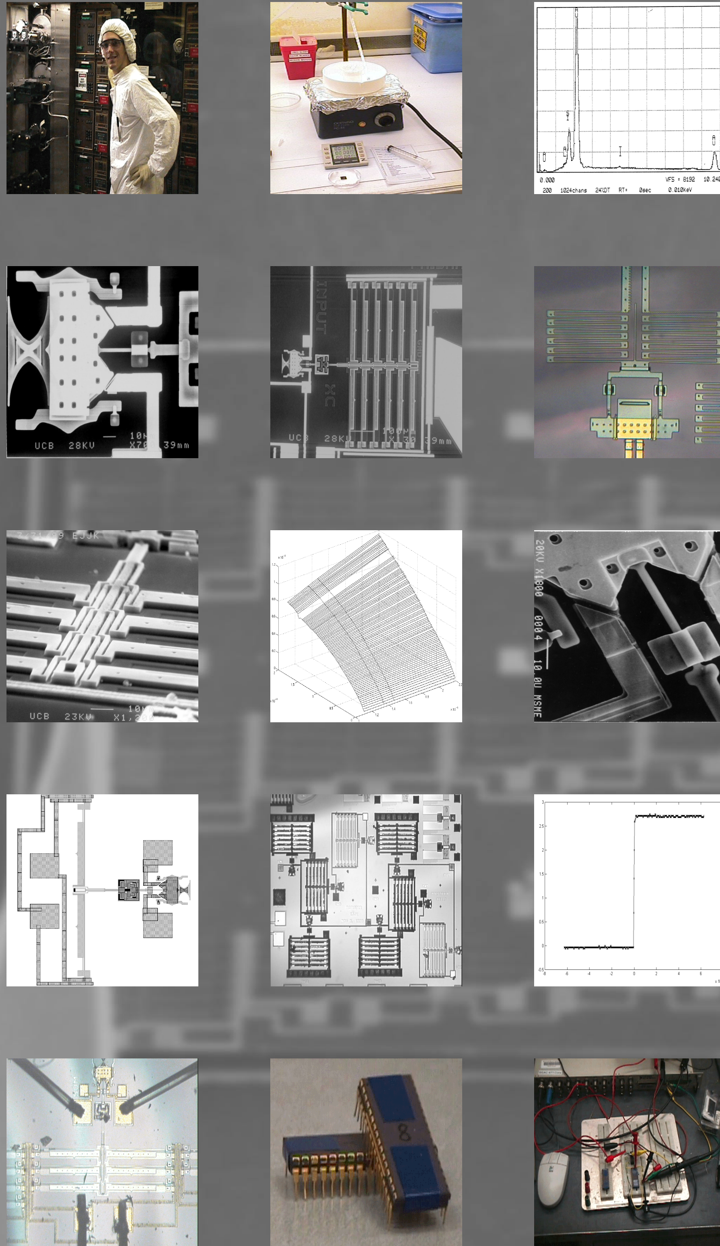


Berkeley Senson and Actuator Center
497 Cory Hall
Berkeley, CA 94720-1770, USA

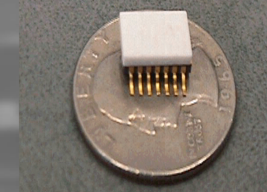
http://www-bsac.eecs.berkeley.edu

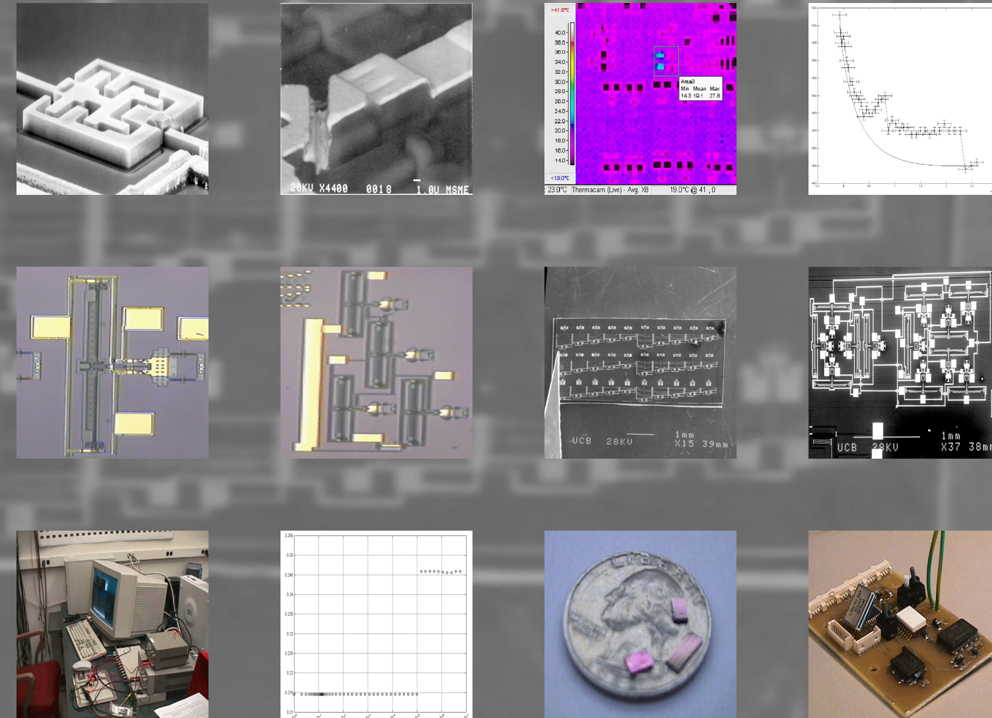
Microrelay design, performance, and systems

*Ezekiel J. J. Kruglick*

# Microrelay design, performance, and systems

**Ezekiel J.J. Kruglick**





**Berkeley Sensor and Actuator Center**

**Advisor: Kristofer S.J. Pister**

# Microrelay design, performance, and systems

by

Ezekiel John Joseph Kruglick

B.S. (University of California, Los Angeles) 1995
M.S. (University of California, Los Angeles) 1996

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy
in

Engineering - Electrical Engineering
and Computer Science

in the

GRADUATE DIVISION

of the

UNIVERSITY of CALIFORNIA, BERKELEY

Committee in charge:

Professor Kristofer S. J. Pister, Chair
Professor Roger Howe
Professor Richard White
Professor Dorian Liepmann

Spring 1999

# *Abstract*

Microrelay design, performance, and systems

by

Ezekiel John Joseph Kruglick

Doctor of Philosophy in Electrical Engineering and Computer Science

University of California, Berkeley

Professor Kristofer S. J. Pister, Chair

Surface micromachined four terminal relays have been fabricated, tested, and packaged. The relays are implemented in a foundry process with minimal post-processing. They show on resistances as low as 20mΩ, current carrying capacities as high as 0.48 Amps, and lifetimes in the millions of cycles while hot switching 80mA. These devices are thus competitive with commercial macroscopic relays.

Achieving these high performance and reliability levels required exploring the physics and tribology of microrelays and using novel physical experiments to outline a design philosophy for microrelays. This philosophy with standard macrorelay design methods by recommending small gaps, light mechanical elements, and low force levels. These differences from the standard relays design philosophy, how-

ever, are supported by theory and experiment and stem from the basic physical differences between designing at the macro scale and the micro scale.

Finally, system design is explored and computation using microrelays is demonstrated as a new application made possible by the tiny scale of these relays.

Professor Kristofer S. J. Pister

Committee Chair

# Dedication

This dissertation is dedicated to my two grandfathers, from whom my middle names come: John Kruglick and Joseph Lorenz. Though I did not finish graduate school quite fast enough to share graduation with them their impact has been vast.

I will always remember Joseph Lorenz as a strongman. During the depression he found work as a circus strongman, lifting barrels of concrete over his head and bursting iron bands riveted around his biceps. He went on to apply his iron will to machining, becoming a legendary tool and die maker and craftsman with parts in every 747 made during his lifetime and parts on every moon landing. Joseph Lorenz was well known for making numerous parts that conventional wisdom said could not be made. When he announced his impending retirement at the age of 79 with a total of 67 years of machine shop experience under his belt Joseph Lorenz

was suddenly deluged with two years worth of orders from desperate suppliers. He continued to work his usual 7-days-a-week 18-hours-a-day schedule until they were done. He was killed by an unknown driver in a hit and run accident shortly after his retirement August 3, 1992.

During my entire lifetime Joseph Lorenz lived in his machine shop (a Quonset hut) inside a single home-built freestanding room the exact size of two sheets of plywood. Whenever I would visit him I would spend my time poking around his machine shop and watching him work with enormous hydraulic equipment. His stoic willpower in the face of the inevitable injuries stemming from decades of machine shop work and the travails of living modestly in a quonset hut in Los Angeles were always a source of inspiration to me. No matter how hard things have gotten, I could always picture my grandpa Joe smiling at the situation and laughing. It's helped. I also can't help but suspect that my predilection for machining and engineering come from growing up with a grandfather like Joe.

John Kruglick was the youngest of seven Jewish siblings who came to the U.S. fleeing pogroms in the Ukraine. As a youth from the wrong side of the tracks John learned English and managed to get through medical school supported by one of the elder siblings, Lewis; a grocer whom my own father was named after. John Kruglick allowed nothing to stand in his way and after serving in world war two he started a medical practice, raised a number of children, and even campaigned for

the Senate. John Kruglick impressed upon his children the opportunity and honor represented in a good education, leading to a family tradition of high educational achievement. It is to this side of the family that I credit my proclivities toward advanced degrees and I regret deeply that John will not be able to watch his grandson receives a Ph.D. as he watched his son get an MD. Unfortunately he was claimed by age in October of 1998.

Do not stand at my grave and weep,
I am not there, I do not sleep.

I am a thousand winds that blow,
I am the diamond glints on snow.

I am the sunlight on ripened grain,
I am the gentle autumn rain.

When you awaken in morning's hush,
I am the swifter uplifting rush

Of quiet birds in circled flight.
I am the soft star that shines at night.

Do not stand at my grave and cry,
I am not there, I did not die.

*From the celebration of life of John Sanford Kruglick.*

# Acknowledgments

I face the usual graduate student problem of being indebted to almost everyone. Every fellow student who ever sat in on one of the late night bull sessions that occur when everyone has been working far too long, every student and staff member who taught me how to use a piece of equipment while listening to inane questions, and every faculty member who patiently explained things deserves a place of acknowledgment.

I would like to express my deep thanks to the member of the research group, both back in the UCLA days and the newer generation at UCB. In particular Richard Yeh stands out as a beacon of sanity and reliability whose exemplary honesty and humility has been reassuring at times when unreasonable people cause chaos. Gisela Lin, at UCLA, deserves special thanks for showing us all just how fun a

research group can be. Jack Judy and Michael Helmbrecht get special mention for helping me survive the highly unusual transition from UCLA to UCB when my advisor moved. James McClurkin also gets a nod for sharing some very cool technical knowledge on wide ranging topics from surface mount soldering to the stability of motorcycles and for knowing how to ask questions that made me rethink my understandings of several issues.

Around the lab a couple people stand out as outstanding staff members whose help regularly went above and beyond the ordinary. I would like to thank Ron Wilson for his excellent SEM work and Marilyn Kushner for her help in the Berkeley microlab.

I would also like to acknowledge the Hertz Fellowship for their funding during my graduate career and for their support network. It's a great fellowship to be a part of and I highly recommend it.

To everyone else (you know who you are) I issue thanks.

## Contents

*Abstract* ***1***

*Dedication* ***iii***

*Acknowledgments* ***vi***

**CHAPTER 1** *Introduction and Motivation* ***1***

1 Introduction **1**
2 Current commercial relays **2**
3 MEMS relays **6**
4 Direct comparison **8**
5 References **10**

**CHAPTER 2** *Contact Design Issues* ***11***

6 Introduction **11**
7 The relay cycle **12**
7.1 Relay sitting open **13**
7.1.1 Open relay wear 13
*7.1.1.1 Chemical diffusion 14*
*7.1.1.2 Defect diffusion 16*
*7.1.1.3 Creep 18*
*7.1.1.4 Dealloying 19*
7.1.2 Corrosion 19
7.1.3 Holding voltage 19
7.2 Relay closing, contact being made **25**
7.2.1 Closing arc 25
7.2.2 Impact energy and contact bounce 28
7.2.3 Surface deformation and contact area 29
*7.2.3.1 Basics of surface contact 29*
*7.2.3.2 Three dimensional origins of hard-*

*ness 34*
7.2.4 Angle of contact 36
7.3 Relay closed, current being conducted **39**
7.3.1 Constriction resistance 39
7.3.2 Supertemperature 44
7.3.3 Temperature induced softening 46
*7.3.3.1 Thermal aging 49*
7.3.4 Welding 49
7.3.5 Fretting 50
7.4 Relay opening **53**
7.4.1 Contact adhesion and separation 53
7.4.2 Particle generation 54
7.4.3 Opening arc 55
8 Considerations for microcontact design **58**
8.1 Materials for contacts **58**
8.2 Force and mass considerations for contact **59**
8.3 Eliminating arcs **60**
9 Comparison of macro- and micro-relay design paradigms **61**
10 Conclusion and caveat **64**
11 References **65**

CHAPTER 3 *Fabrication* ***67***

12 Process selection **67**
13 The MCNC MUMPs process **69**
13.1 The basic fabrication process **69**
14 Electroless gold deposition **72**
14.1 Basic types of electroless plating **73**
14.2 Chemical meta-stability **74**
14.3 Catalytic meta-stable gold deposition **75**
15 Insulating mechanical bridges **84**
16 Dicing and sacrificial release **86**
17 The finished product **86**
18 Packaging **87**
19 Yield and process stability analysis **90**

19.1 Step 1, commercial foundry processing **90**
19.1.1 Gold bridging 90
19.1.2 Misalignment 91
19.1.3 Destructive overetch 93
19.2 Step 2, electroless gold deposition **93**
19.2.1 Unwanted gold deposition 93
19.2.2 Single structure massive overplating 95
19.3 Step 3, photoresist processing **96**
19.3.1 Generation of "goop" 96
19.3.2 Poor lithography 96
19.4 Back end - Bonding and packaging **97**
19.4.1 Bond pulling 97
19.5 Overall yield and yield improvement **98**
20 Conclusion **100**
21 References **100**

CHAPTER 4 *Actuator Design* ***103***

22 The importance of actuator performance **103**
23 Choosing an actuation method **105**
23.1 Electrostatic actuation, basic considerations **105**
23.2 Thermal actuation **106**
23.2.1 Basic considerations 106
23.2.2 Origins of thermal actuation 107
23.2.3 Design considerations of thermal expansion 111
24 Thermal actuator design and characterization **112**
24.1 Thermal actuator design **112**
24.2 Characterization for implementation **115**
25 Reliability and optimization of thermal bimorph actuators **124**
26 Conclusions **126**
27 References **126**

CHAPTER 5 *Physical Contact Characterization* ***129***

28 The need for contact characterization **129**

29 Characterization of plastic deformation point **130**

29.1 Statement of the problem **130**

29.2 Plastic transition characterization using force/current/conductance curves **131**

29.3 Usage of plastic transformation point **137**

30 Number of contact points formed **138**

30.1 Statement of the problem **138**

30.2 Separation mechanism for relays **139**

30.3 Relation between plastic deformation force and surface energy forces **140**

30.4 Using the contact spot population results **148**

31 Conclusion **148**

32 References **149**

CHAPTER 6 *Mechanical Design and Layout* ***151***

33 Considerations for mechanical design **151**

34 Actuator hookup **152**

35 Assuring contact and resisting mechanical lockup **155**

36 Making lateral contacts in MUMPs **157**

37 Conclusions **160**

38 References **160**

CHAPTER 7 *Electrical Testing and Characterization* ***161***

39 Introduction **161**

40 Resistance measurement **161**
40.1 Extending experimental range **162**
40.2 Resolution of basic measurements **164**
40.3 Averaging **166**
40.4 Slope calculation **167**
40.5 Typical relay contact resistances **172**
41 Maximum static current handling **173**
42 Transient characterization **174**
42.1 Methods and equipment **174**
42.2 Looking for bounces and arcs **176**
42.3 Transient performance characterization **178**
43 Cyclic testing **180**
43.1 Test setup **180**
43.2 Long term test results **181**
44 Conclusions **183**
45 References **184**

CHAPTER 8 *Relay Systems using Microrelays* ***185***

46 Introduction **185**
47 Simulation and the lack of CAD tools **185**
48 Digital logic **187**
49 CMOS style complementary logic **189**
49.1 Experimental logic performance **191**
49.2 CMOS style complementary logic analysis and limits **193**
50 ZS logic **200**
50.1 The ZS logic beastiary **200**
50.2 Some fabricated systems **201**
51 Conclusion **203**
52 References **204**

*Complete References* ***205***

FIGURE 1. Some standard “1-Form-A” commercially available relays. Each relay is contained in a 19.6 x 16.5 x 16.5 mm non-hermetic plastic box. Minimum cost is about US$1 (in Q4 1998). 3

FIGURE 2. September 1998 advertisement for relays, emphasizing that the serial actions of coil winding, assembly, and soldering have recently been automated. The tasks are still done one relay at a time. Wafer scale manufacturing of microrelays could entirely change the economics of relay manufacturing. 3

FIGURE 3. The microrelay described in this work. The thermal actuator array on the right pulls a gold crossbar across two gold electrodes, closing a circuit. The function of the relay is basically to close an electric circuit by closing a gap with a gold crossbar. 6

FIGURE 4. A closeup of the non-actuator part of the microrelay developed here. The crossbar and electrodes can be seen more clearly. The isolation structure allows for complete electrical isolation of the actuation from the circuit, making the device a true four terminal relay. 7

FIGURE 5. Dozens of microrelays fit easily into this surface mount package. This package has four microrelays bonded to the 14 pins (two drive pins are shared). Drive power plus dissipated power is well within package limits. 7

FIGURE 6. Historical relay performance chart (from [4]), with indication of microrelay performance. The bands show historical performance regions for electromechanical relays from the 1920s to the early 1990s. The relays presented here are shown along the right edge (time axis value 1999). Size (including demonstrated packaging) is lower than previous solutions by almost an order of magnitude, as indicated by data point at lower right. The industry efficiency metric of these relays is higher by 5 orders of magnitude, putting it well off even this logarithmic chart (value shown at top right). 9

FIGURE 7. Steps of the relay cycle. The operation of any relay can be broken down into these basic four steps. 13

FIGURE 8. Slip lines and grain boundaries form “highways” for defects, which increases slip line motion and results in defect clustering. Diagrams (a), (b), and (c) show some

common line formations and the path taken by defects. Diagram (d) shows a typical line of surface features caused by a slip line intersecting the surface. Diagram (e) shows a row of divots as marks the passing of a slip line below, much like a moving volcano might leave a chain of islands. (From [9]) 16

FIGURE 9. Dependence of defect diffusion on temperature, showing substantial difference in mobility between the volume, grain boundaries, and the surface. This plot is for steel. (From [10]) 18

FIGURE 10. Paschen curves for a number of different gasses. Plot shows breakdown voltage versus the pressure-distance product. The right side, which increases with increasing pd, represents the conventional cascade mechanisms discussed above. The left side, which decreases with increasing pd, represents pd products too low to initiate cascade and stems from other mechanisms. Labelled point on air curve shows that the microrelays presented here are very close to the inflection point for air. Closeness to the inflection point reduces effects of topology or pressure changes. (Background data from [11]) 21

FIGURE 11. A closeup of the left hand side of the Paschen curves from Figure 11 for various gasses. Molecular structure can cause odd shapes in the critical field behavior (such as in He). Choice of gas can have a significant effect on breakdown voltage. (Paschen curve from [11]) 22

FIGURE 12. Low-current arcing damage on a microrelay showing that the arc did not strike at the narrowest gap point. The microrelay dimensions are those represented by the dashed lines on the Paschen curve charts. This arc formed initially under opening with an inductive load and presumably transitioned from Townsend arcing to cascade arcing due to high inductive voltages. 23

FIGURE 13. Breakdown voltages for differently prepared samples at very small gaps and atmospheric pressure. As usual when dealing with sizes in the range of surface effects, behavior becomes very complex and dependent on surface quality. The plot here shows platinum contact arcing data at low spacing for electrodes which were

polished and alcohol cleaned (1), activated with discharges in organic vapors (2), and cleaned by electron emission (3). In this region the theoretically expected breakdown voltage should be 105 to 107 volts. (From [11]) 24

FIGURE 14. Paschen curves explored by varying p then d individually. The solid line shows an experiment done by varying p while keeping d macroscopic, the dotted line shows data collected at microscopic distances for a the same material system.(From [11]) 24

FIGURE 15. Paschen curve showing movement of operating points in macro and micro relays as the relay closes. Macro-relays move into a lower-voltage breakdown region as they close, often starting an arc. If a micro-relay has not arced already, it is unlikely to arc during closing. (Paschen curve redrawn from [11]) 26

FIGURE 16. Schematic representation of the expected voltage behavior in an arcing relay. The so called "arc voltage" develops due to space charge in the arcing region and remains fairly constant over the evolution of the arc (typically 2-20 microseconds). This behavior is very useful because it gives us a characteristic to look for to diagnose arcing. 27

FIGURE 17. Surfaces which appear macroscopically flat nonetheless make initial contact at surface micro-asperities which then begin to deform. The deformation goes through several behavioral zones including elastic, elasto-plastic, and plastic deformation. 29

FIGURE 18. Behavior in the elasto-plastic region where nonrecovered deformation in the center is mixed with deformation around the edges that recovers when force is reduced. Capital R is the original surface radius when compressive force was applied, lowercase r represents the surface as it is after the force is removed. Some area remains in contact due to plastic deformation of the center. (From [14]) 32

FIGURE 19. Hardness of polycrystalline electropolished nickel versus indentation depth. As the indentation depth approaches the crystal size the hardness increases from

the empirical value (approximately 1 GPa) towards the theoretical value of 4.5GPa. (From [17]) 34

FIGURE 20. Behavior of a contact asperity as both perpendicular and tangential force increase. Point Q represents the geometric center of the asperity, it moves as the contact asperity deforms out into the shape represented by the dashed lines. The tangential force increases the area of contact by speeding the onset of plastic deformation. (From [9]) 37

FIGURE 21. Simplified contact system for calculation of constriction resistance. Contact electrodes are infinite half spaces, the contact area has radius of a. 40

FIGURE 22. Resistance plotted versus contact voltage (to follow relay characterization convention established by [13]). Dashed line is theoretical calculation of resistance change due to temperature, solid line is data. 45

FIGURE 23. Thermal microscopy used to verify temperature performance of an active relay. The SEM at left shows a closeup of the contact area of interest as indicated by the small box at right. In this case the current is 30mA and the temperature increase is about 12C, analytically predicted value is 12.2C using dissipated power and fitted Gt from (EQ 17). 45

FIGURE 24. Resistance of a relay upon first application and second application of current while under constant force load. As predicted, the resistance drops off as current increases on the first application as the relay softens and the area of contact increases. This relay is in plastic deformation, so the increased contact area will remain until separated, thus the second current sweep experiences a lower resistance and returns up the heating curve to the same stable point. The dashed lines represent theoretical resistance increase with temperature as predicted by application of (EQ 17) and (EQ 18). Data is presented in terms of contact voltage to conform to convention [13]. 47

FIGURE 25. Plots of relative contact pressure and absolute contact area versus supertemperature (temperature above ambient) for the data of Figure 24. As expected, the contact pressure (which will be equal to the yield

pressure) decreases as temperature increases on the first current application and area increases. On the second current sweep the area does not increase because the effective contact pressure remains below the yield pressure and it does not recover (decrease) because the deformation is all plastic. 47

FIGURE 26. Schematic representation of the process of fretting with corrodible material. As oxide layers or other corrosion products form on the surface they are broken off by the fretting process and become abrasive particles to further enhance the process. Eventually the relay gap can become packed with such particles. Fretting is a main cause of relay failure and the single most prevalent cause of a closed relay suddenly becoming intermittent under vibration. 51

FIGURE 27. As contacts separate fine points (on the order of a nanometer) can be generated which will initiate Townsend arcing. Electrons emitted from the cathode sublimated ions from tiny spots on the anode and a countercurrent is initiated, even in vacuum. If conditions are right this can lead to cascade breakdown by providing the initial arc (which needs far less voltage to sustain than to form). 55

FIGURE 28. Material transfer direction and magnitude in conventional arcs involves at least six conflicting mechanisms which lead to extremely complex wear behaviors. This typical plot of wear direction and magnitude for a particular macro-relay system [21] shows how difficult to characterize this process can be. Microrelays which avoid most arcing face much simpler wear issues. 56

FIGURE 29. Design flow chart for macrorelays. Starting with high force actuators, large gaps, and large masses leads to arcing, bouncing, and fretting. All of these issues reinforce the tendency to use bigger and stronger components and harder and more exotic contact materials. 61

FIGURE 30. The microrelay design flow. Unlike macro-relay design we begin with the assumption of small gaps and masses with limited available force. The only arcing is

Townsend arcing and the physical wear is minimal (though the available wear volume is also minimal, so this cannot be ignored). The most important factor becomes the make and break of good electrical contact. 62
FIGURE 31. Cross sectional representative drawings of a micromotor fabricated by the MUMPs process before and after release by sacrificial silicon dioxide etching in hydrofluoric acid. Step shapes are representative, but approximate. (from [25]) 71
FIGURE 32. Energy diagram for a meta-stable solution. Point A is locally stable, point B is globally stable and has less potential energy than A but is isolated by a potential peak of height U. Unless a species has at least U energy available it will be unable to go from A to B. Uc is the potential peak in the presence of a catalyst. To make the solution self patterning enough heat is provided to put the solution between Uc and U and a catalyst is provided that only adsorbs and works on certain target materials. 74
FIGURE 33. The setup for doing meta-stable catalytic gold plating requires a hot plate, water bath, beaker, and thermometer. A syringe facilitates rehydration and solution storage. The basic recipe used at 73C is shown at right, see the text for more details. 76
FIGURE 34. Scanning electron micrographs of die edges after meta-stable catalytic gold deposition. The one on the top has a substantial gold layer (all of the light topography) due to air-carried organic contaminants, the one on the bottom shows little gold (only the vertical stripe) despite longer deposition processing after being oxygen plasma descummed. While not strictly necessary this sort of cleaning produces chips with less collateral deposition. 79
FIGURE 35. Scanning electron micrograph (left) and electron backscatter image (right) of a contact gap subjected to electroless gold addition. The SEM photo shows texture while the backscatter image has high contrast between materials and shows good coverage. This particular gap

was left with gold bridging it by MCNCs fabrication process. 80

FIGURE 36. EDS spectrum of electroless gold on CMOS 0.25mm process metallization to check coverage. The CMOS metallization (which was spectroscopically examined separately) contained small amounts of titanium and silicon. Both of these diffuse into gold and small peaks of Ti and Si can be seen amongst large peaks of Au. The absence of Al peaks shows that coverage was 100% for this sample, which was examined within a day of fabrication before the gold and aluminum could interact. 80

FIGURE 37. Surface texture of electroless gold after solution has been depleted of approximately 30% of its gold supply. Scale bar shows 1mm. 82

FIGURE 38. Surface texture of electroless gold grown from a solution that was approximately 60% depleted. The roughness here is too unpredictable for reliable relay usage. Photo is same magnification as Figure 37. 83

FIGURE 39. Three dimensional gold “crystals” grown on an aluminum pad using a highly depleted electroless solution. Scale bar is 10mm. 83

FIGURE 40. Microrelay structure (unreleased) with a discontinuity in the polysilicon support in the upper right. This convoluted structure provides plentiful adhesion area for photoresist bridging and has high yield. 84

FIGURE 41. Perspective view of the photoresist bridging structure, showing the ample surface area provided. 85

FIGURE 42. A cross section of the final processed relay, as seen looking into a cleave along the crossbar contact surface. 87

FIGURE 43. A die attached and fully bonded relay. 88

FIGURE 44. Two unsealed packages made by closing identical open cavity ceramic packages together to leave a chamber in the middle. There is no sealing but the method is low temperature and provides effective protection from incidental handling. 88

FIGURE 45. Photos of the final packaging solution. The left photo shows the elements and a finished package after some

time at 260C. The photo on the right shows a surface mount packaged relay die nestling among other surface mount parts on a printed circuit board. The package contains four working wire bonded microrelays, the entire circuit board is not much bigger than a quarter. Bottom photo compares a 4-microrelay package to 2 commercial single relays, the commercial relays have similar performance except that they can conduct more current (6-10 amps). 89

FIGURE 46. Gold bridging across a 6mm gap. 91

FIGURE 47. A SEM showing a typical misalignment. The layout calls for a square 2mm hole in the second polysilicon in the center of a 4mm square hole in the gold. 92

FIGURE 48. Nonideal electroless gold deposition caused by local energy concentrating factors. The same structure is shown before (left) and after (right) electroless gold deposition. The SEMs show that the gold deposited not only on the metal that was present (most clearly visible on the crossbar), but also along the edges of the actuation rod guide and the sliding guide visible near the very bottom of the SEMs where RIE cut edges of polysilicon make sharp angles. 94

FIGURE 49. A pair of relays practically buried in gold. Most of the chip had only a micron of gold deposited, thickness on this structure appears to be approximately 3mm. The reason for this sort of irregularity is not understood. 95

FIGURE 50. Poor wirebond adhesion was pandemic on many chips. Here 8 wirebonds in a row pulled off of a test area made of the layers recommended for bond pads. The wires themselves remain attached to the gold, it is the gold or other layers that delaminate. 97

FIGURE 51. Closeup of the location of a former bond pad. The bond pad itself has separated from the die, taking two layers of polysilicon and the nitride underneath with it. Given that the nitride has peeled away in a rather large patch, this would seem to qualify as a layer adhesion problem. 98

FIGURE 52. A logic latch cell where 2 out of 8 relays survived. This sort of yield makes building large systems problematic. The four relays oriented horizontally are locked up by

misalignment, the remaining two broken relays were overplated. Nonfunctional relays are blackened by current overloading while conducting probe station tests to determine yield and cause of failure. 99

FIGURE 53. Energy versus atomic separation for atoms in a solid and some of the related values. The minimum energy distance is the separation the atoms will have at 0K, and at this location they have negative potential energy related to free atoms by the binding energy. Higher binding energy is equivalent to higher melting and boiling points, 107

FIGURE 54. Atomic separation diagram with energy levels of different temperatures. The quantum mechanical position of the atom is spread over the resulting energy range and the effective atomic separation is the average position. Because of the asymmetry of the energy diagram the average position moves outward with increasing temperature. This particular diagram shows a very high temperature change for clarity. 110

FIGURE 55. Comparison of unconstrained, fully constrained, and typical realization of a thermally expanding beam actuator. The first gives no force but maximum motion, the second gives no motion but maximum force, the third gives both in lesser amounts but does work. 114

FIGURE 56. Some example thermal bimorph geometries, all with equal force amplification in the linear actuation regime for equal R/r. Two geometries change direction of motion by 90 degrees, one by 180. 115

FIGURE 57. Linear model of actuator and load for characterizing thermal bimorph structures. Variables used for the various springs and the generated force are shown. 116

FIGURE 58. A graphical representation of the thermal actuator characterization problem and the solution used here. The solid line represents the performance of a thermal actuator at a given power, with the completely constrained force Fg at one extreme and the unconstrained (zero force) motion at the other. Different measurement springs intersect the performance at different points, limiting their characterization value. If

multiple springs are used at the same power, however, the internal spring constant ki can be found and the system performance can be calculated for any load. 117

FIGURE 59. Thermal actuator test structures with motion indicators and gauges. Calibrated SEM measurement showed that the drawn 2mm structures were 1.70mm wide in the open, growing to 1.84mm as they approached anchors or other structures. 118

FIGURE 60. Raw data from using the same test structure against 2 different external springs. Spring indicated force versus drive voltage is shown for external springs of 0.45 and 2.1 N/m. 119

FIGURE 61. Fg versus drive voltage, calculated from the same data as Figure 60 using correction for internal spring Ki. Note overlapping error ranges. It can be seen that once the internal spring constant is known the internally generated force behaves predictably. Conversely this actuator can now be used with knowledge of Fg and Ki, meaning that the force and displacement for any load can be calculated. 120

FIGURE 62. Second order model of Fg as a function of input voltage. Values are computed for each voltage at which real data was taken. 122

FIGURE 63. A pulling array that uses spring linked thermal actuator pairs. Each additional pair increases the force available but also increases Ki, as long as Ki is significantly greater than Kl the changes in Ki will not be extractable, though the generated force will remain predictable. 122

FIGURE 64. Deflection plots for sets of thermal actuator arrays of varying numbers pushing against a weak Kl. Since Ki increases along with increases in Fg as actuators are added deflection/voltage performance remains approximately the same. 123

FIGURE 65. Perspective view of the actuator linking structure with the “loop” at the end for use at the probe station to free stuck structures. 125

FIGURE 66. Power optimized single layer thermal bimorph actuator from [41], only the section that needs to expand carries

current, reducing power usage, and the leverage system is more efficient for conversion to the desired direction. 125

FIGURE 67. State of yield in elastic, elasto-plastic, and plastic deformation regimes. The most important factor for electrical contact quality is that the surface should have plastic flow, which requires full plastic deformation. 131

FIGURE 68. Force measurement vernier spring from [42] with an early contact crossbar design on it. This contact is flat and wound up providing polysilicon to polysilicon contact. 132

FIGURE 69. Raw data for force, current, and conductance shown with fitted constant-current lines for evaluation. Contact was polysilicon to polysilicon. The gap in the data is from a set of 4 bad data points. 133

FIGURE 70. A single equicurrent line from Figure 69, showing the poor quality of fit for values of b only 10% apart. Thus the fits give good values of b and allow observation of the regime of operation. 134

FIGURE 71. Plot of the fitted values of exponent b. Uncertainties for the value of b are 0.015. This shows strong experimental confirmation of some expected behavior. 135

FIGURE 72. Transient plot of relay opening under a test voltage load. Several zones of behavior are seen. The final exponential voltage decrease may be structural necking or it may be a metal vapor arc, the initial straight regimes would be expected from mechanical necking. 140

FIGURE 73. Necking occurs as the contact point deforms plastically before separation. This increases surface area and thus requires additional energy input. The cross sectional drawing at left represents the appearance of a contact asperity which has deformed to form a contact point, the cross section at right shows a representative view of what the system might look like during separation as the junction necks under an applied separation force. 141

FIGURE 74. Relative magnitude of surface energy force to total separation force as a function of contact resistance in gold. This assumes only a single contact point. It can be

seen that within the commonly observed range of microrelay contact resistance the surface energy term becomes dominant. This also emphasizes that the surface energy term cannot be neglected. 143
FIGURE 75. Plot showing the percentage increase in contact separation force needed as number of contact spots (a-spots) rises above one. All calculations are relative to a single a-spot contact and each solid line represents a resistance from 20mW to 2W, scaled logarithmically. 144
FIGURE 76. Direct observation of microasperities being encountered by a closing contact and transitioning through various ranges of behavior. As each is first encountered it is in elastic deformation and provides less conduction, when it transitions to plastic deformation a drop in resistance is seen. Experimental current is 1mA. 146
FIGURE 77. Mechanical overview of a relay with actuator. The actuators pull the crossbar assembly into the stationary electrodes. The manual pull ring is for freeing the device if it gets stuck down during release. 153
FIGURE 78. General purpose digital logic, this is not required to drive much of a load and has minimum actuation. 154
FIGURE 79. Digital logic output buffer, this is designed to carry a significant current load and so has a large drive array with very high internal spring constant to assure separation. 154
FIGURE 80. Layout and device of the best crossbar/electrode design. The curves and points are designed to reduce stiction problems, the guidepieces are required to maintain planarity and placed to maximize allowable rotation. 155
FIGURE 81. Two types of lateral contact topologies to attempt gold-gold contact in a polysilicon based MEMS process. The top line shows a schematic cross section of the designs, with all major material layers labelled. The second line shows a cross section of how each topology is designed to look and function after chemical removal of the oxide layers. At the bottom is a series of SEMS showing the face of the crossbar and electrodes from each design. The left pair of photos are from a contact perpendicular

to motion, the right pair are from an angled contact. 157

FIGURE 82. Views from above of different relay designs. Note that each relay represents two contact points. Two contact points are required for a crossbar to short out the electrodes. The areas in the ovals correspond to the areas shown in the SEMs of Figure 81. 158

FIGURE 83. Extender for high current, high resolution tests of microrelays. The OPA516 was configured to drive up to 5 Amps, the instrumentation amplifier was configurable for gains of 1, 10, or 100. Actual gains and offsets were characterized. 163

FIGURE 84. Plot of data from a relay showing the effects of low voltage quantization. The upper plot shows the raw current and voltage output with error bars representing only the quantization error. The bottom plot shows the effect of division by plotting the instantaneous resistance at each current as calculated using equation 38. It can be seen that the error bars are very large at low currents and decrease as current increases. This problem is more pronounced at lower resistances where this absolute systematic error is a large multiple of the value to be measured. The plots use actual relay data. 164

FIGURE 85. The effects of a moving average on the quantization uncertainty of data. Larger, lighter error bars represent data taken the conventional way, the circular data points are the results of averaging and the smaller error bars represent the best possible case for this method. The moving average makes points by averaging a sliding window of 8 points, reducing the error by up to a factor of 3. This method obscures some fine detail and will conceal sudden changes in behavior. Additionally it has reducing returns as the window grows. 167

FIGURE 86. A dataset collected using the current extender, amplification, and slope analysis. The wider current range (above 100mA) can be seen. The raw data has a voltage resolution of 0.1mV, which makes for error bars almost on top of each other. This particular run was done in four-point mode to measure the contact resistance with no relay - the resistance is due to a test

set of fixed on-chip traces. The resistance is 0.7497W with a sigma of 9.74mW, calculated using the least squares slope method. 169

FIGURE 87. Calculations of the resistance from figure 86 using the conventional and the slope method with a sliding window of 8 samples, with the best derived value from Figure 86 shown as a dotted line. Plots are in ohms, but to different scales. Error bars could not be calculated for the conventional method due to offset errors in the amplifiers. Remaining error in bottom plot is a mixture of true measurement error and slight heating. Fit to heating is also drawn in and can be seen to correlate strongly to the residual error; another strength of the least squares fit method is that the residuals can be analyzed to detect less significant trends. 169

FIGURE 88. Residuals from the data of Figure 86. Here each point is from the sliding slope fit of figure 87 and the resistance and thermal effects have been subtracted out. The origin of the oscillatory nature of these residuals is unknown. This noise leads to a resolution floor of 5.1mW for this sliding window size (8 V-I points per R point). 171

FIGURE 89. Slope fitting error and overall error values for different sizes of slope fitting windows. The range of values for the errors due to slope fitting show that the added error (which used to be dominant), is now 2% or less of overall experimental error which includes sources such as quantization and amplifier noise. 171

FIGURE 90. Two similar early generation relays, one without damage (left) and one after being burnt out with a heavy static current load (right). Designs of the contact areas are the same, although these devices did have a few minor layout differences. The most recent designs have a bimaterial auto-break function which prevents such burnouts. 174

FIGURE 91. Typical transient testing setup. The bias resistor was usually less than 300W. 175

FIGURE 92. Microsecond scale capture of relay closing across 2.75 Volts (resolution is 25ns). No arc develops and the relay takes about 0.1mS to close fully. There is no bouncing. 176

FIGURE 93. Relay closing a 2.25 Volt source, plot time resolution is 2.5ns, bias resistor is 750W. Timescales and behaviors are wrong for our understanding of arcing, it is possible that the change in slope represents an elastic-plastic deformation transition as discussed in section 29 occurring dynamically. Final resistance is below 3W. 177
FIGURE 94. Transient plot of relay opening under 0.9 Volt load. Several zones of behavior are seen. the Initial straight regimes would be expected from mechanical deformation and necking, the final exponential voltage decrease may be structural necking or it may be a metal vapor arc. 177
FIGURE 95. Plot of relay closing response. Delay from drive voltage application to transition is 12ms. The turn-on transition took 105ns. This test was done with an angled contact type II relay with 6 pairs of actuators. 179
FIGURE 96. Relay opening response, same type of relay as in Figure 95. Delay from deactivation of actuators to transition is 520ms. Turn-off transition took 200ns. 179
FIGURE 97. Short-pulse relay activation. By narrowing the drive signal to a short pulse the relay can be made to just barely close for times on the order of microseconds. This test was conducted with an angled type II relay having 6 actuator pairs. The voltage behavior in the transitions is not as clean as previous tests, the cause for this has not been determined. 180
FIGURE 98. The long term cyclic testing setup and breadboard. Left side photo shows the entire setup including computer and support electronics. Right photo shows a packaged relay under test. 181
FIGURE 99. Long term cyclic testing results for a packaged microrelay. The resistance includes all parasitics such as the breadboard, package, and support structure. Relay was hotswitched at 4 volts and ran 80mA with a 50W bias resistor. Relay was verified as open and closed at every cycle or the program would have aborted, but data was only collected on a logarithmic basis to allow for

file handling. Relay failure was between 5.7 and 9 million cycles. 182

FIGURE 100. A small piece of Difference Engine I built in 1834 at a cost equivalent to that of 20 steam locomotives. The device is about 1 meter high. 187

FIGURE 101. A recently constructed full implementation of Babbage's Difference Engine I. Substantial debugging was conducted. 188

FIGURE 102. Four terminal relays can be wired up to form "N" or "P" three terminal switches. Because the gate voltage is relative to the source voltage P and N switches use the same layout in reverse. Photo at right shows a fabricated implementation. 189

FIGURE 103. I-V plots for various drive values in a microrelay. Linearity better than 2% FS can be seen in the Vds range shown (limited here by the voltage quantization error), attempts to increase the Vds can require very high currents. Device performance is symmetric so drive voltages are shown as absolute values. Drive voltages were converted to force for reference by using the characterization from the actuation chapter. These traces are for a type II angled relay with only a single pair of actuators. 190

FIGURE 104. Input-Output plot for a complementary electromechanical inverter built using relay based three terminal devices. This device operated on a 10.5 volt power supply and the upper parts of the trace show the maximum 10 volt input range of the 4145b. 192

FIGURE 105. Fabricated electromechanical NOR gate and experimentally verified truth table. 193

FIGURE 106. Circuit diagram for analysis of complementary logic design using relays. Diagram indicates input drive resistance Rd, contact resistance Rc, fan-in Fi, and fan-out Fo. 194

FIGURE 107. Turn-on voltage histograms for various parts of a relay lifetime. Top to bottom shows beginning of life, mid-life, and end of life. When evaluated by design equations this shows that the "wear-in" period at the beginning of relay usage can present undesirable

probability of errors for a logic system. Relay repeatability must be high. 197
FIGURE 108. A simple ZS logic beastiary. The “Z” elements and “S” elements would be alternated when implementing logic. 201
FIGURE 109. A noninverting ZS buffer. This can be driven by a relay with a single actuator pair but has a much lower output impedance because of its additional actuators. While this structure has only two relays and thus a 6% chance of working no functional buffers were found. 201
FIGURE 110. A ZS latch, with 8 relays, presently has a 1 in 65536 chance of working. This device has two functional relays out of 8, exactly the statistical expectation. 202
FIGURE 111. A ZS single bit carry-adder. The complexity is such that yield will have to be substantially improved before this device becomes feasible. 203
FIGURE 112. A view of the scale of integration possible with our microrelays. The left photo shows (left to right) a chip with a 6 relay analog system and a one bit carry-adder, a chip with 30 relays in test arrays, and a chip with two 6 relay systems. The right photo shows a scanning electron micrograph of the middle chip with 30 test relays. This will give the reader a feel for potential die sizes. 203

CHAPTER 1

# Introduction and Motivation

*"Less is More"*

-The most quoted design pronouncement of Ludwig Mies van der Rohe, originator of the international style of architecture.

