Development and testing of a micromachined probe card.

Rosamond, Mark

How to cite:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full Durham E-Theses policy for further details.
Development and Testing
of a Micromachined Probe Card

Mark Rosamond, M.Eng.

School of Engineering
University of Durham

01 SEP 2009

A thesis submitted to the University of Durham
for the degree of Doctor of Philosophy

June 2009
Abstract

This thesis is concerned with the design, fabrication and testing of micro scale probes. The probes were designed to act as temporary electrical connections to allow wafer level testing of integrated circuits. The work initially focused on the creation of free standing nickel cantilevers, angled up from the substrate with probe tips at the free end. These were fabricated using a novel method, combining pseudo grey scale lithography and thick photoresist sacrificial layers. Detailed analysis of the fabrication method, in particular the resist processing and lithography was undertaken and the limitations of the method explored. Arrays of cantilevers’ with 13 μm width at 23 μm pitch and tip co-planarities better than ± 2 μm were demonstrated using this method.

The force-deflection behaviour of the cantilevers was measured using a bespoke experimental rig. For modelling purposes the Young’s modulus of the nickel was estimated from separate tension testing experiments. It was found that simple analytical expressions could accurately predicted the cantilever stiffness providing (i) the length of the cantilever was greater than five times the width and (ii) the width was greater than five times the thickness. Outside of these constraints it was found that the geometric simplifications made in the analytical model introduced significant error. In these cases, finite element models could be used to predict the cantilever stiffness.

It was found that the Young’s modulus measured by tension testing did not equate to that determined from the force-deflection experiments. This was attributed to a variation in the properties of electroplated nickel with deposit thickness and/or systematic errors in the measurement of extension caused by mechanical compliances within the tension testing machine.

Further testing of the cantilevers was performed to determine the force-contact resistance behaviour of nickel and gold coated tips, of smooth or cuspidated textures, against a variety of metal films. It was found that for a contact force range of 50 μN to 3.5 mN it was not possible to obtain consistently stable low resistance contacts for any tip-film combination. This behaviour was traced to the presence of contaminating surface films which could not be consistently fractured. It was concluded that it was not possible to fabricate suspended nickel cantilever at very fine pitches (less than 25 μm) that could provide sufficient contact force to make stable electrical contact.
For this reason the probe design was modified to include a silicone elastomer (PDMS) layer beneath the nickel cantilevers. This layer was found to markedly increase the stiffness of the structure. This unique fabrication process utilised the previously developed pseudo greyscale lithography to fabricate a copper mould which in turn was used to define the PDMS layer. The new fabrication process overcame many of the limitations of the first method and resulted in improved yield and tip co-planarity. Mechanical measurements showed that the force increased in an exponential fashion with displacement and simple contact resistance measurements showed that the new design could provide stable electrical contacts. The contact resistance between smooth nickel tips and gold films was typically below 5 Ω. Likewise the contact resistances to copper and aluminium films were below 15 Ω and 30 Ω respectively. Although the new probe design applied much higher contact forces than the suspended structures (up to a maximum of ~ 100 mN), the damage to the device under test was much less than that produced by traditional tungsten needles. This permitted electrical contact to be made to very delicate substrates such as evaporated gold films on top of a photoresist layer. This was not possible using traditional needle technologies.

An alternative probe design, separate but complementary to the main focus of this work, was developed specifically for testing organic field effect transistors (OFETs). The probe was based on a combination of nickel electroplating and bulk silicon etching. No detailed analysis of the mechanical or electrical performance of the probe was conducted. However device testing showed results similar to those taken using traditional needle probes. In addition, observation of the scrub marks left on the OFETs tested showed that the probe was behaving as intended by applying different forces to different pads.
The copyright of this thesis rests with the author.
No quotation from it should be published without their prior written consent
and information derived from it should be acknowledged.
ACKNOWLEDGMENTS

I would like to thank Prof. David Wood for his guidance, support and encouragement throughout my Ph.D. and the University’s Doctoral Fellowship scheme for sponsoring my studies.

I would also like to thank the various members of the Microsystems research group who have helped me over the years and in particular Dr. Andrew Gallant who has been an enthusiastic mentor and friend.

Lastly I would like to thank my family and friends for their love, patience, help and support.
DECLARATION

I declare that no material in this thesis has previously been submitted for a degree at this or any other university and unless referenced otherwise is the author’s own work.
# Table of contents

Chapter One **Introduction** ........................................................................................................... 1

1.1 Integrated circuit testing ........................................................................................................... 2
1.2 Wafer level probing .................................................................................................................... 3
1.3 Probe cards ............................................................................................................................... 4
1.3.1 Limitations of traditional probe cards ....................................................................................... 5
1.4 MEMS ....................................................................................................................................... 6
1.4.1 MEMS fabrication and probe cards ............................................................................................ 6
1.5 Background work ....................................................................................................................... 7
1.5.1 Stressed silicon Cantilevers ...................................................................................................... 7
1.5.2 Gold and SU-8 bridge structures ................................................................................................. 8
1.6 Thesis organisation .................................................................................................................. 9

Chapter Two **Literature review and probe card design** ............................................................... 12

2.1 Introduction ............................................................................................................................... 13
2.2 Probe card requirements .......................................................................................................... 13
2.3 Commercial probe cards .......................................................................................................... 16
2.3.1 Traditional technologies .......................................................................................................... 17
2.3.2 Advanced technologies ............................................................................................................ 18
2.4 Research and prototype probe cards ....................................................................................... 26
2.4.1 Fritting contacts ....................................................................................................................... 26
2.4.2 Membrane designs .................................................................................................................. 27
2.4.3 Suspended structures - Planar designs ...................................................................................... 28
2.4.4 Suspended structures – Out of plane designs .......................................................................... 33
2.4.5 Other probe card design s ....................................................................................................... 40
2.5 Probe card design .................................................................................................................... 42
2.5.1 Design objectives .................................................................................................................... 42
2.5.2 Comparison between cantilevers and bridges ........................................................................ 46
2.5.3 In-plane and out-of-plane designs .......................................................................................... 49
2.5.4 Tip structures .......................................................................................................................... 50
2.5.5 Design summary ..................................................................................................................... 51
2.5.6 Other structures ..................................................................................................................... 52

Chapter Three **Fabrication** ......................................................................................................... 56

3.1 Introduction ............................................................................................................................... 57
3.2 Selection of Sacrificial material ............................................................................................... 57
3.2.1 Positive photoresist ................................................................................................................. 57
3.2.2 SU-8 ....................................................................................................................................... 58
3.2.3 Guided electroplating ............................................................................................................. 61
3.3 AZ9260 processing .................................................................................................................. 62
3.3.1 Spin curve ............................................................................................................................... 64
3.3.2 Pre-bake solvent removal................................................................. 65
3.3.3 Rehydration .................................................................................... 67
3.3.4 Exposure ........................................................................................ 68
3.3.5 Wall profile ................................................................................... 69
3.3.6 Grey scale lithography .................................................................. 73
3.3.7 Reflow characteristics ................................................................... 79
3.3.8 Increasing resist thermally stability .............................................. 80
3.4 Electroplating .................................................................................... 84
  3.4.1 Faraday’s law ................................................................................. 84
  3.4.2 Deposit thickness uniformity ....................................................... 86
  3.4.3 Smooth nickel ............................................................................... 88
  3.4.4 Cuspidated nickel ......................................................................... 91
  3.4.5 Copper ......................................................................................... 93
3.5 Selective etching of copper in the presence of nickel ......................... 94
  3.5.1 Acetic acid and hydrogen peroxide ............................................. 95
  3.5.2 Ammonia .................................................................................... 96
  3.5.3 Sodium persulphate .................................................................. 98
  3.5.4 Electrochemical etching ............................................................ 99
  3.5.5 Etching conclusion ...................................................................... 100
3.6 Sputter deposition and improving adhesion ...................................... 100
  3.6.1 Bonding strength of metals to glass ............................................ 101
  3.6.2 Sputter deposition of titanium and copper ............................... 102
  3.6.3 Oxygen cleaning of resist residues ............................................ 105
  3.6.4 Adhesion results ......................................................................... 106
3.7 Fabrication of cantilevers ................................................................. 106
  3.7.1 Method 1 .................................................................................. 107
  3.7.2 Method 2 .................................................................................. 111
3.8 Fabrication results ........................................................................... 113
  3.8.1 Comparison of method 1 and method 2 .................................... 115
  3.8.2 Sacrificial structure stability and tip co-planarity .................... 115
3.9 Conclusions .................................................................................... 118

Chapter Four Testing rig design .............................................................. 122

  4.1 Testing of MEMS devices ............................................................... 123
    4.1.1 Mechanical characterisation of thin films and MEMS devices .... 123
  4.2 Design of testing rig ...................................................................... 126
    4.2.1 Hardware design ....................................................................... 127
    4.2.2 Software design ....................................................................... 130
  4.3 Testing rig measurement error estimation and balance spring constant 137
    4.3.1 Displacement sensor and transducer ..................................... 137
    4.3.2 Stage positioning ................................................................. 138
Chapter Five Mechanical testing .................................................................................... 151
5.1 Tension testing samples ........................................................................................................ 152
  5.1.1 Fabrication of tension testing samples ........................................................................ 152
  5.1.2 Tension testing results ..................................................................................................... 153
5.2 Cantilever deflection models .................................................................................................. 159
  5.2.1 Bending of straight cantilevers ...................................................................................... 159
  5.2.2 Assumptions for basic cantilever equation .................................................................. 160
  5.2.3 Deflection due to shear .................................................................................................... 161
  5.2.4 Cantilever under bending moment, shear and axial load .............................................. 161
  5.2.5 Relative contribution of bending, axial load and shear in cantilever deflection .......... 165
  5.2.6 Deflection of wide cantilevers ...................................................................................... 166
  5.2.7 Cantilever deflection due to self weight ....................................................................... 167
5.3 Results ................................................................................................................................... 168
  5.3.1 Cantilever geometry used in models ............................................................................ 168
  5.3.2 Test results ..................................................................................................................... 170
  5.3.3 Comparison of results and models ................................................................................ 173
5.4 Discussion of analytical models ............................................................................................ 178
5.5 Finite element modelling ........................................................................................................ 180
  5.5.1 Comparison between FE and analytical models ............................................................. 180
  5.5.2 Geometrically exact FEA ............................................................................................. 183
  5.5.3 Discussion of finite element models .............................................................................. 186
5.6 Cantilever yield ...................................................................................................................... 187
  5.6.1 Finite element models .................................................................................................. 187
  5.6.2 Yield behaviour results .................................................................................................. 188
  5.6.3 Discussion of yield behaviour ....................................................................................... 190
5.7 Conclusion ............................................................................................................................. 191
  5.7.1 Results and specifications ............................................................................................. 192

Chapter Six Electrical testing ......................................................................................... 194
6.1 Introduction ............................................................................................................................ 195
6.2 Contact resistance theory ..................................................................................................... 195
  6.2.1 Contact mechanics ........................................................................................................ 196
8.1.4 Leakage current ................................................................. 283
8.1.5 Wafer level device testing .................................................. 283
8.1.6 Scrub marks ........................................................................ 285
8.1.7 Conclusions ....................................................................... 285

Chapter Nine Conclusions and further work .................................. 286

9.1 Conclusions ........................................................................... 287
  9.1.1 Results and specification ...................................................... 287
  9.1.2 Discussion ........................................................................ 288
  9.2 Further work ....................................................................... 292
    9.2.1 Fabrication .................................................................... 292
    9.2.2 Mechanical testing ............................................................. 293
    9.2.3 Electrical testing ................................................................. 294
    9.2.4 Other ............................................................................. 294
Chapter One

Introduction

This chapter introduces the basics of IC testing, wafer level probing and MEMS. A synopsis of probe card research undertaken at Durham prior to this thesis is given. Finally, the organisation of the remainder of the thesis is presented.
1.1 Integrated circuit testing

Electrical testing of integrated circuits (ICs) is performed before the chips are packaged. Testing is performed at this stage since the costs of packaging are significant and therefore non-functioning ICs must first be discarded. A typical IC test flow could be:

- ‘Burn in’
- Wafer probe
- Laser repair (only for memory)
- Package
- Final packaged test

‘Burn in’ refers to a process in which the ICs are intentionally stressed to identify the devices that were likely to fail during or shortly after packaging, and the purpose of burn in is to prevent the so called ‘infant mortality’ problem [1]. Burn in tests include applying thermal, electrical or mechanical stresses. Thermal stresses are applied by holding ICs at an elevated temperature for a set period of time and then cycling the temperature repeatedly. Temperatures of -50 °C to 150 °C are routinely used [2]. Temperature stressing is particularly useful in identifying micro-chemical flaws, such as contamination [1]. Mechanical stressing is used to help identify problems such as weak bonds and voltage stressing is used to identify problems such as defects in dielectrics [1].

Burn in is followed by wafer probing in which the electrical functionality of the IC is checked. Probing is also sometimes performed during the burn in period, i.e. at elevated temperatures.

In the case of memory, such as DRAM, the wafer probe is sometimes followed by a laser repair stage. When DRAM is manufactured it is common to build redundancy into the circuitry. In the repair stage, the laser is used to cut polysilicon or metal ‘fuses’ to switch out bad memory elements and switch in good elements.

The working devices, often referred to as ‘known good dies’, are then packaged by flip chip techniques or wire bonding to a chip header. The packaged devices are then tested again before the ICs are shipped.
1.2 Wafer level probing

Wafer level testing is the process of testing the functionality of ICs before the wafer is diced and individual ICs packaged. To perform such testing, temporary electrical connections are required between electronic test equipment and electrical terminals on the wafer's surface. The structures that provide these connections are called probes. A probe card is an assembly of probes arranged in the correct geometric configuration to make all electrical connections simultaneously. Connection between probes and testing equipment is made through a printed circuit board (PCB) sometimes called the prober interface board (PIB).

The electrical terminals on the wafer's surface, to which contact is made, are bond pads or solder balls. Some, if not all, of the structures will subsequently form part of the permanent electrical contact in the packaged device.

A bond pad is typically a sputtered aluminium, copper or electroplated gold rectangular terminal. Wire bonding is used to make connections between the bond pads and device header during the packaging stage.

Solder balls are small spheres of nickel-gold, tin-gold-copper or other alloys formed on top of contact pads. After separating an individual IC from the wafer, each chip is turned upside down and aligned onto a substrate. The substrate, usually a ceramic with appropriate conductive paths, provides the interface between the IC and the PCB to which it will be connected. All the required electrical connections between the IC and substrate are formed by heating the pair, causing the solder balls to melt and form metal bridges between the IC and the substrate. Special non-conductive glue called ‘under fill’ is then injected between the die and substrate to improve mechanical strength and aid thermal conduction. This process is known as ‘flip-chip’ bonding.

Bond pads or solder balls are typically arranged either around the perimeter of an IC (sometimes in more than one layer), in an array pattern or in a line as shown in Figure 1.1. Bond pads are most commonly arranged around IC perimeters or in straight lines as used in memory devices. Both patterns are compatible with wire bonding.

Solder balls are often used when arrays are needed as these patterns are very difficult to wire bond. Such arrangements are common on devices such as microprocessors.
1.3 Probe cards

Probes traditionally take the form of precision-engineered tungsten needles. Tungsten wire is machined to a sharp point by chemical etching and then bent into the required probe shape [3]. The individual probes are then manually assembled [4] to form an array with the required geometry called a probe card. The probe card usually consists of a PCB with a hole at its centre, around which the probes are mechanically attached. The PCB is used to provide electrical connections between the probes and test equipment.

Wafer probing is performed by lowering the probe card until the probe tips contact the wafer’s surface, at which point an additional displacement called an ‘overdrive’ is applied [3]. The overdrive causes the probes to deflect, which applies the force needed to create a reliable electrical contact. The deflection also causes the probe tips to skid along the bond pad’s surface which helps fracture insulating layers such as native oxides.

Probe cards are designed to last for millions of touchdowns. Material is often transferred from the bond pad to the probes during testing which degrades subsequent contact resistances [4]. Therefore probe cards have to be periodically cleaned and probe tips re-aligned. A typical probe card is shown in Figure 1.2.
1.3.1 Limitations of traditional probe cards

As IC features continue to shrink there is a demand for decreasing bond pad (or solder ball) pitch and decreasing scrub areas. There is also demand for reducing testing time by testing more devices simultaneously. This means there is a need to increase contacts per touchdown and probe larger areas. Therefore probe cards must be manufactured with increasing probe numbers at decreasing pitch over larger areas. Since traditional technologies require a degree of manual assembly there is a direct relationship between cost and manufacture time and the number of probes. Therefore probe cards are becoming increasingly expensive and time consuming to both manufacture and repair.

The use of manual assembly limits the precision with which probes can be placed and so makes it difficult to achieve a good uniformity of tip planarity across large areas. This means that larger overdrives are required to guarantee all probes are brought into contact with the wafer, which leads to certain probes applying unacceptably high contact forces.

As the number of probes per touchdown increases there is an additional demand to reduce the contact force of each probe. As the number of probes increases, the total force on the wafer increases which causes the wafer to bow and can even result in micro fractures. As the pressure on wafer real-estate increases, designers have started placing bond pads over active circuit areas. The use of low-k dielectric layers between metal interconnects results in fragile structures which can easily be damaged by high probe forces [4, 5]. Damage caused to the bond pads from large scrub marks can decrease the yield of wire bonding processes [4]. Lastly, as probes become finer to accommodate smaller pitches, the probing force can distort the probe tip causing alignment errors.

The use of a PCB as a substrate in traditional probe cards causes electrical and mechanical problems. The coefficient of thermal expansion of a PCB is very different to that of a silicon wafer and therefore alignment errors can occur during burn in testing. The use of multi-layer PCB substrates and the relatively long length of the probes make controlling electrical crosstalk and parasitic inductances difficult. This limits the testing bandwidth of traditional probe cards.
1.4 MEMS

Micro-electro-mechanical systems (MEMS), or Microsystems as they are sometimes known, refer to a class of microscopic structures made by processes similar to those used in IC manufacture. MEMS devices typically act as either sensors or actuators; common examples include accelerometers used in car air bags, gyroscopes used in game controllers, micro-mirror displays used in overhead projectors and ink jet print heads.

MEMS fabrication techniques are a sub class of micromachining technology. Common techniques include etching, thin film deposition, electroplating, bonding and polishing, all of which are batch processes, i.e. they are applied to the entirety of the substrate area.

Lithography, that is the reproduction of patterns into functional layers, is used to control where such processes are preformed. Photolithography, which is the patterning of special polymers by exposing them to UV through a mask made of transparent and opaque regions, is by far the most common technique.

Etching can be chemical or physical, and wet or dry. Critical factors include the selectivity of an etch to different materials, the rate of material removal and the directionality of the etch. Thin film deposition includes evaporation, sputtering and various chemical processes.

Recently serial techniques, such as laser ablation and electro-discharge machining and guided electro-deposition, which are traditionally considered micromachining but not MEMS processes, are being applied to microsystem device fabrication. Since these techniques are serial they are notionally slower than the batch techniques. However the rate of material removal or deposition is so much greater than the batch equivalent that this may no longer be the case.

MEMS designs can be divided into two general approaches. Bulk machining, where regions of a substrate are removed to leave suspended structures, or surface machining, where functional layers are deposited and patterned on top of a substrate to produce devices. One of the most developed surface machining techniques is ‘lithographie galvaniformung abformtechnik’ or LIGA in which X-rays are used to pattern a thick polymer mould which is then filled with metal by electroplating. The metal structure is often subsequently used as a mould insert for precision plastic injection moulding [6].
1.4.1 MEMS fabrication and probe cards

As will be seen in chapter 2, a MEMS approach, that is a batch manufacturing process using lithographically generated patterns, is widely being adopted as the most viable method of engineering ‘cutting edge’ probe cards. Lithographic batch techniques break the proportional relationship between probe count and cost and time to manufacture which lowers the per pin cost. Lithographic patterning allows precise geometric definition of the probes over large areas with good uniformity. Such processes are more scalable than other micro manufacturing techniques.

1.5 Background work

The following work was conducted at Durham University prior to the work presented in this thesis. The work presented in this thesis is not directly derived from these previous investigations. However, knowledge of this material helped inform the direction of this work.

1.5.1 Stressed silicon Cantilevers

In 2001, Kevin Morey conducted a Master’s project into stressed silicon cantilevers. The fabrication process used lightly doped n-type (100) silicon wafers which were bulk machined to fabricate the cantilever.

The fabrication began by oxidising the wafers and then patterning windows in the front side. These windows defined where the cantilevers were to be formed. Boron was then implanted to create p-type regions. The implantation process resulted in a dopant concentration gradient, where the most concentrated region was formed at the surface of the wafer. A window was then etched in the oxide on the back side of the wafer and ethylene diamine pyrocatechol (EDP) was used to etch through the silicon wafer, releasing the p-doped regions. Since the doping density of the p-type regions was greater than $3 \times 10^{19} \text{ cm}^{-3}$, they were not significantly etched by the EDP.

Boron atoms are smaller than silicon atoms and so boron doping causes a tensile internal stress to develop in silicon. It was this tensile stress gradient that caused the cantilevers to curl up, when they were released by the EDP etch. Once released, remaining oxide layers
were removed in dilute hydrofluoric acid. Cantilevers of 250 to 1500 µm width and up to 7 mm length were fabricated.

The scope of the work did not include a method of electrically isolating the cantilevers or creating appropriate conducive tracks. It was also found that the cantilevers were extremely fragile and often snapped when deflected.

1.5.2 Gold and SU-8 bridge structures

During 2003 - 2004 Dr. Michael Cooke investigated adapting a previously developed tunable capacitor structure [7] to produce a MEMS probe card [8-10]. The basic structure was bridge fabricated either from electroplated metals or SU-8 epoxy. The metal probes consisted of a probe tip, fabricated in the centre of the bridge with a ‘self-limiting sensor’ placed beneath the bridge. By measuring the conductivity between the bridge and ‘self-limiting sensor’ it was possible to determine when the probe became fully deflected (Figure 1.3). Metal structures were made by electroplating through resist moulds, the resist also serving as the sacrificial layer.

![Figure 1.3 Metal probe structures](image)

SU-8 structures were made by a double exposure process. A high dose was used to pattern the anchors, exposing the full depth of the film and then a second low dose exposure was used to define the beam. This dose only caused cross-linking in the top layer of the film, such that when developed the process created a free standing bridge (Figure 1.4).

Electroplated metal structures occupying 100 by 100 µm to 1000 by 1000 µm were fabricated. The structures could accommodate a maximum deflection of 3 µm.

Electrical testing of QFN (quad flat no leads) packaged RF switches was performed using gold electroplated probes. The estimated contact resistance between probe and package was 0.5 Ω when a load of 4 mN was applied.
1.6 Thesis organisation

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
</tr>
<tr>
<td>2</td>
<td>Literature review and probe card design</td>
</tr>
<tr>
<td>3</td>
<td>Fabrication</td>
</tr>
</tbody>
</table>
This chapter begins with a general discussion of MEMS testing. Then the development and characterisation of a bespoke testing rig is given. Finally the fabrication of a 'test tip' required for electrical measurement is described.

5 Mechanical testing

This chapter begins with fabrication and tension testing of electroplated nickel samples. Next, models for the force-deflection behaviour of the cantilever probes are derived. Then force-deflection results are given and compared to modelled results. Finally conclusions on the validity of the models are made and the results are compared to the original design specification.

6 Electrical testing

This chapter begins with a review of theory concerning electrical contact resistance. A series of results for the contact resistance-force behaviour of many different metal-tip pairs are then given and discussed with reference to the previously presented theories. The test results for leakage currents are given before conclusions about the nature of the contacting interfaces are summarised. Finally, the results are compared to the original design specification.

7 PDMS device

This chapter begins with a review of the problems encountered with the fabrication and with the mechanical and electrical performance of the probes described in chapters 3, 5 and 6. An alteration to the fabrication process is then presented which overcame many of the problems. Comparison between the original specification and new design is then made.
This chapter describes the fabrication and testing of a probe design very different to the one described by the main body of the thesis. This probe was designed to allow testing of individual organic field effect transistors at the wafer level.

In this chapter an overall assessment of work is presented. Areas where further investigation is required are then identified.

References

Chapter Two

Literature review and probe card design

This chapter is divided into two sections; first a review and then a design section. The review section contains probe card requirements as well as a review of designs from both commercial and academic sources. The design section specifies the probe card objectives and presents the rational of the design approach.
2.1 Introduction

This chapter begins with a general appraisal of probe card requirements taken from the International Technology Roadmap for Semiconductors (2007). A review of probe card designs, taken from both commercial and academic sources, is given. The pros and cons of various designs are discussed, before the design objectives and rational for this work is stated.

