Interpretation and Evaluation of Stylus Profiling Techniques

By

Man Li, BSc.

Submitted for the degree of Doctor of Philosophy
to the Higher Degrees Committee
University of Warwick

Centre for Micro-engineering and Metrology
Department of Engineering
University of Warwick, Coventry, U. K.

Novermber 1991

I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in number, your knowledge is of a meagre and unsatisfactory kind.

— *Lord Kelvin*

To
my parents, and
my motherland, China.

Contents

List of Figures

List of Tables

Declaration

This thesis is presented in accordance with the regulations for the degree of Doctor of Philosophy by the Higher Degree Committee at the University of Warwick. The thesis has been composed and written by myself based on the research undertaken by myself. The research materials have not been submitted in any previous application for a higher degree. All sources of information are specifically acknowledged in the content.

M. Li

Acknowledgements

It is my pleasure to express my gratitude to Prof. D. J. Whitehouse, my supervisor, for his guidance, encouragement, suggestions and comments during the course of research described in this thesis.

My sincere thanks go to Prof. D. K. Bowen, who became my supervisor at the later period of my work because Prof. Whitehouse was in sick leave, for his comments and suggestions about the construction of the thesis. I appreciate very much his dynamic working style.

Thanks to Dr. D. G. Chetwynd and Dr. S. T. Smith for introducing me to the practical aspects of the experiments and tests, also for their enthusiasm and discussions.

To Mr. R. A. Bridgeland & Mr. D. J. Robinson for their careful manufacture and for their help in constructing experimental work.

To Mr. R. D. Dale-Jones for his continuous effort and help in improving the language of this thesis. Without it, this thesis would not be as fluent as it is today.

To all those who kindly gave me their help and support during the time of my research.

Summary

The object of this Ph.D work is to validate stylus profiling techniques for ultra-high precision measurement and positioning in engineering. Stylus methods have been used extensively and successfully in the past in the fields of manufacturing control and components function study, but some problems still exist. Now their role has considerably expanded with the appearance of scanning probe microscopes and the new emphasis on nanotechnology, which shortens the gap between engineering and physics.

The profiling technique is interpreted and evaluated in terms of the mechanical aspects of data collection through a stylus surface instrument. This work contains: (a) 3D digital sampling techniques; (b) effects of the finite dimension of stylus and (c) forces contributing to measurement.

A new plane sampling model - hexagonal model - has been developed to improve the surface 3D data collection to almost 99% of the continuous case for summit height distribution. The dimension effect has been divided into two aspects: the effects of size and the shape. The cut-off effect caused by the size of the stylus on the surface curvature is not correctable. The analysis of the trace formation suggests that the 'deconvolution' of the true profile from the trace is feasible. The simulation using the MAT-LAB computer package confirms this with only computational error. A new method of stylus shape/dimension measurement was proposed based on these principles. Alternatively, a practical method of measuring a stylus shape using a knife-edge was also constructed and further developed. Stylii tips of radii from 1 μm to 50 μm can be measured using this rig to an accuracy of only 5% of the movement of the knife-edge. The physical effect of a stylus is discussed theoretically and experimentally in terms of the static and dynamic stylus loading forces. The dynamic variation is only 2.7% of the static one and it is negligible. Through the study, the lateral resolution and the frictional force within stylus-surface contact are found to be the crucial elements to be tackled so that the profiling technique is able to fulfill its requirements. A general discussion of the scanning probe microscopy, with emphasis on these points of view, presents quantitative problems in 3D measurement.

Chapter 1

Introduction

1.1 A brief history of surface measurement and the advent of stylus instruments

At the end of the 19th century and the beginning of the 20th century, the development of mechanical industry encouraged the transfer of skill from man to machine. More attention began to be paid to the surface finish of components. The interchangeability of components, which started as early as 1902, and the increase in machining precision required a better surface finish. As G. Schlesinger (1942) pointed out:

> "In the production of mechanical components the principal steps involved are in four essential categories: (1) metallurgical, (2) physical, (3) dimensional, (4) surface roughness."

High precision machining led to a great improvement in the surface finish. It was found that surface finish affects both the working life of the component and the speed of manufacture. The direct result of a roughly finished surface is friction. Higher friction demands larger work and more energy from the two mating surfaces. In a bearing system, a rough surface between the contacting parts leads quickly to wear, limited rotational speed and machine noise. Surface finish has also been found to have a marked effect in the region of boundary lubrication (Norton, 1942). In machine tool machinery, if the cutting edges of the tools are sharpened and finished carefully, their life will be prolonged. In other mechanical parts like engines, cylinders, crankshafts and gear manufacture, etc., surface quality and finish are also important (Page, 1948).

In Britain, investigations into surface finish were started soon after the establishment of the Research Department of the Institution of Production Engineers in 1939. It was immediately supported by 19 leading British firms, manufacturers of automobiles, aeroplanes and machine tools, etc. An initial report was issued in March 1940 in 'The Journal' (Vol. XIX, No. 3). In 1950, the surface finish stan-

dard was published (BS 1134: 1950). In America, a similar approach was taken. A preliminary American Surface Standards (B.46) was on trial for about two years in 1940. By 1955, these standards were as sophisticated as the ASA-B46.1 (1955). In other European countries, for example Germany, a book concerning stylus finishing techniques was written by G. Schmaltz, and published as early as 1936. The surface standards DIN 4760-4764 was formed from 1952 to 1954 (Schorsch, 1958).

At the same time, surface measuring instruments (including stylus instruments) were also developed. Among them, there were both quantitative and qualitative types. For example, Abbott's profilometer, the Taylor-Hobson Surfacemeter and the Schmaltz profile microscope are of quantitative type. The qualitative analysis instruments were normally comparison microscopes, e.g. Klemm's Comparoscope and the Vickers projection microscope. Most stylus instruments are of quantitative type, so they have been developed most. They are used not only to insure the quality of the surface finish, but also to analytically study the surface performance.

Over the last 40 to 50 years, surface profiling techniques have progressed. For instance, optical and laser techniques, electronic techniques and computing programming techniques have been developed. The evolution of the stylus instrument can be seen from the development of the sensors used to transfer the signal from the stylus pick-up. Earlier, the signal from the stylus beam was recorded by a photographic method. A beam of light is shone onto a mirror fixed at one end of the stylus beam and then reflected onto a photographic film. Later, electrical sensors such as an inductor or a capacitor replaced the photographic film. Recently, laser techniques like the interferometer have increased the accuracy and stability of the pick-up system. Electronic digital techniques and computer programming have made calculations more comprehensive.

Surface measurement, especially the use of stylus instruments, are the result of engineering requirements. The surface of a component can be used to monitor the performance of the part and to examine the conditions of the machine tool during

manufacture. Most functional studies concern contact situations and changes to the two surfaces after a relative movement.

1.2 Stylus profiling techniques

Stylus profiling techniques check or evaluate things like surface texture, in a similar way to the human fingernail. In fact, the fingernail could be regarded as a primitive stylus, in the same way as the eye can be compared with the interferometer. The stylus instrument is a comprehensive profiling technique.

Stylus profiling techniques can be categorized into three generations or types: mechanical, optical and physical. Their representatives are: (1) stylus instruments, (2) optical profilometers and (3) scanning probe microscopes, respectively. Although they are different generations, none of them can replace completely the previous one. Optical profilometers have some overlap and extension of utilization over stylus instruments. Scanning probe microscopes have been developed recently. They have wider application and great potential. Each of them will be introduced below.

1.2.1 Stylus instruments

The history of the development of stylus instruments is shown in figure 1.1. In the first row, the three basic elements in the construction of the instruments are listed: (1) stylus; (2) transducer and (3) recorder. The techniques become progressively more recent as the columns are descended.

The principle of the stylus instrument's measuring processes can be described in the following manner: a stylus is moved along a surface. It is connected with a pick-up unit which is maintained in a controlled level, yet permits the stylus to follow intimately the physical surface of the part. Variations of the surface in relation to the level of the pick-up translation will cause the stylus to deflect from a reference position. The mechanical deflections of the stylus will be translated by

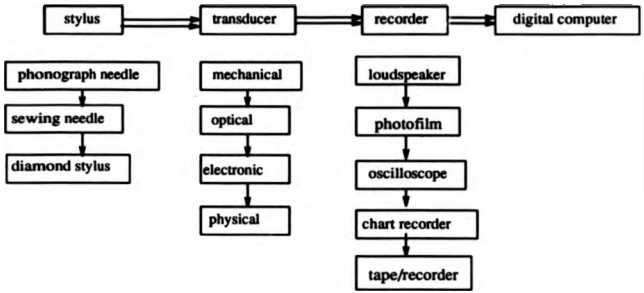


Figure 1.1: The history & development of the stylus type of instrument.

a cantilever with an inductive transducer at the other end, into an electrical signal. This signal is then electronically amplified and either displayed or logged into a computer processor. A gearbox drives the pick-up at a constant speed. A schematic of a stylus instrument is shown in figure 1.2 (Thomas 1982).

Datum reference

The datum in the stylus instrument is the reference for the measurement. The first datum was the mechanical direct reference using a high accuracy straight bar. As the stylus travels across a surface, the surface height is obtained from a comparison with the bar. The other mechanical datum was simulated by a long leaf spring, e.g. the one in the Talysurf-4. The former gives a long range measurement. Both datum systems have their mechanical loops open and, therefore, are sensitive to external vibration. The present datum is the skid. It uses the unknown surface itself as the reference (see figure 1.3). The difference between the stylus and the skid is the

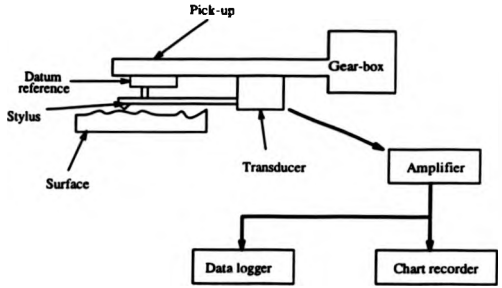


Figure 1.2: Schematic of the stylus instrument.

their spatial bandwidth: the skid has a much smaller spatial bandwidth. In other words, it is only sensitive to much lower frequencies. This system is intrinsic but approximate and dependent on the separation of surface wavelength.

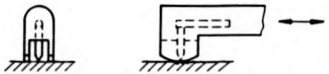


Figure 1.3: Schematic of the skid datum system.

The stylus instrument is generally used for geometrical measurement such as the shapes of engineering components. The above example is for surface roughness measurement. Surface finish is a fine geometric structure embodied in component

form, so a much finer stylus is needed for its measurement in order to neglect the effect of the stylus geometry. The profiling method provides reliable results within its limitations. It is a universal tool for most materials except very soft ones.

1.2.2 Optical profilometers

Optical profiling instruments have been developed to overcome the problem of contact between the stylus and the surface, so that damage to the surface and wear of the stylus tip can be avoided. Another advantage is faster measurement and therefore simultaneous inspection during manufacture.

In optical profilometers the focal point of a light beam is used to replace the stylus. A rough surface causes a change in the optical signal during the profiling procedure. The amount of change can be obtained from the principles of the optical system. The geometry can thus be found.

There are many ways to find the optical change. One simple way is to detect the focus plane through the sensors S1 & S2 by a knife-edge (Thwaite, 1977) or a chopper blade (Jing *et al*, 1987), as shown in figure 1.4.

The defect-of-focus optical technique (Sawatari & Zipin 1979; Mignot & Gorecki 1983) or optical Fourier transformation (Fainman *et al* 1982) uses a comparison of the differences of the intensities obtained by photodetectors P1 & P2 (referring to figure 1.5). Similarly, focus detection is also obtained by the total reflection using a critical-angle prism (Koho *et al* 1985).

An alternative is to use the optical change to generate an electronic signal to control a servomechanism to reposition the optical system to the initial situation. The movement of the objective is a measure of the surface geometry (Dupuy, 1967-68; Arecchi *et al* 1979). In this case, a high precision, stable servo-system is required.

Downs *et al* (1985) have designed a common-path polarimetric interferometer using a bi-refringent lens. The surface profile is calculated from the phase difference of the two polarization beams. This system has a very high accuracy.

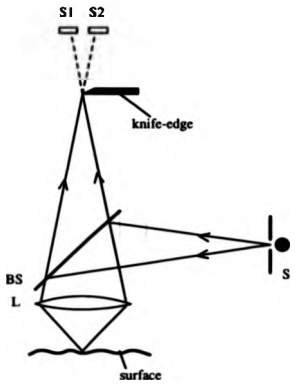


Figure 1.4: Optical system for surface measurement using a knife-edge.

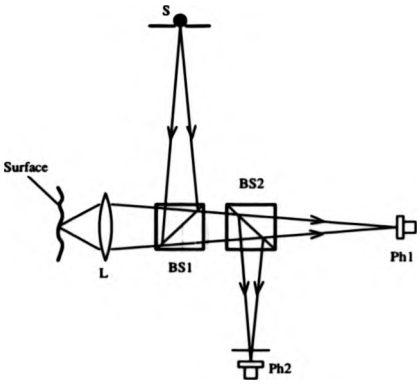


Figure 1.5: Optical system for surface measurement by detecting the change of the intensities.

Most of the optical profilometers have an equivalent or better resolution than stylus instruments. But there are some reasons why they still cannot replace mechanical profilometers. Firstly, the sensitivity of the optical methods depends on the size and the depth of the focus, the magnification of the optical system, and the wavelength of the light source, etc. Their measuring ranges are generally small and limited by the geometry of the systems. For example, the surface gradient has to be within 15° for the system by Dupuy (1967-68). Secondly, optical systems, in general, require more delicate working conditions. They are not suitable for use in the workshop. The third limitation is that surfaces measured by optical profilometers normally require a good finish or, sometimes, a good reflection.

1.2.3 Scanning probe microscopes

Optical methods are principally for speed and non-contact. Their development accelerated with the lasers existing in the 70s and early 80s, and is now growing steadily. The application of lasers to engineering made the stylus methods less important, and consequently their development lapsed somewhat.

The stylus profiling technique has its drawbacks. Its slowness and the destruction it causes to surfaces almost lead to its replacement by other techniques. However, it has become more important since the advent of alternative probes such as the scanning tunnelling microscope (STM) and the atomic force microscope (AFM). The STM and the AFM use a sharp tip to react to a surface through the physical effects. A resolution as high as the nanometer or atomic level can be reached. The tip does not physically contact the surface: rather it travels a very small distance above it. Non-contact and atomic resolution are characteristics of the scanning probe microscopes, the new generation of profiling technique. The possibility of measuring at the atomic level implies:

- Understanding surface structures at the nanometer and Ångström scales;

- Improvement of manufacturing precision;
- Further knowledge of friction and wear;
- Acquisition of detailed functional information of engineering surfaces;
- Shortening the gap between engineering and physical theories; etc.

The first STM, built by Binnig & Rohrer (1985), was inspired by the topografiner described by Young *et al* (1972). Both of them use tunnelling theory. The difference between them is that the tip is kept closer to the sample in the STM, and then the tunnelling current is used rather than the field-emission current in Young's system. This results in an atomic resolution (Binnig & Rohrer 1986). In the STM, the three dimensional variation of the charge density at a surface is probed via electron tunneling between a sharp tip and the sample. The wave nature of electrons means that they are not strictly confined to the interior bounded by the surface atoms. Therefore, the electron density does not drop to zero at the 'surface', but exponentially on the outside with a short decay length of some Ångströms. The two metals approach within a few Å; the overlap of their surrounding electron clouds is substantial, and a measurable current can be induced by applying a small voltage between them. This tunnel current, I, is dependent very strongly on the distance between the two metals, s, as

$$I \propto \exp(-A\bar{\phi}_0 s) \tag{1.1}$$

where $\bar{\phi}_0$ is the average of the work functions of the two metals. When this relation holds, the surface topographic image is obtained by varying the distance s between the tip and the surface to keep the current I constant as the tip is scanned. The electron image is very close to the atomic position (Kuk & Silverman 1989).

1.3 The necessities and requirements for evaluating the profiling technique

The characteristics of profiling techniques can be generalized as follows:

- The two-dimensional representation of a three dimensional surface from the point of view of information acquisition;
- The stylus tip, whether it is a diamond or an optical focus point, is used to interact with surfaces;
- Most applications are of geometrical measurement.

From these characteristics it can be seen that signal recording, storage and programming of the instruments are not significant in an investigation. The similarities between the stylus instrument and the novel microscopes which both use tips/probes, emphasize the importance of an investigation into the nature of the profiling technique. The different mechanisms utilized for data acquisition from the tips seem less important for the surface profiling methods. Moreover, with the decrease of measuring scale down to the atomic level, the tip, especially its dimension, becomes more significant and important.

The scanning probe microscopes have been developed dramatically. Their variety will be briefly listed in chapter 2. Nevertheless, as an instrument, there are still many requirements to be studied, for example, the mechanical and electronic designs, and the detailed reaction between the tip and the surface, etc. It is necessary to evaluate some aspects of this instrument in order to further develop its potential. The tip is one of the important factors. It plays both geometrical and physical roles in the scanning probe microscopes.

The stylus instrument is, after nearly 50 years, still the main tool used to study surface functions. Surface functional studies and this measurement rely on each

other and influence each other. However, the degree of development of surface measurement techniques and the understanding of surface functions have never seemed to be in step. In fact, the stylus profiling technique itself is a contact between a stylus and a surface. In order to see the reliability of this tool, an important point to be considered is the relation between the true surface and the surface revealed by the stylus.

The stylus profiling method is a two-dimensional measurement, but the surfaces are three dimensional. This leads to an essential disadvantage. However, digital techniques enable surfaces to be measured in three dimensions by changing the sampling from a line to a plane. An improvement in sampling on a plane, particularly from the point of view of surface functions, is needed.

The stylus has two effects on the surface: a physical effect and a geometrical effect. The geometrical effect from the stylus is normally considered from the point of view of the lateral resolution. The finite dimension is unable to track the shape of the valleys on a surface which are smaller in size. However, it has not been noticed that the surface information obtained by the stylus is also biased by the shape of the stylus. In this work, the geometrical relationship between the profile, the trace and the stylus will be analyzed. The influence of the shape of the stylus on the local information will be presented. For the stylus instrument the effect of the stylus loading force is not a new consideration. What is new in this work is that not only the static, but also the dynamic forces are used to test for the requirement for the control of the stylus loading force.

In this work, the conventional stylus instrument was chosen as the object for the experiments and tests. As it has been pointed out before, the tip in the profiling technique is the most important factor, not the mechanism. Besides, the present STMs or AFMs are still being developed. The stylus instrument (Talysurf 5) in the Centre for Microengineering and Metrology in the Department of Engineering at the University of Warwick is available and easy to use.

1.4 The layout of the thesis

The stylus profiling technique is interpreted and evaluated in terms of the mechanical aspects of data collection. This subject falls into three areas: (i) 3D digital sampling techniques; (ii) finite dimension of stylus and (iii) forces contributing to measurement. The first aspect is from the point of view of methodology applicable to the optimal recording of any profiling technique. The latter two are the two critical elements inside the techniques. They perform the essential roles. Chapter 2 gives the basic background of the subject on surfaces and their measurement. This is the area where the stylus profiling technique is intensively used. Digital techniques are the tool for three dimensional surface measurement using the stylus profiling technique. Digital data collection is considered in chapter 3 and it is shown that a novel plane sampling model - a hexagonal scheme - significantly improves the surface 3D sampling technique. Chapter 4, 5 and 6 are concerned with the relationship between the geometry of the stylus and data collection. That is, the effect of stylus finite dimension will be discussed in chapter 4. The formation of a stylus trace from a given stylus dimension and a given surface is analyzed in chapter 5. This provides a method to correct the bias caused by the shape of the stylus. As an application of these principles, a method for the measurement of the dimension of the stylus is presented and tested in chapter 6. The last part of the thesis discusses the physical aspect of the stylus. Because of the instrument-dependence, a particular stylus instrument is selected for this analysis. Its characteristics are discussed first in chapter 7. Then, in chapter 8, the measurement of the forces involved in the instrument are presented and discussed. The last chapter contains a general discussion of some of these aspects for the scanning probe microscopes, and suggests future trends.

Chapter 2

Background — Engineering Surfaces and Their Measurement

2.1 Engineering surfaces

2.1.1 What is an engineering surface?

A real surface is defined by its surface separating surrounding space (BS 1134:PART 1:1988). Engineering surfaces are the surfaces of engineering components. They are both a mirror which reflects the manufacturing process, and a window through which the performance of components or parts may be viewed. The important characteristics of engineering surfaces are their geometric features and their physical properties, e.g. thermal, electrical and optical behaviour. Normally, an engineering surface refers to those irregularities, with regular or irregular spacing, that tend to form a pattern or texture on the surface, namely 'surface texture'.

It should be clear that the reason for defining an engineering surface is to study the performance of engineering components or machine parts. Initially, researchers concentrated their efforts investigating surface texture. Now, developments in science and technology show that the structure beneath the surface also plays an important role. For instance, when an X-ray beam illuminates a surface, it will penetrate to a depth of about 10 - 10,000 nm. Even waves with optical frequencies will penetrate about 5 nm in metals. The structure beneath the surface generally determines the surface's optical properties. Thus, the definition of an engineering surface needs to be broadened to include its function: an engineering surface is the boundary of an engineering part, including the surface geometry and the layers beneath.

2.1.2 The structure of a surface

An in depth examination of a metallic surface is illustrated schematically in figure 2.1. The outermost surface is often covered by several layers, for example, an oxide layer, arising from the interaction of a surface material with its environment. The structure contains a quenched layer, a deformed layer and a base material. The

top two crystalline layers are created by the manufacturing processes. The 'Beilby layer' is thought to be a thin, amorphous layer immediately below the oxide layer. It is the result of certain machining/polishing processes. Finally, the surface is usually covered with dust, wear debris and possibly lubricants. Such particles may be of similar size to the intrinsic roughness of the surface boundary.

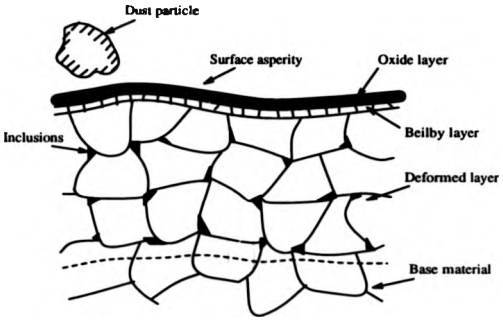


Figure 2.1: A schematic showing the structure of a metallic surface.

The crystallographic structure of the surface layers effects the mechanical properties. Some materials have different strengths in different directions due to the crystalline structure of the materials.

2.1.3 Surface topography

The geometric texture of ordinary surfaces is controlled by the characteristics of the finishing process by which they are produced. Close examination of these surfaces, even after the most careful finishing, shows that they are still rough on a microscopic

scale. In general the surface geometry is described by three parameters: roughness, waviness and error of form. Roughness (micro-roughness) is formed by the fluctuations in a surface. It is characterized by hills and valleys of different amplitudes and spacings. Waviness (macro-roughness) is the variation of the surface over large spacings. It is often caused by surface vibrations during machining, or sometimes by errors in the machine tool itself. Finally, error of form is caused by weight effects, thermal effects and so on. These three parameters characterize a solid surface as shown in figure 2.2. Some examples of engineering surfaces are illustrated in

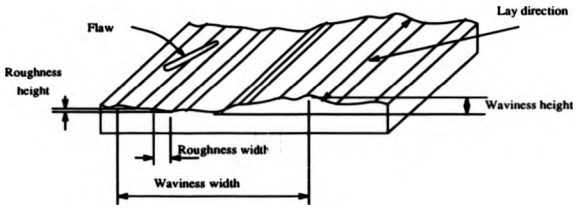


Figure 2.2: The characteristic components of a technical surface.

figure 2.3

A second feature is the machine mark or pattern. Some manufacturing processes such as turning or grinding produce surfaces which have a directional characteristic. This characteristic is known as the 'lay' of the surface. The assessment of a surface and the numerical values obtained are directly related to the lay. For example, if measurements are taken across the lay then the numerical values are likely to be significantly larger than measurements taken with the lay. Surfaces which have a directional lay are called 'anisotropic', those that do not are called 'isotropic'.

Flaws are normally caused by accidental scratches or indentations, or by defects

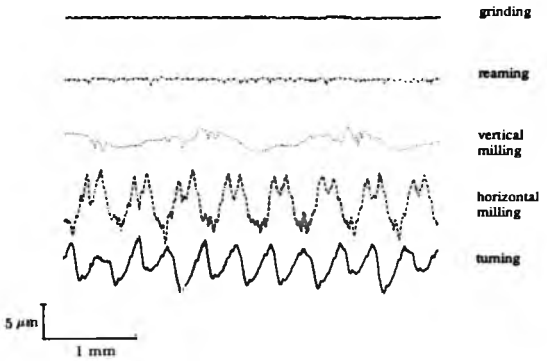


Figure 2.3: Some typical engineering surfaces.

of manufacturing operations such as air barbotage in casting. In general the size or depth of a flaw is larger than the surface texture.

Both lay and flaws are mainly characteristics of size or space within the surface planes, although the depth or height is also important. However, if the stylus type of instrument is used, full spatial information cannot be obtained because only one axis along the surface plane is scanned. At present, there is no existing parameter that can describe the spatial features in a simple way such as a single value. This is still a major problem in surface metrology.

2.2 Surface properties and their function

The surface is a link between the manufacture of an engineering component and its performance (Whitehouse 1978). It provides a way to control the machining process, to diagnose the machine tool conditions, and to predict how well the part functions. In manufacturing processes, the surfaces of components can be regarded as 'fingerprints' of the whole machined condition. The shapes of surfaces are the cumulative results of cutting tool conditions like tool wear, machine vibrations and system errors. Any deviation in manufacture reflects itself in surface texture. So, surface measurement as an effective 'go-gauge' can control the manufacturing processes. This was an early use of surface measurement.

The surfaces of engineering components are formed before being fixed into machines. In conventional engineering, when a machine is operating, its parts are influenced very much by their own surfaces, and the surfaces themselves change. Interaction between surfaces, for example friction, causes the temperature to rise, and consequently leads to changes in shape and, as a result, the accuracy of the machine may deteriorate. Moreover, the life of a machine is influenced by the failure of components due to wear. Although lubrication can take heat away and form a film between the contacting surfaces, separating them, the film may still be penetrated

by surface irregularities, so the performance of the film is related to the surface. Therefore, the functional performance of a surface must be predicted from surface measurement.

As mentioned before, light penetrates into metals, e.g. 15 nm for gold (Hummel, 1985). As a consequence, optical properties are basically measured near the surface which is susceptible to oxidation, deformation (polishing), or contamination by absorbed layers. Conversely, subsurface structures effect the optical properties and then the results of optical methods of surface measurement.

2.2.1 The mechanical properties of a surface

The mechanical properties of a surface refers to its performance under the contacting conditions, such as its strength to resist wear and load carrying capacity, etc. The science of contact is called 'Tribology'. It includes friction, wear and lubrication of interfaces.

The basic law of friction goes back to Lenardo da Vinci, and was later rediscovered by Amontons in 1699. It states that friction is proportional to the load, and is independent of the area of the sliding surfaces. The work done by Bowden & Tabor and their collaborators in the 30's and 40's of this century has advanced the subject greatly. The basic concepts will be briefly outlined in this section. Recent developments will also be summarized.

Because surfaces are rough, the true area of contact is very much smaller than the apparent area in contact. Early studies of friction, in which two surfaces were brought into contact and then slid apart, proved that the frictional force is independent of the apparent area of contact, and is proportional to the applied load. Contact only occurs at the tips of the surface peaks. Since the range of surface forces is extremely small, the surfaces are, in effect, completely separated, and have no interaction with one another. The load is therefore borne on the tips of the asperities (Bowden & Tabor 1967) (see figure 2.4).

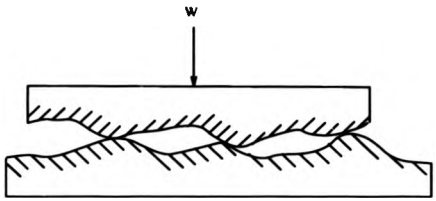


Figure 2.4: Diagram showing areas of real contact between two rough surfaces.

Contact pressure is very high, due to tip contact under an applied load. This will lead, initially, to elastic deformation and then to plastic deformation as the load increases. It is not surprising to find that the junctions formed in ordinary adhesion are as strong as the yield strengths of the materials. The adhesion process leads to friction. It stands to reason that sliding one surface over the other involves shearing of the junctions formed by adhesion. The shearing force F is equal to the shear strength per unit area, S_c (the critical shearing stress), times the area of the junctions A, i.e.

$$F = S_c \cdot A. \tag{2.1}$$

The area is given by the applied load W divided by the plastic flow pressure p, or

$$W = p \cdot A. \tag{2.2}$$

So that the coefficient of friction μ is given by

$$\mu = \frac{F}{W} = \frac{S_c}{p}. \tag{2.3}$$

At the loads normally used in engineering practice, the surface roughness will

only be compressed by about 10%. At loads greater than this value the underlying bulk material reaches the yield stress and deforms plastically (Halling 1976). In addition, as the load increases, the higher asperities deform plastically. New asperities which have lower height, come into contact, and start to deform elastically. In actual contact under an applied load, surface asperities are deformed both plastically and elastically. The ratio of plastic deformation to elastic deformation changes with the increasing load and the height of the asperities.

Recently, knowledge of friction has improved markedly. At low loads, Bowden & Tabor theory is valid. The molecular contact area is dependent upon the normal load, by JKR (Johnston-Kendall-Roberts) theory for adhesive contact and by Hertz theory for non-adhesive contacts. At high loads, Amonton's law of friction contributes to the Bowden & Tabor theory. In summary, the frictional force F is given as (Homola et al 1990)

$$F = C_1 W^{\frac{2}{3}} + C_2 W + C_3 W^{\frac{1}{3}}. \tag{2.4}$$

The first term is under strong adhesive conditions, in which the friction is proportional to the cube root of the normal load squared. The second term is the 'normal' friction described by Amonton's law, where $C_2 = \mu$. The last term always has a small value, according to Homola *et al* (1990). It is influenced by the relative magnitude of surface energy, surface asperities radii, elastic modulus and external load.

As seen in the surface structure represented in section 2.1.2, most surfaces are covered by a variety of layers. The mechanical properties of these layers are, in general, different from their bulk structures. Therefore, the layers play a critical part in a contact situation. The third-body or interface approach is useful as this model presents all of Tribology, from thick film lubrication to dry friction. Goget (1990) has surveyed this topic and described its theories.

Other mechanical features are stress (load intensity, that is, load per unit area)

and strain (deformation characteristics) that occur in the surface and in its ability to sustain such stresses without failure. This information can be determined if we know the elastic constants of the material, its yield strength (which defines the stress at which elastic behaviour changes to plastic deformation), and the hardness of the surface material which is closely related to its yield strength.

When two contacting surfaces undergo a sliding motion, the shear strength of the materials determines performance under static and dynamic situations. The static situation occurs when two asperities are initially in contact. This then gives way to a tangential force. The dynamic situation may be more complex. Two surfaces may shear with a relative velocity in a variety of directions. Two asperities which are not in contact at the beginning, may touch each other with a certain velocity, and then shear occurs.

2.2.2 The physical properties of a surface

This is normally referred to as the physical properties of the materials. The optical, electrical, magnetic and thermal properties of materials can be explained by the electron theory of solids. Hummel (1985) gives detailed information of material properties. Only a brief summary is given in this section.

Thermal properties

Thermal effects are very much influenced by surfaces. Because of the very high pressure on the contact at asperities, the thermal effect is very important during frictional processes. Most of the time during surface contact, energy is lost as heat, and is dissipated either into the surface or into the environment. The heat that goes into a surface causes geometric changes that depend on the thermal conductivities of the materials. Normally, materials which have high electrical conductivity afford high thermal conductivity owing to the large number of free electrons.

Optical properties

When optical techniques are used, a critical role in surface measurement is played by the optical properties of the surface. In particular, metals have very complicated optical properties. Reflection, absorption, and polarization of light by the surfaces must all be considered together. The atomic theory of optical properties contains two parts: free electron theory and dielectric materials theory. At low frequency, electrons in metals behave as free electrons. But, at high frequencies, they become bound electrons, like the ones in insulators. In the latter case, the behaviour of the electrons can be explained by harmonic oscillation. Therefore, the optical properties of metals may be described by postulating a certain number of free electrons, and a certain number of harmonic oscillations. Both the electrons and oscillations contribute to the polarization.

In modern technology, a knowledge of the optical properties of materials is needed in a great number of utilizations and applications. Reflectrometry and spectrometry analyze the composition of metal alloys and the electronic structure of metals. In electrochemical corrosion, optical studies provide a simple, nondestructive tool which can be used to investigate environmentally induced changes in surfaces, e.g. oxide formation.

Semiconductor studies have greatly advanced modern science and technology. A perfect crystal of a semiconductor would become indistinguishable from an insulator at absolute zero, but at finite temperatures electrons may be excited thermally across the relatively narrow forbidden energy 'gap'. Optical absorption measurements provide the most accurate way to determine the gap energy and the energy of localized states.

LASER (Light Amplification by Stimulated Emission of Radiation) action can be explained very clearly by atomic theory. Laser light is highly monochromatic because it is generated by electron transitions between two narrow energy levels. It also has an outstanding feature of strong collimation. The laser's existing applications and

future potential need not be mentioned here.

Electrical properties

The electrical properties of materials are partially described by their resistivity, ρ. The resistivity is influenced by a thermal effect, and by imperfections and defects of the material's structure. As the temperature decreases, the thermal resistivity approaches zero. At very low temperatures, near to absolute zero, superconductivity, i.e. $\rho = 0$, occurs in some materials. Superconductive materials have many potential uses, for example, high magnetic field solenoids with no power dissipation, frictionless bearings and transmission of power at 100% efficiency, etc. Superconductivity can be destroyed by high temperatures, high magnetic fields, high current densities and ferromagnetic impurities.

The electrical properties of surfaces have also been used extensively to study surface functional performance such as contact area under a known load (Holm 1946, Bowden & Tabor 1939). If the surface is free of contamination, the electrical resistance depends only upon the electrical conductivity of the materials (metals), and upon the size of the regions of contact. The resistance is the result of the constriction of the current into the region of contact.

Not only metals but also polymers, ceramics and amorphous materials can be conductive. Polymers have been created that are not only insulators but semiconductors and superconductors as well. However, there are still some problems at present. For example, they are unstable in air at temperatures much higher than room temperature. A diagram of the conductivity of polymers is shown in figure 2.5.

2.2.3 The chemical properties of a surface

Chemical reactivity has to be considered in the study of the functional performance of a surface. Most operations take place in a normal environment. The compo-

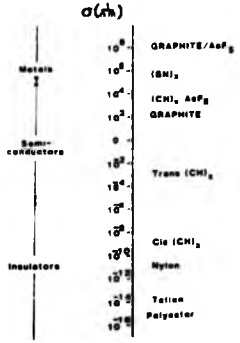


Figure 2.5: The conductivities of polymers in $\Omega^{-1}m^{-1}$ compared with metals, semiconductors and insulators.

nents of the atmosphere, especially oxygen, will react with surface materials. In the structure of a metal, illustrated in figure 2.1, the outermost layer is sometimes an oxide film. The reaction of a bare metal surface with oxygen is extremely rapid. At very low temperatures, it is proportional to the logarithm of the exposure time. When this occurs an oxide film is formed covering the bulk material. Studies have shown that the oxide film is much more thermodynamically stable than the bulk material. The conductivity of oxides is similar to that of semiconductors, with a range of 10^{-9} (BeO) to 10^{2} (FeO) $\Omega^{-1}m^{-1}$. The formation of the oxide depends upon pressure, temperature, supersatuation (no free valencies) and oxide orientation. The structures of the oxides are complicated. Their properties vary through the thickness of the layer. For instance, the amount of oxygen in the layer is higher towards the side of oxygen orientation and less towards the metal side. The lattice structure and the chemical constitution are different as well, for example, the oxides of copper are Cu_2O and CuO. Also, many structural defects occur in oxides. Studies of oxidation have shown that there is either compressive or tensile stress in oxides. This may lead to fractures or breakaway oxidation depending upon the volume ratio of oxide/metal.

The atmosphere or environment may cause corrosion in other ways. These relate to the degree of humidity and the content. Normally the moisture layer (mainly water) is no thicker than two molecules, even when the degree of humidity is about 90%. The more polluted the atmosphere is, the more aggressive the formation of a condensed film. The reaction of a metal surface with a salt deposit, often a sulphate, is referred to as 'hot corrosion'. The effects of this type of corrosion is catastrophic. Studies of oxidation and corrosion are beyond the scope of this thesis. More detailed information can be found in the book by Evans (1976).