## 1 Introduction

The North American relay market alone is US$1.4 billion, and the recent explosive market share of surface mount relays (from 1.2% in 1993 to 19.6% in 1997) demonstrates the value of miniaturization[1]. Within the world relay market (projected to grow 5.3% in 1998 from US$4.99 billion) the surface mount sector is slightly smaller due to delays in technological promulgation[2] but still shows immense growth. Clearly then there are immense advantages available to engineering designs from smaller relays.

This huge market which strongly desires smaller, cheaper, and more reliable product is still being served by devices descended from turn-of-the century electromagnetic solenoid relays. Tribological and MEMS based scientific examination,

meanwhile, suggest that the actual phenomena critical to high-quality relays are nanoscopic in scale. The purpose of this research is to explore the physics in detail using MEMS based experiments to get high resolution characterization and then to use this scientific knowledge to remove the vast overdesign inherent in current relays and transition to a more optimal design regime. If the system can be scaled down, relays should be possible which operate with orders of magnitude less size and power while at the same time reaping the advantages of higher speed and cheap hermetic packaging. Once these relays are built, numerous new systems become possible and the last part of this research is focused on these new systems.

# 2 Current commercial relays

Most current commercial relays rely on solenoids being used to pull two large metal “reeds” together, at which point they conduct. A typical relay is the “1-Form-A” standard, which measures 19.6 x 16.5 x 16.5 mm for a single relay (Figure 1).

**FIGURE 1. Some standard "1-Form-A" commercially available relays. Each relay is contained in a 19.6 x 16.5 x 16.5 mm non-hermetic plastic box. Minimum cost is about US$1 (in Q4 1998).**

These relays started their evolution at the turn of the century with large (several kg) solenoid based power switching devices, and have been growing ever smaller and more efficient as people learned to wind smaller coils and blow smaller glass tubes to contain the contact reeds. While the technology has shrunk well due to the beneficial scaling properties of magnetic actuation this method reached a point of diminishing returns some time ago, largely because the process is largely serial and very labor intensive. Only recently have manufacturers started to automate the process, and the serial nature of relay manufacturing continues (Figure 2).

**FIGURE 2. September 1998 advertisement for relays, emphasizing that the serial actions of coil winding, assembly, and soldering have recently been automated. The tasks are still done one relay at a time. Wafer scale**

**manufacturing of microrelays could entirely change the economics of relay manufacturing.**

As a result of the serial processing, relay prices are changing slowly and prices for a single relay are still in the US$1 range for relays of compatible performance to the MEMS relays presented here. The drive method (magnetic coils) also consumes large amounts of power, 360 mW per relay for the devices pictured in Figure 1 compared to the 3-10mW per relay for the unoptimized thermal microrelays presented here.

Reliability constitutes an additional issue for macrorelay use. Failure rates for macrorelays are typically around 100 PPM even after final test because irregularities from relay to relay are great enough to reduce or eliminate the practicality of burn-

in whereas batch fabricated relays should open up the use of burn-in and other semiconductor reliability techniques.

The fact that most low cost relays are not hermetically sealed also imposes a substantial usage cost as many customers have to do lengthy testing and characterization programs for each environment. Environmental testing for non-hermetic devices is nowhere near as standardized as one might hope; when Volvo was considering using a particular relay in cars for Sweden they had to undergo testing with dozens of material blanks in several locations on a fleet of vehicles for a year[3]. If the same relays are then to go on a car sold in California, the tests will need to be done again in California to get results from the different environment. The results of each test are a set of simulations (usually based on a series of tests from established standards if possible) which can be conducted in the laboratory and might be referred to as a "Sweden relay simulation" or a "California relay simulation." Worse yet, this must typically be requalified fairly frequently as new relay products come out. Cheap hermetic sealing makes MEMS relays even more attractive by removing a massive ramp up cost to the user, once a device is proved hermetic using standard tests environmental testing is almost eliminated. Instead of going through multiple location-specific tests as discussed above, for example, a device which has been proven hermetic must only pass standardized temperature and shock tests.

## 3 MEMS relays

The most visible difference between MEMS relays and current commercial relays is, of course, size. The most commonly used relays presented here take up about 400x200 microns of die area (Figure 3, Figure 4 and Figure 5)- small enough that packaging dominates all dimensions. The overall active volume of a MEMS relay can be approximated at $3x10^{-12}$ $m^3$ as opposed to about $5x10^{-6}$ $m^3$ for a conventional relay. This volume reduction of six orders of magnitude is made possible by extensive experimental characterization and development on the subject of the actual contact point and allows design of actuation methods to start over from basic principles with far easier requirements. The basic principle of the relay is to close or open an electrical circuit by pulling a gold crossbar across two gold electrodes.

**FIGURE 3. The microrelay described in this work. The thermal actuator array on the right pulls a gold crossbar across two gold electrodes, closing a circuit. The function of the relay is basically to close an electric circuit by closing a gap with a gold crossbar.**

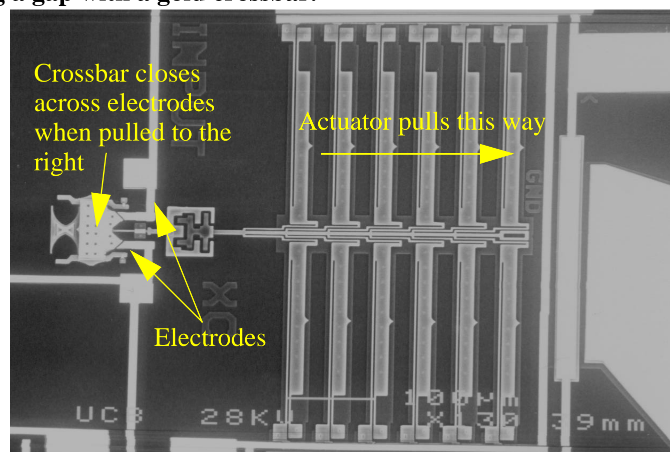


FIGURE 4. **A closeup of the non-actuator part of the microrelay developed here. The crossbar and electrodes can be seen more clearly. The isolation structure allows for complete electrical isolation of the actuation from the circuit, making the device a true four terminal relay.**

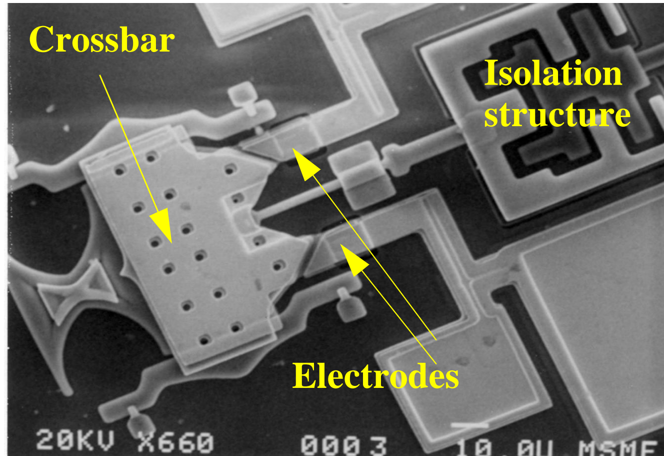
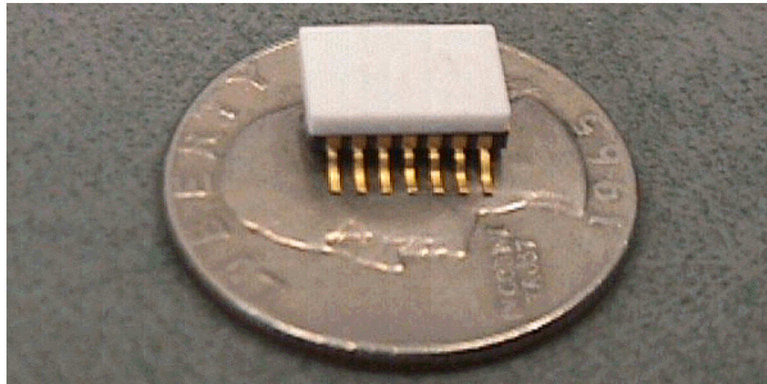



FIGURE 5. **Dozens of microrelays fit easily into this surface mount package. This package has four microrelays bonded to the 14 pins (two drive pins are shared). Drive power plus dissipated power is well within package limits.**



A less externally obvious but very important difference is in the manner of actuation. Whereas the traditional scaling of magnetic relays reduces closing and open-

ing force (the most important part of actuation design), the tremendous scale reduction here allows the efficient use of thermal actuation. Thermal actuation benefits from the small size by taking advantage of the reduced thermal mass and time constants while maintaining high force levels extracted from the basic thermal expansion of structural elements.

Simultaneously, the reduced size and components lowers the amount of extra energy in the system and thus reduces or eliminates some effects such as contact bounce. Since each bounce in a relay provides as much wear (or more) than a switching cycle, this increases lifetime in addition to providing better performance. Removing energy from the system helps reduce the wear tremendously.

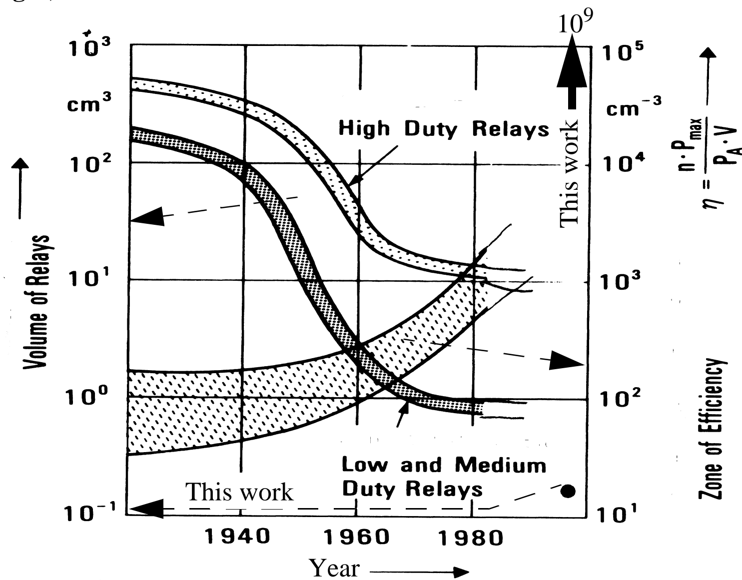
Thus the new surface science perspective has allowed the construction of microscopic relays which can take advantage of more appropriate actuation methods, exhibit higher speeds, and reduce undesirable behaviors.

## 4 Direct comparison

A direct comparison of relay performance can be made using the relay performance metric of [4] (Figure 6) where the metric $\eta$ is computed from the number of relays $n$ in a given package volume $V$ with consideration of their maximum power $P_{max}$ and monitoring power $P_A$ (minimum power to check state). Even for the highly

inefficient DIP packaging used for preliminary testing the package volume is approximately one cubic centimeter for six relays able to switch power in excess of 0.4W (based on 80mA, 5V hot switching long term reliability tests) and which can easily be monitored with 1-2nW. This gives an efficiency number far off the scale shown here.

**FIGURE 6. Historical relay performance chart (from [4]), with indication of microrelay performance. The bands show historical performance regions for electromechanical relays from the 1920s to the early 1990s. The relays presented here are shown along the right edge (time axis value 1999). Size (including demonstrated packaging) is lower than previous solutions by almost an order of magnitude, as indicated by data point at lower right. The industry efficiency metric of these relays is higher by 5 orders of magnitude, putting it well off even this logarithmic chart (value shown at top right).**



## 5 References

[1] *Semiconductor Business News*, Monday, April 14, 1997 Vol. 5, No. 43

[2] *Electronic Buyers' News*, September 14, 1998

[3] Wallinder, I. O.; Eriksson, P; "Characterization of Corrosivity of an Automobile Environment," In Proc. 19th International Conference on Electrical Contact Phenomena; 14-17 Sept 1998 Nuremburg, Germany, pp 157-161

[4] Weiser, J; "Switching Contacts in Relays," In Proc. 19th International Conference on Electrical Contact Phenomena; 14-17 Sept 1998 Nuremburg, Germany, pp 15-21

CHAPTER 2

# Contact Design Issues

*"In the final analysis, every engineering material is rubber."*

-Sir Henry Royce (of Rolls Royce)

## 6 Introduction

As the immortal words of Sir Henry Royce remind us, every material is mutable to some degree. Within relay contacts this becomes critically important as the core behavior within a relay is the plastic contact deformation of micro-asperities to form tiny conducting contact areas. These points are very small (often on the order of nanometers in radius) and almost independent of the apparent surface area but nonetheless critical for relay performance. The primary job of a relay is to press and form these contact points and then to separate them again when contact is no longer needed. Designing relays to accomplish this task, and do so efficiently for a large number of cycles, requires numerous other characteristic behaviors to be understood and considered. This chapter serves to outline the important physical phenomena of microcontacts and describes numerous experiments performed to

establish regimes of behavior and develop a microrelay design philosophy. Phenomena are also presented that are common in macrorelays but not microrelays so that some of the important differences between the two can be illustrated.

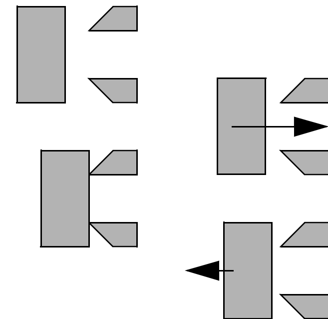
# 7 The relay cycle

The phenomena going on inside an electric relay are varied and complex. During each cycle from open to closed and back again there are literally dozens of mechanisms that contribute to wear and performance. Each mechanism typically has numerous contributing factors that determine its effects as well. In order to properly design and evaluate a relay all these factors and the associated phenomena which draw from materials science, tribology, surface science, electromagnetics, plasma physics, and more must be understood.

As a way to structure the discussion of characterization and design the relay cycle will be presented step by step first (Figure 7). Performance in limited functional environments, characterization, testing, and implications for design will be discussed later. First the relay cycle will be explored step by step and discussions will be made of the physics during each step. The engineering parameters for each factor of each step will be discussed as well as the relative importance and some of the control methods available for both micro and macro relays.

FIGURE 7. **Steps of the relay cycle. The operation of any relay can be broken down into these basic four steps.**

**Steps of the relay cycle:**



1. Relay sitting open (“Relay sitting open” on page 13)
2. Relay closing (“Relay closing, contact being made” on page 25)
3. Relay closed (“Relay closed, current being conducted” on page 39)
4. Relay opening (“Relay opening” on page 53)

## 7.1 Relay sitting open

Often overlooked during the early stages of design, many failures of relays actually start during the time the relay is sitting open and not conducting current.

### 7.1.1 Open relay wear

Although surfaces that are not in direct physical contact are not typically said to “wear,” many failures due to processes that take place in open relays are ascribed to “wear” simply because the material changes that cause them are similar to conventional physical wear. In reality the main mechanisms are material diffusion, creep, and dealloying. While these mechanisms also happen during other parts of the relay cycle they are relatively less important then and are primary sources of wear when

the relay is open. All of these mechanisms are Arhennius based. That is to say that they occur at a rate exponential with temperature and thus may be accelerated by increasing temperature during failure analysis and accelerated aging tests.

#### 7.1.1.1 Chemical diffusion

The materials used in conventional relays are often mixtures of metals and semi-metals made by powder metallurgy. These mixtures were initially developed to decrease arc wear or for other specific purposes and include such interesting combinations as silver tin-oxide[5] and silver iron-oxide[6] (silver cadmium-oxide was in wide use a few years ago before environmental regulations required its elimination). Long term studies of these material systems have begun to recognize that the oxides tend to congregate out into small “nodules” which can have detrimental effects[7]. This congregation presumably comes from the competing forces of diffusion and aggregation. Diffusion forces tend to separate and distribute the different materials into each other while aggregation forces tend to pull together and congregate materials of similar charge properties. Thus evenly distributed insulating particles of oxide within a metal matrix would be expected to congregate into certain sized particles at which point a steady state is reached. This mechanism corresponds well to observed data [7] and predicts that the phenomenon is intrinsic to metal-oxide combinations.

Nodules of oxide within a metal are highly undesirable because they can not only perform as spacers that hold contact surfaces apart but can also rip and tear into opposing metal surfaces to generate particles and arcing protrusions. As the oxides are harder than the raw metals they act as an abrasive, wearing the metal back around other oxide particles until the metals cannot touch because of all the oxide projections.

Chemical diffusion also appears in relays where the contact uses a thin coating. If gold is plated over tin, for example, some limited amount of diffusion takes place. While this results in only impure gold (not a big issue) later corrosion (oxidation) of the tin results in a gold tin-oxide mixture at the surface and continued diffusion flow of tin into the gold now that the initial tin is gone. Now oxide aggregation/diffusion balance takes place as well and the result can be a pebbly oxide-studded surface on what was originally pure gold. Thus the choice of understructure can be crucial. The ideal understructure should bind well to the metallic contact layer but have a very low interdiffusion tendency. During macrorelay design this consideration is usually suborned to the need to harden the surface by making the understructure out of a harder material, resulting in the scenario discussed above. A typical failure generated this way might see the tin-oxide abraded out of the gold so that failure analysis would find tin-oxide “sand” on what was originally thought to be a gold contact.

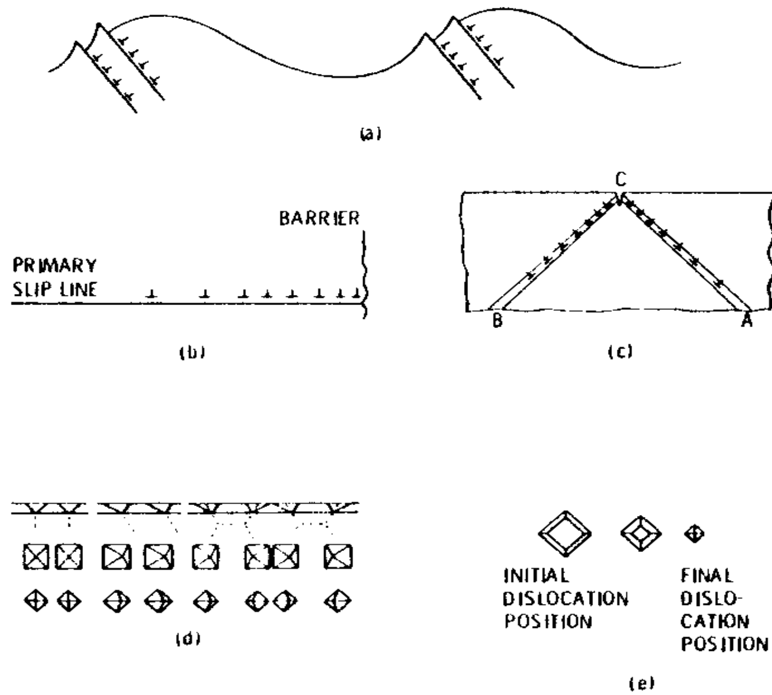
Micromachining material control provides a good way to build contact layers. The contact layers used for the devices presented in this dissertation typically use polysilicon understructure with a chromium adhesion layer/diffusion barrier and a pure gold contact surface. The polysilicon has low interdiffusion with the metals. The chrome adheres well to the polysilicon but has very low diffusion into either of its neighbors (which is good because chromium does form an oxide). And the gold does not oxidize, collect surface layers, or diffuse into the chromium. Materials problems in MEMS may be addressed easily because of the high purity super clean depositions done, and at the same time the low volumes of material mean that if a relay does have materials problems it will be found out quickly. The low volumes also remove cost barriers that might otherwise discourage use of noble metals.

#### 7.1.1.2 Defect diffusion

Contact design cannot be predicated on the assumption that a contact will close in the same shape as it was in at separation. As two surfaces come into contact numerous defects and defect features are generated. Additionally, defects migrate while the contact is open. Point defects (generated in large numbers during relay closing) migrate along slip lines to the surface and accentuate the effects of slip lines on surface topology as well as collecting to form dimples (Figure 8).

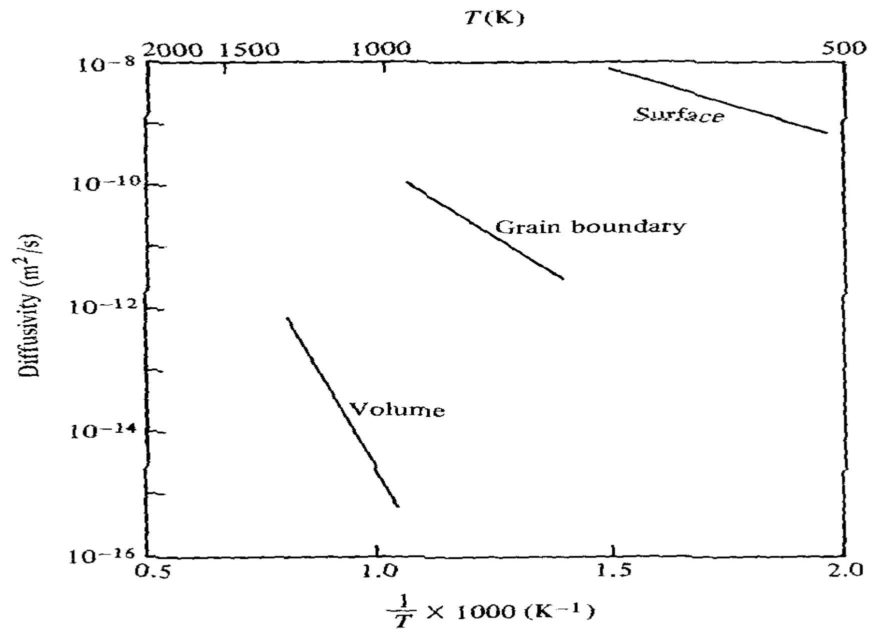
FIGURE 8. **Slip lines and grain boundaries form "highways" for defects, which increases slip line motion and results in defect clustering. Diagrams (a), (b), and (c) show some common line formations and the path taken by**

**defects. Diagram (d) shows a typical line of surface features caused by a slip line intersecting the surface. Diagram (e) shows a row of divots as marks the passing of a slip line below, much like a moving volcano might leave a chain of islands. (From [9])**




The migration of defects, combined with generation during each closure, implies a constantly changing contact surface. Like chemical diffusion, defect diffusion is dependent on temperature (Figure 9), so defect diffusion based aging can also be accelerated by high temperature conditions.

**FIGURE 9. Dependence of defect diffusion on temperature, showing substantial difference in mobility between the volume, grain boundaries, and the surface. This plot is for steel. (From [10])**




### 7.1.1.3 Creep

Relays are typically expected to operate in varying temperature environments, and the large structure of macrorelays combined with variable temperature rates of expansion tend to leave long term stress on many parts. Because of this, material creep has been recognized as a contributing factor in certain long term performance changes[8]. It is unclear how this will effect MEMS relays. The reduced material count seems to indicate much lower stress levels would exist to initiate creep, however it is not uncommon for MEMS layers to contain substantial built-in stress. One

beneficial consideration is that the stress levels in MEMS devices can often be designed for a purpose while macroscopic relay manufacturing issues generally preclude such pre-stress based design.

#### 7.1.1.4 Dealloying

Although generally a minor concern, contacts made of unstable or meta-stable alloys do experience the effects of aging and this can’t be ignored if compound metals are to be used.

### 7.1.2 Corrosion

Corrosion is the most commonly considered and widely feared of the open-contact reliability reduction mechanisms. Testing for corrosion is often done in salt water spray and elevated temperature and humidity over time. This is a major factor because only a few expensive macrorelays can have their bulk hermetically sealed. Since our MEMS relays typically do not have corroding materials and the intention is to hermetically seal them in an inert atmosphere, corrosion is not a concern, removing a large cost and convenience obstacle.

### 7.1.3 Holding voltage

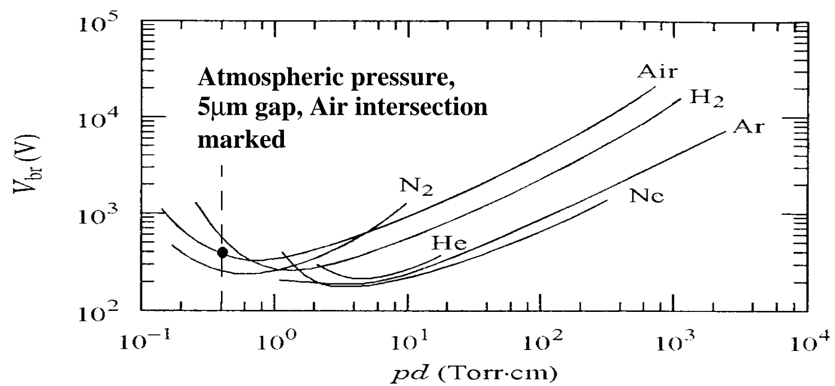
The most dynamic consideration for an open relay is the difference of potential it can withstand without allowing arcing to occur. Holding voltage can be discussed in terms of the maximum transient voltage that can be put across a relay before an

arc begins to strike. This is one of the design facets that has different physics in microrelays than in macrorelays. To understand what is going on requires consideration of arc initiation and the Paschen curve.

In open macrorelays arc initiation typically takes place when the electric field gets high enough that free electrons which find themselves between the contacts (by thermal or photoemission, for example) are accelerated to a sufficient energy to free more electrons from atoms they hit. Each electron then can experience more than one collision and free more than one electron before reaching the anode, a situation which results in large numbers of free electrons passing current in the gap. Once this critical value of electric field has been reached positive feedback (due to the resultant ions forming additional current and higher emission fields at the anode) builds current rapidly and a stable arc initiates which can survive lower fields and can thus be difficult to quench. The limiting field depends on how readily an electron can accelerate and free other electrons and thus on the gas composition, voltage, and mean free path. The mean free path between electron impacts affects the critical field because the electrons lose room to accelerate. Since the limiting electric field/mean free path needed is constant for a given gas the breakdown voltage increases linearly with separation and pressure for a given gas as long as there are sufficient electron mean-free paths for the cascade to build. Thus macrorelays

get open-relay arcs across the **narrowest** part of the gap (Figure 10) because this will be the point of highest field concentration.

**FIGURE 10. Paschen curves for a number of different gasses. Plot shows breakdown voltage versus the pressure-distance product. The right side, which increases with increasing pd, represents the conventional cascade mechanisms discussed above. The left side, which decreases with increasing pd, represents pd products too low to initiate cascade and stems from other mechanisms. Labelled point on air curve shows that the microrelays presented here are very close to the inflection point for air. Closeness to the inflection point reduces effects of topology or pressure changes. (Background data from [11])**

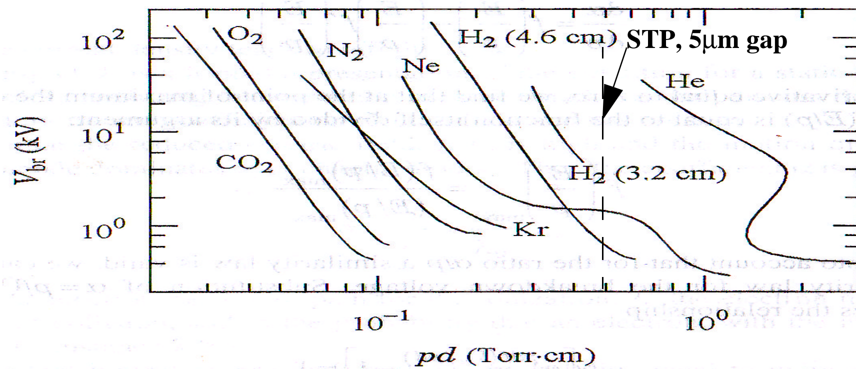



As the gap gets smaller, however, other mechanisms [11] begin to become more important as insufficient opportunities exist for electrons to build up energy and initiate further electron release. Figure 10 shows that the relays described here operate right around the inflection point between the classical arc methods described above and more local non-continuum arc initiation causes.

As the gaps get smaller and there is insufficient space for mean free paths to multiply the number of electrons present the voltage across the gap has to rise in order to

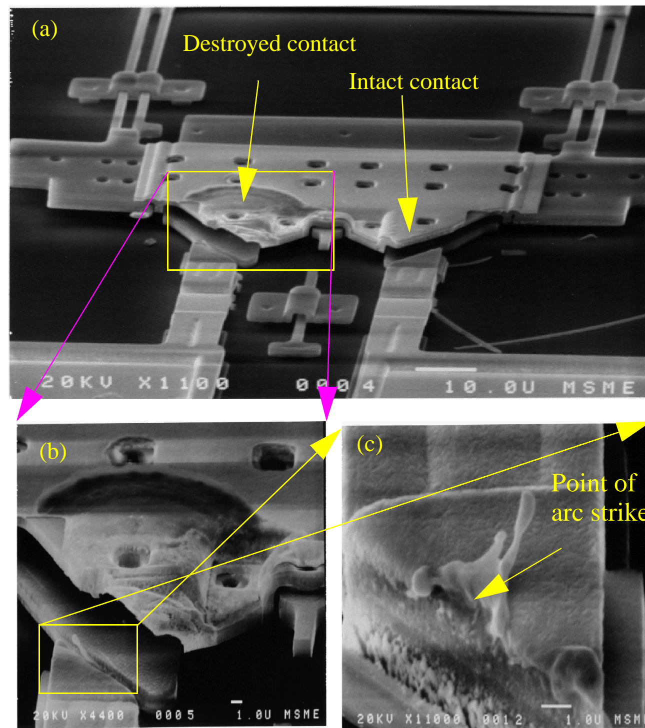
establish an arc by mechanisms that don't benefit from positive feedback. The importance of the electron re-emission decreases slowly and is complicated by particularities of multi-atomic gasses so that the bottom and left hand side of the curves exhibit some significant topology (Figure 11).

**FIGURE 11. A closeup of the left hand side of the Paschen curves from Figure 11 for various gasses. Molecular structure can cause odd shapes in the critical field behavior (such as in He). Choice of gas can have a significant effect on breakdown voltage. (Paschen curve from [11])**




These figures show that the calculated breakdown voltage increases as the gap becomes smaller beyond a certain point. This means that small microrelays may have open-relay arcs across the **widest** part of the gap or at some surface feature which provides a local field concentration. In the case of the relays we have here the openings can span the minima in the Paschen curve (Figure 10) at atmospheric pressure in air so open-relay arcs can tend toward some mid point of the relay. An arc observed in a microrelay while opening an inductive load occurred at a mid-point (Figure 12).

**FIGURE 12. Low-current arcing damage on a microrelay showing that the arc did not strike at the narrowest gap point. The microrelay dimensions are those represented by the dashed lines on the Paschen curve charts. This arc formed initially under opening with an inductive load and presumably transitioned from Townsend arcing to cascade arcing due to high inductive voltages.**




Low gap arcing often stems from electric field concentrations at microasperities and sharp points [11]. Therefore the ideal plots of Figure 10 and Figure 11, based on theoretical conditions, often do not reflect the practical reality. Data collected at very low gaps using surfaces of varying quality can show a wide range of breakdown voltages (Figure 13). This is a recurring theme of microcontact design - as the scale shrinks behavior becomes much more complicated and must take surface sci-

ence into account. A set of measurements conducted in [11] to explore the low *pd* region of the Paschen curve by varying first *p* then *d* in a model relay system produced substantially different results for the two trials (Figure 14), but such behavior has not been observed here. The main lesson to take from these plots is that the Paschen curve is a guideline at best and must be experimentally explored for a given material system.

**FIGURE 13. Breakdown voltages for differently prepared samples at very small gaps and atmospheric pressure. As usual when dealing with sizes in the range of surface effects, behavior becomes very complex and dependent on surface quality. The plot here shows platinum contact arcing data at low spacing for electrodes which were polished and alcohol cleaned (1), activated with discharges in organic vapors (2), and cleaned by electron emission (3). In this region the theoretically expected breakdown voltage should be $10^5$ to $10^7$ volts. (From [11])**

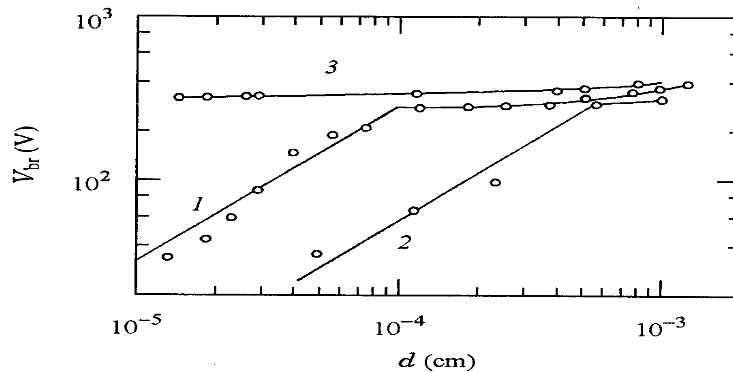
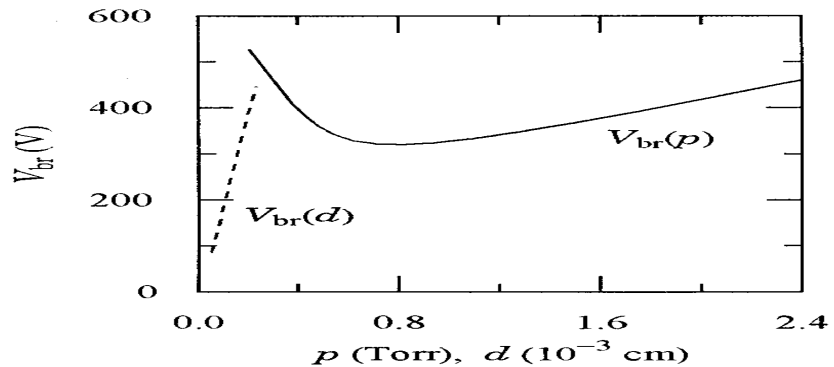



**FIGURE 14. Paschen curves explored by varying *p* then *d* individually. The solid line shows an experiment done by varying p while keeping d**

**macroscopic, the dotted line shows data collected at microscopic distances for a the same material system.(From [11])**




Additionally, experiments have shown[12] that complex electronegative compound gasses such as $SF_6$ can extend the arc point by a factor of two or more by polarizing and therefore "absorbing" some of the potential drop.

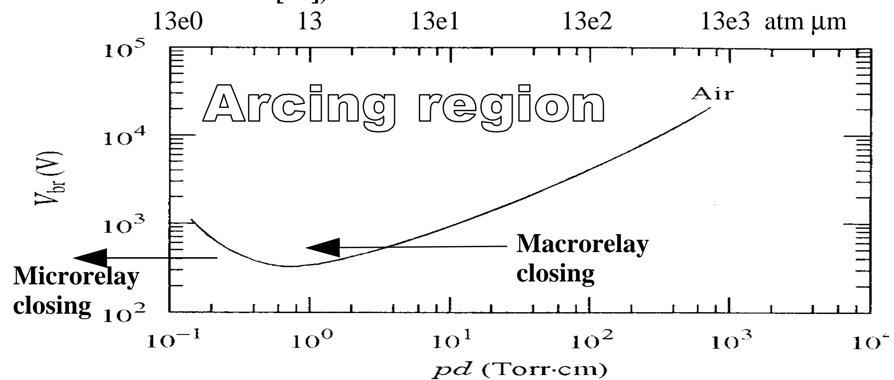
## 7.2 Relay closing, contact being made

### 7.2.1 Closing arc

As the two contact surfaces get closer together the electric field between them increases and the distance decreases. This is equivalent to an operating point moving to the left on the Paschen curve plots and is shown graphically in Figure 15 on an ideal Pachen curve. While it may appear that arcs should only happen above about 300 volts for a closing macrorelay the practical case is that arcs are commonly observed at much lower voltages. The "engineering" value of the Paschen curve can be as much as a factor of ten lower for practical cases due to nonideal

conditions such as finite surface roughness. The result of this is that an arc almost always occurs during macrorelay closure at moderate voltages and yet would not be expected during closure of a microrelay, at least not due to the mechanisms represented on the Paschen curves shown here.

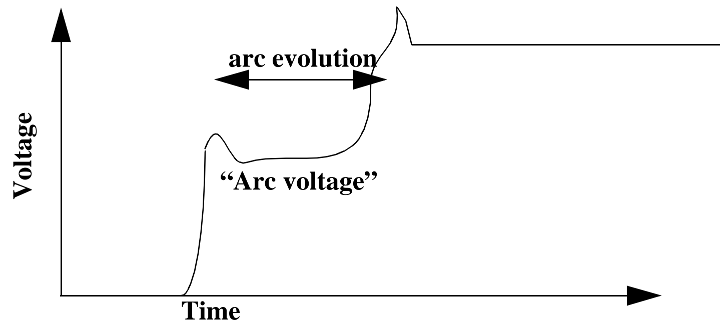
**FIGURE 15. Paschen curve showing movement of operating points in macro and micro relays as the relay closes. Macro-relays move into a lower-voltage breakdown region as they close, often starting an arc. If a micro-relay has not arced already, it is unlikely to arc during closing. (Paschen curve redrawn from [11])**




As the surfaces come very close, of course, various forms of emission can still be expected to occur but these have time scales much shorter than arcing. In experiments conducted closing the relays presented here, below arcing voltage there was no detectable arcing behavior. Twenty-five experiments were carried out with time resolutions from 10μs to 2.5ns and although some transitional behavior was observed it did not have the characteristic behavior of an arc (it was presumably contact point formation). Similar experiments with macrorelays invariably produce

an arc[13]. It should be noted here that the characteristics generally associated with a closing “arc” in the literature are the rapid formation of a stable voltage difference before closure followed by an exponential climb to some overshoot voltage before returning to the true voltage; this behavior is associated with the physical phenomena of developing a space charge of positive ions in the gap due to electron stripping and is described in more detail in [13].

**FIGURE 16. Schematic representation of the expected voltage behavior in an arcing relay. The so called “arc voltage” develops due to space charge in the arcing region and remains fairly constant over the evolution of the arc (typically 2-20 microseconds). This behavior is very useful because it gives us a characteristic to look for to diagnose arcing.**




Thus we have both a theoretical basis and some experimental evidence that arcing, by far the worst wear source of macrorelay operation, is not a factor during the closure of microrelays. It is possible that there is something more fundamental going on as well, given that macrorelays are often reported as experiencing closing arcs even in the 1-2 volt range, a phenomena which is not observed with these microre-

lays. Sufficient to say, for now, that less arcing is expected and less arcing is observed.

### 7.2.2 Impact energy and contact bounce

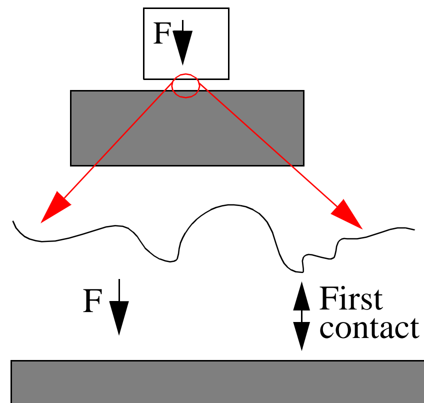
As a relay closes the parts involved develop kinetic energy. As the relay surfaces come into contact part of this kinetic energy goes into deformation of the impacting surfaces and forming a metallic contact patch. If the energy left over after this (and taking into account losses for damping) is enough to separate the two surfaces they will bounce apart and then (if the driving force is still being applied) ram together again. It is not uncommon for a 1-Form-A relay to bounce 3-10 times on a given closure.

Contact bounce does more than just make it difficult to characterize the performance of the relay; each bounce causes the relay to go through the entire close, deform, and opening parts of the relay cycle again. A macrorelay that bounces 5 times, for example, would then experience 6 closing arcs, 6 surface deformation cycles, and 5 opening arcs. Thus a non-bouncing microrelay that does not suffer from closing arcs or bouncing will experience only a single opening arc instead of the 11 various arcs in a macrorelay. Given the arc dominated, and especially closing arc dominated, nature of relay wear this gives an important advantage to microrelays.

### 7.2.3 Surface deformation and contact area

As any two surfaces come into contact they touch first at surface asperities which project above the average surface (Figure 17). These contact points first go through an initial elastic bending period during which the deformation is entirely reversible. As stress gradients build plastic deformation begins first below the surface, surrounded by a region of elastic deformation. This is known as the "elasto-plastic" transition region. Eventually the plastic deformation reaches the surface and the entire contact area goes into plastic deformation which does not reverse when the force is released.

**FIGURE 17. Surfaces which appear macroscopically flat nonetheless make initial contact at surface micro-asperities which then begin to deform. The deformation goes through several behavioral zones including elastic, elasto-plastic, and plastic deformation.**




#### 7.2.3.1 Basics of surface contact

Hertzian theory has established that the relation of area to force $F$ during elastic deformation can be expressed as:

$$Area = kF^{2/3} \quad \textbf{(EQ 1)}$$

where the proportionality constant *k* is dependent on material properties such as hardness. To further complicate the situation the true yield stress of a material under pressure is a "hydrodynamic" quality, which is to say that as flow fields and slip lines begin to move the bulk material below and even lateral to the pressure contribute compressive force to support against the deformation. This can be visualized as pressing against an infinite three dimensional grid of interconnected springs. The overall yielding at a point will be influenced by springs far away and to the sides as well as those directly beneath the point of force application as the overall deformation of the solid is resisted. This leads to an expression for the true stress relative to the yield stress:

$$\sigma = \frac{2\sigma_y}{3}\left\{1 + \ln\left(\frac{Ea}{3\sigma_y r}\right)\right\} [17] \quad \textbf{(EQ 2)}$$

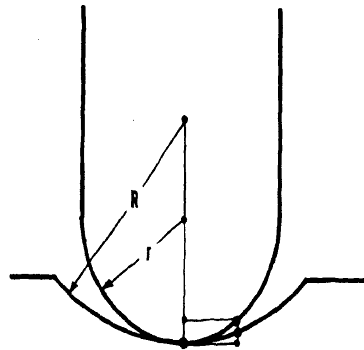
which relates stress $\sigma$ under a microcontact point to the yield stress $\sigma_y$, Young's modulus *E*, radius of contact area *a*, and radius of the contact asperity *r*. This relation serves as a bridge from the elastic to the plastic regime for a spherical indenter pressing into a flat surface. The most important aspect is that plastic deformation begins to occur at $\sigma = 1.1\sigma_y$ and doesn't reach full plastic deformation until $\sigma = 2.8\sigma_y$ for a non-strain-hardening material (reaching 3.3 for maximum strain hardening materials) with some variation for area of contact and radius of the indenter.

These numbers are only approximate for nonideal cases but the important lesson that plastic flow initiates at about three times the yield stress in most metals remains approximately true and very useful.

As the force increases, then, any rigorous analytical treatment must consider the pressure (or stress) under each portion of the contact area in relation to the local stress relative to the yield stress. When the stress at any point passes the yield stresses of a given incremental area, that area will yield. The usual mathematical approximation is that the asperities are hemispherical, and complete theoretical pressure/deformation maps are provided by Hertz theory[15]. Due to the lack of supporting material the edges have a lower yield stress (may be thought of as “softer”) than the center and so move more easily but, because the yield stress does not decrease as much as hardness, the center of the contact area is the first to pass into plastic deformation. Thus when an asperity is pressed against a surface, but not enough to plasticize the entire contact area, reduction in force allows recovery around the edges -- a behavior typical to the elasto-plastic region (Figure 18). While it may seem strange to talk only in terms of hardness and force of deformation a thought experiment shows that two rough surfaces of infinite hardness will meet at only three infinitesimal points. The same would be true of two surfaces meeting with truly zero force, but both are a practical impossibility since materials

have finite hardness and even in the absence of a gravitation field Van der Waals forces are adequate for deformation.

**FIGURE 18. Behavior in the elasto-plastic region where nonrecovered deformation in the center is mixed with deformation around the edges that recovers when force is reduced. Capital R is the original surface radius when compressive force was applied, lowercase r represents the surface as it is after the force is removed. Some area remains in contact due to plastic deformation of the center. (From [14])**




The overall message to take from the discussion of elastic deformation is that it is complicated but can be approximated by (EQ 1) with some value of *k*. This approximation is the usual approach since analytical analysis of any but the simplest, most ideal systems is prohibitively difficult.