2.2 Probe card requirements

The International Technology Roadmap for Semiconductors (ITRS) : Test and Test Equipment [1] defines various categories of device, each with different testing requirements. Table 2-1 summarises the parameters that are relevant to probe card design, for each type of device. Additional challenges, not listed in Table 2-1, identified in the ITRS are:

- Supporting bandwidths of up to 40 GHz (for RF devices) by control of probe inductance and positioning of bypass capacitors
- Probing at fine pitches over a temperature range of - 50 to + 150 °C (for burn in testing)
- Reduction of probe cleaning frequency
- Reduction in force per pin
- Probes compatible with copper bond pads (including oxidation considerations)
- Improving probe tip co-planarity over large areas
- Reduction in the time to manufacture for high pin count probe cards
- Reduction in the cost of high pin count cards
- Support of higher switching currents

Of these, the reduction in probing force, cost and time to manufacture seem the most universal and pressing challenges.
Table 2-1 Summary of ITRS probe card requirements for different device types [1]

<table>
<thead>
<tr>
<th>Table 2-1</th>
<th>Year</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2015</th>
<th>2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microprocessor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond pad - inline pad pitch (µm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solder ball - array pad pitch (µm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond pad X dimension (µm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond pad Y dimension (µm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solder ball diameter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scrub area (% of pad) bond pad</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scrub area (% of pad) solder ball</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area probed (mm²)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of contacts per touchdown</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum contact resistance (Ω)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum series resistance (Ω)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum current (mA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum leakage current (µA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Application specific integrated circuit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond pad - inline pad pitch (µm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solder ball - array pad pitch (µm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond pad X dimension (µm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond pad Y dimension (µm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solder ball diameter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scrub area (% of pad) bond pad</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scrub area (% of pad) solder ball</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area probed (mm²)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of contacts per touchdown</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum contact resistance (Ω)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum series resistance (Ω)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum current (mA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum leakage current (µA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond pad - inline pad pitch (µm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solder ball - array pad pitch (µm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond pad X dimension (µm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond pad Y dimension (µm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solder ball diameter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scrub area (% of pad) bond pad</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scrub area (% of pad) solder ball</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area probed (mm²)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of contacts per touchdown</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum contact resistance (Ω)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum series resistance (Ω)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum current (mA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum leakage current (µA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RF and mixed signal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond pad - inline pad pitch (µm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solder ball - array pad pitch (µm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond pad X dimension (µm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond pad Y dimension (µm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solder ball diameter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scrub area (% of pad) bond pad</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scrub area (% of pad) solder ball</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Area probed (mm²)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of contacts per touchdown</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum contact resistance (Ω)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum series resistance (Ω)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
As mentioned in chapter 1, current probe forces can damage active circuitry beneath the bond pads, reduce wire bonding yields and cause substantial wafer bow. The time to assemble probe cards is becoming a considerable bottle neck in IC manufacture. Probe cards cannot be designed until the final layout of bond pads is known, i.e. the mask design for the IC must be complete. Currently, high pin count probe cards can take longer to design and manufacture than it takes to fabricate the first ICs [1].

Due to long manufacturing times, test engineers must stock numerous spare probe cards since there is a high probability of a card being damaged during use. The use of semi-manual assembly techniques means that there is a proportional relationship between card cost and probe count. This fact is becoming an unbearable cost to the IC industry.

There are other electronic device types, not covered by the ITRS, which have very specific probe card requirements. For instance displays, such as LCDs, require extremely large probe cards, often measuring several metres across. Plastic electronic devices often require high test voltages (10 to 100 V) and only generate signals in the low mA to nA. Therefore a probe card with very low leakage current (low pA) is required for testing plastic electronics. Additionally, the contact force applied by probes should be low since the substrate materials are very soft. However, the test signal bandwidths required are typically less than 100 kHz and therefore impedance control is less critical than in silicon testing.
2.3 Commercial probe cards

Commercial probe cards can be divided into traditional and advanced technologies. The traditional technologies include epoxy needle and ceramic blade designs whilst the advanced technologies include membrane, vertical and MEMS based solutions.

Epoxy type designs still account for a large percentage of the cards in use today. Epoxy designs continue to be economically competitive for low pin counts as they are a robust and well understood technology.

![Graph showing the probe card market](image)

**Figure 2.1 Current and projected probe card market (based on data from 2005)** [2]

<table>
<thead>
<tr>
<th>Company</th>
<th>Revenue $M (including service and spares)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2003</td>
</tr>
<tr>
<td>FormFactor</td>
<td>98.3</td>
</tr>
<tr>
<td>JEM (Japan Electronic materials)</td>
<td>82.3</td>
</tr>
<tr>
<td>Kulicke &amp; Soffa *</td>
<td>56.7</td>
</tr>
<tr>
<td>MJC (Micronics Japan Co., Ltd)</td>
<td>54.1</td>
</tr>
<tr>
<td>TCL (Tokyo Cathode Laboratory Co., Ltd)</td>
<td>23.4</td>
</tr>
<tr>
<td>SV Probe Inc.</td>
<td>15.2</td>
</tr>
<tr>
<td>Wentworth Laboratories, Inc.</td>
<td>14.4</td>
</tr>
<tr>
<td>Technoprobe Corporation</td>
<td>14.3</td>
</tr>
<tr>
<td>Cascade Microtech Inc.</td>
<td>13.4</td>
</tr>
<tr>
<td>Philcom Corporation</td>
<td>-</td>
</tr>
<tr>
<td><strong>Total market value</strong></td>
<td>478</td>
</tr>
</tbody>
</table>

* Sold wafer test division to SV probe in Jan 2006

**Table 2-2 Probe card market by company (2008 data projected)** [3]

Data for the probe card market is shown in Figure 2.1 and Table 2-2. It demonstrates continuing year on year growth of the probe card market; however, the share for traditional technologies is declining.
2.3.1 Traditional technologies

The first design of probe card was the epoxy needle type, first seen in 1969 [4] and still in widespread use today. This design consists of tungsten, tungsten-rhenium or copper-beryllium needles arranged around a hole such that the needles form cantilevers (see Figure 2.2). The tips of the needles are aligned to match the required pattern of bond pads on the wafer to be tested. The needles are fixed to a ceramic ring by epoxy which is then attached to the centre of a probe interface board (PIB) to which the probes are electrically connected. Electrical connection between the probes and test equipment is then made through the PIB. Epoxy probe cards cost between $10 and $20 per probe [5].

A variation of the epoxy probe card is the ceramic blade card (Figure 2.3). This technology is used when superior mechanical robustness or a tightly controlled impedance environment is required. Each probe is mounted on a separate ceramic blade which is then fixed to a PIB. By fabricating micro-strips on the blade it is possible to control the impedance of each conduction path. This is difficult with epoxy designs as it requires the use of special co-axial probes. A disadvantage of the ceramic blade design is that it is not possible to produce high probe count cards.

![Figure 2.2 An epoxy probe card](image)

![Figure 2.3 A ceramic blade probe card](image)
Blade probe cards can only be used for probing perimeter type layouts however epoxy probes can be fabricated in several tiers to allow limited area array probing capabilities. This is shown in Figure 2.4.

Figure 2.4 Conventional perimeter and area array epoxy probe cards

These traditional technologies are becoming increasingly obsolete for cutting edge applications due to the time and cost of manufacturing ultra high pin counts fine pitch cards. However, they represent a simple proven technology which is preferred for less demanding applications.

2.3.2 Advanced technologies

2.3.2.1 Membrane probe card

The membrane probe card was developed in 1988 [4] to meet the requirements of high speed device testing. Impedances are carefully controlled by having a micro-strip type structure. The card consists of nickel probe tips fabricated on a flexible membrane which is pressed into contact with the wafers surface using a plunger or air pressure (Figure 2.5).

Figure 2.5 A membrane probe card
A membrane probe card applies less contact force than either a cantilever or vertical type (see below). A comparison of the pad damage caused by membrane, vertical and cantilever technologies is shown in Figure 2.6.

![Figure 2.6 Pad damage versus technology](image)

Membrane probe cards cost between $40 and $80 per pin [5]. This technology is the most popular choice of high speed testing of RF or mixed signal devices. As described in section 2.2 these types of devices do not require the huge number of probes needed for memory or microprocessor testing.

### 2.3.2.2 Vertical probe cards

The vertical probe card was patented by IBM in 1977 [6]. This type of card is used for probing area array layouts and is commonly used to probe solder balls in flip-chip bonded devices. Vertical probe cards are used with small overdrives that cause the probes to buckle, producing a spring effect without a change in the probe tip X-Y alignment (see Figure 2.7). This is important for solder ball probing or else the probe would skid and loose contact.

![Figure 2.7 A vertical probe card](image)
ceramic metal layered structure which represents the bulk of the manufacturing time and cost of a vertical probe card. Vertical probe cards cost between $30 to $60 per probe [5].

'Traditionally' manufactured vertical probes are starting to be replaced by lithographically defined contacts as will be described next. Using MEMS techniques allows much better geometric and therefore probing force control. Semi-automatic MEMS type production techniques, such as those from Form Factor Inc., (see section 2.3.2.3 below), have become increasingly popular.

2.3.2.3 MEMS based probe cards

There are a wide variety of commercially available designs based on MEMS technology. There are two general approaches, either to use MEMS technology to fabricate individual probes which are later assembled by traditional methods, or to use MEMS technology to build complete cards. A brief review of commercially developed MEMS based designs is given below.

Advantest Corporation

Advantest offer a probe card technology based on MEMS techniques called SiliconFingers (Figure 2.8). SiliconFingers consist of arrays of metal tracks backed by a silicon skeleton. The silicon provides suitable mechanical and thermal behaviour whilst the metal provides the electrical path [7]. The arrays of probes are assembled into a ceramic holder and suitable electrical connections are made by wire bonding to create the final probe card. Probe pitches of 60 µm have been demonstrated.

Figure 2.8 Advantest SiliconFingers [8]
Advantest also use photolithographic techniques to produce individual contactors such as the PhotoFinger shown in Figure 2.9. These are then assembled by pick and place into a carrier substrate to produce a vertical probe card.

![Advantest PhotoFinger](image)

**Figure 2.9 Advantest PhotoFinger [7]**

**Sumitomo Corporation**

Sumitomo use LIGA techniques to define individual nickel alloy probes [9] which are then assembled into a vertical probe card. Each probe is between 2 and 3 mm in length and can be assembled to a 120 μm pitch. SEM images of such probes are shown in Figure 2.10.

![Sumitomo nickel alloy probes](image)

**Figure 2.10 Sumitomo nickel alloy probes**
Japan Electronic Materials (JEM) Corporation

JEM VH series HAWK technology is similar to that of Advantest (see Figure 2.11). Individual contactors are produced using lithographic methods and then assembled into full cards. JEM produce a 80 μm pitch line-on-centre card [6].

![Figure 2.11 JEM HAWK contactor [6]](image)

JEM have also demonstrated a 24 μm straight pitch and 20 μm staggered pitch [10] probe card fabricated entirely by batch MEMS techniques. A two stage sacrificial layer is used to get the cantilever to rise up from the substrate as shown in Figure 2.12.

![Figure 2.12 JEM micro-machined probe card [10]](image)

![Figure 2.13 Mechanical response of a JEM micro-machined probe [10]](image)
The probes are designed to be deflected up to 40 \( \mu \text{m} \) at which they provide contact forces between 3.9 mN and 5.9 mN. Force versus deflection behaviour is shown in Figure 2.13.

**Phicom**

Phicom produce a ‘MEMS unit’ probe card that is suitable for probing contact pads at greater than 65 \( \mu \text{m} \) pitch and produces contact resistances of less than 5 \( \Omega \). The card is produced using a combination of photolithography, wet and dry etching and electroplating [11]. The tip of an individual cantilever is shown in Figure 2.14.

![Figure 2.14 Phicom’s MEMS cantilever [11]](image)

**Advanced Micro Silicon Technology (AMST)**

AMST manufacture a planar silicon probe along which they deposit chrome-gold tracks to carry electrical signals. Their cards can probe at pitches of 35 \( \mu \text{m} \) for staggered line-of-centre pad layouts. An SEM of the silicon probes is shown in Figure 2.15.

![Figure 2.15 AMST silicon cantilevers [12]](image)
SCS Hightech Inc.

SCS have developed a probe card called the ‘Diamond Vespa’. They fabricated specially shaped nickel tips on top of ceramic substrate called the ‘acippiter structure’. They produce probes for area array probing and can probe down to 150 μm solder ball pitch [13]. The nickel probes are shown in Figure 2.16.

![Figure 2.16 SCS Hightech Diamond Vespa probe card [13]](image)

Microfabrica

Microfabrica are a general MEMS company who produce devices using their EFAB™ process. The process is based on repeated through resist electroplating to define structural material, blanket depositing a sacrificial material and planarising [14]. Amongst their portfolio is a 36 layer micro spring (see Figure 2.17) as well as a multi-layered cantilever which can be assembled into a vertical probe card with 50 μm pitch (see Figure 2.18) [14].

![Figure 2.17 Microfabrica 36 layer microsprings [14]](image)
FormFactor Inc.

FormFactor manufacture a large variety of probe cards including area array, line-on-centre and perimeter probe designs. The basis of their probe cards is a contact technology they call ‘Microsprings’. The technology is based on wire bonding combined with electroplating [15-17] and is shown in Figure 2.19.

2.4 Research and Development

2.4.1 Printing contacts

At FormFactor, the printing of contact patterns is performed in a manner that maximizes the performance of the resulting probe cards. The process involves using a wire bonder to define the structure of the MicroSprings using gold wire. Then, a spark is used to cut the gold wire into a free-standing structure. Finally, the gold wires are overcoated with a nickel alloy to enhance their durability and electrical conductivity. This process allows for the creation of high-quality probe cards with consistent and reliable performance.

Figure 2.19 FormFactor's microspring process
Microspring technology is also proposed as an interconnect method for wafer level packaging. Images of fabricated microsprings are shown in Figure 2.20.

In 2007 Formfactor released a probe card called ‘Harmony’ capable of probing an entire 300 mm wafer in one touchdown using 40,000 contactors [18]. FormFactor can probe 40 µm pitch line-on-centre and perimeter layouts and 175 µm pitch area array layouts [19].

![Figure 2.20 Formfactor Microsprings](image)

2.4 Research and prototype probe cards

The following section contains a review of academic publications relating to the design of MEMS based probe cards. The first section explains what is meant by ‘fritting’, a process that is often used in MEMS probe cards. This is followed by a summary of important publications. The summary is divided into membrane, suspended planar, suspended non-planar and other probe card designs.

2.4.1 Fritting contacts

At low contact forces it is often not possible to obtain low resistance electrical connections since native oxides and other insulating layers on the contacting metal surfaces are not broken. Fritting is a process of electrical breakdown that can allow a low resistance contact to be formed even at low contact forces, and is often employed by MEMS based probe cards. Fritting consists of applying a high strength field across a contact surface until a current begins to flow, the maximum value of this breakdown current being limited to a defined value. Generally a high strength electric field is applied by bringing two closely spaced electrodes onto the contact pad and applying a voltage between them. After the fritting has occurred, one of the probes is subsequently used to carry the electrical test signal to the contact pad. The higher the maximum current that is allowed to flow during electrical
breakdown, the lower the subsequent contact resistance [20]. The voltage that is required to cause fritting depends on the thickness and nature of the insulating layers and is found to decrease with increasing contact load [20]. For example, a fritting voltage of approximately 5 V is required to obtain a low resistance contact at 1 mN load whilst a voltage of 20 V is required at a contact load of 10 μN when contacting aluminium films [20, 21].

2.4.2 Membrane designs

In 1995 Beiley et. al. [22, 23] reported on the fabrication and testing of a membrane probe card designed for array probing. The design consisted of a 3.2 μm thick polyimide membrane with embedded aluminium tracks suspended over a hole, etched through a silicon wafer. Tungsten was deposited by chemical vapour deposition (CVD) into vias at the end of the aluminium tracks to form probe tips. The authors found that the polyimide film possessed a tensile stress. This was necessary since a compressive stress would have caused distortion of the membrane on release from the silicon wafer.

Testing of the probe card was performed by using air pressure to deflect the membrane, forcing it into contact with patterned aluminium tracks. The authors found that it was necessary to use fritting to achieve contact resistances less than 2 Ω since the mechanical pressure was insufficient to disrupt the native oxide. A fritting voltage was 1 V per contact and a maximum current of 50 mA was used.

Increasing the probe force by increasing the air pressure was found to have very little effect on the contact resistance and, instead, the fritting parameters determined the electrical area of contact. However, once a low resistance contact was formed it was not possible to increase the area further (continue to lower the resistance) since the fields required to break down the remaining insulating film could not be generated.

The authors found that replacing a single probe tip with an array of smaller tips (of equivalent area) was marginally beneficial in mechanically disrupting native aluminium oxide layers. However, they concluded that mechanical fracture was not as reliable as using the electrical fritting processes.

As stated in section 2.3.2.1 the membrane probe card has found commercial success particularly in high speed testing due to the ability to control the impedance environment right up to the probe tip. However, disadvantages include the complexity required to apply pressure to the back of the membrane and lack of mechanical compliances. Since all probes are mechanically connected to a single membrane it is not possible to accommodate a large
variation in the height of neighbouring probes. For instance, a membrane design is not well suited to probing solder balls since their fabrication process leads to balls of varying size.

2.4.3 Suspended structures - Planar designs

A summary of important characteristics of planar probe designs reviewed are given in Tables 2-3 and 2-4. Table 2-3 contains a description of the designs and fabrication details whilst Table 2-4 list key performance details. The fabrication methods range from very simple surface machining [24] to extremely complex methods based on deep reactive ion etching (DRIE), electroplating, soldering, chemical mechanical polishing and wet etching [25, 26]. Silicon or electroplated metals were the typically used structural materials, whilst silicon was the most popular sacrificial material. A selection of probes is shown in Figure 2.21.
## Table 2.3 Summary of planar probe designs

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Description of probe structure</th>
<th>Probe layout</th>
<th>Sacrificial material</th>
<th>Release etch</th>
<th>Vias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Park et. al. [30]</td>
<td>2000</td>
<td>Aluminium coated, thermally oxidised silicon cantilevers with 35 μm high flat topped pyramid tips</td>
<td>Perimeter</td>
<td>Porous silicon</td>
<td>NaOH</td>
<td>Yes</td>
</tr>
<tr>
<td>Park et. al. [27]</td>
<td>2002</td>
<td>Silicon cantilevers with sharp probe tips</td>
<td>Line-on-centre</td>
<td>Silicon (substrate)</td>
<td>TMAH etching of exposed (100) planes of (111) substrate</td>
<td>No</td>
</tr>
<tr>
<td>Kim et. al. [31]</td>
<td>2002</td>
<td>Nickel – copper – gold coated polysilicon cantilevers</td>
<td>Line-on-centre</td>
<td>Silicon (substrate)</td>
<td>XeF₂ etching to undercut</td>
<td>Yes</td>
</tr>
<tr>
<td>Kim et. al. [32]</td>
<td>2004</td>
<td>Electroplated nickel cantilevers</td>
<td>Line-on-centre</td>
<td>Silicon (substrate)</td>
<td>TMAH etching to undercut</td>
<td>Yes</td>
</tr>
<tr>
<td>Krüger et. al. [24]</td>
<td>2004</td>
<td>Three layer Nickel-tungsten alloy beam structures. All structures were anchored at three points and had a central probe tip (gold coated)</td>
<td>Array</td>
<td>Electroplated copper</td>
<td>“wet etch”</td>
<td>No</td>
</tr>
<tr>
<td>Cho et. al. [33]</td>
<td>2004</td>
<td>Gold coated silicon cantilevers (tungsten coated tips) cut and bonded to a PCB</td>
<td>Line-on-centre</td>
<td>n.a.</td>
<td>n.a.</td>
<td>No</td>
</tr>
<tr>
<td>Kataoka et. al. [28]</td>
<td>2004</td>
<td>Vertical “S” shaped nickel springs</td>
<td>Array</td>
<td>Negative photoresist</td>
<td>“wet etch”</td>
<td>No</td>
</tr>
<tr>
<td>Namazu et. al. [29]</td>
<td>2007</td>
<td>Silicon cantilevers with gold tracks for electrical signals and titanium – nickel alloy shape memory loops for thermal actuation. Sharp pyramid tips</td>
<td>Perimeter</td>
<td>Silicon (membrane)</td>
<td>RIE etching from back</td>
<td>No</td>
</tr>
<tr>
<td>Kim et. al. [25]</td>
<td>2007</td>
<td>Nickel-cobalt alloy cantilevers with pyramid shaped nickel-tungsten alloy probe tips. The cantilevers were contained within a silicon guiding structure</td>
<td>Staggered line-on-centre</td>
<td>Silicon (substrate)</td>
<td>DRIE from back</td>
<td>Yes</td>
</tr>
<tr>
<td>Kim et. al. [26]</td>
<td>2008</td>
<td>Nickel-cobalt alloy cantilevers soldered to a ceramic substrate</td>
<td>Lin-on-centre</td>
<td>Silicon (patterned wafer acts as electroplating mould)</td>
<td>KOH used to dissolve entire wafer</td>
<td>Yes</td>
</tr>
<tr>
<td>Author</td>
<td>Contact resistance (Ω)</td>
<td>Minimum pitch (µm)</td>
<td>Maximum deflection (µm)</td>
<td>Contact force (mN)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------</td>
<td>--------------------</td>
<td>-------------------------</td>
<td>--------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Park et. al. [30]</td>
<td>&lt; 2.0 (measured against metal pads)** stable for 200,000 touchdowns</td>
<td>12.5</td>
<td>80</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Park et. al. [27]</td>
<td>&lt; 0.5 (measured against aluminium)</td>
<td>70</td>
<td>70</td>
<td>16.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kim et. al. [31]</td>
<td>&lt; 0.5 (measured against gold-tin solder balls)</td>
<td>13.5</td>
<td>50</td>
<td>4,000 to 10,000***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Krüger et. al. [24]</td>
<td>0.4 (measured against aluminium)</td>
<td>1000</td>
<td>65</td>
<td>some probe broke during test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cho et. al. [33]</td>
<td>&lt; 3.0 (measured against aluminium)</td>
<td>40</td>
<td>40</td>
<td>20 to 40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kataoka et. al. [28]</td>
<td>&lt; 2.0 (measured against Cu, 2V fretting)</td>
<td>40</td>
<td>10 to 40</td>
<td>30 (multiple tips per cantilever)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Namazu et. al. [29]</td>
<td>2 (measured against aluminium)**</td>
<td>250</td>
<td>30</td>
<td>60 (single tip per cantilever)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kim et. al. [25]</td>
<td>0.24 (measured against gold)</td>
<td>10</td>
<td>30</td>
<td>0.3, ****</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kim et. al. [26]</td>
<td>0.31 (measured against gold)</td>
<td>15</td>
<td>50</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2-4 Summary of planar probe card performance
Cantilevers were not electrically isolated from one another. The probing action was as follows: The probe was brought into contact with the aluminium until it applied a force of 130 μN (no electrical contact). At this point the shape memory alloy was driven with a 1 Hz pulse for 6 seconds thereby applying a pulsed force. After this period the contact resistance was measured. The authors found after this procedure that the contact resistance varied little over a force range of 20 to 350 μN.

Probe was designed to produce 120 mN of force however much higher forces were applied during electrical testing. 300 μN total contact force was made up from 170 μN produced from shape memory alloy (370 mW input power) and 130 μN from probe deflection.

Many authors fabricated their devices on, or from the silicon substrate. These structures were then released by removing the substrate region beneath the device by a variety of etching methods.

In 2000 Park et. al. [30] used a silicon on insulator (SOI) wafer to create probes. The cantilevers were fabricated from the silicon above the insulator whilst a porous region formed in the silicon below the insulator acted as the sacrificial material.

In 2002 Park et al. [27] used a (111) silicon wafer to form suspended cantilevers with sharp probe tips. The fabrication used isotropic and anisotropic dry etching to fabricate the probe tips followed by DRIE to define the cantilevers. Silicon nitride was then blanket deposited and etched from horizontal surface by anisotropic RIE before further DRIE. The exposed vertical side walls produced by the further DRIE defined the depth of the sacrificial region, which was then removed by lateral etching the wafer in tetramethyl ammonium hydroxide (TMAH).

In 2002 Kim et. al. [31] used isotropic xenon difluoride etching to release patterned polysilicon cantilevers from their silicon substrate. Similarly in 2004 Kim et. al. [32] formed suspended nickel cantilevers by undercutting the structures by etching the (100) silicon substrate in TMAH.

In 2007 Namazu et. al. [29] used TMAH etching from the back of a (100) silicon wafer to produce a silicon membrane. On this membrane further wet etching was used (from the front) to define a cantilever with sharp probe tip. Shape memory alloy (SMA) actuators were then deposited, annealed and patterned after which a gold signal track was deposited and patterned. The completed cantilevers were then released from the membrane by RIE thinning from the back side of the wafer. The SMA was annealed at 600 °C before the cantilevers
were released so that it ‘memorised’ that position. Subsequently, when the cantilever was mechanically deflected during probing the SMA could be heated to increasing the contact force of the cantilever.

In 2007 Kim et. al. [25] demonstrated a probe card consisting of cantilevers with a thickness greater than their width. By decreasing the cantilever width the authors were able to achieve finer tip to tip pitches. However the cantilevers became mechanically unstable. To prevent mechanical instability the cantilevers were fabricated inside silicon trenches to guide them such that they could only move vertically.

The fabrication was extremely complex and was based on DRIE, electroplating, soldering, chemical mechanical polishing and wet etching. The cantilevers were fabricated from nickel cobalt alloy with copper vias for back side electrical connection. The cantilevers had a nickel probe bump onto which was soldered a nickel-tungsten probe tip using lead-tin solder. The nickel-tungsten tip was formed by electroplating metal into a potassium hydroxide (KOH) etched pyramid shaped pit in a silicon wafer. The width of the base of the nickel-tungsten tip was greater than the width of the cantilever guiding trench. This prevented the cantilevers being vertically deflected by more than a defined amount.

The authors tested the cantilevers and found that the nickel-tungsten tips eroded by 0.3 μm after 20,000 touchdowns. The leakage current measured at 5 V between adjacent cantilevers was 10 pA. The Probe tip x – y alignment was ± 8 μm between the mask position and measured fabricated position. This variation was due to distortion caused during the polishing steps and wafer bow.

