ABSTRACT

DUEWER, BRUCE ELIOT. A Low-Power, High Performance MEMS-based Switch Fabric. (Under the direction of Paul D. Franzon.)

An approach with the potential for building large low power high performance crossbar networks is presented. Thin film polysilicon MEMS devices are developed to provide crosspoints. These devices are vertically moving plates that serve as variable capacitors. Addressing of large arrays using 2n rather than n-squared lines despite no active circuitry on the MEMS chips is facilitated by bistable device operation. Derivations of equations for bistable device operation are presented. Low power operation is possible as the devices are electrostatically controlled and are stationary except during reconfiguration. Early devices are fabricated using the MUMPS process. The bistability and array addressability properties are demonstrated. The substrate effect on device operation is measured and modeled; methods for utilizing the substrate effect to tune device operation are presented. Later devices are fabricated using the SUMMiT process. Changes in the SUMMiT design rules to increase allowable vertical motion range are proposed and designs using them fabricated. S-parameter characteristics of devices in both 'on' and 'off' states are measured. Addition of metallization after chip fabrication and release is necessary to lower the resistance of interconnect. A self masking method for applying this metallization allowing for decreased resistance at line crossings

is proposed. This method is tested using each of sputtering and evaporation as the deposition technique for a gold and adhesion layer stack. Effectiveness of the method with each technique is evaluated. Chips suitable for providing high voltage control for large MEMS arrays are fabricated in a 2um feature size CMOS process. Architectures suitable for building large crossbars employing variable capacitor arrays are discussed. Optimization of hybrid CMOS/MEMS Clos arrays on the basis of criteria other than minimization of crosspoints is discussed. Array sizings to provide 192*192 and 256*256 crossbars are presented, and software examples for sizing and controlling Clos networks are provided. Evaluation of the suitability of the MEMS devices developed for use as digital or broadband crosspoints is evaluated, and potential future directions are proposed.

A LOW-POWER, HIGH PERFORMANCE MEMS-BASED SWITCH FABRIC

by

BRUCE ELIOT DUEWER

A dissertation submitted to the Graduate Faculty of

North Carolina State University

in partial fulfillment of the

requirements for the Degree of

Doctor of Philosophy

COMPUTER ENGINEERING

Raleigh

2001

APPROVED BY

__________________________ __________________________

__________________________ __________________________

Chair of Advisory Committee

BIOGRAPHY

Bruce Duewer, son of Lawrence and Elaine Duewer, was born on August 21, 1973 in Fairfax, Virginia. He has a younger brother, Trent, and sister, Rachel. He graduated from the Thomas Jefferson High School for Science and Technology in 1991, and became an Eagle Scout the same year. In 1995 he received a BS from Virginia Tech, with a double major in Electrical Engineering and Computer Engineering, and a minor in Computer Science. In 1996 he received a MS in Computer Engineering at North Carolina State University.

Bruce has been an active member of the Lutheran Student Movement throughout his college career, and has served as an officer at the local and regional levels. He has also served as a senior youth advisor at Holy Trinity Evangelical Lutheran Church in Raleigh.

ACKNOWLEDGEMENTS

I would like to thank Dr. Paul Franzon for his service help and guidance through my time at NCSU. I appreciate how despite an incredibly busy schedule, he always made time when I needed to discuss something with him, yet was patient with the various delays and challenges I encountered with the project, and made sure I had the resources necessary to explore solutions. I'd like to also thank the rest of my committee for their support and advice: Dr. Thomas Conte, Dr. Wentai Liu, and Dr. Jacqueline Krim. Also, Dr. Clay Gloster for serving on my committee until he moved out of the state.

David Winick taught me a lot about MEMS, and was a lot of help throughout the project. He helped me scale the learning curves involved in moving from the fairly unrelated areas that had been my focuses in earlier degrees to mechanical design. He also prepared and maintained much of the infrastructure for design and testing of MEMS here, adapting the tools (for example, he wrote an excellent script for cross section extraction that I used extensively), and in specifying and assembling the test equipment we used. We worked together on the bistability for array addressability scheme, which provided the foundation for much of what I developed.

Alan Glaser and Toby Schaffer also provided a lot of help with the design tools, particularly in the design of simulation of electronic circuits.

John Wilson designed the driver and receiver chips, as well as the CMOS crossbars for the hybrid Clos arrays. John also designed the high voltage

driver stage of the control chip. He and Steve Lipa helped a lot with high frequency testing, and with my understanding of high frequency testing issues and test structures. The circuit model for a bank of MUMPS devices discussed in section 3.7 was developed by John Wilson (my earlier models had only resistance and capacitance components), and he did the fitting of component values with the test data.

David Nackashi, Andrew E. Oberhofer, and Jeff Chang all helped me in the cleanroom, preforming the evaporation depositions. JMC Tool & Machine manufactured the steel mounts for the evaporator. David Nackashi and John Damiano provided advice on process issues involved with metallization.

I'd like to thank all my fellow ERL students. They are very talented bunch, good people to either share an office or go rock climbing with.

Finally, I'd like to thank my family for their love and support.

TABLE OF CONTENTS

	Page
LIST OF TABLES	ix
LIST OF FIGURES	x
Chapter 1. Introduction	1
1.1 Motivation	1
1.2 Goals	2
1.3 Overall Approach	2
1.4 Crossbar Scaling	3
1.5 Thin Film MEMS	4
1.6 Challenges & Novel Contributions	6
1.7 Target Crosspoint	8
1.8 Outline of Dissertation	9
Chapter 2. Literature Review	11
2.1 Architectural Issues	11
2.1.1 Switching Networks	12
2.1.2 Clos Networks	13
2.1.3 Programming Clos Networks	16
2.2 Circuit Issues	19
2.2.1 Other crossbars	20
2.2.2 Resistors	20
2.2.3 Automated Bias Control (ABC) Circuit	23
2.2.4 Optical Receiver	25
2.2.5 Quantized Feedback	29
2.2.6 Conclusions Regarding Signaling	31
2.2.7 High Voltage Generation	31
2.3 Device Design	32
2.3.1 Silicon as a Mechanical Material	32
2.3.2 Micromachined Relays	33
2.3.3 Spring Design	34
2.3.4 A Previous Bistable Device	36
2.3.5 An RF and Microwave device	37
Chapter 3. Device Operation & MUMPs	39
3.1 MEMS Switch	39

3.2 Bistability & Array Addressing 41
3.2.1 Derivation of Bistability 42
3.2.2 Array Addressability 44
3.2.3 Stop Gap Height Implications on Bistability 45
3.2.4 Device Throw Implications For Bistability 47
3.2.5 Requirements For Addressibility 47
3.2.6 Experimental Confirmation of Bistability 48
3.2.7 Experiment Details 50
3.3 Substrate Effects on Actuation 53
3.4 MUMPs Designs 55
3.4.1 MUMPs 14 and 18 55
3.4.2 MUMPs 20 58
3.4.3 Later MUMPs runs 60
3.5 Modeling 61
3.5.1 Matlab 61
3.5.2 FastCap 63
3.5.3 MEMCAD 63
3.6 S-parameter results for example device 65
3.7 S-parameter results for a bank of devices 67

Chapter 4. Design of MEMS Structures for Packaging and Test 69

4.1 Introduction 69
4.2 Basic Thin Film MEMS Design Concerns 69
4.3 Electrical Modeling of MEMS Structures 72
4.4 Transmission lines in MUMPs 74
4.5 Testing Considerations: Probing 77
4.5.1 Design for Probing 77
4.5.2 Probe Specifics 78
4.5.3 Two example probe stations 80
4.5.4 Modeling of probing 81
4.5.5 Doing the probing 82
4.6 Packaging 84
4.7 Conclusions 85

Chapter 5. MEMS Crossbar Architecture 86

5.1 System Description 86
5.2 Crossbar Architecture Selection 88
5.3 Characteristics Of A Direct Implementation 89
5.4 Size Limitations 89
5.5 Clos Array Sizes Selection 90
5.6 High Voltage Drivers 93

5.7 The System 96
5.8 Signal Path 97
5.9 Overall Power Consumption 100

Chapter 6. Metallization 101

6.1 Metallization Options 101
6.2 Layout for Metallization 102
6.2.1 Mechanical Concerns 104
6.2.2 Low Resistance Line Crossings 104
6.3 Metallization (Sputtering): First SUMMiT Run 107
6.3.1 Visual Inspection 107
6.3.2 Electrical & SEM Tests 110
6.4 New Metallization Test Structures 113
6.5 Metallization(Sputtering): Second SUMMiT Run 114
6.5.1 Metallization Test Structure Testing 115
6.5.2 “Bubbling” and the SAM 117
6.6 Metallization of chips from no-SAM release 119
6.6.2 Some Probing Abnormalities Explained 121
6.6.3 Testing 122
6.7 Evaporation 123
6.7.1 Evaporator Geometry 123
6.7.2 Synthetic Spread Angle 124
6.7.3 Initial Sanity Check 127
6.7.4 First Four Depositions 128
6.7.5 Another Evaporation Run 130

Chapter 7. Devices Using SUMMiT 132

7.1 SUMMiT Process Features Relative to MUMPs 132
7.2 Initial Planning 133
7.3 Rule Adjustments 134
7.4 Some Basic Design Trade-offs 137
7.5 Device Design 138
7.5.1 Design to Survive metallization 140
7.6 The First SUMMiT Run 141
7.6.1 Devices Included 141
7.6.2 Test Results 145
7.7 Second SUMMiT Run 146
7.7.1 Devices Included 146
7.7.2 Test Results: Sputtering 147
7.7.3 Test Results: Evaporation 149
7.8 Modeling 151

