensive way to generate a high frequency sine wave! Amplitude stability of the waveform depends on the stability of the 5 V supply powering the oscillator and flip-flops. This is generally bad practice, but the use of HC series, CMOS flip-flops enhances stability since the digital output truly swings from "rail-to-rail" due the CMOS output stage. Performance was adequate, but could be further enhanced by using a high stability regulator circuit and discrete MOSFET switching transistors to "chop" at the desired frequency. Alternitively, a ratiometric measurement scheme could be devised to compensate for variations in oscillator amplitude.

The oscillator was constructed on a circuit board separate from the analog portions of the circuit. A solid ground plane extended underneath the circuit board and is kept as a digital ground only, separate from the analog grounds.

Driver amp, transformer and cables

The driver amplifier takes the output from the ceramic filter and boosts the current to drive the low impedance load of the bridge transformer primary. Since the output of the 455 kHz ceramic filter is referenced to the digital ground, a difference amplifier configuration is used for the driver circuit to isolate the grounds. The op-amps used in the driver are Analog Devices[40] AD829. These amplifiers were designed for high DC precision, low noise, high frequency, high current operation for applications such as high-definition television. Being video amplifiers, they were designed to drive 75 Ohm loads with ease. The transformer bridge was simply designed to operate with a

nominal impedance in the 75 Ohm region. One further note, before moving on to the transformer bridge. The drive topology in the prototype was single ended, driving the transformer primary with one side grounded. This caused some problems in that the return currents through the ground caused a small signal to be superimposed on this electrical datum. This, in spite of the fact that the ground in question is a solid "ground plane" (thus proving an old electrical engineer's saying: "There is no such thing as an RF ground"). All this may be avoided by driving the transformer differentially, in a push-pull configuration. In this manner, current enters the circuit on one rail, and leaves through the other, staying off the ground plane (figure 3.19).

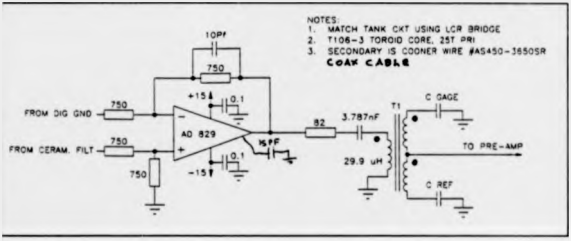


Figure 3.19. Single ended transformer drive topology.

The transformer design is adapted from that of Patterson and Magrab[41]. Two coaxial cable secondaries wound in opposing directions comprise the capacitance bridge. The new design deviates from the literature on at least two counts: The frequency of operation is more than an order of magnitude higher (455 kHz compared to 10 kHz) and the quantity being measured is current rather than voltage. High frequency operation is only possible with core material which exhibits low losses around the frequency of interest. This rules out audio type transformers and most AC standard ratio transformers which are designed to operate in the audio range. A powdered iron core was selected for its low loss properties around 500 kHz. The core material (#3

mix[42]) is a carbonyl iron exhibiting high Q and good temperature stability. Many other mixes are available for other frequency ranges. Applying the manufacturer's winding data as a guide, it was determined that 25 turns on a T-106 core form would give an 89 Ohm impedance. Reflected load impedances from the secondaries have little effect due to the high impedances that the capacitors represent in relationship to the low driving impedance of the secondary. The number of secondary windings was primarily restricted by available space around the core, however, the following observations can be made. More windings give a higher turns ratio and therefore a higher drive voltage at the capacitive sensor. The driving impedance of the secondary is low, reducing the effects of cable bending and triboelectric effects due to charging.

In another departure from previous designs, the output is current summed into an op-amp virtual ground. No bootstrapping guard amplifier is used. As illustrated in figure 3.20, current charging the cable capacitance of the secondary is returned through the

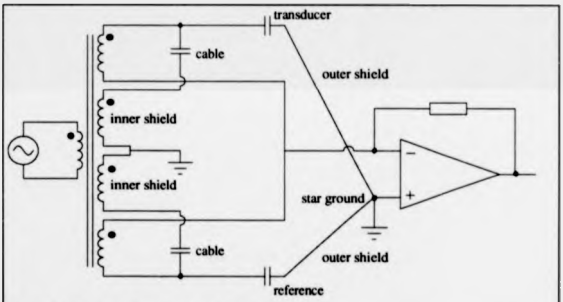


Figure 3.20. Transformer equivalent circuit.

transformer inductor and effectively nullified by Lentz's law.

The current for the capacitance-based force sensor is not returned through the transformer, but on the outside, grounded shield of a triaxial cable. This cable is

effectively double screened since the outer two shields are grounded (in different ways). Touching or bending the cable has almost no effect on the output signal.

Pre-amplifier and filter

As mentioned earlier, the pre-amplifier is a virtual ground input, transimpedence amplifier. An Analog Devices AD745 op amp was selected for this job because of the amplifier's unique combination of high-speed, low-noise, low-input current operation. The signal is amplified as much as possible, given the gain-bandwidth product of the amp. Next, the voltage signal is amplified still further (X 100) by another AD-829. From here the signal passes through another 455 kHz ceramic filter, which also AC couples the signal, removing any DC drifts from the front-end pre-amps. The ceramic filter is buffered by an AD-847 (high speed, but low-noise not needed due to signal amplitude here) and ready for processing by the phase-sensitive detector in the final stage.

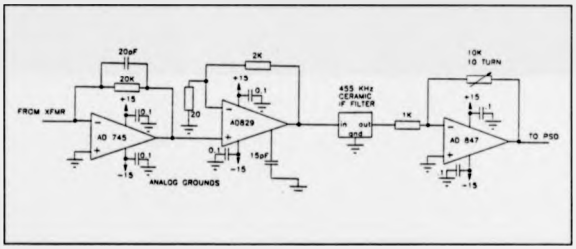


Figure 3.22. Pre-amplifier and bandpass filter.

Phase sensitive detector (PSD) and output stage

A switching PSD was selected for its high frequency operation and low-DC error. DC errors are very important in the PSD circuit because no AC coupling is possible after

this stage. Most linear multipliers (chapter 1) have significant DC errors and hence, are not often applied to phase sensitive detection. A switching phase detector functions as a multiplier in a roundabout fashion. Consider a double pole double throw switch which commutates a signal so that its amplitude changes from -*A* to *A* every time the switch is thrown. This is a sort of digital multiplier which may be expressed as a Fourier series as

$$e = \sum_{0}^{\infty} \frac{2E_1}{(2n+1)} \cos\left\{2\pi\left(f_1 - (2n+1)f_2\right)t + \left(\theta_1 - (2n+1)\theta_2\right)\right\} \qquad (3.21)$$

$$+ \sum_{0}^{\infty} \frac{2E_1}{(2n+1)} \cos\left\{2\pi\left(f_1 + (2n+1)f_2\right)t + \left(\theta_1 + (2n+1)\theta_2\right)\right\}.$$

As in the linear multiplier, when the two frequencies are equivalent, the circuit functions as a phase detector. The upper *n* terms must be removed with a low pass filter. One caveat, however, is that the phase detector is also responsive to inputs at odd multiples of the fundamental frequency (thus adding noise). Generally, a bandpass filter in the preamplifier is enough to remove these potentially problematic, spurious signals. Figure 3.23 illustrates the phase detector concept in a more readily grasped fashion.

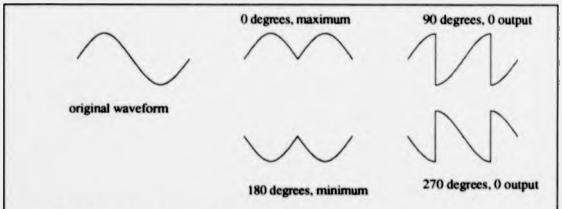


Figure 3.23. Switching phase detector waveforms.

In a variation on the theme presented by Richards[43], the signal is commutated by an analog switch and then differentially filtered to remove any DC offsets introduced by the PSD. Such offsets are most often caused by charging of the switches and their

associated capacitances, resulting a DC offset signal which varies with drive frequency. The analog switches used are Analog Devices ADG201HS, selected for their ability to switch a ±15 V signal with only a +5 V switch drive and their high speed operation. In actual use, the switch outputs are virtually indistinguishable from the illustrations of figure 3.23, even at 455 kHz. These outputs are filtered with a cutoff of 10 kHz and then differenced in an AD620 single-chip instrumentation amplifier. One further 1 kHz filter is placed at the output to further filter the signal, figure 3.24.

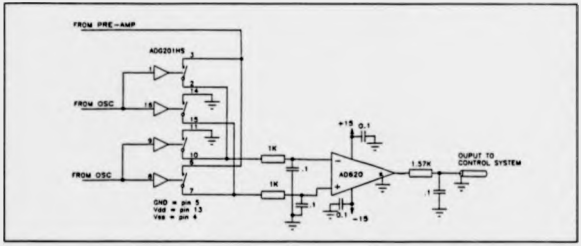


Figure 3.24. PSD and output filter stage.

Performance evaluation

The noise equivalent capacitance (NEC) of the completed capacitance gage was determined in the following manner. Two silver-mica capacitors were prepared and measured on the HP4284A LCR bridge. The capacitances were 11.15 pF and 12.32 pF, giving a ΔC of 1.17 pF. These capacitors were then connected to the force profiler capacitance bridge. Using a 1 Hz lowpass filter (SRS 650), the DC signal and AC RMS noise values were recorded. A noise equivalent capacitance (NEC) of 9×10^{-18} F $Hz^{-1/2}$ was recorded giving a noise equivalent displacement (NED) of 92×10^{-12} m $Hz^{-1/2}$.

Drift of the capacitance gage was determined so that instrumental effects could possibly be isolated when the capacitance gage was integrated into the complete profiler system. Figure 3.25 shows the results of this 4000 second test where the force probe and reference capacitor were connected to the capacitance gage, with the probe not contacting a specimen.

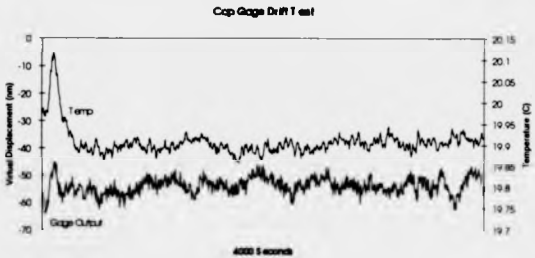


Figure 3.25. 4000 second drift test showing capacitance gage output scaled to simulated displacement units.

Note that the capacitance gage drift is only on the order of 10 nm per hour. This capacitance gage fulfills all requirements for the constant force profiler because the capacitance bridge noise levels are less than the DPT actuator noise levels (0.5 nm RMS) and the electronics drift is orders of magnitude lower than thermal expansion effects in the instrument frame (see Chapter 6).

---

[1] J. Villey, *Nature*, 85, p. 34, (1910).

[2] Arthur R. Von Hippel, Dielectrics and Waves, M.I.T. press, Cambridge, Mass, USA, 1954.

[3] L.S. Fan, Y.C. Tai and R.S. Muller, "IC processed electrostatic micro-motors", *IEDM 88*, p. 666-669, (1988).

[4] J.E. Griffith, G.L. Miller, C.A. Green, D.A. Grigg and P.E. Russel, "A scanning tunneling microscope with a capacitance-based position monitor", *J. Vac. Sci. Technol. B*, 8(6), pp. 2023-2027, (1990).

[5] S. Harb, S.T. Smith and D.G. Chetwynd, "Subnanometer behavior of a capacitive feedback, piezoelectric displacement actuator", *Rev. Sci. Instrum.*, 63(2), pp. 1680-1689, (1992).

[6] M.H. Kwon and R.D. Peters, "Measurement of electrical forces using a modified torsional balance and capacitance sensor", *Rev. Sci., Instrum.*, 62(3), pp. 716-719, (1991).

[7] P.B. Kosel, G.S. Munro and R. Vaughan, "Capacitive transducer for accurate displacement control", *IEEE Trans. Instrum. and Measurement*, IM-30(2), pp. 114-123, (1081).

[8] A.H. Scott and H.L. Curtis, "Edge correction in the determination of dielectric constant", *Journal of Research of the National Bureau of Standards*, pp. 747-775, June, 1939.

[9] "Non-contact dimensional gaging using capacitve sensing", ADE Corporation, Newton, Mass, USA.

[10] ANSI/ASME B89.3.4, "Axis of Rotation".

[11] J.R. Matey and J. Blanc, "Scanning capacitance microscopy", *J. Appl. Phys.*, 57(5), pp. 1437-1444, (1985).

[12] R.D. Cutkosky and L.H. Lee, "Improved ten-picofarad fused silica dielectric capacitor", *J. Res. of the NBS*, Vol. 69C, No. 3, pp. 237-173 - 243-179, (1965).

[13] G.A. Bertone, Z.H. Meiksin and N.L. Carroll, "Elimination of the anomalous humidity effect in precision capacitance based transducers", *IEEE Trans. on Instrum. and Measurement*, **40**(6), pp. 897-901, (1991).

[14] D.A. Grigg and P.E. Russell, "Tip-sample forces in scanning probe microscopy in air and vacuum", *J. Vac. Sci. Technol. A*, **10**(4), pp. 680-683, (1992).

[15] S.M. Huang, A.L. Stott, R.G. Green and M.S. Beck, "Electronic transducers for industrial measurement of low value capacitances", *J. Phys. E. Sci, Instrum.* **21**, pp. 242-250, (1988).

[16] C.D. Bugg and P.J. King, "Scanning capacitance microscopy", *J. Phys. E. Sci. Instrum.*, **21**, pp. 147-151, (1988).

[17] C.C. Williams, W.P. Hough and S.A. Rishton, "Scanning capacitance microscopy on a 25 nm scale", *Appl. Phys. Lett.*, **55**(2), pp. 203-205, (1989).

[18] Matey, *Ibid.*

[19] Y.E. Park and K.D. Wise, "An MOS switched-capacitor readout amplifier for capacitive pressure sensors", Microsensors, IEEE Press, NY, p. 328-333, (1991).

[20] J.T. Kung, H.S. Lee and R.T. Howe, "A digital readout technique for capacitive sensor applications", *IEEE J. Solid State Circuits*, **23**(4), pp. 972-977, (1988).

[21] A. Cichocki and R. Unbehauen, "A switched-capacitor interface for capacitive sensors based on relaxation oscillators", *IEEE Trans. on Instrum. and Meas.*, **39**(5), pp. 797-799, (1990).

[22] D.R. Harrison, W.J. Kerwin and G.L. Schaffer, "A two-wire IC compatible capacitive transducer circuit", *Rev. Sci. Instrum.*, **41**(12), pp. 1783-1788, (1970).

[23] S.M. Huang, J. Fielden, R.G. Green and M.S. Beck, "A new capacitance transducer for industrial applications", *J. Phys. E. Sci. Instrum.*, **21**, pp. 251-256, (1988).

[24] P.D. Chapman, "A capacitance based ultra-precision spindle error analyser", *Precision Engineering*, **7**(3), pp. 129-137, (1985).

[25] J.K. Roberge and M.L. Gray, "A method for the measurement of the ratio of two capacitances, with applications to linear measurement of distance", *Unpublished patent disclosure, MIT*, 1990.

[26] W.C. Heerens, "Microelectronic sensor technology", *J. Phys. E: Sci. Instrum*, **19**, pp. 897-906, (1986).

[27] M.L. Meade, "Advances in lock-in amplifiers", *J. Phys. E: Sci. Instrum.*, **15**, pp. 395-403, (1982).

[28] R.V. Jones and J.C.S. Richards, "The design and some applications of sensitive capacitance micrometers", *J. Phys. E: Sci. Instrum*, **6**, pp. 589-600, (1973).

[29] A.C. Lynch, "Blumlein's transformer-bridge network", Eng. Sci. and Education J., pp. 117-120, (June, 1993).

[30] S.R. Patterson and E.B. Magrab, "Design and testing of a fast tool servo for diamond turning", *Precision Engineering*, **7**(3), pp. 123-127, (1985).

[31] T.R. Hicks, N.K. Reay and P.D. Atherton, "The application of capacitance micrometry to the control of Fabry-Perot etalons", *J. Phys. E: Sci. Instrum*, **17**, pp. 49-55, (1984).

[32] R.D. Peters and P.J. Seibt, "An inexpensive synchronous detector and its application to differential capacitance sensors", *Rev. Sci. Instrum.*, **63**(8(, pp. 3989-3992, (1992).

[33] J. Dratler Jr., "Inexpensive linear displacement transducer using a low power lock-in amplifier", *Rev. Sci. Instrum.*, **48**(3), pp. 327-335, (1977).

[34] J.C.S. Richards, "Linear capacitance proximity gauges with high resolution", *J. Phys. E: Sci. Instrum.*, **9**, pp. 639-646, (1976).

[35] The Author would like to acknowledge a Queensgate Instruments engineer who shall remain anonomyous due to the proprietary nature of this information.

[36] J.C.S. Richards, 'Linear capacitance proximity gauges with high resolution', *J. Phys. E: Sci. Instrum.*, **9**, pp. 639-646, (1976).

[37] G.L. Miller, R.A. Boie, P.L. Cowan, J.A. Golovchenko, R.W. Kerr and D.A.H. Robinson, 'A capacitance-based micropositioning system for x-ray rocking curve measurements', *Rev. Sci. Instrum.*, **50**(9), pp. 1062-1069, (1979).

[38] K. Fritsch, 'Linear capacitive displacement sensor with frequency readout', *Rev. Sci. Instrum.*, **58**(5), pp. 861-863, (1987).

[39] Vectron Laboratories, Inc., 1993 oscillator product catalog, Norwalk, CT.

[40] Analog Devices, Norwood, MASS, USA.

[41] S.R. Patterson and E.B. Magrab, *ibid.*

[42] Amidon Associates, P.O. Box 956, Torrance, California, 90508 USA.

[43] J.C.S. Richards, 'Some aspects of transducer immittance measurement', *J. Phys. E: Sci. Instrum.*, **15**, pp. 1251-1256, (1982).

4 DISTANCE MEASURING LASER INTERFEROMETRY

Chapter summary

Laser interferometry is used on the force profiler to measure the position of the specimen carriage and to clock the profile data into the data acquisition system. By doing so, mechanical hysteresis effects of the instrument carriage drive system are negated and repeatability of the measurement is enhanced. During the design and construction of this interferometer system, many different types of interferometers were considered and rejected, for reasons outlined in this chapter. Ultimately, an inexpensive, yet highly effective polarization interferometer was exploited, using a stabilized HeNe laser as its source. Some of the many types of laser sources and interferometers considered are briefly reviewed herein. At the end of the chapter, the interferometer system components and relevant design and construction details are discussed.

Interferometry basics

There are many hundreds and perhaps thousands of interferometer configurations. One of the best known and historicaly most important is the Michelson interferometer[1]. Many modern interferometers used for distance measuring metrology share some of the same traits of the Michelson, in that an incident light beam, from a laser is split into two beams by a beamsplitter. These two beams shall be called the reference beam and the measurement beam, respectively. Light in the reference beam is reflected off a stationary mirror, back through the beamsplitter. Light in the measurement beam is

reflected off a moving mirror, back through the beamsplitter along a path which is collinear and coincident with the reference beam.

It should be kept in mind that for fringes to be observed in an interferometer, the two light beams must be *coherent, collinear and coincident*. Temporal coherence is a measure of the time interval over which the light wave represents a sinusoid, or the coherence time. Since this sinusoid is a wave traveling at the speed of light through a medium, the distance that this wavetrain covers in the coherence time is called the coherence length. Interferometer optical path differences should not exceed the coherence length. Fringes may still be observed beyond the coherence length, however, but they are less visible and require a venture into partial coherence theory which is beyond the scope of this discussion. The conditions collinear and coincident simply require that the light beams be aligned so that they both reach the detector essentially overlapping all along thier length once they are re-combined in the beamsplitter. The modern remote Michelson interferometer[2], was invented to separate the laser source from the optics, and therefore reduce thermal expansion effects, figure

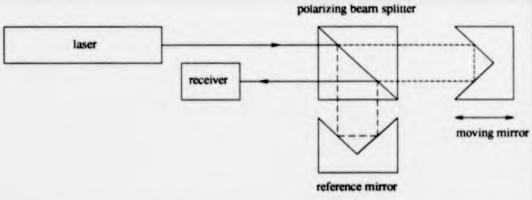


Figure 4.1. Remote Michelson interferometer.

4.1. The remote Michelson configuration also reduces optical feedback into the laser cavity, improving laser stability.

This interferometer will serve as the archetype in all subsequent discussions concerning interferometry. We define the optical path length as the optical path taken by either reference beam or signal beam through the interferometer, starting and ending at the beamsplitter. The optical path difference then, is the difference between the signal and reference optical path lengths. The optical path difference, or OPD represents a phase shift between the two light beams and is given by

$$OPD = \frac{2\pi}{\lambda} 2\Delta L \tag{4.1}$$

where λ is the laser source's wavelength and ΔL is the difference in optical path lengths. For the single frequency, *homodyne* case, (more on this later) the resulting intensity fluctuations observed at a detector are given by

$$I = I_r + I_s + 2\sqrt{I_r I_s}\cos(OPD) \tag{4.2}$$

where I_r and I_s are the reference and signal beam intensities. Fringe formation will be discussed in more detail from an electromagnetic standpoint in the section detailing polarization detector design and construction.

Laser sources

The invention of the laser in the early 1960s paved the way for modern distance measuring interferometry. With coherence lengths in the hundreds of meters and well collimated, high-power beams, many new, previously impossible optical schemes were utilized. The HeNe gas laser is the light source of choice for modern interferometric measurements. The gain curve of this laser is centered about 632.8 nm and is Doppler broadened to approximately 1.4 GHz , figure 4.2.

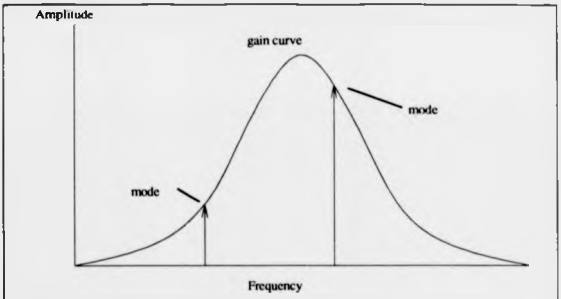


Figure 4.2. Gain curve of unstabilized HeNe laser.

Since the laser itself is a cavity resonator, forcing the light beam to traverse the gain medium many times, more than one longitudinal mode may exist in the cavity at any one time. These modes are spaced in frequency by

$$Mode\ spacing = \frac{c}{2l} \tag{4.3}$$

where c is, of course, the speed of light in the gain medium and l is the length of the resonator tube. In a gas laser, each longitudinal mode is polarized orthogonally with respect to adjacent modes. Normally in lasers used for metrology purposes, the tube length is shortened so that only two modes can "lase", given the gain curve. Such lasers are about 125 mm long and have mode spacings on the order of 700 MHz. Since the tubes are short, power is generally low (less gain medium), on the order of 1 mW. Stabilizing the frequency of such lasers is important because of frequency's relationship apparent displacement. Since coherence length is related to the speed of light multiplied by the coherence time, it follows that the coherence length

$$\Delta x = \frac{c}{\Delta\nu}, \tag{4.4}$$

The laser linewidth, $\Delta\nu$ is about 10 kHz for a typical HeNe laser, giving a coherence length of several Km (for a single frequency laser). The frequency stability, and accuracy of any interferometric measurements made with this laser will be simply

$$stability = \frac{\Delta f}{f} = \frac{\Delta fc}{\lambda}. \tag{4.5}$$

This rather odd mixing of frequency and wavelength measurement units is necessitated by the fact that most laser's center frequencies are specified in terms of wavelength and their frequency stabilities in Hz. For our unstabilized HeNe laser, measurements may be made to a few ppm accuracy. There are at least five different stabilization schemes, one of them the *Lamb dip*,[3] will not be discussed because it is no longer used in commercial metrology lasers. Control of laser frequency basically involves controlling the cavity mirror separation. This may be done by either putting one of the mirrors on a piezoelectric actuator or by heating the tube by either resistance wire or RF induction. The resistance method is very inexpensive; although it is too slow to respond to fast thermal transients, it works amazingly well for most lab situations.