Other films may be formed if a lubricant is used. Such films change the behaviour from that of a pure metallic surface. However, most lubricants do not have a chemical reaction with metal surfaces or oxide films. In boundary lubrication, a

small quantity of certain organic compounds, not themselves lubricants, is sometimes added to the mineral oils. They normally consist of long chain molecules with polar end-groups. The polar end groups attach themselves to the surface of the metal, while lateral attraction between the molecules holds the lubricant together. So, surface films which are strongly attached to the metal surfaces are formed. This type of lubrication is able to resist penetration by surface asperities, and provides low shear strength and low friction.

2.2.4 Surface functions in engineering

Surface texture is important in engineering applications. It has been noticed that components used for a particular purpose may require a specific surface texture to improve their functions. For instance, the inner surface of a cylinder in a motor engine needs a good surface finish to reduce the effects of friction and wear caused by the piston. However, in order to have well lubricated contact between the cylinder surface and the piston, relatively deep valley scratches are needed on the surface texture of the cylinder so that the lubricant can be kept on the surface. In addition, a good surface finish can reduce noise effectively in bearing systems. Examples of applications requiring the control of the surface texture can be found in an engineering handbook. They are summed up below (Farago 1982).

- Resistance to wear;
- Fit clearance or interference;
- Preservation of an uninterrupted lubricant film;
- Load carrying capacity;
- Resistance to chipping;
- Resistance to corrosion;

- Low coefficient of sliding friction;
- Smooth rolling contact;
- Reduced vibration and noise;
- Avoidance of abrasive effects;
- Prolonged service life through increased fatigue strength;
- Assurance of structural strength;
- Dependable dimensional measurement;
- Smooth fluid flow;
- Base for developing ultra thin film;
- Non-functional (aesthetic) considerations.

2.3 Numerical descriptions of surfaces

The object of measurement is to obtain assurance in order to control processes qualitatively. The relationship between measurands and measuring instrument is illustrated in figure 2.6. Metrology is the study of measurement. It includes calibration, traceability and standards. Surface metrology, the science of measuring engineering surfaces, is concerned both with the detection of errors and deficiencies in production processes, and with linking production criteria to component performance etc. (Scott 1986).

The definitions, expressions and standards, including the instruments that access surface textures, are listed in BS 1134:PART 1:1988.

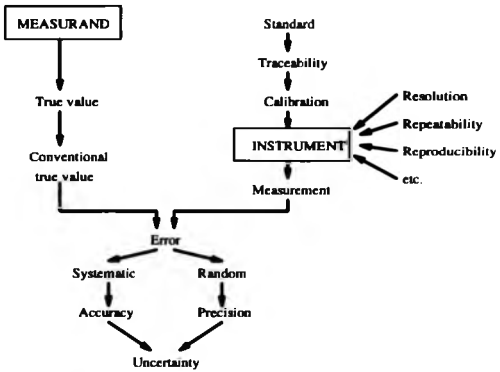


Figure 2.6: Measurement terminology.

2.3.1 Statistical aspects of surface parameters

Surface profiles often reveal both a periodic and a random component in their geometrical variation. The size of the periodic or random component, depends on the machining processes. For example, surfaces produced by turning/diamond turning have a strong periodicity depending upon the tool shape and feeding speed of the tool. Whereas in grinding and electro-discharge machining (EDM), random surfaces are manufactured. Statistical analysis provides a macroview of surface texture. Simple surface statistical parameters are used to ease industrial evaluation of surface finishes (See BS 1134: PART 1: 1988). In this section, some statistics other than those in BS 1134: PART 1: 1988 will be listed.

The height distribution of surface textures

The distribution function of the height describes the surface texture in terms of its profile. It is given as

$$F(y) = \int_{-\infty}^{y} \psi(t)\mathrm{d}t, \tag{2.5}$$

where y refers to the heights measured from the centre-line and $\psi(y)$ is the probability density function of the distribution of these heights.

Skewness and kurtosis

Skewness is a measure of the departure of a distribution curve from symmetry. It is defined as

$$s = \frac{1}{\sigma^3}\int_{-\infty}^{\infty} y^3\psi(y)\mathrm{d}y. \tag{2.6}$$

It is the third central moment of $\psi(y)$ divided by its standard deviation cubed (σ^3), and is a guide to the symmetry or otherwise of $\psi(y)$.

The kurtosis is a measure of the hump on a distribution curve. This is defined as

$$k = \frac{1}{\sigma^4}\int_{-\infty}^{\infty} y^4\psi(y)\mathrm{d}y. \tag{2.7}$$

Figure 2.7: The height distribution of surface textures.

It is the fourth central moment of $\psi(y)$ by σ^4. Figure 2.8 shows representations of them.

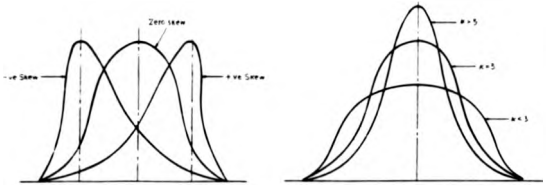


Figure 2.8: Skewness and kurtosis.

The autocorrelation function of a surface profile (ACF)

This is the result of dividing the autocovariance function $R(\tau)$ by σ^2, as follows

$$R(\tau) = \lim_{L\to\infty} \frac{1}{L} \int_{-L/2}^{L/2} y(x)y(x+\tau)\mathrm{d}x, \tag{2.8}$$

$$\rho(\tau) = \frac{R(\tau)}{\sigma^2}. \tag{2.9}$$

where L is the total sampling length of a profile. This function is dependent on the sampling interval. It can be seen that

$$R(0) = R_q^2, \tag{2.10}$$

where R_q is the rms roughness parameter (BS 1134: PART 1: 1988). This relation means that the square of the rms roughness parameter equals the ACF with zero shift.

It should be noticed that the defiritions of all the above functions are based on the surface profile. In three dimensions, the definitions may well change. Some are defined in chapter 3. They are listed because of their present utilization in surface texture analysis.

The power spectral of density function (PSD)

The Fourier transform of $R(\tau)$ is called the power spectral density, i.e.

$$P(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R(\tau)\exp(-i\omega\tau)\mathrm{d}\tau. \tag{2.11}$$

The inverse Fourier relation holds, that is

$$R(\tau) = \int_{-\infty}^{\infty} P(\omega)\exp(i\omega\tau)\mathrm{d}\omega. \tag{2.12}$$

From equation 2.10 and 2.12 it follows that

$$R_q^2 = R(0) = \int_{-\infty}^{\infty} P(\omega)\mathrm{d}\omega. \tag{2.13}$$

This indicates that $P(\omega)$ is a decomposition of R_q into contributions from various spectral components, which are waves with a spatial wavelength $\lambda = \frac{2\pi}{\omega}$.

2.3.2 Surface parameters and surface functions

Existing surface parameters are mostly defined statistically. This means that information about the height of the surface is available as a single value. Therefore, it is easier to control surface finish in the workshop. Studies, for example in bearing technology, have revealed a few relations between the present characterization and functional performance.

The surface parameter used most often, R_a (formerly CLA), is the mean of the absolute profile height (see appendix F). It has no direct effect on static friction, except for ground surfaces (maybe because these surfaces have similar sized fluctuations both laterally and in height). But it could be the right parameter to relate the drop in fatigue strength of a spark-eroded component (Ghabrial 1991). It may be that the maximum depth of grinding 'valleys' is closely related to the fatigue strength.

It is well known that skewness, R_{sk}, illustrates the shape of a surface. A positive value gives sharp peaks, and a negative value produces flat peaks and sharp/narrow valleys. The latter type of surface ensures good stability against wear because it will have a large contact area and, therefore, low bearing pressure.

Less attention has been paid to waviness than roughness, but it is also a very important parameter in surface performance. O'Connor (1991) has studied the causes and effects of waviness. It has been demonstrated that the area of contact is dependent on waviness, and also on how the contact level changes with waviness.

In addition, it characterizes the machining process, for instance, uniformity of the grinding wheel or tool vibration in turning.

Nevertheless, the use of statistical averaging parameters results in loss of local information. More and more research has shown the importance of local surface information. The tendency in surface metrology is to search for more sophisticated parameters related to surface functions. However, because surfaces are used under a variety of conditions, it is almost impossible to cover them all with a single parameter. Additionally, due to drawbacks in existing surface measuring techniques, the information acquired from a surface is incomplete. For example, stylus instruments ignore lay information. The lay direction is functionally important to friction/wear and leakage of a seal (Sulivan *et al* 1991). Intensive studies have to be carried out in the field of surface functions, e.g. friction, wear and lubrication, etc. in order to define functional and geometrical parameters. Surface measuring techniques must also improve in order to assess surfaces in a complete manner.

2.4 Surface measurement techniques and instruments — Panorama

In surface measurement, there are many different techniques to choose from. As mentioned earlier (in chapter 1), the developments in surface measurement techniques have followed the trends in science and technology. For example, the rolling ball and flowing drop methods are based on a simple mechanism. Two early electrical methods are capacitive and inductive sensing. Two non-destructive optical methods, optical sectioning and the glossometer, have a good resolution (refer to Thomas (1982) for details of the old methods). The next revolution occurred with the introduction of digital techniques, and the stylus instrument. Electron microscopes brought a very detailed view. And the discovery of lasers has brought many classic optical theories into practice. Scanning tunneling microscopy (STM) and

atomic force microscopy (AFM) enable a single atom or a molecule to be chased and arranged. Nowadays, surface measuring instruments have become more systematic, with mechanical, optical and digital techniques together.

There are different ways to classify surface measurement. For instance, according to the theory used, there are mechanical, electrical and optical methods. The methods are classified into contact or non-contact methods, depending on the characteristics of the instrument. By virtue of results derived from instruments, parametric and non-parametric methods have been used. In this work, a simple criterion is adapted for the classification. That is profile measurement and area measurement due to the area which the instruments can measure.

2.4.1 Profile measurement

Measuring a profile is a way of representing a surface in two dimensions. The profile measures the heights of hills and valleys, perpendicular to the lay. Because most machined surfaces have very clear lays, profiles from those surfaces contain enough information about the quality of surface finish and the tool conditions. They are obtained by a probe-type instrument such as the Talysurf series or optical probes. The parameters of surface textures, representing height, slope and wavelength, are easily derived from these profiles by digital software, e.g. frequency analysis by Fourier transformation and autocorrelation function. Two typical surfaces, periodic and random, are easily identifiable from their profiles. Much statistical analysis based on these profiles has been developed (Peklenik 1967-8, Whitehouse 1970, Greenwood 1984). The parameters that are normally used in standards are still derived from these profiles. Profiles are measured by stylus instruments, optical probes, optical/taper sectioning methods, scanning electron microscopes and some other methods. A non-contact profile measurement has been developed by Sommargren (1981). It gives a height sensitivity of the order 1 Å. It is based on a common path heterodyne interferometer in which two orthogonally polarized beams of slightly

different frequency are focused on the surface to be measured.

2.4.2 Area measurement

Most measurements using optical principles belong to this group. They represent surfaces in a complete manner, i.e. in three dimensions. The earliest instrument was the glossometer. A light is shone onto an area of a surface and is scattered from it. By taking the light intensities at the specular direction and at a certain diffusive angle from the reflection, the surface is estimated. The most powerful optical measuring instrument is the interferometer. The most popular commercially available interferometer system is produced by the Wyko Cooperation. It uses phase-measurement interferometry principles to obtain data. A projected fringe contour map can be given in a few seconds (Creath & Wyant, 1988). The diffraction and speckle techniques are parametric. The holographic method can give a good three-dimensional picture. Some electronic methods using capacitance and induction gauges also map areas of surfaces, but their scans are very slow compared to optical methods.

2.4.3 Discussion of different measuring techniques

The list of all different kinds of surface measuring techniques is very very long (see Sherrington & Smith 1988). For many years there have been attempts to develop a universal, reliable, non-destructive surface measuring system to suit the requirements of industry. However, because of various difficulties, most of the instruments developed can be used only for a specific purpose. At present, only the stylus instrument is near to the ultimate aim. But it is known that there are two intrinsic drawbacks with this type of instrument: it contacts the surface and its operation is slow. Although optical surface instruments make up the deficiencies of mechanical instruments, they have their own problems.

With the exception of the limitations caused by the wavelength of a light source,

the parameters from the results of optical systems cannot match the existing parameters from the stylus instrument. Many researchers have attempted to bridge this gap between optical and mechanical measurement. One reason for the gap can be seen from the criterion for classification used in this chapter. In profile-type measurements, three-dimensional surfaces are only recorded by two-dimensional features. It loses at least one-third of the information from a surface. For example, it is appropriate to record a difference between a peak on a stylus instrument and an actual summit on the real surface. Since the profilometer will register a peak even when it is traversing the shoulder of an actual summit, it will be clearly recognized that the number of true summits is considerably less than the number of peaks recorded.

Another reason may be due to how the surface reacts to light. As mentioned in the section above, a light beam penetrates a few nanometers into a surface, so that the reflected light will be influenced by sub-surface structures (down to a certain thickness); this thickness may, however, be strongly dependent on the materials because of variation in optical constants. The stylus type of instrument only has mechanical contact between the stylus and the surface, though this too may be influenced by the material to some extent. It is clear that both of them measure different surface properties. The optical profile type of instrument, such as the Wyko interferometer, only has a comparable result to the stylus instrument when the spectrum of the surface has the same range of frequencies for different wavelengths probed, and where the mechanical and optical response of the surface is similar. In practice, this implies that results are only similar for smooth, homogeneous (single phase) surfaces.

A good comparison of the performance of surface measuring instruments has been done in terms of the ability to measure sinusoidal perturbations of varying amplitude or wavelength, namely, 'Stedman Map', (see figure 2.9) The interpretation for the STM and AFM on the map should be noticed. The limitation of 1 nm in surface

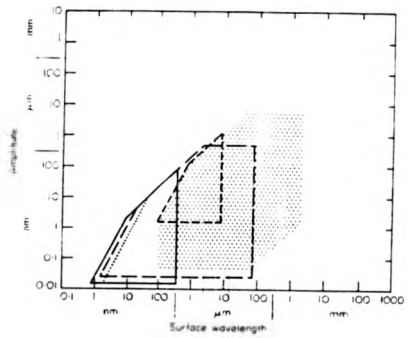


Figure 2.9: The 'Stedman Map'. Talystep: dot screened area; Topografiner: short-dashed line; 'typical' STM/AFM: solid line; dotted line shows limit with gap increased; wide range STM (Tsuda): long-dashed line.

wavelength does not mean that STM and AFM are unable to resolve atoms. As Stedman (1988) explains, there are two reasons: one is that the response from an instrument to sine waves is not linear. (This effect will be explained in chapter 4.) It is purely a geometrical effect. So, the AW map (Amplitude-Wavelength map) will not fully represent the characterization of an instrument. The other is based on one of the limiting conditions for using the AW map. That is, a stylus of a finite radius has to be able to track the bottom of the sine wave valley. For STM or AFM, it would correspond to the ability to trace an atom hole as well as a single atom. A wavelength of 1 nm is normally about the size of two atoms or more, regarded as one atom and one atom hole for each half wavelength. Thus, STM or AFM has the resolving power of one atom.

The scanning techniques recently used in microscopes are reaching a resolution of a few Ångströms on surfaces. The employment of tips means that these microscopes can be categorized into 'scanning probe instruments'. Categorized according to the physical principles used, there are the STM and AFM as mentioned before, also, the laser force microscope (LFM), the magnetic and electrostatic force microscopes (MFM and EFM), the scanning near field optical microscope (SNOM) and the scanning thermal microscope, etc. As their names suggest, these microscopes scan a surface based on their physical principles, such as electron charge density for STM, atomic force for AFM and magnetic force for MFM etc. A general review is given by Wickramasinghe (1989). Therefore, surfaces that have different physical properties, even though they have exactly the same topography, result in different images from each of the microscopes. In a word, a given surface will produce different results depending on the microscope, and surfaces composed of different materials will give different answers by the same measuring method.

This point is clearly illustrated schematically (with some uncertainty) by Bowen (1990). In the figure, a group of materials which have different properties is selected to be measured by different method. The methods range from the conventional -

Talystep - to newly developed microscopes, say, AFM. The measuring methods are arranged from top to bottom in the order of increasing attraction. For example, silicon is a hard, semiconductive material; it is opaque in ultraviolet light and transparent in infra-red light. So, a silicon surface gives a good result if it is measured by either the Talystep or an optical profilometer, but is noisy if a capacitance probe is used. Among the insulators of zerodur (a material that has almost zero expansion to any temperature rise), sapphire and rubber, rubber is soft and the other two are hard. The softest material is indium, then graphite, and lastly tool steel. The dotted horizontal lines next to each methods give the highest levels of the fidelity, in a general sense, to the surfaces of the materials in the bottom of the materials line. It can be seen from the figure that AFM gives the least variation to all the materials. Whilst some of the effects are minimal, and their representation in figure 2.10, is speculative, the figure illustrates the fact that the 'surface profile' depends upon the choice of measuring techniques.

2.4.4 The relationship between surface functions and surface measurement

It can be seen from the measuring techniques discussed above, that the functional properties of a surface plays an important role. An example has already been seen in the optical measuring system. Another example is given by Whitehouse (1988). In contact surface measuring systems, the measurement of surface finish is influenced by friction between the surface specimen and the stylus tip, which means one must take into account the contact property of surfaces, whereas in the friction measurement, the effect of surface geometry can not be ignored. Hence, surface functional properties has to be studied with surface geometry.

Moreover, surface parameters have to be adapted in more complete geometric ways as well as those containing functional meaning. The former work is not likely to be achieved unless some mathematical progress is made. In two-dimensional map-

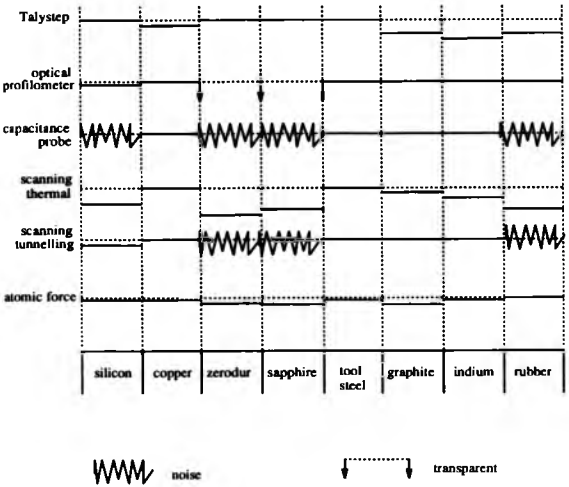


Figure 2.10: Illustration of the interaction between different surface measuring methods with different surface materials.

ping surface theory, different mapping models show different values on some features (Whitehouse & Phillips 1984, Li *et al* 1989). In order to obtain meaningful functional parameters, the behaviour of the surface must be understood. This kind of parameter should not be adopted without serious research. More basic investigation into the functional significance of sophisticated parameters is needed (Whitehouse 1982).

Looking back to the purposes of surface measurements, a similar diagram to Whitehouse's (1987) can be deduced in figure 2.11. On the one hand, in order to

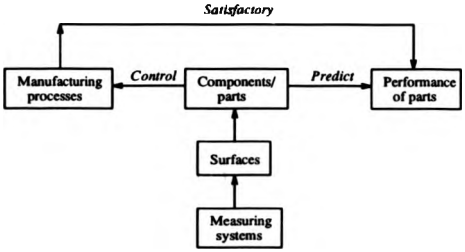


Figure 2.11: The purposes of surface measurements.

control the manufacturing process, a high speed/in-process and non-contact measuring system has to be developed to respond to advanced automation in manufacture. On the other hand, for the prediction of surface performance, not only a measuring system which is sensitive to the surface texture characteristics has to be developed, but also more functionally meaningful parameters have to be adopted. At present, the first purpose has been achieved, to some extent, by optical techniques. However, the realization of the second purpose is still under investigation.

Chapter 3

3D Surface Data Collection Using Digital Techniques

The objectives of surface and its measurement and the information required from it have been examined in the previous chapter. In this chapter, stylus profiling techniques are evaluated in terms of information acquisition with emphasis on the methodology.

3.1 Introduction

The results of a stylus scanning across a three dimensional object is a two dimensional profile. If a careful selection is made of the areas on the surface to be measured, the profile still represents some of the more significant characteristics even though one of the dimensions is missing. Clearly, the disadvantage is that the data required from the surface has to be known before its collection. Consequently, this limits the development of surface functional studies. To overcome this problem, as much information as possible must be known from measuring techniques. This was almost an impossible problem before the appearance of the computer.

Since the stylus instrument was first connected to the computer(Whitehouse & Reason 1965), the transition from analogue data obtained from surface measuring instruments to digital form, has enabled the full power and versatility of digital techniques to be used in surface assessment. In general, the continuous surface information is digitized according to the Nyquist sampling theorem. If the sampling frequency is at least twice the highest frequency in the signal, a continuous signal can be represented by, and reconstituted from, a set of sample values. In surface measurement, as long as the sampling frequency is not less than twice the surface irregularity frequency, the information on the corresponding level of the irregularity, e.g. form, waviness and roughness, will not be lost. In most cases in surface measurement, the sampling distance is used in place of the sampling frequency through the traverse unit of an instrument. Whitehouse (1969) has shown that the sampling distance is adequate with a certain correlation between two adjacent ordinates using

random analysis. For instance, for a profile with a gaussian distribution and exponential auto-correlation functions, the two adjacent ordinates could typically have a spacing of 0.7 correlation (Whitehouse & Archard 1969).

The most significant result of applying digital techniques to surface measurement is that they enable a stylus instrument to measure surfaces in three dimensions. In practice, this is achieved by stylus parallel scanning. The row spacing is equal to the sampling distance. The discretisation is, therefore, applied in the $x-$ and $y-$directions within a surface plane. The digital form of a surface looks like a net of grids. It is known as a two dimensional mapping (2D sampling). In this case, the numerical description of engineering surfaces in three dimensions cannot be replaced simply by two dimensional representation. Nayak (1971) represented the differences between the properties of the surface and the properties of the *section* of the surface, the 'surface profile', which is what is normally measured in engineering metrology.

Most surface functional studies are concerned with surface contact situations. Random analysis is applied intensively in this area (see chapter 2). The properties of the surface summits are critical. Typical properties of interest include the density of summits, their expected heights, slopes and curvatures (Tallian, *et al* 1964; Greenwood & Williamson, 1966, 1977,1984; Whitehouse & Archard 1969,1970). The Nayak random process model of rough surfaces has been studied in the discrete form (Whitehouse & Phillips 1978, 1982 &1985), using digital techniques.

The two-dimensional form of surfaces, i.e. a surface profile, consists of a large sequence of individual surface height ordinates. In this case, a peak on the profile is defined by the three-point method. A peak exists if the two contiguous digital ordinates are lower than the central one. However, moving the two dimensional profile into three dimensions, introduces new problems. A summit[1](2D peak) is now defined when a point on a two-dimensional surface is higher than all its neighbours

[1]A peak is defined only in two dimensions. In three dimension a high point is defined as a summit.

(Nayak 1971). In other words, if an ordinate is a summit it is located at the centre of a circle. The radius of the circle is the sampling distance. The central height is higher than the height of all the ordinates on the circle. Obviously, if the ordinates on the circle are discrete, a different spacing between them forms a different two dimensional mapping pattern, and it is possible that different mapping patterns affect the sampling properties of the instrument.

In this chapter, the digital aspect of data collection using sampling theory is generalized. An improvement in the sampling theory is developed, discussed and compared with other sampling models. The purpose of this chapter is to find a good balance between computer storage, speed of operation and the surface information collected.

It is necessary to be clear about the terminology. The term '1D measurement' means measuring along one dimension on the surface, the height dimension being measured to produce profiles. 2D plane sampling or 2D mapping means that there are two axes perpendicular to each other on a surface, and a height dimension. This represents a surface in three dimensions.

3.2 Review of numerical models of mapping schemes

A number of sampling schemes for sampling surfaces in a 2D plane were discussed by Whitehouse and Phillips (1985). These can best be visualized in figure 3.1 by means of a circle whose centre is an ordinate, and around the circumference of which are k evenly spaced ordinates, with angular spacing $2\pi/k$. Whitehouse & Phillips (1985) discussed the case of $k=2$ (for profile sampling), $k=3$ and $k=4$ (for 2D plane sampling). It is clear that there is a natural inclination to use more and more digital points in the modelling in order to get results which are closer to the theoretical continuous results. But, firstly, a price in the form of data storage or processing time usually has to be paid for the more comprehensive models. Secondly, it is

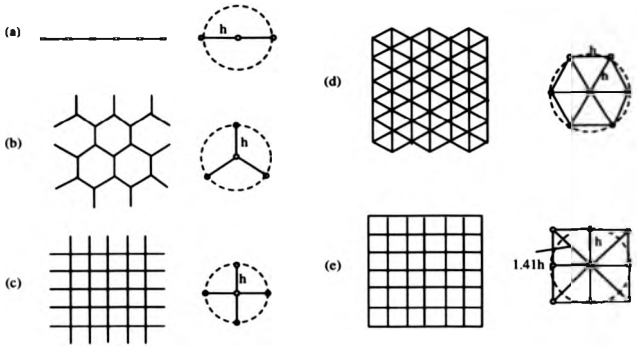


Figure 3.1: Different mapping schemes. (a) 3-point model; (b) 4-point model;(c) 5-point model; (d) 7-point model and (e) 9-point model.

even more difficult that the number of inter-correlations between the ordinates is increased, and no theoretical relationship can be easily deduced to compare with the continuous case.

The conventional 2D mapping scheme is the rectangular grid. One reason for this is the nature of human vision (Preston 1979). There is also less complexity in the calculation of the 2D fast Fourier transformation. In surface contact, the summit slope and curvature are very important elements. The 5-point rectangular sampling grid gives a four-comparison in height to define a summit, but the slope and curvature information have only two directions. In the 9-point rectangular scheme, although it shows four directional information, the height comparison is made with the ordinates having different distances away from the central one, hence there is a different correlation. This not only requires more space for data storage, and consequently the speed of measurement is lower, but also more complexity in the calculations. So, in order to get more information about slope and curvature, a 4-point triangular scheme has been considered (Whitehouse & Phillips 1985). The other advantage of this mapping pattern is that a smaller amount of data is needed. But the penalty is that less information is collected. The 7-point hexagonal sampling scheme is now introduced. It will show a good compromise between the amount of information acquired from the surface, and the storage space in the computer.

3.3 The hexagonal mapping model of surface sampling theory

Before the hexagonal mapping model is discussed in detail, it should be noted that not only surface height, but surface slope and curvature have to be considered owing to the importance of surface slope and curvature information such as summit curvature to its functions. However, surface measuring instruments presently available are generally used to measure height information. It is necessary to consider

the relationship between surface slope, curvature and height so that the gap from theoretical analysis to practical work can be bridged.

3.3.1 Definitions of summit height, slope and curvature

The hexagonal model, i.e. 7-point summit model, is defined as a central ordinate surrounded by six other ordinates, each an equal distance from the centre and evenly spaced around the circumference. Therefore, the seven points are all situated at an equal distance from the nearest points (see figure 3.2). For the purpose of compar-

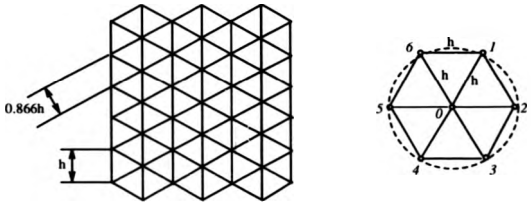


Figure 3.2: Hexagonal sampling scheme.

ison, the notation is the same as Whitehouse & Phillips' (1982) in this chapter. If the m random variables $\mathbf{X} = (x_1, \ldots, x_m)$ have a joint multivariate Gaussian distribution with mean μ and variance-covariance matrix $\mathbf{V}$, then this is denoted by $\mathbf{X} \sim N[\mu, \mathbf{V}]$. The truncated probability with zero mean is denoted by $\Phi^{(m)}(0; \mathbf{V}_m)$, and is known as the orthant probability when truncation is zero.

Considering Gaussian surfaces, the joint distribution between the height of a central ordinate (z_0) and the height differences with its six neighbours $z_i (i = 1, \ldots, 6)$, is given as (see Appendix A)

$$[z_0 \ (2 - 2\rho_1)^{-1/2}(s_1, \ldots, s_6)] \sim N[0; \mathbf{V}_7], \tag{3.1}$$

where ρ_1 is the correlation coefficient between the ordinates with a distance of h (See Fig 3. 2), $s_i = z_i - z_0$, $(i = 1, \ldots, 6)$. $\mathbf{V_7}$ is the variance-covariance of them.

$$\mathbf{V_7} = \begin{pmatrix} 1 & d\mathbf{1}' \\ d\mathbf{1} & \mathbf{V_6} \end{pmatrix}, \tag{3.2}$$

where d is the covariance between z_0 and $s_i (i = 1, \ldots, 6)$, modified by $(2 - 2\rho_1)^{-1/2}$. $\mathbf{1}$ is the unit vector. $\mathbf{V_6}$ is the variance-covariance matrix of s_i $(i = 1, \ldots, 6)$.

The slope (M) of z_0 is defined as the average slope in three directions, i.e.

$$M = \frac{1}{3}(m_1 + m_2 + m_3), \tag{3.3}$$

where $m_i = \frac{1}{2h}(s_i - s_{i+3}), i = 1, 2, 3$. Therefore, the slope M given in terms of s_i is

$$M = \frac{1}{6h}(s_1 + s_2 + s_3 - s_4 - s_5 - s_6). \tag{3.4}$$

Similarly, the curvature C of z_0 is given as

$$C = \frac{1}{3}(c_1 + c_2 + c_3), \tag{3.5}$$

where $c_i = \frac{1}{h^2}(s_i + s_{i+3})$, $i = 1, 2, 3$, or

$$C = \frac{1}{3h^2}\sum_{i=1}^{6} s_i. \tag{3.6}$$

3.3.2 The summit height density function and its expectation

Starting with the results of Whitehouse & Phillips (1984), we find the distribution of 7-point summit height ($k = 6$) derived in terms of correlation coefficients between ordinates. If T_7 is the event $(s_1 > 0, \ldots, s_6 > 0)$, the probability that a central

ordinate is a summit is given by the orthant probability

$$pr(T_7) = \Phi^{(6)}(0; \mathbf{V_6}) \tag{3.7}$$

with

$$\mathbf{V_6} = \begin{pmatrix} 1 & \frac{1}{2} & a & b & a & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} & a & b & a \\ a & \frac{1}{2} & 1 & \frac{1}{2} & a & b \\ b & a & \frac{1}{2} & 1 & \frac{1}{2} & a \\ a & b & a & \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{2} & a & b & a & \frac{1}{2} & 1 \end{pmatrix}, \tag{3.8}$$

where

$$a = \frac{1 - 2\rho_1 + \rho_{\sqrt{3}}}{2 - 2\rho_1}, \tag{3.9}$$

and

$$b = \frac{1 - 2\rho_1 + \rho_2}{2 - 2\rho_1}, \tag{3.10}$$

with the correlation coefficients $\rho_1, \rho_{\sqrt{3}}$ and ρ_2 for ordinates a distance h, $\sqrt{3}h$ and $2h$ apart, respectively. If the surface is isotropic and the autocorrelation function is $\rho(t)$, then $\rho_1 = \rho(h), \rho_{\sqrt{3}} = \rho(\sqrt{3}h)$ and $\rho_2 = \rho(2h)$. So, they vary as t varies, depending on the shape of the autocorrelation function of the surface. The distribution of the 7-point summit height is the conditional distribution of z_0 given that T_7 has occurred. This is given in Appendix B with $m = 6$. Then the density of the 7-point summit height is (Phillips 1984)

$$f(z_0|T_7) = \frac{pr(T_7|z_0)\phi(z_0)}{pr(T_7)}, \tag{3.11}$$

where

$$pr(T_7|z_0) = \Phi^{(6)}(z_0 \left(\frac{1 - \rho_1}{1 + \rho_1}\right)^{1/2}; \mathbf{V_6}). \tag{3.12}$$

Here, $\mathbf{V_c}$ is the variance-covariance matrix and has a similar form to $\mathbf{V_6}$, it is given as (see Appendix B)

$$\mathbf{V_c} = \begin{pmatrix} 1 & d_c & a_c & b_c & a_c & d_c \\ d_c & 1 & d_c & a_c & b_c & a_c \\ a_c & d_c & 1 & d_c & a_c & b_c \\ b_c & a_c & d_c & 1 & d_c & a_c \\ a_c & b_c & a_c & d_c & 1 & d_c \\ d_c & a_c & b_c & a_c & d_c & 1 \end{pmatrix}, \tag{3.13}$$

where

$$d_c = \frac{\rho_1}{1+\rho_1}, \tag{3.14}$$

$$a_c = \frac{\rho_{\sqrt{3}} - \rho_1^3}{1-\rho_1^2}, \tag{3.15}$$

$$b_c = \frac{\rho_2 - \rho_1^4}{1-\rho_1^2}. \tag{3.16}$$

The expected summit height, as $k = 6$, is given by

$$E(z_0|T_7) = \frac{3(\frac{1}{\pi}(1-\rho_1))^{\frac{1}{2}}\phi^{(5)}(0;\mathbf{B_6})}{pr(T_7)}, \tag{3.17}$$

where $\mathbf{B_6}$ is the variance-covariance matrix of the conditional distribution of the differences $s_1, \ldots, s_5$ given s_6. It is given below

$$\mathbf{B_6} = \begin{pmatrix} \frac{3}{4} & \frac{1}{2}-\frac{1}{2}a & a-\frac{1}{2}b & b-\frac{1}{2}a & a-\frac{1}{4} \\ \frac{1}{2}-\frac{1}{2}a & 1-a^2 & \frac{1}{2}-ba & a-a^2 & b-\frac{1}{2}a \\ a-\frac{1}{2}b & \frac{1}{2}-ab & 1-b^2 & \frac{1}{2}-ab & a-\frac{1}{2}b \\ b-\frac{1}{2}a & a-a^2 & \frac{1}{2}-ab & 1-a^2 & \frac{1}{2}-\frac{1}{2}a \\ a-\frac{1}{4} & b-\frac{1}{2}a & a-\frac{1}{2}b & \frac{1}{2}-\frac{1}{2}a & \frac{3}{4} \end{pmatrix} \tag{3.18}$$

and $\mathbf{B_6}$ is the function of h, and it is an orthant probability with degree of five (see Appendix A).

3.3.3 The relationship between slope, curvature and height

From the definitions of slope (M) and curvature (C), the joint probability distribution with height z_0 is

$$(M, C, z_0) \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}; \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix}\right), \tag{3.19}$$

where $c_{i,j}(i,j = 1,2,3)$ are variances and covariances of M, C and z_0, and are given as

$$c_{11} = \frac{1}{18h^2}(3 + 2\rho_1 - 2\rho_{\sqrt{3}} + 3\rho_2), \tag{3.20}$$

$$c_{22} = \frac{2}{3h^4}(7 - 10\rho_1 + 2\rho_{\sqrt{3}} + \rho_2), \tag{3.21}$$

$$c_{33} = 1, \tag{3.22}$$

$$c_{23} = c_{32} = \frac{2}{h^2}(1 - \rho_1), \tag{3.23}$$

$$c_{12} = c_{21} = c_{13} = c_{31} = 0. \tag{3.24}$$

From the variance-covariance matrix $\{c_{i,j}\}$, it can be seen that the slope M is independent of C and $z_0(c_{12} = c_{21} = c_{13} = c_{31} = 0)$. The correlation coefficient of the curvature C and the height z_0 is

$$Corr(C,\ z_0) = \frac{c_{23}}{\sqrt{c_{22}c_{33}}} = \frac{\sqrt{6}(1 - \rho_1)}{(7 - 10\rho_1 + 2\rho_{\sqrt{3}} + \rho_2)^{1/2}}. \tag{3.25}$$

The distribution of a 7-point summit having a height z_0 conditional on a curvature c is normal, and it is given by (Appendix C)

$$(z_0, |C = c) \sim N\left[\frac{3h^2(1 - \rho_1)}{7 - 10\rho_1 + 2\rho_{\sqrt{3}} + \rho_2}c,\ \frac{1 + 2\rho_1 + 2\rho_{\sqrt{3}} + \rho_2 - 6\rho_1^2}{7 - 10\rho_1 + 2\rho_{\sqrt{3}} + \rho_2}\right]. \tag{3.26}$$

In summary, it has been possible to achieve the expectations of the 7-point summit height, curvature and density with the theory of Gaussian truncated random variables (Phillips 1984).

3.4 Results for independent ordinates using different mapping schemes

The different sampling models are now considered. First of all, the relatively simple case, where all the ordinates are independent of each other, is examined. In this example, the correlation distance or sampling interval, h, tends to infinity in terms of mathematical terminology. It is relatively easy because all the cross correlations between the ordinates in the models are zero. They become independent events.