7.9 Capacitance Ratio 151

Chapter 8. Conclusions & Future Work 154

8.1 Achievements 154
8.2 Remainders 155
8.3 Changes 155
8.4 Directions for Future Work 156

References 158

Apprendix 1. Equation Derivations 163

Apprendix 2. Program Listings & Outputs 169

2.1 Clos Array Combination Generator 169
2.2 Clos Programing Code 170
2.3 Macro Code for FastCap 175

LIST OF TABLES

Page

Table 1 Control for Figure 9 Devices 50
Table 2 Control Voltages 51
Table 3 HighVolt Logic 94
Table 4 Pinout for HighVolt Chip (Pad supplies noted with P) 95
Table 5 Resistances of Transmission Lines after Sputtering 115

LIST OF FIGURES

		Page
Figure 1	Signal Path	3
Figure 2	MUMPs Release Example. Source [39]	5
Figure 3	Resistor Circuit	21
Figure 4	Auto Bias Circuit	24
Figure 5	Restoration of Signal	29
Figure 6	Individual MEMS switch	40
Figure 7	Voltage-Displacement Characteristic (flat plate model)	42
Figure 8	Array Addressing Scheme.	45
Figure 9	Microphotographs demonstrating array addressability. The vertical springs for the 'down' switches show fringe lines.	49
Figure 10	Substrate Effects on Pull-in and Release voltages	54
Figure 11	MUMPs 14 Devices (not same scale; springs 2um wide)	55
Figure 12	MUMPs 20 devices (each on different scale)	58
Figure 13	Simulated vs. Measured Actuation	64
Figure 14	S-parameters of an early switch design	66
Figure 15	Bank of MUMPs devices S21 and circuit model	67
Figure 16	The MUMPs Layers. Source: [39]	70
Figure 17	Basic MUMPs Transmission Line Shapes	75
Figure 18	A 192x192 Clos Array using 6 32x32 centers	93
Figure 19	Layout Of Control Chip	95
Figure 20	Unmetallized Model of SUMMiT device 'on' state	99
Figure 21	Basic crossover metallization	105
Figure 22	Dissected Device from first metallization	108
Figure 23	Metallization near back of pad	111
Figure 24	Pad Layout Error	112
Figure 25	Built in pad short	113
Figure 26	Metal Test Lines	114
Figure 27	Bubbling Near Bubble Damaged Device	118
Figure 28	Probe Tip close-ups	121
Figure 29	Evaporation Through 2um Grid Line Below	126
Figure 30	Evaporation 1 (2um recipe, 4um recipe on line, nitride)	129
Figure 31	2um Grid Additional Evaporation	131
Figure 32	JSM Device Family (full layout on left, without P3 on right)	142
Figure 33	DubPo Device Family (full layout on left, w/o P3 on right)	143
Figure 34	3*3 Array Segment using DubPo devices	144
Figure 35	FIB'd Device First SUMMiT Run S12	145
Figure 36	Poly2 Plate Device (with and without top panels)	147
Figure 37	Comparison of Metallized(Sputtering) and Unmetallized	148
Figure 38	Magnitude (dB) of S21 for SUMMiT Device (many voltages)	149

Figure 39 Comparison of Evaporation and Non-Metallized S12 150
Figure 40 SUMMiT Model Fit (single device) 152

Chapter 1 Introduction

1.1 Motivation

Crossbars are a category of interconnect that, for a set of input lines and a set of output lines, can route the input lines to output lines in any combination. High performance crossbars are very useful in many applications such as configurable computing, programmable interconnect, system network routers and switching, cable-network routers, satellite downlink routers, and other telecommunications backbone services. Unfortunately, large high bandwidth low latency crossbars are either infeasible, or prohibitively expensive in terms of power, for many uses that would benefit from a crossbar interconnect. When other interconnects are substituted, there are penalties such as blocking states, arbitrary delays, necessity of buffering (and possible buffer overflow), decreased throughput, and increased latency that can be incurred. Even so, many of the other interconnect strategies utilize smaller crossbars as components, and would operate more efficiently with better crossbars. A low power high performance crossbar technology would therefore be beneficial. This work investigates the possibility of developing such a technology using arrays of variable capacitors manufactured using thin film MEMS processes.

1.2 Goals

The primary goal of this work is to make possible large crossbars which are superior to CMOS crossbars in terms of data rate, latency, and power consumption. In order to achieve these targets, a relatively long reconfiguration delay coming from mechanical operation will be accepted. Circuit switch rather than packet switch applications are targeted; reconfiguration of the switch on the order of 1ms would be acceptable. 'Large crossbars' is taken here to mean crossbars beyond the size at which CMOS crossbars scale well, with a size of 192 inputs and 192 outputs chosen for example purposes. A data rate of at least 1 GHz, and latency on the order of 1ns is sought. Improved power consumption will come from the electrostatic control of the crosspoints, so long as the power consumption of the receivers is kept reasonable. A secondary goal is that the switch point developed have sufficiently good characteristics to be of use not only for digital switching, but for RF switching as well.

1.3 Overall Approach

Crosspoints made using thin film MEMS processes will be organized in a grid to form a crossbar; control is facilitated by use of a novel array addressing scheme. Each row and column will require a control line; these lines will have swings on the order of 5V, but the difference in voltage between the row and column lines will be on the order of 20-30V. These relatively high voltage controls will be supplied by chips manufactured in an old 2um feature size pro-

cess, making them economical while capable of handling the required voltages. A signal going through the crossbar starts in a driver with a fast edge rate on a CMOS chip, goes through interconnect to the MEMS chip, and enters the array on a row. One of the devices in the row will be in the 'on' state; the signal will pass through this device into a column, and back off the chip. A CMOS receiver on another chip will recover the signal.

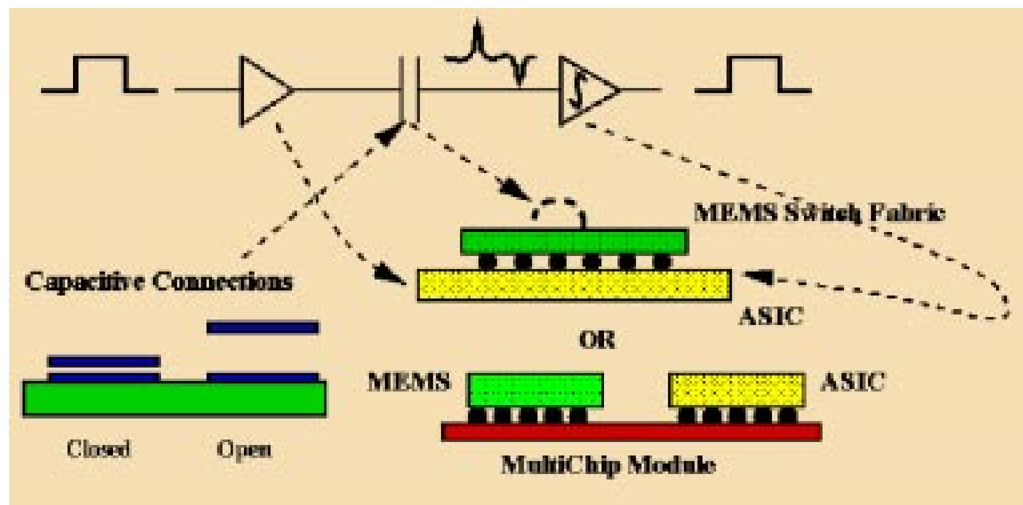

Figure 1 Signal Path

1.4 Crossbar Scaling

A crossbar based on variable capacitors has certain scaling advantages relative to a CMOS crossbar. As a pass-gate CMOS crossbar grows beyond a certain size, the rise and fall time increase with the square of the number of crosspoints in a row. For large CMOS crossbars, this imposes limitations on the throughput and latency achievable. In the case of a crossbar based on vari-

able capacitors, the ‘off’ capacitors do not have such a large effect as they increase in number. For example, consider a ‘on’ to ‘off’ capacitance ratio of 100 to 1. Neglecting parasitic capacitances, the capacitance seen by the driver of a row of 200 ‘off’ devices and one ‘on’ device is the same as that of three ‘on’ devices. The ‘on’ device is 1/3 of the capacitance seen, and gets that share of the signal. As a series capacitance, the voltage across the capacitor does not change quickly, so the high frequency portion of the signal passes through immediately. While parasitics actually become very important if not well controlled, in theory the variable capacitor crossbar should be able to scale much better than a pass-gate CMOS crossbar. Therefore, the use of variable capacitors to create a large crossbar is well worth exploring.

1.5 Thin Film MEMS

Thin film MEMS (MicroElectroMechanical Systems) processes are similar in many ways to the processes for stacking rotating layers in a conventional IC process. Thin layers of materials are deposited and patterned using similar techniques. Instead of selecting layers primarily for their electrical properties, the mechanical properties are paramount. Most layers are either ‘mechanical’ layers or ‘release’ layers. The ‘mechanical’ layers make up the devices present on the result chip; the ‘release’ layers are removed by an etch at the end of processing so that the mechanical layers are movable. One measure of the capabilities of a MEMS process is the number of independently movable

mechanical layers. In order to allow lower layers to shape the vertical shapes of the layers above, planarization is not used as often in MEMS processes as in the metallization stacks of traditional ICs.

Two thin film MEMS processes using polysilicon as the mechanical material, and glass as the release material are used in this work. The MUMPs process has two independently moveable polysilicon layers, as well as a thin gold layer for making reflective surfaces that is constrained to only exist on top of the upper polysilicon layer. The version of the SUMMiT process used here has 3 independently moveable polysilicon layers. Unlike the MUMPs process, there is a planarization of the oxide, immediately below the final mechanical polysilicon layer[39][40].