Zeeman stabilized

If the gas discharge tube is placed into a region of constant magnetic field, oriented in line with the laser beam, the atomic energy levels will be split into two circularly polarized (left and right hand) components, separated by about 2 MHz in most commercial systems. A quarter-wave plate outside the laser tube is used to turn the circularly polarized components into two, orthogonally polarized linear components, which are then separated and sampled by photodetectors. A servo to control tube length (and therefore frequency) is locked to the minimum of the intermode beat frequency which occurs at the atomic line center (figure 4.3).

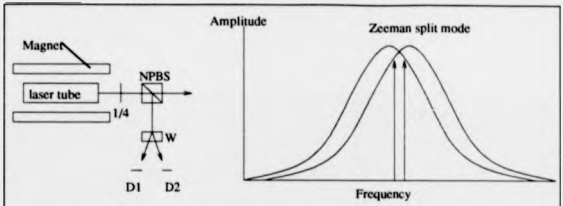


Figure 4.3. Zeeman splitting stabilization scheme.

The stable operating point occurs when the intensities of the orthogonally polarized output beams are equalized. Laser stability can be better than one MHz per year[4], or two parts in 10^{-9}. Furthermore, the output modes are orthogonally polarized and separated in frequency by 2 MHz. This can be fully exploited by AC heterodyne interferometry techniques utilizing polarization coded remote optics. Zeeman stabilized lasers are used in the Hewlett-Packard interferometry systems[5].

Polarization stabilized

The previous example of the Zeeman stabilized laser was a two frequency, or heterodyne system. A DC, or homodyne system requires a single frequency laser source. If a small portion of a laser beam is sampled and passed through either a polarizing beam splitter or a Wollaston prism, the relative amplitudes of the two emission lines may be measured, figure 4.4. A stable servo error signal may be developed by subtracting the amplitudes of the two lines, servoing about the point where they are equal in intensity.

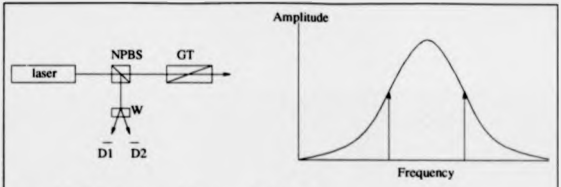


Figure 4.4. Polarization stabilization of a HeNe laser.

One of the modes may be suppressed by passing the output through a good polarizer, such as a Glan-Thompson[6], resulting in a single frequency source. Stability may be on the order of a few parts in 10^{-8}. This stabilization method is used in the Renishaw interferometry system[7].

Iodine stabilized

Iodine stabilized HeNe lasers are used as a reference standard for other lasers. The HeNe tube is placed in-line with a low pressure, temperature stabilized iodine gas cell. By picking-off part of the laser output, ultra-fine structure can be observed due to the absorption of the iodine spectrum, Figure 4.5.

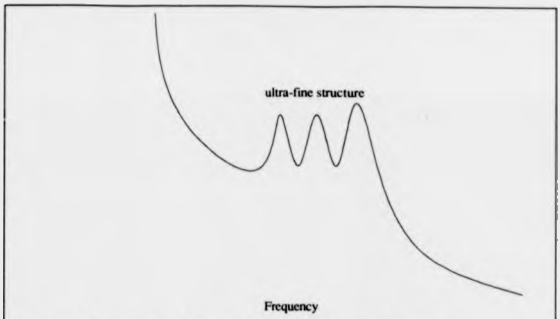


Figure 4.5. Iodine stabilized laser gain curve showing ultra-fine structure.

These ultra-fine peaks are about 4 MHz wide. A variety of locking schemes can be used, but the basic idea is to dither on a peak using a laser mirror mounted on a PZT[8]. Since the PZT is constantly adjusted by a servo to maintain constant laser tube length, PZT drift and hysteresis effects are negated. Stability of these iodine stabilized lasers can be better than 10^{-11}. The Iodine stabilized laser is often used to check the frequency stability of other lasers. This may be done by passing the two orthogonally polarized beams through a polarizing beam splitter and monitoring the resulting beat signal generated by placing a sheet polarizer in front of a high-speed photodetector. High-speed avalanche photodiodes are used for detection with 2 GHz bandwidth signal amplifiers and frequency counters to measure frequency stability.

Diode lasers: Temperature stabilization

Semiconductor diode lasers are becoming increasingly attractive for use in interferometry systems. They are physically very small, produce little heat, are rugged, and with the proper optics can even produce a round, Gaussian beam. However, no analog exists to the Zeeman stabilized laser or the polarization stabilizer. Furthermore,

temperature coefficients are typically 0.25 nm/C. Wavelength is also dependent on the laser's forward bias current. They can be stabilized to some extent by keeping the diode at a constant temperature and bias current. Values on the order of one part per million have been achieved[9], with much better results possible if atomic or molecular absorption lines of external materials are used to produce an ultra-fine absorption spectra for locking purposes.

Detection and fringe splitting

Just *how* the information is extracted from the interferometer makes a great deal of difference to how it may be used for measurement purposes. For the purposes of information, one may regard the number of different signals as an important measure of a interferometer's utility.

Single fringe signal operation

Having access to only one signal severely limits the usefulness and dynamic range of any interferometer. It is very difficult to ascertain direction of motion from only one sinusoidal signal. For this reason, the most common use of such a single signal is to operate it about its *quadrature point*. The quadrature point is the small, linear region about the zero crossing of a sinusoid. If we assume the modulating signal (the motion of the measurement mirror) to be also a sinusoid we then have an intensity function

$$I(t) = I_r + I_s + \sqrt{I_r I_s}\cos\left[A\sin(\omega_m t)\right], \tag{4.6}$$

where A is the modulation amplitude (often called the *phase deviation*) and ω_m is the frequency of the modulation. We then apply the Bessel-Jacobi identity [10], expanding the trigonometric expression into an infinite series of Bessel functions

$$\cos\left[A\sin(\omega_m t)\right] = \sum_{n=-\infty}^{\infty} J_n(A)\cos(n\omega_m t). \tag{4.7}$$

Applying the properties of Bessel functions

$J_n(A) = J_{-n}(A)$ for even n, and

$J_n(A) = -J_n(A)$ for odd n,

the odd terms cancel out and the even terms combine to give

$$\cos[A\sin(\omega_m t)] = J_0(A) + 2\sum_{k=1}^{\infty} J_{2k}(A)\cos(2k\omega_m t). \qquad (4.8)$$

In other words, a DC term with even harmonics. In reality only a few of the terms in the summation are required to give good accuracy. Note that the J_O term drops out when operation is shifted away from the quadrature point to either peak of the sinusoid. At this operating point the signal is doubled in frequency, a useful tool for signal detection. This identity is useful to show how limited the dynamic range of an interferometer is if only one fringe signal is available to the user.

Multiple fringe signal: Quadrature outputs

If two signals are available, which are 90 degrees out of phase with one another, the information situation improves greatly. Two such signals are in a condition known as phase quadrature, often called by engineers an *"a quad b"* signal. Since direction information can be ascertained from an this quadrature signal, fringe counting can be performed to keep a rough count of the moving mirror's relative displacements. Fringe splitting, or interpolation can be performed to further define the mirror's position, down to the nanometer level if care is taken[11]. The "a quad b" signal has the useful property that if displayed on an oscilloscope in the X-Y mode, the phosphor spot will trace a circular path every 360 degrees of phase shift. The direction of rotation changes with the direction the moving mirror takes. With such a setup, it is easy to see where the term *phase unwrapping* comes from.

Fringe counting a quadrature signal is a simple matter if a device known as a 4X quadrature decoder is used. This type of decoder merely ascertains the direction of travel, and issues a clock pulse on every zero crossing of either the *a* or *b* signals.

Since each signal has two zero crossings, the 4X decoder issues 4 pulses every 360 degrees of phase shift (hence its name). Direction and count signals are then used to either increment or decrement a position register. Since the position register may be many bits in length, the fringe counting interferometer can have an enormous dynamic range, the largest of all sensors. The lower resolution limit on this type of measurement depends on what type of, if any, fringe interpolation is used. If no fringe splitting circuitry is used, the position is unknown when the signal is in between quadrants of the circle. Interpolation circuits, also used on moiré grating scales[12], can be used to effectively multiply the resolution of the 4X decoder by factors as high as 256. It should be noted, however, that significant errors can be introduced in the interpolation process. Useful resolution should be limited to about 10 nm for most interferometers operating in air.

DC Interferometry

DC, or *homodyne* interferometry is so-called because of the use of a single-frequency laser source. Fundamentally the same as the textbook Michelson interferometer, these interferometers are somewhat difficult to adapt to multi-signal operation (such as phase quadrature). One way of obtaining a useful signal is to use multiple detectors to analyze the spatial information present in the fringe field, figure 4.6.

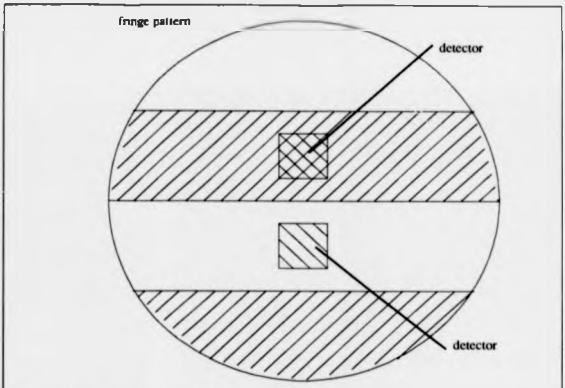


Figure 4.6. Spatial fringe-field analysis by multiple detectors.

Various schemes of this sort have been described by Mertz[13], using three detectors to derive an "a quad b" signal. A commercial version, by Aerotech[14] uses a quad-cell PIN diode placed in the fringe field. Both of these methods are extremely sensitive to fringe spacing; errors are easily introduced by angular and lateral shearing of the laser beam. Angular shearing, which effects fringe spacing the most, occurs when the angles of convergence of the two interferometer beams are not coincident. Angular shearing can occur from errors in the interferometer optics and from laser mis-alignment. Lateral shearing occurs when the interferometer's laser beam is not in-line with the motion of the moving mirror, figure 4.7.

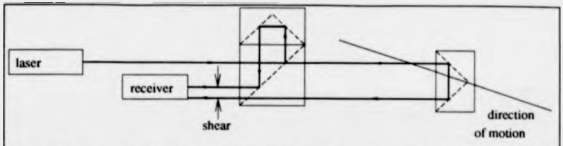


Figure 4.7. Lateral beam shearing causing changes in fringe patterns and errors in DC interferometers.

Since no motion system produces perfect motion and the laser beam can not ever be perfectly collinear with the motion system, beam shearing is always present. The spatial analysis systems are useful and cost-effective however, on short travel stages where the optics can permanently aligned and fixed into place.

AC Interferometry

AC, or *heterodyne* interferometry has become the metric of choice for many metrologists for several reasons:

- The Zeeman stabilized laser seems truly a gift of nature when using this technique.
- Detectors are simple, compact and inexpensive. An important consideration for multi-axis systems.
- Distance measurements are recast into measurements of time intervals, one the most accurate measurements that can be made.

Heterodyne interferometry gets its name from the fact that the Doppler shift[15] from the measurement mirror is used to generate a high frequency beat signal. Note, that the technique described here is more precisely known as polarization coded heterodyne interferometry, because the two-frequency Zeeman stabilized laser source produces beams which are orthogonally polarized. Virtually all modern heterodyne interferometers used for metrology are polarization coded. In the heterodyne technique, distance measurements are transformed into time measurements in the

following fashion: The measurement beam is Doppler shifted according to the mirror's velocity,

$$\Delta f \propto \frac{V(t)}{\lambda} \tag{4.9}$$

Since the beam makes a return trip from the measurement mirror, the shift happens twice

$$\Delta f \propto \frac{2V(t)}{\lambda} \tag{4.10}$$

Recalling that phase is the integral of frequency (frequency is a rotating phase vector),

$$\phi(t) = \int \Delta f \, dt = \frac{2}{\lambda} \int V(t) \, dt \tag{4.11}$$

where the integral of velocity is, of course, displacement

$$\phi(t) = \frac{2\Delta L}{\lambda}. \tag{4.12}$$

The phase in this case is measured between the AC beat signals produced by a reference detector at the laser head and the interferometer receiver. A quadrature signal may be developed with a few extra parts, figure 4.8.

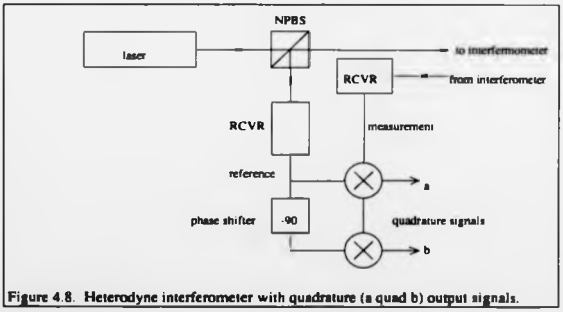


Figure 4.8. Heterodyne interferometer with quadrature (a quad b) output signals.

These quadrature signals may be counted and interpolated in a normal fashion. To generate a beat signal, all that is needed is a high-speed detector and a piece of inexpensive polarizing sheet, placed with its polarizing axis 45 degrees to the horizontal. The polarizer is necessary to mix the two polarization coded reference and measurement beams, which being orthogonal, can not produce fringes on their own. Since the two beams are assumed to be perfectly orthogonal, errors can be introduced into the measurement by optical components which change the polarizations slightly[16]. For this reason, only the highest quality (and therefore most expensive) optics can be used on polarization coded heterodyne systems where accuracy to less than 10 nm is desired. Also, setup and beam orientations are critical at those levels of precision.

The one real drawback to the heterodyne technique prevented its use in the force profiler: Cost. At the time of this writing, a single axis system, with optics costs about $15,000 US Dollars.

DC polarization interferometry

Polarization interferometry uses a DC source in different way from the previously mentioned DC interferometers. By utilizing the same, polarizing optics used with AC polarization encoded heterodyne interferometers, the reference and measurement beams are kept orthogonally polarized until reaching a detector. Upon reaching the detector they pass through a quarter-wave plate which, after superimposing the electric field vectors, results in coding the optical path difference as a rotating polarization state[17]. As the OPD changes by λ, the polarization state rotates through a full 360 degrees. This information may be decoded in a variety of ways to give a displacement output signal. For instance, the polarization axis may be dithered by a Faraday rotator and then 'locked-in' to servo a motor driven linear polarizer. The linear polarizer follows the polarization axis of the rotating vector. This method is obviously of limited bandwidth, but works well down to the nanometer level. Other methods use groupings of detectors and polarizers to derive a quadrature signal. These tend to fall under two

categories: 3 detector and 4 detector systems. In the three detector scheme[18] used by CMX Systems Inc., two signals, 90 degrees out of phase are generated along with an intensity signal. The intensity signal is used to remove any DC levels from the measurement. Each term in parenthesis is a detector signal, where the phase

$$\phi = \tan^{-1}\frac{(I+\sin(OPD))-I}{(I+\cos(OPD))-I}. \tag{4.13}$$

Four detector systems use two pairs of detectors, one for each quadrature signal. The pair signals are differential in nature, being inverses of each other. A simple subtraction removes any DC offsets and doubles signal amplitude. Phase is then

$$\phi = \tan^{-1}\frac{(I+\sin(OPD))-(I-\sin(OPD))}{(I+\cos(OPD))-(I-\cos(OPD))}. \tag{4.14}$$

Polarization interferometry was chosen as the metric for the force profiler. More details will follow later in this chapter, describing the design and construction of such an interferometer.

Phase stepping interferometry

If only one detector is available, how can additional information be gathered as in multi-detector systems? One solution is to add more information to the signal by using the single detector to read fringe amplitudes with different, known phase shifts introduced into the system. If the reference mirror is mounted on a calibrated actuator, such as a PZT stack, controlled phase shifts can be introduced. If the phase shifts are added in 90 degree steps, either four or five readings can be made from the detector and an interpolation algorithm applied[19]. If each reading then is represented by

$$A = I_1 + I_2 + \sqrt{I_1 I_2}\cos(\phi) \tag{4.15}$$

$$B = I_1 + I_2 + \sqrt{I_1 I_2}\cos(\phi + \pi/2) \tag{4.16}$$

$$C = I_1 + I_2 + \sqrt{I_1 I_2}\cos(\phi + \pi) \tag{4.17}$$

$$D = I_1 + I_2 + \sqrt{I_1 I_2}\cos(\phi + 3\pi/2) \tag{4.18}$$

$$E = I_1 + I_2 + \sqrt{I_1 I_2}\cos(\phi + 2\pi), \tag{4.19}$$

the measurements may be reduced by either the "four-bucket" or "five-bucket" algorithms. The four-bucket algorithm is simply

$$\phi = \tan^{-1}\left\{\frac{D-B}{A-C}\right\} \tag{4.20}$$

which, upon inspection is nearly the same as the four-detector scheme used in polarization interferometry. A slightly better result may be obtained by using the five-bucket algorithm,

$$\phi = \tan^{-1}\left\{\frac{D-B}{A+E-2C}\right\} \tag{4.21}$$

which tends to average errors caused by the phase-shifter. Although usually implemented on surface finish and figure interferometers, the method can be adapted to distance measuring interferometry. Bandwidth, limited by the slew rate of the phase-shifter is a limiting factor and likely to be only a few tens of Hertz, not fast enough for a useful measurement.

Frequency-agile or 'chirp' interferometry

If the phase shift in the optical path is changed at a constant rate, a frequency shift is the net result. If, as in the examples set forth in phase-stepping interferometry, the phase shifting element is steadily ramped over 2π while the detector output is integrated over a period of $\pi/2$, the same equations used for phase stepping interferometry will result (recalling that the integral of a frequency shift is a phase shift). Integrating the detector is a simple matter and the phase-shifter performs better for linear ramps than for discrete steps. This method is still extremely limited, however in that the phase shifter limits the overall bandwidth because of the physical motion required by reference mirror.

To be really useful for distance measuring interferometry, the phase-shifting would need to be performed electronically, at high-speed. Modern laser diodes can be frequency-shifted by a modulation of their bias current, at rates into the tens of MHz. Thus, they may be used not only as a light source, but also as a phase shifter[20]. The technique hold great potential for future development in short-range interferometers used for scanning probe microscope metrology.

Environmental effects

Since the optical path difference of light in the interferometer beams depends on the speed of light in the medium (usually air), it is useful to examine the effects of environment on the interferometric measurement. The wavelength of light in a medium is inversely proportional to the index of refraction of that medium. Index of refraction n is defined as

$$n \equiv \sqrt{\frac{\varepsilon}{\mu}} \tag{4.22}$$

Magnetic permeability (μ) of air changes very little with environmental factors. This leaves permittivity to be dealt with, which from Chapter 3 is known to be strongly dependent on temperature. Corrections to the refractive index of air are typically made with the Edlen equation[21], which relates the temperature (T), pressure (P) and relative humidity (RH) of air to refractive index

$$n - 1 = 2.8775 \times 10^{-7} \cdot P \cdot \frac{1 + 10^{-6} \cdot P \cdot (0.613 - 0.00997 \cdot T)}{1 + 0.003661 \cdot T} - 3.03 \times 10^{-9} \cdot RH \cdot e^{0.057627 \cdot T} \tag{4.23}$$

Air turbulence also can add several nanometers of uncertainty to measurements made in open air environments[22]. Enclosing the entire air path in a sealed box helps reduce turbulence effects[23]. While evacuating the sealed chamber would be ideal, it is not normally a viable option. Filling the chamber with helium gas at atmospheric pressure is suggested by Bobroff[24] because helium has a thermal diffusivity some five times

better than air. This environment would influence the calibration of the DPT actuator with its internal capacitance gage, however, and was not pursued for the force profiler.

Polarization interferometer for the force profiler

A single frequency, highly coherent and polarized light source, such as a stabilized HeNe or semiconductor diode laser is positioned with its axis of polarization 45 degrees to the horizontal. A polarizing beamsplitter is placed in the beam path so that the light beam is split into two parts of equal intensity but orthogonal polarizations. If we have arranged the beamsplitter so that light polarized horizontally (We call the electric field intensity E_x) is reflected 90 degrees by the beamsplitter, we may return this light to the direction of the source by means of a fixed, corner-cube retroreflector. The vertical component of the light beam (which we call E_y) passes straight through the beamsplitter and is returned to the direction of the source (along a path collinear-linear to that of the reference beam), by another retroreflector, mounted upon a moving carriage or linear motion system of interest (Figure 4.1). We call the light returned by the fixed reflector the reference beam and light reflected by the moving cube-corner the measurement beam. The electric field equation for the reference beam is then

$$E_x(t) = E_o \frac{\sqrt{2}}{2} \cos(\omega t + \phi) \tag{4.24}$$

and the measurement beam

$$E_y(t) = E_o \frac{\sqrt{2}}{2} \cos(\omega t + \phi + OPD) \tag{4.25}$$

where OPD was defined as being $(2\pi/\lambda)2\Delta L$ where λ is the wavelength of the laser light and ΔL is the change of displacement of the moving reflector. Computations involving polarized light are most easily handled by use of the Jones calculus[25 26]. Whereby expressing the polarizing effects of optical elements in a matrix form, and

expressing incident light as a vector, we may evaluate complex polarization problems with matrix algebra. The Jones vector form of the incident light is then

$$E_i(t) = \begin{bmatrix} E_x(t) \\ E_y(t) \end{bmatrix} = \frac{\sqrt{2}}{2} \begin{bmatrix} \cos(\omega t + \phi) \\ \cos(\omega t + \phi + OPD) \end{bmatrix}. \tag{4.26}$$

By factoring out time dependence, and using complex exponential form, we have

$$E_i = \frac{\sqrt{2}}{2} \begin{bmatrix} e^{j(\phi)} \\ e^{j(\phi + OPD)} \end{bmatrix}. \tag{4.27}$$

This Jones vector represents elliptically polarized light, with the retroreflector's distance information carried as a phase shift between the two light waves. Note, that at this point, there is no interference between the two light beams despite their coincident status. To cause interference between the reference and measurement beams, and convey distance information as a polarization state, it is necessary to somehow mix the two beams so that the horizontal and vertical states may be collinear-linear and interfere with each other. This mixing may be done with a simple linear polarizer, such as Polaroid H-sheet. The following example should provide illumination (bad pun) on the use of the Jones calculus as well as explain polarization mixing. A Jones matrix is composed of the coefficients given by:

$$\begin{bmatrix} E_x \\ E_y \end{bmatrix} out = \begin{bmatrix} a_{11} + a_{12} \\ a_{21} + a_{22} \end{bmatrix} \begin{bmatrix} E_x \\ E_y \end{bmatrix} in \tag{4.28}$$

where

$$a = \frac{E_o}{E_i} \tag{4.29}$$

So, for a perfect linear polarizer with its transmission axis parallel to the horizontal, or x-axis, we have

$$T = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}. \tag{4.30}$$

To determine the Jones matrix for this, or any other element at an arbitrary angle θ to the horizontal axis, we may use a counter-clockwise (anti-clockwise) rotation matrix,

$$S(\theta)=\begin{bmatrix}\cos(\theta) & \sin(\theta)\\ -\sin(\theta) & \cos(\theta)\end{bmatrix} \tag{4.31}$$

to project our optical element onto our coordinate system, and then, project it back onto the optic:

$$[T(\theta)]=[S(-\theta)][T(0^{o})][S(\theta)]. \tag{4.32}$$

It may now be shown, by multiplying matrices, that placing a linear polarizer at an angle of 45 degrees to the horizontal, can cause beam mixing and thus, interference.

$$[E_{o}]=[T\ polarizer_{\angle 45^{o}}][E_{i}]=\frac{1}{2}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}\frac{\sqrt{2}}{2}\begin{bmatrix}e^{j(\phi)}\\ e^{j(\phi+OPD)}\end{bmatrix}. \tag{4.33}$$

We now employ the Jones calculus to determine the effect of a different element, a quarter-wave-retarder, on our incident Jones vector from the interferometer (eq. 4.27). The light beam exiting such a system may be determined by multiplying the Jones matrix for the optical element (or train of elements) by the incident Jones vector. To satisfy our interests of providing a rotating polarization vector, placing the fast axis of the retarder at an angle of 45 degrees to the horizontal

$$T_{\frac{\lambda}{2}}=\frac{\sqrt{2}}{2}\begin{bmatrix}1 & j\\ j & 1\end{bmatrix}. \tag{4.34}$$

Thus the Jones vector of the light exiting the retarder is

$$E=\frac{\sqrt{2}}{2}\begin{bmatrix}e^{j(\phi)}+je^{j(\phi+OPD)}\\ je^{j(\phi)}+e^{j(\phi+OPD)}\end{bmatrix}. \tag{4.35}$$

The description of two counter-rotating circular-polarized light beams. The vector sum of these beams yields a linearly polarized beam. Thus, for the Michelson configuration, we may think of the interferometer as an "optical screw", with the plane

of polarization rotating through one revolution as the measurement reflector moves a linear distance equal to one wavelength of the light source.