3.4.1 The probability of an ordinate being a summit as $h \rightarrow \infty$

If the sampling distance, h, tends to infinity, it means that the ordinates on a surface do not have any correlation. The probability that an ordinate is a summit depends only on the number of ordinates required for height comparison in a sampling model. It is equal to the reciprocal of the number of ordinates in the sampling model, i.e. $\frac{1}{k+1}$. Each sampling model not only has a different number of ordinates, but also covers different area. This area is measured in terms of a unit area of sampling distance squared, for example, h^2. Therefore, the probability of an ordinate being a summit in a given area is equal to the product of the probability of an ordinate being a summit in a sampling model, and the number of the sampling model needed to cover this area.

The number of basic sampling elements (e.g. rectangles or hexangons, etc.) in a given area is also called the density of the sampling model. It is given by the ratio

of h^2 and the area of the sampling model. Results for the different sampling models have been calculated and are listed in table 3.1 below.

Model	k	$\lim_{h\to\infty} pr(T_{k+1})$	Density of sampling model area	$pr(T_{k+1})$ in a given area
3-point	2	$\frac{1}{3}$	—	—
4-point	3	$\frac{1}{4}$	0.770	0.192
5-point	4	$\frac{1}{5}$	0.5	0.1
7-point	6	$\frac{1}{7}$	0.385	0.055

Table 3.1: Probability of an ordinate being a summit in a unit surface area by independent ordinates.

3.4.2 The expected summit density and the expected summit height as $h \to \infty$

The expected summit density, is the product of the probability of an ordinate being a summit and the density of ordinates $\frac{1}{h^2}$. As $h \to \infty$, the result of the expected summit density is shown in table 3.2.

Model	k	Expected density as $h \to \infty$	$\lim_{h\to\infty} E(z_0\|T_{k+1})$
3-point	2	$0.333/h$	0.846
4-point	3	$0.25/h^2$	1.029
5-point	4	$0.20/h^2$	1.163
7-point	6	$0.143/h^2$	1.350

Table 3.2: The expected summit density and the expected summit height as $h \to \infty$.

As h tends to infinity, from equation 3.17 the expected summit height is

$$\lim_{h\to\infty} E(z_0|T_7) = \lim_{h\to\infty} \frac{3\sqrt{1-\rho_1}\Phi^{(5)}(\mathbf{0};\mathbf{B_6})}{\sqrt{\pi}\Phi^{(6)}(\mathbf{0};\mathbf{V_6})} \tag{3.27}$$

The two truncated probabilities, $\Phi^{(5)}(\mathbf{0};\mathbf{B_6})$ and $\Phi^{(6)}(\mathbf{0};\mathbf{V_6})$ become the probabilities in which all the correlation coefficients are equal. They can be found by using

Bacon's (1963) approximation formula. In addition, the ordinates in this situation are independent, i.e. $\rho(h \to \infty) = 0$. Hence, the expected summit height in this case is 1.350 (see table 3.2 & Appendix D). It also can be obtained from David (1981, p.61).

3.5 The limiting behaviour of the hexagonal sampling model as $h \to 0$

The previous section discussed the situation in which all the ordinates on a surface are independent. It has been shown that different sampling models give different results for the summit density and the summit height. This implies that these parameters are not intrinsic to the surface but depend on the spacing of ordinates, number of ordinates per summit etc. And the probability of an ordinate being a summit depends on the number of ordinates involved in one sampling model. However, it is not straightforward to show the characteristic differences in the general case. Whitehouse & Phillips (1985) compared the results of the different sampling schemes, for $k = 2$, $k = 3$ and $k = 4$, using limiting conditions. As the sampling interval, h, tends to 0, the discrete model converges to the continuous case. The results for the distribution of peak and summit heights have been obtained by Rice (1945) and Nayak (1971) for the continuous Gaussian surface with zero mean and unit variance. Now these limiting results will be supplemented by the case of $k = 6$.

Rice (1945) assumed that the correlation function can be expanded near the origin in a Taylor series containing only even terms. So, the autocorrelation function $\rho(t)$ for ordinates a distance t apart will be

$$\rho(t) = 1 + D_2\frac{t^2}{2!} + D_4\frac{t^4}{4!} + o(t^4), \tag{3.28}$$

where $D_2 < 0$, $D_4 > 0$ and $\eta = -\frac{D_2}{\sqrt{D_4}} < \sqrt{\frac{5}{6}}$. D_2 and D_4 are the second and fourth

derivatives of the autocorrelation function at the origin (Whitehouse & Phillips 1985).

The density of peaks or summits is the number of peaks per unit length or of summits per unit area, using the $(k+1)$-point definition of a peak for $k = 2$ and of a summit for $k = 3$, 4 and 6. The expected density of peaks or summits is the product of the $pr(T_{k+1})$ and the density of ordinates $\frac{1}{h^2}$, where T_{k+1} is the event $(s_1 > 0, \ldots, s_k > 0)$.

The density of summits as $h \to 0$

The density of peaks and summits for isotropic continuous Gaussian surfaces are known and were given for peaks as

$$D_{peak} = \sqrt{\frac{D_4}{-D_2}} \cdot \frac{1}{2\pi} \tag{3.29}$$

by Rice (1945) and for summits as

$$D_{sum} = \frac{D_4}{-D_2} \cdot \frac{1}{6\pi\sqrt{3}} \tag{3.30}$$

by Nayak (1971). The probability that a central ordinate is a summit is given as before by the orthant probability (equation 3.7)

$$pr(T_7) = \Phi^{(6)}(0; \mathbf{V_6}).$$

However, there are no general formulae for orthant probability for $k > 3$, except for special cases (Cheng 1968). Plackett (1954) gives the reduction formula that the differential reduces the dimension of the orthant probability by two. When this method was applied, two orthant probabilities of dimension two and four were obtained which could be evaluated. So, from equation 3.7 to equation 3.10, the first

derivatives are given as below

$$\frac{\mathrm{d}}{\mathrm{d}h}\Phi^{(6)}(0;\mathbf{V_6}) = 6\frac{\partial\Phi^{(6)}}{\partial a}\cdot\frac{\mathrm{d}a}{\mathrm{d}h} + 3\frac{\partial\Phi^{(6)}}{\partial b}\cdot\frac{\mathrm{d}b}{\mathrm{d}h}, \tag{3.31}$$

where $\frac{\partial\Phi^{(6)}}{\partial a}$ and $\frac{\partial\Phi^{(6)}}{\partial b}$ can be found by Plackett's formula (1954), which reduces $\Phi^{(6)}(0;\mathbf{V_6})$ from six dimensions to four. So, equation 3.31 becomes

$$\frac{\mathrm{d}}{\mathrm{d}h}\Phi^{(6)}(0;\mathbf{V_6}) = B(h)\Phi^{(4)}(0;\mathbf{V_4}^{(b)}) + C(h)\Phi^{(4)}(0;\mathbf{V_4}^{(c)}). \tag{3.32}$$

It is possible to show that

$$\lim_{h\to 0}\Phi^{(4)}(0;\mathbf{V_4}^{(b)}) = \frac{1}{24}, \tag{3.33}$$

$$B(h) \sim \frac{2}{\pi}\left(\frac{D_4}{-D_2}\right)h, \tag{3.34}$$

and

$$C(h) \sim \frac{3}{2\pi}\left(\frac{D_4}{-D_2}\right)^{\frac{1}{2}}, \tag{3.35}$$

as h tends to 0. However,

$$\lim_{h\to 0}\Phi^{(4)}(0;\mathbf{V_4}^{(c)}) = 0. \tag{3.36}$$

To find out how this last orthant probability behaves as h tends to 0 we differentiate again. This leads to three orthant probabilities of dimension two, whose limits as h tends to 0 are 0.5, 0 and 0.5. Hence,

$$\Phi^{(4)}(0;\mathbf{V_4}^{(c)}) \sim \frac{(\sqrt{3}-1)}{3\pi}\left(\frac{D_4}{-D_2}\right)h^2 \tag{3.37}$$

as h tends to 0. Putting all these results together gives

$$\Phi^{(6)}(0;\mathbf{V_6}) \sim \frac{\pi + 6(\sqrt{3}-1)}{24\pi^2}\left(\frac{D_4}{-D_2}\right)h^2 \tag{3.38}$$

as h tends to 0. Hence, the density of the summit is

$$\lim_{h\to 0} pr(T_7) = \lim_{h\to 0} \Phi^{(6)}(0; \mathbf{V_6}) = 0.0318 \left(\frac{D_4}{-D_2}\right) h^2 \tag{3.39}$$

(see Li *et al* 1989).

The expected summit height as $h \to 0$

For isotropic continuous Gaussian surfaces, the results of expected peak or summit height of a surface with ordinates Z were known and were given for peaks as

$$E(Z|\text{cont.peak}) = \sqrt{\frac{\pi}{2}}\eta \tag{3.40}$$

by Rice (1945) and for summits as

$$E(Z|\text{cont.summit}) = \frac{4}{\sqrt{7\pi}}\eta. \tag{3.41}$$

The expected summit height $E(z_0|T_7)$, as h tends to 0, is given by

$$\lim_{h\to 0} E(z_0|T_7) = \lim_{h\to 0} \frac{3\sqrt{\frac{1-\rho_1}{7}}\Phi^{(5)}(0; \mathbf{B_6})}{pr(T_7)}, \tag{3.42}$$

where $\Phi^{(5)}(0; \mathbf{B_6})$ and $pr(T_7)$ are orthant probabilities of dimension five and six. It can be seen that

$$\sqrt{1-\rho_1} \propto h;$$

$$\Phi^{(5)}(0; \mathbf{B_5}) \propto h.$$

Each of the terms in the numerator in equation 3.42 tend to 0 at the same rate as h as $h \to 0$ because $\rho_1 = \rho(h) = E(z_0 z_*) \to E(z_0^2) = 1$. The denominator, as

mentioned before, $pr(T_7)$ is also zero as h tends to 0 but at the rate of h^2. Therefore, L'Hôspital's Rule is applied. Equation 3.42 becomes

$$\lim_{h\to 0} E(z_0|T_7) = \lim_{h\to 0} \frac{\frac{1}{2}\cdot\frac{-\rho_1'}{\sqrt{1-\rho_1}}\Phi^{(5)}(0;\mathbf{B_6}) + \sqrt{1-\rho_1}\Phi'^{(5)}(0;\mathbf{B_6})}{\Phi'^{(6)}(0;\mathbf{V_6})}. \tag{3.43}$$

Using the reduction formulae of David (1953), for integrals of odd numbers of dimensions, here $m = 5$ for $\Phi^{(5)}(0;\mathbf{B_6})$, which can be expressed by the sum of the joint probabilities whose dimensions are less than five (see Appendix E). The first derivative of the orthant probability can be obtained again using the differential reduction formula of Plackett (1954). Thus,

$$\lim_{h\to 0} E(z_0|T_7) = \frac{3\sqrt{\pi}\left[6\pi - 12\tan^{-1}(\frac{1}{\sqrt{2}} - \sqrt{2}\pi)\right]}{\sqrt{\pi}\left[\pi + 6(\sqrt{3}-1)\right]}\eta \tag{3.44}$$

$$= 0.9884\frac{4}{\sqrt{\pi}}\eta. \tag{3.45}$$

3.6 Comparison of several different mapping schemes

3.6.1 The convergence of different sampling models to the continuous case

To study the behaviour of different sampling models and make comparisons, the limiting results are derived in table 3.3 and table 3.4. The results for $k = 2, 3$ and 4 are taken from Whitehouse & Phillips (1985).

It can be seen from table 3.3 that for profiles ($k = 2$) the expected density of peaks converges to the continuous results, D_{peak}, given by Rice (1945). For the 2D plane ($k = 3$, 4 and 6), the expected density of 'discrete' summits does not converge to the continuous results, D_{sum}, given by Nayak (1971). In all cases, the limit is greater than D_{sum}, by 73% for $k = 3$, by 31% for $k = 4$ and by 4% for $k = 6$.

It can be seen from table 3.4 that the limiting value of the expected summit

Model	k	$\lim_{h\to 0} pr(T_{k+1})$		Expected density as $h \to 0$	
		T^*_{peak}	T_{peak}	D^*_{peak}	D_{peak}
3-point	2	$\frac{1}{6\pi}T^*_{sum}$	$0.0531T_{sum}$	$\sqrt{3}D^*_{sum}$	$1.732D_{sum}$
4-point	3	$\frac{\pi+2\sin^{-1}(\frac{1}{3})+4\sqrt{2}}{24\pi^2}T_{sum}$	$0.0400T_{sum}$	$\frac{\sqrt{3}\left[\pi+2\sin^{-1}(\frac{1}{3})+4\sqrt{2}\right]}{4\pi}D_{sum}$	$1.306D_{sum}$
7-point	6	$\frac{\pi+6(\sqrt{3}-1)}{24\pi^2}T_{sum}$	$0.0318T_{sum}$	$\frac{\sqrt{3}\left[\pi+6(\sqrt{3}-1)\right]}{4\pi}D_{sum}$	$1.038D_{sum}$

$$
\begin{aligned}
{}^*T_{peak} &= \frac{1}{2\pi}\left(\frac{D_4}{-D_2}\right)^{\frac{1}{4}} h \\
T_{sum} &= \left(\frac{D_4}{-D_2}\right) h^2 \\
D_{peak} &= \frac{1}{2\pi}\sqrt{\frac{D_4}{-D_2}} \\
D_{sum} &= \frac{1}{6\pi\sqrt{3}}\left(\frac{D_4}{-D_2}\right)
\end{aligned}
$$

Table 3.3: Expected summit (peak) density as $h \to 0$.

Model	k	$\lim_{h\to 0} E(z_0 \mid T_{k+1})$	
3-point	2	$\left(\frac{\pi}{3}\right)^{\frac{1}{2}}\eta$	$0.555E^*_{cont}$
4-point	3	$2\left(\frac{3}{\pi}\right)^{\frac{1}{2}}\eta$	$0.866E_{cont}$
5-point	4	$\frac{8\sqrt{3\pi}}{\pi+2\sin^{-1}(\frac{1}{3})+4\sqrt{2}}\eta$	$0.938E_{cont}$
7-point	6	$\frac{3\sqrt{\pi}\left[6\pi-12\tan^{-1}(\frac{1}{\sqrt{2}})-\sqrt{2}\pi\right]}{\sqrt{\pi}\left[\pi+6(\sqrt{3}-1)\right]}\eta$	$0.988E_{cont}$

$${}^*E_{cont} = \lim_{h\to 0} E(z_0 \mid T_{k+1}) = \frac{4}{\sqrt{\pi}}\eta.$$

Table 3.4: The expected summit (peak) height as $h \to 0$.

height as $h \to 0$ for $k = 6$ is 0.9884 $\frac{1}{\sqrt{2}}\eta$, i.e. almost 99% of the value of the expected height converges to the continuous surface given by Nayak (1971). So, it is only for profiles that the expected peak height converges to the continuous result, $E(z|cont.peak)$ given by Rice (1945). For the 2D plane, ($k \geq 3$), the expected summit height does not converge to Nayak's result, which is similar to the expected summit (peak) density in table 3.3. But in all cases, the limit is smaller, it is 87% for $k = 3$, 94% for $k = 4$ and nearly 99% for $k = 6$.

3.6.2 Computing time test by different sampling schemes

The computing time test has been simulated on an IBM-PC using Turbo-pascal (see program `smtheit.pas` in Appendix H). A matrix of random data 100 x 100 has been generated by the computer as a surface. It is based on Whitehouse's paper (1983) (see program `cordata2d.m` in Appendix H). The probability of an ordinate being a summit has been calculated. In these tests, the probability which equals the ratio of the total number of summits and the total number of ordinates in each model is given by $\frac{1}{k+1}$, i.e. the theoretical probability. The results from the test are shown in figure 3.3. The figure shows that as the number of ordinates surrounding the central ordinate increases, the computing time also increases. The decrease in the total number of summits in a certain area confirms that the simulation is correct. The computing time in the triangular case can save 30.5%, compared with the 5-point model. Whereas in the hexagonal case, it is 31.4% more. However, for storage, which is determined by ordinates being used in the total data points, the hexagonal case (7-point) is only 3.2% more than the 5-point model. The computing time for the 9-point model is about twice as much as in the 5-point model although the storage is almost the same. In the 4-point model, the storage can save about 31%.

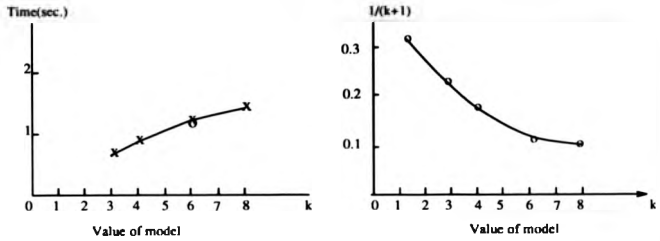


Figure 3.3: The computing time and the $pr(T_{k+1})$ by different 2D sampling models.

3.7 Discussion

3.7.1 The trends of summit parameters with the complexity of numerical models

One object of the 7-point sampling model was to determine the degree of complexity needed to approach the theoretical continuous results within a satisfactory limit. This limit has been chosen to be $\pm$ 5 %. For highly dependent ordinates, the expected summit density and the expected summit height vary with the numerical models. To illustrate this, Figure 3.4 is derived from table 3.3 and table 3.4.

It is obvious that the 4-point and 5-point models are not good enough for the summit density. It is necessary, therefore, to go to model which use more ordinates. For $k = 6$ (the hexagonal case), the digital result for the expected summit density is within $\pm$5 % of the continuous limit. Evidently, as long as the value of the model, k, gets higher, the limiting result tends more closely to the continuous result. When $k = 8$, the limiting result should be almost equal to the theoretical continuous results as h tends to zero though this has still to be verified. It could be argued that the

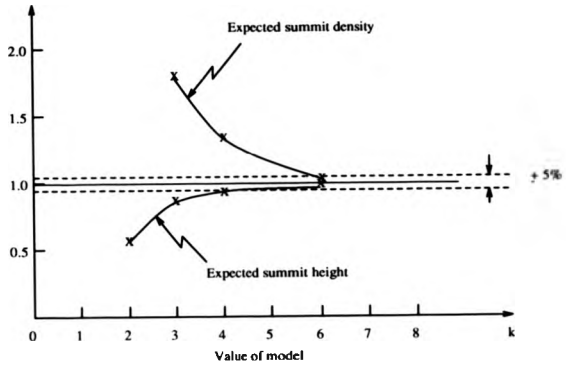


Figure 3.4: Diagram showing the expected summit density and the expected summit height for different models (indicated by their k value) as $h \to 0$.

results should be extended to this case but the indications are that this not only increases the complexity of the calculation but it is also not particularly necessary. In both cases the rate of change is much less dramatic as k gets large so that a law of diminishing returns is possible.

For the summit height, the situation as h tends to 0 for different models are far from clear. Even for $k = 4$ the result is very close to the 5 % acceptable limit. The amount of improvement is very little from $k = 4$ to $k = 6$. This may imply that it might need one more parameter to make a decision about the best model.

For the limiting case as h tends to infinity, the probability that an ordinate is a summit is $\frac{1}{k+1}$ for model k and is shown in figure 3.5 as a function of k. As the numerical model gets more complex, the asymptote of this probability goes to zero. In addition, the density of the sampling model area is absolutely dependent on the numerical models (see table 3.1). Therefore, there must be an agreement between researchers before comparison of results can be made.

3.7.2 Information Collection

It is inevitable that any digital measurement of a surface will cause a loss of information compared with a complete 'map' of the surface. However, any discrete measurement should produce results which converge to the results for the continuous surface as the sampling distance h tends to zero. As show in table 3.3, it is only the digital form of the profile that represents its continuous profile as h tends to 0. But a profile is an incomplete illustration of a 3D surface. In the three-dimensional case, as h tends to zero, the different sampling models converge towards the continuous case but end at different points. As the number of ordinates in a model increase, the point of convergence gets closer to the continuous point. This means that more information is collected.

For instance, it can be seen clearly from table 3.1 that as the number of ordinates involved in one sampling model increases, the probability of an ordinate being a

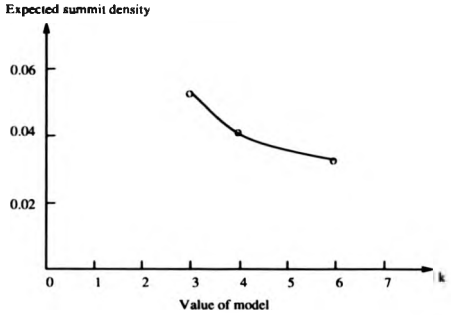


Figure 3.5: The probability of an ordinate being a summit as $h \rightarrow \infty$.

summit becomes lower. This is because the fewer the surrounding ordinates in the height comparison, the higher is the possibility of a central ordinate being regarded wrongly as a summit, especially when the ridge of a plateau runs between the surrounding ordinates. So, the more surrounding ordinates, the more accurate the determination of a summit is.

3.7.3 Correlation between ordinates

When ordinates are sampled digitally from a continuous surface, a certain correlation exists between them. The correlation is dependent on the manufacture of surfaces(Whitehouse 1982). In this chapter, the correlation is represented in terms of sampling interval h. The two extreme cases of $h \rightarrow \infty$ and $h \rightarrow 0$ are discussed, to show the differences between sampling models. In practice, the surfaces in which ordinates are either completely correlated or independent do not exist. They often

have a correlation between a extreme cases. The correlation of engineering surfaces related with their finish is another subject.

3.7.4 Considerations of surface simulation techniques based on fractal theory

As can be seen from the previous sections, the sampling theories based on the random theory are strongly dependent upon the numerical models. Moreover, the statistical parameters such as the variance of the height distribution are related to the length of the sampling. Therefore, for surface simulation, random theory is not advantageous although Whitehouse (1983) and others generated surfaces by correlation functions. Other simulation techniques based on fractal theory have also been used.

Fractal theory developed by Mandelbrot (1977) has given many new insights into nature. Gagnepain & Roques-Carmes (1986) applied these ideas to surface engineering. To explain this concept simply, a surface profile, say, is formed by a basic geometrical curve. As a small section of the curve is taken out and magnified, the same basic geometrical curve appears. This mechanism keeps operating and the surface profile can be magnified infinitely. This model or simulation can be done easily by computer. The fractal approach for characterization of engineering surfaces shows a unique 'fractal dimension' (is also called the *Hausdorff Besicovitch dimension*, a characteristic of a family of curves which is self-similar and repeated refinement) for the sampled surface. Simulation of surface profiles by Weierstrass-Mandelbrot (W-M) fractal function can obtain scale-independent fractal parameters (Majumdar & Tien, 1990).

However, some points need to be considered. Firstly, real surfaces have no 'infinity' with respect to geometry. There is a limit when the atomic level has been reached. The atomic level can be obtained by STM/AFM. At this level, fractal geometry fails. Secondly, in fractal theory, there is no intrinsic value of summit curvature, which is very important in surface contact. A fractal curve is not differ-

entiable.

One of the advantages of simulated fractal profiles is that they are not restricted by instrument-dependent filtering. However, whether a fractal simulation can describe surfaces completely still needs to be proved. Overall, fractal theory only provides a simulation technique in surface studies. The sampling theory in this chapter is from the point of view of the collection of data from surfaces. It is thus appropriate to use random processing theory rather than fractal theory as the tool to analyze the data.

3.8 Application of 2D sampling theory

3.8.1 Application of hexagonal sampling scheme

Two dimensional sampling theory is widely used in the field of image processing. An example is the 2D FFT where a rectangular scheme is used. It is conceptually easy to develop 2D processing algorithms from the 1D case. It is also natural since the electron beam in a TV camera/monitor scans with a rectangular pattern.

Other sampling schemes have been considered in this area. The hexagonal pattern, in particular, can be obtained by altering the scanning line by half of the sampling interval. Mersereau (1979) demonstrated that the hexagonal system requires 13.4% fewer sampling points to maintain the same high frequency information as an image sampled on a square grid system. Staunton (1990) has used the hexagonal sampling structure as a local operator in robot vision, and has shown that the discretization of such a local operator is relatively simple to design and more accurate. Also less digital memory is required to store an image. With various local operator implementations such as edge detectors, less computing time is needed than the square system counterpart. Discretization of raster scanned images on such a grid needs only minimal modification for frame grabber hardware. The aliasing associated with vertical lines in the hexagonal system is not a problem for robot

vision. The filtering algorithms and 2D FFT for the hexagonal sampling system have been developed by Merserean (1979). In conclusion, the hexagonal sampling systems have been shown to be advantageous compared with the conventional rectangular system in image analysis. However, whether or not this will benefit surface digital measurement, still needs to be proved.

3.8.2 Potential application of surface measurement using 2D sampling theory

The purpose of discussing the hexagonal sampling system in this chapter is also to investigate further the digital sampling theory for surface measurement. However, a full study of the potential of the hexagonal case needs an application.

The 2D sampling theory for surface measurement have been studied since 1970's, but the practical uses have not yet been improved much. One problem is the design of a precision mechanism for the moving stage related with other axes. But even if this could be overcome, there would be the drawback of using mechanical stylus measurement, that is, the speed of data collecting.

An alternative way is a photographic method. The ordinate heights referred to in the text would correspond to the intensities of pixels on the image, like SEM, and the surface parameters or other calculations can be achieved by image processing. Certainly, the speed will be improved at least a hundred times. The optical noncontact will give better measurement from the point of view of surface destruction. The key to the possible practical realization is the relationship between the pixel intensity and the local geometrical height on the surface. It includes studies of picture formation, the visual properties of sensors and optical properties of the surfaces, etc. This still remains to be studied.

3.9 Conclusion

The results for summit properties have been derived using a 7-point sampling scheme. The general method used for the extension to the 7-point model is given by Whitehouse & Phillips (1982) and Phillips (1984). The limiting results are obtained by using the reduction formula of Plackett(1954), which reduces the dimension of the orthant probabilities by two. This reduction from six dimensions, for which there are no general formulae, to four dimensions enables the limiting behaviour of the orthant probabilities to be evaluated. It is easy to find out that the orthant probabilities of four dimensions has the same rate of convergence as h. So, the L'Hôspital Rule is applied. The limiting behaviour of the 7-point model can be achieved.

Comparison between the sampling models gives insight into the effect of digital techniques of sampling methods on surface measurement. The application of the hexagonal sampling model gives potential benefits. There is only a small increase in the time needed to sample a surface, but much more information can be collected. For surfaces in which summits are very important elements, the hexagonal sampling model is better because more than 98.8% information can be collected from the surfaces. The amount of information achieved is beneficial in terms of the utilization of random processing techniques for data analysis. Nevertheless, whether this is true for other types of surface analysis must be evaluated, according to the specific application.

Chapter 4

Effects of Tip Dimension on Surface Data Collection

This chapter concentrates on the relationship between the true surface and the stylus trace. We therefore consider the intrinsic drawbacks of stylus profiling techniques. Most drawbacks are instrument dependent. They may, for example, come from datum references, mechanical systems, filtering algorithms and transducers, etc. They are not the object of this work. The most important factor in profiling systems is the stylus, which does not depend on the instrument system. The basic effects on data collection from a stylus are geometrical and physical. The next part of this thesis will deal with these two aspects, beginning with the geometrical aspect.

4.1 Introduction

In this chapter, the emphasis is now on how the geometry of a stylus affects the measured result.

Generally, the dimension of a stylus is a few microns. In the British Standards, it is stated that:

> The nominal value of the tip radius of the stylus shall be one of the following: (a) 2 ± 0.5 μm; (b) 5 ± 1 μm; (c) 10 ± 1 μm.
>
> The nominal value of the stylus angle shall be one of the following: (a) 1.57 radians (90°); (b) 1.05 radians (60°). (BS 1134: PART 1: 1988)

The international standards (ISO) has a similar statement.

It should be noticed that the tips in scanning probe microscopy play the same roles even though they have different sensing mechanisms. For example, in STM, the tip does not contact surfaces, rather, it acts as part of an electrode pair and its position is controlled to hold a constant electronic tunneling current. In AFM also, the tip does not contact surfaces and the atoms on the tip couple the molecules or atoms on the surface with the atomic force at a certain distance. Even though the tips do not contact the surface, but keep a certain distance from it, studies (Kaizuka 1990, Nakagiri & Kaizuka 1990, Musselman *et al* 1990) have shown that

the topographic and work function images of STMs/AFMs depend on tip size. On STM/AFM the effect of the tip geometry can be considered as a stylus in the conventional stylus instrument whose tip is the size of the true tip size plus the constant gap (Stedman 1987a, 1987b). Under this condition, the characterization of a stylus in this chapter is valuable to STM/AFM. The analysis in this chapter is based on a stylus in the conventional measuring instruments.

Before the discussion of stylus tip effects, it is useful to know how stylus tips are manufactured.

4.2 Fabrication of stylus tips

Originally the stylus was a phonograph needle (Abbott & Goldschmidt 1937). Because of rapid wear, a diamond stylus has been used. Today diamond stylii are universally employed: either 60° cones (Williamson 1947) or truncated 90° pyramids. The tip of the pyramid is about 3 μm^2 to 8 μm^2 in area, with the short edge parallel to the direction of travel (Jungles & Whitehouse 1970). In order to achieve atomic lateral resolution, the tip must have an apex with a radius at least below 10 Å. Fabrication of stylus tips can be catalogued into two ways: (i) from small to large scale and (ii) from large to small scale.

4.2.1 Building-up technique

In order to produce a tip so sharp that only a single atom is at the top, it is ideal that atoms are piled up from one atom to form a stylus (see figure 4.2.1). (Piling up is used, rather than lining up, to give good stiffness). This principle is regarded as a way of building up from a small to a large scale. However, because of the interaction between atoms etc., this method, while it is simple in principle, is very complicated in theory and very difficult in practice.

Binh (1988) has tried the 'build-up' technique to produce microtips. The initially

Figure 4.1: The ideal way of fabricating a stylus tip by arranging atoms.

nearly hemispherical field emitter tip surface deforms into a polyhedral shape in a process known as 'build-up', when the emitter is heated in the presence of a high electric field. The build-up process consists essentially of a local rearrangement by surface diffusion of the atoms of some low index facets, leading to an enlargement of these planes (see figure 4.2). This local migration is essentially due to the presence

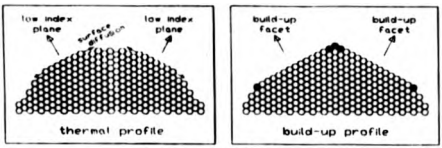


Figure 4.2: The local arrangement of atoms in the build-up method by Binh (1988).

of a gradient of the electric field between the centre of the facet and the vicinal regions. The facet enlargement will stop by itself when two neighbouring facets meet each other, that means when they are separated by a one-atom boundary line. A tip made in this way and observed by field ion microscopy is shown in figure 4.3.

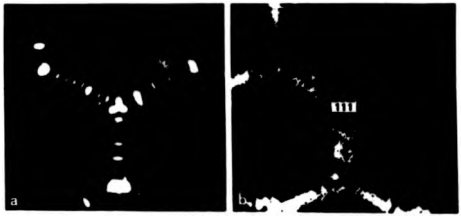


Figure 4.3: The image of the tip observed by field ion microscopy. (Binh 1988)

4.2.2 Techniques of fabrication from large scale to small

Many practical engineering ways have been developed, in which small/sharp ended tips are constructed from large pieces of material. For example, the traditional method of fabricating stylii is by grinding. A piece of diamond is fixed at the end of an arm. The other end of the arm is fixed on a rotational part of a holder. The axis of a grinding wheel is vertical (see figure 4.4). The point A is adjusted along the holder and, therefore, the angle θ matches the required stylus flank angle. The arm can be rotated along its own axis to produce the four flank faces of a stylus. The manufacturing process is inspected by an optical microscope, by turning the arm with A so that its fixed end can turned to the horizontal. Typical tips are shown in figure 4.5 (Jungles & Whitehouse 1970).

The modern techniques are field ion milling/field ion microscopy (Müller & Tsong 1969, David 1989) and electrochemical etching (Musselman *et al* 1990). Because the machining rate in ion beam milling is dependent upon the angle of incidence of the ions and the materials of the specimen, it is possible to predict the resultant profile of a machined stylus which is rotated in the beam at any preset angle. The charac-

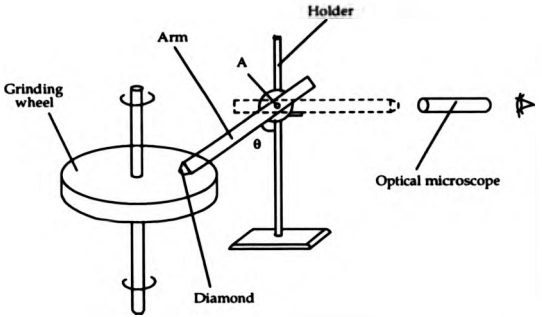


Figure 4.4: The traditional way of manufacturing stylii.

terization of a tip produced by a pseudo-stationary profile technique was studied by Binh (1988). By field ion processing, multiple tips occur easily if a piece of tungsten wire is not fine enough. It has been pointed out that the second tip, recessed by two atoms, carries only about a million times less current on STM (Binnig & Rohrer 1987). But this is only the case for atomically smooth and flat surfaces. The second tip may well join in to generate significant tunnelling current if the surface is not so smooth. Simple etching tends to produce tips of insufficient sharpness. Electrochemical etching can control the tip shape better, especially during the last few seconds of etching, if a control circuit is used to apply a momentary potential to the tip. Musselman *et al* (1990) achieves a Pt/Ir tip with high aspect ratio and a smooth, featureless surface by electrochemical etching. The radius of the tip is as small as 5 nm, ten times larger than the radius of an atom. So, the size of the tip is still very large relative to atomic features on the surface and will have significant

Figure 4.5: (a) Typical tip as used on the Talysurf instruments; (b) Talystep sharp stylus which is used for resolving very fine surface texture.

effects on measurement.

4.3 Review of the geometrical effect of the stylus tip

On the geometrical effect of the stylus, studies have shown that the roughness associated with the smallest and sharpest features is a negligible fraction of the total measured roughness. This is the theoretical basis for the experimental findings of a number of workers (Williamson 1947; Radhakrishnan 1970; Thomas & Salyes 1978), that stylii of the standard dimensions do not significantly misrepresent the average roughness of a surface (Figure 4.6).

However, as the surface geometry is considered in more detail or more locally, the average height parameters are more sensitive to tip size. Nakamura (1966) has discussed the limiting condition of the approximate linear response of roughness curves and an approximate formula of output curve of a cosine wave input. It is

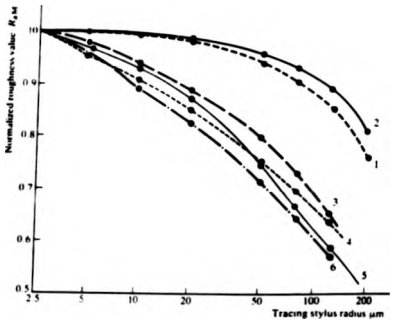


Figure 4.6: Effect of stylus tip radius on measured roughness (Radhakrishnan 1970). (1) Planed; (2) electro-eroded; (3) milled; (4) ground; (5) electrochemically sunk; (6) honed.

clear that this model is only an analytical approximation, with the criterion of R_a value. Surfaces are not cosine waves, nor is the geometrical effect linear.

Whitehouse (1974) has conducted an investigation from the point of view of statistical analysis, and derived the relationship between the true and measured value of the R_q parameter as a function of stylus tip radius and surface roughness. The same conclusion was reached, that the height parameters of surfaces are not varied by more than 2% by using the conventional stylus of size 2μm. McCool (1984) has developed a simulation model for assessing the magnitude of distortion caused by the finite stylus tip, as well as the effect of record length and sampling frequency by using a random process having a specified spectrum. The effects are in terms of the zeroth-, second- and fourth-moments of spectra. The results show that stylus tips vary the even moments of spectra over a large range. Generally, the stylus size increases the moments.

It is believed that the slope of a surface in the neighbourhood of the stylus tip has to be considered in order to assess the distortion caused by the stylus geometry. Sherrington & Smith (1986) carried out a quantitative assessment of the influence. If the surface slope is θ, then for a round tip having radius of R, the error, e, is

$$e = R(\sin\theta\tan\theta + \cos\theta - 1), \tag{4.1}$$

and for a rectangular tip, it is

$$e = \frac{L\tan\theta}{2}. \tag{4.2}$$

Both situations were discussed in the case of straight line surfaces. DeVries & Li (1985) have shown that a better reconstructed profile is obtained when both the slope and the curvature of a profile are considered. However, surprisingly, it is declared that the algorithm based on the simple slope geometry generally gives better compensation results than that based on both slope and curvature, in terms of autocorrelation length.

An algorithm for reconstructing surfaces from the 'apparent' surface seen by a scanning tunneling microscope has also been developed by Chicon *et al* (1987). The limitation of the algorithm is that it works only for continuous surfaces with their analytical functions defined. However, in practice this type of surfaces is rarely found. It was also concluded that an exact result can be given only with a spherical tip.