Once the force reaches the fully plastic deformation regime things get simpler. Once the force is high enough the area of contact can be approximated simply as the area required for the hardness to offset the pressure:

$$Area = \frac{F}{H} \quad \text{(EQ 3)}$$

where H is the hardness in appropriate units. This can be seen simply as the area which satisfies the yield criteria of (EQ 2):

$$2.8\sigma_y = \sigma_{applied} = \frac{F}{Area} = H \quad \text{(EQ 4)}$$

which gives rise to the definition of the so called "universal hardness" of a material expressed as newtons per meter squared. The universal hardness correlates well to most other hardness measurements.

What all of this complex behavior really means for relay design, however, is primarily that force determines area of contact and that the behavior is far more predictable in the plastic deformation region. Characterization done here to aid in relay design showed that in the 10-25 μN force range contacts involving gold and polysilicon heat themselves into plastic deformation at about 0.8 mA [see the chapter on characterization]. This testing also shows, for the first time, experimental observation of the transitional behavior between elastic and plastic deformation.

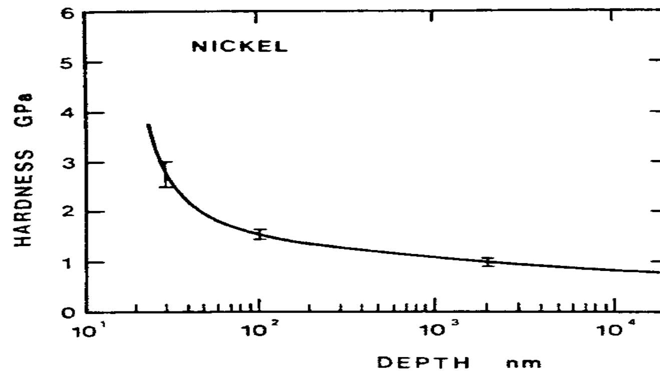
### 7.2.3.2 Three dimensional origins of hardness

The material property of hardness is defined as the resistance of a solid to indentation. This definition can be easily applied to macroscopic experiments involving an almost ideally hard indenter and a large pure block of a substance. The origins of hardness become important as the deformation scale approaches the thickness of the material or the crystalline grain size.

Basic materials theory describes how plastic failure, manifesting at the microscale in dislocations and slip, initiates at grain boundaries and other imperfections. These imperfections can reduce the strength of a material from its atomically derived theoretical value by a factor of up to ten[16]. As the indentation size shrinks to the size of a single crystalline grain, however, the effects of these imperfections vanishes and the strength of the material approaches its theoretical value (Figure 19).

**FIGURE 19. Hardness of polycrystalline electropolished nickel versus indentation depth. As the indentation depth approaches the crystal size the**

**hardness increases from the empirical value (approximately 1 GPa) towards the theoretical value of 4.5GPa. (From [17])**




Another factor which affects the hardness, even in macro-relays, is the fact that hardness is a bulk phenomenon and thus the hardness of a thin film can be influenced by the hardness of its substrate. A classic case of this is the observation that hardness of a thin layer of gold on a hard substrate is closer to the hardness of the substrate than to the hardness of the gold. This is because full plastic deformation of radius *r* is influenced by the hardness of the material within a distance of approximately 3*r* in each direction (including depth)[17]. This means that given the size of deformation region in most macrorelays the hardness of the material must be assumed to be close to that of the film substrate (often steel) with only slight softening due to any surface material. This reduces the effectiveness of macrorelays somewhat, but usually only in cases where an extremely low contact resistance is needed (and this is why such applications often fall to solid gold or mecury relays). Additionally the influence of the support backing reduces the motivation to use soft

materials and thin gold surface layers are often hardened with ruthenium or nickel. Since the resistance range we're targeting here is well above the region where this factor becomes important it is mentioned only as a secondary consideration and example of how macrorelay designs end up with hard metals.

### 7.2.4 Angle of contact

If the contact visualization presented in Figure 18 is considered along with the consideration that hardness is dependent on the material to a depth of $3r$ from the contact point, we might wonder what the effect would be of forces tangential to the surface. The empirically observed result is that tangential forces typically see a lower effective hardness than perpendicular force. To quantify this we first consider the standard vector failure criterion:

$$\sigma_y^2 = \sigma_t^2 + \sigma_p^2 \qquad \text{(EQ 5)}$$

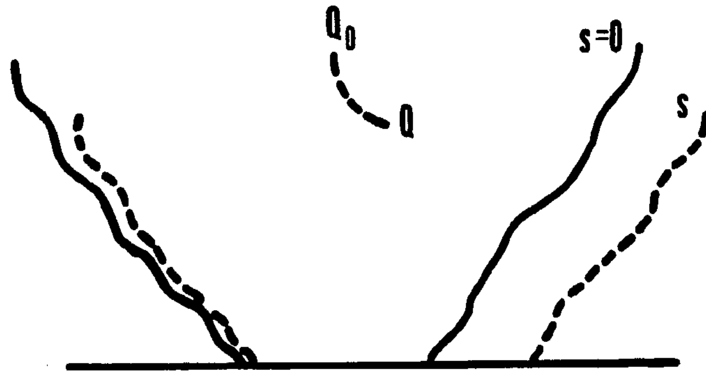
where $\sigma_y$ is the yield stress, $\sigma_t$ is the tangential stress, and $\sigma_p$ is the perpendicular stress. Upon inspection this equation can be seen to assert that failure occurs when the maximum stress vector in any direction is greater than the yield stress and this equation can form the basis of integral volume analysis for bulk materials. When used for two surfaces that are coming into contact, however, the topology and its effect on the effective tangential yield stress must be taken into account. The expression in (EQ 5) doesn't take into account the fact that it is easier to push a

mountain than crush it because there is nothing supporting it in the tangential direction. When the yield point is empirically observed the results are typically expressed:

$$\sigma_y^2 = \alpha\sigma_t^2 + \sigma_p^2 \quad \text{(EQ 6)}$$

where the term $\alpha$ represents the relatively low hardness seen in the tangential direction at an asperity. An $\alpha$ of 2, for example, means that force applied in the tangential direction will be twice as effective at inducing plastic deformation as perpendicular force by causing the contact asperities to “wipe” out in their weakest direction (Figure 20).

FIGURE 20. **Behavior of a contact asperity as both perpendicular and tangential force increase. Point Q represents the geometric center of the asperity, it moves as the contact asperity deforms out into the shape represented by the dashed lines. The tangential force increases the area of contact by speeding the onset of plastic deformation. (From [9])**




This softening in the tangential direction occurs because surface asperities don’t have the structural support behind them in the lateral direction that they do in the

vertical direction. As addressed during the discussion of hardness, the behavior of the material is very dependent on what is supporting it.

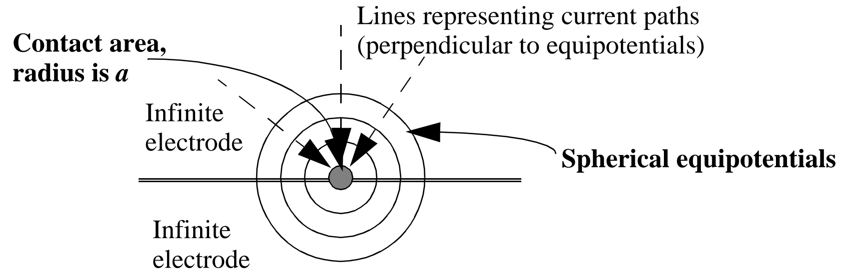
Various authors have reported the value of $\alpha$ anywhere from 2 to 60 depending on the conditions of the test (such as force levels, roughness, and material). The issue is further complicated by the fact that as area of contact is increased by wiping action the perpendicular load must increase to maintain $\sigma_p$ otherwise a state is reached where decreasing perpendicular stress offsets increasing tangential load in the yield criterion (EQ 6). From these elements an optimal contact geometry can be derived if $\alpha$ is known and characterized for a given material. To be optimal this must be done in a force regime where the hardness determining region is entirely within the surface layer, so MEMS is ideal to explore this tribological phenomena. Once solid data is available an optimal geometry can be calculated, thus giving improved performance without any increase in force, area, or complexity. Obviously such data is highly desirable. To date the data remains unavailable and the geometry of the contacts presented in this dissertation, while angled to take advantage of the existence of $\alpha$, may be suboptimal.

## 7.3 Relay closed, current being conducted

### 7.3.1 Constriction resistance

The resistance in a metal-metal contact area is properly known as "constriction resistance," as opposed to contact resistance. This is because the true cause of the apparent resistance is not the presence of some resistive film but rather the effect of the increased current density required to get current through small areas of contact. We can derive the constriction resistance using the approximation that contact area is a single circle (which allows one to use the closed form area calculations given above) and that the metal interconnects leading up to the contact are infinite. We thus represent the system as shown in Figure 21 (after [13]) where the contact surface has been replaced by a contact sphere of radius $a$ and infinite conductivity. The current then comes toward this contact point in radial lines perpendicular to spheres of equipotential as shown.

**FIGURE 21. Simplified contact system for calculation of constriction resistance. Contact electrodes are infinite half spaces, the contact area has radius of a.**




A single side of the contact can be analyzed by integrating the incremental resistance between hemispheres with radii $r$ and $r+dr$. If ρ is the resistivity of the material the resistivity of the unit cube in spherical coordinates is:

$$dR = \frac{\rho dr}{2\pi r^2} \quad \text{(EQ 7)}$$

and the value of resistance on one side of the contact can be integrated out to infinity:

$$R = \frac{\rho}{2\pi}\int_a^{\infty}\frac{dr}{r^2} = \frac{\rho}{2\pi a} \quad \text{(EQ 8)}$$

which makes the constriction resistance of the entire contact:

$$R = \tau\frac{\rho}{\pi a} \quad \text{(EQ 9)}$$

where a temperature factor $\tau$ has been added. And while this general calculation assumes a single circular contact spot (or “a-spot” as referred to in the electrical contact literature), the change due to multiple spot contact can also be calculated. If we divide the area of contact from one spot of radius $a$ into two spots of radius $b$ then the geometric relation is:

$$b = \frac{a}{\sqrt{2}} \quad \text{(EQ 10)}$$

and by (EQ 9) the resultant parallel resistance is approximately 70% of what it would be with a single spot. Unfortunately this is not typically very useful in determining the number of contact points after a contact has closed because their is no way to measure a static resistance and determine whether it represents a single contact point or the collected effects of numerous smaller points.

When (EQ 9) is used in conjunction with (EQ 4), (EQ 3) and (EQ 1) we can develop a general relationship between force $F$ and conductance $G$. We start by using the area of a contact spot:

$$A = \pi a^2 \quad \text{(EQ 11)}$$

where $A$ is the total area. This gives a constriction resistance from (EQ 9) of:

$$R = \tau \frac{\rho}{\sqrt{\pi A}} = \frac{\tau\rho}{\sqrt{\pi}}(A)^{-(1/2)}. \quad \text{(EQ 12)}$$

From this point on it is easier to express values in terms of the conductance $G$; thus we first convert (EQ 12) to:

$$G = \frac{\sqrt{\pi}}{\tau\rho}A^{1/2}. \quad \text{(EQ 13)}$$

From this point we can plug in the area equation for elastic deformation from (EQ 1) to get:

$$G = \frac{\sqrt{\pi}}{\tau\rho}(kF^{2/3})^{1/2} = \frac{\sqrt{\pi}}{\tau\rho}(kF)^{1/3} \quad \text{(EQ 14)}$$

as the relation between force and conductance in elastic deformation. To get the relation for plastic deformation we plug (EQ 3) into (EQ 13) and get:

$$G = \frac{\sqrt{\pi}}{\tau\rho}\left(\frac{F}{H}\right)^{1/2}. \quad \text{(EQ 15)}$$

Both (EQ 14) and (EQ 15) are typically simplified to use a single leading constant, a temperature coefficient, and a force term:

$$G = \frac{k}{\tau}F^{b} \quad \text{(EQ 16)}$$

where k is a constant dependent on microstructure of the contacting zones and the material. Coefficient $b$ is a function of behavior which we expect to be 1/3 in the elastic regime from (EQ 14) and 1/2 in the plastic deformation region from (EQ 15). These behaviors have been directly observed and measured in the relays presented here.

If there is a resistive film in the relay causing a parasitic resistance (as opposed to a constriction resistance) the value of *b* is typically observed as 1. This state behaves empirically as if there is a film being squeezed linearly by the compression force[13]. While F has a larger exponent in this regime, k is generally several orders of magnitude lower because of the resistive film, so performance decreases. Many macro-relays must operate in this regime, however, because they need hard metals which grow films and operate in uncontrolled environments due to the cost of hermetic sealing for large electromechanical systems. This is another factor that often drives the massive force demand of macrorelays.

Thus the three operating regimes are elastic, plastic, and compressive. The plastic regime has the lowest conductance change with force and also typically has a fairly low maximum current. The compressive regime (also known as the “contaminated” regime) has a high conductance increase with force but the parasitic resistance responsible is typically unacceptably large. Neither elastic nor compressive regimes are desirable for relay operation. The plastic regime, however, has a nice combination of current range and shows the lowest resistance of all three regimes and is thus the desirable behavior for relay design.

### 7.3.2 Supertemperature

While the constriction resistance in a contact may stem from the repulsion of like charges it nonetheless dissipates power. The resulting steady state temperature increase with current is simply the temperature rise needed to dissipate the input power through the thermal conductance of the system:

$$\Delta T = \frac{P}{G_t} = \frac{I^2 R}{G_t} = \frac{I^2 \tau\rho}{G_t \pi a} \qquad \text{(EQ 17)}$$

Where $\Delta T$ is the supertemperature (defined as the temperature rise above ambient within the constriction), $P$ is the dissipated power, $G_t$ is the thermal conductance away from the contact points. The terms have been expanded using (EQ 9) to show the feedback which occurs due to the temperature dependence of resistance. The value of the temperature term can be calculated using the same spherical integrals that were used to generate (EQ 9) because each equipotential will also be an isotherm:

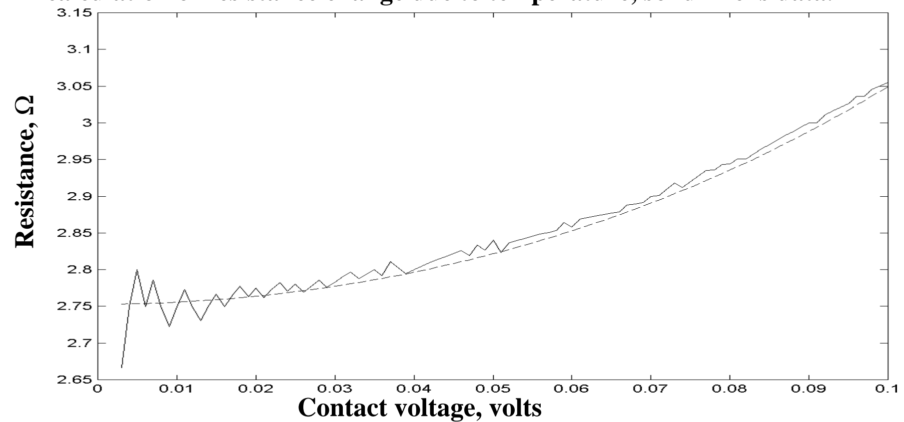
$$\tau = \frac{R(T)}{R_0} \approx \left[1 + \frac{2}{3}\lambda\Delta T\right] \qquad \text{(EQ 18)}$$

where this $\lambda$ is the temperature coefficient of resistance within the metal being used for contact. Derivation of this is presented in [13].

We can use (EQ 17) and (EQ 18) together to calculate the temperature and the expected resistance due to current at any given nominal resistance and current level according to the material properties. Figure 22 shows such a calculation fitted to

data from the relays presented here. The only fitted value here is the thermal conductance, which means we can use the method not only to calculate temperature but also to extract the thermal sinking ability of a given structure in actual operation.

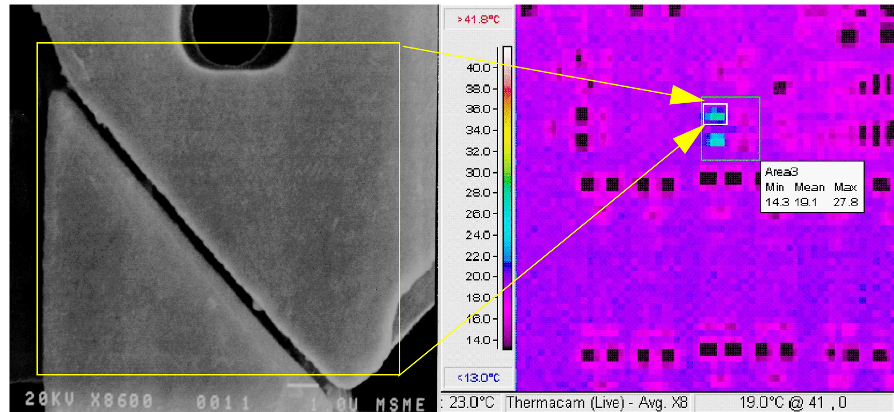
**FIGURE 22. Resistance plotted versus contact voltage (to follow relay characterization convention established by [13]). Dashed line is theoretical calculation of resistance change due to temperature, solid line is data.**




The open design of the relays presented here also allows for direct measurement of the temperature by infrared microscopy (Figure 23). The direct observations confirm the values calculated by combination of (EQ 9), (EQ 17), and (EQ 18) so this method can be used to extract temperature data from current and resistance plots.

**FIGURE 23. Thermal microscopy used to verify temperature performance of an active relay. The SEM at left shows a closeup of the contact area of interest as indicated by the small box at right. In this case the current is**

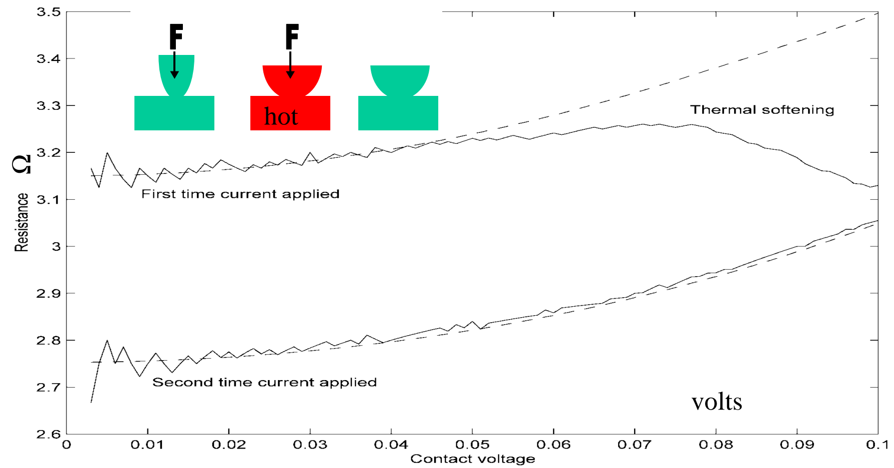
**30mA and the temperature increase is about 12C, analytically predicted value is 12.2C using dissipated power and fitted $G_t$ from (EQ 17).**




### 7.3.3 Temperature induced softening

While current is passing through the constriction point the temperature is elevated and the Young's modulus decreases [18]. The decrease occurs as the temperature increases separation between atoms and therefore reduces interatomic potential gradients [more information pertaining to this appears in the thermal actuation section]. This decrease in Young's modulus directly reduces the yield stress of the material as indicated in (EQ 2). As the yield stress decreases the area of contact must increase proportionally (in the plastic deformation regime) if the force remains constant (EQ 4), causing the contact to experience negative thermal feedback as the thermal softening increases the area of contact and therefore decreases resistance and thermal power dissipated. This process results in stable thermal softening behavior which leads to increased contact area and decreased resistance (Figure 24).

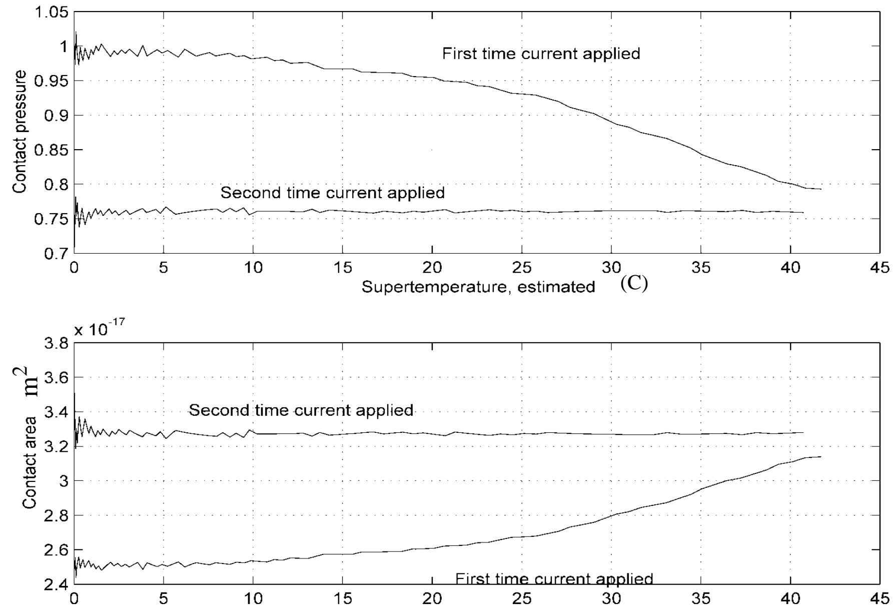
**FIGURE 24. Resistance of a relay upon first application and second application of current while under constant force load. As predicted, the resistance drops off as current increases on the first application as the relay softens and the area of contact increases. This relay is in plastic deformation, so the increased contact area will remain until separated, thus the second current sweep experiences a lower resistance and returns up the heating curve to the same stable point. The dashed lines represent theoretical resistance increase with temperature as predicted by application of (EQ 17) and (EQ 18). Data is presented in terms of contact voltage to conform to convention [13].**




Calculations using (EQ 9) applied to the data of Figure 24 can be used to directly observe the physical behavior in the contact being tested. And since the theoretical predictions in Figure 24 agree so well with the data, we can plot these relative to the predicted supertemperature at the contact point (Figure 25).

**FIGURE 25. Plots of relative contact pressure and absolute contact area versus supertemperature (temperature above ambient) for the data of Figure 24. As expected, the contact pressure (which will be equal to the yield pressure) decreases as temperature increases on the first current application and area increases. On the second current sweep the area does not increase because the effective contact pressure remains below the yield**

**pressure and it does not recover (decrease) because the deformation is all plastic.**




This behavior is very important in that it shows how the operating point can be very influential in determining performance. Thus typical relay specifications can be somewhat misleading. A relay which says "0.1 ohm, 10 Amps" may very well exhibit a resistance several orders of magnitude higher at low currents after each closing action, only to exhibit resistance an order of magnitude lower at low currents after being exposed to high currents during the same cycle. Failure to understand this principle can lead to use of the wrong relays or poor design of an application-targeted relay. This effect is dramatic despite the relatively low temperature increases as seen in Figure 25. Typical temperature increases range from the

tens of degrees for milliamps to hundreds of degrees for significant fractions of an amp.

A range of detailed tribological information is available from current/voltage sweeps. Combined with the low force techniques of MEMS and various techniques to improve resolution, these equations allow new characterization experiments.

#### 7.3.3.1 Thermal aging

Many of the phenomena that have been discussed here are strongly temperature dependent, including the various types of diffusion, tarnish film formation, creep, and corrosion. All of these happen faster at the critical points of contact when the temperature is elevated by the supertemperature and this must be considered during longevity consideration. The amount of time spent at a raised temperature can be a determining factor in failure mechanisms otherwise considered to be environmental or cyclic, and all of these can be accelerated by spending long time periods with heavy current loads.

### 7.3.4 Welding

Considerable discussion has been given to the nature of contact area and the types of deformation involved. Emphasis should be placed on the fact that these contact areas, if they are quality metal-metal contacts, represent areas where the two contacts have joined. No division remains and the metal can be considered structurally

continuous through the contact area. Thus contacts are said to “weld” at the point of contact and if this contact area is greater than that which can be torn apart by whatever actuation will be separating the contacts the failure is called a “weld-failure” or simply a “weld.”

Yet while the surfaces do form a structurally continuous unit the plane of contact does have a disproportionate number of cracks and imperfections. Since these “cracks” (actually just the space between the contact points) amplify separation forces along the plane where the materials first met, contacts often seem to snap apart cleanly without material transfer.

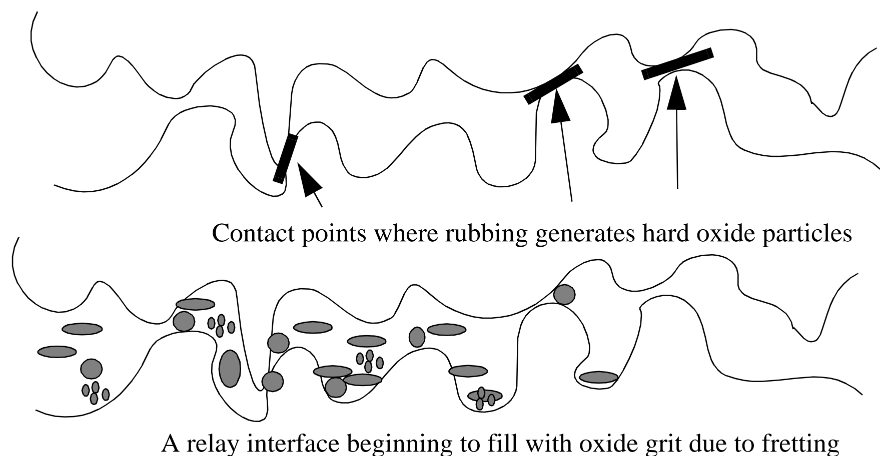
Whatever happens during separation it is important to note that since relays are typically run in plastic deformation welding is a critical and normal part of relay operation. The challenge in relay design is to weld and then separate the contacts repeatably.

### 7.3.5 Fretting

Fretting is the phenomena of vibration induced microscopic wear motion in touching macrorelay contacts. The two contact parts experience cyclic rubbing motion which causes direct physical wear and allows for the growth and fracture of surface tarnishes -- thus generating particles (Figure 26).

Because of the large and separate parts inherent in most relays, ambient vibration or sound is very efficiently coupled to microscopic motion of the contact junction. Numerous studies on typical relays determined that the typical amplitude of motion in macro-relays is in the range of 10 microns with frequencies in the tens to hundreds of Hertz [19].

**FIGURE 26. Schematic representation of the process of fretting with corrodible material. As oxide layers or other corrosion products form on the surface they are broken off by the fretting process and become abrasive particles to further enhance the process. Eventually the relay gap can become packed with such particles. Fretting is a main cause of relay failure and the single most prevalent cause of a closed relay suddenly becoming intermittent under vibration.**




The results of this fretting can be minimal to disastrous and in most practical applications fretting competes with arcing to be the most severe source of relay wear. Fretting generates numerous particles as surface films form and then are rubbed off and these particles, as oxides, are typically harder than the underlying contact mate-

rial. If the materials used in macrorelays were ideal and tarnishless there would be few ramifications to fretting. Unfortunately this is not the case and tarnish/corrosion formation at the surface is amplified by fretting as the fretting breaks off chips of corrosion products; the newly revealed surface begins to corrode as the loose chips form an abrasive driven by fretting to accelerate the process. Under these conditions the relay is soon filled with hard, abrasive, and typically non-conductive chips of corrosion products -- disabling the relay (Figure 26).

Even in relays made with non-tarnishing metals (such as gold) fretting vastly accelerates wear by generating grains of strain hardened metal which are harder and therefore abrasive.

Fretting typically does not occur at all in truly microscale relays. The minute masses of the parts (typically nanograms versus many grams for the smallest macrorelays) reduces the coupling of vibration to negligible levels in microrelays, eliminating fretting effects. The relays presented here have never shown the characteristic scratches or intermittent contact that would accompany fretting. This tremendous advantage inherent in scaling eliminates a scourge which has been the subject of thousands of relay research papers and which is still a primary cause of wear in macrorelays.

## 7.4 Relay opening

### 7.4.1 Contact adhesion and separation

Due to the welding effect discussed earlier the contact can be assumed to simply be two blocks of material joined by connections the size of the contact areas. As the relay is opened the contact areas resist plastic flow until the stress exceeds the yield stress, at which point the material necks down significantly. Failure is then enhanced by the many cracks along the contact interface. This separation is very similar to a standard bar-pulling test, and the yield stress for separation can be obtained in similar ways [10]. Since the yield stress is lower than the hardness by about a factor of 3 the opening force is typically lower than the closing force and the ratio of the two, often called the "adhesion coefficient" is typically less than one. Due to differences in mechanisms and surface energies adhesion coefficients are substantially different for different materials at extremely low loads. Once plastic deformation is entered the values of the adhesion coefficients of different contact metals become approximately equal at one third (the ratio of yield stress to hardness) if surface energy terms are ignored. For macrorelays this is a fairly safe assumption, although it must be more thoroughly explored for microrelays.

Thermal softening can complicate the issue substantially. As we have already seen in section 7.3.3 the effects of current include increased temperature and increased area of contact due to thermal softening. This area counts as welding area and must

be added to the contact area. Thus many studies of adhesion coefficient which use different currents to measure the resistance (in order to determine area of contact by (EQ 9)) produce different results for similar materials. Additionally any contact which has passed current will require increased force to separate it, and by this mechanism the force to separate a contact can exceed that required to close it.

These basic mechanisms are the important elements of macrorelay contact separation. Microrelays, with their lower force operation, turn out to be somewhat more complicated and a more complete development of design guidelines for microrelay separation requires substantial characterization of plastic transition point and surface energy forces.

### 7.4.2 Particle generation

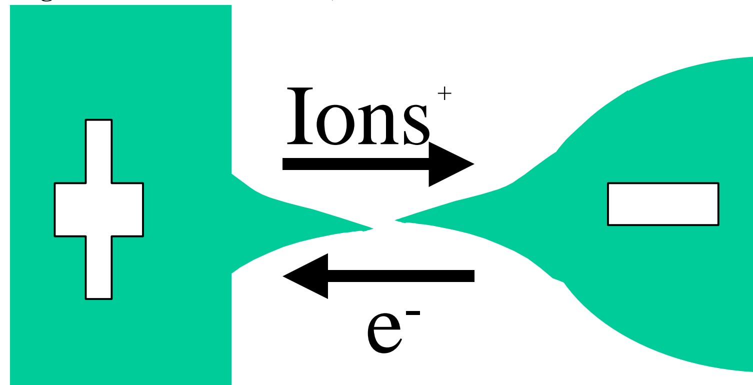
A full analysis of the requirements for particle generation in metal-metal contact can be found in [20] and depends on finding the balance between the energy of newly created surface and the energy stored in deformed material during closing. The important point is that for gold, the contact material used here, particle generation requires a minimum force of 250mN, well above the force levels used in our microrelays. This is another advantage for low force usage.

### 7.4.3 Opening arc

As contact points are separated they can pull and neck out into points. These points are able to initiate an arc by methods completely different than the electron cascade mechanism discussed earlier in association with the Paschen curve. As the points first separate they are extremely close together and extremely sharp, leading to direct electron field emission (also sometimes referred to as ballistic emission). These electrons flow from the cathode to the anode, where they form a tiny spot of great temperature due to energy dissipation and electric field due to a generated space charge of ions. This high energy region continues stripping and liberating ions from the anode to form an ionic counter current. This discharge mechanism, once initiated, is fairly stable and survives extension to several times its initial length. This process by which an arc is always formed in an opening current under load (even in vacuum) is often referred to as the “metal vapor arc” or “Townsend arcing.” In atmosphere and when sufficient mean free paths are available the metal vapor arc typically serves to initiate a conventional electron cascade arc and the system transitions into conventional arcing if the conditions are right. This is presumably what happened in Figure 12.

FIGURE 27. **As contacts separate fine points (on the order of a nanometer) can be generated which will initiate Townsend arcing. Electrons emitted from the cathode sublimated ions from tiny spots on the anode and a countercurrent is initiated, even in vacuum. If conditions are right this can**

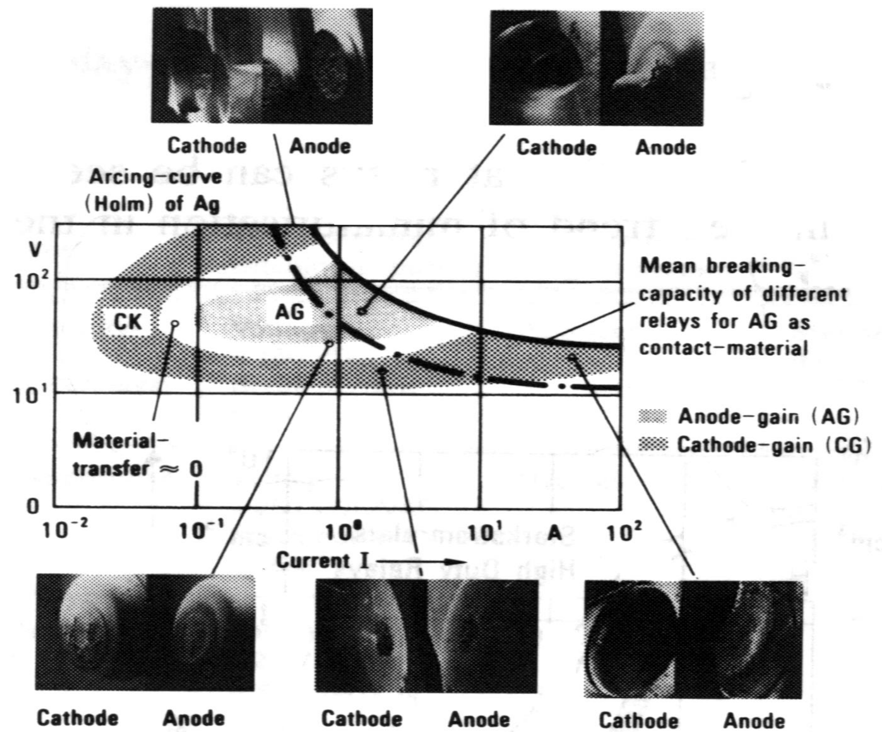
**lead to cascade breakdown by providing the initial arc (which needs far less voltage to sustain than to form).**




Unlike the conventional electron cascade arc (Figure 28) the metal vapor arc has predictable wear behavior, the material transfer of the metal ions goes from anode to cathode and is relatively minimal. This is another dramatic advantage of microrelays, since proper design can eliminate all but the metal vapor arc. Macrorelay design, contrastingly, typically does not worry much about metal vapor arcing because it is inconsequential compared to cascade arcing.

**FIGURE 28. Material transfer direction and magnitude in conventional arcs involves at least six conflicting mechanisms which lead to extremely complex wear behaviors. This typical plot of wear direction and magnitude for a particular macro-relay system [21] shows how difficult to**

**characterize this process can be. Microrelays which avoid most arcing face much simpler wear issues.**




The predictable nature of the Townsend arc material transfer also demonstrates another reliability issue in that it always occurs from anode to cathode. Thus relays will be worn out faster when switching DC in a uniform direction and for cycle testing microrelays should be tested in a single direction to generate the most rapid failure.

# 8 Considerations for microcontact design

## 8.1 Materials for contacts

From the discussion of defects and long term material behavior it is evident that the ideal contact material should be something stable over the long term, preferably with minimal defects and controllable grain size. Grain size control is important because it affects microhardness and defect transport. MEMS is well suited to this task as MEMS films are often super high purity and of controllable structure. This compares favorably with conventional contacts which are usually made with powder metallurgy using powders with a maximum of 99.95% purity and large, unpredictable numbers of defects and grains.

If the designer has a controlled environment without a lot of overly large forces a soft material is also preferable to a hard one. A soft material makes it easier to make and break contact. The primary disadvantage is that a softer material is generally more susceptible to fretting. Historically this has led to harder materials for larger relays and softer materials for smaller relays. For our microrelays fretting is not a significant factor and soft materials can be used.

Having chosen uniformity and softness as initial constraints the issues of thermal stability and surface film formation can be considered. Gold has one of the highest

melting points of soft metals and does not form surface films, making it the preferred choice when all design criteria are considered.

## 8.2 Force and mass considerations for contact

If there is a single greatest advantage inherent in MEMS based microrelay design it is in the elimination of excess force and mass. The minute masses and forces almost completely eliminate the inertial effects typically present in even the most carefully actuated macro-relay because mass decreases as dimensions cubed. This eliminates bounce on contact by reducing the kinetic energy in the moving parts. In dozens of transient tests observed on timescales from 10 microseconds to 2.5 nanoseconds during this work no bounce was ever observed. This is tremendously important because each bounce sends the relay all the way through the relay cycle again. A bounceless relay capable of 10 million cycles might be rated at only 1 million cycles with a reasonable average of 10 bounces. The reality can be even worse than this since the bounces happen close together and dramatically increase the time spent in arcing and high temperature conditions.

Additionally, the elimination of excess force reduces the defects generated at each closing. This in turn leads to more tractable surface behavior.

## 8.3 Eliminating arcs

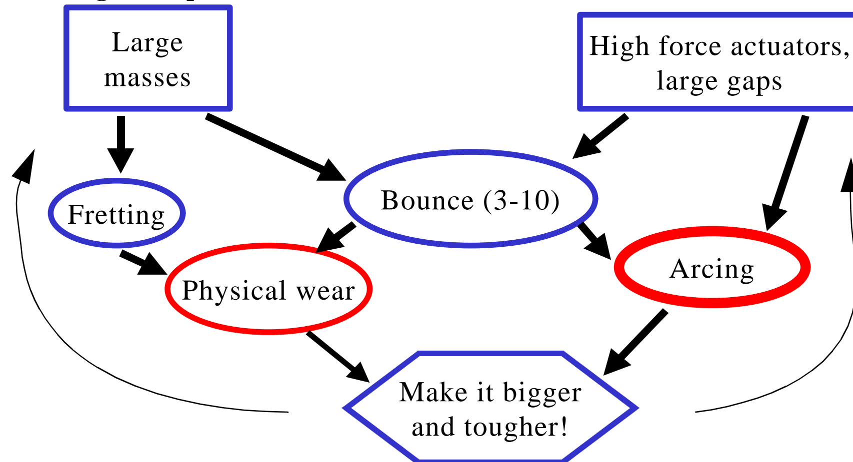
To eliminate arcs for microcontacts we must consider only two types of arcing, maximum hold off voltage breakdown and the opening arc. Operation in vacuum is one solution available to microrelays but not easily to macrorelays which deals nicely with both issues. The lack of any gas at all dramatically increases the holdoff voltage in a microrelay by moving the operation point farther to the left on the Paschen curve. Vacuum also rapidly quenches the opening arc right after the metal-ion stage. Unfortunately it also slows the response of the thermal actuators which are used almost exclusively here. A few macroscopic relays have been vacuum packaged but it dramatically increases cost.

Electronegative gasses (or appropriate choice of gas from Figure 11) minimize arcing while at the same time conducting heat; the pressure can then be set to the lowest value practical for operation of the thermal actuators. It should be recognized that this is the opposite advice than would apply to macrorelays regarding pressure since on the right-hand side of the Paschen curve increasing pressure increases hold off voltage.

# 9 Comparison of macro- and micro-relay design paradigms

Having enumerated many of the issues surrounding relay contact design it is worthwhile to compare the design of macrorelays and microrelays. Such a comparison helps to illustrate the dramatic differences in design and eliminate many of the persistent yet incorrect intuitive tendencies in relay design; following design intuition based on macrorelays can easily lead to non-functional or unreliable microrelays.

**FIGURE 29. Design flow chart for macrorelays. Starting with high force actuators, large gaps, and large masses leads to arcing, bouncing, and fretting. All of these issues reinforce the tendency to use bigger and stronger components and harder and more exotic contact materials.**



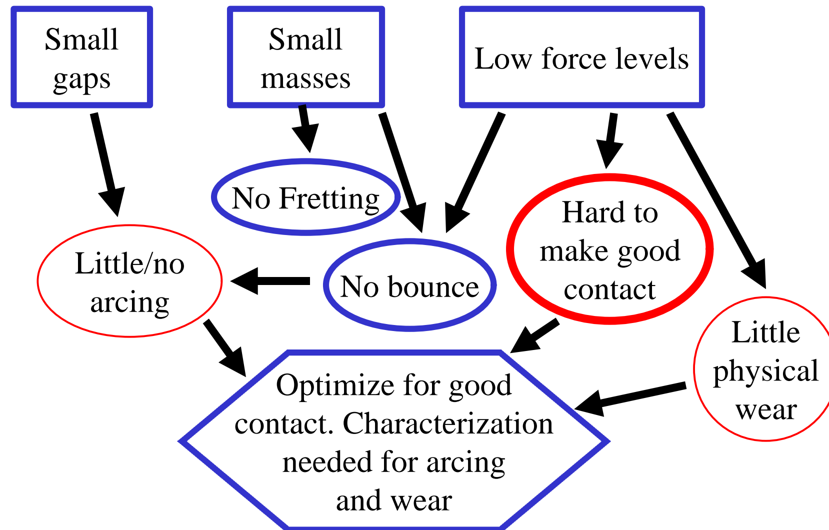

The design cycle of macro-relays (Figure 29) feeds back into making larger and bulkier devices. The large masses inherent in a macro-relay promote fretting and bouncing, which cause substantial physical wear and encourage the use of harder contact materials and thicker starting materials. The large gaps cause very damag-

ing closing arcs, which are magnified tremendously by the bouncing caused by high forces and masses. The arcing, like the physical wear, encourages the use of exotic and hard materials and thicker starting electrodes. While some of this might be mitigated by using lower forces, high forces are needed to get good performance out of the harder contact materials and reducing the force faces diminishing returns because the masses and gaps cannot change much with macro-machined relays. All of this together forms a design feedback that encourages higher forces and stronger subassemblies. The resulting contact metal choices (which often involve suspended oxide particles to increase hardness) lead to the issues of diffusion and wear particle generation discussed earlier - a reliability problem that is often addressed by applying more force or larger contact areas (so that at least some area will prove viable).

Microrelay design, on the other hand, begins with a different set of assumptions and heads in a completely different direction (Figure 30). The tiny masses involved (nanograms in this work) eliminate fretting. The combination of these tiny masses with relatively low force levels (micronewtons in this work) eliminates bouncing. The small gaps eliminate closing arcs so the only arcs causing wear are the relatively benign Townsend arcs that occur each time the relay opens, which is equal to once each relay operation since there is no bouncing.

**FIGURE 30. The microrelay design flow. Unlike macro-relay design we begin with the assumption of small gaps and masses with limited available force. The only arcing is Townsend arcing and the physical wear is minimal**

**(though the available wear volume is also minimal, so this cannot be ignored). The most important factor becomes the make and break of good electrical contact.**




The main issue, then, is one of characterization to discover how low the values of force can go while still providing adequate contact and separation. Substantial characterization is required, for example, to discover the plastic transition point of a micro-relay so that the regime of operation can be established and the proper design equations used. While it might seem trivial to simply plug material properties into the equations presented earlier the complexities of surface science complicate things dramatically: the degree of work hardening going on at the surface interface will dramatically affect behavior, as will the number of actual contact points since the equations presented above generally use a continuum approximation of a single contact point. Once a couple of important establishing experiments have anchored

the behavior, however, designs should be possible which take advantage of all the different behaviors of micro-relays.

# 10 Conclusion and caveat

The principle conclusion from contact physics considerations is that macrorelays and microrelays are completely different devices. Microrelays have masses small enough to eliminate bounce and gaps small enough to eliminate closing arcs. This low mass and non-arcing geometry is the definition of microrelay that will be used here, and is a useful definition in that it prescribes when to use the macrorelay design philosophy (Figure 29) and when to use the microrelay design philosophy (Figure 30).

One caveat which needs to be mentioned is that the information presented here has been presented in a microrelay centric fashion. The contact physics are presented such that if one wants to alter a relay presented in this work for larger holdoff voltage, for example, it will be clear that vacuum or different gaps can be applied and that wider gaps would introduce closing arcs. Meanwhile little information has been presented covering some of the ingenious solutions that have been applied to macrorelays to get around some of the problems presented here. In applications where zero bounce is required (such as super high reliability relays for spacecraft), for example, relays sometimes have dedicated PID controllers which apply care-

fully controlled driving impulses that cause the relay to close softly and remain anchored. These solutions have drawbacks in size, cost, and complexity and such methods are considered outside the scope of this analysis.