Several schemes have been employed to determine the angle of this polarization state vector[27 28]. The objective of most has been to provide signals in phase quadrature, discussed earlier. Many pitfalls present themselves to the designer of a polarization detector. Due to economic constraints, the cost of polarizing optics is possibly the most important consideration of such a design. Typical polarizing optics use either complex thin-film coatings or highly birefringent materials such as calcite. It was determined that to be commercially viable, a receiver design could not make use of traditional polarizing optics such as those described above. Thus the options are reduced considerably.

Polarization detector

The design finally implemented in the force profiler detects light in the form of equation 4.35 (having passed through the quarter-wave-retarder), incident upon a series of thin, glass windows positioned so that the incident light impinges on the surface at the Brewster's angle. A air-glass interface at the Brewster's angle will reflect only light that is polarized perpendicular to the plane of incidence of the window. Thus, the reflected light is completely polarized and represents the intensity of the incident light component at the projected angle that the window makes with the optical axis of the laser beam. The reflection from the back-side of the glass window can be reduced by a thin-film coating. When left uncoated, the additional reflections can be accounted for, but the additional analysis tends only to complicate matters unnecessarily. Therefore, only the front-face reflection will be modeled.

Reflected light from the Brewster window is then incident upon an optoelectronic detector for measurement of intensity, where the detector is of an area large enough to integrate the entire incident beam, or contains a focusing lens to achieve the same effect. Light transmitted by the first Brewster window, being now partially polarized,

is then incident upon a second window, positioned with its angle of incidence at the Brewster angle, making its projected angle with the optical axis 90 degrees to that of the first window. By arranging for the projected angles of each window to be orthogonal to one another, the polarization state of the beam passing through the first two windows is unchanged except for a reduction in amplitude resulting from the polarized, reflected light being split-off to be incident upon optoelectronic detectors.

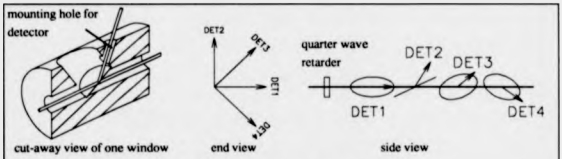


Figure 4.9. Brewster windows for polarization detector.

Light rays incident upon the optoelectronic detectors (PIN diodes in this case) are of a form ideal for differential detection. That is, when one detector is receiving a maximum intensity, its corresponding mate is receiving a minimum intensity. Subsequent electronic subtraction, using the four detector method described earlier, yields a measurement of the vector component intensity which is free from DC offsets.

Transmitted light, having passed through the first set of orthogonal windows, having only its vector magnitude changed and its vector angle unchanged, is then incident upon a second set of orthogonal windows and detectors, identical to the first, except for the entire arrangement being rotated about the optical axis by an angle of 45 degrees. The electronic signal (after subtraction) from the second orthogonal pair of detectors is then in phase quadrature with the signal from the first pair of detectors.

The aforementioned process may be described more quantitatively by application once again of the Jones calculus: First, a Jones matrix is formulated for each of the four glass windows in the polarization detector. Since each window both reflects and

transmits rays of light, two matrices are required for each window, designated R and T respectively. Matrix element coefficients are calculated using the Fresnel equations for dielectric interfaces[29]. The Fresnel equations for dielectric interfaces are

$$r_{\perp} \equiv \left(\frac{E_{or}}{E_{oi}}\right) = -\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} \tag{4.36}$$

$$t_{\perp} \equiv \left(\frac{E_{ot}}{E_{oi}}\right) = +\frac{2\sin\theta_t \cos\theta_i}{\sin(\theta_i + \theta_t)} \tag{4.37}$$

$$r_{\parallel} = +\frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)} \tag{4.38}$$

$$t_{\parallel} = +\frac{2\sin\theta_t \cos\theta_i}{\sin(\theta_i + \theta_t)\cos(\theta_i - \theta_t)} \tag{4.39}$$

where, θ_i is the angle of incidence and θ_t is the transmitted angle through the medium. These angles are determined using Snell's law, setting the angle of incidence to equal Brewster's angle. For a given material, the Brewster angle is by definition

$$\theta_B = \tan^{-1}(n). \tag{4.40}$$

The Jones matrices are then applied for each reflection or transmission of the light beam through the detector along a path from the input of the system, to each detector, in turn. For the four windows, the Jones matrices are:

$$T_{\angle 0^\circ} = \begin{bmatrix} 1 & 0 \\ 0 & .62 \end{bmatrix} \quad R_{\angle 0^\circ} = \begin{bmatrix} 0 & 0 \\ 0 & -.38 \end{bmatrix} \tag{4.41}$$

$$T_{\angle 90^\circ} = \begin{bmatrix} .62 & 0 \\ 0 & 1 \end{bmatrix} \quad R_{\angle 90^\circ} = \begin{bmatrix} -.38 & 0 \\ 0 & 0 \end{bmatrix} \tag{4.42}$$

$$T_{\angle 45^\circ} = \begin{bmatrix} .81 & .19 \\ .19 & .81 \end{bmatrix} \quad R_{\angle 45^\circ} = \begin{bmatrix} -.19 & .19 \\ .19 & -.19 \end{bmatrix} \tag{4.43}$$

$$T_{\angle -45^\circ} = \begin{bmatrix} .81 & -.19 \\ -.19 & .81 \end{bmatrix} \quad R_{\angle -45^\circ} = \begin{bmatrix} -.19 & -.19 \\ -.19 & -.19 \end{bmatrix}. \tag{4.44}$$

Left-multiplying matrices to evaluate the train of optical elements affecting each detector, we have the incident Jones vector for each detector, numbered in-turn,

$$E_{DET1} = [R_{\angle 0^\circ}]\left[T_{\frac{\pi}{2}}\right][E_i], \tag{4.45}$$

$$E_{DET2} = [R_{\angle 90^\circ}][T_{\angle 0^\circ}]\left[T_{\frac{\pi}{2}}\right][E_i], \tag{4.46}$$

$$E_{DET3} = [R_{\angle 45^\circ}][T_{\angle 90^\circ}][T_{\angle 0^\circ}]\left[T_{\frac{\pi}{2}}\right][E_i], \tag{4.47}$$

$$E_{DET4} = [R_{\angle -45^\circ}][T_{\angle 45^\circ}][T_{\angle 90^\circ}][T_{\angle 0^\circ}]\left[T_{\frac{\pi}{2}}\right][E_i] \tag{4.48}$$

Since the intensity at the detector is proportional to the square of the time-averaged, incident electric field, and the current produced by the solid-state detectors is proportional to the incident intensity, we may determine the final output waveforms taking into account the interference effects induced by the polarization mixing. Accordingly

$$I = (E_{DET_x})(E_{DET_x})^* + (E_{DET_y})(E_{DET_y})^* \tag{4.49}$$

where (*) denotes complex conjugation. Alternatively, we may pre-multiply the Jones vector for each detector by its Hermitian adjoint (the complex conjugate of the matrix transpose) to determine intensity in a single, matrix operation. Numerical solutions for the two resulting waveforms are plotted in figure 4.10.

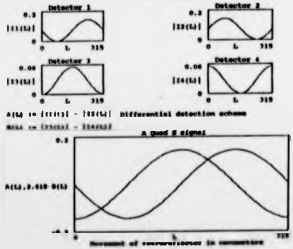


Figure 4.10 Numerical simulation of the polarization detector.

By squaring the waveforms with an electronic comparator circuit, a digital phase quadrature signal may be generated.

Construction

The structure used to fixture all the elements of the detector is machined from 25 mm diameter Delrin rod, chosen for its ease of machining and electrically insulating qualities. Electrical isolation of the PIN diodes used for detectors is convienent due to the diode casings being biased at +5 Volts, and to prevent ground loops. Although the three dimensional representation of the windows shown in figure 4.9 may appear quite complicated, it is, in reality, a very simple matter to produce on a milling machine. The structure was designed so that the Brewster windows locate in the bottoms of holes produced with a square-end milling cutter. As for the windows themselves, their only requirements are that they be thin, non-optically active and reasonably flat. Flatness is not a primary consideration because the beam mixing has already occurred in the quarter-wave-retarder.

The windows are made of 12.5 mm diameter microscope cover slips, approximately 175 microns thick. Such thin windows are necessary for the beam to traverse the length of the receiver tube without taking on significant lateral offsets. The windows were fixed to the bottoms of their holes with a small amount of epoxy, and the hole then covered with black tape to seal out dust and stray light. Figure 4.11 shows the complete laser receiver (the white tube) mounted with the stabilized laser source.

Figure 4.11. Photo of interferometer laser source and receiver system.

Electronics

A simple circuit is used to read the PIN diodes and generate quadrature signals. Each PIN diode is individually amplified and its output current converted into a voltage by four transimpedence amplifiers. The outputs from these amplifiers are then subtracted in pairs using two instrumentation amplifiers, having corresponding gains to match the magnitudes of the resulting two output signals. Each instrument amplifier has a DC offset adjustment, figure 4.12.

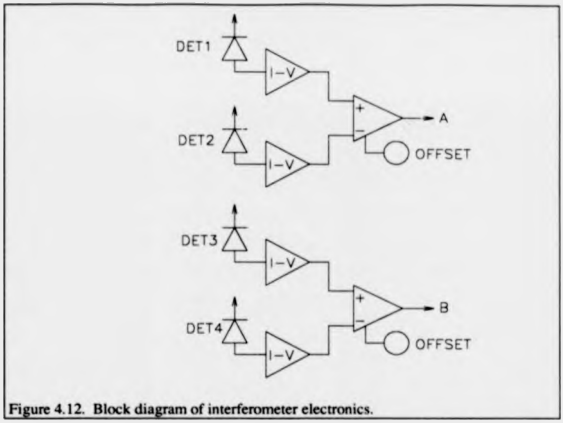


Figure 4.12. Block diagram of interferometer electronics.

Calibration is performed by feeding the outputs to an oscilloscope in the XY display mode. Motion of the measurement reflector should produce, when properly adjusted, a circle centered about the origin. Adjustments are made to offset and gain adjustment potentiometers accordingly and then left alone. A full schematic diagram of the interferometer electronics appears in the Appendix.

Optics

The remote Michelson interferometer archetype depicted in figure 4.1 has some drawbacks that prevented its use in the force profiler. First, the mass and size of the measurement cube corner was too large to fit on the specimen carriage. Also, the resolution of such an interferometer would be only about 79 nm when used without interpolation. For these reasons, a plane mirror interferometer was selected. The flat mirror used by this configuration adds very little mass to the specimen carriage, while at the same time, doubling the resolution to about 39 nm per digital count.

Operation of the plane mirror is as follows: The laser beam, with its plane of polarization at 45 degress to the horizontal, enters the polarizing beam splitter, figure 4.13. Here, the light is split into the reference and measurement beams. The reference beam is turned 90 degrees onto the path designated by the letters (a) through (d). The measurement beam contilues along the numbered paths (1) through (6). Each pass through the quarter wave plate, with a subsequent reflection back through the plate, rotates the beam's plane of polarization 90 degrees. Finally, after two return trips through the quarter wave plate, the measurement beam is coincident with the reference beam, paths (d) and (6).

Two inexpensive cube-corner retroreflectors were obtained from Edmund Scientific. Accuracy specified at 3 arc seconds, λ/4 flatness. These were then affixed to the polarizing beam splitter with small fillets of epoxy around the edges. The whole assembly was then epoxied to a mounting plate on the back of the force profiler. The quarter-wave-retarder, a mica sheet type, was fixed to the other side of the mounting plate with tape, figure 4.13.

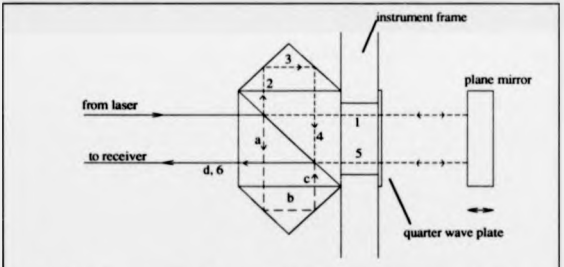


Figure 4.13. Plane mirror interferometer optics on force profiler mounting plate.

As will be shown in the following discussion, stability of the optics was fully adequate.

Results and discussion

With the non-atmospheric compensated laser metrology system used on the force profiler, accuracy on the order of a few parts per million (ppm) per °C can be expected. Given the maximum range of the horizontal traverse carriage of 20 mm, this results in a measurement uncertainty of about 60 nm. Since the minimum clocking interval used by the profiler (more about this in the next chapter) is approximately 80 nm, the expected error is only one count at most. Furthermore, the repeatability of the interferometer is the most important factor since it is being used to compare results from repeatative traces. The repeatability of the measurements in a temperature controlled metrology lab over a time interval of only a few minutes should be on the order of nanometers.

Shown in figure 4.14 are the quadrature signals obtained by digitizing the interferometer output at a sampling rate of 0.02 ms while the profiler carriage was moving at approximately 195 μm/s.

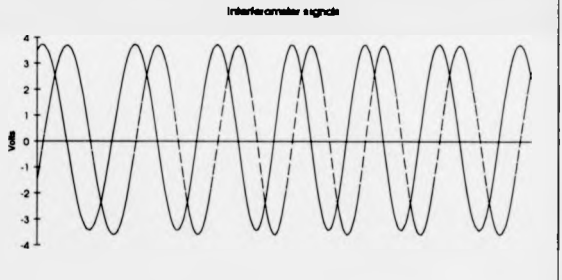


Figure 4.14. Interferometer signals for a carriage velocity of 0.195 mm/s, 0.02 second sampling interval, 256 samples.

To check interferometer noise, the quadrature signals were digitized at a 2 ms rate with the carriage stationary. Phase was recovered by computing the arctangent of the signals and multiplying by the conversion factor 2.27 nm per degree of phase. Shown in figure 4.15 are the results of this noise test.

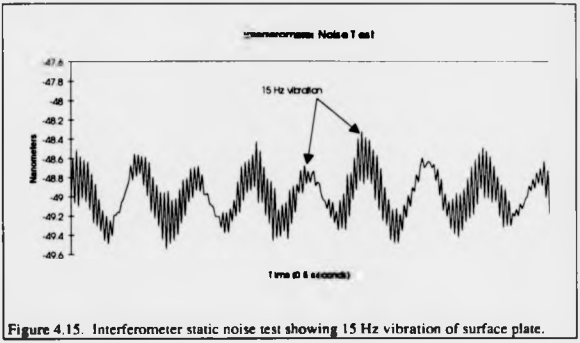


Figure 4.15. Interferometer static noise test showing 15 Hz vibration of surface plate.

Note the presense of the low frequency vibration. This same vibration will later be measured with an accelerometer in Chapter 6(a), environmental testing.

---

[1] M. Born and E. Wolf eds., Principles of Optics, 6th ed., Pergamon Press, NY, pp. 300-302, (1980).

[2] R.R. Baldwin, G.B. Gordon and A.F. Rude, "Remote laser interferometry", *Hewlett-Packard Journal*, 23(4), pp. 14-20, (1971).

[3] J. Koning, P.H.J. Schellekens, P.A. McKeown, 'Wavelength stability of He-Ne Lasers used in interferometry: Limitations and traceability', *CIRP Annals*, 28(1), p. 307, (1979).

[4] G.S. Sasagawa and M.A. Zumberge, 'Five-year frequency stability of a Zeeman stabilized laser', *Applied Optics*, 28(5), pp. 824-825, (1989).

[5] J.N. Dukes and G.B. Gordon, "A two-hundred foot yardstick with graduations every microinch", *Hewlett Packard Journal*, (Aug. 1970).

[6] J. Koning and P.H.J. Schellekens, "Wavelength stability of HeNe lasers used in interferometry: Limitations and traceability", *Annals of the CIRP*, **28**, pp. 307-310, (1979).

[7] Renishaw Plc, Gloucester St., Wotton-under-Edge, Gloucestershire, GL12 7ON.

[8] R.J. Hocken and H.P. Layer, "Lasers for dimensional measurement", *Annals of the CIRP*, **28**, pp. 303-306, (1979).

[9] A. Abou-Zeid, "Diode lasers for interferometry", *Precision Engineering*, **11**(3), pp. 139-144, (1989).

[10] H. Stark, F.B. Tuteur and J.B. Anderson, Modern Electrical Communications: Analog, Digital and Optical, 2nd ed., Prentice-Hall, NJ, p. 313, (1988).

[11] P. Morantz, "The real-time reduction of electronic interpolation errors in precision machine servos", *Proceeding of the Seventh International Precision Engineering Seminar, Kobe, Japan*, pp. 224-228, (1993).

[12] Digital Linear and Angular Metrology, Heidenhain Ltd.

[13] L. Mertz, "Optical Homodyne phase metrology", *Applied Optics*, **28**(5), pp. 1011-1014, (1989).

[14] US Patent #US5018862, (1991).

[15] More properly known as a Lorentz shift due to the relativistic nature of light. Since few engineers are familier with the Lorentz transformations and the mirrors are moving slowly, the shift is uniformly known as the Dopplar shift for light.

[16] A.E. Rosenbluth and N. Bobroff, "Optical sources of non-linearity in heterodyne interferometers", *Precision Engineering*, **12**(1), pp. 7-11, (1990).

[17] J. Dyson, "Very Stable Common-Path Interferometers and Applications", *J. Optical Soc. Am.*, **53**(6), pp. 690-694, (1963).

[18] W. Augustyn and P. Davis, "An analysis of polarization mixing errors in distance measuring interferometers", *J. Vac. Sci. Technol. B*, **8**(6), pp. 2032-2036, (1990).

[19] K. Creath, Progress in Optics, Vol. XXVI, E. Wolf, ed., Elsevier Science Publishing Co., Inc., NY, pp. 351-393, (1988).

[20] K. Tatsuno and Y. Tsunoda, "Diode laser direct modulation heterodyne interferometer", *Applied Optics*, **26**(1), pp. 37-40, (1987).

[21] B. Edlen, "The refractive index of air", *Metrologia*, **2**, p. 71, (1966).

[22] N. Bobroff, "Residual errors in laser interferometry from air turbulence and nonlinearity", *Applied Optics*, **26**(13), pp. 2676-2682, (1987).

[23] R.V. Jones, Instruments and Experiences, Papers on Measurement and Instrument Design, John Wiley and Sons, NY, p. 30, 1988.

[24] N. Bobroff, "Distance measuring interferometry", *ASPE tutorial, Rochester NY, 1990*.

[25] D.S. Klinger, J.W. Lewis and C.E. Randall, Polarized Light in Optics and Spectroscopy, Academic Press, New York, 1990.

[26] R.M.A. Azzam and N.M. Bashara, Ellipsometry and Polarized Light, North-Holland Publishing Co., New York, 1977.

[27] M.J. Downs and K.W. Raine, "An unmodulated bi-directional fringe-counting interferometer system for measuring displacement", Precision Engineering, 1(1), pp. 85-88, (1979).

[28] A. Dorsey, R.J. Hocken and M. Horowitz, "A low cost laser interferometer system for machine tool applications", Precision Engineering, 5(1), pp. 29-31, (1983).

[29] E. Hecht, Optics, Addison-Wesley, Reading, Mass., 1988.

5 SYSTEM INTEGRATION

Chapter summary

This chapter describes the integration of the various profiler sub-systems into a working nanometric instrument. The mechanical structure of the profiler includes the instrument frame, super-precision slideway, probe, specimen approach carriage and drive coupling. The computer interface to the carriage interferometer, the force probe control system and carriage motor drive controller are described.

Profiler mechanical

This section examines individual features of the mechanical system components. First, the overall instrument frame and design philosophy are discussed. The functions of different individual components are then illustrated and examined in greater detail, in light of their contribution to the overall profiler system.

Instrument frame

The instrument frame used in the force profiler is a radical departure from that used on the first prototypes. There are basically two schools of thought regarding materials selection for precision instruments: The use of low thermal expansion materials, where the instrument frame is invariant with environmental changes, and the high thermal diffusivity materials, controlled environment approach. The first instrument, with an all Zerodur glass-ceramic carriage and slideway, coarse approach mechanism and structural loop, paid homage to the former school of thought[1]. This first prototype was based upon a Rank Taylor Hobson Nanostep instrument frame. The original

measuring head was removed and replaced by the DPT with force probe and servo system. The instrument was located in an environment which was only temperature controlled to within a few degrees C (metrology lab at University of Warwick, UK). The second instrument, which is located in an environment which is temperature stabilized to better than 0.1 C (metrology lab at UNC Charlotte, USA), is a good candidate for the latter approach.

The material selected for the force profiler instrument frame was 6061T-6 aluminum stock. It was chosen for its good machinability and high thermal conductivity and diffusivity. Since the high diffusivity approach depends on having any thermal disturbances quickly distributed throughout the instrument frame, the method of joining the various frame members is very important. If glues and epoxies were to be used to bond the frame, they would need a high thermal conductivity to transmit thermal shock to the rest of the structure. A cursory check of various adhesives[2], showed that none of them could even approach the thermal properties of aluminum. Therefore, it was decided to use bolted joints with flat, lapped surfaces to provide maximum contact area and thus, maximum thermal conductivity. The instrument frame joints were lapped into place with silicon carbide compound, and then wiped with a thin layer of silicone heat-sinking grease. Figure 5.1 illustrates the frame components, labelled (F) in the drawing.

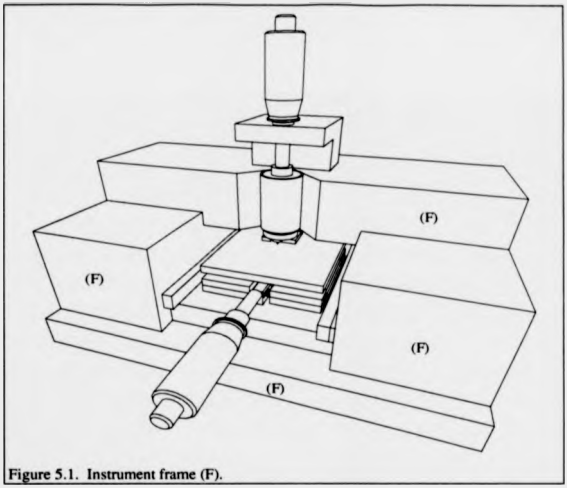


Figure 5.1. Instrument frame (F).

The function of the frame is to provide the metrology loop, hold the coarse positioner, fixture the drive micrometer and the interferometer optics. Upon first inspection, the frame may appear to be over designed for an instrument which experiences very little in the way of dynamic forces. There is some "method to the madness", however. The bridge, 50 by 75 mm, was chosen for the need to support the large vee groove of the coarse positioner. The bridge would also be, structurally speaking, the most important piece of the instrument. External forces from operator motions on the drive micrometer should be negligible, necessitating a stiff spanning member. The stiffness of an idealized, simply supported beam at both ends is given by

$$k = \frac{4Ebh^3}{l^3} \qquad (5.1)$$

Which, for our dimensions is approximately 2.5 MN/m. While this stiffness might seem more appropriate for a diamond turning machine, it is absolutely essential for accurate positioning to nanometer levels with external applied forces. All other frame members are sized to support the bridge piece, this first consideration having a ripple effect throughout the design of the entire instrument.

The bolted joints may also, at first, seem over designed. Previous experience with bolted joints has taught the Author that less rigid joints often rock back-and-forth. They also exhibit creep and thermal expansion problems from mismatches between the bolt and frame materials. The former problem may be attacked by lapping the joint faces flat and spacing the bolt holes for overlapping cones of influence. Thus, 6 M6 bolts with 25 mm spacings on each side of the bridge. An informative review of precision bolted joints is given in Slocum[3]. Thermal creep is not a problem as long as the instrument is assembled in and maintained in a temperature controlled environment. Conical spring "Belville" washers provide inexpensive insurance against gross thermal expansion problems and must be used in any normal shop environment.