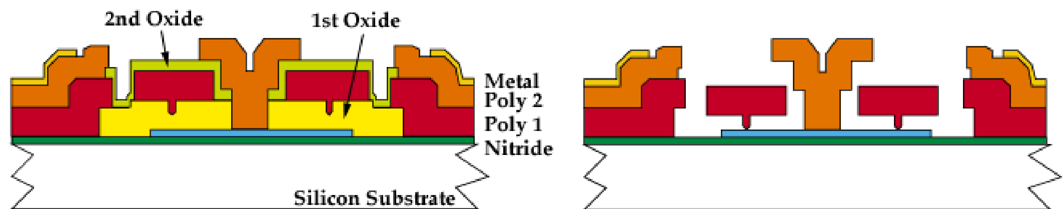

Figure 2 MUMPs Release Example. Source [39]

In order to make low stress polysilicon mechanical layers, significant doping of these layers is generally utilized. Even with heavy doping, it is also necessary to heat the chip to temperatures which preclude the use of active electrical devices. These high temperature anneal steps also preclude the use of metal layers except after the final anneal.

1.6 Challenges & Novel Contributions

In order to achieve a crossbar of acceptable performance, several technical challenges must be overcome. Maintaining signal integrity in the face of processes where routing must be done through polysilicon, while at the same time providing sufficient connections to bring all control and signal lines on and off a chip is an extremely difficult problem. The complete lack of active electrical devices on the MEMS chip to provide any signal regeneration or local configuration memory makes most standard methods of reducing signal count and maintaining signal integrity unavailable. This makes proper driver and receiver design, and controlling non-ideal device characteristics all the more important. Other process issues, such as available gap sizes, lack of planarity, and process variations in layer thicknesses and line widths, constrain the achievable device characteristics. The use of high voltage controls for the MEMS devices introduces the need not only to provide such control levels, but to protect the high speed signaling chips from these levels. Finally, the desired array sizes, the fixed available chip sizes, and the area necessary to create an acceptable interconnect, put severe pressures on acceptable device size, as does the increasing lack of planarity suffered by larger devices.

Several novel contributions were made in the process of facing these challenges. The use of electrostatic attraction countered by spring restoration based bistability to enable array addressing using 2n rather than n*n lines was developed and demonstrated (collective with David Winick's work to use

such techniques for mirror array control) to help meet the interconnect challenges, and it's feasibility was demonstrated in the course of this work. The replacement of landing anchors with a signal pad enabled the use of the devices as variable capacitors, while avoiding destruction at pull-in. Careful study and measurement of the substrate effect provided more reliable (and tunable) control of devices. Improvements in the switch design, and optimization to target devices to the MUMPs and SUMMiT processes were made. As it became evident that improving flatness was important, steps were taken to reduce stress.

At the architectural level, the MEMS switch architecture was optimized to best utilize the characteristics of a variable capacitor crossbar with the available interconnect styles. The space of possible Clos and direct crossbar implementations was explored, and rather than using Clos's criteria of minimizing crosspoints, a hybrid approach of using small CMOS arrays on the first and last stages and large MEMS arrays in the central stage was proposed and explored. In selecting appropriate array sizings, there are trade-offs in the areas of signaling, power, speed, managing parasitics, and overall area constraints.

From the process perspective the existing processes were not sufficient to create interconnects suitable for maintaining signal integrity. A novel post processing metallization approach is introduced for creating multilayer interconnects with reduced penalties for line crossings. Also, adjustments to the

design rules of the SUMMiT process are developed in order to increase the available vertical motion of devices.

While some of the circuit design issues relating to receiver design is discussed in Chapter 2, and some of the background for creating of CMOS crossbars are discussed as appropriate to understanding other issues, the driver and receiver chips will not be discussed in detail here.

1.7 Target Crosspoint

To facilitate comparison, architectural planning, and provide an example of a crosspoint which might be suitable for being part of a large crossbar array, a rough assessment of a possible crosspoint is provided in this section. As will be seen later sections, actually achieving this is quite difficult; this design is highly optimistic for the existing process and area constraints. Assuming flat plates, a flat plate capacitance model, a 2um vertical motion, and surface roughness keeping the signal and moving plates no more than 20nm separated at full actuation, and polysilicon as a conductor, a 100:1 ratio of 'on' to 'off' capacitance would be achievable. If an 'on' capacitance of 0.5pF is considered sufficient for allowing passage of a signal to the drivers, a square signal plate of 34um on each side would suffice. A total device size with interconnect of 100um by 70um is assumed to provide space for actuation, interconnect, springs, and clearances. Unfortunately, devices built according to this plan in MUMPs turn out to be insufficient, as will be seen in Chapter 3.

1.8 Outline of Dissertation

In the next chapter, past work of relevance to this project is examined and discussed. Chapter 3 discusses designs produced using the MUMPs process, and lessons learned there. The basics of bistable device operation and array addressing capability are developed, and the impacts of the substrate biasing effect is explored. Chapter 4 takes a step back and examines some of the more general lessons learned about design and test of devices manufactured in thin film MEMS processes. Chapter 5 returns to the overall architecture. It discusses the other components necessary for creating the system, and the architecture that would have been used if a good enough individual switch point were developed. Chapter 6 investigates metallization, and details methods developed to compensate for the poor suitability of polysilicon for an electrical interconnect. This groundwork is crucial to the use of a process like SUMMiT, which does not include metal layers. Chapter 7 describes the designs in the SUMMiT process, which have significantly improved characteristics relative to the earlier devices. Chapter 8 compares what was achieved with the original goals, suggests alternate approaches that might have been used, and suggests future efforts.

The Appendices provide further detail on some aspects of the work. Appendix 1 presents derivations of equations of the bistability of devices and the effect of substrate biasing. Appendix 2 contains samples and examples of code written to aid in sizing Clos arrays, an example in programming a Clos array,

and an example of using the macro preprocessor of C to simplify the creation of models for FastCap.

Chapter 2 Literature Review

In preparation for designing a large MEMS switch fabric, existing literature from several areas is examined. An exploration of the architecture of large switch fabrics which may be appropriate given the characteristics of mechanical switching elements is made. The need to send and receive high data rate signals through a capacitive coupling network raises circuit issues that are investigated. Earlier work in the physical design of the individual MEMS switches is also examined.

2.1 Architectural Issues

Due to area constraints, it seems impractical to utilize a direct implementation when making very large crossbar arrays, such as a 192x192 array. Since one of the MEMS pistons with best characteristics (described in section 7.7.1) and local routing fit into a 120umx120um square area, this would require a 192*120um=23mm per side array. Also, while driver loading for a full crossbar is not as great a problem as it would be for electrical switches, severe signal integrity problems would occur with the ON:OFF capacitance ratios expected. Therefore, a search of the literature for alternate methods of creating large high performance switch fabrics which use less area was made.

2.1.1 Switching Networks

There are many approaches to building large switching networks. These include Clos, Omega, Delta, Illiac, flip, barrel shifter, data manipulator, shuffle-exchange, generalized shuffle network, and the directly implemented full crossbar. Of these, only the Clos and full crossbar are nonblocking. They provide the highest effective bandwidth, but also require more switches than the other options. An advantage of several of the other options is that they are easier to control [1][2][3].

Two examples of network choices in computers that have been built which require large switching networks are the NYC Ultracomputer and the Cedar from Illinois, both of which use Omega networks[4][5][6]. One interesting finding is that the use of non-blocking crossbars would decrease stall time in the Cedar 12-36%, though an increase in bandwidth to memory and other methods of dealing with traffic convergence yield greater speedup[6].

While the MEMS pistons can be used in making most of these networks, the full crossbar has been eliminated due to area constraints. The various blocking networks would be more inefficient than usual due to the comparatively large reconfiguration time of the mechanical elements; the penalty for a blocked connection would be very high. Since the array being designed is intended for long duration connections, this also increases the penalty a blocked connection would receive. The usual reason for not using a non-blocking network is that it is not feasible to do so due to area, power, or data rate

constraints. This is not true in this case since the MEMS pistons allow large subarrays of suitable performance to make a 3-stage Clos feasible. Therefore, a Clos network is the best choice for the large crossbar.

2.1.2 Clos Networks

The most important work in the area of non-blocking crossbars was done by Charles Clos and is described in his paper “A Study of Non-Blocking Switching Networks” published in 1953[7]. Clos describes a way to create non-blocking arrays utilizing multiple stages each consisting of smaller non-blocking arrays. For an effective crossbar of size NxN, the primary intent is to decrease the number of crosspoints required from the usual N^2 that would be necessary for a direct crossbar implementation. For three stages Clos shows that one method is to use $n=N^{1/2}$ inputs to the subarrays in the input stage, resulting in $6N^{3/2} - 3N$ crosspoints, which saves crosspoints for all $N>=36$ [7].

To determine the number of middle stage subarrays that are necessary, consider that each line from a given first stage switch must pass through a different middle switch. Also, each output array must receive at least as many signals as it outputs from a separate middle switch. In the extreme case, a situation can occur where, for a given input/output array pair, only one signal travels from the one to the other, and all the others must pass through different middle stage subarrays. If there are n outputs in the subarrays in the output stage, and m inputs on each of the subarrays of the input stage, this

means that n+m-1 middle stage subarrays are needed to ensure that blocking will never occur [7].

The general rule for creating a Clos network of 3 stages for a square array (# inputs = # outputs) can be condensed into simple integer equations where n is the size of the square arrays in the center stage (also the numbers of subarrays in the 1st stage), A is the number of inputs in the first stage arrays (also the number of outputs in the last stage arrays), and B is the number of outputs from the first arrays (also the number of arrays in the center stage, and the number of inputs into the last stage arrays).