# 11 References

[5] Walczuk, E.; et al. “Computer Controlled Investigations of the Static Welding of Silver Tin Oxide Contacts” In Proc. 19th International Conference on Electrical Contact Phenomena; 14-17 Sept 1998 Nuremburg, Germany, pp 311-316

[6] Hauner, F.; Jeannot, D.; Pinard, J. “AgFe and $AgFe_2O_3$ Contact Materials for Low Voltage Switchgear” In Proc. 19th International Conference on Electrical Contact Phenomena; 14-17 Sept 1998 Nuremburg, Germany, pp 317-323

[7] Schopf, T.J.; Hundt,I; “Effects of Different Loads on the Surface of Silver Metal Oxide Contacts for General-Purpose Relays” In Proc. 19th International Conference on Electrical Contact Phenomena; 14-17 Sept 1998 Nuremburg, Germany, pp 83-87

[8] Schoft, H; Lobl, H; Kindersberger, J; “Creep-Ageing [sic] of Bolted Electrical Busbar Joints” In Proc. 19th International Conference on Electrical Contact Phenomena; 14-17 Sept 1998 Nuremburg, Germany, pp 269-276

[9] Pashley, M.D.; Pethica, J.B.; Tabor, D “Adhesion and Micromechanical properties of metal surfaces,” Wear, 100, 1984 pp7-31

[10] Shackelford, J.F.; “Introduction to material science for engineers,” Macmillan Publishing Company, 1988

[11] Korolev, Y.D.; Mesyats, G.A. “Physics of pulsed breakdown in gases [sic],” URO-Press 1998

[12] Johler, W; “Extended life of Telecom relays by using $SF_6$” In Proc. 19th International Conference on Electrical Contact Phenomena; 14-17 Sept 1998 Nuremburg, Germany, pp37-44

[13] Holm, Ragnar; “Electric Contacts,” Springer-Verlag 1941

[14] Sharma, S.P.; “Adhesion coefficients of plated contact materials” Journal of Applied Physics, V 47 No. 8, August 1976 pp 3573-3576

[15] Landau, L.D.; Lifshitz, E.M.; “Theory of Elasticity” Pergamon Press, 1986

[16] Barrett, C.R.; Nix, W.D.; Tetelman, A.S.; “The Principles of Engineering Materials,” Prentice Hall, 1973

[17] Pashley, M; Pethica, J; Tabor, D; “Adhesion and Micromechanical Properties of Metal Surfaces” *Wear* 100, 1984, pp 7-31

[18] Collard, S.M.; McLellan, R.B.; “High-temperature elastic constants of gold single-crystals” Acta Metallica Materials, V 39 No. 12, 1991, pp 3143-3151

[19] Antler, M; “Tribology of metal coatings for electrical contacts” Thin Solid Films, Vol 84 No 3, 1981, pp245-56

[20] Rabinowicz, E; “Friction and Wear of Materials”; 1995, Wiley-Interscience

[21] Weiser, J; “Switching Contacts in Relays, ”In Proc. 19th International Conference on Electrical Contact Phenomena; 14-17 Sept 1998 Nuremburg, Germany, pp15-21

CHAPTER 3

# Fabrication

*“Should various misfortunes assail thee, persevere in patience of body, speech, and mind.”*

- Buddhist Scripture

## 12 Process selection

The discussion of contact physics established a number of criteria for building a successful and efficient relay. Working through the design considerations showed that if we can expect some flexibility and options in actuator design the desirable qualities for contact design are minimal mass, soft metal contact, availability of good contact metal chemical isolation, and preferably some mechanical versatility to allow characterization and creative implementations (such as bistable relays).

In this work one more criteria was added: foundry availability. This was desirable not only to maximize repeatability and technology transfer possibilities but also to establish a cost baseline and reduce the amount of effort that must be expended in process development.

Given these criterion and the fact that many actuation methods have been developed for the available foundry processes we used the MCNC MUMPs process[22] for a variety of reasons. The desirability for low mass to eliminate excess force (and thus bounce and excess wear) strongly suggests surface micromachining, and while a hybrid surface-bulk device offers intriguing advantages such processes are not currently foundry available except on a bulk undercut basis[23]. CMOS (with undercut post-processing) and MUMPs are the only processes that provide foundry surface micromachining. After some unsuccessful switch construction efforts using CMOS [24] the gold on chrome on polysilicon contact materials available in MUMPs were determined to be very advantageous and capable of making true 4-terminal relays; chrome is an excellent diffusion barrier and has low intermixability with both of its neighboring materials and the gold is an ideal low-force contact material.

The two major disadvantages in the MUMPs process are that there are no insulating structural layers, which makes drive/signal isolation very difficult, and the thinness of the gold, which makes lateral contact design difficult and can lead to low lifetimes. Both of these required additional processing on the MUMPs dice once they had been returned from the foundry but methods were developed that allowed both additional steps to be done with one mask and no equipment more complicated than a contact aligner, spinner, and hot plate.

The entire process used for fabrication is summed up in Table 1, where it can be seen that most of the fabrication occurs at a foundry. Only three steps, using one mask, are done as postprocessing to enable the relays to have electrical isolation and increase durability.

**TABLE 1. Fabrication process overview for microrelays**

1. Commercial foundry does MUMPs process, dice returned to user.
2. Electroless gold deposition: no mask, thickness controlled by time.
3. Photoresist patterning to allow nonconductive mechanical connections: one mask.
4. HF release and air or critical point drying

# 13 The MCNC MUMPs process

## 13.1 The basic fabrication process

The standard MUMPs process is quoted here from the MCNC MUMPs process web page [25], most of the figures have been removed for brevity but two images from the same site showing a cross section before and after release have been included at the end:

> "The MUMPs process is a three-layer polysilicon surface micromachining process derived from work performed at the Berkeley Sensors and Actuators Center (BSAC) at the University of California. Several modifications and enhancements have been made to increase the flexibility and versatility of the process for the multi-user environment. The process flow described below is designed to introduce inexperienced users to polysilicon micromachining. The text is supplemented

by detailed drawings that show the process flow in the context of building a typical micromotor.

"The process begins with 100 mm n-type (100) silicon wafers of 1-2 ohm-cm resistivity. The surface of the wafers are first heavily doped with phosphorus in a standard diffusion furnace using POCl3 as the dopant source. This helps to reduce or prevent charge feedthrough to the substrate from electrostatic devices on the surface. Next, a 600 nm low-stress LPCVD (low pressure chemical vapor deposition) silicon nitride layer is deposited on the wafers as an electrical isolation layer. This is followed directly by the deposition of a 500 nm LPCVD polysilicon film-Poly 0. Poly 0 is then patterned by photolithography, a process that includes the coating of the wafers with photoresist (Figure 1.2), exposure of the photoresist with the appropriate mask and developing the exposed photoresist to create the desired etch mask for subsequent pattern transfer into the underlying layer (Figure 1.3). After patterning the photoresist, the Poly 0 layer is then etched in an RIE (Reactive Ion Etch) system (Figure 1.4). A 2.0 µm phosphosilicate glass (PSG) sacrificial layer is then deposited by LPCVD (Figure 1.5). This layer of PSG, known as First Oxide, is removed at the end of the process to free the first mechanical layer of polysilicon. The sacrificial layer is lithographically patterned with the DIMPLES mask and the dimples are transferred into the sacrificial PSG layer by RIE, as shown in Figure 1.6. The depth of the dimples is 750 nm. The wafers are then patterned with the third mask layer, ANCHOR1, and reactive ion etched (Figure 1.7). This step provides anchor holes that will be filled by the Poly 1 layer.

"After etching ANCHOR1, the first structural layer of polysilicon (Poly 1) is deposited at a thickness of 2.0 µm. A thin (200 nm) layer of PSG is deposited over the polysilicon and the wafer is annealed at 1050C for 1 hour (Figure 1.8). The anneal dopes the polysilicon with phosphorus from the PSG layers both above and below it. The anneal also serves to significantly reduce the net stress in the Poly 1 layer. The polysilicon (and its PSG masking layer) is lithographically patterned using a mask designed to form the first structural layer POLY1. The PSG layer is etched to produce a hard mask for the subsequent polysilicon etch. The hard mask is more resistant to the polysilicon etch chemistry than the photoresist and ensures better transfer of the pattern into the polysilicon. After etching the polysilicon (Figure 1.9), the photoresist is stripped and the remaining oxide hard mask is removed by RIE.

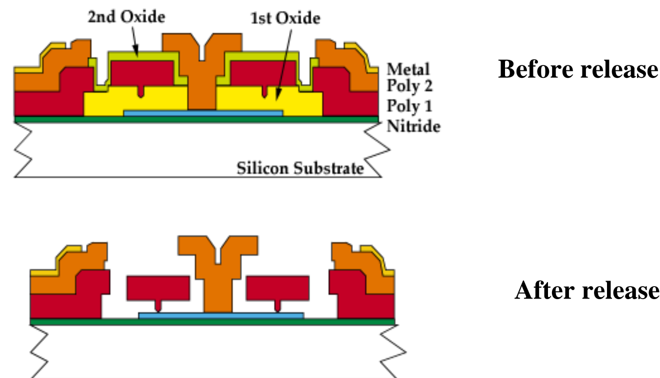
"After Poly 1 is etched, a second PSG layer (Second Oxide) is deposited (Figure 1.10). The Second Oxide is patterned using two different etch masks with different objectives. The POLY1_POLY2_VIA level provides for etch holes in the Second Oxide down to the Poly 1 layer. This provide a mechanical and electrical connection between the Poly 1 and Poly 2 layers. The POLY1_POLY2_VIA layer is lithographically patterned and etched by RIE (Figure 1.11). The ANCHOR2 level is provided to etch both the First and Second Oxide layers in one step, thereby eliminating any mis-alignment between separately etched holes. More importantly, the ANCHOR2 etch eliminates the need to make a cut in First Oxide unrelated to

anchoring a Poly 1 structure, which needlessly exposes the substrate to subsequent processing that can damage either Poly 0 or Nitride (see Section 2.3.3). The ANCHOR2 layer is lithographically patterned and etched by RIE in the same way as POLY1_POLY2_VIA.

“Figure 1.12 shows the wafer cross section after both POLY1_POLY2_VIA and ANCHOR2 levels have been completed. The second structural layer, Poly 2, is then deposited (1.5 µm thick) followed by the deposition of 200 nm PSG. As with Poly 1, the thin PSG layer acts as both an etch mask and dopant source for Poly 2 (Figure 1.13). The wafer is annealed for one hour at 1050C to dope the polysilicon and reduce the residual film stress. The Poly 2 layer is lithographically patterned with the seventh mask (POLY2) and the PSG and polysilicon layers are etched by RIE using the same processing conditions as for Poly 1. The photoresist then is stripped and the masking oxide is removed (Figure 1.14).

“The final deposited layer in the MUMPs process is a 0.5 µm metal layer that provides for probing, bonding, electrical routing and highly reflective mirror surfaces. The wafer is patterned lithographically with the eighth mask (METAL) and the metal is deposited and patterned using lift-off. The final, unreleased structure is shown in Figure 1.15. The wafers are diced, sorted and shipped to the MUMPs user for sacrificial release and test. Figure 1.16 shows the device after sacrificial oxide release. The release is performed by immersing the chip in a bath of 49% HF (room temperature) for 1.5-2 minutes. This is followed by several minutes in DI water and then alcohol to reduce stiction followed by at least 10 minutes in an oven at 110 C

FIGURE 31. **Cross sectional representative drawings of a micromotor fabricated by the MUMPs process before and after release by sacrificial silicon dioxide etching in hydrofluoric acid. Step shapes are representative, but approximate. (from [25])**



# 14 Electroless gold deposition

As discussed in the design of the lateral contact geometries, it would be very desirable to add additional gold to the relays, preferably in such a way that it would hang over the edge of the contact structures.

There are a number of available techniques that could be used to add gold to the contacts. Evaporation and other patterned techniques are undesirable in that they tend to eliminate or minimize lateral growth and overhangs, which are exactly what we want in order to get a good lateral contact. Electroplating is good in that the gold experiences approximately isotropic growth (leading to good overhangs), but bad in that it requires all the gold to initially be wired together. While others have demonstrated methods for wiring together structures temporarily and using mechanically breakable or fusible links to separate them after electroplating [26] it can quickly be seen that applications requiring large numbers of relays combined with the many parts in a relay leads to chips dominated by fusible link layout and very complex fusing routines. Thus it is highly desirable to have a method that adds gold isotropically without requiring wiring and biasing. The method used here combines a meta-stable solution of gold salts with a catalyst that activates at sites with an initial metal layer to accomplish maskless electroless plating.

## 14.1 Basic types of electroless plating

Electroless plating solutions can be grouped into two general categories: chemical replacement solutions and meta-stable solutions.

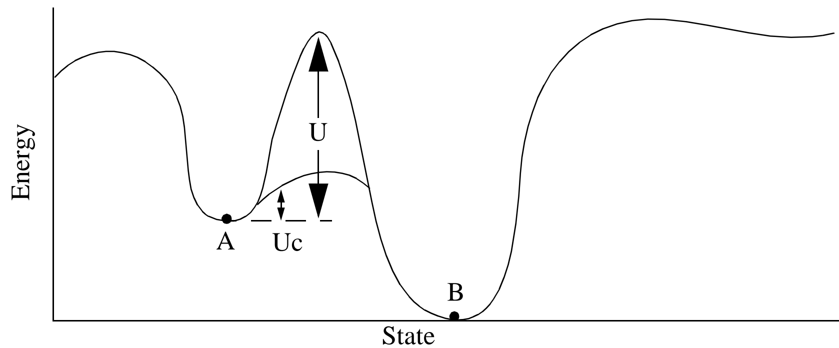
Many chemical replacement solutions, including some typically referred to as "electroless gold" [27], contain metal complexing solutions which can transition to a lower energy state by exchanging the gold they contain for a baser metal like tin or lead. These solutions are typically used for gold coating printed circuit boards or connector pins and accomplish the task very well by stripping off several microns of the metal present and depositing gold in its place (the process is self diffusion terminating at a depth of about 6 microns). As an experiment a MUMPs die was placed in the solution to observe the effects; the result was that the solution stripped the gold off the MUMPs die.

Meta-stable solutions, on the other hand, are supersaturated in at least one component and deposit it wherever conditions are right to propagate the appropriate reaction. Super-saturated electroless nickel solutions are an example of this, but supersaturated gold solutions are typically unstable. Thus, more complicated chemistry is needed.

## 14.2 Chemical meta-stability

Meta-stable solutions can be visualized with Figure 32. Point A may be locally stable but the solution would be at lower energy at state B, this creates a drive for the reaction from A to B to proceed. In this situation the state A is said to be in an energetic potential well or “col” and the depth of the col is U. Until a species has at least U energy it will be unable to exit the col and proceed from A to B, therefore the reaction does not proceed without outside energy even though, if each point were individually considered, it is energetically favored.

FIGURE 32. **Energy diagram for a meta-stable solution. Point A is locally stable, point B is globally stable and has less potential energy than A but is isolated by a potential peak of height U. Unless a species has at least U energy available it will be unable to go from A to B. Uc is the potential peak in the presence of a catalyst. To make the solution self patterning enough heat is provided to put the solution between Uc and U and a catalyst is provided that only adsorbs and works on certain target materials.**



## 14.3 Catalytic meta-stable gold deposition

If the value of U is below the available energy (typically thermal and interfacial) the A to B reaction proceeds spontaneously. This is often the case for supersaturated solutions involving gold, rendering them too unstable to use. If the value of U is above the thermal energy by a small enough value to allow surface energy to supply the rest, the A to B reaction will only occur spontaneously at the interfaces. This is often the case for electroless nickel or supersaturated sugar solutions and thus more nickel or sugar will grow in the presence of appropriate nucleation initiators like a crystal of the appropriate material. The larger number of free atomic bond sites at sharper edges causes these to be encouraging growth sites, similarly once a crystal begins to grow favorable sites are in the directions along which the crystal lattice has the most additional sites available -- and this is what causes dendritic growth.

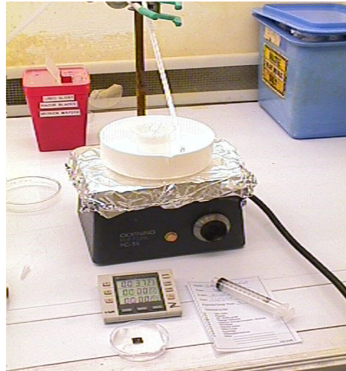
The solution used for gold electroless plating in this work is potassium aurous cyanide ($KAu(CN)_2$) and potassium cyanide (KCN) in a highly basic solution (KOH and water); this solution has a high U value and requires borane dimethylamide ($(CH_3)_2NH{:}BH_3$) as a catalyst, increased temperature, and interfacial energy to allow the reaction to proceed. Previous work from which this solution was derived [28] did not reveal all the ingredients used but the formulas rederived for the

present work are quite effective and deposition rates well above the maximums reported in [28] were achieved.

The stability of the potassium aurous cyanide solution is such that even with the catalyst (dimethylamide borane or DMAB) the temperature needs to be maintained above 60C in order for the gold to precipitate out as a coating on existing gold and other surface features with appropriate atomic energy functions. The processing is carried out within a teflon beaker sitting in a heated water bath, but other than periodic rehydration to combat evaporation no other processing is needed (Figure 33). The recipe given in Figure 33 shows that the molarities used in this solution can be fairly flexible and the specific amounts added for a typical successful solution give good starting points for further development. The order in which these ingredients are added are NOT arbitrary and should be followed.

**FIGURE 33. The setup for doing meta-stable catalytic gold plating requires a hot plate, water bath, beaker, and thermometer. A syringe facilitates**

**rehydration and solution storage. The basic recipe used at 73C is shown at right, see the text for more details.**



Additive gold bath ingredients in the appropriate order for adding to solution. Molarities are in parenthesis, quantities for the specific solution discussed in the text are in brackets:

1. KCN (0.01-0.02) [3.3mg]
2. KOH (set pH 13.5) [173mg]
3. Water [5ml]
4. Potassium aurous cyanide (0.01-0.02) [14.4mg]
5. DMAB (0.07-0.15) [20.6mg]

After the water is added it is best to wait until all the solids have dissolved before adding the gold salt to prevent premature precipitation. Once the first two ingredients are fully dissolved the solution is very stable. After adding the gold salt the solution is still exceptionally stable and can be heated or stirred to any extent desired. The depth of the potential well is such that even after the DMAB is added the solution is fairly stable and requires elevated temperature and appropriate surfaces to deposit. The reaction will still proceed at a very slow rate within temperatures encountered around the lab, however, so storage in a sealed inert container is recommended. A syringe is very handy for handling and storing these small amounts of liquid. The liquid itself is not particularly dangerous (other than being a strong base) but some of the ingredients are and handling of the DMAB, at least, should be done under a fume hood or in a high air flow area. Mixing the solution with acids is also to be avoided as it can release the cyanide in gaseous form.

The 5ml of solution made using the exact recipe in Figure 33 coated 3 chips thoroughly (including around the edges of two) before being sacrificed by running a process with some graphite dust mixed in to precipitate out as much gold as possible - the solution was found to still contain plentiful gold. The solution is tremendously reusable since almost 70% of the gold salt mass (many mg) is gold and the dice consume only micrograms each to coat although, as described below, it is not advisable to use a gold starved solution.

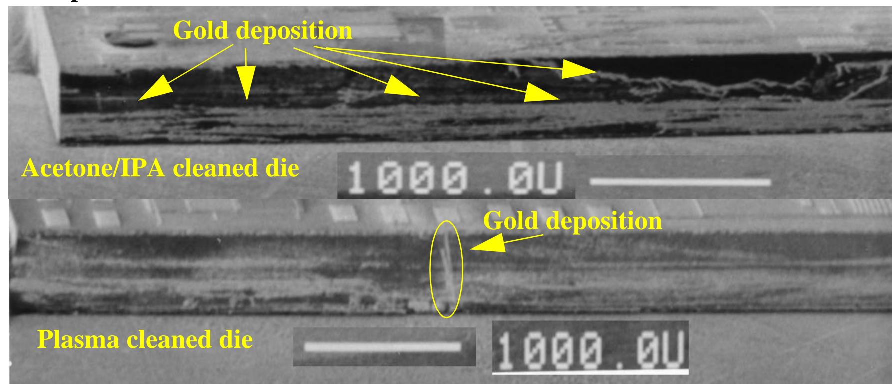
Deposition rate can be increased by increasing temperature, DMAB concentration, or pH. The pH stops increasing deposition rate above 13.5 so this was chosen as the operating point here. The DMAB has been shown to have linear proportional effect on the deposition rate up to at least 0.17 molar. Increasing DMAB concentration, however, also appeared to lead to rougher surfaces so it is not used extensively here.

Increasing the KCN molarity decreases the deposition rate until about 0.07 molar, and increases deposition roughness significantly (apparently by adhering to the surface and decreasing the reaction rate). It is also a necessary component, though, and cannot be eliminated entirely. Minimum molar balance quantities are suggested.

It is known that chops exposed to atmosphere develop organic contaminant layers. These appear to increase deposition probability. Chips processed with and without

oxygen plasma descumming showed increased deposition probability in dice without plasma descum. Die without cleaning grew substantial gold on the die sidewalls, die fresh from a plasma descum grew very little sidewall gold while descummed die again grew plentiful sidewall gold if allowed to sit for a week in dehumidified but unfiltered air (Figure 34). From this it is clear that good cleaning is important if minimal collateral deposition is desired.

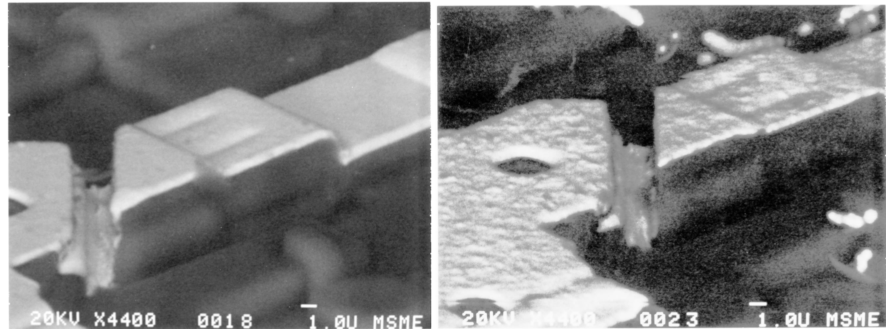
**FIGURE 34. Scanning electron micrographs of die edges after meta-stable catalytic gold deposition. The one on the top has a substantial gold layer (all of the light topography) due to air-carried organic contaminants, the one on the bottom shows little gold (only the vertical stripe) despite longer deposition processing after being oxygen plasma descummed. While not strictly necessary this sort of cleaning produces chips with less collateral deposition.**




As shown by scanning electron microscopy in Figure 35, the gold deposition from the exact recipe listed in Figure 33 (first die using this solution) is fairly smooth and conformal and provides a continuous gold layer over the edge and into the contact

gap. This was achieved without masking and with minimal equipment, nicely achieving the original goals for this processing step.

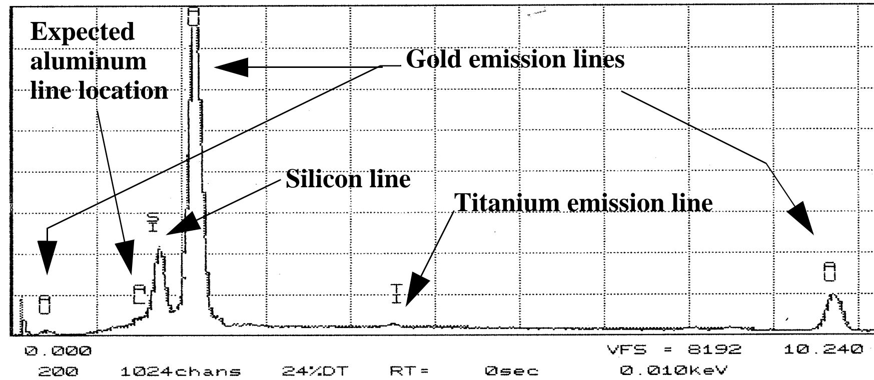
**FIGURE 35. Scanning electron micrograph (left) and electron backscatter image (right) of a contact gap subjected to electroless gold addition. The SEM photo shows texture while the backscatter image has high contrast between materials and shows good coverage. This particular gap was left with gold bridging it by MCNCs fabrication process.**




The purity of the gold deposited using this method was checked using energy dispersive X-ray spectroscopy (EDS) in an SEM. As the electron beam is scanned across a surface the electrons stimulate emissions of X-rays from the atoms impacted at wavelengths characteristic of the atomic elements involved. EDS spectra were done on gold deposited on both relays and CMOS aluminum metallization, the latter because metals are more easily detected by low energy EDS and this provided an excellent way of checking surface coverage.

**FIGURE 36. EDS spectrum of electroless gold on CMOS 0.25μm process metallization to check coverage. The CMOS metallization (which was spectroscopically examined separately) contained small amounts of titanium and silicon. Both of these diffuse into gold and small peaks of Ti**

**and Si can be seen amongst large peaks of Au. The absence of Al peaks shows that coverage was 100% for this sample, which was examined within a day of fabrication before the gold and aluminum could interact.**

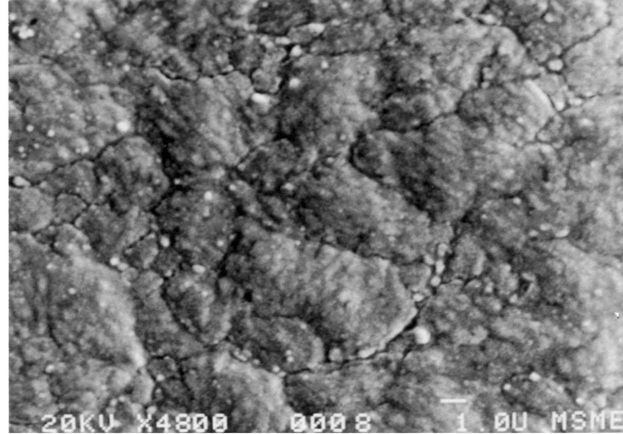



From the mention of deposition of gold deposition onto CMOS aluminum it should be obvious that gold is not the only metal which this solution deposits onto. As a sidenote it is worth mentioning that the process goes much faster with aluminum to plate and upon dropping a CMOS chip into the solution presented there is violent bubbling and immediate plating, within tens of seconds the aluminum surfaces will be heavily plated and the bath at least partially depleted.

The rapid reaction of gold solution with aluminum also gives a way to deplete the bath, and so characterization was done to establish the changes in gold deposition as the gold in the bath is depleted to check the effect this might have on relay preparation. Measuring the deposition rate was somewhat problematic due to the temperature instability of the plating bath and the high dependence of plating rate on temperature, but in general the plating rate did not vary enough to notice during the

depletion tests.The big change was in surface texture of the deposited gold. From the left side photo in Figure 35 it can be seen that the gold from a fresh solution bath was very smooth. After what was calculated to be approximately a 30% gold depletion the surface texture became more rough, Figure 37.

**FIGURE 37. Surface texture of electroless gold after solution has been depleted of approximately 30% of its gold supply. Scale bar shows 1μm.**




After another 30% depletion the solution grew gold with an extremely rough surface texture (Figure 38). This led to the decision to use solution only within the first 30% of its life to make relays in order to keep roughness from varying significantly. Extremely depleted solutions can lead to some very interesting gold growth structures, especially when grown fast on aluminum (Figure 39).

FIGURE 38. **Surface texture of electroless gold grown from a solution that was approximately 60% depleted. The roughness here is too unpredictable for reliable relay usage. Photo is same magnification as Figure 37.**

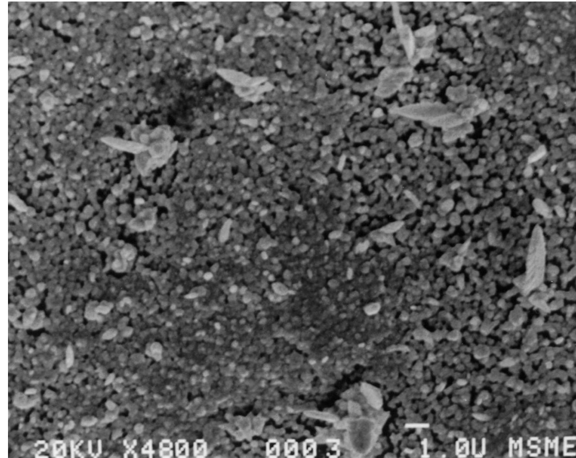
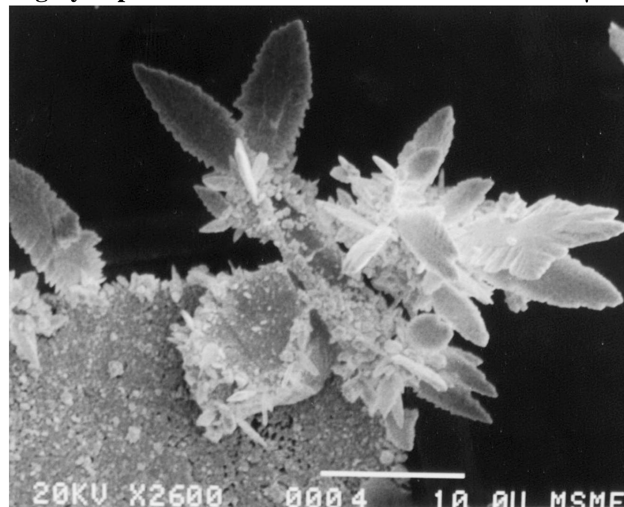



FIGURE 39. **Three dimensional gold “crystals” grown on an aluminum pad using a highly depleted electroless solution. Scale bar is 10μm.**



# 15 Insulating mechanical bridges

Since the MUMPs process builds structures exclusively out of conductors, additional processing must be done to provide electrical isolation between relay electrodes and the actuators. Isolation was provided by using photoresist mechanical bridges to connect the crossbar structure to the actuators (Figure 40). The bridges were realized by breaking the actuation drive line with a convoluted interwoven set of double-poly beams. When photoresist is spun on it connects the beams and provides mechanical connection without electrical connection.

**FIGURE 40. Microrelay structure (unreleased) with a discontinuity in the polysilicon support in the upper right. This convoluted structure provides plentiful adhesion area for photoresist bridging and has high yield.**

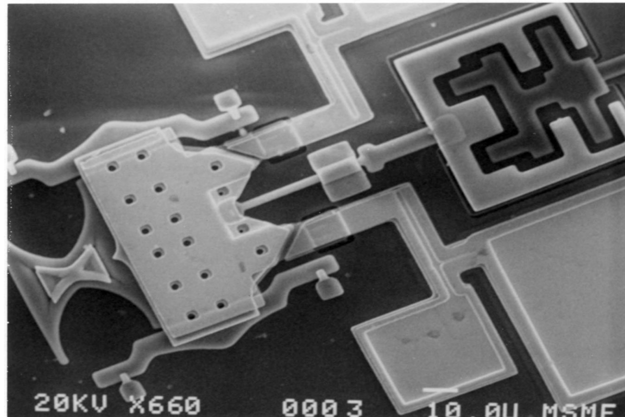
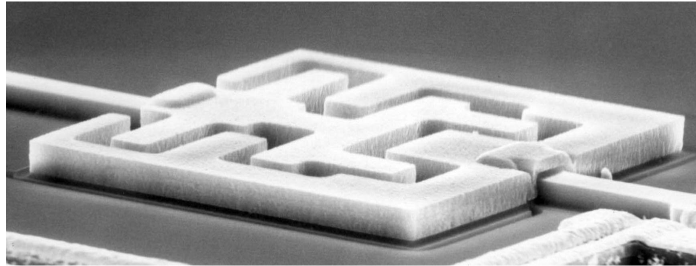


**FIGURE 41. Perspective view of the photoresist bridging structure, showing the ample surface area provided.**



Special care must be taken with the photoresist processing to enable the photoresist to survive the HF release step which follows. The surface was first cleaned with water and isopropyl alcohol followed by a dehydration bake at 120C. OCG 825 photoresist was then spun on at 4000rpm for 30 seconds. A prebake was done at 90C for 18 minutes. Exposure takes from 9 to 15 seconds in an I-line contact aligner, dosage was about 203.2 mJ/cm$^2$ at 28% relative humidity, decreasing by 35% as humidity goes to 32% (both at 21C). The sensitivity of the patterning in this step required monitoring of the environment and running test chips before each batch of device dice. The pattern was developed for 2 minutes in 2:1 OCG 934 developer, then rinsed in DI water and isopropyl alcohol. After a quick oxygen plasma descum the photoresist was cured at 170C for 10 minutes. Photoresist prepared this way survives HF release without degradation.

# 16 Dicing and sacrificial release

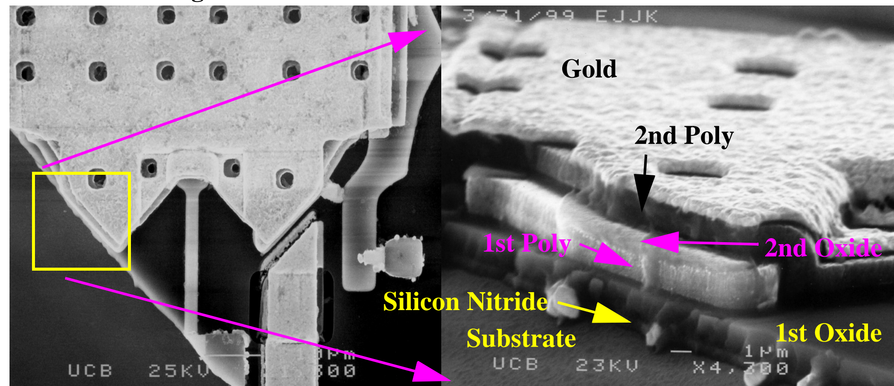
After the photoresist is patterned the dice are ready to be further diced and released. Any required dicing was done using a DISCO wafer saw by placing the dice on Blue Sticky Tape and dicing 90% of the way through the 1cm dice received from the foundry. The chips were then separated using teflon tweezers.

Release was done by soaking the chips in HF for 2 minutes followed by three or more water dips with agitation to dilute out the HF. If critical point release was desired the chips were next moved to methanol and moved to the critical point machine [29]. For air drying the last dip was isopropanol and the chips were dried standing on end on an absorbent clean-wipe with dry nitrogen blowing over the surface.

# 17 The finished product

The final result of this process is shown before release in the cross section of Figure 42.

FIGURE 42. **A cross section of the final processed relay, as seen looking into a cleave along the crossbar contact surface.**

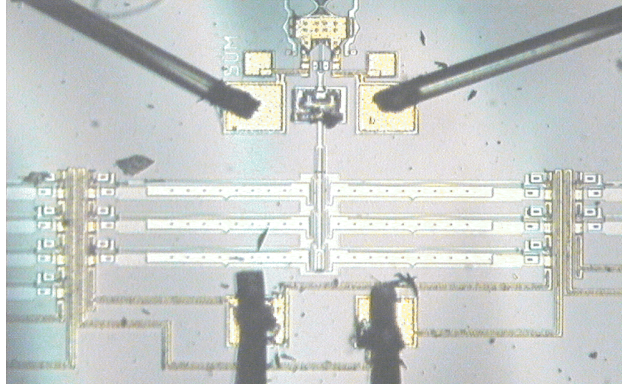



Yield, however, is an issue that needs to be addressed before any assertions can be made as to the viability of these relays for use in integrated systems.

## 18 Packaging

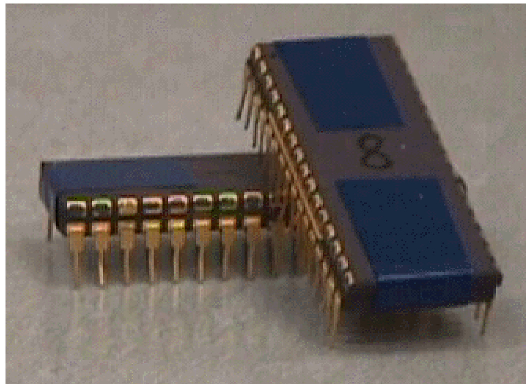
Packaging takes places as a series of steps. Die attach was done with colloidal silver paste, which dries at room temperature. Wire bonding was done with aluminum wire in a Westbond 800 thermosonic bonder and was extremely problematic as the pads had very poor adhesion to the substrate. Dozens of good relays had to be taken to the wire bonder to get a couple of bonded relays, although somewhat higher success rates were achieved with very high looping bonds (Figure 43).

FIGURE 43. **A die attached and fully bonded relay.**



Once the dice were attached and bonded some method had to be found to package the devices with their high looping bonds. The first method was simply to invert a similar open cavity package over the first and tape them together (Figure 44)

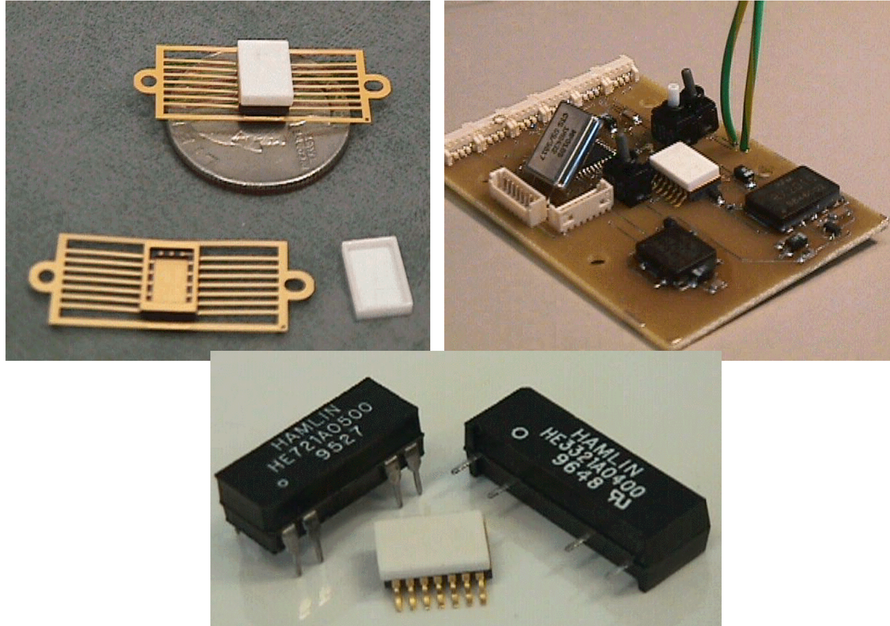
FIGURE 44. **Two unsealed packages made by closing identical open cavity ceramic packages together to leave a chamber in the middle. There is no sealing but the method is low temperature and provides effective protection from incidental handling.**



A more elegant and efficient method can be performed using commercially available lead frames and caps [30] as shown in Figure 45. In order to use this method care must be taken to procure packages which have a open cavity top, in this case the white ceramic tops also have a rim of high temperature epoxy which reflows and bonds above about 250C. Before packaging the silver paste die attach must be dehydrated, however, and this was done by an overnight exposure to vacuum (diffusion pump) in the Berkeley microlab “vacpack” machine, followed by a 30 minute dehydration at 150C. This method has been previously shown to yield long term operation at 2µtorr [31] despite the outgassing tendencies of silver paste so it was used here to dehydrate before sealing. In this case the sealing was performed by clipping the package together with a paperclip and placing it on a 260C hotplate for 20 minutes. Microscopic examination seems to indicate that the epoxy reflowed all the way around. A packaging experiment showed that the RTA could also generate sufficient temperatures to seal packages so if the packaging were done in the RTA the package could be filled with nitrogen or argon, although not enough chips reached this stage to test the effects of such packaging. The resultant surface mount package is strong enough to survive flinging around the lab; light, and tiny enough to make it hard to handle without tweezers (Figure 45).

FIGURE 45. **Photos of the final packaging solution. The left photo shows the elements and a finished package after some time at 260C. The photo on the right shows a surface mount packaged relay die nestling among other surface mount parts on a printed circuit board. The package contains four**

**working wire bonded microrelays, the entire circuit board is not much bigger than a quarter. Bottom photo compares a 4-microrelay package to 2 commercial single relays, the commercial relays have similar performance except that they can conduct more current (6-10 amps).**




# 19 Yield and process stability analysis

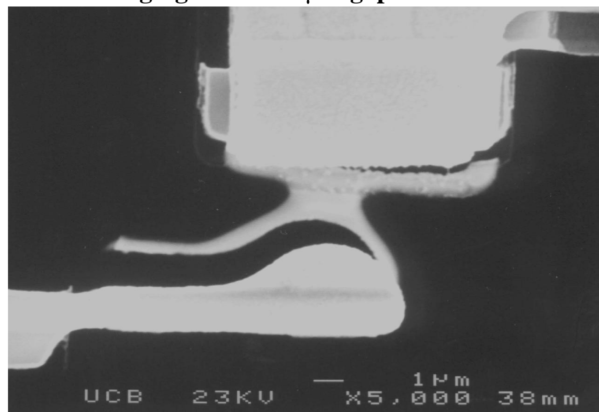
## 19.1 Step 1, commercial foundry processing

### 19.1.1 Gold bridging

Because the MCNC process uses a lift-off process it is not unusual for gold to remain bridging gaps that should be cleared of it. This is a very common occurrence and population counting shows that it happens to approximately 30% of the gaps, even when design rules are followed. Both gaps in a given relay must be clear

for the relay to work because the bridges are sufficiently strong to cause angle lockup (see the mechanical design chapter), leaving a relay survival rate for this mechanism of 49%.

FIGURE 46. **Gold bridging across a 6μm gap.**

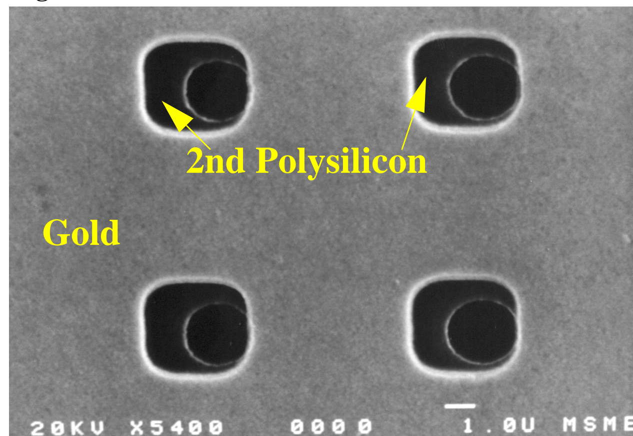



### 19.1.2 Misalignment

The MCNC foundry process uses a large single mask containing the patterns of all the dice on a given run and contact lithography. Achieving high quality alignment across such a large area with contact alignment can be very difficult and from a practical point of view their performance is very good. Nonetheless the typical layer to layer offsets, especially angular, can be very high relative to what most designers are used to in semiconductor processing. Designing structures to survive these misalignments is a necessity. Even with these design efforts, however, fabrication misalignment mortality is approximately 20% (this number is difficult to be

certain of because it is difficult to be certain that a failure should be blamed on misalignment). This number is down from early designs which had misalignment mortality rates in excess of 90%.

**FIGURE 47. A SEM showing a typical misalignment. The layout calls for a square 2µm hole in the second polysilicon in the center of a 4µm square hole in the gold.**




There also tend to be certain misalignments which, if greater than the thickness of the second oxide, cause all structures in a certain orientation (horizontal or vertical) to fail. On chips with this sort of misalignment all devices oriented a certain way will fail, meaning that system design within the MCNC process can require placing all devices in parallel orientation and fabricating two orthogonal copies. This systematic misalignment has accounted for approximately 6% of the failures of the newer designs, usually in systems which were not laid out with all devices in parallel orientation. This number is somewhat lower than the actual failure rate because

once this sort of failure is identified the devices oriented to fail tend not to get tested and don't become part of the failure statistics.

