THE UNIVERSITY OF
WARWICK

University of Warwick institutional repository: http://go.warwick.ac.uk/wrap

A Thesis Submitted for the Degree of PhD at the University of Warwick

http://go.warwick.ac.uk/wrap/56039

This thesis is made available online and is protected by original copyright.

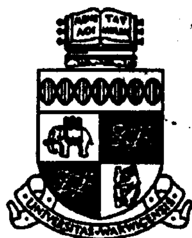
Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.

FIDELITY STUDY IN SURFACE MEASUREMENTS IN NANOMETRE METROLOGY

X. Liu

Ph. D



University of Warwick

To

my parents,

my husband, and my daughter.

FIDELITY STUDY IN SURFACE MEASUREMENTS IN NANOMETRE METROLOGY

By

Xianping Liu, BSc., MSc.

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

Centre for Nanotechnology and Microengineering
Department of Engineering
University of Warwick, Coventry, U. K.

November, 1994

CONTENTS

1. SURFACE MEASUREMENT TECHNIQUES IN NANOTECHNOLOGY 1

1.1 INTRODUCTION 1

1.2 ADVANCES OF SURFACE MEASURING TECHNIQUES 3

1.2.1 From a razor blade to nano-stylus instruments 3

1.2.2 Optical profiling 4

1.2.3 Scanning at atomic level 7

1.3 TECHNIQUE COMPARISON — FIDELITY OF MEASUREMENT 11

1.3.1 Interaction between the surface and the sensing probes 12

1.3.2 Range and resolution 15

1.3.3 Dynamic response 16

1.4 CHARACTERISATION AND CALIBRATION OF INSTRUMENTS 16

1.5 THE LAYOUT OF THE THESIS 19

2. CHARACTERISTICS OF ENGINEERING SURFACES 25

2.1 SURFACE AND ITS DESCRIPTION 25

2.1.1 Concept 25

2.1.2 Solid surfaces in air 26

2.2 ENGINEERING SURFACES 27

2.2.1 The nature of surface 27

2.2.2 Topographies of surfaces 27

2.2.3 Properties of surfaces 28

2.3 TECHNIQUES FOR PREPARING SMOOTH SURFACES 29

2.4 MEASUREMENT OF SURFACES 31

2.4.1 Surface parameters 31

2.4.2 Statistical functions 34

2.4.3 Fractal approach 35

2.4.4 Profiling and imaging 37

3. ANALYSIS OF STYLUS SYSTEM 42

3.1 INTRODUCTION 42

3.2 MATHEMATICAL MODEL OF THE STYLUS SYSTEM 43

3.3 ANALYSIS AND MEASUREMENT OF THE CHARACTERISTIC PARAMETERS 44

3.3.1 The spring rate 44

3.3.2 The damping coefficient 46

3.3.3 System resonance and effective mass 48

3.4 CONTACT FORCES IN STYLUS BASED SYSTEMS 49

3.4.1 Static force 49

3.4.2 Dynamic force 50

3.4.3 Estimation of the magnitude of the dynamic forces 50

3.4.4 The magnitude of the reaction force 53

3.5 EFFECTS OF THE GEOMETRY OF THE STYLUS TIP ON THE MEASURED PROFILE 54

3.5.1 Formation of a stylus trace 55

3.5.2 Effects on the lateral resolution 57

3.5.3 Effects on the parameter measurements 57

3.5.4 Checking of the stylus tip 60

3.5.5 Reconstruction of surface profiles 61

4. EFFECTS OF TRACKING FORCE ON STYLUS MEASUREMENT 72

4.1 VARIABLE FORCE ACTUATOR 72

4.1.1 Principle 73

4.1.2 The construction 74

4.1.3 Design of constant current drive 75

4.2 MODIFICATION OF THE TALYSURF 5 78

4.2.1 Calibration of the variable force transducer 78

4.2.2 Measurement of the spring rate 79

4.2.3 Change of the effective mass and resonance 79

4.2.4 Calibration of the pick up 80

4.3 EXPERIMENTATION 80

4.3.1 Sample trigger 'flag' *80*

4.3.2 Specimen preparation and measurement *81*

4.4 RESULTS AND DISCUSSIONS 82

4.4.1 Profile variations at a nominal force *82*

4.4.2 Profile variation at different tracking forces *83*

4.4.3 Influence of tracking speed *84*

4.4.4 SEM view of surface damage *85*

4.4.5 Conclusions *86*

5. FRICTION FORCES BETWEEN A STYLUS TIP AND SPECIMENS AT LOW LOADS 104

5.1 INTRODUCTION 104

5.2 LITERATURE REVIEW 105

5.3 ANALYSIS OF MICROSCOPIC FRICTION 108

5.3.1 Contact between a tip and surfaces *108*

5.3.2 Brief review of friction models *110*

5.4 INSTRUMENTATION 111

5.4.1 Apparatus for measuring friction *111*

5.4.2 Calibration *112*

5.4.3 Other sources of error *113*

5.5 INDENTATION TESTS ON SPECIMENS 114

5.6 RESULTS AND DISCUSSION 116

5.6.1 Friction measurement *116*

5.6.2 Friction coefficients *117*

5.6.3 Effect of surface finish on friction *119*

5.6.4 The relationship between the friction and surface finish profiles *120*

5.6 FRICTION OF A STEEL BALL ON SPECIMENS 122

5.7 CONCLUSIONS 125

6. IMPROVEMENT ON FIDELITY OF SURFACE MEASUREMENT BY ACTIVE DAMPING CONTROL....139

6.1 INTRODUCTION....139

6.2 DYNAMIC RESPONSE OF A STYLUS SYSTEM....140

6.2.1 System transfer function....140

6.2.2 Spectra of a specimen traversed at different speeds....141

6.2.3 Response to a given displacement....142

6.2.4 Response to a variable force....142

6.3 STATISTICAL APPROACH TO SYSTEM OPTIMISATION....143

6.4 DAMPING FUNCTIONS....145

6.4.1 Viscous damping....145

6.4.2 Dry friction damping....146

6.4.3 Hysteresis....146

6.5 EXPERIMENTAL APPROACH TO SYSTEM OPTIMISATION....147

6.5.1 Construction of an active damping controller....147

6.5.2 Analysis of the velocity sensor....148

6.5.3 Circuitry of the damping controller....149

6.5.4 System response and damping ratio calibration....149

6.6 PERFORMANCE EVALUATION OF THE SYSTEM....152

6.6.1 Effects of damping on dynamic forces....152

6.6.2 Effects of damping on signal fidelity....153

6.6.3 Effects of damping on frictional force....156

6.7 INCREASING RESONANT FREQUENCY....158

6.8 CONCLUSION....159

7. CONSTRUCTION AND ASSESSMENT OF SCANNING TUNNELLING PROFILOMETERS FOR METROLOGICAL APPLICATIONS 181

7.1 INTRODUCTION 181

7.2 ANALYSIS OF STM 182

7.2.1 Principles of STM *182*

7.2.2 Motion control *183*

7.2.3 Vibration isolation *184*

7.2.4 Hysteresis and creep *186*

7.3 DESIGN OF A 2-D STM BASED ON A SPRING TRANSLATION (STM_1) 187

7.3.1 Spring flexure driven by electromagnetic force *187*

7.3.2 Calibration *189*

7.4 A 2-D STM BASED ON A DPT AND A PRECISION SLIDEWAY (STM_2) 190

7.4.1 Close loop compensation of PZT *190*

7.4.2 Nanostep slideway *191*

7.5 INSTRUMENTATION 192

7.5.1 Pre-amplifier *192*

7.5.2 Linearisation circuits *193*

7.5.3 PID controller *194*

7.5.4 Actuator drives *194*

7.6 ANALYSIS OF CONTROL LOOP 195

7.6.1 Mathematical analysis *195*

7.6.2 Steady state response *196*

7.6.3 Parameter tuning and transient response *197*

7.7 EXPERIMENTATION 199

7.7.1 Manufacture of scanning tips *199*

7.7.2 Preparation of specimens *200*

7.8 RESULTS 200

7.8.1 Scanning profiles of gold coated silicon by STM_1 ....*200*

7.8.2 Performance evaluation of STM_2 .... *201*

7.9 Digital Filtering 204

7.10 Summary 206

8 METROLOGICAL APPLICATIONS OF STM **225**

8.1 Problems in Interpretation of STM Data 225

8.2 The characteristic of STM in air 226

8.2.1 Relationship of S-V *226*

8.2.2 Relationship of S-I *227*

8.3 Mechanical Contact Model 228

8.4 Effect of the Tunnelling Tip 231

8.6 Conclusions 232

9. CONCLUSIONS AND SUGGESTIONS FOR FUTURE DEVELOPMENT **235**

9.1 Conclusions 235

9.1.1 Fidelity aspects in contact measurements *235*

9.1.2 Fidelity aspects in non-contact measurement *238*

9.2 Suggestions for Future Development 239

Appendix **241**

A List of Publications 241

B List of Programs 242

References **252**

LIST OF FIGURES

Fig. 1.1 Simplified diagram of a high sensitive LVDT 20

Fig. 1.2 Schematic diagram of an optical follower 21

Fig. 1.3 Schematic diagram of the NPL/CUPE polarisation interferometer focus configuration. O and E correspond to the 'ordinary' and 'extraordinary' components of plane polarisation. 22

Fig. 1.4 Principle of focus detection by critical angle of total reflection 22

Fig. 1.5 Schematic diagram of the scanning tunnelling microscope 23

Fig. 1.6 Estimated range and resolution limits in both the vertical and lateral directions for stylus instruments, optical interferometers, and STM. 23

Fig. 1.7 Traceability ladder of the length 24

Fig. 2.1 Schematic showing typical surface and subsurface microstructures present in metal. 38

Fig. 2.2 Schematic showing selected types of surface deviations relative to ideal surface. 39

Fig. 2.3 Surface profile showing the ISO definition for the peak spacing, S_m. 39

Fig. 2.4 Profiles constructed from fractals: (a) D=1, (b) D=1.28, (c) D=1.5, and (d) D=1.8. The vertical axis is in an arbitrary unit and horizontal axis is the sample points. 40

Fig. 3.1. Schematic of the stylus method; (a) the basic assemblage of a stylus system and (b) the model of the stylus system. 64

Fig. 3.2 The variation about the nominal value of 0.7 mN of the stylus spring rate against displacement at a magnification setting of 2000. 65

Fig. 3.3 Dynamic response of the Talysurf 5 66

Fig. 3.4 Kinematic motion of the stylus. 67

Fig. 3.5 Path of the stylus after separation. 67

Fig. 3.6 Relationship of the dynamic force components 68

Fig. 3.7 (a) The relationship between the surface and the stylus trace and (b) the formation of the contact path. 69

Fig. 3.8 Schematic showing the path traced by a stylus tip traversing over a small, sharp step 70

Fig. 3.9 Changes in the profile form when traced with large stylus radii 70

Fig. 3.10 Height distributions of the original profile and the trace with a large tip radius of 100 μm, bars represent the original. 71

Fig. 3.11 Schematic diagram of the method for characterising a stylus tip. 71

Fig. 4.1 The schematic diagram of the magnet-coil force actuator. 87

Fig. 4.2 Magnetic field gradient against position along axis of coil with different values of exciting current. 87

Fig. 4.3 Experimental arrangement for the variable tracking force using a magnet-coil force transducer. 88

Fig. 4.4 Schematic of current drive 1 89

Fig. 4.5 Diagram of simplified current drive loop 89

fig. 4.6 Improved circuitry of the load cell 90

Fig. 4.7 Calibration of the load cell. 91

Fig. 4.8 Calibration of the variable force transducer. 91

Fig. 4.9 Repeated traces on a steel specimen at 0.69 mN force at a traverse speed of 0.1 mm s^{-1} 92

Fig. 4.10 Repeat stylus traces on a copper specimen at loads ranging vertical downwards from 0.39 to 2.8 mN. 97

*Fig. 4.11(a) Parameter values from repeating tracks at increasing stylus force, normalised to their values at 0.7 mN on a copper specimen. Stylus traverse speed is 1 mm s^{-1} in the top graph and 0.25 and 0.1 mm s^{-1} respectively in the two below. x — R_a, + — R_q, o — Δq, * — Pc.* 98

*Fig. 4.11(b) Parameter values from repeating tracks at increasing stylus force, normalised to their values at 0.7 mN on a steel specimen. Stylus traverse speed is 1 mm s^{-1} in the top graph and 0.25 and 0.1 mm s^{-1} respectively in the two below. x — R_a, + — R_q, o — Δq, * — Pc.* 99

*Fig. 4.11(c) Parameter values from repeating tracks at increasing stylus force, normalised to their values at 0.7 mN on a brass specimen. Stylus traverse speed is 1 mm s^{-1} in the top graph and 0.25 and 0.1 mm s^{-1} respectively in the two below. x — R_a, + — R_q, o — Δq, * — Pc.* 100

*Fig. 4.11(d) Parameter values from repeating tracks at increasing stylus force, normalised to their values at 0.7 mN on an aluminium specimen. Stylus traverse speed is 1 mm s^{-1} in the top graph and 0.25 and 0.1 mm s^{-1} respectively in the two below. x — R_a, + — R_q, o — Δq, * — Pc.* 101

Fig. 4.12 Profiles of a copper specimen traversed at the fastest speed of 1.0 mm s^{-1} under three different loading forces of 0.39, 0.69 and 1.67 mN. 102

Fig. 4.13 Scanning electron micrograph of stylus tracks on a copper surface. The stylus runs from left to right. From the lowest one, loads were 3.3, 2.4, 1.7, 1.3 and 1.1 mN, respectively 102

Fig. 4.14 Scanning electron micrograph of stylus tracks on a steel surface, showing scuff marks. 103

Fig. 4.15 Scanning electron micrograph of the stylus tip prior to the start of the test series. 103

Fig. 5.1 (a) Elastic contact: contact area less than the geometrical area and (b) plastic contact: contact area greater than the geometrical area 126

Fig. 5.2 (a) Photograph of the instrumentation for measurement of frictional forces. 127

Fig. 5.2 (b) Schematic representation of the instrumentation for measurement of frictional forces during the traverse of a diamond stylus over a surface at low loads. *128*

Fig. 5.3 Characteristic of the friction force transducer, (a) output from force gauge for dead weight loading and (b) output from the same gauge when loaded by the stylus. *129*

*Fig. 5.4 Load vs. depth for steel, copper and aluminium surfaces. * — steel, + — copper, and o — aluminium.* *130*

*Fig. 5.5 Coefficients of friction vs. stylus load: — ×, ---- o, ····· * represent values measured as sliding velocities of 1.0, 0.25 and 0.1 mm s^{-1} respectively; (a) mild steel, (b) copper, (c) brass, (d) aluminium, (e) silicon* *131*

Fig. 5.5 (continued) (c) brass, (d) aluminium, (e) silicon *132*

Fig. 5.6 Plots of surface profile and frictional forces as the stylus traverses a complete cycle on a copper specimen at a nominal load of 5.9 mN, (a) surface profile and (b) tractive force parallel to a traverse (frictional drag). *133*

Fig. 5.7 Friction drag (dashed) and surface profiles (solid) from the plot of Fig. 5.6. Plotting units are chosen so that force and profile can be read from the same scale: (a) forward measurement, (b) reverse measurement. *134*

Fig. 5.8 Cross-correlations of surface profile and frictional drag data of Fig. 5.7 (mean line removed): (a) forward traverse, (b) reverse traverse. Horizontal axis in sample point spacing. *135*

Fig. 5.9 Force diagram for the contact of a stylus on the surface of a specimen. *135*

Fig. 5.10 Cross-correlations of first derivative with respect to traverse displacement of surface profile and frictional drag data of Fig. 5.7 (mean line removed): (a) forward traverse, (b) reverse traverse. *136*

Fig. 5.11 Friction coefficients between the steel ball and specimens at an increasing load: (a) the steel ball on a steel specimen for three different tracks, (b) the steel ball on specimens of copper (solid line) and aluminium (dashed line)....... 137

Fig. 5.12 The contact resistance between the steel ball and a steel specimen....... 138

Fig. 5.13 A photograph of the damaged surface of the steel ball used in the experiment....... 138

Fig. 6.1 Bode plots of the transfer function of a stylus system, (a) system gain and (b) phase shift against the normalised frequency....... 160

Fig. 6.2 Spectra of a copper specimen traversed at speed of, from (a) to (d), 0.05 mm s^{-1} to 1 mm s^{-1} corresponding the horizontal magnification, V_h, x100 to x5. The x-axis is in Hertz and y-axis in μm^2... 161

Fig. 6.3 The response of the stylus to a sinusoidal position input at different frequencies: (a) stylus responses to input signals at frequencies of 0.5, 5, 10, 25, and 50 Hz; (b) following errors after subtracting the stylus output from the DPT displacement. 162

Fig. 6.4 Responses of the stylus at simulated force signals of different speeds: (a) input force signal, (b) at a speed of 0.05 mm s^{-1}, and (c) at a speed of 1 mm s^{-1} 163

Fig. 6.5 Schematic of the modified Talysurf 5 with force and damping control. 164

Fig. 6.6 The circuit diagram of the active damping controller. Variable controls adjust: P - proportional gain; I - integral action (with C_i); K_I - the damping coefficient; and R - the static force set-point....... 165

Fig. 6.7 The response of the stylus to release from an initial position: (a) Oscillation of the stylus; (b) fitted result by the simplex searching algorithm. Solid line is the original and the dashed line is the fitted result....... 166

Fig. 6.8 Responses of the stylus to a simulated force signal of an equivalent speed of 0.05 mm s^{-1} at three damping ratios. Graphs arranged from top to bottom: the input force and the outputs of the stylus at damping of 0.6, larger than 1 and 0.022 respectively. 167

Fig. 6.9 Variations of parameters R_a, R_q, Δ_q and P_c of the stylus response to a simulated force signal of equivalent speeds of (a) 0.05 mm s^{-1}, and (b) 0.1 mm s^{-1} as the damping is varied. Parameters are normalised to the values at the damping ratio 0.6 168

Fig. 6.10 Standard deviation of following errors after subtracting the original profile from the stylus responses in figure 6.9. Values are normalised at damping factor 0.6. 169

Fig. 6.11 Responses of stylus to simulated force signals of equivalent speeds of 1, 0.25, and 1 mm s^{-1} at damping ratios of: (a) 0.59 and (b) 0.022. Graphs are arranged from top to bottom in the order as speed increases. 170

Fig. 6.12 Variations of parameter values with damping on a steel surface at traverse speed of 0.05 mm s^{-1} (full curve) and 1 mm s^{-1} (broken curve): (a) R_q normalised to 0.079 μm; (b) Δ_q normalised to 0.082; (c) P_c normalised to 0.34, 0.25 μm^{-1} respectively. 171

Fig. 6.13 Repeated traces on a copper specimen at two damping values with different tracking speeds: (a), (b) at 0.05 mm s^{-1}; (c), (d) at 0.1 mm s^{-1}; (e), (f) at 0.25 mm s^{-1}; (g), (h) at 1 mm s^{-1}. 172

Fig. 6.13 (e), (f) 173

Fig. 6.14 Repeated traces on a mild steel specimen at two damping values with different tracking speeds: (a), (b) at 0.25 mm s^{-1}; (c), (d) at 1 mm s^{-1}; (e), (f) at 0.1 mm s^{-1} 174

Fig. 6.14 (e), (f). 175

Fig. 6.15 Output of the stylus at loads of 1.55 (full curve), 0.84 (broken curve) and 0.46 (dotted curve) mN with a speed of 1 mm s^{-1} and a damping factor of 0.022 176

Fig. 6.16 Profile of a load cell button and the corresponding force measured at damping ratios of 0.022 (broken curves) and 0.59 (full curves) at tracking speeds of: (a), (b) 0.1 mm s^{-1}; (c), (d) 0.25 mm s^{-1}. 179

Fig. 6.17 (a), (b) Surface profiles and friction drag variations at damping ratios of 0.022 (broken curves) and 0.59 (full curves) at a tracking speed of 0.1 mm s^{-1}. 180

Fig. 7.1 Schematic of a typical scanning tunnelling microscope. 208

Fig. 7.2 (a) Piezoelectric effect, and (b) a typical piezoelectric tube scanner. 208

Fig. 7.3 Schematic diagram of the profilometer. 209

Fig. 7.4 Photograph of the complete profilometer. 210

Fig. 7.5 Measured displacement using LVDT gauging of platform displacement against the actuator coil current; (a) Probe axis; (b) specimen axis. 211

Fig. 7.6 Piezoelectric actuator with inbuilt capacitance position sensing (from Queensgate Instrument). 212

Fig. 7.7 Schematic of the long range tunnelling profilometer. 213

Fig. 7.8 Diagram of the sensing and linearisation circuits. 214

Fig. 7.9 Circuit diagram of the PID controller. 214

Fig. 7.10 Current drive circuit for spring flexure actuator. 215

Fig. 7.11 Block diagram of the STM control system. 215

Fig. 7.12 Micrographs of tips prepared in two ways: (a) a tungsten tip with a hoop-like end, (b) a tungsten tip with a cracked end, (c) a more typical tungsten tip, (d) a palladium tip. 216

Fig. 7.12 (c) and (d). 217

Fig. 7.13 Forward and reverse displacement of probe and specimen platforms as derived from the calibration curves of figure 7.5: full curve – probe axis; broken curve – specimen axis. 218

Fig. 7.14 Profile measurement of a gold surface. 218

Fig. 7.15 Noise characteristic of the tunnelling current at I_t = 10 nA: (a) non-scanning tunnelling signal, and (b) its power spectral density. 219

Fig. 7.16 A long cycle measurement on the surface of gold-coated silicon by a tungsten tip: (a) forward scanning and (b) reverse scanning. 220

Fig. 7.17 Three repeated measurements taken at the same position on the gold-coated silicon sample, plotted with artificial vertical shift. 220

Fig. 7.18 A forward and reverse measurement cycle on the surface of platinum using a palladium tip. 221

Fig. 7.19 Details of the plot of in figure 7.18: (a) hill 'A' in original form, (b) filtered at 30 Hz. 222

Fig. 7.19 Details of the plot of in figure 7.18:(c) typical profile at 'B', filtered at 30 Hz. 223

Fig. 7.20 Surface profiles of platinum specimen: (a), (b), (c) repeats using one palladium tip and (d) with the palladium tip shown in figure 7.12 (d). 224

Fig. 8.1 The S-V characteristic of a tungsten tip on a gold-coated silicon surface at constant tunnelling currents of: (a) I_t = 10 nA, (b) I_t = 5 nA, and (c) I_t = 0.2 nA. (Arbitrary zero level, and x-axis in volts). 233

Fig. 8.2 The maximum variation of tip movement as tunnelling current varies from 0.5 to 10 nA at each constant bias voltage on the same surface as in figure 8.1. 234

LIST OF TABLES

Table 3-1 *Dynamic behaviour of a stylus system.* 46

Table 3-2 *The variations in parameters with the different stylus tips.* 63

Table 4.1 *Calibration of the vertical displacement.* 90

Table 4.2 *Parameter variation from repeat traces on steel at 0.69 mN.* 92

Table 4.3 *Parameter variation from three sets of repeated tracks on a copper specimen at increased loading.* 93

Table 4.4 *Parameter variation from three sets of repeated tracks on a steel specimen at increased loading.* 94

Table 4.5 *Parameter variation from three sets of repeated tracks on a brass specimen at increased loading.* 95

Table 4.6 *Parameter variation from three sets of repeated tracks on an aluminium specimen at increased loading.* 96

Table 5.1 *Predicted and measured indent geometry for steel, copper and aluminium specimens.* 130

Table 6.1 *Surface finish parameters and their relative variations for copper and mild steel specimens.* 177

Table 6.2 *Surface finish parameters on a copper specimen at two damping ratios and three different loads.* 178

ACKNOWLEDGEMENT

I would like to take this opportunity to express my sincere thanks to my supervisors, Dr. D. G. Chetwynd and Dr. S. T. Smith, for their support and encouragement during my period of research and their suggestions and comments on this piece of work, and also for their help in correcting the thesis.

Many thanks go to the people in the Nanotechnology and Microengineering Centre, especially, Dave Robinson, Steve Wallace, Rhod Mortimore and Frank Courtney for their general help and brilliant skills which I appreciate very much.

I would also like to express my gratitude to Prof. D. K. Bowen, Prof. D. J. Whitehouse, and Prof. P. Carpenter for their encouragement and discussions.

Finally, my deepest gratitude goes to my husband, Weili Wang, for his general support and understanding.

DECLARATION

This thesis is presented in accordance with the regulations for the degree of Doctor of Philosophy by the High 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 specially acknowledged in the content.

X. Liu

SUMMARY

The object of this Ph.D work is to evaluate fidelity in surface measurements in nanometric metrology for both contact and non-contact methods, namely stylus instruments and scanning tunnelling microscopy. Fidelity is defined, in this thesis, as a measure to which an instrument system reproduces the surface features and thus the parameters of interest. High fidelity measurement has two meanings; less distortion in the measured result and less disturbance to the surface being measured. Interaction at the interface between the probe and the surface is the source of failure to achieve high fidelity.

No instrument measures surface topography alone: all instruments measure a convolution of topography and the geometrical and physical interaction of the measured probe and the surface. In the case of a mechanical stylus, factors extraneous to the topography include (a) the shape and size of the stylus, (b) mechanical properties of the stylus and the specimen such as elastic moduli and hardness, (c) frictional force of the sliding pair, and (d) dynamic interaction forces during the sliding. For the scanning tunnelling microscope, factors which affect measurement in addition to topography include the geometry of the tip, the electronic properties of the surface and mechanical deformation due to electrostatic forces and contamination.

These factors have been investigated in great detail, particularly for the stylus instruments. A specially designed electro-magnetic force actuator has been developed to give a better control on loading force during the experiments. Tracking force effects were evaluated by profiling, statistical parameters, and scanning electron microscopy. Friction between a stylus and specimen has been measured for different loading force, sliding speed, material and surface finish. Improvement on dynamic characteristics of a stylus system has been achieved by active damping control. An optimal damping ratio for stylus instruments is found to be within 0.5-0.7. Through the study, the tracking force and traversing speed are found to be the crucial factors to be tackled so that high fidelity measurement can be obtained. A similar investigation has been also made on two home-built scanning tunnelling microscopes to explore the potential applications of STM on nanometric metrology.

1. SURFACE MEASUREMENT TECHNIQUES IN NANOTECHNOLOGY

1.1 Introduction

In the last decade we have continued to witness dramatic progress in the ability to purposely manipulate our world of natural resources. Precision machining on a large and complex component has achieved an accuracy of 0.1 μm with a tolerance about 10 nm (Franse 1990). Typical applications include optical components for laser interferometer systems, lenses for compact disk players, telescopes, mirrors to reflect x-rays, scanners in copiers and laser printers. Highly controllable and stable microfabrication processes such as lithography have been developed in order to fabricate micron or even submicron structures with tolerances at the nanometre level. Magnetic recording technology has been characterised by a continuous increase in storage density and a corresponding decrease in the head/disk interface or flying height. To increase the recording density in computer disk drives, it is necessary to minimise the spacing between the read-write head and the magnetic medium. 'Zero' spacing between the head and the disk results in the strongest magnetic signal, and therefore contact recording is desirable from a magnetic point of view. However, contact recording causes wear and degradation of the head-disk interface, and thus, from a tribological point of view, a non-zero separation between the head and the disk is desirable. Presently, flying heights of only 100 nm or less are being considered requiring a surface texture of less than 50 nm. In tribology, achievements in engineering have enabled lubricated machine components to enjoy the benefits of fluid-film lubrication, while operating with ever-decreasing effective film thickness. When the thickness of the fluid film comes down the nanometre proportions, where, then, is the transition from fluid-film to boundary lubrication, will the friction coefficient reach its minimum as predicted by Stribeck (1902), and how is the transition affected?.

All of these mark a new era of industrial development known as Nanotechnology, which was so termed by Taniguchi (1983). Nanotechnology is concerned with the design and manufacture of

components with submicrometre dimensions, of larger components with submicrometre tolerance or surface finishes, and of machines with submicrometre precision of positioning and motion. Obviously it will be cross-disciplinary, and will influence virtually all aspects of technology and manufacture. Behind all these achievements, measurement has been playing an important role; the cornerstone of nearly every success. Different measuring techniques provide key tools in controlling precision manufacture, understanding, and ultimately exploiting processes involving nanometre level performance.

Instrumentation for surface dimension and feature measurement has developed rapidly in the decade to meet increasing demands from industry. Conventional stylus instruments have been much improved and a new generation of instrumentation developed. This mainly includes nano-stylus instruments, optical profilometers, and scanning microscope probes such as STM and AFM (Scanning Tunnelling Microscopy and Atomic Force Microscopy). These instruments are all adequate for nanometre resolution, and the STM and AFM can provide even sub-angstrom resolution. Using an STM, it is possible not only to 'see' the atomic structures of a surface but also to manipulate such structures on some surfaces. For example, Eigler and Schweizer (1990) were able to accurately position xenon atoms on a nickel surface to spell out the letters 'IBM'.

Surface measuring techniques are classified in three main groups: profiling, area characterisation, and microscopy. Profiling techniques develop quantitative knowledge of the surface peaks and valleys through point-by-point measurement with a high-resolution probe. This probe may be a focused light beam, some types of proximity sensor, or a contacting probe like a stylus. Area techniques, on the other hand, probe the surface over a finite area and produce a measured quantity that presents some statistical property of the surface. An example of an area technique is total integrated scatter (TIS), in which the integrated intensity of light scattered from a surface may be related to the surface roughness. Microscopy produces surface images whose contrast may depend on the surface topography. Examples of this technique are conventional optical microscopy and scanning electron microscopy (SEM). However, the heights of surface features are difficult to measure in microscopy because the image contrast is rarely a simple function of height or

depth alone. Therefore, microscopy techniques are rarely quantitative. When they are, the results can generally be described as a surface profiling technique.

In the following sections, advances in three main categories of profiling techniques; stylus, optical, and scanning tunnelling microscopy, are discussed. The scope and resolution of each are compared, leading to an outline of this thesis on the fidelity study of surface measurement.

1.2 Advances of Surface Measuring Techniques

1.2.1 From a razor blade to nano-stylus instruments

The stylus profiling is the most widely used technique for measuring surface topography. Despite some disadvantages which will be discussed later it has proved versatile and reliable in both the manufacturing and research environments. It is also the instrument around which current standards are defined (BS 1134: Part 1, 1988).

Stylus technology has more than 50 year history. The first instrument is ascribed to Schmalz (1929) but the first serious investigators of surface texture both practically and theoretically were Abbott and Firestone (1933). They devised the first instrument using a broken razor blade to trace surface textures, and attempted to ascribe a meaningful parameter to the output. It was then called Abbott-Firestone curve and is now called the material ratio curve. It was devised as a curve describing how the ratio of metal to air changed as a lapping plate lapped the surface texture down from the highest peak to the lowest valley. This was the first functional parameter of surface measurement.

Over the last 50 years, this technique has undergone almost continuous development. The stylus tip has been changed from a crude needle to the now well defined diamond tip. Both vertical and horizontal sensitivities have dramatically increased. Integration of a modern micro-computer means that dozens of surface parameters can be produced with one measurement to give information relevant to many applications. However, no matter how it has been refined, the basic principle of this technique remains the same. A schematic diagram of the sensor of a high resolution stylus-based instrument is shown in Fig. 1.1. It is a sensitive linear variable differential transformer

(LVDT). As the carriage holding the sensor traverses the surface, the vertical motion of the diamond stylus over the peaks and valleys causes the ferrite core to be displaced from the central position between the electromagnetic coils. This action changes the balance of an AC bridge, to which the coils are connected. After demodulation and amplification, the resulting voltage signal is proportional to the vertical displacement of the stylus over a wide range (Reason, 1970).

The lateral resolution of a stylus instrument is limited by both the radius and the included angle of the diamond tip itself. This can be as small as 100 nm or less such as Berkovich edges. The vertical resolution is limited by noise in the LVDT but more seriously by acoustic and mechanical noise, accuracy of traversing mechanism and thermal drift. The vertical sensitivity of a high resolution Talystep (Rank Taylor Hobson) stylus instrument was estimated to be 0.1 nm (Bennett and Dancy 1981). Lindsey *et al.* (1988), on the other hand, have reduced the thermal drift by building much of their instrument with Zerodur and other low thermal expansion materials. The rebuilt components include the support structure for both the sample and transducer, a specially designed slideway, and key elements in the transducer assembly such as the stylus support tube. In this the slideway itself has a traverse range of 50 mm, about 20 times as long as a conventional Talystep. It has achieved 4 nm peak-to-valley reproducibility over most of this traverse length in addition to noise resolution at 0.1 nm level. The combination of the Talystep measuring technique and the long range traverse slideway produces a powerful stylus instrument—Nanostep (Garratt and Bottomley, 1990) which is more suitable to applications for smooth surface measurements.

1.2.2 Optical profiling

One big advantage of optical profiling methods is that they are non-contacting and hence avoid the potential for damage to the surface that is present with a mechanical probe. Another advantage is faster measurement and may therefore enable 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. There are mainly two classes of optical profilometers: interferometry and focus detection. The interferometry can be described as the measurement of the phase shift between a measuring beam, reflected from the

surface, and a reference beam. As the measuring beam scans the surface, the changes in surface height produce phase shifts in the reflected beam that are then compared with the reference beam. The phase change of the reflected measuring beam is directly proportional to the path length travelled to and from the reflecting surface and hence the surface height itself. Different types of interferometric profilometers are distinguished by its method for provision of the reference beam. In all cases, electronic phase measuring techniques are used to probe surface height variations of about 0.1 nm. Most approaches use polarising optics to separate and recombine the measuring and reference beams. Fig. 1.2 shows a typical Wyko profiling instrument (Wyant *et al.* 1984, 1985). A beam splitter and a reference surface is located very close to the specimen under examination, and the interference takes place between the light reflected from the specimen and that reflected from the reference surface. Neither the specimen nor the instrument is scanned to develop the surface profile. Rather an interferometric image of the surface is focused onto a photodiode detector array. The phase information on each element of the array is found by vibrating the interferometer with a piezoelectric transducer and specially processing the resulting modulated signals in the photodiode elements. Both linear and two dimensional arrays have been used to yield surface profiles and topographic maps, respectively. The accuracy of the surface profiles or maps does not depend straightness or mechanical noise of any moving stage, but does depend on the quality of the reference mirror. The interferometer has a vertical resolution of about 0.1 nm. The length of scanning is limited by the magnifications of the objective lenses available for the microscope. A typical scanning length is 1.33 mm.

The NPL (National Physical Laboratory) profile interferometer (Downs *et al.* 1985) takes a novel approach to providing a reference surface. A schematic diagram of the focus configuration is shown in Fig. 1.3. A light beam with two components of plane polarisation (O and E) passes through a microscope objective, having a birefringent lens. The birefringent lens causes a difference in the focal positions of the two polarisation components in the beam. The O-polarised component is focused on the specimen with a spot size of about 1 μm and the E-component is slightly defocused there with spot size of about 10 μm. Both components are reflected from the surface. The phase of one component is determined by the surface level at the focused spot of light, whilst the phase of the

other represents a reference level integrated over the large area of the defocused beam. As the surface is scanned across the light beam, the phase-difference between the two components varies according to the surface profile and is measured by a polarimetric system. The resulting profile is then obtained between the probed point and the moving average reference. The vertical sensitivity has been quoted as better than 0.01 nm. However, the lateral spatial wavelength range that is detectable, is limited by the spot size of the reference beam of about 10 μm. Therefore, this interferometer is specialised for high precision profiling of very small surface features.

A disadvantage of these interferometric techniques is that the optical phase shift of the reflected beam depends on surface material as well as height. Hence unambiguous measurements of surface profile require homogeneous materials and contaminant-free surfaces.

The focus detection technique, on the other hand, relies on the degree of focus of a reflected light spot (Sawatari and Zipin 1979). This technique requires intensity measurement and hence avoids the homogeneity problem discussed above. Kohno *et al.* (1985) used a critical angle method of total reflection to get steep reflectivity change around the critical angle. In this design a vertical resolution of 1 nm was obtained. The principle diagram is shown in Fig. 1.4. After passing the microscope objective, the *p*-polarised laser light becomes parallel flux if the surface under test is in focus (at position B). A total reflection prism is positioned to reflect the light at the critical angle. If the surface is out of focus, say close to the lens (position A), the light flux diverges slightly and the condition of the total reflection is lost at the upper side of the optical axis. The reverse situation (say convergence) occurs if the surface lies beyond the focus. Therefore a differential output of two photodiodes gives the amount of surface deviation from the focus position.

Since the optical wavelength in air is sensitive to the environmental effects of air temperature and pressure, all types of optical profilometer designs try to maximise the common path of the measuring beam and the reference beam so that environmental changes will affect both beams equally.

Most of the optical profilometers have a vertical resolution equivalent to or better than stylus instruments, but the lateral resolution is limited by the probe spot sizes on the surface and these have diameters of roughly 1~ 2 μm at best. There are some other reasons why they can not replace the

stylus instruments. Firstly the sensitivity of the optical methods depends on the size and 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. Secondly, optical systems, in general, require a more controlled working environment. They are not suitable for use in workshops. The third limitation is that surfaces measured by optical profilometers normally require a good finish or, sometimes, a good reflection. A detailed discussion can be referred to Whitehouse's handbook of surface metrology (1994).

1.2.3 Scanning at atomic level

The scanning tunnelling microscope (STM) represents an important step forward in profiling technology because it makes possible the direct observation of extremely fine, and under the right conditions even atomic scale surface structure. It provides high resolution profiling or imaging of surfaces over a relatively large dynamic range, both in the horizontal and vertical directions. Even more significantly, this high resolution is achieved in vacuum, air, and liquid environments, which makes this new technology convenient to use for many scientific fields such as physics, materials science, engineering, biology, and chemistry.

The theory of quantum mechanical tunnelling was first developed in the early 1920s. Leo Esaki was the first to exploit tunnelling in a real solid-state device while working at the IBM Yorktown Heights Thomas J. Watson Research Centre. In 1972, he received the Nobel Prize in physics for this invention. At about that time, Russell Young (1972) and his co-workers at the National Bureau of Standards (now the National Institute of Standards and Technology) described an instrument called the "Topografiner", which in many ways is the predecessor of the modern STM. In this system, a tungsten probe is mounted on piezo-electric translation stages that are capable of positioning in three Cartesian axes. The specimen is moved into close proximity to the probe using a differential micrometer. The topografiner was operated in the field emission mode, in which the tip was held about 20 nm away from the surface. Young *et al.* (1971) also moved the tip very close to the surface (~ 1.2 nm) and observed strong evidence of the vacuum tunnelling regime, where the tunnelling current can change by about an order of magnitude with each 0.1 nm change in distance between the emitter and the surface. They achieved a horizontal resolution of 20 nm and

a vertical resolution of 0.3 nm. Vibration and tip instability are the probable reasons that resolution was somewhat limited with this design.

The concept of STM was first described in a patent disclosure in mid-1979 by Heinrich Rohrer and his research assistant, Gerd Binnig. The next few years were spent solving problems related to vibration isolation and both coarse and fine motion controls in all three axes. Their first success occurred in 1981, when images resolving monoatomic steps on the surface of a calcium-iridium-tin crystal were obtained. However, the corresponding paper was rejected because one referee found it "not interesting enough". What caught the attention of the scientific community in 1982 was the paper written by these researchers that described their success in imaging the famous Si(111)-(7x7) surface (Binnig *et al.* 1983). Although this surface structure was first observed in the 1960s, its complexity precluded a structural solution until this work in 1982. In 1986, Binnig and Rohrer were awarded the Nobel Prize in physics for their work.

Since then, many STMs have been built and developed around the world. Commercial versions that operate both in air and in ultrahigh vacuum are also becoming available. Variants, such as the atomic force microscope (Binnig *et al.* 1986), the scanning capacitance microscope (Matey and Blanc 1985; Bugg and King 1988), the magnetic force imaging (Martin and Wickramasinghe 1987), and the scanning thermal profiler (Williams and Wickramasinghe 1986) have also been developed. An excellent review of scanning tunnelling microscope instrumentation is given in the paper of Kuk and Silverman, 1989, who highlight the use of these instruments for either electronic surface properties or biological investigations.

The tunnelling microscope relies on a result of quantum theory in which free electrons, with an energy lower than the potential barrier between the specimen and the emitter probe surfaces, still have a finite probability of "tunnelling" through the barrier, *i. e.* crossing the gap between the probe and the specimen as shown in Fig. 1.5. If the gap size changes, the tunnelling current is altered, producing a signal that can be used in a feedback loop. The system still requires a potential to be applied at the probe tip but it is at the order of 1 V rather than a few tens of volts as in the case of the topografiner. This is accompanied by a reduced specimen-probe separation of approximately 1 nm which is small enough to allow the "electron clouds" of the probe and specimen to overlap

sufficiently to promote tunnelling. As a result, the tunnelling current, I, flowing across a gap, s, can be shown to be

$$I \propto \frac{V_T}{s} \exp(-10.25\sqrt{\phi}\, s) \qquad (1\text{-}1)$$

where ϕ is the effective work function in eV (for most systems, this is on an order of a few eV), s is in nm and V_T is the bias voltage.

The magnitude of the current is extremely sensitive to the distance between the surface and the emitter. One can see that if the tip-surface spacing is increased (decreased) by 0.1 nm, then the tunnelling current will decrease (increase), approximately, by a factor of 10 for an effective work function of 4 eV. As the emitter is scanned over the surface, its height is servo-controlled to maintain a constant current across the gap. The x, y, and z motions are produced by piezoelectric elements; the signals that generate the piezo motion may then be used to generate a z (x, y) map of the surface.

The high sensitivity of the tunnelling current to the spacing gives the STM a very high vertical resolution. It is reported that a vertical resolution of 1~2 pm has been achieved (Hasan *et al.* 1992; Feenstra and Fein 1986). The lateral resolution of the STM is limited by the diameter of the tunnelling electron beam which is estimated by Tersoff and Hamann (1985) in a form of:

$$\Delta = \sqrt{0.2(nm)(R+s)} \qquad (1\text{-}2)$$

for an effective work function of 4 eV, where R is the radius of the tip, and s is the tip-surface spacing, both in units of nm. Therefore, for $R = 0.2$ nm and $s = 0.5$ nm, the electron beam diameter is on the order of 0.4 nm. Park and Quate reported in their paper (1986) that a lateral resolution of 0.2 nm was obtained indicating near atomic resolution.

One important feature of the STM is that as well in air and under liquid as it works in vacuum (Sonnenfield and Hansma 1986; Lindsay and Barris 1988). The reason is that the tip is so close to the surface (0.5 nm) that the volume through which the tunnelling current flows is on the order of 0.1 nm^3. In a normal air environment, there is less than 0.003 air molecules contained in this volume. In water, the corresponding number is about 3. Therefore the tunnelling electrons flow more or less unaffected by ambient molecules.

Generally, image areas of STMs were of a few hundred Å on a side, since atomic resolution of the surface was of interest. But long range STM has been longing for practical surfaces such as surface studies of technological problems on a large scale, engineering surfaces, and semiconductor microfabrication, *etc*.. Instead of seeking large range of piezoelectric tubes, Fu *et al.*(1991) built a STM with 500 μm x500 μm field of view. They use a long range X-Y stage to carry the specimen while the STM head is held stationary.

Over the last ten years, many surfaces ranging from metals and semiconductors to large organic molecules have been imaged (Gerhrtz *et al.* 1988; Sleator and Tycko 1988). An important broadening of STM technology is emerging through the realisation that the STM tip can be employed for purposes other than microscopy, and in particular, that it may be used to modify the surface of the sample in a number ways. Lithography with an STM on a nanometre scale has been tried by various methods. McCord and Pease (1986, 1987) have written lines about 200 nm wide by exposure of thin resist films using the probe tip. Later more complex letters and symbols with linewidths as small as 2 nm were written on a flat Au substrate by electroetching process characterised by a sudden burst in the tunnelling current (Li *et al.*1989). Molecular manipulation was made by Foster *et al.* (1988), in which they "pinned" a complex organic molecule to a graphite base by raising the tip voltage from its normal value of 30 mV to a few volts. In subsequent scan, another voltage pulse disrupted the molecule leaving a benzene ring pinned to the surface. An atomic switch has been realised by Eigler *et al.* (1991) in which a xenon atom is moved reversibly between stable positions of the STM tip and a nickel surface under the control of a voltage pulse. In practical application, Dragoset *et al.* (1986, 1987) have applied STM to optical surfaces, trying to make a direct comparison between STM and a stylus instrument on a well-defined specimen. STM revealed the fine structures of the feed marks (100 nm) but the conventional stylus probe could not.

The major limitation for STM is that the specimen surface must have a reasonably high electrical conductivity ruling out most ceramic and polymer materials. One solution is to put a conduction coating (such as gold or platinum) on such surfaces, assuming that the coating faithfully reproduces the surface topography of the substrate. Atomic force microscopy, on the other hand, provides a solution to the imaging of non-conductive surfaces.

Atomic force microscopy is a sensitive form of stylus profiling, in which a probe tip gently rests on the sample, variations of surface height lead to force changes on the tip that in turn lead to the motion of a delicate cantilever beam. The motion of the cantilever is monitored by either STM or optical interferometry (Erlandsson, *et al.* 1988) and a feedback loop is used to maintain a constant deflection force and thus provides an appropriate signal for imaging purposes. The AFM is directly related to conventional stylus techniques, but with forces in the region of nanonewtons which is six orders of magnitude lower. In theory, the stiffness of an AFM cantilever can be very low as discussed by Binnig and Quate (1986), but the mechanical instability of a soft beam and the interaction between the STM and the beam limit its potential resolution. Using this technique, researchers have been able to image insulators such as cleaved mica and polymers and to image in real time biological processes such as clotting of a protein (Drake *et al.*, 1989). Under appropriate operating conditions, atomic resolution can be achieved (Albrecht and Quate 1988, Binnig *et al.* 1987) equally well with conducting and non-conducting specimens.

The Van der Waals microscope is also known as the attractive mode microscope. It based on the AFM, but in this case a piezo is used to vibrate the cantilever at about 50 kHz. The tip is in a range of 3-20 nm above the surface, and the vibration amplitude is affected by the minute variations in the attractive Van der Waals forces, having magnitudes in the range of 10^{-11} N, when the probe is scanned over the surface. Non-contacting measurements have been found to be particular useful in the microelectronics area (Zong *et al.* 1993, Hansma *et al.* 1993; 1994).

The magnetic and electrostatic force microscopes (MFM and EFM) are similar to the Van der Waals force microscope except that the probes are magnetised or carry an electric charge, and are sensitive to magnetic or electrostatic variations in the specimens under test (*e.g.* magnetic recording media and dopant variation in silicon). Using this technique magnetic domains have been imaged by a magnetised tip (Saenz *et al.* 1987, Martin and Wickramasinghe 1987).

1.3 Technique Comparison — Fidelity of Measurement

We have discussed three types of profiling techniques for measurement of surfaces, primarily in terms of their lateral and vertical resolution. It has shown that each method has certain strengths

and limitations. Although they all claim nanometre or subnanometre sensitivities, there are circumstances or specimens for which it is not at all clear what is being measured, let alone what is the precision of the measurement. In fact any measuring process will, inevitably, introduce interference to the measurand. The question is how to characterise, distinguish and recover the nature of the measurand. In this section, emphasis is addressed on the performance evaluation of each method. This includes the intrinsic limitations such as physical interactions between the surface and the sensing probe, quantum noise, *etc.* and instrumentation limitations such as dynamic range, dynamic response, stability, repeatability, precision and accuracy. For the purpose of this thesis, discussion is restricted mainly to stylus instruments, AFM and STM.

1.3.1 Interaction between the surface and the sensing probes

In most conventional stylus instruments, a leaf spring applies a small force (approximately 1 mN) to the stylus to ensure that it always maintains contact with the specimen. As the stylus presents a very small area to the specimen, high pressure can arise at the contact and this may not only introduce uncertainty in the profile data but also lead to temporary or permanent plastic deformation on the specimen.

The contact between two smooth elastic bodies was investigated by Hertz (1896) who theoretically determined both the size and shape of the contact zone due to elastic deformation of the bodies. For a sphere of radius R pressed against a flat surface under a normal load P, the radius of contact is given by

$$a = \left[\frac{3PR}{4E^*}\right]^{\frac{1}{3}} \tag{1-3}$$

$$\frac{1}{E^*} = \frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2}$$

where ν is the Poisson ratio and E the Young modulus of each material. Hertz theory has governed the contact mechanics for over half century until Roberts (1968) and Kendall (1969) found that, at low loads, contact areas between these bodies were considerably larger than those predicted by Hertz and tended towards a constant finite value as the load reduced to zero. Strong adhesion was

observed between them if the surfaces were clean and dry. At high loads the results closely fitted the Hertz theory. These observations strongly suggested that attractive forces were operating between the solids and, although these 'additional' contact forces were of little significance at high loads, they became increasingly important as the load was reduced towards zero. Later Johnson, Kendall and Roberts (1971) gave more precisely the contact radius for Hertz contact, that is

$$a = \left\{ R / k(P + 3\gamma\pi R + \left[6\gamma\pi RP + (3\gamma\pi R)^2 \right]^{1/2}) \right\}^{1/3} \qquad (1\text{-}4)$$

where γ is the surface energy released by the interaction of the two surfaces and k is related to the elastic modulus by

$$\frac{1}{k} = \frac{3}{4}\left(\frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2} \right) \qquad (1\text{-}5)$$

It is interesting to note that the load across the contact region is higher than the applied load by an amount given by the second term in parentheses in equation (1-4). The surface energy is the consequence of the bonding of two contact points and is often given by the Dupre equation

$$\gamma = \gamma_1 + \gamma_2 - 2\gamma_{12} \qquad (1\text{-}6)$$

where γ_1 and γ_2 represent the free surface energy of the two surfaces and γ_{12} represents the surface energy of the interaction. It can be seen from Eqn. (1-4) that at zero applied load, there will always be a finite contact radius given by the equation

$$a = \left(\frac{6\pi\gamma R^2}{k} \right) \qquad (1\text{-}7)$$

The adhesion force is dependent on the contact area and the conditions of the surfaces. It can be very strong when the surfaces are clean, and can be insignificant in the presence of contaminants. The adhesion force can dominate the friction between the probe and specimen, and causes a finite region of plastic flow at the periphery of the contact. Other factors that cause surface damage by stylus loading are surface hardness, stylus force, stylus tip width and stylus traversing speed. Later it will be shown that for a stylus tip of 2 μm the stylus force has to be kept below 0.5 mN to avoid any detectable damage for some common engineering surfaces.

Another possible major contribution to fidelity of profile measurement will be that of the compliance of the interface between the probe and specimen. Simply, this can be modelled as a frictionless Hertz contact with a total normal load W, giving a contact radius a by Eqn. (1-4). It is shown that the mutual approach of points in the bulk of the two is given by

$$\delta = \frac{a^2}{R} = \left(\frac{9P^2}{16R(E^*)^2} \right)^{1/3} \tag{1-8}$$

Then the contact stiffness, S, can be written as

$$S = \frac{dP}{d\delta} = \frac{3}{2}\left[P(E^*R)^2\right]^{1/3} = 2(\delta R)^{1/2}E^* = 2aE^* \tag{1-9}$$

The contact stiffness is an important factor for evaluating the fidelity of profile measurement. If the stylus-surface interaction is significant, that is the contact stiffness becomes comparable with that of the lever, it is not valid to assume a single stationary value of spring constant. Additionally, compliance will be affected by phenomena such as adhesion, meniscus effects, and local hardness variations which are not relevant to surface geometry. The usual aim of making lever spring constant low to improve force sensitivity can result in the contact stiffness completely dominating spring behaviour. Further, the tip interaction can perturb the deformation behaviour of the mounting spring, so as to make the assumptions about the relevant lever spring constant of dubious worth.

The above discussion applies not only to stylus profiling, but also to atomic force microscopes. Surface forces are classified as short-range and long-range forces. Short-range forces are those that act between atoms and molecules that are essentially in contact, say within 0.1 or 0.2 nm of each other. Long-range forces act between surfaces that are further apart; long range in this context might mean a few nanometres. Examples are Van der Waals, magnetic and electrostatic forces. Most AFMs operate in contact, or repulsive, mode. The process of scanning causes surface deformation, the extent of which is controlled by the generally unknown local mechanical properties at the surface. Thus the source of image contrast may become uncertain. Even those claimed to be non-contact AFMs, when they used in ambient environments, meniscus forces from an adsorbed surface film may cause the tip to jump to contact and may introduce instability in operation. Jarvis *et al.* (1993) have found that the attractive region occurs after the tip jumping toward the surface, in

this region it is possible the AFM operates with an attractive force between the tip and the sample while also being in contact.

Scanning tunnelling microscopy is non-contact in nature but great care has been taken when it is operated in air. Generally an STM scans a surface at a constant tunnelling current, and this will be much reduced if there is a contaminated layer which is more than several nanometres (Mamin *et al.* 1986, Soler *et al.* 1986). To achieve tunnelling current the STM tip will be brought further to the surface, which could either press the surface via the contaminated layer or penetrate into this layer. This can make the STM performance unreliable, sometimes even under vacuum conditions (Main *et al.* 1986). This will be discussed in more detail in Chapter 8.

1.3.2 Range and resolution

A comparison among these methods of their dynamic range for both horizontal and vertical directions is shown in Fig. 1.6. The dynamic range is referred as a ratio of range to resolution. Generally, it depends on two main factors; transducing and mechanical vibration. Stylus instruments typically have a large dynamic range on an order of 10^5 vertically with a range up to 1 mm or a resolution down to 10 nm. Higher resolution is achievable at the cost of range. On the other hand, the resolution is limited by the surface deformation by contact measurement.

The AFM and STM have a small dynamic range in principle, they rely on feedback to drive an actuator (a scan head) to achieve a longer range. Using reasonable geometry (for example, a scan head on an order of a few centimetres long and a scan tube thickness of 1 to 2 mm) and applied voltage, one finds the maximum scan range of an order of 50 to 100 µm. In general, a longer scanner has lower resonant frequency and the scanning speed has to be reduced. In terms of dynamic range, the STM or AFM can not replace the conventional SEM.

The lateral resolution of stylus instruments and AFMs is dependent on the size and shape of tips used. A measured profile or image represents the convolution of the surface structure with the geometry of the stylus (tip). In practice, it is often found that diamond styli are flattened at the end, the width across the flat essentially determining the lateral resolution. A commonly available width is 2 µm, but pyramidal styli with tips as small as a nominal 0.1 x 0.1 µm are also used. In some instances, highly spherical styli are also produced (Vorburger *et al.* 1979; Song and Vorburger

1991). The flank angle of tips also affects its resolution, for it can not accurately profile over slopes greater than the flank angle (usually ± 45° for styli). Very sharp tip for AFM and STM can be made by a technique of Focused Ion Beam (Vasile *et al.* 1991a, 1991b). These tips typically have 5-6 nm radii and cone angles between 12° and 15° over the first two microns from the apex. However, if the tips are too slender they may flex under the influence of lateral forces generated by contaminants on the sample's side walls (Griffith *et al.* 1991) or elongate elastically, which introduces instability in operation and a further uncertainty in the interpretation of profiles.

1.3.3 Dynamic response

Another error source in stylus profiling will be due to dynamic effects. A stylus of finite mass held in contact with a surface by a preloaded spring may, if traversing the surface at a certain high velocity, fail to maintain contact with or press hard against the surface being traced. Where and whether this occurs depends on local surface geometry, the spring constant to mass ratio (the resonance of stylus system) and the traversing speed. As a consequence, the measured profile may produce erroneous information about the surface microgeometry and lead to surface damage due to large dynamic forces.

Analogous to the stylus, an AFM uses the displacement of a soft cantilever to monitor and control the surface forces. Forces are inferred from a nominally known spring stiffness of the cantilever on which the tip is mounted. This is not a problem when the tip is scanned at a very slow speed; the maximum of the equivalent input frequency to the AFM tip (which is the combination of the scanning speed and the wavelengths of surface features) is much lower than the resonance of the cantilever. However, as scanning speed increases, the vibration of the cantilever, expressed by a fourth order differential equation, will also introduce distorted information of the high frequency surface features.

1.4 Characterisation and Calibration of Instruments

An instrument is normally characterised by its range, resolution, precision, repeatability and stability. To clarify those terms, definitions for each one are given below:

Range is the full extent over which the instrument can measure reliably within the confines of its specification.

Resolution or discrimination is the smallest discernible change in the parameter of interest that can be registered by the instrument.

Repeatability is the ability of the instrument to give the same result for the same input, repeatedly applied. Lack of repeatability is due to both random and systematic effects in the instrument and its environment. The standard deviation, or the sample variance, is dependent on the number of tests. Poor repeatability is commonly associated with hysteresis due to backlash or inherent to the transducer as, for example, with piezoelectric gauges. With careful design, the repeatability can approach, but can never be better than, the resolution.

Precision is used with several different meanings in practice. Sometimes it is a synonym for repeatability, as defined above, and other times it may be referred as the standard deviation of a set of results at one operation. In an estimation with the 99.7% of probability, it is given as the percentage of the ratio of the three times of standard deviation, 3σ, to the range.

Stability is the ability of the instrument to keep the given output over a period of time and under certain changes of ambient conditions such as temperature, humidity, *etc*.. It is commonly associated with thermal drifts of mechanical parts and electronic circuits, mechanical creep and relaxation, and ageing effects from both mechanical and electronic components.

Accuracy is the deviation of the measured value from the 'true' value. It is defined in terms of measurement errors. Measurement errors can be classified as random errors and systematic errors. Random errors are associated with 'precision' and systematic errors are related with calibration. The accuracy of an instrument can only be determined after it is calibrated against a standard device. The accuracy of subsequent measurement will then depend on the combination of the calibration error and the precision of the instrument. The calibration error can be large or small depending the grade of the standard device used which is ranging from international agreement; national primary standard, secondary standard to laboratory or industrial standards.

Traceability – At the National Physical Laboratory (NPL), for example, the metre is realised using the wavelength of the 633 nm radiation from an iodine stabilised helium-neon laser. The

reproducibility of this primary standard is about 3 parts in 10^{11} and the wavelength of the radiation has been accurately related to the definition of the metre in terms of the velocity of light. The primary standard is used to calibrate secondary laser interferometers which in turn are used to calibrate precision length bars, or gauges. A simple traceability ladder for length is shown in Fig. 1.7 (Bentley, 1988).

Surface instruments as discussed above are normally calibrated against the step standards produced by evaporation, etching or plating and are traceable to optical wavelengths. However, it has not as yet been possible to produce satisfactorily calibrated standards for steps below about 25 nm. Bowen and Chetwynd *et al.* (1983, 1985, 1988, 1990) have established x-ray interferometry as a metrology tool for the calibration of engineering transducers in the range from 10 pm to a few micrometres (Chetwynd *et al.* 1993). The method of x-ray interferometry relies upon the availability of highly perfect, homogeneous single crystals, of any convenient materials, available in reasonable large sizes. An instrument can be made out of such crystals that, in combination with an x-ray source, can be used to produce sinusoidal variations of x-ray intensity in the detector, with identically the same translation period as that of the crystal lattice itself. If x-ray interferometers and mechanical transducers can be operated simultaneously to measure the same displacement, a reproducible calibration system is thereby achieved. The lattice parameter of the crystal can be determined in absolutely units resulting in a readily portable length standard is also achieved. Semiconductor silicon crystals are normally homogeneous and reproducible, one with another, to a few parts in 10^6. Selected crystals are homogeneous to better than one part in 10^7, and x-ray methods exist to compare the lattice parameter of one crystal with another within 2 to 3 parts in 10^8. The correlation of the silicon lattice spacing to absolute standards has been performed at the National Bureau of Standards (NBS of USA, now the National Institute of Standards and Technology) and the Physikalisch-Technische Bundesanstalt (PTB of Germany) to a precision of about 1 in 10^6 (Deslattes and Henins, 1973; Becker *et al.*, 1982).

1.5 The Layout of the Thesis

The contents of this thesis are arranged in a systematic order. The first two chapters are reviews of the surface measuring techniques and background information on surface formation and description. Three measuring techniques, stylus profiling, optical profiling and scanning probe microscopy such as STM and AFM are compared in their dynamic ranges (range over the resolution), lateral resolutions, and mechanical/physical interactions. Chapter 3 gives a theoretical analysis on a stylus system, which includes system parameter analysis, interaction force calculation and geometrical effect of the stylus on the measurement. Because of the instrument dependence, a particular stylus instrument, Talysurf 5, is selected for the analysis. This instrument has been modified to give a force control at the stylus, and the detail of the instrumentation is discussed in chapter 4. Both chapters 4 and 5 concentrate on the force interactions which are the main concern for measurement fidelity. In chapter 4, the effects of tracking force on measurement are evaluated from profiles, statistical parameters and SEM examination. The design and application of a novel magnetic force actuator is also presented. Frictional force at the interface as the stylus slides over a surface is intensively studied in chapter 5 for a wide range of engineering materials. Instrumentation for simultaneous measurements on both friction and profile is discussed, and this leads to a simple friction model being derived. An experimental approach to increase fidelity or, in another sense, to increase measuring speed without degrading the fidelity has been tried and is reported in chapter 6. This is an active damping control achieved by a velocity sensor and the magnet/coil force actuator. In chapters 7 and 8, topics are mainly on the investigation of the potential applications of scanning tunnelling microscopy to the nanometric metrology. This includes design and construction of two scanning tunnelling profilometers with different translation drives. The characteristics of the two STMs have been evaluated. Problems associated with STM working in normal environment (in ambient conditions) are discussed. The whole thesis is summarised in chapter 9, and some suggestions on future work are proposed.

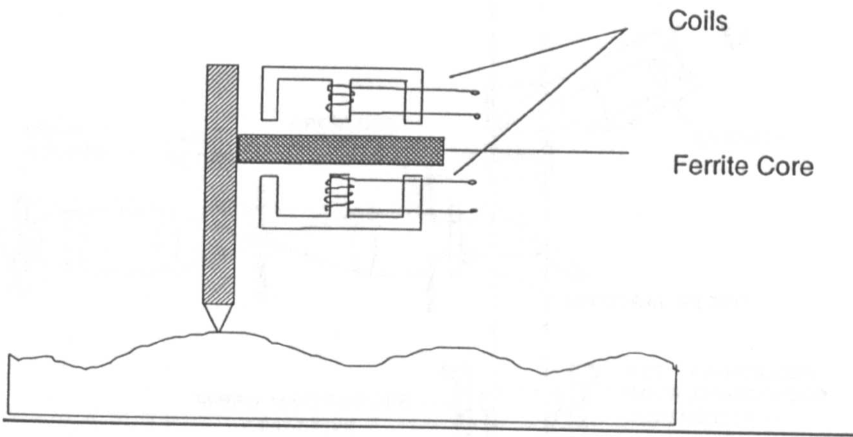


Fig. 1.1 Simplified diagram of a high sensitive LVDT.

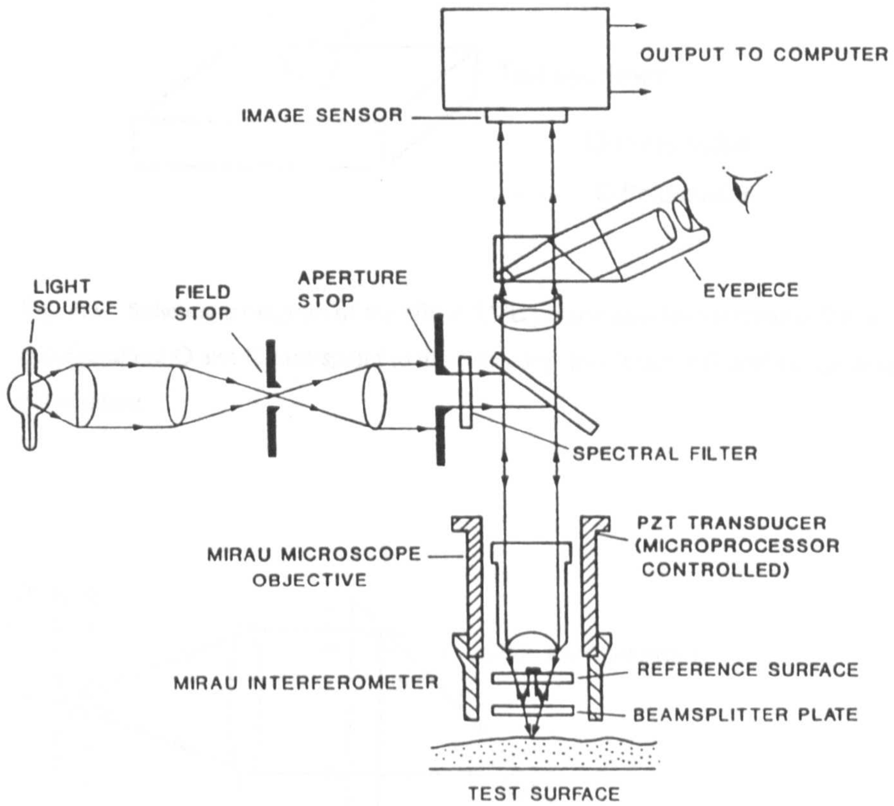


Fig. 1.2 Schematic diagram of an optical follower

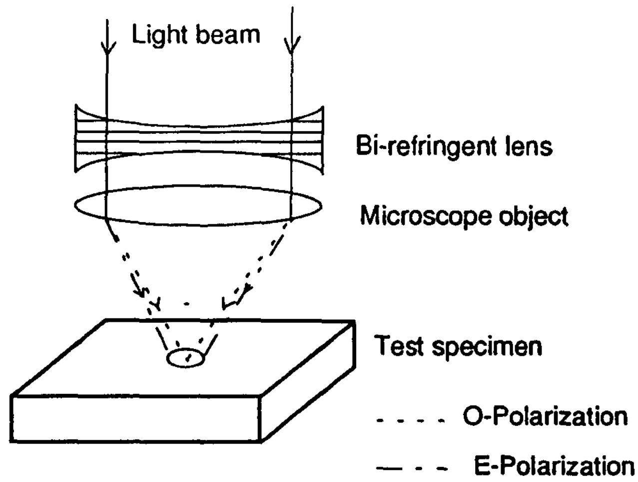


Fig. 1.3 Schematic diagram of the NPL/CUPE polarisation interferometer focus configuration. O and E correspond to the 'ordinary' and 'extraordinary' components of plane polarisation.

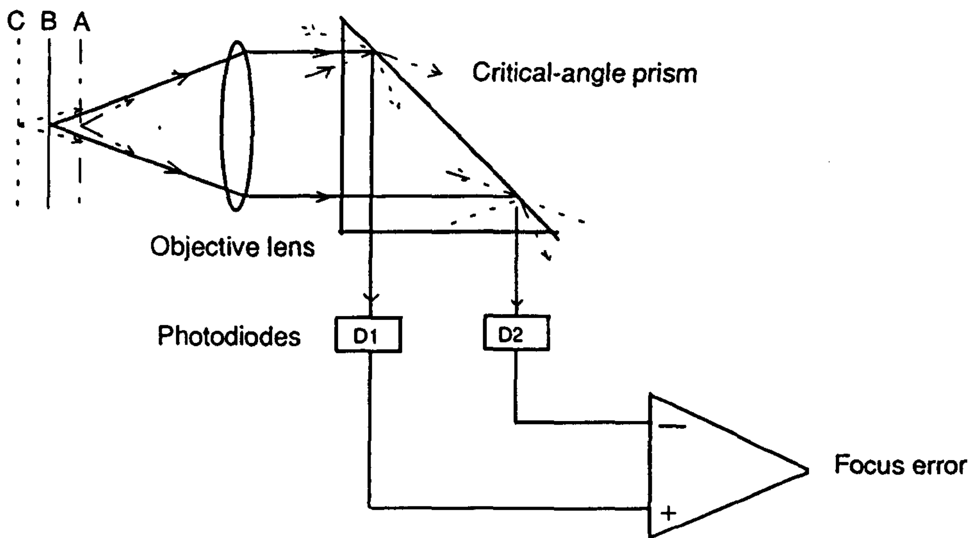


Fig. 1.4 Principle of focus detection by critical angle of total reflection

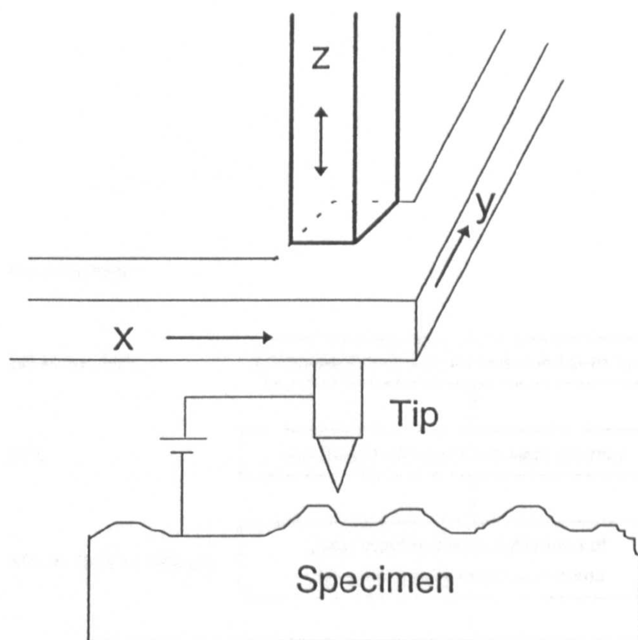


Fig. 1.5 Schematic diagram of the scanning tunnelling microscope

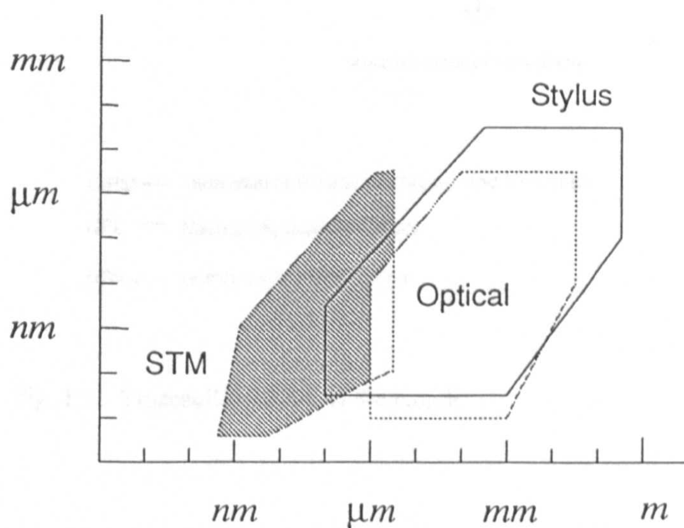
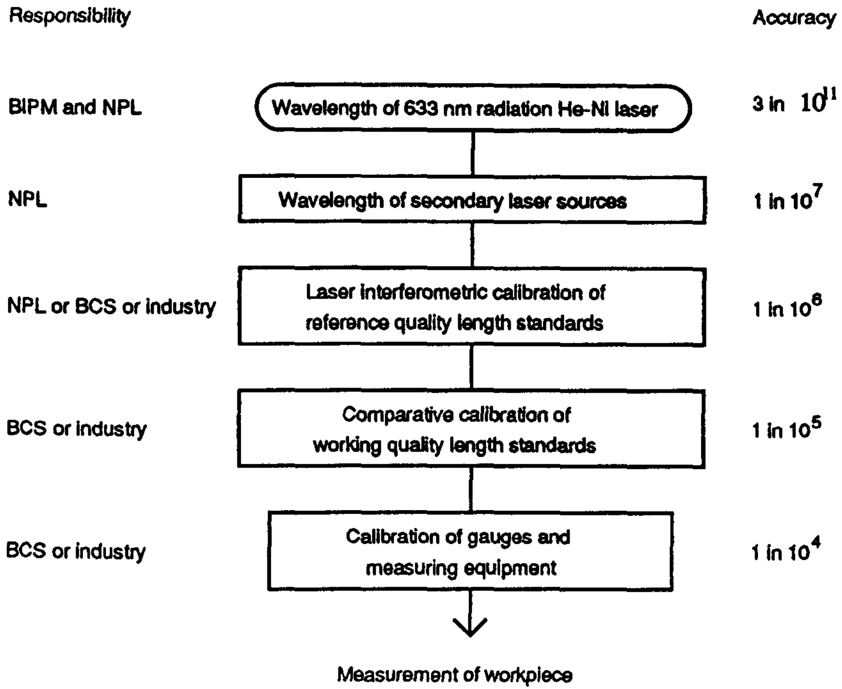


Fig. 1.6 Estimated range and resolution limits in both the vertical and lateral directions for stylus instruments, optical interferometers, and STM.