Subsequently in 2008 Kim et. al. [26] demonstrated a variation on their 2007 design, although this time there was no vertical guiding structure. The nickel-cobalt cantilevers were fabricated by electroplating into a mould made by wet and dry etching a silicon wafer. The fabricated cantilevers were then bonded to a ceramic substrate by gold-tin solder and the wafer dissolved. During testing, the cantilever tips showed 5 μm of wear over 350,000 touchdowns.

Other sacrificial materials included electroplated copper and photoresist. Krüger et. al. [24] used a process of through resist electroplating of nickel or nickel-tungsten in combination with blanket copper electrodeposition followed by mechanical polishing. The copper was selectively wet etched to leave suspended nickel structures. Kataoka et. al. [28] used through resist nickel electroplating and the resist also served as the sacrificial layers. After each nickel electroplating step, copper was deposited by sputtering to provide a seed layer for subsequent electroplating.
2.4.3.1 Remarks and conclusions

All planar designs suffer from the limitation that the maximum device deflection is limited by the height of the tip structure (see section 2.5.3) unless multiple layers are used such as Kataoka [28]. This means that very high aspect ratios or a large number of deposition steps are required in order to allow reasonable levels of structural deflection. High aspect ratios are most simply achieved by DRIE which is why many of the authors discussed use silicon as a structural material. DRIE is an expensive process and although silicon has many desirable mechanical properties, it is not inherently suitable for the formation of probes since it cannot be made sufficiently electrically conductive and has poor fracture toughness. Additionally, fabrication by bulk machining of the substrate precludes the possibility of integration with wafer level electrical circuitry and ultra low leakage substrates.

Recently, reports of high aspect ratio electroplated metal structures fabricated using DRIE patterned silicon moulds have been published [25, 26]. This method allows the production of high aspect ratio metallic structures however is very complex. The process is still limited, since to release the bonded metal cantilevers the silicon mould must be dissolved which imposes restrictions on the choice of substrate.

2.4.4 Suspended structures – Out of plane designs

The methods used to fabricate non-planar probes are generally less complex than the planar designs. There are two categories of designs, those in which all the fabrication is in plane however the probes are designed such that the released device deflects out of plane due to an inbuilt stress gradient [4, 34-43] and those in which a three dimensional sacrificial layer is fabricated [44-48]. A summary of the probe designs and key fabrication methods are shown in Tables 2-5 and 2-6, whilst Table 2-7 contains key performance parameters.
<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Description of probe structure</th>
<th>Probe layout</th>
<th>Sacrificial material</th>
<th>Release etch</th>
<th>Vias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. [4]</td>
<td>1999</td>
<td>Thermally actuated cantilevers (Al-SiO₂)</td>
<td>Perimeter</td>
<td>Silicon (substrate)</td>
<td>KOH and dry etching</td>
<td>No</td>
</tr>
<tr>
<td>Takahiro et al. [44]</td>
<td>2000</td>
<td>Electroplated metal (Pd or Au) beam with TaSi₂ heater and Au, W or doped diamond probe tips</td>
<td>Line-on-centre</td>
<td>Fluorinated polyimide or silicon (substrate)</td>
<td>TMAH</td>
<td>No</td>
</tr>
<tr>
<td>Lee et al. [45]</td>
<td>2005</td>
<td>Angled metal cantilevers (Al-Au) with a trapezoidal or semicircular form</td>
<td>Line-on-centre</td>
<td>Positive photoresist</td>
<td>Acetone</td>
<td>No</td>
</tr>
<tr>
<td>Tanaka et al. [46]</td>
<td>2008</td>
<td>Elongated &quot;Z&quot; shaped cantilevers (p⁺-Si; Ni) bonded to glass substrate</td>
<td>Perimeter</td>
<td>Silicon (substrate)</td>
<td>EDP + H₂O</td>
<td>Yes</td>
</tr>
<tr>
<td>Wang et al. [47-48]</td>
<td>2008</td>
<td>&quot;Hoe&quot; shaped probes (Ni) bonded to ceramic substrate</td>
<td>Array</td>
<td>Silicon (substrate)</td>
<td>TMAH</td>
<td>Yes</td>
</tr>
<tr>
<td>Kim et al. [34-36]</td>
<td>1998 to 2005</td>
<td>Curved epiflaxial Si/SiO₂-Ni:C-Au cantilevers</td>
<td>Line-on-centre</td>
<td>Porous silicon</td>
<td>NaOH</td>
<td>No</td>
</tr>
<tr>
<td>Lee et al. [37]</td>
<td>2005</td>
<td>Curved Mo-Cr-Al alloy overplated with Mo-Cr-Al benzocyclobutene cantilevers</td>
<td>Line-on-centre</td>
<td>Aluminium</td>
<td>HPO₃</td>
<td>No</td>
</tr>
<tr>
<td>Fork et al. [38]</td>
<td>2004</td>
<td>Curved Mo-Cr alloy overplated with Cu, Ni and Au (+)</td>
<td>Line-on-centre</td>
<td>'Metallic layer'</td>
<td>KOH</td>
<td>No</td>
</tr>
<tr>
<td>Ishihata et al. [39-40]</td>
<td>2003 to 2005</td>
<td>Curved, electrostatically actuated Ni cantilevers</td>
<td>Line-on-centre</td>
<td>Aluminium</td>
<td>KOH</td>
<td>No</td>
</tr>
<tr>
<td>Itoh et al. [41-43]</td>
<td>2006 to 2008</td>
<td>Curved Ni cantilevers</td>
<td>Perimeter</td>
<td>Silicon (substrate)</td>
<td>KOH</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 2-5 Summary of non-planar probe designs

* Devices intended as wafer level packaging solution but could be used as probes
<table>
<thead>
<tr>
<th>Author</th>
<th>Cause of out of plane geometry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et. al. [4]</td>
<td>Intrinsic stress differences in cantilever layers and during actuation from bimorph effect and integrated heater</td>
</tr>
<tr>
<td>Takahiro et. al. [44]</td>
<td>Structuring of the sacrificial layer by etching prior to metal deposition</td>
</tr>
<tr>
<td>Lee et. al. [45]</td>
<td>Structuring of the sacrificial layer by over exposure and development prior to metal deposition</td>
</tr>
<tr>
<td>Tanaka et. al. [46]</td>
<td>Structuring of the silicon substrate prior to cantilever formation (cantilever then bonded and silicon substrate dissolved)</td>
</tr>
<tr>
<td>Wang et. al. [47,48]</td>
<td>Structuring of the silicon substrate prior to cantilever formation (cantilever then bonded and silicon substrate dissolved)</td>
</tr>
<tr>
<td>Kim et. al. [34-36]</td>
<td>Bimorph effect by thermal annealing at 500 °C</td>
</tr>
<tr>
<td>Lee et. al. [37]</td>
<td>Internal stress difference between Mo and BCB layers</td>
</tr>
<tr>
<td>Fork et. al. [38]</td>
<td>Stress gradient in Mo-Cr alloy</td>
</tr>
<tr>
<td>Itoh et. al. [39-40]</td>
<td>Internal stress difference between nickel deposited from Sulphamate and Watts sols</td>
</tr>
<tr>
<td>Tsou et. al. [41-43]</td>
<td>Stress in Ti film deposited on SiO₂, cantilever. Both films are ultimately removed</td>
</tr>
</tbody>
</table>

Table 2-6 Summary of non-planar probe methods

- released cantilevers deflected downwards after release due to film stresses. Thermal actuation caused the cantilevers to deflect upwards
<table>
<thead>
<tr>
<th>Author</th>
<th>Contact resistance (Ω)</th>
<th>Minimum pitch (µm)</th>
<th>Contact force (mN)</th>
<th>Maximum deflection (µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang et al. [4]</td>
<td>&lt; 1 (gold on gold)</td>
<td>-</td>
<td>0.1 (** )</td>
<td>150 ( ** )</td>
</tr>
<tr>
<td></td>
<td>&lt; 0.5 (tungsten on tungsten)</td>
<td>~ 100 ( )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Takahiro et al. [44]</td>
<td>0.37 (palladium on gold)</td>
<td>~ 100 ( )</td>
<td>1 to 120</td>
<td>0.5 to 8</td>
</tr>
<tr>
<td>Lee et al. [45]</td>
<td>&lt; 10 (measured against gold)</td>
<td>100</td>
<td>200</td>
<td>60</td>
</tr>
<tr>
<td>Tanaka et al. [46]</td>
<td>&lt; 4500 (measured against a tungsten probe)</td>
<td>&lt; 300 ( )</td>
<td>-</td>
<td>20 (**** )</td>
</tr>
<tr>
<td>Wang et al. [47, 48]</td>
<td>0.8 (against aluminium) 0.3 (against copper or gold)</td>
<td>240 by 160</td>
<td>10 to 50</td>
<td>20</td>
</tr>
<tr>
<td>Kim et al. [34-36]</td>
<td>&lt; 2 (against gold, stable over 20,000 touchdowns)</td>
<td>&lt; 100 ( )</td>
<td>-</td>
<td>170 (for a 800 µm long cantilever)</td>
</tr>
<tr>
<td>Lee et al. [37]</td>
<td>1.38 (against aluminium) 1.25 (against gold)</td>
<td>-</td>
<td>0.231</td>
<td>128 (for a 500 µm long cantilever)</td>
</tr>
<tr>
<td>Fork et al. [38]</td>
<td>-</td>
<td>6</td>
<td>-</td>
<td>60 (for a 160 µm long cantilever)</td>
</tr>
<tr>
<td>Itoh et al. [39, 40]</td>
<td>&lt; 2 (against aluminium using fritting)</td>
<td>&lt; 100 ( )</td>
<td>0.02 (****)</td>
<td>130 (for a 500 µm long cantilever)</td>
</tr>
<tr>
<td>Tsou et al. [41-43]</td>
<td>- ( )</td>
<td>50</td>
<td>1.6</td>
<td>26 (for a 85 µm long cantilever) (****)</td>
</tr>
</tbody>
</table>
A selection of non-planar probe design is shown in Figure 2.22.

Figure 2.22 Various out of plane designs – (a) Huang et. al. [41] (b) Tanaka et. al.[46] (c) Lee et. al [37] (d) Wang et.al. [48]

In 1999 Zhang et. al [4] described a thermally actuated cantilever probe card. The design used a bimorph structure with integrated heater to allow actuation, causing the cantilevers to curl up from their released position. The purpose of the thermal actuation was to allow the card to be reconfigured for testing different devices. The authors also proposed that the card would be useful for testing multi-chip modules or system in package (SiP) devices.

A disadvantage of the bi-morph design is that each probe requires two extra electrical connections for the heater. Since probe cards can often have thousands of probes, a tripling of the required electrical connections represents a substantial problem.

In 2000 Takahiro et. al. [44] described the fabrication of curved metal beams which projected up from the substrate. The beams were electrodeposited over either polyimide or silicon formers which were subsequently etched to leave suspended bridge structures. The polyimide former had a half circle cross section and was fabricated by using RIE to transfer the shape of a refloved photoresist pattern into fluorinated polyimide. The silicon former was fabricated by anisotropic etching of a (SOI) wafer. The buried oxide was used as an etch...
stop. A significant limitation of this bridge design is the very small deflections that are possible before the structure begins to yield. This will be discussed in section 2.5.2.

A similar fabrication method was reported by Lee et. al [45] in 2005. They produced angled metal cantilevers by using an overexposed and overdeveloped layer of thick positive photoresist as a sacrificial layer. After dissolving the sacrificial resist layer in acetone the silicon substrate beneath the cantilevers was recessed by isotropic dry etching to increase the separation between the cantilever and substrate.

In 2008 Tanaka et. al. [46] reported on probes fabricated from heavily doped silicon and electroplated nickel which were anodically bonded to a thick glass substrate. The fabrication used anisotropic silicon etching to structure a silicon wafer to define the vertical geometry of the probe and probe tip. Probes were then created by selectively implanting boron (to act as an etch stop) followed by through resist electroplating nickel and gold. Finally, the probes were bonded to a glass substrate containing the required conducting paths before the lightly doped silicon substrate was dissolved in ethylene diamine pyrocatechol (EDP). The probes are shown in Figure 2.22 (b).

In 2008 Wang et. al [47,48] reported on an array probe card. The probes were created by a combination of defined nickel electroplating and anisotropic silicon etching. The fabricated nickel structures were flip chip bonded to a ceramic substrate before the silicon substrate was dissolved in TMAH. Each probe was designed to have a ‘hoe’ shape as shown in Figure 2.22 (d). The tips were fabricated with a variable width so that when deflected the stress was uniformly distributed throughout their length. A ceramic substrate was chosen since it has a coefficient of expansion very similar to that of silicon and is therefore useful for ‘burn in’ testing.

Various authors report on cantilevers that curled up from the substrate. In each case the cantilevers were fabricated in-plane. However they deformed on release due to a stress gradient in the direction of their thickness.

One of the earliest published works was Kim et. al. in 1998 [34]. They used a highly doped silicon region as a sacrificial layer to release an epitaxial silicon, oxide and metal cantilever. The cantilevers were then annealed at 500 °C which caused them to curl up from their released positions due to the difference in thermal expansion coefficients of the various films. In 2005 the same research group republished the work with a slight adjustment to the fabrication process [35] and again including vias for back side contact [36].

In 2005 Lee et. al. [37] reported on molybdenum, chrome, gold and benzocyclobutene (BCB) composite beams released using a sacrificial aluminium layer. Electrical testing was
performed with gold coated molybdenum probes without BCB layers. Despite the absence of BCB, the probes still curled up from the substrate due to a stress gradient in the molybdenum attributed to a variation in grain size and the formation of native oxide.

A very similar technology to that described by Lee et. al. called Stressed-Metal ClawConnect™ was developed by Fork et. al. at the Palo Alto research centre (Xerox corporation) [38]. They used molybdenum-chromium alloy cantilevers covered with electroplated copper, nickel and then gold to increase their mechanical strength and improve their electrical conduction. The technology has been proposed for forming interconnects in wafer level packaging.

Another example of stressed metal cantilevers is the work of Itoh et. al. [39, 40]. In their work two different electroplating bath formulations (Watts and Sulphamate) were used to deposited nickel with different internal stresses. By using an gold electrode, encapsulated on the wafer surface, Itoh et. al. demonstrated electrostatic actuation of the cantilevers. They propose that such a facility would allow the same probe card I/O to be used for multiple probes (if they do not need to be driven at the same time) since the probes not in use could be switched away from the substrate under test.

In 2006 and 2008 Tsou et. al. [41, 42] reported on out of plane nickel probes in 2D arrays and perimeter layouts. These were fabricated by releasing patterned silicon dioxide cantilevers by KOH etching the silicon substrate and then evaporating titanium. The cantilevers were caused to curl up by the residual stress in the titanium. Nickel was then electroplated before the silicon dioxide and titanium etched. This fabrication dose not allow for electrical isolation of one cantilever from another although the authors propose laser cutting (which they demonstrated in an earlier publication [43]) as a method of isolating cantilevers.

2.4.4.1 Conclusions

Inclined devices remove the need for high aspect ratio machining (see 2.5.3) and therefore offer the possibility of more varied, simpler and cheaper fabrication methods. Methods that rely on electroplating and KOH bulk etching to provide an inclined surface (such as Tanaka et. al. and Wang et. al. [46, 48]), require the silicon wafer to be dissolved to release the fabricated probes. This limits the choice of substrate to which the probes are bonded. Also, there is no choice in the angle of inclination since this is fixed by the crystallographic properties of silicon.
A greater degree of design freedom is afforded by engineering a stress gradient to cause cantilevers to curl up from a substrate. If the stress is a result of the deposition method there is little restriction on the choice of substrate. If however the stress gradient is caused by thermal annealing then this requires a substrate compatible with annealing temperature.

Very fine pitches have been fabricated by this method which makes it suitable for wafer level packaging technology when used in combined with solder re-flow. However, its use as a probe card is limited by the stiffness of the cantilever structures which are often very low since they consist only of thin evaporated or sputtered metals. Therefore, it is common to electroplate the released structure to provide mechanical reinforcement [34, 38, 41, 42] when using the cantilevers as probes. Since the probes are released during the electroplating there is a significant danger of damage occurring as the probes pass through the liquid surface. As well as potential stiction issues there is also a loss of dimensional control, since the electroplating is isotropic. Finally, controlling stress gradients in deposition is often poorly understood and therefore repeatability, especially over large areas, could be difficult.

Using shaped polymers as formers as demonstrated by Takahiro et. al. [44] and Lee et. al. [45] seems to offer a greater degree of control than using engineered stress gradients since the x, y and z profiles can all be controlled. A complication of this method is that the lithography (to define the cantilever) must be performed over the 3D former. The chemistry required to remove sacrificial polymer formers is compatible with a wide range of substrates (e.g. unpackaged IC's, ceramic, PCB etc.) as are the processing temperatures.

2.4.5 Other probe card designs

In 2003 Kataoka et. al. [49] presented a probe card that used cantilevers designed to buckle. Their device consisted of an array of nickel-iron cantilevers that extended 1400 µm beyond the end of a substrate. The authors found that when they brought the cantilevers into contact with a surface at an angle of 80° or more from the horizontal, it caused the cantilevers to buckle rather than bend. Using a buckling mode allowed much higher contact forces to be applied than was possible by bending. Additionally, once the cantilevers had buckled the contact force remained almost constant as the deflection was increased. This allowed a greater vertical compliance to be accommodated without applying undesirably high contact loads.

Cantilevers were fabricated in pairs such that a fritting contact could be used. One probe was slightly longer than the other so that after fritting had occurred the probe card could be
slightly raised, meaning only one probe (the signal path) was in contact. The probes applied a contact force of 5 mN after buckling and, with a maximum fritting current of 300 mA, the contact resistance to either copper or aluminium was approximately 1 Ω.

An obvious limitation of this design is that it is only possible to probe line-on-centre pad layouts since it is not possible to create a 2D array of probes. Also, the length of the probes means that they will be extremely fragile and would be easily damaged during handling.

In 2004 Kim et. al. [50] reported on a blade type probe card designed for flat panel display testing. Flat panel displays require huge probe cards (e.g. 170 by 200 cm). To make such a large card the authors created a 3.1 cm long 'probe block' that consisted of an array of 406 metal blades at 40 μm tip to tip pitch. Several of these ‘probe blocks’ were then bolted to a frame to create the full probe card.

Individual metal blades were fabricated by etching 20 μm thick nickel plate and then coating them with gold. Each blade was designed to provide 850 mN at a 160 μm deflection. The blades were then manually assembled into a slotted holder and glued in place. The holders were fabricated by DRIE etching a silicon wafer. The wafer was first etched from the front to form a slot the length of the metal blade and then from the back to create a hole for the tip of the blade to protrude from. The patterned silicon was then oxidised so that the assembled blades would be electrically isolated. Electrical connection between the PCB and blades was made using pogo pins. The x-y alignment of the assembled probe tips was ± 6 μm and the contact resistance was less than 0.5 Ω.

In 2006 Ryu et. al. [51] reported on the fabrication of individual vertical probes that could be assembled into a vertical probe card (not demonstrated). The probes were fabricated by electroplating nickel-cobalt alloy into moulds made by DRIE etched silicon. The etched silicon was bonded to glass wafers coated with metal seed layers to facilitate electroplating. The probes had a meandering shape which allowed the stress from deformation to be distributed throughout the structure. Mechanical testing showed the probes could apply forces of 10 to 80 mN and tolerate deflections of 10 to 40 μm depending on the design.

The design of both Kim et. al. [50] and Ryu et. al. [51] require manual assembly of the probes into a card. These methods are therefore not suitable for very fine pitches. They are however the only practical method of assembling the very large probe cards needed for display testing.
2.5 Probe card design

2.5.1 Design objectives

A successful probe card design had to meet most of the criteria specified in section 2.2 and be different, from those designs discussed in sections 2.3 and 2.4. Since the chosen design had to be fabricated, there were restrictions imposed by the facilities available. For this reason designs that relied on DRIE, bonding (e.g. by soldering or anodic processes), silicon doping or CVD were not considered.

It was decided that the probe card design should have the following objectives:

- Fine probe pitch (less than 25 μm).
- Low leakage current (less than 1 nA).
- A stable contact resistance of less than 2 Ω.
- Low manufacturing time and cost.
- Minimise the contact force.
- Have a fabrication method that was as versatile as possible.
- Produce probes that could accommodate between 30 and 40 μm of deflection without damage.

The rational of each objective is discussed in the following sections along with any implications they impose.

2.5.1.1 Fine probe pitch

Area array layouts are used exclusively with solder balls since it is not possible to wire bond such geometries. Solder balls are relatively large and so the pitch of area array layouts is greater than that of bond pads (see section 2.2). Conversely, very fine pitches are required for line-on-centre and perimeter bond pad layouts. Since a MEMS approach is well suited to the generation of fine pitch probes it seemed most logical to concentrate on designs for fine pitch line-on-centre and perimeter layouts.

As was seen in section 2.4, through wafer connections or flip-chip bonding techniques are critical to the fabrication of an area probe card. Since neither DRIE nor bonding techniques were available in house this was an additional reason to restrict the design to perimeter and line-on-centre geometries.
The aim was to develop a probe card capable of 25 \( \mu m \) pitch. This is the minimum pitch required, according to ITRS estimates, up to 2020 (see Table 2-1).

2.5.1.2 Low leakage current

It was decided that the probe card design should provide high levels of inter-probe electrical isolation. The aim was to develop a probe card with less than 1 nA leakage between probes (measured at 100 V). This far exceeds the requirements of the ITRS (see Table 2-1) but would make the technology useful for testing plastic electronic devices.

2.5.1.3 Low contact resistance

It was decided the contact resistance to gold should be less than 2 \( \Omega \) to be comparable to other MEMS probe cards. The contact resistance to aluminium or other metals should ideally be a similar value. Note that 2 \( \Omega \) is somewhat higher than the 0.5 \( \Omega \) stated in the ITRS (see Table 2-1) however many commercial probe cards continue to use 2 \( \Omega \) as the contact resistance benchmark. In general, the contact resistances should be made as small as possible to prevent heating of the contact and to allow large currents to be delivered.

2.5.1.4 Low manufacturing time and cost

In order to keep the manufacturing time and cost down it was important that the whole probe card could be made by MEMS type batch fabrication techniques (as opposed to probe by probe assembly for example). By using such techniques it was hoped there should be no significant increase in manufacturing time or cost associated with scaling up the card to a higher number of probes.

This approach limited the probe card size to that of the substrates that could be processed. In the case of this work, all fabrication was conducted on 2 inch diameter substrates. However it was intended that the fabrication process should be such that it could be applied to the current industry standard of 300 mm diameter substrates.

2.5.1.5 Minimise the contact force

As already stated, the contact force during probing should be minimised to reduce the risk of physical damage to the device being tested and to reduce the total load on the wafer under test. However, a certain level of contact force is needed to provide a stable electrical connection, the exact value of which depends on the physical and electrical nature of the contacting surfaces. Since this was a critical experimental parameter, the fabrication had to be able to produce a wide range of contact forces.
Epoxy needle type designs apply a contact load of 100 mN per probe. Conversely, MEMS switches apply contact forces of less than 100 μN between contacts. Therefore the aim was to fabricate probes that could apply contact forces in the range 100 μN to 100 mN.

2.5.1.6 General fabrication method

It was intended that the fabrication method should have as few limitations, in terms of substrate and probe material, as is possible. If the fabrication method was compatible with a wide range of substrate materials then this would allow the use of ceramic substrates for ultra low leakage current or oxidised silicon substrates for burn in testing (since they would have matched thermal expansion properties to the wafer under test) or pre-fabricated strip lines on PCB substrates for RF testing.

Making the fabrication process substrate independent also permitted the possibility of using an IC as the probe card substrate. By integrating test circuitry onto the probe card it was anticipated it would be possible to achieve higher test speeds (due to the close proximity between circuit and probe) and to reduce the number of electrical connections required between the testing equipment and probe card.

An application of where such an approach might be used is in the testing of system in package (SiP) devices. In SiP products it is common to fabricate the active and passive components on separate dies. If probes could be fabricated on a known good passive die, then this could be used to test the active dies. Since the probe length would be small, this testing environment would accurately represent the final packaged device.

It was also intended that the fabrication should offer as much choice in probe material as possible. This was to allow refinement, for instance using an alloy rather than pure metal for its hardness or wear resistance.

These requirements were best meet by choosing an additive surface fabrication process as opposed to a bulk machining process. To fabricate structures capable of applying the required contact forces (100 μN to 100 mN) it was predicted that structural layers would need to be several microns thick. This negated the use of sputtered or evaporated layers and instead polymers (such as SU-8) or electroplated metals were considered. Ultimately, electroplated metals were chosen as a structural material since they provided the required electrical connectivity in addition to the ability to deposit thick layers.

In MEMS fabrication nickel, copper and gold electroplating are common. Of these metals, nickel was the best choice for a structural material since it has a higher yield stress than either gold or copper. Nickel alloys, such as nickel cobalt or nickel tungsten were also
considered. However, for simplicity, nickel electroplating was used to demonstrate the fabrication.

2.5.1.7 Overdrive

Generally, probes have to be able to compensate for the variation in probe tip planarity (which is a property of the probe card and is a result of the manufacturing process) and wafer bow (which depends how many contact there are, how thick the wafer under test is and over what area contact is made). Additionally, solder balls typically have an appreciable size variation (± 8 µm variation on 120 µm diameter balls [52]) which must be accommodated for by the probe card.