Carriage and slideway

Kinematic slideways based upon thin-film PTFE pads bearing against a polished glass counterface have been extensively tested and proven by Lindsey et al.[4], and Smith et al.[5]. Noise levels of such bearing systems are on the order of tens of picometers, making them an excellent (and at this time the only) choice for ultra-precision instrument slideways. Such bearings are unique and of unusual construction. They are based on "DU" bearing material, which is a PTFE and lead mixture keyed into a sintered bronze matrix on a steel backing, approximately 1 mm thick. PTFE is a very poor conductor of heat, and when heated . If the bearing material were used in its natural condition, heat buildup at the sliding interface would cause bubbling in the PTFE material, degrading its utility as a precision bearing. To be truly useful, the

PTFE layer must be very thin, on the order of micrometers, so that heat may be rapidly sinked away from the interface to the bronze and steel backing. The polymer layer is lapped down using a piece of roughened float glass as "sandpaper". Once thinned, the bearing pads are cut out into circular blanks and pressed into spherical shape using a large diameter steel ball bearing. The complete bearing need only be on the order of 1 mm thick.

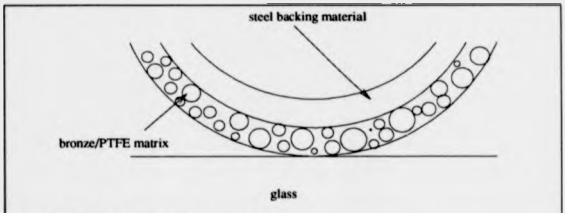


Figure 5.2. Thin film PTFE against glass bearing.

Slideway and bearings

The glass flat used as a counterface for the polymer bearings is a departure from previous embodiments. Polished Zerodur glass-ceramic had been the material of choice for the bearing systems of Lindsey, Smith and Chetwynd. Also, the prismatic shape used on the Nanosurf 2 instrument was discarded for the inexpensive simplicity of a flat slideway with a flat guidebar on one side of the carriage, figure 5.3.

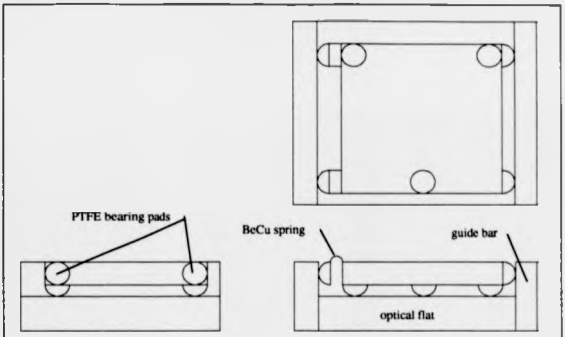


Figure 5.3. Slideway concept.

Since the force profiler design does not call for low-expansion materials, inexpensive BK-7 optical glass was used. An optically flat, $\lambda/10$ square window was obtained[6] with dimensions 100 by 100 mm by 20 mm thickness. This window was then measured in-house on a Zygo MK IV interferometer. The back side of the window was painted black to reduce reflections and fringes off the rear face. Peak to valley flatness over the optic was measured as 460 nm,. The measurement remains somewhat suspect however due to small ghost reflections from the back face. The coating of black paint was left in place so that the slideway flatness could be again ascertained once secured to the profiler frame, figure5.4.

Figure 5.4. Slideway interferogram after gluing to instrument frame.

The flat window was attached to the profiler base by gluing (two-part epoxy) in three locations, corresponding to the Airy points. Finally, although the black layer of paint is very bad for thermal coupling to outside sources, no problems were noted in operation.

The guidebar forming the other face of the slideway is attached to the profiler frame. An opposing guide is attached to the other side of the frame to provide a guide for the preloading springs. These side guides are made of strips of float glass. Since horizontal straightness is normally not paramount in a stylus instrument carriage, the flatness of the float glass, about 2 μm, is considered adequate. The PTFE bearing pads are glued on the aluminum carriage with small amounts of epoxy. Three pads located on the bottom, and two on the side. BeCu preloading leaf-springs are bent in a "U" shape and glued to the opposing side of the carriage. PTFE pads mounted on the leaf spring bear against the opposing strip of float glass.

Normally, for precision applications, the drive point for a carriage must be carefully selected to be at or near the centers of friction and inertia. Since the carriage in this instrument application is moving very slowly and undergoing only small accelerations, no problems are expected if the drivepoint is not located at the center of inertia. The center of friction is the point where the sum of the moments about each of the bearings is equal to zero. Due to the symmetry of the current slideway design, the center of

friction is very near to the center of the carriage. The BeCu leaf springs tend to shift the center somewhat from the geometrical centerline, but because the springs are narrow, the shift is negligible.

Carriage leveling platform

The carriage must contain a leveling mechanism to tilt the specimen, making the plane defined by the specimen surface collinear with the path of carriage traverse. Such adjustments are most easily made by pushing wedges underneath a kinematically located specimen platform. Since the user only has access to the front of the carriage and the drive coupling takes up the center portion, the arrangement shown in figure 5.5 was devised.

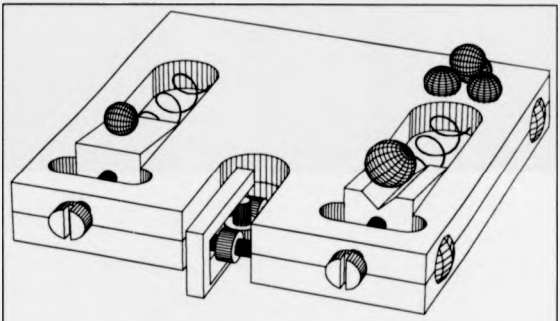


Figure 5.5. Specimen carriage with sample platform removed to show leveling wedges and adjustment screws.

The top specimen mounting plate locates onto the carriage to form what is essentially a Kelvin clamp (or 3-2-1 clamp) with sliding, wedged contacts for the vee and flat. Recessed rubber o-rings, stretched between the top and base plate act as springs to

preload the clamp into place. The three point contact at the rear acts as the pivot point and was deliberately made collinear with the vee contact and the line of travel. This eases leveling somewhat, reducing dependence on the other flat wedge. Software algorithims also ease leveling, directing the user how far, and which direction to turn the leveling screws. One further reason for locating the center of rotation in line with the vee groove is for thermal stability. Thermal expansions tend to be minimized as the balls can slip along the vee and flat, producing only a small amount of tilt. Allowing the vee groove to point in any direction other than the fixed, three point contact reduces the system to a non kinematic constraint.

The wedges are lapped flat on the bottom bearing surface and the interface lubricated with a thin film of paraffin wax. Sideways motions of the vee wedge are constrained by a ball-ended spring-plunger, preloading the wedge against the side wall of the traverse slot. As originally envisaged, the wedges would be coupled to the adjusting screws by hysteretic drive linkages, but space constraints forced the use of low rate springs to preload the wedges against the adjusters. In testing, this arrangement proved entirely satisfactory. Each wedge has an angle of approximately 1.5 degrees, which when coupled with the 40 threads-per-inch nature of the adjusting screws, gives and adjustment of about 37 mrad per turn. Finally, the wedges were shimmed to ensure that the condition of a level carriage was within the adjustment range of the screws. This condition was achieved by setting the profiler slideway and carriage with wedges adjusted to their midpoint, upon a surface plate and running the surface of the carriage against a dial indicator. Appropriate shims were carefully glued to the bottoms of each wedge until the specimen carriage ran true.

Motor drive and couplings

The motor drive system consists of six components.

1. A small DC gearmotor and mounting bracket.

2. Sliding-coupling connecting the motor to the driveshaft

3. Driveshaft tube.

4. Flexible spiral coupling connecting the driveshaft to the micrometer.

5. Non-rotating spindle micrometer head.

6. Crossed radial ball-bearing, non-influencing drive coupling.

The arrangement of these components is outlined in figure 5.6.

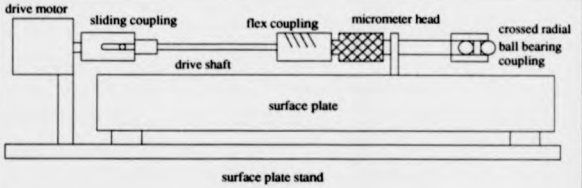


Figure 5.6. Drive system.

This type of drive system was outlined in the paper of Smith *et.al.*[7] and proven in an extensive testing program at the University of Warwick. Due to an unusually high degree of confidence in the drive system, and given the nature of the interferometrically controlled data acquisition system, a low-cost DC gearmotor[8] was selected for motive power. This is a 3600 RPM permanent magnet, 12V DC motor, reduced by a 65.1 to 1 planetary gearbox. Shaft revolutions can range from 6 to 60 RPM depending on drive voltage, corresponding to traverse speeds ranging from roughly 50 to 500 µm/s.

The sliding-coupling was turned from Delrin stock, with a 303 stainless steel drive pin loosely fitting into the Delrin groove for a sliding fit. The driveshaft is simply press-fit into the Delrin. At the other end of the driveshaft, an aluminum spiral coupling fits

over the micrometer thimble and attaches with a set screw. The micrometer head, a Mitutoyo 153-204, has a 25 mm range and a non-rotating spindle. Errant spindle motions and drive axis mis-matches are de-coupled by a non-influencing drive coupling. The coupling is essentially a variation of the crossed cylinders theme. Radial ball-bearings of 6 mm outer diameter are located on 3 mm diameter ground stainless steel shafts. The ball bearings carry almost no axial load, and so are fixed to the shaft by a drop of lacquer. (Lacquer is a good adhesive to use when mistakes are anticipated; a dip into acetone quickly releases the parts). About 0.5 mm, or one turn of the micrometer, is designed into the coupling for hysteresis. Hysteresis is non-consequential in this application because of the laser interferometer and the fact that carriage position is not under closed-loop control. Drive accuracy will be discussed in a Chapter 6 where system testing is presented.

Coarse approach mechanism

The coarse approach mechanism used in the first prototype force profiler utilized an already existing Zerodur glass-ceramic carriage. The coarse carriage on the Nanostep was located by a long vee groove and flat, with three alumina balls providing the kinematic linkage. The design made use of the stick/slip properties of the alumina/Zerodur interface. The carriage was positioned into place by a micrometer head driving a hysteretic coupling. When the desired position was reached, the micrometer was simply withdrawn half a turn to remove it from the measurement loop. While the design had many strong points, it was originally designed to carry a Talystep measuring head. This head is relatively heavy, requiring an intricate spring arrangement to reduce frictional forces against the carriage guideway and provide probe stability.

A small, compact package was desired for force profiler coarse approach mechanism due to the smaller size of the DPT actuator and force probe holder. Several clamping designs were studied and discarded for the same reason: The "Poisson pusher"[9] effect

upon tightening the clamp would have been unacceptable for this application. For this reason, the forces of stick/slip were again utilized in a much more compact design, specially adapted for the DPT actuator. A fused silica tube, 24 mm in outside diameter was selected with an inside diameter just large enough to pass the DPT without touching. The tube was cut slightly longer than the actuator to allow an Invar™ slug to be glued firmly into the end of the glass tube, Figure 5.7.

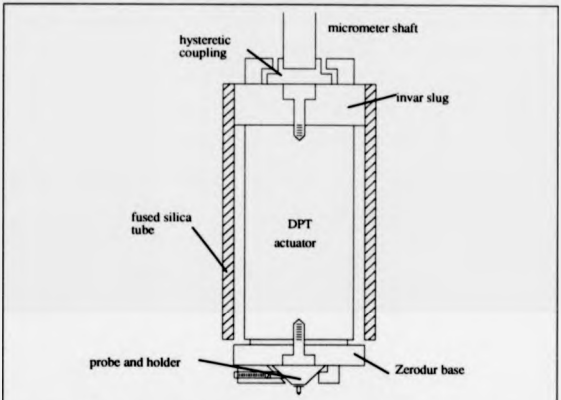


Figure 5.7. Cut-away view of the coarse approach positioner.

The DPT is attached to this slug by a single M3 screw. To allow passage of the DPT control cable, a slot was milled into one side of the tube using the same diamond grinding technique used to manufacture the probes. A hysteretic coupling was attached to the top of the invar slug using cyanoacylate adhesive. Alignment of the coupling is aided by monitoring electrical continuity between the two metal parts. The entire assembly locates into the vee groove of the profiler bridge. Using small pieces of float glass and silicon carbide grit, the sides of the vee were lapped smooth and flat.

Graphite dry lubricant was then keyed into the sides of the vee along the contact lines with the glass tube. A dry lubricant is important here as fluid films, such as paraffin wax tended to flow during experiments. The tube is held into the vee by a strong leaf spring, bearing on the glass with a thin film PTFE pad similar to those used on the specimen carriage. When the probe is adjusted into contact with the specimen, the coarse drive micrometer screw is withdrawn from contact, and static friction holds the assembly in place. Figure 5.8 illustrates the entire coarse approach mechanism.

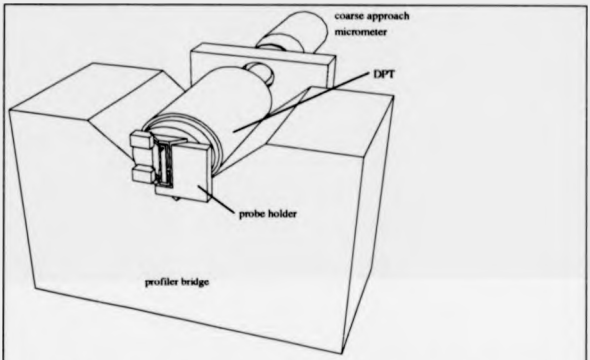


Figure 5.8. Coarse approach mechanism isometric view (retaining spring not shown).

Shown in figure 5.9 is a photo of the profiler system resting on a granite surface plate.

Figure 5.9. Photo of second force profiler instrument.

Profiler data acquisition

A Data Translation[10] DT2823 data acquisition card is used with a 25 MHz "386" personal computer to control the profiler and acquire data. The DT2823 is a full 16 bit card with 4 differential input A/D channels and 2 differential D/A channels. Two unidirectional 8 bit digital ports are available, and are configured for 8 bits in, 8 bits out. Full DMA is implemented on the card and internal crystal controlled clocks may be used, or external clocking and triggering. The ribbon cable connecting the screw terminal connector board and the computer is wrapped several times around a split ferrite core to reduce RF emissions from the PC. The ±10 Volt range of the input and output channels matches perfectly with that of the Queensgate DPT controller. The DPT also accepts an input of ±10 Volts, driving it over its 15μm range of motion. Thus, for the A/D channels, the equivalent least significant bit value is about 0.25 nm. Digital and analog port assignments are listed in the Appendix.

Interferometer phase-unwrapping

Since the interferometer output is an analog phase-encoded signal, it "wraps around" every 360 degrees. Phase unwrapping (also known as fringe counting) is necessary to determine the position of the specimen carriage. The dynamic range of the interferometer is determined by the size of the counter register. In this case, 32 bits gives a maximum equivalent displacement count of 679 m. Although this may seem

ridiculously large, the circuitry may some day be used by others to measure much larger displacements. Figure 5.10 illustrates the interferometer/acquisition controller block diagram.

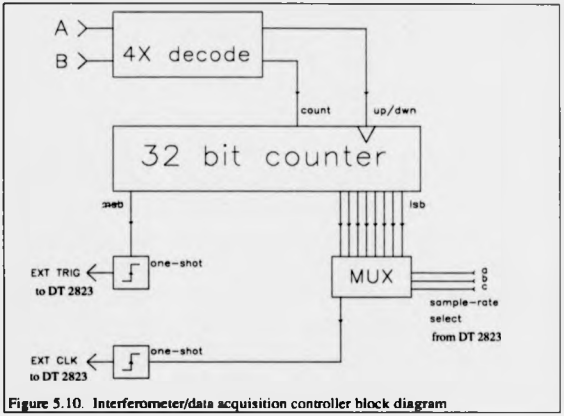


Figure 5.10. Interferometer/data acquisition controller block diagram

The analog quadrature signals generated by the interferometer electronics are carried to the profiler controller chassis via two coaxial cables. Once inside, the analog signals are turned into square waveforms by a comparator circuit. A few millivolts of hysteresis is essential in these comparators to prevent comparator "ringing" at the transitions. The digital phase quadrature signal is then decoded by a 4 state quadrature decoder. The Hewlett-Packard HCL-2020 circuit is a complete quadrature decoder-on-a-chip, incorporating such features as digital filtering and outputs of both count and direction pulses. Each count pulse equals an equivalent displacement of approximately 39.55 nm. Count and direction pulses are then sent to a 32 bit binary up/down counter register. This register consists of eight four-bit counters chained together in cascade fashion. The register is reset via a digital output line from the DT-2823. Register

reads are accomplished by multiplexing the 32 bits onto an 8 bit output bus, connected to the DT-2823 input port. Two bits from the output port are used to control the multiplexing and are set by software to reconstruct the original 32 bit interferometer register in the computer.

Data clocking and triggering

Data from a profile sequence is loaded into the DT2823 automatically, under the control of the interferometer. Since the interferometer register encodes displacement data in a binary offset scheme, the home or "zero" position, is indicated by having all bits in the register cleared except for the most significant bit (MSB). If this MSB is then monitored, we have an accurate signal to begin data clocking into the computer. Although the resolution of the interferometer is some 39 nm, the repeatability of the clocking pulses is several nanometers. The MSB then, is connected to a monostable multivibrator with a 500 ns output pulse, per specifications of the DT2823. Each rising edge of the interferometer MSB issues an external trigger command to the data acquisition system.

Once triggered, the profile data is sampled by using count pulses from the interferometer to clock data into the DT2823. The 39 nm count pulses however, occur too often for long profile scans. To keep profile data files at a reasonable size, say 1024 points, a division of the count pulses is necessary. This divided clock signal is readily available on the least significant bits of the interferometer clocking register. The lowest eight significant bits are multiplexed to a monostable multivibrator connected to the DT2823 external clocking input. Table 5.3 shows the clocking rates available and the profile length for a 1024 point trace.

Scan number	Sampling interval	Profile length (1024 pts)
0	79.1 nm	80.998 um
1	158.2 nm	161.997 um
2	316.4 nm	323.994 um
3	632.8 nm	647.987 um
4	1.266 um	1.296 mm
5	2.531 um	2.592 mm
6	5.062 um	5.184 mm
7	10.125 um	10.368 mm

Table 5.1 Spatial sampling intervals and profile lengths for 1024 points file size.

Sampling interval control is achieved by routing three bits from the DT2823 output port to the multiplexer.

Profile sequence

The profile sequence, once initiated is completely automatic. First, the specimen carriage is positioned into place and a software command issued to clear the interferometer register and reset the position counter. Next, the specimen carriage is backed up approximately 0.5 mm to allow for motor "run-up" and to take up the hysteresis in the carriage coupling. The DT 2823 is then set into full DMA , external trigger, external clock mode. The drive motor is then activated, and upon crossing the "zero" position, the interferometer issues a trigger command and begins data acquisition. Although the DMA mode requires no software intervention, for simplicity's sake the software waits in a loop until the DMA cycle is finished before

shutting off the traverse motor. More sophisticated software could display the profile data as it is clocked into the system.

Profiler software

Compiler selection

Profiler software was written in Asyst 3.0[11] scientific programming language. Selection was made on its several merits: The software effectively isolates the programmer from tedious I/O card programming calls. Asyst uses a system of "templates" for various I/O systems. Programming DMA data acquisition sequences take only a few lines of code. Second, the package has a fairly extensive set of matrix math and statistical tools available. Third, results may be presented in graphical form with just one line of code.

Asyst does exact a relatively high price from the programmer in terms of frustration and obscure documentation. All operations are stack oriented and math is implemented in reverse Polish notation (RPN). While the Author is proficient with RPN calculators, programming software with stack notation requires a major mental shift. Furthermore, system error trapping is extremely poor, requiring the user to exit the program and re-enter again to be sure of removing old commands from the stack. In summation, the program was state-of-the-art back in 1989 when it was purchased.

Profiler commands

<RESET.COUNTER> Resets the interferometer counter register and establishes the "home" position for the profile.

<R> Reads the interferometer counter register and displays carriage position on the CRT.

<REMOTE> Allows remote control motion of the carriage drive motor via the remote pendent. Exit via any keystroke from console.

<SETUP> Reads output of force probe capacitanace gage, takes user force setpoint and issues an analog setpoint value to the force servo system. Useful for negating the effects of capacitance gage offsets.

<PROFILE.GO> Initiates profile sequence. Exit upon completion of trace.

<SAVE123> Saves profile data in a Lotus123 file format.

<SCREEN.PRINT> Screen dump to printer.

<HOME> Returns specimen carriage to the "zero" position plus 0.5 mm for carriage "wind-up" and hysteresis take-up. Useful on repeat traces.

<GOTO> Move carriage to user input position.

<DELTA> Relative 'delta' measurements on screen cursors. Used for step height measurements.

<SPEED> Determines carriage speed by measuring carriage displacement for one second under full speed.

<HARD.LEVEL> Hardware leveling assistance for the carriage leveling screws. A flat surface such as an optical flat is first profiled. The computer then fits a least-squares line to the profile and prompts user to turn carriage leveling adjustment screws.

A screen dump of the profiler control screen is shown in figure 5.11.

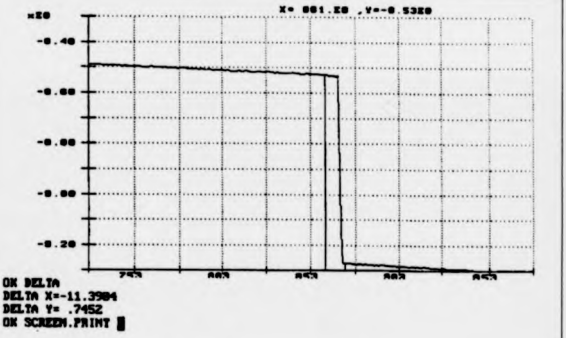


Figure 5.11. Screen dump from profiler computer control system.

Shown in the figure are the measurement cursors which implement data "windowing" functions. Using these cursors an area may be selected for closer study, then "zoomed" in for detailed measurement.

Program listing

A full listing of the software required to operate the profiler is included in the Appendix to provide a more complete record of system integration. Readers should note that while much of Asyst's command structure is cryptic, tasks such as data array graphing and display cursors are easily implemented using very little code.

---

[1] L.P. Howard and S.T. Smith, "Long range constant force profiling for engineering surfaces", *Rev. Sci. Instrum.*, 63(10), pp. 4289-4295, (1992).

[2] CRC-Elsevier Materials Selector, Vol. I, N.A. Waterman and M.F. Ashby, eds., 1991 CRC press, Boca Raton, Fla, p. 29.

[3] A.H. Slocum, Precision Machine Design, Prentice-Hall, NY, 1992, pp. 366-381.

[4] K. Lindsey, S.T. Smith and C.J. Robbie, "Sub-nanometre surface texture and profile measurement with Nanosurf 2", *Annals of the CIRP*, 37, pp. 519-522, (1988).

[5] S.T. Smith, S. Harb and D.G. Chetwynd, "Metrological characterization of polymeric bearings at the nanometre level", *J. Phys. D. Appl. Phys.*, pp. A240-A248, (1992).

[6] Rolwyn Optics, Rochester, NY.

[7] S.T. Smith, S. Harb and D.G. Chetwynd, *ibid.*

[8] Pittman Motor, model # GMT9414B874, 65.5:1 gearbox, Harrisville, Pa.

[9] R.V. Jones and J.C.S. Rivhards, "The design and some applications of sensitive capacitance micrometers", *J. Phys. E: Sci. Instrum.*, 6, pp. 589-600, (1973).

[10] Data Translation, Inc., 100 Locke Dr., Marlboro, MA, 01752-1192.

[11] Asyst Software Technologies, Inc., 100 Corporate Woods, Rochester, NY, 14623.

6(A) PERFORMANCE CHARACTERIZATION

Chapter summary

Force profiler system performance will be characterized in this chapter. Environmental characterization, system dynamics, noise, drift and repeatability have been assessed. The significance of the results and limitations to profiler performance are discused. A precision balance for calibrating the cantilever force probes is introduced. Force balance operation and integration with the force profiler is presented.