### 19.1.3 Destructive overetch

During patterning of the polysilicon layers the plasma etching is run longer than would be needed to cut through the nominal thickness of the polysilicon. This is done to cut through thicker areas where polysilicon goes over topology, but also results in plasma etching under the photoresist which thins beams. While MCNC does bloat their polysilicon masks in an attempt to compensate a nominal 2μm drawn beam was observed to vary from 1.70 to 1.89μm over three runs. Occasionally there will be a section of a chip (usually no more than a quarter of the 1cmx1cm die) where beams of 3μm or below will be overetched so badly that they are no longer usable. When this happens it destroys relays because the actuators wither and adhere to the surface when driven. This only happens on a few dice and so accounts for approximately an 8% failure rate overall.

## 19.2 Step 2, electroless gold deposition

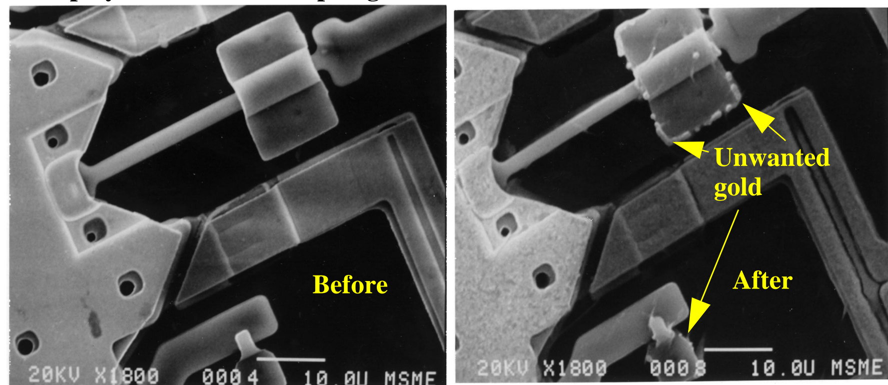
### 19.2.1 Unwanted gold deposition

Although flexible and simple the electroless gold deposition technique is not without its flaws. The use of a catalyst which adsorbs to metals makes the energy of deposition lower for metals and thus causes selective deposition on the metal sur-

faces. From a practical point of view this means that ideally with a system-wide energy level above $U_c$ but below U in Figure 32 deposition will only occur on surfaces already patterned with metal. Unfortunately the energy distribution over the surface of the chip is not constant and contaminants, junctions between materials, and angles in material (which present increased surface energy concentrations) all can act to cause deposition where it is not wanted by raising the local energy levels. Typically this means that extra deposition happens at the edges of cut polysilicon where it meets other materials as shown in Figure 48. This happens even on chips which have been carefully cleaned, and can occasionally cause mechanical difficulties if the particles jam guides or short actuators, leading to a mortality rate of about 10%.

**FIGURE 48. Nonideal electroless gold deposition caused by local energy concentrating factors. The same structure is shown before (left) and after (right) electroless gold deposition. The SEMs show that the gold deposited not only on the metal that was present (most clearly visible on the crossbar), but also along the edges of the actuation rod guide and the**

**sliding guide visible near the very bottom of the SEMs where RIE cut edges of polysilicon make sharp angles.**

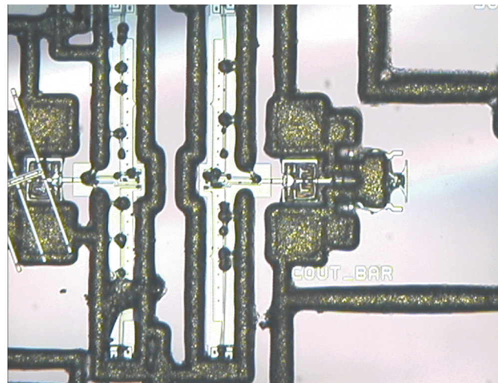



### 19.2.2 Single structure massive overplating

Another phenomenon observed with the electroless gold growth is massive overplating of single structures. In this behavior a single connected structure somehow grows gold at a massively enhanced rate. The mechanism is unexplained although since it typically effects large areas of electrically interconnected devices and may not affect a structure within the effected area which is electrically isolated the cause is presumably electrochemical in nature. This phenomenon causes a mortality rate of approximately 21%.

FIGURE 49. **A pair of relays practically buried in gold. Most of the chip had only a micron of gold deposited, thickness on this structure appears to be**

**approximately 3µm. The reason for this sort of irregularity is not understood.**




## 19.3 Step 3, photoresist processing

### 19.3.1 Generation of "goop"

The involvement of photoresist at this late stage of processing, and especially the generation of photoresist designed to withstand HF, tends to leave some unintended photoresist around the chip. This photoresist can stick down or hold open a relay and accounts for a 2% mortality rate.

### 19.3.2 Poor lithography

Poor alignment and over or underexposure typically accounts for less than a 1% failure rate. This number is very low because many dummy samples are run each day before any real samples and any time the lithography looks bad it is stripped

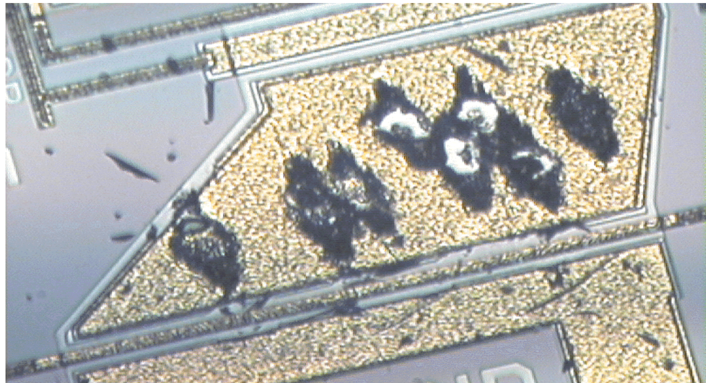
and redone. To be blunt: there are so many other failure mechanisms earlier on that this step had to be developed to a very high yield level before functional samples could be made with reasonable regularity.

## 19.4 Back end - Bonding and packaging

### 19.4.1 Bond pulling

One problem that seems pandemic to certain fabrication runs through MCNC is that wirebonding gets a miserable success rate. In the structures tested here is was not unusual to have 80% of wire bonds pull up during bonding.

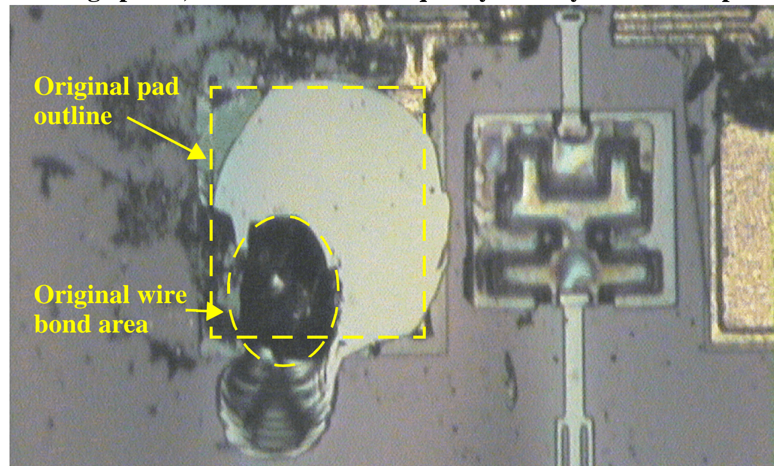
**FIGURE 50. Poor wirebond adhesion was pandemic on many chips. Here 8 wirebonds in a row pulled off of a test area made of the layers recommended for bond pads. The wires themselves remain attached to the gold, it is the gold or other layers that delaminate.**



The problem is more serious when it is considered that the failure is not the typical type where a bonded wire pulls off the metal. Rather the failures usually occur very

destructively as the entire bonding area separate from the die, often taking large surrounding areas with it.

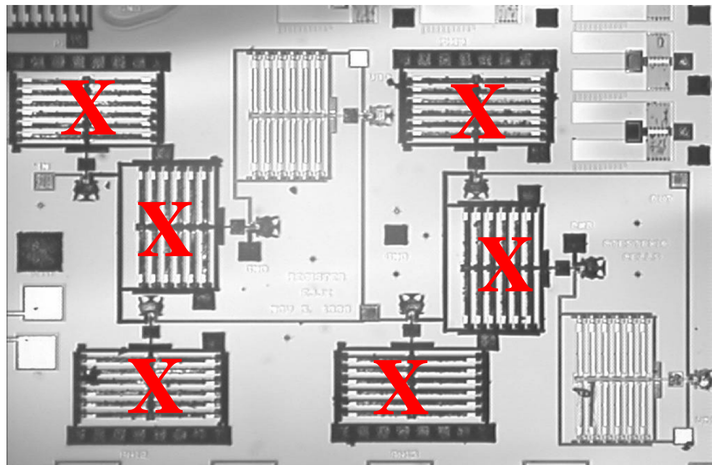
**FIGURE 51. Closeup of the location of a former bond pad. The bond pad itself has separated from the die, taking two layers of polysilicon and the nitride underneath with it. Given that the nitride has peeled away in a rather large patch, this would seem to qualify as a layer adhesion problem.**




## 19.5 Overall yield and yield improvement

Combining all these factors we get a non-systematic before-bonding yield for current devices of about 24% versus 6% for earlier designs that weren't so fault tolerant. Even without the orientation dependent failure mechanisms (so assuming the correct direction is chosen of that multiple orthogonal copies are made) a 24% yield rate means that a system of only 4 relays will work only 0.33% of the time, or about once for every 300 full systems fabricated.

**FIGURE 52. A logic latch cell where 2 out of 8 relays survived. This sort of yield makes building large systems problematic. The four relays oriented horizontally are locked up by misalignment, the remaining two broken relays were overplated. Nonfunctional relays are blackened by current overloading while conducting probe station tests to determine yield and cause of failure.**



Some methods to improve this are suggested by the descriptions of the problems themselves. Going to stepper based lithography should reduce the misalignments and eliminate the angular lockup failures typical of whole wafer contact lithography. Changing the gold deposition from lift-off followed by postprocessing to a single step evaporation based deposition with a planetary mount followed by chemical etch may eliminate not only the electroless deposition difficulties but also much of the bridging, although this requires some process development. Even using very conservative estimates for the improvements, yields of 70% seem possible with these changes. This would change survival rate of a four relay system from 1

in 300 to 1 in 4 with fairly minor process changes. Further development would be needed to get yields appropriate for much larger systems.

# 20 Conclusion

In conclusion we can say that although much useful process development has been done, much more needs to be done to provide higher yields and enable larger systems. While the steps used here are versatile and quick, lending themselves to research applications, they also lack much of the stability of a main line fabrication process and therefore the yield.

Two of the process steps, electroless gold plating and photoresist insulating bridges, were developed to overcome deficiencies in the capabilities of the available foundry processes for building relays. Both achieve their goals with minimal added difficulty, although yield issues remain for the electroless gold plating.

# 21 References

[22] MCNC Corp, Now Cronos Integrated Microsystems, Inc. 3021 Cornwallis Road, P.O. Box 12889, Research Triangle Park, NC 27709-2889, USA; Ph: (919) - 248 - 1983 FAX: (919) - 248 - 1434; http://mems.mcnc.org

[23] CMP [Circuit Multi-Projects]; 46 avenue Felix Viallet; 38031 Grenoble Cedex; France; FAX: +33 4 76 47 38 14; email: cmp@imag.fr; http://tima-cmp.imag.fr/CMP/CMPInfo.html [english version]

[24] Kruglick, E.J.J.; “Aluminum hinges and Electrostatic Actuators in Standard CMOS,” Masters thesis, available for reorder through UMI; 300 North Zeeb Road, PO Box 1346, Ann Arbor, MI 48106-1346, USA; Ph: (800) 521-0600; http://www.umi.com

[25] http://mumps.mcnc.org

[26] Kovacs, G.T.A.; “Micromachined Transducers Sourcebook” McGraw-Hill, 1998

[27] “Gold Plating Solution - Bright Electroless Gold”; Transene Company, Inc.; Danvers Industrial Park; 10 Electronics Ave; Danvers, MA 01923; Ph: (508) 777-7860

[28] Okuna, K; Otsuka, K; Hattori, N; Torikai, E; “Auto catalytic electroless gold plating,” The Electrochemical Society Proceedings, Vol 94 No 32, pp 164-185

[29] Mulhern, G.T.; Soane, D.S.; Howe, R.T.; “Supercritical Carbon Dioxide Drying of Microstructures” In Proc. 7th International Conference on Solid-State Sensors and Actuators (Transducers ‘93), June 7-10, 1993, Yokohama, Japan pp 296-299

[30] Spectrum Semiconductor Materials, Inc.; 2027 O’Toole Ave.; San Jose, CA 95131; Ph (408) 435-5555 Fax (408) 435-8226; http://www.spectrum-semi.com

[31] Method developed by Kevin Flynn, passed along by Jim Bustillo, UC Berkeley microlab

# CHAPTER 4 Actuator Design

*"...it's automatic; it's SYStematic; it's HYDROMATIC!"*

-from the movie Grease

## 22 The importance of actuator performance

The choice and implementation of actuator must be driven by the properties of the contact physics problem. It's also highly desirable to have a structure which can be driven by conventional electronics without high voltage adaptation. One of the most important factors for this experimental work, however, is simply robustness -- it is very difficult to troubleshoot a system if the actuation is only marginally adequate. For the purposes of developing micro-relay technology providing low power actuation has been completely ignored in favor of robustness and an ability to determine the appropriate design trade-offs. From previous discussion of the properties of microcontact two elements stand out which become primary design drivers for MEMs actuator design.

The first design goal is to produce raw force. We have seen that contact performance is directly influenced by the force applied. From this and basic calculations applied to the design equations (EQ 3) and (EQ 9) with the resistivity $2.3 \times 10^{-8}\ \Omega m$ of gold and the approximate universal hardness $174 \times 10^{7}\ N/m^{2}$ (1.74 GPa) we can calculate that a micronewton of force (if it causes plastic deformation) sets a resistance of one half ohm. Thus we need forces in the micronewton range to establish contact of the quality desired for conventional relay usage.

The second design goal is an actuator which will apply a greater force to separate the contacts than to close them. Contact separation force will often be approximately a third of the closure force(EQ 2), thermal softening and the surface energy force (EQ 34) can require more force to open a relay than was applied to close it. In this sense many actuation designs become insufficient due to the low return force provided by biasing springs. Until we have designed and thoroughly characterized the actuators we can't do the appropriate experiments to determine the relative importance and the exact separation force we should expect, but from the first we know that the separation force will sometimes be greater than the closing force.

These main considerations must be coupled with the design requirements to remain within the microrelay regime of low mass (to avoid bounce) and small gap (to avoid arcing).

Since we wish to use a standard foundry process without any assembly we will only consider electrostatic and thermal actuation.

# 23 Choosing an actuation method

## 23.1 Electrostatic actuation, basic considerations

Electrostatic actuation is common in MEMs. The high force requirements for microrelays, coupled with the desire for small masses in the actuators, makes it difficult to solve the problem with any electrostatics other than gap-closing actuators operating at pull-in. Even with electrodes in pull-in, large arrays such as those in [32] would be needed to apply sufficient forces laterally. In most available foundry processes such structures typically have significant problems with stiction. Single large electrostatic plates can be used to provide adequate force, and have been in previous work such as [33] and [34]. These large parallel plates tend to increase mass and may result in bouncing as well as proposing a significant challenge to design such that assembly is not a problem.

An additional concern is that electrostatic structures lack any way to propel themselves apart, amd so separation must be done by a bias spring. While such bias springs do not have to be weaker than the closure force it is a considerable exercise in design and operation to provide a system with greater separation force than closure force.

This is not to say that electrostatic actuation cannot produce microrelays! Some such devices were built during this work. Designing such actuation simply places more design requirements on the devices being designed and typically leads to larger layout, lower yield, and lesser performance.

## 23.2Thermal actuation

### 23.2.1 Basic considerations

Thermal actuation has been shown to provide more than enough force to meet our design goal of tens of micronewtons in multiple foundry processes and in a reliable fashion [35]. The actuators in question are typically compact and low in mass as well.

Additionally, thermal bimorph structures have a useful asymmetric force application profile. As heat is introduced thermal actuation overcomes internal springs to provide force to the relay. As the system cools the separation force is equal to a recovery force as strong as the applied force due to heating plus the internal spring force which was overcome to close the relay in the first place. While many actuators have some sort of built in return spring, the linear response of thermal actuators combined with built-in returning springs ensures that the return force will always be greater than the closure force (unlike possible pull-in scenarios with electrostatics). This is accomplished with compact structures that have low mass.

Thus thermal actuators appear to meet our initial requirements. In depth analysis and testing, however, is required to verify linear behavior, characterize reliability, and determine what forces will be involved in the contact.

### 23.2.2 Origins of thermal actuation

Thermal actuation has numerous advantages. It gives a high force with predictable and linear behavior at the cost of high power required to heat up the system. In order to establish linearity throughout the system to justify later characterization, we must begin with the basic origins.

The two primary energies related to atomic separation distance *r* are the repulsive energy:

$$E_{repulsive} \propto \frac{1}{r^m} \qquad \text{(EQ 19)}$$

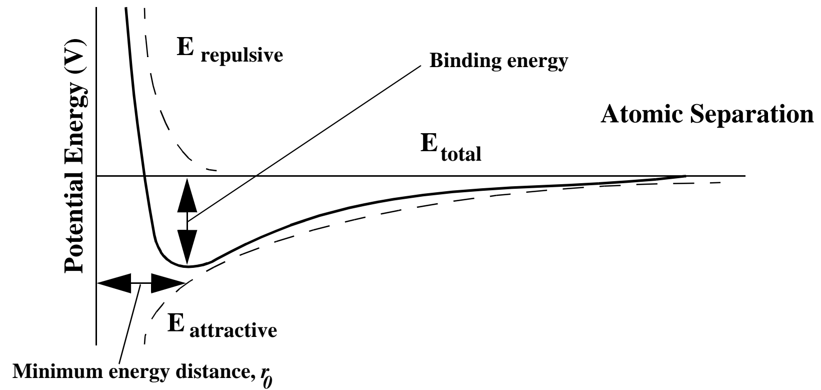
where *m* is approximately ten, and the attractive energy:

$$E_{attractive} \propto \frac{1}{r} \qquad \text{(EQ 20)}$$

which dominates at long range. A plot of the overall energy of two neutral atoms in a solid with respect to separation (discarding gravitation, which is insignificant) can be made by summing these two forces (Figure 53).

**FIGURE 53. Energy versus atomic separation for atoms in a solid and some of the related values. The minimum energy distance is the separation the atoms will have at 0K, and at this location they have negative potential**

**energy related to free atoms by the binding energy. Higher binding energy is equivalent to higher melting and boiling points,**




This plot can tell us a lot about a substance[36]. The binding energy, for example, gives a direct measurement of the energy required to disassociate a solid and as it increases the melting and boiling points increase. We can also use this plot to derive the elastic modulus of a material by using the fact that force is the gradient of energy with displacement. In order to find the value of potential energy *V* (from Figure 53) with respect to displacement one can use a Taylor expansion:

$$V(r_0 + \Delta r) = V_{(r_0)} + \left(\frac{dV}{dr}\right)_{r_0} \Delta r + \frac{1}{2}\left(\frac{d^2 V}{dr^2}\right)_{r_0} \Delta r^2 + HOT \quad \textbf{(EQ 21)}$$

where *HOT* refers to the presence of **h**igher **o**rder **t**erms. Assuming that the value of $\Delta r$ is "small" (a 1% strain or $\Delta r/r$ of 0.01 makes nonlinear terms insignificant and yet is also a very large deflection, so "small" is an appropriate assumption) we can

discard the higher order terms. We also note that $dV/dr = 0$ at $r = r_0$ so for deriving the elastic modulus we can rewrite (EQ 21) as:

$$V(\Delta r) = V_{r_0} + \frac{1}{2}\left(\frac{d^2V}{dr^2}\right)_{r_0} \Delta r^2 . \quad \text{(EQ 22)}$$

And since the atomic force is the negative of the slope of the potential energy we can get the force $F$ directly by:

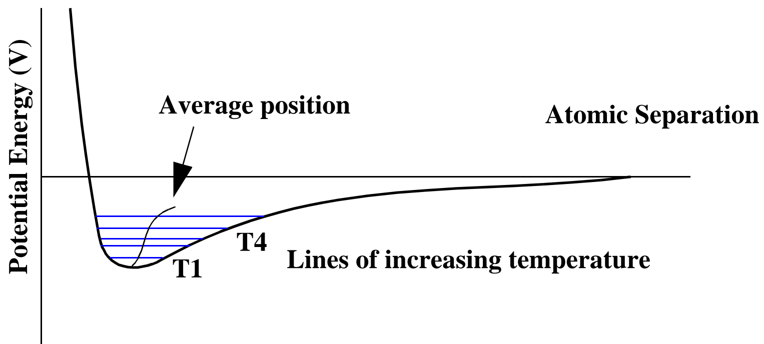
$$F = -\frac{dV(\Delta r)}{d(\Delta r)} = -\left(\frac{d^2V}{dr^2}\right)_{r_0} \Delta r \quad \text{(EQ 23)}$$

and we can see that the force $F$ is linearly related to the displacement $\Delta r$. When this atomic elastic modulus ($-(d^2V/dr^2)$) is integrated over an area being deformed it forms the Young's Modulus $E$, which is used extensively for macroscopic analysis of the thermal actuators. We will see how this is important in section 24 after we have derived more behavior of thermal expansion, for now it is enough to establish that a linear term can be used to characterize this behavior.

Thermal expansion can be intuitively visualized by considering that increasing temperature adds kinetic energy to each atom. In order to visualize this within the context of Figure 53 we must remember that the atomic position is a probability governed by quantum mechanics; as the energy level increases the average separa-

tion is higher because the probability of being farther apart increases faster than the probability of being closer together (Figure 54).

**FIGURE 54. Atomic separation diagram with energy levels of different temperatures. The quantum mechanical position of the atom is spread over the resulting energy range and the effective atomic separation is the average position. Because of the asymmetry of the energy diagram the average position moves outward with increasing temperature. This particular diagram shows a very high temperature change for clarity.**




Thus we expect to see most of the important responses of a material from its' atomic energy diagram. A deep, sharp plot has a high stiffness, low temperature coefficient of expansion, and high melting point. A shallow, gentle curve has a low stiffness, high temperature coefficient of expansion, and low melting point. This can be generalized and means that if one has some of the numbers available for a material, others can be estimated. We can also see that the thermal expansion of materials is a basic principle stemming from quantum mechanics and may be considered a new "natural length" which resists any deformation with the full material

modulus (which is effected by temperature). This is why mechanical expansion analysis is based on relative temperature changes and not absolute temperature.

### 23.2.3 Design considerations of thermal expansion

Thermal expansion has a number of important advantages stemming from its functional mechanism. Since the expanded condition can be considered a natural length the material exerts all of its elastic modulus against any restraint, this means that for the design scales and materials available to MEMS we can easily produce forces in the millinewton range - far higher than by most electrostatic actuation. Additionally, this force can be generated in a very small area.

Perhaps more important is the fact that when the heat is removed thermal actuators return to their previous shape with all the considerable force that they originally used to deform. Indeed, practical applications use biasing springs which typically act against the actuation such that initial actuation is the thermal force minus the internal spring forces and return actuation is the thermal force plus the internal spring forces. On a practical level this means that thermal actuators for relays by nature have a somewhat higher opening force than closing force, and this difference can be tailored. From the relay design point of view this is very convenient and in the work presented here thermal actuation is used as the mechanism of choice.

The main disadvantage of thermal actuation is that it is very power inefficient. The process of heat generation using current is relatively smooth, but the amount of input power required is very high and the thermal energy is dissipated to the environment, requiring constant energy input. Additionally, there is a lengthy cooling time associated with releasing a thermal actuator which defines the speed at which a thermally actuated relay can operate.

# 24 Thermal actuator design and characterization

## 24.1 Thermal actuator design

The empirical performance of various polysilicon thermal actuators have been considered in a number of different places, notably by Comtois [35] (section 3). Most thermal actuators employ the same basic principle of using heat induced material expansion to provide force and displacement.

Unconstrained macroscopic thermal expansion can be expressed by the relation:

$$\Delta l = l \cdot \kappa \cdot \Delta T \quad \text{(EQ 24)}$$

where $\Delta l$ is the change in length, $l$ is the overall length of the element in question, $\kappa$ is the material's thermal coefficient of expansion (derived from the previously described quantum behavior and normally indicated with $\gamma$, $\kappa$ is used here to avoid conflict with the surface energy $\gamma$), and $\Delta T$ is the change in temperature over which

the elongation is measured. This lengthening comes directly from the increasing space between the atoms of the material. The interatomic length increases as the increasing thermal energy changes the equilibrium point between the long range attractive and short range repulsive atomic forces; and therefore ideal thermal expansion is completely elastic [37].

The exact opposite case, fully constrained thermal expansion, can also be considered. Since the atomic expansion is entirely an elastic phenomena constrained thermal expansion can be considered as a simple compression of the material by the same distance $\Delta l$. Thus the force exerted by a perfectly constrained heated bar will be:

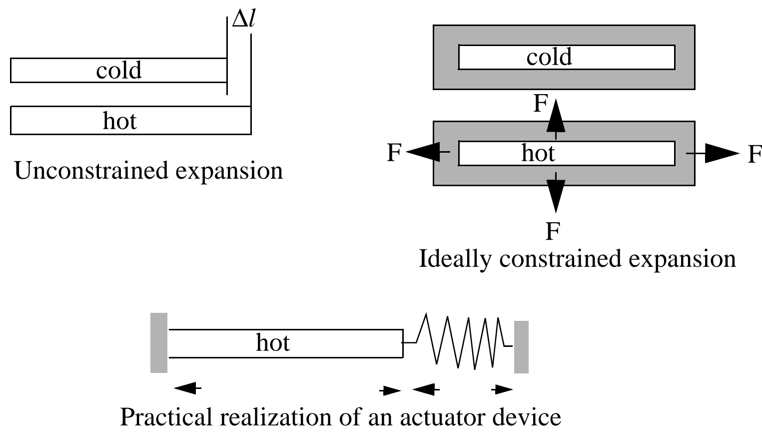
$$F = l \cdot \kappa \cdot \Delta T \cdot A \cdot E = A \cdot E \cdot \Delta l \quad \text{(EQ 25)}$$

where $F$ is the force and the $A$ and $E$ are the area of the bar and the Young's modulus of the bar respectively.

Unconstrained thermal expansion, however, is of limited use since it provides motion but no work. Fully constrained thermal expansion, if practically realizable, would be similarly limited because it provides force but no motion. In most practical systems the reality is somewhere in between and in order to do meaningful design an understanding of the force/motion trade-off is required (Figure 55). By allowing the thermally expanded beam to provide force against some sort of spring,

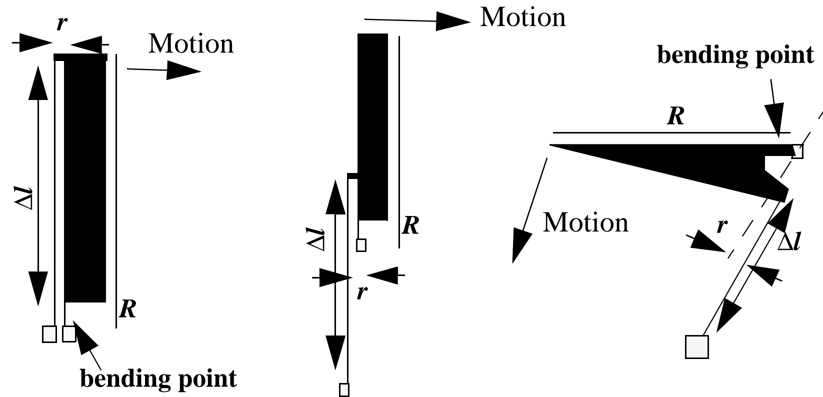
both force and deflection are extracted from the heating power, converting it to useful work.

**FIGURE 55. Comparison of unconstrained, fully constrained, and typical realization of a thermally expanding beam actuator. The first gives no force but maximum motion, the second gives no motion but maximum force, the third gives both in lesser amounts but does work.**




Complications arise when attempting to predict the output force and deflection of a practical thermal actuator. Typical thermal actuators use different materials (bimetal, bimaterial actuators) or differently heating geometries (bimorphs) in the design of a compliant structure that converts the thermal expansion to longer travel but weaker force[38]. Many of these devices are based on lever or moment principles (Figure 56) and can be generalized to networks of simpler structures [39].

FIGURE 56. **Some example thermal bimorph geometries, all with equal force amplification in the linear actuation regime for equal R/r. Two geometries change direction of motion by 90 degrees, one by 180.**

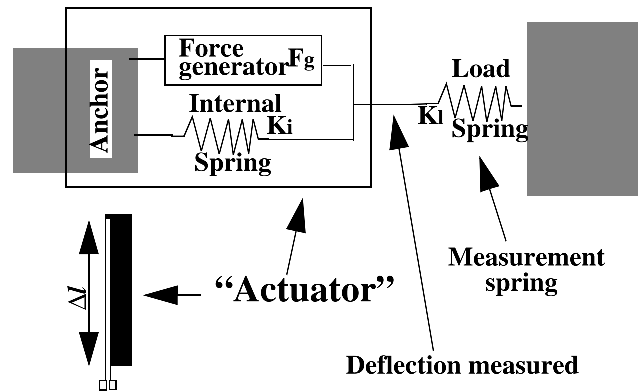



## 24.2 Characterization for implementation

As a practical matter, then, the realized “thermal actuator” as described in [36] has its own internal parasitic springs as part of the structure that turns small motion into larger motion. These internal springs are important because if they are stronger than the external spring they determine the motion, and at best represent a parasitic load on the system. Thus, force measurements to date using external reference springs provide a measurement of an available force/deflection point in a complicated design space and not measurements of an absolute force production level. This can be useful if the important part of the system is its performance against a given bias spring, but against a changing load (such as a closing contact) the information is less useful. As the contact closes the other springs in the system get less important

as the motion goes to zero. In an effort to find this “zero-travel” force a linear superposition model was created (Figure 57).

**FIGURE 57. Linear model of actuator and load for characterizing thermal bimorph structures. Variables used for the various springs and the generated force are shown.**




If the linear model of Figure 57 is considered adequate an equation can be written for the observed deflection $x$ in terms of the generated force $F_g$:

$$F_g = xK_i + xK_l \qquad \text{(EQ 26)}$$

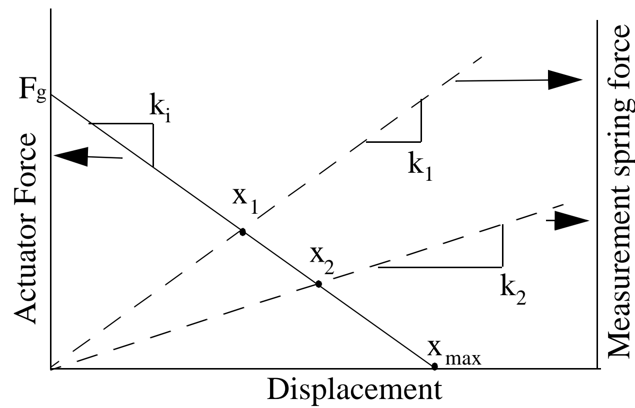
and this equation can be used to design an experiment which will give the zero-travel force. If we use the fact that internal force production will be identical for identical temperature changes (EQ 25) and therefor identical power inputs in similar structures, we can determine the value of $K_i$ by testing two structures at the same power but with different load springs as shown graphically in Figure 58:

$$F_g = x_1K_i + x_1K_{l1} = x_2K_i + x_2K_{l2} \qquad \text{(EQ 27)}$$

$$K_i = \frac{x_2 K_{l2} - x_1 K_{l1}}{(x_1 - x_2)} \quad \text{(EQ 28)}$$

Here the values of $x_1$ and $x_2$ are directly measured from the experiment, and $K_{l1}$ and $K_{l2}$ can be determined from basic spring theory and beam measurements.

**FIGURE 58. A graphical representation of the thermal actuator characterization problem and the solution used here. The solid line represents the performance of a thermal actuator at a given power, with the completely constrained force $F_g$ at one extreme and the unconstrained (zero force) motion at the other. Different measurement springs intersect the performance at different points, limiting their characterization value. If multiple springs are used at the same power, however, the internal spring constant $k_i$ can be found and the system performance can be calculated for any load.**




Once $F_g$ and $k_i$ are known the system in Figure 57 is fully characterized such that against any given load both force and motion may be calculated.

For experiments the thermal actuator from [40] was used. The actuator was hooked up to a motion amplification testing rig as in [36] to increase resolution and a gauge was put adjacent to the measurement indicator (Figure 59). The indicator and gauge

was magnified to a video screen where tape was put over the image and marked to record the gauge spacing and the motion of the indicator. By doing this and measuring multiple gauge spacings along with the indicator motion separate error bars were generated for each series of measurements. All of these steps were necessary to reduce the error in x values to around 100nm. This error reduction was needed to reduce the error bars in applied $F_g$ to the levels shown.

**FIGURE 59. Thermal actuator test structures with motion indicators and gauges. Calibrated SEM measurement showed that the drawn 2μm structures were 1.70μm wide in the open, growing to 1.84μm as they approached anchors or other structures.**

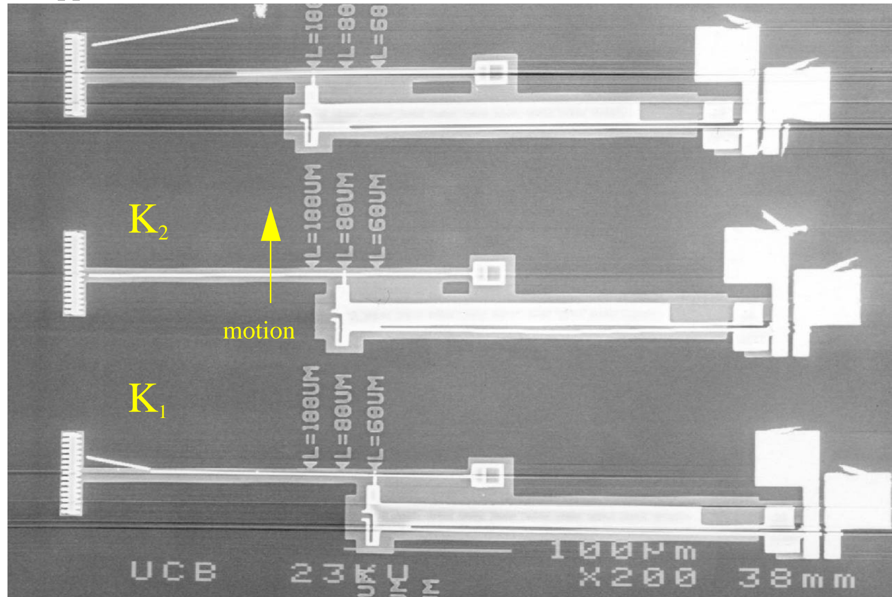
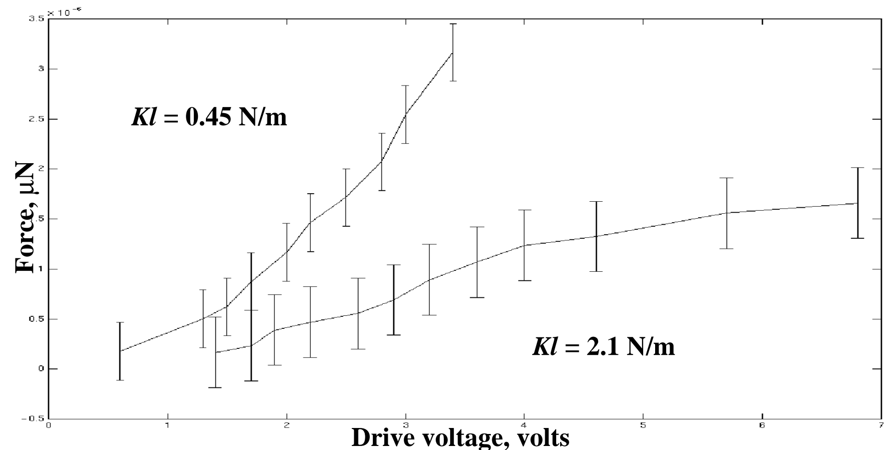



Figure 60 shows the raw data from the test device of Figure 59 with 2 different external springs of 0.45 and 2.1 newtons per meter. The calculated internal spring

constant according to (EQ 28) is 6.0 newtons per meter. Since the internal spring constant is higher than both the external spring constants the deflection should be dominated by internal spring constants and we would expect inconsistencies in the two traces of Figure 60. It can be seen that the two sets of data separate as the test progresses and above approximately 1.5 volts the error bars of the two sets of data no longer overlap. This illustrates the behavior shown in Figure 58 and is the result of the different spring constants involved. The divergence will grow worse as $K_l$ becomes larger in relation to $K_i$ and as the values of $K_l$ involved grow further apart.

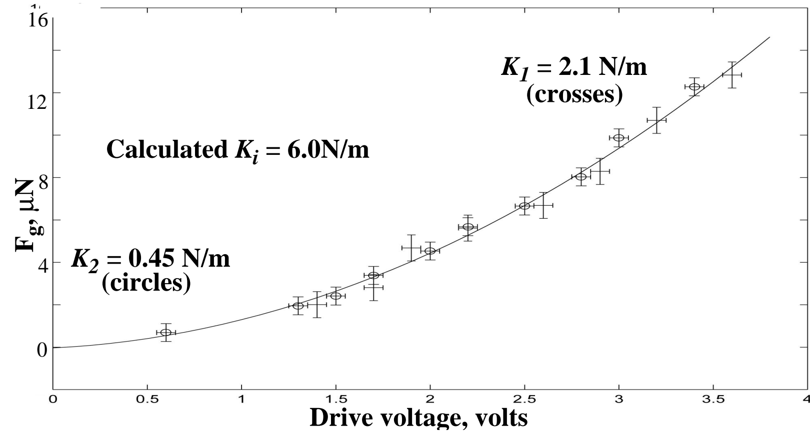
**FIGURE 60. Raw data from using the same test structure against 2 different external springs. Spring indicated force versus drive voltage is shown for external springs of 0.45 and 2.1 N/m.**




The extracted value of $K_i$ was used to calculate $F_g$ from both curves in Figure 60, the result is shown in Figure 61. We expect that with $K_i$ and $K_l$ calculated out the values of $F_g$ from the two experiments should correspond, and they do. This dem-

onstrates that the linear model allows for the effective removal of variations in $K_l$ in calculation and measurement of closure force.

**FIGURE 61.** **Fg versus drive voltage, calculated from the same data as Figure 60 using correction for internal spring $K_i$. Note overlapping error ranges. It can be seen that once the internal spring constant is known the internally generated force behaves predictably. Conversely this actuator can now be used with knowledge of $F_g$ and $K_i$, meaning that the force and displacement for any load can be calculated.**




The ultimate application of this method should be to calculate the force of closure within a contact. As the structure we are using is closed the $K_l$ goes from near zero to very large and the closure force becomes approximately:

$$F = F_g - (K_i \times gap) \quad \text{(EQ 29)}$$

because the closed gap no longer allows significant movement.

The quadratic power-voltage relation is defined by:

$$P = \frac{V^2}{R} \qquad \text{(EQ 30)}$$

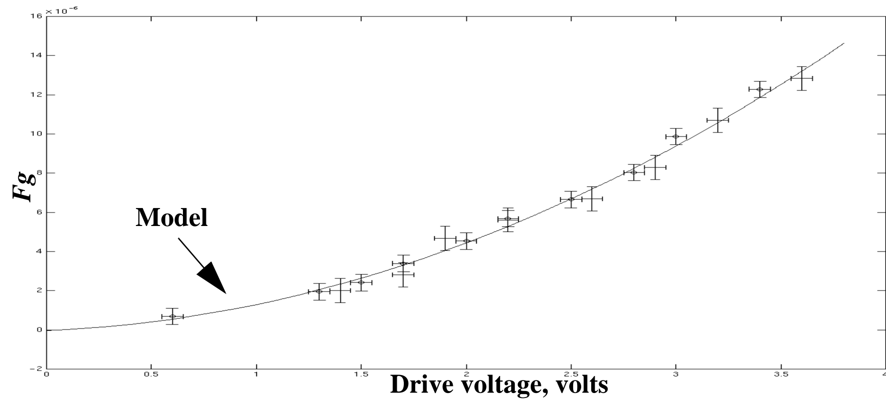
where $P$ is the power dissipated, $V$ is the voltage applied, and $R$ is the resistance of the system. There are numerous higher order effects including thermal coefficient of resistance, piezoresistivity, thermal coefficient of piezoresistivity, etc. but one would expect from (EQ 30) and (EQ 25) that the primary voltage to force relation would be a square-law. Models of first to seventh order were fitted to the data of Figure 61 with and without the 4 highest voltage points. Without the 4 higher voltage points (where the structure could be seen to be plastically deforming) a second order model of the generated force was found to be most accurate (Figure 62). The best fit model comes out to:

$$F_g(V) = (0.9020V^2 + 0.4318V - 0.0375) \cdot 10^{-6} \qquad \text{(EQ 31)}$$

with units of Newtons, and is the function shown in Figure 62. It should be noted that the zero-voltage value of this is -37nN. The standard deviation of the $F_g$ points being fitted is 519nN and 18 points were used for a least-squares fitting technique yielding an expected error for the model of 129nN assuming the square law relation is indeed appropriate. Residuals from this model were plotted and showed no significant bias, further indicating that the model choice is a good one. Thus the zero-voltage value is well within the model error bounds of zero and we can accept this model as predictive of the value of $F_g$ for use in the model in Figure 57. This model

reconciles differences in the observed data from Figure 60 using only the assumption of linearity, which was established in section 23.2 to have basic physical origins.

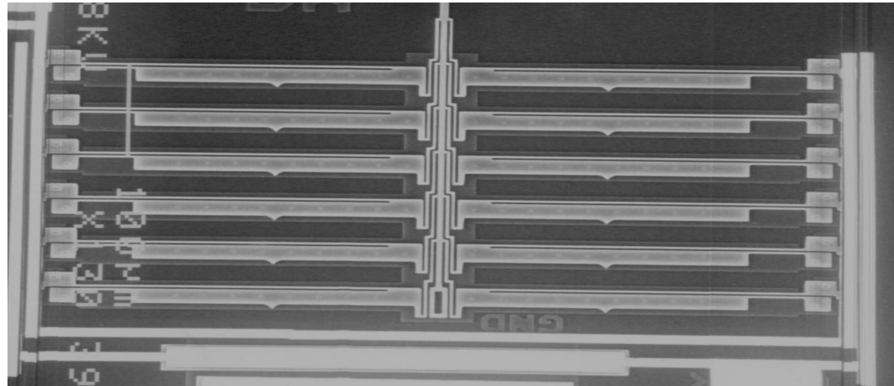
**FIGURE 62. Second order model of *Fg* as a function of input voltage. Values are computed for each voltage at which real data was taken.**




Measuring internal spring constants becomes more complicated with larger arrays of thermal actuators (Figure 63) because of the parasitic springs required to link them together into systems. It can be seen from Figure 57 and (EQ 26) that the output motion *x* and thus the measured values for deriving $F_g$ in (EQ 28) will be primarily dependent on the stiffer of the two springs. Thus the ratio of spring constants determines how much $K_i$ can be resolved and if $K_i$ is much larger than $K_l$ (as is the case in the structures in Figure 63) the value of $K_i$ will not be visible.