BIPM — International Bureau of Weights and Measures

NPL — Natinal Physical Laboratory

BCS — British Calibration Service

Fig. 1.7 Traceability ladder of the length

2. CHARACTERISTICS OF ENGINEERING SURFACES

An engineering surface is a link between the manufacture of an engineering component and its performance. The recognition that the vast majority of engineering components can potentially degrade or catastrophically fail in service through such surface-related phenomena as wear, corrosion and fatigue, led in the early 1980s to the development of the interdisciplinary subject of surface engineering (Bell 1987, 1990). It involves the application of traditional and innovative surface technologies to engineering components and material in order to produce a composite material with properties unattainable in either the base or surface material. This indicates that the study on surfaces comes to a stage of not only to investigate but also to manipulate the properties of surfaces.

2.1 Surface and its Description

2.1.1 Concept

In physics, the surface is defined as the transition region between the three dimensional periodic structures of bulk crystal and vacuum. Thus the surface includes all the layers of atoms that do not have the three dimensional periodicity of the bulk. The wave functions of the electrons are modified in the vicinity of the surface. The resulting relaxation of electron charge density creates a new self-consistent potential in the surface region of which the principal compound is a surface dipole; those electrons tunnelling into the potential barrier at the surface and the uncompensated positive charge inside the surface create a surface dipole layer. The extent to which the electron gas penetrates into the vacuum can be estimated from the uncertainty principle as follows

$$\begin{aligned} \Delta x \Delta p &\approx \hbar \\ \Delta x &\approx \hbar / \sqrt{2m_e E_f} \approx 1\text{Å} \end{aligned} \qquad (2\text{-}1)$$

as the Femi energy $E_f \approx 5$ eV = 5 x 1.6 x 10^{-19} J for metals, here $\hbar \approx 1.05 \times 10^{-34}$ J s, $m_e = 9.11 \times 10^{-31}$ kg.

For the semiconductors and insulators, the formation of surfaces allows polarisation of bonds in the direction of surface normal, again creating a dipole layer.

The simplest type of solid surface is the perfect cleavage face of a single crystal such as mica or highly oriented pyrolytic graphite (HOPG). The surface consists of a regular array of atoms. In surfaces of many semiconductors as well as a few metals, more complex atomic rearrangements occur which results the so called 'reconstructed surfaces'. The cleavage faces are rarely flat; they are covered with steps and blocks, often many hundreds of atoms high.

2.1.2 Solid surfaces in air

The above surfaces are ideal or clean surfaces and they do not exist in the conventional environments. Most metallic surfaces, even if they are noble metals, will immediately adsorb oxygen and water vapour if they exposed to air; this film may not be more than a few molecules thick (Bowden and Throssel, 1951). With more reactive metals there is initially some chemisorption that soon gives way to chemical reactions and the formation of oxides or hydroxides. During the early stage of oxide growth there may be rearrangement of the surface atoms of the metal that, in a more exaggerated form, can lead to faceting. The oxide may then grow to a thickness of the order of 50-100 Å by a process of diffusion of metal ions through the oxide to the oxide-gaseous interface. Some oxides are epitaxial and firmly linked to the substrate; they can impede diffusion. In that case the oxidation soon comes to an end, and the material appears to be noncorrodible or even 'noble' (*e.g.*, chromium, rhodium). In other cases the oxide *is in marked mismatch* with the substrate and cracks as it grows, so exposing fresh metal for further oxidation. Oxides may be amorphous or crystalline; they may be smooth or they may grow as crystals and whiskers.

Ionic solids and covalent solids are usually materials of high surface energy and the adsorption rarely leads to the presence of more than a few molecular layers of oxygen or water vapour. However, contaminant films will condense from the atmosphere and remain on the surface.

In the study of surface structures, surface cleaning usually involves either heating in high vacuum, ion bombardment and/or preferential chemical attack. These steps are needed to remove

the last traces of surface contaminant and even so small quantities of bulk impurity may diffuse to the surface during or after the cleaning process.

2.2 Engineering Surfaces

2.2.1 The nature of surface

Engineering surfaces are mostly prepared by mechanical methods such as cutting, grinding, abrasion, polishing or chemical etching to meet certain purposes. These processes generate surfaces over a large scale of roughness typically ranging from millimetres down to submicrometres. In addition they produce distortions and modifications in the outermost layers. Fig. 2.1 is a schematic diagram of a typical polished metal specimen showing topography and sub-surface structure. On the top of the normal crystalline structure of the base material lies a layer of deformed material resulting from the processes used in the manufacturing of the surface. It may consist of a smeared 'fudge' of greatly deformed metal, metal oxide, and polishing powders with a typical depth of about 40 µm (Larsen-Basse, 1992). This deformed layer is itself covered by a compound layer resulting from chemical reaction of the metal with its environment. For a metal in air, it will usually be oxide, and may introduce organic compounds. Normally it is about 10 nm thick (Halling 1975). This reacted layer will, in a normal environment, often be covered with a layer of contaminants such as condensed oil vapour and particles of smoke and dust. The thickness of the layer may come up to 1 µm, and in a normal operation of cleaning a surface using organic solvents the formation of the chemisorption layer can be about a few nanometres. Because of their shape and chemical reactivity and because of their composition and microstructures, surfaces are extremely complicated and may be very different from those of the bulk solids.

2.2.2 Topographies of surfaces

The geometric shape of any surface is determined by the characteristic of the finishing process used to produce it. There will be undulations of wavelengths that range from atomic dimensions to the length of the component. These often result from the dynamics of the particular finishing process or machine used. There may be additional peaks and valleys caused by local

microevents, such as uneven deformation of hard microstructural constituents, local fracture, or corrosive pitting. Even after a surface has been carefully polished, it will still be rough on an atomic scale. In general the surface geometry is described by three main parameters: error of form, waviness and roughness. They are defined as follows, shown in Fig. 2.2.

Errors of form are *macrodeviations* — *irregular surface departures from the design profile*, often caused by lack of accuracy or stiffness of the machining system.

Waviness is periodic deviations from the geometric surface, often sinusoidal in form and caused by low frequency vibrations of the machine-tool-workpiece system during the machining.

Roughness is the deviations from the wavy surface itself, caused by geometry of the cutting tool and its wear, machining conditions, microstructure of the workpiece and vibrations in the system. Examining the details of surface roughness will reveal a finer roughness superimposed on the surface roughness. It may extend down to the near-atomic scale and may be caused by internal imperfections in the material, non-uniform deformation and or fractures of the individual grains at the surface, or corrosion and oxidation processes occur while the surface is being generated or during its exposure to the environment.

2.2.3 Properties of surfaces

The mechanical, physical and chemical properties at the surface are normally different from those of the bulk material. The complexity of the formation of a surface makes it difficult to determine its characteristics. This has led to rapid development of the science of Tribology which studies surfaces in contact and in relative motion. Tribology is interdisciplinary concept that involves many sciences such as mechanics, physics and chemistry. Surfaces are thus evaluated through contact effects, friction, wear and chemical reactivity.

The topographies of surfaces, especially those of peaks often called asperities, are of primary concern in sliding friction and wear of materials, because these processes usually involve contacts between asperities on opposing surfaces. The true contact area, the very outer portions of asperities, is very much smaller than the apparent area in contact. The strength of a surface to withstand loading and resist wear is much determined by the composition of the surface which may, as we mentioned early, vary significantly from the base material and will be affected by the different

manufacturing process. The shear strain in the deformed layers may increase hundred times and thus leading to a very heavy dislocation concentration. Evidences show that polished glass ceramic can reduce their strength and the strength of a single crystal silicon degraded after grinding and regained when the surface was etched away a few micrometers (Chetwynd and Smith 1990a, Smith *et al.*, 1991).

The chemical constituents at the surface play another important role on its mechanical performance. This compound can be formed either during the surface processing or when the surface comes into contact with a different solid, a gas, or chemisorbed species. It may be a layer of oxide or hydroxide due to reactions with ambient moisture and oxygen. When two surfaces rub against each other, they may adhere at local spots that can reach elevated temperature by frictional heating and interdiffusion may take place, resulting in local compound formation in the surface layers which leads to higher friction (Rabinowicz 1971).

It is obvious that the characteristic of a surface is determined by a number of mechanisms that may act in combination and in different proportion under different circumstances, but the real process of the surface performance such as friction remains unclear. However, the recent development of sensitive and powerful technologies for measuring and modelling surfaces and even manipulating individual surface atoms is creating a new future for surface studies.

2.3 Techniques for Preparing Smooth Surfaces

Smooth and sometimes supersmooth surfaces are needed for critical applications such as semiconductor wafers, magnetic disks, and optical components for laser mirror. In order to obtain good performance from these surfaces proper preparation and cleanings are essential. Most smooth surfaces are produced by mechanical polishing after being machined into shapes, or by a combination of mechanical and chemical polishing. However, other techniques have been developed and these include single-point diamond turning, non-contact float polishing, hydrodynamic polishing, and ion and laser beam processes. Details are reviewed by Bennett (1986) and here only a brief discussion is listed.

MECHANICAL POLISHING:

Mechanical polishing has been used for centuries. In its simplest form it involves contact between a rotating pitch lap, which has been pressed to the inverse of the shape desired for a specimen, and the surface being polished. The polishing is accomplished by abrasive particles, generally under 2 μm in diameter suspended in a water slurry that is either added incrementally to the polisher (fresh feed polishing), or continually bathing the lap and substrate (submerged, bowl-feed, or recirculating polishing). Random motion between the lap and the surface assures that the material removal is uniform. The purpose of the polishing is to remove any fractured and damaged material produced by machining such as grinding. A good rule of thumb is to remove a layer whose thickness is equal to ten times the diameter of the abrasive used in the preceding operation. A surface produced by so-called fresh-feed process will generally contain fine scratches and will be improved to scratch-free by different final polishings—polishing under water or polishing to dryness (Bennett 1970). As an example, sometimes a weak acid is added to the slurry when mechanically polishing metals, to prevent the formation of a surface oxide layer.

DIAMOND TURNING

The term "diamond turning" describes a process by which a special very stiff lathe usually having an air bearing spindle and a single point cutting tool (a diamond) is used to machine components. The parts are usually mounted in a vacuum chunk and are cut on-centre or by interrupted cut, similar to the way metal parts are faced on a conventional lathe. Since the tool-part interaction occurs in a very small contact area, vastly different amounts of material can be removed from adjacent regions of the part. This makes the fabrication of small parts with complex-contoured surfaces possible. High quality diamond turned surfaces appear shiny to the unaided eye, but show a grooved pattern from the cutting tool when observed in a differential interference contrast microscope (Decker *et al.*, 1978). The diamond is in conjunction with mechanical polishing, however, chemical polishing can help to prevent the formation of the surface layers (amorphous in nature) produced by mechanical polishing, and thus preserve the structure of a crystalline material and also get rid of the tool marks.

PRECISION GRINDING

Precision grinding ranks in between diamond turning and polishing in many aspects. Compared to diamond turning, the position of the cutting edge of the tool is less certain. At any

time, anything from one to many grains are in contact with the part. Grinding wheels tend to be compliant and wear (Shaw, 1972), the effects which make it more difficult to achieve the desired form accuracy than with diamond turning. Besides these disadvantages, there are some notable advantages of precision grinding over diamond turning. With small wheels and depths of cut it can be used to work brittle materials such as ceramic and glass in a ductile fashion (Bifano *et al.* 1988).

FLOAT POLISHING

Float polishing is a form of hydrodynamic polishing that consists of a combination of fluid erosion and chemical reaction. During polishing there is no physical contact between the specimen and the polishing lap. Float polishing has been shown to produce very smooth flat surfaces with low subsurface damage, and to have a very high removal rate if the surface figure matches that of the lap. The disadvantages of the process are that only flat surfaces can be polished, they need to be nearly optical flat before starting, and only small pieces can be polished because high pressure is needed between the specimen and the lap.

2.4 Measurement of Surface Microtopography

The purpose of surface measurements is on one hand to control the quality of products and to evaluate the manufacturing processes, and on the other for tribologists, it is to assess surface finish in order to set up a relationship between the surface topography and its tribological property to obtain optimised surface textures and manufacturing processes for various surface function needs.

In general, surface textures are very complex, because surface finishing processes such as polishing, grinding, and shot blasting, are statistical by nature. To characterise such surfaces, two types of statistical descriptions are used: parameters, which attempt to quantify some aspects of surface with a single number, and statistical functions, which give, by their nature, an array of information about the surface. Details of some commonly used descriptions are discussed as follows.

2.4.1 Surface parameters

There are a great variety of surface parameters, many of which have been developed to characterise the function of engineering surface for particular applications. Some typical parameters appear in the national standard (BS 1134: PART 1: 1988). Nevertheless, surface parameters can

generally be classified as height parameters, wavelength parameters, shape parameters, and combinations of these, known as hybrid parameters.

Height parameters

The most common statistical descriptors of surface height are the mean line average roughness, R_a, and the rms (root-mean-square) roughness, R_q. They are given by the following formulas, shown in integral and digital forms

$$R_a = \frac{1}{L}\int_0^L |y(x)|dx = \frac{1}{N}\sum_{i=1}^{N}|y_i| \tag{2-2}$$

$$R_q = \left[\frac{1}{L}\int_0^L y^2(x)dx\right]^{\frac{1}{2}} = \left[\frac{1}{N}\sum_{i=1}^{N}y_i^2\right]^{\frac{1}{2}} \tag{2-3}$$

where $y(x)$ is the surface profile relative to mean, sampled by the set of N points y_i over the length, L. The R_a and R_q are closely related, and $R_q \approx 1.25\ R_a$ for Gaussian distributed surfaces (Thomas 1982). R_q is more sensitive than R_a to large deviations from the mean line, it is more commonly used in control of optical components and bearing surfaces. And also R_q represents the standard deviation of the height distribution, and it links between the random process techniques such as autocorrelation and power spectrum. This will be discussed later.

Wavelength parameters

Wavelength parameters are used to characterise the spacings of the peaks and valleys of the surface. Spacings or wavelengths are often characteristic of the process that formed the surface, such as the shot size used for abrasive blasting, the grit size of a grinding wheel, or the feed of a tool. A typical wavelength parameter, recognised by the International Organisation for Standardisation (ISO 1984), is the mean peak spacing S_m, defined a surface profile as the average spacing between two successive negative crossings of the mean line, as shown in Fig. 2.3.

$$S_m = \frac{1}{n}\sum_{i=1}^{n} S_{mi} \tag{2-4}$$

Spragg and Whitehouse (1970) have shown that, for most profiles encountered in practice, the average wavelength of a profile may be calculated from the equation: $\lambda_q = 2\pi R_q / \Delta_q$, where Δ_q is the rms slope. The calculation of the slope will be given later.

Shape parameters

Some surfaces may have same R_a and average wavelength, but have different shapes and thus may perform differently in the same application. Other parameters such as the skewness and kurtosis involving higher order moments about the mean line help to quantify the difference between these surfaces. The skewness, R_{sk}, is a measure of the symmetry of the profile about the mean line. It is defined as

$$R_{sk} = \frac{1}{NR_q^3}\sum_{i=1}^{N} y_i^3 \tag{2-5}$$

A positive value indicates sharp peaks, and a negative value indicates flat peaks but sharp/narrow valleys. The latter type of surfaces ensures good stability against wear because it will have a large contact area and thus low bearing pressure.

Kurtosis, k, is a measure of the peakedness of the surface which is given by

$$k = \frac{1}{NR_q^4}\sum_{i=1}^{N} y_i^4 \tag{2-6}$$

A symmetric Gaussian distribution has a kurtosis of $k = 3$. Higher values are more violently peaked.

Hybrid parameters

The slopes and peak curvatures of surface profiles are concerned in tribology for assessing the contact area between surfaces. The slope is defined as the spatial derivative of the height profile while the curvature is the second derivative of the profile.

Normally these two parameters are described by their rms values. There are many algorithms for differentiation such as three, five or seven point polynomial fitting algorithms and some mathematical approaches, but the most appropriate algorithm depends on the nature of the signal and the reason for making the measurement (Chetwynd 1978, 1991). For equi-spaced data, the three and five point formulae are

$$_3\dot{y}_i = (y_{i+1} - y_{i-1})/2h \tag{2-7}$$

$$_5\dot{y}_i = (-y_{i+2} + 8y_{i+1} - 8y_{i-1} + y_{i-2})/12h \tag{2-8}$$

where h is the spacing and y_{i-m}, y_{i+m} (m =1, 2) are the mth data points before and after the point at which the derivative is taken. A peak is defined as any point higher than its immediate neighbours,

i.e. $y_i > y_{i-1}, y_{i+1}$, the peak curvature, $\ddot{y}_i$, can be calculated by fitting a parabola through the three points as:

$$\ddot{y}_i = (2y_i - y_{i-1} - y_{i+1}) / h^2 \tag{2-9}$$

There is strong influence of the sampling interval, h, on the calculated values of the rms slope and rms peak curvature, especially the latter. The number of counting peak points may be increased by oversampling because of the quantisation error from an analogue to digital converter. A solution is either following the Nyquist sampling criterion or using the contact diameter calculated from the Hertzian contact formula.

2.4.2 Statistical functions

There is much research on statistical descriptions of the properties of surface profiles (Nayak 1971, Bendat and Piersol 1971, Whitehouse and Achard 1969, 1970). Four important statistical functions are the amplitude density function or height distribution, the bearing area curve, the power spectral density, and the autocorrelation function. A brief discussion is given here.

Height distribution and its relationship to the parameters

Surface textures can be adequately described in terms of the height distribution function of the profile. It is given by

$$F(y) = \int_{-\infty}^{y} p(y)dy \tag{2-10}$$

where y refers to the heights of the profile measured from the centre and $p(y)$ is the probability density function of the distribution of these heights.

The nth moment of the distribution curve $p(y)$ about the mean is defined as

$$M_n = \int_{-\infty}^{\infty} y^n p(y)dy \tag{2-11}$$

therefore, we have the following relationships

$$\left[\int_{-\infty}^{\infty} y^2 p(y)dy\right]^{\frac{1}{2}} = \sigma = R_q \tag{2-12}$$

$$2\int_{0}^{\infty} yp(y)dy = R_a \tag{2-13}$$

$$R_q^{-3}\int_{-\infty}^{\infty} y^3 p(y)dy = R_{sk} \qquad (2\text{-}14)$$

$$R_q^{-4}\int_{-\infty}^{\infty} y^4 p(y)dy = k \qquad (2\text{-}15)$$

Autocovariance and Autocorrelation (ACF)

The autocovariance is calculated by an overlap integral of shifted and unshifted profiles by

$$R(\tau) = \lim_{L\to\infty}\frac{1}{L}\int_0^L y(x)y(x+\tau)dx \qquad (2\text{-}16)$$

and the autocorrelation function is simply the $R(\tau)$ normalised by the profile sample variance, σ^2 or the square of rms roughness, R_q. Here L is the total sampling length.

Power spectral density (PSD)

PSD decomposes the surface profile into its spatial Fourier component wavelengths in a form by

$$PSD(\omega) = \lim_{L\to\infty}\frac{1}{L}\left|\int_0^L y(x)\cdot e^{-i\omega x}dx\right|^2 \qquad (2\text{-}17)$$

and also it can be calculated from the autocovariance function, $R(\tau)$ by

$$PSD(\omega) = \frac{1}{\pi}\int_0^{\infty} R(\tau)\cdot e^{-i\omega\tau}d\tau \qquad (2\text{-}18)$$

Therefore, as measures of the lateral structure of surfaces, the PSD and the ACF are useful in different ways. The PSD is useful for studying the strengths of various periodic components in the surface profile and for comparing these with the strength of the broad spectrum of random components. The ACF is useful for observing directly the lateral extent of the random structures by studying the decay in the function near zero shift.

2.4.3 Fractal approach

Fractal geometry was found by Mandelbrot (1967) when he showed that for decreasing unit of measurement, the length of a natural coastline does not converge but, instead, increases monotonically. On plotting the length L as a function of the unit of measurement ε on a log-log plot, he found a simple relation of the form $L \sim \varepsilon^{1-D}$. He finally made an interesting conclusion that the

real number D associated with every coastline is the dimension of the coastline (The coast of Britain is of the dimension of 1.24). This study marked the origins of fractal geometry which has now found numerous applications in characterising and describing a wide range of phenomena in science and engineering (Mandelbrot, 1982).

It takes some time for scientists to accept the idea that ordinary engineering surfaces are fractal in nature; that to present a profile correctly, one should draw a curve which does not possess a derivative at any point. The non-differentiability arises from the fact that if the profile is repeatedly magnified, more and more roughness keeps appearing and the profile is never smooth. In addition, when the profile is magnified, there is striking resemblance in the shape of the probability distribution of the magnified and the original surface. A fractal curve need only be 'self-affine' — the magnifications along and normal to the curve need not be equal. To illustrate these properties, the Weierstrass-Mandelbrot function is given as

$$y(x) = G^{(D-1)} \sum_{n=n1}^{\infty} \frac{\cos(2\pi\gamma^n x)}{\gamma^{(2-D)n}}; \quad 1 < D < 2; \gamma > 1 \qquad (2\text{-}19)$$

where G is a scaling constant and the frequency modes γ^n correspond to the reciprocal of the wavelength of roughness as $\gamma^n = 1/\lambda_n$. This function has properties that the series for $y(x)$ converges whereas that for dy/dx diverges. Although the W-M function appears similar to a Fourier series the difference is that the frequencies increase in a geometric progression instead of an arithmetic progression. The parameter n_1 corresponds to the low cut-off frequency of the profile. Since a surface is a non-stationary random process (Sayles and Thomas, 1978), the low cut-off frequency depends on the length L of the sample by the relation $\gamma^{n1} = 1/L$. The phases of the frequency modes of $y(x)$ are randomised by choosing a non-integer γ and taking its powers to form a geometric series. Research (Majumdar and Tien, 1989) shows that a choice of $\gamma = 1.5$ provides both a phase randomisation and high spectral density.

Since $y(x)$ is composed of a superposition of infinite frequency modes, it is a multiscale function. The multiscale nature of this function is displayed by the power spectrum of the function which provides the amplitude of the roughness at all length scales. The power spectrum of the W-M function is given as (Berry and Lewis, 1980)

$$S(\omega) = \frac{G^{2(D-1)}}{2\ln\gamma} \frac{1}{\omega^{(5-2D)}} \qquad (2\text{-}20)$$

The power spectrum is the most fundamental quantity of a multiscale function because most statistical parameters can be obtained from the spectrum. Details can be referred to Majumdar and Bhushan (1990). It is noted that the form of the fractal spectral density is much like that of the Sales and Thomas (1978) spectral density, especially when the fractal dimension D=1.5.

To have a view of the graphs generated by the W_M function, a programme was written and processed in 'Matlab'. The results are shown in Fig. 2 4 with four different fractal dimensions and the parameters, G, γ and n, are chosen as 1, 1.5 and 1000. It will be seen that the graphs are much similar to some profiles of ground surfaces taken by a stylus instrument.

Using a fractal dimension approach to characterise surfaces is still in its infancy with the link depending heavily on matching with the power spectral density of the profile. In addition, it has been found that not all machined surfaces can be described by the fractal dimensions, the surface produced by random and undirectional process such as grinding and polishing can be reasonably described by a fractal dimension but surfaces with deterministic textures have no fractal dimensions.

2.4.4 Profiling and imaging

The use of statistical parameters results in loss of local information. More and more research has shown the importance of local surface features. Profiling a surface is the basic step of measurements. Most surface parameters that are derived from profiles with deterministic and random textures can often be visually identified from their profiles. Comparison can be easily made from profiles with similar parameter values. Imaging a surface in three dimensions provides a complete measurement of a surface. However these techniques are limited by the range and resolution; probes with large ranges are normally incapable of measuring fine details whereas those scanning microscopes such as STMs and AFMs can only measure a small range of a few to hundreds of micrometers. With the development of new technologies, profiling and imaging will play more and more important role in surface measurements, more and more features will be revealed and new parameters will be generated or a complete new description would be generated.

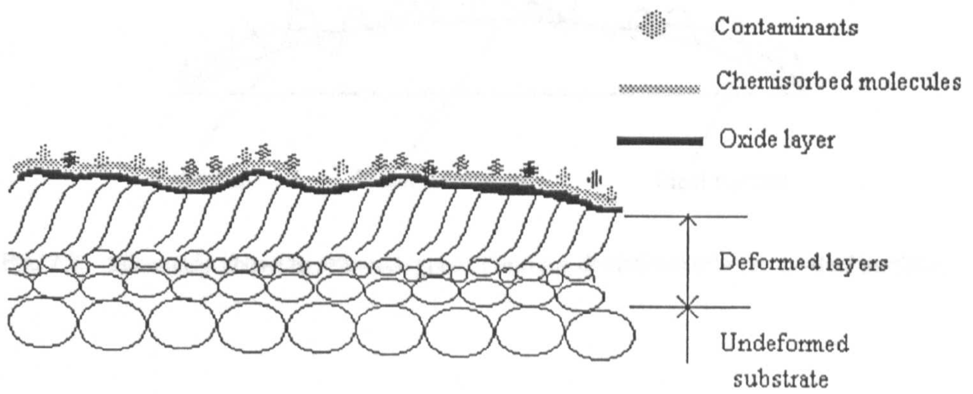


Fig. 2.1 Schematic showing typical surface and subsurface microstructures present in metal.

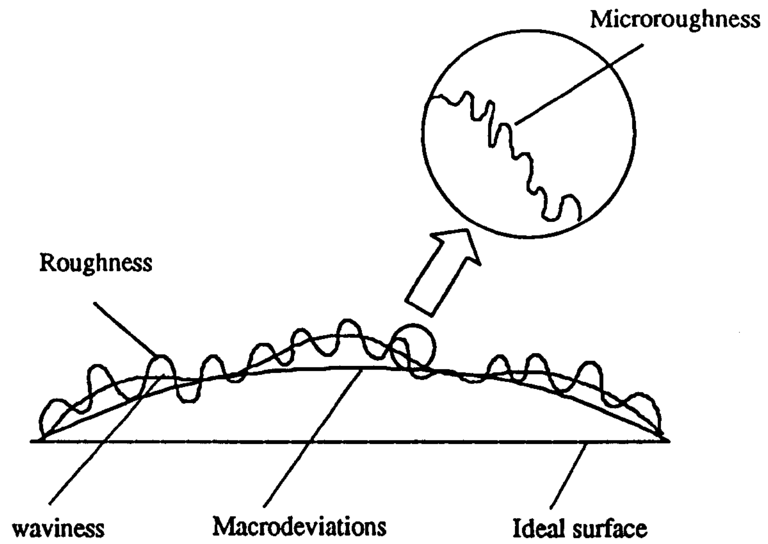


Fig. 2.2 Schematic showing selected types of surface deviations relative to ideal surface.

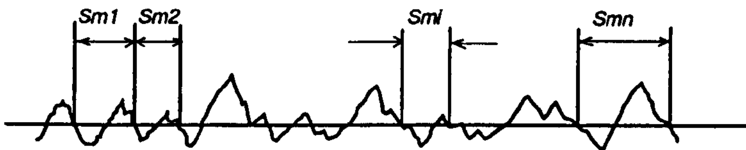
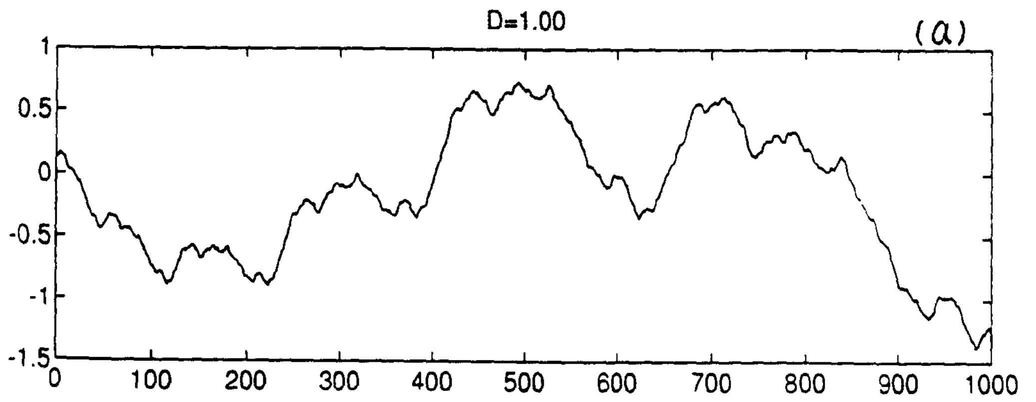
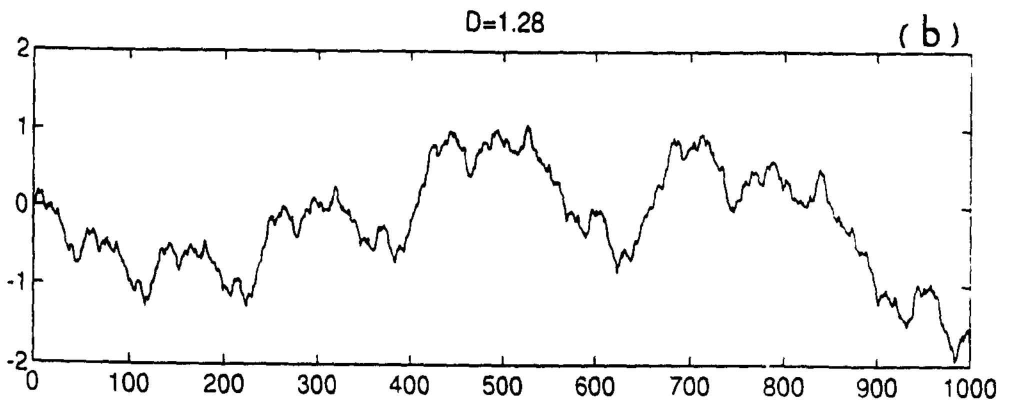


Fig. 2.3 Surface profile showing the ISO definition for the peak spacing, S_m.


D=1.00
(a)
1
0.5
0
-0.5
-1
-1.5
0
100
200
300
400
500
600
700
800
900
1000


D=1.28
(b)
2
1
0
-1
-2
0
100
200
300
400
500
600
700
800
900
1000

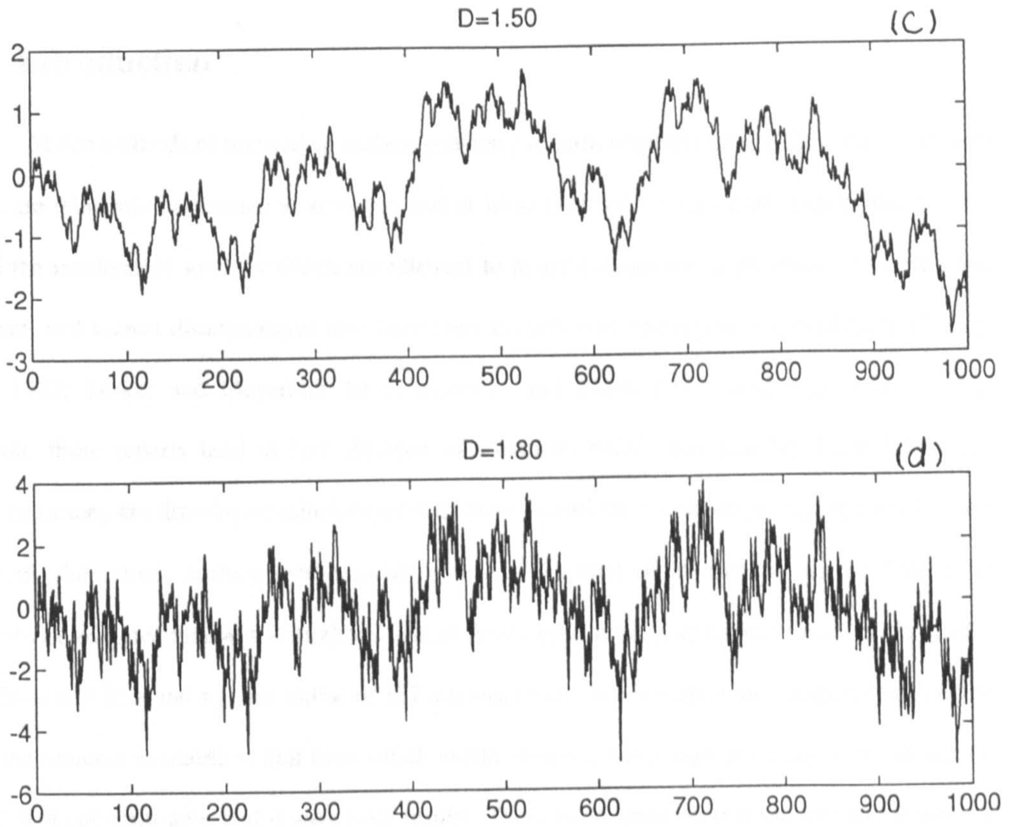


Fig. 2.4 Profiles constructed from fractals: (a) D=1, (b) D=1.28, (c) D=1.5, and (d) D=1.8. The vertical is in an arbitrary unit and horizontal axis is the sample points.

3. ANALYSIS OF STYLUS SYSTEM

3.1 Introduction

Stylus methods of measuring surface geometry remain popular because they are convenient to use, are relatively insensitive to vibration and to some types of surface contaminants and directly record the mechanical surfaces which are relevant to many tribological applications. Nevertheless, there are well-known disadvantages that have been investigated and reported sporadically (Quiney *et al.* 1967; Turker and Meyerhoff 1969; Guerrero and Black 1972; Vorburger *et al.* 1979). However, these reports tend to lack detailed information, which was possibly limited by their facilities, or they are drawing conclusions from extreme situations. For example, Quiney *et al.* made a general claim that "surface measurement with a stylus-type of instrument can no longer be considered a non–destructive test method." But in their experiment, a stylus instrument which had a stylus load of 1.25 g and a stylus radius of 127 μm was used. This is very crude compared with other stylus instruments available at that time which would typically use a load of 60 mg and a radius of 10 μm; it is not surprising that it gave poor results. There is no question that on very soft materials the stylus tip can distort the surface and in some cases damage it. However, it is not clear at all at what conditions the damage occurs and to what degree it occurs. Further it seems a little harsh to call a technique leaving tracks on the surface which are very shallow compared with most random surface features already there "destructive". To rework a well-worn analogy, although a skier leaves tracks in the snow she closely follows the topography of the hillside. To determine whether these tracks constitute functionally significant damage requires a knowledge of the application.

It is quite possible that the stylus damage is due to the combination of direct load and friction at the tip, and also the dynamic behaviour of the stylus mechanism. From the point view of

surface distortion or damage, the stylus load should be as small as possible, but this may lead to the loss of the fidelity of measurement as the stylus may bounce over the surface due to its dynamic and frictional behaviours. 'Fidelity' here has a twofold meaning in that a stylus instrument should both truthfully reflect the surface geometry and not damage it significantly. In this and the following chapters a theoretical and experimental approach is developed to investigate the detailed behaviour of the stylus mechanism, the force interaction and the friction, and finally, a method of optimising the stylus instrument for fidelity of measurement is proposed and evaluated. It is intended that this will provide a framework for stylus type instruments and may also be useful for other cantilever type probes such as scanning force and lateral force microscopes. First of all, in this chapter, the characterisation of the stylus system will be discussed in detail. A typical stylus instrument, a Rank Taylor Hobson Talysurf 5 has been chosen to carry out the research for this purpose. This gives a common context to all the work and so aids evaluation. However, the analysis and proposals require at most minor adjustment to detail for applications to other instruments.

3.2 Mathematical Model of the Stylus System

A typical configuration for a stylus-type instrument is shown in Fig. 3.1(a). It represents a conventional side-acting system. 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 end is the ferromagnetic core for the inductance transducer. The leaf spring system has two functions: one is to contribute the contact force to the measured surfaces, and the other is to constrain the beam at the end-stop when the stylus is not used, to protect it from damage. In an alternative axial gauge system the transducer is directly in line with the stylus, as shown in Fig. 1.1. In both cases the system have been modelled as a second order system (Whitehouse, 1988), see Fig. 3.1(b). For small angular deflections, the movement of the stylus mechanism can be described as

$$m_e\ddot{y}(t)+b\dot{y}(t)+\lambda y(t)+f_0=f_v(t) \tag{3-1}$$

where $y(t)$ is the vertical movement of the stylus and f_0 is the nominal static force. $f_v(t)$ is the vertical reaction force on the stylus, m_e , b and λ are the effective mass, the viscous damping coefficient and the spring rate, respectively. Equation (3-1) can be written in a general form

$$\ddot{y}(t)+2\zeta\omega_n\dot{y}(t)+\omega_n^2 y(t)=K[f_v(t)-f_0] \tag{3-2}$$

where $\omega_n=\sqrt{\lambda/m_e}$ is the natural angular frequency, $\zeta=b/(2\sqrt{\lambda m_e})$ the damping ratio, $K=1/m_e$ the system gain. The time derivatives can be expressed in terms of surface spatial derivatives by

$$\dot{y}(t)=V_h y'(x) \qquad \ddot{y}(t)=V_h^2 y''(x) \tag{3-3}$$

providing that the system traverse velocity V_h remains constant. The prime notation indicates differentiation with the normal displacement, x. Equation (3-2) can then be written as

$$y''(x)+\frac{2\zeta\omega_n}{V_h}y'(x)+\frac{\omega_n^2}{V_h^2}y(x)=\frac{K}{V_h^2}[f_v(x)-f_0] \tag{3-4}$$

The right side of the equation (3-4) represents the resultant force between the tip and the surface and the left side is the response of the stylus mechanism in which $y''(x)$ is considered to be local surface curvature and $y'(x)$ is local slope.

3.3 Analysis and Measurement of the Characteristic Parameters

3.3.1 The spring rate

Stylus instruments generally use the deflection of either a cantilever or a helical spring to bias against the surface. In either case behaviour is close to that of a linear spring and there will be a variation on force proportional to its deflection. A low stiffness spring system is desirable as it exerts small force to the surface and also enables high sensitivity and thus, the high resolution. However, it may also lead to a low resonant frequency that is undesirable for high speed measurements.

The spring force is proportional to the stylus deflection and is independent the velocity at which it is traversed. Standards for conventional stylus instruments require a spring rate of less than 35 N m^{-1} (BS1134, Part 1 1988). This force can be positive and negative about its balance position, and it will add to or subtract from the nominal force. For a typical nominal force of 0.7 mN the reaction force affected by the spring force alone (it happens when a stylus is in standstill or traversing at a very slow speed) would change from 0.3 mN to 1 mN as the stylus is deflected 10μm. The spring rate should be a constant value within its specified range, however, the combination of a tension spring and a leaf spring in some instruments adds complexity and sometimes it shows non–linear behaviour.

The spring rate of a stylus system can be determined simply by measuring the stylus deflection at a given force or weight. If a force transducer or a strain gauge load cell is used, it is necessary that the stiffness of the force transducer is much higher than the stylus so that the deflection of the force transducer is negligible. To measure the spring rate, a typical type of the stylus instrument, Talysurf 5, was lowered onto the surface of a precision load cell (Maywood Instruments, type 4934) and then moved vertically using the manual control on the instrument so that it was displaced across its full range. As indicated in Fig. 3.1, the stylus arm is pulled against a knife-edge pivot by a bias spring that consists of a leaf spring and an adjustable coil spring for fine tuning of the spring force at the zero output of the displacement transducer. Consequently, the actual contact force will vary in relation to the position of the stylus about this zero point. There is a range of magnifications available on the instrument and, because the zero point may not correspond to the same relative displacement of the stylus for each setting, the test was performed for each magnification. A typical plot of the output from the load cell as the stylus is translated across its full range and back again at a magnification of 2000 is shown in Fig. 3.2. The forward scan from a full scale deflection of 12.5 μm to −12.5 μm is represented by the full line and the return is shown dashed. Although the instrument is within the specification, there is clear hysteresis in this plot and a rather steep change in force between zero and +10 μm. Measurements at different magnifications indicate a similar trend. The exact reason for this is not quite clear but it could arise from non-linear

behaviour of the leaf spring support which may carry compressive forces and so snap between equal energy geometries or hysteresis with the knife edge pivot or perhaps minor misalignments of the pickup mechanism accumulated over its considerable life. On the other side of the zero point of the pickup, there is a much slower variation of the force, indicating a spring rate of approximately 10 $\mu N\ \mu m^{-1}$ or 10 $N\ m^{-1}$. As will be confirmed later in section 4.3.2 by a more precise method in which a magnet-coil force transducer is used, the value is 11.2 $N\ m^{-1}$ over this region.

3.3.2 The damping coefficient

The damping effect is normally expressed by a ratio of the damping coefficient to the critical damping value in a form of

$$\zeta = \frac{b}{b_{cr}} = \frac{b}{2\sqrt{\lambda m_e}} \tag{3-5}$$

where b and b_{cr} are damping coefficient and critical damping value respectively. For a second order system such as the stylus beam, the damping ratio determines the dynamic response of the system. It is summarised in table 3.1.

Table 3.1 Dynamic behaviour of a stylus system.

Value of damping ratio	Response of the system
$\zeta = 0$	*No damping* Continual oscillation at frequency and constant amplitude, $\omega = \omega_n$ [1]
$\zeta < 1$	*Underdamping* Exponentially decaying oscillation at frequency ω, $\omega < \omega_n$
$\zeta = 1$	*Critical damping* Oscillation just ceases, $\omega = 0$
$\zeta > 1$	*Overdamping* No oscillation, exponential settling

[1] ω_n is the undamped natural frequency of the system.

A system with no damping or large damping ($\zeta > 1$) is not suitable for measuring purposes because it is either unstable or too sluggish. A stylus system, normally, has a very low damping ratio of much less than 0.1. This may be merely convenient design since no discrete damping element is included or may be to obtain a higher followable frequency for $\omega \approx \omega_n$. But the low damping system invariably introduces large overshoots (about 94% over for a step input when $\zeta = 0.02$), resulting in a distorted profile.

Experimentally, the damping ratio can be determined from the response of the stylus system to a free oscillation about its balance position. This was done by lifting the stylus to its balance, or midpoint position using a home made force transducer (details are discussed in the following chapter). The stylus was then slightly displaced and released manually and its consequent oscillatory response was recorded. The successive peak magnitudes of the response can be expressed by

$$\exp\left(-\frac{n\zeta\pi}{\sqrt{1-\zeta^2}}\right) \tag{3-6}$$

where n is the nth peak. The damping ratio can then be calculated using two successive peak values, p_n and p_{n+1}, by using log decrement, δ, defined as

$$\delta = \ln\left(\frac{p_n}{p_{n+1}}\right) = \frac{2\pi\zeta}{\sqrt{1-\zeta^2}} \tag{3-7}$$

$$\zeta = \frac{\delta}{\sqrt{(2\pi)^2 + \delta^2}} \tag{3-8}$$

For a lightly damped second order system, such as the stylus system, the above method gives a good estimation. This method has the advantage that it is purely ratiometric measurement and, as such, the magnitude of the *input* (which can be by an initial release or by lightly taping the stylus) does not have to be measured.

3.3.3 System resonance and effective mass

System resonance determines the bandwidth of its dynamic response, and thus the maximum followable frequency. An ideal design aim for a stylus system would be to scan as rapidly as possible without the sensor dynamics unduly distorting the fidelity with which the surface is measured. This requires that the resonance of the stylus assembly be as high as possible given other constraints such as the unavoidable mass of the sensor element and the spring stiffness being low enough that the contact force does not vary significantly with stylus position.

There are many ways to measure the resonance of a mechanical system. The system can be excited by an impulse applied using a hammer or simply by flicking it with a finger, and the decaying response is monitored. It is important that the initial impulse is not too sharp, in which case higher harmonics are accidentally excited, resulting in a rather confused combination of the basic and harmonic signals. For the stylus system the initial release method used in measuring damping ratio is used with which a near perfect decay oscillation can be easily obtained. The damped natural frequency, ω_d, can be determined by counting the time interval, Δt, between the two successive peaks as

$$\omega_d = \frac{2\pi}{\Delta t} \tag{3-9}$$

Thus the resonance is given by

$$\omega_n = \frac{\omega_d}{\sqrt{1-\zeta^2}} \tag{3-10}$$

The recorded system response of the Talysurf 5 is shown in Fig. 3.3(a). The resonant frequency can also be obtained from the FFT analysis, as shown in Fig. 3.3(b). The stylus pickup of the Talysurf 5 has a resonance of 178 rad s^{-1} and a damping ratio of 0.01.

Since the natural angular frequency ω_n and the spring constant λ are known, the effective mass is easily calculated from the following equation

$$m_e = \frac{\lambda}{\omega_n^2} \tag{3-11}$$

In the case of $\omega_n = 178$ rad s^{-1} and $\lambda = 11.2$ N m^{-1}, the effective mass is 0.353 g. The effective mass here is defined as the mass of a weight held at one end of a weightless beam. A typical stylus beam is a hollow tube made of aluminium. The distributed mass of the beam does not contribute much to the effective mass and the major contribution comes from the concentrated mass such as stylus tip and the ferrite core. A typical value for the ferrite core is 0.185 g, which is about half of the total effective mass. Therefore, to reduce the effective mass so as to increase the system resonance, a different transduing element with a light weight, for example a capacitive transducer, may be a solution.

3.4 Contact Forces in Stylus System

The contact forces between a stylus tip and the surface being measured are classified into static and dynamic forces. Ideally, for a contact type displacement sensor the contact force should remain constant during measurements. However, in reality, the contact force varies due to the spring rate and dynamic reactions.

3.4.1 Static force

The nominal loading force at the stylus is the force applied by the bias spring when the sensor is stationary and positioned at its balance point, that is, $\ddot{y} = 0$, $\dot{y} = 0$, and $y = 0$, thus the static reaction force is equal to the nominal force, $f_w = f_0$. When a stylus is deflected from its neutral position, an extra force from the surface is added to the stylus, and the static reaction is given by

$$f_w = f_0 \pm \lambda |y| \tag{3-12}$$

The plus sign is for upward movement of the stylus.

The static force here is defined to be time-independent and is normally evaluated when a stylus is stationary or is traversed at a very slow speed at which the dynamic reaction can be negligible.

3.4.2 Dynamic force

Generally, the response of a stylus system to the reaction force between the stylus tip and the surface has three components corresponding to acceleration, damping and spring forces. The dynamic characteristic of a stylus is determined by the inertial and the damping aspects because they are varying with time. The damping force is assumed to be proportional to the vertical velocity of the stylus which is given by the product of the stylus traverse speed V_h and the first derivative, or slope, of the surface. The acceleration force is proportional to the product of the square of the traverse speed and the curvature of the surface profile, providing that the traversing speed of the stylus remains constant. The magnitude of the dynamic force depends on the dynamic input from the surface being measured. This input is the combination of the stylus traverse speed and the surface features. Typically, the traverse speed range for Talysurf 5 is from 0.05 to 1 mm s^{-1}, and thus we can expect a factor of 20 in the variation of the damping force and 400 in that of the acceleration force over its speed range for nominally similar profiles. These dynamic effects are sometimes so significant that they cause the stylus to jump from the surface. On the other hand, when the dynamic force is in the same direction as the nominal loading force, the stylus will exert a much larger than expected force on the surface, which may lead to its damage.

3.4.3 Estimation of the Magnitude of the Dynamic Forces

The magnitude of the dynamic force can be evaluated in two particular situations, lift off and recontact with the surface. The former gives a false measurement of the surface and the latter may cause surface damage.

(1) Point of separation of the stylus

Fig. 3.4 shows the kinematic analysis of the motion of the stylus over the surface profile. The stylus vertical motion is assumed under an elastic spring with a preload δ to give the nominal force and the stiffness of the spring is λ. According to the profile of the surface the stylus will move

with a speed $\dot{y}$ and an acceleration $\ddot{y}$ in the vertical direction. The equation of the motion of the stylus is

- When the stylus moves upwards

$$m_e\ddot{y}+b\dot{y}=f_v-\lambda(\delta+y) \quad \text{(3-13)}$$

- When the stylus moves downwards

$$m_e\ddot{y}+b\dot{y}=\lambda(\delta+y)-f_v \quad \text{(3-14)}$$

The separation of the stylus from the measured surface occurs during the upward motion, when the reaction force $f_v=0$. Fig. 3.5 shows the path of the stylus over the surface profile in which the stylus is lifted off at the point s, where

$$m_e\ddot{y}+b\dot{y}+\lambda(\delta+y_s)=0 \quad \text{(3-15)}$$

If the measured surface is of a form $y=g(x)$, then, $\dot{y}=\dot{x}g'(x)=V_h g'(x)$ and $\ddot{y}=V_h^2 g''(x)$, and the above equation becomes

$$m_e V_h^2 g''(x_s)+bV_h g'(x_s)+\lambda(\delta+y_s)=0 \quad \text{(3-16)}$$

The values of y_s and hence x_s represent the point on the measured surface at which separation occurs. It depends on V_h^2, λ, m_e and the geometry of the surface.

(2) Path of the stylus after separation

The path of the stylus after separation is governed by the equation

$$m_e\frac{d^2\bar{y}}{d^2t}+b\frac{d\bar{y}}{dt}=-\lambda(\delta+y_s+\bar{y}) \quad \text{(3-17)}$$

The solution to this motion is generally given by

$$\bar{y}(t)=-(\delta+y_s)+\{A\cos(\omega_n\sqrt{1-\zeta^2}\cdot t)+B\sin(\omega_n\sqrt{1-\zeta^2}\cdot t)\}e^{-\zeta\omega_n t} \quad \text{(3-18)}$$

where, the $\bar{y}$ represents the false displacement of the stylus, A and B are arbitrary constants depending on the initial positions of the separation.

(a) at $t=0$ $\quad \bar{y}=0$

Therefore $A=\delta+y_s$.

(b) at $t=0$ $\quad \dot{\bar{y}} = \dot{y}_0 = \dot{y}_s$

$\dot{y}_s$ is the vertical velocity of the stylus at the point of separation, $\dot{y}_s = V_h\, g'(x_s)$. Therefore

$$B = \frac{\dot{y}_s}{\omega_n[\sqrt{1-\zeta^2} - \zeta(\delta + y_s)]}$$

The general solution, *i.e* the equation of the path of the stylus is

$$\bar{y} = -(\delta + y_s) + \left\{ (\delta + y_s)\cos(\omega_n \sqrt{1-\zeta^2} \cdot t) + \frac{\dot{y}_s}{\omega_n[\sqrt{1-\zeta^2} - \zeta(\delta + y_s)]} \sin(\omega_n \sqrt{1-\zeta^2} \cdot t) \right\} e^{-\omega_n \zeta t}$$

(3-19)

where, $t = \dfrac{\bar{x}}{V_h}$.

(3) The maximum lift $\bar{y}_m$

The maximum overshoot of the stylus after separation can be obtained from the conservation of energy. There will be no external force acting on the stylus system shown in Fig. 3.5 after the separation. At the start point of the lifting, the system has a kinetic energy and a potential energy which are expressed by terms of $\frac{1}{2} m_e\, \dot{y}_s^2$ and $\frac{1}{2}\lambda(\delta + y_s)^2$. At the maximal height point the kinetic energy reaches zero and the potential energy becomes $\frac{1}{2}\lambda(\delta + y_s + \bar{y}_m)^2$. There is some energy loss due to the damping in the system, it is defined by the product of the magnitude of the damping force and the distance through which the stylus moves, that is $b\dot{y}_s\bar{y}_m$. According to the law of conservation of mechanical energy, we have

$$\frac{1}{2} m_e \dot{y}_s^2 + \frac{1}{2}\lambda(\delta + y_s)^2 = \frac{1}{2}\lambda(\delta + y_s + \bar{y}_m)^2 - b\dot{y}_s\bar{y}_m \tag{3-20}$$

Therefore

$$\bar{y}_m = \frac{-\lambda(\delta + y_s) + b\dot{y}_s + \sqrt{[\lambda(\delta + y_s) - b\dot{y}_s]^2 + \lambda m_e \dot{y}_s^2}}{\lambda} \tag{3-21}$$

The same result can also be obtained by differentiating the equation (3-19).

(4) Determination of $\bar{x}_m$

At the point of maximum lift, $(\bar{x}_m, \bar{y}_m)$, the stylus speed is zero, then by differentiating Eqn. (3-19) the time interval can be obtained and hence $\bar{x}_m$. A rough estimation of $\bar{x}_m$ can be achieved from the free oscillation relationship, thus

$$\bar{x}_m = \frac{1}{4} V_h \frac{2\pi}{\omega_n \sqrt{1-\zeta^2}} \tag{3-22}$$

3.4.4 The magnitude of the reaction force

The acceleration, damping and static forces of the stylus will be supported by the surface during the stylus traversing on it. However, these forces do not all reach their maxima at the same time due to the phase differences. The damping force experiences the maximum and minimum values at the steep upward and downward slopes, while at these positions the acceleration force is close to zero. The acceleration reaches the maxima and minima at valleys and peaks, at which there is no damping force. The relationships of position, velocity and acceleration of a sinusoidal signal are illustrated in Fig. 3.6.

The magnitude of the dynamic reaction force depends mainly on the acceleration, thus the curvature of a surface because this value can be very large under certain circumstances, for example where at very sharp peaks and narrow valleys. The minimum of the reaction is zero when the stylus is lifted from the surface, and it occurs when the dynamic force is equal to the static force. The maximum reaction force, on the other hand, occurs when the dynamic force adds to the static force during the downward tracking. There is no direct method to measure this force. Thus many researchers have assumed that the maximal reaction is twice the static force, Ajioka (1974). However, there is no real justification for this estimation, and the maximum dynamic force is not really limited by the static force. It depends on the traversing speed of a stylus and the local curvature of the surface valleys. To evaluate the magnitude of this maximal reaction force for a particular surface, a simple way is to vary the tracking force until the lift–off behaviour ceases. Then the maximal reaction force will be the tracking force plus the static loading force. Experiments show that even at a loading force of 1.55 mN lift-off still exists when a stylus is traversing on a fine

ground copper specimen (discussed in the following chapter). As a consequence, the maximum reaction force has been estimated as high as 2.25 mN (0.7 mN of the nominal load) during the stylus traversing on this copper specimen.

3.5 Effects of the geometry of stylus tip on measurements

The fidelity of surface measurement by stylus instruments is affected by another source; the finite size and shape of a stylus tip. In any stylus measuring instrument the measured profile will not be an exact representation of the true surface profile (Whitehouse and Archard, 1970). The so called 'effective profile' (BS1134: 1988) recorded by the stylus instrument will be the locus of some point on the stylus, or worse than this, the contact point changes throughout the traverse and there is no way to know where it was. As a result, the radius of curvature of a peak may be exaggerated, while a valley may be represented as a cusp.

Theoretically, to provide a true reproduction of the surface features, the stylus tip radius should be zero and the cone angle should be as small as possible. However, the pick-up cost increases with the decreasing tip radius and, also, as the stylus radius becomes smaller maintenance becomes critical. Another important aspect of stylus instruments is the low force with which the stylus traces the surface. Usually this force is in a range of 0.7 – 1.0 mN. Even with this low force, there is a certain amount of wear on the stylus tip because the area of the contact between the tip and the asperities of the surface is very small and the contact pressure will be very high, for example, a stylus with a tip radius 1 μm will produce a contact pressure about 9 GPa on a copper specimen. This can eventually cause a change in the shape of the stylus tip. The cost of producing a very small radius at the stylus tip needs to be considered together with the errors produced in the measured values such as dynamic effects, tip wear and also the deformation of the surface. A present, the smallest tip size commercially available is 0.1 μm.

The assessment of surface roughness is based on the profile of the surface traced by a stylus. For a very rough surface, the geometrical effect of the stylus on the measurement is negligible

fraction of the total measured roughness, as shown by many studies (Williamson 1947; Radhakrishnan 1970; Thomas & Sayles 1978). However, as the demand for smooth surfaces increases, surfaces are considered in more detail or more locally and such an effect as the stylus geometry has to be taken into account.

It should be noticed that the tips in STM (scanning tunnelling microscopy) and AFM (atomic force microscopy) play the same roles even though they have different sensing mechanisms. This effect is more emphasised because of the relatively small features of surfaces being measured. The effect of the tip geometry for non contact STM/AFM can be considered as a stylus in the conventional stylus instrument whose tip is the true tip size plus the constant gap (Stedman 1987,1988; Keller 1991; Nakagiri & Kaizuka 1990). Therefore the characterisation of a stylus is equally useful to the STM/AFM techniques.

3.5.1 Formation of a stylus trace

Normally, according to the national standard (BS1134: Part 1, 1988), a stylus may have an included angle of 60 or 90° and tip radius of curvature of 2 ± 0.5, 5 ± 1, 10 ± 1 μm. The 60° tip is in a shape of cone and the 90° tip is in a truncated pyramid shape. Those tips are normally made of diamond to reduce wear rate.

When the stylus tip scans over a smooth and slowly changing surface, *i.e.* the tip radius is negligible compared to the surface curvatures, only the very end of the tip makes contact with the surface, and the trace of the tip will give a faithful profile. However, real surfaces are rough, and the radius of curvature is comparable to the tip and sometimes even smaller. The trace obtained will then be quite different from the true profile of the surface. In the extreme case, where the surface is a step with a sharp angle much smaller than the flank angle, the stylus trace is actually the profile of the tip rather than the surface. The process of the trace formation is very complicated. Often it is regarded as 'convolution' between the tip and the surface, which is inappropriate terminology because convolution is a linear process while the probe-sample intermixing is not.

As shown in Figure 3.7 (a), as the stylus tip scans over the sample, the contact point of the tip varies due to the slopes of the sample surface. When the slope of the sample surface is high, the tip touches the sample surface at a point other than the nominal tip position. As a consequence, the apparent edge of the surface feature is shifted with respect to the true edge, and the corners of the trace are more rounded than those in the true profile.

At the contact point, both the tip and sample surface have the same tangent line, that is the slope of the tip surface equals the slope of the true surface. At this moment, the whole body of tip is moving in this direction because the stylus does not roll. This means that the corresponding trace at the end of the tip has same slope or tangent line as the contact point. When the stylus is moved a sufficiently small section on the sample surface, *i.e.* the sample interval, *h*, is small, the relative geometrical positions of the most vicinal points on the trace *A*, *B* and *C*, seen in the Fig. 3.7 (b), should be very similar to those of the points on the tip *a*, *b* and *c* which correspond to the surface slopes.

Let ΔZ_{AB} denote $Z_A - Z_B$, and *h* the horizontal separation of *A* and *B* be a constant in the measurement. The above relationship is then: $\Delta Z_{BA} \approx \Delta Z_{ab}$ and $\Delta Z_{CB} \approx \Delta Z_{bc}$. For a tip with a known shape, the position of the contact point on the tip surface can be estimated from the quantities of ΔZ_{BA} and ΔZ_{CB}. This can be described as:

$$\Delta z = f_Z(\Delta \overline{Z}) \qquad \Delta x = f_X(\Delta \overline{Z}) \tag{3-23}$$

where the $\Delta \overline{Z}$ is a vector composed by the variables ΔZ_{BA} and ΔZ_{CB}. The functions of f_Z and f_X are depended on the shape of the tip. From the figure 3.7(b), it is easy to see that the co-ordinates of the true point of contact are:

$$Z = Z_B + \Delta z \qquad X = X_B + \Delta x \tag{3-24}$$

The equations (3-24) represent the true surface profile, the first term on the right side is the measured profile or apparent trace and the second term is the distortion due to the tip.

3.5.2 Effects on lateral resolution

Because a stylus profile represents a 'pseudo-convolution' of the surface structure with the end form of the stylus tip, the lateral resolution depends critically on the size of the stylus tip and to a lesser degree on the flank angle. Styli are often flattened at the end, the width across the flat essentially determing the lateral resolution. For a spherical tip, the lateral resolution can be smaller than the tip radius, r. Fig. 3.8 shows the geometrical construction when a spherical stylus first contacts a step, which is assumed to be infinitely sharp. The initial contact occurs at a position, Δ, to the left of the step edge, and Δ can be determined by

$$\begin{aligned} \Delta &= \sqrt{r^2 - (r-h)^2} \\ &= \sqrt{2rh - h^2} \end{aligned} \qquad (3\text{-}25)$$

If $h \ll r$, the width of the transition $\Delta = \sqrt{2rh}$. This calculation assumes that no deformation occurs at the single point of the contact. In the case of a 1 μm radius stylus and a step of height 10 nm, the calculated transition width is 0.14 μm. For a truncated pyramid stylus, the transition width is equal to the tip width.

The flank angle of the stylus also affects its resolution. The stylus can not accurately profile over slopes greater than the flank angle (usually ± 45°). This is not much of a problem for most types of smooth surfaces where the average surface slopes are only a few degrees, but it does complicate the measurements of surfaces with steep cracks or holes, such as those often found on ceramics.

3.5.3 Effects on parameter measurements

The effects of the tip geometry on measurements depend on the nature of irregularities on the surface. Generally, for a smooth, slowly changing surface with an average slope smaller than the flank angle of the tip, the effect is only minor distortion which can be negligible. However, some manufacturing processes give deep and narrow grooves whilst others give sharply projecting peaks,

both of which cause the stylus to make distorted profile traces. When surface features are comparable to the tip geometry, the effects can be great.

Practical evaluation of the tip effect on measurement needs different tip radii and specified sample surfaces. There are many difficulties obtaining such tips on specific instruments and surfaces and also the measured effects will be a combination of different causes such as dynamic behaviour, surface deformation *etc*. To simplify the process, a simulation based on a 'true' profile is pursued to estimate the tip effect on the measurement. The 'true' profile is obtained from a copper specimen traced by a Talysurf 5 with a tip radius of 2 μm.

To emulate the motion of the stylus as it traverses the 'true' profile, the process is limited to a two-dimensional representation and the stylus is represented by a circle of radius, r, that moves across the crests of the profile. Dynamic effects are not considered and neither is the deformation of the profile due to the contact stresses exerted by the stylus. It is assumed that the stylus moves across the profile in equal horizontal increments, Δx. As is shown in Fig. 3.7 (a), the stylus tip contacts the surface at the point (x_i, z_i), and the apparent point is (x_i', z_i'). Physically, the stylus makes at least one point contact within the horizontal distance bounded on either side by the stylus radius r, that is the range of $(x_i - r) < x_j < (x_i + r)$. The resulting stylus position would have the maximum ordinate value z_i' of all possible contacts. Mathematically, if the true profile ordinates are considered a function of $z_j(x_j)$, then determing the ordinate of the traced profile is a maximisation process, it can be expressed by

$$z_i' = \max[z_j(x_j) + \sqrt{r^2 - (x_j - x_i)^2} - r] \qquad (3\text{-}26)$$
$$(x_i - r) < x_j < (x_i + r)$$

where the subscript j refers to the points on the profile and the subscript i to the traced profile co-ordinates.

Simulations have been performed with seven different size of tips, they are 2, 5, 10, 20, 50, 70 and 100 μm. When the profile is traced with a 2 μm tip there is no noticeable difference between the true and traced profiles. Slight variation from the true profile is noticed with the 5 μm tip. Large

distortions are observed with the two larger tips. The results with tips of 5, 50 and 100 μm are illustrated together with the original profile in Fig. 3.9. As the stylus radius became larger, the minute irregularities present in the true profile gradually disappeared and the traced profile transformed itself into a curve made up of circular arcs when the stylus radius was large enough to touch only the peaks of the true profile. Any further increase in the radius only reduced the number of contacting peaks and did not alter the basic form of the profile. The peaks and valleys of the true profile vary from 0.1 to 0.2 μm, the resolution for the largest tip of 100 μm radius can be estimated using the equation (3-25) to be 4 ~ 6 μm. The selection of this large tip for evaluation is firstly to emphasise the tip effect and secondly to represent the situations where very fine surfaces are traced by a normal stylus tip.

From the traces it can be seen that the 'convolution' of the stylus with the surface reduces the widths of valleys and increases the widths of peaks in the output profile. The degree of the distortion depends on the nature of the surface irregularities; being greater for locally convex features and less for concave features. To show how this affects parameter measurement, a set of surface parameters was computed from these traces. The results are shown in Table 3.2. The roughness value such as R_a and R_q did not change much with the increase of the tip radius. But looking back the traces shown in Fig. 3.9 the differences do exist. This indicates that the roughness parameters are not sensitive to the detailed variations of the trace as long as the basic form is remained. The rms slope and rms peak curvature, however, do change a lot with the increase of the tip radius. The reduction on two parameters is between 30 and 40% of the original parameters for tips up to 50 μm. As it is expected, the reduction is due to the stylus tip effect. To further investigate the characteristic of the true and traced profiles, the skewness and kurtosis of each profile were calculated. The skewness of the original profile is a positive value and the kurtosis is slightly larger than 3. When the stylus tip is smaller than 10 μm, there are slightly increase on skewness and decrease on the kurtosis. An interesting effect is that with the larger tip size the trends on parameters is reversed, an increase rather than decrease on parameters of R_a, R_q, Δq and Pc was observed. The skewness decreased to near zero, which means the height distribution of the profile

became more symmetrical. The kurtosis varied between 3.4 and 3.5. The shift of symmetry on height distribution can also be seen from Fig. 3.10 in which the true profile is represented by bars, while the other is traced with the tip radius of 100 μm.

In summary, the effect of variation of the stylus radius on measurement is not very great on roughness values measured provided that the variation in the stylus radius is within permissible limits. Among the roughness values, the height parameters showed least reduction as the stylus radius increased. The hybrid parameters such as rms slope and rms peak curvature are decreased with an increase in tip radius. However, this conclusion is limited to certain surfaces. For a very fine surface which surface structures dominant the roughness, the tip size effect will be more significant on measurements.

3.5.4 Checking of the stylus tip

When a stylus tip is applied to measure a fine surface structure such as an integrated circuit in which certain patterns may have steep steps, the effect of the tip geometry has to be removed to obtain an accurate measurement. In this case the geometry of the stylus tip must be known to recover the true profile. Moreover experimental results (Elewa & Koura, 1986) have shown that a worn stylus gave a false indication of smoother surfaces. Further, the calibration of surface roughness instruments by the use of precision reference specimens depends on the stylus radius. It is, therefore, important that experimental methods and algorithms for determing stylus radii be standardised (Vorburger, *et al*, 1979) so as to monitor and control the manufacture of styli.

The stylus size and shape can be measured with an electron microscope, an optical microscope, and mechanically by a sharp edge. Scanning electron microscopes (SEM) have been used intensively for examination of styli. This method can provide submicrometer resolution, which is useful for all available stylus tips. However, the SEM method normally gives a visual image and rarely provides any corresponding data. This can only help us to estimate the size and shape of the stylus tip rather than the quantitative information.

An alternative method for checking the stylus tip profile, quantitatively, is to use a sharp edge step. Early works were using a sharp knife edge to trace the profile of a stylus tip (Vorburger, *et al* 1979, Li 1991), but there were some problems due to the finite stiffness of the knife edge used. Griffith *et al* (1991) proposed a solid edge step, as illustrated in Fig. 3.11. In this approach, a known measurement structure was used on the probe microscope. The structure is a array of polysilicon pillars with undercut sidewalls. As the tip encounters the side, the proximal point stops at the cusp of the upper edge and generates a scan of the probe tip itself. In the resulting image the sides of the pillar form an extended image of the tip. A circular measurement structure yields the simplest data analysis, though other shapes such as squares, are acceptable as long as the shape is well known. This probe tip characteriser consists of large lithographically patterned arrays deposited on an oxidised silicon wafer. This method can be readily applied for determination of a stylus tip. The pillar height can be processed to cater for different ranges of measurements, generally, a few hundred micrometers will be suitable for most applications. The pillar specimen can be made as a standard, therefore, a routine checking would be available.

This method gives very accurate information of the stylus tip used and is necessary for either evaluating a measured result or reconstructing the true profile.

3.5.5 Reconstruction of surface profiles

The problem of extracting or reconstructing the true surface profile or image from the distorted scans has been attempted by several researchers (DeVries & Li, 1985; Keller, 1991, Griffith *et al*, 1991). These discussions were mainly based on STM/AFM images distorted by a finite-size of a scanning tip. On the stylus measurement, there are some discussions on the effects but not much attention has been paid on the reconstruction of true profiles. The reason is, probably, the different applications; an STM/AFM measures the structures of a surface, the local detail information is important, while a stylus instrument measures the roughness, the average or statistical features, and the stylus distortion has not attracted much attentions in a practical sense. However, it should be noted that the recent developed stylus instruments such as the Nanostep (Rank

Taylor Hobson) are capable of measuring nanometer or even subnanometer features. For such an application, reconstruction becomes important.

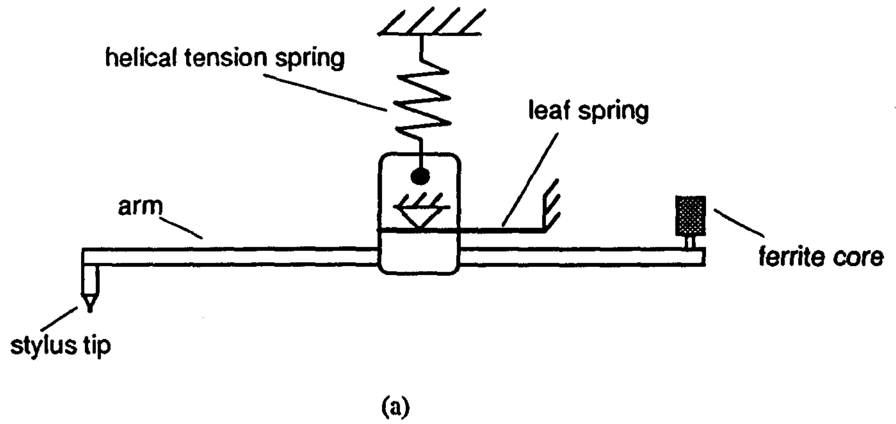
Reconstruction or 'deconvolution' as often called is a very complicated process, if not impossible. Up till now, the progress has been made with a theoretical analysis and computer simulation. There are several algorithms based on different approaches; compensation based on only slope, and compensation based on both the slope and curvature approach. Of these methods, they all invariably require calculations of the first and the second derivatives at each point of the measured profile. It is well known that these derivatives are very sensitive to high frequency noise. The high frequency noise may come from variable sources, for example, the mechanical vibrations, electronic noise, and quantization noise in the data acquisition system. Sometimes, the compensation even gives worse results. Therefore, different smoothing algorithms have to be applied to the measured profile before the computation of the compensation. This is one reason that the simplest method based on only the estimating slope normally gives better overall compensation than the slope plus curvature (DeVries & Li, 1985), this is probably the calculation of second derivative introduces greater errors. The choice of the sample interval in the calculation is also an important factor that could affect the generality of the result. Recent development on the reconstruction has been achieved without requiring the calculations of surface slope and curvature (Wang and Whitehouse, 1995; Keller and Franke, 1993; Sarma *et al.*, 1992).

Finally, it should be noted that the reconstruction approach can only recover part of the distorted profile because, in some cases, the true information has been simply lost completely during the measurement. For example, the portion of the step shape is not profiled in the measurement when the stylus tip traverses over a sharp edge as shown in Fig. 3.11. The fraction of a surface that is unrecoverable will depend on the topography of the surface and the sharpness of the tip.

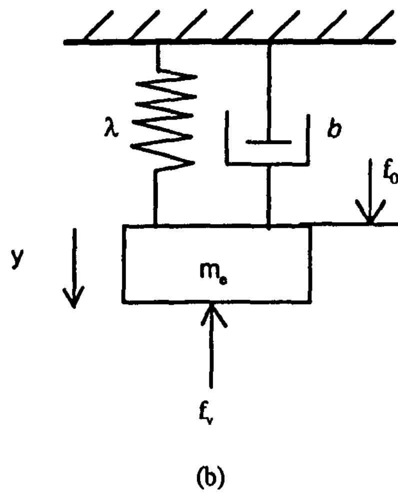
In general, the effect of a finite tip on surface measurement is tolerable to most applications, since the minor distortions are probably within the roughness range. However, there are some applications of which certain patterns needs accurately measured such as precision steps, gratings, and atomic flat surfaces, upon which 'deconvolution' has to be considered.

Table 3.2 The variations in parameters with the different stylus tips.