Environment

Because of the relatively large scale of the mechanics with respect to the resolution of measurements, environmental variations tend to have a large effect on the performance of nanometric instruments. External and internal vibrations couple into a precision instrument's frame, forcing the designer to either isolate it from the source of vibration and/or design it with a stiff enough metrology loop so that specimen and force probe vibrate together, as one, resulting in little or no relative motion. Temperature effects the dimensions of the metrology loop, necessitating either low expansion materials, a steady-state ambient temperature and "normal" materials and/or a symmetric design where thermal expansions tend to calcel each other. Commercial profiling instruments specify performance over a certain, narrow temperature range along with a maximum level of external noise and vibration permissible.

Vibration

The force profiler is situated on a 1 m square, granite surface plate supported by a standard steel support stand. Between the steel stand and the granite plate are three 25 mm thick rubber mounting pads. The steel stand rests on four rubber levelling feet. The floor underneath the stand is concrete, located over bedrock. No external housing is used over the instrument, but diurnal temperature variations of the metrology lab are less than 0.1 C. Vibration pickup into this instrument comes from two external sources: Airborne and groundborne. Groundborne vibrations are measured using seismic accelerometers, while airborne vibrations can be measured by calibrated microphones. Wilcoxan Research seismic accelerometers (model 731) were used to measure groundborne vibrations. First, the accelerometers were mounted vertically, one on the surface plate and one on the floor, directly beneath it. A Hewlett-Packard signal analyzer (HP35665A) was used to record and analyze the outputs, which were scaled to 1000 V/g. The first record, figure 6.1, shows that the rubber feet and mounting of the surface plate stand are doing a reasonable job of attenuating energy above some 40 Hz.

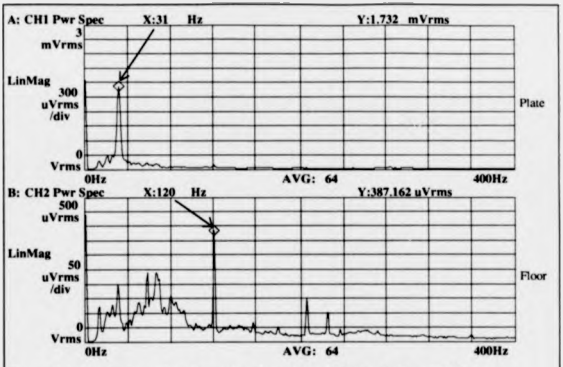


Figure 6.1 Vertical accelerations of surface plate (CH1) and floor (CH2) (note different scales).

The large spike at 31 Hz, is probably the resonance of the plate/stand/mounting system. Next, accelerations in the transverse directions were recorded, with the accelerometer mounted to a steel block, 125 mm on a side, in line with the drive direction of the profiler. A 14 Hz resonance was observed, with another small peak at 31 Hz, figure 6.2 (a). This 14 Hz vibration compares well with the 15 Hz signal observed on the interferometer noise test in Chapter 4.

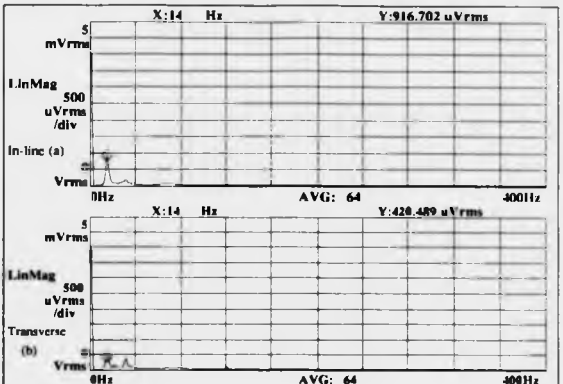


Figure 6.2. Transverse accelerations along the drive axis (a) and perpendicular (b).

Turning the accelerometer 90 degrees, a smaller 14 Hz resonance was observed, figure 6.2 (b), while the 31 Hz peak grew slightly larger. The 14 Hz vibration is most probably due to parallelogram distortions of the metal surface plate stand and shear compliance of the rubber mounting feet.

Temperature, humidity and pressure

The instrument is located in a temperature and humidity controlled metrology lab, kept nominally at 20 °C and 40 % RH. Diurnal temperature variations are kept below 0.1 ° C, with much better surface temperature variations, 0.01 °C is often observed over a typical working day. The room exhibits a cyclic oscillation of approximately 0.03 °C, with a 15 minute period. Air enters the room evenly throughout the entire ceiling by diffusers and is returned through a duct near the floor that spans the length of the room. Effectively, the space is a class 100,0000 clean room. Atmospheric pressure is

monitored by a mercury manometer. When operating the profiler, cotton gloves were worn to lessen the thermal shocks transmitted to the instrument during loading and setup.

System dynamics:

Dynamic characteristics of each device in the feedback loop were determined to aid design of the control system and to verify the operation of the force servo.

DPT

The integral capacitance gage of the DPT provides unique opportunities for system characterization. Since the internal capacitance gage monitors piezo-stack extension with a bandwidth much higher than the closed-loop actuation bandwidth, the internal gage signal may be used to measure the DPT's system bandwidth. If the DPT controller is treated as a three terminal network, a transfer function will result if taken from the internal gaging output ("integrator out" in DPT terminology) and the input from a demand signal. Results were recorded on the HP 35665A signal analyzer system, figure 6.3.

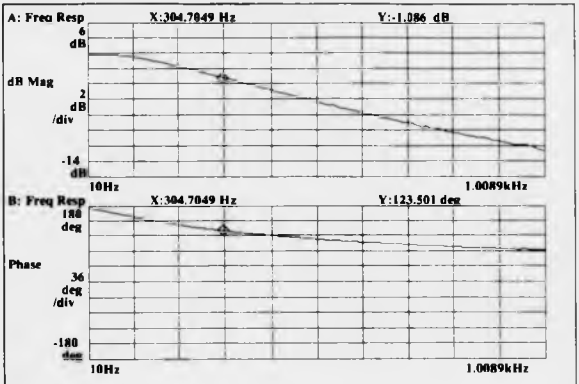


Figure 6.3. Closed-loop response of the DPT actuator.

Note that the -3 dB bandwidth of the system is approximately 300 Hz. Precisely where the bandwidth was set at the factory.

Probe

Having validated the internal capacitance monitor of the DPT for dynamic measurements, the system was then used to measure the transfer function of the cantilever beam force probe. For this measurement, the phase lag of the DPT was effectively removed by measuring between the DPT internal capacitance gage output and the force probe capacitance gage output, figure 6.4.

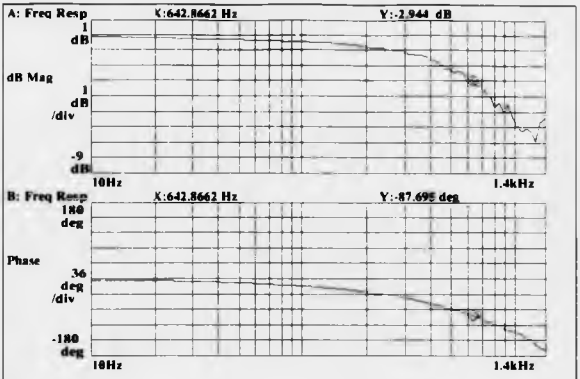


Figure 6.4. Dynamic response of the rapid-replacement force probe.

Note that the response appears to be 2nd order, with a bandwidth of some 640 Hz. Faster probe responses are not necessary due to the fact that the DPT actuator is slew-rate limited to 300 Hz. From figure 6.4 it is estimated that the damping factor ζ of the force probe is approximately 1. Happily, this is considerably less than the well-overdamped condition predicted by the primitive parallel-plate damping model of Chapter 2. Work continues on the development of an accurate model for cantilever damping at nanometer displacements.

Servo, closed-loop

Finally, a closed-loop transfer function of the force servo was recorded. This was measured taking into account all of the phase lags around the servo loop, according to the methods outlined in Chapter 2.

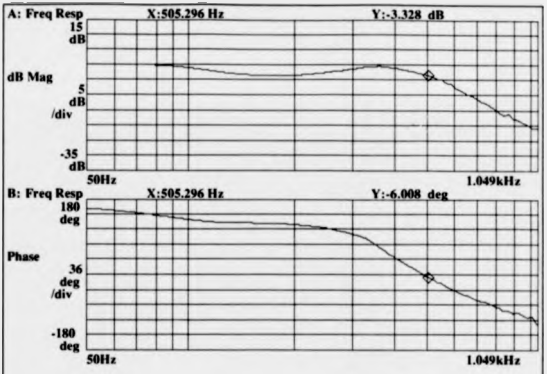


Figure 6.5. Dynamic response of force servo in contact with Zerodur specimen.

This Bode plot shows the system bandwidth to be approximately 500 Hz with a damping coefficient around 0.7. Note that the gain flatness up to the rolloff point is only about 3 dB. This particular measurement was made with a replacement DPT controller, the original at the factory for repair. Earlier measurements with the original controller were flat to better than 1 dB. The point should be stressed that any changes in any of the elements in the feedback loop can bring about undesired changes in instrumental accuracy. It is also important to note that the system response bandwidth is limited primarily by the DPT actuator and that the system damping is controlled by the control system settings, most notably, the derivative, or "D" term. It was also discovered that the response of the closed loop transfer function of the force servo could be radically altered by simply changing the specimens. Specimens with high internal damping decreased the high frequency response of the servo while specimens with low internal damping increased the high frequency response of the servo. This effect and its potential ramifications will be discussed in Chapter 7.

Noise

A thorough evaluation of the contribution of individual elements on the total system noise was performed. In this test, system components were brought into the measurement, one-at-a-time until the complete system was active. The noise measurement was made over a period of 100 seconds with a sampling interval of 100 ms. The measurements were performed in the following order:

1. The A/D system input was terminated with a 50 Ω resistor to represent self-noise from the data acquisition system.

2. The capacitance gage was connected to the A/D system, with the probe not in contact and the DPT actuator off.

3. A Zerodur specimen was introduced and the probe brought into contact some 100 nm.

4. Next, the probe was withdrawn from the specimen, the DPT actuator energized and the noise level recorded.

5. The force probe was once again contacted with the surface of the Zerodur specimen and the noise digitized.

6. The servo-control loop was energized, with a setpoint setpoint value corresponding to some 0.3 mg force. With no filtering present, the results were measured.

7. A 200 Hz low-pass filter was inserted between the control system and the A/D system.

Each measurement adds another noise source so, by superposition they may each be ascertained separately. Figure 6.6 illustrates the various noise levels plotted in units of

apparent profile height. Each trace has been separated an additional 5 nm from the baseline to avoid ambiguity.

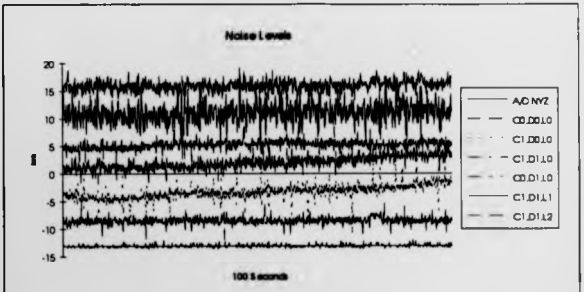


Figure 6.6. Superposition of noise sources in the force profiler.

The combinations in the figure are represented by "C" for the capacitance gage, "D" for DPT actuator and "L" for low pass filter. The "on" state is represented by "1" and the "off" by a "0". It may be seen in the figure that the total peak-to-peak noise level of just the A/D converter is on the order of 1 nm. The unfiltered, p-p noise of the force probe is some 2 nm, with the DPT having no effect on the measurement. Bringing the probe into contact with the specimen adds another nanometer of noise and some instrumental drift, while monitoring the unfiltered servo output brings the noise up to some five nanometers, peak-peak. Filtering the output to the servo bandwidth (300Hz) reduces the noise to a level commensurate with the noise of the capacitance gage only, some 2 to 3 nm peak-to-peak.

Drift

Long-term drift was measured over 4000 seconds, with samples taken every 4 seconds. Four thousand seconds was considerably longer then the maximum profiler

measurement cycle, which was on the order of several hundred seconds. To be truly useful, such drift measurements should also include the ambient temperature and/or the temperature of the instrument frame itself.

Temperature measurements of the instrument frame were made with the National Semiconductor LM35 precision centigrade temperature sensor. More traditional thermistor sensors could have been used, but would have required a more complex data acquisition setup with IEEE-488 bus interfacing to a DMM. The LM-35 provides a 10 mv/°C output which makes for easy A/D interfacing. The specifications guarantee 0.5 °C accuracy with a maximum non-linearity of ±0.25 °C over full range from -55 to 150 °C. Since the range of interest in this application is only one degree range around 20 °C, linearity can be expected to be much better. The sensor was connected as shown in figure 6.7.

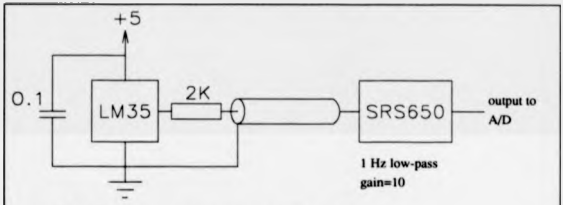


Figure 6.7. Temperature measurement for drift testing.

Since calibration concerned only the range around 20 °C, the sensor was compared to a thermistor thermometer used to monitor the metrology lab and also a precision, mercury thermometer (NIST traceable) used as the lab's temperature standard. Sensor output was amplified by a factor of 10 and low-pass filtered to 1 Hz (12 dB/octave slope) using a Stanford Research Systems SRS650 instrument amplifier. Results were very encouraging, within 0.1 °C of the calibrated thermometer and a noise level of only 0.16 mv AC RMS at 1 Hz bandwidth. The minimum resolvable temperature is then

1.6 milli-degrees, fully adequate for drift testing. The circuit's plastic package was coated with a thin layer of silicone heat-transfer grease and attached to the center portion of the force profiler bridge, near the coarse approach mechanism. Data was recorded by the computer and is shown in figure 6.8.

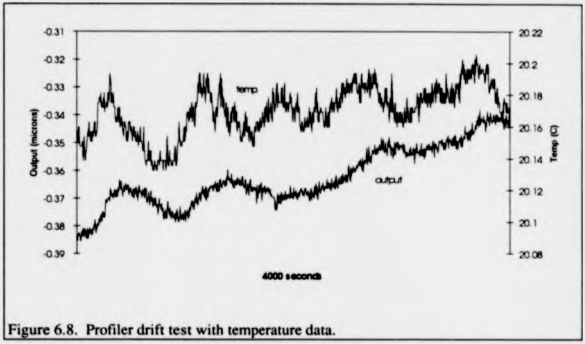


Figure 6.8. Profiler drift test with temperature data.

As can be seen in the figure, the room temperature exhibits cyclical variations with a 15 minute period. Clearly, the profiler follows the frame temperature with a slight lag, as expected for a first-order system. Any attempt at thermally compensating the system, however would require a more detailed study of temperature effects, with multiple sensors on the instrument frame. Even so, without compensation, the profiler exhibits a thermal expansion coefficient of about 320 nm/°C, consistent with a piece of aluminum some 15 mm long. This compares with the 7 nm/°C of the Nanostep instrument used for the prototype profiler. The larger drift here is acceptable because of better temperature control and a greatly reduced measurement cycle time.

Carriage velocity

Velocity characteristics of the specimen carriage were examined using the laser interferometer. The specimen carriage was traversed at a nominal speed of 200 µm/s while the carriage position was sampled every 20 ms for a total of 1024 points. These data were then transferred to a spreadsheet (Microsoft Exel 4.0) where it was backwards-time differentiated to determine velocity. Position and velocity records were then subjected to fast Fourier transforms. Velocity -vs- time data were graphed as percentage variations from nominal (figure 6.9), while the FFT of these data are displayed in "cycles per turn" format to determine relationships to the micrometer leadscrew (figure 6.10).

Figure 6.9. Velocity variations of the specimen carriage.

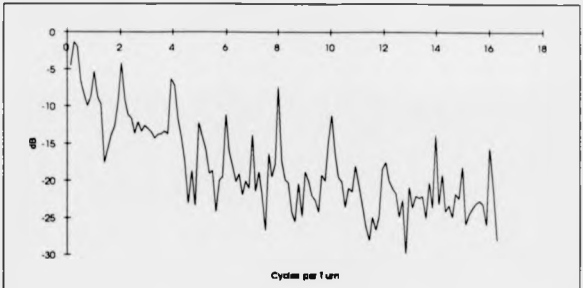


Figure 6.10. Fourier transform of velocity variation displayed in "turns-space".

The first figure shows velocity variations of some 20 % from nominal. Examining the second figure, the Fourier transform, reveals strong variations at integer periods of cycles per turn. It is difficult at this point to separate the errors due to gearbox noise from the errors due to leadscrew inaccuracies. However, since the leadscrew in question is a fairly high accuracy device (a few microns per turn), with a pitch of 500 μ m per turn, the velocity variations caused by the micrometer could be at most 0.01 %. So, it is concluded that most of the errors visible in figure 6.10 are caused by the motor gearbox. The gearbox in question is a 65.5 to 1 planetary type, not designed for precision applications. If the interferometric data acquisition system is truly effective, these velocity variations should not manifest themselves in the profiler measurements (although some second-order effects from rate dependent damping of the cantilever may be present). Experiments described in Chapter 7 will examine these velocity dependent effects. Note that the velocity variations caused by the use of the low-quality drive motor serve to prove a point about the effectiveness of a laser interferometer controlled data acquisition system.

Step height characterization

An old and rather well-used step height standard was used to calibrate the profiler. The standard is a chromium pattern on a 50 by 50 by 4 mm thick glass substrate. The specified height of the steps were 740 nm, which were checked and verified on a commercial stylus instrument[1]. The step specimen was then measured with the force profiler and the step height evaluated using the measurement cursors in the profile display software. A small correction factor (-0.9326) was then incorporated in to the software to calibrate the measurements. Note that the last digit in the correction factor represents the tenths of nanometers place. It is only expected that the profiler linearity be as good as the linearity of the DPT actuator, specified as 0.1% over a 15 μm range.

Repeatability:

Repeatability was examined by measuring successive profiles over nominally the same trace at least twice, then subtracting the raw, unfiltered data to determine a worst-case repeatability.

Steps

Figure 6.11 shows the same 740 nm step standard was profiled with a 10 μm stylus at a speed of 80 μm/s. After subtracting raw, unfiltered data, the profile of the repeatability is computed, figure 6.11.

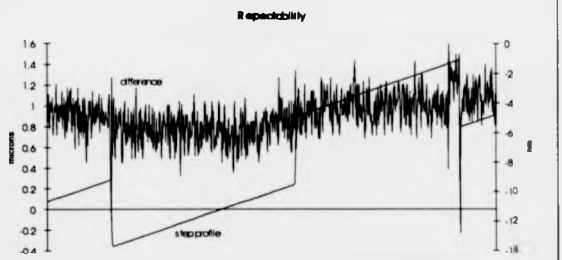


Figure 6.11 Repeatability of profiles showing subtraction of raw, unfiltered data. Step profile axis on left (μm), difference axis on right (nm)

Note that at the steps there is sometimes a small spike, a result of this, a worst-case test artifact. From the figure, we may deduce the worst-case repeatability to be about 3 nm. If the data were filtered and step heights extracted, repeatability should be on the order of system noise, about 0.5 nm RMS.

Table 6.1 summarizes force profiler performance at the time of this writing:

Probe tips	2, 5, 10 μm ISO spherical diamond. Berkovich 20 nm cleaved diamond.
Contact force	between 30 μN and 10 nN (3 mg and 0.01 mg force)
Force servo bandwidth response	300 Hz
Vertical range	15 μm
Vertical resolution	0.25 nm
Noise	0.5 nm RMS @ 1 Hz, (DPT limited)
Horizontal traverse range	20 mm
Horizontal traverse speed	between 50 and 500 μm/s

Table 6.1. Profiler performance summary.

It may be seen that the speed of operation of the force profiler in its current embodiment is some two orders of magnitude faster than typical stylus instruments, which often traverse at a speed of 5 μm/s. The implications of high speed operation are clear for large-area three-dimensional profilers. Such instruments typically take tens of minutes to measure flat panel displays. An increased operational speed would reduce profile errors due to thermal distortions in the instrument frame.

---

[1] Tencor Alpha Step 200, Tencor Instruments, Sunnyvale CA.

6(B) FORCE PROBE CHARACTERIZATION

This section is paraphrased from a paper co-written by the candidate[1].

CHAPTER SUMMARY

Precision balances have been applied to a variety of problems in analytical chemistry, experimental gravitation, mass artifact standards and *in-vacuo* thin film thickness measurement. Balances in use at national standards laboratories[2] have mass spectral densities of approximatley 10^{-8} Kg $Hz^{-1/2}$, with typical frequency ranges of between 0.03 to 2 Hz. Beam balances for gravitational wave experiments have demonstrated resolutions of approximately 30 pg in 1.5 g although the natural period in these experiments is of the order 800 seconds therefore requiring long measurement times[3]. Described herein is an inexpensive precision balance having a 30 Hz bandwidth and microgram resolution for application to the in-situ calibration of modern force microscope probes.

Current techniques for the calibration of such atomic force probes typically involves determining the spring constant of the cantilever beam used to position a sharp tip in contact with a specimen. In practice, the linear spring constants for such cantilever beams are predicted from a determination of its critical dimensions which are then used as the basis of a thoeretical model. Recently Stewart and Parker[4], and Cleveland et. al.[5], have derived a technique from which the spring stiffness is determined by measuring the change in resonant frequency of the cantilever when masses are added at the end. This however relies on a accurate knowledge of both the position and value of the applied mass, Maltbaek[6], therefore making it difficult to assess the accuracy of the measured results. A first and important assumption is that the cantilever behaves

according to simple, uniform beam theory. For small displacements, the theoretical stiffness at a distance, L, along a prismatic cantilever beam is given by

$$F(\delta) = \frac{EBH^3\delta}{4L^3} \tag{6.1}$$

where E is Young's modulus of the material, B is beam width, H is beam thickness, δ is deflection perpendicular to its major axis and L is the shortest length of the beam between the rigid support and the line of action of the probe contact. Clearly the inverse cube power law makes an estimation of the true length very dependent on having perfect boundary conditions. That is, zero strain at the base of support, and no residual stresses in the beam itself. In reality, there exists considerable stresses in the

Figure 6.13. Microinterferogram of a commercial silicon nitride AFM cantilever (UNC Charlotte, Zygo Maxim photo).

thin films used to build microcantilevers, figure 6.13.

Differences in thermal expansion coefficients and the heat used in many thin-film processes ensures that the force probe microstructures have considerable warpage. If the metrology cantilever is of the more recent single crystal silicon designs, the internal stresses may be low, but the material has well known anisotropies along the crystallographic planes. Furthermore, most cantilever designs, often employing both

cross section and axial shape variations deviate considerably from the simple rectangular beam, compromising the validity of simple models still further. Ignoring these effects, the fractional error in the predicted force on a cantilever beam due to a deflection, δ, can be estimated[7] by applying the chain rule to 6.1,

$$\frac{dF}{F} = \frac{dE}{E} + \frac{dB}{B} + 3\frac{dH}{H} + \frac{d\delta}{\delta} - 3\frac{dL}{L} \qquad (6.2)$$

Assessing each term in the above equation enables an evaluation of the possible fractional error in the predicted stiffness as a function of geometric and materials property measurement tolerances. Methods attributed to Peterson and Guarnieri[8] to measure E for thin film cantilever resonators depend on determining the dimensions of the structure and it's resonant frequency to produce an estimate of this value. This method is subject to virtually all of the limitations of the cantilever force measurement i.e geometrical uncertainties and validity of simple models. Estimated uncertainties of such techniques can be as high as 50%. To accurately measure the geometric parameters of a micro-cantilever, the beam thickness may be measured ellipsometrically to within 0.2%, at best, width and length may be measured in a Metrology SEM to about 1% [9] (at best) and deflection considerably less accurately. However, in most laboratories these facilties are not readily available and more commonly the contact force of an AFM can only be estimated with confidence to within, at best, a few percent and more generally 20 to 50 %. Clearly, an independent means of force probe calibration is desired.