**FIGURE 63. A pulling array that uses spring linked thermal actuator pairs. Each additional pair increases the force available but also increases $K_i$, as**

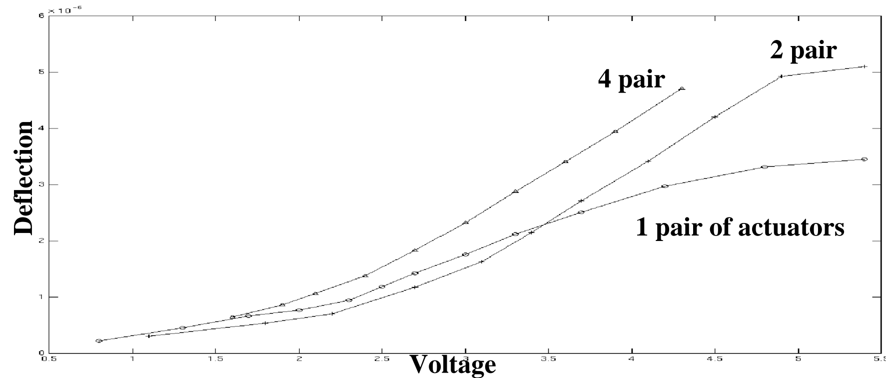
**long as $K_i$ is significantly greater than $K_l$ the changes in $K_i$ will not be extractable, though the generated force will remain predictable.**




The structures in Figure 63 are not only dominated by $K_i$, but have increasing $K_i$ with increasing numbers of actuators because each added element includes additional springs in addition to added actuation. Thus we would expect the output **motion** with small $K_l$ to be approximately equal for varying numbers of pairs of thermal actuators. This is indeed what is observed (Figure 64), but should not invalidate our earlier analysis of the internal force (and zero-travel force). Additional test structures with much larger $K_l$ values could be used to measure the internal parasitic spring constant of the arrays.

FIGURE 64. **Deflection plots for sets of thermal actuator arrays of varying numbers pushing against a weak $K_l$. Since $K_i$ increases along with increases**

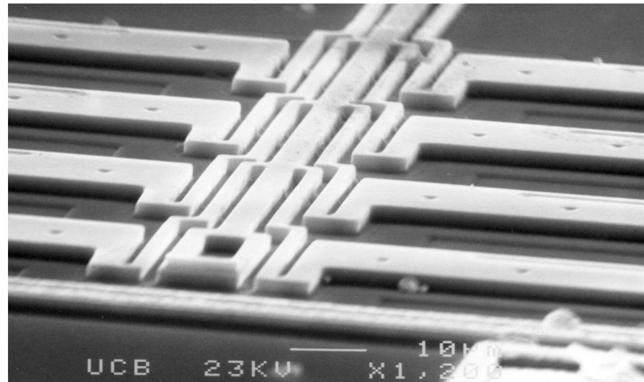
**in $F_g$ as actuators are added deflection/voltage performance remains approximately the same.**




# 25 Reliability and optimization of thermal bimorph actuators

The basic bimorph layout shown here was chosen for the fact that it has been extensively tested by others for basic behavior [35] and long term cycling limits (shown to be in the billions or more when run correctly) [40]. The devices showed very good robustness to processing problems, with the exception of stiction and thus the "loop" visible in Figure 63 at the lower end of the actuator linking spring was introduced (detail in Figure 65). The loop can be pulled by a probe to free the actuator from surface stiction. If the actuators are stuck and this is not done usually at least one will bend the wrong direction, shorting itself against the wide bar and melting the actuator.

FIGURE 65. **Perspective view of the actuator linking structure with the "loop" at the end for use at the probe station to free stuck structures.**




Other researchers have also demonstrated that optimization routines can produce thermal actuators with far lower power consumption [41] (Figure 66), and a logical step would be to implement such actuators to reduce power requirements and area.

FIGURE 66. **Power optimized single layer thermal bimorph actuator from [41], only the section that needs to expand carries current, reducing power usage, and the leverage system is more efficient for conversion to the desired direction.**

# 26 Conclusions

Actuator design requirements stemmed directly from the phenomena of electrical contact. These requirements and practical considerations lead us to the use of thermal actuation. By starting at the basic quantum principles we can see the range of linearity in all the mechanical parts and an expected square law voltage relationship for the force generated by a given voltage. Modeling and experimentation bear this expectation out and we are able to characterize the thermal actuators to the point that we will be able to not only use them to close contacts, but also to provide known force levels for characterization experiments.

# 27 References

[32] R. Yeh, E.J.J. Kruglick, K.S.J. Pister; "Surface-Micromachined Components for Articulated Microrobots"; Journal of Microelectromechanical Systems, vol. 5, no. 1, March 1996, pp 10-17

[33] Majumder, S; McGruer, N.E.; Zavracky, P.M.; Adams, G.G.; Morrison, R.H.; Krim, J; "Measurement and modeling of surface micromachined, electrostatically actuated microswitches" 1997 International Conference on Solid-State Sensors and Actuators, June 16-19, 1997 (Transducers 97), pp 1145-1148

[34] Schiele, I; Huber, J; Evers, C; Hillerich, B; Kozlowski,F "Micromechanical Relay with Electrostatic Actuation" 1997 International Conference on Solid-State Sensors and Actuators, June 16-19, 1997 (Transducers 97), pp 1165-1168

[35] Comtois, J.H. "Structures and techniques for implementing and packaging complex, large scale microelectromechanical systems using foundry fabrication processes." Ph.D. dissertation, Air Force Inst. of Technology, 12 May 1996

[36] Barrett, C.R.; Nix, W.D.; Tetelman, A.S. "The Principles of Engineering Materials," Prentice Hall, 1973

[37] Courtney, T.H. "Mechanical behavior of materials", McGraw Hill 1990

[38] Guckel, H.; Klein, J.; Christenson, T.; Skrobis, K.; Laudon, M.; Lovell, E.G. “Thermo-magnetic metal flexure actuators” Technical Digest. IEEE Solid-State Sensor and Actuator Workshop, Hilton Head Island, SC, USA, 22-25 June 1992

[39] Ananthasuresh, G.K.; Kota, S.; Gianchandani, Y. “A methodical approach to the design of compliant micromechanisms” Technical Digest. Solid-State Sensor and Actuator Workshop, Hilton Head Island, SC, USA, 13-16 June 1994

[40] Conant, R.A.; R.S. Muller: “Cyclic Fatigue Testing of Surface-Micromachined Thermal Actuators.” Proceedings of the 1998 ASME International Mechanical Engineering Congress, Nov. 15-20, 1998, Anaheim CA, pp. 273-277

[41] J. Jonsmann, O. Sigmund, S. Bouwstra “Compliant Electro-thermal Microactuators” Proceedings 12th annual international conference on microelectromechanical systems, Orlando FL, USA, 1999

# CHAPTER 5 Physical Contact Characterization

*"Form follows function."*

-Louis Sullivan

## 28 The need for contact characterization

The contact physics chapter described, as is currently understood, much of the physics behind electrical contacts. This chapter details new experiments performed to determine regimes of behavior and outline the design philosophy for microrelays.

Considerations of contact physics allowed us to reduce the design space somewhat. Elimination of fretting and closure arcs has enabled the beneficial use of soft metals, and initial gaps should be 5µm for arcing reasons. Forces should also be "low," but what, exactly, is "low?" It has been shown that materials soften as they heat but how does this interact with the minimum force for plastic deformation and can the new plastic deformation point be observed directly to anchor these design equa-

tions? How many contact spots typically form and what are the implications of this, taking surface energy into account, for separation requirements? A key set of experiments must be done to provide the missing physical knowledge to convert what is already known into sufficient design information for doing rational micro-relay design.

The experiments presented in this chapter finalize the design space we can use. Experiments in onset of plastic deformation set the minimum force levels for a given signal current. Characterization of separation forces sets the requirements for separation force versus closure force. Implementing the results of these two sets of experiments allowed for increases in relay lifetime from the range of 10-30 cycles to the range of $10^7$-$10^8$ cycles

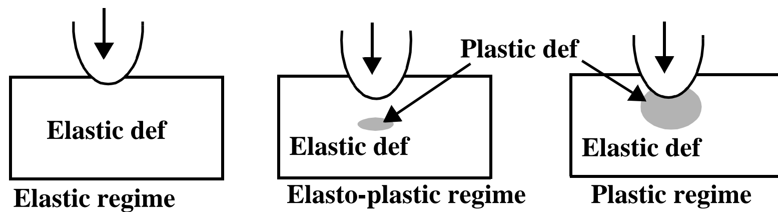
# 29 Characterization of plastic deformation point

## 29.1 Statement of the problem

The best electrical contact occurs when the surfaces making contact have experienced plastic flow. This becomes complicated because, as expressed in (EQ 2), plastic deformation sets in over a range of stress that depends on several factors.

The situation is complicated by the fact that maximum stress occurs at a depth below the surface of the material and so plastic deformation does not actually start at the surface until fully plastic behavior is reached (Figure 67).

FIGURE 67. **State of yield in elastic, elasto-plastic, and plastic deformation regimes. The most important factor for electrical contact quality is that the surface should have plastic flow, which requires full plastic deformation.**

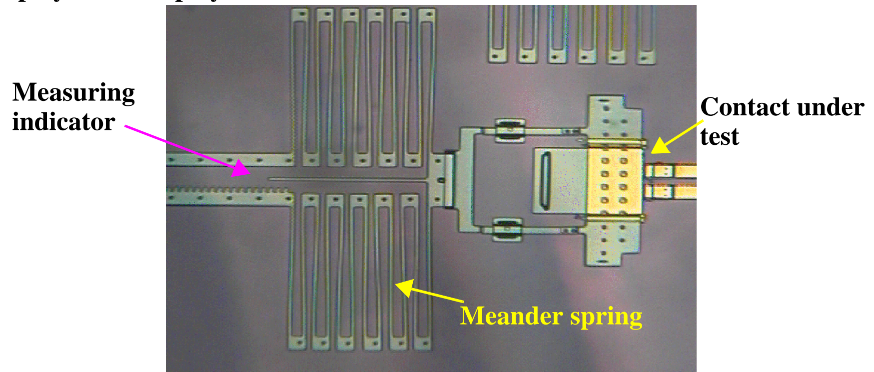



With the appropriate data the relation of (EQ 16) can be used to directly observe the transition point while varying the hardness of the material using the supertemperature heating effect showed graphically in Figure 25.

## 29.2 Plastic transition characterization using force/current/conductance curves

To characterize plastic deformation of contacts structures were designed supporting a crossbar on a vernier with a predictable spring constant as in [42] (Figure 68). The device shown in Figure 68 is a crossbar with two contacts like the devices shown previously, but is an earlier design with flat contacts as in Figure 82.

**FIGURE 68. Force measurement vernier spring from [42] with an early contact crossbar design on it. This contact is flat and wound up providing polysilicon to polysilicon contact.**



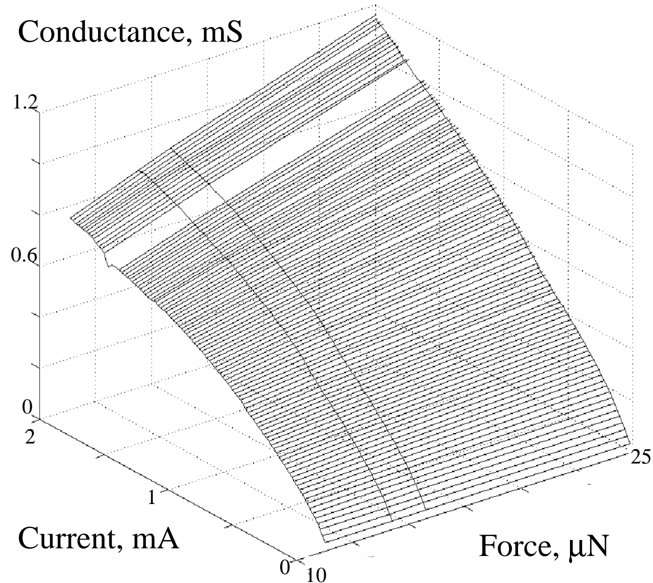

The structure was pushed into contact using probes and the force was measured by measuring the displacement of the vernier springs on a video screen. The resolution of the system is 0.1μN after magnification, measurement, and calculation. Great care must be taken not to let stiction interfere with the measurements - the vernier is only accurate when all the rovings are deflecting together in a coordinated fashion and the crossbar is sliding freely.

In order to characterize the mechanical mechanisms occurring at various current levels, I-V characteristics were taken at multiple forces and the data was fit to the established relationship of (EQ 16), reproduced here:

$$G = \frac{k}{\tau}\tau F^{b} \qquad \text{(EQ 32)}$$

Where it will be recalled that *b* is the exponential relation between force and conductance. We would expect *b* to remain at 1/3 during the entire elastic regime, 0.5 during plastic deformation and unity when resistance is dominated by surface contamination. Current, conductance, and force data is shown in Figure 69.

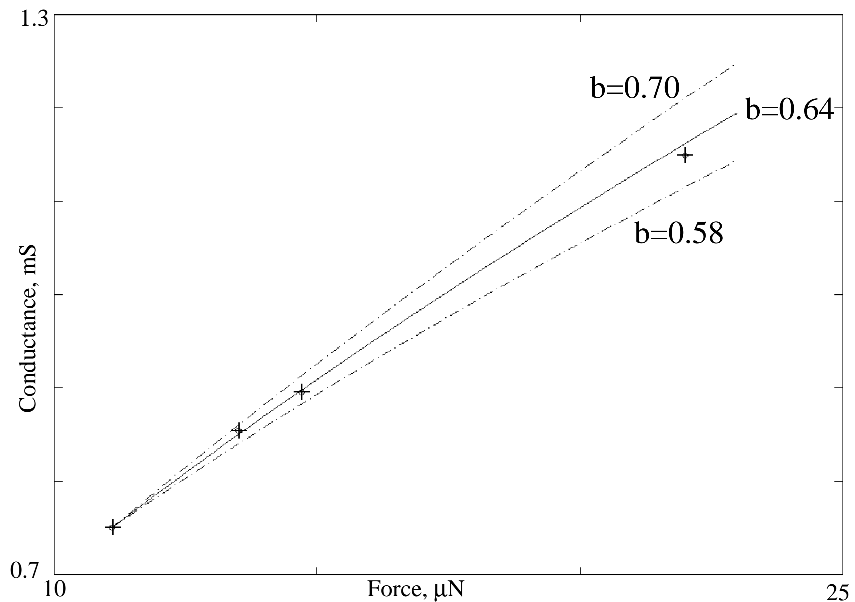
**FIGURE 69. Raw data for force, current, and conductance shown with fitted constant-current lines for evaluation. Contact was polysilicon to polysilicon. The gap in the data is from a set of 4 bad data points.**




Lines were fitted to the data perpendicular to the current axis (in the constant current direction). Once the effect of temperature is removed using (EQ 18) the first point gives *a* and fitting the shape of all the points gives *b* (Figure 70). We expect the equicurrent fits to show different values of *b* as the gold contact asperities soften with temperature, similar to the behavior observed in Figure 25. If the exper-

iment behaves according to theoretical ideals we should observe an initial plateau at $b$=1/3. The value of $b$ should then undergo some sort of transition toward a value of 1/2, eventually ending up at some value from 1/2 to 1 depending on the condition of the surface.

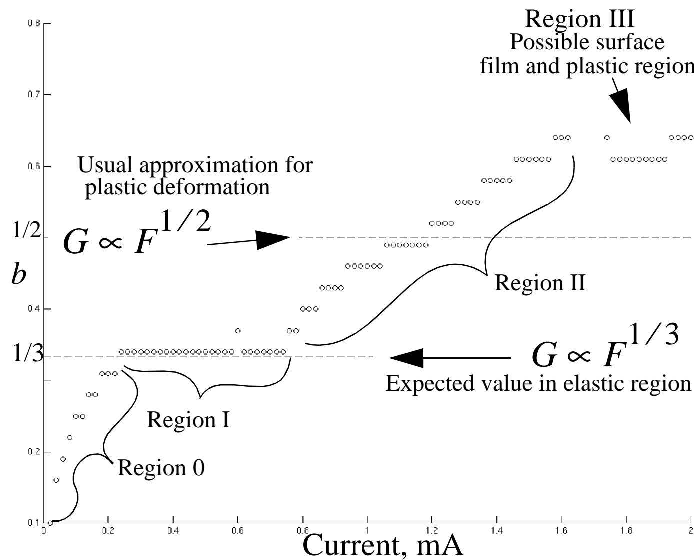
**FIGURE 70. A single equicurrent line from Figure 69, showing the poor quality of fit for values of *b* only 10% apart. Thus the fits give good values of *b* and allow observation of the regime of operation.**




The resulting values of $b$ from the data fits are shown in Figure 71. The expected elastic region (region I, $b$ = 1/3) is clearly visible. At a certain current (and associated temperature) the value of $b$ begins to increase again, signalling a transition toward the expected value for plastic yielding (region II). As the current continues

to increase the value of *b* begins to change more slowly and tends toward a limit which may represent a combination of surface films and plastic behavior. Earlier publication that region III models were less statistically significant [43] were based on a mathematical error. The initial rise to 1/3 in Figure 71 contains statistically valid models but is not well understood. The models show gaussian residuals, even distributions of differences from prediction which indicate that the model is of the correct form.

**FIGURE 71. Plot of the fitted values of exponent *b*. Uncertainties for the value of *b* are 0.015. This shows strong experimental confirmation of some expected behavior.**




Previous work has indicated that at extremely small deformation levels the Young's modulus of gold increases by a factor of 10 and approaches the ideal crystalline

strength [44]. It is possible that region 0 represents transition from this much higher modulus to bulk behavior as previously experimentally observed in nickel (Figure 19), but this is still speculation. Currents beyond those shown here caused nonrepeatable degradation in contact behavior for this particular contact geometry and were therefore excluded from this test so that all data could be verified on the same structures. All I-V data sets were taken 100 times and any that did not prove consistent within our error bounds were discarded as unstable. Very good vibration isolation was required.

One possible interpretation of this data is that the structure is undergoing elastic behavior at low temperatures, plastic behavior at higher temperatures, and approaching limits due to contaminating films as the softening continues. If this is the case then the transitions from region to region deserve increased scrutiny for understanding the physics of electric contacts. Indeed, preliminary modeling which uses Hertz theory to consider the influence of both a plastic and elastic zone does predict a departure of b from 1/3 which looks very similar to the experimental data. Similarly, simulations of softening with contaminating films generate similar behavior going from 1/2 to higher b values. However these simulations necessarily still have many fit parameters and cannot be considered rigorous until more experimentation is done.

Also, the data shown in Figure 71 is now presumed to be polysilicon contact. The absolute magnitude of the resistance can be used to calculate the size of a metal-metal contact area using (EQ 9), and if calculated here the result is on the order of picometers. Even if this range were mechanically plausible, the behavior of gold-gold contacts becomes mean-free-path limited below 42nm [45] and behaves quite differently. Calculations based on a polysilicon contact are more speculative due to the unknown effect of contaminating oxides but result in more reasonable contact size ranges.

The final conclusion here is that in the 10-25µN force range approximately 1mA is needed to transition into plastic deformation with polysilicon to polysilicon contact.

## 29.3 Usage of plastic transformation point

Once the plastic transition point is known for the system explored here we can extrapolate to guide other relay design. Softer materials mean roughly proportional decrease in force requirements by (EQ 2), (EQ 3), and (EQ 4), while harder materials increase force requirement likewise. By (EQ 9), (EQ 17), and (EQ 18) we would expect that an increase in current lowers the required force proportional to the square of the current change. Thus this experiment enables design in the microrelay regime by anchoring already known relations and can be used to generate minimum force approximations for a wide variety of microrelay systems.

From this we can extrapolate, for example, that a gold to gold contact (approximately 100 times softer than poly/poly) can transition to the fully plastic regime in the 1µN force range with 10µA of current, or in the 3µN force range with 1µA of current. Experiments in which force was applied in a controlled fashion (using the characterized thermal actuators) and the current was swept showed changes in behavior within 40% of each of these points.

# 30 Number of contact points formed

## 30.1 Statement of the problem

Separation force is typically estimated by multiplying the contact area in a relay (found using (EQ 3) or (EQ 9)) by the yield stress of the material. This is a reasonable approximation for large contact areas such as those generated in macrorelays, although it does not necessarily apply to microrelays. Macrorelays are also typically designed with large excesses of force so that accurate force approximations are not important.

The difficulty in estimating separation force arises from the numerous phenomena that determine the separation behavior, each one of which is important in different area and material regimes. The issue is further clouded by the fact that entirely different regimes might be involved if, for example, a given contact area is composed of 5 a-spots (as each contact point is commonly referred to) instead of a single one.

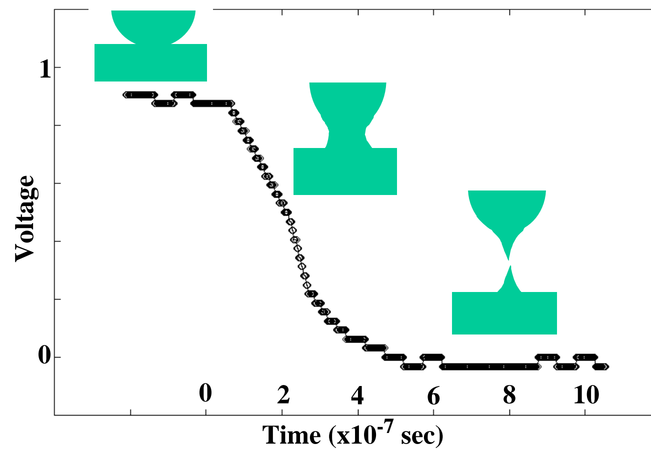
There is no static way of measuring the number of a-spots without somehow directly observing the contacts. Direct observation has been done with tools such as X-ray microscopy [46] but this is difficult to do with microrelays *in situ*. Equations first touched upon in the contact physics chapter can be used to show that the number of contact points is in fact important in the regime of operation of many microrelays using gold. The high resolution force prediction ability developed in the actuator design chapter can then be used to make transient charts that allow for the development of design rules of thumb for separation force levels.

## 30.2 Separation mechanism for relays

There are a couple of possibilities for the mechanism involved when two contact electrodes separate. The electrodes could simply crack along paths approximating the surfaces before they were pressed together as cracks originating at contact voids propagate quickly through the worked region where the materials pressed together. Alternately the contact spots could neck and deform, pulling ductilly as in a standard mechanical engineering pull-bar test. The distinction is important as the first case will require an energy balance for separation energy analysis and the second requires a more complex combination of stress and surface energy during deformation. To determine the case a super high time resolution trace was taken of the opening of a microrelay (Figure 72). If the mechanism in these microrelays is rapid crack propagation we would expect no effect at all until a sudden and linear fracture

separation. If the mechanism for separation is ductile pulling as in a pull-bar test we would expect several discrete domains of geometry change, as is experimentally observed in Figure 72.

**FIGURE 72. Transient plot of relay opening under a test voltage load. Several zones of behavior are seen. The final exponential voltage decrease may be structural necking or it may be a metal vapor arc, the initial straight regimes would be expected from mechanical necking.**



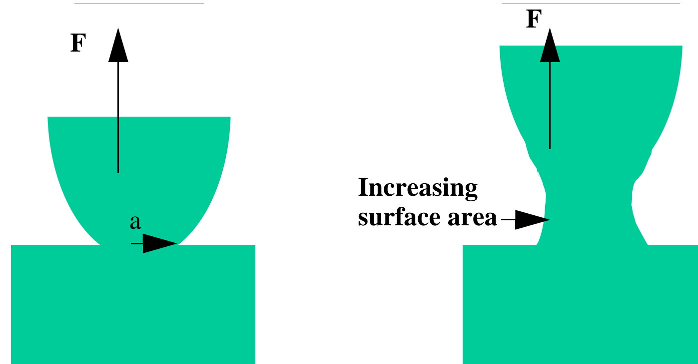

The resultant plot shows, at a 2.5ns resolution, what appears to be area reduction during mechanical necking. This is perhaps not surprising given that the electrodes involved are gold and therefore both ductile and relatively soft.

## 30.3 Relation between plastic deformation force and surface energy forces

When separation occurs it resembles a standard mechanical engineering pull-bar test with contact spots pulling and deforming into narrower cross-sections until sep-

aration occurs. This entails not only physical deformation forces but also requires additional energy input to form new surface area (Figure 73).

**FIGURE 73. Necking occurs as the contact point deforms plastically before separation. This increases surface area and thus requires additional energy input. The cross sectional drawing at left represents the appearance of a contact asperity which has deformed to form a contact point, the cross section at right shows a representative view of what the system might look like during separation as the junction necks under an applied separation force.**




As deformation occurs there are two main forces involved, one is due to the stress needed to plastically deform the material and is:

$$F_p = \pi r^2 \sigma_y \qquad \text{(EQ 33)}$$

where $r$ is the radius of the connection being pulled (and is distinct from $a$ - the effective areal radius if there were only a single contact point) and $\sigma_y$ is the yield stress of the material (for gold this is about 580 MPa). This force can be recognized as the tensile strength times the area. The second force is the force due to genera-

tion of new surface area. Since making a surface requires energy input to a system the change in surface area due to deformation can be used to calculate the force input needed to supply the new energy:

$$F_s = \pi r \gamma \qquad \text{(EQ 34)}$$

where $\gamma$ is the surface energy of the material in question (for gold this is 1120 erg/cm$^2$ or 1.12 J/m$^2$).

These two forces have some interesting interactions. The plastic deformation force is typically dominant on the macro and meso scale due to its increase with the square of the radius of contact, but it comes close to and is then less than the surface energy force at small contact radii. For a given material the two become equal at a point defined by combining (EQ 33) and (EQ 34):

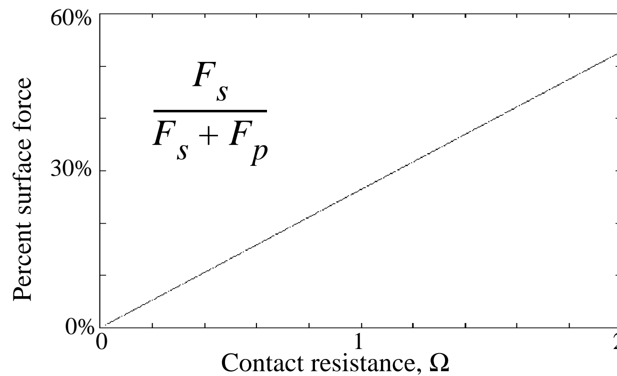
$$r_{crit} = \frac{\gamma}{\sigma_y} \qquad \text{(EQ 35)}$$

which works out to be 0.19μm for gold. For a single contact spot this occurs at a contact resistance of 38mΩ (for a total relay resistance of about 80mΩ), right in the range of performance seen here. Larger contact points (lower resistances) are dominated by plastic deformation forces, smaller contact sizes (higher resistances) are dominated by surface forces. It is important to note that while the geometry

changes during separation the force decreases monotonically so it is sufficient to evaluate the separation force starting from contact to see if a constant force will be able to separate two surfaces.

The resistance can be calculated in terms of the contact radius by using (EQ 9). Together with (EQ 34) and (EQ 35) the relative importance of the surface energy term for force of separation at a given resistance can be calculated(Figure 74).

**FIGURE 74. Relative magnitude of surface energy force to total separation force as a function of contact resistance in gold. This assumes only a single contact point. It can be seen that within the commonly observed range of microrelay contact resistance the surface energy term becomes dominant. This also emphasizes that the surface energy term cannot be neglected.**



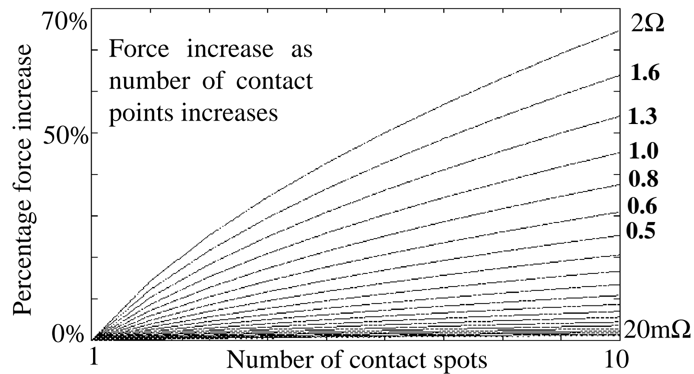

This analysis gives a raw magnitude feeling for the importance of the surface energy force if the contact is assumed to be a single a-spot. The practical case, however, is that most electrical contacts are presumed to form from multiple actual material contact points due to surface roughness. The calculations which led to Figure 74 assumed a single contact point and therefore probably understate the

importance of surface energy terms. It should be more accurate to take the area calculated from resistance measurements, divide it up into a number of separate connection spots, and use their individual radii of contact to compute the force required to separate the two samples. Even if the actual number of contact spots is unknown this will give us a feeling for the magnitude of separation force overdesign that must be included in a real application. Since our goal is design guidelines for functional microrelays a conservative approach (too much separation force) is preferable.

Figure 75 shows the percentage increase in surface energy separation force required as the number of contact spots increases. The comparison is made by calculating the total plastic and surface energy force with a single spot or with a number of identical spots equal to the x-axis value which make up an identical area and therefore an identical resistance. This plot emphasizes that underestimates of tens of percent can be made in calculating separation force if the number of actual contact points is not characterized in some way, even at sub-ohm resistance levels.

FIGURE 75. **Plot showing the percentage increase in contact separation force needed as number of contact spots (a-spots) rises above one. All**

**calculations are relative to a single a-spot contact and each solid line represents a resistance from 20mΩ to 2Ω, scaled logarithmically.**



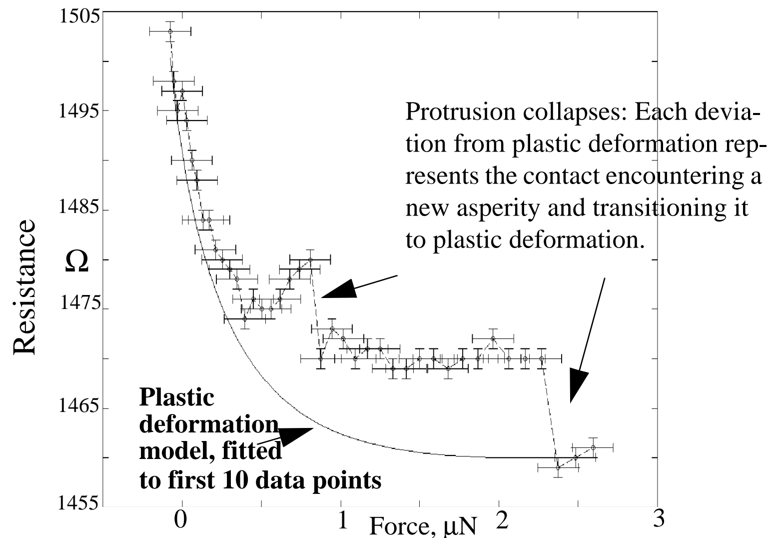

An interesting ramification of Figure 75 is that if plastic deformation takes place to join two materials an increased roughness can increase the adhering force by increasing the number of contact points. This contrasts with the conventional wisdom that rougher surfaces stick less. While more contact points do increase the force required to transition to plastic behavior they also increase the surface energy required to separate materials once plastic flow has occurred. The previously presented behavior of Figure 71, for example, represents a sum of the behaviors of many points and may represent many individual points going into plastic deformation -- which would require significantly more separation force.

Thus if we can use the characterized actuators to come up with an approximation for the rate of contact spot formation within our application we can adjust our performance expectations and design parameters for the number of actual contact spots we expect.

Now we know contact spot population has a significant effect on separation force. We also have a way of measuring the force that is applied during relay or surface closure by using the characterization presented in chapter 4, so we can measure the number of spots that form during known force ranges. Such data allows significant refinement of contact and separation models. The data in Figure 76 shows a test contact made on a test chip using electroless gold. The data shows the formation of at least three contact spots within a 3μN force range (parasitic structural resistance was approximately 1450Ω). The asperity point formation rate shown here of approximately one per μN appears fairly consistent across our structures and relays in multiple tests. This means that as we design for a given contact force we can estimate separation forces with much higher accuracy by including an estimate of the number of active contact asperities in the calculations.

**FIGURE 76. Direct observation of microasperities being encountered by a closing contact and transitioning through various ranges of behavior. As each is first encountered it is in elastic deformation and provides less**

**conduction, when it transitions to plastic deformation a drop in resistance is seen. Experimental current is 1mA.**




It is interesting to note that if we apply the conclusions of section 29.3 to the consideration of Figure 76 we would expect that in the 1mA current range an incremental force input of approximately 0.3μN would be predicted to collapse each gold asperity to plastic deformation. Given that the observed rate of protrusion collapse is about one per micronewton this seems to be well within the realm of believability. The electroless gold used here is fairly rough (see the fabrication chapter for photos) and the observed behavior seems to indicate that little more than the minimum plastic transformation conditions are required to encounter the next asperity. Similar tests to Figure 76 run with lower currents show larger intervals between protrusion collapses, although the decreased resolution makes it very hard to see.

Higher currents appear to transition through asperities too quickly to get unambiguous data.

### 30.4 Using the contact spot population results

The simplest way to use these results is to consider both (EQ 33) and (EQ 35) when designing separation force and to use an approximation of one contact spot per micronewton. Preliminary analysis and observation, however, indicates that this rate of contact spot formation should be altered depending on the current levels through the contact to take microthermal effects into account. Thus, for example, if the current levels are only going to be 10μA, we would expect to have a protrusion collapse only each 10μN, although quantitative data for this has not yet been observed.

It should also be noted that, unlike the earlier plastic transition experiments, the contact spot experiments may have an extreme dependence on surface texture and hardness and need to be done for each new system.

## 31 Conclusion

Novel physical contact characterization has allowed for design guidelines to be generated which bridge the gaps left by classical relay physics. We can now design in a fully specified environment. If a relay is desired for a given current load the

minimum actuation force for plastic deformation can be calculated, providing knowledge of the minimum force that needs to be applied. Earlier equations allow for calculation of area and resistance so the relay can now be designed within the force regime desired. Now the effects of the extra force in terms of contact point generation and separation force can be calculated. This allows for actuator selection with adequate internal spring constant to provide the extra force needed for positive separation. In general it can also be said that as a rule of thumb higher currents generate significantly higher areas and separation forces for identical closing forces.

Knowing these design regimes makes a tremendous difference in relay success. Designs using the force of plastic deformation estimated here and the force necessary to ensure clean contact separation found for this system resulted in devices with six orders of magnitude better switching lifetimes.

# 32 References

[42] Yeh, R; Pister, K.S.J. “Measurement of static friction in mechanical couplings of articulated microrobots” Proceedings of the SPIE - The International Society for Optical Engineering vol 2642, 1995, pp40-50

[43] Kruglick, Ezekiel J.J.; Pister, Kristofer S.J. "Micronewton Contact Characterization for MEMS Relays,"; 19th International Conference on Electric Contact Phenomena, Nuremberg Germany, Sept. 1998, pp 53-57

[44] Oliver, W.C.; Hutchings, R.; Pethica, J.B. “Measurements of hardness at indentation depths as low as 20 nanometers”, Microindentation Techniques in Materials Science and Engineering, ASTM STP 889, 1986, p.7637-42

[45] Beale, J.P.; Pease, R.F.W. “Apparatus for studying ultrasmall contacts” Electrical Contacts - 1992. Proceedings of the Thirty-Eighth IEEE Holm Conference on Electrical Contacts, Philadelphia, PA, USA, 18-21 p.45-9

[46] Buckley, Donald H. “The use of analytical surface tools in the fundamental study of wear” Wear 46, 1978, pp 19-53

CHAPTER 6

# Mechanical Design and Layout

*"Miracles sometimes occur, but one has to work terribly hard for them"*

-Chaim Weizmann

## 33 Considerations for mechanical design

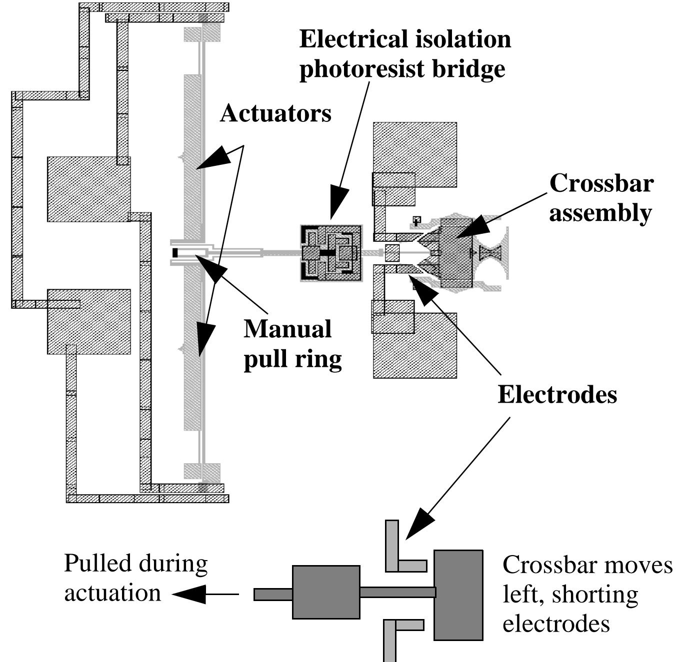
The contact mechanics and design trade-offs investigated so far have outlined a well defined design space. The actuator should be hooked up to provide greater separation force than closure force, the contact should close with tangential and normal force on the contact points, and the relay must be tolerant of as much process variability as possible. This must all be done while satisfying the force/current balance for plastic deformation.

General purpose relays, then, have some freedom in design because the test current can be whatever is convenient. Relays to operate within a specific system such as relay based logic, however, must be designed with considerations of the anticipated current levels.

# 34 Actuator hookup

The first decision is whether to set the actuator to pull or push the relay closed. Pulling action was selected because it has fewer failure modes, actuation rods in tension cannot buckle. While the probability of buckling at the actuation forces here may seem remote, the fact that the crossbar is allowed some rotation significantly reduces the buckling threshold. Thus the device is hooked up to achieve a positive pulling contact force (Figure 77). Later calculations have indicated that the separation force will in many cases exceed the contact force in high current applications so reconsideration of the actuator positioning may be in order for those devices. Devices subjected to low currents, however, should be arranged to be pulled closed.

FIGURE 77. **Mechanical overview of a relay with actuator. The actuators pull the crossbar assembly into the stationary electrodes. The manual pull ring is for freeing the device if it gets stuck down during release.**




Placement of the electrical isolation bridge was somewhat arbitrary, motivated more by the need to have it in the clear for photolithography than any mechanical needs.

The design equations developed in chapter 5 also showed that the level of expected current plays a part in the design of a relay for issues of minimum closing force and maximum separation force. The limiting factor (using the careful design parameters espoused here) becomes separation force for many applications so different sub-system parts were often designed with different actuation levels. Examples of this

can be seen in Figure 78 and Figure 79 which show two different system sections designed with different actuation forces to carry different loads.

**FIGURE 78. General purpose digital logic, this is not required to drive much of a load and has minimum actuation.**

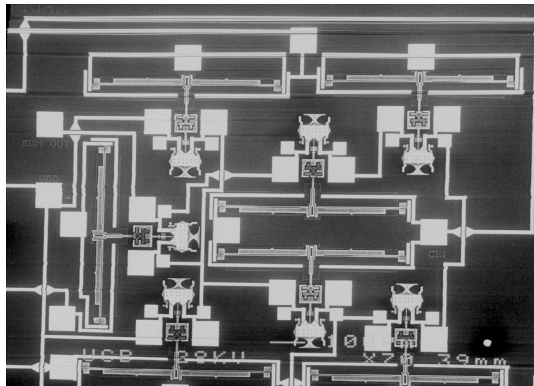
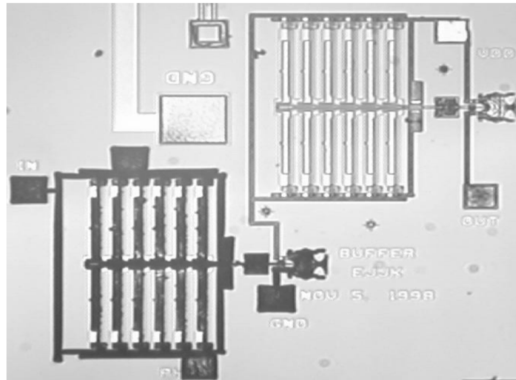


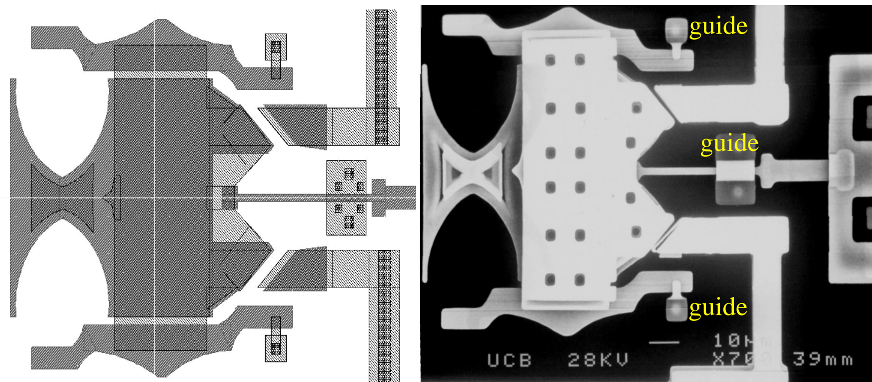
**FIGURE 79. Digital logic output buffer, this is designed to carry a significant current load and so has a large drive array with very high internal spring constant to assure separation.**



# 35 Assuring contact and resisting mechanical lockup

Dozens of different designs for the crossbar and electrodes were tested before the best one was determined. The final design with the highest yield is shown in Figure 80.

**FIGURE 80. Layout and device of the best crossbar/electrode design. The curves and points are designed to reduce stiction problems, the guidepieces are required to maintain planarity and placed to maximize allowable rotation.**




The gentle concave curves, alternating with relatively sharp convex points where needed, are shaped in an attempt to guide water meniscus formation away from the structure wherever possible and to cause the fastest outside-to-inside meniscus transition possible[47]. While ordinary air-dry yield did seem to increase after the addition of these elements (such that critical point release was no longer necessary) control structures were not fabricated so no certain statements can be made about their usefulness.

The guides seen in Figure 80 turn out to be one of the most important and difficult to design portions of the relay. It is critical that planarity be maintained in order to achieve lateral contact, without guides the crossbar pivots upward upon contact with the electrodes and little electrical contact occurs. At the same time the guides should allow as much freedom as possible so that when one of the two contact sides hits first the crossbar can pivot about it, pulled by the central rod, and be levered onto the other electrode. This is problematic because the clearance between the guide and the first polysilicon underneath it will always be equal to the thickness of the second oxide both laterally and vertically.

This clearance means that the maximum angular displacement that can be allowed in the guided arm is 0.5µm. Each guide is this design 35µm from the centerline and therefore 44µm from the center of rotation which will occur if the opposite electrode hits first, about as close as they can be placed. Additionally the guides are located on a straight line with the electrodes to minimize the component of rotation perpendicular to the allowed sliding direction. The yield using earlier guide structures was extremely low and provided motivation for failure analysis. Due to common misalignments one side inevitably hits first and structures with guides placed, for example, on the back set of arms (those extending to the left of the crossbar in Figure 80) would frequently bind.

# 36 Making lateral contacts in MUMPs

The MUMPS process [48] has two structural polysilicon layers and a final gold metallization. We want to allow a lateral relay contact design using the gold metallization. Thus an important contact design consideration is the necessity for metal-metal contact. Two different classes of design variation were tested and the basic types are shown in Figure 81 and Figure 82. Each contact topology is defined by a vertical (or cross-sectional) and horizontal topology. The basic horizontal topologies are perpendicular contacts, which apply normal contact force, and angled contacts, which approach such that the closure force has both normal and tangential components (see Figure 82). Two different cross-section topologies for bringing gold to gold are shown in Figure 81. These are classified here as type I and type II.