Tip	Ra (μm)	Rq (μm)	Δq (rad)	Pc (μm^{-1})	Skewness	Kurtosis
original	0.061	0.078	0.068	0.240	0.2312	3.496
2 μm	0.060	0.077	0.066	0.234	0.260	3.454
5 μm	0.059	0.076	0.058	0.187	0.281	3.441
10 μm	0.058	0.074	0.052	0.150	0.315	3.360
20 μm	0.056	0.072	0.046	0.143	0.292	3.430
50 μm	0.057	0.074	0.042	0.157	0.168	3.487
70 μm	0.057	0.074	0.044	0.173	0.042	3.580
100 μm	0.059	0.076	0.047	0.198	0. 022	3.405



(a)



(b)

Fig.3.1. Schematic of the stylus method; (a) the basic assemblage of a stylus system and (b) the model of the stylus system.

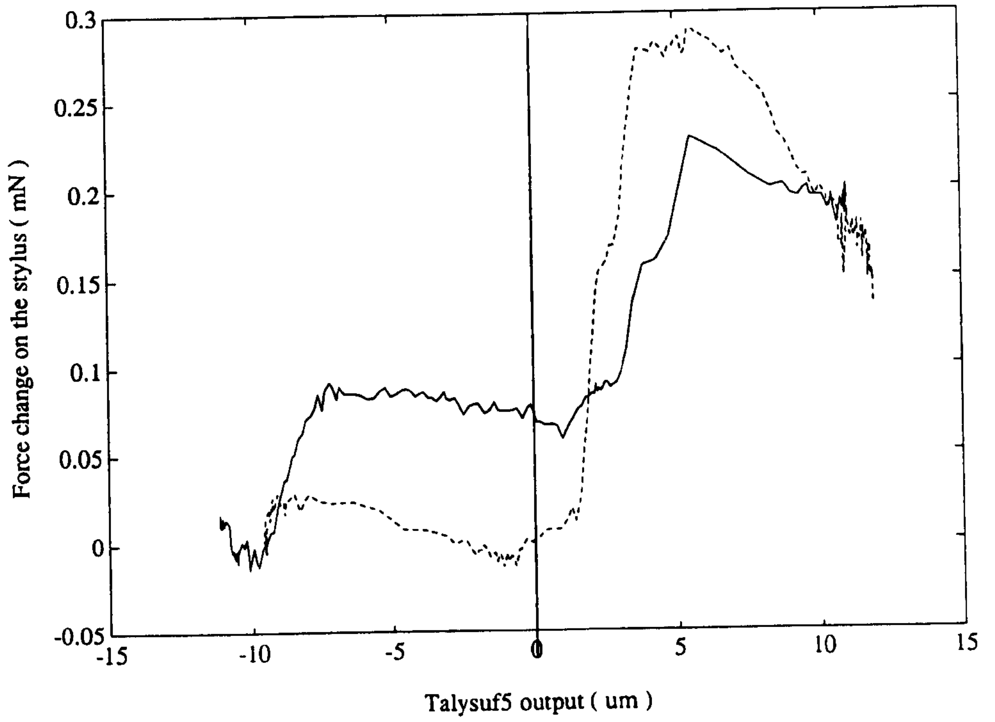


Fig. 3.2 The variation about the nominal value of 0.7 mN of the stylus spring rate against displacement at a magnification setting of 2000.

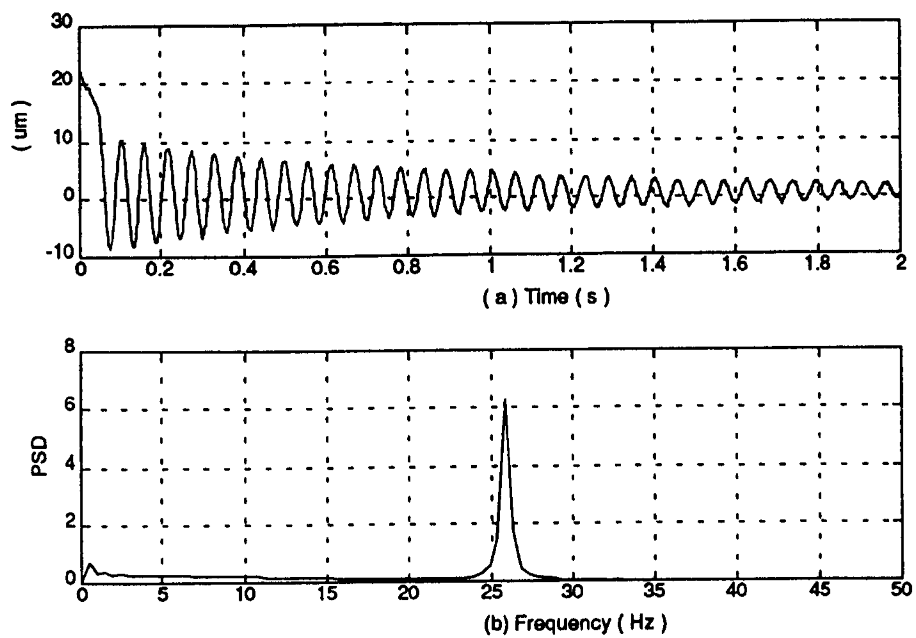


Fig. 3.3 Dynamic response of the Talysurf 5.

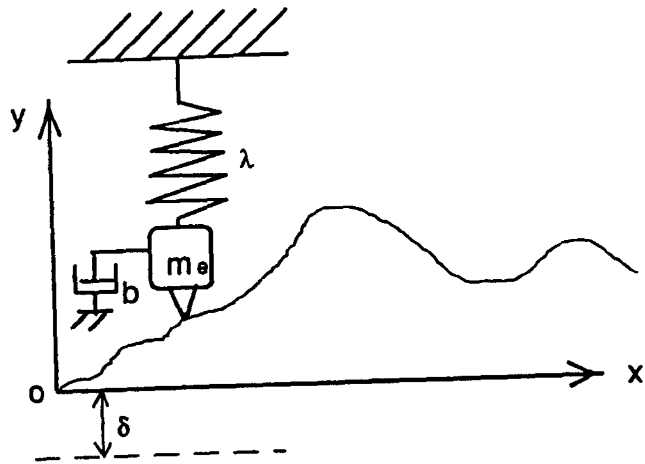


Fig. 3.4 Kinematic motion of the stylus.

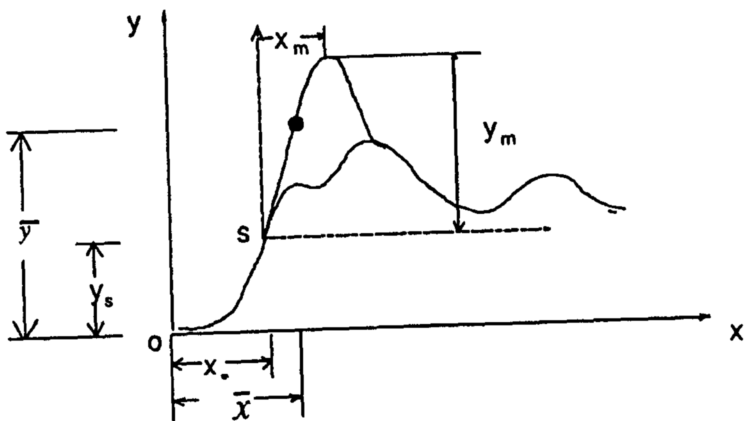


Fig. 3.5 Path of the stylus after separation.

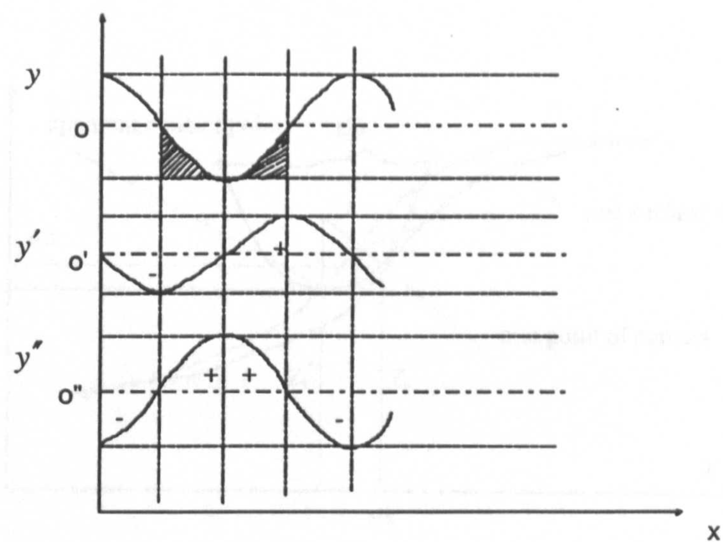
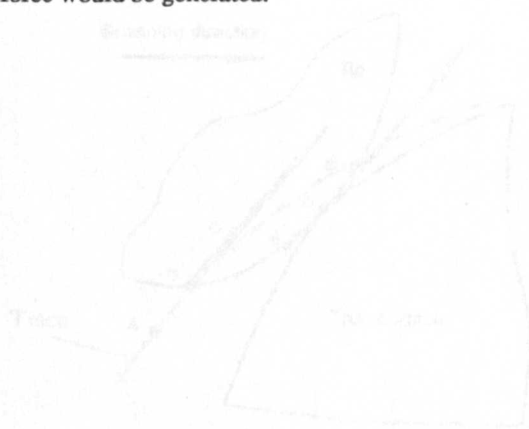
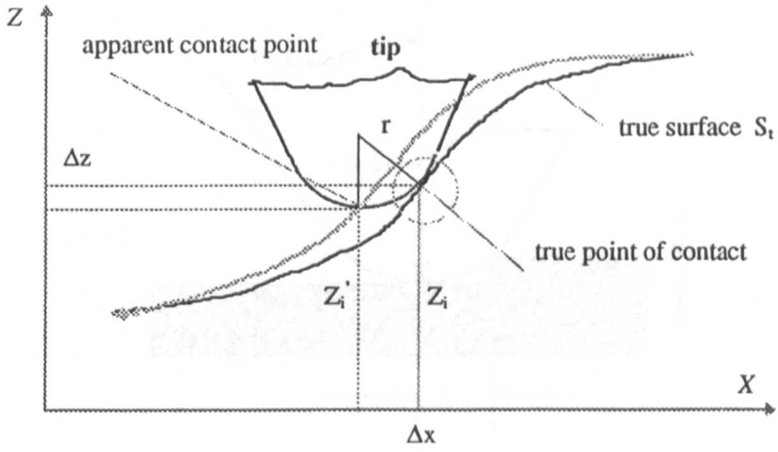
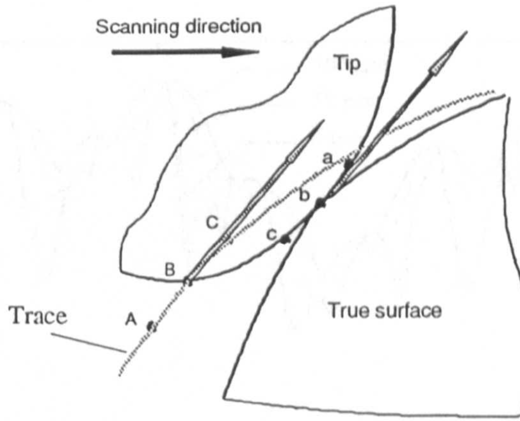


Fig. 3.6 Relationship of the dynamic force components, the shaded area indicates where a large dynamic force would be generated.





(a)



(b)

Fig. 3.7 (a) The relationship between the surface and the stylus trace and (b) the formation of the contact path.

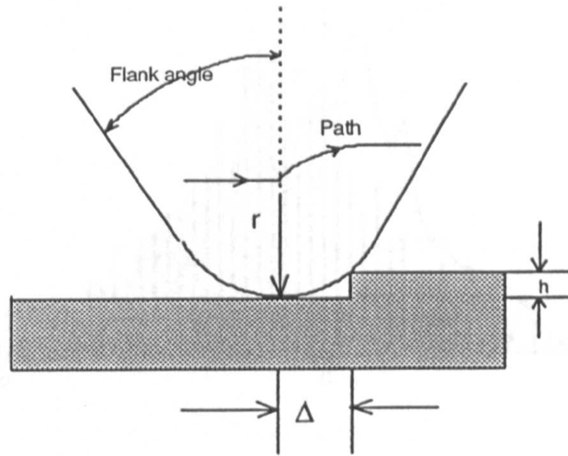


Fig. 3.8 Schematic showing the path traced by a stylus tip traversing over a small, sharp step.

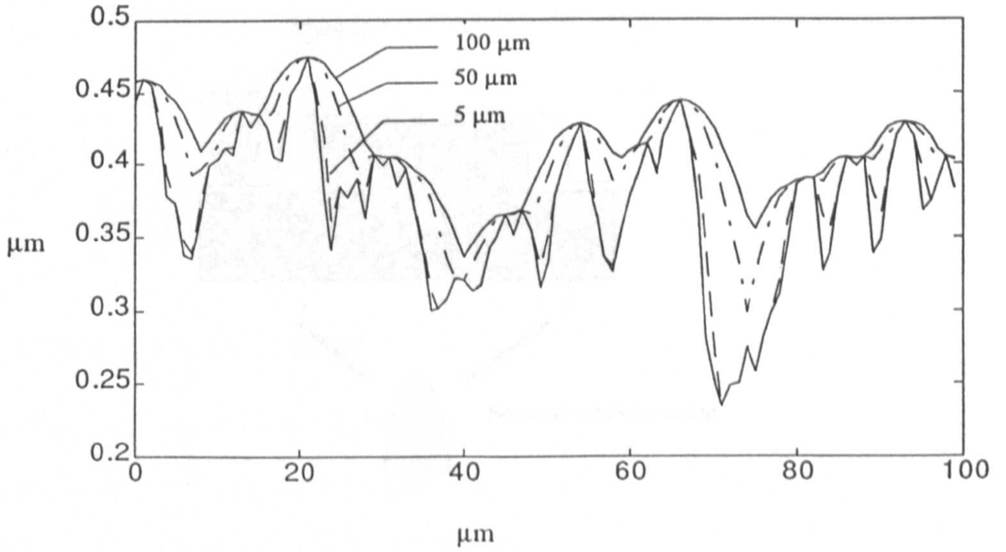


Fig. 3.9 Changes in the profile form when traced with large stylus radii

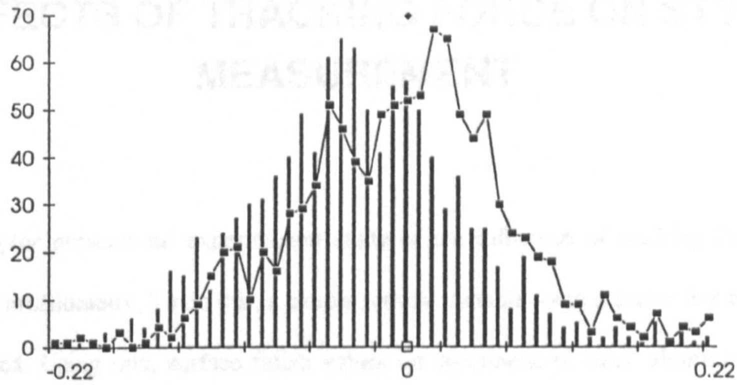


Fig. 3.10 Height distributions of the original profile and the trace with a large tip radius of 100 μm, bars represent the original.

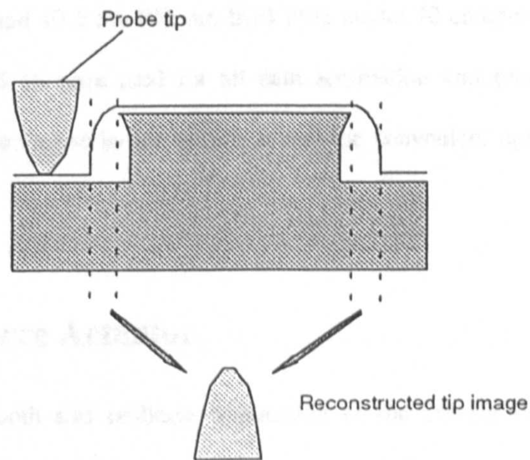


Fig. 3.11 Schematic diagram of the method for characterising a stylus tip.

4. EFFECTS OF TRACKING FORCE ON STYLUS MEASUREMENT

This chapter presents an experimental study of the influence of tracking force on signal fidelity in stylus profilometry. To do this, a simple retrofit variable force actuator has been designed, built and assessed. Using this, surface finish values on specimens of steel, aluminium, brass and copper have been measured over a range of stylus loads and traverse speeds. Analysis of these traces has been carried out by evaluation of the surface finish parameters, R_a, R_q, rms slope and rms peak curvature. Additionally, a qualitative analysis of surface damage was performed using scanning electron microscopy.

All experiments cited in this thesis were carried out in a temperature and humidity controlled laboratory at 20 ± 1 °C and 40 ± 5% RH. An IBM PS/2 model 70 computer and a data acquisition board of Metrabyte DAS-16 were used for all data acquisition and processing. The acquisition software was specially developed by the author to provide convenient operation under the typical profilometry conditions.

4.1 Variable Force Actuator

An accurate, smooth and co-linear application of the driving force is essential for the investigation of the influence of tracking force. Using a standard mass has disadvantages that, first, the availability of milligram masses is limited and, second, the effective mass of the system being tested will vary accordingly. Therefore, an electromagnetic drive is chosen for the purpose of this research. It has the distinct advantages of linear operation, low hysteresis, direct electrical control and the absence of a mechanical linkage between the force actuator and the stylus beam. It has been successfully used for many precision force drive applications (Smith & Chetwynd, 1990; Bowen *et al.* 1990; Whitehouse *et al.* 1988).

4.1.1 Principle

A simple solenoid-magnetic configuration is shown in Fig. 4.1. A uniformly wound circular cylindrical coil is fixed to a base and the magnet is positioned inside the coil with the axis of its magnetic field co-linear with the coil axis. Upon the excitation of the coil, a force between magnet and the solenoid is produced. The field strength along the axis of the coil is given by (Montgomery 1969).

$$H\left(\frac{x}{a_1}\right) = H(0)\left[\frac{F\left(\alpha,\beta+\frac{x}{a_1}\right)+F\left(\alpha,\beta-\frac{x}{a_1}\right)}{2F(\alpha,\beta)}\right] \tag{4-1}$$

where x is the distance from the centroid of the coil, a_1 is the inside radius of the coil (m), a_2 is the outside radius of the coil (m), $\alpha = a_2 / a_1$, $\beta = L / (2a_1)$. $F(\alpha, \beta)$ is the well known Fabry factor given by

$$F(\alpha,\beta) = \beta \ln\left(\frac{\alpha+(\alpha^2+\beta^2)^{1/2}}{1+(1+\beta^2)^{1/2}}\right) \tag{4-2}$$

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

$$H(0) = \frac{NI\chi F(\alpha,\beta)}{2a_1\beta(\alpha-1)} \tag{4-3}$$

where L is the length of the coil (m), N is the number of turns on coil, χ is the packing factor, and I is the current through the coil (A). The detailed analysis can be found from, for example, Montgomery (1969). For circular wires the packing factor is typically in the region of 0.7.

It is assumed that radial field inside the coil is uniform along the length, and that, with the short magnet considered here, only a small torque would be produced by slight misalignment and the mechanism being driven is much stiffer in torsion than translation. Therefore, from electromagnetic theory, the x component of force F experienced by a magnet of moment M within a field of strength, H, is given by

$$F_x = M\frac{dH_x}{dx} \tag{4-4}$$

For a permanent magnet in moderate field strengths, the intrinsic magnetisation is approximately equal to its remanence B_{rem}. Thus for a magnet of volume V, the magnet moment is given by

$$M = B_{rem} V \qquad (4\text{-}5)$$

Substituting equations (4-1), (4-2) and (4-3) into (4-4), the force on the magnet positioned at the point on the axis a distance x from the centroid of the coil can be evaluated. This force will be proportional to the coil current. For a given current, the force varies with its position along this axis, being zero at the centre and also decaying to zero in the far field. By definition, at the point along the axis at which the force is maximum its gradient with respect to axial displacement will be zero. This shows an important feature of this type of force actuator that the position stability of the coil has only a second-order influence on the force and may be ignored for relatively small displacements of magnet or coil. The optimum geometry for the coil is $\alpha = 1.75$, and $\beta = 0.75$ (Smith and Chetwynd, 1990). However, in many instrument designs a large surface area is desirable to aid heat dissipation and so $\alpha = 2.0$ and $\beta = 1.0$ is considered to constitute a good design. Fig. 4.2 shows the map of the axial field gradient against the position along the axis of coil corresponding to coil excitation currents from 10 mA to 400 mA. The parameters are $\alpha = 2$, $\beta = 1$, $a_1 = 1.0$, $\chi = 0.77$ and N = 1000. It can be seen that the maximum axial field gradient occurs at a point just beyond the end of the winding, $x / a_1 = 1.1$.

4.1.2 The construction

In principle, it is quite possible to design and construct a sensor head that incorporates a stylus, a force transducer and a displacement gauge. For initial experimentation a standard instrument pick-up has been modified. To do this, a magnet is attached onto the stylus pick-up while the coil is fixed to the frame of the instrument. There are two possible positions for the attachment of the magnet: above the stylus tip or beneath the LVDI. The former is chosen for its convenience and space, and importantly it is placed well away from the inductive transducer.

A schematic diagram of the experimental arrangement of the variable stylus force transducer is shown in figure 4.3. A long-reach stylus arm has been used to provide more working space. Clamped to the stylus shank is an aluminium connector that supports the magnet just above the height of the stylus arm. The coil assembly is supported over the magnet on a bracket clamped by set

screws to the body of the pick-up. This is spring mounted to the bracket and held against an adjustment screw so that the optimum magnet-coil position can be obtained. Because of the force is independent of small displacements, slight vibrations of the coil are not coupled into the measurement loop and so such a simple construction is permissible. A saturated permanent magnet of Neodymium-boron-iron (NdBFe) is used. It is 3 mm diameter and 2 mm long. The coil assembly follows the typical rule that the outer diameter of the coil is about equal to its length and twice the inner diameter. The coil has internal and external diameters nominally of 4 mm and 8 mm, respectively, and a length of 8 mm. The winding uses 38 gauge copper wire and has a total resistance of 5.3 Ω.

With this arrangement, all features of the commercial equipment remain available while the force system is readily available for manually setting and fine adjustment. A big drawback of this construction is the use of aluminium mounting piece for clamping the magnet to the stylus arm. It has a mass of about one gram, which obviously increases the second moment of mass of the arm. Consequently the sensor resonant frequency will be reduced, sacrificing some high frequency performance. The aluminium piece could be removed by gluing the magnet directly onto the stylus shank and this will much reduce the added mass. The present design is useful for retrieval of the instrument, leaving no permanent modification to it. The weight of the added mass must be backed off using the force transducer, as otherwise it would have to be supported directly by the stylus. This means that the force transducer must carry always a relative large bias current so that the net downward force is in the desired range. Moreover, the magnet is well shielded from above by the coil, but there are problems of both space and weight in shielding its underside. Care must be taken to prevent ferromagnetic workpieces coming into too close proximity with the transducer in case its calibration is shifted. Tests show that steel workpieces considerably larger than the specimens used in this work do not introduce significant error when at, or below, the level of the stylus.

4.1.3 Design of constant current drive

When the magnet is placed in its optimal position to the coil, the force exerted on the magnet is proportional to the current input to the coil. For a stable force it is necessary that the current is controlled. The coil can be driven by two sources, a voltage controlled source or a current controlled

source. The former maintains the voltage rather than the current constant, current will vary with coil resistance which is subject to temperature change due to heat dissipation. The current controlled source can keep current constant over a fairly large variation of the coil resistance.

The current stabilisation is typically obtained by comparison of a reference voltage with the voltage drop produced by the current through a sensing resistor which is in series with the coil. Through a careful selection of circuit components and or controlling of the circuit environment, current drifting can be reduced to 1 μA/h or better (Retsky & Wall 1972, Cafferty & Thompson 1989). To get further low drift a temperature controller is commonly employed in which a sensor in a feedback loop is used to regulate the power given to a heating/cooling element (for example, the solid-state heat pump). With such temperature control, milliKelvin temperature stability can be obtained (Van baak, 1992).

The current drive regulates the coil current by comparing a set voltage with a feedback voltage proportional to the coil current. Consideration must be given to possible drifts of these voltages since the current stability results directly from them. For the present work, a task of 5 μA controllability was set, and a relatively simple and inexpensive circuit was designed to meet it. The circuit diagram is shown in Fig. 4.4. Drift in the set and feedback voltages results from changes in the ambient temperature of the relevant circuit. The critical components include the voltage reference, the sensing resistors and amplifiers. Each of these components should have a low temperature coefficient. A precision band-gap voltage reference housed in metal package, ZN423N, is used for its low temperature drift and low noise. The temp. coeff. is 30 ppm/°C, and the drift of reference voltage is typically 10 ppm/1000 hrs. ZN423N is an industrial type of voltage reference. To go for further low temperature coefficient the precision reference, type LM299A, is a good candidate with its extremely low drift of 0.5 ppm/°C. Its use here is considered unnecessary, given the additional cost. Except for the sensing resistor R_s, precision metal film resistors are used in the circuit with temp. coeff. of 15 ppm/°C. The sensing resistor R_s is a wirewound precision resistor with a temperature coefficient of 3 ppm/°C. The amplifiers U1 and U3 have to be of low offset and low drift. An AD524AD precision instrumentation amplifier is chosen as the input amplifier and the output amplifier for current reading uses an OPA121KP for its low offset current and low noise. The amplifier U2 is in a feedback loop comparing the demand voltage with the current sensing voltage

from R_s. This amplifier itself working in an open loop mode together with a pair of Darlington power transistors gives this voltage to current converter a fast response and a minimal following error.

It is worth noting that for an inductive load, the loop response, *i. e.* the time of current setting up at a voltage demand, is determined by the time constant of the loop. This is an intrinsic feature of an inductive load. For a steady current drive the effect can be ignored but for a dynamic voltage - current follower the effect has to be considered.

The current drive to an inductive load can be simplified as shown in Fig. 4.5. The inductive load, the coil, is normally consists of a pure inductor L and a resistor r. The current in the coil is determined by the demanding signal, V_i, and the following error is determined by the loop gain, K. Then we have

$$K(V_i - iR_s) = L\frac{di}{dt} + i(r + R_s) \qquad (4\text{-}6)$$

The transfer function can be written as

$$\frac{i}{V_i} = \frac{\frac{K}{L}}{s + \frac{r + (1+K)R_s}{L}} \qquad (4\text{-}7)$$

and the time constant of the loop is

$$\tau = \frac{L}{r + (1+K)R_s} \qquad (4\text{-}8)$$

It can be seen that the time constant is determined by the load, L and r, the loop gain and the sensing resistance. For a given load, the r and L are fixed. To reduce the time constant either the loop gain or the sensing resistance or both of them has to be increased. The scope for increasing R_s is limited by the power supply, thus a high gain loop has to be pursued. The high gain in this design is obtained by the open loop gain of the operational amplifier LM741 ($\sim 10^5$) and the gain of the Darlington power transistors ($\sim 10^2$), which gives a total gain of 10^7. Considering the inductance of the coil is about a few millihenry, the loop time constant is a sub-millisecond which is more than enough for present static current/force setting and has potential for dynamic current/force regulation on instruments that operate with bandwidths of a few tens of hertz.

With this circuit, the drive can deliver a current up to ± 250 mA to the coil. The current drift is less than 10 μA/hr, and the noise is less than 5 μA.

4.2 Modification of the Talysurf 5

Due to the attachment of the magnet-coil force actuator the performance of the Talysurf 5 is expected to vary from its original configuration. Thus the behaviour of both the force actuator and the stylus instrument have to be calibrated before carrying out any experiments.

4.2.1 Calibration of the variable force transducer

The force of this magnet-coil actuator, now referred as the stylus force, was calibrated against a miniature strain-bridge load cell (Maywood Instruments, type 4934). It has a nominal range of 50 mN with an output range of 0 ~ 100 mV. It has a resolution of 5 μN and the non-linearity and hysteresis is less than 0.1% of full range. The load cell was calibrated by standard weights. For the purposes of these studies an absolute calibration to within, say, 5% would be perfectly acceptable in the final results. However, close calibration must be undertaken in order to establish the linearity and stability of the force transducer which is in an open loop control during actual measurements. To improve the performance of the load cell, the line power supply was replaced by a battery supply and an instrumentation amplifier with a gain of 10 connected to the output. The improved load cell circuitry is shown in Fig. 4.6. Having done this the noise level of the load cell reduced from 50 mV to less than 1 mV. The calibration curve of the load cell is plotted in Fig. 4.7, the sensitivity is 19.4 mV mN^{-1} or 51.6 mN V^{-1}. The solid line is a least squares fit line, and the difference between the measured and the calculated values indicates quite good linearity with a standard deviation of residuals, $\sigma = 0.3\%$. This load cell was then used for calibration of the variable force transducer, During calibration, the stylus was lowered onto the button of the load cell and adjusted in height until the pick-up signal was in mid-range. The screw adjustment on the coil height was then used to maximise the recorded force at a constant coil current. Having thus set the optimum magnet position, a calibration of force against current could be carried out.

A set of control signal was given by a computer through a D/A channel to the current drive to exert a force on the stylus, then the output readings from the load cell and the stylus were recorded.

The results are shown in Fig. 4.8. The measured data were fitted by a least squares straight line. The relationship between the force and current could be expressed by

$$F(mN) = -0.1365I(mA) + 10.538 \qquad (4\text{-}9)$$

with a maximum deviation of residuals from the least squares fit of less than 0.6%. A 'lifting' current of 74 mA is needed to get zero force on the stylus, with the maximum force from the self weight of the stylus pick-up plus the mounting piece is about 10.5 mN.

Maintaining constant current in the coil ensures a constant force at the stylus to the limits of the measurement (principally caused by thermal effects of the load cell). There is high confidence that the coil current may be used as a measurement of the static force imposed upon the tip by the system. It does not, however, monitor the variation of force at the stylus caused by its deflection against the bias spring of the conventional pick-up, and this will be given in the following section.

4.2.2 Measurement of the spring rate

With the facility of the variable force transducer the spring rate of the stylus pick-up can be measured easily. This was done by applying a certain current to deflect the stylus within a given magnification. First the stylus was suspended making no contact with any surfaces. Then a ramp voltage, either from a channel of D/A converter or set manually by adjusting the potentiometer at the front panel of the current driver, was given to the current driver to deflect the stylus against its spring suspension. The current variation and the corresponding output readings of the stylus at each magnification were monitored. For most useful magnifications above X5000 the spring rate varied from 8 to 12 $\mu N\ \mu m^{-1}$ which is much better than specified in the National Standard. Thus for a deflection within 5 μm, the maximal variation of the spring force is 60 μN.

4.2.3 Change of the effective mass and resonance

The mass added by the force transducer changes the effective mass and thus the resonance of the stylus pick-up. The new resonance of the stylus was tested by lifting it to its neutral position, deflecting it a little further and then releasing it to give a step change. The stylus then oscillated around its neutral position and this response was recorded. The resonant angular frequency was now measured as 96 rad s^{-1} with a slightly changed damping ratio of 0.02. The effective mass is 1.2 g. As

mentioned earlier, the large mass is mainly due to the aluminium mount because the magnet itself has a mass of only 0.1 g. With the attachment of the force transducer, the current driver has been always left on at an appropriate force setting that compensates for most of this weight in case a large force acting on the stylus might cause deterioration of the pick-up. Therefore, the change on the damping was thought to be due to the force transducer in which a brass tube was employed as the coil frame. The movement of the stylus, and thus the magnet, would generate an eddy current in the brass tube in which a reverse force would be produced in return. It can be seen from the measurement that this effect is of little significance although it doubles the previous value.

4.2.4 Calibration of the pick up

A standard plate with specified steps was used to calibrate the stylus instrument. Typically a stylus instrument gives a design maximum output of ± 1V at each magnification generally accurate to 10% or better. However, a slight variation on the scale factor would be expected due to the non-linear characteristic of the stylus pick-up mechanism, *i.e.* the arc movement of the stylus tip. Therefore, a close calibration of the stylus instrument at the magnifications of interested is needed. This is often done by measuring the recorded displacement produced by traversing a step etched into a glass flat whose height has been calibrated by interferometric measurement. The results are shown in table 4.1 in which a step standard of 2.46 μm depth was used for the magnifications from 1000 to 10000, and a step depth of 0.36 μm was used for the magnification of 20000.

It should be noted that absolute measurement is not necessary in this investigation, the calibration operated here is mainly for the consistency of measurements, therefore no further error analysis is conducted.

4.3 Experimentation

4.3.1 Sample trigger 'flag'

In this research a comparison of surface profiles over a range of different tracking forces has been carried out. Profile comparison requires that the tested surface be relocated closely to the same place from one measurement run to the next. Discrete, recognisable surface features, either natural

or artificial, could be used for relocation, but identifying them are quite time consuming. Here a simple position trigger was constructed. It consisted of a brass shim with a well-defined edge and an optoswitch. The brass shim was fixed to the stylus pick-up and moved together with the horizontal traverse of the pickup. The optoswitch was attached to the instrument frame to provide a datum position on the sample being tested. Once the brass shim intercepted the light beam of the optoswitch, a trigger signal was produced. Repeat tracks could be taken by running the instrument in auto-return mode over a traverse rather longer than that being measured so that incremental creep in the gearbox did not affect the trace. The flag provided a synchronisation signal to initiate profile data-logging at a fixed point, after the traverse had established steady motion.

The track repeatability is governed laterally by the ligament stiffness of the stylus arm and in the traverse direction by the repeatability of the optoswitch. Tests to monitor the trigger repeatability indicated variation of less than 2 μm and expected within ± 1 μm of its average position which is of the order of the stylus dimension.

4.3.2 Specimen preparation and measurement

Mild steel, copper, 30/70 brass and aluminium materials were chosen for their different properties. They were cut from standard supplies into blocks of dimension 10 x 10 x 1 mm, then ground, polished on conventional metallographic equipment and then coarsely abraded with a 1200 grit size abrasive to give a distinctive lay and enough roughness. Specimens after such a process have surface roughness suitable for two purposes; they are small enough that damage would be readily detected and also visible in electron microscope examination.

Two basic forms of test were undertaken to examine the variation of profile with tracking force. In one, as nearly as possible, a single track was repeatedly scanned, each time at an increased stylus force. In the other, single tracks at different stylus force were taken parallel to each other but spaced at about 10 stylus widths apart, *i e.* about 20 μm spacing between them. The first method allows ready comparison of variations in profile, but the behaviour at higher stylus forces might be influenced by the earlier traces through, for example, a work hardening process. The latter method avoids this by using a pristine surface for each trace, but provides traces which would not be expected to be identical, so comparison is more difficult.

Generally, 0.8 mm active tracks were used, traversed at speeds of 1.0, 0.25 and 0.1 mm s^{-1} corresponding to instrument horizontal magnifications of X5, X20 and X50. Three sampling speeds were used, they were 2000, 500 and 200 points per second for the traverse speed of 1.0, 0.25 and 0.1 mm s^{-1}, respectively, which gave a nominal sampling interval in space 0.5 μm. Vertical magnifications were typically X10000, giving a discrimination of about 1.5 nm in the digitisation for the 12-bit A/D converter used.

Trace to trace variation was assessed in two ways. Simple and effective is to level the traces by subtracting out the linear least squares straight line and to overplot them. Visual inspection, which can readily adjust for small phase errors caused by trigger variation, then detects differences that appear to be significant. Alternatively, parameterizations of the levelled profiles may be compared. Three parameters are used here. The R_q value should be stable and hardly affected by distortion of small features and so is used as control. Distortion of profile flanks is sought by examining the rms slope, Δ_q. To follow up the observation that profile peaks are most affected, rms peak curvature, Pc, is used. The profile slope is calculated by a symmetrical (Lagrangian) formula using five points, to give a reasonable compromise between the sensitivity and noise rejection, Chetwynd 1978. The peaks are detected by the simple rule that a point is peak if it is larger than the points on either side, and the curvature found by three point second derivative formula centred on the peak point. This algorithm was used because it is known to be very sensitive to sampling and stylus conditions, Chetwynd 1979, and so might emphasise relatively small profile distortions.

4.4 Results and Discussions

4.4.1 Profile variations at a nominal force

The level of significance in detecting changes in the profile and its parameter values was determined by repeating measurements under a standard loading of 0.7 mN on a steel sample. With a traverse speed of 0.1 mm s^{-1}, six traces were measured over the same track of the specimen. Three of these are shown in Fig. 4.9. It is clearly shown that the repeatability is very good with a maximum variation of less than 7% in the height of peaks. From trace to trace, features appear to drift horizontally by up to 2 - 3 μm inconsistently. The direction of this drift can vary over a period

of one traverse. Because of the accuracy of the hardware time in the computer, this implies that there is variability in the drive velocity and this will be further discussed in chapter 6.

The parameters of R_a, R_q, Δq and Pc were calculated from these traces and those at different traversing speeds. The results are shown in the table 4.2. At the traverse speed of 0.1 mm s^{-1} a peak variation of 1.4%, 2.7%, and 2.5% was observed for parameters R_q, Δq and Pc respectively. As the stylus traverse speed is increased, it can be seen from table 4.2 that, at the standard load, the surface finish parameters remain relatively constant. However, at the fastest speed of 1 mm s^{-1} there is a significant increase in the parameters R_a and R_q with Δq becoming more variable and the peak curvature reducing. Although it was not confirmed at that time, it was suspected that this might be caused by the stylus bouncing over the surface. This dynamic effect may be enhanced by the added inertia of the magnet and clamp in this experiment.

4.4.2 Profile variation at different tracking forces

Figure 4.10 shows a set of eleven traces on the same track of a copper specimen at loads of 0.39, 0.51, 0.68, 0.8, 1.01, 1.18, 1.37, 1.69, 1.96, 2.28 and 2.79 mN with a traverse speed of 0.1 mm s^{-1}. The slowest speed is chosen here to avoid stylus dynamic effect in order to simplify the process of comparison. Curves are arranged in sequence from top to bottom and from (a) to (d) as the load increases, *i. e.* the top profile in (a) is at the lightest loading of 0.39 mN and the bottom one in (d) is at the heaviest loading of 2.79 mN. It can be seen that the basic form of the profile remains unchanged, but details were disappearing as the load increased. At the lighter loads few features seem significantly different, but as this increases definite smoothing occurs with some of the sharp peaks disappearing completely. A big peak shown at low loads was removed at the load 1.18 mN, the bottom curve in figure (b). Later SEM examination showed that a defined trace mark appeared as the load increased to 1.3 mN on the same specimen. Therefore, at the load 1.18 mN occasionally damage on some weak parts had occurred. It is also noted that at very heavy loads, the downward slope and valleys of these traces become progressively steeper and wider.

The parameters for these traces were calculated and listed in track 3 of the table 4.3. The parameter evaluation confirms the pattern observed in the profiles. The amplitude parameters R_a and R_q remain fairly stable with variation of 3 ~ 5% from that at the nominal load. However, the

rms slope and curvature reduced by up to 25% and 43%, respectively. Values are quite consistent in the region of the standard stylus force of 0.7 - 1 mN. A similar trend was observed on other materials. Results are listed in tables 4.4, 4.5, and 4.6 for specimens of mild steel, brass and aluminium, respectively. At the traverse speed of 0.1 *mm* s^{-1}, the variations (or reductions) on parameters of Δq and Pc were observed to be about 16% and 22% for mild steel, 23% and 31% for bronze, 41% and 44% for aluminium, indicating the differences of material strength and stiffness. Over the same range of load forces the soft material, the aluminium in our case, showed more damage than the others.

To give a clearer view, the parameter values for different materials are plotted as a function of the loading force in Fig. 4.11(a) to (d). Each parameter has been normalised to its value at about 0.7 mN load. Each set of three graphs represents stylus traverse speeds of 1.0, 0.25 and 0.1 mm s^{-1}, the fastest being the top graph with the slowest at the bottom. They all confirm the pattern mentioned above.

4.4.3 Influence of tracking speed

It can be seen from parameter measurements, shown in tables 4.3 to 4.6, that at a traverse speed of 1.0 mm s^{-1}, considerable profile variation is observed with light loads. Looking at the profiles traced on a copper specimen at the loads of 0.39 mN, 0.69 mN and 1.67 mN at the traverse speed of 1.0 mm s^{-1}, figure 4.12, it appears that the stylus was overshooting and lifting off at the lighter load of 0.39 mN, giving the effect of the profile following the surface features at an exaggerated amplitude. Both the amplitudes and the lateral spacings were distorted, but the number of peaks seemed to be preserved. As a result, the amplitude parameter, R_q, and the rms slope, Δq reduced by up to 114% and 36% for copper, 93% and 58% for steel, 192% and 33% for brass and 11% and 25% for aluminium respectively as the tracking force increased from 0.39 mN to 0.69 mN. These variations are much larger than those at increasing loads tested. The variations in amplitude parameters, R_a and R_q, are most significant since these values are relatively stable in other tests. This provides strong evidence that they were caused by the stylus dynamic effect. As the load increased to 1.67 mN the most features followed the profile at the force of 0.69 mN but some sharp peaks were trimmed off.

The stylus dynamic effect depends on the spatial features of a surface and the speed at which they are traversed. The combination of these two gives the spectrum of the interaction force between the stylus and the surface. When the spectrum of the interaction reaches the resonance of the stylus system, large dynamic force, thus large variation on measurement would be expected. The large variation is mainly from magnitude parameters due to the stylus lifting off behaviour, and the magnitude of this variation can be very high about 100% or higher. The reason that the aluminium specimen gave a small variation is probably due to its large surface features which gave a lower frequency spectrum than the other materials. Further discussion on stylus dynamic effect will be given in chapter 6.

4.4.4 SEM view of surface damage

The sets of parallel traces made at each different tracking force and traversing speed on each material, were examined under a scanning electron microscope.

A typical micrograph of the stylus tracks on a copper specimen is shown in Fig. 4.13. They are, as in previous observations, more difficult to see than would be expected from their visibility to the naked eye. The surface is viewed obliquely at an angle of approximately 27°. Each track has been produced by repeat parallel traces, applying a different constant force to the stylus during each.

Comparison of tracks made on steel with those of copper highlighted the major differences between materials of different hardness. Those made in copper at traverse speeds of 0.25 and 0.1 mm s^{-1} had clearly defined tracks as the load is increased above 1.3 mN. These were only visible to the experienced eye and cease to be discernible as the load reduced to less than 0.7 mN. The same effect was also observed on the steel specimens only in this case the tracks cease to be discernible as the load reduced below approximately 2.5 mN, probably reflecting the higher hardness of this material. As the values of load were further reduced, no anomalous markings or deformations were observed. Curiously, as the traverse speed was increased to 1.0 mm s^{-1}, the force at which tracks cease to be discernible on the same copper specimen rose approximately to 1.3 mN. Below this load, only occasional ploughing of extreme asperities could be seen.

Another observation was the appearance of jumping or hopping at high loads with steel, as shown in Fig. 4.14. This could be seen as a series of equaspaced scuff marks that commence

suddenly and systematically reduce in length. As the loads increase from 3.5 up to 8 mN the period of these scuff marks was markly reduced. This is thought to be caused by the large friction force between clean and ploughed surfaces. The width of the ploughed tracks at high load corresponds to that of the stylus to within approximately 10%. The micrograph of the stylus used throughout this research is shown in Fig. 4.15, it was taken before in use in the series tests.

4.4.5 Conclusions

The important conclusion from these initial tests is that the test system itself is reliable and convenient. The force transducer technique can be firmly recommended for research requiring measurements under difficult conditions.

Perhaps unsurprisingly, given the earlier evidence, working at loads of a factor of two or three from the normal does not appear to be problematical with a 2 μm stylus. Even on copper, the worse case tested since it is the softest, the profile divergence seems unlikely to matter in practical terms. The small amount of plastic damage detected may not matter either. Note, however, that copper, in particular, is likely to be machined for special applications such as precision optical surfaces and here the damage could be important. Reducing the tracking force to below 0.4 mN and traversing the stylus at a slow speed seems to be an adequate procedure. Conversely, the results of this study imply that rapid estimation of surface finish parameters may be obtained by using a fast traverse speed and an increasing the contact force. Clearly the ability to automatically vary tracking forces will enable present stylus instruments to be used for a variety of measurements over a wide range of applications.

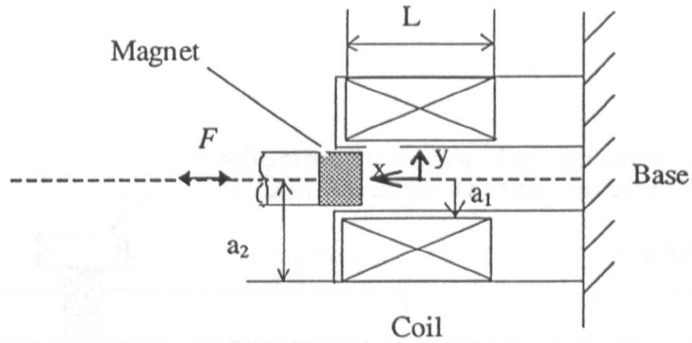


Fig. 4.1 The schematic diagram of the magnet-coil force actuator.

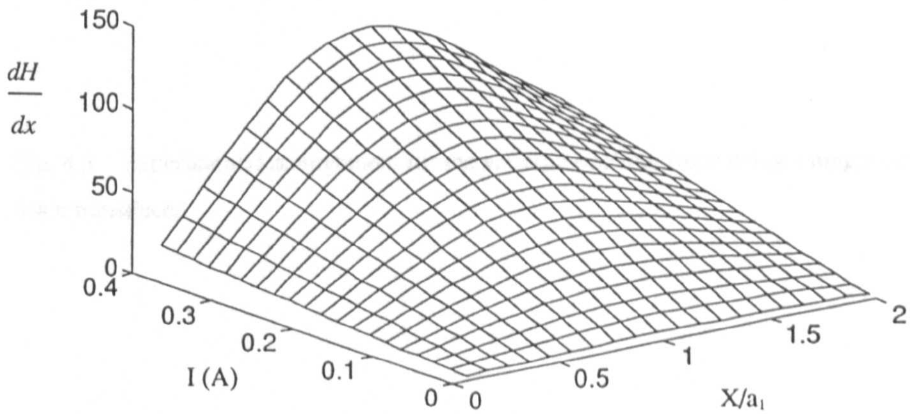


Fig. 4.2 Magnetic field gradient against position along axis of coil with different values of exciting current.

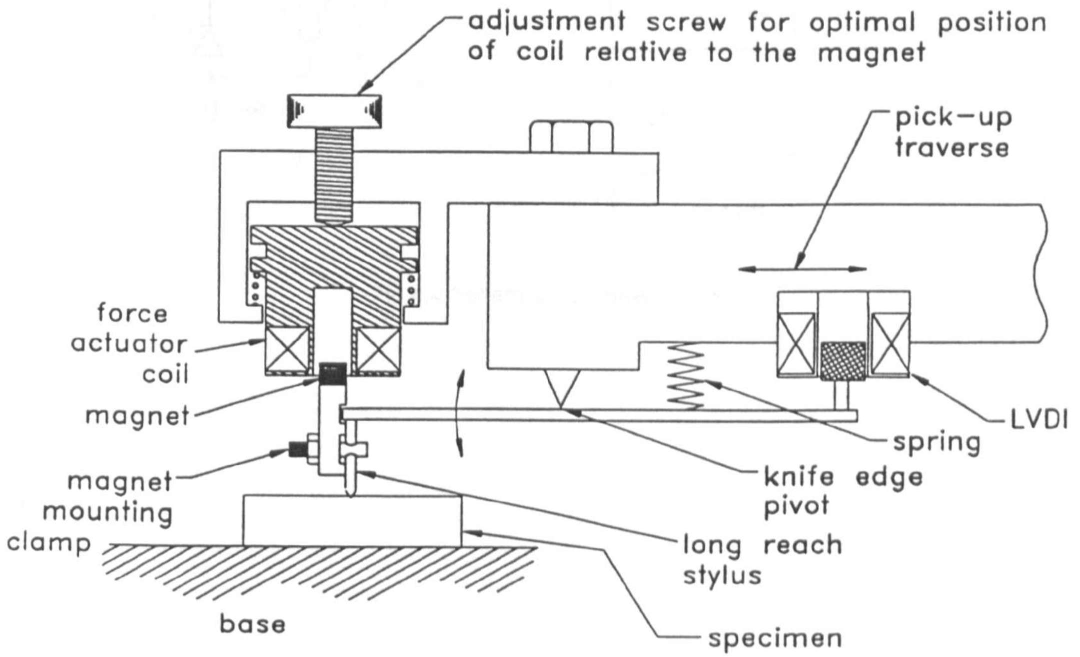


Fig. 4.3 Experimental arrangement for the variable tracking force using a magnet-coil force transducer.

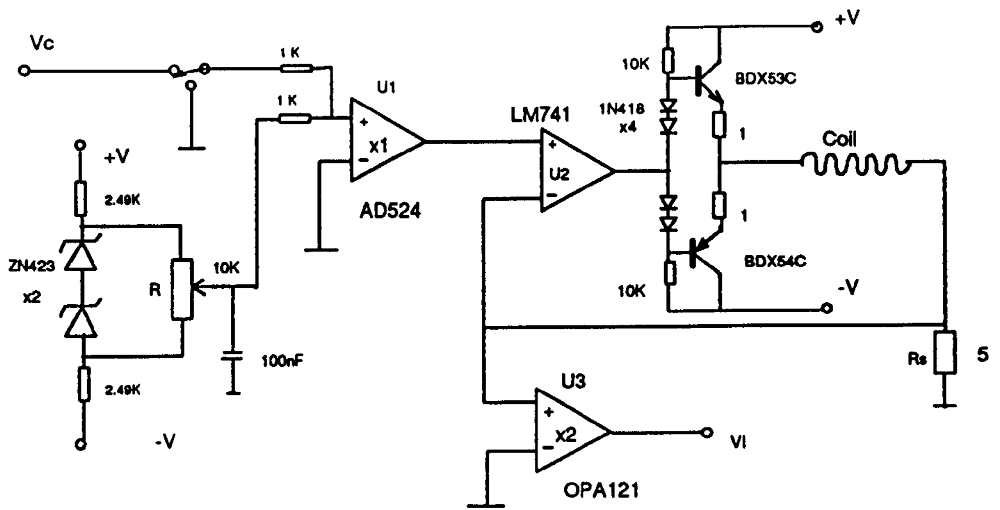


Fig.4.4 Schematic of current driver 1

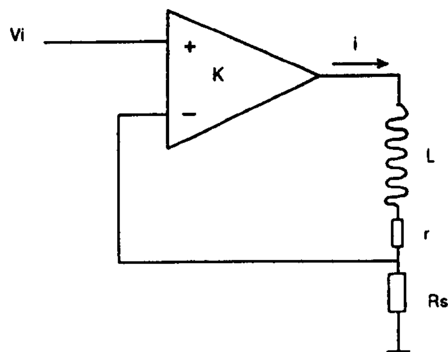


Fig. 4.5 Diagram of simplified current drive loop

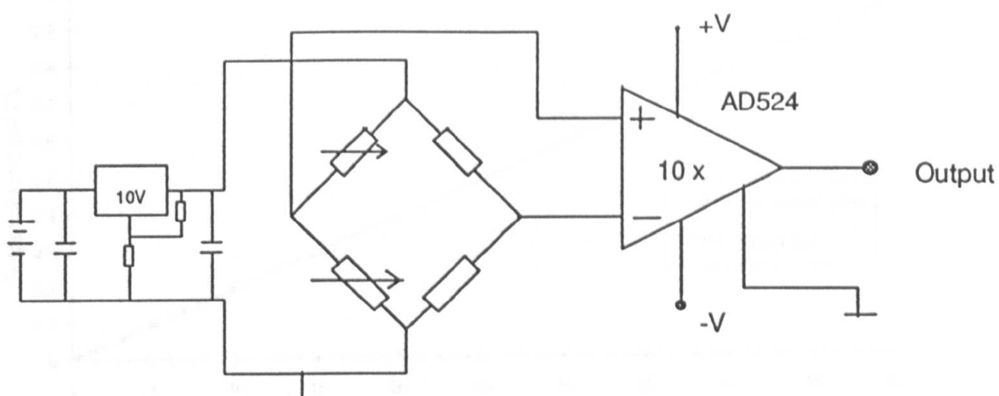


Fig. 4.6 Improved circuitry of the load cell

Table 4.1 Calibration of the vertical displacement.

Magnification	Specified scale $\mu m\ V^{-1}$	measured scale $\mu m\ V^{-1}$
x 1000	25	24.21
x 2000	12.5	12.20
x 5000	5	4.84
x 10,000	2.5	2.41
x 20,000	1.25	1.02

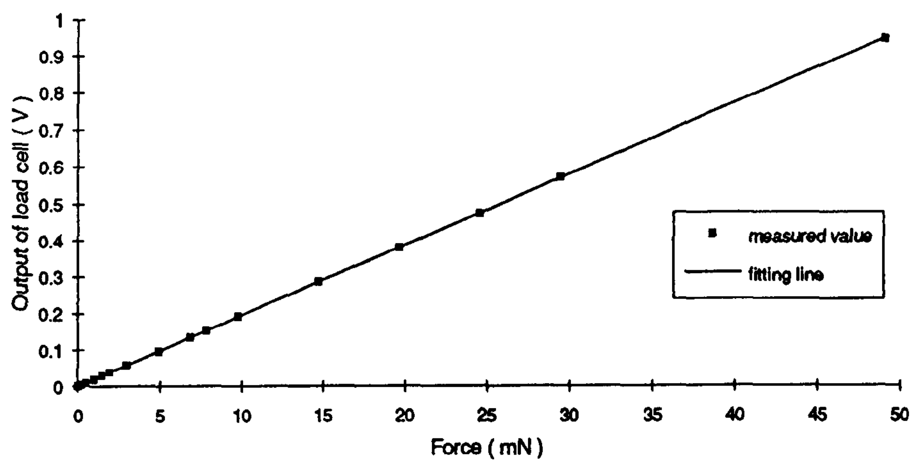


Fig. 4.7 Calibration of the load cell.

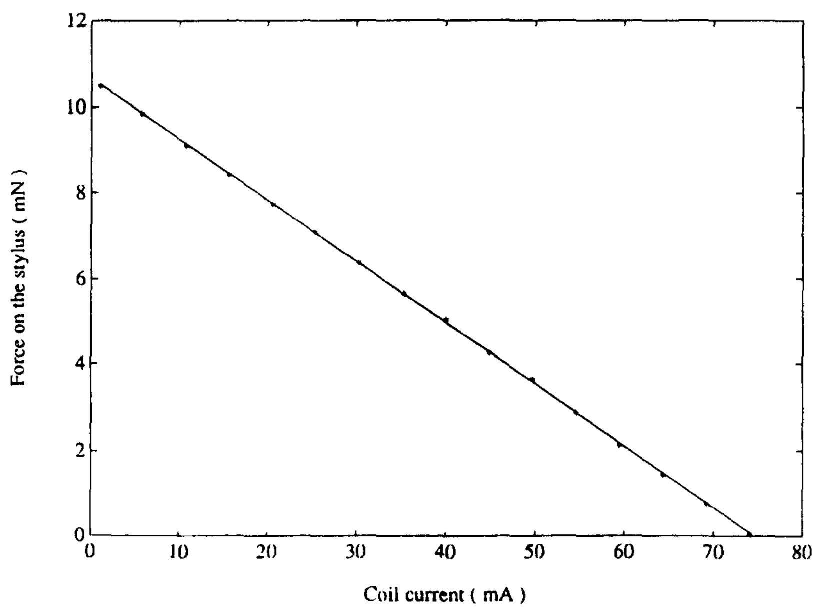


Fig. 4.8 Calibration of the variable force transducer.

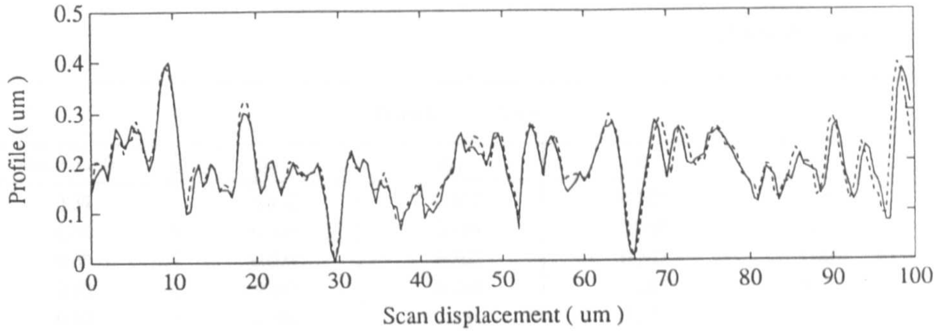


Fig. 4.9 Repeated traces on a steel specimen at 0.69 mN force at a traverse speed of 0.1 mm s^{-1}.

Table 4.2 Parameter variation from repeat traces on steel at 0.69 mN.

Material: Steel

Force (mN)	Ra	Rq	Dq	PC
Track 1 Vh = 5				
0.69	0.061	0.085	0.070	0.150
0.69	0.061	0.081	0.066	0.144
0.69	0.057	0.075	0.065	0.137
0.69	0.071	0.102	0.082	0.159
0.69	0.069	0.112	0.081	0.147
Track 1 Vh = 20				
0.69	0.053	0.069	0.076	0.215
0.69	0.052	0.067	0.074	0.216
0.69	0.053	0.068	0.074	0.221
0.69	0.052	0.067	0.074	0.207
0.69	0.052	0.067	0.073	0.212
0.69	0.052	0.067	0.073	0.213
Track 1 Vh = 50				
0.69	0.052	0.067	0.074	0.236
0.69	0.052	0.067	0.074	0.234
0.69	0.052	0.066	0.074	0.230
0.69	0.051	0.066	0.074	0.235
0.69	0.052	0.067	0.072	0.232
0.69	0.051	0.066	0.072	0.231

Table 4.3 Parameter variation from three sets of repeated tracks on a copper specimen at increased loading.

Material: Copper

Force (mN)	Ra	Rq	Dq	PC
Track 1 Vh = 5				
0.39	0.462	0.652	0.282	0.47
0.50	0.333	0.484	0.249	0.51
0.69	0.225	0.305	0.207	0.46
0.80	0.211	0.281	0.201	0.46
0.99	0.192	0.254	0.178	0.42
1.18	0.183	0.239	0.165	0.44
1.37	0.184	0.235	0.164	0.45
1.68	0.181	0.231	0.156	0.42
1.96	0.179	0.228	0.148	0.43
2.28	0.179	0.227	0.138	0.42
2.82	0.168	0.215	0.132	0.37
Track 2 Vh = 20				
0.40	0.192	0.248	0.162	0.50
0.51	0.187	0.239	0.163	0.48
0.69	0.185	0.239	0.163	0.47
0.81	0.183	0.235	0.158	0.45
1.00	0.182	0.233	0.154	0.41
1.18	0.180	0.233	0.150	0.38
1.37	0.181	0.232	0.147	0.37
1.67	0.178	0.229	0.137	0.34
1.96	0.177	0.227	0.131	0.34
2.28	0.176	0.224	0.127	0.30
2.79	0.172	0.221	0.119	0.27
Track 3 Vh = 50				
0.39	0.213	0.272	0.161	0.49
0.51	0.211	0.269	0.162	0.48
0.68	0.210	0.267	0.162	0.46
0.80	0.206	0.258	0.156	0.46
1.01	0.208	0.259	0.154	0.43
1.18	0.210	0.261	0.147	0.40
1.37	0.210	0.259	0.141	0.39
1.69	0.213	0.264	0.138	0.37
1.96	0.210	0.260	0.131	0.35
2.28	0.214	0.262	0.126	0.32
2.79	0.212	0.259	0.121	0.29

Table 4.4 Parameter variation from three sets of repeated tracks on a steel specimen at increased loading.

Material: Mild steel

Force (mN)	Ra	Rq	Dq	PC
Track 1 Vh = 5				
0.387	0.124	0.185	0.125	0.251
0.519	0.095	0.147	0.112	0.255
0.659	0.069	0.096	0.079	0.226
0.796	0.066	0.089	0.072	0.214
0.973	0.063	0.082	0.066	0.213
1.188	0.063	0.081	0.065	0.202
1.360	0.063	0.081	0.063	0.208
1.682	0.062	0.080	0.061	0.201
1.726	0.061	0.078	0.059	0.221
2.164	0.061	0.080	0.060	0.212
2.806	0.062	0.080	0.058	0.188
Track 2 Vh = 20				
0.397	0.054	0.070	0.066	0.223
0.517	0.054	0.069	0.066	0.226
0.678	0.054	0.069	0.066	0.224
0.796	0.054	0.070	0.067	0.226
0.974	0.054	0.069	0.067	0.221
1.184	0.054	0.069	0.066	0.221
1.359	0.054	0.069	0.066	0.216
1.675	0.054	0.069	0.065	0.211
1.943	0.054	0.069	0.064	0.204
2.291	0.054	0.069	0.064	0.203
2.795	0.053	0.069	0.062	0.186
Track 3 Vh = 50				
0.387	0.052	0.068	0.070	0.239
0.502	0.052	0.067	0.068	0.234
0.682	0.052	0.067	0.068	0.241
0.792	0.051	0.067	0.066	0.228
0.982	0.052	0.067	0.067	0.236
1.184	0.051	0.066	0.065	0.226
1.360	0.051	0.066	0.065	0.228
1.682	0.051	0.065	0.064	0.207
1.950	0.051	0.066	0.062	0.208
2.285	0.052	0.066	0.061	0.204
2.795	0.052	0.067	0.059	0.187

Table 4.5 Parameter variation from three sets of repeated tracks on a brass specimen at increased loading.

Material: Bronze

Track 1 Vh = 5				
Force (mN)	Ra	Rq	Dq	PC
0.393	0.546	0.847	0.279	0.433
0.693	0.196	0.290	0.209	0.432
0.897	0.166	0.237	0.181	0.439
1.192	0.153	0.212	0.166	0.426
1.690	0.145	0.199	0.151	0.428
2.284	0.145	0.194	0.141	0.403
3.132	0.139	0.187	0.133	0.404
Track 2 Vh = 20				
0.409	0.180	0.236	0.171	0.495
0.699	0.177	0.233	0.172	0.467
0.897	0.169	0.223	0.166	0.447
1.190	0.162	0.211	0.153	0.406
1.706	0.157	0.204	0.143	0.406
2.318	0.140	0.177	0.118	0.329
Track 3 Vh = 50				
0.392	0.182	0.238	0.165	0.432
0.697	0.170	0.219	0.160	0.475
0.874	0.164	0.208	0.151	0.438
1.199	0.156	0.198	0.145	0.412
1.696	0.152	0.193	0.132	0.367
2.309	0.148	0.191	0.126	0.360
3.110	0.146	0.188	0.118	0.327

Table 4.6 Parameter variation from three sets of repeated tracks on an aluminium specimen at increased loading.

Material: Aluminium

Track 1 Vh = 5				
Force (mN)	Ra	Rq	Dq	PC
0.402	0.571	0.706	0.321	0.554
0.694	0.429	0.563	0.290	0.527
1.058	0.382	0.509	0.247	0.554
1.345	0.415	0.545	0.250	0.494
1.701	0.420	0.551	0.248	0.490
2.295	0.414	0.543	0.224	0.453
3.155	0.395	0.514	0.201	0.463
Track 2 Vh = 20				
0.399	0.422	0.511	0.227	0.567
0.699	0.427	0.520	0.233	0.660
1.070	0.438	0.532	0.224	0.562
1.350	0.438	0.536	0.214	0.527
1.706	0.438	0.540	0.206	0.486
2.312	0.436	0.543	0.196	0.503
3.089	0.438	0.547	0.182	0.445
Track 3 Vh = 50				
0.409	0.413	0.527	0.258	0.694
0.698	0.401	0.514	0.249	0.606
1.082	0.411	0.523	0.243	0.598
1.370	0.429	0.537	0.195	0.509
1.684	0.416	0.518	0.182	0.478
2.279	0.415	0.523	0.172	0.448
3.118	0.409	0.518	0.156	0.428

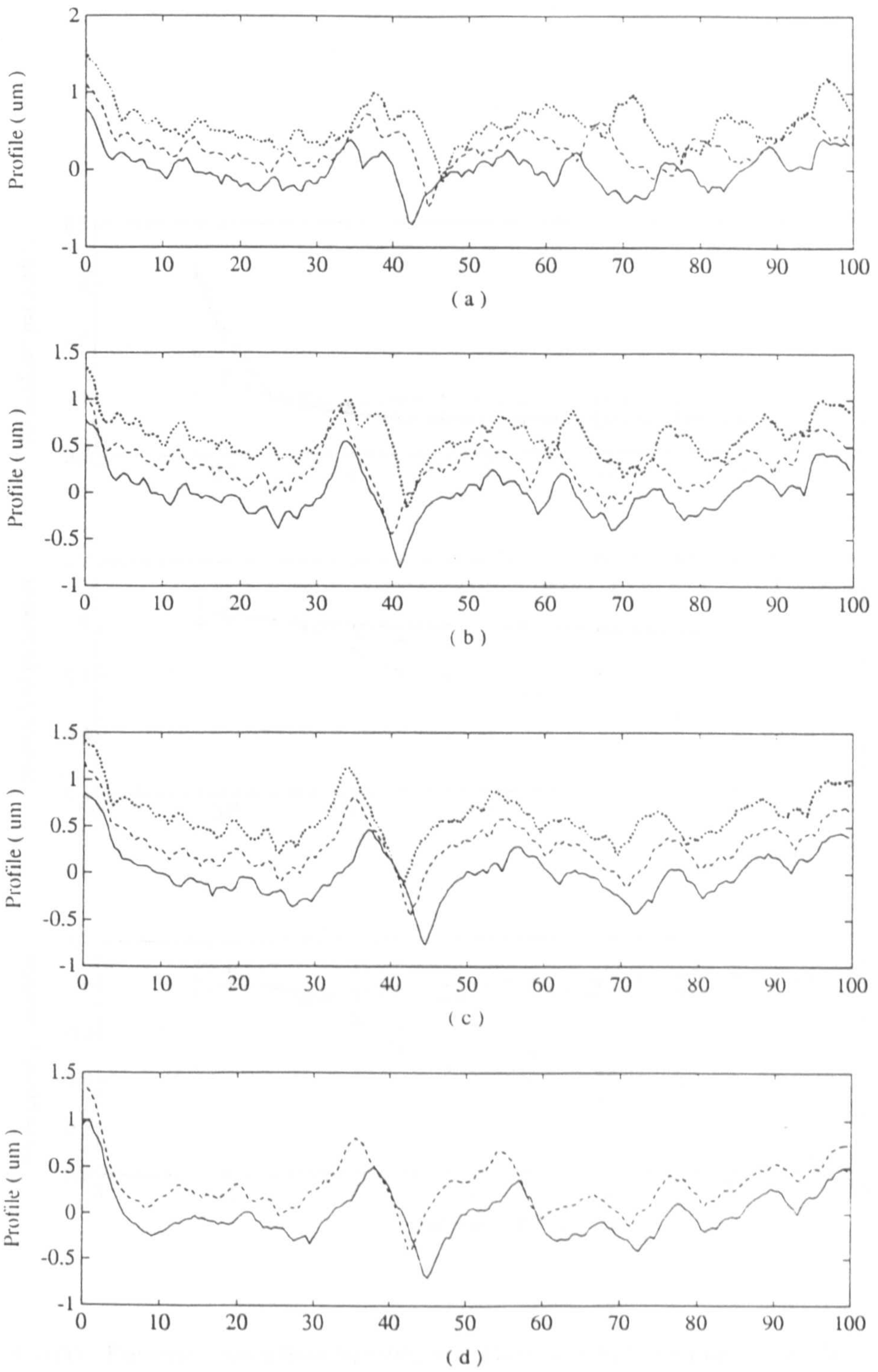


Fig. 4.10 Repeat stylus traces on a copper specimen at loads ranging verticall downwards from 0.39 to 2.8 mN

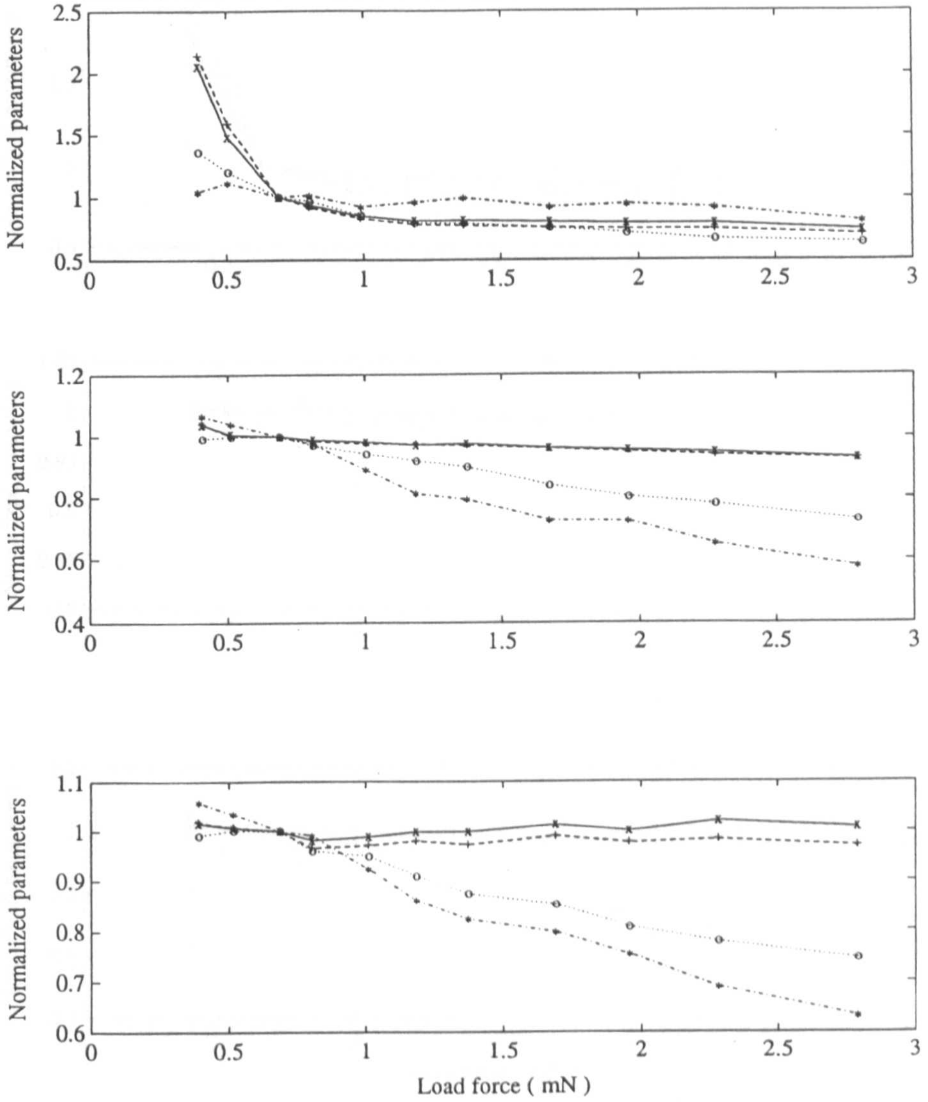


Fig. 4.11(a) Parameter values from repeating tracks at increasing stylus force, normalised to their values at 0.7 mN on a copper specimen. Stylus traverse speed is 1 mm s^{-1} in the top graph and 0.25 and 0.1 mm s^{-1} respectively in the two below. x — R_a, + — R_q, o — Δq, * — Pc.