Traditional epoxy probe cards are used with 50 to 80 µm of overdrive. This overdrive is sufficient to accommodate the height variations mentioned above as well as being required to deform the probes, causing them to apply the contact load which is needed to establish a reliable area of contact. In the case of bond pads, the overdrive is also needed to get the probe tip to skid over the bond pad which helps break native oxide layers, ensuring a good electrical connection.

Instead, MEMS probe cards (section 2.4) were found to be designed for a large range of overdrives. Deflections ranged from only a few microns to 250 µm.

Since it was intended to develop a probe card designed for probing bond pads, not solder balls and to develop probes that applied a minimum of contact force, it was felt that 30 to 40 µm overdrive would provide an adequate compliance. Therefore the aim was to develop a probe structure that could be deflected by 40 µm, apply the required contact force and yet not be damaged.
2.5.2 Comparison between cantilevers and bridges

Two basic structures were considered for as a probe design; a cantilever and a bridge. Let us first consider the application of a bridge structure to the probing of a perimeter layout. As can be seen from Figure 2.23 the geometry of bridge is directly constrained by the pitch of the bond pads.

![Diagram showing bridge structures probing a perimeter layout](image)

From Figure 2.23 it can be seen that the diagonal pitch $D$ and pad pitch $P$ are related as shown by equation 2-1.

$$D = \sqrt{2P^2}$$  \hspace{1cm} (2-1)

Further, if we consider section X-X we can see that the bridge length, $L$ anchor size $a$ and inter-bridge gap $g$ are related to the diagonal pitch $D$ as follows:

$$D = g + \left(2 \times (a + \frac{L}{2})\right) = g + 2a + L$$  \hspace{1cm} (2-2)

Suppose the bond pad pitch $P = 25 \, \mu m$, then from equation 2-1 the diagonal pitch is calculated to be $D = 35.4 \, \mu m$. Let the gap $g = 5 \, \mu m$, and anchor $a = 5 \, \mu m$, then from equation 2-2 the bridge span, $L = 20.4 \, \mu m$.

Similarly, the maximum value of beam width, $w$, is also constrained by bond pad pitch. It can be seen from Figure 2.23 that $w$ should be less than half the diagonal pitch; $D/2 = 17.7 \, \mu m$. 

46
The simple theory of bending [53] can be used to calculate the maximum stress that occurs in a built-in beam for a given loading condition. For a point load, \( F \), in the centre of a beam with rectangular cross section, the maximum stress, \( \sigma_{\text{max}} \), can be calculated from:

\[
\sigma_{\text{max}} = \frac{3FL}{4wt^2}
\]  

(2-3)

where the beam has thickness \( t \). Additionally, the vertical deflection, \( v \), of the beam, due to point load \( F \), can be calculated from:

\[
v = \frac{FL^3}{16Ewt^3}
\]  

(2-4)

where the beam has a Young’s modulus \( E \). Let us assume that the probes are fabricated from nickel (electroplated nickel values: Young’s modulus, \( E = 160 \) GPa and yield stress, \( \sigma_y = 600 \) MPa [54]) and must generate a maximum contact load, \( F = 1 \) mN. For a beam of length \( L = 20.4 \) \( \mu \)m, \( w \) and \( t \) can be calculated from equation 2-3 by equating maximum beam stress to yield stress. Additionally, the maximum deflection can then be calculated from equation 2-4. Calculated values of \( w \), \( h \) and \( v \) are shown in Table 2-8.

<table>
<thead>
<tr>
<th>( w (\mu m) )</th>
<th>( h (\mu m) )</th>
<th>( v_{\text{max}} (\mu m) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3.57</td>
<td>0.0364</td>
</tr>
<tr>
<td>4</td>
<td>2.52</td>
<td>0.0515</td>
</tr>
<tr>
<td>6</td>
<td>2.06</td>
<td>0.0631</td>
</tr>
<tr>
<td>8</td>
<td>1.79</td>
<td>0.0728</td>
</tr>
<tr>
<td>10</td>
<td>1.60</td>
<td>0.0814</td>
</tr>
<tr>
<td>12</td>
<td>1.46</td>
<td>0.0892</td>
</tr>
<tr>
<td>14</td>
<td>1.35</td>
<td>0.0964</td>
</tr>
<tr>
<td>16</td>
<td>1.26</td>
<td>0.1030</td>
</tr>
</tbody>
</table>

*Table 2-8 Maximum deflection of a bridge type probe*

It can be seen in Table 2-8 that the maximum vertical deflection is \( \sim 100 \) nm which is much lower than the 40 \( \mu \)m required. This calculation indicated that a bridge type probe was not suitable for probing perimeter layouts.
Now let us consider a cantilever structure. Cantilevers of varying width, such as those designed by Wang et. al [47, 48], would evenly distribute stress throughout their length. However using such cantilevers limits the pitch at which they can be fabricated. For this reason only cantilevers of uniform width were considered as shown in Figure 2.24.

If we assume a bond pad pitch, \( P = 25 \, \mu \text{m} \), and gap between cantilevers, \( g = 5 \, \mu \text{m} \), then it can be seen the maximum cantilever width, \( w = 20 \, \mu \text{m} \).

As for the bridges, let us assume that the cantilevers are fabricated from nickel and must generate a maximum contact load, \( F = 1 \, \text{mN} \). The maximum stress in a cantilever, \( \sigma_{\text{max}} \), with point load \( F \) applied at the free end, can be calculated as [53]:

\[
\sigma_{\text{max}} = \frac{6FL}{wh^2} \quad (2-5)
\]

The vertical deflection, \( \nu \), of the cantilever, due to point load \( F \), can be calculated from:

\[
\nu = \frac{4FL^3}{Ewt^3} \quad (2-6)
\]

Suitable cantilever thickness and lengths can then be calculated, as can the corresponding maximum deflection.
Calculated values for $w = 20 \mu m$, $F = 1 mN$, $\sigma = 600 \text{ MPa}$ and $E = 160 \text{ GPa}$ are shown in Table 2-9.

<table>
<thead>
<tr>
<th>$L$ ($\mu m$)</th>
<th>$h$ ($\mu m$)</th>
<th>$v_{max}$ ($\mu m$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>7.07</td>
<td>3.54</td>
</tr>
<tr>
<td>200</td>
<td>10.0</td>
<td>10.0</td>
</tr>
<tr>
<td>300</td>
<td>12.2</td>
<td>18.4</td>
</tr>
<tr>
<td>400</td>
<td>14.1</td>
<td>28.3</td>
</tr>
<tr>
<td>500</td>
<td>15.8</td>
<td>39.5</td>
</tr>
<tr>
<td>600</td>
<td>17.3</td>
<td>52.0</td>
</tr>
<tr>
<td>700</td>
<td>18.7</td>
<td>65.5</td>
</tr>
<tr>
<td>800</td>
<td>20.0</td>
<td>80.0</td>
</tr>
<tr>
<td>900</td>
<td>21.2</td>
<td>95.5</td>
</tr>
<tr>
<td>1000</td>
<td>22.4</td>
<td>112</td>
</tr>
</tbody>
</table>

Table 2-9 Maximum deflection of a cantilever type probe

As can be seen from Table 2-9, cantilevers greater than 400 $\mu m$ length can accommodate the required deflection. When compared to the bridge type structure (Table 2-8) it is clear that a cantilever structure is better suited than a bridge structure. Therefore it was decided to use a cantilever structure to design the probe card.

2.5.3 In-plane and out-of-plane designs

Consider a planar cantilever design such as that shown in Figure 2.25. It can be seen that the maximum deflection is either determined by the height of the cantilever tip (Figure 2.25 (a)) or the height of the anchor (Figure 2.25 (b)), depending on which is greater.

Therefore to permit a deflection of 40 $\mu m$, such a design would require both the anchor and tip height to be 40 $\mu m$ or greater. Recall the design objective is to fabricate a probe card with a 25 $\mu m$ pitch. Therefore it can be assumed the cantilever must be between 10 and 15 $\mu m$
wide. This implies that the tip and anchor must be fabricated with an aspect ratio greater than 2.7 for a 15 μm wide cantilever or 4 for a 10 μm wide cantilever.

If the pitch between probes was further decreased or the required deflection was increased then this would imply that the aspect ratio of both anchor and tips has to increase further. This has implications for both the structural stability of the anchor and tip as well as being difficult to fabricate.

Now consider a non-planar cantilever such as shown in Figure 2.26. Here the whole length of the cantilever can contribute to deflection and in this instance the maximum deflection is determined by the yield strength of the cantilever. This eliminates the requirement for high aspect ratio fabrication methods whilst permitting large deflections. Therefore non-planar designs are better suited to reducing scales than planar designs.

![Figure 2.26 Inclined cantilevers used as probes](image)

### 2.5.4 Tip structures

Following the work of Beiley et. al. [22, 23] it was decided to investigate the effect of tip design on contact force-resistance behaviour. Beiley found by replacing a single large tip with an array of smaller tips, of equal area, it was possible achieve low resistance contacts at low forces.

It was known that the probe tips would have to be fabricated at the end of inclined cantilevers which would make it difficult to achieve fine structures by lithography. Instead, two different electroplating baths were used to deposit the nickel tips, one that produced smooth and the other pointed or cuspidated deposits. By measuring the contact force-resistance behaviour, it was hoped to obtain an indication as to whether the tip design played an important role in low force probing. In particular, the contact behaviour to metals with native oxides, such as aluminium, was of interest.
2.5.5 Design summary

Having conducted a review of probe card requirements and design approaches (sections 2.2, 2.3, 2.4) it was decided to develop a probe card suitable for probing fine pitch bond pads in line-on-centre and perimeter layouts. It was decided to use inclined cantilevers made from electroplated nickel as the probe structure. Calculation showed that a cantilever was best suited to fine pitch probing (see section 2.5.2) and it was shown that an inclined structure was more scalable than a planar design.

It was decided that the fabrication method had to be independent of the substrate material and should be capable of producing cantilevers at 25 µm pitch. To be useful as an electrical probe, the cantilever had to withstand 30 to 40 µm displacements without damage. Additionally, the cantilevers had to be highly electrically isolated from one another with a leakage current of less than 1 nA at 100 V and the target contact resistance to gold of less than 2 Ω.

The force requirements were unknown and had to be determined by experimentation. Traditional epoxy probe cards and MEMS switches suggested that a sensible range of forces to investigate was 100 mN to 100 µN. It was also decided to investigate the effect of tip surface texture on the force-contact resistance behaviour of the probes.
2.5.6 Other structures

Early during the development of the fabrication method (see chapter 3) it became clear that by adjustment of the mask pattern it was possible to create a wide variety of structures. Therefore it was decided to also fabricate devices that were suitable for solder ball probing. Illustrations of the various types of structures envisaged are shown below in Figure 2.27.

![Various probe designs](image)

**Figure 2.27 Various probe designs**

The ‘hoop’ and ‘spring’ type structures are simple variations on the basic inclined cantilever structure. One problem with using cantilevers to probe solder balls is the tendency of the cantilever to slip off the ball during overdrive. The ‘hoop’ structures are designed to ‘capture’ solder balls and prevent this happening. The ‘spring’ structures should deflect vertically unlike the cantilevers. Therefore they could also be used for probing solder balls.

References


18. *FormFactor Ships Industry's First One-Touchdown 300-mm Wafer Probe Solution for Burn-In Test Applications*. 2007, Nano Science and Technology Institute.


Chapter Three

Fabrication

This chapter is concerned with the fabrication of the micromachined probes. As described in chapter 2 the probes were designed to be angled up from the substrate. It was found that the most important aspect of the fabrication was the choice of sacrificial layer which is discussed in detail. Relevant information on fabrication techniques, namely electroplating, selective etching and on factors affecting adhesion are presented. Descriptions of the fabrication methods are then given. Finally, the capabilities and limitations of the fabrication processes in terms of geometric control are detailed.
3.1 Introduction

This chapter will detail two separate methods of fabricating inclined cantilevers. The methods are similar in concept but differ in their order of processing. One method allows greater design freedom (method 1), and the other was found to provide a better yield (method 2).

In both cases the structural material used to form the cantilevers was electroplated nickel. Nickel was an attractive choice since it has good mechanical properties, and by using electroplating it was possible to deposit sufficiently thick layers to form mechanically robust cantilevers. Another advantage of using metal as a structural material was that no additional layers were required to carry electrical signals.

3.2 Selection of Sacrificial material

It was argued in chapter 2 that a surface machining approach was preferred to bulk machining. This type of fabrication relies on depositing sacrificial layers which are subsequently removed to leave suspended parts.

To allow the process to be compatible with the widest range of substrates it was important to choose a sacrificial material with the correct properties. These included the temperatures required for processing, whether the material could be removed selectively to the structural material (nickel), how thick the material could be deposited, etc. Three possible candidates for forming 3D sacrificial layers were considered. These were:

- Positive Novolac type resist (AZ9260)
- Guided copper electroplating
- Negative epoxy resist (SU-8)

3.2.1 Positive photoresist

Positive resists are attractive as a sacrificial layer since they can be easily dissolved in acetone, which is selective against a wide range of materials, including nickel. They can be lithographically defined directly and can be used to form 3D features by using greyscale lithography [1]. A thick positive resist, AZ9260, can be processed up to 150 μm using a single exposure and thicker still using multiple exposures [2]. There is negligible swelling
during development (as occurs with some negative resists), and the final resist image contains very little internal stress and demonstrates excellent adhesion to many substrate materials.

In Novolac type resists the polymer that remains after patterning is not cross-linked, and so the layers are not thermally resilient. Both thermal and chemical stability can be improved by cross-linking the resin [3].

### 3.2.2 SU-8

SU-8 was considered as a sacrificial material due to its excellent thermal stability and ability to be patterned at a wide range of thicknesses (sub-micron to several millimetres). SU-8 has an extremely high contrast, allowing structures with very high aspect ratios to be patterned. SU-8 is a negative resist, and therefore it is difficult to fabricate structures with positive sidewalls. Two approaches are discussed in sections 3.2.2.1 and 3.2.2.2.

Exposed SU-8 is cross-linked so has a high chemical stability which makes it difficult to remove. This is a drawback from a sacrificial layer consideration.

Common wet removal techniques rely on swelling, crazing and peeling the resist [4]. However, these methods are not suitable for removing SU-8 from beneath structures. Other techniques include various dry etching approaches such as reactive ion etching (RIE) and downstream chemical etching (DCE). Dry etch techniques often leave a residual scum and etch rates can be affected severely by the presence of metal structures. SU-8 can be removed by oxidation in liquid salt baths or thermal decomposition in an inert atmosphere [4]. However, because both techniques require high temperatures, very large stresses are introduced by difference in coefficients of thermal expansion at interfaces.

SU-8 processing is also hampered by its poor adhesion, which partly results from the shrinkage that occurs during post exposure bake cross-linking [5].
3.2.2.1 Inclined exposure

The concept of an inclined exposure is shown in Figure 3.1. It can be seen that structures with a positive sidewall are obtained from an inclined exposure from the front through an inclined mask (Figure 3.1(a)). If, however, a transparent substrate with a patterned metal is coated with SU-8, and exposed from the back, then structures with negative sidewalls are formed (Figure 3.1(b)).

![Figure 3.1 Inclined exposure of SU-8. (a) exposure from the front, (b) exposure from the back](image)

Structures with rotational symmetry about an axis normal to the substrate can be formed by rotating the substrate during exposure with a constant angle to the exposure source [6].

Figure 3.1 does not show the effect of refraction at the SU-8 - glass interface. In the case of exposure from the front, if mask contact is poor, refraction will also occur at glass - air and air - SU-8 interfaces. It has been shown by Han et al. [6] that for a front side exposure, with air gap between the mask and SU-8, the maximum angle between a normal to the substrate and a SU-8 structure is approximately 39°. Ling et al. [7] showed that by performing the exposure in water it was possible to increase the maximum angle to 50°.

An angle of inclination of 40° from the substrate is still a very steep gradient meaning that the horizontal length of the cantilever would be approximately the same as the thickness of the sacrificial layer. This greatly limits the geometry of any probes fabricated using a sacrificial layer patterned by this method.
3.2.2.2 SU-8 greyscale lithography

To fabricate 3D structures with a positive sidewall using greyscale lithography it is necessary to expose the SU-8 layer from the back (substrate side). Therefore this process is limited to transparent substrates.

Because the exposure is from the substrate side it complicates the mask process. If a mask is simply brought into contact with the back of the substrate there will be a gap, equal to the thickness of the substrate, which affects the optical performance of the system. This can be avoided by fabricating the mask on the front side of the substrate before the SU-8 is spun and baked. However, because the mask required must be greyscale, this is not easy to achieve (Further discussion of greyscale masks is given in section 3.3.6).

Figure 3.2 shows a SU-8 layer exposed from the back through a dithered chrome mask which was patterned on the substrate. As can be seen from Figure 3.2 the binary mask was not successful in creating a gentle positive slope.

![Figure 3.2 SU-8 exposed from the back using a binary mask](image)

The contrast of a resist describes the range of exposure energies over which the resist layer changes from being totally unexposed to exposed throughout its full thickness. SU-8 has a very high contrast (hence can produce very high aspect ratio features) which results in the individual binary features of the mask being reproduced (see Figure 3.2), rather than creating a gradual gradient. This shows that a binary mask is not suitable for greyscale lithography of SU-8.

Any mask fabricated on the substrate would also have to be removed after the sacrificial SU-8 was removed. This further complicates the fabrication.
3.2.3 Guided electroplating

It was shown by Maciossek [8] that by patterning appropriate electrodes on an insulating substrate it is possible to use electroplating to grow a 3D structure with a controlled profile. This 3D structure can then be used as a sacrificial layer to form inclined suspended structures [9].

A diagram of the electroplating scheme is shown below in Figure 3.3. The left hand part of the diagram shows how the electrodes are patterned. Only the large electrode is connected to the cathode at the start of the electroplating, with the other electrodes being electrically isolated from it. The right hand side of the diagram shows how the electroplated metal isotropically grows from the cathode. As it grows it bridges the gaps between the electrodes, electrically connecting them to the cathode and hence guiding the deposition.

![Diagram of guided electroplating](image)

*Figure 3.3 Guided electroplating*

This method does not allow complete geometric control, unlike greyscale lithography, and the theoretical steepest inclination possible is 45°. It is extremely difficult to produce smooth slopes by this method (as opposed to a terracing type effect) and to do so requires extremely narrow electrodes to be patterned.

If guided electroplating is used to produce a sacrificial layer then the metal electroplated must be able to be etched selectively to the structural material (nickel). Copper is an obvious choice of sacrificial material as it can be electroplated easily and etched selectively against nickel (as discussed in section 3.5).

Copper is typically electroplated from a sulphate solution. Sulphuric acid is used in sulphate baths to help dissolve the copper anode (amongst other reasons). The inclusion of sulphuric acid therefore forbids the use of copper as a seed layer on the substrate since,
whilst it is not connected to the cathode (before the electroplate bridges the gaps). It will be dissolved. Therefore the seed layer electrodes must be made from a metal such as gold which is not dissolved by the electroplating solution and does not form any surface oxide. Unfortunately, gold is very difficult to remove selectively from the nickel structures.

Finally, the properties (such as internal stress) of a metal deposited through electroplating are often highly dependent on current density. Since the area of the cathode varies as the deposit grows, this is difficult to control.

3.3 AZ9260 processing

AZ9260 is a viscous positive photoresist manufactured by AZ electronic materials. It is based on the diazonaphthoquinone (DNQ)/Novolac system and uses propylene glycol monomethyl ether acetate (PGMEA) as a coating solvent. AZ9260 is similar to the popular AZ4562 thick film resist, but is formulated using a less absorbing Novolac component. This makes it possible to expose thick films (up to 150 μm) in one go (compared to a maximum of 90 μm thick films with AZ4562) and achieve features with higher aspect ratios [2, 10].

A difference in the chemistry of the photoactive component of the two resists leads to the need to include an important step in processing AZ9260 which is not necessary with AZ4562. To explain this we must first consider how positive resist operates.

Novolac is a phenolic resin which, due to the presence of the OH groups, is hydrophilic and as such is readily dissolved by aqueous alkaline solutions. The photoactive component of the resist is DNQ which is a hydrophobic molecule. When Novolac is impregnated with DNQ the mixture becomes insoluble in alkaline solutions. On exposure to UV radiation DNQ undergoes photolysis to form a carbene (a reactive compound containing divalent carbon atoms) and a nitrogen molecule. The carbene is unstable and undergoes ring contraction to form a ketene (an organic molecule containing the functional group C=\(\text{C}=\text{O}\)). The ketene then reacts with water (present in the resist) to form a carboxylic acid. The carboxylic acid is hydrophilic and allows the Novolac resin to be dissolved in the regions exposed to UV radiation (i.e. developed).

If no water is available in the resist film then the ketene can react with the Novolac resin to form an insoluble cresol ester. If this reaction occurs, it is not possible to develop the resist image.
Water absorption into a resist film is a diffusion limited process, and hence as the thickness of the resist film increases the time taken for water to penetrate the film thickness increases. Usually the reaction between the ketene and Novolac is slow and so there is sufficient time for water to diffuse into the resist during the development stage: this applies to the development of AZ4562. However, AZ9260 uses 2,1,4-DNQ as a photosensitiser which forms a highly reactive ketene. The ketene is sufficient reactive such that at the bottom of the resist layer the majority of it will have reacted with the Novolac resin before water can diffuse to the region [11]. Therefore water has to be present in the resist film before exposure to UV to allow the resist to be developed.

It was found at a 40 to 60% relative humidity (such as the clean room) it was necessary to leave the baked resist overnight to hydrate. This process was accelerated by placing the baked resist in a beaker of deionised water for 15 minutes prior to exposure. Lehar et al [11] suggest hydrating the resist in water at elevated temperatures. The proposed elevated water temperatures reduced the time taken to hydrate the film, improved the removal of coating solvent and so improved the thermal flow stability of the developed resist image.

It was found that hydrating at elevated temperatures resulted in no noticeable improvement in thermal flow resistance but did seem to adversely affect the resist adhesion (see Figure 3.4).

Figure 3.4 Developed AZ9260 (a) resist hydrated at 21 °C (b) resist hydrated at 40 °C
3.3.1 Spin curve

The resist film thickness as a function of spin speed was determined experimentally (Figure 3.5). The resist was spun coated on to 2 inch diameter silicon wafers, baked, exposed and developed in a standard way. A Talystep stylus profilometer (~1 nm vertical resolution) was then used to measure the photoresist film thickness.

As can be seen, to achieve a resist thickness greater than 25 μm the resist had to be applied in multiple layers. Figure 3.6 shows the thickness of the total resist film as a function of the number of layers. Each layer was coated at 1000 rpm (with an initial 300 rpm spin to spread the resist evenly) and the wafer was baked at 95 °C for 5 minutes in between each coat.

![Figure 3.5 Spin curve of AZ9260 thickness on silicon](image)

![Figure 3.6 Film thickness when spinning multiple layers at 1000rpm](image)
3.3.2 Pre-bake solvent removal

An experiment was conducted to determine the optimum baking conditions for a 55 µm thick layer of AZ9260. The pre-bake should remove as much solvent as possible without causing decomposition of DNQ, cross-linking the Novolac or introducing defects in the resist layer such as cracks.

It has been shown [12] that the DNQ in AZ9260 exhibits no evidence of decomposition when baked at temperatures up to 122 °C. Novolac resin starts to undergo oxidative cross-linking at temperatures above 150 °C [13] and is completely insoluble in acetone if baked above 250 °C [3]. Therefore it should be possible to bake AZ9260 at temperatures of up to 122 °C without degrading the resist chemistry.

In practice, however, it was found that baking the resist at temperatures above 110 °C resulted in brittle layers that cracked, even if the resist was cooled slowly (several hours) to room temperature. It was found that layers baked at 95 °C never cracked (if cooled slowly), whereas layers baked at 110 °C sometimes cracked on cooling, and could easily be caused to crack by small mechanical strains. The film cracking was attributed to stress in the film, which was most likely caused by non-uniform shrinkage throughout the film thickness when significant casting solvent was removed. Un-cracked and cracked patterned resist is shown in Figure 3.7 (a) and (b) respectively.

![Patterned 55 µm thick AZ9260. (a) resist prebaked at 110 °C. (b) resist prebaked at 120 °C](image)

Form these experiments it was decided that the 55 µm thick AZ9260 layer could be processed most easily at 95 °C and therefore this bake temperature was fixed and the time was varied.
To investigate solvent removal as a function of bake time, gravimetric assessment was used. A 2 inch diameter silicon wafer was weighed before a 55 μm thick layer of AZ9260 was spin coated and baked. Three layers of resist were applied to achieve the required thickness. Each layer was spun at 1000 rpm (with a 300 rpm spreading stage), and between the first and second, and second and third, layers the wafer was baked at 95 °C for 5 minutes. After the final layer was applied the wafer was weighed before being baked at 95 °C on a vacuum hotplate. Over the next hour the wafer was periodically removed and weighed. The percentage mass loss of the resist film, relative to its mass after the third layer was spun, is shown in Figure 3.8.

![Figure 3.8 Percentage mass loss with bake time](image)

As can be seen the rate of solvent removal has slowed to zero after 1 hr, and therefore no more solvent can be removed by baking the resist longer. As already discussed it is not desirable to increase the bake temperature, and therefore this demonstrates the maximum mass of solvent that can be removed.