$A = (N - 1) / n + 1$

$B = 2A - 1$

This results in an array with the actual number of inputs and outputs equal to n*A, which will be greater than or equal to N. A program has been written to generate a table of valid three stage Clos networks for a target array size, which will be discussed further Chapter 5 and is included in Appendix 2. Clos derives in his paper that $N \sim 2n^2$ will achieve the best reduction in crosspoint total[7]. There are situations where another choice for n may be better due to other considerations, and this program may be of value in determining whether this is true in any given situation.

After his explanation of 3 stage networks, Clos continues with a discussion of 5 stage networks, 7 stage networks, and a discussion of how to achieve even longer networks. He provides a graph showing the growth of the number of

crosspoints required as N increases for various numbers of stages. It appears that a 3 stage network is most useful for this work, since it achieves most of the size reduction at a lower latency and complexity cost than the networks with more stages require[7].

Benes includes Clos's work in his textbook along with a few enhancements others have suggested for special cases. The most interesting improvements involve the case of N mod n != 0. For example, I. G. Wilson discovered that for single input (or single output) subarrays, it is necessary to connect each to only n (or m) of the middle arrays, and this allows some of the middle arrays to be smaller, saving crosspoints. J. Riordan has found that in some cases instead of using one switch of size N mod n, making the last several switches (n-1) will often save crosspoints[8]. These techniques from Benes's textbook may save some area, but not enough to warrant significant increases in control complexity.

One technique sometimes used to save crosspoints over a true Clos network is to make a "wide-sense" Clos network. A true Clos network is strictly nonblocking; it allows connections to be maintained during rearrangement. In a wide-sense nonblocking network, connection requests may require breaking old connections in order to set up the network. When the first stage of the array is made up of n*m arrays (note these variable are different here than my earlier uses of n), the requirement to be wide-sense nonblocking is that

$m \geq \left\lfloor \left(2 - \frac{1}{F_{2r-1}}\right) n \right\rfloor$ where F is the Fibonacci number[10]. This property seems most useful for small central stages. For the large center stages used in this work, this does not allow me to reduces the number of stages from m=2n-1 in most cases.

2.1.3 Programming Clos Networks

Many approaches to programming Clos networks have been proposed; some are described in [9][11][12][13]. Heath and Disch have developed and implemented a control algorithm for a three-stage Clos network of 256x256 described in [13]. A delay model for their algorithm is also developed. An implementation of this algorithm utilizing custom logic chips is discussed in [14].

Their approach is to split control into four basic building blocks: a RAM module and a module type for each of the three stages of the array. The RAM module keeps track of information about the outputs of the Clos array, and connects to each first stage module. The first stage module communicates with the input device, and determines whether the requested output port is available. It executes opcodes and queries second stage modules as to whether they can provide connection (or disconnection) to the final stage. The second stage is responsible for accepting address and control data from the first stage and responding as to whether it has access to the output port. The third stage sim-

ply executes connect or disconnect commands from the first stage. The communication between stages is done using bi-directional communication along the links used by the modules to connect to the output, with the exception of the last stage which never sends a message back[13].

The opcodes are presented for each of these stages. They include the information necessary to set up connections of a specified duration. After this duration, the connection will be removed unless this time is extended by a subsequent command. Since the control is based on each stage controlling the next along the existing lines, this control scheme does require the duration based control. A model is derived for the time taken to set up and remove connections; the largest delay is the maximum of 2823 cycles it may take to set up a connection. The estimated cycle time for their hardware using MOS was 500ns, for a delay in this case of 1.41ms [13].

Despite the circuitry being based on 1988 fabrication technology, this provides some interesting information. The effective algorithm the modules implement is (except for the auto-turnoff to save control lines) the same as the obvious one used to put together a computer program to simulate programming a generic three-stage Clos network before the literature search (see Appendix 2 for listing). The main innovation in [13] therefore appears to be the ability to handle the control without additional lines. However, in the case of the large MEMS based crossbar, those additional lines are required anyway due to the special signaling nature of the data lines, and the necessary high

voltage controls for the MEMS devices. Also, the algorithm relies on timed ending of connections, while for this dissertation it has been decided that connections remain until explicitly ended. If the method in [13] is representative of those available for programing, the speed of device switching will not be as significant a portion of overall programing time as was initially anticipated, even with the faster cycle times now possible.

The main drawback of the algorithm used is its reliance on a timer to end connections, and repeated commands to add time to keep connections open. A circuit-switch network is usually used for connections of relatively long duration, and the long reconfigure time also fits long duration connections. Therefore it seems better to utilize a method in which connections remain until they are explicitly removed.

Another algorithm for programing a 3 stage Clos network is described in [15]. Here memories are used to keep track of possible configurations of the middle networks, and a couple lookups are sufficient determine a usable middle switch to connect the ends together. The algorithm is relatively fast (6 cycles per connection, pipelinable to N+5 cycles for N connections). It does however essentially perform a large vector multiplication in parallel in order to achieve this, so the implementation costs may be significant. In deciding whether to use this algorithm it will be necessary to evaluate what the cost in area and power will be for a given network configuration, as well as the suit-

ability of the available process for making the memories. Also, the paper does not present the algorithm for removing connections[15].

2.2 Circuit Issues

One of the challenges to overcome in utilizing switchable capacitors as a part of a crossbar is the design of drivers and receivers which can operate at high data rates on the sorts of signals which pass through such a crossbar. Because the signal passes through at least one series capacitor, there is no DC path. Low frequency components of the signal are filtered out- the switch is effectively a high pass filter when on. In the case of low resistance lines, all signals through the switch are converted to pulses. Special circuits for receiving pulses have been suggested [20][21], but the parasitics of the array being developed mean that the signal is more like the output from a highpass filter than simply pulses. After a quick look at a commercially available crossbar, this section examines a few papers which utilize self-biasing techniques or related techniques which may be useful in creating a self-biasing receiver. The way these techniques work is summarized, and there is some discussion of their usefulness in providing a bias for a high speed receiver for capacitively-coupled signals.

2.2.1 Other crossbars

Some examples of circuit approaches to the design of electrical large crossbars include [22][23][24]. The available commercial 32x32 crossbar which seems most comparable in performance to the middle stage being developed is [25] by Applied Micro Circuits Corporation. It utilizes PECL and differential signaling in order to achieve a 1.5 Gbit/s data rate. It has a latency of 3.0ns for data which is longer than the approximately 1.5ns of the MEMS approach, but it can reconfigure much faster (6 ns). The main drawback for applications where the data latency is acceptable is the much higher power consumption. The nominal power consumption is 1.95 Amps, and a 5V supply is used, leading to a rough power estimation of 9.75 Watts[25].

A more recent commercial crossbar, designed to carry electrical signals in large optical routers, is discussed in the preliminary data sheet [26]. This is a 140x140 switching matrix, which uses 2.5V and 1.8V supplies and consumes 17W at 2.5Gbps. It allows multi-rate connections, and is suitable to serve as part of a larger Clos array. The maximum data rate is 3.2 Gbps; latency is unspecified[26]. Based on the power supply levels, it probably utilizes a more advanced process than the ones used for the circuitry in this work.

2.2.2 Resistors

In designing a receiver, a common need is for termination. For a large number of signals, an external termination resistor for each would be inconve-

nient. A simple technique for creating this component discussed is the one described by Gabara in [16]. This paper describes how to make a controlled resistance on-chip despite temperature and process variations. The intended use is to minimize the need for off-chip termination resistors. The technique makes use of one off-chip 100 Ohm resistor, a -2V supply, an Op-Amp, and on-chip polysilicon resistors of 100 Ohms and 50 Ohms. The polysilicon resistors are used to create a reference voltage of -1.3V when used as a voltage divider. A voltage divider is also created between the external 100 Ohm resister and a PMOS transistor; the Op-Amp is used to control the gate voltage of the transistor[16].

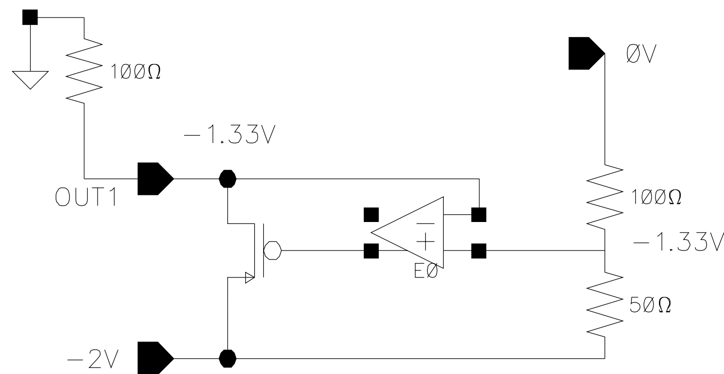

Figure 3 Resistor Circuit

So far it would seem a significant amount of resistors and circuitry were used to make a single resistor. Actually, the gate voltage to the PMOS transistor can be used on many other PMOS transistors, creating a large number of resistors of the same value. So all the 50 Ohm lines for the chip could in theory be terminated by transistors biased in this manner. One potential difficulty

for higher accuracy applications than the one described in the paper is that if the bias voltage must travel a long distance from the reference area to other areas of the chip, process variations along with noise picked up along the way could lead to different resistance values of the PMOS transistors. To deal with these difficulties, it may be necessary to use multiple reference generators close to where the references will be used if resistances will be tightly controlled[16].

The paper claims the error of the match to 50 Ohms is l5 to 20%. It then presents a method for bringing the match to within 5%. An n-plus resistor which varies from 50-150 Ohms depending on process variations is added to the circuit in parallel to the transistor. If the resistor is closer to 50 Ohms, the transistor is turned off, while if the resistor is of significantly higher resistance the transistor acts as the appropriate parallel resistance to make a 50 Ohm resistance. Of course, now an extra on-chip resistor is necessary for each usage. However, since the transistor is only being used as trim now, it need not be as large as before. The paper does not detail the sizes of the transistors for either case[16].