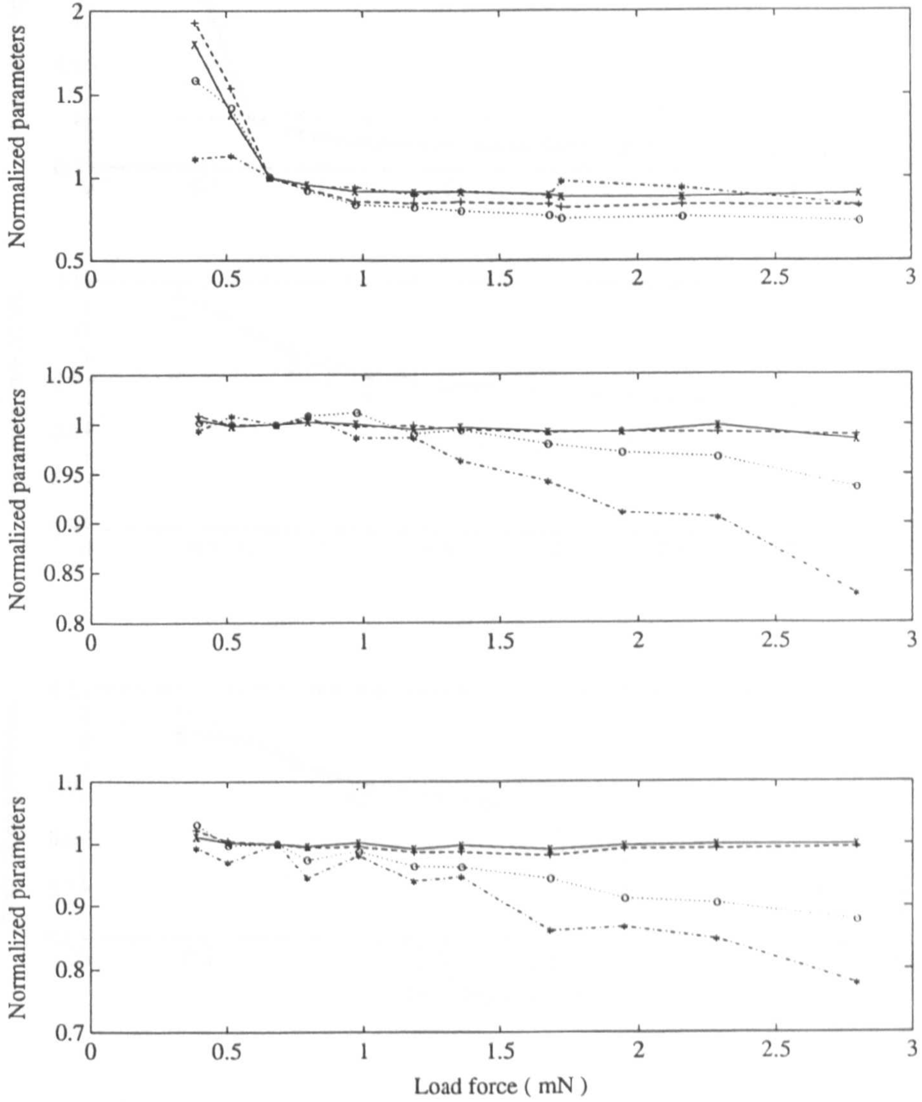


Fig. 4.11(b) Parameter values from repeating tracks at increasing stylus force, normalised to their values at 0.7 mN on a steel specimen. Stylus traverse speed is 1 mm s^{-1} in the top graph and 0.25 and 0.1 mm s^{-1} respectively in the two below. x — R_a, + — R_q, o — Δq, * — Pc.

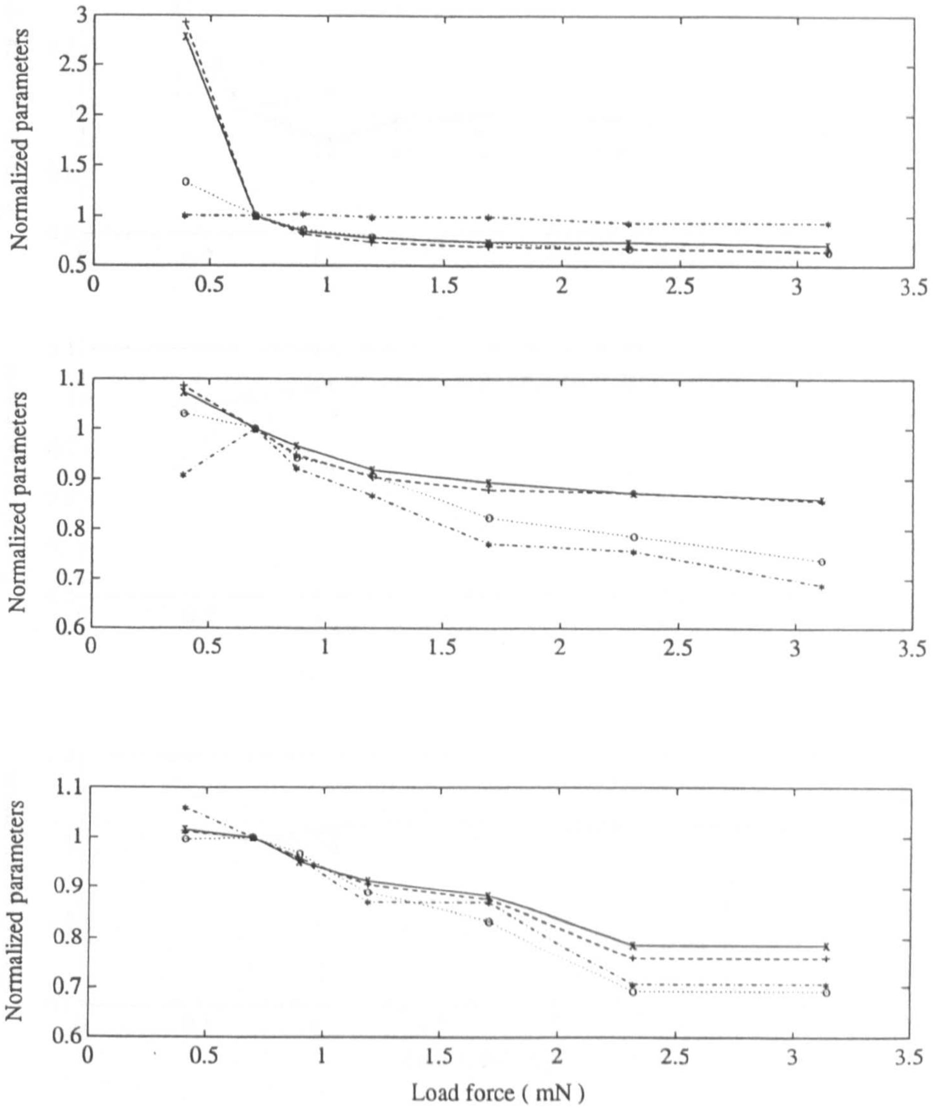


Fig. 4.11(c) Parameter values from repeating tracks at increasing stylus force, normalised to their values at 0.7 mN on a brass specimen. Stylus traverse speed is 1 mm s^{-1} in the top graph and 0.25 and 0.1 mm s^{-1} respectively in the two below. x — R_a, + — R_q, o — Δq, * — Pc.

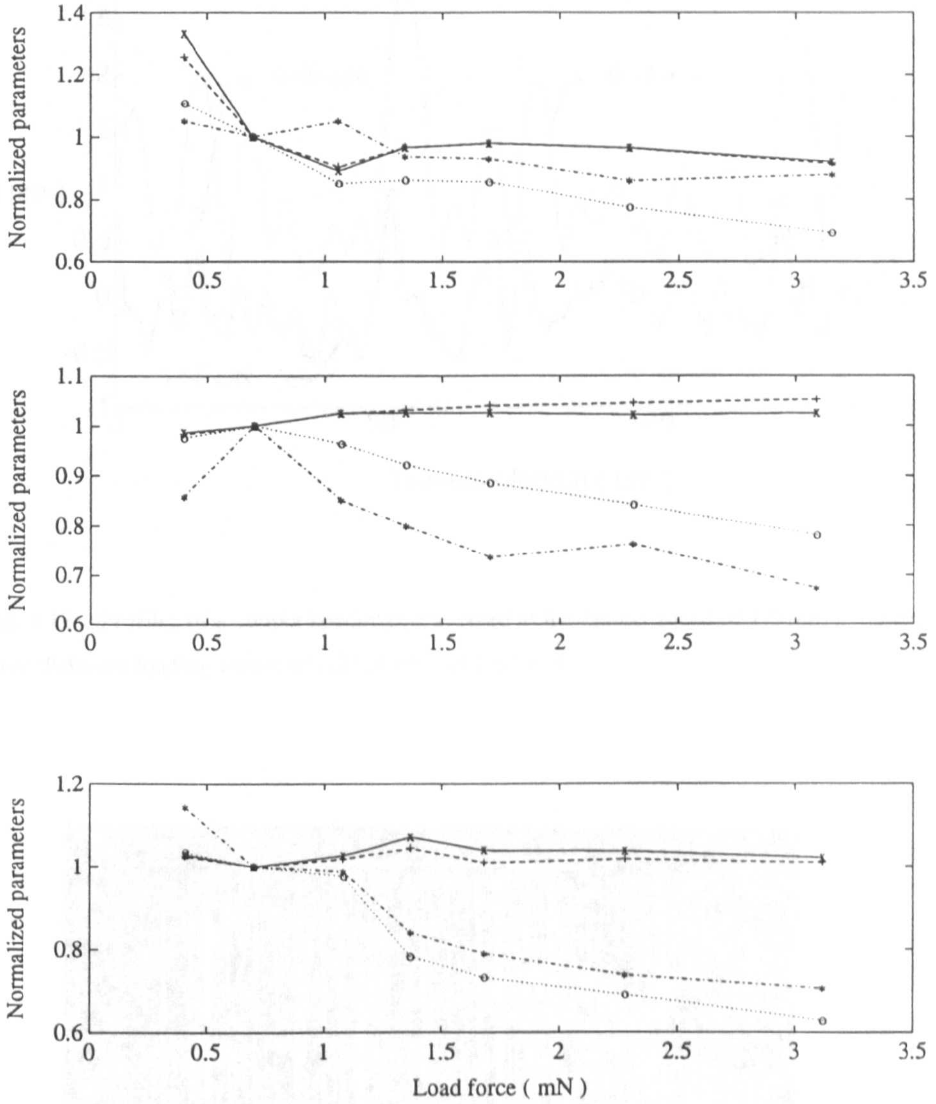


Fig. 4.11(d) Parameter values from repeating tracks at increasing stylus force, normalised to their values at 0.7 mN on an aluminium specimen. Stylus traverse speed is 1 mm s^{-1} in the top graph and 0.25 and 0.1 mm s^{-1} respectively in the two below.x — R_a, + — R_q, o — Δq, * — Pc.

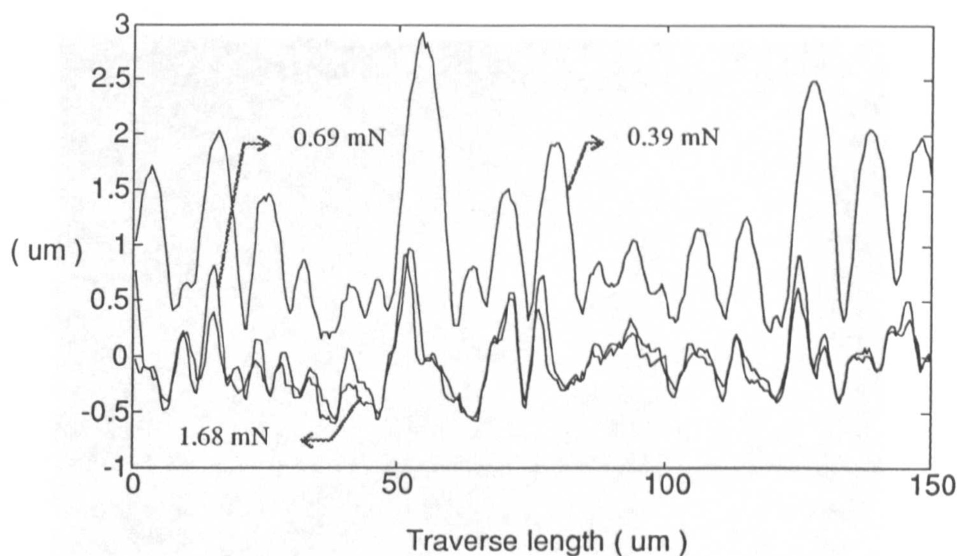


Fig. 4.12 Profiles of a copper specimen traversed at the fastest speed of 1.0 mm s^{-1} under three different loading forces of 0.39, 0.69 and 1.67 mN.

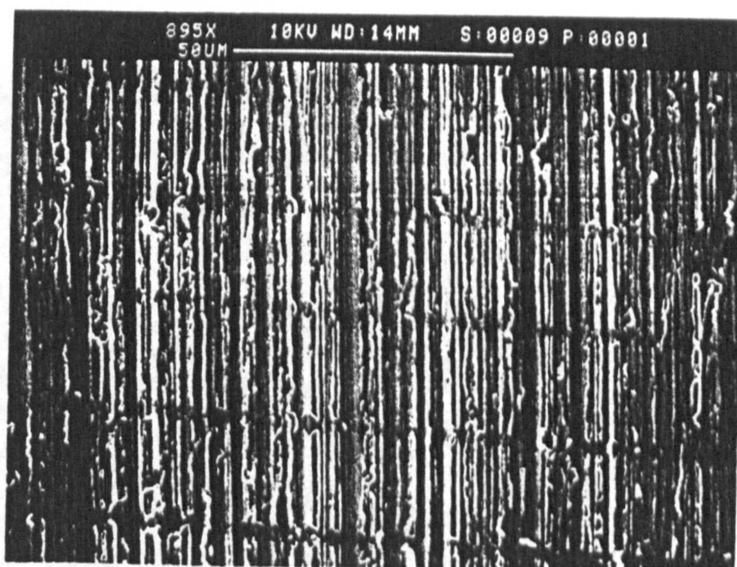


Fig. 4.13 Scanning electron micrograph of stylus tracks on a copper surface. The stylus runs from left to right. From the lowest one, loads were 3.3, 2.4, 1.7, 1.3 and 1.1 mN, respectively.

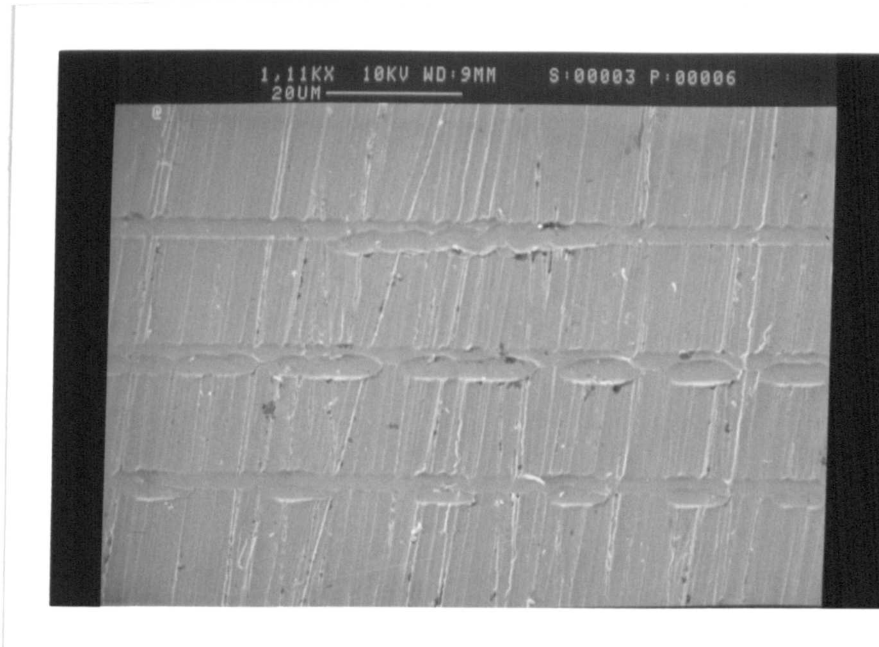


Fig. 4.14 Scanning electron micrograph of stylus tracks on a steel surface, showing scuff marks.

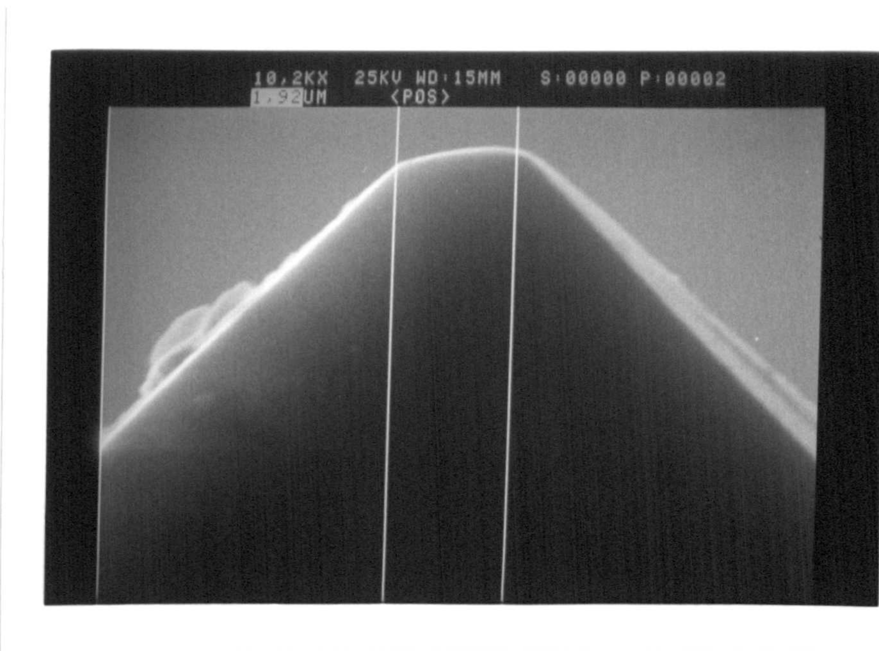


Fig. 4.15 Scanning electron micrograph of the stylus tip prior to the start of the test series.

5. FRICTION FORCES BETWEEN A STYLUS TIP AND SPECIMENS AT LOW LOADS

5.1 Introduction

Analysis of frictional forces is important for assessment and control of wear with any sliding pair. Extensive studies of macroscopic behaviour have been performed and are well documented. From these studies, the macroscopic nature of friction is becoming clearer. Recently, attempts have been made to measure friction and wear behaviour at the molecular level (Marti *et al.* 1990). However, friction of individual microscopic contacts has not been so vigorously investigated. For indentations below 1 μm, it has been reported that the apparent hardness of a material increases dramatically (Gane 1970 and Pethica *et al.* 1983). Because of this and other possible scaling effects it is by no means clear whether conventional frictional behaviour occurs with nominal contact areas as small, or smaller, than the dimensions typical to many styli.

The quest to understand and observe natural phenomena on refined microscopic scales has led to the development of conceptual and technological devices allowing the interrogation of materials with increasing resolution. On the experimental front, the importance of investigating single asperity contact in order to study the fundamental micromechanical response of solids has been long recognised. Such conditions are usually associated with tip on flat configurations, with a tip radius of 1-2 μm or less (Pashley *et al.* 1984, Skinner & Gane 1972, Maugis *et al.* 1976). In the present work, the contact between a stylus tip and a specimen is investigated. Frictional forces on specimens were measured by attaching them to a characterised linear spring mechanism and monitoring its deflection during a stylus traverse. The load between the stylus and the specimen could be continuously varied using the force transducer, as reported in chapter 4. Materials of wide

applications in mechanical, electrical and electronic engineering, brass, copper, aluminium silicon and mild steel were chosen to provide a range of mechanical properties for this initial investigation.

5.2 Literature Review

It is well known that friction is caused by forces between the two contacting bodies, acting at their interface, but the exact nature of friction is still unclear. Over several hundred years, different methods have been developed to build up the basic mechanism of friction, and each has had its proponents among scientists and engineers. Interestingly, the overall situation has changed relatively little. With some modifications, the same general basic mechanisms are still thought to be responsible for friction. The general consensus seems to be that all the various mechanisms may be involved in the generation of friction but that the dominant mechanism in each case depends on the particular situation. In general, friction is considered to depend on the nature of the two surfaces, the materials, the environment, the application conditions, and certain characteristics of the apparatus, such as vibrations and specimen clamping. The basic mechanisms that are involved, to varying degrees, in generating friction are (1) adhesion, (2) mechanical interactions of surface asperities, (3) ploughing of one surface by asperities on the other, (4) deformation and/or fracture of surface layers such as oxides, and (5) interference and local plastic deformation caused by third bodies, primarily wear particles, trapped between the moving surfaces.

In history, basically, there were two schools of thoughts: a French school, which emphasised mechanical interaction of surface roughness or asperities, and an English school, which emphasised 'cohesion' or adhesion between the materials (Dowson 1979, Bowden & Tabor 1964, Ludema 1987). The French scholar, Amontons in 1699, began the experimental study by setting up the two main 'laws' of friction, often called Amontons' laws:

- The friction force is proportional to the applied load
- The friction force is independent of the apparent area of the contact

The same relationship had been observed by Leonardo da Vinci 200 years earlier. But unfortunately, Leonardo's notes and manuscripts were hidden away in private collections and were

only discovered and printed fairly recently (NATO ASI, 1991). The great contribution of the French school was to emphasise that contact occurs only at discrete points. Its major failing was its belief that the contact was determined solely by the original geometry of the asperities and its exclusion of plastic deformation and asperity shape change from the model.

The English school was actually started by a Frenchman, Desaguliers, in 1724. He introduced the concept of cohesive force (now called adhesion). Desaguliers observed that if two lead balls were pushed together with a light twist, they would stick together and that took significant force to separate them again. He then suggested that friction can be largely attributed to the adhesion between asperities that come into intimate contact with one another.

Research in friction accelerated and reached a firm foundation with the work of Bowden and Tabor in the mid 20th century (Bowden & Tabor 1950). Their work focused on adhesion as a major cause of friction, but also showed that more than the outmost layers is involved — that is, that both adhesion and deformation of subsurface material provide important contributions to the energy dissipation in friction. In their earlier work, Bowden and Tabor assumed that the contacting asperities would deform to the point of plastic flow and reach a contact pressure equal to the indentation hardness of the material. The real area of contact, A_r, is then determined from:

$$A_r = \frac{F_N}{H} \tag{5-1}$$

where F_N is the normal load in Newtons and H is the flow hardness in N m^{-2}. If it is further assumed that friction is due to the shearing of bonds, then the friction force would simply be A_r times the relevant shear stress, s. In that case:

$$\mu = \frac{F_k}{F_N} = \frac{A_r s}{A_r H} = \frac{s}{H} \tag{5-2}$$

This expression satisfies both of Amontons' laws in that contact area and load are eliminated. Because for most metals $H \sim 3Y$, where Y is the flow stress and $s \sim 0.5$ to $0.6\ Y$. Thus a value for μ of 0.17 to 0.2 implies a constant value for coefficient of friction. Indeed, this value is often found for clean metals in air, but as later discovered, much higher values are found in vacuum

when the surfaces do not have a protective oxide film. It was also suggested that shearing could also take place below one of the contacting asperities, especially if one of the materials was substantially weaker than the other. In that case, the weaker material would wear.

Tabor and his Cambridge students have continued work on friction and wear over the past half century. Much of our present understanding is due to their dedicated efforts. It is well accepted that for metals, friction is a process of welding, shearing, and ploughing. High pressures developed at individual contact spots cause local welding, and the junctions thus formed are sheared subsequently by relative sliding of the surfaces. Ploughing by asperities of harder surface through the matrix of the soft material contributes the deformation component of friction (Desmond 1975).

From an engineering point of view, strong adhesion between sliding surfaces becomes important only for very clean surfaces in a very high vacuum, such as space applications where the phenomena of adhesion and attendant seizure are extremely important (Buckley 1981). The contacting surface asperities are cold weld together and form intimate atomic bonds across the interface, and the friction coefficient can be very large.

However, in recent years, there are a couple of interesting findings which can not be readily explained by present theories. For example, the friction coefficient for a wide range of material combinations decreases somewhat as the loads increase. This behaviour has been attributed to a transition from elastic to plastic loading, the influence of surface energy, and various other factors (Blau 1987, Bowden & Tabor 1964). In addition, problems have arisen from observations of the friction of micromechanisms, such as motors, gears and other engineering components with masses in the microgram range, and of atomic force microscopes, in which sliding surfaces operate with normal loads down to nanonewton level. The discrepancy of the results obtained by these devices is perhaps most easily seen if we note that at very light loads micromechanism tend to stall, indicating infinite friction (Lim *et al.* 1990), while atomic force microscopes often give friction coefficient values tending to zero (McClelland *et al.* 1988). There is no comprehensive explanation for these discordant friction coefficient values.

On the theoretical front, recent advances in the formulation and evaluation of the energetics and interatomic interactions in materials coupled with modern molecular simulation techniques (Landman *et al.* 1990, 1991) open new avenues for theoretical investigations of the microscopic origins of friction. Probably, one day, large scale molecular dynamics computer simulation may help us to see what is actually taking place at the interface. On the other hand, the energy dissipation approach to friction is beginning to be recognised. It may bring us a new picture of friction as addressed by Tabor (NATO ASI, 1991):

"*.....I came to see that all frictional processes - or most of them - involve a distortion of atomic arrangement, an instability, a flicking back to a new position of equilibrium while the distorted material vibrates and the vibrations ultimately degrade into heat.*"

5.3 Analysis of Microscopic Friction

5.3.1 Contact between a tip and surfaces

The contact between a stylus tip and a surface can be simplified as a smooth ball with a radius of the tip on a flat surface, because the local surface seen by the tip with a typical size of 2 μm is relatively flat. As mentioned in the first chapter, such contact can be simply expressed by Hertzian theory. The radius of the contact circle, a, and the mutual approach, δ, are

$$a = \left\{ \frac{3WR}{4} \left[\frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2} \right] \right\}^{\frac{1}{3}} \tag{5-3}$$

$$\delta = \frac{a^2}{R} = \left(\frac{9W^2}{16R(E^*)^2} \right)^{1/3} \tag{5-4}$$

where ν is the Poisson ratio and E the Young modulus of each material. The suffices 1, 2 refer to the two bodies in contact, R is the radius of the tip and W is the loading force. The contact area is given by

$$A = \pi a^2 = \pi R \delta \tag{5-5}$$

where δ is the mutual approach or elastic deformation made by the tip under a load of W. Note that the relation indicates that the surface outside the contact region is displaced in such a way that the actual area of the contact is only one half of the geometrical area of a sphere truncated by δ, which is clearly equal to $\pi a_g^2 = 2\pi R\delta$. In a situation of plastic deformation the contact area will be larger than the geometrical area, as demonstrated in Fig. 5.1.

More accurately, the contact radius can be calculated from JKR (Johnson, Kendall and Robert) theory. The significant adhesive force depends surface energy released by the interaction of the two surfaces, and is dependent on the contact area and the conditions of the surfaces. It can be very strong when the surfaces are clean, and can be insignificant in the presence of contaminants. The JKR contact radius is given by

$$a = \left\{ \frac{3R}{4E^*} (W + 3\gamma\pi R + \left[6\gamma\pi RW + (3\gamma\pi R)^2 \right]^{1/2}) \right\}^{1/3} \qquad (5\text{-}6)$$

where γ is the surface energy released by the interaction of the two surfaces and E^* is effective elastic modulus, given by equations (1-6) and (1-3b) respectively.

For metals, the magnitude of surface energy is ~ 0.2 J m^{-2} and for Van de Waals solids such as polymers or rubber is ~ 0.04 J m^{-2} (Johnson 1982). Such adhesion will not be seen on real surfaces due to contamination and their surface roughness.

In static Hertzian contact the peak shear stress lies centrally below the surface at a distance equal to $0.67a$ (Halling, 1975). Furthermore, we find that as the load is increased this peak shear stress will reach its critical value, S_c, when the maximum pressure at the centre of the contact zone is $3S_c$. Further increasing the load indicates plastic deformation but the plastic zone is constrained by elastic material on all sides. Fully plastic flow will take place only when the mean contact pressure is about $6S_c$. The mean pressure under these conditions is essentially the indentation hardness, H, and this is why for metals we have

$$H \approx 6S_c \approx 3Y \qquad (5\text{-}7)$$

where Y is uniaxial tensile yield strength of the material.

In the presence of a friction force, the location of the maximum value of shear stress is much nearer to the surface, thus the plastic deformation can take place more readily than in the previous case. In other words, macroscopic plastic deformation is facilitated by the presence of such friction tractions. The maximum tensile stress occurs on the circumference of the circle of contact. The effect of friction is to add a compressive stress to the front edge of the contact and to intensify the tensile stress at the back edge. Thus for brittle materials, the region near the back edge is the most likely point of failure (Hamilton, 1983).

5.3.2 Brief review of friction models

At present, microscopic friction is classified into three levels which are described by three different models: normal friction, boundary friction and interfacial friction. Each has particular applications.

Most friction processes occur with surfaces 'damaged' in one form or another. This is generally referred to as 'normal' friction. In some cases, the surfaces will slide past each other while separated by large, almost macroscopically sized particles of wear debris. In other cases, usually under high loads, but also depending on the smoothness and hardness of surfaces, the damage or wear may be localised within a much narrower interfacial region of plastically deformed nanometre -sized asperities.

There are also situations where sliding can occur between two perfectly smooth undamaged surfaces. The term 'boundary friction' is mostly restricted to a thin boundary region, which, as in a grain boundary, would typically not extend more than a few nanometres on either side of it. The term 'boundary lubrication' is commonly used to denote the sliding of two surfaces separated by thin (possibly monomolecular) layers of some suitable boundary lubricant, though here too it is presumed that plastic deformations of isolated asperity contacts are also occurring during sliding.

Interfacial friction is restricted to a molecularly thin interfacial region having a uniform gap thickness and a well defined contact area, such as friction between an atomic force microscope and a surface. Clearly, the interaction forces associated with interfacial friction would be much more localised than in the case of boundary friction or normal friction (Israelachvili *et al.* 1988). It has

been found that during interfacial sliding, the friction depends critically on the intermolecular forces between surfaces, their contact area, the precise distance between the two surfaces (at the angstrom level), but it depends only weakly on the applied load (Krim *et al.* 1990, 1991, Landman *et al.* 1990).

In brief, in the case of unlubricated sliding of 'dry' surfaces, the friction force at low loads is described by the equation originally proposed by Bowden and Tabor: $F = S_c A$, where S_c is the critical shear stress and A is the molecular area of the contact. The variation of A with applied load W is well described by JKR theory, even during sliding and even at negative values of W. At higher loads there is an additional contribution to F that is proportional to W, given by $F = CW$. This contribution is analogous to Amontons' law, but C has a different origin and exists even in the absence of adhesion. When damage occurs, Amontons' law $F = \mu W$ becomes applicable where μ is considered to be the traditional coefficient of friction.

5.4 Instrumentation

5.4.1 Apparatus for measuring friction

In the investigation of this chapter, friction between the stylus tip and a specimen was measured in a typical laboratory environment during a stylus traverse in order to monitor the interaction and evaluate the effects. To do this, a specially designed friction force transducer is used (Smith, 1989). It consists of mainly three parts: a specimen platform, a mechanical translation stage and a displacement transducer. The complete instrumentation is shown in Fig. 5.2. A specimen is mounted onto the platform of the simple, notch type linear spring which is deflected by the drag force of the stylus on the specimen. Connected to the translation stage is a Rank Taylor Hobson Talystep LVDI transducer with its measurement axis aligned to the deflection direction of the spring stage. The normal profile of a specimen is measured by the modified Talysurf 5 at a constant setting force applied by the variable force actuator.

Because of its important role in the measurement, it is necessary to mention the notch type spring system. This spring is an aluminium monolith and its compliance is a result of the thin web that remains after two holes have been drilled closely together. The stiffness of the spring is an important factor that has to be chosen with care. High stiffness will loose the sensitivity of friction measurement because of the resolution of the displacement transducer, and low stiffness will, obviously, introduce lateral distortions on the profile measurement. The measured stiffness of the complete spring system is 1786 N m^{-1}. If the maximum normal force is limited to less than 5 mN during the friction tests, then the worst lateral deflection of the specimen would be less than 0.9 μm for a typical friction coefficient of 0.3. This is judged to be near the maximum that can be tolerated without significant lateral errors for the 2 μm stylus tip. The natural resonant frequency of the spring flexure is approximately 70 Hz, which is much higher than the stylus pick-up, and the dynamic effect can be considered to be insignificant. The natural damping of the spring system is very small because of its monolith construction. To avoid a long settling time of the system after handling a specimen on it, for instance, cleaning the sample, or loading and unloading the stylus to and from it, an external damper is added to the system. A paddle of dimension 10 mm by 16 mm immersed in a bowl of oil of viscosity 320 cst was attached to the platform to provide damping. The damping can be roughly adjusted by changing the level of the oil.

The LVDI transducer plus a standard Talystep amplifier provides the measurement of the deflection of the spring system. The transducer was calibrated to within 2% and has a resolution of better than 0.5 nm, giving the frictional force measurement with a resolution of better than 1 μN.

5.4.2 Calibration

The frictional force transducer was calibrated against both standard weights and the variable stylus force actuator. The friction force transducer was mounted to a cast iron block with its measurement axis in the vertical plane. The initial offset due to gravitational loading was then compensated by the micrometer adjustment on the inductive transducer mount, see the photograph in figure 5.2(a). Standard weights ranging from 0.1 to 1 gram were then placed onto the end of the spring platform and the output from the Talystep instrument was monitored. This is a direct

measurement of applied force to the spring platform and will include effects of stiffness of the inductance gauge and distortion of other components within the force loop. For the second calibration, the force was provided from the variable stylus force actuator. This was done by lowering the stylus onto the end of the platform of the friction force transducer which was again mounted with its axis vertical. Known coil current and thus force, was then given from a digital to analogue converter, and the deflection of the spring system was recorded. The results from both calibrations are shown in Fig. 5.3. Both give a consistent output corresponding to a compliance of 0.56 $\mu m\ mN^{-1}$, with a maximum deviation of residuals from a least squares straight line fit of less than 0.3% for the standard weight and 1% for the stylus loading. The second calibration using the stylus force actuator demonstrates once again the good performance of this force actuator. This calibration includes all of the system errors associated with both the friction test device and the stylus force actuator.

5.4.3 Other sources of error

As indicated by the schematic diagram of the system in Fig. 5.2 (b), the stylus arm is pulled against a knife edge pivot by a bias spring that consists of a leaf spring and an adjustable coil spring for fine tuning the spring force at the zero point of the displacement transducer output. Consequently, the actual contact forces will vary in relation to the position of the stylus about this zero point. The magnitude of this variation was investigated in chapter 3. Based on those results, a maximum force variation of less than 30 nN is expected for the displacement range of about 3 μm used in all subsequent experiments.

Another source of error is drift, due mainly to thermal variation of electrical and mechanical components of the test system. The warming up time for electronic components is generally much less than the settling time for the mechanical parts, especially when a combination of different materials were clamped together. After setting up a test, it is better leave it for a period of time at least an hour with the instrument switched on to allow the whole system to reach its thermal equilibrium. The drift was monitored after the warm-up. For the stylus force actuator the drift rate was 0.5 μNs^{-1}. The drift rate of the friction device is dominated by the LVDI transducer. The friction

forces were obtained by measuring the step change in force between forward and reverse motions. Linear drifts were removed by eye (since least squares fitting across a step is misleading) but are considered to introduce negligible errors in these experiments.

5.5 Indentation Tests on Specimens

Specimens of mild steel, copper, brass and aluminium used for previous experiments were used for friction measurement. Before doing any such experiments, contact and penetration of the stylus into the surface of a specimen was measured.

The combination of Talysurf 5 with the variable force actuator readily provides instrumentation for indentation testing. The specimen was mounted on a flat steel block, and then the stylus was initially contacted with the surface under a very light load. The load was then increased using the electromagnetic force transducer while the resultant displacement was monitored from the output of the Talysurf 5.

Typical results for mild steel, copper and aluminium undergoing a steadily increasing load up to 6 mN are shown in Fig. 5.4. It shows the indentation characteristic typical of many metals at these low loads. The shape of this curve is determined by the geometry of the interface. In this case the stylus had a flat end of approximately 2 μm width. Consequently, the initial portion of the curve can be associated with elastic deformation. This is verified by looking at the initial slope of each of these curves. The ratio of slopes of steel to copper to aluminium is 3:1.6:1 which corresponds very closely to the ratio of elastic moduli of these materials. It is clear from these curves that steel is by far the hardest material with a total indentation depth of approximately 80 nm followed by copper at 250 nm and aluminium at 340 nm. Quantifying hardness values is not possible for these tests owing to the uncertainty in the shape of the stylus probe. However, it has been shown that the ratio of the flow pressure P to the yield stress Y of any material is constant for a given material and indenter geometry and dependent upon the ratio of elastic modulus to yield stress (Johnson 1970).

The stylus was a pyramid with a truncated end of width of 2 μm, but wear on the top of the stylus results in rounding of the corners with the flattening towards a spherical shape, see Fig. 4.15.

Consequently, the contact zone may reasonably be considered circular of, say, radius a. Based on these assumptions, it is possible to derive a simple expression for the contact radius as a function of the applied load

$$a = \left[\left(\frac{F_n}{\pi Y}\right)\left(\frac{Y}{P}\right)\right]^{\frac{1}{2}} \qquad (5\text{-}8)$$

where F_n is the normal force.

As far as we know, the values for the ratio of P/Y do not exist for the case of a rigid, flat punch but it may be considered reasonable to assume that these would be not dissimilar to those for a very blunt wedge. Using values for a blunt wedge having a 170° included angle (Hirst & Howse 1969), the corresponding contact radii at the load of 6 mN were calculated and are shown in Table 5.1. From this, it can be seen that there is strong correlation between the expected contact radius and measured depth of indent. The predicted value is larger than the width of the flat tip for both the copper and aluminium specimens. There are a number of possible reasons for this discrepancy. Firstly, the contact region may well be a mixture of rectangular and spherical which would increase the actual area resulting in a 10% reduction relative to the calculation of the radius for a square region and even greater if it were elongated to a rectangular contact. Also the yield stress is based on that for an annealed bulk material. Consequently, any work hardening of the surface would cause a local reduction in the ratio P/Y up to a maximum factor of approximately 3, as well as direct influence on the radius through the denominator of Eqn. 5.8 (Johnson 1970). Finally, equation 5.8 uses the projected contact radius which will be close to real value for large radius probes and shallow indentations. Although these errors will result in a discrepancy of the contact radius, possibly by 50% or more, the above analysis indicates that reasonable estimates of the static indent shape may be made using these relatively crude assumptions and it is envisaged that further refinement of the model will increase the accuracy of this prediction.

It can be seen that the elastic deformation in these materials was in the range of 10 nm for steel, 20 nm for copper and 40 nm for aluminium specimens for an applied load of 1 mN. At this load, no surface damage has been found during the previous investigation. Therefore, an elastic

contact is assumed for these materials at the load of 1 mN. Upon this, the calculated Hertzian contact radius for specimens of steel, copper and aluminium is 0.18, 0.22 and 0.27 μm respectively, while the measured depth is 10, 20, 40 nm, and thus the corresponding radius is 0.14, 0.2 and 0.28 μm for steel, copper and aluminium according to Eqn.5.4. A spherical tip shape is assumed in the calculations. The values correspond closely, and the slight difference between the two sets may be due mainly to uncertainty in the depth measurement as they are close to the resolution of the instrument. The adhesion effect can be insignificant due to the presence of surface contaminants because the specimens were prepared and operated in a normal environment.

5.6 Results and Discussion

5.6.1 Friction measurement

Five types of materials were chosen for this investigation. Besides the mild steel, copper, brass and aluminium used previously, silicon was used because of its wide applications in electronics industry. The silicon specimen was obtained by breaking a small portion from a single crystal wafer that had been previously polished for microelectronics applications. The surface finish of this specimen was better than the realisable precision of the measuring instrument. Before mounting onto the platform of the friction device, each specimen was flooded with acetone and wiped. This is characteristic of cleaning performed in most engineering measurements of surface finish. The specimen was glued to the platform using a silver loaded-paint which could be easily removed by acetone.

During each measurement, the set force was determined from the coil current, the stylus output and friction output were recorded simultaneously by a computer through the Metrabyte DAS-16 I/O board. In each run, the stylus was traversed in auto-return mode. Thus we have forward frictional force and reverse frictional force. The friction coefficient was obtained by summing the mean force in each direction and dividing by twice the normal load. Ideally, variations in normal force caused dynamically or statically by the vertical stylus movement are averaged over the surface,

so that no independent instantaneous measurement of normal force is undertaken. This approach assumes that friction is independent of traverse direction. The arcuate action of the stylus arm means that the dynamics are not identical in both directions of traverse but this is unlikely to affect friction to a significant extent with the roughnesses and traverse speeds used here.

Sets of repeated tests were collected for each specimen at three different traverse speeds. At each speed there are about 15 to 20 different loading forces in a range from 0 to 6 mN, and at each loading force the specimen was traversed three to five times to get a fair reading of the measurement. For each data file, an effective length of surface profile and frictional force has to be selected and levelled by a least squares line, and then the standard deviation and mean values were calculated.

5.6.2 Friction coefficients

The calculated friction coefficients for the five materials as a function of load are plotted in Fig. 5.5 (a) — (e). The curves on the graph correspond to stylus traverse speeds of 1.0, 0.25 and 0.1 mm s^{-1}. At this level, the surfaces gave a friction coefficient below that normally encountered with these materials. The friction coefficients for these materials are around 0.1 and 0.2, with an exceptionally low value for silicon of about 0.05 to 0.06.

Generally, there is little change in the overall friction coefficients with loads at the levels investigated. The increased scatter at loads below 1 mN may be attributed to measurement errors since the friction forces are becoming close to the uncertainty of the instrumentation. Dynamic variations in the normal force due to the stylus accelerations are expected to be no more than a few tens of micronewtons under most test conditions. However, the worst combination of high traverse speed and short wavelength surface detail could cause occasional peaks to around 0.25 mN and so might contribute to the increased scatter seen with low impressed forces. The fastest sliding speed appears to give the highest coefficient of friction, which tends to reduce as the load increases. It falls below the values for the other speeds at around 3 mN for copper, and 4 mN for the other metals and becomes comparable with that at the other speeds with silicon. As loads are raised above this transition, the friction coefficients of the metals tend to rise slightly. It is possible that the same drop

occurs for mild steel at higher loads, beyond the range of the investigation. These results are consistent with the observations of stylus tracks using a scanning electron microscope in the chapter 4, which showed less visible damage with stylus measurements at the higher sliding speeds. The reasons for this rate effect are unclear but may relate to the strain rate sensitivity of the hardness of the surface asperities. In general, the friction coefficient of both metal and non-metals in sliding contact decreases with increasing speed, and this decrease is believed to attributable to frictional heating (Bowden & Tabor; Part II 1964). At a high speed, the temperature at the contact spots could be very high because of a sudden release of heat, and the immediate contact area may experience considerable softening or even a change from solid to liquid, thus reducing the shear stress and so the coefficient of friction. However, in our case, the sliding speed is relatively low compared to those quoted by Bowden and Tabor. It is more appropriate to assume that the adhesion junction formed at the contact is time dependent as originally proposed by Sampson (1943), the longer the time for the contact to form the stronger the adhesion. As a result, the lower the sliding speed the stronger the adhesion force and thus the higher the frictional force.

The general reduction in coefficients of friction encountered with these specimens is probably due to two reasons, the increase in hardness at the surface of these materials and the existence of a contaminant film. That hardness at surface is different from the bulk material has been observed with a wide range of materials (Bowden & Tabor; Part II 1964). The difference can be several times higher than at light loads than at higher ones, Gane (1970) and Pethica *et al.* (1983). In our case the increase in hardness is mainly due to the work-hardening of the surface layer during the polishing of specimens. However, it not clear whether the shear stress would be increased, decreased or remain unchanged by this work-hardening. But surely a reduction of the shear stress to hardness ratio would help to reduce the coefficient of friction.

The contaminant film plays an important role in the reduction of frictional force. For instance, a friction coefficient for copper on copper was found to be 0.5 in air and it increased to nearly 5 in vacuum after prolonged heating (Bowden & Tabor; 1950 Part I). This shows that the removal of only surface film leads to a very great increase in friction. This is due largely to the ease

with which the metal surfaces can weld together in the absence of contaminant films. It would also seem that the shearing action of the sliding process itself produces an increase in the area over which metallic junctions are formed. When specimens are cleaned in air, their surfaces are still covered by a thin film of oxide, water vapour, and other adsorbed impurities. This contaminant film is usually at least several molecular layers in thickness. In the presence of the contaminant film the coefficient of friction is dominated by two factors, the hardness of the substrate *i.e.* the specimen surface and the shear stress of the film. For most cases, the shear stress of a contaminant film is normally low, thus leads to the reduction of the friction coefficient.

5.6.3 Effect of surface finish on friction

On the basis of experimental observations and theoretical work, the frictional force for metal sliding is generated by adhesion, ploughing by wear particles, and asperity deformation. It has been suggested that the ploughing component contribute more to the frictional force than the adhesion. This means that mechanical interactions at the sliding interface are primary causes of friction. If the mechanical interactions between the two surfaces are reduced, the frictional force would be significantly lowered.

To investigate the effect of surface roughness on frictional force, surface finish of mild steel and copper specimens was improved by further polishing to give an order of magnitude reduction on roughness and experiment was repeated. There was no significant difference in either the trends or the values of the coefficients measured. However, the variations of the drag forces about their mean values were much reduced. For example, on the copper specimen, at a nominal load of 0.7 mN and a tracking speed of 0.1 mm s^{-1}, the standard deviation of the drag force was 49 μN for a surface roughness value 0.18 μm, and it reduced to 3.2 μN for the surface roughness of 0.013 μm. This indicated that the frictional force was dominantly determined by adhesion and deformation. Also the localised contact between the stylus tip and the surface, which is in submicrometre level, was little affected by the improving of the surface finish as the average surface features were, comparatively, already larger than the stylus size. This suggests that the ploughing component of friction is not significant under the condition of this test.

5.6.4 The relationship between the friction and surface finish profiles

The present experimental arrangement enables simultaneous measurement of the friction force and the surface profile. Fig. 5.6 shows a typical plot for a forward and reverse traverse over a distance of about 1 mm on a copper specimen at a nominal load of 5.9 mN. Clearly for a perfect measurement in which no damage is incurred and the stylus travels over the same track in both directions, the reverse profile measurement would be a perfect mirror image of the forward trace. Behaviour closely consistent with this pattern is clearly visible in Fig. 5.6 (a). The 1.2 mrad slope of the profile is caused by a slight misalignment between the specimen surface and the axis of motion of the stylus instrument. From the previous error analysis, a variation of only about 2 μm in the stylus height will not induce significant errors in the normal force calculation. The alignment was considered adequate in a view of the increasing difficulties in levelling to better precision. The changes in force output before sliding occurs are due to the start up characteristics of the instrumentation. A typical start process shows two small step height changes on the profile and corresponding changes on frictional force. This is thought due to the start up of the stylus traversing unit and the well known stick-slip behaviour. Mechanical coupling backlash in the unit caused the first drop, in which the stylus was in pull stress. Stick-slip between the stylus and the specimen surface caused the second change in force and thus in profile. Stick-slip occurred at the start and the end of sliding. The starting spikes are due to the static frictional force which tends to be higher than the kinetic one and the stoping ones occur when two sliding surfaces are brought to rest over a finite time during which strong adhesion is generated before the surfaces stop moving (Bowden and Leben 1939; Bowden and Tabor Part II, 1964).

To examine local variations the profiles were extracted and levelled by subtracting out a linear least squares straight line. Plots of surface profile and tractive force variation for both traverse directions are shown in Fig. 5.7. Two data sets of surface profile and tractive force were processed by cross-correlation in 'Matlab' and the results are shown in Fig. 5.8, in which the maximum coefficient of the cross-correlation is approximately 0.58. These graphs show an apparently negative correlation in the forward traverse and a positive correlation for the reverse direction.

The behaviour of this correlation can be explained as follows. The stylus contact force is essentially perpendicular to the trend line of the surface and the stylus drag force measured parallel to this. The simplest model of the interaction between these forces is that a horizontal component caused by local surface slope and a friction force (governed conventionally by Amonton's laws) opposes the motion tangentially to the local contact angle. The force balance at the stylus tip is shown schematically in Fig. 5.9. Since the stylus force F is impressed, the normal traction F_n at the surface must support it and so, in the absence of friction,

$$F_n = \frac{F}{\cos\beta} \tag{5-9}$$

Once friction is included, it is important to take care over the sign convention and consistency of the forces being examined. The instrumentation measures forces imposed onto the specimen and only these forces will be considered in this discussion. The directions considered positive are shown in Fig. 5.9. The normal force on the surface is $-F_n$ and this may be resolved into vertical and horizontal components $-F_n \cos\beta$ and $F_n \sin\beta$ respectively. The local frictional force acting on the surface is simply $|\mu F_n|$ acting in the same direction as the stylus motion and may be similarly resolved into vertical and horizontal components. If Eqn. (5-9) is used to model the normal force, the overall horizontal (apparent frictional) force F_h acting on the surface will be

$$F_h = F(\tan\beta \pm \mu) \tag{5-10}$$

where the positive sign is taken when the traverse is in the positive x direction. However, consideration of the stylus equilibrium shows that the vertical component of the local friction force will cause F_n to vary if F is truly constant. The interaction between friction and normal forces can be modelled by an equivalent static equilibrium to give

$$F_n = \frac{F}{\cos\beta \pm \mu \sin\beta} \tag{5-11}$$

where the positive sign now corresponds to a traverse in the negative x direction. Generally we expect β to be small and μ is also small, so the error of using Eqn. (5-9) is reasonably estimated as a

modulation by $(1 \pm \mu \tan\beta)$. Given that the model is simple, the few per cent variation typically given by this modulation will be neglected and Eqn. (5-9) is used.

The first term in Eqn. (5-10) expresses a variation about the mean horizontal friction drag that is dependent upon local surface slope. As shown in Fig. 5.9 with the traverse proceeding from left to right (*i.e* positive), both terms act to the right. On the other flank of the asperity the terms act in opposition since the sign of β is reversed. However, with the traverse from right to left, the true friction drag will be negative while the slope remains unaltered. The effect on the flank shown is thus for the horizontal drag to be increased above its mean value when traversing to right but decreased for a traverse in the opposite direction. It is this that causes the correlation of drag to the profile to be sensitive to the direction of traverse in Fig. 5.8. Furthermore, if the simple model is adequate, the profile related term, being slope dependent, should correlate with the derivative of the profile with respect to traverse distance. Fig. 5.10 shows the cross-correlation between the friction profile and the displacement profile differentiated using a five-point symmetrical Lagrangian formula that has proved useful for roughness assessment in other contexts (Chetwynd 1991).The correlation now shows a single peak of small width and a maximum coefficient of 0.89. Given, particularly, that the noise levels will be exacerbated by differentiation, this correlation is very high and suggests that this simple model describes well the behaviour in the regime being studied. To verify further that our system is behaving consistently, the profile was also cross correlated with its own derivative. This resulted in a curve very similar to those of Fig. 5.8 with a maximum value of 0.62, giving only 4% in difference.

5.7 Friction of a Steel Ball on Specimens

From above investigation, the contact between a stylus tip and an engineering surface can be reasonably considered as a ball sliding on a locally smooth and flat surface. Therefore, a Hertzian elastic contact model can be directly used. However, in a situation where the tip size is relatively larger than the average features of a surface, the surface seen by the tip is no longer smooth but rough. The topographical roughness of such surfaces ensures that they make true contact over an

area which is generally small compared with the apparent area of contact. The real contact pressure at the true contact spots will be much higher than the pressure given by the Hertz theory. In this section the friction behaviour for rough contact surfaces is discussed and initial results are given.

To emulate such a situation where the tip size is relatively larger than surface features, it is simpler to use a large tip size than to make a surface with fine features. Here a steel ball with 1 mm diameter was used for the experiment. The steel ball was embedded in resin and ground down till only a 'cap' remained. This was glued to the end of the aluminium magnet holder, see Fig. 5.2(b). Then using a standard stylus arm, rather than the long reach one, the ball made contact with the surface and the instrument behaved exactly as before except for having effectively a 0.5 mm radius stylus of steel. Frictional force on three types of specimens was measured at a fixed traversing speed of 0.25 mm s^{-1} under an increasing load up to 25 mN. The coefficient of friction was then calculated and plotted in Fig. 5.11 in which (a) shows friction coefficients of the steel ball on a steel specimen for three different tracks, and (b) shows friction coefficients of the steel ball on copper and aluminium specimens. The results are quite different from those with the stylus of 2 μm except for the aluminium which gives a similar curve as in Fig. 5.5 (d). The curve of friction coefficient for copper started with a very high value of about 0.4 at low loads, decreased rapidly as the load increases, then stabilised to 0.24 when the load reached 5 mN. There might be a slight decreasing trend afterwards. On the steel specimen, the three curves are different even though they were obtained from the same specimen. These three curves were obtained in a time order that the bottom line (solid) was first, the middle one (dashed) the second and the top one (dot and dashed) the last. The first run test gave the lowest friction coefficient and the last run gave the highest. There are two distinct transitions on each curve. The first occurs at the load about 3.5 mN for all three tracks and the second occurs at a different loading force; 13 mN for the first run test, 9 mN for the second and 7 mN for the last. The two transitions probably mark different contact states; the first indicates the end of non-metallic contact between the ball and the surface and the second indicates the start of considerable metallic contact and the occurrence of damage. Between these, the contact is a mixture between non-metallic and metallic. At low loads, where the sliding is completely within the oxide

film, the friction coefficient is marked by a slight decrease due to the effect of film thickness. As the load increases some metallic contact is generated and the friction increases slightly. Further increasing the load probably generates damage on either sides of sliding pair and this is marked by a reduction on the friction force which could be due to the debris lubrication. The damage on the ball could be the main reason for the increase on friction and the earlier appearance of the second transition for the other two tracks. On the aluminium specimen there seems no change in the contact which leads to a constant coefficient of friction at all loads. This could be due to the thick layer of oxide film which dominated the friction behaviour at all loads. On the copper specimen the transition from non-metallic contact to some extent of metallic contact occurs at about 10 mN. Again the oxide film causes the behaviour that the friction coefficient reduces as the load increases, and there is no sight of a subsequent rise in friction, which means no second transition, within the loading range.

To confirm the above discussion the contact state between the ball and a steel surface was measured via electrical contact resistance. To do this a pair of copper wires were silver-glued to the steel specimen and the aluminium piece (the steel ball is glued to the end) and connected to a circuit which was originally designed for tunnelling current measurement (detail of the circuit will be discussed in Chapter 7). During each sliding test the contact resistance was measured simultaneously along with surface profile and frictional force. This dynamic contact resistance was then averaged at each different loading. A typical result of the averaged contact resistance against the loading is shown in Fig. 5.12. Three contact regions are clearly marked: non-metallic contact with high resistance above 2 kΩ, some metallic contact with resistance about 1 kΩ, and considerable metallic contact with the resistance less than 100 Ω. Two transitions from non-metallic contact to some metallic contact and then to the considerable metallic contact occur at about 3.5 mN and 12 mN respectively.

After the friction experiments, the steel ball was investigated under both a scanning electron microscope and an optical microscope. A photograph taken from the optical microscope is shown in

Fig. 5.13. There are obvious damage marks on the top side of the ball. The size of the bright spot is about 100 μm, and the damage size is about 30 μm.

5.8 Conclusions

The novel arrangement of electromagnetic force transducer attached to a standard stylus based surface measuring instrument has been demonstrated and shown to provide a versatile method for assessment of static and dynamic contact forces at low loads. Also, the simple flexure used with a high sensitivity displacement sensor has proved to be capable of measuring drag forces to well below 10 μN while deflecting by amounts always small compared with the stylus size.

The stability and controllability of the system have been demonstrated by using it as a microhardness indenter, a task for which was not intended nor its design optimised. Results obtained were, nevertheless, highly consistent with those of other workers.

Friction coefficients on four typical engineering surfaces were measured with a contact size in both micrometer and sub-millimetre levels. For a typical 2 μm diamond tip, the friction coefficient varies little either with load over a range 0.5 – 5 mN or with speed in the region 1.0 – 0.1 mm s^{-1}, except for a slight tendency to decrease at the fast speed. Measured values vary from about 0.25 for aluminium, which is softer than the copper, to around 0.1 on steel and 0.07 on single-crystal silicon. Horizontal drag on the specimens shows fluctuations about a mean level which correlate very strongly with the derivative of the surface profile. This is consistent with a model in which friction is proportional to normal load and modulated by surface slope. For a large scale of contact with a 1 mm steel ball the friction coefficients on these specimens are much higher, 0.3–0.6 for steel, 0.2 – 0.4 for copper and a fair constant value of 0.2 for aluminium. Surface hardness and contaminant film are believed having played an important role in each case. On a large scale contact the hardness of surface is dominated by the bulk properties while on a small scale contact the hardness could be much increased due to the work-hardening process. The reason that aluminium has quite close results for the two cases is probably that the sliding was happening within the thick oxide layer in both situations.

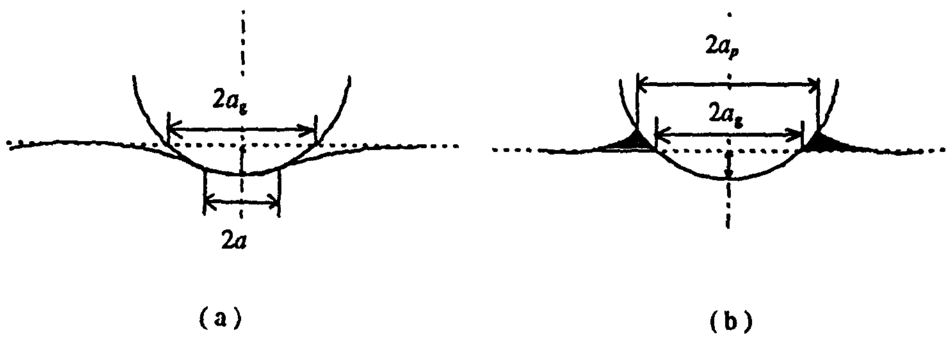


Fig. 5.1 (a) Elastic contact: contact area less than the geometrical area
(b) Plastic contact: contact area greater than the geometrical area

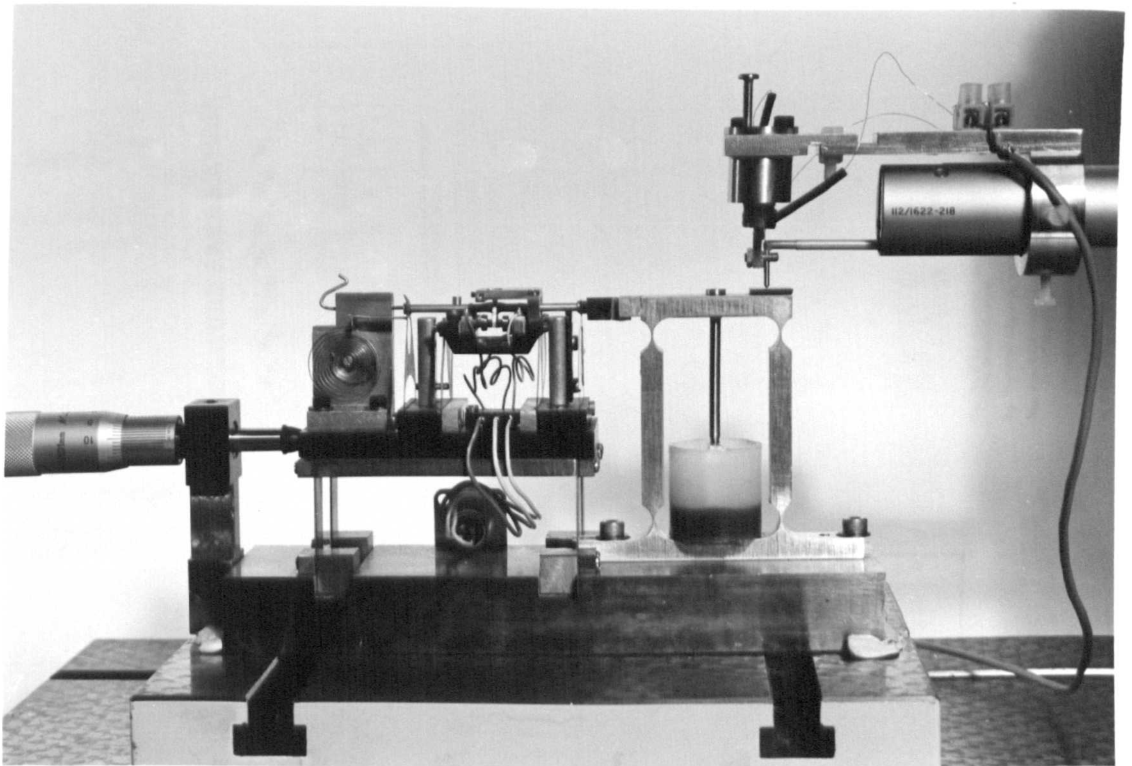


Fig. 5.2 (a) Photograph of the instrumentation for measurement of frictional forces.

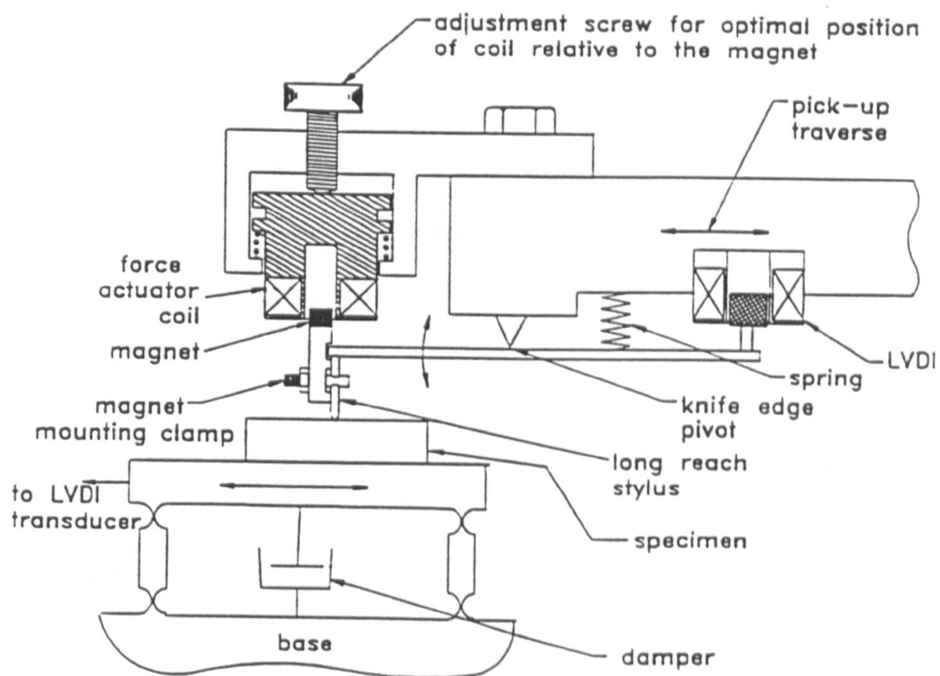


Fig. 5.2 (b) Schematic representation of the instrumentation for measurement of frictional forces during the traverse of a diamond stylus over a surface at low loads.

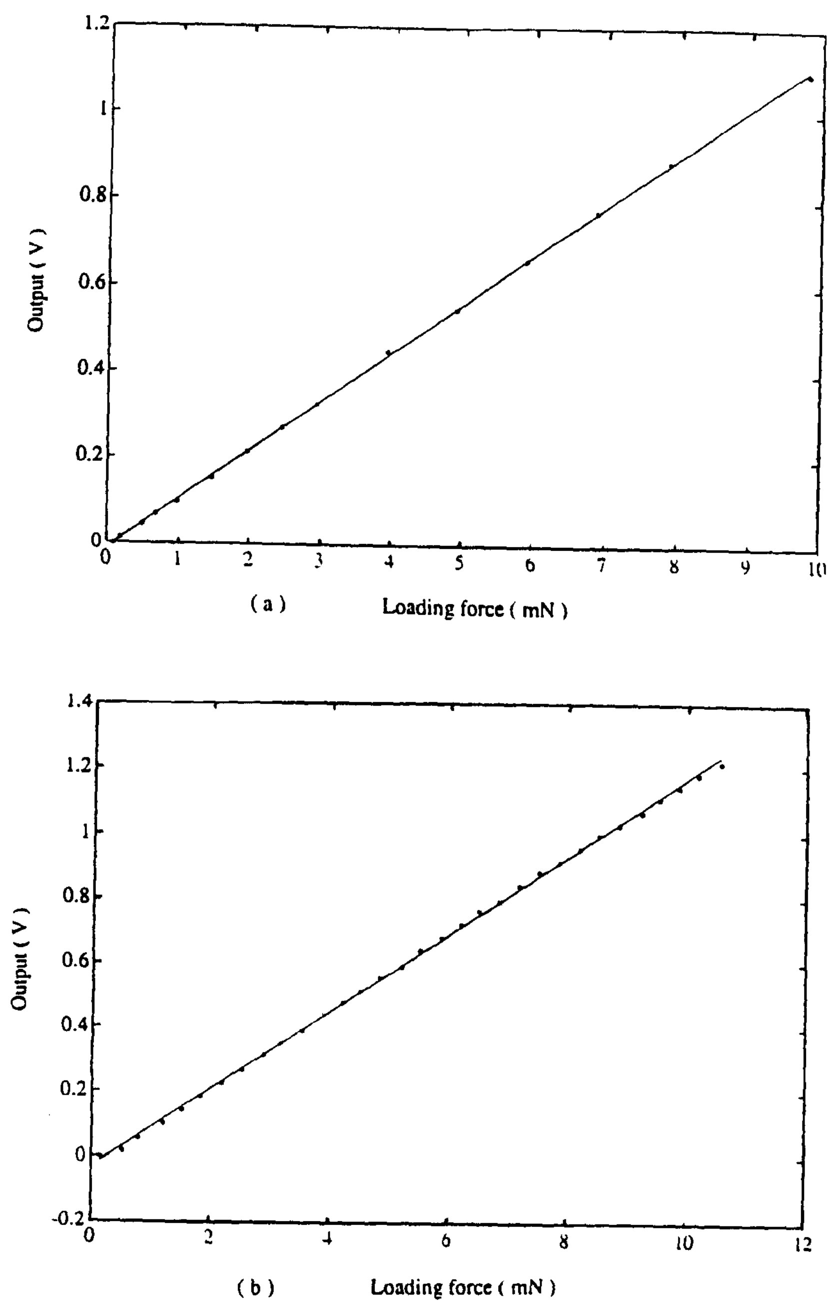


Fig. 5.3 Characteristic of the friction force transducer, (a) output from force gauge for dead weight loading and (b) output from the same gauge when loaded by the stylus.

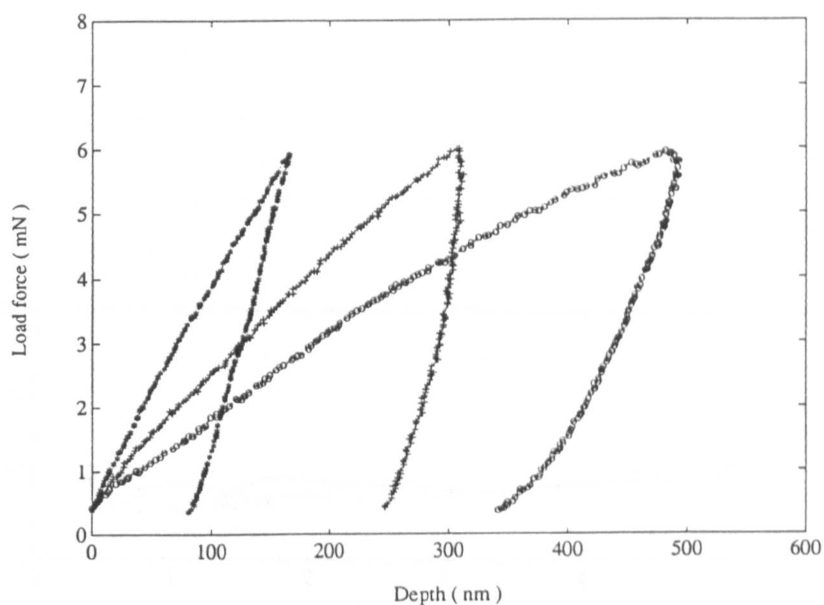


Fig. 5.4 Load *vs* depth for steel, copper and aluminium surfaces. * — steel, + — copper, and o — aluminium.

Table 5.1 Predicted and measured indent geometry for steel, copper and aluminium specimens.

Material	E (GPa)	Y (GPa)	P/Y	a (μm)	δ (nm)
Steel	210	0.79	2.9	0.9	80
Copper	126	0.4	2.0	1.54	250
Aluminium	70	0.15	2.25	2.4	340

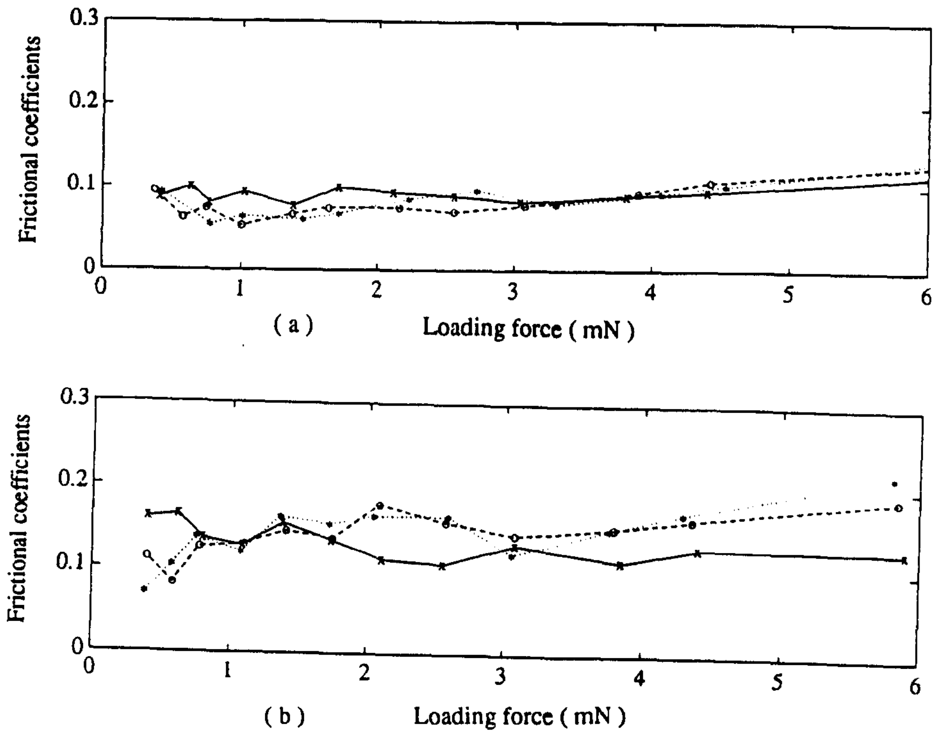


Fig. 5.5 Coefficients of friction *vs.* stylus load: — x, ---- o, ····· * represent values measured as sliding velocities of 1.0, 0.25 and 0.1 mm s^{-1} respectively; (a) mild steel, (b) copper, (c) brass, (d) aluminium, (e) silicon.

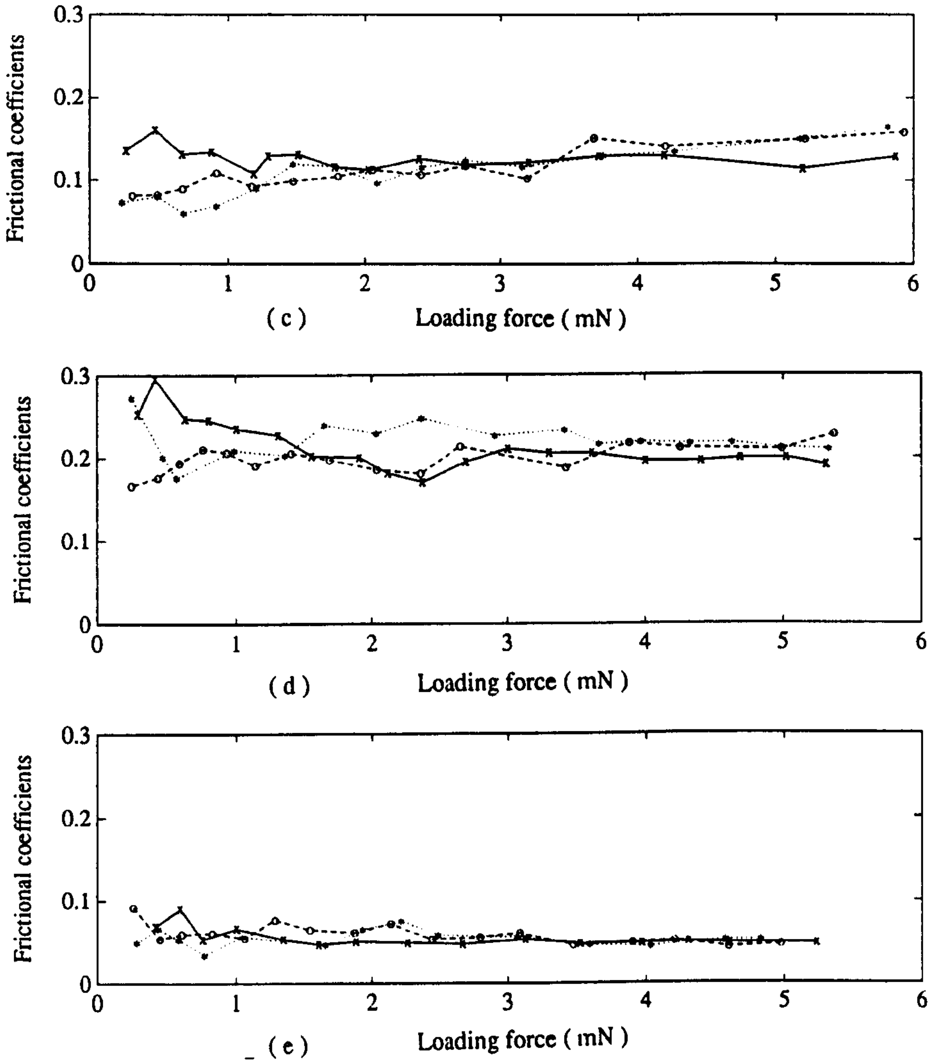


Fig. 5.5 (continued) (c) brass, (d) aluminium, (e) silicon.