4.4 Effect of the stylus dimension

There are two considerations for the stylus geometry: (a) stylus size and (b) stylus shape. The stylus flank angle and the curvature are classified as (b). This study on stylus effects has been carried out by computer simulation. To start with, a surface height profile is generated randomly with exponential correlation (Whitehouse, 1983) with 3070 data points, relative to a length of 3 mm (refer to program `cordata1d.m` in Appendix H. The profile has an R_a value of 3.30 μm or R_q value of 4.17 μm. The arithmetical mean deviation of the slope, Δ_{a1} is 0.97, and the wavelength, λ, is 21.33 μm. It should be noticed that these roughness results are only artificial values generated by computer. A stylus with a specified shape tracks the surface by computer to form the trace (see figure 4.1).

Three types of stylus tips have been considered here, shown in figure 4.7. The

Figure 4.7: The three types of stylus discussed in the paper.

reason for this selection is that the triangular tip has a sharp end and is ideal. The most commonly used stylus, at present, has a round tip with a straight flank. It can be regarded as the combination of a circular and a triangular shape as a normal

tip (refer to program `round-tip.m`, `triangle-tip.m` & `normal-tip.m` in Appendix II).

4.4.1 The effect on surface measuring parameters

The surface profile is traversed by the tips of these three shapes with sizes 5 μm, 10 μm and 20 μm, respectively. The roughness parameters R_a and wavelength λ_a are listed in table 4.1.

size (μm)	Round Tip		Triangular Tip		Normal Tip	
	R_a (μm)	λ_a (μm)	R_a (μm)	λ_a (μm)	R_a (μm)	λ_a (μm)
0	3.30	21.33	3.30	21.33	3.30	21.33
5	3.31	33.67	3.15	24.79	3.36	34.22
10	3.51	49.12	3.09	25.22	3.33	36.63
20	3.95	80.09	3.06	25.08	3.31	38.23

Table 4.1: The surface roughness parameters of tips with different sizes and shapes.

Generally, surface roughness R_a does not vary so much with stylus tip size and shape, although for the round tip and for the normal tip the roughness increases slightly whereas for the triangular tip the roughness value is reduced. However, the wavelength value, after the stylus has traced out the profile, increases dramatically, sometimes by at least 16% to more than double.

In addition, the normal tip gives height roughness values closer to the original profile than the others do. It implies that the trace by this stylus is insensitive to the size. It confirms the effectiveness of stylii currently used for surface roughness measurement.

4.4.2 Filtering characteristics of the stylus

The effect of the stylus dimension on surface parameters shows the explicit relationship in the previous subsection. The filtering characteristic is now presented in this

subsection.

A stylus is often referred to as a 'mechanical low-pass filter'. It is unable to sense the irregularities of a surface whose dimensions are smaller than the stylus tip size. So, these high frequencies are filtered out.

However, this low-pass filter has a non-linear filtering characteristic. The conventional Fourier transform has been tried to obtain an approximation to the character of the filter. For example, Nakamura (1966) used a sum of a series of cosine waves with different frequencies to represent the trace of a stylus. For a cosine wave profile as an input $g(x) = a\cos\omega x$, the output trace is

$$e(x) = r + a\cos\omega x + \frac{a\nu}{4} - \frac{a\nu}{4}\cos 2\omega x + \frac{\delta^2}{2}(r + a\cos\omega x) - \delta x a\sin\omega x - \frac{1}{2}\delta\lambda^2 a\sin 2\omega x, \tag{4.3}$$

where r is the radius of a tip. The first four terms are due to the effect of the stylus tip size. The other terms are due to the effect of the tilting arm, $\nu = \alpha\lambda$ and $\alpha = \frac{2\pi R}{T}$. A simulated trace using the method mentioned before and Nakamura's trace calculation is shown in figure 4.8 with $r = 2.5$ μm. The calculation is based on the simulation which gives approximately the same values of surface roughness parameters like R_a and R_q. With this criterion, a trace can be regarded as a linear response to the profile when $\nu \geq 0.4$ with 10% tolerance. Clearly, Nakamura's formula for the trace distorts the local information severely. In the E-system, the filtering character of a stylus is artificially defined by using the covering area of a stylus travelling across a triangle as in figure 4.9 (Weingraber & Abou-Aly 1989). Two stylii which have different sizes cross the triangle and give two different traces 1 & 2. The ratio of the area covered by the stylus trace to the whole triangle area quantifies the filtering character.

Most of the previous work on the filter shows that this mechanical low-pass filter is not linear due to the geometrical combination of the stylus and the surface. This can been seen from the power spectrum functions (PSF) of three pairs of profiles

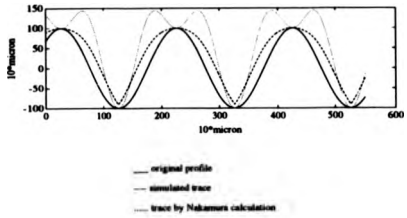


Figure 4.8: A comparison of a sine wave profile, its simulated trace and the trace calculated by Nakamura's formulas. (refer to programs `form-trace.m` & `nakamura.m` in Appendix H.

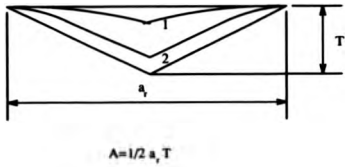


Figure 4.9: The triangle for determination of the stylus filtering character in the E-system.

and traces with the transfer functions shown in figure 4.10, 4.11 and 4.12. The three traces are formed from a round, triangular and normal tip, respectively. The transfer function is calculated from the ratio of the PSF of the trace to its profile.

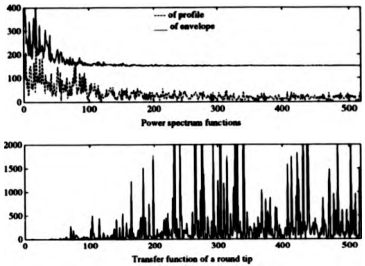


Figure 4.10: The PSF of a profile and the trace by a round tip and the transfer function. (In these figures, 'envelope' is the trace.)

For the convenience of comparison, the PSFs of the traces in each figure are lifted up. As can be seen, the trace from the triangular tip has the closest spectrum to that of the profile[1]. All the tips with rounded ends filter out the high frequencies.

Each of the three stylii shows its filtering characters in the frequency domain. Overall, they, to a certain extent, filter out the high frequencies, but modify the low frequencies in different ways. The filtering characteristics of a stylus tip shown above are merely a display of the detailed effects of the size and shape. In order to develop better ways to represent the filter character, tracking mechanism of the stylus needs to be known.

[1]Noticed that the vertical ranges of the transfer functions in figure 4.10, 4.11 & 4.12 are different.

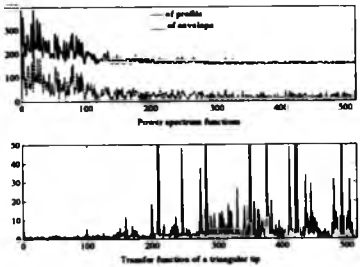


Figure 4.11: The PSF of a profile and the trace by a triangle tip and the transfer function.

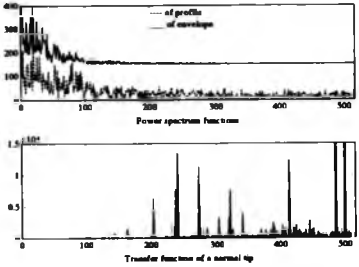


Figure 4.12: The PSF of a profile and the trace by a normal tip and the transfer function.

4.4.3 Effect of the stylus size

In cross section, most stylii have a circular shape at the tip end. Here, circular shaped stylii with different radii are used. The effect of the stylus size gives the low-pass filtering character, as shown in the previous section. On the other hand, the finite size stops the stylus reaching the bottoms of the valleys that have smaller sizes than it. So, the information revealed by the stylus is, to some extent, truncated at the waists of the valleys.

Firstly, let us view this effect on the height distribution of a surface. The Gaussian profile height distribution is shown in figure 4.13. The trace height distributions

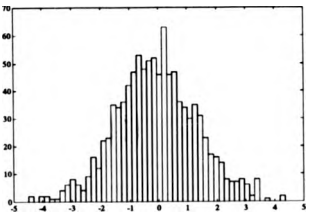


Figure 4.13: The height distribution of a profile.

with different size stylii can be seen in figure 4.14. From figure 4.14 (a) to (c) it is clear that as the stylus radius becomes larger, the height distribution skews towards the positive and the left end of the distribution range is increased. This means that more peaks and less valley points are picked up by the stylus. It follows that the trace is sensitive to peaks. However, as the stylus size become even larger, (c) to (d), the distribution tends to be symmetrical about a positive mean, and the skew reduces. This is because most height is contributed by the exaggerated peaks on the trace. This effect is caused by the stylus shape. The distributions of the trace

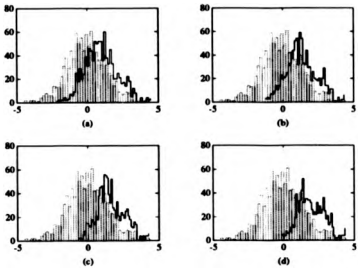


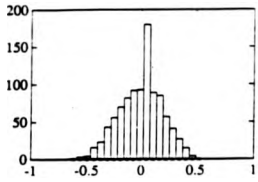
Figure 4.14: The height distributions of a trace by circular tips of different sizes of (a) 5 μm, (b) 10 μm, (c) 20 μm and (d) 40 μm. The background is the distribution of the profile.

slope and curvature by different sizes of circular tips are displayed in figure 4.16. The circular stylus tip does not change its slope so much but the distribution of the trace curvature.

From figure 4.14 to figure 4.16, it can be seen that the stylus size has a stronger effect on the curvature information, and less effect on the height and slope. Owing to the round tips used in this case, if the curvature radius of a valley part is smaller than the tip radius, the valley will not be tracked by the tip. Therefore, the surface curvature information is the critical factor in a discussion of the filtering effect in this case.

After eliminating of the effect of the stylus shape (that will be analyzed in detail in the next chapter), the curvature distribution of this modified trace is presented in figure 4.17. This is most that can be done in the way of correcting the trace.

In this diagram, the original profile is produced as a random series with an exponential correlation between the continuous three points. Then, the series is interpolated twice with the cubic spline function. The resultant profile, with a



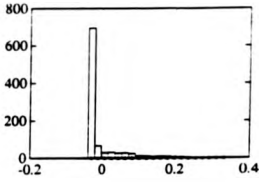


Figure 4.15: The slope distribution of profile and its trace by circular tip of 10 μm. (Left)

Figure 4.16: The curvature distribution of profile and its trace by circular tip of 10 μm. (Right)

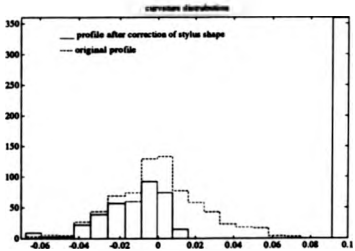


Figure 4.17: The comparison of the curvature distributions of the profile after correction of the stylus shape with the original profile.

total number of 737 ordinates, is referred to in figure 5.9 in the next chapter. The background in figure 4.17 is the curvature distribution. The curvature distribution of the corrected profile is calculated only for the sections tracked by the stylus. The curvatures of the rest of the ordinates is set to an artificial value outside the curvature range. In this case, it is 0.1. The stylus used is round with a radius of 50 ordinates relative to the profile unit point. The curvature histogram of the corrected profile, therefore, presents the cut-off at the value of 0.02 (= 1/50). Notice that a valley gives a positive curvature value and vice versa for a peak.

To sum up, a round stylus gives the upper cut-off at its radius on the curvature distribution, i.e. the second order differentiation. This may imply that a triangular or a flat stylus has its cut-off on the slope or the height distribution to the profile. However, these two types of stylii do not exist in practice. They are not discussed in detail here. The triangular tip influencing the slope will be shown in the next subsection.

4.4.4 Effect of the stylus shape

The effect of the shape of the stylus is caused by the contact with the surface at its flank. The stylii used here were mentioned in the previous subsection. Let us see how these stylii effect the profile in the distributions of the height, slope and curvature. The calculations of the slope and curvature are shown in programs `slope5.m` & `curvt5.m` in Appendix H.

Figure 4.18[2] shows the height distribution of traces produced by the three different tips. The triangular tip gives more or less the same height distribution (figure 4.18 (b)) to the profile (figure 4.18(a)). Both the circular tip and the normal tip give positive skewness owing to the exaggerated peaks (figure 4.18(c), (d)).

The tip shape effects are strongly shown in the slope and curvature distributions, as in figure 4.19 and figure 4.20. The round tip and the normal tip, which is with

[2]Noticed that the traces of (b) & (c) in this figure is other-way-round in the two figures later.

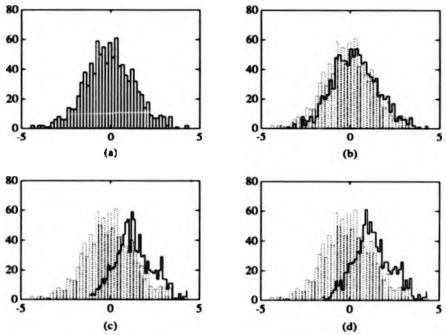


Figure 4.18: The height distributions of (a) a profile, and (b) of its traces by a triangular tip, (c) a circular tip and (d) a normal tip.

a rounded end, cause most of the peaks on the profile to change to the tip shapes, so that curvature distributions give high peaks over 500 data points on (b) and (d). The distribution is truncated near to zero for any values less than it. In fact, this value is 0.1 (= 1/10 unit). This means that these two tips are unable to reach the bottom of the valley on the profile. The triangular tip has a severe effect on

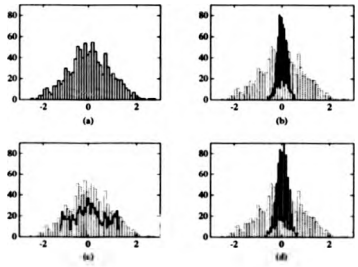


Figure 4.19: The slope distributions of (a) the profile, and of its traces by (b) a circular tip, (c) a triangular tip and (d) a normal tip.

the slope distribution. This may be due to the constant slope angle of the tip (see figure 4.19 (c)). However, the slopes of the traces by the round and normal tips are distributed in a complex manner. The 'comb' shapes of the slope distribution is hardly explained, and might be the effect of the discretisation error.

In summary, the shape of the stylus causes the cut-off effect in the distribution according to its shape.

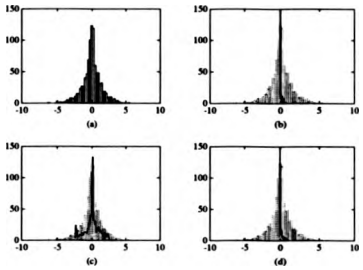


Figure 4.20: The curvature distributions of (a) the profile, and of its traces by (b) a circular tip, (c) a triangular tip and (d) a normal tip.

4.5 Conclusion

The effect of the dimension of the stylus cannot be analyzed through the frequency information. A further division of the dimension into size and shape has to be made. The causes of the effect become more evident. The stylus size causes the truncating value on the appropriate distribution based on the stylus shape. The loss of the information cannot be recovered. However, the effect of the stylus shape is presented as the exaggerated peaks and the narrowed valleys. It is believed that this effect should be correctable. A detailed analysis will be presented in the next chapter.

Chapter 5

Geometrical Analysis of the Formation of A Stylus Trace

There are two aspects of the geometrical limitation of the stylus. One is that the valley information cannot be revealed when its curvature radius is smaller than the stylus size. The other is the modification to the original profile geometry by the stylus shape. This chapter studies the formation and the nature of the stylus trace.

5.1 Introduction

Stylii are employed widely in the field of geometrical measurement in engineering. For example, a co-ordinate measuring machine (CMM) is used to obtain the dimensions of a workpiece, a roundness measuring instrument, e.g. the Talyrond, is used to measure the circularity of a cylindrical shape, and a surface measuring instrument, e.g. the Talysurf, is used to assess surface roughness. The stylus instrument, when used to measure the geometrical shape of a component, has a high reliability relative to other measuring methods such as optical methods. This may be because contact is one of the commonest human senses, and the most relevant to tribology performances. In conventional engineering, component sizes are, generally, larger than 0.01 mm. The sharpness, or size of a stylus (say, 10 μm) is sufficiently accurate for measuring them. However, as the size of the component decreases, the size of the stylus has a significant influence on the measured results. This happens particularly in surface finish measurement.

The surface finish or surface texture is probably the smallest geometry in general dimensional measurement. Its range is, normally, from a few mms to a few microns, or even nanometers. The trace of a stylus is regarded as the geometry of a surface. For this measurement, the stylus is not smaller than 2 μm in radius. At the smaller end of the range, therefore, the similar sizes of the stylus and the measurand undoubtedly results in a severe bias in the measurement caused by the stylus dimension.

The relationship between the profile, the stylus and the trace is analyzed in this

chapter. Computer simulations of the tracing process are presented. The terms 'profile' and 'trace' used here, mean the true surface shape and the shape resulting from the measurement of the profile by the stylus, respectively.

5.2 The formation of a stylus trace

The finite dimension of the stylus tip implies that the resultant trace will not follow exactly the profile of the real surface: it may not enter steep valleys, and it may record sharp peaks as blunt cusps, as shown in figure 5.1. Due to the contact of the

Figure 5.1: Diagram showing a surface profile and its trace by a finite stylus.

stylus and the surface, their respective geometries are the most important factors effecting the formation of the trace. The stylus trace can, therefore, be regarded as the 'convolution' of the stylus and the surface geometry, i.e.

$$\boxed{\text{Trace}} = \boxed{\text{Surface}} \otimes \boxed{\text{Stylus}}.$$

This 'convolution' is not the same as the general mathematical convolution. The word used here is meant in the sense of the process of combining two geometries.

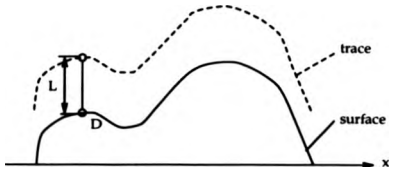
Clearly, it is not just a simple summation. It should be noted that the analysis in this chapter is only for the two dimensional case.

5.2.1 Analysis of stylus trace formation

In the analysis below, it is assumed that both the shapes of the stylus and the surface profile are smooth and continuous. The stylus can be in contact at any point on the surface. Both materials are rigid. No elastic or plastic deformation exist.

(i) The trace of a mathematical dot and a line.

To begin with, imagine that a dot D which has no size, or a finite line L which is perpendicular to the x-axis, travels across a given surface profile. The traces of the dot D and the top end of the line L do not differ in shape from the surface. The trace is only lifted upwards a distance L by the line, shown in figure 5.2. Of course, it would be perfect if an engineering stylus were like such a dot or line.



(a)

Figure 5.2: The traces by a dot and a line.

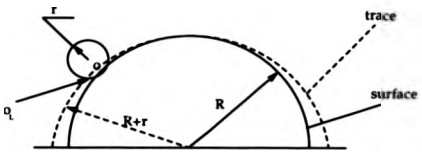
(ii) The trace of a round stylus with radius r.

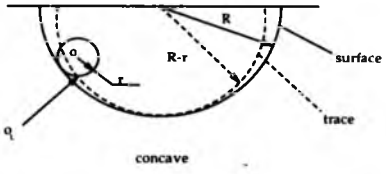
Consider that the dot D is now expanded into a circular disk with radius r, and used to measure concave and convex semicircular profiles of radius R (Figure 5.3). The resultant trace is undoubtedly a circle which has the same centre as the profile, but with radius $R+r$. The trace of the lowest end, o_L, of the stylus (without rolling) is a trace with the same radius and the centre shifted down a distance r. Therefore, the trace of the stylus (defined by the trace of o_L) is not the same as the profile itself. The geometrical information of the stylus, such as r, is added or 'convoluted' into the trace (see figure 5.3). A similar situation occurs on a concave profile, but the resultant trace has a radius of $R-r$.

(iii) The trace of the same stylus over a smooth profile

To see in more detail how the trace is formed, a smooth profile such as a sine curve is used instead of the circular profile. As shown in figure 5.4, $A(x,y)$ is the contact point, and it is always on the profile. $B(\alpha,\beta)$ is the centre of the stylus. Because the tangent line is shared by both the stylus and the profile at the contact point, the relationship between A and B, corresponding to the stylus r is

$$\begin{array}{ll} y' \geq 0 & \begin{cases} \alpha = x - r\cdot\cos\nu \\ \beta = y + r\cdot\sin\nu \end{cases} \quad \nu = 90^\circ - \theta \\ y' < 0 & \begin{cases} \alpha = x + r\cdot\cos\nu \\ \beta = y + r\cdot\sin\nu \end{cases} \quad \nu = \theta - 90^\circ, \end{array} \tag{5.1}$$





(b)

Figure 5.3: The trace of a round stylus over a convex and a concave circular profile.

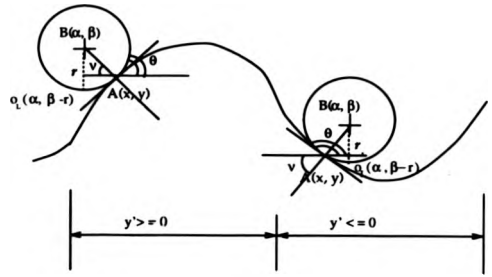


Figure 5.4: Analytical diagram of trace formation.

where $0 \leq \nu \leq \frac{\pi}{2}$, $y' = \tan\theta$ and θ is .he surface profile slope angle. Replacing ν by θ, equation 5.1 becomes

$$\begin{cases} y' \geq 0 & \begin{cases} \alpha = x - r \cdot \sin\theta \\ \beta = y + r \cdot \cos\theta \end{cases} \\ y' < 0 & \begin{cases} \alpha = x - r \cdot \sin\theta \\ \beta = y + r \cdot \cos\theta. \end{cases} \end{cases} \tag{5.2}$$

Therefore, for all cases of y', $B(\alpha, \beta)$ can be given as

$$\begin{cases} \alpha = x - r \cdot \sin\theta \\ \beta = y + r \cdot \cos\theta, \end{cases} \tag{5.3}$$

where $0 \leq \theta \leq \pi$ and $r \geq 0$. Equation 5.3 implies that the trace, $B(\alpha, \beta)$, is formed by the surface profile, $A(x, y)$, the contact point, and the stylus tip r. The second part on the right side of the equation contains the stylus size r and the slope angle.

Because the tangent line is shared, this angle is the slope angle of both the profile and the stylus at this contact point. Thus, θ can be regarded as the bridge between the stylus and the profile. Equation 5.3 illustrates the 'convolution' by the profile and the stylus to form the trace. If the angle θ is replaced by the profile slope, y', equation 5.3 becomes

$$\begin{cases} \alpha = x - r \cdot \frac{y'}{\sqrt{1+y'^2}} \\ \beta = y + r \cdot \frac{1}{\sqrt{1+y'^2}}. \end{cases} \tag{5.4}$$

It can be seen that the variation of $B(\alpha, \beta)$ depends upon y', as shown in figure 5.4.

(iv) The trace of a smooth stylus over the same smooth profile

The relation (eqn. 5.1) is also valid for a smooth and non-circular stylus. However, in this case, both r and the centre of curvature vary with position on the stylus. The stylus trace results from the displacement of a fixed point on the stylus, e.g. o_L. In order to fully construct the relationship between the trace, the profile and the stylus shape, the relation of the tip end o_L with each curvature centre of the stylus must be known. This makes the problem much more complicated and impractical.

5.2.2 Discussion of the relations

Some of the relationships in this approach need to be discussed. Referring to equation 5.4, let

$$g(y') = \frac{y'}{\sqrt{1+y'^2}},$$

taking the differential of equation 5.3, the first equation becomes

$$\frac{d\alpha}{dx} = 1 - |r| \cdot \frac{dg(y')}{dx}. \tag{5.5}$$

Now

$$\frac{dg(y')}{dx} = \frac{dg(y')}{dy'} \cdot \frac{dy'}{dx} = \frac{dg(y')}{dy'} \cdot y'', \tag{5.6}$$

and

$$\frac{dg(y')}{dy'} = \frac{1}{(1+y'^2)^{\frac{1}{2}}}. \tag{5.7}$$

Hence,

$$\frac{d\alpha}{dx} = 1 - |r| \cdot \frac{y''}{(1+y'^2)^{\frac{3}{2}}}. \tag{5.8}$$

The curvature of the surface, K, is defined as

$$K = \left| \frac{y''}{(1+y'^2)^{\frac{3}{2}}} \right| = \frac{1}{R}. \tag{5.9}$$

Therefore,

$$(\frac{d\alpha}{dx} - 1)^2 = \frac{r^2}{R^2}. \tag{5.10}$$

Similarly, the second equation in 5.3 will be

$$(\frac{d\beta}{dy} - 1)^2 = \frac{r^2}{R^2}. \tag{5.11}$$

Equations 5.10 and 5.11 imply that:

1. The resultant trace of a stylus will have a linear relationship with the original surface profile when both of them have a constant curvature radii. This means that only the trace of a round stylus over a circular profile, such as a ball and a cylinder, or a flat surface, gives the correct form of the surface profile.

2. Both of the differentials in the x- and the y-axis are equal, and they are only related to the curvature ratio of the stylus and the profile, i.e.

$$\frac{d\alpha}{dx} = \frac{d\beta}{dy}, \tag{5.12}$$

or

$$\frac{d\beta}{d\alpha} = \frac{dy}{dx}.$$

This means that the slopes on the trace and the profile are the same when the stylus is a round shape.

5.3 Geometrical analysis of the trace formation

Another approach to this problem is the use of geometry. The assumptions in the previous section still stand, i.e. the shapes of the stylus and the surface profile are geometrically smooth and continuous. The relationship between the stylus, the profile and the trace is displayed in figure 5.5.

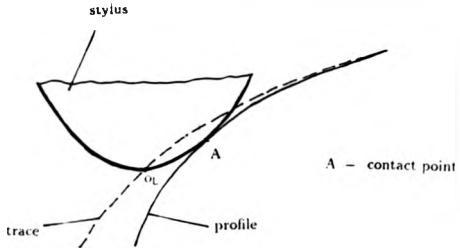


Figure 5.5: The relationship between the profile, the stylus and the trace.

The stylus is moved a sufficiently small section along the profile such that the movement can be considered to be a straight line. The contact point A moves to B, as shown in figure 5.6. This is equivalent to taking the first order of the Lagrange expansion. The previous contact point A on the stylus is now located at A_B. Similarly, the contact point B on the stylus is at B_A when the contact point is at A. o_L^A and o_L^B give the trace points when the stylus moves from A to B. Because the stylus does not roll, i.e. no rotational movement occurs, the points o_L^A, A and B_A on the stylus, at the contact position A, move equal displacements to o_L^B, A_B

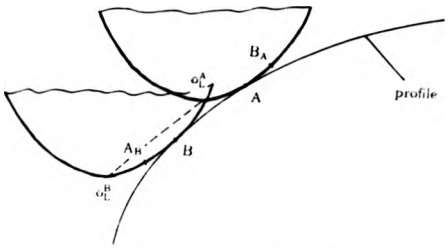


Figure 5.6: Differential analysis of the formation.

and B at the contact position B. That is,

$$\overset{\frown}{o_L^A o_L^B} = \overset{\frown}{AA_B} = \overset{\frown}{B_A B}, \tag{5.13}$$

and they are parallel to each other. Also, it is obvious that

$$\overset{\frown}{o_L^A o_L^B} = \overset{\frown}{AB} + \overset{\frown}{B_A A}, \tag{5.14}$$

or

$$\overset{\frown}{o_L^A o_L^B} = \overset{\frown}{AB} + \overset{\frown}{BA_B}. \tag{5.15}$$

Therefore, the points B_A, A, B, and A_B are located on the same line $\overset{\frown}{B_A A_B}$. In other words, the slope of $\overset{\frown}{AB}$ on the profile, is equal to the slope of $\overset{\frown}{B_A A}$ or $\overset{\frown}{BA_B}$ on the stylus flank and the slope of $\overset{\frown}{o_L^A o_L^B}$ on the trace. Hence, with these relations, it can be stated that:

1. The slope of a point on the trace is equal to the slope of the appropriate contact point on the profile and on the stylus.

2. The length the stylus moves is equal to the sum of the distances on the stylus and on the profile that the contact points moves.

Statement 1 confirms equation 5.12 geometrically.

5.4 Computer simulation of a stylus trace formation

The two-dimensional trace is obtained by moving a circle with a specified radius over a two-dimensional profile. This trace is controlled by the peaks and by their flanks, by the crest spacing, and by the radius of the circle (see above section). Hence, the procedure for computing this trace, for various conditions, should take into consideration each and every point of the profile.

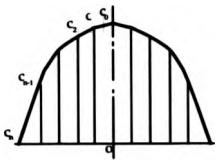


Figure 5.7: Diagram showing digitized stylus tip.

A stylus shape is digitized into a series of data points, for example, $c_0, \ldots, c_n$ for a symmetrical stylus in figure 5.7. The profile is also a series of ordinates. Both of them have an equal spacing between the ordinates. For a circular tip, n should be not less than 4 so that the second order differentiation is continuous. Additionally, the sampling space is determined according to the correlation coefficient of two adjacent ordinates on a profile (referring to the introduction section in chapter 3).

When the stylus is travelling over the profile, the point of contact with the profile is changing and moving between c_0, where the slope is zero, and c_i $(i = 1, \cdots, n)$, at the flank. The tracing process is obtained by determining the trace heights at each profile ordinate. The stylus ordinates are placed with the centre c_0 over the ith profile ordinate in figure 5.8 (Shunugam & Radhakrishnan 1974). The contact

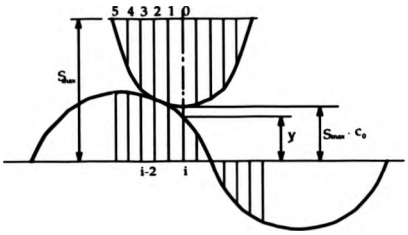


Figure 5.8: Computation of a stylus trace formation.

point, i.e. the $(i-2)$th in this case, gives the maximum of the height summations of the corresponding profile ordinates covered by the $(2n+1)$ stylus ordinates. That is,

$$S'_{max} = \max_{i}(e_j + y_j) \quad j = -n, \ldots, 0, \ldots, n. \tag{5.16}$$

Then this maximum, S'_{max}, is transferred to the position i. The trace ordinate at i, i.e. E_i, is given by

$$E_i = S'_{max} - c_0. \tag{5.17}$$

Then, the stylus tip c_0 moves to the next profile ordinate. The above procedure is repeated until all the ordinates on the profile are taken. Hence, each trace ordinate

height is obtained, and the trace is formed over this profile. That is,

$$\max\left[\begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n+1} \end{pmatrix}\right] = E_{n+1} - c_0$$

$$\max\left[\begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{pmatrix} + \begin{pmatrix} y_2 \\ y_3 \\ \vdots \\ y_{n+2} \end{pmatrix}\right] = E_{n+2} - c_0 \qquad (5.18)$$

$$\vdots$$

$$\max\left[\begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{pmatrix} + \begin{pmatrix} y_{N-n-1} \\ y_{N-n-2} \\ \vdots \\ y_N \end{pmatrix}\right] = E_{N-n-1} - c_0$$

It should be noticed that if the profile has N, i.e. $y_1, \ldots, y_N$ ordinate points, the trace is formed only by the $(n+1)$th to the $(N-n-1)$th points.

The computing simulation was performed using the MATLAB software package. The general programmes for the computation of the trace are given in program `form-trace.m` in Appendix H.

5.5 Discussion

5.5.1 Determining the profile from the stylus and the trace

The profile can be worked out using the statements mentioned in section 4.3, when the trace and the stylus geometry are known. The slope of a series of ordinates can

be calculated using the five-point numerical differentiation formulae, as

$$slp = \frac{1}{12h^2}(-y_{i-2} + 16y_{i-1} - 30y_i + 16y_{i+1} - y_{i-2}), \tag{5.19}$$

where h is the sampling spacing (see program `slope5.m` in Appendix H. If the slope of one point on the trace is known, the point on the stylus which has the same slope can be found. The distance between this stylus point and the tip end gives the relative position of the contact point to the trace point in this situation (referring to figure 5.5 & 5.6). Therefore, all the true profile ordinates can be found by searching for contact points. This is a very important issue.

The computer simulation for this correction is shown in program `true-profile.m` in Appendix H. The result is now displayed in figure 5.9. The sections on the trace missed by the stylus are set to an artificial value. The comparison of the height distributions is also exhibited. The agreement of the corrected trace with its original profile confirms that the 'deconvolution' process is correct. It is necessary to point out that the true profile cannot be revealed if the stylus cannot reach the bottoms of the valleys on the profile. Therefore, the resultant trace become sharp corners of the first order discontinuity.

In practice, this relationship can be used to modify measurements such as form and roundness. In surface measurement, this modification can only be performed for all the sections of the profile that the stylus has touched.

5.5.2 The stylus geometry from the inside and the outside traces of a profile.

Another issue is to find the stylus geometry from the two known traces. These two traces must be the outside and inside traces from the same profile. The inside trace of the profile is obtained by the stylus as it travels underneath the profile as shown in figure 5.10. The contact distance is defined (see figure 5.11) when the stylus

___ profile corrected from its trace
--- original profile

Figure 5.9: Diagram showing the corrected trace of the original profile with their height distributions.

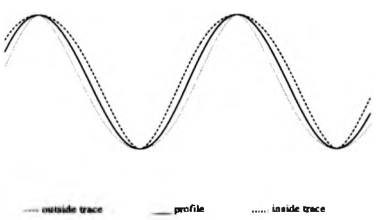


Figure 5.10: The traces formed as the stylus travels over the profile from the outside and beneath.

contacts the same point on the profile from the inside and the outside. It is the distance between the two positions of the tip end. The contact distance line does not lie in either the x- or the y-axis, and goes through the contact point and the two stylus ends. According to the statements in the previous section, the trace slopes of the two tip positions from the inside and outside should be the same. They should also be equal to the slope of the according profile point, i.e. the contact point. The different slopes imply that the stylus and the profile contact at different points. It can be seen from figure 5.11 that the profile point lies in the middle of the contact distance. Therefore, the profile can be found from the inside and outside traces. In addition, the half contact distance also gives the relation of the stylus point in this contact position with the stylus end. All the stylus flank points can be found from the different contact situations, so, the stylus tip shape can be known. The process simulated by computer shows that the stylus tip shape generated by the algorithm

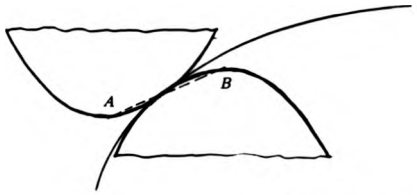


Figure 5.11: The definition and relationship between contact distance.

has a good agreement with the given stylus. A method based on the relation for measuring the stylus geometry, therefore, can be proposed.

5.6 Applications

5.6.1 A method for measuring the stylus dimension based on the trace formation

Obviously, the stylus cannot travel inside the surface. In practice, this can be done by finding a mirror matching pair of surfaces - the 'positive' and 'negative' pair. The advantage of this method is that the true surface shape itself is not necessarily known when a pair of 'positive' and 'negative' surfaces are given. Moreover, the way the principle works is the same as the way the stylus tip performs. So, the stylus shape can be checked without removing it from its instrument, especially the stylus type of surface measuring instrument. Therefore, checking the stylus shape is simple and easy.

This method also works for unsymmetric stylus tip shape. In this case, there are

two more traces required, which are obtained after turning the stylus 180° on its own vertical axis. The outside trace without turning and the inside trace with turning form a pair and generate the algorithm above, and so do the other two traces.

It should be noted that the stylus tip shape worked out by this method is only that part which is involved during contact when the trace is formed, and it is normally more important than the rest of the tip.

In practice, however, there are some problems that must be solved in order to apply this method. The basic considerations are categorized in the three following points:

(i) The checking surfaces

The pair of 'positive' and 'negative' surfaces is the key to this method. The surfaces should have a smooth change of form and relatively large wavelength, so that the stylus tip can keep contact at every point of the surfaces. One of the best possibilities to obtain a 'negative' surface is by using a replica. A replica gives a good matching profile. But it normally does not have the same rigidity as the 'positive' surface. If the hardness of a replica is increased, it may be difficult to take it off from the 'positive' surface.

Another way is to use a sinusoidal surface with a certain amplitude and a certain wavelength. In this case, the 'negative' surface is not necessary. The half wavelength shift gives the 'negative' surface. The ratio of the maximum of the surface curvature radius to the maximum of the stylus curvature radius should be kept less than 100 : 1 so that the stylus traces contain enough stylus shape information. But the difficulty is to obtain a very good sinusoidal surface. Lapping can produce sinusoidal surfaces. Unfortunately, the finish is too rough to use in this method.

(ii) Turning mechanics

To check for an unsymmetric tip shape requires turning either the tip or the checking surface. It is very important to keep the same surface trace after turning. The turning accuracy should be as good as 0.5μm.