In order to use this technique to provide the bias for a capacitively coupled receiver, two resistors could be created to make a voltage divider at the given voltage. Of course, while the resistances may remain sufficiently close to their proper values to hold the bias voltage steady, this approach would not account for changes due to process variations and temperature of the proper bias

needed for the receiver. Also, since two on-chip resistors are used to generate the voltage bias necessary in making the resistor, it is unclear whether these resistors and the voltage divider they form would be superior to making a voltage divider directly with onboard resistors. Finally, the voltages and resistances used would have to be appropriately adjusted to generate the appropriate resistances, but this is a comparatively minor problem since each of the components can be created in a standard CMOS process.

In summary, while this simple method is probably sufficient for Gabara's target application of creating 50 Ohm termination resistors, it is probably insufficient by itself for the needs of a capacitively coupled receiver. It may be useful for creating terminators if utilizing a scheme that requires them.

2.2.3 Automated Bias Control (ABC) Circuit

One interesting technique for dealing with a small amplitude signal with some DC wander is described in [17]. In this approach, the receiver circuit current is controlled to affect the biasing of its output. This technique utilizes two copies of the unaugmented receiving circuit, an Op-Amp, and three CMOS inverters. By making two copies of the sense-amp portion of the receiver circuit and using the current source NMOS transistor between each circuit and ground, the current source of the sensing circuit can be controlled. The replica circuit's output is compared with an inverter tied back on itself with an Op-Amp to force the replica to be properly biased; the output of the Op-Amp goes

to the current source on both circuits, so both are properly biased even though only one is checked[1]. It appears that the slowness of the Op-Amp is counted on to assure that the current variation is not fast enough to cause difficulties in operation of the sense-amp. The first inverter after the sensing circuit should be a duplicate of the one used in the biasing circuit, except that it can be intentionally modified to create a delta-V from Vth for the bias of the output. Figure 4 shows this circuit[17].

Figure 4 Auto Bias Circuit

---

1. One somewhat confusing portion of the description states "By feeding the same input signal (In0) on the replica circuit as that for the BiCMOS sensing circuit before the input transitions, Vsense0 can be monitored"[17]. One possible interpretation would be that In0 should be delayed be one transition; another is that the biasing occurs before transitions. The later interpretation makes more sense due to the lack of any mention of delay circuitry, and the fact that at the speeds under consideration an Op-Amp wouldn't be able to respond quickly enough for it to matter much anyway.

This method looks promising for the capacitive coupling receiver application because of its ability to use a sense-amp based receiver without changing the receiver itself, or requiring large area passive components, but only assuming the main problem of creating a bias to compare to the incoming single signal is taken care of some other way. One drawback to using it is that the paper assumes a BiCMOS process. A CMOS sense-amp, or other receiver circuitry, may not perform as well, or will at least require further design. A greater problem is that with the delta-V between Vth and the bias point about 200mV, the 1.25ns sensing time, which though significantly better than the non-ABC circuit may not be good enough for the target application (depending on the how the CMOS receiver in current technology size would behave). And of course, this still does not take care of the part of self-biasing sought to begin with: a mechanism to properly bias the input, or in a pseudo-differential case to create the reference level. Alternately, the capacitively coupled signal must be sent full-differential. Still, this method looks like a promising way to adapt for process and temperature variations and cut down on additional amplification stages, decreasing latency. Some other solution to the DC wander at the input will need to be considered.

2.2.4 Optical Receiver

This paper [18] describes a design for a CMOS receiver for optical signals, and thus must deal with many of the same challenges faced in designing a

capacitive coupling receiver: The average value can move around, the amplitude of the input may vary significantly depending on the path the signal has taken, and a low latency response is desired. In order to meet these challenges, the approach used is to keep track of the high and low values, and then use their average as the bias. Since in the optical application, the variance of the DC mostly occurs in different packets rather than within one packet, a reset is provided to be used between packets so that if the amplitude of the next packet is lower it is still received properly[18].

This method requires the use of two resistors and two capacitors, as well as some CMOS circuitry, for each ABC section; it recommends multiple sections for each receiver. Each section contains a top-hold and bottom-hold circuit to track the expected high and low values of the current packet, and sets the bias to the value midway between these using the resistors as a voltage divider.

The top-hold and bottom hold circuits use a voltage follower going into a diode which charges a capacitor. Since the current can only flow one way through the diode, the capacitor only charges through the diode. A transistor is added in parallel to the diode to facilitate a reset by opening an additional discharge path. A voltage follower on the output prevents the capacitor from needing to discharge while supplying current to the voltage divider. The paper neglects to explain the selection of Vbias for current source circuit, but inspection would seem to indicate it should be set one or two hundred mV above Vt of the transistors it is connected to; this should be checked by simulation. The

authors may have made a mistake in labeling their figures; the bottom hold circuit they provide is not sufficient for providing current to the voltage divider. What is more likely intended is a bottom hold circuit similar to the top hold circuit, but with the diode reversed, and this description assumes this. When not in reset the capacitors can only discharge through parasitic leakage currents, thus the circuit is not nearly as susceptible to errors in the average due to the same symbol repeating itself many times, like a BJT-based version of the receiver might be (since the base current would discharge the capacitor)[18].

The paper suggests the use of a pre-amplifier followed by more stages, each with its own ABC circuits, rather than doing all the amplification in one stage. This way the first stage need only be close enough to optimal bias to get some amplification, not provide all the amplification necessary to completely restore the signal. When using this multi-stage approach, it is good to use a limiting amplifier to keep the output within the operative range of the top-hold and bottom-hold circuits so that the reference voltage generated is correct. A circuit to do this is also provided in the paper, as well as a pre-amp design for low noise. The power consumption of the preamp was measured to be 50mW in a 0.8u technology with a 4.5V supply. The pre-amp test chip was 0.8mm by 1 mm, while the GCA was 2mm x 2.5mm[18].

One thing of interest is that the pre-amp circuit does not seem to utilize an ABC circuit, though it requires one on its output. It does however use some

feedback, so the stability would need to be examined. Also, simulation should be done to check what range of inputs it can use, and whether it can still be used as a front end with the expected amount of variation in the target application of this work.

If the pre-amp is not used, it may be possible to use this approach to ABC directly. In order to use this ABC approach for a capacitive coupling receiver without the pre-amp as a front end, a few things need to be taken into consideration. One is that unlike the optical signal, the capacitively coupled signal may have some DC drift within configurations. A trade-off will thus develop in choosing the capacitor size in the top and bottom hold circuits. With the optical circuit, letting them discharge if not recharged frequently is a bad thing (thus leading to large C_{hold} values to prevent repeated symbols from not working). In the capacitive coupling case some decay is useful for tracking a wandering DC value. Thus a trade-off must be made in choosing the capacitor to be smaller, and some other method for dealing with too many repeated symbols may have to be used. In addition other concerns need to be addressed. Area is a concern since capacitors and resistors are used. Also, power consumption must be minimized to preserve the low power savings of using capacitive coupling.

2.2.5 Quantized Feedback

An interesting approach in [19] is to use the quantized feedback method, long used to deal with low frequency interference in transatlantic communications, to deal with the DC wander in capacitively coupled situation as well. In this method, the output data is fed through a low-pass circuit and added back into the input data. If the low-pass circuit is based on the same RC as high-pass circuit which occurs due to the capacitive coupling, then in theory the input to the receiver is restored to the original input, except for the delay required to go through evaluation and filtering. A diagram of the approach is shown in Figure 5.

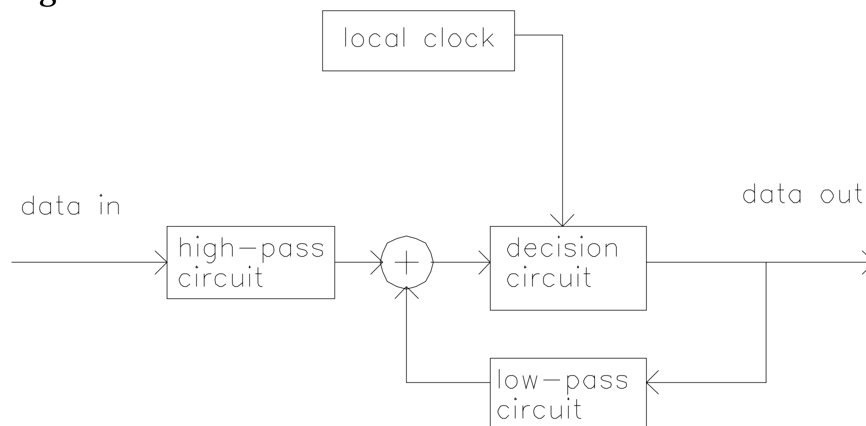

Figure 5 Restoration of Signal

Note the need for a local clock; an alternate for this is to use an edge detector based on the output of the highpass circuit. Schematics of the circuits are given in the paper. These are for differential signals[19].

This approach offers several advantages. Since a 0.5um CMOS version was tested at 800Mb/s with 750mV input voltage swing, as well as long same symbol strings (2kHz signal)[19], the 0.35um CMOS process we have access to can probably be used to make a version that can operate at the desired data rates. There are still many challenges to meet this application however. The highpass circuit that is the MEMS piston switching array may vary in resistance significantly depending on the path taken through it, and there may also be significant variation in the effect of the parasitic capacitance. This means that matching the lowpass filter properly to the highpass filter is not as easy as in the applications the paper is targeted at where the receiver provides both the highpass and lowpass filters. Possible solutions to this include using a decision section and lowpass that will work with any of the highpass configurations, or to make the lowpass section adaptable to different configurations. Also, since the delay through the crossbar will vary dependent on the route taken, the local clock is probably not a good option. Thus the edge detector will have to successfully detect the difference between an edge from the signal and noise from other paths, despite the possibility of significant differences in input signal amplitude dependent on crossbar configuration. Finally, it would be preferable to develop single-sided versions of these circuits so that differential signaling will not be necessary.