### 3.3.2.1 Estimated solvent content after baking

To estimate the solvent remaining in the resist (after a 1 hour 95 °C bake) the wafer described in section 3.3.2 was subjected to a further high temperature bake in a vacuum oven. It was hoped that this would remove the remaining coating solvent, and hence by measuring the change in mass an estimate of solvent content could be made.

The thickness and area of the resist film after the 1 hour 95 °C bake were measured, and from this the volume of the baked resist was calculated as 7.96 x 10⁻⁸ m³. The mass of the film was measured as 102.7 mg, and so the baked resist film had a density of 1290 kg/m³.
The wafer was then baked under vacuum at 110 °C for 3 hours. It was clear that some residual solvent was removed during this bake since the escaping solvent formed bubbles (see Figure 3.9).

Figure 3.9 Resist film after baking in vacuum oven at 110 °C for 3 hours

It is known that casting solvent in the resist films can hydrogen bond to the Novolac resin [14]. If this was the case the solvent would be very hard to remove, and therefore it is not possible to be sure that all the solvent had been removed after the vacuum oven bake. Higher temperatures could ensure greater solvent removal, but were not used due to the risk of cross-linking the Novolac which would produce erroneous results.

After the vacuum oven bake the wafer was found to have lost 12 mg which corresponds to a 11.7% mass reduction. Therefore 11.7% can be used as a minimum estimate for the percentage solvent (by mass) in the baked resist film. This is of similar order to published results that used gas chromatography to determine solvent content of baked AZ9260 films. Lehar et al. [12] showed for a film thickness of 22 μm there remained 8 to 15% casting solvent (by weight), and for a film thickness of over 40 μm there remained 20% casting solvent (by weight).

3.3.3 Rehydration

As explained in section 3.3 it was necessary to place the baked resist layer in deionised water to ensure the film was hydrated prior to exposure. It was found that a 55 μm thick layer of AZ9260, baked at 95 °C for 1 hour, gained 3.6% by mass (relative to the baked film) after 15 min immersion in water. No further increase in mass was observed by immersing the wafer for longer. A mass gain of 3.6% is of similar magnitude to results obtained by Berger et al. [15], who found 0.8 to 1.4 μm thick Novolac films absorbed 2 to 3% mass of water.
3.3.4 Exposure

The exposure dose required to correctly pattern a 55 μm thick film of AZ9260 was determined experimentally. All exposures were performed using an EVG 620 mask aligner with a broadband near UV source (350 - 450 nm), which irradiated the wafer's surface with a power density of 13.2 mW/cm². Exposure could be performed in proximity, soft, hard or vacuum contact modes.

To determine the optimum exposure dose, resist films were exposed using a wide range of energies and the developed profiles were examined by SEM. A mask of varying line widths (from 70 μm to 2.5 μm) was used to pattern the resist. This allowed the minimum printable feature size to be established. The resist was developed in 400K developer, diluted 1:3 with deionised water. Development continued until all large areas were clear.

An electroplated copper film on a silicon wafer was used as a substrate for the resist experiments. This was chosen to match the circumstances of the resist layer used to define the cantilevers during fabrication of the devices. It was important to optimise the processing conditions of this resist layer to permit high density cantilever fabrication. It was felt that the electroplated copper's roughness would influence resist adhesion, as well as the required exposure dose, due to its low reflectivity (compared to the smooth surface of evaporated metals).

It was found that at low exposure doses small gaps could not be cleared (resist remained at the substrate-resist interface). At high exposure doses the (masked) resist at the mask-resist interface started to be exposed, causing rounding of the top of resist features (Figure 3.10).

A similar effect to overexposure was observed if the contact between the resist layer and the mask was poor. This can be seen in the case of the 1300 mJ/cm² exposure (Figure 3.10), as even large features have rounded tops.

An ‘optimised’ exposure was defined as the highest possible dose that did not degrade the resist near the mask-resist interface. It was determined that a dose of 1530 mJ/cm² produced the best exposure. As mentioned, mask contact was found to be a critical aspect of the exposure, hence ‘hard contact’ (which applies a pressure of 0.3 MPa between mask and wafer) was used during the exposure. A 55 μm thick layer of AZ9260 exposed at 1530 mJ/cm² can be seen in Figure 3.11, section 3.3.5.
3.3.5 Wall profile

In thick film lithography the width of a feature exposed using broadband UV is not constant throughout the thickness of the resist. This has been investigated by Yao Cheng et al [16] who show resist wall profiles follow a line of constant dose. They showed that horizontal diffraction of the light in combination with absorption of the resist leads to wall profiles similar to those of Figure 3.11. They also showed theoretically that 'smaller and uniform proximity gap reduces diffraction error' which is why mask contact is so critical to resist definition.

The width $w$ of the developed resist trench was measured relative to height $h$ above the substrate (Figure 3.11) for various mask aperture widths $A$. The results are shown in Figure 3.12. The widths of the apertures are marked directly on the horizontal axis as vertical lines to allow direct comparison to the resist feature width.

Figure 3.10 Effect of exposure dose
Figure 3.11 AZ9260 exposed at 1530 mJ/cm² showing varying feature width with resist thickness

Figure 3.12 Width of feature in resist as a function of resist depth for different mask slit aperture widths (A)
In the fabrication of the cantilevers AZ9260 resist was used to define an electroplating mould over a 3D layer. Since the variation in developed resist width (as a function of height) was known, it was possible to modulate the width of the cantilever feature on the mask to counter variation. This allowed the width of the developed resist feature, at the resist-3D layer interface, to remain constant with height and allowed a more dense cantilever fabrication.

It was assumed that mask contact would be poor and the resist moulds would have rounded tops (as seen in Figure 3.10) rather than the hour glass shown in Figure 3.11. Therefore it was acceptable to approximate the resist sidewalls to a straight line as shown in Figure 3.13. This allowed a simple design rule that any feature width should taper by 6 µm from start to end to produce a resist image with the same width when printed over a linear varying 3D layer 55 µm high.

![Figure 3.13 Error in resist feature width as a function of resist height](image)

The results of using this design rule as compared to a simple constant width feature are shown in Figure 3.14. It can be see that despite the approximation of the resist sidewalls as linear, and the non-linear variation in 3D layer (discussed in section 3.3.6.4), the improvement in electroplated track width control is easily visible.
Electroplated tracks vary in width

Electroplated tracks with near constant width

Figure 3.14 Fabrication results from tapered and non-tapered mask designs

The variation in linewidth was estimated using equation 3-1. Without tapering the mask, the variation in linewidth was 32%; with feature tapering, the variation reduced to 12%.

\[
\frac{2 \times (\text{max} - \text{min})}{(\text{max} + \text{min})} \times 100 \quad (3-1)
\]
3.3.6 Grey scale lithography

Greyscale masks can be categorised into two groups: analogue and binary masks.

3.3.6.1 Analogue grey scale masks: High energy beam sensitive glass and Laser direct write masks

High energy beam sensitive glass (HEBS glass) is a commercial material available from Canyon Materials Inc. in San Diego [17]. The top 3 \( \mu \)m of the glass has silver ions diffused into it in the form of silver-alkali-halide nanocrystals which are generally transparent. These can be reduced to opaque metallic silver by exposure to an electron beam. The density of the reduced silver metal depends on the exposure dose.

The beam power required is high (>10 keV) and so long write times are required. However the technique is capable of very high spatial resolution.

Laser direct write (LDW) greyscale masks are usually based on bimetallic thermal resist systems such as Sn-In or Bi-In [18, 19]. A laser mask writer is used to heat the (opaque) metals which caused the bilayer to be converted into a transparent mixed oxide. Therefore by controlling the laser exposure dose it is possible to create greyscale masks.

An emulsion photographic plate can also be used as an economical greyscale mask [20]. An image is first printed onto a transparency using an ink-jet printer. The printer driver dithers the output so that grey levels can be created by different densities of black dots. This transparency can then be photoreduced on to an emulsion plate to create a mask. If the density of the black dots is below the resolving power of the lens system in the camera, then the dots are not reproduced and a greyscale mask is created. A major drawback with this process is control of the development of the emulsion plate in order to produce repeatable results and rendering a high enough resolution printout on the transparency.

HEBS, LDW and greyscale emulsion masks create regions of controlled UV absorption and so can be used to produce ‘true’ greyscale exposure in both projection and contact lithography systems.
3.3.6.2 Binary grey scale masks

Binary mask, usually made from patterned chrome, consist of opaque and transparent areas which are patterned in such a way as to produce the appearance of a ‘grey’ region. There are many coding algorithms that could be used to create ‘greys’ which were developed for use in other applications such as electronic displays, printing and image processing.

In photolithography common approaches to coding of grey levels include pulse width modulation and pulse density modulation [21]. In pulse width modulation the opaque dots size is varied to modulate the grey level. In pulse density modulation the dots have the same size, but the density changes to modulate the grey level.

A binary mask and a stepper projection system can produce ‘true’ greyscale exposure since the lens acts as a spatial filter [22, 23], preventing the binary pattern being reproduced on the wafer. In a contact exposure system local undesired oscillations in the image intensity occur: however, the average exposure dose is still spatially controlled.

3.3.6.3 Implementation of greyscale masks

The original mask designs were created in Corel Draw II using 24-bit RGB space which was capable of defining 256 different grey levels. Once completed the design was exported as an uncompressed 8-bit bitmap which was converted to a binary (2-bit) image in Jasc Paint Shop Pro 8 using an ‘ordered dither’ reduction method.

By experiment it was determined that ‘ordered dither’ command applied an Bayer algorithm [24] based on an 8 x 8 grid. This produced 64 different binary patterns representing the 256 levels of grey i.e. one pattern represented 4 different grey levels in the original file. The Bayer algorithm is a form of pulse density modulation with pixels fixed on a constant spacing orthogonal grid.

An n x n grid should be able to represent n² + 1 different patterns, which in the case of a 8 x 8 grid is 65. The software only produced 64 patterns, as for reasons of symmetry it dithered black as a pattern of mainly black with some white pixels (Figure 3.15). This was manually corrected after the dither so that black regions in the original image remained solid black in the dithered image and hence increased the number of different patterns to 65.
To make the dithered mask useful it needed to be converted into a format that could be used to control an E-beam or laser exposure tool to write the chrome mask. To do this, the image was converted to a gerber file format (commonly used in PCB manufacture) using a Matlab script provided by Dr Richard McWilliam. The conversion was performed such that each pixel represented a $1 \times 1 \mu m$ block, i.e. each pattern grid was $8 \times 8 \mu m$.

Finally the gerber file was converted into a Graphic Data System 2 (GDS-II) format using CleWin v2.87 before being externally written using an ALTA 36 laser exposure system by Compugraphics photomask solutions, Glenrothes, Fife.

### 3.3.6.4 Grey level to resist height mapping

#### 3.3.6.4.1 Preliminary experiment

At the time of the designing the mask set it was necessary to conduct an experiment to determine how the developed height of resist corresponded to grey level in the design. This would then allow the profile of the developed resist to be controlled.

A mask of total area $1 \text{ mm}^2$ with various dithered patterns was written on a Heidelberg DWL66 maskless lithography system by Dr Richard McWilliam. This was used to expose a
55 μm thick AZ9260 resist layer. A Zygo Newview 100 interferometer was used to measure the height after development and the results of this are shown in Figure 3.16.

Figure 3.16 SEM of exposed and developed resist and measured resist height as a function of grey value

From these results it was believed that a linear variation in resist height could be achieved by spatially controlling the grey level in the manner shown by the black line of Figure 3.17.

Figure 3.17 Expected resist profile for designed grey level profile

This spatial variation (in grey level) was used to design linear resist features in the final mask set. However, it became apparent when this mask set was used to fabricate devices that the mask design did not produce the desired linear profiles.

This discrepancy is likely to have resulted from poor processing of the resist in the preliminary experiment (ostensibly the resist development) and from recording too few data points. It is also believed that the feature size on the mask written on the Heidelberg system was unlikely to be exactly 1 μm due to the lack of control in the chrome etching step.
3.3.6.4.2 Grey level effect of resist height from the final mask set

A typical selection of developed resist profiles produced using the final mask set can be seen in Figure 3.18.

![Figure 3.18 Typical developed resist profiles from dithered mask](image1)

A section was taken through a selected resist profile and the resist height was measured using SEM (see Figure 3.19).

![Figure 3.19 Resist profile](image2)

---

Figure 3.18 Typical developed resist profiles from dithered mask

Figure 3.19 Resist profile
From this it was possible to determine the relationship between grey level and the resist height which is shown in Figure 3.20. It is clear from Figure 3.20 that there are two regions. One type of behaviour prevails for grey values between 0 and 127, and another for grey values above 127. In RGB colour space 127 represents the mid point between black and white.

![Figure 3.20 Resist height as a function of grey level](image)

The significance of a grey value of 127 can be seen if we consider how grey values correspond to binary dithered chrome patterns. A grey value of 127 is represented by a pattern that looks like a chessboard made from 1 μm² pixels.

Grey values more than 127 are represented by patterns where more than 50% of the area is opaque. These are dark field mask with isolated 1 μm² apertures.

Grey values below 127 represent patterns where less than 50% of the area is opaque. These are light field mask with isolated opaque regions of 1 μm². Therefore in these patterns the transparent regions consist of a mixture of apertures, most of which are greater than 1 μm².

Since the exposure source is 350 to 450 nm, 1 μm² apertures will cause significant diffraction of the light which passes through them. It is believed this diffraction by patterns corresponding to grey values greater than 127 causes a dose attenuating effect that is greater than that anticipated from the ratio of opaque to transparent pixels.

Patterns that correspond to grey values less than 127 have a number of apertures greater than 1 μm². Therefore it is anticipated diffraction effects will be less pronounced for these patterns. The resist height is assumed to be proportional to average exposure dose. Therefore the discontinuity shown in Figure 3.20 demonstrates that the exposure dose increases suddenly at grey values lower than 127. This is consistent with the above explanation.
The trends of both regions shown in Figure 3.20 can be accurately fitted by exponential relationships. This is shown in Figure 3.21. The trend shown in black (which passes through the origin) is given by equation 3-2 and the trend shown by the red line is given by equation 3-3.

\[
y = 0.0525 \cdot e^{y/56.83} - 0.0525 \quad \text{(black line)} \tag{3-2}
\]
\[
y = 0.0527 \cdot e^{y/94.53} - 0.224 \quad \text{(red line)} \tag{3-3}
\]

Since the resist can be assumed to have a constant absorption coefficient the intensity of the UV in the resist will take a first order exponential relationship. If it is assumed a critical dose is needed to make the resist soluble this would imply that the resist height should follow an exponential relationship with dose. Literature suggest that first and second order exponentials as well as Gaussian tail relationships between transmittance and resist height are possible [22, 23, 25, 26].

### 3.3.7 Reflow characteristics

To test the thermal stability of the 3D resist profiles they were subjected to post development bakes on a 95 °C hotplate for different durations. The reflowed profiles can be seen in Figure 3.22.

As can be seen substantial reflow has occurred after 20 s of baking. Also it can be seen that a loss of adhesion occurs at the thin end of the profile after only 5 s baking. The origin of this phenomenon is unknown but it may suggest that there is a high level of stress in this part of the resist that causes the resist to delaminate on heating. Notice also that the glass...
transition temperature of the resist is very low. This is an effect of the residual casting solvent which acts as a plasticiser [12].

From this experiment it was concluded that post development baking had a detrimental effect on the 3D resist layer and should not be used as a method of thermally stabilising the resist.

**Figure 3.22 Developed resist profiles after baking at 95 °C**

### 3.3.8 Increasing resist thermally stability

Various attempts were made to increase the stability of the developed resist profile. This was important since subsequent patterning steps during the fabrication process involved baking of resist, and hence subjected the 3D resist layer to elevated temperatures.

It was known from experiments on solvent removal (see 3.3.2) that it was not possible to increase the prebake temperature because this introduced too much stress into the resist layer, which caused the resist to shatter on cooling. Also it has been shown that prebaking in a vacuum could not be used to remove more solvent (see Figure 3.9). Therefore, efforts were concentrated at stabilising the developed 3D layer.

Three approaches were investigated to increasing the resist’s thermal stability. The first was to cross-link the surface to prevent it from deforming in a subsequent baking step which would be used to fully cross-link the resist. The second was to cover the resist with a conformal mould and similarly use this to prevent the resist deforming during a thermal...
cross-linking step. The third was to diffuse a cross-linking agent into the resist and then cross-link it at a low temperature.

Surface cross-linking was attempted using deep UV (254 nm) curing and an argon plasma. The deep UV approach was based on work by Hartung et al. [27] who used a combination of UV and thermal curing to stabilise resist images for transfer into silica by ICP etching.

It was found that 19 hrs of exposure to a 254 nm source (EPROM eraser) caused no noticeable increase in the thermal stability of a 55 μm thick layer of AZ9260. Hartung et al. worked with only 7 μm thick resist, which would have a much lower solvent content than the 55 μm thick resist. This and the very low power of the UV source might explain the failure to affect the thermal stability. Hiraoka et al. [13] working with a medium pressure mercury lamp, showed that in the best case the maximum cross-linking depth was only 3 μm. If the resist was 7 μm thick this would represent a substantial percentage; however, with 55 μm thick resist it is less influential.

Argon hardening was performed in an Oxford Plasmalab 80plus in RIE mode. The processing conditions were a chamber pressure of 400 mT, RF power of 150 W and argon flow rate of 40 sccm. The resist was then baked at 80 °C in a vacuum oven for 40 minutes to test its stability. During this baking stage, the solvent in the resist layer boiled causing bubbles (see Figure 3.23). From these bubbles it was clear that a ‘skin’ had been formed from the plasma treatment; however, it was not sufficiently robust to prevent reflow.

![Figure 3.23 Argon hardened resist after baking in a vacuum oven at 80 °C](image)

The second approach was to use a conformal mould to prevent reflow during thermal cross linking. This followed Lai et al. [28] who, by casting a PMMA layer over patterned AZ1350 photoresist and thermally cross-linking, produced resist images that were stable up to 500 °C under nitrogen. In Lai’s work the PMMA layer was cast to be twice the thickness of the resist, which resulted in a conformal covering with a plain surface. This was then baked at
200 °C to harden the resist during which the PMMA mould reinforced the resist image preventing reflow. Since the AZ1350 was cross-linked the PMMA could then be stripped in acetone.

With a 55 μm thick layer of resist it was not easy to cover with PMMA to twice the thickness. Instead a ‘skin’ of PMMA was used to coat the AZ9260. This was achieved by spinning PMMA (9% solids dissolved in chlorobenzene) over the AZ9260 layer at 400 rpm for 5 s, and then leaving the wafer for 10 minutes to allow the chlorobenzene to evaporate before baking at 40 °C for 1 minute. This procedure was repeated before the wafer was baked to cross-link the resist.

The cross linking bake consisted of a 10 minute 65 °C bake followed by 10 minute 90 °C bake, 20 minute 125 °C bake, 60 minute 160 °C bake and finally 15 minute 250 °C bake. The PMMA was then removed by sonicing in acetone; the resulting resist features can be seen in Figure 3.24. The PMMA mould largely prevented reflow; however, the cross-linking has caused the resist to shrink (vertically by 51%). This means that the resist-substrate interface is subjected to a tremendous amount of stress which makes the process unreliable.

Another attempt was to use Dow Coming Sylgard 184 PDMS as a conformal mould. Since PDMS could be peeled away (rather than having to be removed by acetone) it was possible to use a lower temperature which would improve resist solvent removal whilst only partially cross-linking the resist. It was hoped that this would not result in acceptable resist shrinkage.

The AZ9260 3D layer was first covered with a thin layer of sputtered gold to ensure the PDMS could easily be removed. Then PDMS (with cross-linking agent) was cast over the AZ9260 to a layer several millimetres thick. The wafer was then transferred to a hotplate at 180 °C and baked for 10 minutes. During the bake, solvent bubbles were observed to form in the larger resist areas. The PDMS was then peeled away and the resist imaged in the SEM.
This temperature produced a vertical shrinkage of 36% (the vertical height after the bake was measured as 35 μm) which, although less than occurred using a higher bake temperature, was still too great to form a reliable process. Also, the SEM showed solvent evaporation from the resist was problematic causing certain areas to rupture. A substantial degree of reflow also still occurred (see Figure 3.25).

![Figure 3.25 AZ9260 after 180 °C bake covered with a PDMS conformal mould](image)

The third approach involved trying to diffuse a cross-linking agent into the patterned AZ9260 to allow low temperature cross-linking to occur without reflow. Hexamethylenetetramine (HMTA) is a well known cross-linking agent used in the metal casting industry to harden Novolac. HMTA and p-tert-butyl phenol dissolved in ethanol has been used to crosslink Novolac resin at 80 °C [29].

Patterned AZ9260 was flood exposed (1600 mJ/cm²) to make the resist hydrophilic and then the resist was immersed in a 7% by weight solution in water of HMTA for 24 hrs. The sample was then baked at 50 °C for 30 minutes to promote cross-linking.

There was no noticeable improvement in thermal stability when the sample was baked at 95 °C.

Although many other methods of hardening resist are reported [30] it was clear from the above experiments that the high solvent content and large thickness of the AZ9260 layer meant that it would not be possible to harden. Instead, layers of other materials were used over the resist to provide a stable enough base to process on. Details of this are given in section 3.7.
3.4 Electroplating

Many metals, alloys, semiconductors and conducting polymers can be deposited by electroplating [31]. The properties of an electroplated deposit such as internal stress, yield strength, Young's modulus, resistivity, chemical resilience etc. depend on many different variables including: current density (or over-potential), bath temperature and agitation, bath chemistry (including impurities), bath pH, the type of power supply (pulsed or DC), the substrate material, anode-cathode spacing etc. Therefore it was important to control the electroplating parameters to obtain deposits with consistent properties. In practice this meant controlling current density, bath temperature and agitation, and regularly replacing the electroplating solution to limit the effects of impurities, pH variations and depletion of chemicals.

In order to demonstrate the fabrication process it was considered prudent to choose widely used and documented electroplating baths. Nickel was chosen as a structural material (as discussed in chapter 2) and is a very common choice in Microsystems for electroforming [32]. Copper, which is commonly used in the PCB and IC industries, was also used. In total three different electroplating formulations, one bath to electroplate copper (see section 3.4.5) and two baths to electroplate nickel (see sections 3.4.3 and 3.4.4) were used during the fabrication of the cantilevers. One of the nickel baths deposited nickel with a low surface roughness (see 3.4.3) whilst the other produced nickel deposits with a highly textured surface (see 3.4.4).

3.4.1 Faraday's law

In electroplating, an estimated for the rate of metal deposition can be calculated, for a given current density, from Faraday's law, which states that the mass of material deposited is proportional to the 'charge consumed' and the electrochemical equivalence of the substance. The electrochemical equivalence describes the amount of substance liberated by one Coulomb of charge. Faraday's law can be expressed as shown in equation 3-4.

\[ M = \frac{AQ}{Fn} \]  

(3-4)

where \( M \) is the mass deposited (in grams), \( A \) is the atomic mass of the metal (in grams per mol). \( Q \) is the charge 'consumed' (in Coulombs), \( n \) is the number of electrons added to each
metal ion as it is reduced to a metal atom and \( F \) is Faraday's constant (in Coulombs per mol). Faraday's constant is equivalent to Avogadro's constant multiplied by the charge of an electron and is equal to 96484.5 C/mol.

Nickel sulphamate and nickel chloride were chosen as the source of nickel ions in the two electroplating baths (see sections 3.4.3 and 3.4.4). In both of these salts, nickel is in a divalent state and therefore we can substitute \( n = 2 \) into equation 3-4. The atomic mass of nickel is \( A = 58.71 \) g/mol. From equation 3-4 it is now possible to determine the mass of nickel deposited in terms of 'charge consumed' as shown in equation 3-5.

\[
M_{\text{Nickel}} = 3.04 \times 10^{-4} \cdot Q
\]  

(3-5)

Not all the current that flows through the cell is used to reduce nickel ions to nickel atoms (for instance some charge is consumed reducing hydrogen ions). Therefore the useful charge, \( Q \) in equation 3-5, can be written in terms of the current that flows, \( I \) (in amps), the time it flows for, \( t \) (in seconds) and an electroplating efficiency \( \eta \) as shown in 3-6.

\[
M_{\text{Nickel}} = 3.04 \times 10^{-4} \cdot I t \eta
\]  

(3-6)

The mass of metal plated can be written in terms of the density, thickness and area of the deposit as is shown in equation 3-7.

\[
M = \rho d A
\]  

(3-7)

where \( \rho \) is the density of nickel (in g/cm\(^3\)), \( d \) is the thickness of the deposit (in cm) and \( A \) is the electroplated area (in cm\(^2\)). By substituting equation 3-7 into equation 3-6 and dividing both sides by area, \( A \) and time, \( t \) it is possible to calculate the electroplating rate, \( r \) (cm/s) from an applied current density, \( i \) (A/cm\(^2\))

\[
r_{\text{nickel}} = \frac{3.04 \times 10^{-4}}{\rho} \cdot i \eta
\]  

(3-8)

Substituting the density of nickel as 8.902 g/cm\(^3\) into equation 3-8 and writing the rate, \( r \), in \( \mu \text{m/min} \) and the current density, \( i \), in mA/cm\(^2\) gives:

\[
r_{\text{nickel}} = \frac{3.04 \times 10^{-4}}{8.902} \cdot i \eta \cdot \left(\frac{10,000 \times 60}{1000}\right)
\]  

\[
\Rightarrow r_{\text{nickel}} = 0.02049 \cdot i \eta
\]  

(3-9)

For copper electroplated from a copper sulphate solution (\( n = 2 \), \( A = 63.55 \) g/mol and \( \rho = 8.93 \) g/cm\(^3\)), using the above procedure gives a deposition rate \( r \) (\( \mu \text{m/min} \)) for a current density \( i \) (mA/cm\(^2\)) as shown in equation 3-10.

\[
r_{\text{copper}} = 0.022 \cdot i \eta
\]  

(3-10)
3.4.2 Deposit thickness uniformity

Thickness uniformity of an electroplated deposit can be improved by making the fluid boundary layer on the cathode (substrate) as small as possible. This reduces the possibility of localised metal ion depletion or changes in pH. Therefore, electroplating solutions must be agitated. This is achieved by the use of high pressure jets [33], paddles or stirrers [34], blowing gas through the solution or ultrasonic agitation [31]. In this work a simple magnetic stirrer was used to agitate the solution.