PRINCIPLE OF OPERATION.

A cross section through the low force balance is shown in figure 6.14.

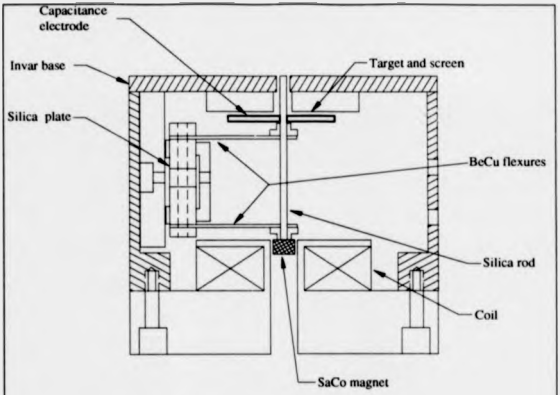


Figure 6.14. Cross section of force balance instrumentation.

Apart from the berylium copper leaf springs, all components are circularly symmetric and the outer diameter of the body of the instrument is 40 mm. The working mechanism consists of a silica rod that is suspended by a solenoid/magnet force actuator consisting of a samarium cobalt permanent magnet 2 mm diameter by 2 mm long surrounded by, but not in contact with, a copper coil having inside and outside diameters 4 and 10 mm and a length of 9 mm. These dimensions were chosen for ease of manufacture and because they are reasonably close to the optimum ratio for maximum force per unit power input[10]. Using 0.2 mm diameter windings the coil resistance is approximately 4 Ω and, from the proceeding calibration, the actuator sensitivity is approximately 0.014 mA μN^{-1}.

Position of the rod is monitored by a flat plate circular capacitance gauge with one electrode consisting of an aluminum coated 12 mm diameter glass cover slide 100 μm thick bonded to the rod and the other, an aluminum coated Zerodur optical flat,

attached to the top plate of the balance. Two beryllium copper flexures having an effective length of 5 mm, a width of 2 mm and 25.4 µm thickness have been bonded to the silica rod to form a simple, leaf type linear spring. Beryllium copper was chosen because of its high yield strain and therefore its resilience to accidental damage. From simple beam bending theory the force or mass to be measured is then applied to the top of the rod. Thus causing a deflection of the springs which will be monitored by the capacitance electrode and this signal amplified and fed into the force actuator through a proportional plus integral plus derivative controller (PID controller) and voltage to current amplifier. A block diagram of the complete system is shown in figure 6.15.

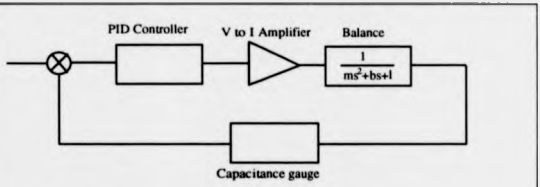


Figure 6.15. Block diagram of balance control system.

Essentially this is a null balance in which the displacement of the central rod due to any load or applied mass is nulled by the force actuator. This ensures that the forces acting on the platform other than those from the applied load and actuator reaction are unchanged during the measurement process. With the magnet positioned at the optimal point along the axis of the coil, the applied current is a direct monitor of force on the magnet. This actuator consists of a magnet attached to the specimen platform with its poling axis colinear with the platform displacement axis. Surrounding the magnet, but not in contact, is a solenoid coil attached to a rigid datum. Energization of the coil will result in a force on the magnet that is proportional to the current. Optimal design of this type of actuator is discussed in the paper of Smith and Chetwynd. If the magnet is positioned at the optimal point along the axis of the coil, to

a first order approximation the force on the magnet does not change with relative position. Therefore for small displacements, due to instrument instability, thermal effects etc, the force characteristic of the actuator remains constant. In mathematical terms the actuator has infinite compliance. This removes the requirement for high positional stability between the magnet and coil. Therefore the voltage signal to the current amplifier can be monitored and this then proportionally related to the external force on the rod. There are a number of issues with such a device at the extreme sensitivities required for measurements at nano Newton levels. Firstly, the rod is floating in ambient air and will thus be susceptible to environmental changes as also will be the capacitance gauge. Temperature will also degrade the remanent values of the magnet which, for samarium cobalt, will be of the order 0.04% per degree Celsius and therefore considered negligible. Ground and airborne vibrations limit bandwidth performance. To attenuate this influence the balance is mounted on an isolation stage consisting of a cardboard honeycomb/wood sandwich mounted on a concrete block of dimensions 0.7 x 0.7 x 0.04 m which is in turn resting on inner tubes of inside and outside diameters of 100 and 220 mm respectively. Transmission of vibration between the floor of the building and the upper surface of the isolation stage have been measured using siesmic accelerometers. The frequency response is plotted in figure 6.16.

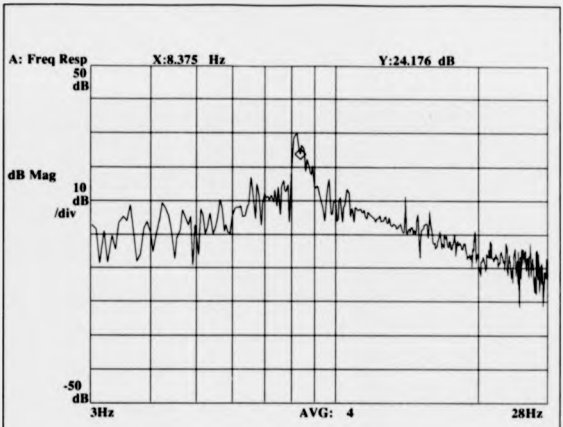


Figure 6.16. Frequency response of vibration isolation stage to groundborne noise.

This plot shows a resonant peak at approximately 8 Hz with attenuation over frequencies of greater than 14 Hz. Electronic and acoustic noise is isolated by housing the complete instrumentation in a 12 mm thick foam enclosure with a conducting foil coating that can be grounded to form a Faraday cage.

All experiments have taken place in a temperature and humidity controlled laboratory maintained at 20±0.05 °C and 40±5 %RH respectively.

DESIGN CONSIDERATIONS

The silica rod is constrained by a simple flexure of stiffness k introducing an uncertainty in our measurement that is related to the resolution of the null sensor, r, by

$$\delta F = kr \qquad (6.3)$$

Simple beam bending theory (Eq. 6.1) gives a value for the stiffness of approximately 36 N m^{-1}. Combining this with a sensor noise equivalent displacement of approximately 1 pm Hz$^{1/2}$, theoretically it should be possible to resolve sub nano-Newton loads at relatively low bandwidths. However, the output signal is taken from the servo controller output to the current amplifier and there will be a large number of noise sources contributing to the signal. Typically, the electronic components will impose shot noise as well as flicker noise that is often observed to be inversely proportional to frequency. Superimposed on the signal will be noise due to mechanical disturbances. Before looking at the system noise level, it is informative to assess the mechanical characteristics of the balance. Because of the high stiffness of the rod and magnet relative to the flexure, it is reasonable to use a lumped second order model of the form

$$m_e \ddot{x}(t) + d\,\dot{x}(t) + kx(t) = f_v(t) \tag{6.4}$$

where $x(t)$ is the vertical movement of the rod, m is the equivalent mass and f_v is the applied vertical force from probes etc. **Note that while the purpose of this instrument is to reduce reliance on such second-order models for AFM cantilevers, the use of such a model here is justified because the system will later be calibrated using dead weights.**

Using Rayleigh's criteria, the equivalent mass can be computed from that of the rod, magnet and capacitance electrode plus a third contribution from the springs. Calculating these values from products of density and volume the equivalent mass is approximately 0.25 g giving a natural frequency of 52 Hz. Based on previous experience it was expected that squeeze film effects between the capacitor electrodes would provide the a major contribution to damping properties of this system. The damping force between two circular plates separated by a small nominal distance, h, can be computed from the equation[11]

$$d = \frac{3\eta\pi(r_1^2 - r_2^2)}{2h^3 P_a} \tag{6.5}$$

where η is the viscosity of air (taken to be 17×10^{-6} Ns m^{-1}), Pa is the absolute pressure (taken to be 10^5 N m^{-2}) and r_1 and r_2 are the outside and inside electrode radii (taken to be 0.006 and 0.0015 m). Measuring the capacitance, C, between the electrodes, it is possible to estimate a value for h from the equation

$$h = \frac{\varepsilon_o \pi (r_1^2 - r_2^2)}{C} \tag{6.6}$$

For zero current through the coil, a capacitance of 18 pF was measured using a Hewlett Packard 4284A Precision LCR meter. It is noted however, that this value includes stray and cable capacitance and the real value for the average separation is likely to be larger than the predicted value of 50 μm. However, substituting this value into equation (2) gives a damping constant of approximately 0.2 N s m^{-1}. To test these predictions, the dynamic response of the balance was measured using a Hewlett Packard 35665A Dynamic Signal Analyzer. Open loop performance was measured by feeding a swept sine signal into the current amplifier while monitoring power spectrum of the subsequent output, figure 6.17.

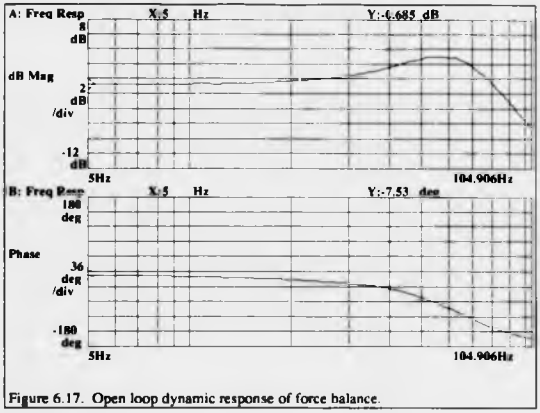


Figure 6.17. Open loop dynamic response of force balance.

From this figure it may be seen that there is a 90° phase shift at 61.1 Hz and a second order response characteristic with a critical damping factor of 0.34. Consequently, a 30 Hz bandwidth was considered a reasonable upper bandwidth limit. Using appropriate time constants and with additional fine tuning using the Zeigler-Nichols criteria the closed loop response of figure 6.18 was achieved.

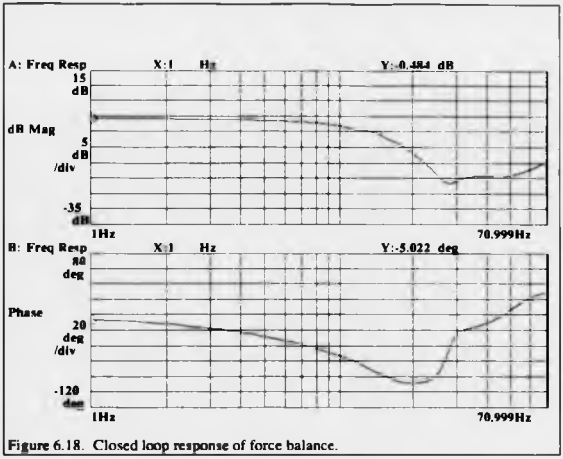


Figure 6.18. Closed loop response of force balance.

This is the response of the output to a desired change in the intput set point to the controller. The peak amplification occurs at a frequency, ω, given by

$$\omega = \omega_n \sqrt{1-2\zeta^2} \tag{6.7}$$

An undamped natural frequency, ω_n, of approximately 500 rad s^{-1} is predicted. Damping ratio may be computed from the equation

$$\zeta = \frac{b}{2\omega_n m} = \frac{b}{4\pi f_n m} \tag{6.8}$$

Substituting into (6.4), a critical damping ratio of 0.8 is predicted (compared to the predicted 0.34). It is felt that the difference between the measured and predicted value is reasonable considering the number of assumptions within the model and also the extreme sensitivity of this value to the predicted separation and electrode geometry and

the uncertainty in many other parameters. For example, a reasonably small stray capacitance of 3 pF would change the predicted separation by 12 μm resulting in a factor of two reduction in the predicted critical damping ratio to a value of 0.4.

Having experimentally determined the second order characteristics of the balance, the closed loop response was then assessed. In tuning the controller for this application, our main objective is the reduction of noise. Consequently the closed loop response is designed for a relatively sharp cut off at around 30 Hz. This is shown in figure 6.18 which shows the magnitude and phase response to a swept sine applied to the set point. Because of the unknown characteristics of the noise sources, the PID controller was manually tuned while continuously observing the noise output from the balance after which the balance noise spectrum shown in figure 6.19 was achieved.

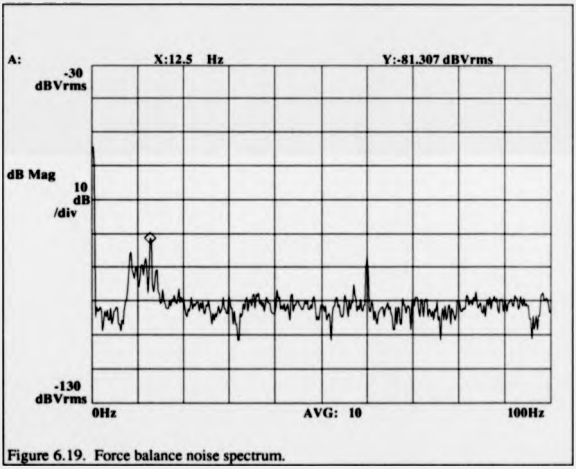


Figure 6.19. Force balance noise spectrum.

During tuning it was possible to obtain a reduction of the broad band noise levels to approximately -90 to -100 dB which is considered to be close to that for the combined electronic sources from amplifier, controller and sensor. However, it can be seen that there is a -80 dB noise source at approximately 12.5 Hz which is thought to be mechanical disturbance. This could not be attenuated further and it would probably be easier to employ more elaborate vibration isolation from both air and groundborne sources. For longer term monitoring and the determination of sampling strategies, the general stability of the balance was assessed. Figure 6.20 shows the gage output over a period of over one hour.

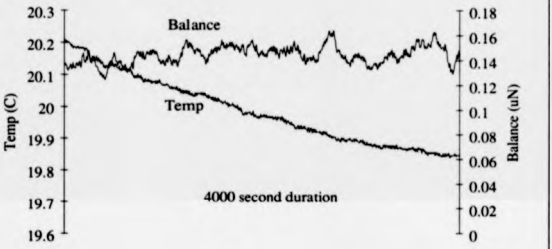


Figure 6.20. Drift measurement over 67 minute period.

Apart from an average drift of around 0.01 μN with a temperature change of 0.35 K or 0.03 μN K^{-1}, there is another component characteristic of flicker, or 1/f noise, which can be seen from the spectrum (not shown). Above a frequency of approximately 60 mHz or for time periods of less than 17 s the noise level reduces to below -120 dB indicating that data should be sampled at a frequency much higher than 12.5 Hz and for a time period in each signal of less than 17 s. At this level resolution of 1 μV is implied although this is likely to be difficult to achieved practice. Typically a sampling

rate of 100 Hz for 10 seconds is considered to be reasonable for each data point from the balance readout.

EXPERIMENTATION

Two experiments are presented, one being the calibration of the balance using dead weight loads. In the other, the variation in force as a function of the distortion of a force probe is presented.

In the first experiment a series of rectangular strips were sheared from brass foil of 27 μm thickness. Their mass was then calculated as the product of density and volume. The largest uncertainty occurs in the measurement of the length between the two parallel faces. From many measurements, a repeatability of better than 0.1 mm could be achieved using manually operated calipers. Being a straightforward product of these dimensions, the error in mass is simply the sum of ratios of individual errors to their dimensions. Consequently, for the smallest mass of 1.2 mg having dimensions of 1.7 by 3.1 mm the predicted error is less than 10%. A total of eight masses were prepared covering a range from 1.23 to 8.33 mg and their relevant geometries and predicted error margins are reproduced in table 6.1.

Table 1: Major dimensions of brass foil masses

Mass number	Dimensions (mm)	Mass (mg)	Predicted error (%)	Experimental error (rms %)
1	4.45x8.00	8.33	3.5	1.045
2	4.65x7.70	8.28	3.4	0.745
3	2.75x8.00	5.12	4.9	2.358
4	4.20x4.00	3.90	4.9	1.606
5	3.55x3.70	3.01	5.5	1.528
6	1.55x8.00	2.88	7.7	2.659
7	1.70x5.05	2.00	7.9	2.657
8	1.70x3.10	1.23	9.1	2.960

Table 6.1. Dimensions of calibration masses.

These were then placed on the balance and the corresponding change in output monitored by a Keithley 199 digital multimeter. Measurements were then repeated at least six times for each mass and two of these masses were measured after an elapsed time of 7 days.

The objective of this second experiment was to measure the interfacial forces between the force probe and the balance. The stiffness values are derived by contacting the force probe against a solid surface and reciprocating the DPT actuator a known distance. The slope of the force probe output to displacement can then be used as a

measure of gage sensitivity. Alternatively, a method not used in these experiments would be to store the subsequent calibration results in a look-up table for future reference.

RESULTS.

A. Balance calibration

Results from the calibration are shown in figure 6.21.

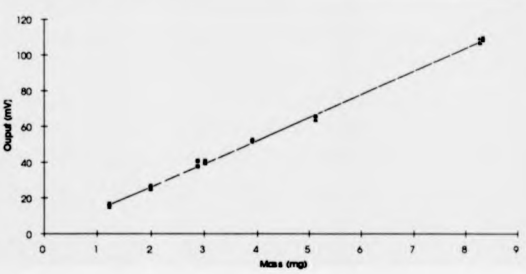


Figure 6.21. Dead weight calibration of force balance.

Each data point was obtained by carefully dropping the dead weights onto the balance and noting the subsequent change in output voltage. At least six repeats were performed for each mass. It should be noted that there appear to be only six data sets on this graph. This is because, although each mass was constructed from foil sections of different dimensions, two pairs have very similar mass values and are therefore indistinguishable on the graph. Also, masses 2 and 5 were measured after an elapsed time of 7 days. These data points have then been fitted to a least squares straight line of slope 13.05 mV mg^{-1} and an offset of 0.0012 mV. These values yield a correlation coefficient of 0.998 and a standard error of 1.1 mV implying a reasonably linear

relationship between mass and output voltage. The magnitude percentage error of each measurement from the least squares line has been computed and the average value for each experiment is also produced in table 1 are within the predicted error band. These values for the mass are not traceable masses but represent the precision capability of the balance. However, the density is known to better than 1% and combined geometric errors to better than 9% resulting in an uncertainty of approximately 10%. From the central limit theorem, it would be expected that the calibration is accurate to considerably better than this over the calibration range. For the quoted standard error we may again quote a resolution of better than 0.1 mV for an average of 100 samples which corresponds to an applied mass of 7 μg or 70 nN and a balance precision of better than 1 part in 10,000. During this calibration procedure, manual positioning and placing of the masses required considerable dexterity. A further calibration was performed using class I calibrated masses of 1, 2 and 5 mg (supplied by Cooper Instruments, VA). From this sensitivity of 13.01 mV mg^{-1} was measured which has been used for all subsequent plots.

It is also of interest that the balance would, on occasion, be accidentally contacted with excessive force resulting in visible displacement of the balance rod. These large perturbations did not seem to cause significant changes in the balance set point implying an extremely good recovery of the flexure and electronic system. Also, even longer term stability is observed in that the balance can be left for many days and when switched on is usually within 5 to 10 mV of the set point value.

B. Force probe characteristic

In this second experiment a cantilever force probe was contacted with the balance and displaced using the piezoelectric actuator. Extension of this actuator was measured using its internal capacitance gage. A graph of the force applied to the balance as a function of the force probe output for a ramp displacement of approximately 330 nm is shown in figure 6.22.

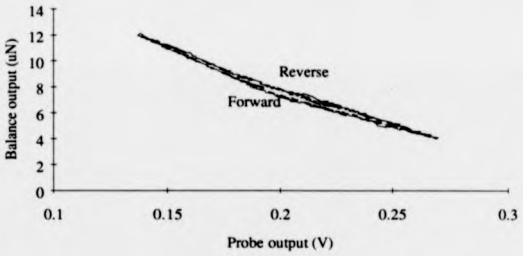


Figure 6.22. Balance output against setpoint control of constant force probe.

Two distinct slopes can be observed within the forward and return probe displacements. The slope at which contact has just been made varies from between 11.9 and 14.6 mV/μN at initial contact and 19.5 and 21.1 mV/μN towards the middle of the ramp cycle. All that is required to determine the stiffness of the probe system is to calibrate its sensor output against relative displacement. This is far from trivial at these levels. By contacting the probe against a solid Invar surface and displacing the actuator a distance of 500 nm the force probe output as a function of displacement is obtained, figure 6.23.

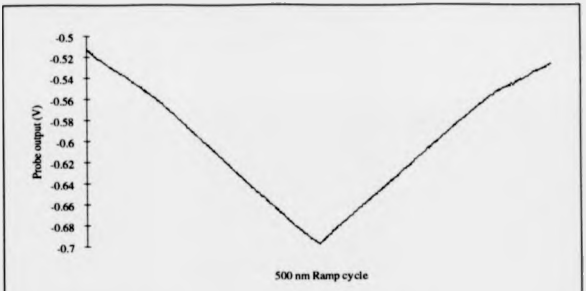


Figure 6.23. Force sensor output response to a 500 nm ramp cycle.

From this graph two slopes of output against displacement are observed and these correspond to sensitivities of 3.8 and 2.6 μm/V. Using these sensitivities in the two corresponding regions in figure 8, four stiffness values of 45, 56 and 51 & 55 N/m can be derived. This average stiffness of 52 N/m is encouragingly close to the value of 53 N/m predicted from simple beam theory for the above dimensions. The change in slope of the sensor output with displacement could be a combination of effects. The probe tip in this case is a Berkovich diamond of with a tip radius of less than 20 nm. This could easily be indenting the surface. Other factors include; capacitance value being a nonlinear function of the tip displacement[12], the possibility of "tin canning" or twisting of the cantilever if it has any curvature along its axis or any nonlinear and relatively low stiffnesses in the metrology loop.

As an example of the diagnostic value of the balance for this application, figure 6.24 shows the balance output from a ramp of approximately 500 nm with a force probe having an anomalously high stiffness.

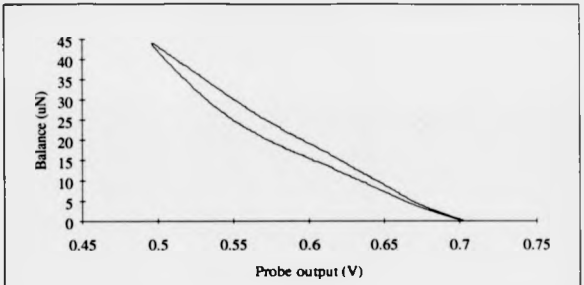


Figure 6.24. Force voltage characteristic of a probe showing anomalously high stiffness.

The measured stiffness of approximately 90 N/m is much higher than the predicted value of 40 N/m. Subsequent inspection of the probe revealed that the cantilever and base were in physical contact.

---

[1] S.T. Smith and L.P. Howard, "A precision force balance with application to the calibration of atomic force probes", Rev. Sci. Instrum.. Accepted for publication, Sept. 1993.

[2] C.C. Speake, *Proc. Roy. Soc. Lond.*, A414, p. 333, (1987).

[3] C.C. Speake and G.T. Gillies, *Proc. Roy. Soc. Lond.*, A414, p. 315, (1987).

[4] A.M. Stewart and J.L. Parker, *Rev. Sci. Instrum.*, 63(12), p. 5626, (1992).

[5] J.P. Cleveland, S. Manne, D. Bocek and P.K. Hansma, *Rev. Sci. Instrum.*, 64(2), p. 403, (1993).

[6] J.C. Maltbaek, *Int. J. Mech. Sci.*, 3, p. 197, (1961).

[7] S.T. Smith and D.G. Chetwynd, Foundations of ultraprecision mechanism design, Gordon and Breach, London, (1992).

[8] K.E. Peterson and C.R. Guarnieri, *J. Appl. Phys.*, 50(11), p. 6761, (1979).

[9] M.T. Postek, A.E. Vlader, S. Jones and W.J. Keery, 'Report on the NIST low accelerating voltage SEM magnification standard interlaboratory study', *SPIE*, at press, 1993.

[10] S.T. Smith and D.G. Chetwynd, 'Optimisation of a magnet/coil force actuator and its application to linear spring mechanisms', *Proc. Inst. Mech. Engr.*, 204(C4), pp. 243-253, (1990).