(iii) Data collection

In most engineering measurement, a stylus normally has a size of greater than 1.5 μm. In order to find the stylus shape, there are at least 4 or 5 data points needed on the stylus tip. Therefore, the sampling space should be kept at 0.5 μm or lower.

5.6.2 The use of the 'deconvolution' in a micro-hardness test

The geometrical relationship between a profile, a stylus and a trace was analyzed. The profile can be found from a consideration of the effects of the stylus geometry modified on the trace. This relationship can be used to study micro-hardness. The tests of micro-hardness are conducted using a stylus. The stylus has two functions. It is used not only as a hammer, but also to profile the depth or the shape of the indentation after hammering. Figure 5.12 illustrates the bias due to the shape of stylus. Using the relationship analyzed in chapter 4, the geometry of the indentation

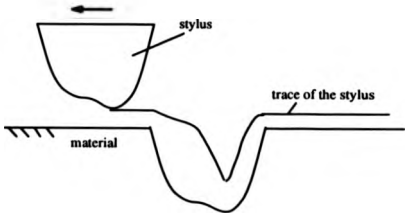


Figure 5.12: The bias by the stylus on the depth of the intentation in micro-hardness test. (Smith 1990)

can be found.

5.7 Conclusion

The relationship between a stylus and the measured engineering surfaces is purely geometrical. The trace is the profile modified by the geometry of the stylus. This is sometimes called the 'integration' or the geometrical 'convolution'. However, the 'differentiation' or the 'deconvolution' has been left out since the surface parameters are generally statistically averaged. In this chapter, the relationship between the stylus, the profile and the resultant trace are analyzed from the geometrical point of view. The central point of this relationship is the share of the same slopes in the stylus and the profile at the contact point. According to this, the detail in the tracing process is exposed. The computer simulation of the 'deconvolution' process confirms that the mechanism is correct. Based on the principle, some applications are intended. A method for checking a stylus dimension is proposed. Its advantages are that the checking operation is relative simple and does not require removing the stylus from the instrument. The practical difficulty in this method is the fabrication of the special checking surface pair. In general, this method is feasible. Another application is in the micro-hardness test.

Chapter 6

Measurement of the Stylus Tip Dimension Using A Knife-edge

6.1 Introduction

The stylus type of instrument suffers from the effect of the tip acting as a low-pass filter and consequently the distortion of the measured results by the stylus, as discussed in the previous chapter. The tip effect is significant since the surface measurement is at the nanometer level. However, if the stylus geometry is known, the true surface profile can be found by 'deconvoluting' the trace, as pointed out in chapter 5. First, however, the stylus dimension must be measured. In chapter 5, a method is proposed by means of a mirror pair of surfaces. But the practical difficulties demand more time and effort. In this chapter, a knife-edge method for measuring the tip dimension is designed. The rig is experimentally tested.

The knife-edge can be regarded as a surface that has a special shape. When it travels underneath a stylus, the knife has a much sharper edge than the stylus, so that the contact point on the knife can be thought of as unchanged. Therefore, the resultant movement represents the geometry of the stylus. A knife-edge is much easier to make than a special pair of surfaces.

6.2 Importance of checking stylii

From the point of view of metrology, the instrument, in particular its stylus, has to be calibrated regularly because damage to the stylus varies the measured results. There is a certain amount of wear on the stylus tip, even with the hardest material (diamond) owing to the stylus force. This is because the area of contact between the stylus and the asperities is very small, and it can eventually cause a change in the shape of the stylus tip. Moreover, the impact between the stylus and a surface can occur occasionally by incorrect operation of the instrument, causing damage to the tip. This can be seen by examining the stylus tip by SEM (Hilmann, 1984) in figure 6.1. Elewa & Koura (1986) have shown that the results obtained by measuring with a worn stylus falsely indicated smoother surfaces. The bearing area

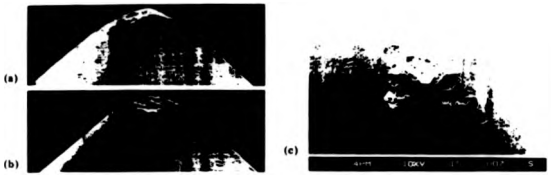


Figure 6.1: Scanning electron micrographs of a new stylus tip: (a) side view; (b) side view at 90° to that in (a); (c) top plan view. The tip radius is found to be about 4 μm.

of a surface is affected significantly by 25% - 100%, a greater percentage being found for smoother surfaces. Thus, the authors claimed the importance of establishing a method for checking the tip radius.

In addition, the calibration of surface roughness instruments by the use of precision reference specimens depends on the stylus radius (American National Standards ASNI B46.1-1962). It is therefore important that the experimental methods and algorithms for determining stylus radii be standardized (Vorburger, *et al* 1979).

Finally, manufacture of stylii for these instruments has to be monitored and the quality has to be controlled.

6.3 Review of stylus checking methods

Stylus checking is usually performed using a shadow projector or a visual microscope. In both cases high resolving power is required. With projectors, even the best quality instruments at high magnification are barely adequate, while with microscopes a

magnification of at least 100 × and a high numerical aperture is required.

An alternative method for checking the stylus is to use a special fine grating, a 'stylus check' (see figure 6.2), which gives a specified value of a roughness parameters

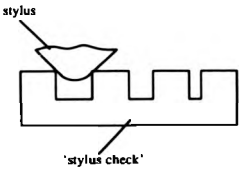


Figure 6.2: A grating for checking stylus - 'stylus check'.

only when the stylus is not too badly worn. When the stylus is worn it does not penetrate the structure of the grating fully and a low reading results. This method has been advocated for many years, but as yet is not widely used due to a lack of availability.

Scanning electron microscopes have been used intensively for more accurate examination of stylii. In SEM, the emission of secondary electrons from the specimen surfaces is collected. The amplified current is used to modulate the brightness of the cathode ray tube (CRT), and consequently an image of the surface is built up on the screen. The image magnification is determined solely by the ratio of the sizes of the rasters on the CRT screen and the specimen surface. Magnification of > 10,000 can be obtained. Because of the emission of electrons, the specimen surfaces have to be conductive. For insulators like diamond, a material generally used for stylii, the surfaces need normally to be coated with conductive layers. The complicated preparation results in a major drawback for examining stylii by SEM. In addition, stylii have to be removed from their instruments. It is therefore a time consuming method. Moreover, if a stylus arm cannot be fitted into the microscope, replication

techniques must be used.

Current measuring techniques are restricted in their use because of a general lack of portability, high capital expenditure and inability to measure the tip profile *in-situ*. The other disadvantage is that these optical systems only give qualitative information which is not suitable for numerical analysis. The 'stylus check' does give a numerical value in terms of the width and the depth, that is an indication of wear, however, no information relating to the shape of the tip can be obtained from this method.

Two measurement systems have been developed by the present author to overcome these problems and to obtain rapid shape measurement. They can be readily mounted on any stylus instrument, whether to check wear or to measure diamond stylii at the manufacturing stage. In one method, the stylus is moved across a knife-edge such as a razor blade (knife-edge method). Another method is based on the function of engineering surfaces, called the 'checking-surface method'. The former method is now in use in industry.

6.4 The knife-edge method

Using a knife-edge to check a stylus shape or size is not a new technique. The critical element in this technique is the traverse unit. It has to provide a stable, linear and controlled movement. In this case, the whole system consists of a knife-edge that is traversed in a linear path beneath the stylus under examination. It results in a motion that can be directly monitored by using the stylus instrument on which the device is mounted. That is, the height information is obtained from the stylus instrument. The horizontal movement of the knife-edge is by a controlled linear system. Combining the output from the stylus with the knife-edge position, a detailed profile of the tip can be obtained (see figure 6.3).

In this method, the knife-edge should profile stylii with tip radii between 1μm

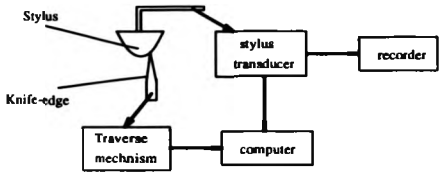


Figure 6.3: The system for stylus tip measurement.

and 50μm to an accuracy of at least less than 5% of the movement. The linear movement is achieved by a low friction high stiffness mechanism which includes a spring system and an electro-magnetic generating system. The spring system is designed from monolithic elastic hinge mechanisms. The system was originated by Chetwynd (1988) with detail design and later development by the author. The design rules of combination of a monolithic block incorporating flexure elements with coil/magnet actuator have been discussed by Smith *et al* (1987).

6.4.1 The theoretical basis of the design

Theoretically, in the monolithic system, the displacement has a relationship with the applied force F and the spring stiffness (λ) given by

$$d = \frac{F}{\lambda}. \tag{6.1}$$

Coil/magnetic drive

For a magnet of moment (M) in a field of strength (H), the force (F) is given by

$$F = M\frac{dH}{dx}, \tag{6.2}$$

where x is the distance from the centroid of coil. The field strength along the axis of a uniformly distributed circular cylindrical coil is given by

$$H(\frac{x}{a_1}) = H(0)\left[\frac{F(\alpha,\beta+\frac{x}{a_1})+F(\alpha,\beta-\frac{x}{a_1})}{2F(\alpha,\beta)}\right], \tag{6.3}$$

where $F(\alpha,\beta)$ is the well known Fabry factor given by

$$F(\alpha,\beta) = \beta \ln\left(\frac{\alpha+(\alpha^2+\beta^2)^{\frac{1}{2}}}{1+(1+\beta^2)^{\frac{1}{2}}}\right). \tag{6.4}$$

The field at the centroid of the coil ($H(0)$) can be expressed by the equation

$$H(0) = \frac{NI\chi F(\alpha,\beta)}{2a_1\beta(\alpha-1)}, \tag{6.5}$$

where a_1 is the inside radius of coil (m), a_2 is the outside radius of coil (m), $\alpha = \frac{a_2}{a_1}$, $\beta = \frac{l}{2a_1}$, l is the length of coil (m), N is the number of turns on coil, χ is the space factor, and I is the current through coil (A). The detailed relationship between the magnetic field strength H and the position x has been analyzed by Montgometry (1969). Substituting equations 6.3 and 6.5 into 6.2, it is possible to compute the force on a magnet of moment (M) measured from the centroid of a coil along its axis. The detailed characteristics of such systems have been discussed by Smith & Chetwynd (1990). The power that is put into the coil is also an important parameter which influences the environment because of heat from the coil. It depends on the current and the coil resistance determined by the dimension of the coil and the material of the winding wire as

$$R = \frac{N^2\pi\rho(\alpha+1)}{2\chi a_1\beta(\alpha-1)} \tag{6.6}$$

where ρ is the resistivity of winding material.

The spring system

For the spring system, the stiffness (λ) is

$$\lambda = \frac{EI}{KrL^3}(\mathrm{Nm}^{-1}) \tag{6.7}$$

where E is the modulus of elasticity, $I = \frac{1}{12}ht^3$, is the second moment of area of the cross-section, L is the length of the spring legs (m), r is the radius of the holes for the web (m), h is the depth of the web (m), and t is the thickness of the web (m). The constant K is given approximately by

$$K \simeq 0.589\sqrt{\frac{t}{R}}. \tag{6.8}$$

Thus, the displacement (d) is the function of the geometries of the spring and coil/magnet system, and is linearly related to the current through the coil and its number of turns.

The accuracy of the displacement is influenced by the operating conditions, mainly by the temperature change. Moreover, the deflection resulting from the stylus force also influences the linearity of the platform motion.

6.4.2 The design

For knife-edge application of this mechanism, both coarse and fine movements are needed. A simple compound rectilinear spring which comprises platform A and platform B as shown in figure 6.4, is used for motion of up to 100 μm. This is settable (open-loop) to 25 nm under reasonable operating conditions. It has a linear rate of traverse of 1 μm movement per second with an accuracy better than 5%. A single rectilinear spring which can be coarse -positioned via a micrometer is used for 1 mm rough adjustment. The platform A upon which the knife-edge is mounted is driven by a coil/magnet force actuator (see figure 6.4) controlled by computer.

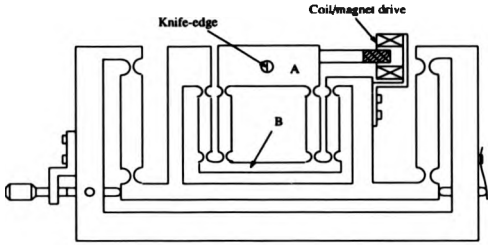


Figure 6.4: The device for form measurement of stylus tip.

With respect to the material of the knife-edge, although it is capable of giving higher performance than other materials, it is not appropriate to use silicon or glass ceramics etc. at the initial testing because of the extra manufacturing costs. So a razor blade or scalper blade is used as the knife-edge.

From the force analysis of the tip in Figure 6.5, the horizontal force (F_h) is

$$F_h = -W(\mu + \tan\theta) \tag{6.9}$$

where θ is the angle between stylus force (W) and the applied force (N) by the knife-edge, and μ is the friction coefficient between them.

For the worst case, $\theta_{max} = 45°$ and $\mu = 0.3$, a stiffness, λ, of a value at least 3.97×10^4 N/m is needed to limit to 5 nm the deflection caused by a maximum stylus force of 0.7 mN (BS 1134: PART 1:1988). Therefore, the length of the spring leg L is less than

$$L \leq \sqrt{\frac{EI^2}{\lambda K_T}}. \tag{6.10}$$

The parameters of the monolithic spring system driven by the coil/magnet system are calculated based on the optimizing program made by Smith (1988) and

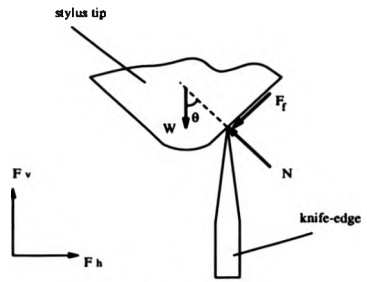


Figure 6.5: Stylus force analysis

listed below:

For coil/magnet,

- Inner radius $a_1 = 4.1$ mm;
- Outer radius $a_2 = 8.2$ mm;
- Length $l = 8.2$ mm;
- Coil packing factor $\chi = 0.77$;
- Turns on coil $N = 1300$;
- Wire diameter = 0.15 mm;
- Magnetic moment $M = 3 \times 10^{-7}$ Nm/H;
- Wire resistivity (copper) $\rho = 1.72 \times 10^{-8}$ Ωm;
- Inductance = 7 mH;
- Resistance $R = 43\ \Omega$.

For spring system,

- Notch thickness $t = 0.75$ mm;
- Hole radius $r = 4$ mm;
- Hinge depth $h = 10$ mm;
- spring leg length $L = 25$ mm;
- Young's Modulus (Aluminium) = 72 GPa;
- Natural frequency of the system = 159 Hz.

According to the design, the components were manufactured and assembled.

6.4.3 Experimentation

Experiment setting up

By using an IBM-PC computer and Bede Minicam interface to operate and control the movement of the knife-edge (a razor blade), the complete device has been set up on the Talysurf-4 to examine the stylus tip in the temperature-controlled laboratory of the Micro-engineering Centre. Photographs of the system are shown in Figure 6.6 and Figure 6.7. The motion of the stylus during translation of the knife-edge has been monitored from the voltage output from the stylus amplifier using a Metra Byte DASH-16 A/D Converter. This data was stored on a floppy disk for further analysis.

Results

The form of the stylus tip is plotted either using the chart recorder from the Talysurf-4 or storing the data into a disk and plotting out from the computer data (see figure 6.8). These have then been compared with a micrograph of the stylus tip imaged using a scanning electronic microscope (SEM) (see figure 6.9). Although

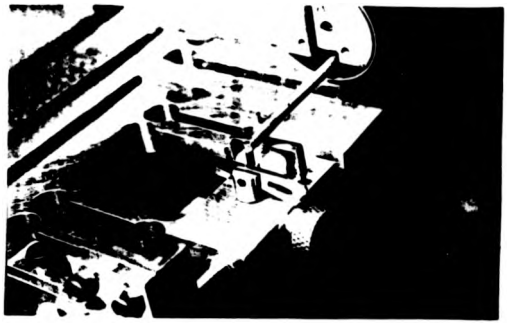


Figure 6.6: Photograph of the spring traverse system with razor blade.

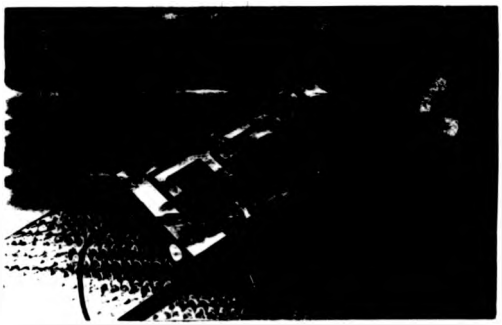


Figure 6.7: The whole construction of the knife-edge checking stylus shape.

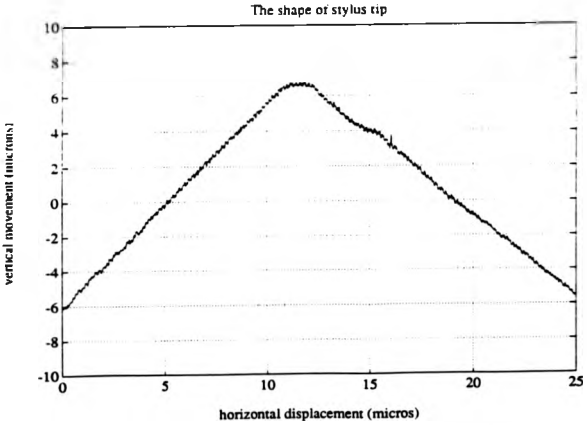


Figure 6.8: The graph of the stylus tip obtained from the computer.

Figure 6.9: A photograph of the stylus tip from a SEM.

only a visual impression is available from SEM picture, these are entirely consistent with the profiles.

In this test, a coil power(P) of 5.32 W was required. This gives a maximum force of 3.25 N per ampere. When such a power is applied to the translating system, the stress on the webs is 7.42 MPa and a displacement of 28.6 μm has been achieved.

6.5 Discussion

The most important factor in the measurement of the shape of a stylus using a knife-edge is that the knife-edge must follow a straight linear movement, otherwise the reading from the instrument in which the stylus is being tested gives only the combination of stylus shape and the knife-edge movement. In this design, the monolithic spring system offers good straight movement and linearity to the force applied. The coil/magnet force actuator is able to control the force in terms of current. Therefore, the knife-edge is under the control of the translation movement. The whole design

gives a reliable reading of the stylus shape.

6.5.1 Selection of a material for the knife-edge

From the tests, the material of the knife-edge has to be chosen carefully. The harder material of a diamond stylus tip causes severe wear on the softer steel blade after one test. The damage on a scalper blade was observed by SEM and is shown in figure 6.10. A comprehensive study of materials for precision engineering has

Figure 6.10: A SEM photograph of the damaged blade after use.

been conducted by Chetwynd (1987 & 1989) and the classification of materials and criteria of material selection are given. It is pointed out that diamond offers good mechanical performance. Therefore, a diamond knife-edge has been applied. The diamond was checked by SEM after several traverses. The observation is illustrated in figure 6.11. The collision damage is clearly seen and may be caused when two similar hard and brittle materials come into contact at the beginning of the test.

Therefore, from the tests, it can be concluded that frequent changes of the knife-

Figure 6.11: A photograph of the diamond edge from SEM after use.

edge are required if a metal blade such as a razor or a scalper is used, and if a diamond knife-edge is used, the edge and the tip must be protected from collision.

One possible method to protect from collision is to utilize a slope stage. The testing knife-edge is sat inside a groove on the stage. The depth of the groove has the same height as the knife-edge (see figure 6.12). To use the stage, the stylus is set lower than the slope end. When it crosses the knife-edge, the stylus is gradually lifted up by the stage, moving closer to the knife-edge and then sudden collision can be avoided.

6.5.2 Improvement of traverse displacement

A straight forward method of increasing the displacement is to increase the power. This, however, results in undesirable heat input. A better way is to fix two magnets at each end of the coil with opposite poles (see figure 6.13). This utilizes the field gradient at both ends of the coil and will double the force and given power. Using

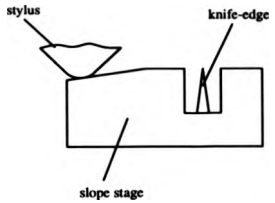


Figure 6.12: A possible method to protect of the knife-edge from collision.

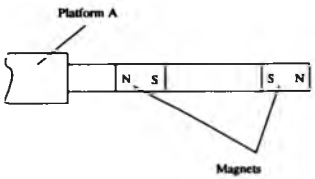


Figure 6.13: Improved coil/magnet system — dual magnet system.

this, the present simple setup will easily obtain 100 μm displacement.

6.6 Conclusion

The success of the device for stylus form measurement indicates that this method of measuring the shape of a stylus tip using a knife-edge is feasible. High quality diamond knife-edges with a radius of less than 50 nm have been successfully produced using ion beam milling (Miyamoto 1987). It is a portable method easily set up on any stylus instrument *in-situ*. With a chart recorder, qualitative information can be obtained for the quick assessment of the wear. With a computer, the numerical information can be used for the detailed quantitative dimensions of stylus tips suitable for numerical analysis.

Chapter 7

Mechanical Analysis of the Stylus Instrument

7.1 Introduction

The previous three chapters deal with the geometrical effects of a stylus. In this chapter, the other intrinsic effect of a stylus, i.e. the physical effect, will be discussed.

The geometry of a measured surface is obtained from the physical interaction between the surface and the tip. Then, the reaction of the tip is transferred through a transducer to an electrical signal, as illustrated in figure 7.1. The transducer is

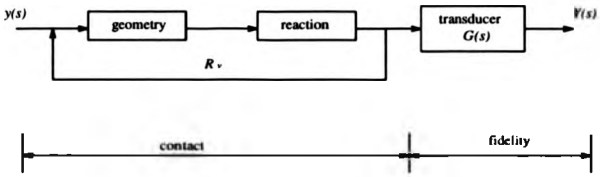


Figure 7.1: A flow diagram of a stylus system.

characterized by its translation function, either in the time or in the frequency domain, as the relationship of the input to output. Intensive studies of the translation function have been accomplished in control theory. It is not the purpose of this thesis. It focuses on the reaction, which is closely related to the characteristics of the transducer.

Based on the different principles, the media, or the reaction, between a measurand and a result is various, as chapter 1 discussed for the different profiling techniques. For example, the reaction bridge in the scanning t tunnelling effect. The reaction, in this case, is the overlap of the work functions between the tip and the surface. The reliability of an electron density image in representing the atomic topography of the surface is the fidelity of the measurement, which is not discussed in detail here.

In the above example, the tip scans within a very small area, possibly only a few tens of nanometers squared, and its speed is quite low. Because of the small gap of a few nanometres between the tip and the sample, the vibration isolation is a critical factor in the STM design. This is well discussed by Kuk & Silverman (1989). The dynamic effect within the system, therefore, is negligible. While, in the stylus instrument, the range of the measurement is very large, compared with STM, the speed of the stylus traverse has to be kept high enough to ensure the measuring time of a few seconds. The characteristic of the stylus pick-up system may varies the loading force through the stylus to the surface during the measurement. Hence, the dynamics of the stylus pick-up system become more important. The physical effect depends upon the mechanism used to sense surface information by the stylus. That is, it is instrument-dependent. Threfore, inorder to discuss this aspect, the stylus instrument - Talysurf 5, in the laboratory of the Centre for Micro-Engineering & Metrology, is selected. The physical aspect of the stylus in this case is, then, mechanical. The interaction is the stylus contacting force. First of all, the mechanical analysis of the stylus pick-up system will be presented in detail in this chapter. The dynamic characteristics will also be experimetally tested.

7.2 Stylus pick-up system

The basic stylus assemblage of the Talysurf 5 instrument is shown in figure 7.2. The tension spring is used to keep the anvil in contact with the knife edge pivot and to provide a static load. At one end of the stylus arm is the stylus tip and at the other is the ferromagnetic core for variable inductive transducer. The spring system has two functions: one is to contribute the contact force to the measured surfaces, and the other is to press the beam at the stop-end when the stylus is not used, to protect it from damage.

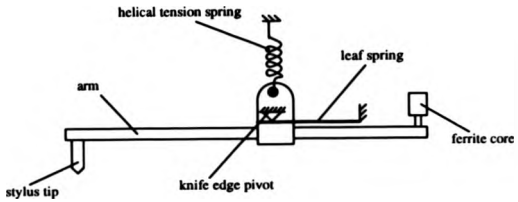


Figure 7.2: The basic stylus assemblage of the Talysurf 5 instrument.

7.3 Static characteristics of the system

The static characteristics of a measuring system is the relationship between the input and the output when input is either at a constant value or changing very slowly. These characteristics provide information that must be considered when choosing an instrument for a particular application. They include accuracy, range, linearity, sensitivity and resolution, etc.

7.3.1 Accuracy

Accuracy is very important in measurement. It is the extent to which a reading corresponds to the 'true' value. There are two forms of it: absolute and relative values. Only absolute accuracy is closely related to the range of measurement. The relative accuracy is a percentage of the full-scale reading. Among the stylus type of instruments, the absolute accuracy depends upon the dimension of measurement. For example, an instrument for form or shape measurement, needs less absolute accuracy, say a few microns, than a surface finish instrument, even though the former has a similar or higher accuracy than the latter. The absolute value has been steadily improved especially in surface measurement. Nowadays, it almost

hits the physical limit, which is down to the atomic level. On the other hand, for the same level of accuracy, the range of measurement has also been increased. For instance, the 'Talystep' has a resolution of less than 0.1 nm and a vertical measuring range of 10 μm, but the horizontal traverse range has improved from 5 mm to 50 mm and it becomes 'Nanostep'.

7.3.2 Linearity

Linearity is quantified in terms of the maximum non-linearity between the input and reading. It is sometimes named as 'non-linearity'. It is always desirable to have a straight-line (linear) relationship between I & O. It is dependent on the mechanism and also on environmental changes. Change in environment cause a non-linearity of reading in most cases. To avoid this, high precision instruments are normally kept in the laboratories in which temperature, pressure and humidity are controlled.

The non-linearity caused by the mechanical design can sometimes be compensated for by software. For example, the Talysurf series of surface instrument all use stylus arms. The consideration of the non-linearity of the stylus arm is displayed in figure 7.3, with a displacement of x and $-x$, respectively from the mean positions. The lateral movement of the stylus over the surface being measured will be $|e_2| - |e_1| = 2l\theta$ for small angles of θ. For $\theta = \pm 2°$, which is $\pm$ 1 mm measurement height for a 30 mm cantilever, the lateral movement will be 0.51 mm, as l = 7.3mm.

The stylus shank length l introduces a non-linearity such that

$$x = L\theta \pm \frac{l\theta^2}{2} \tag{7.1}$$

for small values of θ, where '+' is taken when the stylus moves upwards and '−' when downwards. Therefore for x positive, θ will be smaller than if no non-linearity existed, the transducer indicating a low reading, proportional to θ^2, likewise for x

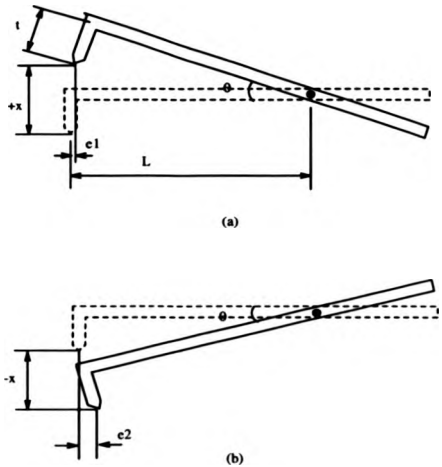


Figure 7.3: Nonlinear effect of the stylus tip length when the stylus arm moves upwards and downwards.

negative, the transducer indicates a high reading. For $\theta = \pm 2°$ and $t = 7.3$ mm the non-linearity will be $\pm 4.5\mu$m.

Other non-linearities depend upon the principles of the transducers based. For example, the Form Talysurf instrument uses a cube corner at the other end of the stylus arm, instead of an inductive gauge. The transducer is the interferometer. When the stylus arm is tilted, the change in glass path introduced into the beam and the lateral movement of the beam from the cube corner across the face of the mirror can introduce non-linearity.

7.3.3 Sensitivity

The sensitivity of measurement is a measure of the change in instrument output which occurs when the quantity being measured changes by a given amount. Sensitivity is the ratio of the change in output with respect to input. On the Talysurf series, a sensitivity as high as 4 volts/μm for Talysurf and 0.4 Volts/nm for Talystep can be obtained. But the sensitivities of environment and disturbance should be kept as low as possible.

7.3.4 Resolution

Resolution is also an important static characteristic. It is defined as the largest change in input that can occur without any corresponding change in output. This parameter is also dependent upon the physical principles used. In stylus instruments, for example, the vertical resolution is mainly determined by the electronic system. Talystep can be as good as 5 pm (1 pm $= 10^{-12}$ m). The lateral resolution is processed by the stylus tip dimension. In order to improve lateral resolution, a sharper tip is needed ultimately of atomic dimensions (see chapter 4). However, the physical contact between the stylus and a surface requires a strong enough tip so that the stylus will not be worn off. Thus, the sharpness of the tip is limited. For non-contact type of stylus instruments, the sharpness of the tip may be very high,

just a few atoms.

Other static characteristics such as span, zero drift, hysteresis, wear and durability, will not be discussed here. An example of the specifications of the stylus type instrument can be found in Appendix G.

7.4 Dynamic characteristics of the system

The dynamic characteristics of a measuring instrument describe how the instrument responds to the changes of the measured values. The important element is the time for the instrument to move from one steady state to another, namely, transient time, or settling time. Therefore, elements which are related to the velocity and the acceleration are the parameters to describe the instrument's dynamics.

In order to analyze the relationship, most instruments are regarded simply as linear, time-invariant systems of no more than the second order. In this case, a second order system of mass-damper-spring is used to model the system, as shown in figure 7.4 (referring to figure 7.2).

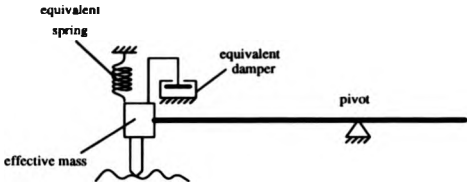


Figure 7.4: Equivalent analytical system of the stylus assembly.

The equation representing the system is given in differential form by

$$R_v(t) = m^* \frac{d^2y}{dt^2} + T\frac{dy}{dt} + ky + W, \tag{7.2}$$

where $R_v(t)$ is the vertical reactional force at the stylus tip end (referring to the next chapter). m^*, T and k are the effective mass, the damping ratio and the spring rate, respectively. Because of a constant traverse speed V, equation 7.2 in terms of the distance x is given as ($x = Vt$)

$$R_v(x) = m^* V^2 \frac{d^2y}{dx^2} + TV\frac{dy}{dx} + ky + W, \tag{7.3}$$

where $y = f(x)$ is the profile. The first and second order differentials, i.e. $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$, are representations of the slope and the curvature of the profile. The relationship between the reaction and the measured geometry is thus contained in equation 7.3.

The intrinsic dynamic characteristics of the system are determined by the coefficients of each term in equation 7.3, i.e. m^*, T and k. The fidelity of the stylus instrument is determined by these parameters. Measurements of these parameters were undertaken, and results are reported in the following sections.

7.5 Analysis and tests of characteristic coefficients on the pick-up

7.5.1 The spring rate

The spring rate of the stylus system can be tested simply by measuring the stylus movement by a given weight. Because of the end stop on the stylus arm and the leaf spring (see figure 7.2), the stylus arm is not in its equilibrium position when it is not in use. Therefore, the test was carried out by turning the stylus pick-up up-side-down as in figure 7.5, putting a weight of 0.320 g at the stylus end on the

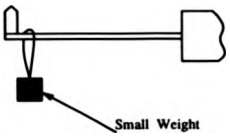


Figure 7.5: The stylus position for the spring rate test.

arm so that the stylus spring unit is in the equilibrium position. Loading a small weight of 0.165 g, the pen signal movement from the chart is as below in table 7.5.1 with a magnification of 100×. Thus, the spring rate k is given as (if the load is

Weight	Movement
0.165 g	21 mm/100×

Table 7.1: The spring rate test.

assumed to be within the linear range of the spring.)

$$k = \frac{\Delta W}{s} = 7.7 \text{ N/m}. \tag{7.4}$$

The result is well under the limit of the British Standards of 35 N/m for the stylus radius of 2 μm.

The accuracy of the result is estimated from the accuracies of the raw data. That is, the weights have a minimum estimated error of 0.001 g and the accuracy of the pen displacement is half of the division on the recording paper, which is 1 mm. Therefore, the estimated error of the spring rate is 0.3 N/m.

7.5.2 The natural frequency f_n

The stylus pick-up system is that of a lever arm. A small weight is loaded on the beam end after turning the stylus pick-up system up-side-down so that the beam is free from the stop-end and stays at an equilibrium position, as in the spring rate test. The natural frequency in this case is changed by the small weight. The relationship between the natural frequency of one with a small weight and without is as follows

$$\frac{1}{\omega_n'^2} = \frac{1}{\omega_n^2} + \frac{1}{\Delta\omega_n^2}, \tag{7.5}$$

where ω_n' is the natural frequency with a small weight, and ω_n is the natural frequency without it. And

$$\Delta\omega_n = \sqrt{\frac{k}{\Delta m}} \tag{7.6}$$

where Δm is the mass of the small weight.

The natural frequency is found from the hammer test. The vibrating signal from the stylus is transferred into and recorded by the Advanced Digital Spectrum Analyzer TR9403, shown in figure 7.6. The spectrum shows clearly that the resonance frequency, which is equal to the natural frequency of the system, is 16.6 ± 10^{-4} Hz. The spring constant k is found by loading a known weight and measuring the displacement of the stylus from the chart recorder, as 7.7 N/m. The small weight used was of 0.320 g. The natural frequency without the small weight, which is the natural frequency of the pick-up system, is calculated from equation 7.5 as 22.4 Hz. Other frequencies caused by, for example, torsional and longitudinal effects etc., are considered negligible. The error of f_n is estimated to be about 0.33 Hz.

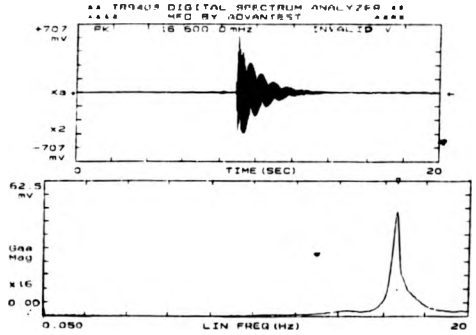


Figure 7.6: Time record and its spectrum of the stylus vibration.

7.5.3 The effective mass m^*

Since the natural frequency ω_n and the spring constant k are known, the effective mass m is easily concluded from equation 7.7 as

$$m^* = \frac{k}{\omega_n^2}. \tag{7.7}$$

In the case of f_n = 22.4 Hz and k = 7.7 N/m, the effective mass is 0.396 g. The estimated error of the effective mass is 0.7 mg.

The effective mass can be estimated from the geometry of the pick-up beam and the mass of each different components (Frampton, 1974). The pick-up beam is shown in figure 7.7.

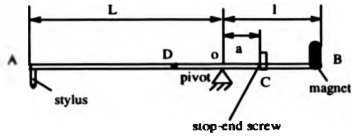


Figure 7.7: Analytical diagram of the pick-up beam.

The effective mass here is defined as the mass of a weight held at one end of a weightless beam with length L. From the point of view of kinetic energy, the relationship between real beam and effective beam is found as

$$\frac{1}{2}Mv_A^2 + \frac{1}{2}I_G\omega^2 + \frac{1}{2}mv_B^2 + \frac{1}{2}m_s v_C^2 = \frac{1}{2}m^* v_A^2, \tag{7.8}$$

where M, m, m_s and m^* are the masses of the stylus tip, ferrite core, stop-end screw and the equivalent mass, respectively. v_A, v_B and v_C are the velocities of end A, B

and C respectively, and

$$I_G = \frac{1}{3} m_D \left[\frac{1}{2}(L-l)^2\right], \tag{7.9}$$

where m_D is the mass of the beam, $m_D = \rho(L+l)$, ρ is the beam mass per unit length, the beam is assumed to be uniform. With $v_A = L\omega$, $v_B = l\omega$ and $v_C = a\omega$, the equation 7.8 is simplified as

$$m^* = M + m\frac{l^2}{L^2} + m_s\frac{a^2}{L^2} + \frac{1}{12}\rho(L-l)(1-\frac{l^2}{L^2}). \tag{7.10}$$

Let $\frac{l}{L} = r$ and $\frac{a}{L} = u$, hence,

$$m^* = M + mr^2 + m_s u^2 + \frac{1}{12}\rho(L-l)(1-r^2). \tag{7.11}$$

m^* in this method is in grammes.