If the fluid boundary layers across the cathode are uniform then the localised rate of deposition is determined either by mass-transport limitations or (more commonly) by localised field strength. Which of these factors dominates the localised deposition rate depends on the applied current density, metal ion concentration and solution conductivity [35].

If the current density is very high then it is likely the electroplating is operating under mass-transport limitations. The typical width of the ion diffusion boundary layer in electroplating is 1 to 100 µm [36]. The feature sizes used in Microsystems are of similar order to this diffusion length (for instance when electroplating into small diameter holes in thick photoresist layers) and therefore it is likely that diffusion limitations will lead to a non-uniform deposition rate for features of different sizes. In this situation, rate uniformity can be improved by using a pulsed power supply (typically 1 ms or shorter cycle time) which allows metal ion concentration at the cathode surface to replenish in the off cycle [31]. Additionally, pulse plating can result in deposits with improved mechanical and chemical properties compared to those obtained by constant current conditions [37]. Pulse plating is recommended when depositing alloys since alloy composition is strongly dependent on the local ion concentrations [38].

In practice, electroplating is not often operated under mass-transport limitations since current density needs to be controlled in order to control internal stress and other properties of the deposit. When operating at lower current densities the localised deposition rate is dependent on the local field strength which determines local current density (recall that the rate of deposition is proportional to current density, see section 3.4.1). The resistance of the substrate (which depends on the nature of the seed layer), geometry of the plating mould and
the geometry of the plating set up and solution conductivity are all known to affect the local field strength [35].

Various refinements in the electroplating set up have been shown to improve the uniformity of current density across the cathode. These include the use of insulating shields placed between the anode and cathode [35, 39], the use of auxiliary electrodes [31] and placing the electrodes such that their edges are in contact with a dielectric (usually the wall of the electroplating tank) [34]. In order to be optimised each of these methods needs careful numerical modelling which was outside the scope of this work. However, some procedures were used to improve deposit uniformity.

In practice it was found on a unpatterned wafer (i.e. a wafer with a seed layer but no resist mask) that the thickness of the deposit varied in an approximately linear fashion such that it was thinnest where electrical contact to the wafer was made, and thickest at the point on the wafer furthest from the electrical connection. Therefore to improve the thickness uniformity of the deposit the wafer was rotated 180° half way through any electroplating step.

When electroplating was performed through a resist mask it was found that the geometry of the features affected the uniformity of the deposit. The most dramatic effect was observed when concentric structures were electroplated.

If a continuous ring was placed around a central electrode (central electrode to ring spacing in the order of tens of microns) it was found nearly impossible to deposit onto the central electrode. Moreover, when a ring was positioned a large distance away around an array of electrodes (several millimetres gap), it was still found to have a marked effect on their localised current density. This effect can be seen in Figure 3.26 which shows nickel deposited through a resist mask. The black areas of the diagram indicate holes in a resist mask through which nickel was deposited. The red shapes indicate where electrical connection to the wafer was made.

The texture of the electroplated nickel from the cuspidated bath depends on current density [40]. It is clear from Figure 3.26 there is a difference in texture (hence localised current density) if there was a complete or incomplete ring around the array of electrodes. The deposit texture was as expected when the ring was broken. To permit control of localised current density, no concentric structures were allowed. In practice this entailed covering half of the alignment marks (which formed a continuous strip around the cantilevers) with PMMA before electroplating.
Figure 3.26 Effect of feature geometry on the texture of nickel deposited from the cuspidated bath using the same average current density.

### 3.4.3 Smooth nickel

**Formulation**

- Nickel sulphamate tetrahydrate, Ni(SO₃NH₂)₂·4H₂O 400 g/l
- Nickel chloride hexahydrate, NiCl₂·6H₂O 1.5 g/l
- Boric acid, H₃BO₃ 40 g/l
- Sodium dodecyl sulphate (SDS), C₁₂H₂₅NaO₄S 1.5 g/l
- Saccharin Sodium salt dihydrate, C₇H₁₄O₃NSNa·2H₂O 4 g/l

(All chemicals have purities greater than 98%)

**Operating conditions**

- Anode: Nickel
- Temperature: 55 to 60 °C
- Current density: 5 to 15 mA/cm²

The formulation of this bath is an amalgam from various sources [31, 32, 40]; however the composition in each source is very similar. Sulphamate baths are very common in Microsystems fabrication since they produce deposits with a low internal stress at high rates of deposition [31]. It was found that a cathode current density of 10 mA/cm² produced a deposition rate of approximately 0.18 μm/min. By comparison to equation 3-9 it can be seen
that this rate corresponds to a cathode efficiency of 87.8% which is close to the expected efficiency of 90% for bright nickel plating solutions [31].

The bath produced deposits with a low surface roughness of $R_a = 43.4$ nm (measured by Digital Instruments Dimension 3100 atomic force microscope (AFM) with IVa controller) and fine grain texture (see Figure 3.27). The levelling power, that is the ability to produce deposits which form flat surfaces over a rough substrate, was very high as can be seen in Figure 3.28.

![Figure 3.27 Surface of electroplated nickel from 'Smooth nickel' bath](image1)

**Figure 3.27 Surface of electroplated nickel from 'Smooth nickel' bath**

![Figure 3.28 Electroplated copper and nickel showing different surface textures](image2)

**Figure 3.28 Electroplated copper and nickel showing different surface textures**

### 3.4.3.1 Nickel chloride

If chlorine is not present in a sulphaminate solution and a pure nickel anode is used then when current is passed through the bath a thick passive layer of nickel oxide forms on the anode. This oxide prevents the anode dissolving and leads to rapid depletion of nickel ions from the solution [31]. If, however, chloride ions are present in the solution, then although the oxide layer still forms, its thickness is limited and dissolution of the anode is possible
The erosion, by the chloride ions, of the passive oxide layer occurs in a non-uniform manner and leads to characteristic pitting of pure nickel anodes.

Unfortunately the presence of chloride ions in a sulphamate solution is known to cause internal stress in the deposited nickel [31]. Ideally, therefore, chloride ions would be eliminated from the solution. This is only possible by using an anode containing approximately 0.025% sulphur (sulphur depolarised nickel). Sulphur depolarised nickel anodes are known not to form any passive layers until very high current densities are applied [31]. During the electroplating sulphur depolarised nickel becomes covered in a precipitate of insoluble black nickel sulphide. To prevent this entering the solution it is necessary to place the anode metal in a titanium basket covered with a fine polypropylene cloth bag [31].

Because of the unattractive complexity of this setup, a pure nickel anode and nickel chloride was used. However, as described in section 3.4.3.2 additives were included to counter the nickel chloride stress inducing effect.

3.4.3.2 Stress

It is known that a sulphamate bath with a similar composition to the one given in section 3.4.3 (but without SDS or saccharin) produced deposits with a tensile internal stress up to 55 MPa [31]. Saccharin was included in the electroplating bath to act as a stress reducer [31]. It is thought that sulphur from the saccharin is co-deposited with the nickel and that this incorporation tends to reduce tensile stress (or increase compressive stress) [31]. It is also believed that saccharin prevents hydrogen being included in the nickel deposit which is a known cause of stress [31]. Therefore it was hoped that the saccharin would cancel the stress caused by the use of nickel chloride.

Although no systematic study of the internal stress of the nickel deposit was made, no adhesion issues were encountered suggesting that there was a low level of stress in the nickel deposits. A disadvantage of the use of saccharin is that the incorporation of sulphur into the nickel increases the reactivity of the deposit and makes it more vulnerable to corrosion (for instance during etching) [31].

Other factors that can increase the internal stress of a deposit are the uses of high current densities, a reduction in the bath temperature and the increase or decrease in pH [31]. In sulphamate solutions the efficiency of the cathode is typically lower than that of the anode because of the discharge of hydrogen. Therefore over time the pH of the solution and nickel ion concentration increases and the solution must be replaced [31].
3.4.3.3 Other additives

Sodium dodecyl sulphate was included as a surfactant. This was necessary to prevent hydrogen gas bubbles forming and sticking to the cathode (substrate) surface. These gas bubbles locally prevent metal deposition, causing pitting.

Since hydrogen is nearly always evolved at the cathode, there is a tendency for the pH of the solution to increase over time [31]. Boric acid was included as a buffer to help stabilise the pH of the solution. This was important since the rate of deposition, internal stress and many other properties are strongly affected by the pH [31].

3.4.4 Cuspidated nickel

Formulation

Nickel chloride hexahydrate, NiCl₂·6H₂O 235.4 g/l
Boric acid, H₃BO₃ 30.92 g/l
Sodium dodecyl sulphate, C₁₂H₂₅NaO₄S 0.3 g/l

(All chemicals have purities greater than 98%)

Operating conditions

Anode
Temperature 70 to 90 °C
Current density 10 to 15 mA/cm²

Nickel

The formulation of this electroplating bath was taken from ‘Nano-plating: microstructure control theory of plated film and data based of plated film microstructure’ [40]. SDS was also included to help reduce the risk of pitting caused by sticking of hydrogen gas bubbles to the cathode surface. This bath was used to produce very rough nickel deposits, Ra = 359 nm with jagged surfaces (see Figure 3.29).
This plating bath was used to deposit nickel for (some of) the cantilever tips. As discussed in chapter 2 it was hoped that this toothed texture might be beneficial in forming low resistance contacts to metals pads with thick native oxides.

At a cathode current density of 12 mA/cm² this bath was found to deposit at an approximate rate of 0.16 μm/minute. By comparison to equation 3-9 it can be seen that this corresponds to a cathode efficiency of 65.1%. This is much lower than the sulphamate bath but is within the reported range of 60 to 80% for additive free nickel plating baths [40]. If a current density higher than 15 mA/cm² was used, then the jagged texture of the surface was lost and the deposit became smoother.

It was anticipated that the high concentrations of the chloride ions in this bath would lead to deposits with a very high internal stress. Experimental evidence seemed to support this since deposits easily delaminate from their substrate. The bath formulation given at the beginning of this section (but without the SDS) is known to produce deposits with internal stresses in the range 275 to 340 MPa [31].

*Figure 3.29 Surface of electroplated nickel from 'Cuspidated nickel' bath*
3.4.5 Copper

Formulation

Copper sulphate pentahydrate, CuSO$_4$·5H$_2$O (99+ %) 60 g/l
Sulphuric acid, H$_2$SO$_4$ (95+ %, sg = 1.83) 110 ml/l
Hydrochloric acid, HCl (35.4%, sg=1.18) 123 μl/l

Operating conditions

Anode Copper
Temperature 18 to 30 °C
Current density 10 to 50 mA/cm$^2$

The formulation for this bath is taken from Modern electroplating fourth edition [31]. The sulphuric acid is used to increase the conductivity of the solution allowing high current densities to be used without significant cathode or anode polarisations. If there is no free acid then only very low current densities can be used without getting spongy deposits that contain salts. Higher acid levels also prevent precipitation of basic salts [31].

The bath contained 123 μl/l of hydrochloric acid (equivalent to 50 mg/l of chloride ions) which is known to reduce the internal stress of copper films. Despite this, very thick copper films (100 μm) demonstrated a tensile internal stress. The magnitude of this stress was not measured, however the internal stress of copper deposits from acid solutions is known to depend on current density, temperature and impurities. A range of different solutions demonstrated both tensile and compressive stresses up to a maximum of 53 MPa [31].

It was found that at a cathode current density of 23 mA/cm$^2$ the deposition rate was approximately 0.51 μm/minute. By comparison to equation 3-10 this corresponds to an approximate cathode efficiency of 100%. Very high efficiencies are common for acid copper plating solutions [31].

The surface texture of the copper deposits was rough, Ra = 204 nm, and displayed a toothed appearance (see Figure 3.30 and Figure 3.28). The surface roughness did not seem to be dependent on the cathode current density: however, it did increase with increasing deposit thickness.

This copper electroplating bath was capable of being used at high current densities (50 mA/cm$^2$) without forming nodal growths or other undesirable defects. Neither of the nickel plating chemistries previously described could be used at such high current densities without the growth of widespread defects.
3.5 Selective etching of copper in the presence of nickel

During the fabrication of the cantilevers a wet etch was required that could be used to remove copper without attacking the nickel cantilevers. The degree of undercutting the etchant produced was important since this affected the released length of the cantilevers (see fabrication process flows in section 3.7) and determined some minimum feature sizes (specifically the size of the anchor area of the cantilevers produced using fabrication method 2, see section 3.7.2).

The majority of wet metal etches rely on oxidation of the metal or the formation of aqueous metal ion complexes [41]. Both methods of etching were investigated.

The standard reduction potential of nickel (II) (Ni^{2+}(aq) + 2e^- \rightarrow Ni(s)) is - 0.257 V and the standard reduction potential of copper (II) (Cu^{2+}(aq) + 2e^- \rightarrow Cu(s)) is + 0.342 V. Therefore if an oxidising agent is capable of oxidising copper, it should also oxidise nickel. Hence to etch nickel in the presence of copper it was necessary to form a passive layer on the nickel to protect it from the etchant. In most metal etches such layers are typically metal oxides or hydroxides [41].

Four different chemistries were investigated to etch copper in the presence of nickel. The first three, acetic acid-hydrogen peroxide, ammonia and sodium persulphate were outer current-less solutions (i.e. no external electrical supply was required). The fourth solution, made from nickel sulphamate, copper sulphate and boric acid was used as an electrochemical etch.
It was found that the acetic acid-hydrogen peroxide, ammonia and the electrochemical etch were very selective, whereas the sodium persulphate etch caused the electroplated nickel to discolour. However, each etchant had its own characteristics including undercutting profiles, rate and process compatibility. The characteristics of each etch is discussed in the following sections.

3.5.1 Acetic acid and hydrogen peroxide

This etchant consisted of equal parts hydrogen peroxide (29.8% weight per volume) and acetic acid (100% glacial) diluted between 1:1 to 1:4 parts with water. Even when concentrated, this etch was relatively slow but showed very good selectivity against nickel. The etchant was compatible with photoresist (DNQ-Novolac type) and the etchant could be removed by rinsing in water. Various groups have used the etch to remove copper selectively from nickel [42-44].

Acetic acid and hydrogen peroxide can react to form per-acetic acid and water but this reaction is very slow and (even in the presence of a sulphuric acid catalyst) takes 90 hrs to reach equilibrium [45]. Therefore, we can assume that the etch consists of un-reacted acetic acid and hydrogen peroxide.

The standard reduction potential of hydrogen peroxide (H$_2$O$_2$(aq) + 2H$^+$ (aq) + 2e$^-$ → 2H$_2$O(l)) is 1.776 V which is high enough to oxidise copper and nickel (if no passive layers were formed). Therefore it is likely that the copper will be oxidised by the hydrogen peroxide as shown in equation 3-11. The overall cell voltage under standard conditions is 1.434 V.

$$
\text{Cu(s)} + \text{H}_2\text{O}_2(\text{aq}) + 2\text{H}^+(\text{aq}) \rightarrow \text{Cu}^{2+}(\text{aq}) + 2\text{H}_2\text{O}(\text{l}) \quad (3\text{-}11)
$$

The hydrogen ions shown in equation 3-11 are provided by the acetic acid. During etching the solution turns blue due to the presence of Cu$^{2+}$ and it is likely that the copper is in the form of a hydrated complex. It was noted that the etch attacked both copper and nickel if the hydrochloric acid was used instead of acetic acid.

An interesting property of this etch was that it formed shallow angled sidewalls. A copper layer that was etched through a resist mask can be seen in Figure 3.31. The origin of the undercutting behaviour is not known but it is not due to resist delamination or some similar physical phenomenon since the same behaviour occurred when an electroplated nickel mask was used.
A layer of electroplated copper (grown on an evaporated copper seed layer) was covered with evaporated titanium and copper and then electroplated to deposit a second thick layer of copper. When this two layer deposit was etched (through a resist mask) both of the electroplated layers formed angled walls. This suggests the side wall is dependent on some property of the electroplated copper and not mass-transport limitations etc. This can be seen in Figure 3.32.

![Figure 3.31 Copper etch by acetic acid-hydrogen peroxide](image)

**Figure 3.31 Copper etch by acetic acid-hydrogen peroxide**

![Figure 3.32 Two layers of copper etched in acetic acid-hydrogen peroxide](image)

**Figure 3.32 Two layers of copper etched in acetic acid-hydrogen peroxide**

### 3.5.2 Ammonia

This etchant consisted of undiluted ammonia (30% by weight). Copper in high concentrations of ammonia is known to dissolve forming an aqueous complex. The exact ligands in the complex depend on the pH but consist of a mixture of ammonia and water. Various authors have used a mixture of ammonia and hydrogen peroxide to etch copper from nickel [46, 47]. This etch is much faster than one consisting of only ammonia: however, it aggressively attacks resist (DNQ-Novolac type).

Two disadvantages of the concentrated ammonia etch were that the ammonia is an alkali, and as such slowly attacked photoresist (DNQ-Novolac type). Also, if the etched substrate was rinsed in water then the drop in ammonia concentration caused pale blue copper (II)
hydroxide to precipitate out of solution. Therefore it was necessary to first rinse the substrate in several beakers of concentrated ammonia, to dilute the copper to a very low level before rinsing in water.

This etchant produced near vertical sidewalls but also a large undercut (see Figure 3.33). The etch was much slower than the acetic acid-hydrogen peroxide etch. It appeared as though the etch proceeded from numerous discreet points across the copper surface rather than uniformly. This meant that the substrate had to be etched for a long time (over etched) to clear unmasked areas of copper. It is this requirement which may explain the vertical sidewalls. If two layers of electroplated copper (separated by an evaporated titanium copper layer as previously described) were etched through a resist mask, both layers demonstrated vertical sidewalls (Figure 3.34).

![Figure 3.33 Copper etch in ammonia](image)

![Figure 3.34 Two layers of copper etched in ammonia](image)
3.5.3 Sodium persulphate

This etch consisted of 250 g/l sodium persulphate (98%), Na₂S₂O₈, which is a common etch in the PCB industry. This etch was much faster than the ammonia or acetic acid-hydrogen peroxide etches but caused a discolouring of the electroplated nickel indicating some attack (see Figure 3.35). Sodium persulphate was compatible with photoresist (DNQ-Novolac type) and the substrate could be rinsed in water without the formation of precipitates.

Sodium persulphate is an oxidising agent, the persulphate ion has standard reduction potential \( \text{S}_2\text{O}_8^{2-} \rightarrow \text{SO}_4^{2-} \) of 2.010 V. This means that sodium persulphate can oxidise both copper and nickel (if no passive layers are formed). The overall reaction showing the dissolution of copper is given by equation 3-12 which has an overall cell voltage under standard conditions of 1.668 V.

\[
\text{Cu(s)} + \text{S}_2\text{O}_8^{2-} \rightarrow \text{Cu}^{2+} + \text{SO}_4^{2-}
\]  

(3-12)

The blue colour of the etch after use is due to aqueous copper (II) ions. In this etch the nickel is attacked slowly, suggesting any passive layer that forms is not stable. The discoloration of the nickel is unlikely to be nickel sulphate which, although the correct colour (yellow), is soluble in water and so would not deposit on the nickel surface. Instead the origin of the colour is likely to be a change in surface roughness of the nickel.

This etch produced isotropic etch profiles as can be seen in Figure 3.35.

![Figure 3.35 Copper etched by sodium persulphate. Optical micrograph shows discolouring of nickel](image)
3.5.4 Electrochemical etching

This etch was based on the work of Sun et al. [48] who use a solution of 1.6 M nickel sulphamate and 0.1 M copper sulphate (buffered with boric acid to pH 2.5) as both a electroplating bath and electrochemical etch. From this solution they deposited a copper nickel alloy and then used the same solution, with the electrodes reversed, to selectively remove the copper. In their work it is stated that ‘the formation of a passive oxide film on nickel in sulphamate solutions allows the selective electrochemical etching of copper’. The composition of this passive film is not given.

To perform the etch, the substrate containing the copper and nickel was made the anode while platinum was used as the cathode. The etch was performed at a constant current density of 10 mA/cm² and at 20 °C with a magnetic stirrer providing agitation.

It was found that the etch produced vertical profiles (see Figure 3.36) which again is attributed to the need to ‘over-etch’ the substrate to clear the exposed areas. It was evident that etching was highly affected by local field patterns and if a nickel mask was used, etching was faster around its edges than elsewhere. The effect of encircling areas of copper with conductive nickel masks was to reduce the copper etch rate in those areas. This can be seen in Figure 3.37.

![Figure 3.36 Copper etch by electrochemical method](image1)

![Figure 3.37 Electric field effect on etching](image2)
3.5.5 Etching conclusion

Electrochemical etching was found to be highly dependent on geometric factors and so of limited use. The large undercut and slowness of the concentrate ammonia etch combined with the complexities of rinsing after the etch meant it was also an unattractive choice. The other two etches (acetic acid-hydrogen peroxide and sodium persulphate) were both used. The acetic acid-hydrogen peroxide etch was the best choice if undercut was less important since it was the most selective. However, if undercutting was paramount, then the sodium persulphate etch offered less undercutting at the expense of selectivity.

3.6 Sputter deposition and improving adhesion

The adhesion of the sputtered metal layers that were deposited over the 3D AZ9260 layer (to the substrate and to the resist) was critical to the successful fabrication of the cantilevers. Strong adhesion was required to prevent the sputtered metal delaminating from the substrate. Delamination occurred when the substrate was baked by the pressure of the residual casting solvent in the resist layer.

Figure 3.38 shows devices that have been fabricated on a substrate where solvent has delaminated the sputtered (and then electroplated) metal from the substrate, causing catastrophic bubbles to form in the resist layer (used to define the cantilevers) that was being baked. This indicates that the pressure of the solvent must be resisted for a successful fabrication to occur.

Figure 3.38 The effect of metal delamination caused by rupture of resist casting solvent
Adhesion was improved by the inclusion of an oxygen cleaning step to ensure the substrate was clean of resist residues prior to metal deposition, by the use of a titanium seed layer which forms strong bonds to the glass substrate and by minimising the internal stress of the sputtered metals.

3.6.1 Bonding strength of metals to glass

It is known that metals that are oxygen-active (have a large heat of formation of oxide) tend to form strong interface bonds to both silicon and glass, whereas noble metals demonstrate poor adhesion. Noble metals such as gold only form van der Waals bonds to the substrate. Most oxygen-active metals form chemical bonds to the substrate via an intermediate oxide layer. If the atmosphere in which the metals are deposited has a very low oxygen content, then this process of intermediate metal oxide formation is dependent on the diffusion of oxygen to the substrate-metal interface. The oxygen can be trapped in the film or diffuse through the film from the atmosphere. Therefore an increase in interface adhesion can occur over time. The rate of diffusion can depend on both the film structure and the density of any surface oxide [49].

Low melting point metals (such as cadmium) with a moderate or sometimes high affinity for oxygen tend to have weak adhesion forces. This is due to the instability of the film with high surface mobility leading to a large degree of film aggregation. The cohesional forces causing aggregation are greater than those of attraction to the substrate [49].

Benjamin et al. [49] measured adhesion between thermally evaporated films and soda-glass microscope slides. Their work involved drawing a loaded point across the substrate, increasing the load until the film was removed. Since all films were deposited on the same substrate it was possible to get a relative measure of adhesion by comparing the loads at which the films delaminated. Figure 3.39 shows a selection their results. (The x-axis shows the time since evaporation in hours. In-between measurements, the samples were stored in a desiccator).
Therefore it was important to include oxygen-active adhesion layers (such as titanium or chrome) before depositing noble or low melting point metals. This is why a titanium layer was sputtered before depositing the copper electroplating seed layer.

3.6.2 Sputter deposition of titanium and copper

The pressure at which a film is sputtered is known to influence its internal stress [50]. Films were deposited over a range of pressures and their stress measured in order to determine the conditions required for a minimum internal stress. The internal stress of a film was measured using the wafer bow method [32, 50]. (100) orientated, 100 mm diameter, 500 μm thick silicon wafers were used as substrates. The biaxial modulus of the silicon was assumed to be 180.5 GPa [51].

A Moorfield MiniLab sputter system was used to deposit the films. The system consisted of three 3" magnetrons (2 DC and 1 RF) each with a corresponding shutter and quartz crystal monitor, QCM. The system had a PID controlled gate valve to manage sputter pressure. Gas admission (argon and/or nitrogen) was controlled by mass flow controllers and the system was evacuated by a rotary backed turbo pump. The pumping system achieved base pressures of 4 x 10^{-7} mbar. The substrate was mounted on a rotating platen to improve the uniformity of the deposit. All experiments were conducted using argon as the sputter gas with a flow rate of 5 sccm. Titanium was sputtered from the RF magnetron using a constant power of 204 W. Copper was sputtered from a DC magnetron using a constant current of 0.7 A.