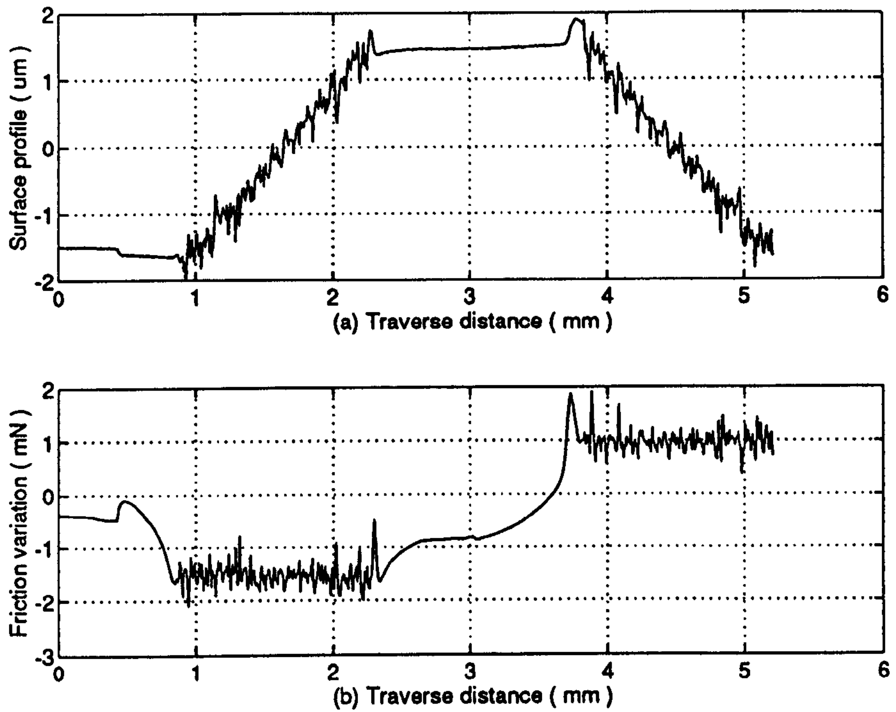


Fig. 5.6 Plots of surface profile and frictional forces as the stylus traverses a complete cycle on a copper specimen at a nominal load of 5.9 mN, (a) surface profile and (b) tractive force parallel to a traverse (frictional drag).

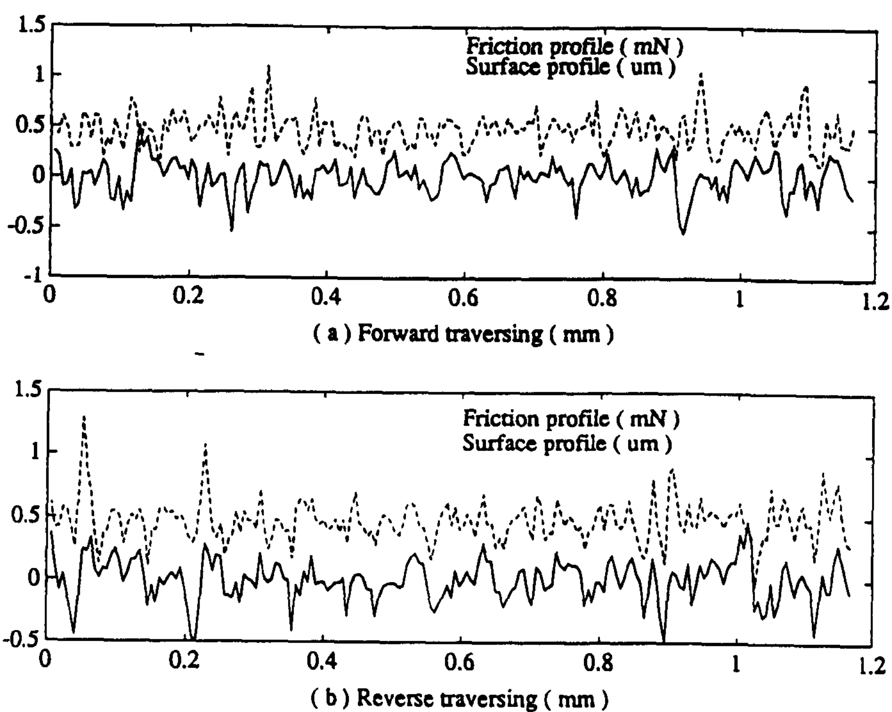


Fig. 5.7 Friction drag (dashed) and surface profiles (solid) from the plot of Fig. 5.6. Plotting units are chosen so that force and profile can be read from the same scale: (a) forward measurement, (b) reverse measurement.

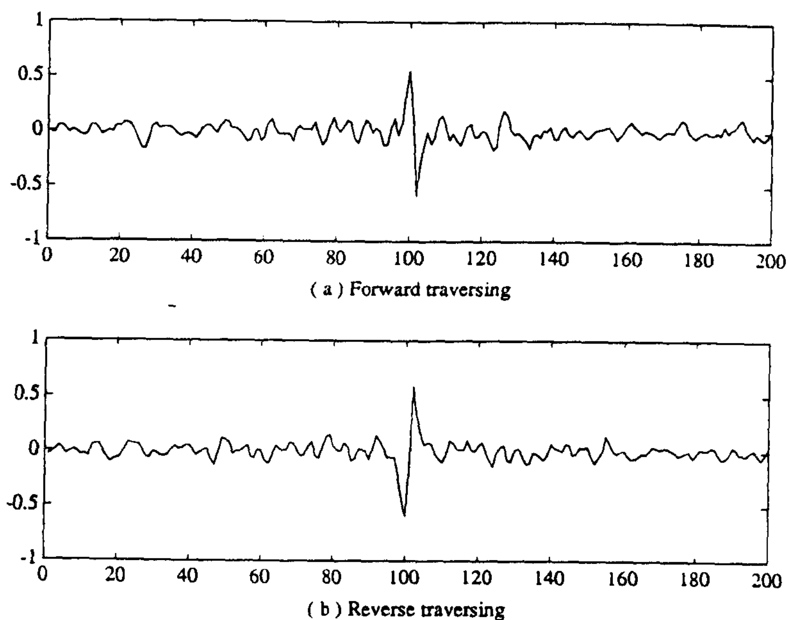


Fig. 5.8 Cross-correlations of surface profile and frictional drag data of Fig. 5.7 (mean line removed): (a) forward traverse, (b) reverse traverse. Horizontal axis in sample point spacing.

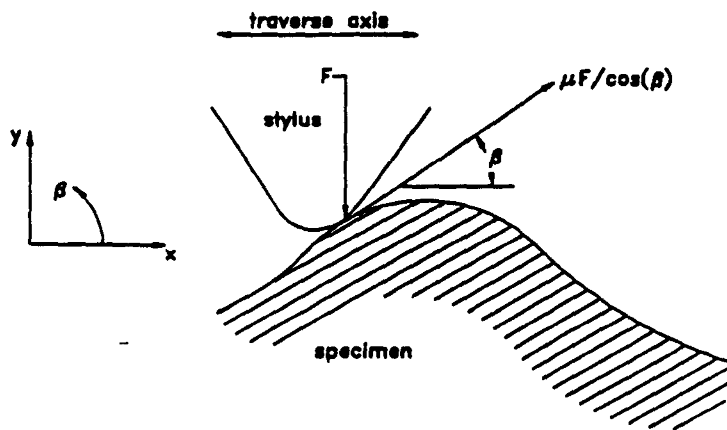


Fig. 5.9 Force diagram for the contact of a stylus on the surface of a specimen.

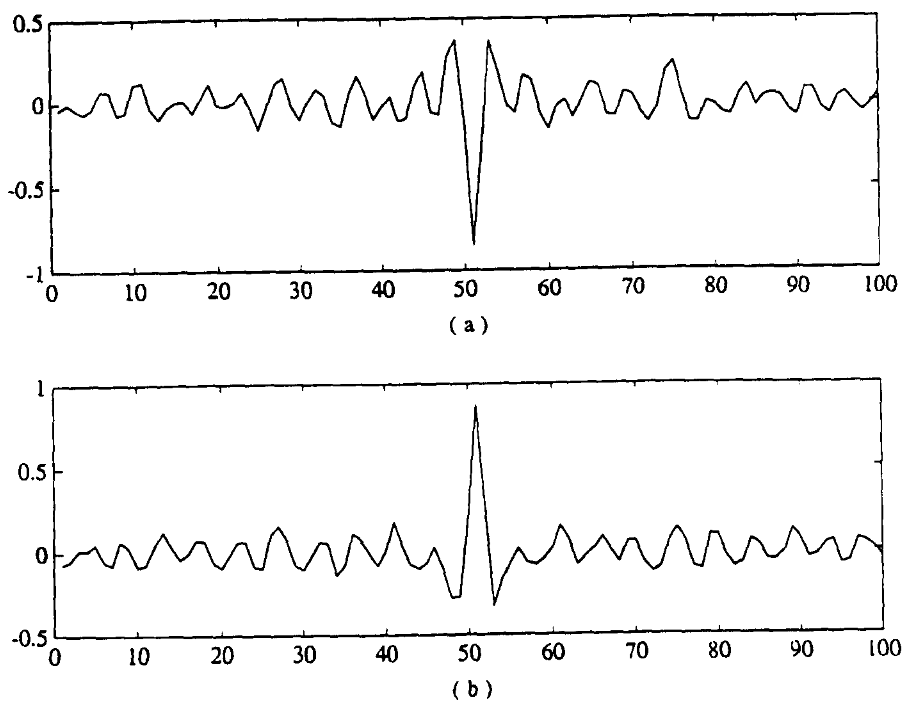


Fig. 5.10 Cross-correlations of first derivative with respect to traverse displacement of surface profile and frictional drag data of Fig. 5.7 (mean line removed): (a) forward traverse, (b) reverse traverse.

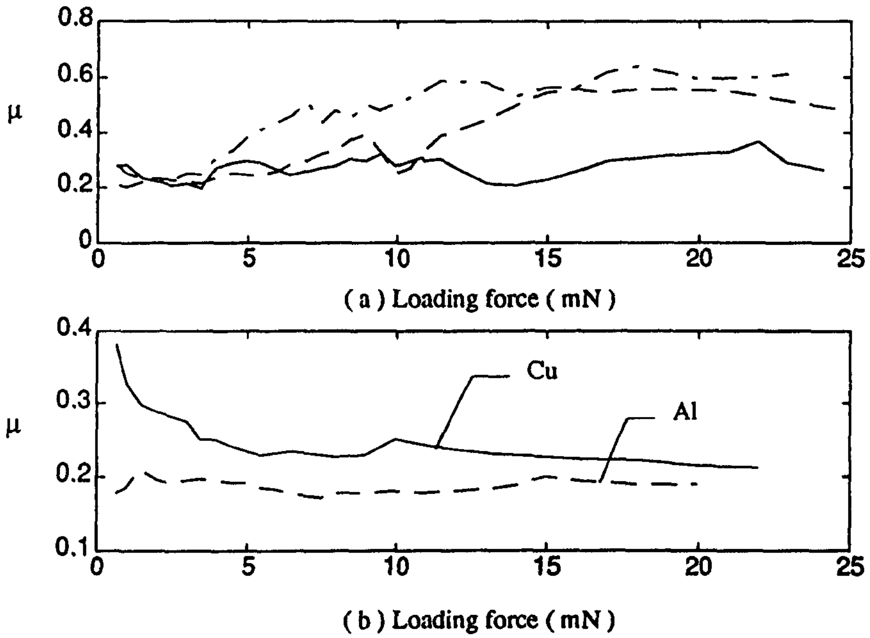


Fig. 5.11 Friction coefficients between the steel ball and specimens at an increasing load: (a) the steel ball on a steel specimen for three different tracks, (b) the steel ball on specimens of copper (solid line) and aluminium (dashed line).

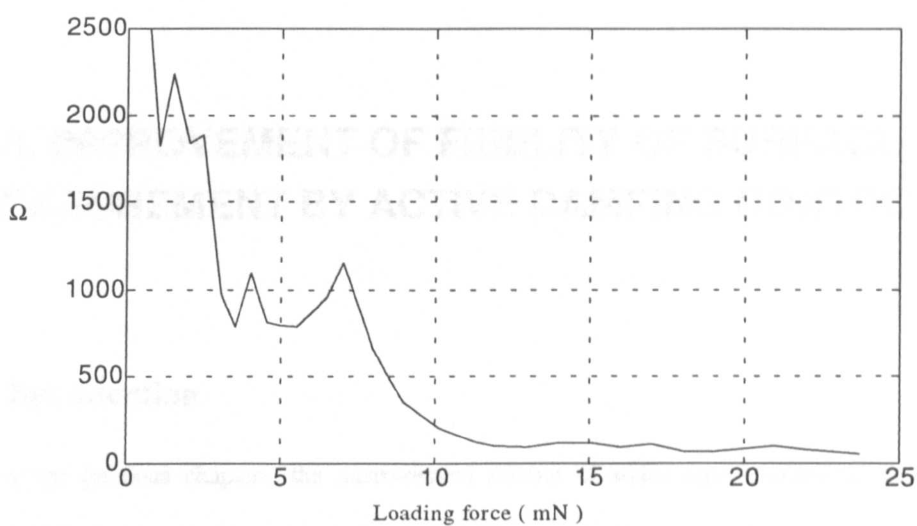


Fig. 5.12 The contact resistance between the steel ball and a steel specimen

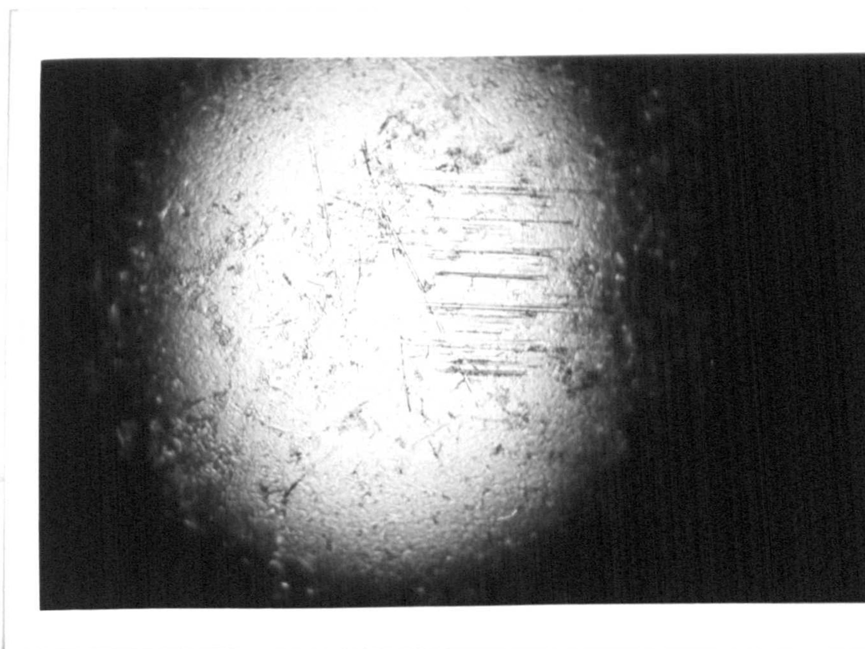


Fig. 5.13 A photograph of the damaged surface of the steel ball used in the experiment.

6. IMPROVEMENT OF FIDELITY OF SURFACE MEASUREMENT BY ACTIVE DAMPING CONTROL

6.1 Introduction

In the previous chapters, the measurement fidelity of stylus type instruments has been investigated both theoretically and experimentally. Measurements on a variety of specimens show that around the nominal load of 0.7 - 1.0 mN, the high stress does not lead either to unacceptable surface damage or a lack of integrity in following a surface profile. On some surfaces of soft materials the contact force, for a typical 2 μm truncated pyramidal tip, has to be kept below 1 mN, or better below 0.7 mN, to avoid any damage. However, experiments have shown that dynamic effects caused significant variations on measurements at different traverse speeds. The dynamics of a stylus mechanism may lead to the stylus bouncing over the measured surface at a light load when it is traversed at above a certain speed. Generally this speed is quite high, in the present context, but often is within the range offered on commercial instruments. As a consequence, the measured result can indicate false information on its roughness. For example, R_q may be reported by over 100% larger than it should be. The dynamic behaviour of a stylus system is dependent on its input frequency which depends on both the spatial wavelength of surface structure and the speed at which the probe is traversed. For many machined surfaces, the structures or irregularities vary over a wide range with the high-frequency components corresponding to sharp peaks or edges and narrow valleys. With sufficiently high forces this causes stylus jumping at the peaks and pressing hard at the valleys, both of which may lead to either a distorted profile or surface damage.

Theoretical work on stylus dynamics by Whitehouse (1988, 1990) based on study of reaction force at the contact suggests that the fidelity of the recorded profile might be enhanced by using a

damping factor of about 0.59 rather than the very low value used in the conventional instruments. By using this damping value a stylus pick-up can be operated at a frequency up to its resonance without introducing a significant distortion, thus the measurement speed can be increased. His definition of fidelity was statistically based and it is not clear how closely such ideas relate to the issue of faithfully recording specific and individual features on a profile. The basis of the argument was that the damping should be set such that the total static and dynamic load at the surface is as near constant as possible over all input frequencies up to those approaching the resonance of the stylus system. If this is possible it addresses strategies for obtaining higher speed data collection and for minimisation of damage with small styli while maintaining acceptable data collection rates. The work described in this chapter investigates the practicability of Whitehouse's proposal.

6.2 Dynamic Response of a Stylus system

6.2.1 System transfer function

As discussed in chapter 3, a stylus system can be modelled as a second order system in which the system transfer function can be expressed by

$$G(s) = \frac{y(s)}{f(s)} = \frac{K}{s^2 + 2\zeta\omega_n s + \omega_n^2} \qquad (6\text{-}1)$$

The dynamic response of the pick-up is characterised by the natural angular frequency, ω_n, and the damping ratio, ζ. The system gain, K, determines the sensitivity of the stylus pick-up and it has no effect on the system dynamic behaviour. The bode plots of the system (Talysurf 5 with the force actuator attached) are shown in Fig. 6.1, in which the damping ratio was 0.02 and the natural resonance 96 rad s^{-1}. The system response can be divided in three regions; low frequency up to about 0.1 ω_n, around the natural resonance and higher than the resonance. In the low frequency region,

the system gives a constant gain*, and this gain is actually the compliance of the stylus pick-up which means the deflection of the stylus is proportional to the contact force. As the frequency rises the gain becomes variable, becoming high around the resonance and then decreasing steadily at higher frequencies. This means that, for a given deflection of the stylus, less force is needed at a frequency around the resonance and larger force at a higher frequency. The low frequency region is the designed working region for stylus instruments. A good fidelity of measurement can be obtained when the stylus is operated in this region. However, in practice, it is difficult to know whether or not a stylus is operating well below its resonance during the measurement because of a wide range of wavelengths encountered on a specimen.

6.2.2 Spectra of a specimen traversed at different speeds

To demonstrate a variation in its spectrum with traverse speed a copper specimen was traversed by the stylus (Talysurf 5) at four different speeds, then the measured profiles were processed in the 'Matlab' to obtain their power spectra. The spectrum at each traverse speed was plotted and the results are shown in Fig. 6.2 (a-d). Each spectrum was fitted by the autoregression curve (AR model with an order of 20) to indicate the general trend. For this particular specimen, the spectrum ranges from 0.1 Hz to 20 Hz at a traverse 0.05 mm s^{-1}, giving a span range of 200. At a higher traverse speed the spectrum shifts towards high frequency parts but the span remains at about 200. At the fast speed 1 mm s^{-1}, the spectrum ranges from 2 Hz to 400 Hz. It shows that the shift corresponds to the traverse speed with the same proportion. It can be seen that the fine features or the features with high frequencies were only resolved at the slowest speed and disappeared at higher speeds due to the low system gain at higher frequencies. In addition, it will be shown later in this chapter that a high contact force is generated when the stylus is traversed at a high speed.

*Note that in this chapter 'gain' is used in its normal control sense and should not be confused with the instrument magnification.

6.2.3 Response to a given displacement

The difficulty in directly detecting any distortion in profiles for a random surface has been considered, here an alternative approach was used. The basic response of the pick-up in contact was tested by driving it by displacement from a digital piezoelectric translator (DPT, from Queensgate Instruments Ltd). This DPT operates in a closed servo-loop by a built-in position sensor (a capacitance gauge), characterised with low drift and low noise of less than 1 nm in a range of 10 µm. With a slew rate of 3 mm s^{-1}, the DPT is quite adequate for the tests. A piece of alumina was glued to the end face of the moving part of the DPT to reduce indentation by the stylus. The DPT was driven by a sinusoidal signal from a computer through the Metrabyte 12-bit digital to analogue converter. The frequency of the signal was set in turn to 0.5, 5, 10, 25 and 50 Hz, while the displacement was maintained at a peak value of 2.9 µm. The response of the stylus system to the DPT movement at each frequency is plotted in Fig. 6.3(a) which show them increasingly shifted to the right indicating a progressively increasing phase lag. The two traces at the lowest frequencies are overlapped. The graphs are plotted against cycles, giving equal spacing for ease of comparison. Subtracting the output of the stylus system from the DPT demand displacement gives the 'following' error at each frequency, Fig. 6.3(b). The two lowest frequencies are indistinguishable from the input, but by 10 Hz a noticeable phase lag has arisen. As the frequency rises further the phase lag increases and the measured amplitude reduces. The positive regions of the plot of the 'following' error indicate that the system assembly and contact are distorting while negative regions indicate loss of contact. There will be also a transition associated with surface distortion due to the nominal contact force.

6.2.4 Response to a variable force

The stylus system is characterised by its response to a force input rather than a displacement input at the contact. However, the real situation is not controllable. With a given displacement, as discussed above, the system response is obtained with no knowledge of the contact force which could be very high at certain frequencies. Here the system response to a variable force input over a wide range of frequency is explored. An ideal spring based stylus sensor (or a real one traversing very

slowly) will show a displacement response proportional to the contact force between the stylus and the surface. We take a profile (either real or computer-generated) and compute the force that should have occurred under such conditions. A signal representing this is used to modulate the force transducer drive about a static level that freely supports the stylus in its mid-range. The force signal was generated from a previous stylus measurement obtained from a copper specimen, and then output, via a digital to analogue converter, to the current amplifier. The data transfer rate was set equivalent to the tracking speeds of 0.02 and 0.2 mm s^{-1} on the original measured surface. The subsequent responses of the stylus system to such force signals were recorded and shown in Fig. 6.4. The stylus system followed closely to the original signal when it was driven by a force signal at an equivalent speed of 0.02 mm s^{-1}, but failed to do so at the higher speed of 0.2 mm s^{-1}. At the higher speed, the stylus responded the force signal in an exaggerated manner so that both the magnitude and lateral spacing were distorted, resulting in great loss of the original features. Obviously, this distortion is caused by the dynamic behaviour of the stylus system.

6.3 Statistical Approach to System Optimisation

For a contact, spring-based instrument the measurement or the deflection of a sensing beam is dependent of its spring rate, inertia, internal damping, and the compliance of the contact between the probe and the surface. Ideally, the measured displacement should be related only to the height changes of the surface at a constant contact force. Next best, the contact force should be determined only by the spring rate, there would be no dynamic reaction and thus the compliance is constant (ideally, zero). The dynamic effect can be minimised by operating the stylus at very slow speed but this may introduce other problems such as mechanical and electrical drifts which may degrade the measurement. Alternatively, as Whitehouse proposed (1988), the dynamic reaction forces at the stylus tip can be minimised by optimising the system parameter, the damping factor. The variation of the damping factor, ζ, has a great effect on system response. Making $\zeta \rightarrow 0$ introduces problems at the transducer end of the system if very sharp features are to be measured. Sensitivity to shocks

and long settling time are system features which should be avoided. Increasing ζ to a large value is obviously counterproductive because of the increase in reaction force and the sluggish response. What is required is an optimum value of ζ to improve fidelity and/or reduce damage. First, consider the power spectrum of the transfer function which has the following form

$$P_y(\omega) = |G(\omega)|^2 \, P_f(\omega) \tag{6-2}$$

where $P_y(\omega)$ and $P_f(\omega)$ are the power spectra of the output (displacement) and input (reaction force), and $|G(\omega)|$ is the magnitude of the transfer function of the stylus system. Rearranging the above equation in terms of P_y, we have

$$P_f(\omega) = \frac{1}{|G(\omega)|^2} P_y(\omega) = g(\omega) P_y(\omega) \tag{6-3}$$

$|G(\omega)|^2$ can be obtained from Eqn. (6-1) by substituting $s = j\omega$ and then multiplying with its conjugate. Thus

$$g(\omega) = \frac{1}{K^2}\left\{1 + \left(\frac{\omega}{\omega_n}\right)^2 (4\zeta^2 - 2) + \left(\frac{\omega}{\omega_n}\right)^4\right\} = \frac{1}{K^2} H(\omega) \tag{6-4}$$

where K is the system gain and $H(\omega)$ is named as weighting factor.

Ideally, the weighting factor $H(\omega)$ should be unity at all frequencies. Because random surfaces have a spread of frequencies contained in them attempts to minimise the effect of $H(\omega)$ must encompass a band of frequencies. Here the maximum bandwidth of the system is taken up to $\omega = \omega_n$. Then the damping ratio is picked so that the integrated weighting factor is unity over the frequency range. That is

$$\int_0^{\omega_n} H(\omega) d\omega = 1 \tag{6-5}$$

Whitehouse gives three values of damping ratios for three criteria; $\zeta = 0.59$ for equal area (area below the unity is equal to the area above), $\zeta = 0.57$ for the least squares, and $\zeta = 0.54$ for the mini.-max. criterion. All three criteria produce damping ratios within a very small range of values.

Hence if the criteria for fidelity and possible damage are to be used the damping ratio should be anywhere between 0.59 and 0.54. Within this range the statistical fidelity is assured.

6.4 Damping Functions

6.4.1 Viscous damping

Natural damping derives from both internal energy dissipation within the material and interface energy losses due to deformation and friction (Burton, 1958). Therefore, to increase damping is to dissipate more energy. Mechanically, the drag of a fluid on a solid always opposes the relative motion of the body, and its magnitude is a function of velocity of this motion. At low speeds, drag is directly proportional to the velocity and is called viscous drag. At higher speeds, it becomes more nearly proportional to the square of the velocity. For a spring-mass-dashpot system, damping is represented as a dashpot where fluid is drawn in or squeezed out through small slots as shown in Fig. 3.1(b). The resistance of such a device can be made proportional to the velocity of the piston in the dashpot. Without external forces the free vibration of this system can be found by

$$m\ddot{y} + b\dot{y} + \lambda y = 0 \tag{6-6a}$$

or

$$\ddot{y} + 2\zeta\omega_n\dot{y} + \omega_n^2 y = 0 \tag{6-6b}$$

where $\zeta = b/(2\sqrt{\lambda m})$, $\omega_n = \sqrt{\lambda/m}$, m, b, and λ are mass, damping, and spring rate respectively. A general solution to this differentiation equation, when damping $b < 1$, is given by

$$y = e^{-\alpha t}(A\cos\omega t + B\sin\omega t) \tag{6-7}$$

where the constants A and B can be evaluated from the initial conditions of the system. The characteristic of the viscous-damped vibration is that the vibration decays in an exponential way, and $\alpha = \zeta\,\omega_h$.

6.4.2 Dry friction damping

Dry friction damping appears when contacting parts shift relative to each other. In many mechanical systems, dry friction dampers are used because of their convenience and mechanical simplicity. Conventionally, the frictional force, f_T, is proportional to the normal force, f_N on the plane of contact, and is expresses as

$$f_T = \mu f_N \tag{6-8}$$

where μ is friction coefficient. For the above spring-mass-dashpot system, if the dashpot is replaced by a dry friction damper, the equation of motion will be

$$m\ddot{y} + \mu f_N + \lambda y = 0 \tag{6-9}$$

and the solution is given by

$$y = A\cos\omega t - \frac{\mu f_N}{\lambda} \tag{6-10}$$

If the initial amplitude is y_0, at the end of the half-cycle the amplitude will be $y_0 - 2\mu f_N / \lambda$. The reduction of amplitude per cycle is $4\mu f_N/\lambda$, and this gives a straight line envelope of the decay which is different from the exponential envelope for viscous damping. For dry friction damping there is no quantity equivalent to critical damping, but $\mu f_N = \lambda y_0 / 2$ represents the most rapid return to rest position. If the friction force is greater than this, the system will come to rest after disturbance at some position other than the static rest position of zero damping.

6.4.3 Hysteresis

Elastic materials (even metals, though the effect is much larger in rubber and elastomers) yield different curves on a stress-strain plot for loading and unloading. When such a substance is subjected to a cyclic load reversal, a hysteresis loop appears on the stress-strain diagram. The area of this loop is interpreted as energy dissipated per cycle. In the spring-mass system, hysteresis requires more work to be done in compressing the spring than is recovered upon expansion. The internal

hysteresis damping is, in general, treated as viscous damping due to the similar shape of their force-displacement loops.

6.5 Experimental Approach to System Optimisation

6.5.1 Construction of the active damping controller

In a typical stylus system, damping appears as viscous damping as it gives a closely exponential decay when it is in vibration. This is mainly due to spring hysterisis and air damping. A typical damping ratio for stylus instruments is 0.01 to 0.02. There are many ways to increase the viscous damping. However, mechanical viscous damping such as a dashpot will invariably add unwanted mass, and the damping factor will not be easily controlled. Instead, an active or electrical damper is sought. For example, when an aluminium plate moves between the poles of a magnet, the force acting on the plate is proportional to the velocity, and can therefore be called viscous damping. Alternatively, a velocity feedback method may be used to control the damping.

Theoretically, the velocity feedback signal can be achieved by differentiating the output of a stylus system. This approach was tested using a proportional and derivative (PD) controller having a frequency bandwidth matched to that of the Talysurf 5. A battery power supply was used to reduce the noise effects which are often met in derivative actions. The output from the PD controller was then fed to the magnet/coil force actuator. Using this, the damping factor of the system could be continuously varied over the range 0.022 to 0.2. However, phase shifts primarily associated with low pass filters in the instrumentation introduced considerable problems with high frequency signals.

Alternatively, a separate velocity sensor is used to avoid this difficulty. As shown in Fig. 6.5, a small magnet of neodymium boron iron, 3 mm in diameter and 2 mm long, is glued to one end of a glass tube whose the other end is attached to the stylus arm beneath the LVDI. Glass is used to avoid any coupling interference between the LVDI and the sensor and also because of its low density. The induction coil is placed around the magnet, mounted to the body of the stylus unit by

set screws to enable adjustment the coil position relative to the magnet. For best sensitivity in the design, a large number of turns is needed for the coil and the magnet should be located half exposed from the coil when the stylus is in balance. The induction coil has the same dimensions as the force coil, but its former is made from Tufnol instead of brass to avoid generating edge current which will decrease the sensitivity of the sensor. The winding uses 42 gauge copper wire and has a resistance of 27.6 Ω

6.5.2 Analysis of the velocity sensor

As the magnet is moved with speed $\dot{y}(t)$, an electromotive force (EMF) is induced in the coil according to the Faraday law

$$E = -N\frac{d\varphi}{dt} \tag{6-11}$$

where φ is the mean magnetic flux and N is the number of turns in the coil. The flux φ in the coil is a function of time through the displacement of the magnet; therefore

$$E = -N\frac{d\varphi}{dy}\frac{dy}{dt} \tag{6-12}$$

For small displacement of the magnet it can be shown that (Mavroyannakis, 1982)

$$\frac{d\varphi}{dy} = \text{constant} \tag{6-13}$$

Finally, we find that the EMF is directly proportional to the velocity of the stylus,

$$E = -CN\frac{dy}{dt} \tag{6-14}$$

where C is a constant depending on the characteristics of the coil and the magnet. Because the magnetic flux and the speed of the stylus are both low, a large number of turns N is needed to produce an appreciable EMF. (For precise calculation, further analysis is necessary because C is not truly constant and the EMF is not exactly a linear function of the speed).

The induced current i is then given by

$$i = \frac{E}{R_c + R_i} \tag{6-15}$$

where R_c and R_i are resistances of the coil and the current to voltage converter, respectively. R_i is small compared to R_c, and the induced current is mainly determined by the coil resistance. Therefore, to increase the sensitivity of this velocity sensor, the following points have to be considered.

1. The coil must have a large number of turns, the smallest possible resistance. The coil form has to be non-conductive.

2. The magnet must have a high remanence, be permanent and of suitable size. Also the displacement of the magnet must be small.

6.5.3 Circuitry of the damping controller

The induction coil of the velocity sensor was electrically connected to a low-impedance amplifier consisting of a low noise, low bias current FET amplifier OPA111 configured as a current-to-voltage converter. This combination has advantages of zero dc output and very low internal noise. The reverse effect of the force produced by the induced current in the sensor coil itself causes slight damping but this was found not to be significant. The voltage output of the sensor was then fed via a proportional amplifier into the transconductance amplifier to exert a feedback force on the stylus. The circuitry of this feed-back loop damping controller is shown in Fig. 6.6.

6.5.4 System response and damping ratio calibration

The combination of the force actuator and the velocity sensor with the Talysurf 5 provides this stylus system with a continuously variable force and damping control. With damping control the equation of the stylus motion becomes

$$\ddot{y}(t) + 2\zeta\omega_n \dot{y}(t) + \omega_n^2 y(t) + Kf_o = K(f_v(t) - f_d(t)) \tag{6-16}$$

here $f_d(t)$ is the feedback control signal given by $f_d(t) = k_1 dy/dt$, where k_1 is an adjustable gain coefficient. Rearranging this equation, we have

$$\ddot{y}(t) + 2\omega_n(\zeta + \frac{k_1}{2\sqrt{\lambda m_e}})\dot{y}(t) + \omega_n^2 y(t) + Kf_o = Kf_v(t) \qquad (6\text{-}17)$$

It is readily shown that the damping ratio is increased by an amount $k_1 / (2\sqrt{\lambda m_e})$. By varying the amount of feedback signal to the force actuator, different damping ratios can be obtained. This has been done by adjusting the coefficient, k_1, via a potentiometer on the damping control box.

The modified dynamic characteristic of this system was then measured by adjusting the static force set-point to bring the stylus to its balance, or midpoint, position using the force actuator with the velocity sensor in operation. The stylus was then slightly displaced and released manually by a piece of plate held on a micrometre stand. This would also be done by hand. The subsequent oscillatory response of the stylus was recorded. A typical response is shown in Fig. 6.7(a).

After recording the responses at different settings of the damping feedback coefficient, k_1, the damping ratios were determined by the log decrement method mentioned in chapter 3.3.2. The damping value was estimated from the successive logarithm amplitudes of the recorded signal. However the accuracy of the estimation depends heavily on the noise levels of both the signal and the data acquisition board, since it uses only few samples at peaks of a signal. Alternatively, a best fitting to the measured response using a standard simplex method is employed with the help of Weili Wang at Warwick University. A brief discussion of it is given below.

A general solution to a second order differential equation is expressed by

$$y(t) = \left[A\cos(\omega_n\sqrt{1-\zeta^2}.t) + B\sin(\omega_n\sqrt{1-\zeta^2}.t)\right]e^{(-\zeta\omega_n t)} \qquad (6\text{-}18)$$

or

$$y(t) = [\theta_3\cos(\theta_2 t) + \theta_4\sin(\theta_2 t)]e^{-\theta_1 t} \qquad (6\text{-}19)$$

where

$$\omega_n = \sqrt{\theta_1^2 + \theta_2^2}$$
$$\zeta = \frac{\theta_1}{\sqrt{\theta_1^2 + \theta_2^2}} \qquad (6\text{-}20)$$

and θ_i, i =1, 2,··· 4 are the variables to be determined. Let $\{Y_1, Y_2, \cdots Y_N\}$ be the samples of the recorded data, and N the recording length. Then an objective function is formed as

$$J(\theta) = \sum_{i=1}^{N}[Y_i - y(t_i)]^2$$
$$= \sum_{i=1}^{N}\left\{Y_i - [\theta_3 \cos(\theta_2 t) + \theta_4 \sin(\theta_2 t)]e^{-\theta_1 t}\right\}^2 \quad (6\text{-}21)$$

Now the problem is how to adjust the variables θ_i, (i = 1, 2, ··· 4) to make the function, $J(\theta)$ reach its minimum. Here a direct searching algorithm, the simplex method of Nelder and Mead (1965) is used. The method does not require any partial derivatives of the objective function, it relies only on the values of the objective function and the information from earlier iterations. The fitting program was written by Weili Wang and performed in Matlab. By setting an error limit, we can get a very accurate estimation. In Fig. 6.7(b) the solid curve is from the original data and the dashed line is calculated from Eqn. (6-18) using the parameter estimation. The standard deviation of the fitting is better than 0.001. The original pickup resonated at 178 rad s^{-1} with a damping factor of 0.012. After adding only the force actuator magnet, the values changed to 96 rad s^{-1} and 0.021. When the velocity magnet was added they further changed to 86 rad s^{-1} and 0.022. The effective mass of the stylus system was 1.5 g with both magnets attached. Damping ratios were then calibrated by repeating the above procedure at each damping feedback coefficient.

These characteristics were determined without contact between probe and specimen. The figures are probably little different under normal operating conditions. Since the model assumes that the surface forces the stylus, the stiffness of the contact is absorbed into the reaction force term. Local friction at the stylus as it slides over the surface is in direction perpendicular to the displacement and so is not expected to have a great effect on the damping term used in the simple model at light contact loads. But it is unclear whether it may be totally neglected.

6.6 Performance Evaluation of the System

6.6.1 Effects of damping on dynamic forces

There are no standardised or calibrated surfaces suitable for testing damping effects. Even if a suitable specimen were available, these dynamic effects would be difficult to identify in the presence of features such as the contact stiffness. Thus a simple alternative approach has been introduced here that exploits the presence of a force transducer. As discussed in 6.2.4, a simulated force signal is applied to the force actuator to represent an ideal profile. If the output of the displacement sensor corresponds closely to the original profile, the sensor system must be behaving close to the ideal. Differences between the original and recorded profile signals indicate modification caused by dynamic effects in the pick-up. Obviously, the larger the differences are, the poorer the fidelity. This time, the data transfer rate of the force signal was set equivalent to the tracking speeds of 0.05 and 0.1 mm s^{-1} on the original measured surface, which generated significant amplitudes at input frequencies up to 12.5 and 25 Hz, respectively. The subsequent responses of the stylus system to such force signals were examined for damping ratios ranging from 0.022 to 1.8. Fig. 6.8 shows an example of the original profile and the outputs of the stylus at a simulated speed of 0.05 mm s^{-1} with the damping ratios of 0.6, greater than 1 and 0.022, respectively. With damping of 0.022, the stylus overshot considerably at sharp peaks and valleys, giving a much distorted profile. Increasing damping to 0.6, the stylus followed faithfully the input, while a further increase caused a sluggish response with some high frequency features clearly filtered out. The response was also evaluated by means of statistical parameters: the magnitude parameters, R_a, and R_q, rms slope, Δ_q, and rms peak curvature, P_c. The variations of the parameters over the range of damping ratios are shown in Fig. 6.9 where values are normalised to those at the damping ratio of 0.6. Values varied wildly with damping ratios below 0.4 for both the equivalent speeds. More consistent behaviour occurred with damping ratios from 0.4 or 0.5 to 0.8. Above 0.8, the variations tended to become a little larger. The standard deviations of following errors, calculated by subtracting the output of the stylus from the original profile, are plotted in Fig.6.10,

again normalised to the values at damping ratio 0.6. At the slower speed, the minimum following error occurs with damping between 0.5 and 0.7, with error increasing away from this region because of overshoots or lost details. At the higher input speed the minimum is not seen, presumably because in the presence of large amplitudes at high frequency the system behaviour tends to be dominated by a failure to follow them. There might be other instrumentation-based objections to using an over-damped system, so taken together these curves suggest that a damping ratio in the region 0.5 to 0.8 best suits this measure of fidelity, a result which vindicates Whitehouse's arguments. To further check behaviour with higher frequency drives, limited tests were also performed at data rates corresponding to equivalent traverse speeds of 0.25 and 1.0 mm s^{-1} where there will be significant excitation at frequency above the resonance. Results are shown in figure 6.11. There is always a considerable loss of detail and the situation rapidly arises in which it is difficult to see any sensible correlation of the output with the drive. As an illustration of second order system behaviour this is unremarkable. However the instrument operated in its normal configuration is capable of producing a good fidelity profile if it is traversed over the specimen used to generate the drive signal at 1 mm s^{-1}. This can only arise because the system contact has a stiffness much higher than that implied by the present drive conditions. High forces must be generated at the contact to force it to follow and the example illustrates how far from the ideal the contact condition may become. The conditions are less severe at the higher damping factor.

Note that we take the closeness of the force profile to the original measurement profile as a direct indication of the dynamic force variation, that is identical profiles imply zero dynamic interaction.

6.6.2 Effects of damping on signal fidelity

Specimens of mild steel and copper similar to those used in previous research were chosen for this set of experiments. To test the effect of damping on measured profiles, repeated measurements were taken, nominally over the same track with traverse speeds of 0.05, 0.1, 0.25 and 1 mm s^{-1}. The position trigger (described in chapter 4) was used again. For each experiment, the

specimen was cleaned by swabbing with acetone and a couple of traces were made at low speed before collecting data to avoid any ambiguity caused by contaminants or debris from the polishing. The specimens were prepared by polishing using a 1200 grit abrasive. From stylus profiles and SEM (Scanning Electronic Microscope) examination, this process generated structures in which the minimum detected lateral features were about 2 μm for mild steel and 4 μm for copper, giving equivalent frequencies of 25 and 12.5 Hz when traversed at 0.1 mm s^{-1}.

All of the measurements, whether evaluated using statistical parameters or by visual inspection of profile details indicate that increased damping reduces the scatter between nominally repeated measurements. For example, with a mild steel specimen varying the damping over the range of 0.02 to 0.76 caused no significant change on parameter evaluation of the surface measurements when the stylus was traversing at a slow speed such as 0.05 mm s^{-1}, since the frequency of the equivalent input to the stylus system is well below the resonance and damping has little effect on the measurements. At the fast speed of 1 mm s^{-1}, visible smoothing occurred as the damping increased and the evaluation showed a slight decrease on magnitude parameters R_a and R_q, and *rms slope*, Δ_q, but an increase on the rms peak curvature, P_c, as shown in Fig. 6.12. This is consistent with the view that with a lightly damped stylus, traversing at a fast speed, the measurement will normally overestimate R_a, R_q and Δ_q and underestimate P_c due primarily to stylus bounce at lighter loads. The most consistent results are obtained with damping ratios in the region of 0.5 -0.7.

Repeat profile measurements were taken on nominally the same tracks on the specimens of copper and mild steel, at damping ratios of 0.022 and 0.59 with traverses at increase speed. Results are shown in Fig. 6.13 and 6.14. Figure 6.13 shows repeat traces on a copper specimen with plots (a-h) arranged in order of increasing traverse speed. Sampling rates have been chosen such that profile points are spaced at 0.5 μm/point. Each plot has four to five curves superposed. The profiles showed very consistent features of the specimen for a range of speeds. Some lateral distortions may be due to slight speed fluctuations in the traverse drive. At slow speeds of 0.05 and 0.1 mm s^{-1}, the dynamic force is relatively small and no significant difference can be seen between profiles at the two damping values. The profiles with damping ratio of 0.59 are perhaps more consistent in both amplitude and

wavelength than those using a value of 0.022, suggesting that force variations on both horizontal and vertical directions are smaller with higher damping. The horizontal force here is frictional force, which is dependent on the vertical force, and may have an effect on the stability of axial stiffness of the pick-up pivot. Four features indicated by ' x ' in figure 6.13(a) are seen to be lost in figure 6.13(c) during traversing at the speed of 0.1 mm s^{-1}. Although it is unusual to see damage of this scale caused by stylus at these loads and speeds, the consistency of later traces suggests that some fracture occurred. Since no further such changes were observed it is thought that the cause was weakly adhered particles not removed by the specimen preparation process. Imperfect consistency between traces is a little more noticeable, and slight overshoots up to 0.1 µm at sharp peaks and valleys occur at 0.25 mm s^{-1}. Large differences are seen at the fastest speed of 1 mm s^{-1}, Fig. 6.13(g, h). Lift-off behaviour can be seen with low damping. With the damping of 0.59, profiles are much more consistent and more like those obtained at low traverse speeds. For the mild steel specimen, profiles were obtained successively at tracking speeds of 0.25, 1, and 0.1 mm s^{-1}. The average lateral peak-valley spacing is between 5-10 µm. When such a surface is traversed at 1 mm s^{-1}, the equivalent frequency of reaction force at the surface for average features is around 100-200 Hz. Even with a speed of 0.1 mm s^{-1}, those peaks are still scanned at a frequency near the resonance of the stylus system. In these situations, the stylus will not follow the surface faithfully because of dynamic forces and significant differences are seen on the profiles, Fig. 6.14 (a-d). Again better results are obtained with the damping factor of 0.59. The steepest slope shown by ' + ' in Fig. 6.14(e) caused the stylus to jump when it was traversed at fast speeds, which giving a distorted profile. Since the slowest, and most consistent, traverse was this time done last, the smaller features it shows must have been present at the time of the fast traverse. This is evidence of poor fidelity recording without surface modification at higher speeds on materials of moderate hardness.

To quantify the profile information a statistical parameter evaluation was undertaken. Parameters R_a, R_q, rms slope, Δ_q, and rms peak curvature, P_c, have been averaged for eight profiles in each set. These average values and their relative standard deviations (evaluated as $\pm$ σ%) are compared for two damping ratios in Table 6.1. It can be seen that there are no significant differences

among the parameters when the stylus is traversing at speeds below 0.1 mm s^{-1}. At a traverse speed of 0.1 mm s^{-1}, where the typical feature frequency is close to the stylus resonance, parameter variations with the damping ratio 0.59 are reduced by 4 to 10 times on R_a, R_q and Δ_q but there is only a slight reduction on the peak curvature. When the tracking speed is further increased to 1 mm s^{-1}, the stylus is sometimes forced to vibrate at a frequency beyond resonance, the scatter generally increases and damping has little effect.

To give a qualitative view of dynamic effects on the stylus measurements, the same track of the copper specimen was traced with three different loads of 0.46, 0.84 and 1.55 mN at the fastest tracking speed of 1 mm s^{-1}. The profiles are shown in Fig. 6.15. The results clearly show the stylus lift-off at sharp peaks under the loads up to 0.84 mN. The profile at the large load of 1.55 mN was compared with the profile obtained at the slowest tracking speed, which showed nearly the same features, but slight lift-off still exists at very sharp peaks on the faster trace, indicating the magnitude of dynamic forces. The calculated parameters are given in Table 6.2, which are listed in the order in which they were measured for the consideration of the effect of a large loading force. The effect of damping is clearly seen at lighter load. With a damping ratio of 0.022 at a load 1.55 mN a reduction on the parameters was caused by the large loading force.

6.6.3 Effects of damping on frictional force

Friction between a stylus and a specimen at low load has been investigated in a previous chapter. It was demonstrated that a simple model in which the instantaneous frictional force (apparent drag) depends on the local slope of the surface could be used with reasonable accuracy. The overall vertical and horizontal forces were expressed as

$$f_v(t) = f_n(t)\,(\cos(\beta) - \mu\sin(\beta)) \qquad (6\text{-}22)$$

$$f_h(t) = f_n(t)\,(\sin(\beta) + \mu\cos(\beta)) \qquad (6\text{-}23)$$

where β is a local slope of the surface, $f_n(t)$ is the force normal to the surface and μ the friction coefficient. With active damping control, the vertical force will be

$$F_v(t) = f_v(t) - k_1 \dot{y}(t) = f_v(t) - k_1 V_h \tan(\beta) \tag{6-24}$$

Then the frictional (horizontal drag) force will be

$$F_h(t) = \frac{(f_v(t) - k_1 V_h \tan(\beta))\ (\mu + \tan(\beta))}{(1 - \mu \tan(\beta))} \tag{6-25}$$

This can be rearranged in terms of the friction angle α , $\tan(\alpha) = \mu$, to give

$$F_h(t) = (f_v(t) - k_1 V_h \tan(\beta))\ \tan(\beta + \alpha) \tag{6-26}$$

The measured horizontal force depends on the vertical force, surface slope and the friction angle. Because damping reduces variations of the vertical force there will be a corresponding reduction in variation of the horizontal force and thus in friction force.

The frictional drag force reflects indirectly the variation of the vertical force. To measure directly this vertical force, a precision load cell, from Maywood Instruments, type 4934, was used. Its resonant frequency was about 200 Hz, much higher than that of stylus. The stylus was lowered onto the load cell button (which had a surface profile caused by face turning) and then traversed over it with the outputs of both the stylus and the load cell being recorded. Because of its high resonance, the dynamic effect of the load cell itself is negligible and its output will represent the force generated between the stylus and the load cell surface. The stylus tracking speed was kept below 0.25 mm s^{-1} to avoid excitation of the load cell. The results are shown in Fig. 6.16, in which, to aid comparison, the plots are deliberately shifted vertically. As the damping increased from 0.022 to 0.59, the standard deviations of the measured profile and the force reduced from 2.477 μm and 0.079 mN to 2.417 μm and 0.060 mN for speed of 0.1 mm s^{-1}, and from 2.419 μm and 0.232 mN to 2.358 μm and 0.134 mN for speed of 0.25 mm s^{-1}, respectively. Traversing at the speed of 0.25 mm s^{-1}, the interaction force between the stylus and the surface of the load cell is about three times

higher than that at speed of 0.1 mm s^{-1}. This confirms that, as expected, high frequency features on the profile impose a larger force at the contact.

Using a load cell to detect the vertical force provides a relative quantitative evaluation although there is coupling from the horizontal frictional force which is of unknown magnitude.

Using the same device as reported in chapter 5, friction measurements were taken on the copper specimen at two damping factors with the stylus tracking at speeds of 0.1 and 0.25 mm s^{-1} under a nominal load of 0.8 mN. As damping only affects dynamic force, the static friction coefficient is unaltered. The frictional force varied less with higher damping ratio. The magnitude of the reduction is around a few tens micronewtons on average. Details of the profiles and the variation of frictional force at the speed of 0.1 mm s^{-1} are plotted in Fig. 6.17. It can be seen that large variations of frictional force at sharp peaks and narrow valleys have been reduced with the higher damping value. The relationship between the profile and the force variation is not simple. Near dominant peaks and valleys of the profile the signals are roughly 180° out of phase, indicating that acceleration effects are dominant. Away from these features there is a strong correlation between profile and the force as previously reported.

6.7 Increasing Resonant Frequency

By combining the active feedback of both velocity and position to the force actuator it is possible to increase the resonant frequency of a stylus system. With this state feedback control the governing system equation becomes

$$\begin{aligned} &\ddot{y}(t) + 2\zeta\omega_n\dot{y}(t) + \omega_n^2 y(t) + Kf_o = K(f_v(t) - f_d(t) - f_p(t)) \\ &f_d(t) = k_1\dot{y}(t) \\ &f_p(t) = k_2 y(t) \end{aligned} \qquad (6\text{-}27)$$

where control variables f_d and f_p are damping and position force, and k_i are control constants. Rearranging equation (6-27) gives

$$\ddot{y}(t) + 2\omega_n(\zeta + k_1 / 2\omega_n)\dot{y}(t) + \omega_n^2(1 + k_2 / \lambda)y(t) + Kf_o = Kf_v(t) \qquad (6\text{-}28)$$

State feedback control increases the resonant frequency for all positive values of k_2. This is accomplished by adding another channel to the current drive of the force actuator and feeding back a fraction of the displacement signal, $k_2y(t)$. By adjusting the coefficient k_2 via a potentiometer, the resonant frequency of the system was increased from 13.7 to 44 Hz. However, such position control also increases the effective spring constant of the system which will consequently introduce large force variations at the contact. This is a disadvantage from the point view of surface damage during a measurement, however this approach may be useful for applications in which speed of measurement is of primary importance and surfaces are reasonably hard. It also, in practice, allows the simultaneous tuning of spring constant and damping ratio to optimise the fidelity at a given traverse speed with minimal risk of surface damage.

6.8 Conclusion

A relatively simple system has been designed and constructed that allows the damping of a stylus profilometer to be varied via force feedback. Theoretical and practical evaluations confirm that there is some merit in adjusting damping factors in the region between 0.5 and 0.7, rather than the value well below 0.1 present in most stylus instruments, to obtain improved fidelity at high traverse speeds. Evidence has been obtained to show that dynamic loading effects increase the variability in both vertical and horizontal axes of the profile and that higher damping reduces these effects.

In practice, the optimal damping value between 0.5 and 0.7 should be considered in designing of stylus instruments and it may also be applicable to other force following systems such as atomic force microscopes and force transducers in which displacements are used to monitor forces.

The active force feedback appears to be a very promising method of improving the behaviour of future system of lightly loaded contact gauges.

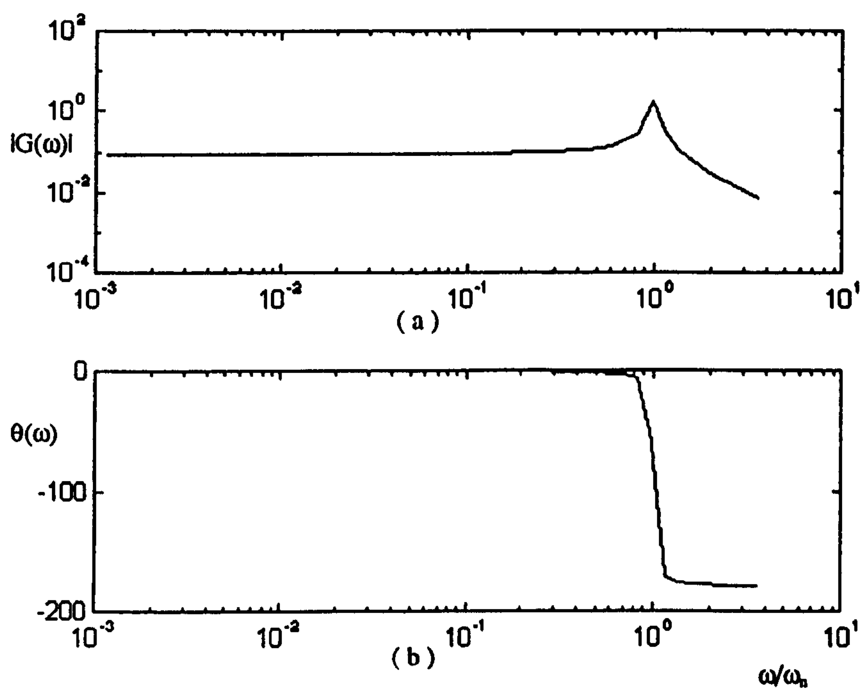


Fig. 6.1 Bode plots of the transfer function of a stylus system, (a) system gain and (b) phase shift against the normalised frequency.

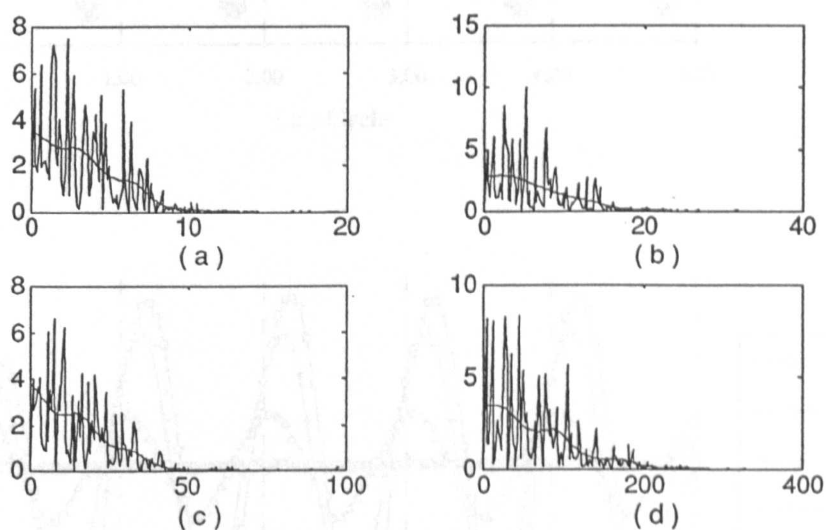


Fig. 6.2 Spectra of a copper specimen traversed at speed of, from (a) to (d), 0.05 mm s^{-1} to 1 mm s^{-1} corresponding the horizontal magnification, V_h, x100 to x5. The x-axis is in Hertz and y-axis in μm^2.

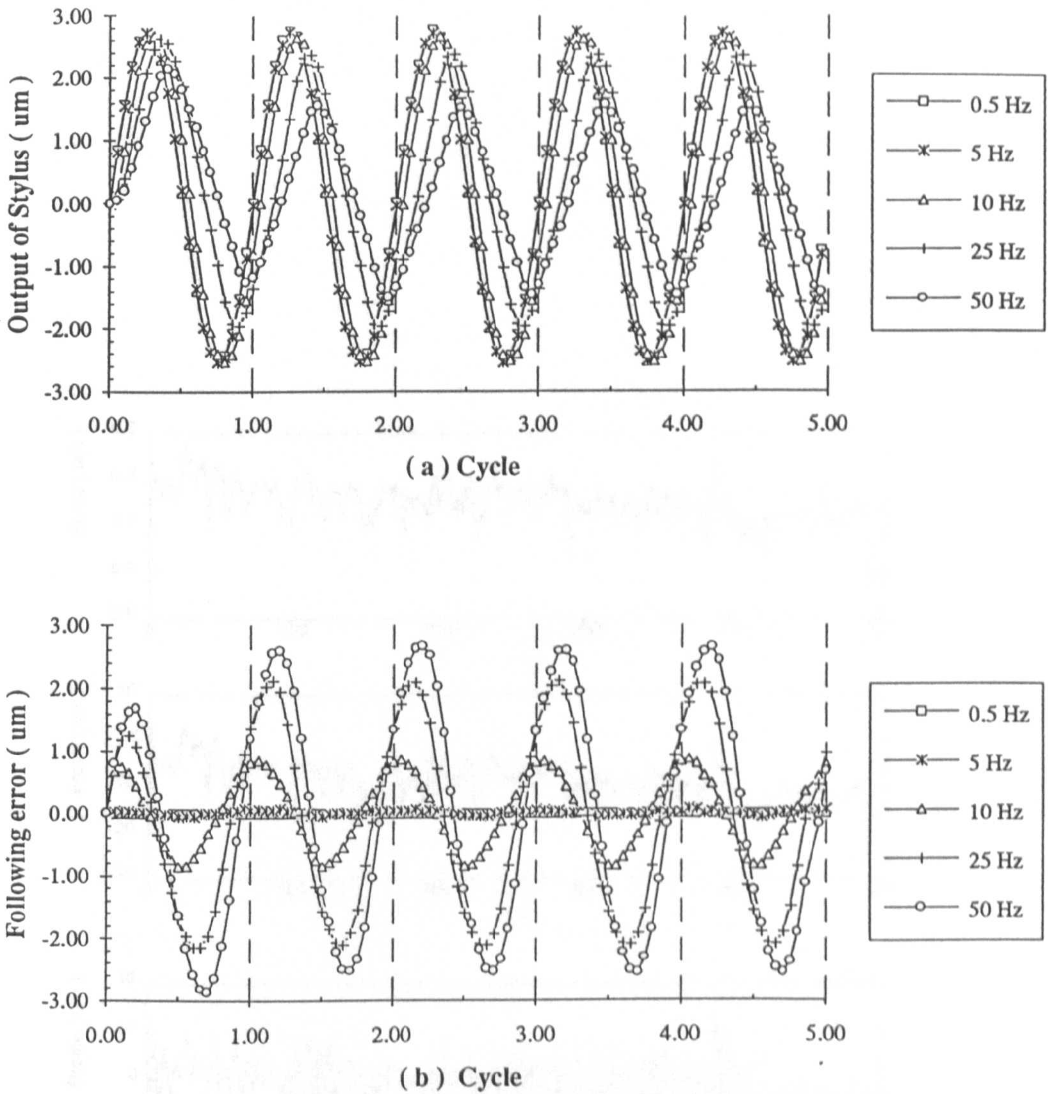


Fig. 6.3 The response of the stylus to a sinusoidal position input at different frequencies: (a) stylus responses to input signals at frequencies of 0.5, 5, 10, 25, and 50 Hz; (b) following errors after subtracting the stylus output from the DPT displacement.

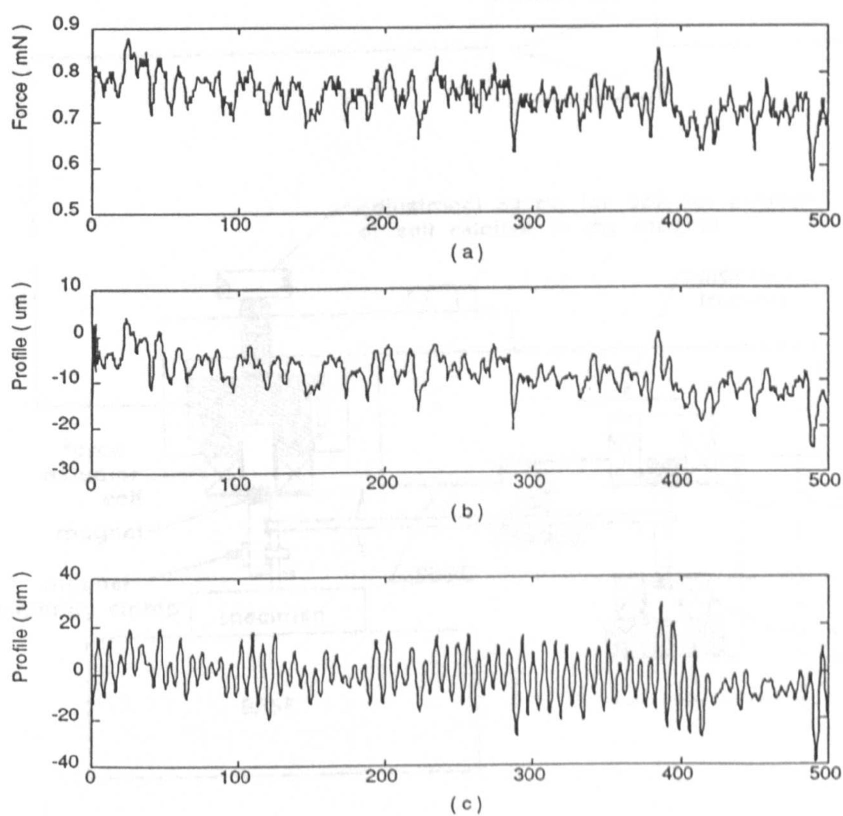


Fig. 6.4 Responses of the stylus at simulated force signals of different speeds: (a) input force signal, (b) at a speed of 0.02 mm s^{-1}, and (c) at a speed of 0.2 mm s^{-1}.

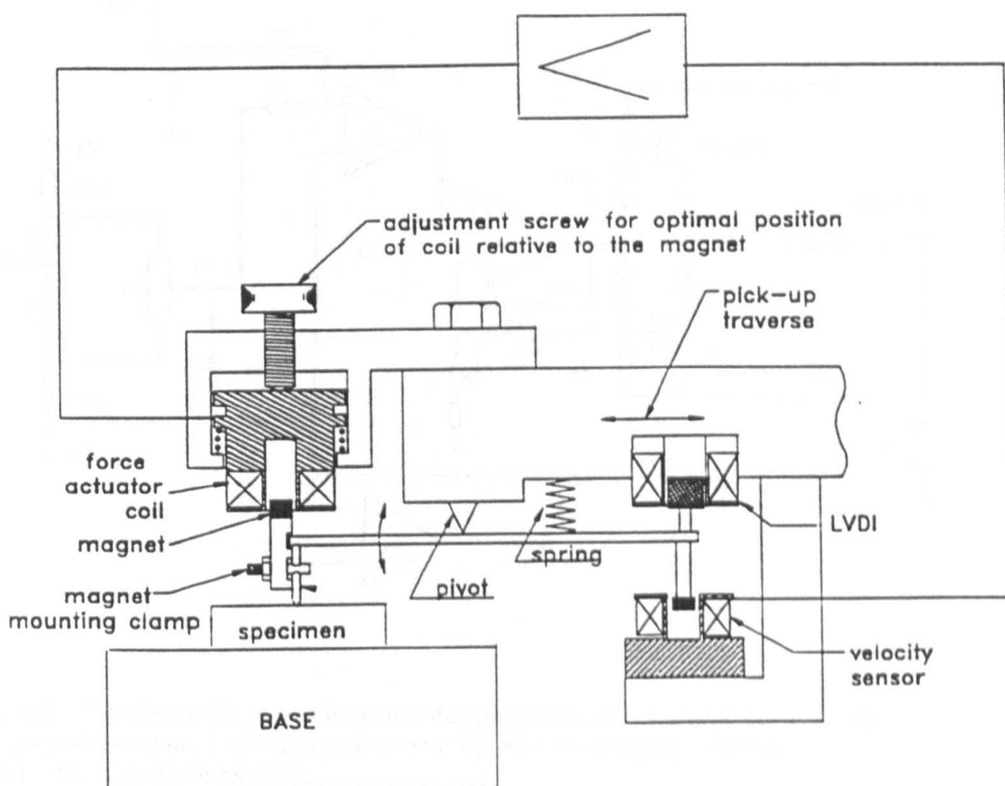


Fig. 6.5 Schematic of the modified Talysurf 5 with force and damping control.

Fig. 6.6 The circuit diagram of the active damping controller. Variable controls adjust: P - proportional gain; I - integral action (with C_i); K_1 - the damping coefficient; and R - the static force set-point.

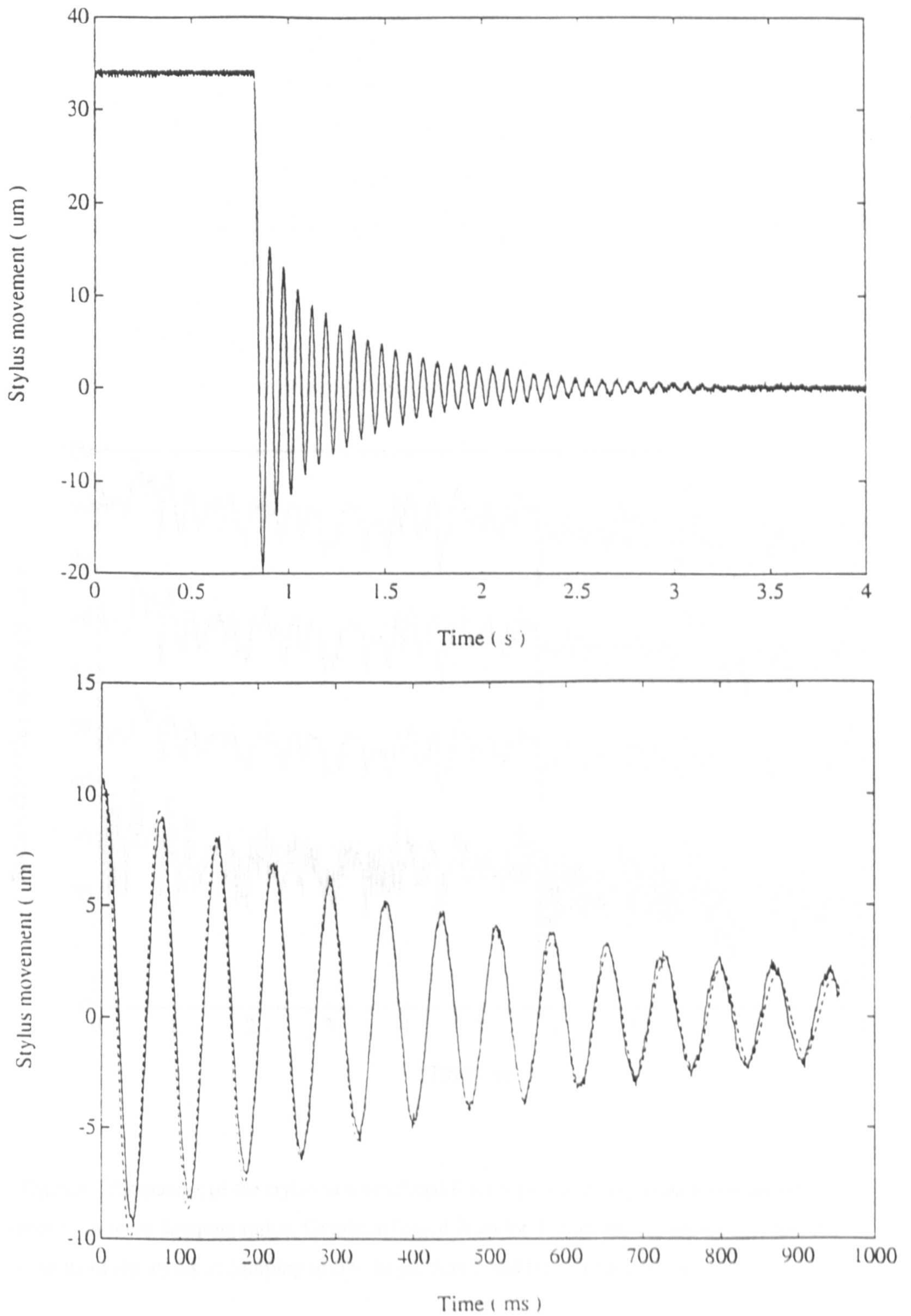


Fig. 6.7 The response of the stylus to release from an initial position: (a) Oscillation of the stylus; (b) fitted result by the simplex searching algorithm. Solid line is the original and the dashed line is the fitted result.