2.2.6 Conclusions Regarding Signaling

The papers by Kurodo et al.[17], and Nakamurea[18] provide some interesting methods for biasing receivers. Each method would require some adjustments to utilize with the target application. The former would require the use of differential signaling, or some other method (perhaps even the one presented in [18]) to create the reference for comparison. The latter method would require adaptation for careful power and area management to preserve some of the advantages of capacitive coupling, as well as making sure that the input signal was appropriate for working with the preamplifier, or modifying the reference voltage creation circuit so the system can work without a preamp. The controlled resistance method described by Gabara[16] may be useful for creating many resistors of the same size, but alone is not sufficient for making a capacitively coupled receiver. The single best method for signaling through a MEMS array appears to be the one presented by Gabara and Fischer in [19], if the problems in matching the RC characteristics of the crossbar can be overcome.

2.2.7 High Voltage Generation

In order to control electrostatic MEMS devices, it is sometimes necessary to generate higher voltages than are commonly used in supply voltages for logic or signals. Generating higher voltages than those supplied can be done by various methods, one method being through the use of a charge pump [27].

Charge pumps are commonly used to increase voltages in EEPROMS and flash memories for programming, as described in [28][29]. Because the MEMS devices to be designed should not consume significant current, the current limitations of charge pumps should not be a problem.

2.3 Device Design

While there is not much related to the design of the devices developed for this work, it is instructive to examine foundational work in MEMS, and look at a few other MEMS devices targeted at provide forms of electrical switching. Some details of design, such as predictable spring design, have been investigated in the past and so are examined here.

2.3.1 Silicon as a Mechanical Material

This paper [30] examines the properties of silicon and its suitability as a material for mechanical uses, and provides a good survey of many of the devices in which it has been employed. While much of it is not directly relevant here since polysilicon is actually the material used for the devices developed in this dissertation, and silicon is only the substrate material, much of the reasoning in favor of the use of silicon is valid for polysilicon as well. In summary, the semiconductor industry has already developed the fabrication techniques which are necessary for the sort of MEMS process utilized, and due to the nature of batch fabrication can economically make large quantities

of a design once it has been developed. Also, some early MEMS switches are discussed, mostly based on bulk fabrication; in this section several of the advantages (high on-state to off-state impedance ratios, low power, high speed relative to relays, and integration) and disadvantages (high switching voltages) of MEMS switches are discussed. It is predicted that the ideal applications would be in systems requiring large arrays of switches, which is what is being developed here [30].

2.3.2 Micromachined Relays

There are many papers which describe contacting micromachined relays based on magnetic or electrostatic actuation including [31][32][33]; one of the most recent is [34] which is examined now for comparison purposes.

This [34] relay utilizes magnetic attraction to actuate, pulling a plate suspended on springs into contact with two contacts. The fabrication is based on polyimides and electroplating, rather than polysilicon layers used here. This allows the use of a nickel-iron alloy core as well as metal for other parts. Gold was used for the contacts. The size of the upper plate alone is 3.5mm by 1.95mm. The specifications provided deal with the amounts of current and contact resistances; one design was capable of switching more than 1.2A. The contact resistances varied from cycle to cycle during the first few hundred operations, with maximums approaching 7 Ohms, but eventually settling down to about 2.5mOhms. The switching speed was in the range of 0.5ms-

5ms. 850,000 cycles have been tested without failure. Since the testing appears geared to DC characteristics, no high frequency information is given. Switching power of 33mW was observed[34].

Due to the size and switching power, such a relay could not be used to build large low power crossbars, but it has the advantages for other applications of low resistance and the ability to carry large currents, while the devices developed in this dissertation are DC isolated. The 850,000 cycles without failure using metal contacts was impressive. The switching speed is on the order of the switching speed for a large array of the devices developed here. Regarding the merits of the presented relay relative to other relays, the most significant aspects are the relatively simple construction, avoiding the need for multiple coil layers, and not requiring the attachment of additional components such as separate springs or magnetic cores for operation after the fabrication; also no coil-suppression diodes were used[34].

2.3.3 Spring Design

One issue when designing the bistable MEMS pistons is the design of the springs. In [35] and [36] one aspect of spring design is discussed. Springs made by connecting a long beam to the substrate with a cut in the sacrificial oxide vary in performance from the ideal equations due to the step-up which occurs beside the anchor point. Each of these papers presents a different method of dealing with this problem.

The first method is to derive equations which take into account the effect of elasticity of the step-up in modeling the behavior of beams. Equations are derived for a variety of cases, and finite element modeling was used to verify the results; the largest difference between model predictions and finite element analysis was 5%. There is also an analysis of buckling force [35].

One concern is the lack of fabrication to check these results, since the finite element analysis and derivations seem to make simplifying assumptions regarding the shape and thickness of the step-up. While the equations to account these variations would likely be complex, some test results would at least provide some information as to the degree of error these assumptions introduced.

The second method is to try to remove the effect of the step-up. By adding an extra bump in the polysilicon, the spring acts much more like the ideal case of a fixed-end beam. The method of verifying this was to find the fundamental frequency for a spring with and without the additional bump using finite element modeling, and comparing these to the predicted fundamental frequency. The traditional step-up had a larger error (a few percent), while the modified anchor had a 0.1% error in frequency. A 300um beam was fabricated. It is interesting that the test results varied significantly more from the FEM and analytical predictions than the two cases varied from each other. Because of this, whether or not the suggested refinement effectively makes the spring behavior much more predictable is unclear, but the experimental cases do

have the predicted variation in frequency from each other. The inclusion of experimental results so that this could be noticed was good[36]. It is interesting to note that similar problems with matching theoretical behavior happened in this work, as well be apparent from chapter 3.

An adaptation of the method described in [36] for anchoring spring will be used since it appears they will make things more predictable by simpler equations, and this method takes very little additional area. It appears however that, compared to other factors this may not matter, as shown by the experimental results of [36].

2.3.4 A Previous Bistable Device

A previously developed bistable device is reported in [37], and is discussed here as an example of a typical previous use of bistability in a MEMS device. This device operates utilizing a buckling cantilever and a tension band. Fabrication involves both bulk and surface micromachining, since a recess is made under the device to allow movement range. Thermal expansion of polysilicon is used to actuate the cantilever, which is built from two layers of polysilicon with silicon dioxide between them, and the tension band made of high residual tensile stress silicon nitride is used to hold it into position in either state. The heating is done by running current through the upper or lower polysilicon layers. After preheating both polysilicon layers with 3mA at 7.5V for 50ms,

the device state can be changed with 7mA at 25V for 12usec. Total device length excluding pads is 200um. The vertical movement at the end of the cantilever is +/- 6um[37].

This device is not useful for the intended application since the curved cantilever would probably not achieve high enough capacitance with another surface, and adding metal to make a contacting switch would interfere with device operation. It is however very interesting. The use of heating rather than the usual electrostatics or magnetics in actuation was of particular note, and it is a good feature that the design retains its state due to its mechanical properties.

2.3.5 An RF and Microwave device

A MEMS switch targeted at RF and Microwave applications is discussed in [38]. This switch is fabricated using a thin film MEMS process which uses polyimide as the sacrificial layer, silicon dioxide as the structural material, and metals (gold and platinum) for the contacts and electrical signals. The device is simple- a plate suspended by two silicon dioxide springs, which is electrostatically actuated. The design takes advantage of the fact that the structural material is an insulator to prevent shorts during actuation. Metal is present on top of the silicon dioxide in the actuation region below in the switch area. The switch is large, though smaller than the relays- the actuation capacitor alone is 200um by 200um[38].

The performance is quite good; the electrical isolation is >50dB at 4GHz and >25dB at 40GHz, and insertion loss is <0.5dB. The current handling capacity is 200mA, and power consumption for actuation of 1.4uW for the lowest voltage actuation. The silicon dioxide beam is reported to have been tested for 65 billion cycles without fatigue, but no reliability data on the contacting portion of the switch is reported; this is a potential area for problems due to the metal-metal contact[38].

This switch appears to have a lot of promise for use with high frequency signals; it is not suitable for the making of large arrays in its current state due to the lack of bistability (each switch would need to be individually controlled) and the large amount of area each switch takes up.

Chapter 3 Device Operation & MUMPs

3.1 MEMS Switch

The basic device used at each crosspoint is a programmable capacitor made using a polysilicon thin film MEMS process. This switch consists of a platform with springs attached. The platform has sections at two different heights; the areas closer to the matching electrodes are signal areas and the ones farther away are used for actuation. A cross-section of a single MEMS switch made using MCNC's MUMPs process [39] is shown in Figure 6. In this example, the programmable capacitor is formed by the center plates; when they are in close proximity the switch is 'on'. The outer plates form the actuators that pull the capacitor down. A voltage is placed across the actuator plate:moving plate gap causes an attractive electrostatic force and thus move the capacitor plate

down to the 'on' position. The flexure spring restores the plate up to the 'off' position when the voltage is removed.