FIGURE 81. **Two types of lateral contact topologies to attempt gold-gold contact in a polysilicon based MEMS process. The top line shows a schematic cross section of the designs, with all major material layers labelled. The second line shows a cross section of how each topology is designed to look and function after chemical removal of the oxide layers. At the bottom is a series of SEMS showing the face of the crossbar and**

**electrodes from each design. The left pair of photos are from a contact perpendicular to motion, the right pair are from an angled contact.**

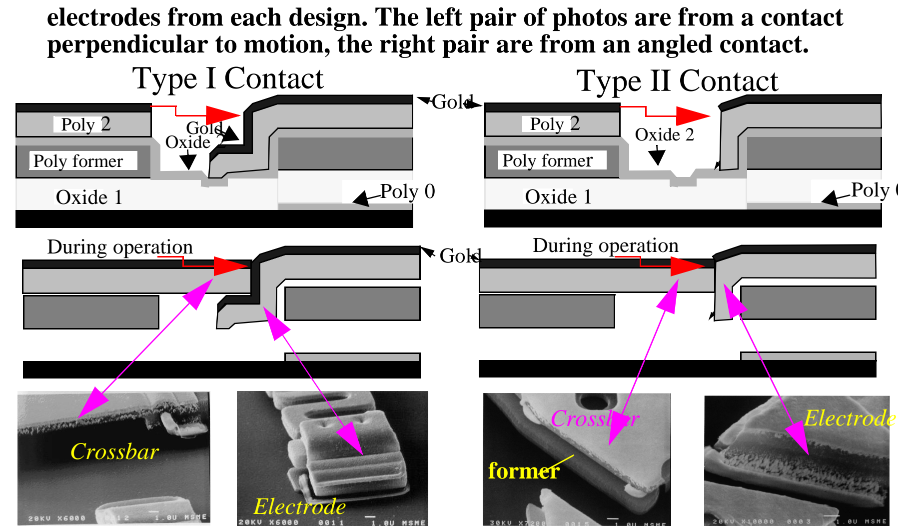
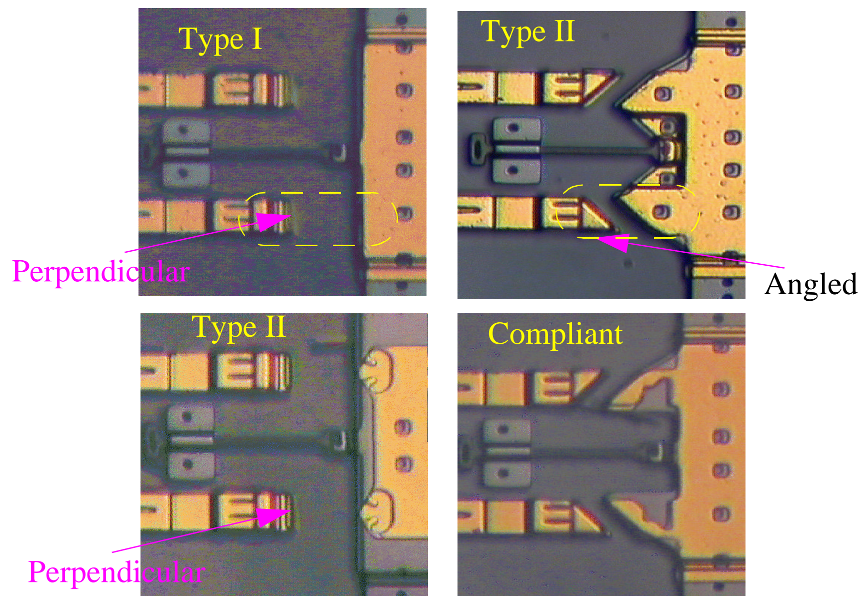



FIGURE 82. **Views from above of different relay designs. Note that each relay represents two contact points. Two contact points are required for a crossbar to short out the electrodes. The areas in the ovals correspond to the areas shown in the SEMs of Figure 81.**



Angled structures using the type II topology exhibited the best performance without supplementary processing. The angle laid out was chosen arbitrarily at 45 degrees for ease of layout because (as detailed in section 8.2) the optimal angle is not known. The device shown in the upper right of figure 82 can exhibit a sub-ohm contact resistance without any additional gold. The contact resistance was stable when cycled by hand over 80 times with a 5V open circuit, 400μA load and forces near 500μN. Fatigue testing of this sort has been limited by the inevitable eventual mistakes made at a manual probe station so structures were designed to test the different relays automatically.

A thermally actuated relay using the structure in the upper left of Figure 82 achieved 2.4Ω contact resistance (for a total relay resistance of 4.8Ω) and 80mA current maximum when actuated at 7-12V; and contacts implemented using the type II angled contacts shown in the upper right of Figure 82 have conducted 0.45 amps with sub-ohm resistances without post processing. These performances are not typical, however, and often do not stand up to cycling as the amount of gold hanging over the edge to make contact is minimal and quickly eroded. For this reason it is highly desirable not only to deposit additional gold but to do so in an isotropic fashion so that there will be considerable lateral gold protrusion to make contact with. Once additional gold is added the relays perform better and more reliably.

# 37 Conclusions

Once sufficient characterization was done, most of the design details were a matter of simply having enough constraints. While the angle of the contacts was chosen arbitrarily, almost all other aspects, from guide necessity and placement to number of actuators, was dictated by the operational specifications of each device.

This section has outlined the many choices made during mechanical design of the relays. Actuator placement and crossbar design was shown to be governed by mechanical considerations. Number of actuators sets force and different numbers are used for different load needs. The electrical isolation bridge was placed for lithography consideration. Different contacts were tried and a final design chosen for yield and durability.

# 38 References

[47] Abe, T; Messner, W.C.; Reed, M.L.; "Effective Methods to Prevent Stiction During Post-Release-Etch processing" In Proc. IEEE Micro Electro Mechanical Systems Conference, Jan 29 - Feb 2 1995, Amsterdam, The Netherlands, pp 94-99

[48] Provided by MCNC Corp, Now Cronos Integrated Microsystems, Inc. 3021 Cornwallis Road, P.O. Box 12889, Research Triangle Park, NC 27709-2889, USA; Ph: (919) - 248 - 1983 FAX: (919) - 248 - 1434

CHAPTER 7

# Electrical Testing and Characterization

*"The only certainty is that nothing is certain"*

-Pliny the Elder

## 39 Introduction

This chapter details the methods, limitations, and considerations of electrically testing relays in several ways while presenting more comprehensive results as well.

## 40 Resistance measurement

Voltage and current were simultaneously measured under computer control by a Hewlett Packard 4145b parameter analyzer. Our 4145b allowed us to drive with a sweep of current in the range 1µA to 0.1A while simultaneously capturing the required voltage up to 10 volts. This method was used for most testing at low currents but was limited in resolution by our 4145b's voltage resolution of 1mV.

## 40.1 Extending experimental range

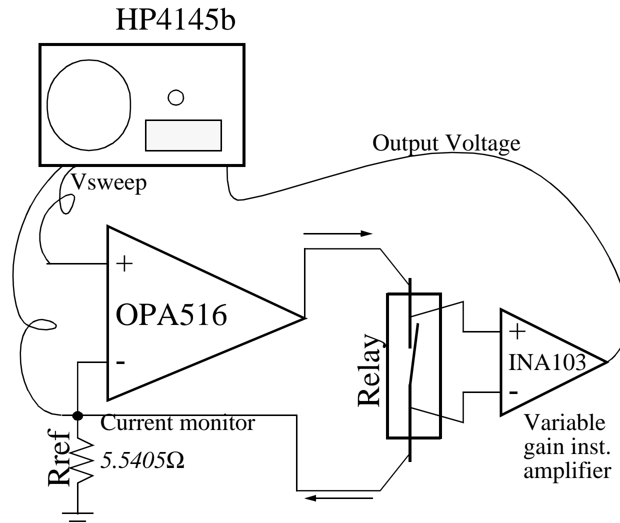
For high current testing an extension setup was built with higher current range and adjustable voltage sensitivity. The higher current range was needed to explore the maximum performance and modes of failure of the devices. The adjustable voltage sensitivity extended the effective dynamic range of testing by allowing best raw resolutions on the order of 10 μΩ.

The high current test extension takes a swept voltage input from the HP4145b and converts it to current using the circuit shown in Figure 83. The OPA516 was configured to provide up to 5 amps DC current load. The feedback circuit, if ideal, would ensure that the current through the relay was:

$$I = \frac{Vsweep}{Rref} \quad \text{(EQ 36)}$$

and this was very close to true. However, thorough characterization showed that while the current was very stable and linear at any given relay resistance the exact relation of the current to the sweep voltage varied slightly with load. Thus open loop sweep control was deemed unacceptable and the current was monitored by measuring the voltage across Rref. The resistor Rref was implemented with a temperature stable structure and the entire circuit was actively cooled to avoid temperature problems associated with the high power in this circuit.

FIGURE 83. **Extender for high current, high resolution tests of microrelays. The OPA516 was configured to drive up to 5 Amps, the instrumentation amplifier was configurable for gains of 1, 10, or 100. Actual gains and offsets were characterized.**




The instrumentation amplifier was implemented using an INA103 and was adjustable to have designed gains of 1, 10, or 100. Characterization showed the actual gains to be 0.992, 9.86, and 111.44 and these values were reconfirmed before any testing in which the system was used.

Using this system 4 wire measurements could be made with a voltage resolution of 10μV while driving up to 5 A. This gives a theoretical single point resistance measurement limit for the system of 1.8 μΩ, which can be made better with slope fitting and data averaging techniques. In practice most measurements were made at lower currents, which reduces the resolution linearly. Most measurements also did

not use the maximum amplification so that a wider range of behavior could be observed.

## 40.2 Resolution of basic measurements

The simplest way to measure the resistance of a relay is to hook a 4145b channel (a Source/Measurement Unit or SMU) to one pad of a relay and ground the other. Then the SMU current is swept and the voltage is plotted to monitor the resistance as current changes. With the basic relation:

$$V = IR \tag{EQ 37}$$

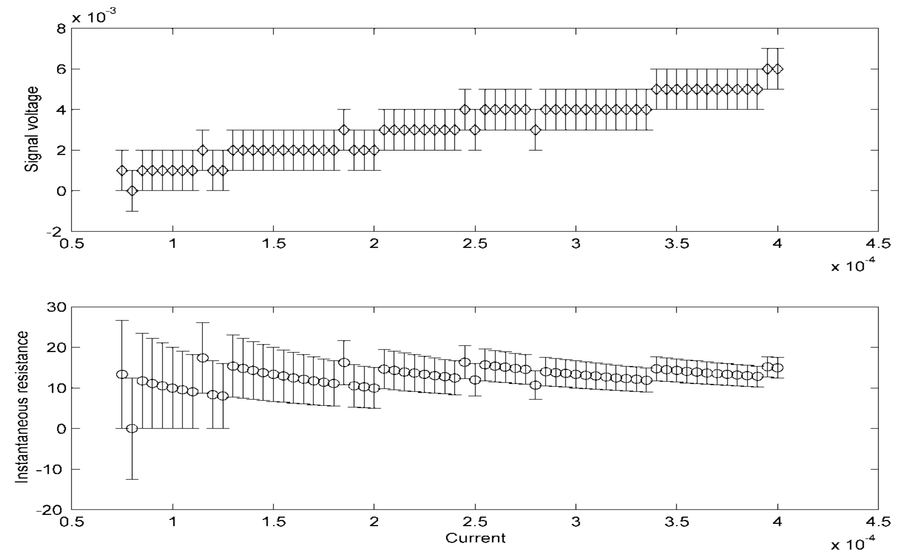
we can calculate the instantaneous value of the resistance at any value of I by transformation,

$$R = \frac{V}{I} \tag{EQ 38}$$

and we can apply the 1mV measurement uncertainty to see what the resistance uncertainty is at various currents. During a sweep from 1μA to 0.1mA (the range of the 4145b) the resistance uncertainty goes from 1kΩ down to 0.01Ω. This leads to resistance versus current plots that look jagged with systematic error at low currents and smoother at high currents (Figure 84).

FIGURE 84. **Plot of data from a relay showing the effects of low voltage quantization. The upper plot shows the raw current and voltage output with error bars representing only the quantization error. The bottom plot shows the effect of division by plotting the instantaneous resistance at each current as calculated using equation 38. It can be seen that the error bars**

**are very large at low currents and decrease as current increases. This problem is more pronounced at lower resistances where this absolute systematic error is a large multiple of the value to be measured. The plots use actual relay data.**




Despite its drawbacks this method provides valuable data as long as the errors are taken into account and most of the resistance/current plots in this dissertation were produced using this direct method; the errors are acceptable as long as the low current regimes are avoided.

Yet while this method is sufficient for measuring high resistance relays it becomes less and less accurate as the resistance becomes lower. The best performing relays can easily have resistances below the measurement noise until very high currents are reached. Similarly, even the moderately good relays can experience enough quantization jaggedness to obscure behavioral qualities.

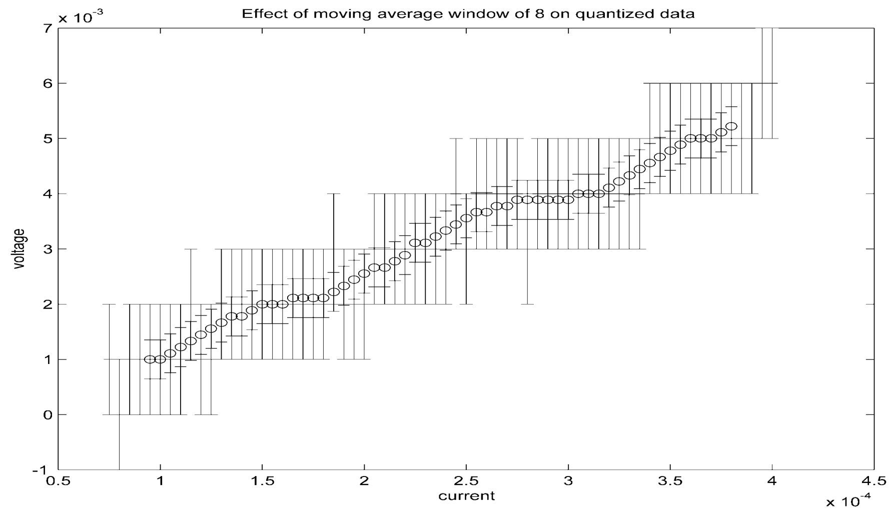
### 40.3 Averaging

When the signal is sweeping monotonically one way to smooth out the data quantization is averaging (Figure 85). This method reduces the uncertainty by up to the square root of the number of points used to compute each averaged value, but has the unfortunate side effect of shortening the data stream by a number of points equivalent to the size of the window. It also removes fine detail by combining neighboring points and suffers from decreasing efficiency with increasing window size because the best improvement possible is proportional to the square root of the number of samples.

Figure 85 was generated by replacing each point with an average of the 8 points surrounding its location on the x-axis. The first and last 4 points, therefore, have no value because there is insufficient data to one side and the resulting dataset is shorter by eight points. In this case the input data is from a monotonically increasing output so averaging does improve quantization noise, though this would not be the case if the output were a single stable value. Artifacts from quantization can still be seen in plateaus at certain values, indicating that we are getting less than optimal improvement from this method. The error bars are approximations calculated by dividing the original error by the square root of the number of samples and would be accurate if the errors being addressed were gaussian. In practice the error will be

somewhere between the two values shown on the plot because the quantization error is not gaussian.

**FIGURE 85. The effects of a moving average on the quantization uncertainty of data. Larger, lighter error bars represent data taken the conventional way, the circular data points are the results of averaging and the smaller error bars represent the best possible case for this method. The moving average makes points by averaging a sliding window of 8 points, reducing the error by up to a factor of 3. This method obscures some fine detail and will conceal sudden changes in behavior. Additionally it has reducing returns as the window grows.**




## 40.4 Slope calculation

From equation 37 it is apparent that the slope of these voltage and current plots is the resistance. This gives us the opportunity to convert from I-V plots to resistance without doing the noise increasing division of equation 38 by using statistical mathematics to exploit the fact that we have a large number of points. The least squares method can be used to fit a slope for each group of points, giving the most statisti-

cally accurate possible rendition of the slope from these points. This is different from averaging in that the least squares method fits the best slope possible instead of averaging the points then converting to resistance by doing a noise-injecting division by current. The algorithm (which uses series to minimize the sum of the error squares) gives the provably best linear slope estimation for the points involved [49]. The equation for calculating slope *m* from a set of *x* and *y* points is:

$$m = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2} \quad \text{(EQ 39)}$$

where the bar values are the averages. This equation is derived to fit to linear data and, although we do not need it for resistance calculation the intercept *b* is needed for calculating the error of each point. The intercept is found by:

$$b = \bar{y} - m\bar{x} \quad \text{(EQ 40)}$$

and is needed to use the equation for the error at slope fit [49]:

$$\sigma_m = \frac{\sum ((mx+b)-y)^2}{\sum (x-\bar{x})^2} \quad \text{(EQ 41)}$$

The slope method also eliminates all influence of DC offset from any amplification used.

An example application of this method is shown in Figure 86, where data taken with the improved system was used to measure the background trace resistance (which will be slope *m* in the I/V plots) as 0.7497Ω with a resolution of 9.74μΩ.

This number was arrived at by using all 51 voltage-current pairs to get a single value from the least squares fit method - provably the most statistically accurate method while the data is linear. This method is contrasted with instantaneous resistance calculation in Figure 87.

FIGURE 86. **A dataset collected using the current extender, amplification, and slope analysis. The wider current range (above 100mA) can be seen. The raw data has a voltage resolution of 0.1mV, which makes for error bars almost on top of each other. This particular run was done in four-point mode to measure the contact resistance with no relay - the resistance is due to a test set of fixed on-chip traces. The resistance is 0.7497Ω with a sigma of 9.74μΩ, calculated using the least squares slope method.**

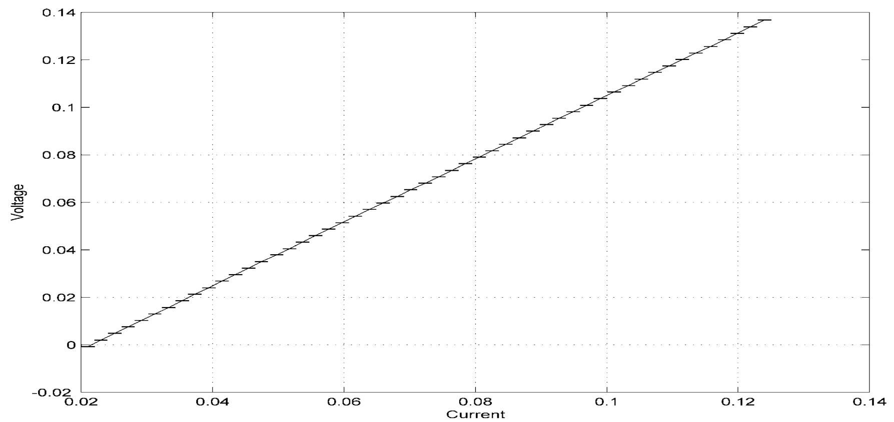
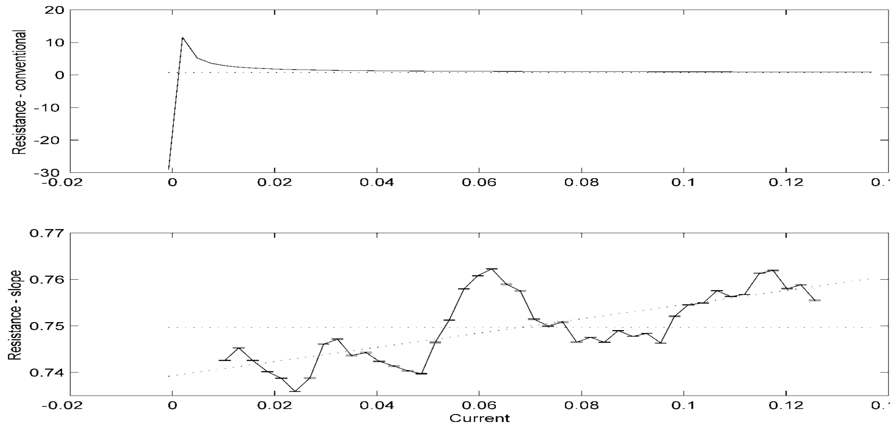



FIGURE 87. **Calculations of the resistance from figure 86 using the conventional and the slope method with a sliding window of 8 samples, with the best derived value from Figure 86 shown as a dotted line. Plots are in ohms, but to different scales. Error bars could not be calculated for the conventional method due to offset errors in the amplifiers. Remaining error in bottom plot is a mixture of true measurement error and slight heating. Fit to heating is also drawn in and can be seen to correlate strongly**

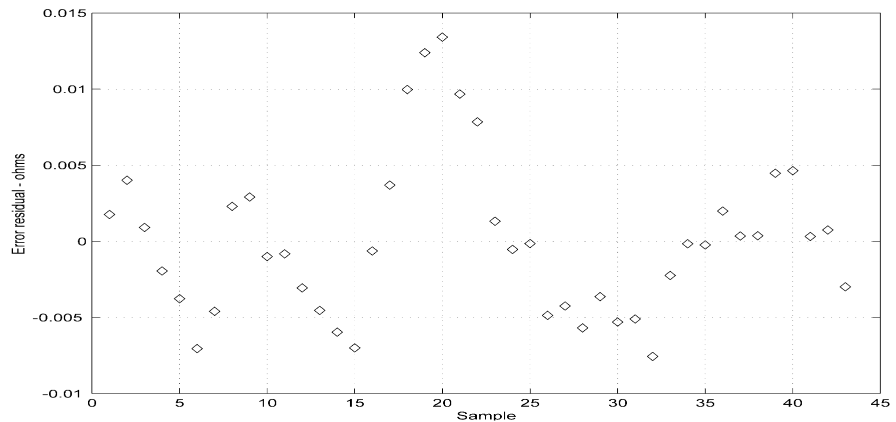
**to the residual error; another strength of the least squares fit method is that the residuals can be analyzed to detect less significant trends.**




The slope fitting method is not so advantageous if the data is not linear, although it can still provide substantial gains when used on a sliding window of points. The top plot in Figure 87 shows the results of using equation 38 on the data in Figure 86, while the bottom plot in figure 87 is a conversion of the same data using the slope fitting method with sliding groups of 8. The slope fitting method leads to far greater resolution as well as known error bars. The slope fitting method gives sharp enough results that some other systematic nonlinearities can be seen. These include (presumed) thermal heating and an oscillation which becomes visible in the residuals after the resistance value and thermal slope are removed (Figure 88). These residuals ultimately limit this resistance measurement accuracy to 5.1mΩ when using an 8 point sliding window to do least squares fit conversion (a more typical application method). By comparison the standard deviation of the division method here is

4.6Ω, with a best case error of 1.6Ω if an 8 point averaging method is used. Thus the resolution gain going from averaging the voltage and dividing to slope fitting is about 300.

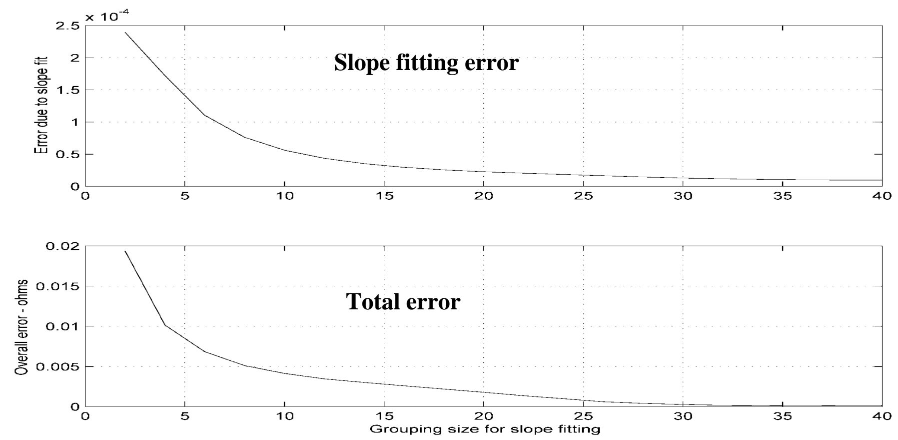
**FIGURE 88. Residuals from the data of Figure 86. Here each point is from the sliding slope fit of figure 87 and the resistance and thermal effects have been subtracted out. The origin of the oscillatory nature of these residuals is unknown. This noise leads to a resolution floor of 5.1mΩ for this sliding window size (8 V-I points per R point).**




The slope method can be used with varying sizes of window to gain resolution at the expense of fine detail. The effects of various sizes of window are shown in figure 89. The error added by the conversion from V-I data to resistance, which used to be the dominant error source, can be made to contribute less than 2% to overall experimental error.

**FIGURE 89. Slope fitting error and overall error values for different sizes of slope fitting windows. The range of values for the errors due to slope fitting show that the added error (which used to be dominant), is now 2% or less**

**of overall experimental error which includes sources such as quantization and amplifier noise.**




## 40.5 Typical relay contact resistances

Using all the methods above might be expected to yield solid numbers for relay resistances. Unfortunately any resistance must, for reasons given in the chapters on contact physics and physical characterization, be couched in terms of actuator drive voltage and conducted current due to the effects of changing force and thermal softening. Unless otherwise stated the test current was arbitrarily chosen as 1mA and the drive input was 5 volts. Variability in performance is attributed mostly to unevenness in the electroless gold and structural offsets which cause the relays to close different gaps, thus leaving them with different applied forces. Typical results for the relay resistance were anywhere from 10mΩ to 1.5Ω, with most occurring between 30-200mΩ.

# 41 Maximum static current handling

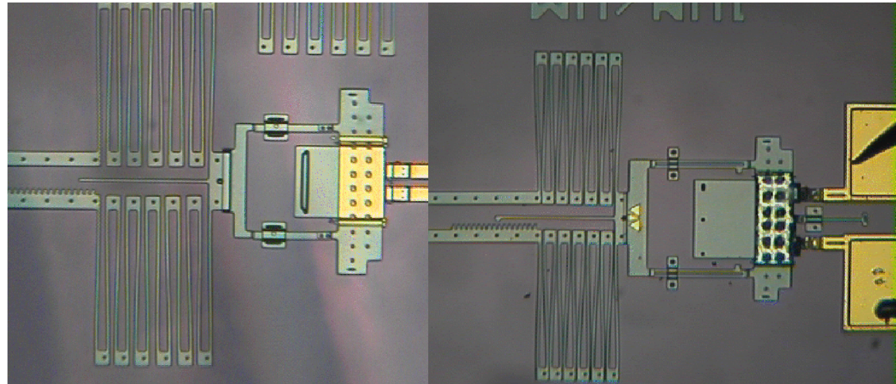
The static current limits were tested by closing contacts and running increasing current through them until burnout. Figure 90 shows a type I perpendicular contact after being burnt out at 30mA. Thermal failure over the entire crossbar and electrodes shows that in this static situation the relay undergoes widespread damage.

Similar tests with a type II angled relay required the current extender. The angled type II relay was able to pass 448mA while remaining below 750mΩ. The lower resistance (which reduces Joule heating) leads to lower operating temperature in these designs. The design of the newer type II angled relays also meant that this was not a destructive failure point. Instead, once the temperature gets too high (at around half an Amp) the gold on top expands more than the polysilicon below, flexing the crossbar and breaking contact. When the current is removed the crossbar can be observed to return to its previous shape but gets wedged into a position where the contact points are no longer in contact. Thus the relay must be cycled to recover from such an overload. This mechanism was observed on multiple relays at the probe station. This can be avoided if desired by using a slightly different crossbar design with corrugations for increased rigidity.

By measuring voltage as the current was increased during maximum loading tests we were able to view the resistance behavior as thermal failure approached. Behavior showed the expected softening transitions and followed the theoretical thermal

response to a high degree, confirming our understanding of the effects of thermal softening. It is expected that enhanced thermal sinking from the crossbar will allow for higher static current loads.

**FIGURE 90. Two similar early generation relays, one without damage (left) and one after being burnt out with a heavy static current load (right). Designs of the contact areas are the same, although these devices did have a few minor layout differences. The most recent designs have a bimaterial auto-break function which prevents such burnouts.**



# 42 Transient characterization

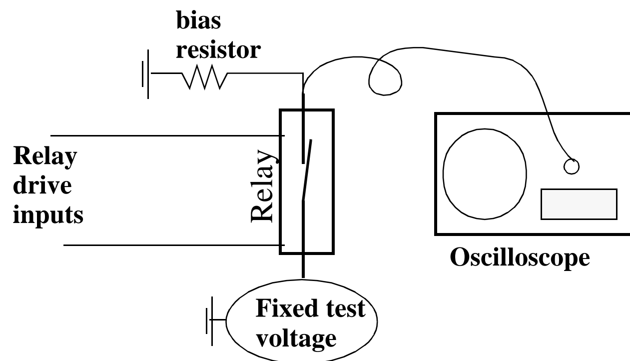
## 42.1 Methods and equipment

Transient characterization was done by recording waveforms while a load was switched by the relays. A laptop computer was prepared with a GPIB PCMCIA card and Benchlink HP34810A software was used to gather data from an HP 54600B digitizing oscilloscope. This setup was time resolution limited by the abil-

ity of the oscilloscope to catch transient events effectively. The finest time resolution usable was 2.5 nanoseconds, but this was rapid enough to observe closing and opening transitions, which typically took approximately a hundred nanoseconds.

Most tests were performed with a bias load of 300Ω or less between the relay and the power supply (Figure 91). This load simulates the drive requirements of another microrelay and thus allows the results to be used in the design of systems where relays drive each other.

FIGURE 91. **Typical transient testing setup. The bias resistor was usually less than 300Ω.**

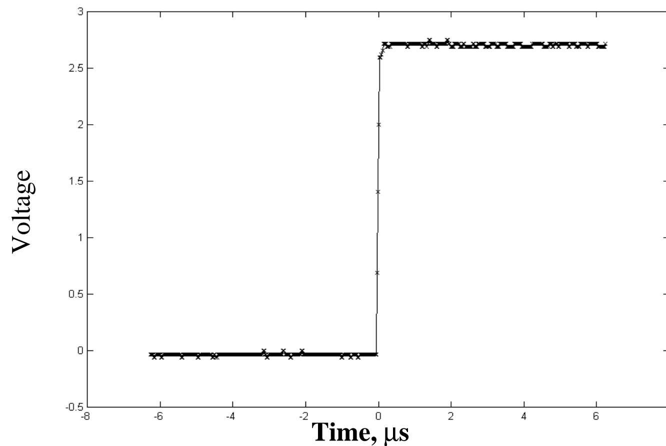


## 42.2 Looking for bounces and arcs

The way to recognize an arc during closing is by the characteristic “arcing voltage” formed during the cascade. This voltage remains fairly constant over the development of the arc and is due to a space charge of ions between the relay plates (Figure 16 in section 7.2.1) and gives us a characteristic behavior to look for to detect closing arcs.

Since closing arc space charge regions typically evolve over the scale of many microseconds this was the range first explored. Neither arcing nor bouncing were observed (Figure 92).

FIGURE 92. **Microsecond scale capture of relay closing across 2.75 Volts (resolution is 25ns). No arc develops and the relay takes about 0.1μS to close fully. There is no bouncing.**




Relay opening proved similarly abrupt and devoid of arcing. Great efforts were required to eliminate noise and drive jitter, as these were the main sources of error

at higher time resolutions. Transient captures with time resolutions of 2.5nS were then possible (Figure 93 and Figure 94). Neither trace shows the characteristics of arcing that have been shown in the literature for macro or MEMS relays.

**FIGURE 93. Relay closing a 2.25 Volt source, plot time resolution is 2.5ns, bias resistor is 750Ω. Timescales and behaviors are wrong for our understanding of arcing, it is possible that the change in slope represents an elastic-plastic deformation transition as discussed in section 29 occurring dynamically. Final resistance is below 3Ω.**

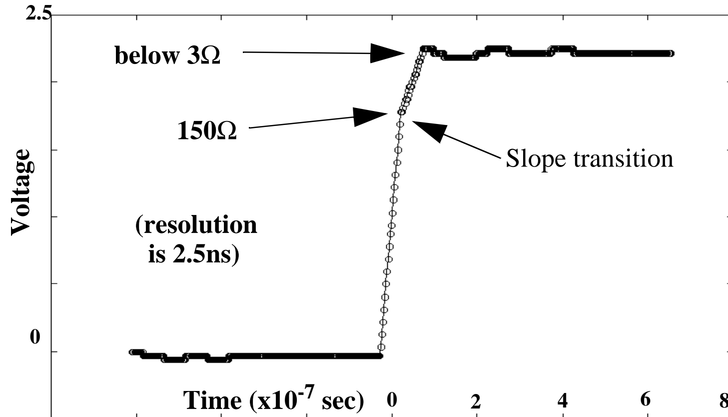
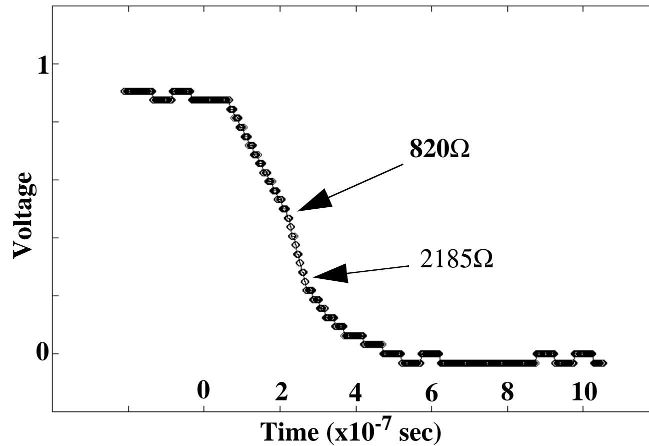



**FIGURE 94. Transient plot of relay opening under 0.9 Volt load. Several zones of behavior are seen. the Initial straight regimes would be expected**

**from mechanical deformation and necking, the final exponential voltage decrease may be structural necking or it may be a metal vapor arc.**




## 42.3 Transient performance characterization

Throughout the transient testing effort was made to track the delay between onset of drive voltage and relay function. The best results were achieved by setting the system up so that separating two banana plugs by hand caused the drive transition, this method (essentially using a bounceless relay transition) gave fairly consistent results for each relay. Fabrication differences, particularly in the thickness of the electroless gold, led to turn-on times for different relays anywhere from 5-100 µs. Turn off times were longer but more consistent, averaging 0.5mS. A typical relay closing time plot is show in Figure 95 and a similar plot of relay opening is shown in Figure 96.

FIGURE 95. **Plot of relay closing response. Delay from drive voltage application to transition is 12µs. The turn-on transition took 105ns. This test was done with an angled contact type II relay with 6 pairs of actuators.**

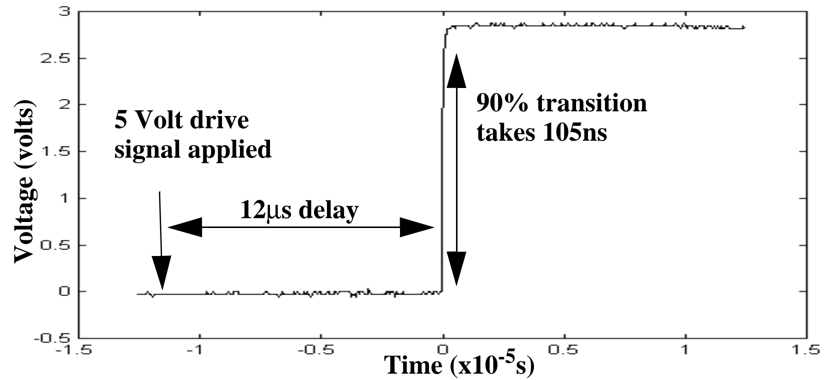
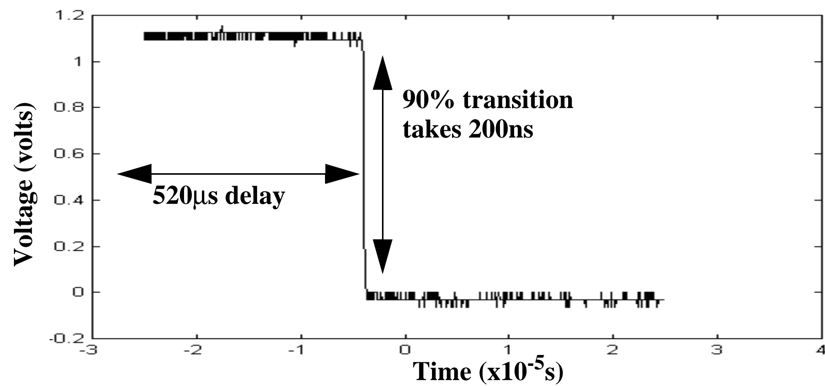


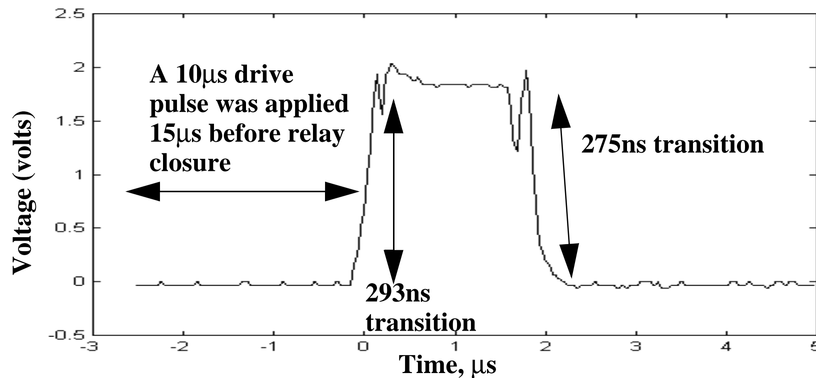

FIGURE 96. **Relay opening response, same type of relay as in Figure 95. Delay from deactivation of actuators to transition is 520µs. Turn-off transition took 200ns.**




Additional tests were conducted using progressively shorter drive pulses. The plot in Figure 97 shows the results of a ten microsecond drive pulse was to have the relay close for two microseconds. It is assumed that this represents a contact just barely making contact and the behavior around turn-on and turn-off presumably are

due to this. The voltage behavior at turn-on shows some similarities to what we would expect from a closing arc (Figure 16) but the time scale is to short for this to correspond to our understanding of arc evolution. At this time these features remain unexplained.

**FIGURE 97. Short-pulse relay activation. By narrowing the drive signal to a short pulse the relay can be made to just barely close for times on the order of microseconds. This test was conducted with an angled type II relay having 6 actuator pairs. The voltage behavior in the transitions is not as clean as previous tests, the cause for this has not been determined.**



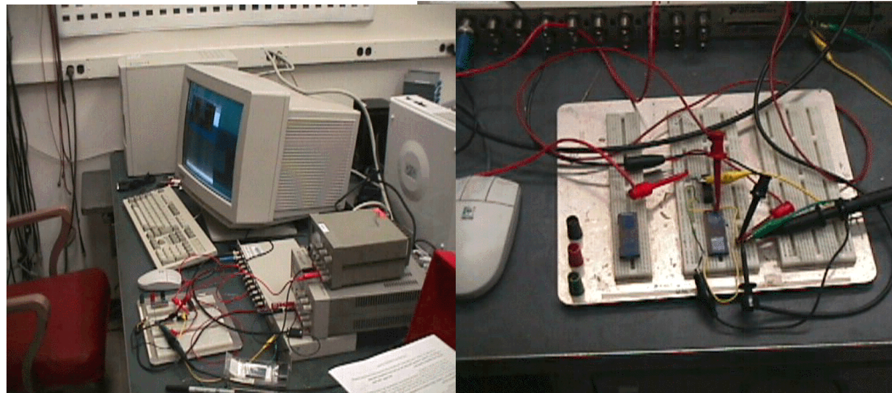

# 43 Cyclic testing

## 43.1 Test setup

In order to test the long term behavior of the relays a computerized test setup was built to cycle the relays and record their performance. The system was built using a Windows NT platform with a National Instruments AT-MIO-16E card with shielded BNC-2090 breakout box. Code was written which closed the relay, veri-

fied it was closed, opened the relay, then verified it was opened - generating about 50 bytes of data per measured cycle. It was quickly realized that quality relays would yield files too large to be analyzed if data were taken every cycle so the code was programmed to take data on a logarithmic basis. Care also had to be taken to give the program appropriate behavior during relay failure as the file buffering functions of Windows NT can lose a lot of data if the program is not shut down promptly.

FIGURE 98. **The long term cyclic testing setup and breadboard. Left side photo shows the entire setup including computer and support electronics. Right photo shows a packaged relay under test.**

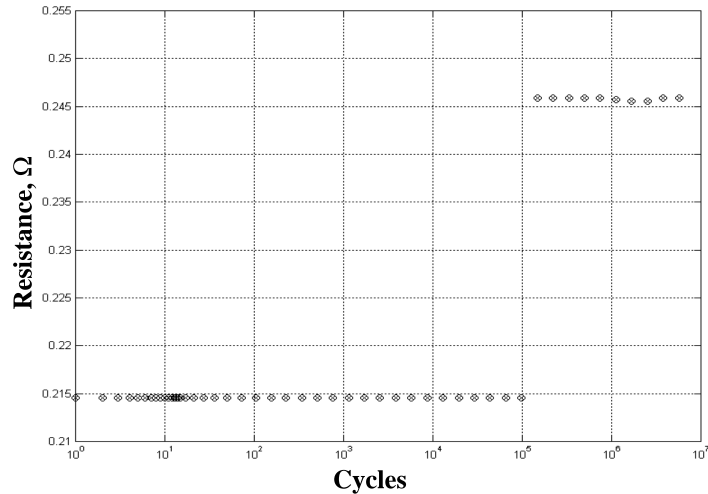


## 43.2 Long term test results

At each of the logarithmically spaced recorded data points taken during testing the cycle number, reference voltage, and a measurement voltage were recorded for both the open and closed state. A bias resistor of 50Ω was used, but resolution was still high due to the excellent quality of the National Instruments equipment. For

reasons of yield that have been discussed few devices survived all the way through fabrication and packaging to be tested for the long term, and much data was lost before the necessity for special file handling within Windows NT was understood. Nonetheless very good long term hot-switching data was captured for two packaged relays. One demonstrated a lifetime greater than 19 thousand cycles (thus it failed between there and about 30 thousand cycles, which would have been the next recorded cycle) while the other lasted between 5.7 and 9 million cycles (Figure 99). Both of these were angled contact type II relays with 6 pairs of thermal actuators.

**FIGURE 99. Long term cyclic testing results for a packaged microrelay. The resistance includes all parasitics such as the breadboard, package, and support structure. Relay was hotswitched at 4 volts and ran 80mA with a 50Ω bias resistor. Relay was verified as open and closed at every cycle or the program would have aborted, but data was only collected on a logarithmic basis to allow for file handling. Relay failure was between 5.7 and 9 million cycles.**



It should be noted that the data shown in Figure 99 was taken over the course of four days on a relay in an unsealed package, making for an average cycle time of 100ms (limited by the software and data collection). The device was exposed to ambient conditions with temperatures from 14C to 23C and humidities from 15% to 28%.

This performance forms an interesting contrast with earlier devices. The very first lateral designs with perpendicular contact areas lasted around 10-20 cycles and the early angled contacts lasted 80-300 cycles. After doing much of the analysis presented in the contact physics chapter it was determined that raw wear and Townsend arcing would be expected to be the dominant wear mechanisms and thicker gold would help both of these. After thickening the gold with electroless deposition relays have now been observed operating for many millions of cycles. This improvement demonstrates the importance and power of doing in-depth physical analysis of the functional mechanisms.

## 44 Conclusions

The electrical test methods presented here were used to test the performance, limits, and durability of fabricated microrelays. This becomes the ultimate test of the physics and design experiments that were done earlier as theory is faced with blunt reality. In this case the results were good with the relays exhibiting sub-ohm

resistances, almost half an amp of maximum static current carrying, and multi-million cycle lifetimes under 80mA hot switching conditions. This represents a great leap forward from the kilo-ohm resistances and twenty cycle lifetimes of the first relays constructed and puts the newer generation of relays squarely into the performance regime of commercial relays at much lower cost, size, and power.

# 45 References

[49] G.P. Box; W.G. Hunter; J.S. Hunter; "Statistics for Experimenters" Wiley Interscience, 1978

CHAPTER 8

# *Relay Systems using Microrelays*

*"Anything is possible if it happens."*

-The Angry Beavers, Nickelodeon cartoon show

## 46 Introduction

The main differentiating quality of microrelays is that they are small. This allows them to fit into smaller packages and new applications, but the microrelays presented here are so much smaller that hundreds of them can be built together as systems. The implications of this are that previously impossible or unusable systems become possible, making this an enabling technology.