In this case, the masses of the stylus tip, the stop-end screw and the core are approximately 60 mg, 30 mg and 185.4 mg respectively. The stylus beam is a hollow tube made of aluminium. The mass per unit length is about $\rho \simeq 2.64$ mg/mm. The beam has a total length of about 65 mm with the position of the stop-end screw and the core at $r = 0.50$ and $u = 0.30$. The mass and the inertia at the pivot position is neglected. Thus, the effective mass calculated from the equation 7.11 is 0.114 g. This effective mass is much lower than the one calculated from the natural frequency. It may be because of the inertia of the large weight at the pivot position and frictional losses at the pivot. The effective mass obtained from the calculation using the natural frequency and the spring rate is more reliable, due to the difficulty of estimating them.

7.5.4 The damping coefficient T

Damping mechanism other than simple viscous damping are generally difficult to predict, but they can be present to some measurable degree in real systems. Damp-

ing effects can be categorized into these three groups: (a) viscous, (b) hysteretic and (c) dry friction. Their effects are not much different when not at resonance. At resonance, viscous and hysteretic damping limit resonant amplitudes but dry friction damping does not. Of the three, only viscous damping affects the frequency of the resonant peak. Dry friction damping can be interpreted in terms of viscous damping when sliding speed is low. In the case of the stylus system, damping is assumed to be viscous, which is related to the first order differential of the system.

Normally, the damping effect is described by a quantity known as the 'damping ratio' and it is given as ξ,

$$\xi = \frac{\text{Actual value of damping coefficient}}{\text{Value of critical damping}}, \tag{7.12}$$

or

$$\xi = \frac{T}{T_{cr}}. \tag{7.13}$$

The meaning of the damping ratio ξ is determined can be found in most books of the vibrational system. Here, it is summarized in table 7.2, and the instrument responses are shown in figure 7.8.

Magnitude of Damping Ratio	Transient Response of Movement
$\xi = 0$	*No Damping* Continual oscillation at frequency ω_n and constant amplitude. ($\omega = \omega_n$)
$\xi < 1$	*Underdamping* Dampen oscillations at frequency ω. ($\omega < \omega_n$)
$\xi = 1$	*Critical Damping* oscillation just cease. ($\omega = 0$)
$\xi > 1$	*Overdamping* No oscillations.

Table 7.2: The damping ratio ξ. (Pearson, 1957)

Experimentally, the damping ratio ξ can be determined from a vibrating record

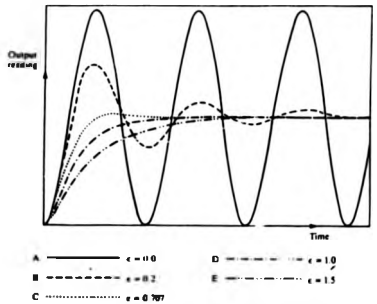


Figure 7.8: The instrument responses with different damping ratios.

by measuring two successive amplitudes. The logarithm of the amplitude ratio is known as 'logarithmic decrement' (η) and given as below (see figure 7.9)

$$\eta = \ln \frac{x_1}{x_2}. \tag{7.14}$$

The relationship between η and the damping ratio ξ is as follows,

$$\eta = \frac{2\pi\xi}{\sqrt{1-\xi^2}}, \tag{7.15}$$

or

$$\xi^2 = \frac{\eta^2}{(2\pi)^2 + \eta^2}. \tag{7.16}$$

In summary, the damping coefficient T is obtained from experimental record $\rightarrow$ logarithmic decrement $\eta \rightarrow$ damping ratio ξ.

In the experiment, the stylus was kept up-side-down and the same small weight is loaded to ensure the equilibrium position of the stylus with the springs. The record from the TR9403 analyzer is shown in figure 7.10. From the figure, it can be seen that the stylus pick-up system is undamped, i.e. $\xi < 1$. Therefore, the equation 7.15

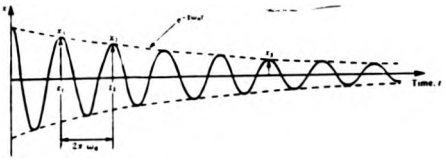


Figure 7.9: Logarithmic decrement.

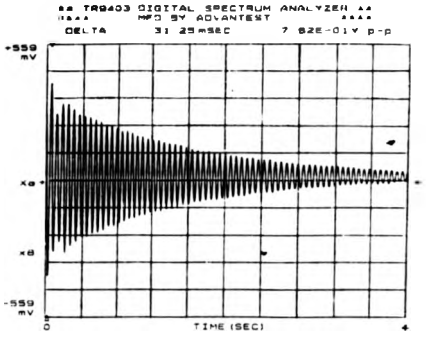


Figure 7.10: The record from the Digital Spectrum Analyzer TR9403.

can be simplified as

$$\xi = \frac{1}{2\pi} \ln \frac{x_1}{x_2}. \tag{7.17}$$

Practically, a more accurate calculation of ξ is obtained by measuring n successive peaks and the relation holds as

$$\xi \approx \frac{1}{2\pi n} \ln \frac{x_1}{x_n}. \tag{7.18}$$

In this case, three damping ratios from three different ranges of the record were averaged to obtain the damping ratio. The results are shown in table 7.3. The

Range	Damping Ratio ξ
$0 \rightarrow 16$	0.015
$16 \rightarrow 61$	0.012
$0 \rightarrow 61$	0.013
Average	0.0136

Table 7.3: Table of results for the damping ratio.

averaged damping ratio ξ is 0.0136 with its estimated error of 5.3×10^{-6}. The fact that ξ is much less than 1 implies that the system is almost complete undamped.

The critical damping coefficient T_{cr} can be given in terms of the spring rate and the effective mass as

$$T_{cr} = 2\sqrt{km^*}, \tag{7.19}$$

or

$$T_{cr} = \frac{k}{\pi f_n}. \tag{7.20}$$

With $k = 7.7$ N/m and $f_n = 22.4$ Hz, the critical damping coefficient $T_{cr} = 0.11$ kg/s[1]. With a small weight, the critical damping coefficient will be changed and the

[1]The dimensional unit of the damping coefficient T_{cr} is as: $T = \frac{F}{V}$, $[T] = \frac{[F]}{[m/s]} = \frac{[kg \cdot m/s^2]}{[m/s]} = kg/s$.

relation is given as

$$T_{cr} = \frac{T'_{cr}}{\sqrt{1 + \frac{\Delta m}{m^*}}}, \tag{7.21}$$

where T'_{cr} is the damping coefficient with the small weight, and T_{cr} is the one without it. The small weight Δm is 0.320 g and $m^* = 0.396$ g. Therefore, $T_{cr} = 0.74\ T'_{cr}$.

The damping ratio ξ of the system is assumed to be unchanged when without the small weight. Thus, the actual value of the damping coefficient T of the stylus pick-up system is

$$T = \xi T_{cr} = 1.76 \times 10^{-3}\ \text{kg/s} = 1.76\ \text{g/s}. \tag{7.22}$$

The estimated error of the damping coefficient T is 0.6%.

Summary

The coefficients of characteristics for the stylus pick-up system are achieved from the hammer test. In the case of Talysurf-5, they are listed below:

- Effective mass $m^* = 0.396$ g;
- Damping coefficient $T = 1.76$ g/s;
- Spring constant $k = 7.7$ N/m;
- Natural frequency $f_n = 22.4$ Hz.

7.6 Discussion of the results

The dynamic characteristic coefficients are determined by stylus instrument structures and rely upon measuring methods. A lot of studies have been carried out to investigate the dynamic characteristics of stylus instruments and the stylus kinematic effects. The tracibility of stylus instruments discussed by Frampton (1974) includes

the maximum velocity that a stylus can travel without loss of contact between the stylus and a surface, and the maximum surface amplitude at which contact remains. It is pointed out that the natural frequency of the instrument should be well outside the surface frequency range and the damping ratio kept as low as possible.

There are many ways to design a system having its natural frequency based on these conditions. For example, to keep the frequency outside the surface frequency range (about 300 Hz), the instrument frequency can be very low or very high. With an instrument of very low natural frequency, it can be designed by increasing the system's mass or decreasing the spring rate. Reversing the rules of the design, very high natural frequency can be obtained. Among them, reduction of the mass is limited by the need for certain physical sizes of the components. Increasing the mass or the spring rate will cause the stylus pressure to increase and consequently increase the risk of surface damage. So, a better dynamic design for the natural frequency is to decrease the spring rate to an acceptable stylus pressure. Therefore, the natural frequency is set a little higher than the top of the surface frequency range.

Frampton's work only outlined a preliminary dynamic investigation. There is no experimental confirmation. In the case of the tests conducted in this chapter, the resultant natural frequency f_n of 22.4 Hz confirms that the natural frequency is small. In addition, the equivalent wavelength with this frequency is 2.2 μm when the stylus travels at the standard measuring speed of 0.05 mm/s. This implies that a surface wavelength smaller than this will not detected, or, to be precise, the stylus may separate from the surface. Moreover, this data also matches the lateral resolution of the stylus instrument by its stylus tip radius of 2.5 μm.

The damping ratio ξ in Frampton's work is an assumption (it is approximately 0.2), because of the practical difficulty of determining it. The analysis in McCool's (1984) has also neglected this term. Nevertheless, this term, as Frampton (1974) pointed out, should be kept as low as possible and is believed to have a small value

(at least less than 1) in most of dynamic analysis. The test in this chapter shows the value of the damping ratio ξ is as low as 0.0136. In the design of the stylus system, the main element affecting its value is the pivot. From the manufacturer, the designs of the pivot, either the knife edge or the use of ball race pivot bearings, do not alter the dynamic performance significantly, even with $\xi = 1.0$ (Steven 1991).

The calculation of the effective mass recommended by Frampton (1974) has been tried in this case and resulted in 0.114 g which is smaller than the estimation in the chapter. This is discussed in the section 7.5.3.

Concerning the tests in this chapter as a whole, there are only two independent tests: the spring rate k and the damping ratio ξ. Although the natural frequency test is from the Digital Spectrum Analyzer TR9403, the resultant f_n is calculated through the small weight effect involving k. So, it is also k-dependent. Therefore, the test of the spring rate is very significant. Some other testing methods of k were conducted. A method of utilizing a load cell gave a slightly lower value. And the calibration of the spring rate by Liu *et al.* (1991) shows an insignificantly higher value and a non-linear rate near the equilibrium position. To compromise, the result of 7.7 N/m is acceptable. A set of these coefficients is also given by McCool (1984). The method of deduction is based on the data given by the manufacturer with the regression technique. The spring rate turns out to be a much larger value than the standards recommended (> 35 N/m, BS 1134: PART 1: 1988). And the effective mass is also too low, at 0.2 mg. Similar results obtained with a regression technique have also been tried by the author. The values of the coefficients vary a lot with different sets of raw data. One reason could be that the involvement of subtraction and division of small values in the programming processes causes the results to be very unstable.

Chapter 8

Contact Forces in the Stylus Instrument

8.1 Introduction

Contact is a feature of stylus instruments. To maintain the contact between the stylus and the surface, a force has to be introduced. The force is the means of communication between the sensors and the measurand — here, the surface. On the other hand, as the stylus travels along a surface, its movement leads unavoidably to frictional force.

In conventional stylus instruments, the contacting force is obtained by means of a spring passing through the stylus in the stylus pick-up system (referring to figure 7.2). The force, namely 'loading force', is normal to the surface of the specimen and is in the same direction as the measured surface height direction. Because of the horizontal movement of the stylus and dynamic effects, the loading force must be high to ensure contact with the surface (see the discussion in chapter 2). Ideally, the necessary loading force would be constant. However, the movement of the stylus will dynamically affect the vertical force. The load cannot remain unchanged during the measuring process. It is dependent upon the characteristics of the pick-up system, i.e. the effective mass, the spring constant and the damping ratio, as is shown in equation 7.2 or 7.3,

$$R_s(t) = m^* \frac{d^2y}{dt^2} + T\frac{dy}{dt} + ky + W. \tag{8.1}$$

The frictional force as well as the loading force is involved in the contact movement. Intensive studies of friction have shown that the frictional force is perpendicular to the normal to the surface, in the opposite direction to the movement, and dependent upon the normal force and the properties of both the contacting materials. They have also shown that the travelling speed influences the frictional force. Does the frictional force influence measurements of the surface? In the author's opinion, this force cannot be ignored in the analysis of a stylus instrument. Friction will be discussed in section 6.5.

In other types of stylus systems, the forces can be given in terms of other effects,

for instance, the electron charge force due to the tunnelling effect in STM, or the van der Waal force in AFM. In these cases, the forces are naturally attractive or repulsive between electrons or atoms. No mechanical contact exists. Therefore, no frictional force exists. However, when a tip moves horizontally along a surface, the physical force fields will be sheared. How this affects the results is not the topic of this thesis. As an example, the normal and shearing force - the frictional force in this case - in the conventional stylus instrument will be studied.

After a literature review these forces will be analyzed and discussed later in this chapter. Experimental tests of these forces in terms of reactional forces have been conducted. It will be shown that the frictional force plays a critical role in the analysis.

8.2 Literature review

The stylus load is regarded as one source of error in the instrument (Thomas, 1982). Several workers have shown the scratch caused by a stylus due to its loading force. Quiney *et al* (1967) show that a scratch about 1.7 μm deep is formed on an aluminum surface by using a stylus force about 10 times higher than the recommended standard. Guerrero & Black (1972) have also examined scratches as deep as 0.05 μm formed by SEM, when a stylus load of about 2 mN was applied on a steel surface. These scratches show that damage can be directly attributable to the load. However, a question arises: will similar damage occur if a standard force of 0.7 mN, much lower than these researchers have used, is applied.

Thomas (1978) attempted to estimate the deformation of a specimen surface due to the load of a pyramidal flat tipped stylus, using elastic analysis and the bulk properties of the steel specimen. Under a load of 1 mN, the surface deformation was found to be 8×10^{-4} μm and Thomas asserted that the effect was negligible.

Sherrington & Smith (1986) have pointed out that pyramidal stylii are likely

to cause less permanent damage to a specimen than hemispherical stylii, as they present a more uniform load to the surface.

All the above workers have investigated the theoretical and experimental relationship between load and deformation, using either an elastic model or a plastic model. The overall conclusion is that deformation under the standard recommended load is negligible. However, these studies have only been carried out under static conditions. Under dynamic condition, the applied load will vary. At some point the load may be large enough to cause severe damage. To find if this is indeed the case, the dynamic force must be tested experimentally.

Such an investigation has been carried out at PTB (Physikalisch Technische Bundesanstalt) in Germany. They found that for sapphire, steel and glass, the mean values of the elastic deformation depths, of the order of 10-30 nm, are in confirmation with the values calculated according to Hertz's theory, with a maximum uncertainty of ±30%. The plastic deformation is estimated on the basis of the Vickers hardness. For instance, with steel of Vickers hardness of 150 HV, a stylus tip of radius 5 μm and a static load of 1 mN, the elastic deformation is about 0.015 μm and the plastic deformation about 0.02 μm (Hillmann, *et al* 1984). These experimental results remain unconfirmed because of the difficulties associated with measuring the dynamic load.

8.3 Static loading force and surface deformation

Materials vary in hardness and strength. Under a certain load, they deform either elastically or plastically.

The elastic behaviour of a surface under a stylus loading force can be calculated (Timoshenko & Goodier 1951). In the model the surface is regarded as a homogeneous isotropic elastic half-space. The average vertical deflection δ caused by a rigid indenter of rectangular cross-section, is given as 8.3×10^{-4} μm for steel under a load

of 1mN. If the deformation is elastic, the measurement will not cause damage to the surface. The results, on the other hand, will be biased.

If the loading pressure is at least three times larger than the yield stress of the surface, plastic deformation will occur. In the standard stylus instrument, the tolerance of the loading force is less than 0.7mN for a nominal tip radius of 2 μm, and the maximum load of 16mN for a tip radius of 10 μm(BS1134: PART 1: 1988). The loading force is very small, but the pressure could be very high because the area of contact is very small.

If the contact is considered to be Hertzian, the stylus tapers to a spherical end. The contact region is a circular disk. The radius of the disk a is given by the well known Hertzian expression

$$a = \left\{ \frac{3\pi P R_1 R_2}{4(R_1 + R_2)} \left[\frac{1-\nu_1^2}{\pi E_1} + \frac{1-\nu_1^2}{\pi E_2} \right] \right\}^{\frac{1}{3}}, \tag{8.2}$$

where E and ν are Young's modulus and Poisson's ratio and R is the radius of curvature. In this case, $R_1 = R$, $R_2 = \infty$, so,

$$a = \left\{ \frac{3}{4}\pi P R_1 \left[\frac{1-\nu_1^2}{\pi E_1} + \frac{1-\nu_2^2}{\pi E_2} \right] \right\}^{\frac{1}{3}}. \tag{8.3}$$

For a loading force of 0.7mN and a stylus radius of 2 μm, a is 0.25 μm according to equation 8.3. The true contact area is about 0.19 μm^2. Therefore, the pressure in this area is about 3.7 GPa. This is much greater than the yield stress of metals. Therefore, the deformation is plastic. Figure 8.1 shows typical damage on a surface caused by a stylus, taken by Bowen (1987).

8.4 Force analysis

R_t, the reactional force, is the force given by the stylus after it preloads onto the surface. It is equal to the loading force if the stylus is stationary. When the stylus

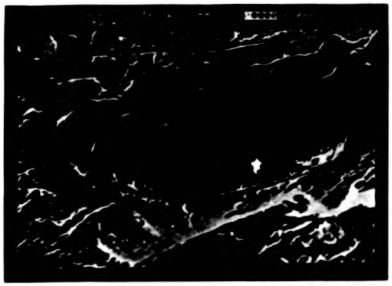


Figure 8.1: A surface damaged by a stylus.

does not travel across a surface, the reactional force is equal but in the opposite direction to the stylus loading force W. The direction of the reactional force is towards the stylus at the normal to the contacting point. However, when the stylus moves across a surface it is lifted up and down by the irregularities of the surface, and the reactional force will change accordingly (figure 8.2).

The force between the stylus and a surface at the contact point is shown in figure 8.3. The vertical reactional force

$$R_v = R_t \cos\theta. \tag{8.4}$$

When the stylus is not moving, except for the force caused by the stylus resting on the table, the frictional force is given by

$$F_f = \mu R_t = 0. \tag{8.5}$$

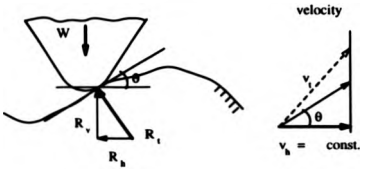


Figure 8.2: Diagrams showing the reactional force of a stylus and the relative velocity between a stylus and a surface.

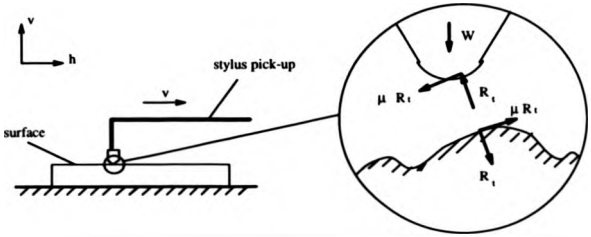


Figure 8.3: Analysis of the force at the contact point between the stylus and a surface.

For the stylus,

$$R_t \cos\theta - W = 0 \tag{8.6}$$

Therefore,

$$R_t = \frac{W}{\cos\theta} \tag{8.7}$$

or $R_v = W$. When the stylus moves across the surface towards the right, for the surface

$$\begin{cases} \sum F_v = -R_t \cos\theta + \mu R_t \sin\theta = R_t(\mu \sin\theta - \cos\theta) \\ \sum F_h = R_t \sin\theta + \mu R_t \cos\theta = R_t(\sin\theta + \mu \cos\theta), \end{cases} \tag{8.8}$$

and, similarly, for the stylus

$$\begin{cases} \sum F_v = -W + R_t \cos\theta - \mu R_t \sin\theta \\ \sum F_h = -R_t(\sin\theta + \mu \cos\theta). \end{cases} \tag{8.9}$$

Examining the movement of the stylus, it can be seen that

$$\sum F_v = m^* \frac{d^2y}{dt^2} + T\frac{dy}{dt} + ky. \tag{8.10}$$

Substituting $\sum F_h$ & equation 8.4 in equation 8.9, it becomes

$$R_v = \frac{1}{1 - \mu \tan\theta}(m^* \ddot{y}_t + T\dot{y}_t + ky_t + W). \tag{8.11}$$

Compared with equation 7.2, the vertical reactional force in equation 8.11 contains the frictional term. For normal engineering surfaces, the slope angles, θ, are no greater than 10°, and $\mu \tan\theta$ is much less than 1, so equation 8.11 becomes

$$R_v \simeq m^* \ddot{y}_t + T\dot{y}_t + ky_t + W. \tag{8.12}$$

It can be seen clearly that the dynamic reactional force is no longer constant at W. It varies according to the dynamic coefficients and surface geometrical features.

In addition, friction is involved.

8.5 Experimental method

As the vertical reactional force R_v is in the same direction as the surface height, it is difficult to measure the dynamic force from the stylus directly. Therefore, R_v is calculated from the total horizontal force $\sum F_h$ on the surface. The method is represented schematically in figure 8.4. The relationship between R_v and $\sum F_h$ is,

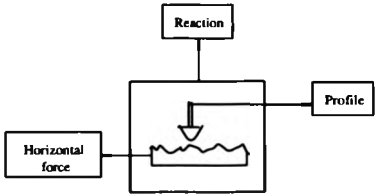


Figure 8.4: The diagram of indirect method of measuring the forces.

from equation 8.8,

$$R_v = R_t \cos\theta = \frac{\sum F_h}{\tan\theta + \mu}, \tag{8.13}$$

and in the dynamic case by equation 8.12.

In order to measure the total horizontal force $\sum F_h$, a linear vibration displacement table (LVDT) constructed by Smith (1989) was used. The LVDT has a notch type linear spring, with its displacement axis parallel to the traverse axis of the stylus. The drag forces during the traverse cause submicron deflections of the spring platform which are measured by an inductive transducer, similar to the one in 'Talystep'. The LVDT with a surface under the stylus is illustrated in figure 8.5.

Figure 8.5: The LVDT used in the experiment constructed by Smith (1989).

8.5.1 Calibration of the system stiffness of the LVDT

The spring stiffness of the LVDT system plays a critical role in the system. The stiffness was tested by loading low weights by means of a cotton thread, as shown in figure 8.6. The displacements of the LVDT caused by the tension of the horizontal

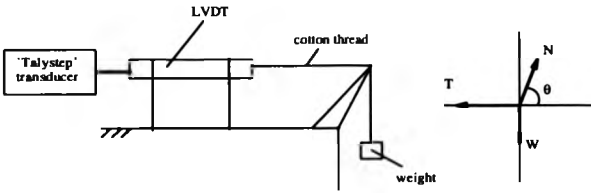


Figure 8.6: The construction of LVDT stiffness test.

part of the thread were recorded. Analyzing the thread tensions in figure 8.6, the tension of the horizontal part is not equal to the weight, and the relation is

$$T = \frac{W}{\tan\theta}. \tag{8.14}$$

So,

$$\begin{cases} T < W, & \theta > 45^\circ \\ T = W, & \theta = 45^\circ \\ T > W, & \theta < 45^\circ. \end{cases} \tag{8.15}$$

In this case, $\theta \simeq 70^\circ$. Hence, $T = 0.36W$.

Five different weights varying from 6 g to 2 g were used. Each weight was loaded 10 times and then averaged. The results are shown in table 8.1. The error for the stiffness in each weight was estimated by differentiating the relation, i.e. $\lambda = W/d$. Errors in W and d were evaluated from the accuracies of the weights and

the displacement measurement.

Weights (mN) W	56.5	47.6	37.8	28.2	18.3	Average
Averaged displacement (mm) 5000 ×	40.1	33.9	24.3	20.8	12.8	
Stiffness λ (N/m)	2564.4	2554.2	2828.5	2467.6	2608.2	2604.6
$\lvert\Delta\lambda\rvert$ (N/m)	7.01	7.02	8.67	11.85	18.17	10.54

Table 8.1: Testing results of the LVDT stiffness.

Thus, the averaged system stiffness of the LVDT is 2604.6 N/m with an averaged error of 10.54 N/m.

8.5.2 Calibration of the coefficient of friction

By tilting a flat specimen

The frictional coefficient μ can be calibrated in several ways. One way is to utilize a well finished flat specimen by tilting it with a small angle α. When the specimen is tilted to such an angle that there is no indication on the LVDT,

$$\mu = \tan\theta. \tag{8.16}$$

The angle can be measured directly by means of geometrical adjustment. Three ways of adjustment are shown in figure 8.7. In (a) & (b), the adjustment is achieved by varying the height with a screw in (a) in (b) the adjustment is achieved by varying the distance with a rod, and

$$\mu = \tan\theta = \frac{b}{a}. \tag{8.17}$$

In (c) the adjustment is obtained by varying the stylus tracking direction angle β on a specimen tilted with a fixed angle α. The frictional coefficient in this case is

$$\mu = \tan\theta = \frac{\sin\beta}{\cos\alpha}, \tag{8.18}$$

(see figure 8.7 (c)). Method (b) was used for the experiment. A rod with a diameter

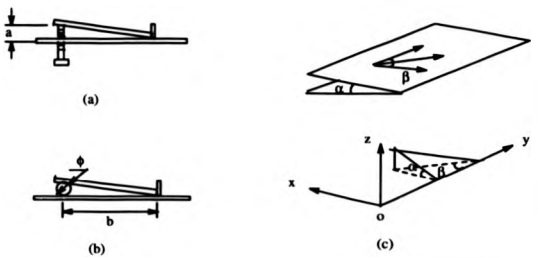


Figure 8.7: Three examples for the adjustment in the friction coefficient test.

of 1.5 mm was placed under a piece of silicon wafer which had been polished smooth for microelectronic applications. The displacement b is approximately 28.5 mm when the LVDT had a zero reading. Therefore,

$$\mu = \tan(2\tan^{-1}\frac{\phi}{2b}) = 0.053. \tag{8.19}$$

with an estimated error of 6.8×10^{-6}.

The frictional coefficient μ between silicon and diamond can also be obtained directly by a calculation from the displacement of the LVDT, as long as the static loading force is known. In our case, the value of μ will be used later in the calculation

of the static force. Therefore, the latter method was not used.

8.6 Static loading force test

8.6.1 The principle

Generally, the static loading force is tested or calibrated simply by loading the stylus onto a scale. This was executed by two type of scale. One was a conventional scale - OERTLING scale, and the other was a load cell with a nominal range of 50 mN, and very high sensitivity. The result was 3.9 mN from the OERTLING scale and 4.05 mN from the load cell.

Another method, based on dynamic analysis by Whitehouse (1989a), was also conducted. The idea is to use the the LVDT with a very well finished step specimen. If the specimen is placed flat on the LVDT, it can be seen from the proceeding force analysis that the first two terms in equation 7.2 disappear because $\bar{y} = 0$ & $\dot{y} = 0$. The horizontal force is entirely due to friction which is assumed to have a linear relationship with the normal force, i.e.

$$\mu(ky + W) = \lambda d. \tag{8.20}$$

For the LVDT displacement, the two different heights of the step refer two different readings. So,

$$\begin{cases} \mu(ky_1 + W) = \lambda d_1 \\ \mu(ky_2 + W) = \lambda d_2. \end{cases} \tag{8.21}$$

Solving the equation 8.21 for W and k, it gives

$$W = \frac{\lambda(d_2 y_1 - d_1 y_2)}{\mu(y_2 - y_1)} \tag{8.22}$$

and

$$k = \frac{\lambda(d_1 - d_2)}{\mu(y_1 - y_2)}. \tag{8.23}$$

8.6.2 Fabrication of the step specimen

The material of the step specimen was silicon in order to keep the material-dependent parameter μ unchanged. Chemical etching was used as the fabricating process as it gives a good finish and it is easily controlled. The amount of material etched off is determined by the amount of time the material spends immersed in the etching liquid. The etching liquid is composed of Hydrofluoric acid (HF), Nitric Acid and Acetone. Fast etching can be obtained with high concentrations of HF. For example, 1% HF gives an etching depth of 0.2 μm per minute, and 10% HF 6.4 μm. An etching liquid composed of 5% HF was used, and the unetched part was covered by lacquer. After etching the step specimen was profiled by the Talysurf-5 instrument, illustrated below, with a step height of about 3.2 μm. With the etching technique,

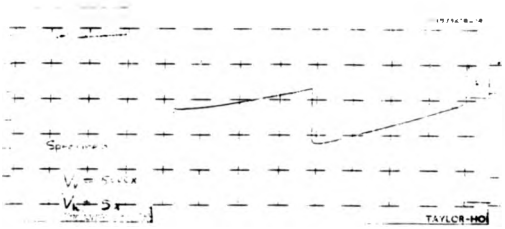


Figure 8.8: The record of the step specimen from the Talysurf-5.

the longer the etching time is, the more the corner is etched off and the bigger the arc given (see figure 8.8). These are the disadvantages of the etching process.

8.6.3 The experiment and the results

The specimen was placed on the LVDT and traversed by the stylus on Talysurf-5 at the lowest speed (about 0.01 mm/s), as shown in figure 8.9.

Both the Talysurf-5 and the LVDT were recorded through an A/D converter into a computer. The test was conducted 49 times, to ensure accurate measurement. The static force W was calculated from equation 8.21 for each independent test. All the results are listed in table 8.2.

No.	W (mN)	No.	W (mN)	No.	W (mN)	No.	W (mN)	No.	W (mN)
1	2.44	11	2.27	21	2.39	31	3.13	41	3.05
2	9.28	12	3.26	22	3.48	32	2.95	42	2.52
3	2.67	13	3.12	23	3.58	33	3.33	43	2.65
4	2.45	14	2.98	24	3.65	34	3.04	44	3.61
5	2.73	15	3.10	25	3.53	35	2.81	45	3.13
6	2.24	16	2.55	26	3.06	36	1.63	46	3.57
7	2.77	17	2.47	27	3.54	37	3.55	47	3.99
8	2.67	18	2.50	28	3.45	38	3.19	48	2.44
9	2.81	19	3.28	29	2.64	39	3.02	49	3.33
10	3.21	20	2.71	30	3.06	40	2.78		

Table 8.2: The result list of the static force test.

Except test No.2, which gave a poor value, the other 48 data give an average of 2.97 mN with a standard deviation of 0.44.

In order to check whether the tests follow the Gaussian distribution, a χ^2-test was carried out. Seven groups of force, from the minimum of 1.63 mN to the maximum of 3.99 mN, were divided into ranges of 0.33 mN. The results of the χ^2-test are listed in table 8.3. With the seven groups and two assumed parameters, i.e. the mean and the standard deviation, the dimension of the χ^2-test is 4. From the standard χ^2-test table,

$$\chi^2_{0.05}(4) = 9.488 > 7.7346. \tag{8.24}$$

Hence, the observed values are considered to be Gaussian distribution with mean of

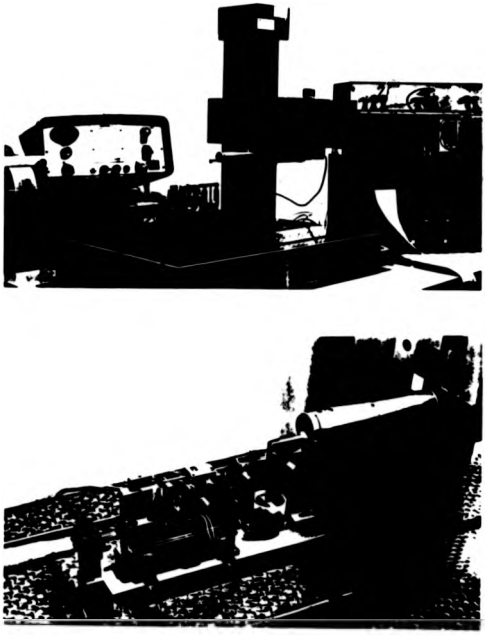


Figure 8.9: The experimental construction of the force test.

Group	Observed $freg_i$	Theoretical np_i	$\frac{(freg_i - np_i)^2}{np_i}$
1	1	0.2832	1.8143
2	2	1.4784	0.1841
3	9	5.1792	2.8187
4	11	11.2176	0.0042
5	13	13.4784	0.0170
6	11	10.2624	0.0530
7	1	4.6272	2.8433
sum	—	—	7.7346

where $n = 48$ and p_i is the theoretical probability from the standard distribution table.

Table 8.3: The resulting table of χ^2-test on the static force.

2.97 mN and the standard deviation of 0.44 with a confidence of 95%.

8.6.4 Discussion of the results

The static loading force is 2.97 mN. The values measured by the scales are slightly different; the load cell is slightly higher. A possible reason for this is that since the stylus loads on surfaces, the spring has contributed to the total force. Therefore, the measurement by the scale and the load cell involved with the static loading force and the spring force together, gives a higher results. In addition, all the results are higher than the value recommended by standards. This may be due to utilizing the instrument for a variety of experiments and a lack of calibration. However, the higher value of the load does not matter in the dynamic performance test. The purpose of the experiment is to show that it is possible to test the dynamic loading force.

8.7 Dynamic force test

The previous sections have introduced the LVDT, and explained how it is used to measure the static reactional force. The main point of utilizing the LVDT is to

monitor the dynamic force indirectly.

The experimental dynamic force can only be tested at the same time as the stylus is measuring. Because the direction of the stylus force and the profile height is the same, it is very difficult to test the dynamic force directly. Since the stylus travels on the surface, the frictional force along the measured surface exists and it is related to the applied load. It may be possible to measure the reaction of the stylus during its motion using friction. Friction, therefore, is the bridge to test dynamic force *in-situ*.

8.7.1 The relationship

The relationship between the total (measured) horizontal force $\sum F_h$ and the vertical reactional force has been given in equation 8.13. The surface slope is $\tan\theta$. It can be deduced from the height record in the Talysurf-5. The frictional coefficient μ was calibrated before as 0.053. $\sum F_h$ is the result of multiplying the stiffness of the LVDT and the displacement from the Talystep recorder.

To avoid division by zero or a very small denominator of $\mu + \tan\theta$ in the computer program, logarithms are applied to both sides of equation 8.13. After the calculation the inverse function was taken to obtain R_v. That is,

$$\log R_v = \log \sum F_h - \log(\tan\theta + \mu), \tag{8.25}$$

then,

$$R_v = \exp\left[\log \sum F_h - \log(\tan\theta + \mu)\right]. \tag{8.26}$$

8.7.2 The experimental results

In the test, a ground silicon surface was profiled. The records from the Talysurf-5 and the LVDT are illustrated in figure 8.10. In these graphs, it is more convenient to use data points on the x-axis because they contain the recording data when

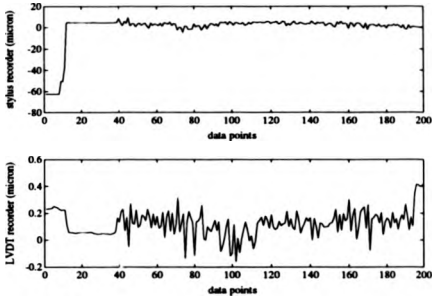


Figure 8.10: The records from the Talysurf-5 (upper) and the Talystep (down) for a ground silicon surface.

the stylus was not triggered. After the stylus moves, the actual sampling distance between data points is 2 μm, as the stylus speed was 0.01 mm/s. At this speed, the influence of the stylus size, which is about 2 μm (referring the SEM picture in figure 6.9), can be ignored.

The first graph in figure 8.10 clearly shows three sections related to the three stages of the operation. The first stage occurred when the stylus was unloaded. The stylus rested on the stop-end by the spring, so the data in both graphs is of no value. The horizontal straight line section in the top graph (the second stage), occurred when the stylus had just been loaded onto the surface. The loading force was equal to 2.97 mN as in the section before. The LVDT reacted to this load and gave an ambiguous amount of movement. It may have been caused by side contact between the stylus and the rough surface, or, more likely, by bending of the stylus beam due to the loading force, as shown in figure 8.11. The latter case produces a force to the surface towards the left, as in the figure. The largest section in figure 8.10

Figure 8.11: The horizontal force given by the bending beam of the stylus.

occurs when the stylus was triggered (the third stage of the operation). From this operation, the data explicitly illustrates the effect of friction in the LVDT record between the end of the second stage and the third stage. The horizontal force is reduced when the stylus was moved from left to right in figure 8.10. This also confirms the force analysis in figure 8.3.

In order to find out whether the two measurements were obtained simultaneously, the cross-correlation function between them was calculated, as shown in the figure 8.12. The maximum of the correlation of the profile to the LVDT displacement is

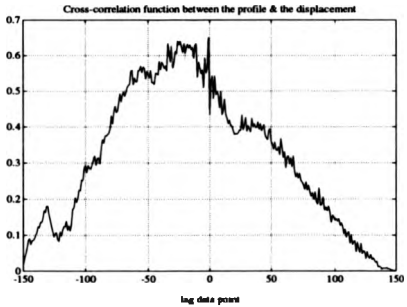


Figure 8.12: The cross-correlation between the profile and the LVDT displacement.