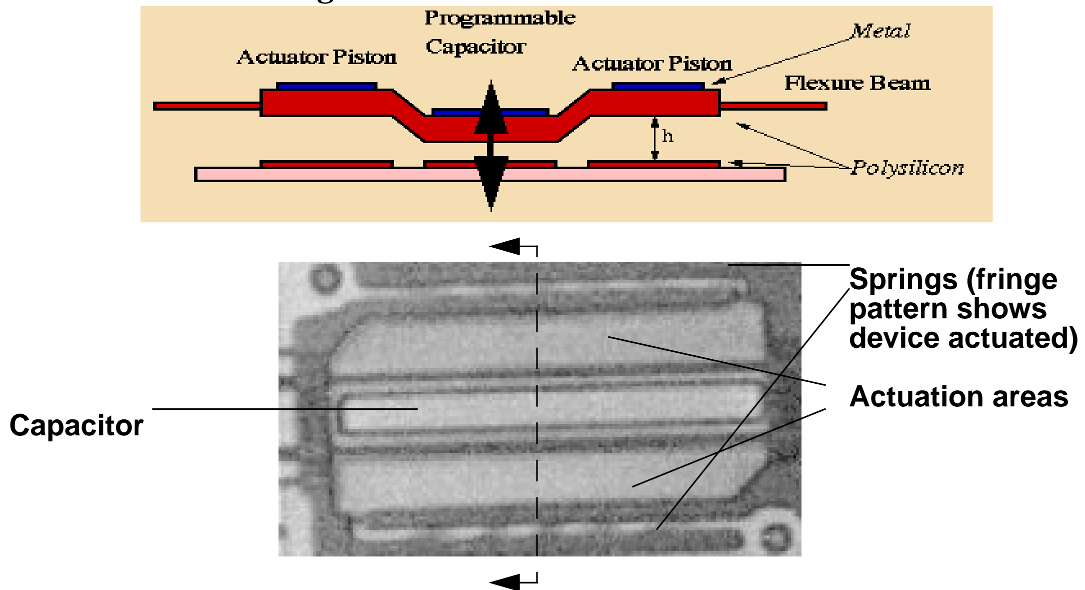

Figure 6 Individual MEMS switch

The entire device, except for the optional metal at the top, is polysilicon. When the device is down (in the 'on' position), the capacitance dielectric consists only of the native oxide of the polysilicon (about 20nm) and any air gap created by a lack of flatness at the plate interface. When this device is up (in the 'off' position), the capacitance dielectric includes an additional approximately 2um of air gap, so that ideally there would be a 100:1 capacitance ratio between the 'on' and 'off' states. Actual capacitance ratio results for this device were not nearly as good, which is evident from the s-parameter measurements shown in section 3.6.

The plate is shaped so that when the device is actuated, the actuation control surfaces never touch the top plate. Despite the high resistance created by the native oxide, the control voltages are high enough that a short would cause the destruction of the device. Common failure modes when such shorts occur include melting of the springs and spot welding of the moving plate to the actuation plates. Because of this, the signal plate voltage is kept the same as the moving plate, so that when they contact there is no DC current flow. Some of the AC signal may pass through the high resistance oxide, but most of the signal should pass through the capacitive connection.

3.2 Bistability & Array Addressing

Bistable device operation facilitates a novel control scheme, which lowers the number of connections necessary to control a large MEMS array to manageable levels. The reason for the bistability is derived, it's utility for controlling an array is explained, and experimental confirmation of this property and control scheme is reported. In this dissertation, bistability is taken to mean that there are two states in which a device will remain in indefinitely for some range of inputs. These states need not be a specific device position, but can be sets of positions with certain characteristics.

3.2.1 Derivation of Bistability

The relationship between plate displacement and the applied voltage necessary to create a force balancing the restoring spring force at a given displacement for a basic switch is given in Figure 7.

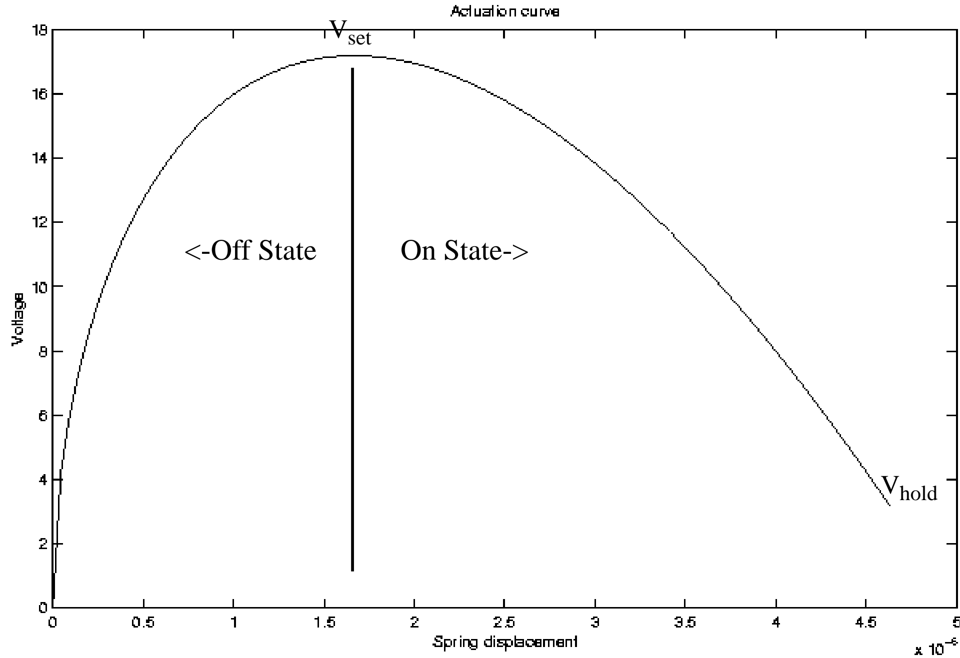

Figure 7 Voltage-Displacement Characteristic (flat plate model)

This can be derived from the equations for these forces assuming a flat-plate model for the capacitances between plates:

$$F_c = \frac{C \cdot V^2}{2 \cdot d} \qquad F_s = \frac{y \cdot 6 \cdot I \cdot E}{L^3} \qquad C = \frac{\varepsilon \cdot A}{d} \qquad I = \frac{b \cdot h^3}{12}$$

where d is the distance between plates, V is the applied voltage, y is the spring displacement, L is the length of the spring, and I is the moment of inertia for the spring. Note that the F_s equation is necessarily the equation of a beam with a guided end, rather than the free end equation commonly used for

cantilever beam devices. In a design with two springs, setting the forces equal yields:

$$y = \frac{\varepsilon \cdot A \cdot V^2 \cdot L^3}{2 \cdot b \cdot E \cdot h^3 \cdot d^2}$$

The position of the plate is stable in the region of the curve with a positive slope (to the left of the maxima), while the segment with negative slope (to the right of the maxima) is characterized by a "pull-in" effect, in which the device moves to full actuation and remains there until the voltage is reduced below the level V_{hold}. In other words, once enough voltage (V_{set}) is applied, the device turns on and remains on until it is released. In order to return to the left side of the maxima it is necessary to apply a voltage lower than the hold voltage necessary for the maximum displacement ($V_{release} < V_{hold}$).

In other words, there are two forces acting on the plate. The electrostatic force due to the applied voltage difference is trying to move the plate down. This is offset by the force of the spring trying to bring the plate back to it's initial position. For the first region of the device operation, the spring force is sufficient to keep the device from being pulled all the way to the mechanical stops. As the device moves and the distance between the plate and control surfaces decreases, the force pulling the device down increases. At a certain point, simply moving the plate lower increases the force exerted on it faster than the restoring force from the springs grow with their deflection. The plate is therefore pulled all the way to the mechanical stops. Now the plate is close enough

to the control surfaces that the voltage must be lowered to a value significantly lower than was required to cause pull-in. This difference in voltages for actuating and releasing is what make the device bistable.

While the “on” state consists of a single physical position which the device snaps into when a sufficient voltage is applied, the “off” state consists of the range of positions the device can occupy before pull-in. These should all be fairly low capacitances relative to the “on” state.

It is important to note that the effective applied voltage is simply the difference between the voltage on the plate, and that on the control surfaces.

3.2.2 Array Addressability

This bistability property can be used to provide array addressability. Consider a simple two by two array of devices as illustrated in Figure 8. Remember that the applied voltage potential is the difference between the X and Y control voltage levels, so the state of the device is affected by the relation between the two control lines which pass that device. For this illustration, assume the voltage on the X control lines is the higher one. If one wishes to set a device in row X1 to a given state, control X1 can be set low enough so that all the switches in that row are released (return to their “off” state). The voltage on line X1 is then set just to a few volts less than V_{set}. Appropriate voltages are then applied to individual columns (Y1 or Y2) so as to bring the desired

individual switches down into the “on” state while other switches in the array remain in the “off” state.

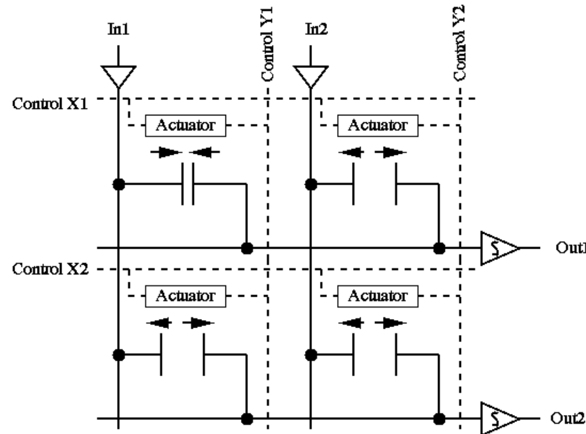

Figure 8 Array Addressing Scheme.

In other words, by moving the voltages on a pair of control lines apart, the device at their intersection can be actuated without affecting other devices. By moving the voltages on the pair of control lines closer together, the device at their intersection can be released without affecting other devices.