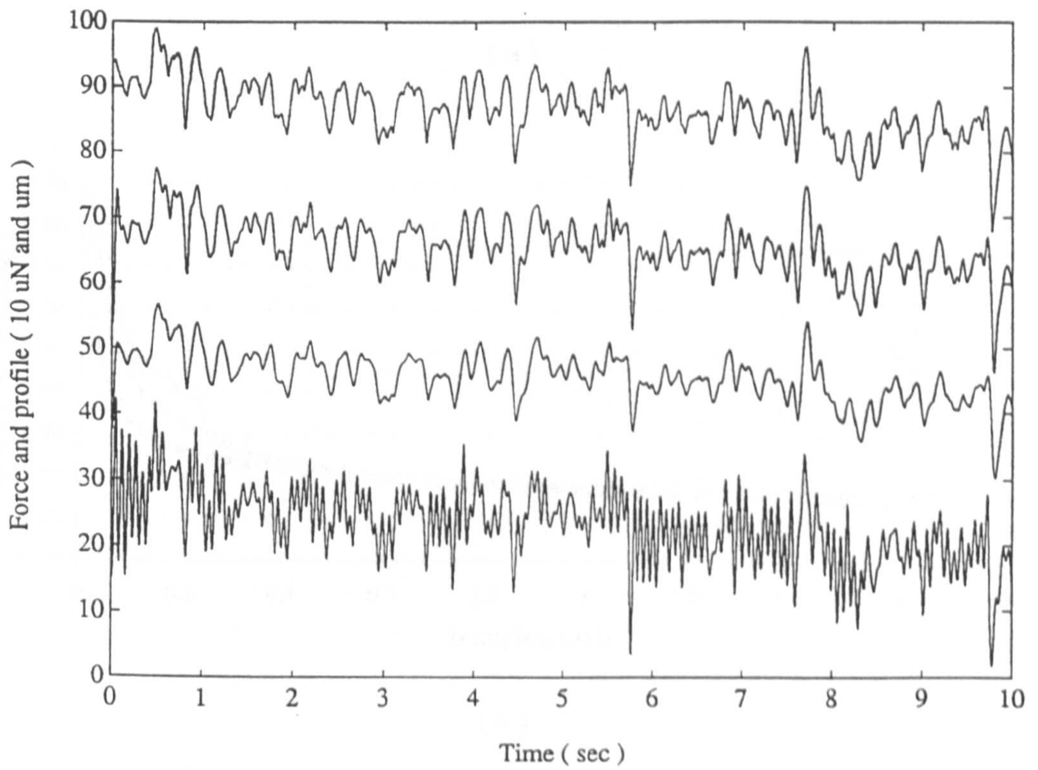
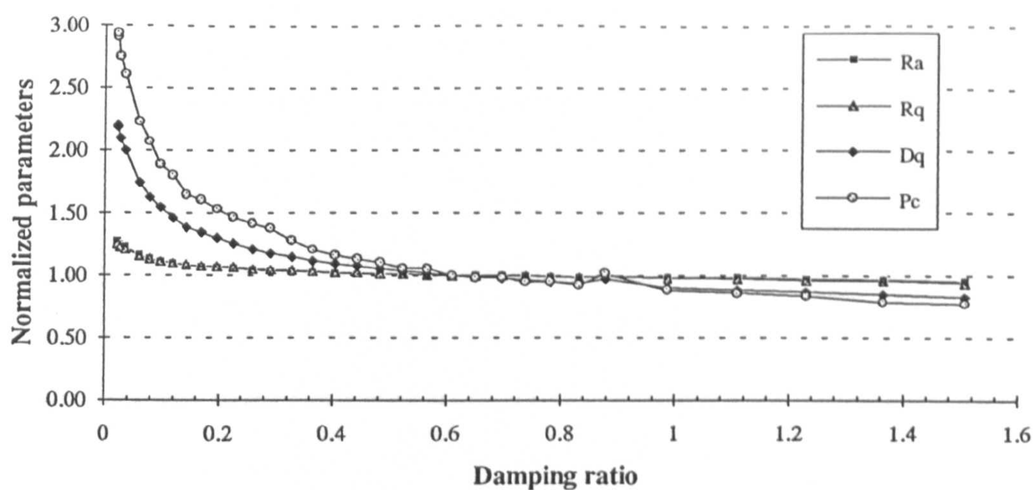
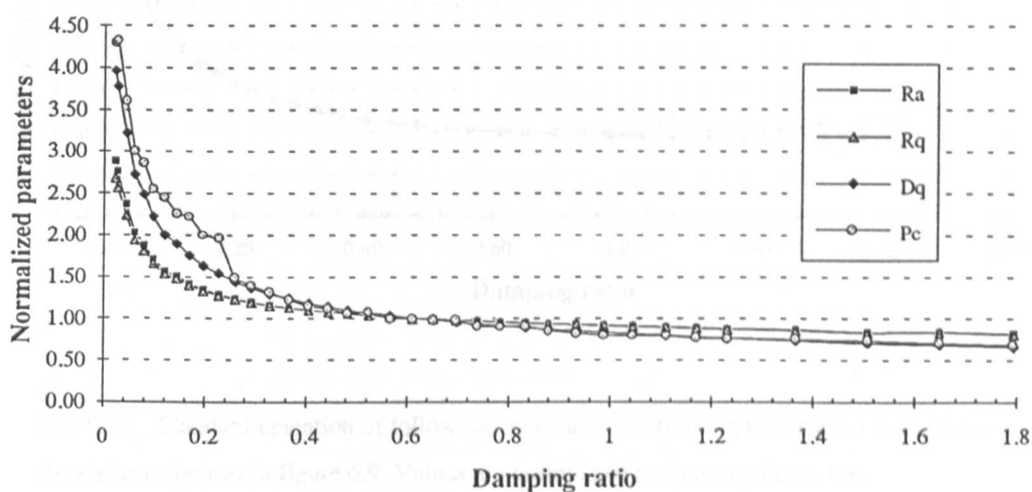


Fig. 6.8 Responses of the stylus to a simulated force signal of an equivalent speed of 0.05 mm s^{-1} at three damping ratios. Graphs arranged from top to bottom: the input force and the outputs of the stylus at damping of 0.6, larger than 1 and 0.022 respectively.



(a)



(b)

Fig. 6.9 Variations of parameters R_a, R_q, Δ_q and P_c of the stylus response to a simulated force signal of equivalent speeds of (a) 0.05 mm s^{-1}, and (b) 0.1 mm s^{-1} as the damping is varied. Parameters are normalised to the values at the damping ratio 0.6.

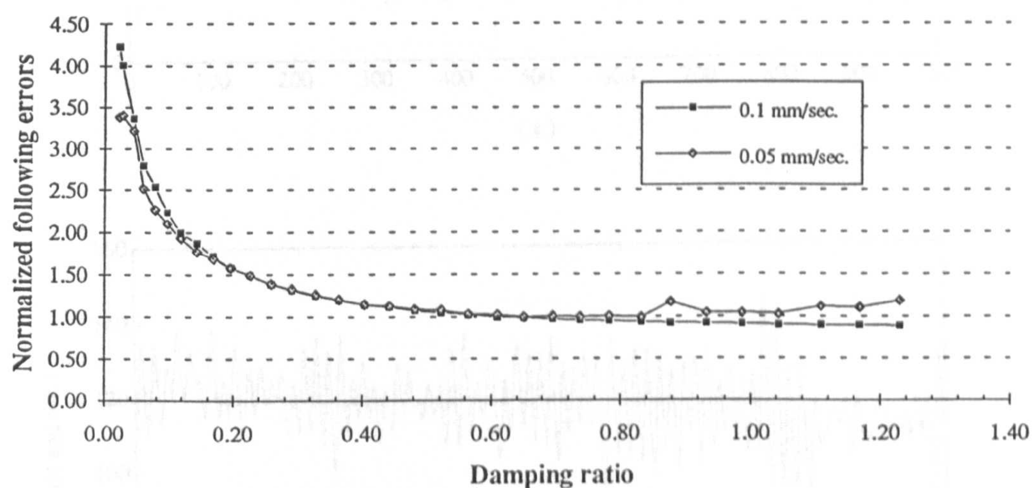
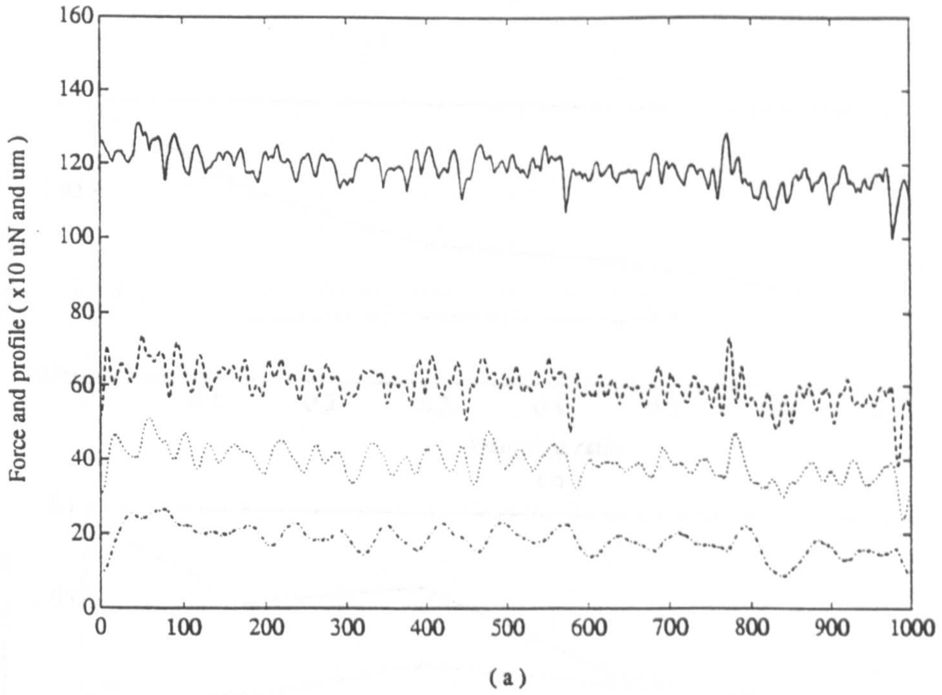


Fig. 6.10 Standard deviation of following errors after subtracting the original profile from the stylus responses in figure 6.9. Values are normalised at damping factor 0.6.



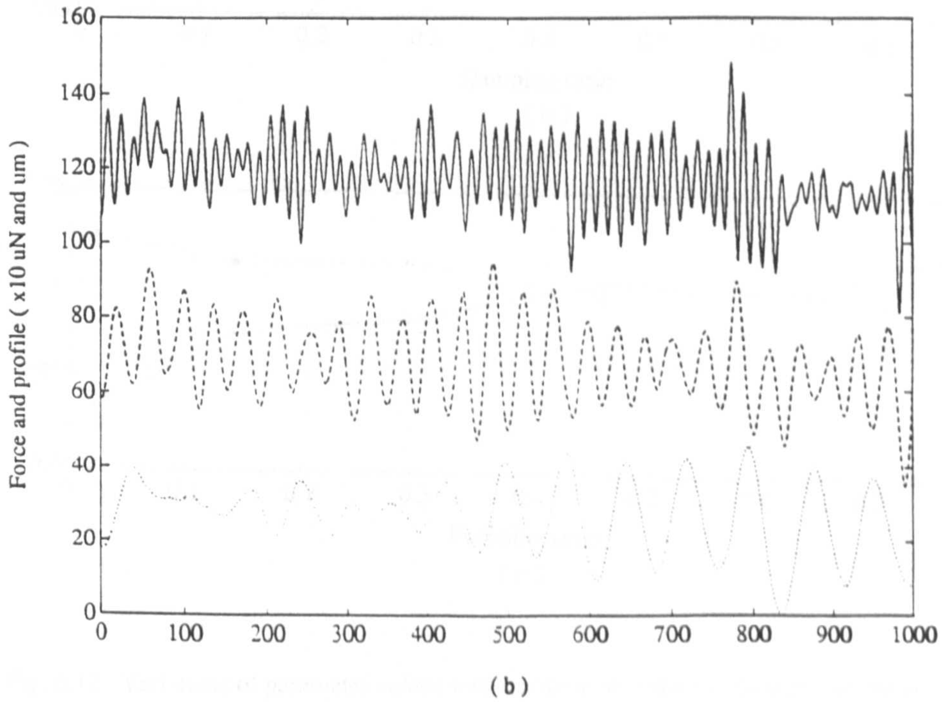


Fig. 6.11 Responses of stylus to simulated force signals of equivalent speeds of 0.1, 0.25, and 1 mm s^{-1} at damping ratios of: (a) 0.59 and (b) 0.022. Graphs are arranged from top to bottom in the order as speed increases. The top curve in (a) is the original profile.

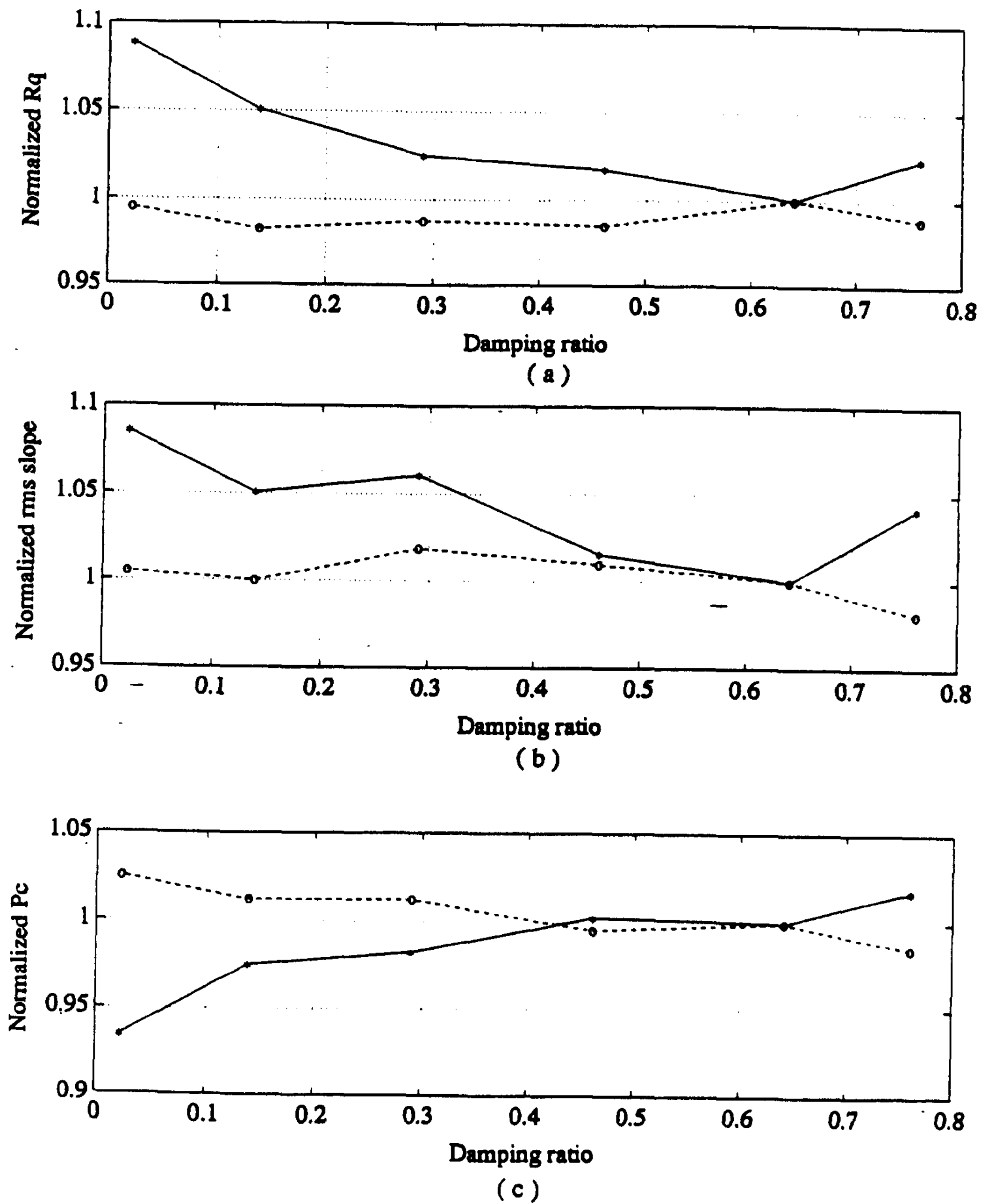


Fig. 6.12 Variations of parameter values with damping on a steel surface at traverse speed of 0.05 mm s^{-1} (full curve) and 1 mm s^{-1} (broken curve): (a) R_q normalised to 0.079 μm; (b) Δ_q normalised to 0.082; (c) P_c normalised to 0.34, 0.25 μm^{-1} respectively.

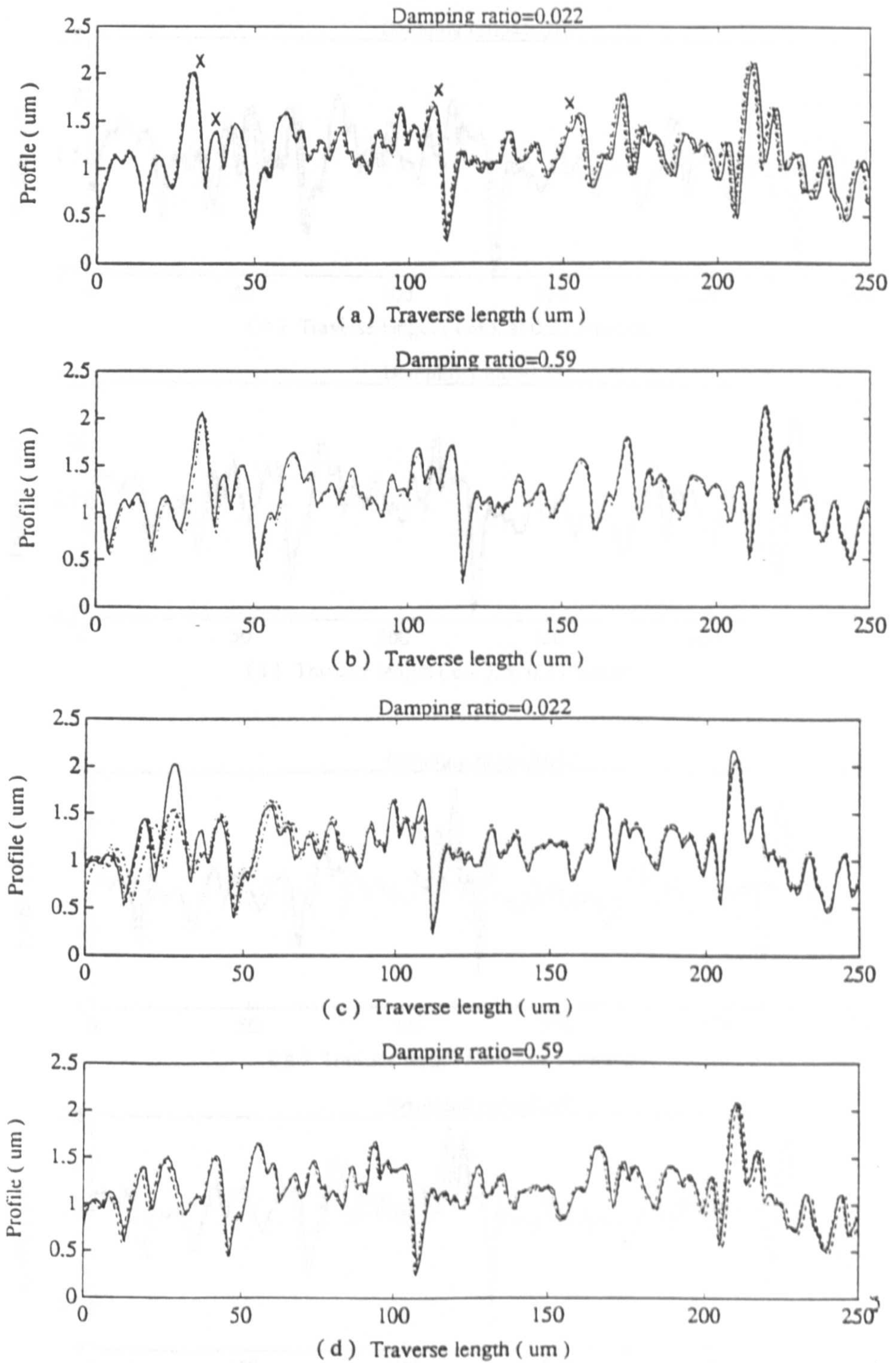
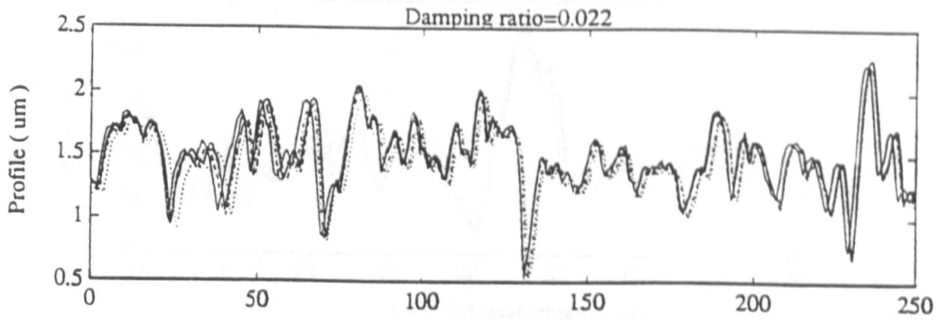
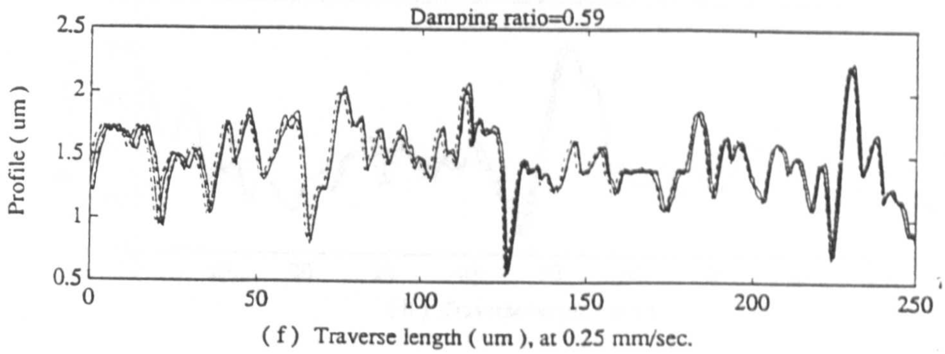


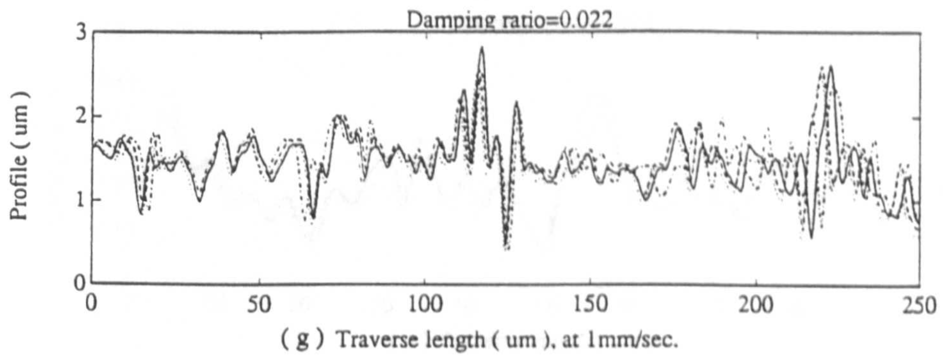
Fig. 6.13 Repeated traces on a copper specimen at two damping values with different tracking speeds: (*a*), (*b*) at 0.05 mm s^{-1}; (*c*), (*d*) at 0.1 mm s^{-1}; (*e*), (*f*) at 0.25 mm s^{-1}; (*g*), (*h*) at 1 mm s^{-1}.



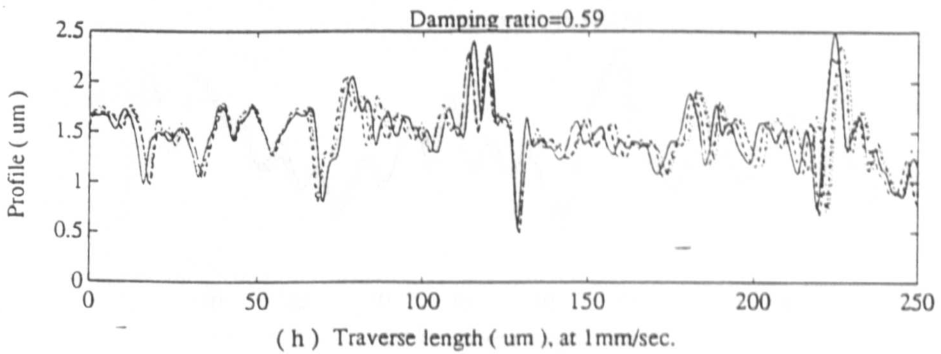
(e) Traverse length (um), at 0.25 mm/sec.



(f) Traverse length (um), at 0.25 mm/sec.

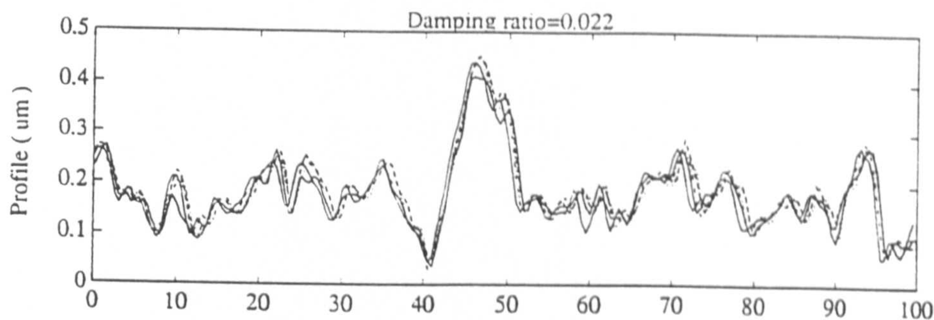


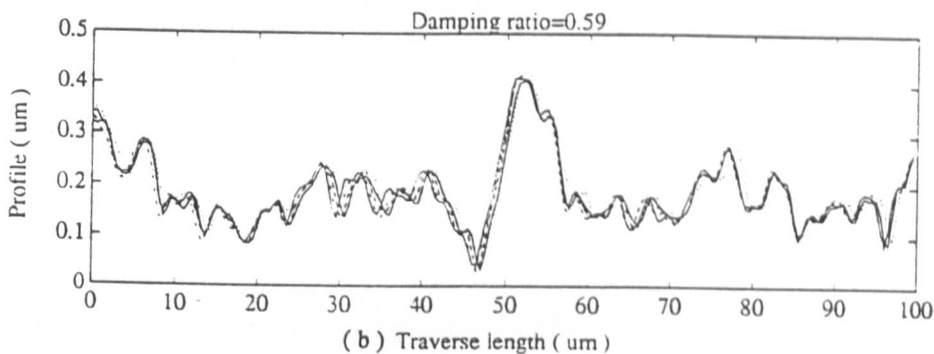
(g) Traverse length (um), at 1mm/sec.

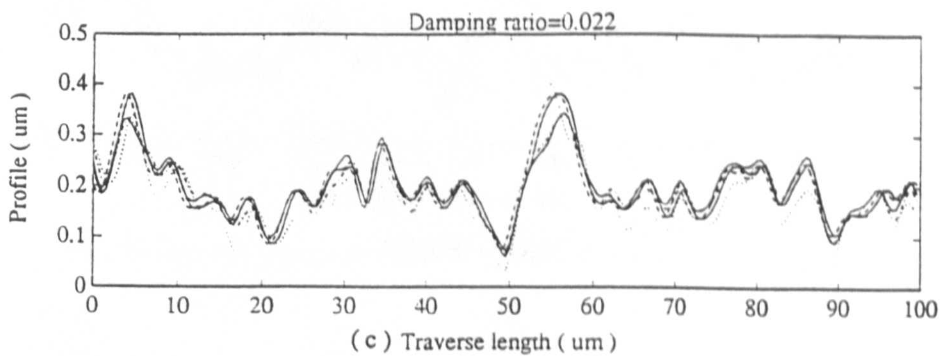


(h) Traverse length (um), at 1mm/sec.

Fig. 6.13 (continued) (*e*), (*f*), (*g*) and (*h*).







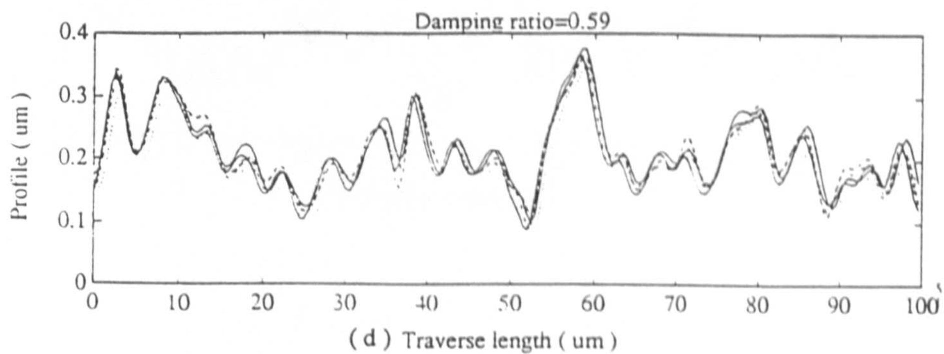


Fig. 6.14 Repeated traces on a mild steel specimen at two damping values with different tracking speeds: (*a*), (*b*) at 0.25 mm s^{-1}; (*c*), (*d*) at 1 mm s^{-1}; (*e*), (*f*) at 0.1 mm s^{-1}.

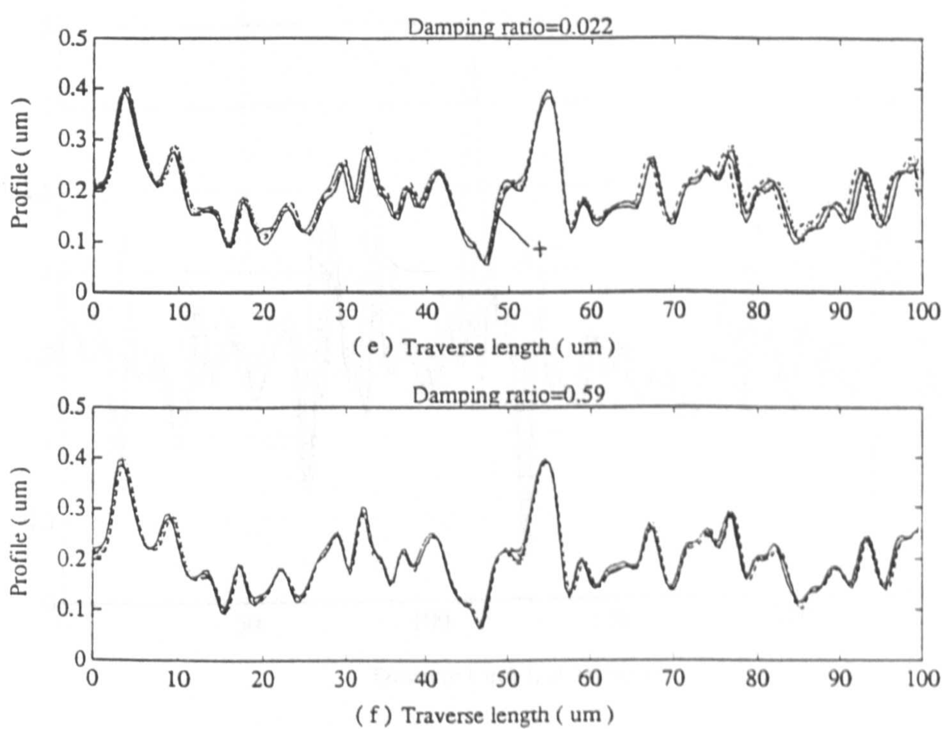


Fig. 6.14 (continued) (*e*), (*f*).

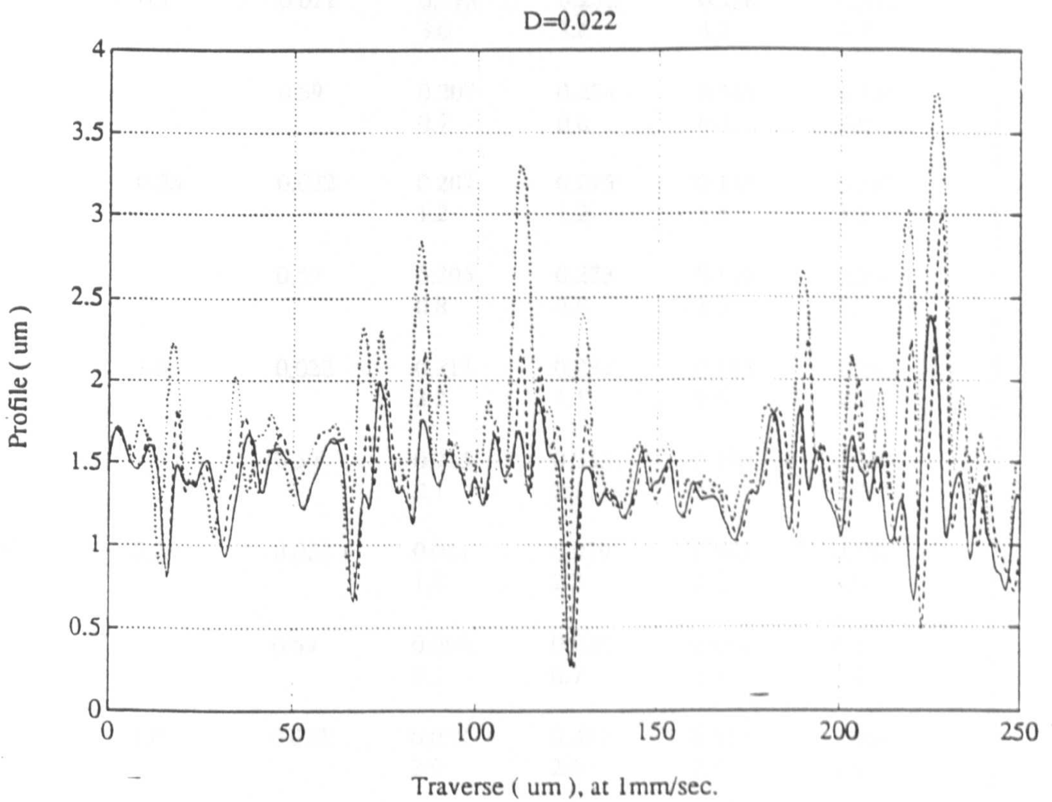


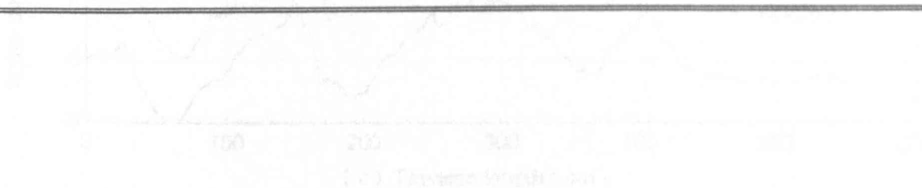
Fig. 6.15 Output of the stylus at loads of 1.55 (full curve), 0.84 (broken curve) and 0.46 (dotted curve) mN with a speed of 1 mm s^{-1} and a damping factor of 0.022.

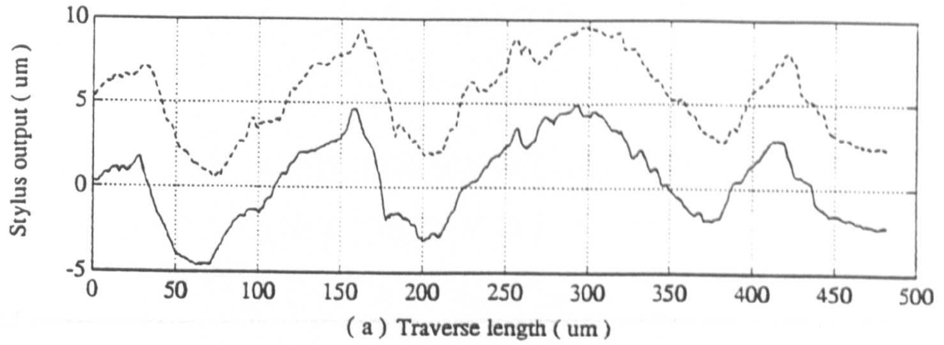
Table 6.1 Surface finish parameters and their relative variations for copper and mild steel specimens.

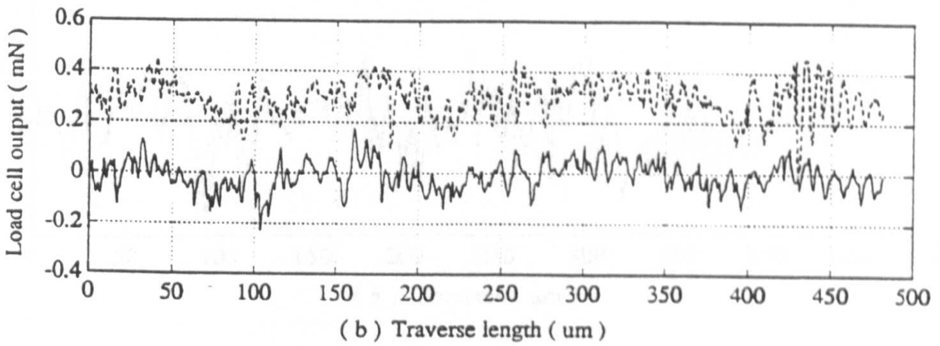
Materials	Speed mm s^{-1}	ζ	R_a (μm) ±σ%	R_q (μm) ±σ%	Δ_q (rad) ±σ%	P_c ±σ%
Copper	0.05	0.022	0.224	0.299	0.164	0.359
			0.3	0.3	2.1	8.7
		0.59	0.222	0.295	0.159	0.336
			0.3	0.4	0.5	5.1
	0.1	0.022	0.213	0.282	0.150	0.331
			3.0	3.6	4.2	4.7
		0.59	0.207	0.274	0.145	0.324
			0.7	0.6	0.4	4.0
	0.25	0.022	0.207	0.275	0.141	0.292
			1.2	1.2	1.4	3.2
		0.59	0.205	0.273	0.140	0.291
			0.8	0.6	1.3	2.2
	1.0	0.022	0.213	0.291	0.165	0.252
			3.2	4.1	9.5	12.0
		0.59	0.225	0.313	0.196	0.309
			2.1	2.1	3.7	3.7
Steel	0.25	0.022	0.061	0.079	0.062	0.181
			1.8	2.3	2.2	4.8
		0.59	0.060	0.078	0.061	0.183
			0.5	0.7	1.1	1.9
	1.0	0.022	0.058	0.077	0.053	0.104
			2.9	2.6	2.7	3.5
		0.59	0.058	0.078	0.055	0.107
			0.8	0.9	1.4	3.7
	0.1	0.022	0.061	0.079	0.068	0.244
			0.98	1.0	0.9	2.1
		0.59	0.061	0.078	0.067	0.238
			0.8	0.7	1.0	3.1

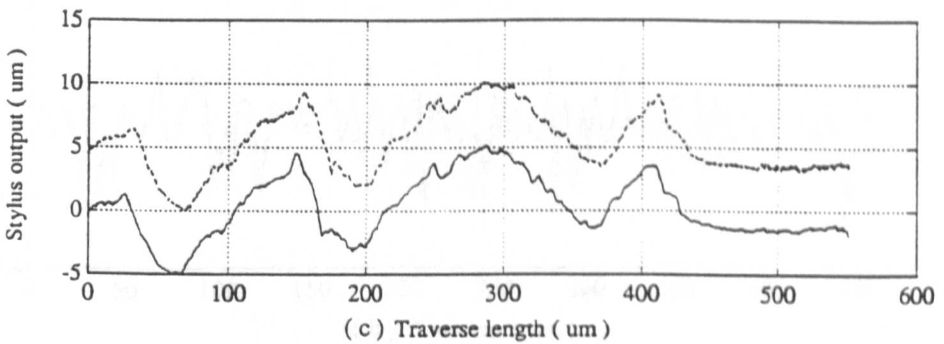
Table 6.2 Surface finish parameters on a copper specimen at two damping ratios and three different loads.

Force (mN)	ζ	R_a (μm)	R_q (μm)	Δq (rad)	Pc
0.46	0.59	0.318	0.454	0.252	0.337
	0.022	0.350	0.503	0.259	0.342
0.84	0.022	0.246	0.345	0.224	0.339
	0.59	0.244	0.340	0.224	0.338
1.55	0.59	0.222	0.308	0.180	0.250
	0.022	0.219	0.300	0.172	0.232









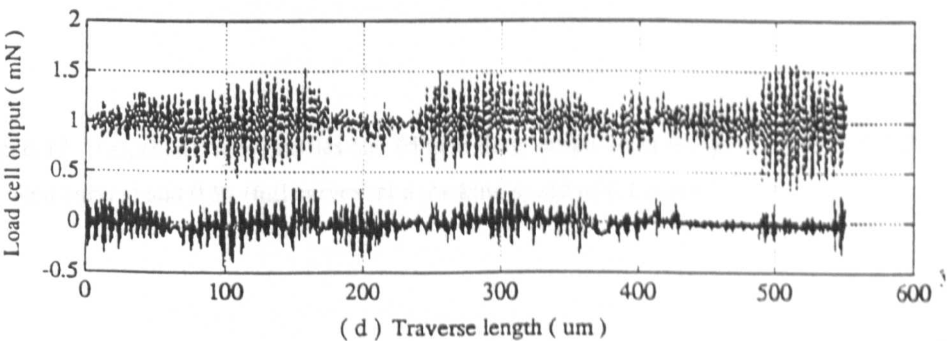


Fig. 6.16 Profile of a load cell button and the corresponding force measured at damping ratios of 0.022 (broken curves) and 0.59 (full curves) at tracking speeds of: (*a*), (*b*) 0.1mms^{-1}; (*c*), (*d*) 0.25 mm s^{-1}.

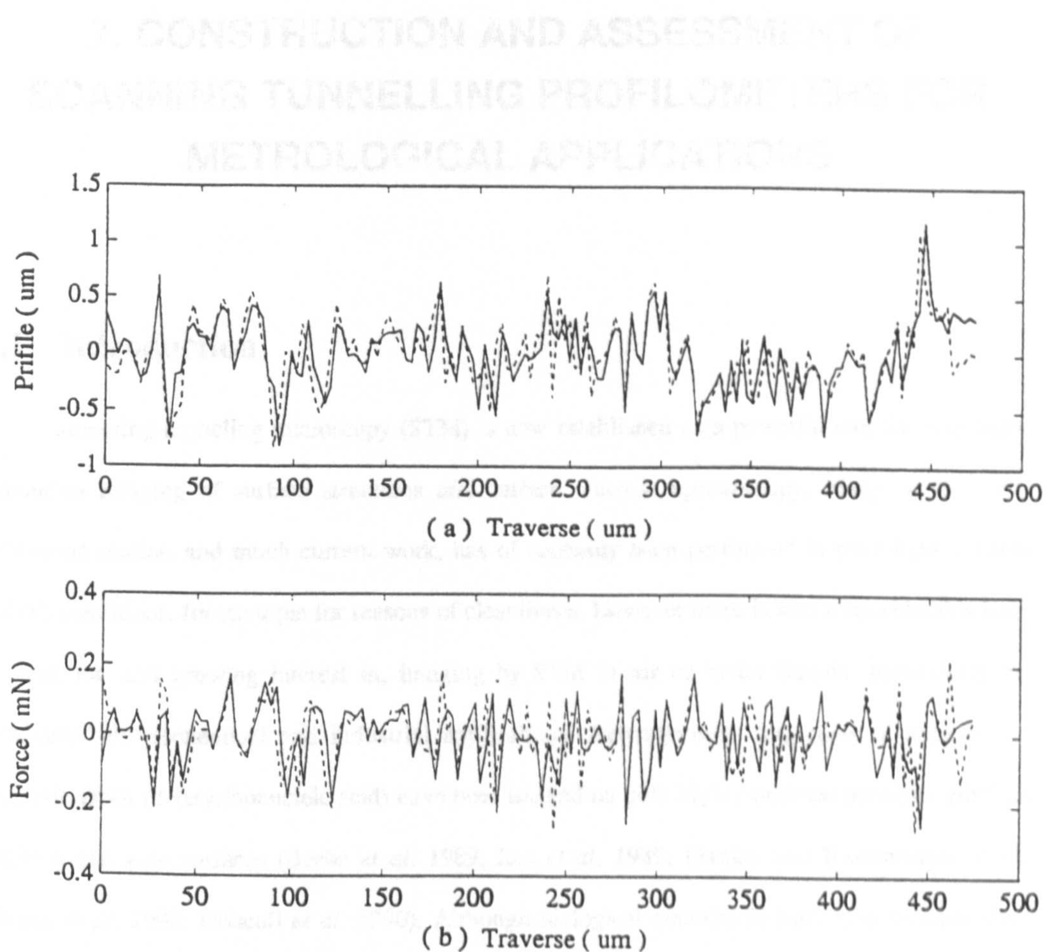


Fig. 6.17 (*a*), (*b*) Surface profiles and friction drag variations at damping ratios of 0.022 (broken curves) and 0.59 (full curves) at a tracking speed of 0.1 mm s^{-1}.

7. CONSTRUCTION AND ASSESSMENT OF SCANNING TUNNELLING PROFILOMETERS FOR METROLOGICAL APPLICATIONS

7.1 Introduction

Scanning tunneling microscopy (STM) is now established as a powerful tool for very high-resolution imaging of surface structures and surface electron spectroscopy. Many of the well-publicised studies, and much current work, has of necessity been performed in ultra-high-vacuum (UHV) conditions, for example for reasons of cleanliness. However there is also a considerable body of work on, and growing interest in, imaging by STM in air or under liquids, particularly for biological, electrochemical and industrial applications. Heterogeneous fragments of natural and synthetic DNA (deoxyribonucleic acid) have been imaged on both highly oriented pyrolytic graphite (HOPG) and gold surfaces (Beebe *et al.* 1989; Lee *et al.* 1989; Dunlap and Bustanmante 1989; Allison *et al.* 1990; Driscoll *et al.* 1990). Although biological material in bulk is non-conductive, when isolated molecules are placed on a conductive surface, high-quality STM images have been obtained through a mechanism which is not yet fully understood (Bottomley *et al.* 1992). STM techniques have also been demonstrated as a means of manipulating surface structures at the nanometre scale.

These studies are useful, but they tend, naturally, to approach the subject from a surface physics background. Here the potential uses of STM in an engineering context, that is operating on relatively large specimens in air and with well defined displacement metrology, are explored. For this purpose, two scanning tunnelling profilometers have been designed, constructed, and their performances assessed. Based on these two STMs, characteristics of STM system have been

investigated and some metrological implications of variation in the tip to specimen separation are then discussed in chapter 8.

7.2 Analysis of STM

7.2.1 Principles of STM

In scanning tunnelling microscopy, the three-dimensional variation of the charge density at a surface is probed via electron tunnelling between a sharp tip and the sample, as shown in Fig. 7.1. This vacuum tunnelling phenomenon has been known since the introduction of quantum mechanism in the 1920's (Fowler and Nordheim, 1928). It has been shown that the one-dimensional tunnelling theory of Simmons (1963) gives a good fit to the observed behaviour in both low-voltage and field-emission regimes, and the tunnelling current can be written

$$I = \frac{c}{s}\left\{(\bar{\phi} - \frac{eV}{2})\exp\left[-A(\bar{\phi} - \frac{eV}{2})^{\frac{1}{2}}s\right] - (\bar{\phi} + \frac{eV}{2})\exp\left[-A(\bar{\phi} + \frac{eV}{2})^{\frac{1}{2}}s\right]\right\} \tag{7-1}$$

where $A = 1.025\ (\mathrm{eV})^{-1/2}\ \text{Å}^{-1}$, s is the tip-sample gap spacing, c is a constant, $\bar{\phi}$ and V are the average of the barrier height and the bias potential between the sample and the tip,respectively. Equation (7-1) is valid even when the gap voltage $V > 5$ V, in the field emission regime. When the gap is small and the voltage low, the current-gap distance relation can be simplified to

$$I \propto \left(\frac{V}{s}\right)\exp(-A\bar{\phi}^{\frac{1}{2}}s) \tag{7-2}$$

With a barrier height $\bar{\phi} \sim 4$ eV a change of the separation, s, by a single atomic step of 2-3 Å changes the current by up to three orders of magnitude.

As the tip horizontally scans a specimen, any change in the tip-surface spacing results in a large change in current I. Some feedback mechanism can be used to move the tip up or down to maintain a constant value of I. According to equation (7-2) this implies that $\bar{\phi}^{\frac{1}{2}}s$ = constant. Thus

the s displacement gives the surface topography only for constant work function. Because of surface heterogeneity a normal STM image contains both geometric and electronic information about the surface. One way to differentiate between changes in topography and work function is to modulate the gap s, at a frequency higher than the cut-off of the control system or with the feedback loop disabled momentarily. A ramp voltage is then applied to measure the tunnelling current as a function of applied voltage. This measurement can be performed at each point of a topography scan, resulting in spatially resolved I (V) relation. Some electronic structural information, such as the energy-band gap of semiconductors or superconductors, can be directly obtained. Since the area probed by the STM is typically so small, the uncertainty principle may play a role in determing the energy resolution of this spectroscopy. Therefore, a sharp tip is desirable for topographic scanning but undesirable for tunnelling spectroscopy on a surface with a dispersive band structure (Kuk and Silverman 1989). In general, for an absolutely flat surface with an effective work function of (4 ± 1) eV and a nominal tip-surface spacing of 0.5 nm, the STM topography would reveal an apparent height variation of less than 0.1 nm. Hence, for surface roughness variations that exceed 0.1 nm, STM images obtained in the constant current mode reveal primarily surface topography.

7.2.2 Motion control

To bring the tip to within tunnelling range, it must be moved over a macroscopic distance (perhaps hundreds of microns). During the tunnelling and feedback control, the tip must be positioned with a precision better than 0.1 nm. The task is usually segregated into coarse and fine motion control.

• Coarse motion control can use several methods (Smith, 1991). As described in their original work, Binnig and Rohrer used a piezoelectric inchworm, a 'louse' as they called it. This device achieves motion in a manner similar to some caterpillars. A piezoelectric actuator is mounted onto either a slideway or an optical flat with a conducting surface. Attached to the base of the actuator are two or more conducting pads separated from contact with the slideway by a thin layer of insulating material. The inchworm is actuated by first clamping one pad to the slideway by an

electrostatic force and then extending the piezoelectric material. This feed end is then clamped and the fixed end is released. By then contracting the piezo and repeating these steps, it is possible to obtain unlimited motion. This type of device was first proposed by Joyce and Wilson (1969). More recent designs are based on purely mechanical means, using fine-thread screws to get several millimetre range with sub-micron precision

• Fine motion control is required in the feedback control during tunnelling. The precision of tip positioning relative to the specimen surface must be better than 0.1 nm. This is usually achieved by using piezoelectric positioners. Lead zirconium titanate (PZT) is the material of choice in the STM community. Most STMs are designed with a response of 1 to 10 nm V^{-1}, although a response of up to 500 nm V^{-1} can be achieved for large-scale scanning. Because voltages can be easily controlled and monitored in the submillivolt level, subnanometre control can be readily attained. In the initial development of STMs, three-axis control was accomplished using three separate piezoelectric bars held together in an orthogonal arrangement. To improve rigidity, especially for long-range scanners, most designers opt for a design based on a piezoelectric tube scanner, shown in Fig. 7.2(b) (Binnig and Smith, 1986). The inside of the piezoelectric tube is completely metal-coated, whereas the outside is metal-coated in four separate quadrants. By applying appropriate voltages to one pair of diametrically opposite quadrants, one causes the piezoelectric tube to bend along that direction, thus achieving x or y scanning motion. Application of voltage to the inner surface causes the tube to expand or contract (z-axis motion). Therefore, three-axis motion can be attained with a single tube.

7.2.3 Vibration isolation

The exponential dependence of the tunnelling current on gap distance makes vibration isolation an extremely critical factor in good STM design. For many surfaces, especially metals, the atomic corrugations observed in a constant current STM scan will typically have amplitudes of around 0.1 Å, so one must set a vibration isolation goal of ~ 0.01 Å or less (Park and Quate 1987;

Pohl 1986; Okano *et al.* 1988). Even better vibration isolation may be required for the detection of inelastic tunnelling signals (Binnig *et al.*, 1985).

There are two common types of vibration isolation systems in STM. The first one is coiled spring suspension with magnetic damping, and the second is a stack of stainless-steel plates with Viton dampers between each pair of steel plates. The second system is easy to build and relatively rigid so that it is convenient to manipulate the internal components. On the other hand, it may need additional pneumatic suspension system to improve the vibration isolation at low frequencies. The first system is more complicated to build, but it has proven to be more efficient for vibration isolation. Both systems have been used in STM instruments capable of recording images with atomic resolution.

Assume that the STM sits on a platform that is coupled to the outside world via a spring to provide a resonant frequency, f, and that the lowest resonant frequency of the STM is F, which is much greater than f. The external vibration has a frequency, f' and amplitude, A. With such a system, the vibration amplitude transmitted to the STM depends on the frequency f' of the vibration. There are four regimes to consider.

The first regime is $f' < f$. The platform spring does nothing to attenuate the external vibration. The vibration amplitude entering into the microscope causes a tip-surface spacing change by

$$a = A(f'/F)^2 \qquad (7\text{-}3)$$

The second regime is $f' \sim f$. The vibration amplitude entering into the microscope is actually amplified, depending on the amount of damping in the platform spring.

The third regime is $f << f' << F$, where the vibration amplitude, a, entering into the microscope is independent of f' and given by

$$a = A(f/F)^2 \qquad (7\text{-}4)$$

In the fourth regime, $f' >> F$, only the platform spring does the attenuation and the transmitted vibration amplitude, a, is given by

$$a = A(f/f')^2 \qquad (7\text{-}5)$$

In examining these four cases, it becomes clear that one should support the STM on a soft platform (small f value) and design a microscope with high rigidity (large F value). For example, for the third regime, with $f = 1$ Hz, $F = 10$ kHz, and $A = 10$ μm, the transmitted vibration amplitude, a, can be shown to be around 10^{-4} nm for intermediate frequencies. Therefore, a rigid STM not only allows fast image acquisition, but also more effective vibration isolation, as well.

7.2.4 Hysteresis and creep

Hysteresis and creep (Gallego-Juarez 1989) are undesirable characteristics common to all piezoelectric actuators (piezos for short) used in STMs. At a given applied voltage, the piezo's position depends on whether the applied voltage is increasing or decreasing. In addition, the piezo will make an initial instantaneous response to the voltage change, followed by slow continued movement over next few hours. This slow motion is creep and shows up as a drift in the STM probe position.

Control of piezo movement by controlling the applied charge (charge control) instead of the applied voltage has been proposed by Newcomb and Flinn (1982) and later simplified by Kaizuka and Siu (1988) as a way to reduce hysteresis and creep. They believe that the hysteresis and creep arise from the piezo's own capacitance changing with its extension. By inserting another capacitor in series with the piezo's capacitor the percentage of hysteresis is reduced by up to one fifth with, however, a similar reduction in the sensitivity.

A more effective method of reducing the hysteresis and creep is to use a feedback control loop in the actuator. By doing this the precision of positioning a piezoelectric actuator will be determined by the feedback transducer such as a capacitance gauge which can provide higher precision and low drift. Commercial devices are available now, for example, the digital piezoelectric translator (Queensgate Instruments Ltd) and this type actuator will be discussed in detail later.

Temperature has a great effect on piezoelectric materials. The extension of a piezo element for a given applied voltage generally increases with temperature. Thermal expansion coefficients of PZT materials are typically 1×10^{-6} to 5×10^{-6} K^{-1}. Operation at elevated temperatures (close to the

Curie temperature, which is above 300 °C for hard PZT, but less than 200 °C for soft PZT, King *et al.* 1990) will partially or completely depolarise the material, causing a degradation of its piezoelectric properties.

7.3 Design of a 2-D STM Based on a Spring Translation (STM_1)

7.3.1 Spring flexure driven by electromagnetic force

To reduce the hysteresis and creep associated with the PZT drives, an alternative approach is exploited by using electro-magnetic drive. One such design, where a permanent magnet is attached to the moving platform of a spring mechanism which is, in turn, surrounded by a solenoid actuation coil, has been investigated by Smith and Chetwynd (1990). In this study, open loop precision of better than 0.5% could be obtained over a displacement range of 100 nm to 50 μm with a single design. In this design, the displacement is linearly proportional to the applied force which is, in turn, proportional to the coil current. The invariance of force with displacement indicates a drive of extremely high compliance and is one of the key differences between electromagnetic and piezoelectric actuators. This high compliance effectively decouples parasitic forces caused by manufacturing error, thereby exploiting the rectilinearity of flexures (Jones 1956), and introduces the possibility of assessing tip-to-surface interactions from the observed tunnelling characteristic. A similar technique using a single crystal silicon flexure spring has been used for the control of the analyser in an x-ray interferometer where open loop accuracy of better than approximately 30 pm linear and 10^{-9} angular over a displacement range of 100 nm has been observed (Bowen *et al.* 1990).

2-D STM, a profilometer in another words, has been constructed in which a tunnelling probe is mounted on a linear spring flexure driven by electromagnetic force. A schematic diagram of the profilometer system is shown in Fig. 7.3 and a photograph of the complete profilometer shown in Fig. 7.4. The tunnelling probe is mounted to a moving platform of the z-axis linear spring. Attached to the base are three drive coils, with inside diameter 11 mm, that surround, but are not in contact

with, the six permanent magnets. In each coil there are two magnets separated by an aluminium spacer so that the centre of the poles is coincident with the ends of the coil (*i.e.* at the position of maximum field gradient), see Smith and Chetwynd (1990). A similar spring flexure for specimen translation is positioned perpendicular to the drive axis. This is driven by two coils rigidly mounted to a base which forms the upper platform of a coarse positioning spring with its axis collinear with the probe. A differential micrometer enables manual positioning in which it is possible, with dexterity, to position the probe to within about 50 nm.

Both spring flexures are type of notch hinges which are 3 mm and 2.5 mm thick and are separated by distances of 50 and 35 mm for the probe and specimen springs respectively. Both are 30 mm deep and manufactured from a single monolith of duralumin. Assuming an elastic modulus of 70 GN m^{-2}, the theoretical stiffness values of the two flexures are 1.5×10^6 and 2.1×10^6 N m^{-1} (based on formulae given by Smith and Chetwynd 1992).

A number of constructional considerations were required for the instrument to function correctly. Due to the small displacement range of the probe (around 2 μm) the thermal stability becomes important if the probe is to be maintained with the range for long periods of time (in practice it is possible to use this instrument continuously for 5 - 6 hrs). Because of the high thermal expansion coefficients of the aluminium, brass and steel materials used in this design, any component parallel to the probe axis will affect the steady-state thermal stability of this instrument. Although the measurement loop is relatively large going from the probe, through the drive flexure and up through the specimen flexure via the coarse feed micrometer, it is constructed almost entirely from aluminium, so thermal effects will tend to balance out in the steady-state. Additionally, to avoid twisting effects, the whole instrument is symmetrical about the probe axis. The effect on the thermal loop due to long length (200 mm) of the probe axis flexure has been effectively reduced by bonding the front end to a glass base and lightly clamping the rear. Consequently, any changes in dimension will tend to emanate from the bonded region reducing the influence of any parts of the structure towards the probe.

The clearance between the outside edges of the magnets and the inside core of each coil should be as small as possible to give good sensitivity. Here the clearance is about 150 μm and it is very difficult to ensure that they are not touching. This is done by electrically isolating the coils from the base and then checking for a short circuit as an indicator of mechanical contact. By slightly distorting or aligning the structure a non-contacting condition can be obtained.

The number of coils has been chosen so that the mass of magnet material is comparable with that of the moving platform. In the case of the specimen platform two coils have been used to reduce space while maintaining the probe axis symmetry of the structure. A glass plate has been chosen for the instrument foundation because of its high stiffness and dimensional stability.

7.3.2 Calibration

Two current amplifiers were built for driving the z and x-axis spring stages using a circuit similar to that shown in Fig. 4.4. For the z-axis drive, a battery powered reference voltage was used to reduce electronic noise. Calibration of the current/displacement characteristics of the translation stages has been performed using a Tesa LVDT gauge with a stiffness that is insignificant compared with that of the springs. These are accurate to within 5% for the range and time taken for each calibration. The drive voltage to the current amplifiers is supplied from the Metrabyte DAS 16 I/O driven from the PS/2 microcomputer. The accuracy of this part of the system, including the constant of proportionality of the current amplifier is known to within 1%. Fig. 7.5 (a) and (b) show plots of input current against output displacement for z and x-axis translation stages. The sensitivities for the two translation stages are 7.2 and 12.9 nm mA^{-1} respectively. Errors from linearity are thought to be due mainly to the misalignment between the stage and the LVDT. The linearity also depends on the fractional demagnetisation which, in turn, closely related to the recoil permeability. In the application above where a relatively weak field is generated and the permanent magnet NdFeB has a very flat characteristic curve in the presence of a low field strength among those other permanent magnets such as NdDyFeBNb and NdDyFeB, thus the effect should be very small.

The resonance of each stage was tested by tapping lightly using soft stick and measuring the response from the stylus pick-up with a spectrum analyser, TR 9430 (ADVANTEST). For z-axis stage the resonance was found 337.5 Hz. There is a second peak at a frequency of 395 Hz, and the reason for this behaviour is not clear. It may be due to the differences in stiffness within the four notches which is equivalent to a mechanism of two series springs with stiffness λ_1 and λ_2. The resonance of the x-axis stage is quite high about twice that of the z-axis because of its relative small size.

Tunnelling experiments have been performed with the instrument mounted on a large concrete block that is, in turn, mounted on a pneumatic isolation stage to attenuate ground borne vibration. All of this is encased in a lead-lined cabinet to provide high mass-damping for acoustic attenuation. It should be noted, however, that even with these precautions the best results were obtained after normal working hours and with minimal acoustic disturbances from nearby conversations. The results are shown in section 7.8.

7.4 A 2-D STM Based on a DPT and a Precision Slideway (STM_2)

7.4.1 Close loop compensation of PZT

To overcome the nonlinearity associated with the piezoelectric actuators, closed loop control strategies have been developed. Using capacitive feedback techniques, Hick *et al.* (1984) were able to control Fabry Perot etalons (parallel optical flats for use in spectrometry) at a dynamic gap separation maintained constant to 1 pm $Hz^{-1/2}$. Later on they developed a series commercial piezoelectric actuators with built in feedback. The principle is shown in Fig. 7.6, where a stack of disk piezoelectric elements is driven by a high voltage amplifier. Each end of this stack is connected to an Invar cap, to which external devices may be attached. An ultra-low expansion glass-ceramic ('Zerodur') rod is positioned in the centre of the tube, attached to one cap. An electrode deposited on

its free end forms part of a capacitance gauge that monitors the extension between the two end plates. Any error between the position sensor and a specified reference value (or demand) is used to provide a drive signal to the actuator via an integrator. The integrator is necessary because the field applied to the piezostack must be maintained if it is to hold a constant extension. If a simple proportional control is used, the drive signal would go to zero as the actuation error decreased and the device would relax. Another advantage with an integral control is that the stiffness of the actuator will be much increased (up to infinity in principle). For example, a load is applied to the actuator resulting in a distortion monitored by the capacitance sensor, the integrator will continue increasing the drive over time until it fully compensates the force and returns the capacitance electrodes to the desired separation. It is usual to combine proportional and integral control for an optimum response. The relative position of the two end plates can be set with sub-nanometre precision over a range of 15 μm under applied loads up to 10 N or more and with a temperature coefficient of less than 15 nm K^{-1}, for example the digital piezoelectric translator (DPT from Queensgate Instruments Ltd). The potential of this type of devices for displacement transfer standards of sub-nanometre accuracy has been explored by Harb *et al.* (1992).

In this 2-D STM, a DPT from Queensgate Instruments Ltd is used as *z*-axis position control. As shown in Fig. 7.7 the tunnelling tip is mounted in a small aluminium holder which is, in turn, glued to the lower face of the DPT, separated by a 0.5 mm thick alumina plate for electrical isolation. The top face of the DPT is attached to a Rank Taylor Hobson 'Nanostep', in which the mounting bracket for the stylus-based inductive gauge has been replaced. Coarse adjustment of the complete probe assembly over a 25 mm distance could be achieved manually using the standard micrometer-based adjuster of the Nanostep.

7.4.2 Nanostep slideway

The Nanostep slideway (Garratt and Bottomley, 1990) is used for the specimen translation. It has a prismatic slideway providing the traverse surface and two carriages that rest on it, held by gravitational force. One carriage serves as a slave to transmit forces and the other is the specimen

carriage. Above this is a frame to which originally a stylus-base displacement transducer LVDI and now the tunnelling probe is attached. To reduce the thermal expansion of the measurement loop, all of these components are constructed in the low expansion ceramic Zerodur. Attached to the underside of each carriage are five PTFE (poly-tetra-fluoro-ethylene) bearing pads, two to one side and three to the other to provide smooth and low friction bearing. These pads are PTFE/lead composites contained within a sintered bronze matrix and lubrication is believed to be through oriented films being drawn from the matrix and deposited onto both the bronze and counterface material. This mechanism has intrinsic advantages that thermal conductivity and stiffness of the bearing are not adversely affected by the polymer and wear is in part compensated by the reservoir held in the matrix. The detail of polymer bearing is not discussed here and can be referenced to Lindsey *et al.* (1988), Smith *et al.* (1992 and 1993). Using this slideway system, a traversing length of 50 mm with the straightness better than 15 nm can be readily achieved. The vertical fluctuation due to the traversing is within 1 nm (Lindsey *et al.* 1988).

The specimen carriage is translated by a micrometer driven by a DC motor/gearbox system. This is mounted on a cast iron block which is, in turn, isolated from the instrument base by soft anti-vibration mounts. To reduce electromagnetic coupling between DC motor and the STM probe, the complete motor drive was enclosed by a tin-plated steel box. This material is chosen here for its screening of both electric and magnetic interferences.

7.5 Instrumentation

7.5.1 Preamplifier

As shown in Fig. 7.8, tunneling current sensing is obtained via a 10 MΩ cermet film resistor in the feedback path of the preamplifier. The tip is connected to the inverting input of the preamplifier and the sample is biased via an adjustable power supply. The non-inverting input of the preamplifier is grounded. The requirement for a preamplifier with low offset input current, low

noise and low drift is not difficult to meet with modern electronics. Both the metal canned bi-polar amplifier LM11 and the FET amplifier OPA111 were adequate and an OPA111 was used for the results given here. Traditionally, STM designs have applied bias to the tip and sensed current from the specimen. However noise pick-up at the sense input depends upon the amount of exposed electrode and the present design, with 50 mm scan capability, requires large paths around the specimen. Hence in all the experiments the tip was used as the sensing element with the bias applied to the specimen. The tip bias voltage is provided by a battery powered reference circuit through a 10-turn potentiometer, giving a low noise and stable bias voltage. The preamplifier and tip bias circuits were built in separate tin-plate box placed close to the sensing probe.

7.5.2 Linearisation circuits

After pre-amplification, the tunneling signal is further amplified using an AD524 instrumentation amplifier inside the main electronics box, giving adjustable gains of x1 to x500 by selecting the two switches, sw1 and sw2. In order to improve the dynamic range, the non-linearity in the current-versus-gap relationship, the signal is then passed to an AD759N logarithmic amplifier. It offers closely logarithmic operation over three decades of voltage (1mV to 2V). Its output becomes zero when the input is 100 mV. Thus by changing the gain of the AD524, we can vary the equilibrium tunnelling current and gap distance without using a reference voltage. Rectification of the signal, when it is negative, prior to the log amplification is obtained using a single IC, AD630 modulator/demodulator, rather than the common op-amps/diodes inverting and noninverting rectifier circuit. The AD630 is made to perform rectification when operated in the modulator mode with both modulation and carrier inputs connected to the signal to be rectified. An output voltage equal in magnitude to the input regardless of the sign of the input voltage is then produced. The tunnelling current can be obtained at this output and this is collected through another AD524 configured with unity gain. The output of the log amplifier is then fed into the PID controller to adjust the tip-specimen separation.

7.5.3 PID controller

A stand-alone proportional, integral and derivative (PID) controller was designed and constructed, details are shown in Fig. 7.9. A reference signal or a set point value is provided, for a general purpose, by a referencing circuit which includes a pair of precision zener diodes ZN423, a multi-turn potentiometer and an AD524 (U4). For STM control, the reference is by-passed by switching 's1' to ground. The controller consists of an input or comparison amplifier U1 (AD524), a passive proportion net with a multi-turn potentiometer, a buffer U2 (OPA121), a derivative amplifier U3 (OPA121) which is optional for optimising system performance by selecting 's3', an integral amplifier U5 (OPA121), the summing amplifier U6 (AD524), and a reverse amplifier U7 to give optional output with different sign. The feedback loop can be switched on or off by 's2'. To avoid over-saturation of the integration action, a push button is inserted across the integrating capacitors for discharging. The variations of three terms are P: 0.001 to 1; T_i: 0.1 msec. to 0.1 sec.; T_d: 0.1 msec. to 10 msec.

The requirement for the operational amplifiers in PID control is low drift, low input bias current and high open-loop gain to minimise low frequency errors. Most of the op amps are type OPA121 from Burr-Brown, chosen for its low cost and high performance (low bias current 5 pA, low noise 1.6 μVp-p between 0.1 and 10 Hz, low offset voltage drift 3 μV/°C). For better performance, ultra-low noise and drift of OP27 from Analogue Devices can be used but the OPA121 has been proven to be adequate for our purposes. AD524 is a precision instrumentation amplifier with very good performance of low noise 0.3 μVp-p between 0.1 and 10 Hz, and low drift 0.5 μV/°C.

7.5.4 Actuator drives

For the DPT actuator, the demand signal from the controller is input to the AX100 drive (from Queensgate Instruments) in which a high voltage signal is generated to drive the piezoelectric actuator and move the tip to maintain the required tunneling current. Details of the AX100 can be found in its manual.

For the spring flexure translator, the demand signal from the controller is input to the current drive amplifier. The circuit is similar to the current drives used for the Talysurf 5 with a little modification to the specified requirement. The details are shown in Fig. 7.10. The pre-set current is provided via a reference circuit which is powered by a pair of batteries. The current flowing through the coils (three coils for z-axis) is monitored via an OPA121 with a gain of 10. It is proportional to the topography of the surface being measured.

7.6 Analysis of Control Loop

7.6.1 Mathematical analysis

The control process, incorporating three term PID control can be schematically depicted in Fig. 7.11. The primary function of the control system is to maintain a constant tip-sample spacing, that is to make (Z_i - Z_0) independent of the sample height Z_i and the position of the tip actuator Z_0. Thus the tip will follow the topography of the surface. The following relationships can be written:

$$
\begin{aligned}
&I(t) \propto = K_1 V_T \exp(K_0 h(t)), \\
&h(t) = Z_i(t) - Z_0(t) \\
&K_0 \approx -1.025\sqrt{\bar{\phi}}(eV)^{-1/2}\ \text{Å}^{-1}
\end{aligned}
\qquad (7\text{-}6)
$$

$$
\begin{aligned}
&V_1(t) = R_f I(t) \\
&V_2(t) = K_2 V_1(t) \\
&V_3(t) = K_3 Log\left(\frac{V_2(t)}{E_r}\right) \quad K_3 = 1
\end{aligned}
\qquad (7\text{-}7)
$$

where R_f is the feedback resistance of the I/V converter, V_T is the tip bias voltage, $\bar{\phi}$ is the average work function of the tip and the specimen, and E_r is the effective reference voltage from the logarithmic amplifier, E_r = 100 mV and K_i (i=0, 1, 2, 3) are constants. Then

$$V_c(t) = K_p\left(V_3(t) + \frac{1}{R_I C_I}\int V_3(t)\,dt + R_d C_d \frac{dV_3(t)}{dt}\right) \tag{7-8}$$

$$Z_0 = K_A\ V_c(t) \tag{7-9}$$

7.6.2 Steady state response

Equation (7-9) may be differentiated to give

$$\frac{dZ_0}{dt} = K_A K_p\left(\frac{dV_3}{dt} + \frac{V_3}{T_i} + T_d\frac{d^2V_3}{dt^2}\right) \tag{7-10}$$

$$V_3 = Log\frac{K_1K_2R_fV_T}{E_r} + K_0K_2R_f(Z_i - Z_0) \tag{7-11}$$

where $T_i = R_I C_I$, $T_d = R_d C_d$. The steady state solution of equation (7-10) is

$$(Z_i - Z_0) = \frac{-1}{K_0K_2R_f}Log\frac{K_1K_2R_fV_T}{E_r} \tag{7-12}$$

and therefore the tip-specimen gap will be independent of the sample height and the displacement of the actuator. This is due to the use of the integrator in the control loop.

If a low-pass RC is used instead of an integrator, the steady state solution for the response of the resulting control system is

$$Z_i - Z_0 = -\frac{K_A Log\frac{K_1K_2R_fV_T}{E_r} - Z_0}{K_AK_0K_2R_f} \tag{7-13}$$

where the tip-sample spacing is a function of the displacement of the actuator. There is a steady state error in (Z_i - Z_0) of the magnitude Z_0 / (open loop gain). Thus the tip-sample spacing will change during a scan for a sample with a large height variation.

The use of an integrator can lead to problems if the STM is used in a spectroscopic mode for which the control system is momentarily disabled. The integrator may drift while the control system is disabled.