0.65. But it occurs at the -1 lag data point. From the definition of cross-correlation

function (Zhejiang University, 1979),

$$R_{XY}(\tau) = \mathrm{E}[X(t)Y(t+\tau)], \tag{8.27}$$

where $\tau > 0$ means that $Y(t)$ advances $X(t)$ and vice versa. Therefore, the LVDT recorder has a delay of one data point in this case, compared with the stylus instrument.

The vertical reactional force was calculated from equation 8.13. It is shown in figure 8.13 (top).

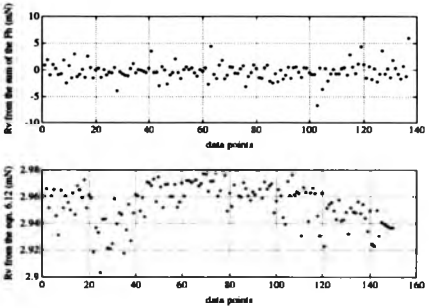


Figure 8.13: The experimental result of R_v from $\sum F_h$ and from the eqn. 6.12.

8.7.3 Analysis of the results

The vertical reactional force R_v

The R_v value obtained from $\sum F_h$ in 8.13 has a maximum cross-correlation of 0.05 with the profile. After taking out the outliers ($|\text{values}| \geq 9\text{mN}$), the maximum value

was increased to 0.22. It is still not very high. This value of the cross-correlation implies that the relationship between R_v and $\sum F_h$ in equation 8.13 is not reliable.

As the dynamic coefficients of the stylus instrument are known (see equation 8.12, as shown in figure 8.13 (bottom) with a travelling speed of 0.05 mm/s (equivalent to the $V_h^{\cdot} = 20\times$). The static force is 2.97 mN (in the previous section). The R_v value varies around the static force with the surface profile. Its maximum is 3.02 mN, its minimum is 2.94 mN and its mean is 2.99 mN. So, the variation is caused by the stylus dynamic loading force. The dynamic variation of the stylus force is 2.7% and negligible on surfaces where the roughness R_a is 2.125 μm or R_q is 0.229 μm.

However, this dynamic force is also influenced by the static force value. If the static force decreased to 0.7 mN as in the Standards, the ratio of the dynamic to the static forces on the same surface with the same travelling speed will be 3.7%, and it can still be ignored.

On the other hand, the dynamic force also depends on the speed, especially the second power in the first term of equation 8.12. In the Standards for surface measurement, the travelling speed is less than 1mm/s. A high speed will lead to kinematic problems, e.g. the stylus cannot keep in contact with the surface. The relationship between the speed and R_v was calculated and shown in figure 8.14. As can be seen from the figures, the mean and the maximum values of R_v increases dramatically when the travelling speed is greater than 1 mm/s. In this case, the stylus may sometimes lose touch with the surface. Noise caused, for example, by vibrations in the stylus pickup, are added into the recorder. The R_v value, therefore, may not be accurate.

In addition, the dynamic force also depends upon the surface geometry measured. A systematic analysis of the relationships between the speed, the surface amplitude and the stylus tip size has been given by McCool (1984).

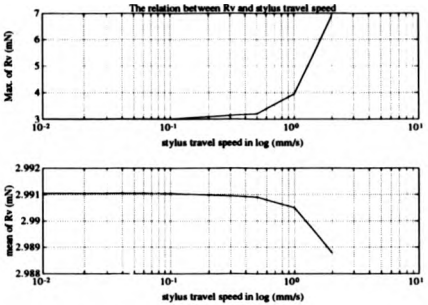


Figure 8.14: The relationship between the travelling speed and the R_v.

Correlations

From figure 8.12, it can be seen that both recorders are fairly highly correlated (Max. = 0.65). The auto-correlation functions of the records are displayed in figure 8.15, respectively.

8.8 Discussion

Because of the difficulties in measuring the dynamic reaction directly, the indirect method is favoured. The indirect technique shown here is, in principle, feasible. Experiments have been carried out. However, the results for R_v (vertical reaction) look inconclusive. The probable reason is that the frictional theory is still not well understood. That is, the frictional coefficient, μ, may not be constant or predictable. In practice, this method of measuring the dynamic reaction using friction as a bridge is almost impossible, unless a better understanding of stylus friction is reached.

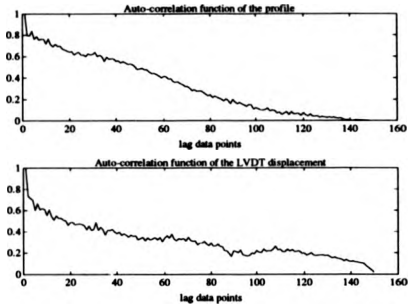


Figure 8.15: The auto-correlation functions of the profile and the displacement.

8.9 Consideration of friction in the stylus type of instruments

8.9.1 Discussion of local friction

In general, when friction is considered in surface measurement instruments, the frictional force, due to the normal load, W, and the relative movement between contact surfaces, is given by μW, where μ is the coefficient of friction between a diamond stylus and a specimen surface. It depends on the contact materials. The direction of the frictional force is perpendicular to the direction of W. The total work done by the frictional force when a stylus travels a distance L, is

$$P = -\mu W L, \tag{8.28}$$

where the negative sign means that the frictional force has the opposite direction of the stylus traverse. However, this is too simple. At the microscale it is easy to see that the actual contact is not only on the tip of the stylus, but also on the side because of the roughness of the surface (see figure 8.3). If the reactional force is R_t, which is normal to the point of contact, the tangential force, i.e. frictional force, is μR_t. In this case, the frictional force varies according to the slope angle of the surface at the contact point. The total reactional forces in the horizontal and vertical directions are given in equation 8.8. Therefore, the effective coefficient of friction is as below,

$$\mu_{eff} = \frac{\sum F_h}{\sum F_v} = \frac{\tan\theta + \mu}{1 - \mu\tan\theta}. \tag{8.29}$$

As $\mu\tan\theta \ll 1$,

$$\mu_{eff} = \mu + \tan\theta. \tag{8.30}$$

The actual distance traveled by frictional force is

$$L_i = L\sqrt{1 + \tan^2\theta(t)}. \tag{8.31}$$

Thus, the work done by friction force is

$$P_i = \mu R_t L\sqrt{1 + \tan^2\theta(t)}. \tag{8.32}$$

In this case, the instantaneous travel length is certainly longer than the nominal length L. The work done by frictional force is therefore greater than the simple case.

8.9.2 The effective friction coefficient

The effective frictional coefficient μ_{eff} can be calculated from the ratio of the mean of the total horizontal force and the vertical reactional force. It should be noticed that this parameter is a statistical average. As the equilibrium position of the stylus

system was not known, the third section of the profile in figure 8.10 only refers to the variation of the stylus. The mean of the total vertical force, including the static part, is 2.955 mN. The mean of the total horizontal force was calculated from the LVDT displacement and its stiffness of 2604.6 mN/mm, as 0.1955 mN. Thus,

$$\mu_{eff} = \frac{\sum F_h}{R_v} = 0.0662. \tag{8.33}$$

The summation of the local frictional coefficient μ_i and the mean of the profile slope is 0.0646. This represents a relative difference of 2.4% only between the both sides of the equation and a good agreement is revealed. Hence, the parameter of the effective frictional coefficient μ_{eff} has been confirmed experimentally.

8.9.3 Cross-correlation of surface information and friction

From equation 8.30, it can be seen that the effective friction coefficient μ_{eff} is related to the surface slope information, and the slope angle θ can be expressed in terms of $\dot{y}/V$, where V, the speed, is assumed to be constant. The cross-correlation of R_v with y (surface profile height information) is given by Whitehouse(1988)

$$C_{R_v \cdot y} = M\frac{\mathrm{d}^2 A_y(\beta)}{\mathrm{d}\beta^2} + kA_y(\beta), \tag{8.34}$$

where $A_y(\beta)$ is the autocorrelation function of y, and β is the lag distance. From equation 8.13 and 8.30, the cross-correlation of the total horizontal force $\sum F_h$ with the surface height y is

$$C_{\sum F_h \cdot y} = E[(M\ddot{y} + T\dot{y} + ky + W)(\mu + \frac{\dot{y}}{V}) \cdot y(t+\beta)] \tag{8.35}$$

$$= \mu M E[\ddot{y} \cdot y(t+\beta)] + \mu k E[y \cdot y(t+\beta)] \tag{8.36}$$

$$= -\mu[M\sigma_s^2 - kA_y(\beta)], \tag{8.37}$$

or in other words,

$$C_{\sum F_{h\cdot y}} = \mu[M\frac{\mathrm{d}^2 A_y(\beta)}{\mathrm{d}\beta^2} + kA_y(\beta)]. \tag{8.38}$$

Therefore,

$$\mu = \frac{C_{\sum F_{h\cdot y}}}{C_{R_{v\cdot y}}} \tag{8.39}$$

This implies that friction is related to the surface geometry and the reactional force. In the study of friction, the surface geometry has to be taken into account, and conversely, in the study of surfaces, friction can not be ignored.

8.10 Conclusion

In this chapter, in order to study the effects of the stylus force on a surface, the static and dynamic forces of the stylus have been analyzed theoretically. For practical assessment, an indirect method of force measurement by the LVDT has been proposed. The static loading force has been tested by a conventional scale or a load cell, and also using a step silicon specimen was fabricated and used. The tested static force is much higher than that recommended by the Standards, due to long experimental use without calibration.

The aim of the experiment using the LVDT is the measurement of the stylus dynamic loading force. However, its dependence on the frictional force between the stylus and the surface leads to an inconclusive result for R_v. To sum up, the idea behind this method is that the frictional force in the horizontal direction is used to estimate the vertical reaction. The vertical stylus force can be measured indirectly. It might have the disadvantage of being less accurate but would, nevertheless, allow a measurement to be made in a convenient manner without disturbing the normal instrument operation. The failure of the dynamic force measurement through friction also indicates that the surface measurement and the stylus-surface friction cannot be separated.

Some further work on friction has been conducted by Chetwynd *et al* (1991) and

Liu *et al* (1991). A capacitive force controller has been built and fixed onto the stylus. The friction has been studied with a fixed loading force.

Chapter 9

General Discussion and Conclusions

9.1 General discussion

The purpose of this work is to investigate stylus profiling techniques with respect to the requirements of development of instrumentation and measurement. It concentrates particularly on some problems of the mechanical aspect of data collection. The conventional stylus instrument was used throughout this research to study this aspect. There are two underlying themes throughout this work. One is that the stylus instrument is used as a tool for the tribological study of engineering surfaces. The other is a consideration of the requirements of the new instruments - scanning probe microscopes - with a view to their further development. The former aspect is the main subject of the work.

The choice of the stylus instrument does not mean a loss of generality. The methodology and characteristics of the stylus profiling techniques have one thing in common with the new generation of scanning probe microscopes - the stylus. On the other hand, in order to obtain a thorough understanding of the effect that the stylus has on measurement, a specific instrument has to be chosen. This is also necessary for a discussion of the aspects that are instrument-dependent, for example, the physical effect of the stylus. Some more detailed discussion on these two lines are presented below.

9.1.1 A tool for surface functional study

The original requirements of the stylus instrument were to improve the performance of engineering components. Through this work, however, it has been shown that the instrument has some disadvantageous factors when it is used for the tribological analysis of surfaces. The geometry of a component's surfaces is important to the tribological study and should be measured as accurately as possible. It has been known for a long time that the stylus trace is not exactly the same as the true surface profile, and their relationship was unknown. Generally, accuracy is improved with a

sharper stylus, though this is obviously limited. So, the effect of the finite dimension of the stylus has to be examined in more detail. It has been found that the low-pass filtering characteristic is caused by the size of the stylus, and the influence on the measurement is not correctable. On the other hand, the stylus shape distorts the local information on a profile. This bias can be corrected when the stylus is given (chapter 5). This also implies that the geometry of the stylus is an important factor. It should be noticed that this modification is limited by the fact that the stylus profiling technique is unable to reveal surface valley information properly in some cases. The effects of the stylus are intrinsic. It is difficult to remove them as long as the stylus tip is employed. The scanning probe microscopes reduce the effects to a certain extent.

The physical aspect of the tip in the stylus instrument is mechanical, i.e. the stylus contact force. It is known that the contact force has to be kept as low as possible to reduce damage to the surface. However, the dynamic character of the system will vary this force. Experimental results from testing the dynamic loading force in chapter 7 and 8, show that the dynamic variation of the loading force of the stylus can be neglected during the measurement recommended by standards. On the other hand, as shown in chapter 8, friction in the contact between the stylus and the surface may change the results of measurement through contact forces.

The digital sampling techniques used in the stylus instrument have some practical difficulties. The mechanical design of a $x - y$ moving table with very high accuracy is the main problem. The relocation technique for the starting points at each row is also critical. Therefore, the photographic image mapping, as mentioned in chapter 3, is advantageous. In this case, the relationship between the image and the surface geometry has to be solved. No matter which method is used for 3D surface measurement, the sampling technique discussed in this thesis is applicable.

9.1.2 Scanning probe microscopes

One of the new profiling techniques is scanning probe microscopy. Their proliferation shows how important they have become. The most attractive aspect of this type of microscope is its molecular and atomic resolution. They are a powerful tool for the study of nanostructures in numerous areas of science and technology, such as the study of metallic surfaces, semiconductors, micro-machining and biology (Hamann *et al* 1988).

Following the previous evaluation in this thesis, a brief discussion of this profiling technique will now be presented.

Scanning patterns of the microscopes

In general, an image is formed by parallel tracking in the scanning tunnelling microscope. But the distance between rows is much longer than the ordinate sampling space within a row. Figure 9.1 shows a STM topography of $Au(001)$ (Kuk & Silverman, 1989). At the moment, the scan is only used to achieve a image rather than quantitative analysis. So, it is still in the early stage.

The $x-$, $y-$ and $z-$movements on the STM, for example, are mostly obtained using piezoelectric material with a good electronic loop control. It is easy to shorten the distance between the rows, if necessary (using the control system), so that it equals the sampling distance within a row. In this case, the sampling pattern would be the rectangular scheme, discussed in chapter 3. Similarly, it could also be triangular. However, it is not clear whether this is needed for image formation by the microscopes: the criterion for selecting the schemes should be studied. It will differ from the study of an engineering surface.

Scanning tip effects

The effects of the tip in conventional stylus instruments are geometrical and physical, or mechanical in this case. The mechanical effect influences the measuring results

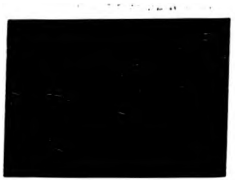


Figure 9.1: STM topography of $Au(001)$. Rows along the close-packed $(1\bar{1}0)$ are separated by $\sim$ 14 Å. Scales along the $(1\bar{1}0)$, (110), and (001) indicate 10, 10, and 2 Å, respectively.

by means of the contact force, with respect to the mechanical properties of surface materials. In a scanning microscope, the tip also plays these two roles. The tip in the STM, for instance, influences not only the resolution and shape of the scan, but also the measured electronic structures since it is one of the tunnelling electrodes (refer to chapter 5).

The physical role of the tip can be used for spectroscopic studies. The geometry of the tip determines the energy resolution. The uncertainty principle is normally used because of the small area covered by the tunnelling. This is given as

$$\Delta k \Delta x \sim 1, \tag{9.1}$$

where Δk is momentum uncertainty and Δx is the radius of the tip. The equation assumes that the radius of the tip and the lateral resolution are approximately equal. Kuk & Silverman (1989) have reviewed this aspect.

In the STM, the tunnelling effect is used to form an image of the geometric

information at the atomic level. Not many experiments using these microscopes have considered the geometry of the tip. By taking the FIM (Field Ion Microscope) image of the tunnelling tip before and after a STM scan, the character of the STM topography can be correlated to the tip structure. The corrugation amplitude was estimated theoretically (Tersoff & Hamann 1983) as

$$\Delta \propto \exp[-\beta(R+d)], \tag{9.2}$$

where $\beta \simeq \frac{1}{4}k^{-1}G^2$, R is the radius of curvature of the tip (approximated as a hemisphere), d is the gap distance, k^{-1} is the electron decay length in the vacuum, and G is the smallest surface reciprocal wave factor. This is presented in figure 9.2. It is pointed out by Kuk & Silverman (1989) that the corrugation of $Au(100)$,

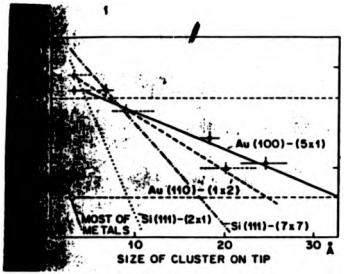


Figure 9.2: Dependence of measured corrugation on size of tip for $Au(001)-(5\times1)$ (solid line and filled circle), $Au(110)-(1\times2)$ (broken line and triangle), $Si(111)-(7\times7)$ (—.—), $Si(111)-(2\times1)$ (dotted line), and most metals without reconstruction (—..—). (Kuk & Silverman 1989).

$Au(110)$ and $Si(111)-(7\times7)$ surfaces could be detected by a tip of 20 Å, rather

than a single atom tip, with a STM noise level of 0.05 Å. However, most metal surfaces with small reconstruction unit cells require a very sharp tip and low STM noise level.

The tip geometry affects the resultant image through the tunnelling reaction. Several atoms on the tip take part in the phenomenon. If the relationship between the tunnelling current and the gap is certain, the purely geometric effect of the dimension of the tip, discussed in this thesis, is suitable and may also be so in a more complicated case, e.g. 3D effect and the physical reaction, etc.

9.2 Future development of measurement and instrumentation

9.2.1 The measuring object

The extension of precision measurement to the atomic level, and the increase in accuracy, raises some questions about the objects measured by present instruments. For instance, a question about surface measurement using current techniques is: what exactly is measured and where a surface is, as has been mentioned in chapter 2 (referring figure 2.10). In order to find the answer, further differential surface functions need to be taken. For a specific functional study, a particular surface measuring technique may be used to match the requirements. On the other hand, the techniques at the atomic level are not only the serving tools, but also become operators which could position and arrange atoms in a controllable manner. Therefore, the dream of the famous physicist, R. Feynman, could come true. That is, if we can arrange atoms and molecules in the way we want, "we will get an enormously greater range of properties that substances can have, and of different things that we can do" (Gillbert, 1990). This means that artificial materials could be made. The gap between physics and engineering could shorten.

9.2.2 Area assessment

Stylus profiling is a 2D measuring technique. It can be improved to 3D measurement using digital techniques. Area assessment requires a search for and identification of the spatial correlation of features (Whitehouse, 1991). In the conventional stylus instrument, the spatial calibration is determined by using the electronic clock. The bridge between an electronic clock and the spatial measurement is the traverse unit in the system. This does the temporal length calibration sampling. It is only needed if analogue devices are used. An alternative is spatial length calibration sampling. This method may typically be used to find crystalline structure, the order of molecules, isotropy, corrugations. The height and spatial information achieved by the instruments are schematically shown in the feature scale of size by Whitehouse (1991) below. It has been pointed out from this figure that the metrology gap is

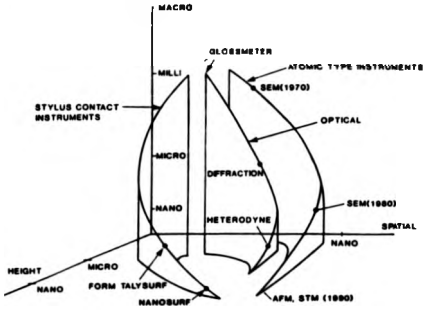


Figure 9.3: Feature scale of size.

the inability to quantify the visibility. This problem should be tackled in the future

development of measurement and instrumentation.

The use of the scanning electronic microscope (SEM) has increased enormously in recent years, but the data obtained is mainly for visualization or for mapping. Is this likely to continue? The reason for the question is that mapping can also be achieved with the stylus instruments, but it is time-consuming. However, it is inefficient if it only produces a visualization as does an electronic microscope, which operates very fast. On the other hand, it accesses less surface information and also indicates some other features, e.g. subsurface structures. This will not happen in the mapping by the stylus instrument. The stereo electronic microscope is a likely method for area access.

Scanning probe microscopes are more advanced than SEM in that quantitative information can be achieved through $x - y - z$ movement of the tip by keeping a constant gap. This has to be done to reduce the row spacings to the sampling distance. Since the full surface information is known, the problem left in area assessment is the area parameters. So, the metrology gap in figure 9.3 can be shortened by scanning probe microscopes.

9.2.3 Measurement integration

Building up the power of instruments is also an important task. One way is by increasing the speed of measurement. Some success has been achieved by optical techniques. An alternative is to improve the versatility of the instruments, or the integrating measurements. The combination of the form and the roughness measurement, for example, is obtained by one instrument - Talyform.

Another possibility is a unification of geometrical measurement with tests of the mechanical properties of a component. For instance, friction existing in the stylus surface profiling measurement cannot be neglected and separated, as has been pointed out at the end of chapter 8. An idea exists for measuring them together at the atomic level (see figure 9.4). A mirror is fixed at the top of the cantilever

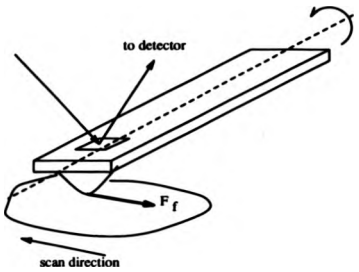


Figure 9.4: A scheme for combining measurement of the surface profile and the friction.

spring in an atomic force microscope. The incident and the reflecting beams are in the plane where the tip traverses, and perpendicular to the surface plane. The reflecting angle changes according to the tilting angle by the frictional force while the surface height is recorded.

9.2.4 The size of instruments

When the versatility of an instrument is increased, the size of the instrument may become a problem. The larger the size is, the larger the internal effects will be within the instrument. This can lead to less accuracy in measurement. In smaller sized instruments, the forces generated will be less important. Besides, the balance between external (environmental) and internal forces changes with size. For example, the original stylus instrument has a large mechanical loop of about 400 mm. The loop of the Talystep which has much higher accuracy is approximately one tenth of this. An AFM reduces the loop size by another one tenth. Very recently, a

group of people at Cornell University in USA, have even modified and miniaturized STMs onto a silicon chip. They "are trying to collapse the whole package - including springs, drives, and gears - into a silicon speck that could fit on the cut end of a human hair" (Amoto, 1991). This is still in the trial stage. But if it is successful, the instrument will be very small and highly integrated.

9.3 Conclusion

There are three main points in this thesis: (i) the two-dimensional measurement of three-dimensional objects; (ii) finite stylus size and (iii) stylus contact and the loading force. From these starting points, the main achievements have been: (i) the improvement of the 3D digital sampling techniques by introducing a new sampling scheme - the hexagonal grid and comparing it with other existing models; the new model improves data collection on the summit height density to 99% compared with the continuous case. The slope curvature information is also naturally obtained in one more direction. (ii) a further understanding of the geometric effect of the stylus by dividing the size and the shape. The 'deconvolution' mechanism of the profile from the trace is simulated successfully by the computer. Using this relationship, a method for measuring the dimension of the stylus is proposed by using a pair of 'checking surfaces'. From the practical point of view, a rig using a knife-edge to measure the dimension of the stylus is designed and experiments were carried out; (iii) an experiment was undertaken to assess the dynamic loading force of the stylus instrument during its function for the requirements of control this force. The dynamic variation of the stylus loading force is only 2.7% about its mean and it is negligible for the standards.

Appendix

Appendix A

Some basic concepts

A.1 Multivariate Gaussian distribution

If the m variables $\mathbf{X} = (x_1, \ldots, x_m)$ have a joint multivariate Gaussian distribution with mean μ and variance-covariance matrix $\mathbf{V}$, then this is denoted by: $\mathbf{X} \sim N(\mu, \mathbf{V})$. If there is a constant correlation coefficient between the random variable z_0 and each of the m random variables in $\mathbf{X}$, then the correlation matrix of $\mathbf{X}$ has a constant row (or column) sum. So,

$$\begin{bmatrix} z_0 \\ \mathbf{X} \end{bmatrix} \sim N\left[\mathbf{0}, \begin{pmatrix} 1 & d\mathbf{1}' \\ d\mathbf{1} & \mathbf{V} \end{pmatrix}\right], \tag{.3}$$

where $\mathbf{V}$ is the variance-covariance matrix of $\mathbf{X}$, is given as: $\mathbf{V} = E(\mathbf{X}\mathbf{X}')$, and d is the constant correlation coefficient. $\mathbf{1}' = (1, \ldots, 1)$.

In chapter 3, the central ordinate height is z_0. $\mathbf{X}$ is a vector of the height differences between z_0 and the heights $z_i (i = 1, \ldots, m)$, which are the same distance away from z_0. That is:

$$s_i = z_0 - z_i \ (i = 1, \ldots, m). \tag{.4}$$

Let the variance of all heights z is 1, i.e.

$$E(z_0^2) = E(z_i^2) = 1. \tag{.5}$$

Then, the variance of the height difference is as below:

$$E(s_i^2) = E\left[(z_0 - z_i)^2\right] \tag{.6}$$

$$= 2 - 2E(z_0 z_i). \tag{.7}$$

Let $E(z_0 z_i) = \rho$, which equals to correlation coefficient between z_0 and z_i. So,

$$E\left[\left(\frac{s_i}{\sqrt{2(1-\rho)}}\right)^2\right] = 1 \tag{.8}$$

and

$$d = E\left[z_0 \frac{s_i}{\sqrt{2(1-\rho)}}\right] = \frac{1}{\sqrt{2}}(1-\rho)^{\frac{1}{2}} \tag{.9}$$

So, the joint multivariate Gaussian distribution of z_0 and $s_i (i = 1, \ldots, m)$, which all have unit variances themselves is given as:

$$\left[z_0, 2(1-\rho)^{-\frac{1}{2}}\mathbf{S_m}\right] \sim N(0; \mathbf{V}) \tag{.10}$$

where $\mathbf{S_m} = (s_1, \ldots, s_m)$, and $\mathbf{V}$ is the $(m+1)$th variance-covariance matrix.

A.2 The truncated probability

The truncated probability with zero mean is denoted by $\Phi^{(m)}(0; \mathbf{V_m})$. If the truncation is from zero, the truncated probability is also known as orthant probability.

Results are given for truncated Gaussian random variables. Let $\mathbf{X}$ be truncated below, from 0, and the event $\mathbf{X} > 0$ be denoted by Y. Then

$$pr(Y) = \Phi^{(m)}(0; \mathbf{V}) \tag{.11}$$

where

$$\Phi^{(m)}(0; \mathbf{V}) = \int_0^\infty \int_0^\infty \cdots \int_0^\infty \phi^{(m)}(\mathbf{X}'; \mathbf{V}) dx_1 dx_2 \cdots dx_m \tag{.12}$$

A.3 The conditional multivariate Gaussian distribution

Let $\mathbf{X}$ be distributed normally according to $N(\mu; \Sigma)$ (with Σ nonsingular). Let us partition:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}; \begin{pmatrix} \mathbf{\Sigma_{11}} & \mathbf{\Sigma_{12}} \\ \mathbf{\Sigma_{21}} & \mathbf{\Sigma_{22}} \end{pmatrix}\right] \tag{.13}$$

where $\mathbf{\Sigma_{11}}, \mathbf{\Sigma_{22}}$ and $\mathbf{\Sigma_{12}} = \mathbf{\Sigma_{21}}$ are the covariance matrices of $\mathbf{X}^{(1)}$, of $\mathbf{X}^{(2)}$, and of $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$, respectively. Then, the conditional distribution of $\mathbf{X}^{(1)}$ given $\mathbf{X}^{(2)} = x^{(2)}$ is normal with mean as:

$$\mu_{1|2} = \mu^{(1)} + \mathbf{\Sigma_{12}\Sigma_{22}}^{-1}(x^{(2)} - \mu^{(2)}) \tag{.14}$$

and the covariance matrix as:

$$\mathbf{\Sigma_{1|2}} = \mathbf{\Sigma_{11}} - \mathbf{\Sigma_{12}\Sigma_{22}}^{-1}\mathbf{\Sigma_{21}}. \tag{.15}$$

(See Anderson 1984).

Appendix B

Appendix by Whitehouse & Phillips (1982)

Results are given for truncated Gaussian random variables. There is a constant correlation coefficient between the random variable $\mathcal{Z}_0$ and each of the m random variables $\mathbf{X}$, and the correlation matrix of $\mathbf{X}$ has a constant row (or column) sum. So

$$\begin{bmatrix} \mathcal{Z}_0 \\ \mathbf{X} \end{bmatrix} \sim N \left[\mathbf{0}, \begin{pmatrix} 1 & d\mathbf{1}' \\ d\mathbf{1} & \mathbf{V} \end{pmatrix} \right], \tag{.16}$$

where $\mathbf{V}$ is a correlation matrix such that

$$\mathbf{V}\mathbf{1} = \lambda \mathbf{1}, \tag{.17}$$

where $\lambda > md^2$ (and $\mathbf{1}' = (1, 1, \ldots, 1)$). From (.16)

$$U \equiv \mathbf{1}'\mathbf{X} \sim N[0, m\lambda], \tag{.18}$$

$$(\mathcal{Z}_0 | \mathbf{X} = \mathbf{x}) \sim (\mathcal{Z}_0 | U = u)$$

$$\sim N[(d/\lambda)u, 1 - md^2/\lambda] \tag{.19}$$

$$(\mathbf{X} | \mathcal{Z} = z_0) \sim N[dz_0\mathbf{1}, \mathbf{V} - d^2\mathbf{J}] \tag{.20}$$

(where $\mathbf{J} = \mathbf{1}\mathbf{1}'$). So the conditional distribution of $\mathcal{Z}_0$ given $\mathbf{X}$ only depends on $\mathbf{X}$ through U ($= \mathbf{1}'\mathbf{X}$).

Let $\mathbf{X}$ be truncated below, at $\mathbf{0}$, and the event $\{\mathbf{X} > \mathbf{0}\}$ be denoted by $\mathcal{Y}$. Then

$$\text{pr}\,(\mathcal{Y}) = \Phi^{(m)}(\mathbf{0}; \mathbf{V}), \tag{.21}$$

where

$$\Phi^m(\mathbf{y}'; \mathbf{V}) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \cdots \int_{-\infty}^{y_m} \phi^{(m)}(\mathbf{x}'; \mathbf{V}) \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_m \tag{.22}$$

(with $\Phi^{(m)}(d\mathbf{1}';\mathbf{V}) \equiv \Phi^{(m)}(d;\mathbf{V})$, $\Phi^{(1)}(y;\mathbf{V}) \equiv \Phi(y/\mathbf{V}^{\frac{1}{2}})$ and $\phi^{(1)}(y;\mathbf{V}) \equiv \phi(y/\mathbf{V}^{\frac{1}{2}})/\mathbf{V}^{\frac{1}{2}}$). Now as $\mathcal{Z}_0$ is not truncated,

$$(\mathcal{Z}_0|\mathbf{X} = x, \mathcal{Y}) \sim (\mathcal{Z}_0|\mathbf{X} = \mathbf{x}) \sim (\mathcal{Z}_0|u), \tag{.23}$$

as given by (.19), while the probability density function of $(\mathbf{X}|\mathcal{Y})$ is given by

$$f(\mathbf{x}|\mathcal{Y}) = \phi^{(m)}(\mathbf{x}';\mathbf{V})/\text{pr}\ (\mathcal{Y}) \tag{.24}$$

and, from (.20)

$$\text{pr}\ (\mathcal{Y}|\mathcal{Z} = z_0) = \Phi^{(m)}(dz_0\mathbf{1}';\mathbf{V} - d^2\mathbf{J}). \tag{.25}$$

Hence the probability density function of $\mathcal{Z}_0$ given $\mathcal{Y}$ is given by

$$f(z_0|\mathcal{Y}) = [\text{pr}\ (\mathcal{Y}|z_0)/\text{pr}\ (\mathcal{Y})]\phi(z_0), \tag{.26}$$

with the use of (.24) and (.25). Therefore from (.19) and (.23)

$$\begin{aligned} E(\mathcal{Z}_0|\mathcal{Y}) &= (d/\lambda)E(U|\mathcal{Y}) \\ &= md\phi(0)\,\Phi^{(m-1)}(0;\mathbf{B})/\text{pr}\,(\mathcal{Y}) \end{aligned} \tag{.27}$$

as Tallis (1961) has shown that using (.24)

$$\text{pr}(\mathcal{Y})\,E(\mathbf{X}|\mathcal{Y}) = \lambda\phi(0)\Phi^{(m-1)}(0;\mathbf{B})\mathbf{1}. \tag{.28}$$

where $\mathbf{B}$ is the variance-covariance matrix of $(X_1, X_2, \ldots, X_{m-1})$ given X_m. Hence from (.27) it is seen that the expectation of the non-truncated random variable $\mathcal{Z}_0$ given $\mathbf{X}$ (which is truncated) equals the product of m, the number of truncated random variables, d, the common correlation coefficient between $\mathcal{Z}_0$ and any of the truncated random variables, $\phi(0)$, and the ratio of two orthant probabilities, one of

which is for $m-1$ and the other for m random variables.

The probability density function of the conditional distribution of U given $\mathcal{Y}$ is given by

$$f(u|\mathcal{Y}) = [\text{pr}\ (\mathcal{Y}|u)/\text{pr}\ (\mathcal{Y})]f(u) \tag{.29}$$

where $f(u)$ is the probability density function of the untruncated distribution given by (.18). Because of the result given by (.23) the probability density functions of both the distributions of $(\mathcal{Z}_0, U|\mathcal{Y})$ and of $(U|\mathcal{Z}_0, \mathcal{Y})$ are given by the product of $\text{pr}(\mathcal{Y}|u)$ and the respective probability density function of the truncated distribution, suitably normalized. So they are of the same form as (.29).

Appendix C

The expected summit curvature of 7-point model

The curvature of 7-point model is defined as equation 3.6 in chapter 3:

$$C = \frac{1}{3h^2}\sum_{i=1}^{6} s_i, \; i = 1, 2, \ldots, 6 \tag{.30}$$

The distribution of a 7-point summit having a height z_0 conditional on a curvature $C = c$ is as below.