It was found that the density of the films varied with sputter pressure and therefore the QCM could only be used to deposit the same mass on each wafer. After wafer bow measurements had been taken, the thickness (and hence relative density) of the deposited film was measured by AFM. Titanium was sputtered to an expected thickness of 30 nm.
(thickness ranged from 21 to 46 nm depending on pressure) and copper was sputtered to an expected thickness of 60 nm (thickness ranged from 47 to 69 nm depending on pressure). Wafer bow measurements were taken using an Ionic Systems stress gauge.

### 3.6.2.1 Titanium

Figure 3.40 shows the film stress of sputtered titanium. The error bars are calculated as one standard deviation of the wafer bow measurements. A measured average film thickness, calculated from five readings, was used for each individual wafer. A positive value indicates compressive while a negative stress indicates tensile internal stress.

![Figure 3.40 Internal stress of titanium film as a function of pressure](image)

Usually a QCM is used to estimate film thickness, although the sensor is in reality measuring the mass of material deposited (via a frequency shift). To calculate film thickness, the mass, area and density of the deposited film must be known. The area of the deposit is constant and defined by the QCM geometry whilst the density is also often assumed to be constant.

As already mentioned, the density of the deposited films was found to vary with pressure. By combining a post deposition film thickness measurement with the QCM measurement an average density of the film could be calculated. For a constant deposited mass it can be shown that the density of the film is proportional to the inverse of the deposited thickness. Inverses of the film thickness were calculated and the values were normalised to the minimum recorded value (Figure 3.41). The error bars represent the calculated range of normalised density based on ±1 standard deviation of film thickness.
As can be seen from Figure 3.40 a minimum internal stress was obtained by sputtering between 5 and 10 mT although a denser film was obtained by sputtering at higher pressures. This suggests that argon is being incorporated into the film at lower pressures. 6.5 mT was decided to be the optimum sputter pressure for titanium.

3.6.2.2 Copper

Figure 3.42 shows the film stress of sputtered copper. The error bars are calculated as one standard deviation of the wafer bow measurements. A measured average film thickness, calculated from three readings, was used for each individual wafer. As can be seen from Figure 3.42 all film stresses were found to be tensile.
3.6.2.1. Again it can be seen that the density of the sputtered film increases with increasing sputter pressure (Figure 3.43).

A minimum internal stress was obtained from sputtering at 20 mT (see Figure 3.42) and so films were deposited at this pressure.

The variation seen in the calculated stress values (Figure 3.42) are most likely due to errors resulting from the wafer bow measurement method. However, in absolute terms the standard deviations are similar to that shown in the stress measurements for titanium (Figure 3.40).

3.6.3 Oxygen cleaning of resist residues

Atomic force imaging showed that resist residue remained after developing the 3D AZ9260 layer. This greatly reduced the adhesion of the metal to the substrate. To remove this residue an oxygen plasma was used. Oxygen cleaning was conducted in a Yield Engineering Systems R3 plasma cleaner at a chamber pressure of 2.4 T, on a floating electrode (not electrically driven or earthed), with an RF power of 100 W for 20 minutes. Experiments were conducted on both glass and silicon substrates. The results shown in Figure 3.44 are of a silicon substrate since the very low roughness of the silicon makes the difference between the oxygen treated and the as-developed surface more pronounced.

At a power of 100 W no thermal damage to the resist occurred and SEM analysis showed no change in sidewall profiles. It was found that the higher the soft bake temperature of the resist, the more likely it was to leave a residue after development.
3.6.4 Adhesion results

No quantitative analysis of adhesion was conducted. However the ‘scotch tape’ test indicated a marked improvement. With no oxygen cleaning and electron beam deposited (not stress controlled) metal films it was found that scotch tape removed the metal film from both the glass substrate and resist layer. Including an oxygen clean step improved adhesion such that no metal could be removed from the glass: however, it could still be removed from the resist. Using sputtered metals (with controlled stresses) as well as oxygen cleaning resulted in no metal being removed by the tape from either the glass substrate or resist.

3.7 Fabrication of cantilevers

A full description of the process steps for fabricating the cantilevers is given in appendix 3-1. In the following sections only a brief description of the cantilever fabrication is given. The detail of AZ9260 processing, sputtering, electroplating and etching, that made the fabrication possible, have already been discussed and are referred to where appropriate.

Two fabrication methods are described. Method 1 was the first process developed and allows the maximum possible cantilever design freedom. In this method the sacrificial structure is defined and then cantilevers are fabricated on top. The entire sacrificial structure is then removed releasing suspended inclined cantilevers. However, this method suffered
from low yield caused by the pressure of residual casting solvent in the 3D resist layer. This problem was partly mitigated by changing the order of processing such that part of the sacrificial structure remained after the cantilevers were released – this is method 2. It will be shown however, that this restricts the cantilever design in that the anchor areas of the cantilevers must be large enough not to be released by undercutting during a copper etching step.

3.7.1 Method 1

The first stage in the fabrication was to etch alignment marks into the borosilicate substrate as in Figure 3.45(a). Etching alignment marks made the fabrication easier since many of the subsequently deposited layers, for instance thick photoresist which could easily deform at elevated temperatures, were not suitable for accurate alignment.

The next stage in the fabrication was to deposit the sacrificial structure. This began with the spinning and patterning, using a dithered binary mask, of a 55 μm thick layer of AZ9260 photoresist (Figure 3.45(b)). This was used to form the basis of the 3D layer over which the cantilevers were later deposited. The details of the resist processing were given previously in section 3.3.

As discussed in section 3.3, the resist was not thermally stable and so could not be used as a sacrificial layer on its own. To overcome this it was necessary to cover the resist with a metal film. The overall resist-metal structure had the same shape as the resist but was much more thermally stable. This was achieved by sputter coating the substrate with a titanium adhesion layer followed by copper seed layer (Figure 3.45(c)) and then electroplating copper to form a thick (5 μm) non-porous covering (Figure 3.45(d)).

Before the metal was sputtered over the resist, an oxygen plasma was used to remove any resist residue from developed areas of the substrate. This ensured good adhesion between the titanium and glass (as described in 3.6.3). The titanium adhesion layer and copper seed layer were also sputtered at pressures to minimise film stress to further discourage delamination (see section 3.6.2).

Before copper was electroplated, PMMA (dissolved in chlorobenzene) was painted over the alignment mark areas which were placed several millimetres from the device areas. This provided a simple method of preventing copper electroplating over the alignment marks. Since the copper surface was rough (see section 3.4.5) it was not very reflective which caused difficulties when trying to perform subsequent mask alignments.
Novolak (the main constituent of AZ9260) is not soluble in chlorobenzene. Since the sputtered metal films were found to be gas permeable, the use of chlorobenzene as a solvent for the PMMA eliminated the risk of damage to the AZ9260 resist from evaporating fumes. Also, it was found that chlorobenzene was sufficiently volatile that the PMMA could be left
to dry without the need for baking. After the copper had been electroplated the PMMA was removed in acetone and the wafer was washed in IPA and then water.

The next stage in the fabrication was to provide areas where the cantilevers would be attached to the substrate. These anchor points were defined by patterning a thick layer of AZ9260 resist (Figure 3.45(e)) and etching the sacrificial structure (Figure 3.45(f)) to reveal the borosilicate substrate. Sodium persulphate was used to etch the copper (see 3.5.3) and dilute hydrofluoric acid was used to etch the titanium. The resist etch mask was then removed in 400T resist stripper which was found to leave no resist residue. This completed fabrication of the '3D substrate' (Figure 3.46(a)) and the cantilevers could now be fabricated.

Figure 3.46 (a) shows a wafer after the sacrificial structure was completed. (b) shows a wafer after the cantilevers were fabricated but before release.
The first step in fabricating the cantilevers was to sputter a thick titanium (50 nm) layer followed by thin nickel (10 nm) seed layer (Figure 3.45(g)). The nickel seed layer provided a good surface for subsequent nickel electroplating. The titanium layer was used as both an adhesion layer to the borosilicate wafer and as the current carrying layer for subsequent nickel electroplating steps. The sputtered nickel seed was porous which allowed the titanium layer beneath to be etched without first etching the nickel layer. This was important since it was found very difficult to etch the very pure sputtered nickel in the presence of less pure electroplated nickel (see 3.4.3).

A thick layer of AZ9260 was coated and patterned to define electroplating moulds for the formation of the cantilevers (Figure 3.45(h)). Nickel from the ‘smooth nickel’ bath was then electroplated to form the cantilevers (Figure 3.45(i)). This resist mould was then stripped and a final thick layer of AZ9260 was coated and patterned to define a mould for the cantilever tips (Figure 3.45(j)). Nickel from either the ‘smooth’ or ‘cuspidated’ formulations was then electroplated to form the tip (Figure 3.45(k)). The resist mould was then stripped completing the deposition of the probe structure (Figure 3.46(b)).

For some devices a gold covering was desired on the cantilever tip. This was achieved by evaporating a titanium adhesion layer followed by gold over the substrate before the tip plating mould was removed. By performing evaporation at this stage it prevented metal depositing anywhere except on the surface of the cantilever tip (see Figure 3.47).

![Ti/Au film on tips plated from “smooth nickel”](image1)

![Ti/Au film on tips plated from “cuspidated nickel”](image2)

*Figure 3.47 Ti/Au films on cantilever tips patterned using ‘lift-off’ technique*

Having fabricated the cantilevers, the sacrificial resist-metal structure had to be removed. The sputtered titanium and nickel was first removed by etching in dilute hydrofluoric acid. This attacked the titanium beneath the 10 nm nickel layer and has the effect of ‘lifting off’
the nickel. The copper was then selectively etched using acetic acid-hydrogen peroxide (see 3.5.1) (Figure 3.45(m)). Then hydrofluoric acid was used to remove the titanium adhesion layer (Figure 3.45(n)) and the 3D resist layer was remove using 400T stripper (Figure 3.45(o)).

It was found that if rinsed cantilevers were allowed to dry in air then some of the slender and finely space cantilevers suffered lateral stiction. No stiction to the substrate was observed since the gap between the cantilever and substrate was relatively large.

To prevent stiction the cantilevers were dried in a critical point dryer (Autosamdri 815B manufactured by Tousimis, Rockville, Maryland, USA). To use critical point drying it was necessary to replace the water of the final rinse with IPA. To do this the wafer was soaked in several different baths of IPA before being placed in the critical point dryer.

3.7.2 Method 2

This fabrication method begins in the same way as method 1. Fabrication stages Figure 3.48(a) to Figure 3.48(d) are the etching of alignment marks and the fabrication of the sacrificial structure and are the same as described in section 3.7.1. Instead of then etching holes through the sacrificial layer, AZ9260 resist was used to define cantilevers plating moulds (Figure 3.48(e)). Cantilevers were electroplated using the ‘smooth nickel’ bath (Figure 3.48(f)). AZ9260 resist was then used to define tip plating moulds and the tip as electroplated using either the ‘smooth’ or ‘cuspidated’ nickel baths (Figure 3.48(g)). As in method 1 if gold was desired over the cantilever tips it was evaporated over the electroplated tips before the resist mould was removed in 400T stripper (Figure 3.48(h)).

The fabricated cantilevers were then ready for release. Since the cantilevers were not attached directly to the substrate the release was staggered to allow all the copper to remain in areas where the cantilever was to be anchored to the substrate, but removed everywhere else. First AZ9260 was spun and patterned to protect anchor regions (Figure 3.48(i)). This required a mask of inverse polarity to the one used to define anchor points in method 1. Copper was then selectively etched using acetic acid-hydrogen peroxide and the titanium seed layer was etched in dilute hydrofluoric acid (Figure 3.48(j)). Then 3D resist layer and resist protecting the anchor regions was removed in 400T stripper (Figure 3.48(k)).
Figure 3.48 Fabrication method 2

(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l) (m)

Nickel — Titanium — AZ9260 — Copper

Titanium — Glass substrate

112
The copper was etched in sodium persulphate to remove the material from between the cantilevers in the anchor region (Figure 3.48(l)). This etch had to be carefully controlled since over-etching caused undercutting of the cantilever anchor and eventually the release of the entire cantilever from the substrate. For this reason sodium persulphate rather than acetic acid-hydrogen peroxide was used as it produced smaller undercuts (see section 3.5). Finally the titanium adhesion layer was etched in dilute hydrofluoric acid (Figure 3.48(m)) and the released cantilevers were dried in a critical point dryer.

3.8 Fabrication results

Figure 3.49 shows SEM images of various structures. Figure 3.49 (a) shows a cantilever array which was fabricated using method 2, with all other structures fabricated using method 1. The spiral and hoop structures (Figure 3.49 (c) and (d) respectively) were designed for probing solderballs and could be created by simply changing the geometry of the mask used to create the cantilever.

Cantilevers of a variety of widths, lengths and thickness were fabricated. The minimum cantilever width fabricated was 13 μm and the minimum pitch (tip centre to tip centre) demonstrated was 23 μm. The maximum horizontal length fabricated was 700 μm and the thickest cantilever deposited was 18 μm.

For the aligner and resist used the (minimum) values given for width and pitch represent a limit of the fabrication method. A thickness of 18 μm was found to be the limit before over-plating started at the tip end of the cantilever.
Figure 3.49 A selection of fabricated structures
3.8.1 Comparison of method 1 and method 2

In method 2, the maximum thickness to which copper could be electroplated was limited by two factors. The first was that the thickness of the copper had to be less than half the width of the cantilever anchor. If this was not the case even an isotropic etch would cause the cantilever to be released from the substrate when the copper between the cantilevers was etched. If the etch had a faster undercut rate then the anchor must be wider that this constraint. The second constraint on copper thickness is that sodium persulphate slowly attacks the electroplated nickel. This demands the thickness of the electroplated copper is minimised, since this in turn minimises etching time and hence damage to the nickel.

Therefore in method 2 there is a trade off between geometric stability of the sacrificial structure (which is favoured by thicker electroplated copper) and minimising the width of the cantilever’s anchors and deterioration of the nickel surface (which is favoured by thinner electroplated copper). This trade off does not exist in method 1 where there is complete freedom in the anchor geometry and electroplated copper thickness (within limits of the resist processing).

An advantage of method 2 is that there is a much smaller chance of the residual solvent in the 3D resist causing delamination at the interface between the substrate and titanium-copper layers (see Figure 3.38). This delamination occurred whilst baking resist layers, and caused bubbles to form in these layers, destroying the devices in the surrounding regions and affecting the quality of the mask contact across the entire wafer. The chance of delamination was decreased because the area of the interface was much greater if holes for the anchors are not etched.

3.8.2 Sacrificial structure stability and tip co-planarity

As already described, the resist in the sacrificial structure contained a large percentage of residual casting solvent. As the wafer was heated (during resist bakes) this could cause either catastrophic failure (as shown in Figure 3.38) or the distortion of the structure’s shape. The distortion increased as the size (volume) of the sacrificial structure increased. This was because the reinforcement provided by the electroplated copper film (which was the same thickness for each sacrificial structure) decreased as the structure’s area increased.

Since the released cantilevers follow the shape of the sacrificial structure, the distortion of the sacrificial structure could be measured from the shape of the released cantilevers. As
expected it was found that the distortion was most noticeable in the longer cantilevers; this can be seen in Figure 3.50.

The devices shown in Figure 3.50 were fabricated using a 9 µm thick electroplated copper film to reinforce the sacrificial resist. It can be seen that this was effective up to a horizontal cantilever length of between 300 and 400 µm after which a ‘bulge’, positioned between the middle and tip of the cantilever, can be seen. The effect on longer cantilevers is to cause the cantilevers to form an ‘arc’ where their tips are no longer the highest point from the substrate. An example of this can be seen in the ‘hoop’ structures shown in Figure 3.49 (d).

*Figure 3.50 SEM images showing effect of distortion of sacrificial structure.*

Measurements of selected cantilevers were taken using a scanning optical profiler (Stil Micromasure system manufactured by Stil SA, Aix en Provence, France), shown in Figure 3.51.
The brown dotted lines indicate where the section profiles (shown on the right of the figure) were taken. As can be seen the bulge caused by solvent pressure affects the longer cantilevers the most.

![Figure 3.51 Surface profiles of selected devices after release](image)

The sacrificial structures can be approximated as (non-symmetric) trapezoidal prisms (see Figure 3.18). The short inclined side of the trapezoid (non-cantilever side) was a fixed gradient (horizontal length 100 μm, angle of inclination 29° from the substrate). Irrespective of the cantilever length, the tips of the cantilevers were positioned close to this steep gradient. Therefore, the tips were positioned on a relatively stiff region of the sacrificial structure and so were less affected by solvent pressure than the centre region of the cantilevers. This can be seen in the section profiles as there is no large variation in the tip heights from the centre of the section to the edges (where the sacrificial structure is even stiffer).

The data shown in Figure 3.51 was used to determine the tip co-planarity of the cantilevers; the results are shown in Figure 3.52. The box plots define the median, 25th and 75th percentiles whilst the whiskers represent the 5th and 95th percentiles. The mean is
indicated by a square marker, the 1st and 99th percentile by cross shaped markers. The maximum and minimum values are shown by small horizontal dashes.

The general downward trend in tip height with increasing cantilever length is due to a residual strain gradient in the electroplated nickel and is described in chapter 5. The variability in tip height across a section is on average better than 2 μm. Tip-tip height variations for the 400 μm long sample are greater than the 200 or 700 μm long samples. This is attributed to mechanical damage during or after release. It is unlikely such variation could be caused by resist non-uniformity (during spinning) or resist distortion.

![Graph showing tip height variation as a function of cantilever length](image)

Figure 3.52 Tip height variation as a function of cantilever length

### 3.9 Conclusions

The fabrication method described in this chapter can be used to produce inclined suspended nickel cantilevers as specified in chapter 2. The minimum cantilever pitch demonstrated was 23 μm while the tip co-planarity was better than ± 2 μm. 23 μm pitch satisfies the design objective of less than 25 μm specified in chapter 2. Cantilevers of horizontal lengths 100 to 700 μm and widths of 13 to 50 μm have been fabricated. The fabrication was based on batch processing and was independent of the substrate choice. This satisfies the low manufacturing cost and time and also the independent fabrication method specified in chapter 2.
References

19. Chapman, G.H., et al., Creating precise 3D microstructures using laser direct-write bimetallic thermal resist grayscale photomasks. in Photon Processing in


Chapter Four

Testing rig design

This chapter is concerned with the development of custom built test equipment. First, some relevant testing techniques used to assess thin films and MEMS structures are introduced. Development of a custom testing rig that was used to measure both mechanical (load-deflection) and electrical (contact-resistance) performance is then given. The errors in terms of accuracy and drift of the rig are then discussed. Finally, the fabrication of a test tip structure which was required to measure contact resistance is described.
4.1 Testing of MEMS devices

MEMS are tested to determine functionality, environmental resilience and in-service reliability. Such testing can be complex since it could involve electrical, mechanical, thermal, audible, chemical, optical or other stimuli and measurements [1].

Testing is often facilitated by the inclusion of dedicated test structures fabricated in parallel with the devices. These allow the mechanical testing of thin film properties such as adhesion, internal stress, Young’s modulus and yield stress. This type of testing is most important in surface machined devices where the material properties of the structural films are a function of deposition parameters.

As well as thin film properties, test structures can also help in the testing of device functionality. For instance an extra resistor can be included to provide a heat flux for testing of a temperature sensor.

Mechanical properties such as Young’s modulus and yield stress are very important to the design of a MEMS probe card since they, in conjunction with the geometry of the probe, determine how stiff the probe is and how robust it will be. Below are summarised some of the important methods of testing the mechanical properties of thin films.

4.1.1 Mechanical characterisation of thin films and MEMS devices

Various techniques have been developed to determine important material properties (Young’s modulus, Poisson’s ratio, yield strength, fracture strength, hardness, internal stress etc.) and surface properties (coefficient of friction, morphology, interface strength, wear rate etc.) of thin films [1]. Some of these techniques can be used to directly to record load-displacement behaviour of MEMS structures. Below are listed some of the more common techniques for measuring these properties.

4.1.1.1 Resonant frequency measurements

The resonant frequency of a cantilever can be used to determine its elastic properties if the dimensions are known [2]. The resonant frequency is measured by exciting the cantilever over a range of frequencies and measuring the amplitude of response. A piezoelectric actuator is often used to vibrate the cantilever although, if the MEMS device can be
electrically driven, for instance by a comb drive, then this can be used to excite the structure directly.

The amplitude of response can be determined by measuring the deflection of a laser reflected from the end of the cantilever [3] or laser Doppler vibrometry [4]. A more detailed analysis of the deflection of a structure can be obtained using stroboscopic interferometry [5] or real-time holography [6]. By these methods the movement of all points on a surface can be measured simultaneously.

4.1.1.2 Atomic force microscope (AFM)

The AFM was developed in the mid 1980s for very high resolution measurement of surface topography [7]. At its most basic an AFM scans a cantilever, at the end of which is an extremely sharp tip, across a surface in a raster scan such that the same level of force is applied throughout the scan. To do this the cantilever must be raised and lowered following the surface topography. The level of force being applied is determined by the deflection of the cantilever, which is measured by the deflection of a laser reflecting from the cantilever end. A control loop is used to move the cantilever up and down in response to a change of deflection [8].

The technique has been expanded, by use of specially designed tips, to allow simultaneous mapping of topography and other surface properties. Such properties include thermal, magnetic, surface potential, capacitance, spreading resistance, electrostatic force, stiffness and friction properties, amongst others [9, 10].

An AFM can be used to perform force-deflection measurements. The cantilever can be brought into contact with the suspended structure causing it to deflect. Since the spring constant of the AFM cantilever is known, by measuring its deflection it is possible to determine the force being applied at any given time. This technique can only be applied to structures of similar or lower stiffness than the AFM cantilever. Typically an AFM can apply deflections less than 10 μm, and measure forces in the $10^{-8}$ to $10^{-5}$ N range.

It is also possible to perform nano-hardness or wear tests using an AFM. Hardness can be measured by indenting a surface using the tip of a nitride cantilever with a known force and then using the AFM to measure the resulting impression. Nano-wear test can be performed by scanning a tip back and forth across a surface whilst applying a known load. A limitation of this technique is that, if using standard AFM cantilevers, there is a very high contact pressure resulting from the extremely small radius of curvature of the tip.
4.1.1.3 Nano-indentation

Nano-indentation involves bringing a hard tip, with a radius of curvature of a few microns, into contact with a surface. The load on the tip is ramped from zero to the set value, held for a defined period before being ramped back to zero. During this load cycle the vertical displacement of the tip is continuously measured. After indentation the resulting impression is measured by an AFM. This technique can be used to measure the local hardness, fracture toughness and elastic modulus of very thin films [11]. Typical nano-indenters apply forces in the range 0.025-500 mN and can measure deflections of 100 μm with sub nanometre resolution [12]. This type of instrument can be used directly for load-deflection testing of suspended structures.

4.1.1.4 Scratch testing

Scratch testing is primarily used to measure the adhesion of a thin film to a surface. A diamond tip is brought into contact with a sample; an increasing load is then applied as the tip moves across the surface of the sample at a set velocity. During the movement the frictional load (lateral load) and vertical displacement of the tip are recorded. An acoustic sensor attached to the tip picks up sounds associated with chipping and delamination of the film [1, 13].

The normal load that is applied depends on the thickness of the film being tested, with thicker films requiring higher the loads. Micro-scratch testing is used for films greater than 0.5 μm thickness and applies loads of up to 30 N. Nano-scratch testing is used for films of less than 1 μm thickness and applied loads of up to 1 N. Nano-scratch instruments have normal load resolutions between 15 and 0.15 μN, vertical displacement resolutions of 1.5 nm and frictional force resolutions of 6 to 30 μN. Typical loading rates vary from 5 to 100 N/min whilst scan velocities are in the range 0.2 to 20 mm/min [14].

This technique measures lateral force and so can be used to perform tension testing of in-plane tension testing samples. The film to be tested is patterned into a suspended beam anchored to the wafer at one end with a hoop at the other. The tip of the nano-scratcher is positioned into the hoop and used to apply an axial load, measuring force and displacement as it moves.

4.1.1.5 Bulge testing

Bulge testing typically consists of depositing a film on an appropriate substrate and then removing a region of the substrate such that the film forms a suspended membrane. By
applying a pressure difference across the film and measuring its deflection the residual stress, biaxial modulus and yield stress of the film can be calculated. This method works well for low modulus materials such as polymers [15].

It is also possible to assess the adhesion between the film and substrate by applying sufficient pressure to cause delamination between the two whilst measuring the injected volume and pressure [15].

The displacement of the membrane is usually measured by interferometry and therefore has sub-nanometre resolution.

4.1.1.6 Film stress measurement techniques

Bulge testing is one method of measuring film stress but it is not the most common. Film stress is most simply measured by depositing the film of interest onto a substrate of known mechanical properties and then measuring the resulting deformation. If various geometric criteria are met then it is possible to calculate the stress in the deposited film [15]. Other techniques for measuring film stress include determining the critical buckling length of a suspended beam anchored at both ends. The critical length is found by fabricating arrays of beams of different lengths [15]. X-ray techniques can also be to measure material strain and hence imply stress if appropriate elastic material properties are known [16].

Stress gradients in thin films can be measured by observing the out of plane deflection of suspended cantilevers [15].