7.6.3 Parameter tuning and transient response

The transient response of the tunnelling probe depends on the setting of parameters of the control loop, and it is limited by the resonance of the system. The tuning of the PID controller is very important for the quality of the control action, and therefore for the accuracy of the measurement. The resonant frequency of the actuator either the DPT or the spring flexure, which would normally be low (especially for spring flexure), limits the response time of the feedback control loop. Generally speaking, increasing P, the proportional action, reduces the transient and steady state errors but introduces increasingly unacceptable overshoot, while increasing integral action, I, removes steady state error but a strong integral action (that is a short integral time) leads to system instability. Derivative action, D, is useful when fast response of the feedback is required, especially when the controlled system, the actuator, has a dead time. Potentially the derivative action helps the stability of the feedback loop (acting as a damper) but in practice it tends to increase high frequency noise which is especially undesirable for the measurement. Therefore most STM systems employ no derivative action. Here the derivative action is provided only to cover any unusual situation that may arise and also to investigate whether the performance of the tunnelling probe can be improved.

Determining by trial and error the optimum values of controller gain, integral time, and derivative time for a particular system would be very time-consuming because of many possible combinations of settings. In practice, values close to the optimum can be obtained quite easily by testing the closed-loop system with only proportional action on the controller. This can be done by setting the integral time to infinity, and the derivative time to zero. A step input can be provided by a step change in set point, or in this case by changing the gain setting of K_2. Then adjusting the proportional, P, results in oscillation with a constant amplitude at the gain K_{pmax}. The period of the oscillation is called 'the ultimate period' T_u. Usually only a few tests are needed to establish K_{pmax} and T_u, since the decay ratio of the first response curve shows whether the gain is close to or far

from the maximum value. Using the rules recommended by Ziegler and Nichols (1942), we have the following relations.

For P control,

$$K_p = 0.5K_{p\max} \tag{7-14}$$

For PI control,

$$K_p = 0.45K_{p\max}$$
$$T_i = \frac{T_u}{1.2} \tag{7-15}$$

For PID control,

$$K_p = 0.6K_{p\max}$$
$$T_i = \frac{T_u}{2.0} \tag{7-16}$$
$$T_d = \frac{T_u}{8}$$

Note that, with proportional-integral control, the recommended gain is 10 per cent lower than with only proportional control. Integral action makes the system less stable because of phase lag in the controller. When derivative control is added, the phase lead of the controller helps to stabilise the system and higher gain and lower reset time are recommended.

The above method is used for most applications where sustained oscillations for testing purposes are allowed. However, in STM operation, such an oscillation may damage both the tunnelling tip and the specimen. Therefore, a more conservative method is introduced here. By using only the proportional action and starting with a low gain, the gain is increased slowly until the transient response of the closed loop shows a decay ratio of $^1/_4$, that is a response with 2 to 3 peaks. The gain is then reduced slightly, and the integral and derivative actions are increased until overshooting is just ceased in its step response. To avoid any touching between a tip and a sample, a step input from the setting point was given that made the actuator move the tip away from the sample.

7.7 Experimentation

7.7.1 Manufacture of scanning tips

Tunneling tips were prepared in two ways. Tungsten tips were electrochemically etched from a 1 mm diameter wire by masking all but a small portion and immersing it in a 1 N potassium hydroxide solution. A carbon rod electrode was also placed in the solution and a constant dc voltage applied between the two resulting in the gradual erosion and eventual detachment of a sharp tip pulled from the end of the wire by its own weight. These tips were then plasma coated with a nominal 5 nm or 10 nm gold film. Palladium tips were produced from 250 μm diameter wire by mechanical cutting with a pair of surgical scissors.

The tip preparation by electro-chemical etching procedure frequently produces reasonably good probes with tip dimensions of a few tens of nanometres. Many reporters, for example Lbe *et al.* (1990) have found that the length (or the weight) beneath the etching zone (i.e. the zone exposed to the electrolyte) plays an important role in tip formation. The longer the length of wire in the solution, the larger the radius of the tip. At the final stage of the etching, the tensile strength of the neck can no longer hold the lower part and breaking occurs. This results in some tips having shapes far from the ideal, with micro-cracking, bending and recoiling caused by residual stresses. Scanning electron microscope (SEM) examination of tips etched for this study shows this behaviour, as seen in Fig. 7.12 (a,b,c). However these undesirable effects can be minimised by carefully managing the etching current by control of the dc supply voltage and the concentration of the solution to affect the shape and the radius of a tip. Tungsten tips produced by electrochemical etching are normally covered by a layer of contamination, mainly solid carbon and oxides, which should be removed prior to plasma coating. More consistent performance was obtained for tips first immersed in an industrial solvent.

The mechanically cut tips are easily prepared and, although variable in shape, they can provide high resolution images on relatively smooth samples (Musselman and Russel 1990). The cut tips normally have larger radius than etched tips, although it appears that there are commonly sub-

tips on the end. Fig. 7.12 (d) shows a micrograph of the palladium tip after being used for experiments.

7.7.2 Preparation of specimens

Two different sets of specimens were prepared. Both consisted of pieces of 'Syton' polished silicon wafer as substrates. One set was coated with between 10 and 20 nm gold, and the other with sequential depositions of 1000Å films of copper, palladium, and platinum using electron beam evaporation (provided by J. Miller of the University of North Carolina at Charlotte, NC, USA).

7.8 Results

7.8.1 Scanning profiles of gold coated silicon by STM_1

The tunnelling tip was produced from 1 mm diameter tungsten wire by electrochemical etching from 1 N KOH solution by applying a DC voltage of 12 V with a variable current. The tip was then flushed with distilled water and cleaned with a solvent called 'Micro'. A gold film of 2 ~ 5 nm was then coated onto the tip. The specimen was a portion of a Syton polished silicon wafer covered by a gold film, approximately 20 nm thick. Setting up the instrument is a relatively quick procedure. The specimen is initially moved to within approximately 0.2 mm of the probe by hand and clamped in position by a locknut, see the Fig. 7.4. Further adjustment is then achieved by manually adjusting the differential micrometer until a tunnelling current is monitored. This procedure takes approximately ten to fifteen minutes. The whole system is then left under control to settle for about half hour whereby the instrument is ready for use.

Fig. 7.13 shows a plot of both the specimen platform drive current and subsequent servo feedback to the probe drive against time. These have been converted into units of displacement from the calibration curves of figure 7.5. The results shown in figure 7.13 are typical of a large number of tests but have been chosen to indicate some of the characteristics of this system. The parameter

settings for these results are: tunnelling current of 0.2 nA, tip voltage of -104 mV and specimen traverse rate of 0.35 nm s^{-1}. The probe has traversed a distance of 35 nm and then reversed across nominally the same path. For a perfect trace, a mirror image would be expected. However, it can be seen that there is a slow specimen axis drift of approximately 3 nm giving a drift rate of 0.02 nm s^{-1}. This is probably caused by twisting of the specimen drive support which, being a fabricated flexure spring, is known to possess inherent instabilities.

The probe axis drift which will result in a vertical shift of any features is of much reduced magnitude and in this case is only a few nanometres over a complete traverse. It has been observed that the probe noise reduces after a large number of traces, the reason for this is, as yet, unknown but is likely to be the removal of insulating contaminant films or damage induced by probe contact. A further profile at a tunnelling current of 10 nA, and a tip voltage of -260 mV over the different portion of the same surface is shown in Fig. 7.14. The forward and reverse profiles over a longer scan range of 800 nm are plotted on the same axes for comparison. It is clear that the probe follows the same overall profile although some of the finer scale features do not repeat at the nanometre level. It should, however, be noted that there is only a 4 nm discrepancy in the vertical scales between the forward and reverse traces after a scan time of approximately five minutes. This drift is commonly encountered as an artefact of the instrument design. The main cause is likely to be due to thermal expansion of the manual adjustment micrometer although it was found that physical contact between the probe and specimen during the initial approach would sometimes cause an increase in the drift rate. It is thought that this might have been due to some additional instability of the spring stresses in the various elements of the fabricated flexure stage (a common problem at this level, Jones 1967).

7.8.2 Performance evaluation of STM_2

STM_1 was mounted on a large concrete block which sits on a pneumatic isolation stage to attenuate ground borne vibration. STM_2 has to be used with the Nanostep slideway. The slideway base is a granite epoxy moulding block which is, in turn sitting on rubber feet on a solid table. These

precautions are adequate for the Nanostep probe but not good enough for the tunnelling probe due to its higher sensitivity. An air damping (air reservoir) table was not available so the system had to be operated in a quiet period time, for example, in evenings when most machines and ventilations are switched off and there are no people walking around. Another way to reduce the vibration is to narrow the bandwidth of the system, and this has been achieved by setting the time constant in the DPT drive AX100. On the other hand, the high resonance of the DPT makes the system less sensitive to vibration than that STM_1. The cover of the Nanostep is made from glass-reinforced plastic which provides damping for airborne noise but is inadequate for screening purposes. Therefore, another cover made of aluminium was added. The DPT and pre-amplifier were also screened in a small tin-plate box.

Fig. 7.15 shows the output signal from the DPT and its power spectral density about 20 minutes after switching on the instrument. The tip was maintained at a constant tunneling current of 10 nA without scanning the sample. At low frequency, the graph shows a 1/f type noise spectrum which is related to the tunneling current, and several peaks at about 12, 47, 50 and 76 Hz which were associated with the ground vibration and electro-magnetic disturbance. Better vibration isolation between the instrument base and ground and a better screening of the whole instrument may further reduce this noise level, but the limit will be determined by the DPT. A drift of 0.06 nm s^{-1} was recorded in this test, but it normally settles to better than 0.02 nm s^{-1} by one hour after setting up.

To assess the system repeatability, cyclic measurements over a scan length of 1.4 mm at a surface velocity of 5 μm s^{-1} were obtained by traversing the sample in forward and reverse directions. Digitisation of the profiles was performed at equal time intervals, taken to represent adequately equal spatial intervals since the drive is nominally constant speed. Figures 7.16 (a and b) show a typical measurement with a tungsten electrode and a gold coated silicon specimen. A tip to sample bias of 200 mV was used with the tip controlled to maintain a constant tunneling current of 10 nA. There is a clear match between many of the surface features and a consistent slope of approximately 7.2 minutes of arc between the specimen surface and the line of action of the

slideway. The relatively large features on this particular specimen were probably introduced during preparation of the silicon substrates which were obtained by breaking a large silicon wafer: slight damage could be seen with the naked eye on some rejected pieces. Several traces were made at the same location and three typical results are shown in figure 7.17, in which profiles are plotted with a vertical shift for clarity. The horizontal axis is plotted in scan time since each trace shows a traverse, dwell and reverse traverse. Agreement vertically is excellent. There is a consistent drift in the horizontal axis magnification which appears to be due to a slight variation in traverse speed. Most of the apparent discrepancy between features in the traces will be seen, on closer examination, to be caused by their consequent lateral misalignment.

Closer sampling intervals and increased lateral resolution were obtained by introducing a higher ratio gearbox in the micrometer drive mechanism of the Nanostep to reduce the minimum traverse speed from 5 $\mu m\ s^{-1}$ to 312.5 $nm\ s^{-1}$. Further experiments were carried out using a palladium tip and platinum specimen surface. At this speed a duration of 5.34 minutes was required for backlash to be removed between driving in the forward and reverse directions. To avoid the recording of redundant data, data logging was suspended during this period of time. A forward and reverse trace over a 4.69 μm length of the surface are shown in figure 7.18. Because of the pause while backlash was taken up, there is no true horizontal register between the forward and reverse profiles. However, closely agreeing features are present in both and there is high confidence that they correspond. The repeated measurements revealed a raised portion on the surface around the start point of the scan denoted by region A in the figure. This region is plotted for both traverse directions in figure 7.19 (a,b), which shows the original profile and that after smoothing with a 30 Hz cut-off digital filter. A more representative profile taken from the region B of figure 7.18 is shown in figure 7.19(c), again with the 30 Hz filter. This shows small variations with an average amplitude 1.5 nm from peak to valley and a period of around 30 nm. A repeatability of better than 0.2 nm can be seen from this measurement cycle.

The shorter range profiles are comparable with those obtained using a Digital Instruments Nanoscope on the same platinum specimen and with a similarly produced probe with a sample bias

voltage of 750 mV and a constant tunneling current of 0.1 nA (images and profiles provided by J. Miller, UNC). Measurements of a single profile are shown in figure 7.20 where (a), (b), and (c) were taken with one tip and (d) was taken with a sharper tip. The profiles show roughly spheriodal grains with characteristic heights between 5 and 10 nm and characteristic diameters of around 30 nm. Examination of the first profile shows lower slopes to the left of peaks and higher slopes to the right. This asymmetry suggests that there is a similarly asymmetric slope each side of the tip or, equivalently, that a symmetric tip is inclined to the surface. This is partially confirmed from SEM micrographs of the probe. After slightly bending the probe, the profile (c) was taken and shows symmetric profiles around the spheroidal particles and better definition in the valleys between them. By trial and error, sharper tips could be found, eventually resulting in the profile shown in figure 7.20(d) which reveals more detailed surface features giving height and period variations of 5 to 15 nm and 20 to 30 nm respectively. Obviously, poor tips produce lower heights and broader features because of the convolution between tip and sample. This is of considerable importance if they are to be used for metrological purposes, and some form of standardisation will be needed. The research on the standardisation of scanning tips has been reported recently by a number of authors (Nyyssonen *et al.* 1991, Griffith *et al.* 1991, Vasile *et al.* 1991).

7.9 Digital Filtering

Most STM data require digital processing such as filtering or smoothing because of the high frequency noise involved in the data. High frequency noise comes from electronic circuits, data acquisition, and significantly from improper tuning the parameters of the closed feedback loop. To avoid excitation of the system resonance, there are two factors to be considered. One is that there is no signal with a phase shift reaching -180°, that is no positive feedback signal involved within the frequency bandwidth. If this fails then the second has to be guaranteed which is that the signals at or near the resonance must be so small as not to cause any significant change in the system. It is easy to state these two rules but, in practice, it is difficult to maintain one or another. Narrowing the

bandwidth of the control loop and scanning at a slow speed can reduce the noise to a certain extent but this is not desirable because other influences such as thermal drift may become significantly. Therefore, there is always a compromise between the scanning speed and the resolution. By applying a proper filtering technique to the STM data, a better resolution can be obtained.

There are four steps to implementing the digital filtering process used here, all executed in 'Matlab'. The first is to obtain the spectrum of the data, a profile in our case, through the FFT (fast Fourier transform). All that needs to be known about the data is its sampling frequency. Before taking any measurements, the sampling frequency has to be set properly to avoid signal aliasing, a problem caused by sampling too slowly. The upper limit on the frequencies that should be contained in the data or digital signal is half of the sampling frequency, that is the Nyquist frequency. By examining the spectrum of a profile, one should be able to determine the cut-off frequency although this always requires some experience accumulated by comparing the results before and after the filtering. Second, the pass band and the stop band windows are designed. Because a low-pass filter is wanted, the pass band is from DC to the cut-off frequency and the stop band is above the cut-off frequency. Third, an appropriate form and order of filter must be chosen from those available in Matlab. There are two main types of filters, FIR (finite-impulse-response) and IIR (infinite-impulse-response). They have different characteristics, yet they can often meet the same specifications. The FIR filter has a transfer function with an all-zero filter in a polynomial form, that is, the denominator of the transfer function is equal to 1, while the IIR filter has both zeros and poles. The combination of a pole near the pass-band edge with a zero near the stop-band edge can give an IIR filter a very short transition region between the pass-band and the stop-band. Generally, an IIR filter can give a sharper cur-off than an FIR filter of the same order because both poles and zeros are present. However, an IIR filter cannot achieve exactly linear phase shift while FIR filter can. The IIR filter is a recursive filter which requires less memory than an FIR filter. On the other hand, the IIR filter, when implemented in fixed-point arithmetic, may have instabilities (limit cycles) and may have large quantisition noise, depending on the number of bits allocated to the coefficients and the signal variables in the filter. In our applications, a sharp cut-off is important and so the IIR type

filter is used. Matlab supports four types of IIR filters: Butterworth, Chebyshev (I and II), elliptic, and Yule_Walker. Since the filters are optimal in all cases, it is necessary to understand in what sense they are optimal. The Butterworth filter is optimal in the sense that it is the best Taylor series approximation to the ideal low-pass filter magnitude at both $\omega = 0$ and $\omega = \infty$, which has maximally flat passband and stopband. The Chebyshev filter gives the smallest maximum magnitude error over the entire passband. Chebyshev I filter has ripple in the passband, Chebyshev II or inverse Chebyshev filter has ripple in the stopband. The elliptic filter has ripple in both passband and stopband, but for a given filter order, the elliptic filter has the sharpest transition of all these filters. Chebyshev filters have a sharper transition than a Butterworth filter with the same design specifications. The Yule_Walker filter can be used to design an arbitrarily shaped, possibly multiband, frequency response. Note that all of these filter designs are magnitude approximations and do not address the phase frequency response. Among these filters, the Butterworth filter has the smoothest phase curve followed by Chebyshev II, Chebyshev I, and the elliptic filter. Finally, the filter function is implemented and the data is processed. After filtering, the result is compared with the raw data. The first a few points in the result are not valid, due to the uncertainty in the initial values of the filter. The number of the invalid data points depends on the filtering order and the type of the filter, and it is easily identified by its large variation. For a Butterworth filter with an order of 5, the number of the invalid data points is no more than 20. Details of the program for filtering are given in Appendix B.

7.10 Summary

STM_1, a profilometer driven by force actuators via linear springs, has prove to be adequate for nanometre level control and measurement. The linear spring actuator does not have significant drift or hysteresis and thus has potential for accurate surface metrology. Moreover, the tip position is directly proportional to the current in the coils. Variations in the coil current should result in a known change in tip to specimen separation. Thus modulation of tip-specimen separation can be

used to diagnose the presence of contaminants between the surface and the probe. For example, an insulating contact will cause an apparent lowering of the measured value due to the added stiffness of the interface (Coombs and Pethica 1985). In reality, it is likely to be very difficult to measure changes in the interface stiffness for such small forces and displacements using the present design. The major limitation of this present design is that of its large size. This results in a low dynamic response and a rather high thermal susceptibility. However, these are not fundamental limitations. In the extreme, the springs could be manufactured from single crystal silicon which would be approximately one quarter of the present scale and the system resonance would be much increased. Research on x-ray interferometers in this Centre has achieved, using a single crystal silicon spring, a 10-μm range with controllable sub-Ångstrom resolution (Chetwynd *et al.* 1993). The materials of the prototype built here were chosen for low cost and easy machining.

The STM_2, driven by a digital piezoelectric translator (DPT) and traversed by a precision slideway, provides a long scanning range of 50 mm and a vertical positioning range of 15 μm. The high resonance of the DPT allows the system to be operated at a higher scanning speed and with relatively poor isolation of vibration. The exceptionally low drift of the slideway and the DPT makes the long scanning distance possible. The results over a distance of 1.4 mm measured here is not the fundamental limit, potentially it can go as long as 50 mm. But at present the computer's memory size and the flatness of a sample limit its potential applications.

Compared with the STM_1, the STM_2 is more convenient to use, less demanding on surroundings; sometimes good results can be obtained during the working hours. A unique feature of STM_2 is its long scanning range. On the other hand, STM_1 has its own character, without relying on commercial parts, and can be optimised to give better performance.

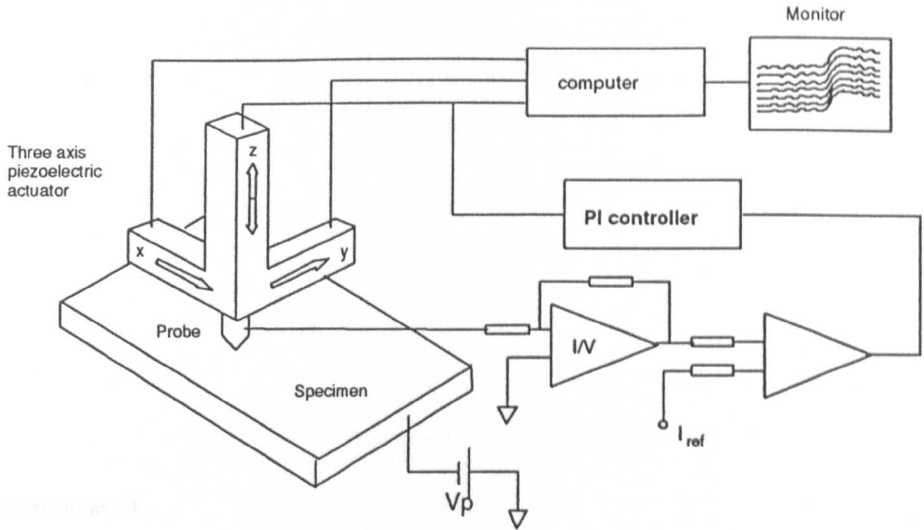
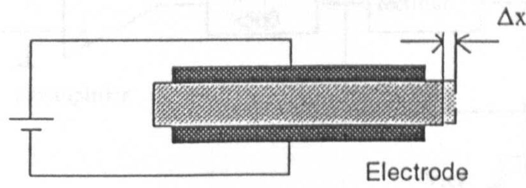
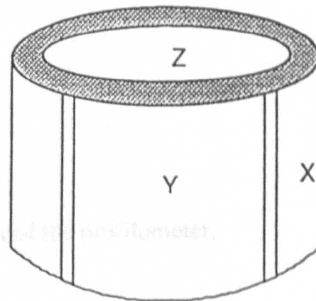


Fig. 7.1 Schematic of a typical scanning tunnelling microscope.



(a)



(b)

Fig. 7.2 (*a*) Piezoelectric effect, and (*b*) a typical piezoelectric tube scanner.

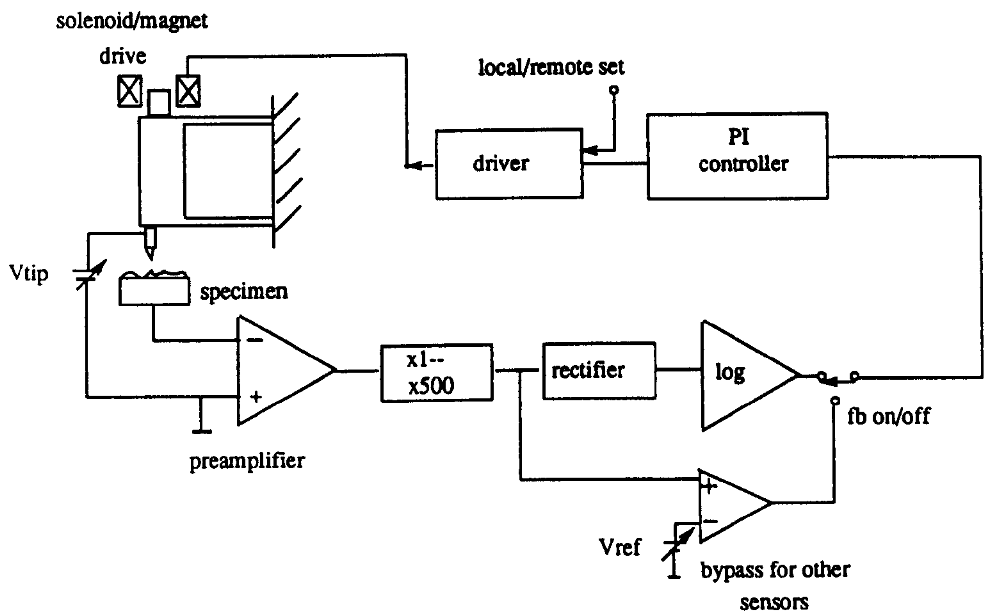


Fig. 7.3 Schematic diagram of the profilometer.

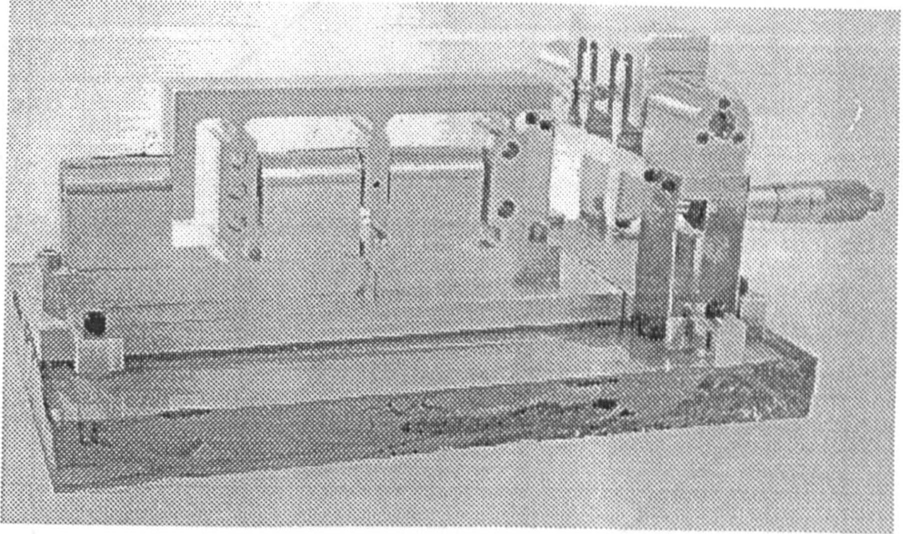
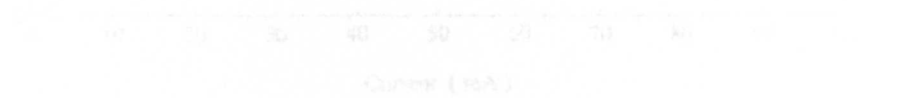


Fig.7.4 Photograph of the complete profilometer.



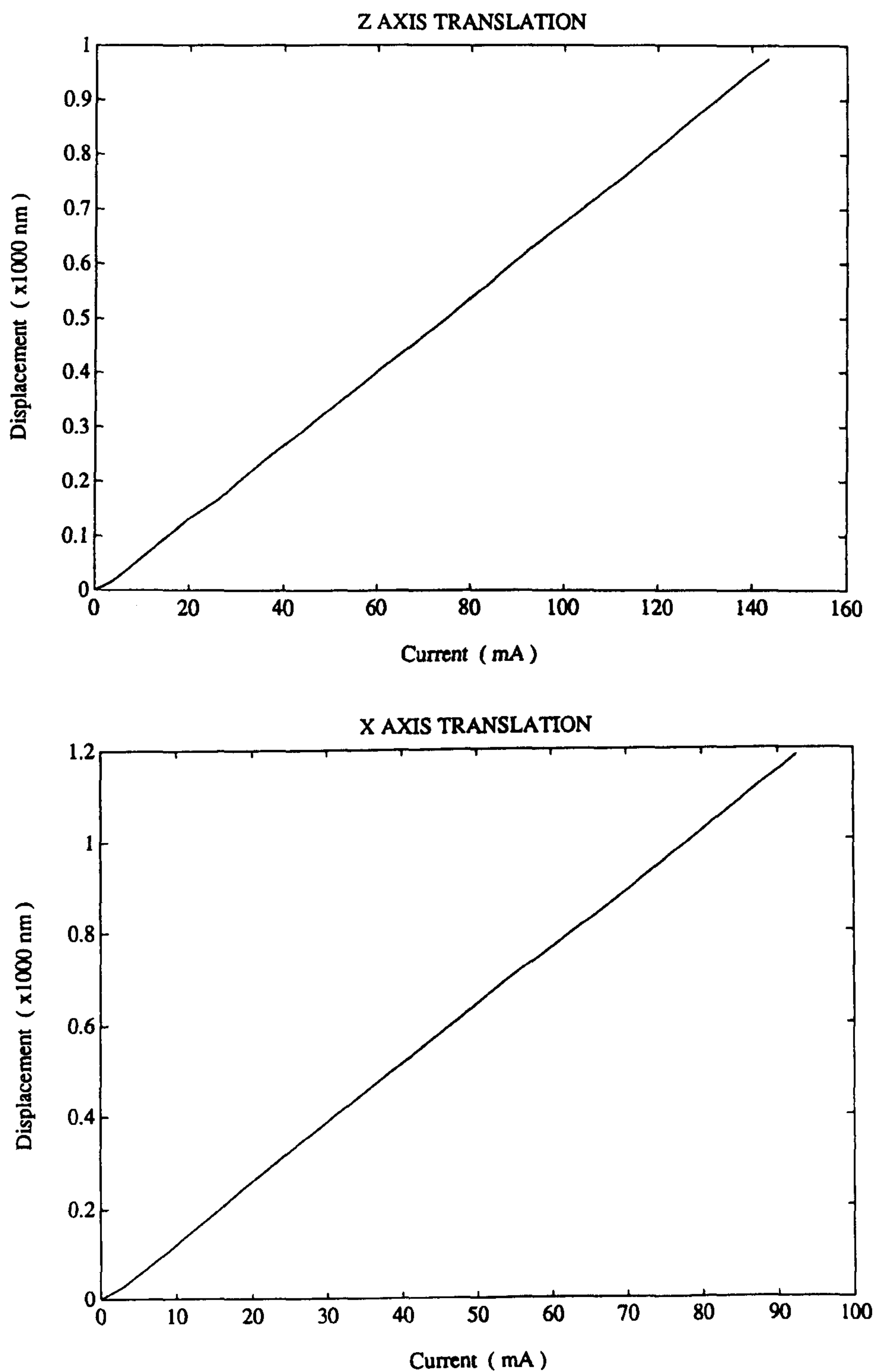


Fig. 7.5 Measured displacement using LVDT gauging of platform displacement against the actuator coil current; (*a*) Probe axis; (*b*) specimen axis.

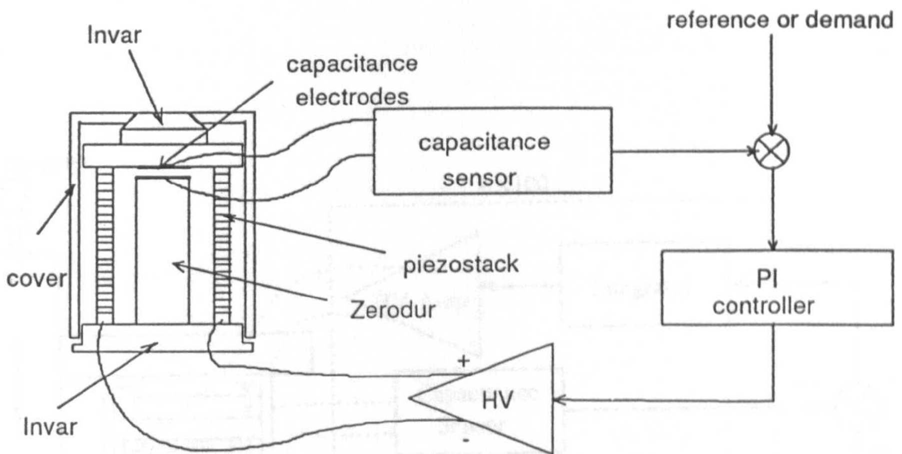


Fig. 7.6 Piezoelectric actuator with inbuilt capacitance position sensing (from Queensgate Instrument).

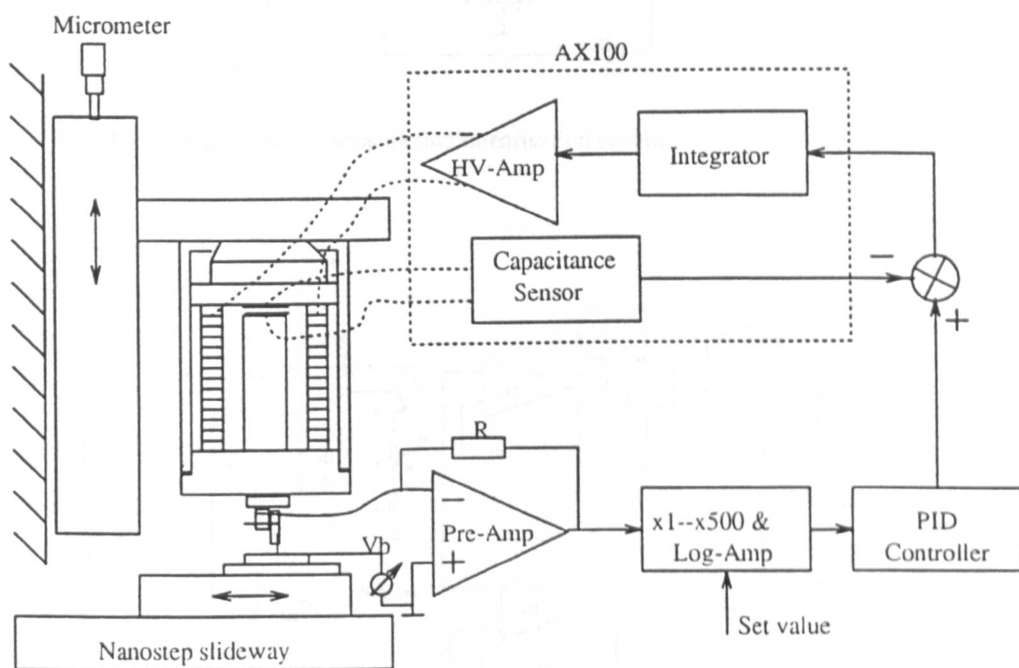


Fig. 7.7 Schematic of the long range tunnelling profilometer.

Fig. 7.8 Diagram of the sensing and linearisation circuits.

Fig. 7.9 Circuit diagram of the PID controller.

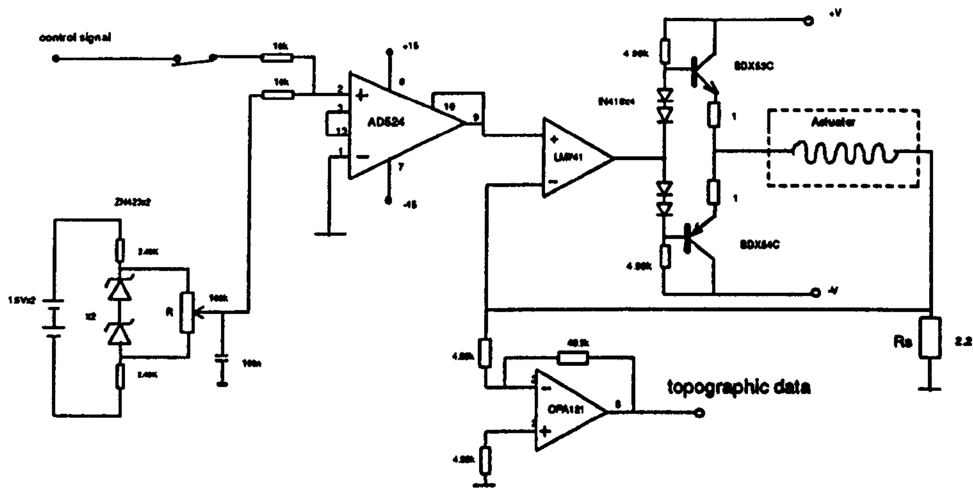


Fig. 7.10 Current drive circuit for spring flexure actuator.

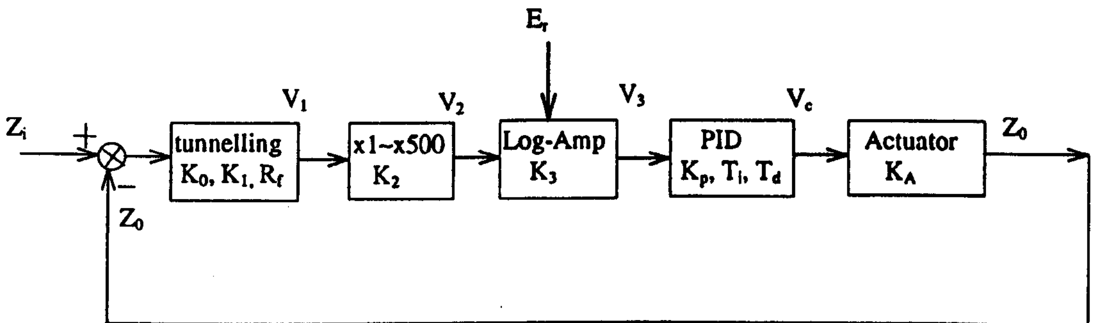
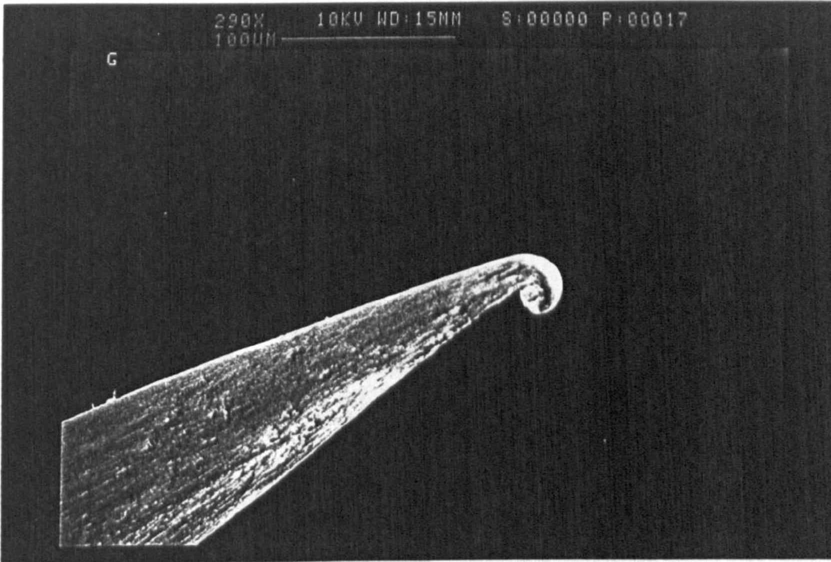
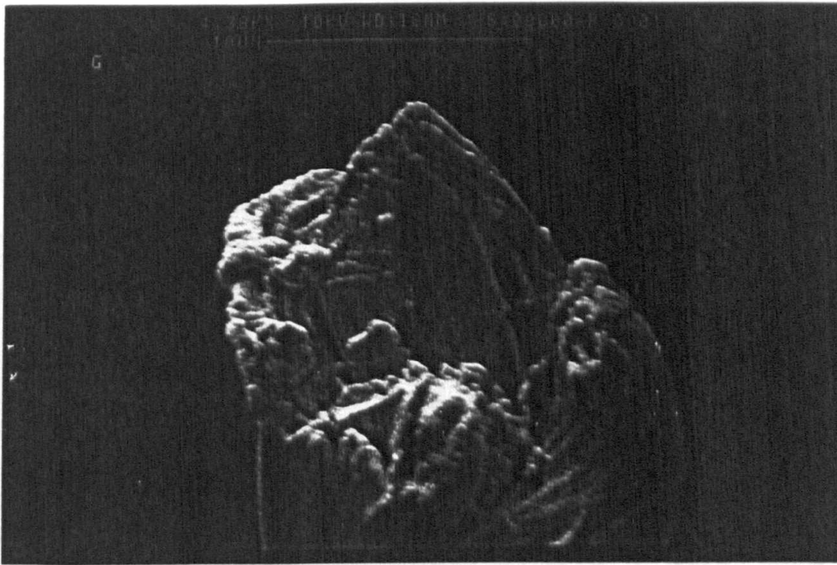


Fig. 7.11 Block diagram of the STM control system.

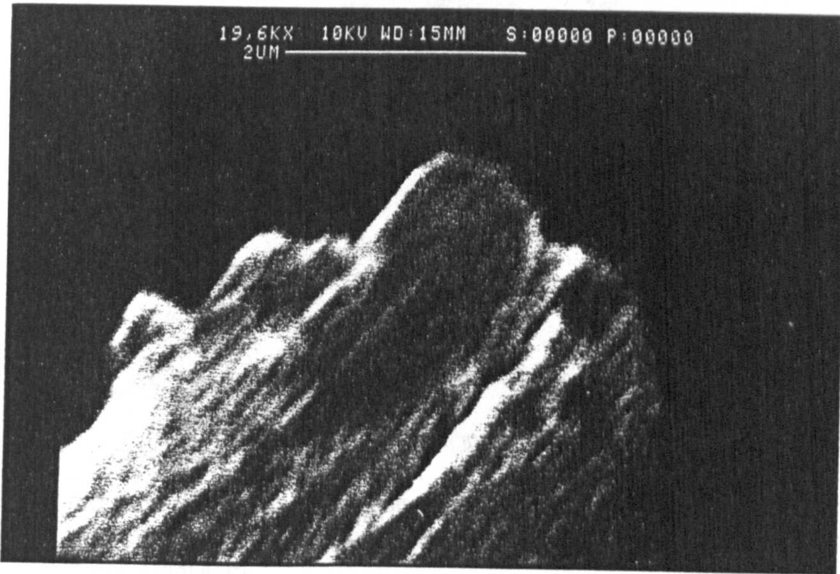


(a)



(b)

Fig. 7.12 Micrographs of tips prepared in two ways: (*a*) a tungsten tip with a hoop-like end, (*b*) a tungsten tip with a cracked end, (*c*) a more typical tungsten tip, (*d*) a palladium tip.



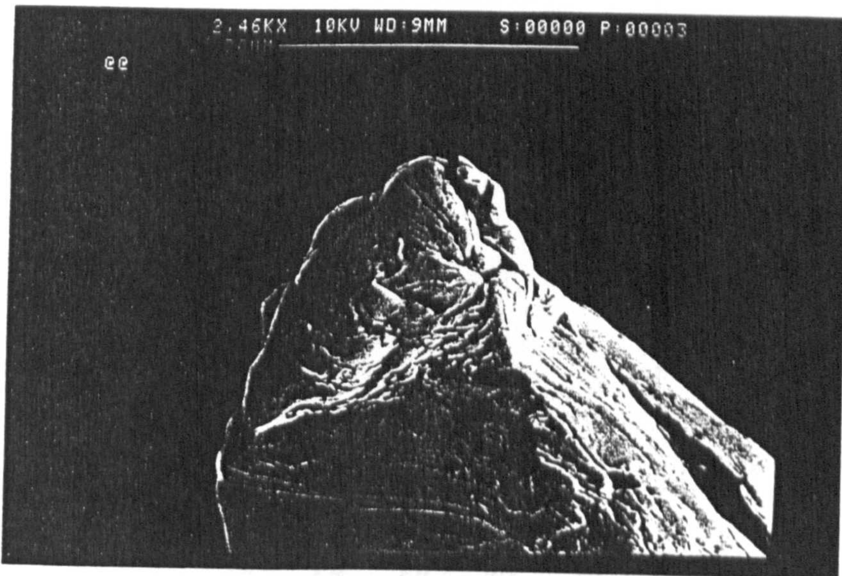


Fig. 7.12 (continued) (*c*) and (*d*).

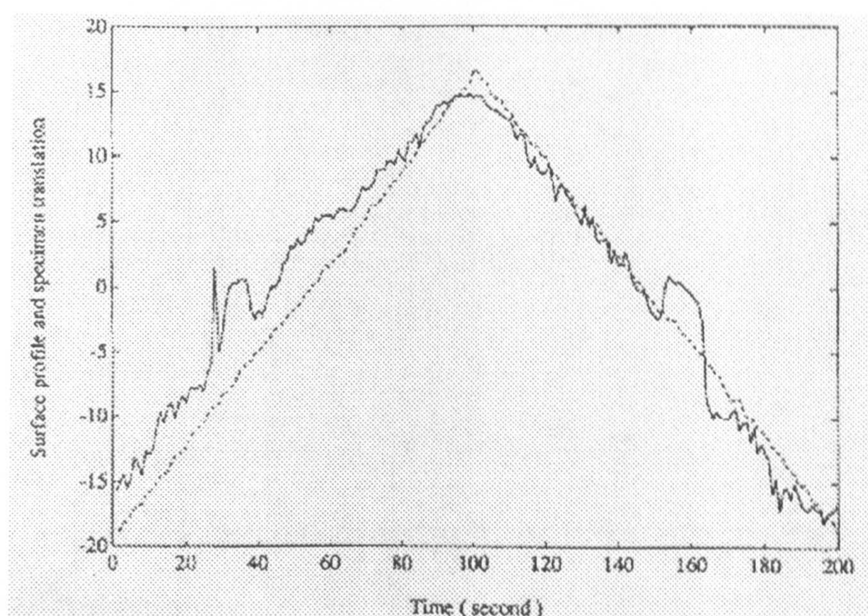


Fig. 7.13 Forward and reverse displacement of probe and specimen platforms as derived from the calibration curves of figure 7.5: full curve – probe axis; broken curve – specimen axis.

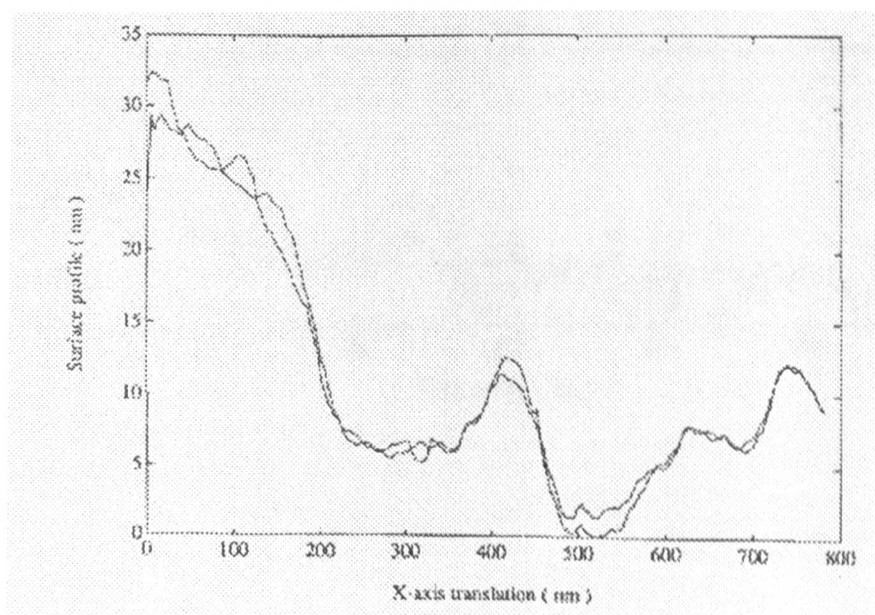


Fig. 7.14 Profile measurement of a gold surface.

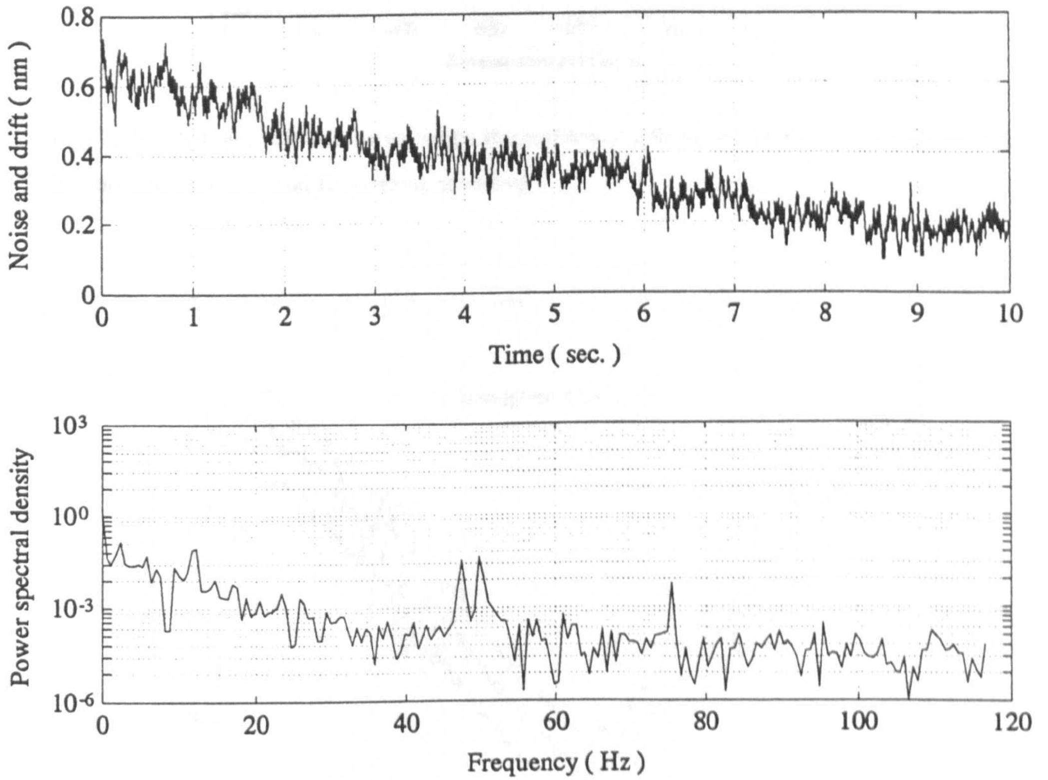


Fig. 7.15 Noise characteristic of the tunnelling current at I_t = 10 nA: (*a*) non-scanning tunnelling signal, and (*b*) its power spectral density.

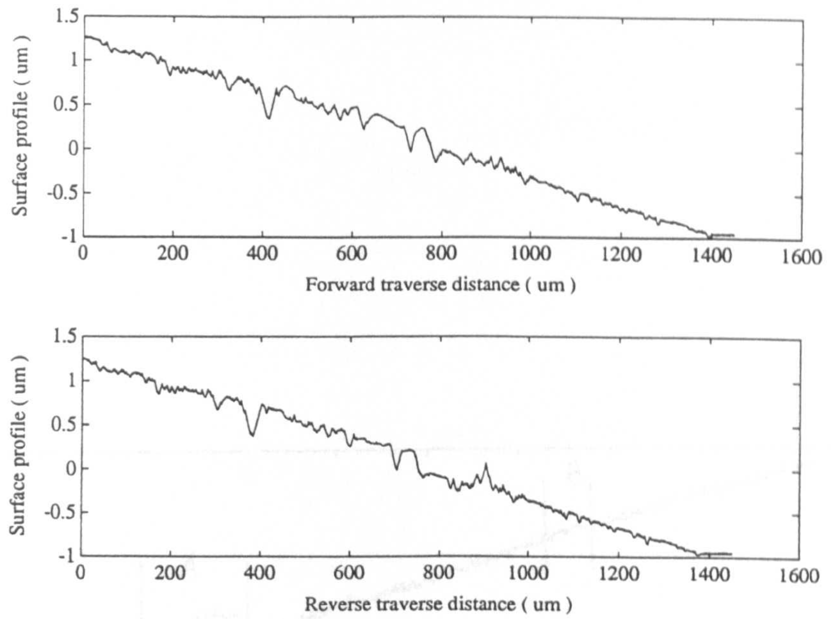


Fig. 7.16 A long cycle measurement on the surface of gold-coated silicon by a tungsten tip: (a) forward scanning and (b) reverse scanning.

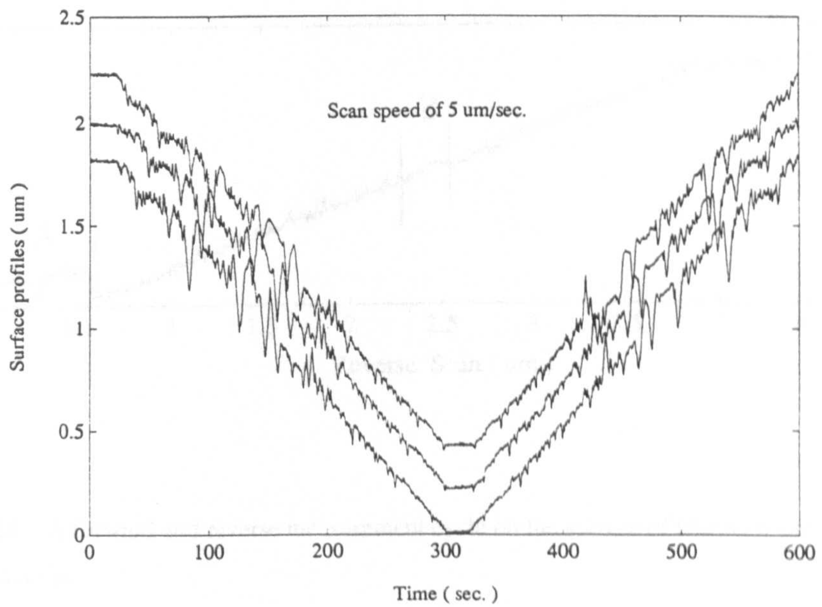


Fig. 7.17 Three repeated measurements taken at the same position on the gold-coated silicon sample, plotted with artificial vertical shift.

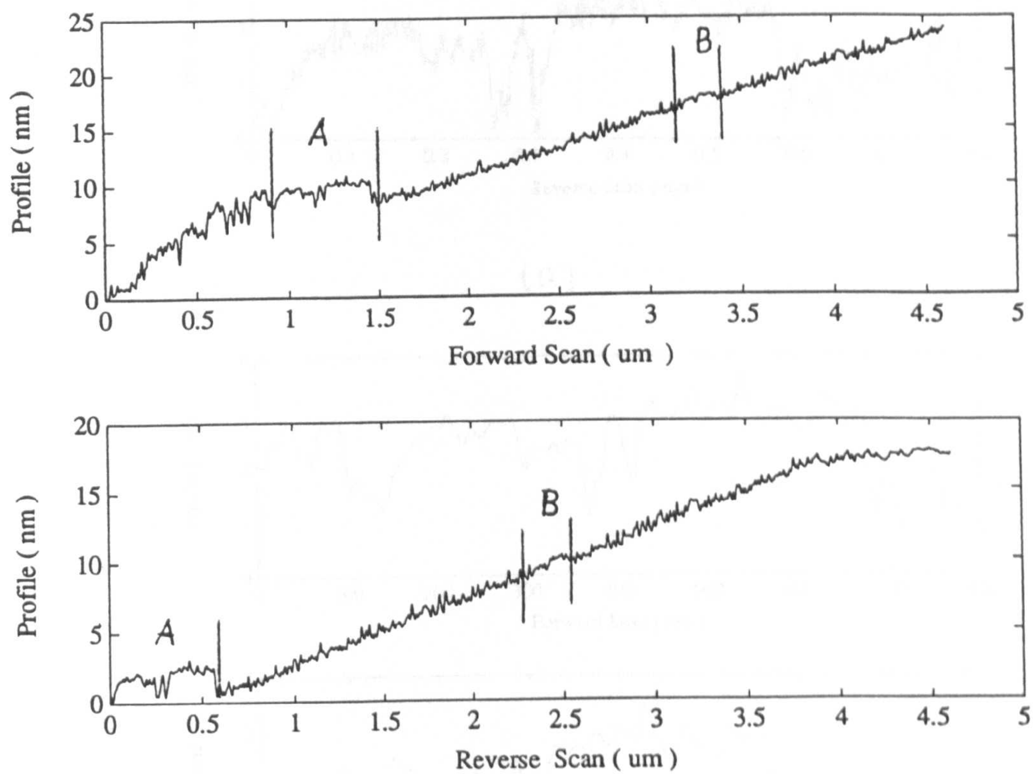
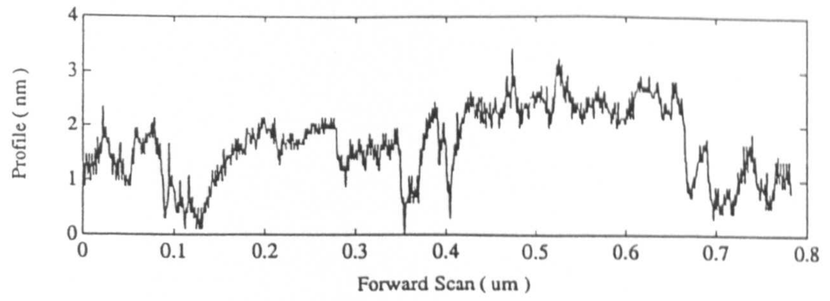
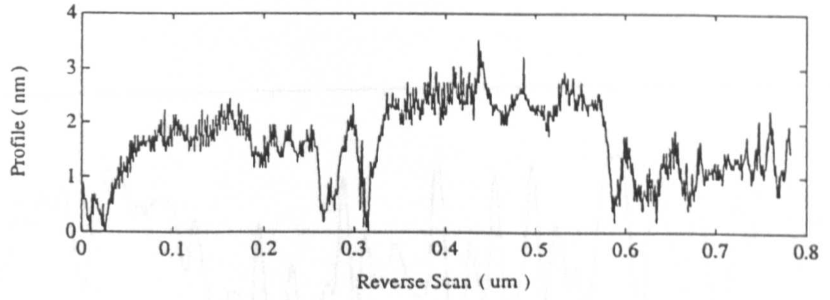
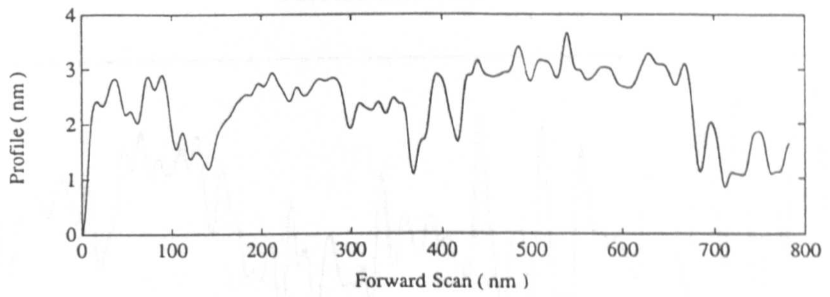


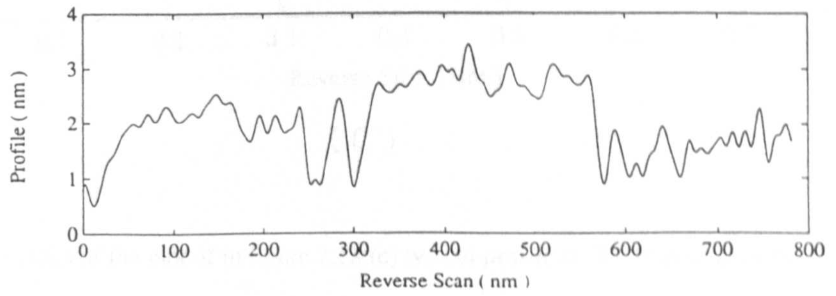
Fig. 7.18 A forward and reverse measurement cycle on the surface of platinum using a palladium tip.





(a)





(b)

Fig. 7.19 Details of the plot of in figure 7.18: (*a*) hill 'A' in original form, (*b*) filtered at 30 Hz.

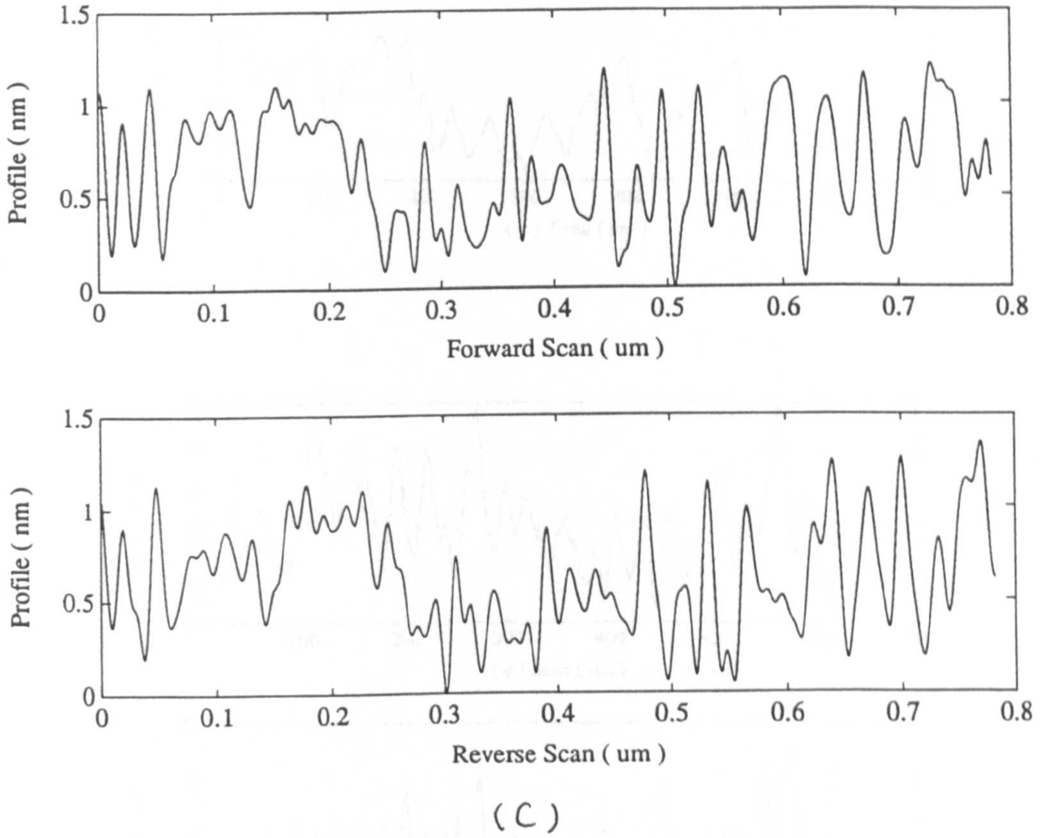


Fig. 7.19 Details of the plot of in figure 7.18:(c) typical profile at 'B', filtered at 30 Hz.

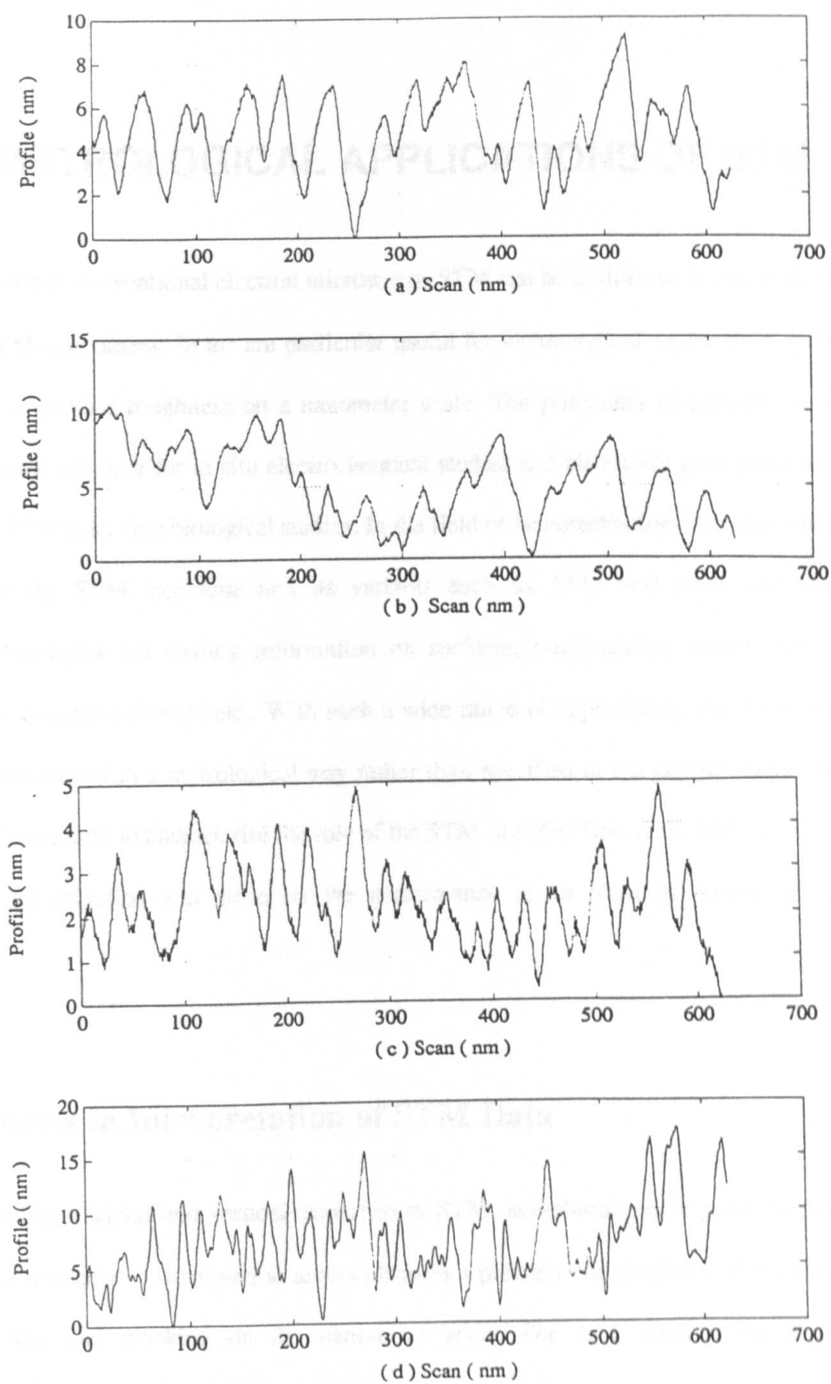


Fig. 7.20 Surface profiles of platinum specimen: (*a*), (*b*), (*c*) repeats using one palladium tip and (*d*) with the palladium tip shown in figure 7.12 (*d*).

8. METROLOGICAL APPLICATIONS OF STM

In contrast to conventional electron microscopy, STM can be performed in air, in liquids and in vacuum. STM experiments in air are particular useful for technological applications such as the determination of surface roughness on a nanometre scale. The possibility of performing STM in liquids has opened the door for in situ electrochemical studies and also holds great promise for the application of STM to in vivo biological studies. In the field of Nanotechnology dramatic change has been made by the STM technique and its variants such as AFM and other scanning force microscopes. Examples are writing information on surfaces, manipulating matter down to the atomic scale in a controlled way, etc.. With such a wide range of applications, the STM technique needs to be charaterised in a metrological way rather than specified in the physics study. Attempts have been made recently to characterise the role of the STM tips (Griffith, *et al.* 1991, 1992). Here a more general investigation was made on the performance of an STM in normal engineering environment.

8.1 Problems in Interpretation of STM Data

The combined lateral and vertical resolution of STM (sub-atomic under good conditions) is now starting to be used to control and to assess ultra-high precision fabrication and to characterise surface roughness and cracking on the nanometer scale. For such applications a precisely characterised metrological framework is essential. Note that many scientific studies can be undertaken successfully with less precision in this respect because theoretically known patterns may provide a register for the measurements. Good metrology will also be needed for the reliable application of STM as a nanotechnological 'machine tool' or assembler. This aspect of the

instrument design has been recognised as important for a decade or more, but only in recent years has its study been pressing. Metrological interpretation of STM data, which contains both geometrical and electrical information, requires a separate knowledge both the current-voltage (I-V) and separation-voltage (S-V) characteristics. Detailed studies of tunneling characteristics in vacuum have been carried out in both constant current mode and constant separation mode (Binnig and Rohrer 1982, 1986). Recently more attention has been paid to the characteristic of STMs operated in air, following the observation of anomalies while imaging atomically flat surfaces of highly oriented pyrolytic graphite (HOPG) in ambient air (Coombs and Pethica 1986; Soler *et al* 1986; Mamin *et al.* 1986). HOPG has been used as a standard because it can easily be cleaved to give an atomically flat surface and is relatively inert in air. However, some peculiarities have been reported indicating that, in air, the tunneling current is only relatively weakly dependent on the separation distance of a tip relative from a graphite surface and that there are sometimes unexpected corrugations in images of up to 2 nm or more (Mamin *et al.* 1986). A bias voltage of 50 mV and 1.5 V was required for operating the STM in vacuum and in air, respectively. These anomalies were explained through a model of mechanical contact between the tip and the specimen via either a contaminate layer or a graphite flake. The contaminant layer on the surface increases the resistance of the tunnelling junction, therefore, a higher bias voltage has to be applied to get the required tunnelling current.

8.2 Characteristics of STM in Air

8.2.1 Relationship of *S-V*

To investigate the metrological capability of the STM system, the voltage sensitivity of the STM_2 was tested using the constant tunneling current mode. Again a gold coated tungsten tip and a similarly coated silicon specimen were used, and the bias voltage V_T was varied and the subsequent displacement of the probe monitored. In this mode it is not possible to pass through the zero bias

point and it was therefore necessary to investigate each polarity separately. The magnitude of the sample to tip voltage was ramped from 1.5 to 0.05 V and back again in each polarity at a rate of 0.05 V s^{-1}. Measurements were taken at constant tunneling currents of 10, 5, 2, 1, 0.5 0.2 and 0.1 nA which were obtained by changing the magnification K_2. Observations of the data, shown in Fig. 8.1 (a,b,c) indicated a tendency for discontinuous `jumps' of the tip position to occur, in some instances by up to 2 nm. There was distinct hysteresis, with the separation remaining relatively greater when reducing the voltage at constant current than when increasing it again. It is possible that this is an artefact of the actuator control, but experience of the quality of these systems in other applications suggests that it is more likely that it is caused by a tip-specimen interaction. The effect is shown in each plot of *S-V* (separation-voltage), which have been measured at constant tunneling currents of 10, 5 and 0.2 nA respectively. It is important to note that the separation is measured relative to that obtained with a bias of 0.05 V and that the tip was moving towards the surface as the magnitude of the bias voltage decreased.

Although the laboratory is controlled for temperature and humidity and is relatively clean, it is highly likely that surface contamination will occur under ambient conditions and that the contaminants will be insulators. It is widely held that tunneling occurs after the tip has contacted this layer and that the subsequent characteristics are dominated by the need to deform this layer to affect a change in the conductivity of the junction. The observed steps in the *S-V* curves suggest that increases in current may be due to the expulsion of material or fracture of the surface layer under varying loads. The large apparent displacement required for maintenance of the current could be caused by microscopic surface deformations of the contaminate layer and the specimen or by macroscopic displacements due to finite stiffness around the measurement loop.

The difference in the *S-V* characteristic with different polarity of bias voltage could be explained by tunneling at different bands (Freenstra *et al.* 1987; Berghaus *et al.* 1988). For a positive sample voltage the unoccupied states, and for a negative bias the occupied states, make

different contributions to the local tunneling current, which results in different absolute proximities between the probe and surface.

From the S-V curve, it is only possible to know the relative changes in separation: the change in position reflects a change in the tunneling current.

8.2.2 Relationship of *S-I*

The relation between the variation of the sample-tip distance and the tunneling current at the fixed bias voltage can be obtained by changing the set point value in the constant current mode. At each bias voltage, the current value was set to 0.5 nA and then increased to 1.0, 2.0, 5.0 and 10 nA, while the corresponding change in the tip position was recorded. A typical result is shown in Fig. 8.2. Each point represents the maximum displacement of the actuator when tunneling current changed over a range from 0.5 to 10 nA. This movement is different at each bias voltage. Interestingly, large displacements were found to occur between 400 and 500 mV at both polarities of bias voltage. Below 400 mV, the actuator displacements were relatively small, although still large compared to the classical tunneling range, with a sensitivity between the tunneling current and the tip-sample separation of 2 nA nm^{-1}. This implies that the contaminant layer may have been penetrated, resulting in a quite stable tunneling current over the same bias voltage range. Referring back the *S-V* characteristic, the rate of 2 nA nm^{-1} suggests that had the separation been held constant during such a voltage change, the current would have changed by 5 to 10 nA, which is a much smaller variation than that is observed in UHV.

In each series of experiments a duration of 20 minutes was allowed after the initial tunneling current was established with a complete series of tests requiring approximately two hours. Initial and final positions of the DPT corresponded to a 93 nm shift giving an average drift of 13 pm s^{-1} in a direction away from the surface. This is a considerable improvement on the drift rate of 60 pm s^{-1} shown in the Fig. 7.15(a) indicating a temporal dependence of stability after initial setting of the probe.