[11] W.A. Gross, Gas Film Lubrication, J. Wiley and Sons, Inc. NY, 1962, p. 318.

7 EXPERIMENTAL RESULTS

Chapter summary

In this final chapter, the results of studies using the force profiler are presented to illustrate a variety of nanometer-scale phenomena and indicate some possible future applications of this technology. It will be demonstrated that the force profiler exhibits superior traversing speeds in comparison with conventional stylus instruments and that nanometer features may be examined with traverse rates up to one hundred times faster. New and unexpected uses for the technology were found, in particular the potential for measurement of material damping properties of the microscopic contact region. Finally, a study to determine damage thresholds for profiler contact force levels is proposed.

Variable traverse speed repeatability

To determine the effects, and limits of profiler traverse velocity on the information in the profile data, a series of tests were conducted on a variety of samples at dissimilar speeds.

Step height standard

The (by now thoroughly tested) step height standard was once again employed for a series of repeat profiles. Using a 2 μm stylus with contact forces of 5 μN, the traverse speed was varied from 0.077 mm/s to 0.548 mm/s.

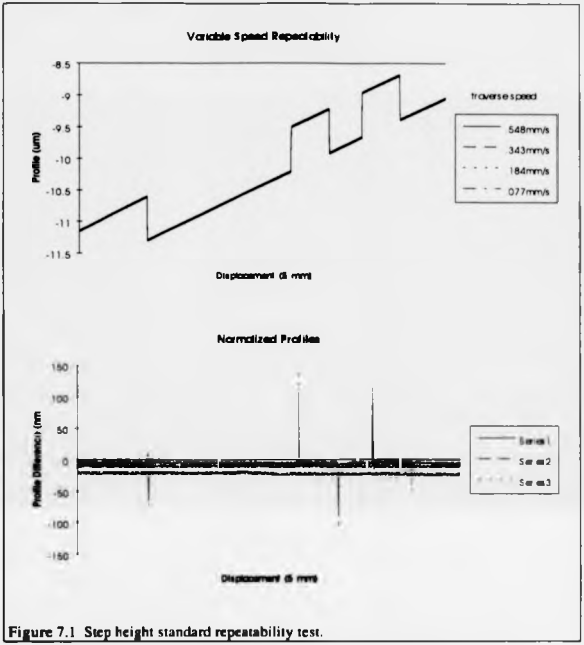


Figure 7.1 Step height standard repeatability test.

As shown in the figure, a few differences can be seen between the traces, even at high (.548 mm/s) traverse speeds. The lower trace is composed of direct subtractions of raw data, normalized to the first profile. The nature of the "spikes" in this plot could be the result of the profile not tracing the same exact path. No conventional stylus instrument can match this performance at high speeds and low forces.

Compact disc

A PMMA compact disc was traced radially outward, again using a 2 μm stylus with 5 μN (0.5 mgf) contact force. System bandwidth was limited to 800 Hz by a low-pass filter in the control system. Profiles were taken at traverse speeds of 52 μm/s and 182 μm/s, the results are displayed in figure 7.2.

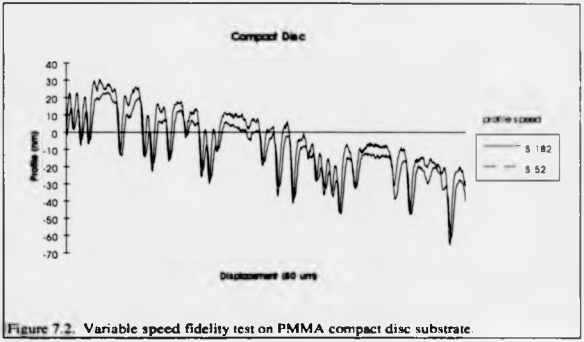


Figure 7.2. Variable speed fidelity test on PMMA compact disc substrate.

Note that this test is especially difficult for a stylus instrument. The Pits on a compact disk are some 20 to 30 nm deep and only slightly larger than 1 μm across. Furthermore, the stylus tip traces a path through many different pits, tracing straight across some, while only skirting the edges of others. Lateral repeatability of the instrument must be extremely good to measure such structures. As shown in the figure, the shallow pits are easily resolved. However, some loss of high frequency information is evident in the higher speed profile. This frequency dependence is expected when measuring surfaces with micrometer-sized features at high speeds. The effects are likely to be more profound when using Berkovich tips with much smaller tip radii; large radius stylus tips act as mechanical low-pass filters. Also, if the relatively flat regions of the profiles are studied carefully and compared, it is seen that the noise

level is slightly higher for the higher speed scan. The exact nature of this additional noise is unknown at the time of this writing.

Optical flat

A Zerodur optical flat was profiled at high speeds to test the ability of the profiler to measure nanometer form errors over long traverse distances. Using a 10 µm stylus probe at 1 µN (0.1 mg) force, repeated 15 mm profiles were obtained at speeds of 80, 250 and 500 µm/s. The resulting profiles are shown in figure 7.3, each staggered 5 nm to provide separation on the vertical scale.

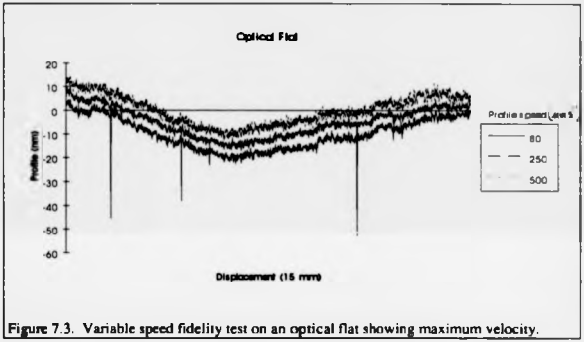


Figure 7.3. Variable speed fidelity test on an optical flat showing maximum velocity.

the top speed of 500 µm/s (the top trace), the micrometer drive shaft is completing two revolutions per second. The low tolerance gear motor emitted audible noise during this high speed traverse. Needless to say, the Author was pleasantly surprised with the results. While there is some degradation of the high frequency information (which is to be expected), the measured form is still consistent with the lower speed traces.

Note added in proof: In recent experiments, a 6.4 nm step height standard was accurately measured at a traverse speed of more than 0.5 mm/s. No bouncing was evident in the trace.

Repeatability anomalies

While performing the step height standard repeatability tests, some curious, and as yet unexplained anomalies appeared. While not reproducible, these may be of some importance to future experimenters.

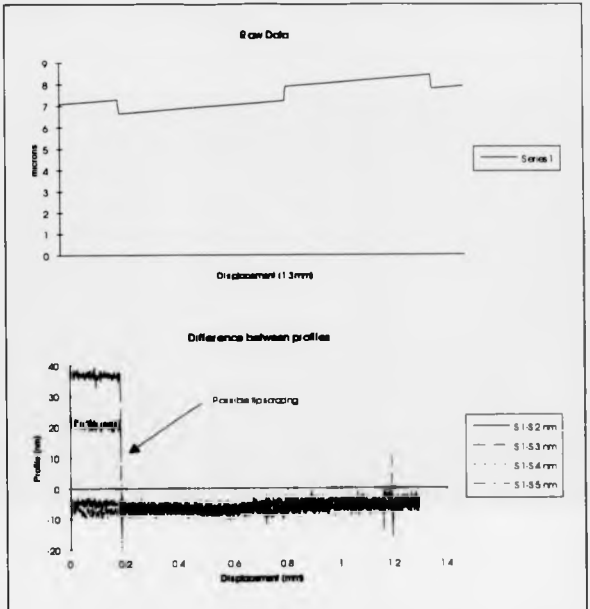


Figure 7.4. Possible surface contaminant layers causing anomalous tip scraping effects.

Shown in Figure 7.4, note that all of the traces shown repeat *after* the first step edge has been traversed. Before the first edge has been reached, there is some non-repeatability in the form of a "DC" offset. The nature of this offset is possibly a surface contaminant film, which could build up on the leading edge of the stylus tip during the retrace cycle, while returning to the home position. Once a profile is underway, this former leading edge now becomes the *trailing* edge. Quite possibly,

this edge is being wiped clean by passing over the falling edge, much like a boot scraper. While subsequent experiments were performed on cleaned samples, the results were not repeatable. No correlation could be made between cleaning procedures and surface preparations and the anomaly. This effect has been observed twice during the profiler tests. Other causes such as controller setpoint following, A/D converter errors and specimen motion errors were considered but thought to be unlikely sources of this error. No further investigations were undertaken, although the effect is likely a fruitful topic for further study.

Force repeatability

To examine repeatability under varying contact forces, a series of profiles were measured using the same specimen, under different loads. The step height specimen was chosen because its step edges give the profiler control system a severe test (although an impulse function would be the worst case scenario). Seven profiles were traced, at a traverse speed of 77 μm/s, with decreasing forces from 16 μN to 250 nN (1.6 mg to 0.025 mg force). Three of the traces (the other four nearly identical traces have been removed for clarity) are displayed graphically in figure 7.5. The 2 and 0.5 μN traces are closely superimposed, giving the appearance of only two traces in the figure.

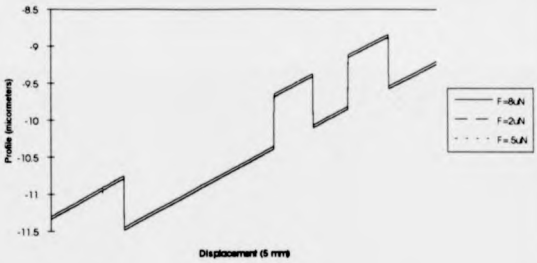


Figure 7.5. Variable contact force repeatability.

By visual inspection, differences between profiles are minimal with no bouncing or loss of contact evident even at low contact forces. This quality is likely an attribute of the inherent squeeze film damping in the force probe.

Contact stiffness studies

As hinted at in Chapter 2, the stiffness of the mechanical contact between the force probe and the specimen plays a crucial role in the dynamics of the closed-loop force servo system. Perhaps less obvious, but equally important, is the role of internal damping in the specimen and its effect on the contact dynamics. To examine the role of this damping, theoretical analysis of two different types of solids is presented.

The first type of solid is known as a Voigt[1] solid, having its internal damping properties modeled as a damper in parallel with a spring, figure 7.6.

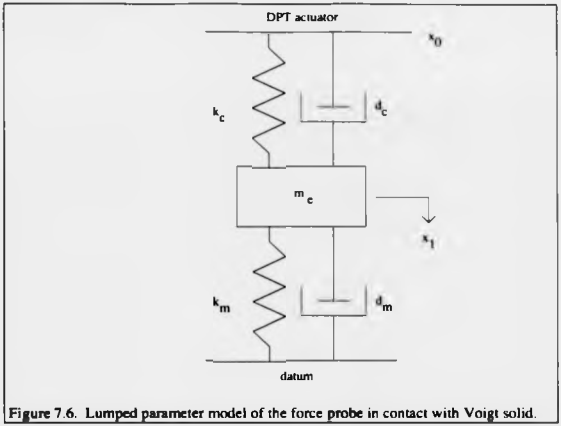


Figure 7.6. Lumped parameter model of the force probe in contact with Voigt solid.

In this case, the spring in question is the non-linear contact stiffness examined in Chapter 2. For small displacements of x_1, the equations of motion for this two-degree-of-freedom system are

$$\begin{aligned} d_c(\dot{x}_o - \dot{x}_1) + k_c(x_o - x_1) &= F_o \\ m_e\ddot{x}_1 + d_c(\dot{x}_1 - \dot{x}_o) + d_m\dot{x}_1 + (k_m + k_c)x_1 - k_c x_o &= 0 \end{aligned} \tag{7.1}$$

Where d_C and k_C are the cantilever damping and stiffness respectively, d_m and k_m are the material damping and stiffness respectively, m_e is the effective mass of the cantilever/probe system and F_O is the static force setpoint.

Dividing the second equation of 7.1 by the equivalent mass, it can be rewritten as

$$\ddot{x}_1 + 2\zeta\omega_n\dot{x}_1 - \frac{d_c}{m_e}\dot{x}_0 + \omega_n^2 x_1 - \frac{k_c}{m_e}x_0 = 0 \tag{7.2}$$

where

$$\zeta = \frac{d_c + d_m}{2\sqrt{m_e(k_c + k_m)}} \text{ and } \omega_n = \sqrt{\frac{k_c + k_m}{m_e}}. \tag{7.3}$$

Assuming sinusoidal excitation at frequency ω, equation 7.2 may be solved for the frequency response, H(jω), of the probe tip to a displacement of the DPT actuator head given by

$$H(j\omega) = \frac{x_1(j\omega) - x_0(j\omega)}{x_0(j\omega)} = 1 - \frac{x_1(j\omega)}{x_0(j\omega)} = 1 - \frac{\dfrac{k_c}{k_c + k_m} + j\left(\dfrac{\omega d_c}{k_c + k_m}\right)}{1 - \dfrac{\omega^2}{\omega_n^2} + j\left(2\zeta\dfrac{\omega}{\omega_n}\right)}. \qquad (7.4)$$

The influence of the material damping, d_m on the pole in eq. 7.4 is difficult to determine due to several factors. Friction at the point of contact is extraordinarily difficult to assess, and is an additional mechanism for lost energy in the system. The effects of surface roughness also affect contact stiffness. Influences from the control system and actuator response make it extremely difficult to obtain useful internal damping information from the solid. It should be noted, however, that although unambiguous data are not available at this time, the technique could be applied to the constant force profiler in two different ways: First, the transfer function of the actual tip in contact with the specimen could be used for *in situ* tuning of an adaptive servo controller. Such a controller could tune itself for maximum flatness of response and communicate to the horizontal traverse motor controller the optimum traverse speed under such conditions. Second, the technique could be used as a contrast mechanism for imaging. A modulation signal could be applied to the force setpoint and either the magnitude or phase information extracted with a lock-in amplifier. Such an instrument would operate in the "repulsive force modulation mode" and could be extremely useful.

The magnitude of the transfer for a variety of different damping values d_m is plotted in figure 7.7 where frequency has been normalized to ω_n in equation 7.3.

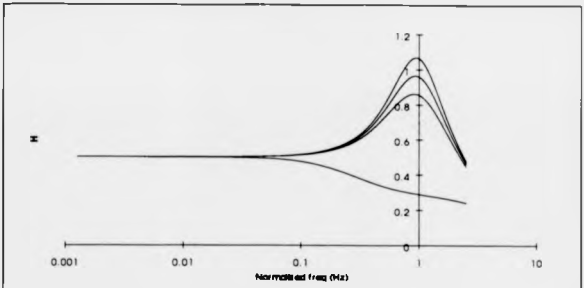


Figure 7.7 Theoretical response of a Voigt solid in contact with a force probe.

As might be expected, the magnitude response will peak for materials with low internal damping.

To experimentally investigate these dynamic effects, a series of substances were placed on the force profiler carriage and tested using the methods outlined in Chapter 6 for assessing control system closed-loop dynamic response. Because of the nature of the point contact, samples need not be the same size (implied by St. Venants principle). In each case, a 2 μm stylus was used with a nominal contact force of 3 μN. The modulated disturbance waveform represents a ±1 μN disturbance. In all experiments, the total applied forces are well under 0.5 mgF. The first sample to be tested was 6061T-6 aluminum alloy, flycut and hand-lapped, figure 7.8.

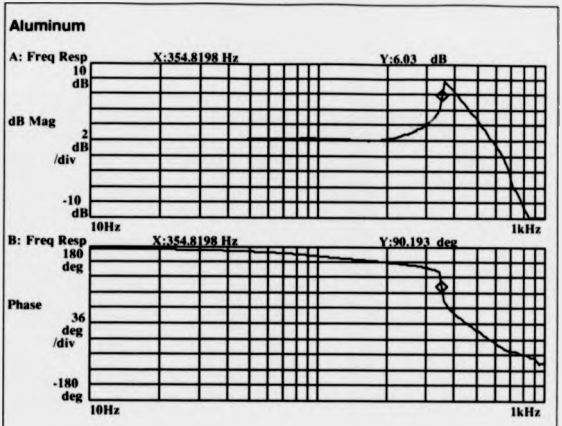


Figure 7.8. Dynamic response of aluminum alloy showing low internal damping.

Note the presence of the resonance peak in the figure. This peak is a combination of control system and actuator responses convolved with the dynamics of the tip/specimen interaction. However, in spite of these other effects, the response could be indicative of low internal damping in the material. To test this hypothesis, a similarly prepared copper specimen was next tested in the repulsive force modulation mode, figure 7.9.

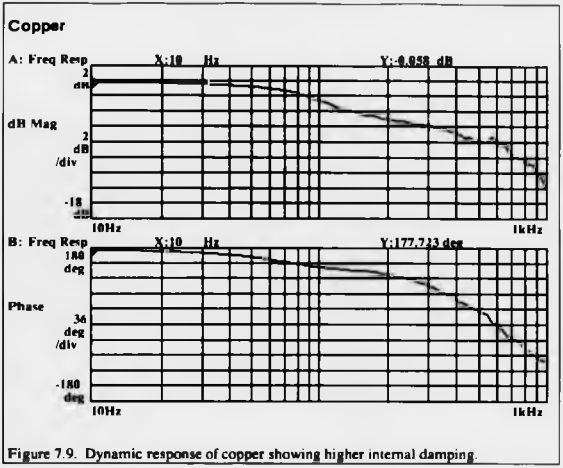


Figure 7.9. Dynamic response of copper showing higher internal damping.

This figure shows what appears to be an increased level of damping measured from the point contact in repulsive force modulation mode. Note, that at this point of development, no attempt has been made to obtain any quantitative data from these analyses. Nevertheless, the correlation between the measurements shown in figures 7.8 and 7.9 and the theoretical results of figure 7.7 indicate that damping force data may be experimentally derived. A sample of tool steel from a blank used as a lathe cutting tool has been measured, (figure 7.10)

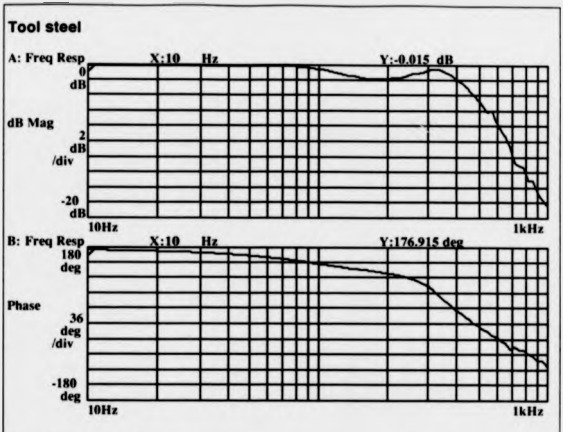


Figure 7.10. Dynamic response of tool steel.

and exhibits significant damping characteristics. This damping is by design, as cutting tool vibration is a serious problem for modern manufacturing. Figure 7.11 to 7.13 show similar measurements on polished Zerodur, tungsten carbide and brass.

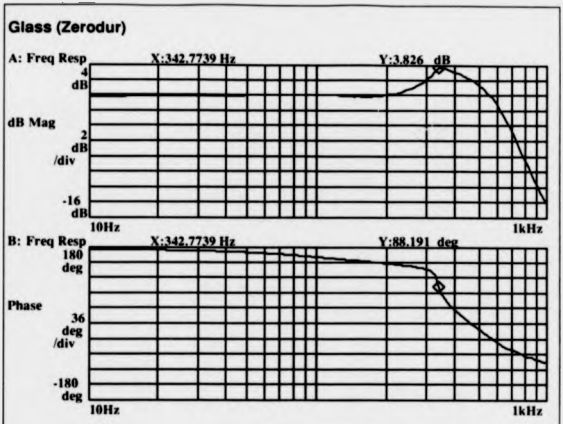


Figure 7.11. Dynamic response of polished Zerodur glass ceramic.

As might be expected, Zerodur exhibits about 4 dB of "gain peaking" in its response curve; expected due to the material's low internal damping. Polished quartz would have been an interesting selection due to its ultra-low internal damping. No sample was available at the time of this writing, however. An 18 mm diameter tungsten carbide ball bearing was tested for its dynamic interaction with the probe tip, figure 7.12.

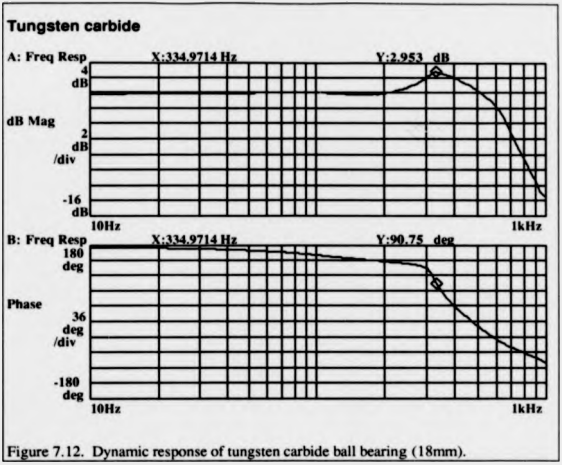


Figure 7.12. Dynamic response of tungsten carbide ball bearing (18mm).

Note that the tungsten carbide exhibits more damping than both the aluminum and the Zerodur glass ceramic. Brass was measured and found to have a characteristic similar to that of tungsten carbide, figure 7.13.

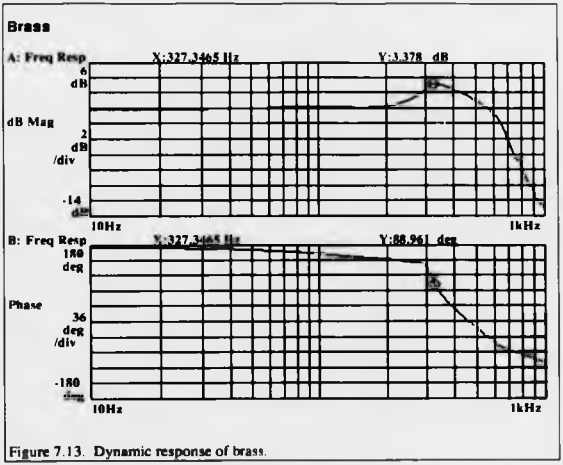


Figure 7.13. Dynamic response of brass.

Some materials expected to exhibit higher damping properties were tested. The characteristic of rubber is shown in figure 7.14.

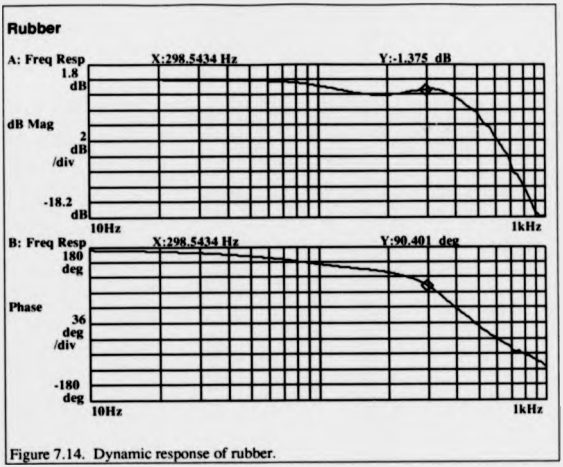


Figure 7.14. Dynamic response of rubber.

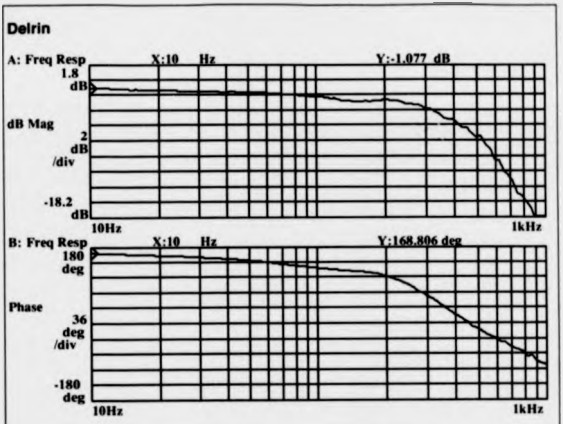


Figure 7.15. Dynamic response of Delrin .

Delrin seems to have more damping than the rubber, an expected result for a solid. It is extremely difficult to arrive at any general conclusions however, given the wide variety of rubber compounds available. The adhesion at the point contact between the Delrin and the diamond stylus should be relatively low, reducing the effects of contact friction.