From equation 3.19, the distribution of height z_0 and curvature C is given as:

$$(z_0, C) \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \frac{2(1-\rho_1)}{h^2} \\ \frac{2(1-\rho_1)}{h^2} & \frac{2}{3h^4}(7 - 10\rho_1 + 2\rho_{\sqrt{3}} + \rho_2) \end{pmatrix}\right] \tag{.31}$$

The 7-point summit height distribution conditional on a curvature c is given by (Appendix A.3):

$$\mu_{1|2} = \frac{3h^2(1-\rho_1)}{7 - 10\rho_1 + 2\rho_{\sqrt{3}} + \rho_2} c \tag{.32}$$

$$\sigma_{1|2} = \frac{1 + 2\rho_1 + 2\rho_{\sqrt{3}} + \rho_2 - 6\rho_1^2}{7 - 10\rho_1 + 2\rho_{\sqrt{3}} + \rho_2}. \tag{.33}$$

Hence

$$(z_0|C = c) \sim N\left(\frac{3h^2(1-\rho_1)}{7 - 10\rho_1 + 2\rho_{\sqrt{3}} + \rho_2} c, \frac{1 + 2\rho_1 + 2\rho_{\sqrt{3}} + \rho_2 - 6\rho_1^2}{7 - 10\rho_1 + 2\rho_{\sqrt{3}} + \rho_2}\right) \tag{.34}$$

Appendix D

The expected summit height as $h \to \infty$ in the 7-point model

The expected summit height as h tends to infinity is given as:

$$\lim_{h\to\infty} E(z_0|T_7) = \lim_{h\to\infty} \frac{3\sqrt{1-\rho_1}\,\Phi^{(5)}(0;\mathbf{B_6})}{\sqrt{\pi}\,\Phi^{(6)}(0;\mathbf{V_6})}. \tag{.35}$$

As h tends to infinity, the variance matrices of the two truncated probabilities $\Phi^{(5)}(0;\mathbf{B_6})$ and $\Phi^{(6)}(0;\mathbf{V_6})$ become the matrices of equal correlation coefficients, i.e.

$$\mathbf{B_6} = \begin{bmatrix} 1 & & \frac{1}{3} \\ & \ddots & \\ \frac{1}{3} & & 1 \end{bmatrix} \tag{.36}$$

and

$$\mathbf{V_6} = \begin{bmatrix} 1 & & \frac{1}{2} \\ & \ddots & \\ \frac{1}{2} & & 1 \end{bmatrix} \tag{.37}$$

Bacon (1963) gives the resultant probability in this case in which all the correlation coefficients are equal to be denoted by $P_n(\rho)$. Hence,

$$P_n(\rho) \approx \left(\frac{1}{2}\right)^n \left[1 + \frac{n(n-1)}{\pi}\arcsin\rho + \frac{n(n-1)(n-2)(n-3)}{\pi^2} I_2^* \right. \tag{.38}$$

$$\left. + \frac{n(n-1)\cdots(n-5)}{\pi^3} I_3^* + \cdots \right] \tag{.39}$$

where

$$I_K^*(\rho) = \frac{(\arcsin\rho)^K}{K!(1+4\theta)(1+8\theta)\cdots[1+4(K-1)\theta]} \tag{.40}$$

and

$$\vartheta = \arcsin(\frac{\rho}{\pi}). \tag{.41}$$

Therefore, in the case of $\Phi^{(5)}(0;\mathbf{B_6})$, $\rho = \frac{1}{3}$, $n = 5$, $P_6(\frac{1}{3}) = 0.1143$. In the case of $\Phi^{(6)}(0;\mathbf{V_6})$: $\rho = \frac{1}{2}$, $n = 6$, $P_6(\frac{1}{2}) = 0.0064$. Substituting them into .35, thus

$$\lim_{A\to\infty} = 1.350 \tag{.42}$$

Appendix E

The reduction formulae by Plackett (1954) and David (1953)

E.1 Plackett reduction formula (1954)

Supposed that the vector $\mathbf{X}$ of n random variables with zero means and unit variances has a non-singular normal multivariate distribution with probability density function

$$\phi_n(x_1, x_2, \ldots, x_n; c_{11}, c_{12}, \ldots, c_{nn}) = (2\pi)^{-\frac{1}{2}n}|\mathbf{C}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{X}'\mathbf{C}\mathbf{X}\right) \tag{.43}$$

Here $\mathbf{C}$ is the inverse of the variance matrix of $\mathbf{X}$, and has elements $\{\rho_{ij}\}$. The integral is:

$$\Phi_n(a_1, a_2, \ldots, a_n; c_{11}, c_{12}, \ldots, c_{nn}) = \int_{a_1}^{\infty}\int_{a_2}^{\infty}\cdots\int_{a_n}^{\infty}\phi_n \mathrm{d}x_1\mathrm{d}x_2\ldots\mathrm{d}x_n. \tag{.44}$$

The partial differential equations of equation (.43) are represented by:

$$\frac{\partial\phi_n}{\partial\rho_{ij}} = \frac{\partial^2\phi_n}{\partial x_i \partial x_j} \tag{.45}$$

To differentiate Φ_n with respect to all $\{\rho_{ij}\}$, it can be obtained as:

$$\begin{aligned}\frac{\partial\Phi_n}{\partial\rho_{12}} &= \int_{a_1}^{\infty}\int_{a_2}^{\infty}\cdots\int_{a_n}^{\infty}\left(\frac{\partial\phi_n}{\partial\rho_{12}}\right)\mathrm{d}x_1\mathrm{d}x_2\ldots\mathrm{d}x_n \qquad (.46)\\ &= \int_{a_3}^{\infty}\int_{a_4}^{\infty}\cdots\int_{a_n}^{\infty}\{\int_{a_1}^{\infty}\int_{a_2}^{\infty}\left(\frac{\partial^2\phi_n}{\partial x_i\partial x_j}\right)\mathrm{d}x_1\mathrm{d}x_2\}\mathrm{d}x_3\mathrm{d}x_4\ldots\mathrm{d}x_n\\ &= \int_{a_3}^{\infty}\int_{a_4}^{\infty}\cdots\int_{a_n}^{\infty}\phi_n(a_1, a_2, x_3, x_4, \ldots, x_n; c_{11}, c_{12}, \ldots, c_{nn})\mathrm{d}x_3\mathrm{d}x_4\ldots\mathrm{d}x_n\end{aligned}$$

To bring this into more convenient form, equation (.47) becomes:

$$\frac{\partial \Phi_n}{\partial \rho_{12}} = (2\pi)^{-1}(1-\rho_{12}^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\cdot\frac{a_1^2 - 2\rho_{12}a_1a_2 + a_2^2}{1-\rho_{12}^2}\right) \times \Phi_{n-2}(a_3 - b_3, a_4 - b_4, \ldots, a_n - b_n; c_{33}, c_{34}, \ldots, c_{nn}), \quad (.47)$$

where

$$b_r = \frac{(\rho_{1r} - \rho_{2r}\rho_{12})a_1 + (\rho_{2r} - \rho_{1r}\rho_{12})a_2}{1-\rho_{12}^2}, (r = 3, 4, \ldots, n). \quad (.48)$$

E.2 David reduction formula (1953)

The multivariate normal integral when the truncation of the integral is from zero is denoted as:

$$P\{x_1 > 0, x_2 > 0, \ldots, x_n > 0\} \quad (.49)$$

where $x_1, x_2, \ldots, x_n$ are normal variables. Let the event E_i consist in the random variable x_i being positive. Thus

$$P\{x_1 > 0, x_2 > 0, \ldots, x_n\} = P\{E_1 E_2 \ldots E_n\} \quad (.50)$$

For n odd, we have

$$P\{E_1 E_2 \ldots E_n\} = \frac{1}{2}\left[1 - \sum_i P\{E_i\} + \sum_{i<j} P\{E_i E_j\} - \cdots \text{etc.}\right] \quad (.51)$$

the last term on the right-hand side being the sum of the joint probabilities of he E's taken $n-1$ at a time. For n even we get an identity.

Appendix F

BS1134: PART 1: 1988

Section three. Instrumentation

6 Stylus-type measuring instruments

6.1 Stylus

6.1.1 *Tip radius of the stylus.* The nominal value of the tip radius of the stylus shall be one of the following:

(a) 2 ± 0.5 µm.
(b) 5 ± 1 µm.
(c) 10 ± 2.5 µm.

See also appendix C.

6.1.2 *Stylus angle.* The nominal value of the stylus angle shall be one of the following:

(a) 1.57 radians (90°).
(b) 1.05 radians (60°).

6.1.3 *Static measuring force.* The static measuring force shall be sufficient to ensure continuous contact between the stylus and the surface being measured and shall be not greater than that given in table 2.

Table 2. Static measuring force of the stylus

Nominal tip radius of stylus	Maximum static measuring force at mean level of stylus	Maximum rate of change of measuring force
µm	mN	N/m
2 ± 0.5	0.7	35
5 ± 1	4.0	200
10 ± 2.5	16.0	800

6.2 Skid

6.2.1 *Skid dimensions.* If a skid is employed, its radius in the direction of the traverse shall be not less than 50 times the meter cut-off used.

If two simultaneously operative skids, as shown in figure 10, are used, their radii shall be not less than eight times the meter cut-off.

NOTE. Although the use of the skid may, when applied under suitable conditions, introduce no error of any great practical significance, external datum units should be used in all serious metrological work such as, for example, calibration procedures, and in the case of surfaces of limited area or requiring the use of cut-off values of 2.5 mm or greater.

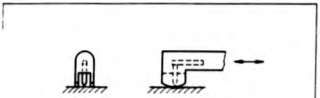


Figure 10. Stylus acting midway between two skids

6.2.2 *Skid surface roughness.* The surface roughness of the skid as determined by the ten point height of irregularities, R_z, shall be not greater than 0.1 µm when measured in the direction of traverse.

6.2.3 *Skid force.* The force exerted by the skid on the surface to be measured shall be not greater than 0.5 N.

6.3 Traverse

In profile instruments with predetermined or running evaluation lengths, the length shall depend on the meter cut-off value λ_s within the limits given in table 3.

Table 3. Evaluation lengths

Type of profile meter	Cut-off λ_s	Evaluation length	
		Min.	Max.
	mm	mm	mm
Predetermined evaluation length	0.08	0.4	2
	0.25	1.25	5
	0.8	2.4	8
	2.5	5	15
	8	16	40
Running evaluation length	0.25	2.5	16
	0.8	5	16

6.4 Values of vertical and horizontal magnification

The values of vertical and horizontal magnification for profile recording instruments shall be selected from the following series:

Vertical (V_v): 100, 200, 500, 1 000, 2 000, 5 000, 10 000, 20 000, 50 000, 100 000, 200 000, 500 000, 1 000 000.

Horizontal (V_h): 10, 20, 50, 100, 200, 500, 1 000, 2 000, 5 000, 10 000, 20 000, 50 000.

6.5 Transmission characteristics in the long wavelength

6.5.1 *Rate of attenuation.* The rate of attenuation shall be equivalent to that produced by two independent C-R networks of equal time constant in series. This describes a system in which the maximum slope of the transmission curve is 12 dB per octave and in which the phase shift at the 75 % cut-off λ_s is 90°.

The transmission coefficient of such a system shall be given by the equation:

$$\frac{\text{output to indicating system}}{\text{input to stylus}} = \frac{1}{\left(1 - j\,0.577\,\frac{\lambda}{\lambda_s}\right)^2}$$

where

$j = \sqrt{-1}$;

λ is the wavelength;

λ_s is the meter cut-off.

The effective cut-off wavelengths shall be taken at 75 % transmission. These are deemed to be equivalent to the sampling lengths in table 1.

NOTE. In a practical determination, the values of the transmission coefficients for the characteristics shown are measured relative to the flat part of the transmission curve (see figure 20).

6.5.2 *Cut-off values.* The cut-off values (in mm) to be used in instrument construction shall be selected from the following series:

0.08, 0.25, 0.8, 2.5, 8.0.

NOTE 1. A cut-off of 0.2 mm is found adequate for most of the finer surfaces.

NOTE 2. Nominal sinusoidal frequency response characteristics for a profile instrument are shown by the values given in table 4 (see also figure 20).

The permitted deviations from the nominal values of the transmission coefficients shall be as given in table 5, and graphically plotted in figure 21, and these allow the cut-off to be assessed at between 70 % and 80 % of maximum transmission.

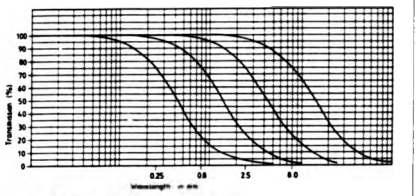


Figure 28. Profile instrument frequency response

Table 4. Nominal sinusoidal frequency response characteristics for a profile instrument

Wavelength	Percentage transmission			
	Cut-off 0.25 mm	Cut-off 0.8 mm	Cut-off 2.5 mm	Cut-off 8.0 mm
mm	%	%	%	%
0.025	98.7	–	–	–
0.05	98.7	–	–	–
0.08	98.7	99.7	–	–
0.10	94.9	99.5	–	–
0.25	75.0	96.8	99.7	–
0.5	42.9	88.5	99.7	–
0.8	22.7	75.0	94.7	99.7
1.0	15.8	65.8	94.9	99.5
2.5	2.9	23.5	75.0	96.8
5.0	0.75	7.1	42.9	88.5
8.0	–	2.9	22.7	75.0
10.0	–	1.8	15.8	65.8
25.0	–	–	2.9	23.5
50.0	–	–	0.75	7.1
80.0	–	–	–	2.9

NOTE. Because of practical difficulties in measurement at the very short wavelengths involved, the electrical transmission characteristic for 0.08 mm cut-off, although nominally of the same form as for the longer cut-off values, has not been measured.

7 Accuracy

7.1 Statement of basic error of calibration of R_a instruments

The basic error of profile instrument reading (as defined in 2.3.14) given within the cut-off by an instrument in optimum adjustment and use (see C.3), and expressed as a percentage of the designated value of the surface roughness parameter of an instrument calibration specimen complying with BS 6393, shall be determined from the formula:

$$\frac{p}{x} + q$$

where

x is the fraction of the range indicated by the instrument;

p is a percentage of full range;

q is a percentage of reading.

NOTE The permissible basic error of calibration thus expressed does not include the effect of deviations in the transmission characteristic which will be additional thereto.

7.2 Deviations of transmission coefficients

The permissible deviations of the amplitude transmission coefficient (see table 5 and figure 21) of a profile instrument from the nominal transmission coefficient shall be given by the equations:

$$\text{upper limit} = \frac{1.03}{1 + 0.29\left(\frac{\lambda}{\lambda_c}\right)^2}$$

$$\text{lower limit} = \frac{0.97}{1 + 0.39\left(\frac{\lambda}{\lambda_c}\right)^2}$$

where

λ is the wavelength;

λ_c is the meter cut-off.

Table 5. Upper and lower limits of transmission coefficients

Wavelength, λ / Cut-off, λ_c	Transmission coefficient			
	Lower limit		Upper limit	
	%	dB	%	dB
0.1	96.6	−0.30	102.7	+0.23
0.2	95.5	−0.40	101.8	+0.15
0.3	93.7	−0.56	100.4	+0.03
0.5	89.4	−1.57	98.0	−0.26
0.7	81.4	−1.78	90.2	−0.80
1.0	69.8	−3.13	79.8	−1.86
1.5	51.7	−5.74	62.3	−4.12
2.0	37.9	−8.43	47.7	−6.44
3.0	21.5	−13.5	28.5	−10.9
5.0	9.0	−20.9	12.5	−18.1
10.0	2.4	−32.3	3.4	−29.3

NOTE An explanation of the method divergence of the instrument reading (see 2.3.16) is given at appendix B and factors affecting the assessment of accuracy are explained at appendix C.

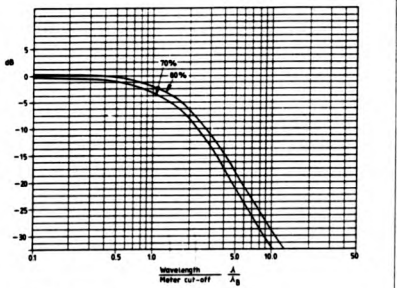
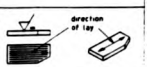
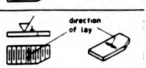


Figure 21. Permissible deviations of the transmission coefficient

Symbol	Interpretation		
=		Parallel to the plane of projection of the view in which the symbol is used	
⊥		Perpendicular to the plane of projection of the view in which the symbol is used	
X		Crossed in two slant directions relative to the plane of projection of the view in which the symbol is used	
M		Multi-directional	
C		Approximately circular relative to the centre of the surface to which the symbol is applied	
R		Approximately radial relative to the centre of the surface to which the symbol is applied	

NOTE. Should it be necessary to specify a direction of lay not clearly defined by these symbols, this may be done by a suitable note on the drawing.

Figure 22. Symbols for the direction of lay

Appendix G

The Specifications of the Talysurf-5

51.4 120 mm TRAVERSE UNIT

51.4.1 - Specification

Maximum length of traverse:	120 mm (4·72 in) between adjustable end stops
Traverse speed:	1 mm/s at Vh ÷ 5 and inversely proportional for other magnifications
Horizontal magnification (Vh):	÷ 2 (for setting up only), ÷ 5, ÷ 20, ÷ 50, ÷ 100
Skid force (adjustable):	0·5 N (50 gf) maximum for all pick-up attitudes
Accuracy of traverse (at stylus position):	
Over full traverse:	
Stylus up to 150 mm (6 in) from left-hand face of drive bracket:	Within two parallel planes having a spacing of: 0·5 µm (20 µin)
Stylus more than 150 mm (6 in) from left-hand face of drive bracket:	0·5 µm + (1 µm for every 50 mm extension beyond 150 mm) or 20 µin + (20 µin for every inch extension beyond 6 in)
Over any 60 mm traverse:	
Stylus up to 150 mm (6 in) from left-hand face of drive bracket:	Within two parallel planes having a spacing of: 0·25 µm (10 µin)
Stylus more than 150 mm (6 in) from left-hand face of drive bracket:	0·25 µm + (0·5 µm for every 50 mm extension beyond 150 mm) or 10 µin + (10 µin for every inch extension beyond 6 in)
Maximum penetration of stylus from drive bracket:	265 mm
Levelling range:	± 10° from mean position
Electrical supply:	..ow voltage, derived from processor
Overall dimensions (drive shaft fully retracted):	330 × 140 × 195 mm (13·25 × 5·5 × 7·7 in)
Weight:	Approx. 9·5 kg (21 lb)

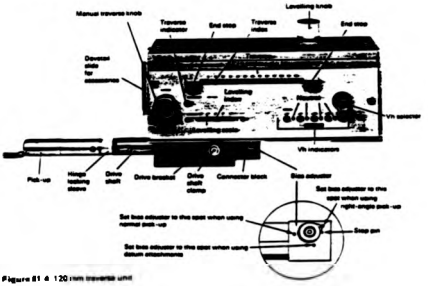


Figure 51.4 120 mm traverse unit

Appendix H

The Computer Programmes

The programes are arranged in the alphabatical order.

survid.m

```
```

form_trace.m

```
```

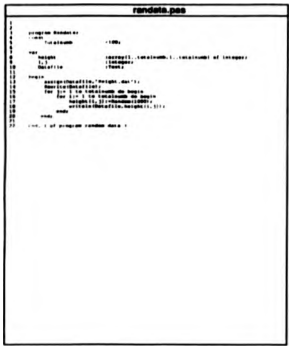
fdiamond_dr

```
```

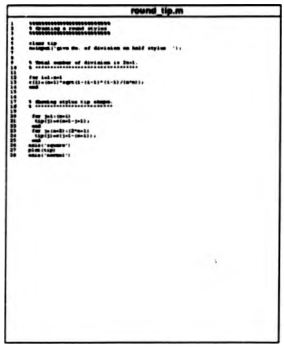
normal_tip.m

```
```

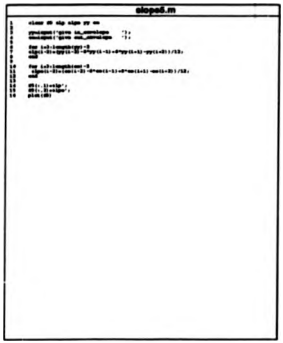
randata.pas



round_tip.m



slope5.m



triangle_tip.m


smithsort.pas

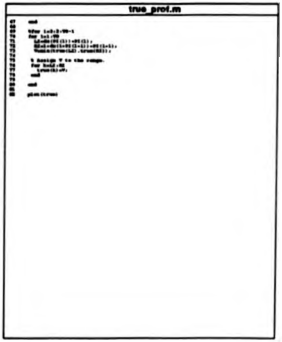

smithsort.pas


smithsort.pas


smithsort.pas


smithelt.pas


true_prof.m


true_prof.m

PAGINATION ERROR

Pg. 219

References

Abbott, E. J. & Goldschmidt, E. *1937* Surface quality. *Mech. Engng.* **59**, pp.813.

Anderson, T. W. *1984 An Introduction to multivariate statistical analysis* 2nd Ed., John-Wiley & Sons.

Arecchi, F. T., Bertani, D. & Ciliberto, S. *1979* A fast versatile optical profilometer. *Optical Communication* **31**, No. 3, pp.263.

Bacon, R. H. *1963* Approximations to multivariate normal orthant probabilities. *Ann. Math. Statist.* **34**, pp.191.

Binh, V. T. & Marien, J. *1988* Characterization of microtips for scanning tunnelling microscopy. *Surf. Sci.* **202**, pp.L539.

Binnig, G. & Rohrer, H. *1985* Scanning tunneling microscopy. *Surf. Sci.* **152.153**, pp.17.

Binnig, G. & Rohrer, H. *1987* Scanning tunneling microscopy — from birth to adolescence. *Rev. Mod. Phys.* **59**, No. 3, Pt. 1, pp.615.

Bowden, F. P. & Tabor, D. *1950 The Friction and Lubrication of Solids* Oxford at the Clarendon Press.

Bowden, F. P. *1967 Friction and lubrication* Methuen & Co. Ltd.

Bowen, D. K. *1990* SERC Advanced vacation school in instrumentation & Nanotechnology, 16th - 21st, September, Warwick University, U. K. and private communication.

Bowen, D. K., Chetwynd, D. G. & Schwarzenberger, D. R. *1990* Sub-nanometre displacements calibration using X-ray interferometry. *Meas. Sci. Technol.* **1** pp.107.

Cheng, M. C. *1968* The orthant probabilities of four Gaussian variables. *Ann. math. Statist.* **40**, pp.152.

Chetwynd, D. G. *1987* Selection of structural materials for precision devices. *Prec. Eng.* **9**, pp.3.

Chetwynd, D. G. *1988* A monolithic traverse for a stylus measuring instrument. *Micro-engineering Centre Report* ME67.

Chetwynd, D. G. *1989* Materials classification for fine mechanics. *Prec. Eng.* **11**, pp.203.

Chetwynd, D. G., Liu, X. P. & Smith, S. T. *1991* Signal fidelity and tracking force in stylus profilometry. *Inter. Conf. Metrology & Properties Eng. Surf.*, 10th - 12th, April 1991. Leicester Polytechnic, U. K.

Chicon, R. et al *1987* An algorithm for surface reconstruction in scanning tunneling microscope. *Surf. Sci.* **181**, pp.107.

Creath, K. & Wyant, J. C. *1988* Measurement of ultraprecision components using non-contact interferometry based instruments. *Proc. Inter. Conf. Ultraprecision Technology*, Aachen, May'88, pp.287.

David, F. N. *1959* A note on the evaluation of the multivariate normal integral.

Biometrika **40**, pp.458.

David, H. A. *1981 Order Statistics* John Wiley, New York.

David, S. T. *1989* Ion beam machining of diamond. *Industrial Diamond Review* **5**, pp.201.

DeVries, W. R. & Li, C. *1985* Algorithms to deconvolve stylus geometry from surface profile measurements. *J. Eng. Ind., Trans. ASME* **107**, pp.167.

Dupuy, O. *1967-68* High-precision optical profilometer for the study of micro-geometrical surface defects. *Proc. Instn. Mech. Engrs.* **182**, Pt. 3K. pp.255.

Elson, J. M. & Bennett, J. M. *1979* Relation between the angular dependence of scattering and statistical properties of optical surfaces. *JOSA* **69**, pp.1.

Evans, U. R. *1976 The Corrosion and Oxidation of Metals* Edward Arnold.

Fainman, Y., Lenz, E. & Shamir, J. *1982* Optical profilometer: a new method for high sensitivity and wide dynamic range. *Applied Optics* **21**, No. 17, pp.3200.

Farago, T. F. *1982 Handbook of Dimensional Measurement* Industrial Press Inc., New York.

Frampton, R. C. *1974* A theoretical study of the dynamics of pickups for the measurement of surface finish and roundness. *MR/2-5-1974*.

Gagnepain, J. J. & Roques-carmes, C. *1986* Fractal approach to two-dimensional

and three-dimensional surfaces roughness. *Wear* **109** pp.119.

Ghabrial, S. R. *1991* A methodology for relating surface processing in characterization to functional performance. *5th Intern. Conf. on Metrology & Properties of Engineering Surfaces*, 10-12, April, 1991. Leceister Polytechnic.

Gilbert, H. D. *1961 Miniaturization* ed. Reinhold, New York. pp.282.

Goget, M. *1990* Third-bodies in Tribology. *Wear* **136**, pp.29.

Greenwood, J. A. & Williamson, J. B. P. *1966* Contact of nominally flat surfaces. *Proc. R. Soc. Lond.* **A295**, pp.300.

Greenwood, J. A. & Williamson, J. B. P. *1977* Surface roughness effects in lubrication. *Proc. 4th Leds-Lyon Symp. on Tribology* ed. D. Dowson, C. M. Taylor, M. Godet & D. Berthe. London: Mechanical Engineering Publications.

Greenwood, J. A. *1984* A unified theory of surface roughness. *Proc. R. Soc. Lond. A*, **393**, pp.133.

Guerrero, J. L. & Black, J. T. *1972* Stylus tracer resolution and surface damage as determined by scanning electron microscopy. *J. Eng. Ind. Trans. ASME* Nov. pp.1087.

Halling, J. *1976 Introduction to tribology*, Wykeham Pub. Ltd.

Hansma, P. K., Elings, V. B., Marti, O. & Bracker, C. E. *1988* Scanning tunneling microscopy and atomic force microscopy: application to biology and technology.

Science **242** pp.209.

Hilmann, W., Kranz, O. & Eckolt, K. *1984* Reliability of roughness measurements using contact stylus instruments with particular reference to results of recent research at the Physikalisch Technische Bundesanstalt. *Wear* **97**, pp.27.

Homola, A. M., Israelachvili, J. M., McGuiggan, P. M. & Gee, M. L. *1990* Fundamental experimental studies in Tribology: The transition from 'interfacial' friction of undamaged molecularly smooth surfaces to 'normal' friction with wear. *Wear* **136**, pp.65.

Hummel, R. E. *1985* Electronic Properties of Materials Springer-Verlag.

Kaizuka, H. *1990* Application of the capacitor insertion method to STMs and reliability of STM images. *Joint Forum/ERATO Symposium on Nanotechnology*, 21 & 22 Aug. 1990, University of Warwick. U. K.

Kohno, T., Ozawat, N., Miyamoto, K. & Musha, T. *1985* Practical non-contact surface measuring instrument with one nanometre resolution. *Prec. Eng.* **7**, No. 4, pp.231.

Kuk, Y. & Silverman, P. J. *1989* Scanning tunneling microscope instrumentation. *Rev. Sci. Instrum.* **60** No. 2, pp.165.

Jing, Fang-Sheng, Hartman, A. W. & Hocken, R. J. *1987* Noncontacting optical probe. *Rev. Sci. Instrum.* **58**, No. 5, pp.864.

Jungle, J. & Whitehouse, D. J. *1970* An investigation of the shape and dimen-

sion of some diamond stylii. *J. Phys. E: Sci. Instrum.* **3**, pp.437.

Li, M., Phillips, M. J. & Whitehouse, D. J. *1989* Extension of two-dimensional sampling theory. *J. Phys. A: Math, Gen* **22**, pp.5053.

Lindsey, K., Smith, S. T. & Robbie, C. J. *1988* Sub-nanometre surface texture and profile measurement with 'Nanosurf-2'. *Annal CIRP*, **37**, No.1, pp.519.

Liu, X. P., Smith, S. T. & Chetwynd, D G. *1991* Friction forces between a diamond stylus and specimens at the low load. *The poster on the 6th IPES, 1991*, Braunschweig. Germeny.

Longuet-Higgins, M. S. *1957a* The statistical analysis of a random, moving surface. *Phil. Trans. R. Soc. Lond. A* **249**, pp.321.

Longuet-Higgins, M. S. *1957b* Statistical properties of an isotropic random surface. *Phil. Trans. R. Soc. Lond. A* **250**, pp.157.

Majumdar, A. & Tien, C. L. *1990* Fractal characterization and simulation of rough surfaces. *Wear* **136**, pp.313.

McCool, J. I. *1984* Assessing the effect of stylus tip radius and flight on the surface topography measurement. *Trans. ASME* **106**, pp.202.

Merserean, R. M. *1979* The processing of hexagonally sampling two dimensional signals. *Proc. IEEE* **67**, No. 6, pp.930.

Mignot, J. & Gorecki, C. *1989* Measurement of surface roughness: comparison be-

tween a defect-of-focus optical technique and the classical stylus technique. *Wear* **87**, pp.39.

Miyamoto, I *1987* Ultra fine finishing of diamond tools by ion beams. *Prec. Engng.* **9**, No. 2, pp.71.

Montgomery, D. B. *1969 Solenoid Magnet Design.* Wiley Interscience.

Morris, A. S. *1988 Principles of Measurement and Instrumentation* Prentice Hall.

Musselman, I. H. *et al 1990* Fabrication of tips with controlled geometry for scanning tunneling microscopy. *Prec. Eng.* **12**, No. 1, pp.3.

Müller, E. W. & Tsong, T. T. *1969 Field ion microscopy principle and applications*, American Elsevier Publishing Company, Inc. New York, pp.119.

Nakagiri, N & Kaizuk, H *1990* Simulation of STM images ad work function for rough surfaces. *Japan. J of Appli. Phys.* **29**, No. 4, pp. 744.

Nakamura, T. *1966* On deformation of surface roughness curves caused by finite radius of stylus tip and tilting of stylus holder arm. *Bull. Japan. Soc. Proc. Engg.* **1**, No.4, pp.240.

Nayak, P. R. *1971* Random process model of rough surfaces. *Trans. ASME: J. lub. Tech.* **93**, pp.398

Norton, A. E. *1942 Lubrication* ed. by Muenger, J. R. and pub. by McGraw-Hill Book Co. Inc., New York and London.

O'Connor, R. F. *1991* The use of a complete surface profile description to investigate the cause and effect of surface features. *5th Intern. Conf. on Metrology & Properties of Engineering Surfaces*, 10-12, April, 1991. Leceister Polytechnic.

Page, S. F. *1948 Fine Surface Finish* pub. by Chapman & Hall ltd., U. K.

Pearson, E. B. *1957 Technology of Instrumentation* The English Universities Press Ltd.

Peklenik, J. *1967/68* New developments in surface characterization and measurements by means of random process analysis. *Proc. Instn. Mech. Engrs.*, **182**, Pt 3K, pp.108.

Peres, N. J. C. *1959* Geometrical considerations arising from the use of square wave calibrations standards of stylus finish. *Aust. J. Appl. Phys.* **4**, pp.380.

Phillips, M. J. *1984* Truncated random variables with application to random surfaces. *Math. Meth. in Appl. Sci.* **6**, pp.248.

Preston, K. *1979* Basics of Cellular logic with some applications in medical image processing. *Proc. IEEE* **67**, pp.826.

Quiney, R. G., Austin, F. R. & Sargent, L. B. Jr. *1967* The measurement of surface roughness and profiles on metals. *ASLM Trans.* **10**, pp.193.

Radhakrishnan, V. *1970* Effect of stylus radius on the roughness values measured with tracing stylus instruments. *Wear* **16**, pp.325.

Reason, R. E., Hopkins, M. R. & Garrod, R. I. *1944* Report on the measurement of surface finish by stylus methods. *Rank Taylor Hobson, Leicester, U.K. Report.*

Reason, R. E. *1951* Surface finish. *Australasian Engr.* **44**, pp.48.

Reason, R. E. *1954* The trend of surface measurement. *J. Inst. Prod. Engrs.* **33**, pp.263.

Reason, R. E. *1956* Significance and measurement of surface part 2: how transducers affect instrument performance; how to select proper cut-off values. *Grinding & Finishing* **2**, pp.32.

Rice, S. O. *1945* Mathematical analysis of random noise. *Bell. Sys. Tech. J.* **24**, pp.46.

Roques-Carmes, C., Wehbi, D., Quiniou, J. F. & Tricot, C. *1988* Modeling engineering surfaces and evaluating their non-integer dimension for application in material science. *Surface Topography* **1**, pp.435.

Sawatari, T. & Zipin, R. B. *1979* Optical profile transducer. *Optical Engineering* **18**, No. 2, pp.222.

Schlesinger, G. *1942 Surface Finish - report of the Research Department of the Institution of Production Engineers* pub. by the institution, London.

Schorsch, H. *Gütebestimmung an Technischen Oberflächen* pub. by Wissenschaftliche Verlagsgesellschaft MBH. Stuttgart, Germany.

Scott, P. J. *1986* Surface Metrology, a new philosophical approach. *Wear*, **109**, pp.267.

Sherrington, I. & Smith, E. H. *1986* A quantitative study of the influence of stylus shape and load on the fidelity of data recorded by stylus instruments. *2nd National Conf. Production research, 10-12 Sep. 1986.* Napier College, Edinburgh, UK, pp.762.

Sherrington, I. & Smith, E. H. *1988* Modern measurement techniques in surface metrology: Part I: Stylus instruments, electronic microscopy and non-optical comparators. Part II: Optical instruments. *Wear* **123**, pp.271.

Smith, S. T., Chetwynd, D. G. & Bowen, D. K. *1987* Design and assessment of monolithic high precision translation mechanisms. *J. Phys. E: Instrum.* **20**, pp.977.

Smith, S. T. *1989* The design and construction of the LVDT. (unpublished)

Smith, S. T. *1990* Private communication.

Smith, S. T. & Chetwynd, D. G. *1990* An optimized magnet-coil force actuator and its application to linear spring mechanisms. *Proc. Inst. Mech. Eng.* C (in process).

Sommargren, G. E. *1981* Optical heterodyne profilometry. *Appl. Opt.*, **20**, No.4, pp.610.

Staunton, R. C. *1989* Hexagonal image sampling: a practical proposition. *Image*

and Vision Computing, **7**, pp.162.

Stedman, M. *1987a* Basis for comparing the performance of surface-measuring machines. *Precis. Eng.* **9**, pp.149.

Stedman, M. *1987b* Mapping the performance of surface-measuring instruments. *Proc. SPIE* **803**, pp.138.

Stedman, M. *1988* Limits of topographic measurement by the scanning tunneling and atomic force microscopes. *J. Microscopy* **152**, Pt. 3, pp.611.

Steven, D. M. G. (1991) *private communication*, Rank Taylor Hobson Ltd. U. K.

Sullivan, P. *et al 1991* An application of a three-dimensional surface analysis system to the prediction of asperity interaction in metallic contacts. *5th Intern. Conf. on Metrology & Properties of Engineering Surfaces*, 10-12, April, 1991. Leceister Polytechnic.

Tallian, T. E., Chiu, Y. P., Huttenlocher, D. F., Kamanshine, J. A., Sibley, L. B. & Sindlinger, N. E. *1964* Lubrication films in rolling contact of rough surfaces. *Tran. ASLE* **7**, pp.109.

Tersoff, J. & Hamann, D. R. *1983* Theory and application for the scanning tunnelling microscope. *Phys. Rev. Lett.* **50**, No. 25, pp.1998.

Thomas, T. R. *1982 Rough Surfaces* Longman.

Thwaite, E. G. *1977* The roughness of surfaces. *The Australian Physicist* **November**, pp.170.

Tsukada, T. & Anno, Y. *1974* An evaluation of machined surfaces topography. (1st report, on slope distribution of projections). *Bull. Jap. Soc. Prec. Eng.* **8**, pp.141.

Tsukada, T. & Anno, Y. *1975* An evaluation of machined surfaces topography. (2nd report, on statistics of surface asperity heights). *Bull. Jap. Soc. Prec. Eng.* **9**, pp.1.

Vorburger, T. V. et al *1979* Measurements of stylus radii. *Wear* **57**, pp.39.

Weingraber, H. V. & Abou-Aly, M. *1989 Handbuch Technische Oberflächen* . Published by Vieweg. (in Germany)

Whitehouse, D. J. & Reason, R. E. *1965 The equation of the mean line of surface texture found by an electric wave filter.* Rank Taylor Hobson.

Whitehouse, D. J. & Archard, J. F. *1969* The properties of random surfaces in contact *ASME conference on Surface mechanics, Los Angeles* Nov. 1969, pp.36.

Whitehouse, D. J. & Archard, J. F. *1970* The properties of random surfaces significance in their contact. *Proc. R. Soc. Lond. A*, **316**, pp.97.

Whitehouse, D. J. *1974* Theoretical Analysis of stylus integration. *Ann. CIRP* **23**, pp.181.

Whitehouse, D. J. *1978* Surfaces - a link between manufacture and functions. *Proc.*

Instrn. Mech. Engrs., **192**, pp.179.

Whitehouse, D. J. & Phillips, M. J. *1978* Discrete properties of random surfaces. *Phil. Trans. R. Soc. Lond.* **A290**, pp.267.

Whitehouse, D. J. *1982* The parameter rash - is there a cure? *Wear*, **83**, pp.75.

Whitehouse, D. J. & Phillips, M. J. *1982* Two-dimensional discrete properties of random surfaces. *Phil. Trans. R. Soc. Lond.* **A305**, pp.441.

Whitehouse, D. J. *1983* The generation of two dimensional random surface having a specified function. *Ann. CIRP* **32** Pt. 1, pp.495.

Whitehouse, D. J. & Phillips, M. J. *1985* Sampling in a two dimensional plane. *J. Phys. A: Math. Gen.*, **18**, pp.2465.

Whitehouse, D. J. *1987* Surface Metrology Instrumentation. *J. Phys. E: Sci. Instrum.* **20**, pp.1145.

Whitehouse, D. J. *1988* Friction and surface measurement. *Surface Topography*, **1**, pp.427.

Whitehouse, D.J. *1989a* Stylus force analysis - Assessment of dynamic forces. *Manufacturers Meeting May 10th, Coventry, U.K.*

Whitehouse, D.J. *1989b* A revised philosophy of surface measuring systems. *Proc. Instn. Mech. Engrs.* **202**, No. C3, pp.169.

Whitehouse, D.J. *1991* Nanotechnology instruments. *Meas. & Control* **24**, No.2, pp.37.

Wickramasinghe, H. K. *1989* Scanned-probe microscopes. *Scientific American* Oct. pp.74.

Williamson, D. E. *1967* Tracer-point sharpness as affecting roughness measurements. *Trans. ASLE* **69**, pp.319.

Young, R., Ward, J. & Scire, F. *1972* The topografiner: an instrument for measuring surface micro-topography. *Rev. Sci. Instrum.* **43**, No. 7, pp.999.

Zhejiang University *1979 Engineering Mathematics: Probability and Statistics.* People's Education Publisher. (in Chinese)