8.3 Mechanical Contact Model

The measured sensitivity of the tunneling current to the tip-sample separation of less than 2 nA nm^{-1} implies an extremely low barrier height between the tip and the specimen. As Binnig *et al.* (1984) predicted that the work function lowers by only a volt or so when separation is reduced to 4 Å, it is felt that the observed variations are best explained by the contact model. In this, the change in the tip position cannot reasonably be equated to a change in the sample-tip gap. The apparent change will contain contributions due to the deformation of the contamination layer, of the tip, and of components around the measurement loop of the instrument, particularly in a small region around the tip, and these are unlikely to behave linearly. Considering Hertzian deformation of elastic contacts, a power law relationship between probe distortion and displacement would be expected. If the actuator or the tip moves Δs, the tip-sample spacing changes Δd and the deformation is δ, the relationship between the Δs and Δd can be expressed by

$$\Delta s = \Delta d + \delta \qquad \text{(8-1a)}$$

$$\delta = \left[\frac{9F^2}{16RE^{*2}}\right]^{1/3} \qquad \text{(8-1b)}$$

$$\frac{1}{E^*} = \frac{1-\nu_T^2}{E_T} + \frac{1-\nu_s^2}{E_s} \qquad \text{(8-1c)}$$

where R is the tip radius, E_T is the Young's modulus and ν_T the poisson's ratio of the tip material, while E_s and ν_s are the Young's modulus and the poisson' ratio of the specimen material or, exactly, the material of the contaminant film, respectively. The loading force F is unknown, and it depends on the demand of the tunnelling current and the condition of the contaminant film. The contact radius due to the tip-sample deformation is given by

$$a_s = \left[\frac{3RF}{4E^*} \right]^{1/3} \tag{8-2}$$

It can be seen that a given change Δd thus requires a correspondingly greater value of Δs, giving rise to an amplification of the measured values as the tip is moved in and out to maintain a constant tunnelling current. This amplification is not linear and dependent of local properties of the contaminant layer.

Metrologically this is unsatisfactory as it will introduce significant variations in measured position on different engineering surfaces when using a tunneling probe in air. It is likely that, in many instances, tips are scanning in contact with a contaminated layer or ploughing through it depending upon the thickness of the contaminant and the mode of operation. Perhaps this may be useful in some applications such as monitoring the growth of oxidation or characterisation of both mechanical and electrical properties of surfaces for use in controlled engineering environments.

The deformation caused by the contact limits the resolution of the measurement in both vertical and horizontal directions. The level of this uncertainty can be estimated numerically through the deformation and the contact radius by the Hertzian expressions of Eqn.(8-1b) and (8-2).

The method used to obtain data for Fig. 8.2 is a quick way to check if a very low barrier height exists. The rate of change of tip position with tunneling current will give diagnostic information for the evaluation of barrier height.

The contact model can successfully explain the anomalous low barrier height, and is consistent with many experimental studies (Mate *et al.* 1989; Yamada *et al.* 1988). However there is as yet no solid experimental data to verify it and to demonstrate the existence or the depth of the deformation made by a tip.

8.4 Effect of the Tunnelling Tip

A key element in STM is the tip. It is now well known that the electronic states and chemical composition of the tip play as large a role in determing STM results as does the sample (Lang 1986). Further, the tip must be atomically sharp to achieve high resolution, but a blunt tip may be more desirable for spectroscopic work (Huk and Silverman 1989). Additional concerns include tip shank shape and possible occurrence of multiple tips. If a tip is composed of a wide flat surface with a small protrusion, high-resolution scans of flat surface can be obtained, but images of rough surfaces will be strongly distorted by its shank, or shape of the tip above the protrusion. If the end of the tip has several nearly equal height protrusions, then multiple point tunnelling (either sequentially or simultaneously) will be observed. This often results in noisy images, or in multiple images, or both.

The dimension effect of a tunnelling tip on STM images is much similar to that of a stylus tip except that the effective size of the tip is the combination of its real size plus the gap distance between the tip and the sample. The dependence of STM images on tip radius is more significant than the dependence on gap because of the short range of tunnelling gap. If the gap distance becomes shorter than 0.5 nm, the tip may touch the sample, while if it is larger than 2 nm, the emission may become more significant than the tunnelling.

To reduce the tip size effects, efforts have been made to produce sharp tips, for example, by focused ion beam (FIB) milling (Vasile *et al.* 1991a, 1991b; Griffith *et al.* 1991), by oxidation sharpening of silicon tips (Ravi *et al.* 1991), by direct electro-beam fabrication (Lee *et al.* 1991). The last two techniques are particularly used for making sharp AFM tips.

Schneir *et al.* (1986) have reported a tip preparation method of crashing the tip repeatedly against the surface. Applying a higher bias voltage momentarily may help to remove the contamination on the tip. Sometimes, however, merely waiting a sufficient period of time allows

high resolution imaging. This has been actually experienced during the STM tests. It was found that more details were resolved after several scans at the same track.

A sharp tip is desirable for high resolution measurement but it tends to be too slender at the end which may cause the tip to bend when scanning rougher specimens due to the lateral attractive forces such as electrostatic attraction caused by the tunnelling bias, van der Waals attraction, and capillary attraction caused by adsorbed gases in the gap between the tip and the sample (Griffith *et al.* 1992).

8.6 Conclusions

The combined high lateral and high spatial resolution of STM makes it a strong candidate for the reference gauge for transferring displacement standards at the nanometre level. However this can only be realised when a full understanding of the interaction between a tunneling tip and a sample is obtained. It would be useful to set up a fully characterised standard specimen, for instance HOPG, as a basis for obtaining the calibrated characteristic of different tips under different operating conditions.

The present construction and combination of the STM probe with a Nanostep slideway gives us a quick and efficient way to look into some problems and potentials in building and operating a STM in ambient air in a metrological laboratory. A better insight into the characteristics of STM operating in air has been obtained, and features of a few nanometres have been resolved. Improvements are being undertaken to increase the vertical resolution by further reducing the noise level. A method of testing experimentally the existence of the contaminate layer is under investigation.

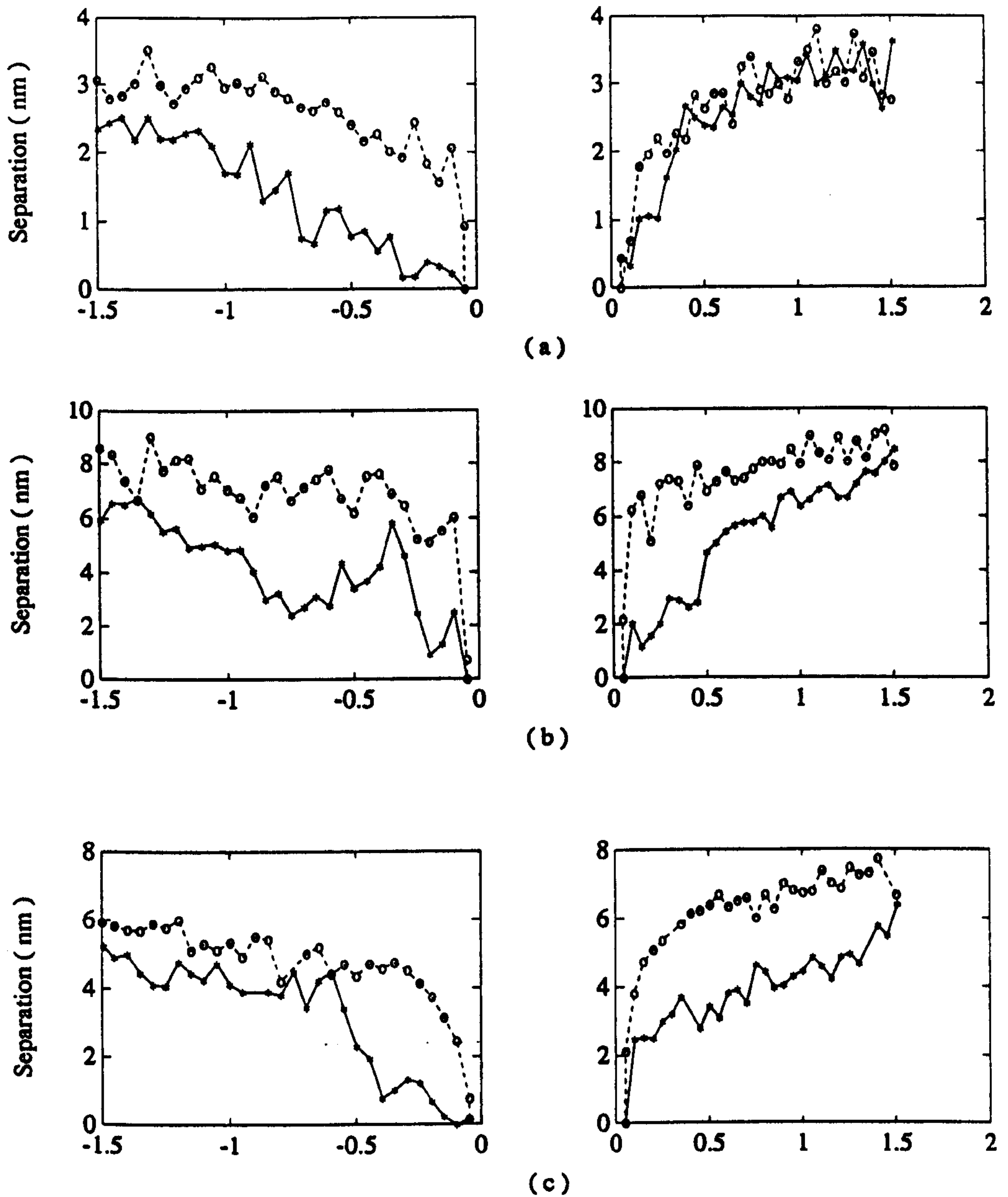


Fig. 8.1 The *S-V* characteristic of a tungsten tip on a gold-coated silicon surface at constant tunnelling currents of: (*a*) I_t = 10 nA, (*b*) I_t = 5 nA, and (*c*) I_t = 0.2 nA. (Arbitrary zero level, and x-axis in volts).

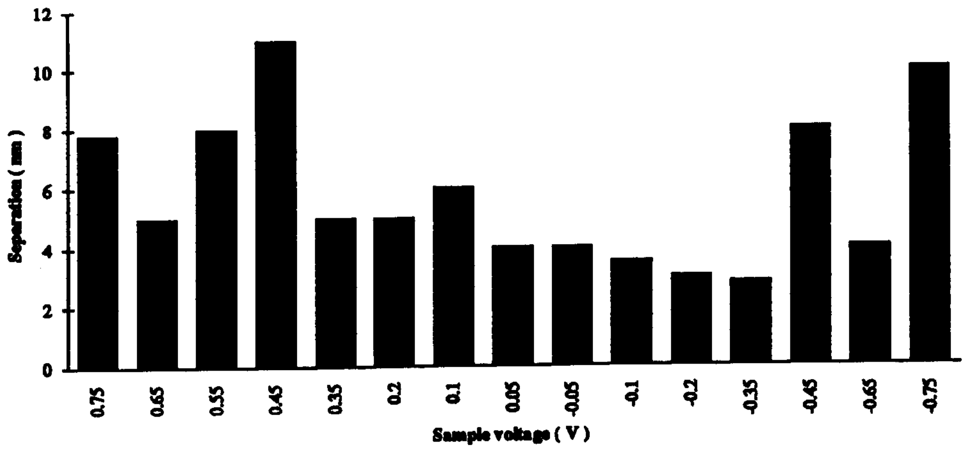


Fig. 8.2 The maximum variation of tip movement as tunnelling current varies from 0.5 to 10 nA at each constant bias voltage on the same surface as in figure 8.1.

9. CONCLUSIONS AND SUGGESTIONS FOR FUTURE DEVELOPMENT

9.1 Conclusions

Performance evaluations in a sense of measuring fidelity have been carried out, theoretically and experimentally, first on a typical stylus instrument and then extended to one of the most advanced measuring techniques; the scanning tunnelling microscopy. The theme runs into two regimes; the contact one represented by stylus instruments and the non-contact one by scanning tunnelling measurements. Limitations in both principle and application have been investigated particularly in the interests of submicrometre and nanometre levels. Conclusions are drawn as follows.

9.1.1 Fidelity aspects in contact measurements

The fidelity of a stylus instrument is affected, physically and geometrically, by the interactions between the stylus tip and the surface being measured. The physical or mechanical interactions are mainly from tracking force, frictional force and dynamic force, and the geometrical interaction is from the dimension of the stylus tip.

A linear variable force actuator based on a magnet-coil was built, evaluated, and has been successfully used throughout the experiments for studying the force interactions between the stylus tip and the specimen.

At the nominal contact force of a typical stylus instrument, around 1 mN, the small contact area leads to high Hertzian stress. Measurements on a variety of specimens show that around this nominal load, the high stress does not lead either to unacceptable surface damage or a lack of integrity in following a surface profile. Generally, it was not until loads increased to about 2 or 5 mN that definite smoothing was observed, with sharp peaks disappearing and downward slopes and valleys becoming progressively steeper and wider, on a range of common engineering materials.

This was also reflected in the statistical parameters of profiles where root mean squares (rms) slope and peak curvature both tend to decrease slightly. On some surfaces of soft materials the contact force, for a typical 2 μm truncate pyramidal tip, has to be kept below 1 mN or better below 0.7 mN to avoid any damage. The combination of reducing the tracking force to about 0.4 mN and traversing the stylus at a slow speed seems to be an adequate procedure. Conversely, the results of this study imply that rapid estimation of surface finish parameters may be obtained by using a fast traverse speed and an increased contact force. Clearly the ability to automatically vary tracking forces will enable present stylus instruments to be used for a variety of measurements over a wide range of applications.

However, the experiments of this investigation show that dynamic effects can introduce significant variations on measurements at different traverse speeds. The dynamics of a stylus mechanism cause the stylus to bounce over the measured surface at a light load when it is traversed above a certain speed. As a consequence, it has been observed that the measured roughness value may increase by more than 100%. The dynamic behaviour of a stylus system is dependent on its input frequency which depends on both the spatial wavelength of surface structure and the speed at which the probe is traversed. For many machined surfaces, the structures or irregularities vary over a wide range with the high-frequency components corresponding to sharp peaks or edges and narrow valleys. With sufficiently high forces this causes stylus jumping at the peaks and a large increase in contact forces at the valleys both of which may lead either to a distorted profile or surface damage.

A relatively simple system has been designed and constructed that allows the damping of a stylus profilometer to be varied with force feedback. Theoretical and practical evaluations confirm that there is some merit in adjusting damping factors in the region between 0.5 and 0.7, rather than the value well below 0.1 present in most stylus instruments, to obtain improved fidelity at higher traverse speeds. Evidence has been obtained to show that dynamic loading effects increase the variability in both vertical and horizontal axes of the profile and that higher damping reduces these effects.

In practice, the optimal damping value between 0.5 and 0.7 should be considered in designing of stylus instruments and it may also be applicable to other force following systems such

as atomic force microscopes and force transducers in which displacements are used to monitor forces. Recent studies of such points are showing promising developments (Smith and Howard 1993, Xu and Smith 1993). Active force feedback appears to be a very promising method of improving the behaviour of future systems of lightly loaded contact gauges

Another aspect of fidelity in contact measurement is the friction at the interface. It was demonstrated that the frictional force could be modelled as a mean frictional force superimposed with an instantaneous frictional force. The mean frictional force is dependent on properties of the two surfaces and contact conditions. The instantaneous one is dependent on the local slope of the specimen surface. It has been shown that frictional force affects measuring fidelity in a way that it affects both the vertical force and the horizontal force due to the variation of the surface slopes, which leads to distortions both in height and in lateral distance. Friction coefficients on four typical engineering surfaces were measured with a contact size in both micrometer and sub-millimetre levels. For a typical 2 μm diamond tip, the friction coefficient varies little either with load over a range 0.5 – 5 mN or with speed in the region 1.0 – 0.1 mm s^{-1}, except for a slight tendency to decrease at the fast speed. Measured values vary from about 0.25 for aluminium, which is softer than the copper, to around 0.1 on steel and 0.07 on single-crystal silicon. Horizontal drag on the specimens shows fluctuations about a mean level which correlate very strongly with the derivative of the surface profile. This is consistent with a model in which friction is proportional to normal load and modulated by surface slope. For a large scale of contact with a 1 mm steel ball the friction coefficients on these specimens are much higher, 0.3 – 0.6 for steel, 0.2 – 0.4 for copper and a fairly constant value of 0.2 for aluminium. Surface hardness and contaminant film are believed to have played an important role in each case. On a large scale contact the hardness of the surface is dominated by the bulk properties while on a small scale contact the hardness could be much increased due to the work-hardening process.

The above information provides a background knowledge of stylus measurements. For a particular surface, the tracking force (adjustable for new stylus instruments) has to be set on the consideration of surface hardness and stylus tip radius, the traversing speed should be set as low as possible to avoid dynamic effect. The resolution on height measurement, if not limited by the instrument, can be evaluated by the elastic deformation made by the stylus, and the lateral resolution

will be limited by the tip radius. The effect of frictional force on both height and lateral measurement can be estimated through the average slope of the surface and the frictional coefficient.

9.1.2 Fidelity aspects in non-contact measurement

Fidelity in non-contact measurement is investigated based on the scanning tunnelling microscopy. Two STM systems have been constructed and evaluated for this purpose. Each STM uses a linear drive for probing; a spring flexure for STM_1 and a closed-loop piezo-electric translator for STM_2, for the exploration of metrological capability of STMs in engineering applications.

Although scanning tunnelling microscopy is non-contact in nature, great care has to be taken when it is operated in normal environment. It was observed that the sensitivity of the tunnelling current to the tip-sample separation for STM_2 was less than 2 nA nm^{-1}; a value which is far lower than that normally encountered for STMs operated in vacuum. The measured relationship between the tip-sample separation and the bias voltage shows a tendency for discontinuous 'jumps' of the tip position to occur, in some instances by up to 2 nm. There was distinct hysteresis, with the separation remaining relatively greater when reducing the voltage at constant current than when increasing it again. The observations provide strong evidence that the surface follows as the tip is retracted, thereby implying that the surface is deformed by the tip. And this deformation is mainly caused by the surface contaminant layer under ambient conditions. It is more likely that tunnelling occurs after the tip has contacted this layer and the subsequent characteristics are dominated by the need to deform this layer to affect a change in the conductivity of the junction. The observed steps in the *S-V* (separation-bias voltage) curves suggest that increase in current may be due to the expulsion of material or fracture of surface layer under varying loads. The large apparent displacement required for maintenance of the current could be caused by microscopic surface deformations of the contaminate layer and the specimen or by macroscopic displacements due to finite stiffness around the measurement loop.

From metrological point of view, this contact situation will introduce significant variations in measured position on different engineering surfaces when using a tunnelling probe in air. In many instances, tips are probably scanning in contact with a contaminated layer or ploughing through it

depending upon the thickness of the contaminant and the mode of operation. A method was proposed for checking the contact situation by the change rate of tip position with tunnelling current, which can give diagnostic information for the evaluation of barrier height.

9.2 Suggestions for Future Development

To cater for an ever increasing demand for surface topography measurement at nanometre level, the stylus technique has been pushed to measure height variation to 0.1 nm and lateral resolution down to 0.1 μm in conjunction with a very slow traverse speed and a small stylus force. Under a small stylus force, the spring rate of the stylus system makes a large difference in the contact force. For instance, the force of the Nanostep can be set as small as 10 μN at the neutral position for a range of 1.2 μm, then the contact force will vary from zero to 20 μN. The measured profile will be distorted at deep valleys as a consequence of loss of contact.

A solution to this and other problems is an optimal stylus system that should meet four requirements: (a) constant but adjustable stylus force, (b) optimal damping, (c) high measuring speed and (d) stable transducing. The system would consist of a force actuator, a ligament suspension with stylus assembly and a displacement transducer. The force actuator can be built based on the method of the magnet/coil combination used in this thesis to provide different force settings, and a constant contact force can be achieved by feedback compensation for the ligament support system. The feedback will allow a much stiffer support to be used than is allowable in conventional designs, giving advantages in terms of cross-axis stiffness and resonance without unacceptable variation in contact force during a measurement. The stylus traverse speed may then be increased due to the increase in stiffness thus in resonance, supposing that the effective mass of the stylus system does not increase proportionally. The displacement transducer may employ a conventional LVDI gauge or, alternatively, a capacitive transducer which can be made out of low thermal expansion materials such as Zerodur glass ceramic to give better stability. Capacitive sensing allows very precise practical measurement since a capacitor does not suffer Johnson noise or shot noise associated with the use of resistors and semiconductors. Modern electronics provides high resolution on the measurement of capacitance, of the order of 10^{-18} to 10^{-19} F. One major source of

noise is the variation of stray capacitance of the sensor plates, circuit and wiring cable layout, thus proper care has to be taken to the detail of shielding and grounding. However, a capacitance gauge with resolution better than 0.1 nm over a range of 10 μm is easily obtainable (for example, NS2000 from the Queensgate instrument Ltd).

The system damping can be designed to have a fixed value between 0.5-0.7 by the feedback derivative control as discussed in chapter 6 or other mechanical approaches. This is a linear force feedback, giving a constant damping ratio at all frequencies. Ideally, a stylus system needs lightly damping when it is traversed at a frequency far below its resonance. The optimal damping value derived from the experiments around 0.6 is best applied when the stylus is traversed at a frequency in the region of the resonance. Therefore, a non-linear force feedback for damping control would be more appropriate for a stylus system, that is there is no extra damping added at low frequency but optimal damping is applied at high frequency. In practice, this can be accomplished based on the damping control device used in chapter 6 by comparing the derivative signal with a certain set value before passing it to the feedback loop; when the derivative signal is smaller than the set value there will be zero feedback, otherwise, a certain signal will be fed into the feedback loop to achieve the required damping. The selection of the set value has to be determined according to the dynamic characteristics of the stylus system.

Future work on STM will be focused on force interactions at the interface between the tunnelling tip and the specimen as these forces play an important role in fidelity of measurements. The present research has only brought up some problems and initial results, the sources of these problems and the extent of their effects need further investigations. When an STM is operated in ambient environments, there much likely exist the following forces such as Van der Waals, electrostatic, capillary, repulsive, adhesive, and frictional forces. The latter three forces occur when the tunnelling tip is scanning in or beneath a contaminant layer on the specimen. The variation of the resultant force could be from 10^{-8} to 10^{-6} N. The deformations of both the tip and the specimen under such forces will have large effects on imaging and metrology.

APPENDIX A

Publications in the period of PhD:

1. Liu, X. and Smith, S. T., A scanning tunnelling profilometer for surface measurement applications, *The 2nd Biennial Joint Warwick/Tokyo Nanotechnology Symposium*, Warwick, July, 1990.

2. Chetwynd D. G., X. Liu and Smith, S. T. 1992, Signal fidelity and tracking force in stylus profilometry, *Int. J. Mach. Tools Manufac.*, **32**, 239-245.

3. Liu,. X., Smith, S. T. and Chetwynd, D. G. 1992, Frictional forces between a diamond stylus and specimens at low load, *Wear*, **157**, 279-294.

4. Smith, S. T. and Liu, X. 1992, A profilometer for surface proximity probe applications, *Nanotechnology*, **3**, 37-41.

5. Smith, S. T., Howard, L. P., Liu, X., Xu, Y. and Chetwynd, D. G., Practical design considerations in the applications of atomic force profilometry for engineering surface measurements, *The 3rd Biennial Joint Warwick/Tokyo Nanotechnology Symposium*, Tokyo, Septempber, 1992.

6. Liu,. X., Chetwynd, D. G. and Smith, S. T. 1993, Improvement of the fidelity of surface measurement by active damping control, *Meas. Sci. Technol.*, **4**, 1330-1340.

7. Liu, X., Smith S. T. and Chetwynd D. G. 1993, Tunnelling probes for metrological applications using a long-range profilometer, *Rev. Sci. Instrum.*, **64**, 3161-3168.

APPENDIX B

- **The following programs are written in Matlab format and can be run in Matlab.**

```
function fy = Dfilter(yy, fs),
% Filter design;
%fs = input( '  Sample Freq  =  ');
write;
echo off;
Option = input( '    ','s');
if Option == 'l' | Option == 'L',
fc = input( ' Low Pass Filter:  Fc =  ');
H=[1 1 0 0 ];
f=[0 , 2 * fc / fs , 2 * fc  / fs , 1 ]; typ = 0;
end
if Option == 'b' | Option == 'B',
fc1 = input( ' Band Pass Filter:  Fa  and  Fb '); fc2 = input('   ');
H=[0 0 1 1 0 0];
f=[0 , 2 * fc1 / fs , 2 * fc1 / fs , 2 * fc2 / fs ,2 * fc2 / fs , 1]; typ = 1;
end
if Option == 'h' | Option == 'H',
fc = input(' High Pass Filter:  Fc = ');
H=[0 0 1 1];
f=[ 0 , 2 * fc  / fs ,  2 * fc / fs , 1];
typ = 2;
end
if Option == 's' | Option == 'S',
fc1 = input( '  Band Stop Filter: Fa  and  Fb  ');
fc2 = input('   ');
H=[1 1 0 0 1 1];
f=[0 , 2 * fc1 / fs , 2 * fc1 / fs , 2 * fc2 / fs , 2 * fc2 / fs , 1]; typ = 3;
end
% Dtype = input( ' The Methods Choice: Yule-Walk/Butterworth: YY/BB ','s');
% if Dtype == 'B' | Dtype == 'b',
fy = buttfilt(f, H, yy, fs, typ);
% else,
% fy = filter4L(f, H, yy, fs);
end
```

```
function [ F, f ] = DampF(x, y, t)
% For filter design;
%q=1+(0.01*t).^2;
n = max(size(y));
m = max(size(x));
f = y(1) * exp ( - x(1) * t ) .* cos ( x(2) * t ) ;
F = sum((f-y).^2);
```

```
function y = buttfilt(f, H, x, fs, typ),
% For filter design;
clc;
clg;
NN = max(size(x));
fhz = f*fs/2;

% N = input(' Input the order of the filter (number of poles and zeros )  ');
 N=5;
if typ == 0,
wn = f(2);
[Bh, Ah] = butter(N, wn);
end
if typ == 1;
wn = [f(2),f(4)];
[Bh, Ah] = butter(N, wn);
end

if typ == 2,
wn = f(2);
[Bh, Ah] = butter(N, wn, 'high');
end
if typ == 3,
wn = [f(2), f(4)];
[Bh, Ah] = butter(N, wn, 'stop');
end

plot(fhz, H), title('Desired Frequency Response'),
xlabel('Frequency (Hz)'),
ylabel('Magnitude'), clc;
 format compact
 clc
Ah, Bh   % Here are the filter coefficients:
% Strike any key to continue.
 format
 clc
 %  response of the filter:
 n = 256;
hh = freqz(Bh,Ah,n);
 hy  = abs(hh);
% Strike any key for plot of  the frequency response.
 ff  = fs/(2*n) * (0:n-1);
 plot(fhz, H, ff, hy), title('Actual vs. Desired Frequency Response'),
 xlabel('Frequency (Hz)'), ylabel('Magnitude'), pause
clc
clg
y = filter(Bh, Ah, x);
% Strike any key for plot of the filtered series.
L = 0:1:NN-1;
t = 1/fs*L;
% subplot(211),plot(t,x),title('Input of The Filter'),xlabel('Time'),ylabel('Magnitude');
% subplot(212),
% plot(t,y), title('Output of The Filter'), xlabel('Time'), ylabel('Magnitude'); %pause;
end;
```

```
function p = parameter(y1, y2, l1, l2)
% Calculation of parameters
[x,diffy] = fittmean(y1, l1, l2);
g1 = input('input the first sensitivity in um/v ');
v1 = input('input the first scan speed in um/step '); aa=[x,diffy];
xx1 = aa(:,1)*v1;
yy1 = aa(:,2)*g1;
l = l2-l1;
Pra1 = sum(abs(yy1))/l;
Prq1 = sqrt(yy1'*yy1/l);
ypp2 = 0.;
h = v1;
n = size(xx1);
count = 0;
for k = 3:n-2 yp(k) = (yy1(k-2)-8*yy1(k-1)+8*yy1(k+1)-yy1(k+2))/12/h;
 if (yy1(k)-yy1(k-1))*(yy1(k)-yy1(k+1))>0 yp2(k)=(yy1(k-1)+yy1(k+1)-2*yy1(k))/h/h; count=count+1;
ypp2 = ypp2+yp2(k)*yp2(k);
else
yp2(k) = 0.;
end;
end;
DetaQ1 = std(yp);
fprintf('PRa= %g\n',PRa1);
fprintf('PRq= %g\n\n',PRq1);
fprintf('DetaQ1= %g\n\n',DetaQ1);
fprintf('Peak Curvature= %g\n\n',sqrt(ypp2/count));
[x,diffy] = fittmean(y2,l1,l2);
g2 = input('input the second sensitivity in um/v ');
v2 = input('input the second scan speed in um/step'); aa1=[x,diffy];
xx2 = aa1(:,1)*v2;
yy2 = aa1(:,2)*g2;
Pra2 = sum(abs(yy2))/l;
Prq2 = sqrt(yy2'*yy2/l);
ypp22 = 0.;
h = v2;
count1 = 0;
for k=3:n-2 ypp(k)=(yy2(k-2)-8*yy2(k-1)+8*yy2(k+1)-yy2(k+2))/12/h;
if (yy2(k)-yy2(k-1))*(yy2(k)-yy2(k+1))>0 yp2(k)=(yy2(k-1)+yy2(k+1)-2*yy2(k))/h/h;
ypp22=ypp22+yp2(k)*yp2(k);
count1=count1+1;
else
yp2(k)=0.;
end;
end;
DetaQ2=std(ypp);
fprintf('PRa2= %g\n',PRa2);
fprintf('PRq2= %g\n',PRq2);
fprintf('DetaQ2= %g\n',DetaQ2);
fprintf('Peak Curvature= %g\n',sqrt(ypp22/count1)); clg
subplot(211),plot(xx1,yy1,xx2,yy2);
xlabel('Scan displacement ( um )');
ylabel('Profile displacement ( um )');
dd=yy1-yy2;
av=std(dd);
```

```
fprintf('std of difference= %g\n',av); ylabel('Profile displacement ( um )');
title('Difference in height ');
p=[PRa1 PRq1 DetaQ1 PRa2 PRq2 DetaQ2];

function [py1,f1]=corrloop(y),
% Non-linear fitting program;
y=y*1.24;
t=1:1000;
a=polyfit(t',y,1);
yp=a(1)*t'+a(2);
yy=y-yp;
subplot(211);
plot(yy);
print4;
ss=[ ];
N=max(size(y));
[x,y1,pyy,f]=dprf(y,.5);
save too y1;
!xcorr.out
load bar
c=xx;
for i=1:27,
[a,r,m,s]=correlat(c,.5,i);
s1=s*(N+i+1)/(N-i-1);
ss=[ss,s1];
end;
[si,p]=min(ss);
[a,r,mod,s]=correlat(c,.5,p);
a1=[1,a'];
[m,w]=freqz(1,a1,100);
py1=s*(abs(m).^2);
f1=w*2/2/pi; subplot(211);plot(f(1:512),pyy(1:512)); subplot(212);plot(f1,py1);
clg
dir *.dat
filename = input( ' input your data filename ','s'); s=[' load ' , filename , '.dat ' ];
disp([' Loading The Data File ' ,filename,' ... ']); eval(s);
yy= 25*eval(filename);
subplot(211),plot( yy );
str = 'n';
while str == 'n',
st = input( ' input the start point ');
sf = input( ' input the final point ');
y1= yy(st:sf)-mean(yy(3000:4000)); [x4(1),n1]=max(y1);
y=y1(n1:length(y1));
clg;
subplot(211),plot( yy );
subplot(212),plot(y);
str = input( ' Is it that you want:  yy/nn ' ,'s'); end
fs=input(' input sample frequency  '); yy=[ ];y1=[ ];
nn=max(size(y)); t=0:nn-1;t=t';
k = y < 0;
k1 = filter([1 -1],1,k);
k2=sum(abs(k1));
x(2)=70;
```

```
%x(2)=800*pi*k2/nn;
x(1)=0.015*x(2);
x=x';
t=t/fs;
%[X1,p]=s_shap(y,x,t,50);
X=fmins('DampF',x,[ ],[ ],y,t);
[FF,ff]=DampF(X,y,t);
plot(t,ff,t,y);
wn=sqrt(X(1)^2+X(2)^2);
Wn=wn
z=X(1)/wn
```

function u= cc2dd(y,fs,z,wn,K),

```
bb=K;
aa=[1 2*z*wn wn*wn];
[a,b,c,d]=tf2ss(bb,aa);
[P,G]=c2d(a,b,1/fs);
[dn,dp]=ss2tf(P,G,c,d,1);
u=0;
n=max(size(y));
for i=3:n,
u0=u(length(u));
k=i:-1:(i-2);
un=0.5*dp*y(k)/dn(2);
u=[u;un];
end;
u(1)=u(2);
```

function [x,Y]=fit_xy(xx)

```
% Least squares fillting program;
m=input('fitting order=');
n=size(xx);
x=xx(:,1);
y=xx(:,2);
for j=1:m+1
A(:,j)=x.^(m+1-j);
end
c=A\y;
n=size(xx);
Y  = A * c;
diffy=y-Y;
sigma=std(diffy);
c
sigma
plot(x,y,'*',x,Y);
```

function y = frac(g, r, D, x, n)

```
% Fractal calculation;
N  = 1:n;
R  = r .^(N);
R=R';
P  = R * x;
G  = (R .^ (2-D)) * ones(size(x));
clear R;
```

```
y  = sum( cos( 2 * pi * P ) ./ G );


function b = nlfilter(a,nhood,FUN,P1,P2,P3,P4,P5,P6,P7,P8,P9,P10)
%NLFILTER Local non-linear filtering.
%           B = NLFILTER(A,[M N],'fun') applies the m-file 'fun' to
%           each M-by-N sliding neighborhood of A. The m-file 'fun'
%           should return a scalar result, c = fun(x), which is
%           the filtered value for the center pixel in the M-by-N
%           neighborhood, x. 'fun' will be called for each value in A.
%
%           Up to 10 additional parameters can be passed to the
%           function 'fun' using
%             B=NLFILTER(A,[M N],'fun',P1,P2,P3,...)
%           in which case 'fun' is called using
%             c = fun(x,P1,P2,P3,...).
%
%           At the edges, the M-by-N neighborhood is formed by padding
%           with ones if A is an indexed image or with zeros otherwise.
error(nargchk(3,13,nargin));

% Form call string.
    params = [ ];
    for n=4:nargin
    params = [params,',P',int2str(n-4)];
end
if ~any(FUN<48), fcall = [FUN,'(x',params,')'];
else fcall = FUN; end
% Expand A
[ma,na] = size(a);
if isind(a),
    aa = ones(size(a)+nhood-1);
else
    aa = zeros(size(a)+nhood-1);
end aa(floor((nhood(1)-1)/2)+[1:ma],floor((nhood(2)-1)/2)+[1:na]) = a;

% Apply m-file to each neighborhood of a
b   = zeros(size(a));
rows = [0:nhood(1)-1]; cols = [0:nhood(2)-1];
f   = waitbar(0,'Applying non-linear filter...');
for i=1:ma,
    for j=1:na,
        x = aa(i+rows,j+cols);
        b(i,j) = eval(fcall);
    end
    waitbar(i/na)
end
close(f)
clg
        f1=input('input the first ch. of force data file ','s');
        f2=input('input the first ch. of profile data file ','s')
            n1=input('from ')
            n2=input('to ')
            for i=n1:n2,
```

```
            s=num2str(i);
          s1=[f1 s];
          s2=[f2 s];
i
 s=['load ',s1];
eval(s);
l=ff;
    s=['load ',s2,'.dat'];
    eval(s);
m=eval(s2);
n=max(size(l));
t=20:n;
tt=(t-min(t))*0.5;
subplot(211)
plot(tt,m(t)*1.02)
xlabel('Traverse length ( um )') ylabel('Profile ( um )')
title(s2)
subplot(212)
plot(tt,l(t))
xlabel('Traverse length ( um )') ylabel('Dynamic force ( uN )') title(s1)
print;
clg;
eval(['clear ',s1])
eval(['clear ',s2])
end;

function [f,pyy]=freq(y,fs);
% Spectral analysis; fs -sample frequency
n=max(size(y));
t=0:(n-1);
a=polyfit(t',y,1);
yy=a(1)*t'+a(2);
yy1=(y-yy);
Fy=fft(yy1);
pyy=Fy.*conj(Fy)/1024;
f=(0:1023)/1024*fs;
plot(f(1:511),pyy(1:511));
title('Power spectrum density');
pause;

function y = simutip(x, r, n),
% Simulation of the tip effect;
X = -r: r/n : r;
[M,N] = size(X);
[mm, nn]= size( x );
N2  = fix( N / 2 ) + 1;
z = [ ];
y = x(1:N2-1);
for i = N2: nn - N2,
z = x((i-N2 + 1):(i - N2 + N) ) + sqrt( r * r - X .* X ) - r ; y = [ y, max( z ) ];
end
y = [ y, x(nn - N2 + 1 : nn )];

function [Ra, Rq, Rsk, K]=sk_ku(y)
```

```
% Calculation of the parametersl;
[M,N]=size(y);
MM = max([M,N]);
y1 = detrend(y);
Ra =sum( abs(y1) )/MM;
Rq = sum( y1 .* y1 )/MM;
Rq = sqrt(Rq);
RR = Rq * Rq * Rq * MM;
Rsk = sum( y1 .* y1 .* y1 )/RR;
K = sum( y1 .* y1 .* y1 .* y1)/RR/Rq;
```

- **The following program is written in Pascal and can be run in conjunction with the Metrabyte DAS-16 board.**

```
PROGRAM of DATD_SAMPLE;
uses
   dos,crt;
{$L tpucdasg}
type
         parray=array[0..15] of integer;
         data  =array[0..4000] of integer;
var
  mode,flag,i,j,samples,gain,addelay           :integer;
  d,d1,result,k,channel,N,freq,i1,output       :integer;
  param                                        :parray;
  volt1                                        :array[0..8000] of real;
  outfile                                      :text;
  outfilename                                  :string[20];
  answ,cc                                      :char;
  mm,tt                                        :real;
  p                                            :^data;
  pdata                                        :data;
FUNCTION TP_DASG(var mode:integer; var param:parray; var flag:integer):
              integer;external;
PROCEDURE INITIALISE;
begin
   clrscr;
   gotoxy(10,10);
   writeln('choose the channel');
   readln(channel);
   writeln('Choose the gain: 0 for x1 or 1 for x10');
   readln(gain);
   writeln('Enter the number of samples');
   readln(samples);
end;
PROCEDURE INITIALISE_METRABYTE;
begin
   mode:=0;
    param[0]:=768;
    result:=tp_dasg(mode,param,flag);
    mode:=1;
    param[0]:=channel;
    param[1]:=channel;
```

```
    result:=tp_dasg(mode,param,flag);
    mode:=21;
    for j:=0 to 15 do param[j]:=0;
    param[channel]:=gain;
    result:=tp_dasg(mode,param,flag);
end;
PROCEDURE TIMER;
begin
      writeln('input the sample frequency in times/sec.');
      readln(freq);
      tt:=10000000 div freq;
      N:=round(sqrt(tt));
      mode:=17;
      param[0]:=N;
      param[1]:=N;
      result:=tp_dasg(mode,param,flag);
 end;
PROCEDURE GET_DATA;
begin
       { mode:=15;
        param[0]:=0;
        writeln(' input D/A output');
        readln(output);
        param[1]:=output;
        result:=tp_dasg(mode,param,flag);  }
        writeln('to start A/D sampling, press any key');
        readln(cc);
        mode:=4;
        param[0]:=samples+2;
        new(p);
       param[1]:=Ofs(p);
       param[2]:=1;
       result:=tp_dasg(mode,param,flag);
     end;
PROCEDURE STORE_THEM;
begin
   gotoxy(10,18);
    writeln('Enter output file name');
    readln(outfilename);
    writeln('insert discette and press r for ready');
    readln(answ);
     if (answ='r') then
      assign(outfile,outfilename);
      rewrite(outfile);
      if gain=0 then
      mm:=204.8;
      if gain=1 then
      mm:=2048;
    j:=0;
    repeat
      writeln(outfile,pdata[j]/mm);
      j:=j+1
   until (j>samples-1);
   writeln('finished');
```

```
end;
{---------------------------------------------------------------------}
BEGIN  {main program}
    INITIALISE;
    INITIALISE_METRABYTE;
    TIMER;
    GET_DATA;
    STORE_THEM;
    close(outfile)
END.
```

REFERENCES

Abbott, E. J. and Firestone, F. A. (1933), Specifying surface quality, *J. Mech. Eng.*, **55**, 569-572.

Ajioka, Shigeyasu (1974), The dynamic response of stylus, *Bull. the Japan Soc. of Prec. Engg.*, 1 (4), 228-233.

Albrecht, T. R. and Quate, C. F. (1988), Atomic resolution with the AFM on conductors and non-conductors, *J. Vac. Sci. Technol.*, A **6**, 271-274.

Becker, P., Seyfried, P. and Siegart, H. (1982), The lattice parameter of highly pure silicon single crystals, Z. *Phys.* B **48**, 17-21.

Beebe, T. P., Wilson, T. E., Ogletree, D. F., Katz, J. E., Balhorn, R., Salmeron, M. B., and Siekhausr, W. J. (1989), Direct observation of native DNA structures with the scanning tunnelling microscope, *Science*, **243**, 370-372.

Bell, T. (1987), Surface engineering: a rapidly developing discipline, *Eur. J. Eng. Educ.*, **12**, 27-32.

Bell, T. (1990), Surface engineering – past, present and future, *Surf. Eng.*, **6**, 31-40.

Bendat, J. S. and Piersol, A. G. (1971), Random data: analysis and measurement procedures, Wiley-Interscience.

Bennett, J. M. (1986), Surface preparation and cleaning techniques, *Proceedings of the 1986 Society of Vacuum Coaters Annual Technical conference*, 1-12.

Bennett, J. M. and Dancy, J. H. (1981), Stylus profiling instrument for measuring statistical properties of smooth optical surfaces, *Appl. Opt.*, **20**, 1785-1802.

Bennett, J. M. and King, R. L. (1970), Effect of polishing technique on the roughness and residual surface film on fused quartz optical flats, *Appl. Opt.*, **9** 236-238.

Bentley, J. P. (1988), Principles of measurement systems, Longman Scientific & Technical, U. K..

Berghaus, Th, Brodde, A., Neddermeyer, H., and Tosch, St (1988), On the interpretation of current images in scanning tunnelling spectroscopy of Si (111) 7x7, *J. Vac. Sci. Technol.*, A **6** (2), 483-487.

Berry, M. V. and Lewis, Z. V. (1980), On the Weierstrass-Mandelbrot fractal function, *Proc. Roy. Soc.* (London) **370**, 459-484.

Binnig, G. and Rohrer, H. (1982), Scanning tunnelling microscopy, *Helvetica Physica Acta*, **55**, 726-735.

Binnig, G. and Rohrer, H. (1986), Scanning tunnelling microscopy, *IBM J. Res. Develop*, **30**, 355-369.

Binnig, G. Smith, D. P. E. (1986), Single-tube three-dimensional scanner for scanning tunnelling microscopy, *Rev. Sci. Instrum.*, **57**, 1688-1689.

Binnig, G., Gerber, Ch., Stoll, E., Albrecht, T. R., and Quate, C. F. (1987), Atomic resolution with the atomic force microscope, *Surface Science*, **189/190**, 1-6.

Binnig, G., Quate, C. F., and Gerber, Ch. (1986), The atomic force microscope, *Phys. Rev. Lett.*, **56**, 930-933.

Binnig, G., Rohrer, H., Gerber, Ch., and Weibel, E. (1983), 7 x 7 reconstruction on Si(111)resolved in real space, *Phys. Rev. Lett.*, **50**, 120-123.

Blau, P. J. (1987), A model for run-in and other transitions in sliding friction, *J. of Tribology*, **109**, 537-544.

Bottomley, L. A., Haseltine, J. N., Allison, D. P., Warmack, R. J., Thundat, T., Sachleben, R. A., Brown, G. M., Woychik, R. P., Jocobson, K. Bruce, Ferrell, T. L. (1992), Scanning Tunnelling microscopy of DNA: The chemical modification of gold surfaces for immobilization of DNA, *J. Vac. Sci. Technol.* **A 10 (4)**, 591-595.

Bowden, F. P. and Leben, L. (1939), The nature of sliding and analysis of friction, *Proc. Roy. Soc. A*, **169**, 371-379.

Bowden, F. P. and Tabor, D. (1950), *The friction and lubrication of solids, Part I*, Clarendon Press, Oxford.

Bowden, F. P. and Tabor, D. (1964), *The friction and lubrication of solids, Part II*, Clarendon Press, Oxford.

Bowden, F. P. and Throssel, W. R. (1951), Adsorption of water vapour on solid surfaces, *Proc. Roy. Soc.* (London) **A209**, 297-308.

Bowen, D. K., Chetwynd, D. G. and Schwarzenberger, D. R. (1990), Sub-nanometre displacements calibration using x-ray interferometry, *Meas. Sci. Technolo.*, **1**, 107-119.

Bowen, D. K., Chetwynd, D. G., and Schwarzenberger, D. R. (1990), Subnanometre displacements calibration using x-ray interferometry, *Meas. Sci. Technol.*, **1**, 107-119.

Bowen, D. K., Chetwynd, D. G.,and Davies, S. T. (1985), Calibration of surface roughness transducers at ångstrom levels using x-ray interferometry, *Proc. SPIE*, **563**, 412-419.

Buckley, D. H. (1981), Surface effects in adhesion, friction, wear and lubrication, Elsecier, 315-427.

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

Burton, R. (1958), Vibration and impact, Addison - Wesley Publishing Company, Inc..

Cafferty, M. S. and Thompson, E. D. (1989), Stable current supply with protection circuits for a lead-salt laser diode, *Rev. Sci. Instrum.*, **60**, 2896-2901.

Chetwyd, D. G., Smith, S. T. (1990a), Strength degradation of ground and etched Si surfaces, *Proc. 2th Symposium on New Materials and their Application* (ed D. Holland) IOP Conf. Series No. 111, IOP, London, 493-495.

Chetwynd, D G., Harb, S., Krylova, N., and Smith, S. T. (1993), The feasibility of extended range monilithic x-ray interferometric calibrators, *Nanotechnology*, **4** (4), 183-193.

Chetwynd, D. G. (1978), Slope measurement in surface texture analysis, *Proc. I. Mech. E: J. Mech. Eng. Sci.*, **20**, 115-119.

Chetwynd, D. G. (1979), The digitisation of surface profiles, *Wear*, **57**, 137-144.

Chetwynd, D. G. (1991), Algorithms for computer aided precision metrology, in *From Instruments to Nanotechnology*, (Ed.) Gardner, J. W. and Hingle, H. T., Gordon and Breach, London.

Chetwynd, D. G. and Schwarzenberger, D. R. (1988), Accuracy estimation for fringe division in x-ray interferometry, *Microengineering Report ME70*, University of Warwick.

Chetwynd, D. G., Siddons, D. P., and Bowen, D. K. (1983), X-ray interferometer calibration of microdisplacement transducers, *J. Phys. E: Sci. Instrum.*, **16**, 871-874.

Coombs, J. H., Pethica, J. B. (1986), Properties of vacuum tunnelling currents: anomalous barrier heights, *IBM J. Res. Develop*, **30**, 455-459.

D. M. Eigler, C.P. Lutz , and W. E. Rudge (1991), An atomic switch realised with scanning tunnelling microscope, *Nature*, **352**, 600-603.

D.S.I.R. (1920), 'Report of the Lubricants and Lubrication Inquiry Committee', Advisory Council of the Department of Scientific and Industrial Research, HMSO, London.

Decker, D. L., Bennett, J. M., Soileau, M. J., Porteus, J. O., and Bennett, H. E. (1978), Surface and optical studies of diamond-turned and other metal mirrors, *Opt. Engr.*, **17**, 160-166.

Deslattes, R. D. and Henins, A. (1973), X-ray to visible wavelength ratios, *Phys. Rev. Lett.*, **31**, 972-975.

Desmond, F. M. (1975), Principles and applications of tribology, Pergamon Press.

Downs, M. J., McGivern, W. H., and Ferguson, H. J. (1985), Optical system for measuring the profiles of super-smooth surfaces, *Prec. Eng.*, **7**, 211-215.

Dowson, D. (1979), *History of Tribology*, Longman.

Dowson, D. (1991), Friction and traction in lubricated contacts, *NATO ASI*, 325-349.

Drake, B., Prater, C. B., Weisenhorn, A. L., Gould, S. A. C., Albrecht, T. R., Quate, C. F., Cannell, D. S., Hansma, H. G., and Hansma, P. K. (1989), Imaging crystals, polymers and processes in water with the atomic force microscope, *Science*, **243**, 1586-1589.

Driscoll, R. J., Youngquist, M. G., and Baldeschwieler (1990), Atomic-scale imaging of DNA using scanning tunnelling microscopy, *Nature*, **346**, 294-296.

Dunlap, D. D. and Bustamante, C. (1989), Imaging of single-stranded nucleic acids by scanning tunnelling microscopy, *Nature*, **342**, 204-206.

Eigler, D. M. and Schweizer, E. K. (1990), Positioning single atoms with a scanning tunnelling microscope, *Nature*, **344**, 524-526.

Eigler, D. M., Lutz, C. P., and Rudge, W. E. (1991), An atomic switch realized with the scanning tunnelling microscope, *Nature*, **352**, 600-603.

Elewa, I. and Koura, M. M. (1986), Importance of checking the stylus radius in the measurement of surface roughness, *Wear*, **109**, 401-410.

Erlandsson, R., McClelland, G. M, Mate, C. M., and Chiang, S. (1988), Atomic force microscopy using optical interferometry, *J. Vac. Sci. Technol.*, A **6**, 266-270.

Feenstra, R. M. and Fein, A. P. (1986), Scanning tunnelling microscopy of cleaved semiconductor surfaces, *IBM J. Res. & Dev.*, **30**, 466-471.

Foster, J. S., Frommer, J. E., and Arnett, P. C. (1988), Molecular manipulation using a tunnelling microscope, *Nature*, **331**, 324-326.

Fowler, R. H. and Nordheim, L. (1928), *Proc. R. Soc. London*, **A119**, 173.

Franse, J. (1990), Manufacturing techniques for complex shapes with submicron accuracy, *Rep. Prog. Phys.*, **53**, 1049-1094.

Fu, J., Young, R. D., and Vorburger, T. V. (1991), Long - range scanning for scanning tunnelling microscopy, *Rev. Sci. Instrum.*, **63** (4), 2200-2205.

Gane, N. (1970), The direct measurement of the strength of metals on a submicrometre scale, *Proc. R. Soc. London, Ser. A*, **317**, 367-391.

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

Gerhrtz, M., Strecker, H., and Grim, H. (1988), Scanning tunnelling microscopy of machined surfaces, *J. Vac. Sci. Technol.* A **6**, 432-435.

Gewirth, A. A, Graston, D. H., and Bard, A. J. (1989), Fabrication and characterisation of microtips for *in situ* scanning tunnelling microscopy, *J. Electroanal. Chem.*, **261**, 477.

Greenwood, J. A. and Tripp, J. H. (1967), The elastic contact of rough spheres, *Trans. ASME, J. Appl. Mech.*, **35**, 153.

Greenwood, J. A. and Williamson, J. B. P. (1966), The contact of nominally flat surfaces, *Proc. Roy. Soc. A*, **295**, 300-319.

Griffith, J. E., Grigg, D. A., Vasile, M. J., Russell, P. E. and Fitzgerald, E. A. (1991), Characterisation of scanning probe microscope tips for linewidth measurement, *J. Vac. Sci. Technol. B* **9** (6), 3586-3589.

Griffith, J. E., Grigg, D. A., Vasile, M. J., Russell, P. E. and Fitzgerald, E. A. (1992), Scanning metrology, *J. Vac. Sci. Technol*, A **10** (4), 674-679.

Guerrero, J. L. and Black, J. T. (1972), Stylus tracer resolution and surface damage as determined by scanning electron microscopy, *J. Eng. Ind.*, 1087-1093.

Halling, J. (1975), *Principles of Tribology*, The Macmillan Press Ltd.

Hamilton, G. M. (1983), Explicit equations for the stresses beneath a sliding spherical contact, *Proc. Instn. Mech. Engrs*, **197** C, 53-57.

Hansma, H. G., Sinsheimer, R. L., Groppe, J., Bruice, T. C., Elings, V., Gurley, G., Bezanilla, M., Mastrangelo, I. A., Hough, P. V. C., and Hansma, P. K. (1993), *Scanning*, **15**, 296.

Hansma, P. K., Cleveland, J. P., Radmacher, M., Walters, D. A., Hillner, P. E., Bezanilla, M., Fritz, M., Vie, D., and Hansma, H. G (1994), Tapping mode atomic force microscopy in liquids, *Appl. Phys. Lett.* **64**, 1738-1740.

Harb, S., Smith, S. T., and Chetwynd, D. G. (1992), Sub-nanometre behaviour of capacitive feedback piezoelectric displacement actuators, *Rev. Sci. Instrum.*, 63 (2), 1680-1689.

Hasan, Z., Andsage, D., Saltz, D, Cartwright, K., and Nayfeh, M. H. (1992), Scanning tunnelling microscope as a high sensitivity radiation detector, *Rev. Sci. Instrum.*, 63 (3), 2099-2100.

Hertz, H. (1896), *Miscellaneous papers*, P146, London: Macmillan.

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

Hirst, W. and Howse, M. G. J. W. (1969), The indentation of materials by wedges, *Proc. R. Soc. London, Ser. A*, 311, 429-444.

ISO 1984, Surface Roughness —Terminology—Part 1: Surface and its Parameters. ISO 4287/1, International Organisation for Standardisation.

Israelachvili, J. N., McGuiggan, P. M. and Homola, A. M. (1988), Dynamic properties of molecularly thin liquid films, *Science*, 240, 189-190.

Jarvis, S. P., Oral, A., Weihs, T. P., and Pethica, J. B. (1993), A novel force microscope and point contact probe, *Rev. Sci. Instrum.*, 64, 3515-3520.

Johnson, K. L. (1970), The correlation of indentation experiments, *J. Mech. Phys. Solids*, 18, 115-126.

Johnson, K. L. (1975), Non-Hertzian contact of elastic spheres, In *The mechanics of the contact between deformable bodies*, Edited by de Pater and Kalker, Delft Univ. Press.

Johnson, K. L. (1982), One hundred years of Hertz contact, *Proc. Instn. Engrs*, 196, 363-378.

Johnson, K. L., Kendall, K. and Robert, A. D. (1971), Surface energy and the contact of elastic solids, *Proc. R. Soc. Lond.*, A 324, 303-313.

Jones, R. V. (1956), Some parasitic deflexions in parallel spring movements, *J. Sci. Instrum.*, 33, 11-15.

Jones, R. V. (1967), The measurement and control of small displacements, *Phys. Bull.*, 18, 325-336.

Joyce, G. C. and Wilson, G. C. (1969), Micro-step motor, *J. Phys. E: Sec. Instrum.*, 2(2), 661-663.

Kaizuka, H. (1989), Application of capacitor insertion method to scanning tunnelling microscopes, *Rev. Sci. Instrum.*, 60 (10), 3119-3122.

Kaizuka, H. and Siu, B. (1988), A simple way to reduce hysteresis and creep when using piezoelectric actuators, *Japaness Journal of Applied Physics*, 27, L773-L776.

Keller, D. (1991), Reconstruction of STM and AFM images distorted by finite-size tips, *Surface Science*, 253, 353-364.

Keller, D. and Franke, F. S. (1993), Envelope reconstruction of probe microscope images, *Surface Science*, 294, 409-419.

Kendall, K. (1969), Ph'D dissertation, Cambridge University, UK.

King, T. G., Preston, M. E., Murphy, B. J. M., and Cannell, D. S. (1990), Piezoelectric ceramic actuators: A review of machinery applications, *Precision Engineering*, **12** (3), 131-136.

Kohno, T. , Ozawa, N., Miyamoto, K., and Musha, T. (1985), Practical non-contact surface measuring instrument with one nanometre resolution, *Prec. Eng.*, 7, 231-232.

Krim, J., Solina, D. H. and Chiarello (1991), Nanotribology of a Kr monolayer: a quartz-crystal microbalance study of atomic-scale friction, *Phys. Rev. Lett.*, **66**, 181-184.

Krim, J., Watts, E. T. and Digel, J. (1990), Slippage of simple liquid films adsorbed on silver and gold sunstrates, *J. Vac. Sci. Technol.*, A 8, 3417-3420.

Kuk, Y. and Silverman, P. J. (1989), Review: Scanning tunnelling microscope instrumentation, *Rev. Sci. Instrum.*, **60**, 165-180.

Landman, U and Luedtke, W. D. (1991), Nanomechanics and dynamics of tip-substrate interactions, *J. Vac. Sci. Technol.*,B **9**, 414-423.

Landman, U, Luedtke, W. D., Burnham, N. A. and Colton, R. J. (1990), Atomistic mechanisms and dynamics of adhesion, nanoindentation, and fracture, *Science*, **248**, 454-461.

Larsen-Basse, J. (1992), Basic theory of solid friction, in ASM handbook, Friction, Lubrication, and Wear Technology, Volume **18**, 27-37.

Lazan, B. J. (1968), Damping of materials and members in structural mechanism, London: Pergamon.

Lee, G., Arscott, P. J., Bloomfield, V. A, and Evans, D. F. (1989), Scanning tunnelling microscopy of nucleic acids, *Science*, **244**, 475-477.

Lee, K. L., Abraham, D. W., Secord, F., and Landstein, L. (1991), Submicro Si trench profiling with an electro-beam fabricated atomic force microscope tip, *J. Vac. Sci. Technol.*, B **9** (6), 3562-3568.

Li, M (1991), Ph'D thesis, Warwick University.

Li, Y. Z., Vazquez, Z., Piner, R., Andres, R. P., and Reifenberger, R. (1989), Writing nanometre - scale symbols in gold using the scanning tunnelling microscope, *Appl. Phys. Lett.*, **54**, 1424-1426.

Lim, M. G., et al. (1990), Polysilicon microstructures to characterise static friction, *Proc.* 3rd *IEEE Workshop on micro electron Mechanical Systems*, Napa Valley, CA.

Lindsay, S. M. and Barris, B. (1988), Imaging deoxyribose nucleic acid molecules on a metal surface under water by scanning tunnelling microscopy, *J. Vac. Sci. Technol.* A **6**, 544-547.

Lindsey, K., Smith, S. T., and Robbie, C. J. (1988), Sub-nanometre surface texture and profile measurement with Nanosurf 2, *Ann. CIRP*, **37**, 519-522.

Ludema, K. C. (1987), Friction, a study in the prevention of seizure, *ASTM, Stand. News*, 54-58.

Majumdar, A. and Bhushan, B. (1990), Role of fractal geometry in roughness characterisation and contact mechanics of surfaces, *ASME J. of Tribology*, **112**, 205-216.

Mamin, H. J., Ganz, E., Abraham, D. W., Thomson, R. E., and Clarke, J. (1986), Contamination-mediated deformation of graphite by the scanning tunnelling microscope, *Phys. Rev B* **34**, 9015-9018

Mandelbrot, B. B. (1967), How long is the coast of Britain? statistical self-similarity, and fractional dimension, *Science*, **155**, 636-638.

Mandelbrot, B. B. (1982), The Fractal Geometry of Nature, W. H. Freeman, New York.

Marti, O., Colchero, J. and Mlynek, J. (1990), Combined scanning force and friction microscopy of mica, *Nanotechnology*, **1**(2), 141-144.

Martin, Y and Wickramasinghe, H. K. (1987), Magnetic imaging by "force microscopy" with 1000 Å resolution, *Appl. Phys. Lett.*, **50**, 1455-1457.

Mate, C. M., Erlandsson, R., McClelland, G. M., and Chiang, S. (1989), Direct measurement of forces during scanning tunnelling microscope imaging of graphite, *Surface Science*, **208**, 473-486.

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

Maugis, D., Desatos-Andarelli, G., Heurtel, A. and Courtel, R. (1976), *ASLE, Trans.*, **21**, 1.

Mavroyannakis, E. G. (1982), Inductive sensor for the detection of low velocity, *J. Phys. E: Sci. Instrum.*, **15**, 80-82.

McClelland, G. M. et al. (1988), Atomic scale friction measured with an atomic force microscope, From *Engineered Materials for Advanced Friction and Wear Applications*, Ed. Smidt, F. A and Blau, P. J., ASM International.

McCord, M. A. and Pease, R. F. W. (1986), Lithography with the scanning tunnelling microscope, *J. Vac. Sci. Technol.*, B **4**, 86-88.

McCord, M. A. and Pease, R. F. W. (1987), Exposure of calcium fluoride resist with the scanning tunnelling microscope, *J. Vac. Sci. Technol.*, B **5**, 430-433.

Montgomery, D. B. (1969), Solenoid magnet design, Wiley Interscience, New York.

Musselman, I. H. and Russell, P. E. (1990), Platinum/iridium tips with controlled geometry for scanning tunnelling microscopy, *J. Vac. Sci. Technol.*, A **8** (4), 3558-3562.

Nakagiri, N. and Kaizuka, H. (1990), Simulations of STM images and work function for rough surfaces, *Japanese Journal of Applied Physics*, **29**, 744-749.

NATO ASI series, *Fundamentals of friction: Macroscopic and Microscopic Processes*, held in July/August 1991, Ed. Singer, I. L. and Pollock, H. M., 569-572.

Nayak, P. R. (1971), Random process model of rough surfaces, *J. Lubr. Thenol. (Trans. ASME)*, **93**, 398.

Nelder, J. A. and Mead, R. (1965), A simplex method for function minimization, *The computer Journal*, **8**, 308-313.

Newcomb, C. V. and Flinn, I. (1982), Improving the linearity of piezoelectric ceramic actuators, *Electronic Letters*, **18** (11), 442-444.

Nyyssonen, D., Landstein, L., and Coombs, E. (1991), Two-dimensional atomic force microprobe trench metrology system, *J. Vac. Sci. Technol.*, B **9** (6), 3612-3616.

Park, Sang-II. and Quate, C. F. (1986), Tunnelling microscopy of graphite in air, *Appl. Phys. Lett.*, **48**, 112-114.

Park, Sang-iL and Quate, C. F. (1987), Theories of the feedback and vibration isolation systems for scanning tunnelling microscope, *Rev. Sci. Instrum.*, 58, 2004-2009.

Pashley, M. D., Pethica, J. B. and Tabor, D. (1984), Adhesion and micromechanical properties on metal surfaces, *Wear*, **100**, 7-31.

Pethica, J. B., Hutchings, R. and Oliver, W. C. (1983), Hardness measurement at penetration depths as small as 20 nm, *Philos. Mag.*, 48 (4), 593-606.

Petrov, N. P. (1883), Friction in machines and the effect of the lubricant, *Inzh. Zh.*, St. Petersburg, **1**, 71-140; **2**, 227-279; **3**, 377-436; **4**, 535-564.

Pohl, D. W. (1986), Some design criteria in scanning tunnelling microscopy, *IBM J. Res. Dev.*, **30**, 417.

Quiney, R. C., Austin, F. R. A., and Sargent, L. B. Jr (1967), The measurement of surface roughness and profiles on metals, *Trans. ASLE*, **10**, 193-202.

Rabinowicz, E. A. (1971), Determination of compatibility of metals through static friction tests, *ASLE Trans.*, 14, 198-205.

Radhakrishnan, V. (1970), Effect of stylus radius on the roughness values measured with tracing stylus instruments, *Wear*, *16*, 325.

Ravi, T. S., Marcus, R. B., and Liu, D. (1991), Oxidation sharpening of silicon tips, *J. Vac. Sci. Technol.*, B 9 (6), 2733-2737.

Reason, R. E. (1970), The measurement of surface texture, in *Modern workshop technology*, Part 2 — Processes. Ed. Baker, H. W., 3rd edition, Macmillan, London.

Retsky, M. W. and Wall, J. (1972), One ampere current supply stable to one part per million per hour, *Rev. Sci. Instrum.*, 43, 384-385.

Roberts, A. D. (1968), Ph'D dissertation, Cambridge University, UK.

Sampson, J. B., Morgan, F., Reed, D. W., and Muskat, M. (1943), Studies in lubrication XII. friction behaviour during the slip portion of the stick-slip process, *J. Appl. Phys.*, 14, 689-700.

Sarma, Copal, Pingali and Jain Ramesh (1992), Surface recovery in scanning probe microscopy, The conference of Machine Vision Applications, Achitectures, and systems Integration, *SPIE*, **1823**, 151-162, Boston, MA, USA.

Sawatari, T. and Zipin, R. B. (1979), Optical profile transducer, *Opt. Eng.*, **18**, 222.

Sayles, R. S. and Thomas, T. R. (1978), Surface topography as a nonstationary random process, *Nature*, **320**, 431-434.

Schneir, J., Sonnenfeld, R., Hansma, P. K., and Tersoff, J. (1986), Tunnelling microscopy study of the graphite surface in air and water, *Phys. Rev.*, B 34, 4979-4984.

Shafei, T., Arnell, R. O. and Halling, J. (1983), *Trans. ASLE*, **26**, 481.

Simmons, J. G. (1963), Generalised formula for the electric tunnel effect between similar electrodes separated by a thin insulating film, *J. Appl. Phys.*, 34, 1793-1803.

Skinner, J. and Gane, N. (1972), Sliding friction under a negative load, *J. Phys. D: Appl. Phys.*, **5**, 2087-2094.

Sleator, T. and Tycko, R. (1988), Observation of individual organic molecules at a crystal surface with use of a scanning tunnelling microscope, *Phys. Rev. Lett.*, **60**, 1418-1421.

Smith, S. T , Chetwynd, D. G. and Jackson, D. (1991), Manufacture of large scale mechanical devices in single crystal silicon by high speed, *SPIE* vol. **1573** (Commercial Applications of Precision Manufacturing at the Sub-Micro Level), 53-61.

Smith, S. T. (1991), Nanoactuators for controlled displacements, in *From Instrumentation to Nanotechnology*, (Ed.) Gardner, J. W. and Hingle, H. T., Gordon and Breach, London.

Smith, S. T. and Chetwynd, D. G. (1990), An optimised magnet-coil force actuator and its application to precision elastic mechanisms, *Proc. Instn. Mech. Engrs.*, **204C**, 243-253.

Smith, S. T. and Chetwynd, D. G. (1992), *Ultroprecision mechanism design*, Gordon and Breach Science Publisher.

Smith, S. T. and Howard, L. P. (1994), A Precision, low force balance and its application to force probe calibration, *Rev. Sci. Instrum.*, **65** (4), 903-909.

Smith, S. T. Harb, S., and Chetwynd, D. G. (1992), Metrological characterisation of polymeric bearings at the nanometre level, *J. Phys. D:Appl. Phys.*, **25**, A240-A248.

Smith, S. T. Harb, S., Eastwick-filed, V., Yao, Z. Q., Bartlett, P. N., Chetwynd, D. G., and Gardner, J. W. (1993), Tibological properties of electroactive polymeric thin film bearings, *Wear*, **169**, 43-57.

Soler, J. M., Baro, A. M., Garcia, N., and Rohrer, H. (1986), Interaction forces in scanning tunnelling microscopy: giant corrugations of the graphite surface, *Phys. Rev. Lett.*, **57**, 444-447

Song, J. F., and Vorburger, T. V. (1991), Stylus profiling at high resolution and low force, *Appl. Opt.*, **30**, 42.

Sonnenfield, J. and Hansma, P. K. (1986), Atomic-resolution microscopy in water, *Science*, **232**, 211-213.

Spalvins, T. and Buzek, B. (1981), Frictional and morphological characteristics of ion-plated soft metallic films, *Thin solid films*, **84**, 267.

Spragg, R. C. and Whitehouse, D. J. (1970), A new unified approach to surface metrology, *Proc. Inst. Mec. Engrs*, **185**, 47-71.

Stedman, M. (1987), Basis for comparing the performance of surface-measuring machines, *Precis. Eng.*, **9**, 149-152.

Stedman, M. (1988), Limits of topographic measurement by the scanning tunnelling and atomic force microscopes, *J. Microscopy*, **152**, (3), 611.

Steidel, R. F. (1989), An introduction to mechanical vibrations, 3rd Ed. John Wiley & Sons.

Stribeck, R. (1902), *Z. Ver. Dt. Ing.* **46**, 1341-1348.

Taniguchi, N. (1983), Current status in, and future trends of, ultroprecision machining and ultrofine materials processing, *Annals CIRP*, **32/2**, 573-582.

Tersoff, J. and Hamann, D. R. (1985), Theory of the tunnelling microscope, *Phys. Rev. B*, **31**, 805-813.

Thomas, T. R. (1982), Ed., *Rough Surface*, Longman, London.

Thomas, T. R. and Sayles, R. S. (1978), Some problems in tribology of rough surfaces, *Tribology Int.*, **11**, 8.

Tower, B. (1883), First report on friction experiments (Friction of lubricated bearings), *Proc. Inst. Mech. Engrs*, Nov. 1883, 632-659.

Turker, R. C. Jr and Meyerhoff, R. W. (1969), An SEM study of surface roughness measurement, *Proc. 2nd Annual SEM Symp.*, Chicago, 389-396.

Utsugi, Yasushi (1990), Nanometre-scale chemical modification using a scanning tunnelling microscope, *Nature*, **347**, 747-748.

Utsugi, Yasushi (1992), Chemical modification for nanolithography using scanning tunnelling microscopy, *Nanotechnology*, **3**, 161-163.

Van-Baak, D. A. (1992), Temperature servomechanism using thermoelectric modules, *Am. J. Phys.*, **60**, 803-815.

Vasile, M. J., Grigg, D. A., Griffith, J. E., Fitzgerald, E. A., and Russell, P. E. (1991a), Scanning probe tips formed by focused ion beams, *Rev. Sci. Instrum.*, **62**, 2167-2171.

Vasile, M. J., Grigg, D. A., Griffith, J. E., Fitzgerald, E. A., and Russell, P. E. (1991b), Scanning probe tip geometry optimized for metrology by focused ion beam milling, *J. Vac. Sci. Technol.*, **B 9**, 3569-3572.

Vorburger, T. V., Teague, E. C., Scire, F. E., and Rosberry, F. W. (1979), Measurements of stylus radii, *Wear*, **57**, 39-49.

Wang, W. L. and Whitehouse, D. J. (1995), Application of Neural Networks to the reconstruction of scanning probe microscope image distorted by finite-size tips, submitted to *Nanotechnology*.

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

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

Whitehouse, D. J. and Achard, J. F. (1969), The properties of random surfaces in contact, in *Surface Mechanics*, (Ed.) Ling, F. F., ASME, New York.

Whitehouse, D. J. and Archard, J. F. (1970), The properties of random surface of significance in their contact, *Proc. R. Soc. London, Ser. A* 316, 97-121.

Whitehouse, D. J., Bowen, D. K., Chetwynd, D. G. and Davies, S. T. (1988), Nano-calibration for stylus-based surface measurement, *J. Phys. E: Sci. Instrum.*, **21**, 46-51.

Whitehouse, D. J. (1994), Handbook of surface metrology, Institute of Physics Pub., Philadelphia & London.

Williams, C. C. and Wickramasinghe, H. K. (1986), Scanning thermal profiler, *Appl. Phys. Lett.*, **49**, 1587-1589.

Williamson, D. E. (1947), Tracer-point sharpness as affecting roughness measurements, *Trans. ASLE*, **69**, 319.

Wyant, J. C., Koliopoulos, C. L., Bhushan, B., and George, O. E. (1984), An optical profilometer for surface characterisation of magnetic media, *ASLE Trans.*, **27**, 101-113.

Wyant, J. C., Koliopoulos, C. L., Bhushan, B., and George, O. E. (1985), Measurement of surface topography of magnetic tapes by Mirau interferometry, *Appl. Opt.*, **24**, 1489-1497.

Xu, Y. and Smith, S. T. (1993), The determination of squeeze film damping in capacitance based cantilever force probes, *Proc. ASPE annual Meeting*, Seattle, Washington, 326-329.

Yamada, H., Fujii, T., and Nakayama, K. (1988), Experimental study of forces between a tunnel tip and the graphite surface, *J. Vac. Sci. Technol.*, A **6** (2), 293-295.

Young, R. (1971), Surface microtopography, *Phys. Today*, **24** (11), 42-49.

Young, R. , Ward, J., and Scire, F. (1971), Observation of metal - vacuum - metal tunnelling, field emission, and the transition region, *Phys. Rev. Lett.*, **27**, 922-924.

Young, R. , Ward, J., and Scire, F. (1972), The topografiner: An instrument for measuring surface microtopography, *Rev. Sci. Instrum.*, **43**, 999-1011.

Ziegler, J. G. and Nichols, N. B. (1942), Optimum settings for automatic controllers, *Trans. ASME*, **64**, 759.

Zong, Q., Inniss, D., Kjoller, K., and Elings V. B. (1993), Fractured polymer/silica fiber surface studied by tapping mode atomic force microscopy, *Sur. Sci. Lett.*, **290**, L688-L692.