Viscoelastic solids

In the previous section, the internal damping of the solids could be described using the Voigt model. Some solids however, conform to the Maxwell model of viscoelastic solids[2]. In this model, the internal viscous damper is shown in series with the elastic stiffness of the solid. Maxwell solids include viscoelastic polymers such as Silly Putty ™ and metals which exhibit cold-flow or "creep" characteristics over extremely long time scales. The aluminum which was modeled as a Voigt solid in the previous section

could be modeled as a Maxwell solid if cold-flow considerations were being assessed.

A Maxwell solid, lumped parameter model is shown in figure 7.16

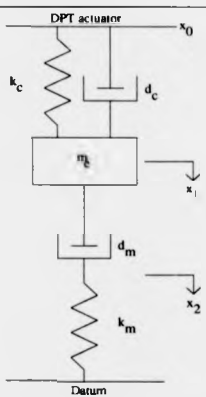


Figure 7.16. Lumped parameter model of a Maxwell viscoelastic solid.

The equations of motion for this 3 degree-of-freedom system are given as

$$\begin{aligned} m_t\ddot{x}_1 + d_m(\dot{x}_1 - \dot{x}_2) + d_c(\dot{x}_1 - \dot{x}_0) + k_c(x_1 - x_0) &= 0 \\ d_c(\dot{x}_1 - \dot{x}_0) + k_c(x_1 - x_0) &= F_0 \\ d_m(\dot{x}_2 - \dot{x}_1) + k_m x_2 &= 0 \end{aligned} \tag{7.5}$$

Where all variable are the same as eq. 7.1. The transfer function is determined by assuming a sinusoidal solution, solving the third equation of 7.5 for x_2, and substituting this into the first relation, which, after some algebraic manipulation yields

$$H(j\omega) = 1 - \frac{x_1(j\omega)}{x_0(j\omega)} = 1 - \frac{-d_c d_m\omega^2 + k_c k_m + j(d_c k_m\omega + d_m k_c\omega)}{-m_t k_m\omega^2 - d_c d_m\omega^2 + k_c k_m + j\left(m_t d_m\omega^3 + d_m k_m\omega + d_m k_c\omega + d_c k_m\omega\right)}, \tag{7.6}$$

A theoretical plot of the response of the Maxwell solid Silly Putty™ to a force probe is shown in figure 7.17. Note that the specimen acts as a mechanical high pass filter.

Figure 7.17. Theoretical magnitude response of a Maxwell solid in contact with the force probe.

Measuring a viscoelastic solid presents some unusual problems. Upon first contact with the stylus tip, the solid flows until the tip and probe cantilever are returned to their nominal position through a combination of viscous and surface tension forces. The contact force used for the Silly Putty was the same as those used for the previous samples.

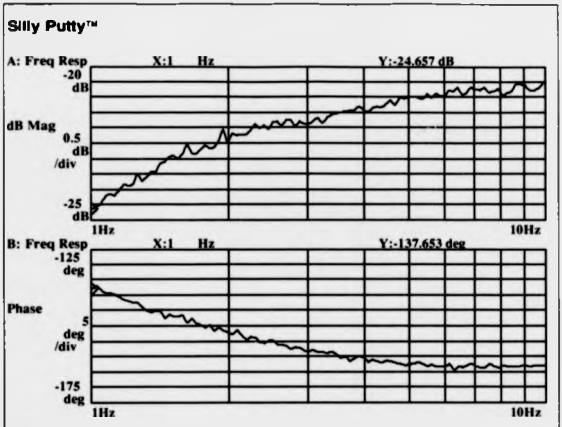


Figure 7.18. Dynamic response of Silly Putty, a Maxwell viscoelastic solid.

As expected, the putty behaved as a mechanical high-pass filter, as explained by RV Jones[3]. Silly Putty is more than just a children's toy, however. Similar viscoelastic polymers, when loaded with abrasives are essential to the extrusion honing process. When forced through internal cavities under pressure, the "putty slurry" increases velocity (and stiffness) as it flows around corners and sharply changing geometries; removing burrs and polishing the surface in the most important places for high-speed flow applications like automotive engine intake manifolds. While it appears that we can measure the dynamic properties of such putties at the point contact level, the real purpose of the experiment was to illustrate the strong dependence of the interaction of the force probe with the specimen material.

Scanning dynamics microscope/profiler

Clearly, the opportunity has presented itself for the development of a scanning dynamics microscope or profiler. Such an instrument could be used to map the dynamic properties of thin films or bulk specimens at the surface. To make use of the contact stiffness as an imaging mechanism would require some form of phase and/or amplitude measurement scheme. One method would be to disturb the control system with an repulsive force modulation signal while monitoring phase shift with a two-phase lock-in amplifier.

An interesting test specimen for such a microscope might be a piece of diamond turned, polished copper, having large grain size and hence, large anisotropy. The force profiler could profile the specimen surface in the normal mode of operation to determine the topography. If well polished, very little topography would exist on such a specimen. Next, the same trace would be profiled in the repulsive force modulation mode. If successful, the instrument should be able to discriminate grain boundaries in the material. Another specimen of interest would be an unpassivated integrated circuit. The identification of various features such as metal, polysilicon or SiO_2, would be of great value. Furthermore, since internal damping is affected by internal strains, the technique may even prove useful for diagnosing thin-film depositions, surface damage and residual stresses. Many possibilities exist for such a measurement technique.

Future work: Specimen damage study

A study of the mechanics of contact between a microscopic probe and a surface has applications to the fields of atomic force microscopy, tribology and brittle materials machining. Many fundamental issues governing the mechanics of contact between an AFM probe and a specimen are poorly understood as of this writing. Probe damage mechanisms of brittle materials such as Si, GaAs, InP and CdTe are of critical importance to the microelectronics and optoelectronics industries. Already, manufacturers of stylus instruments and atomic force microscopes are advertising the

"low damage" aspects of their instruments. Many makers of AFMs claim to operate at zero contact force and produce no specimen damage. Clearly, a more basic understanding of any damage mechanisms will be essential as these instruments find their way into the mainstream of surface metrology. A basic understanding of the physics of contact forces produced by AFMs and stylus profilers could also lead to a better understanding of the nanometric cutting process by studying the transitions from sliding to burnishing to cutting.

New Technology and Plan of Experiment

A proposed experimental methodology divides the problem into two distinct parts. First, the construction of metrologically characterized instrumentation necessary for the production of artifact crystals. Second, the analysis of these crystals using X-ray techniques commonly used by materials scientists.

By making a profile, starting at low forces and gradually increasing to higher loads, an artifact may be produced which maps contact force to distance from a fiducial mark. By starting the trace with a contact force of only a few nanoNewtons and progressing linearly to a force of a few tens of microNewtons (a few milligrams), the entire range of contact forces encountered with both AFMs and stylus instruments may be examined.

This artifact will then be analyzed by X-ray diffraction methods to produce a map of the damage introduced into the brittle material. The Lang topograph interposes a scanning slit between a collimated X-ray beam and a specimen mounted at/or near to the Bragg angle. Scanning the slit across sample produces an image capable of exposing microcracks in the sample's surface layer. Locations of damaged sites revealed by the Lang camera may then be mapped directly to corresponding contact force levels.

Repeating the experiments for several sized styli, a 20 nm radius tetrahedral (Berkovich) diamond indenter, a 2 μm spherical diamond profiler stylus and a 5 μm diamond stylus would be used to generate a four dimensional experimental parameter space: Stylus radius, force, friction and damage. Clearly, quantifying the parameter for damage will be beyond the scope of such an experiment. Most likely, thresholds will be noted where damage is first indicated on the X-ray analysis and where damage is first visible as "tracks" on an SEM image. These parameters should be sufficient to establish damage criteria for atomic force microscopes and stylus profilers.

Impact on the Engineering Community

As atomic force microscopes find their way into the mainstream of metrology, it will become increasingly important to fully understand the nature of the tip/specimen interaction. The microelectronics community is already planning to apply AFMs to sub-micron line width metrology and angstrom level roughness measurements for 16MB DRAM production. One Japanese company is already using the AFM to certify sub-nanometer level roughness for its ultra smooth wafers[4]. The current level of end-user awareness regarding probe/sample interactions is somewhat akin to the early use of the SEM in the 1960s. Compounding the complexities of using scanning probe microscopes, developments in image processing have often preceded developments in physical imaging mechanisms. It is very possible to consistently produce images of surfaces that contain erroneous information. Clearly, the current situation is somewhat less than desirable from a metrological standpoint. Considerable research is required if sufficient knowledge of contact interactions enabling unambiguous interpretation of constant force profiles is to be achieved. It is hoped that the developments presented in this dissertation have provided a useful foundation for further study.

---

[1] H. Kolsky, Stress Waves in Solids, Dover Publishing, Inc., NY, 1963, p. 115.

[2] H. Kolsky, *ibid.*

[3] R.V. Jones, Instruments and Experiences. J. Wiley and Sons, NY, (1988).

[4] C.F. Quate, "Keynote speakers address, *The American Society of Precision Engineering, 1992 Annual Meeting, Orlando, Fla.*, (1992).

APPENDIX

Data Translation DT-2823 digital port assignments:

Bit number	Port 0, output	Port 1, input
0	counter select 'A'	interferometer LSB
1	counter select 'B'	interferometer
2	counter reset (rising edge)	interferometer
3	motor on/off	interferometer
4	motor direction	interferometer
5	sampling interval LSB	interferometer
6	sampling interval	interferometer
7	sampling interval MSB	interferometer MSB

Table 5.1. Digital port assignments.

Analog I/O

Port number	A/D input	D/A output
0	force servo output	force servo setpoint
1		
2	capacitance gage output	---
3	remote control pendent	---

Table 5.2. Analog port assignments

Software listing, 1st June, 1993.

```
\ ASYST PROGRAM "PROFILE.TXT" TO OPERATE FORCE PROFILER
FORGET.ALL
ECHO.OFF
NORMAL.DISPLAY
DT2820

\ VARIABLES

REAL SCALAR VALUE
REAL SCALAR CAPGAGE
REAL SCALAR SETPOINT
REAL SCALAR USERIN
REAL SCALAR NEW.ZERO
REAL SCALAR P1
REAL SCALAR P2
REAL SCALAR POSITION
REAL SCALAR CALIBRATION
-0.9326 CALIBRATION !

INTEGER SCALAR SAMPLING

DP.REAL SCALAR POSITION.COUNTER

40 STRING PATH
40 STRING FILE

INTEGER DIM[ 1024 ] DMA.ARRAY DATA
0 DATA !
```

```
INTEGER DIM[ 4 ] ARRAY COUNTER.BYTES

REAL DIM[ 1024 ] ARRAY SCALED.DATA
REAL DIM[ 1024 ] ARRAY TRAVERSE.DATA
REAL DIM[ 4 ] ARRAY CURSORS
REAL DIM[ 2 ] ARRAY LSQ.LINE


\ TEMPLATE DEFINITIONS

1 DIGITAL.TEMPLATE MOTOR.BITS
BINARY
00011000, DIGITAL.MASK
DECIMAL
DIGITAL.INIT

1 DIGITAL.TEMPLATE DIGIN
DIGITAL.INIT

1 DIGITAL.TEMPLATE COUNTER.BITS
BINARY
00000111, DIGITAL.MASK
DECIMAL
DIGITAL.INIT

1 DIGITAL.TEMPLATE SAMPLE.BITS
BINARY
11100000, DIGITAL.MASK
DECIMAL
DIGITAL.INIT

2 2 A/D.TEMPLATE IN2
A/D.INIT

3 3 A/D.TEMPLATE IN3
A/D.INIT

0 0 A/D.TEMPLATE IN0
DATA DMA.TEMPLATE.BUFFER
EXT.TRIG
IN0 A/D.INIT

0 0 D/A.TEMPLATE OUT0
D/A.INIT

\ WORD DEFINITIONS

: SETUP
IN2
CR ." CONNECT CAP GAGE OUTPUT TO A/D CHANNEL #2"
CR ." INPUT DESIRED SETPOINT IN VOLTS:  " #INPUT USERIN !
A/D.IN -10 10 A/D.SCALE CAPGAGE !
CAPGAGE USERIN - SETPOINT !
OUT0
SETPOINT -10 10 D/A.SCALE D/A.OUT
CR ." CAPGAGE=" CAPGAGE .
CR ." USERIN=" USERIN .
CR ." SETPOINT=" SETPOINT .
```

```
CR ." DONE"
;

: MOTOR.+
BINARY
MOTOR.BITS 11, WRITE.BITS  \ ENERGIZE RELAY AND MOTOR
DECIMAL
;

: MOTOR.-
BINARY
MOTOR.BITS 01, WRITE.BITS  \ ENERGIZE MOTOR
DECIMAL
;

: MOTOR.OFF
BINARY
MOTOR.BITS 00, WRITE.BITS  \ CLEAR MOTOR CONTROL BITS
DECIMAL
;

: REMOTE
IN3
CR ." PRESS ANY KEY TO EXIT"
BEGIN
    A/D.IN -10 10 A/D.SCALE VALUE :=  \ READ REMOTE CONTROL
    VALUE 1.00 >
    IF MOTOR.+ THEN
    VALUE -1.00 <
    IF MOTOR.- THEN
    VALUE 1.00 <
    VALUE -1.00 > AND
    IF MOTOR.OFF THEN
    ?KEY
UNTIL
;

: WAIT
BEGIN
?DMA.ACTIVE
WHILE
REPEAT
;

: RESET.COUNTER
BINARY
COUNTER.BITS
100, WRITE.BITS
000, WRITE.BITS
100, WRITE.BITS
DECIMAL
;

: READ.COUNTER
0 COUNTER.BYTES :=
BINARY
COUNTER.BITS 111, WRITE.BITS
DIGIN DIGITAL.IN COUNTER.BYTES [ 1 ] :=
```

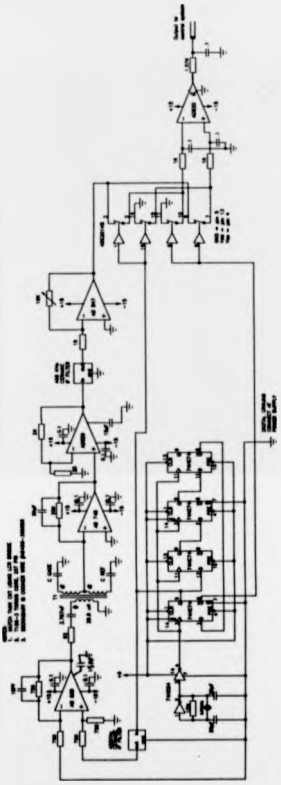
```
COUNTER.BITS 110, WRITE.BITS
DIGIN DIGITAL.IN COUNTER.BYTES [ 2 ] !=
COUNTER.BITS 101, WRITE.BITS
DIGIN DIGITAL.IN COUNTER.BYTES [ 3 ] !=
COUNTER.BITS 100, WRITE.BITS
DIGIN DIGITAL.IN COUNTER.BYTES [ 4 ] !=
DECIMAL
;

: CONVERT.COUNTER
0 POSITION.COUNTER !=
COUNTER.BYTES [ 4 ] POSITION.COUNTER !=
256.D COUNTER.BYTES [ 3 ] * POSITION.COUNTER + POSITION.COUNTER !=
65536.D COUNTER.BYTES [ 2 ] * POSITION.COUNTER + POSITION.COUNTER !=
16777216.D COUNTER.BYTES [ 1 ] * POSITION.COUNTER + POSITION.COUNTER
!=
POSITION.COUNTER 2147483648.D - POSITION.COUNTER !=
;

: R1
READ.COUNTER
CONVERT.COUNTER
POSITION.COUNTER 0.00003955 * POSITION !=
;

: R
R1
POSITION .
;

: HOME
STACK.CLEAR
CR ." RETURNING CARRIAGE TO HOME POSITION"
MOTOR.-
BEGIN
  R1
  POSITION -0.5 <
UNTIL
MOTOR.OFF
;

: GOTO
STACK.CLEAR
CR ." ENTER DESIRED POSITION (mm):  " #INPUT NEW.ZERO !=
MOTOR.-
BEGIN
  R1
  POSITION NEW.ZERO <
UNTIL
MOTOR.OFF
;

: SPEED
MOTOR.+
BEGIN
  R1
  POSITION 0 >
UNTIL
```

```
R1 POSITION P1 !
1000 MSEC.DELAY
R1 POSITION P2 !
MOTOR.OFF
CR ." CARRIAGE VELOCITY IS " P2 P1 - . ." mm/sec"
100 MSEC.DELAY
HOME
;

: DELTA
CR ." DELTA X=" CURSORS [ 3 ] CURSORS [ 1 ] - .
CR ." DELTA Y=" CURSORS [ 4 ] CURSORS [ 2 ] - .
;

: PROFILE.GO
STACK.CLEAR
IN0 EXT.CLOCK
A/D.INIT
NORMAL.DISPLAY
CR ." SCAN NUMBER      SAMPLING INTERVAL      TRACE LENGTH"
CR ." <0>              79.1   nm              80.998  um  "
CR ." <1>              158.2  nm              161.997 um  "
CR ." <2>              316.4  nm              323.997 um  "
CR ." <3>              632.8  nm              647.987 um  "
CR ." <4>              1.267  um              1.296   mm  "
CR ." <5>              2.531  um              2.592   mm  "
CR ." <6>              5.062  um              5.184   mm  "
CR ." <7>              10.125 um              10.368  mm  "
CR ." INPUT CHOICE:  " #INPUT SAMPLING !
SAMPLE.BITS SAMPLING @>MASK WRITE.BITS
1024 RAMP .0791 * 2 SAMPLING ** * TRAVERSE.DATA !
MOTOR.BITS MOTOR.+
IN0 A/D.IN>ARRAY(DMA)
CR ." PROFILE IN PROGRESS..."
WAIT
MOTOR.BITS MOTOR.OFF
DATA 0 15 A/D.SCALE SCALED.DATA !
SCALED.DATA CALIBRATION * SCALED.DATA !
CURSORS READOUT>ARRAY
NORMAL.COORDS
.5 .975 READOUT>POSITION
WORLD.COORDS
TRAVERSE.DATA SCALED.DATA XY.AUTO.PLOT
ARRAY.READOUT
CR ." PRESS <DELETE> KEY TO EXIT CURSOR MODE"
;

: HARD.LEVEL
STACK.CLEAR
LOAD.OVERLAY MATFIT.SOV
TRAVERSE.DATA SCALED.DATA 1 LEASTSQ.POLY.FIT
LSQ.LINE !
CR ." NEGATIVE, TURN SCREW ANTI-CLOCKWISE"
CR ." POSITIVE, TURN SCREW CLOCKWISE"
CR ." TURNS = " LSQ.LINE [ 1 ] 0.0008 / .
;

: SAVE123     \ SAVES SCALED DATA AS A LOTUS FILE
```

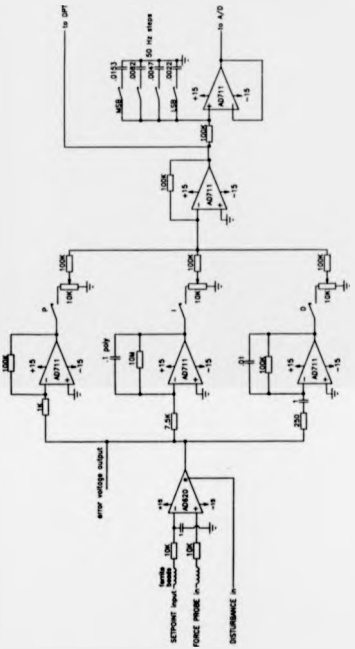
```
" C:\ASYST\DATA\" PATH "!=
CR ." ENTER 123FILE NAME AND EXT. TO BE SAVED IN " PATH "TYPE
"INPUT FILE "!=
LOAD.OVERLAY C:\ASYST\123IO.SOV
PATH FILE "CAT DEFER> 123FILE.CREATE
PATH FILE "CAT DEFER> 123FILE.OPEN
1 1 123WRITE.DOWN
SCALED.DATA
ARRAY>123FILE
PATH FILE "CAT DEFER> 123FILE.CLOSE
;

\ PROGRAM STARTS HERE
MOTOR.BITS MOTOR.OFF
NORMAL.DISPLAY
CR ." <SETUP>          Adjusts setpoint of control system"
CR ." <REMOTE>         Remote operation of drive motor"
CR ." <PROFILE.GO>     Takes profile and displays graph, must be in
home posit."
CR ." <SAVE123>        Saves profile data in a Lotus file format"
CR ." <RESET.COUNTER>  Resets interferometer to zero"
CR ." <R>              Reads interferometer and displays position"
CR ." <HOME>           Returns to home position for repeat trace"
CR ." <SCREEN.PRINT>   Graphics screen dump to printer"
CR ." <GOTO>           Move carriage to specified position"
CR ." <DELTA>          Performs delta measurement on cursor positions"
CR ." <SPEED>          Determines carriage speed from home position"
CR ." <HARD.LEVEL>     Level specimen carriage hardware"
```

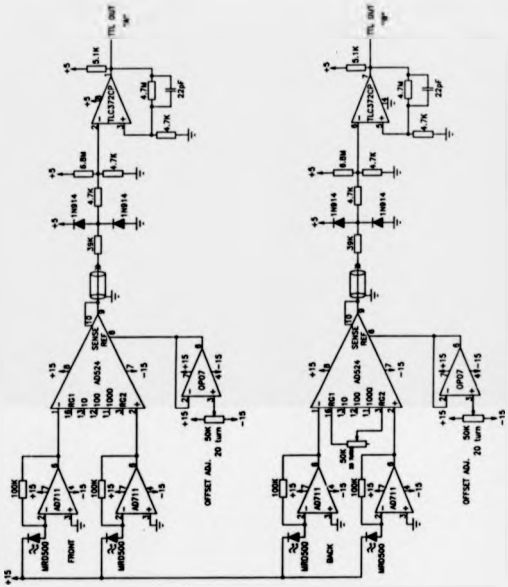
Electronics hardware

Capacitance gage

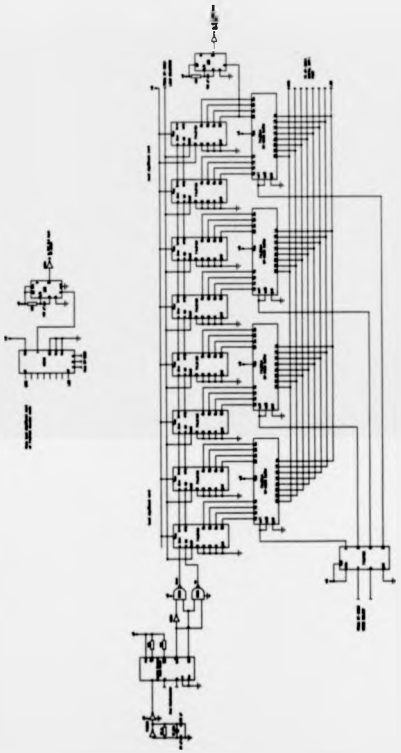


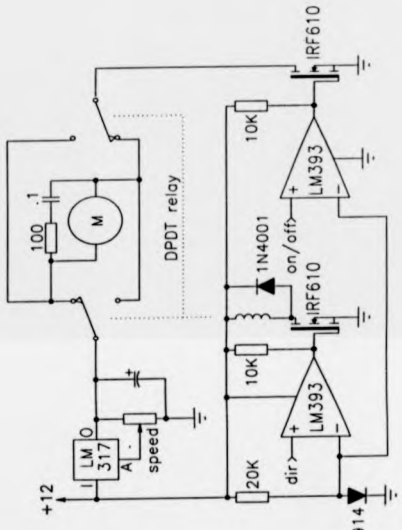
PID servo


to DPT
to A/D
50 Hz steps
MSB
LSB
AD711
error voltage output
SETPOINT input
FORCE PROBE
DISTURBANCE


TTL OUT
5.1K
+5
TLC3702
4.7K
22pF
4.7K
6.8M
1N914
39K
SENSE
REF
AD524
+15
-15
OP07
50K
20 turn
OFFSET ADJ.
100K
AD711
MRD500
FRONT
BACK

Counter and controller




+12
LM 317
I
O
ADJ
speed
100
.1
M
DPDT relay
20K
10K
dir
LM393
IRF610
1N4001
on/off
10K
LM393
IRF610
1N914