## 47 Simulation and the lack of CAD tools

Before large systems are fabricated it is wise to simulate them. In this case the simulation proves difficult because of the many different mechanisms relating to material softening and force. Given the uncertainty of the absolute values of some of

these behaviors no attempt was made to build a physically correct SPICE model, instead a simple linear model (based somewhat on Figure 103 and similar plots) was used. Various models were constructed to represent relays with different numbers of actuators and with or without insulating bridges. The models were fairly good except that the transient behavior could not be modeled well because the closing times vary too much. Thus the simulations were not good at modeling transient behavior but were very useful in analyzing various logic topologies. All systems were simulated this way before layout.

Layout itself, however, can be very challenging as no commercial tool tried offered appropriate place and route or layout versus schematic capabilities. Cadence, L-Edit, and the OCT toolset were all tried at length with considerable programming. In the case of Cadence and L-Edit the existing tools seem to be hard coded with limits that seem to preclude their use in place and route or LVS of arbitrary structures, although the efforts cannot be considered exhaustive. The OCT toolset did not have such hardcoded requirements but the autorouter tended to make mistakes in our trials. This made layout of large systems difficult and LVS was done by hand numerous times by different individuals. Even in a 12-relay system with interconnect it was not unusual to still be finding errors on the third cycle.

# 48 Digital logic

The idea of performing computation through mechanical means has been familiar to engineers since it was proposed by Charles Babbage in 1834, but when he built a small piece of his Difference Engine I it was at a cost equivalent to that of almost 20 steam locomotives (Figure 100) [50]. Recent work has managed to build the entire Difference Engine [50], providing a subset of the capability of a standard ALU within a volume of approximately twelve cubic meters (Figure 101).

**FIGURE 100. A small piece of Difference Engine I built in 1834 at a cost equivalent to that of 20 steam locomotives. The device is about 1 meter high.**

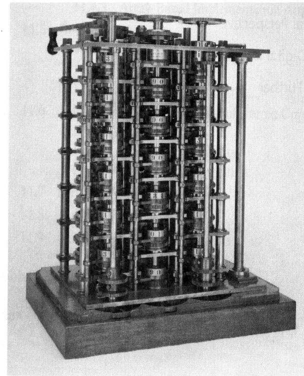
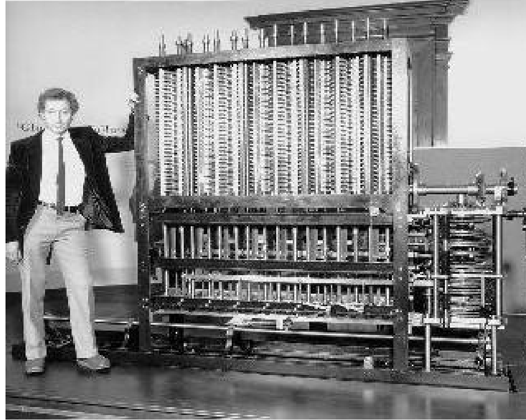


**FIGURE 101.** **A recently constructed full implementation of Babbage's Difference Engine I. Substantial debugging was conducted.**

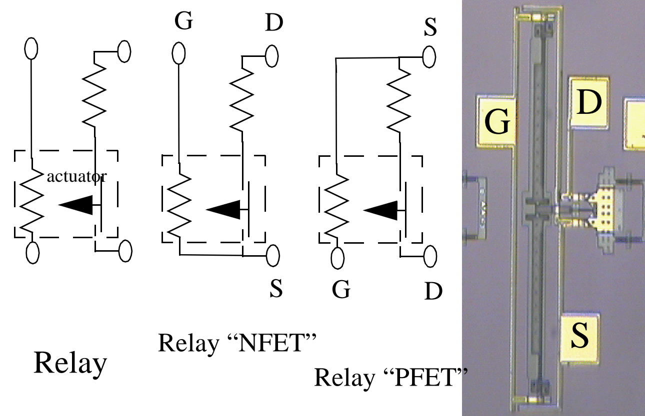


The possibility of high levels of integration and batch fabrication in MEMS allows for reevaluation of mechanical computation to overcome some of the limitations of CMOS such as its low temperature and radiation robustness. Most MEMS devices presented as being able to do logic, however, do so only in one stage when driven by external electronics. In order to do computation and provide a complete logic system a relay based logic gate must be able to drive at least two more of itself (fan-out of 2) and should be repeatable enough to be usable for logic with many gates without generating unacceptable error rates.

# 49 CMOS style complementary logic

The four terminal relays reported here can be used to directly replace standard three terminal FETs in complementary logic merely by wiring one of the inputs to one of the drive electrodes and calling it the source (Figure 102).

**FIGURE 102. Four terminal relays can be wired up to form "N" or "P" three terminal switches. Because the gate voltage is relative to the source voltage P and N switches use the same layout in reverse. Photo at right shows a fabricated implementation.**




These "MEMS-FETs" can be thought of loosely as FETs but they exhibit several differences.

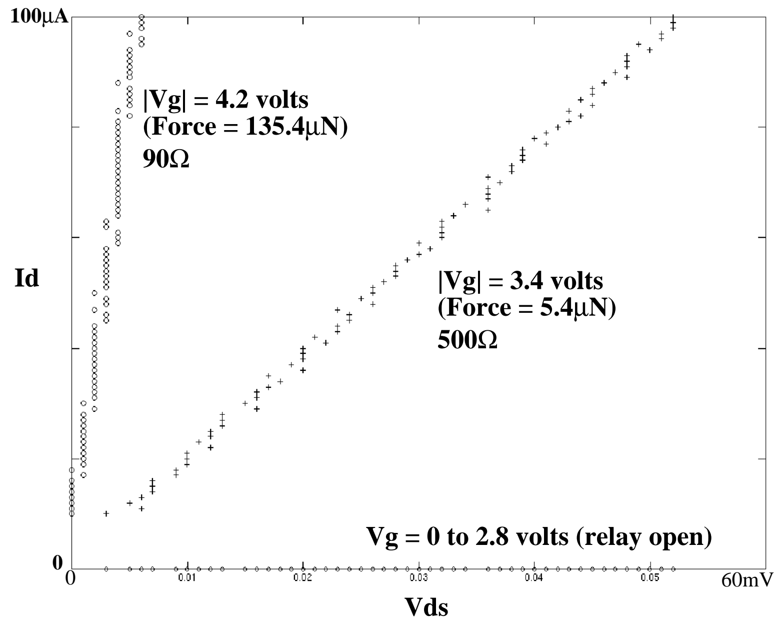
Due to the structure of the device the difference between the layout for an N and a P style device is purely in how it is hooked up. While this makes layout easier and means that fabrication does not get more complicated in order to provide comple-

mentary devices, it also means that great care must be taken when wiring the devices.

Another distinction from FETs is that the on-resistance of the device is a true resistance. Although non-linearities are introduced by the effects of heating, the devices typically exhibit a linear behavior better than 1% full scale at currents below 10mA (which is plenty for most digital logic). Taking a tradition IV plot on the 4145B draws as much current as the 4145b can provide while still showing a relatively small Vds (Figure 103).

FIGURE 103. **I-V plots for various drive values in a microrelay. Linearity better than 2% FS can be seen in the Vds range shown (limited here by the voltage quantization error), attempts to increase the Vds can require very high currents. Device performance is symmetric so drive voltages are shown as absolute values. Drive voltages were converted to force for**

**reference by using the characterization from the actuation chapter. These traces are for a type II angled relay with only a single pair of actuators.**



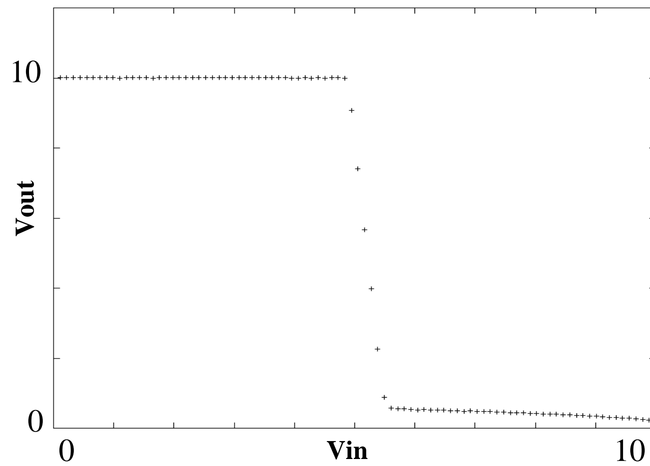

Another very important difference is that the input to the device is resistive in nature. The resistance seen between the source and gate is approximately 780Ω divided by the number of actuator pairs in the devices being used (more actuators typically means higher yield). This has implications for power consumption.

## 49.1 Experimental logic performance

Construction of single gates using these three terminal devices can be done following the same designs as are used for CMOS. Figure 104 shows a typical output trace from a simple complementary device (an inverter) built using this technique.

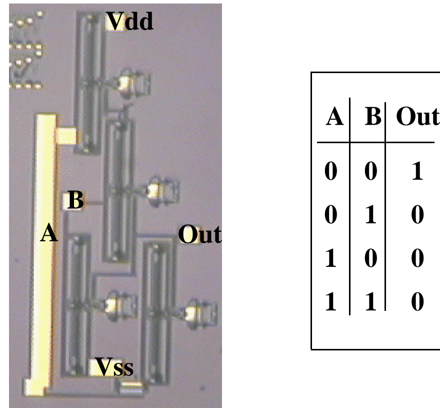
The output was loaded to simulate driving another stage and the high and low values represent voltage dividers between the on resistance of the active device and the input resistance of the simulated next stage. The transition can be observed to happen over a suitably short input range.

**FIGURE 104. Input-Output plot for a complementary electromechanical inverter built using relay based three terminal devices. This device operated on a 10.5 volt power supply and the upper parts of the trace show the maximum 10 volt input range of the 4145b.**




A more complicated gate, a two input NOR gate, is shown in Figure 105. The logical performance was experimentally verified to be as expected.

FIGURE 105. **Fabricated electromechanical NOR gate and experimentally verified truth table.**




| A | B | Out |
|---|---|---|
| 0 | 0 | 1 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 0 |

## 49.2 CMOS style complementary logic analysis and limits

For electromechanical computation each gate must be able to drive one or more equivalent gates. This is an element missing in MEMS based relay logic presented so far, limiting functionality to a single logical operation and making computation impossible [51]. Typical MEMS relay drive problems occur because most MEMS relays small enough to be desirable for logic cannot handle enough current or high enough voltages to drive themselves.

The relays presented here have sufficiently low resistance and high enough current capacities to provide significant fan-in and fan-out performance. The current capability of each relay is typically about 0.48 amps, compared to a drive requirement of 3 to 18mA (approximately 3mA per actuator pair, with various designs using 1 to 6 pairs). This means that if current drive were the only requirement a fan-out of 180

might be expected. The topology of resistively driven complementary logic and constraints of variance in turn-on voltage between devices, however, complicate the issue and calculations are needed to outline a suitable design space.

**FIGURE 106. Circuit diagram for analysis of complementary logic design using relays. Diagram indicates input drive resistance $R_d$, contact resistance $R_c$, fan-in $F_i$, and fan-out $F_o$.**

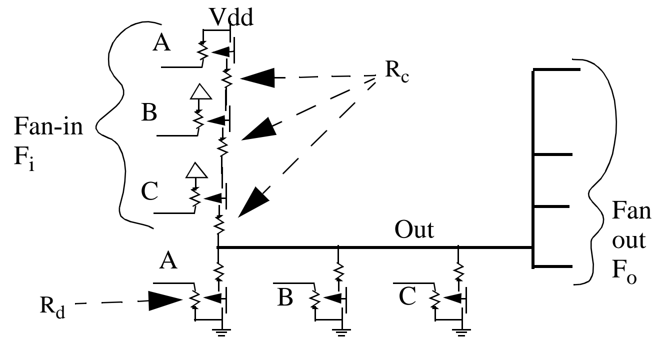
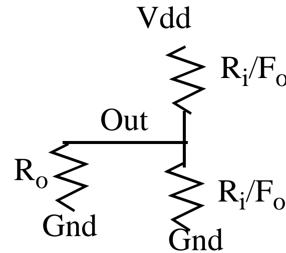



Figure 106 shows an analysis schematic for computing performance limits of relay based complementary logic. One of the biggest problems with directly implementing logic this way with resistively driven relays is that the worst case logic input A in the diagram is hooked through one drive resistor $R_d$ to Vdd and one drive resistor

$R_d$ to ground for each gate that is being driven. This means that the output voltage in the low state (for example) is determined by an equivalent circuit:




where $F_o$ is the fan-out and $R_o$ is the output resistance of the preceding stage and must be assumed to be the worst case and equal to $F_i$ times $R_c$. This voltage will be:

$$V_{out} = \frac{\left(Vdd \cdot \left(\frac{R_i}{F_o} \parallel F_i \cdot R_c\right)\right)}{\left(\frac{R_i}{F_o} \parallel F_i \cdot R_c\right) + \frac{R_i}{F_o}} = \frac{Vdd}{2 + \frac{R_i}{F_i \cdot R_c \cdot F_o}} \quad \textbf{(EQ 42)}$$

Which can be subtracted from the undriven voltage at A:

$$V_{undriven} = \frac{Vdd}{2} \quad \textbf{(EQ 43)}$$

to get one half of the voltage swing. The full voltage swing (difference between a digital one and zero) at node A can be calculated from equation 43 and equation 42:

$$V_{swing} = \frac{Vdd}{2 \cdot \left(\frac{F_i \cdot R_c \cdot F_o}{R_i}\right) + 1} . \quad \textbf{(EQ 44)}$$

If the voltage swing is less than the deviation in the turn-on voltage during any cycle, an incorrect output will be generated. To reduce errors acceptably the voltage swing must be many times the standard deviation of the turn-on voltage for the relays being used.

Another consideration is that the power supply voltage cannot be more than twice the turn-on voltage of the relays or it may cause undefined outputs. This can be seen by observing that point A in Figure 106 will be half the supply voltage if it is not driven at any moment. For robustness this must be modified by the variance of the turn on voltage, otherwise the power rails could drive a particular gate to a fixed value despite input signals. The resultant design restriction becomes:

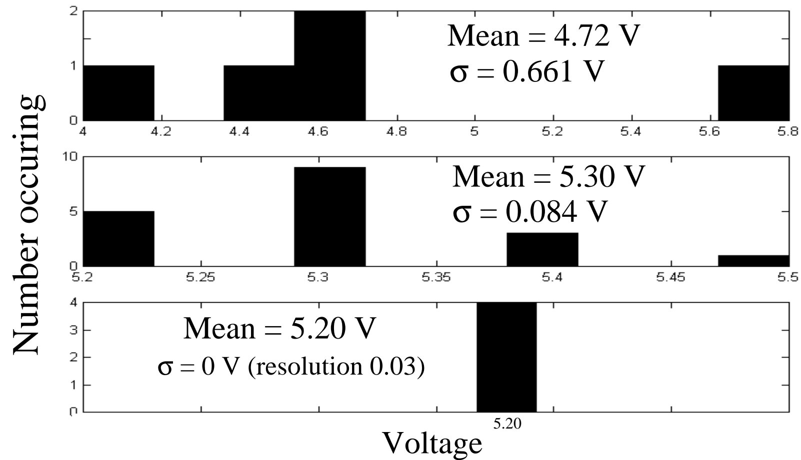
$$Vdd < 2(V_{on} - N\sigma_{von}) \quad \textbf{(EQ 45)}$$

where $V_{on}$ is the voltage to turn on a relay, $\sigma_{Von}$ is the standard deviation of $V_{on}$, and N is a safety factor. As N is increased larger systems become feasible and we use 6 as a safety factor which should allow for systems of hundreds of thousands of gates.

We have characterized the values of $V_{on}$ and $\sigma_{Von}$ in the relays presented here. From tribological considerations we would expect that if $\sigma_{Von}$ is due to asperities in the contacting surfaces it will decrease over time and either remain constant if we have picked forces appropriately to avoid pulling particles (which we have, see section 7.4.2) at the surface or increase again if the forces are excessive. In order to test

this we ran a relay, tracking its $V_{on}$ value at the beginning, middle, and end of its life. The results are shown in Figure 107 where the histogram, mean and standard deviation value $\sigma_{von}$ are indicated for each segment of life.

**FIGURE 107. Turn-on voltage histograms for various parts of a relay lifetime. Top to bottom shows beginning of life, mid-life, and end of life. When evaluated by design equations this shows that the "wear-in" period at the beginning of relay usage can present undesirable probability of errors for a logic system. Relay repeatability must be high.**




The beginning of life is the first few cycles after fabrication, it is separated by several thousand cycles from the mid-life data which is again separated by several thousand cycles from the end of life data which occurred immediately before failure. The measured standard deviation of turn-on voltage for these relays is 0.661 Volts immediately after fabrication, 0.084 Volts after wear-in, and below 0.03 Volts by end-of-life (where resolution was only 0.03 volts due to small number of

measurements). These values can be evaluated for design by rearranging equation 45:

$$N\sigma_{von} < V_{on} - \frac{Vdd}{2}\,. \quad \text{(EQ 46)}$$

If equation 46 is evaluated with the measured values it can be seen that while the relays are acceptable once they have worn-in, their initial performance is such that Vdd cannot be sufficient to drive the system and still satisfy the reliability requirement. Relay based logic systems must exhibit exceptional repeatability and process control before they can be used for large scale computation with six sigma error rejection.

Combining previous equations together to form a design space we find that a given design must exhibit behavior such that:

$$N\sigma_{von} < \frac{Vdd}{\left(\frac{F_i \cdot R_c \cdot F_o}{R_i}\right) + 2} < \frac{V_{on} - N\sigma_{von}}{2\left(\frac{F_i \cdot R_c \cdot F_o}{R_i}\right) + 1}\,. \quad \text{(EQ 47)}$$

In practice the value of $\sigma_{von}$ appears to be set by the smoothness of the final gold deposition (based on the change with cycling), N is set by bit error requirements -- 6 being the number we are using. Once these two values are set Vdd can be chosen using equation 45 or lowered to reduce power expenditure which is:

$$P_{gate} = \frac{Vdd^2 \cdot F_i}{2 \cdot R_i} \quad \text{(EQ 48)}$$

due to the currents on the input resistors (this power value is somewhat gate dependent for gates that implement functions more complicated than a simple NOR or NAND). The resulting equation gives a budget for fan-in and out as a function of contact and input resistance. If this equation is evaluated for an example case of $\sigma_{von}$=0.25 Volts (slightly better than current as fabricated devices), $R_c$= 5Ω (conservative for current devices), and $R_i$ = 187Ω (corresponding to 4 actuator pairs), with Vdd = 9 Volts we get allowable fan-in and fan-out values of 9.7 each. This represents a viable architecture for computational logic. This example is within the implementable range of the relays presented here as long as the relays are allowed to “wear-in” for a while before computation is done.

The example given above demonstrates both the viability of a system based on the technology demonstrated and that it still needs development. The example system just calculated would burn approximately 0.9 watts per 4-input 4-output gate! However, this calculation serves to demonstrate the technological viability and we have made no effort yet at optimizations such as power reduction. Other work [52] has shown that a geometric redesign based on optimization algorithms can reduce thermal actuator power by a factor of 60, for example, and such a change would feed directly into increased freedom with other design variables (including power reduction, of course).

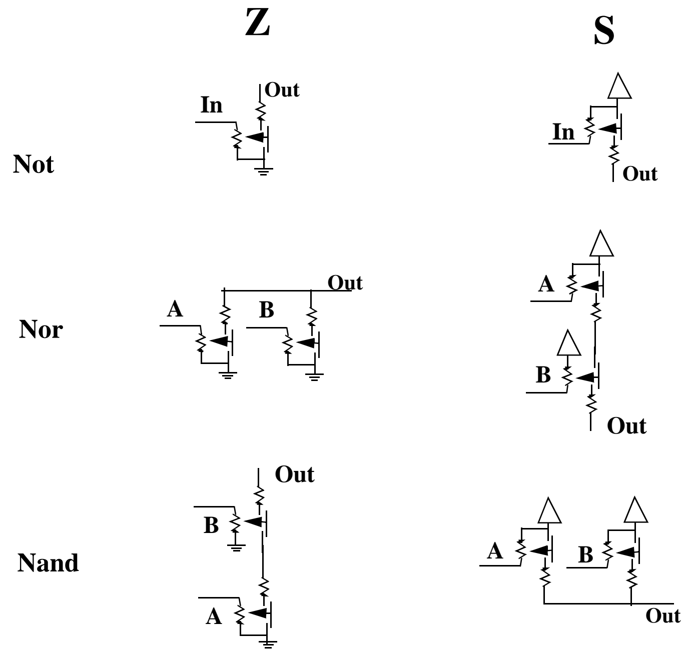
# 50 ZS logic

Architectural changes must also be considered to reduce power and improve performance. One can see from figure Figure 106 that many of the problems in optimization come from the way that each input can find itself resistively coupled to both ground and the supply. The most direct way to reduce power is to use an architecture which does not have this dual-sinking problem. We have been working on such an architecture which essentially builds alternating half-gates which provide only alternately the pull-up or pull-down section of the logic gate and use the resistive input of the next stage as a biasing resistor. Used in proper sequence this eliminates much of the power waste while simultaneously extending the available range of the other design variables. It is referred to here as “ZS” logic for zero and source, the two rails to which a gate can be hooked. This logic seems a more natural and native way of making use of four gate relays for digital logic.

## 50.1 The ZS logic beastiary

ZS logic is designed so that there is a “Z” version of each gate which hooks the next gate to zero (or ground) when activated and an “S” version which hooks the next gate to the source. These gates are then alternated when layout is done to provide a full equivalent to complementary logic. The only difficulty is in hooking together separate pieces of ZS logic where it can sometimes be tricky to arrange the ZS staging so that each element sees the appropriate next element.

FIGURE 108. **A simple ZS logic beastiary. The “Z” elements and “S” elements would be alternated when implementing logic.**

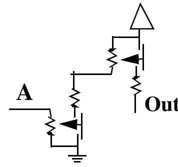



## 50.2 Some fabricated systems

Numerous systems were simulated and then fabricated using ZS logic, unfortunately this is one case where the poor yield was so bad that no systems were successfully tested. This is largely because the ZS logic, being a recent and complicated design, has only been placed on a single fabrication run and therefore not many structures exist to be tested.

FIGURE 109. **A noninverting ZS buffer. This can be driven by a relay with a single actuator pair but has a much lower output impedance because of its**

**additional actuators. While this structure has only two relays and thus a 6% chance of working no functional buffers were found.**



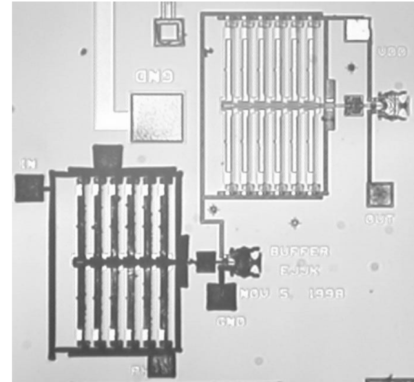




FIGURE 110. **A ZS latch, with 8 relays, presently has a 1 in 65536 chance of working. This device has two functional relays out of 8, exactly the statistical expectation.**

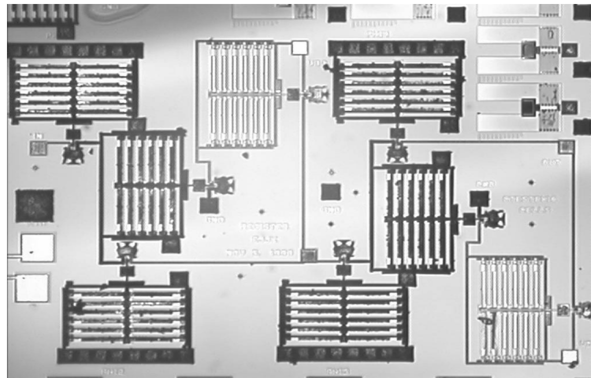
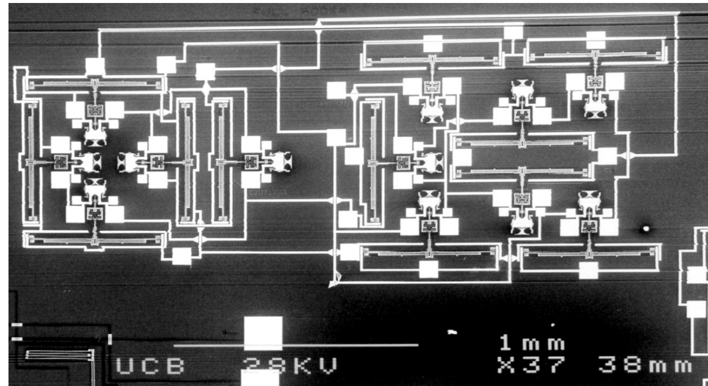


FIGURE 111. **A ZS single bit carry-adder. The complexity is such that yield will have to be substantially improved before this device becomes feasible.**

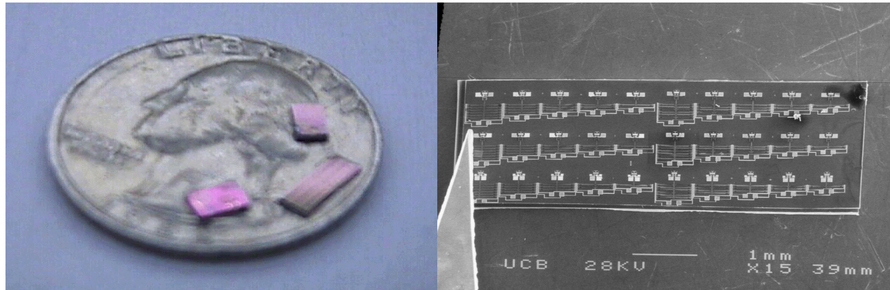



# 51 Conclusion

The photos in Figure 112 serve to give an intuitive feel for the size of systems constructed using the microrelays presented here. Despite their competitive performance with conventional relays they are so small that even large groups of them such as the test chip shown require microscopy to photograph. This gives motivation for further work in this field and improvement of process quality (almost the only requirement to move these relays and their systems to practicality).

FIGURE 112. **A view of the scale of integration possible with our microrelays. The left photo shows (left to right) a chip with a 6 relay analog system and a one bit carry-adder, a chip with 30 relays in test arrays, and a chip with two 6 relay systems. The right photo shows a scanning electron**

**micrograph of the middle chip with 30 test relays. This will give the reader a feel for potential die sizes.**




## 52 References

[50] D. Swade, “Redeeming Charles Babbage’s Mechanical Computer,” Scientific American, pp86-91, February 1993

[51] A. Hirata, K. Machida, H. Kyuragi, M. Maeda; “A Micromechanical Switch as the Logic Elements for Circuits in Multi Chip Module on Si (MCM-Si),” Proceedings of the 1999 International conference on Micro Electro Mechanical Systems, Orlando FL, January 17-21 1999 pp582-587

[52] J. Jonsmann, O. Sigmund, S. Bouwstra; “Compliant Electro-Thermal Microactuators” Proceedings of the 1999 International conference on Micro Electro Mechanical Systems, Orlando FL, January 17-21 1999 pp588-593

# *Complete References*

[1] Semiconductor Business News, Monday, April 14, 1997 Vol. 5, No. 43 10

[2] Electronic Buyers' News, September 14, 1998 10

[3]Wallinder, I. O.; Eriksson, P; "Characterization of Corrosivity of an Automobile Environment," In Proc. 19th International Conference on Electrical Contact Phenomena; 14-17 Sept 1998 Nuremburg, Germany, pp 157-161 10

[4]Weiser, J; "Switching Contacts in Relays," In Proc. 19th International Conference on Electrical Contact Phenomena; 14-17 Sept 1998 Nuremburg, Germany, pp 15-21 10

[5]Walczuk, E.; et al. "Computer Controlled Investigations of the Static Welding of Silver Tin Oxide Contacts" In Proc. 19th International Conference on Electrical Contact Phenomena; 14-17 Sept 1998 Nuremburg, Germany, pp 311-316 65

[6]Hauner, F.; Jeannot, D.; Pinard, J. “AgFe and AgFe2O3 Contact Materials for Low Voltage Switchgear” In Proc. 19th International Conference on Electrical Contact Phenomena; 14-17 Sept 1998 Nuremburg, Germany, pp 317-323 65

[7]Schopf, T.J.; Hundt,I; “Effects of Different Loads on the Surface of Silver Metal Oxide Contacts for General-Purpose Relays” In Proc. 19th International Conference on Electrical Contact Phenomena; 14-17 Sept 1998 Nuremburg, Germany, pp 83-87 65

[8]Schoft, H; Lobl, H; Kindersberger, J; “Creep-Ageing [sic] of Bolted Electrical Busbar Joints” In Proc. 19th International Conference on Electrical Contact Phenomena; 14-17 Sept 1998 Nuremburg, Germany, pp 269-276 65

[9]Pashley, M.D.; Pethica, J.B.; Tabor, D “Adhesion and Micromechanical properties of metal surfaces,” Wear, 100, 1984 pp7-31 65

[10]Shackelford, J.F.; “Introduction to material science for engineers,” Macmillan Publishing Company, 1988 65

[11]Korolev, Y.D.; Mesyats, G.A. “Physics of pulsed breakdown in gases [sic],” URO-Press 1998 65

[12]Johler, W; “Extended life of Telecom relays by using SF6” In Proc. 19th International Conference on Electrical Contact Phenomena; 14-17 Sept 1998 Nuremburg, Germany, pp37-44 65

[13]Holm, Ragnar; “Electric Contacts,” Springer-Verlag 1941 65

[14]Sharma, S.P.; “Adhesion coefficients of plated contact materials” Journal of Applied Physics, V 47 No. 8, August 1976 pp 3573-3576 65

[15]Landau, L.D.; Lifshitz, E.M.; “Theory of Elasticity” Pergamon Press, 1986 66

[16]Barrett, C.R.; Nix, W.D.; Tetelman, A.S.; “The Principles of Engineering Materials,” Prentice Hall, 1973 66

[17]Pashley, M; Pethica, J; Tabor, D; “Adhesion and Micromechanical Properties of Metal Surfaces” Wear 100, 1984, pp 7-31 66

[18]Collard, S.M.; McLellan, R.B.; “High-temperature elastic constants of gold single-crystals” Acta Metallica Materials, V 39 No. 12, 1991, pp 3143-3151 66

[19]Antler, M; “Tribology of metal coatings for electrical contacts” Thin Solid Films, Vol 84 No 3, 1981, pp245-56 66

[20]Rabinowicz, E; “Friction and Wear of Materials”; 1995, Wiley-Interscience 66

[21]Weiser, J; “Switching Contacts in Relays, ”In Proc. 19th International Conference on Electrical Contact Phenomena; 14-17 Sept 1998 Nuremburg, Germany, pp15-21 66

[22]MCNC Corp, Now Cronos Integrated Microsystems, Inc. 3021 Cornwallis Road, P.O. Box 12889, Research Triangle Park, NC 27709-2889, USA; Ph: (919) - 248 - 1983 FAX: (919) - 248 - 1434; http://mems.mcnc.org 100

[23]CMP [Circuit Multi-Projects]; 46 avenue Felix Viallet; 38031 Grenoble Cedex; France; FAX: +33 4 76 47 38 14; email: cmp@imag.fr; http://tima-cmp.imag.fr/CMP/CMPInfo.html [english version] 100

[24]Kruglick, E.J.J.; “Aluminum hinges and Electrostatic Actuators in Standard CMOS,” Masters thesis, available for reorder through UMI; 300 North Zeeb Road, PO Box 1346, Ann Arbor, MI 48106-1346, USA; Ph: (800) 521-0600; http://www.umi.com 101

[25]http://mumps.mcnc.org 101

[26]Kovacs, G.T.A.; “Micromachined Transducers Sourcebook” McGraw-Hill, 1998 101

[27]“Gold Plating Solution - Bright Electroless Gold”; Transene Company, Inc.; Danvers Industrial Park; 10 Electronics Ave; Danvers, MA 01923; Ph: (508) 777-7860 101

[28]Okuna, K; Otsuka, K; Hattori, N; Torikai, E; “Auto catalytic electroless gold plating,” The Electrochemical Society Proceedings, Vol 94 No 32, pp 164-185 101

[29]Mulhern, G.T.; Soane, D.S.; Howe, R.T.; “Supercritical Carbon Dioxide Drying of Microstructures” In Proc. 7th International Conference on Solid-State Sensors and Actuators (Transducers ‘93), June 7-10, 1993, Yokohama, Japan pp 296-299 101

[30]Spectrum Semiconductor Materials, Inc.; 2027 O’Toole Ave.; San Jose, CA 95131; Ph (408) 435-5555 Fax (408) 435-8226; http://www.spectrum-semi.com 101

[31]Method developed by Kevin Flynn, passed along by Jim Bustillo, UC Berkeley microlab 101

[32]R. Yeh, E.J.J. Kruglick, K.S.J. Pister; "Surface-Micromachined Components for Articulated Microrobots"; Journal of Microelectromechanical Systems, vol. 5, no. 1, March 1996, pp 10-17 126

[33]Majumder, S; McGruer, N.E.; Zavracky, P.M.; Adams, G.G.; Morrison, R.H.; Krim, J; “Measurement and modeling of surface micromachined, electrostatically actuated microswitches” 1997 International Conference on Solid-State Sensors and Actuators, June 16-19, 1997 (Transducers 97), pp 1145-1148 126

[34]Schiele, I; Huber, J; Evers, C; Hillerich, B; Kozlowski,F "Micromechanical Relay with Electrostatic Actuation" 1997 International Conference on Solid-State Sensors and Actuators, June 16-19, 1997 (Transducers 97), pp 1165-1168 126

[35]Comtois, J.H. "Structures and techniques for implementing and packaging complex, large scale microelectromechanical systems using foundry fabrication processes." Ph.D. dissertation, Air Force Inst. of Technology, 12 May 1996 126

[36]Barrett, C.R.; Nix, W.D.; Tetelman, A.S. "The Principles of Engineering Materials," Prentice Hall, 1973 126

[37]Courtney, T.H. "Mechanical behavior of materials", McGraw Hill 1990 126

[38]Guckel, H.; Klein, J.; Christenson, T.; Skrobis, K.; Laudon, M.; Lovell, E.G. "Thermo-magnetic metal flexure actuators" Technical Digest. IEEE Solid-State Sensor and Actuator Workshop, Hilton Head Island, SC, USA, 22-25 June 1992 127

[39]Ananthasuresh, G.K.; Kota, S.; Gianchandani, Y. "A methodical approach to the design of compliant micromechanisms" Technical Digest. Solid-State Sensor and Actuator Workshop, Hilton Head Island, SC, USA, 13-16 June 1994 127

[40]Conant, R.A.; R.S. Muller: "Cyclic Fatigue Testing of Surface-Micromachined Thermal Actuators." Proceedings of the 1998 ASME International Mechanical Engineering Congress, Nov. 15-20, 1998, Anaheim CA, pp. 273-277 127

[41]J. Jonsmann, O. Sigmund, S. Bouwstra "Compliant Electro-thermal Microactuators" Proceedings 12th annual international conference on microelectromechanical systems, Orlando FL, USA, 1999 127

[42]Yeh, R; Pister, K.S.J. “Measurement of static friction in mechanical couplings of articulated microrobots” Proceedings of the SPIE - The International Society for Optical Engineering vol 2642, 1995, pp40-50 149

[43]Kruglick, Ezekiel J.J.; Pister, Kristofer S.J. "Micronewton Contact Characterization for MEMS Relays,"; 19th International Conference on Electric Contact Phenomena, Nuremberg Germany, Sept. 1998, pp 53-57 149

[44]Oliver, W.C.; Hutchings, R.; Pethica, J.B. “Measurements of hardness at indentation depths as low as 20 nanometers”, Microindentation Techniques in Materials Science and Engineering, ASTM STP 889, 1986, p.7637-42 149

[45] Beale, J.P.; Pease, R.F.W. “Apparatus for studying ultrasmall contacts” Electrical Contacts - 1992. Proceedings of the Thirty-Eighth IEEE Holm Conference on Electrical Contacts, Philadelphia, PA, USA, 18-21 p.45-9 150

[46]Buckley, Donald H. “The use of analytical surface tools in the fundamental study of wear” Wear 46, 1978, pp 19-53 150

[47]Abe, T; Messner, W.C.; Reed, M.L.; “Effective Methods to Prevent Stiction During Post-Release-Etch processing” In Proc. IEEE Micro Electro Mechanical Systems Conference, Jan 29 - Feb 2 1995, Amsterdam, The Netherlands, pp 94-99 160

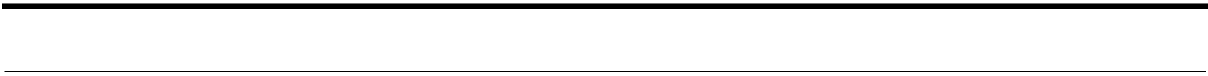
[48]Provided by MCNC Corp, Now Cronos Integrated Microsystems, Inc. 3021 Cornwallis Road, P.O. Box 12889, Research Triangle Park, NC 27709-2889, USA; Ph: (919) - 248 - 1983 FAX: (919) - 248 - 1434 160

[49]G.P. Box; W.G. Hunter; J.S. Hunter; “Statistics for Experimenters” Wiley Interscience, 1978 184

[50] D. Swade, “Redeeming Charles Babbage’s Mechanical Computer,” Scientific American, pp86-91, February 1993 204

[51]A. Hirata, K. Machida, H. Kyuragi, M. Maeda; “A Micromechanical Switch as the Logic Elements for Circuits in Multi Chip Module on Si (MCM-Si),” Proceedings of the 1999 International conference on Micro Electro Mechanical Systems, Orlando FL, January 17-21 1999 pp582-587 204

[52]J. Jonsmann, O. Sigmund, S. Bouwstra; “Compliant Electro-Thermal Microactuators” Proceedings of the 1999 International conference on Micro Electro Mechanical Systems, Orlando FL, January 17-21 1999 pp588-593 204



# Index


**Numerics**

1-Form-A 2
1-Form-A relay 28
4145b 161

**A**

abrasive 52
actuation 7
Actuator design requirements 126
actuator performance 103
adhesion layer 16
aggregation 14
air drying 86
air-dry 155
aligner 85
aluminum wire 87
ambient conditions 183
angled contacts 157
Angry Beavers 185
arc initiation 20
arcing behavior 26
arcing voltage 176
a-spot 41
a-spots 138
asymmetric force application profile 106
atomic separation distance 107
averaging 166

**B**

Babbage 187
beastiary 200
bias resistor 175
bias spring 105
bimaterial auto-break 174
binding energy 108
bloat 93
Bond pulling 97
borane dimethylamide 75
buckling 152
Buddhist Scripture 67
burnout 173

**C**

CAD tools 185

Cadence 186
calculating slope 168
Chaim Weizmann 151
chemical replacement solutions 73
col 74
commercial relays 2
complementary logic analysis 193
compliant structure 114
Constriction resistance 39
contact bounce 28
contact points 138
contact resistance 39
contact resistances 172
contact spot population 146
contact spots 130
contact topology 157
contaminants 94
cooling time 112
Corrosion 19
corrosion products 52
crack propagation 139
Creep 18
critical field 20
critical point release 86, 155
cross section 86
current density 39
Cyclic testing 180

**D**

defect feature 16
defects 16
dehydration bake 85
delay 178
dendritic growth 75
design space 129
developed 85
Die attach 87
Difference Engine I 187
diffusion 15
diffusion barrie 16
Digital Logic 187
dissipated power 44
DMAB 76
dual-sinking problem 200
ductile pulling 139

**E**

EDS 80
efficiency metric 9

elastic modulus 108
elastic region 134
electrical isolation bridge 153
Electroless gold deposition 72
Electroless plating solutions 73
electromechanical computation 193
Electroplating 72
Electrostatic actuation 105
emissions of X-rays 80
enabling technology 185
energy dispersive X-ray spectroscopy 80
Environmental testing 5
epoxy reflow 89
equicurrent line 134
error at slope fit 168
estimating separation force 138
Exposure 85

**F**
Failure rate 4
fan-in 193
fan-out 193
FETs 189
Fg 116, 119
force due to generation of new surface area 141
force of closure 120
foundry 67
Fretting 50
fully plastic behavior 131

**G**
gap-closing actuators 105
gold 16
Gold bridging 90
gold depletion 82
gold resistivity 104
gold to gold contact 138
gold universal hardness 104
GPIB 174
guides 156

**H**
hermetically sealed 5
Hewlett Packard 4145b 161
Holding voltage 19
humidities 183

**I**
I-line 85

INA103  163
increasing numbers of actuators  123
inductive load  22
Insulating mechanical bridges  84
interatomic length increase  113
internal spring constants  122
internal spring force  106
inverter  191
ions  20
isopropanol  86
IV plot  190

**J**
jitter  176

**K**
Ki  116
kinetic energy  28

**L**
lateral contacts  157
layout versus schematic  186
lead frames and caps  89
L-Edit  186
lifetime  182
lift-off process  90
linear superposition model  116
long term cycling limits  124
Long term test results  181
Louis Sullivan  129
Ludwig Mies van der Rohe  1
LVS  186

**M**
massive overplating  95
maximum stress  131
MCNC MUMPs  68
measurement indicator  117
mechanical computation  187
mechanical lockup  155
mechanical necking  140
MEMS-FETs  189
meniscus formation  155
meta-stable solutions  73
methanol  86
micro-asperities  11
microscopic wear motion  50
misalignment  156
model of the generated force  121

moment principles 114
motion amplification testing 117

**N**

National Instrument 180
Nickelodeo 185
non-hermetic 5
NOR gate 192

**O**

OCG 825 photoresist 85
OCG 934 developer 85
OCT toolset 186
oscilloscope 174
overetch 93
overshoot voltage 27

**P**

paperclip 89
parameter analyze 161
Particle generation 54
Paschen curve 20
perpendicular contacts 157
photoresist mechanical bridges 84
photoresist processing 85
place and route 186
plastic contact deformation 11
plastic deformation 104, 129
Plastic transition characterization 131
potassium aurous cyanide 75
potassium cyanide 75
powder metallurgy 14
power supply voltage 196
prebake 85
pre-stress 19
process development 67
Protrusion collapses 147
pull-bar test 139
pull-in 105
pulling particles 196

**Q**

quadratic power-voltage relation 120

**R**

radiation robustness 188
rate of contact spot formation 145
reeds 2
reference springs 115

relative humidit 85
relay cycle 12
relay market 1
Reliability 4
resistive film 43

**S**

safety factor 196
sand 15
separate connection spots 144
separation force 138
Sir Henry Royce 11
size 6
slip lines 16
slope method 171
SMU 164
solenoid based power switching 3
Source/Measurement Unit 164
space charge 27
spectroscopy 80
SPICE model 186
standard deviation of Von 196
steam locomotives 187
stepper based lithography 99
Supertemperature 44
surface energy of gold 142
surface mount package 89
surface mount relays 1
surface texture 82

**T**

technology transfer 67
temperature coefficient of resistance 44
temperature dependence of resistance 44
temperature factor t 41
Temperature induced softening 46
temperatures 183
Thermal actuation 106
Thermal aging 49
thermal bimorph structures 106
thermal conductance 44
Thermal failure 173
thermosonic bonder 87
three terminal devices 191
time resolution 177
transient behavior 186
Transient performance characterization 178
transient testing setup 175

**U**
Unconstrained macroscopic thermal expansion 112
Unconstrained thermal expansion 113
Unwanted gold deposition 93

**V**
vernier 131
voltage swing 196

**W**
wafer saw 86
wear-in 199
weld-failure 50
Welding 49
Westbond 800 87
window size 166
Windows NT 180
Wire bonding 87
worn-in 198

**X**
X-ray microscopy 139

**Y**
Yield 90
yield stress 46
yield stress of gold 141

**Z**
zero-travel force 116
ZS logic 200