4.1.1.7 Uni-axial tension testing

Uni-axial tension testing is performed by anchoring one end of a sample whilst applying a known rate of displacement to other. As the sample is stretched the force applied to the sample is measured. From the load-deflection data stress-strain data is calculated and Young’s modulus, yield stress and ultimate tensile stress (UTS) can be extracted. By combining uni-axial testing with bulge testing (section 4.1.1.5) it is possible to calculate Poisson’s ratio, since both the bi-axial and Young’s modulus are known [15].

4.2 Design of testing rig

In this work, once cantilevers had been fabricated it was necessary to determine if they would be suitable for use as electrical contacts in a probe card. In order to determine this,
both their mechanical (load-deflection and yield behaviour) and contact resistance behaviour had to be assessed.

A commercial prober typically applies 2 - 3 mils (51 - 76 μm) of overdrive, at which each probe typically applies 10 gf (~100 mN). The only commercial instrument described in section 4.1 which would be capable of measuring this range of forces and displacements is a nano-indenter. Since these instruments are very expensive, it was decided that satisfactory performance could be obtained by building a specialised instrument to perform the testing. For this reason a custom testing rig was developed which allowed the load-deflection behaviour of the probes to be measured whilst simultaneously measuring the contact resistance between the probe and a specified surface.

The design of the testing rig is discussed in the following sections.

### 4.2.1 Hardware design

![Diagram of the testing rig](image)

**Figure 4.1 Diagram of the testing rig**

The testing rig is shown in Figure 4.1. A wafer containing cantilevers to be tested is placed on a balance. A rigid object, discussed later, is brought into contact with the end of a single cantilever causing it to deflect. The balance is used to measure the load that is applied.

Since the surface of the balance moves as a load is applied, a correction must be made to determine the deflection of the cantilever. The cantilever and balance can be visualised as simple springs, as shown in Figure 4.2.
From Figure 4.2 it can be seen that if the spring constant of the balance, $K_{balance}$, the applied total displacement, $\Delta_{total}$, and applied force, $F$, are known then the true displacement of the probe can be calculated from:

$$\Delta_{device} = \Delta_{total} - \Delta_{balance}$$

where $\Delta_{device}$ represents the deflection of the cantilever, $\Delta_{balance}$ represents the displacement of the balance, $F$ represents the applied force and $K_{balance}$ represents the spring constant of the balance.

It can be seen from equation 4-1 that if the balance spring constant is much stiffer than the cantilever being tested, then any error incurred by not adjusting for the balance movement will be small. Similarly, it can be seen that it would not be possible to deflect a cantilever if its stiffness was much greater than that of the balance. The spring constant of the scales could be determined by measuring the load-deflection behaviour directly of the balance measuring pan.

If only load-deflection data for a cantilever was required then a blunt tungsten needle was used to apply the known deflection. If contact resistance data was also required then a specially designed “test tip” was used. The operation, design and fabrication of this “test tip” are given in section 4.4.

The tungsten needle (or test tip) could only travel in Z and so the balance was placed on a stack of micrometer stages that allowed adjustment of X, Y, Z and $\theta$ to align the needle to the cantilever. A binocular microscope was mounted such that a user could view the wafer to perform this procedure. The whole rig was contained in a large cabinet to reduce the effects of draughts and was placed on an anti-vibration table. Finally the rig was situated in a clean
room to allow control of ambient temperature and humidity and was placed on an air table to reduce the affects of vibration. A photo of the testing rig is shown in Figure 4.3.

![Figure 4.3 Photo of testing rig](image)

4.2.1.1 Selection of displacement transducer and sensor

It was decided that the most versatile method of moving the tungsten needle (or test tip) was to use a nano-positioning stage. A Melles Griot nanostep PTS 1000 and 17MDU002 nanostep control unit were chosen. The nanostep PTS 1000 provides a range of travel of 100 mm and can move at a maximum velocity of 6 mm/s. The stage consists of a stepper motor which turns a threaded shaft, which in turn drives the movement of a plate onto which objects can be mounted.

The stage can be used with a load of up to 100 N when moving the plate away from the motor, 50 N when moving the plate towards the motor and up to a 500 N load in directions orthogonal to the direction of travel. This means that relatively large and heavy fixtures could be moved with high precision. This would not have been possible if, for instance, a piezo actuator had been selected.

The 17MDU002 controller has an open loop positioning resolution of 0.5 μm, a positional linearity of 0.02% and a positional accuracy of ±2 μm. This level of placement was not thought to be sufficiently accurate, and therefore the position of the stage was measured using a linear variable differential transformer (LVDT).

An LVDT manufactured by Solartron Metrology model DG / 2.5 was selected. This LVDT had a built-in oscillator circuit permitting use of a DC supply. It could measure ± 2.5 mm of
travel and had a non-rotating, spring loaded shaft. This allowed the stage position to be measured without having to attach the LVDT permanently to the plate. The stiffness of the spring loaded shaft was 13 g/mm, which represents a maximum load of 0.64 N, well below the limit of the stage. The response time of the LVDT is less than 3 ms.

The use of different length tungsten needles (or test tips) meant that the position of the stage, relative to the wafer, was not always the same. By mounting the LVDT on a Z micrometer stage it was possible to adjust its relative position to account for this. This stage was of course not adjusted during actual measurements.

A Bench-top DC supply manufactured by Isotech, model IPS2303D was used to power the LVDT. The DC output of the LVDT was measured using an Agilent 34401A 6½ digit multimeter.

4.2.1.2 Selection of balance

It was decided to use an A&D GF-300 series balance with a measuring resolution of 0.001 g (9.81 μN). This balance could weigh up to a maximum of 310 g, which allowed ample range to place more equipment than just the wafer being tested on the balance. One source of extra weight was a wafer chuck. This helped prevent unwanted movement of the wafer whilst manipulating the balance position as well as providing extra mass to dampen any vibrations.

It was thought that 9.81 μN resolution would be adequate, since a traditional probe applies up to 100 mN. In retrospect a higher resolution balance might have been preferable for measurement of the longest cantilevers. At resolution higher than 9.81 μN, however, it was universally found that the draught guard around the balance measuring pan was permanently attached and would have to had been cut away to permit use in the rig. This was not the case with the GF-300 which had a removable draught guard.

The GF-300 was calibrated using a 100 g weight before each measurement session. The balance had repeatability of 0.001 g, a linearity of ± 0.002 g, a stabilisation time of less than 1 second and a temperature drift of less than ± 2 ppm/°C.

4.2.2 Software design

Software to control the movement of the nanostep stage, record the LVDT output voltage, load on the balance and reading from the ohmmeter was written using National Instruments
LabVIEW 7.0™ Express. Screen captures of the two most important user interfaces are shown in Figure 4.4 and Figure 4.5.
The software allowed a user to define a step in terms of distance, direction, and repetition value. After each movement, the software paused for a defined period, before measuring the balance, LVDT (stage position) and (if selected) resistance. The pause allowed values on the instruments to settle before readings were taken.

It was possible to set the software to take multiple readings of the balance for each step and then calculate the average and standard deviation to obtain a more accurate reading for force.

Once the desired number of steps had been completed the software could output the data as a text file for further analysis. The data was also displayed graphically in real time as it was captured. Coloured indicators and text boxes showed which part of a cycle the software was currently in, as well as which step it was performing.

Facilities were included to allow independent control of the stage which was used when aligning the tungsten needle (or test tip) to the end of a cantilever to be tested.

Other screens (not shown) allowed debugging of the balance and nanostep controller by accessing their communication buffers, an option to zero the balance, an option to cycle the stage up and down a set distance a set number of times (with a defined pause between movements) and an interface for directly obtaining the output voltage of the LVDT.
Because LabVIEW™ is a visual programming language it is difficult to document. Instead of full documentation, simplified flow diagrams of the important software operations are shown in Figures 4.6 to 4.10.

Figure 4.6 Principal control loop

Figure 4.6 shows how the software performed a load-deflection cycle. After instrument initialisation the starting position of the stage was recorded from the LVDT voltage. The stage was then moved a set distance and the software paused to allow the sensors to settle. The stage position was then recorded and the distance moved relative to the starting position was calculated. If selected, the resistance was then measured followed by multiple reads of the balance. The average and standard deviation of mass recorded by the balance was then calculated and converted into a force. The force, displacement and resistance data were then plotted on separate graphs and the cycle was repeated until the set number of moves had been completed.
Figure 4.7 shows how the communication channels to the ohmmeter, balance and stage controller were configured. It represents the "set up balance, ohmmeter and stage controller" icon shown in Figure 4.6.

Initialisation parameters are shown in Figure 4.7. A critical initialisation step included turning off "zero tracking" on the balance. This was a built-in function that is usually used to combat drift by dynamically reassigning "zero" if a slow drift was seen. This obviously caused problems when the balance was used to detect gradual small increases in load.

Figure 4.7 Device initialisation steps
Figure 4.8 represents more detail of the “move stage” algorithm, Figure 4.9 represents more detail of the “measure resistance” algorithm and Figure 4.10 represents more detail of “read balance” algorithm referred to in Figure 4.6. Further details of reading from the DVM are not given since this device had publicly available programming objects which allowed simple communication.

Figure 4.8 Routine to move stage
Figure 4.9 Routine to record resistance value from ohmmeter

Figure 4.10 Routine to record weight on balance
4.3 Testing rig measurement error estimation and balance spring constant

4.3.1 Displacement sensor and transducer

The position of the LVDT was calculated as follows:

\[ \text{Position} = \frac{\text{measured output voltage}}{\text{supply voltage} \times \text{LVDT sensitivity}} \] (4.2)

The LVDT had a calibrated sensitivity of 73.219 mV/V·mm. The supply voltage was set to 20 V and the output voltage was measured using a DVM set on the 1 V or 10 V scale. The 1 V scale has a resolution of 1 μV which corresponds to a displacement resolution of 0.7 nm and the 10 V scale has a resolution of 10 μV which corresponds to a displacement resolution of 7 nm.

At 23 ± 5 °C the variation of the voltage supply was less than 0.01% + 3 mV with less than 1 mV (RMS) ripple. This corresponds to a possible absolute error of ± 6 mV or a relative error of 0.03%. At 23 ± 5 °C the DVM on the 1 V scale had a relative error of 0.003% + 7 μV. This corresponds to a possible absolute error of ± 37 μV or a relative error of 0.0037%. Similarly, if the DVM is used on the 10V scale and measures a maximum signal of 3.7 V (full deflection of the LVDT) then the maximum possible absolute error was 124 μV or a relative error of 0.00127%

The LVDT has less than 0.07% non-linearity error.

Combining these errors to calculate bounds on the measurement error of position gives a relative error of less than 0.104%. The maximum deflection range measured was less than 100 μm and so over that range the position was known to ± 0.1 μm.
4.3.2 Stage positioning

Figure 4.11 shows the measured stage movement versus the position sent by the stage controller. It can be seen there is a local phenomenon of an oscillating error with a period of about 6 μm caused by the construction of the stage. Additionally there is the global error trend which over a larger sample (millimetres of travel) averages to zero.

![Figure 4.11 Plot showing LVDT measured stage position vs. controller instructed position](image)

Using the LVDT it was clear that large positional errors occurred when the stage changed direction (that is the difference between the instructed position and measured position).

Although the position of the stage could only be controlled to a relatively coarse degree (± 2 μm) it has been shown that its position could be measured with good accuracy.
4.3.3 Drift

Mechanical settling and the use of materials with varying coefficients of expansion for different parts of the testing rig caused relative movement between different parts over time. The drift of the (stationary) stage and balance weighing pan relative to the LVDT mounting (as measured by the LVDT) are shown in Figure 4.12 and Figure 4.13.

The fact that both drifts show a net value with decreasing rate suggests that mechanical settling dominates the drift. The temperature of the clean room is controlled and so if it dominated the drift would average to zero over time.

The drift shown above introduces the largest uncertainty in the measurement of displacement. It was found that rates of drift of up to 0.2 μm/min occurred. A typical cantilever deflection of 60 μm may take 5 minutes to perform which corresponds to an uncertainty of 1 μm or less than 2%.
4.3.4 Balance spring constant

The spring constant of the balance was calculated by pressing a tungsten needle directly into the pan of the balance and measuring load-deflection. From Figure 4.14 the spring constant of the balance was calculated to be 11.7 mN/μm. A typical commercial probe has a stiffness of less than 2 mN/μm which was considerably less than the stiffness of the balance (and as expected the cantilevers tested). Therefore the balance was a suitable method of measuring the load-deflection behaviour of the fabricated probes (see equation 4-1).

![Figure 4.14 Load-deflection of balance using tungsten needle](image)

4.4 Contact resistance measurement

4.4.1 Concept of parallel contact resistance and load-deflection testing

One of the critical parameters of a probe is the electrical contact resistance between the probe and the bond pad it is contacting. To understand this contact resistance it is important to know the load being applied since this largely determines the contact area. The contact load determines whether the native oxide layers, that can exist on both the bond pad and probe, are fractured, providing a low resistance contact. For this reason it was decided that the contact resistance should be measured in parallel with the load-deflection behaviour of the cantilevers.
To measure the contact resistance a circuit had to be created that included the contact of the cantilever tip to a specified metal surface representing a bond pad. To achieve this, cantilevers were fabricated and electrically connected in pairs to form a circuit (red line in Figure 4.15) without the need to have direct electrical connections from the wafer containing the cantilevers to external testing equipment. This could be done by bringing two electrically isolated metal surfaces with external electrical contact to an ohmmeter into contact with two adjacent cantilevers. To calculate the contact resistance all the bulk resistances had to be calculated and subtracted from the recorded total resistance value.

Not having any external electrical connections from the wafer containing the cantilevers was important not only because there were 1000s of cantilevers on each wafer (which is a very large number of connections) but also because any physical contact to the wafer (other than the test tip) would cause an erroneous reading of on the balance.

As can be seen in Figure 4.15, the cantilevers to be tested were fabricated in large arrays. Therefore it was not possible to use a flat wafer covered with a patterned metal film to represent bond pads (since if the two wafers were brought into contact every cantilever would be deflected simultaneously). Therefore a specially fabricated “test tip” was developed that allowed deflection of two adjacent cantilevers.
4.4.2 Test tip geometry

Initially it was considered that a suitable geometry could be achieved by structuring the surface of a wafer, for instance with KOH crystallographic etching, to give a protrusion such as shown in Figure 4.16.

![machined wafer providing electrical contact to selected cantilevers](image1)

Figure 4.16 Using metallisation on bulk etched wafer as test tip

It was realised that the tolerance of parallelism between the two wafers, to prevent unwanted areas of contact, would be very hard to achieve. Consider a protrusion depth (on the top wafer) of 100 μm and that the tips of the cantilevers were 60 μm above the surface of the bottom wafer. If both wafers had a 2 inch diameter then the maximum angle of offset between the two wafers would be 0.04°.

Instead a high aspect “needle” type of structure as shown in Figure 4.17 was developed. Because this structure had small surfaces “at the end of a needle” it allowed a much greater angular offset than the surface machined wafer idea.

![wafer containing probes](image2)

Figure 4.17 Needle type test tip
In order to measure the effect of different surfaces on contact resistance it was important to fabricate the tip in such a way that any metal could be used on the connecting surface.

### 4.4.3 Fabrication of test tip

The fabrication process for the test tip is shown in Figure 4.18.

![Fabrication of test tip](image)

Figure 4.18 Fabrication of test tip

The process started with the thermal oxidation of a (100) silicon wafer (a). A 5 nm titanium adhesion layer followed by 200 nm copper layer was evaporated, S1813 photoresist was spun coated and patterned and the copper layer was etched using dilute nitric acid (b). A 55 \( \mu \)m thick layer of AZ9260 photoresist was then spun coated and patterned (c) with the sample then subjected to oxygen plasma to ensure a clean seed layer for electroplating. Any copper oxide that was formed by exposure to the plasma was removed by a short etch in dilute nitric acid, and nickel was electroplated into the mould to a thickness of 40 \( \mu \)m (d). The thick resist mould was then removed using acetone (e) and the wafer was immersed in dilute hydrofluoric acid to etch the exposed titanium layer (which was masked by the patterned copper and electroplated nickel layers). The exposed oxide layer was then removed by reactive ion etching in a CHF\(_3\)/Ar plasma (f). The silicon wafer was then etched through
using XeF₂ dry chemical etching (g). The nickel electroplated areas were then isolated by etching the remaining exposed copper selectively using a mixture of acetic acid and hydrogen peroxide followed by etching the titanium layer is dilute hydrofluoric acid (h). Finally a 15 nm chrome adhesion layer and 200 nm gold layer (or aluminium or copper layer) was evaporated at an angle coating the end of the test tip. A shadow mask was used to prevent the evaporated metals from shorting the nickel areas (i). The tip was then mounted onto a glass carrier for ease of handling and electrical connections made between the ohmmeter and the nickel electroplated areas using conductive silver paste and fine wire. SEM and optical images of a test tip before and after the angled evaporation are shown in Figure 4.19 and Figure 4.20.

![Figure 4.19 Fabricated test tip before angled evaporation](image1)

![Figure 4.20 Test tip after angled evaporation of a chrome gold coating](image2)

Nodular growths can be seen around the edge of the electroplated nickel. These are caused by the increased field strength that exists at the edge of the electroplating mould. Care had to be taken to ensure that none of these nodular growths caused shorts between the two areas of
nickel. If a nodular growth occurred at the tip, where the contact is made to the cantilevers, then the test tip was rejected. Nodular growths in other regions could be removed by grinding the surface of the metal using an abrasive. This was performed before the photoresist mould was removed as it was found to help protect the vertical surfaces from damage.

The quality of the contacting surfaces was determined by the roughness of the photoresist mould which itself was determined by the line roughness of the mask and any deformation of the mould that occurred during the electroplating step. Of these factors, deformation of the mould was found to be the biggest problem and had to be combated by the inclusion of additional plating areas in the mask. These areas altered the local electric field reducing nodular growth and helped mechanically enforce the resist by equally compressing critical resist walls from both sides.

SU-8 photoresist was investigated as a plating mould since it is mechanically and thermally much more stable than the AZ9260. However, it was found difficult to reliably remove it.

4.4.4 Differences between test tip and bond pads

A bond pad's mechanical response (to an applied load from a probe) is determined by the hardness and thickness of the bond pad metal and the hardness and thickness of any native oxide on the surface. The silicon substrate is unlikely to deform under the bond pad since silicon is much harder than the bond pad metal (typically sputtered aluminium or electroplated gold). Electrically, the contact resistance of the bond pad probe combination is determined by the resistivity of the metals, the area of contact and nature of any oxide layers and the area of contact. The area of contact is dependent on deformation of the surfaces and surface roughness. This is discussed in detail in chapter 6.

The electroplated nickel used in the test tip fabrication was a lot softer than single crystal silicon. Also the thickness of the evaporated metal used in the test tip was 200 nm, compared to several microns used in commercial bond pads. Therefore the electroplated nickel mechanical properties will dominate the deformation of the contact surface (for larger force ranges). It is possible at lower forces the area of contact could still be determined by the evaporated metal mechanical properties. The Vickers hardness of the electroplated nickel was measured at 476 kgf/mm² (measured using a MVK-H2 micro hardness tested manufactured by Mitutoyo with a 25 g indentation load). This compares to a Vickers
hardness of 80 kgf/mm² for electroplated gold [17] and 64.5 to 83 kgf/mm² for a 1 µm thick sputtered aluminium bond pad [18]. The Knoop hardness of silicon is 11270 N/mm² [19] which corresponds to an approximate Vickers hardness of 1493 kgf/mm².

The roughness of a typical bond pad was found to be comparable to the electroplated nickel. For instance a commercial electroplated gold bond pad had a measured roughness of 45 nm, a commercial sputtered Al bond pad of 15 nm whilst the roughness of the electroplated nickel was measured at 44 nm. Note that the electroplated nickel dominates over the roughness of the evaporated metal which is about 1.3 nm (for a 150 nm thick evaporated copper layer on 0.3 nm Rₐ silicon).

However due to the problems discussed in section 4.4.3 the flatness of the test tip surfaces was much worse than commercial bond pads. An AFM scan of a commercial aluminium bond pad (that has been probed) and SEM of the contacting surfaces of a test tip (before XeF₂ etch) are shown below in Figure 4.21 (a) and (b) respectively.

![Figure 4.21 (a) AFM scan of commercial Al bond pad after probing (2.5 µm vertical scale) (b) SEM of nickel plated test tip before XeF₂ release.](image)

It is clear that the testing using the described test tip represents only an approximation to the situation of probing bond pads. This limits the applicability of contact resistance data but should be sufficient to indicate the likely order of the contact resistance.
4.4.5 Ohmmeter specifications

A TTi 1906 computing multimeter was chosen to measure the circuit resistance. This instrument has 5½ digit resolution and when used as an ohmmeter can take 1 reading a second. The accuracy depends on the range and is shown in Table 4-1.

<table>
<thead>
<tr>
<th>Range</th>
<th>Accuracy</th>
<th>Maximum measuring current</th>
</tr>
</thead>
<tbody>
<tr>
<td>200 Ω</td>
<td>0.025% + 4 mΩ</td>
<td>1.5 mA</td>
</tr>
<tr>
<td>2 kΩ</td>
<td>0.019% + 30 mΩ</td>
<td>1.5 mA</td>
</tr>
<tr>
<td>20 kΩ</td>
<td>0.019% + 300 mΩ</td>
<td>400 μA</td>
</tr>
<tr>
<td>200 kΩ</td>
<td>0.019% + 3 Ω</td>
<td>40 μA</td>
</tr>
<tr>
<td>2 MΩ</td>
<td>0.022% + 30 Ω</td>
<td>4 μA</td>
</tr>
<tr>
<td>20 MΩ</td>
<td>0.07% + 300 Ω</td>
<td>400 nA</td>
</tr>
</tbody>
</table>

Table 4-1 Accuracy of ohmmeter

The maximum open circuit voltage was 3.5 V.

4.4.6 Calculation of contact resistance

To calculate the contact resistance, the resistance of the cantilevers and test tip needed to be known and subtracted from the total resistance. The resistance of the wires between the ohmmeter and test tip were eliminated by using a four terminal measurement method as shown in Figure 4.22.

Using this arrangement if the input impedance of the voltmeter, \( R_v \) is very high (so that practically no current flows through it) then the voltmeter accurately measures the potential...
drop across $R_x$. For the low resistances being measured ($R_{wire}$, $R_A$ and $R_x$ are all much smaller than $R_v$) then the input impedance of the voltmeter can be considered sufficiently high that the resistance of the wires do not affect the measurement.

The test tip had a complex shape making it difficult to calculate the resistance by hand. A model of the 40 μm thick electroplated nickel deposit was created in FEMLAB® 3.1 and a potential field simulated. Using the bulk value for the resistivity of nickel (6.84 x 10^{-8} Ωm at 20 °C) the calculated resistance of each electroplated path was 0.055 Ω. Therefore the test tip introduces a resistance load of 0.109 Ω (if the wires are connected to the very end of the nickel plated region). The error incurred by not connecting to the end of the nickel area is low since most of the resistance is attributed to the narrow contacting end of the tip. The first centimetre of the nickel track only accounts for 0.014 Ω.

Because the cantilevers were fabricated from rectangular sections it was simple to calculate a reasonable estimate of their resistance.

![Figure 4.23 Geometry of cantilever](image)

The resistance of the path can be approximated to:

$$R_{path} = \rho \left[ \frac{S}{w \times t} + 2 \times \left( \frac{l' + \sqrt{l^2 + h^2}}{w \times t} \right) \right]$$

$$R_{path} = \rho \left[ \frac{50 \times 10^{-6}}{(50 \times 10^{-6} \times t)} + 2 \times \left( \frac{325 \times 10^{-6} + \sqrt{l^2 + (60 \times 10^{-6})^2}}{w \times t} \right) \right] \quad (4-3)$$

where $S$ is the distance between the cantilevers, $w$ is the cantilever width, $t$ is the cantilever thickness, $l$ is the cantilever length, $h$ is the tip height of the cantilever, $l'$ is the horizontal length of the track in region B (see Figure 4.23), $w'$ is the width of the current path in region
A (perpendicular to the cantilevers) and \( \rho \) is the nickel resistivity. Table 4-2 shows the calculated resistance value of the total cantilever path for varying geometries.

<table>
<thead>
<tr>
<th>( w (\mu m) )</th>
<th>50</th>
<th>40</th>
<th>30</th>
<th>20</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( l (\mu m) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>700</td>
<td>0.281</td>
<td>0.351</td>
<td>0.469</td>
<td>0.703</td>
<td>1.406</td>
</tr>
<tr>
<td>600</td>
<td>0.254</td>
<td>0.317</td>
<td>0.423</td>
<td>0.635</td>
<td>1.269</td>
</tr>
<tr>
<td>500</td>
<td>0.227</td>
<td>0.283</td>
<td>0.378</td>
<td>0.567</td>
<td>1.134</td>
</tr>
<tr>
<td>400</td>
<td>0.200</td>
<td>0.249</td>
<td>0.333</td>
<td>0.499</td>
<td>0.998</td>
</tr>
<tr>
<td>300</td>
<td>0.173</td>
<td>0.216</td>
<td>0.288</td>
<td>0.432</td>
<td>0.863</td>
</tr>
<tr>
<td>200</td>
<td>0.146</td>
<td>0.183</td>
<td>0.243</td>
<td>0.365</td>
<td>0.730</td>
</tr>
<tr>
<td>100</td>
<td>0.121</td>
<td>0.151</td>
<td>0.201</td>
<td>0.302</td>
<td>0.604</td>
</tr>
</tbody>
</table>

Table 4-2 Resistances of cantilever path in \( \Omega \)

References