3.2.3 Stop Gap Height Implications on Bistability

Consider again the basic curve presented in Figure 7. Without mechanical stops, the curve would continue on to the x-axis, where the plates touch. At that point, the plates would short, destroying the device. If there were an arbitrarily thin insulator preventing device destruction at this point, the device

would still be difficult to release, as a very tiny voltage difference would be sufficient to provide a huge amount of force. Thus, the mechanical stop, which is also the signal plate, is a very important part of the mechanical system.

The gap between the height of the stop, and the height of the portion of the plate over the actuation area, determines the gap when the device is actuated (using the graph, it determines how far back from the d=0 to look up the release voltage). Thus, for all other design parameters held equal, this gap determines the release voltage, and the size of the hold voltage region, which has significant implications for controllability. The hold voltage region must be large enough so that all devices in an array can be controlled (this is discussed further in section 3.2.5). On the other hand, if the gap is too small, the force on the device becomes large enough to bend or unbalance the device so that a short occurs and the device is destroyed.

Unfortunately, for a given supplied process, there are usually not many choices available in selecting the gap height. For the MUMPs process, barring sacrificing much of the potential actuation distance to use both POLY1 and POLY2 for the plate, the DIMPLE cut is the only reasonable method of creating this gap. Therefore, the gap for all MUMPs devices is 0.75um.

3.2.4 Device Throw Implications For Bistability

The primary concern regarding device throw that the greater the device throw, the better the “off” state of a device can be. It is beneficial however to understand how this design parameter relates to bistability as well. Referring again to Figure 7, notice that the steepest portion of the curve is at the beginning, where the actuation capacitance is still increasing slowly with device motion. The greater the initial distance between the plate and the actuation areas, the more displacement the spring experiences before reaching a given distance from the actuation plates. Thus, the restoring force from the spring is higher, and the plate must get closer to the actuation area before the restoring force and actuating force balance. In other words, the necessary set voltage is increased by increasing the initial gap height. This aids in increasing the size of the hold voltage region. Indeed, total device throw is one of the few design parameters which (at least within the bounds of the available processes) it is beneficial to maximize from both the mechanical controllability and electrical switch quality perspectives.

3.2.5 Requirements For Addressability

In order for the array Addressability scheme to work, a certain amount of consistency in device operation is necessary. This is analogous to examining a circuit for proper operation within process corners, except that the process corners in most MEMS processes and their effect on a given structure are less

well understood. Three regions have been presented: the actuation voltage (and any higher voltage), the hold region, and the release voltage range (all voltages from the release point down to zero). Due to process variation, the actuation and release threshold may vary from chip to chip; they even vary from device to device within the same array. In addition, any signal going through a device will add an additional attractive force. Finally, there are other forces involved, such as the substrate effect discussed in section 3.3. All of these cut into the usable hold region. In order for a voltage to be successfully used as the hold voltage in a control scheme, it must be an acceptable hold voltage for all the devices in the array; if the control voltages are not tuned on a chip by chip basis it must be an acceptable hold voltage for all devices of that design produced by the process. It is therefore necessary to design devices with a large theoretical hold voltage range, just as designing circuits with large noise margins helps to achieve working designs.

3.2.6 Experimental Confirmation of Bistability

The bistability and addressability properties have been verified experimentally. Figure 9 shows a optical microphotograph of the top of 4 switches. The central 'capacitor' plates and the outer 'actuator' plates can be clearly seen on each switch. The restoring springs run parallel to the actuator plates' long edge. This photograph is taken using 546nm light and a Mireau objective lens, resulting in fringe patterns appearing on the sloped spring surfaces for those

devices that are ‘down’. The corresponding programming voltages are given in Table 1. Since a 5V swing on each line is an appropriate choice, programming can be performed by a 5V capable CMOS process, once a DC bias source of around 20V is provided. In the photograph on the left in Figure 9 two devices diagonal to each other have been actuated using x-y addressing while leaving the surrounding devices unactuated. Then these devices were released and two neighboring devices on the opposing diagonal were actuated. Optical devices based on the same principles are described in [47], and analog control of device state is described in [48].

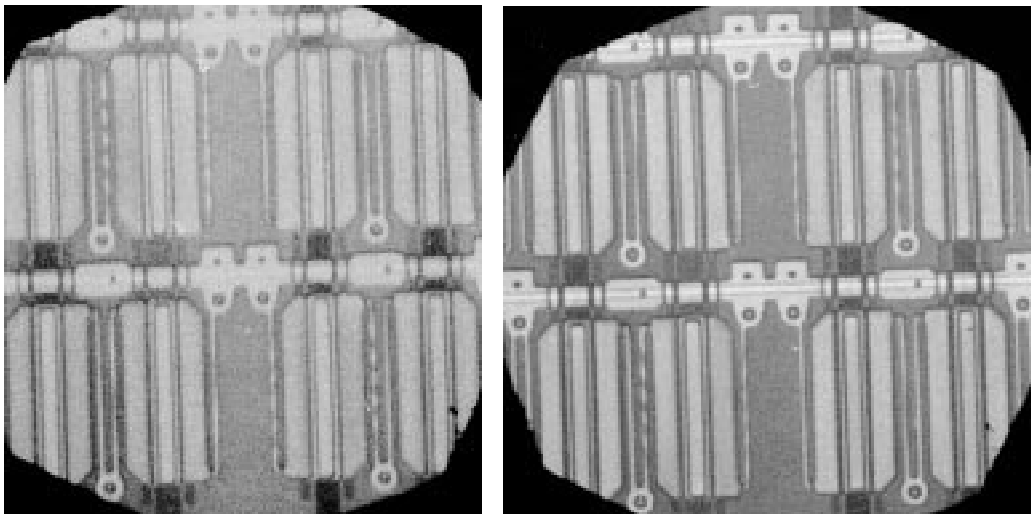

Figure 9 Microphotographs demonstrating array addressability. The vertical springs for the ‘down’ switches show fringe lines.

Table 1 Control for Figure 9 Devices

Pull-in/Release (Volts)	Column 4	Column 5
Row G	29.7/20.6	29.1/22.9
Row H	30.0/22.6	29.7/22.3

3.2.7 Experiment Details

This test was performed on part of an 8x8 array of the device F_JS2_Base and was fabricated in the MUMPs20 run. F_JS2_Base is the base form of one of the flat spring devices. Six 50 Ohm probes were used. Two were connected to the bottom signal lines, two to top-plate signal/actuation lines, and two to actuation lines. The substrate and the other 18 lines in the array were left unconnected since 6 probes is the maximum that can be fit around a chip in the probe station utilized.

A 20x diffractive index lens was used for observing the devices during the test; the green filter was used so that the bending of the springs could be seen by counting the number of dark-light brightness transitions on the springs.

Four power supplies were used. Two were connected to the horizontal actuation probes. The other two were connected to the top plate probes. The remaining two probes (the signal lines) used were connected to the same supply as the top line that would be actuated at the given stage in the experiment. This was to prevent large quantities of charge from flowing from the top

to bottom plates when contact was made. The supplies for the top plates were configured to produce negative voltages, while the actuation lines would have positive voltages applied.

Since this was the first set of measurements for this particular device, it was necessary to start by measuring the pull-in and release voltages of the devices and getting an idea of the process variation. The probes were landed so as to be working with rows G and H of the array (lettering top to bottom looking from above) and columns 4 and 5 (looking right and left). Since the substrate was not controlled, some variation was expected in the voltages; if variation had been too large a 7th probe to the substrate would have been needed to prevent excessive charge accumulation. The pull-in and release voltages were measured and are given in Table 1. Based on these results, control voltages were chosen and are shown in Table 2.

Table 2 Control Voltages

(Volts)	Actuation	Moving Plate
Set	28	-4
Hold	24	-4 when doing releases, 0 when doing sets
Release	0	0

The following operations were done using these control voltages:

1. Setting device G4 without moving other elements in array.
   The signal plate for G4 was connected to the supply for the top plate. The other signal plate was hooked to the other top plate supply at this time. The lines not being used to set were put at the hold states. Then the voltage of

the top plate was slowly ramped the top plate to the Set value and with actuation plate under test; the set occurred when the actuation voltage reached 25.2.

2. Setting device H5 without moving other elements in array.
   The voltage on the actuation line was returned to the Hold voltage, and returned the top plate voltage to zero for that device. The device held down. Then the top plate was set to -4V then ramped the voltage for the actuation line for H5 up to the set voltage. The set occurred when the actuation voltage reached 25.4V.
3. Returning array to signalling mode. The actuation plate voltage on H5 was set to hold voltage, and the top plate voltages to 0V. The signal lines were disconnected from the supplies.
4. Both devices released in preparation for setting G5 and H4.
   They were released the same time, as the array would normally be released to save time in a full reprogram.
5. Set G5 and H4.
   Following the same procedure these two elements were set. Setting occurred at 24.8V for G5 and 25.2V for G4. As before, the other devices retained their state during these operations.
6. Release G5
   G5 was now released without releasing H4 by ramping down the actuation voltage while holding the top plate at zero. Release occurred at 21.7V.
7. Other tests and pictures.
   Each supply was individually moved to set and release while keeping the other lines at hold, and the devices were observed to keep their states. Pictures were taken of several of the above experiments (see Figure 9, as well as the states during this test (since the devices move a little when not set, and that movement can be discerned from the diffraction patterns).

3.3 Substrate Effects on Actuation

The voltage on the substrate has a significant effect on device operation and the substrate cannot be ignored without leading to unreliable device operation. This lack of reliability is because a charge is induced on the substrate when voltages are applied to control the device. This problem can be overcome by intentionally biasing the substrate. The bias affects the substrate on a chip-wide level.

If differential signaling is used, with a pair of devices at each crosspoint instead of a single device, the selection of the bias level is very important. In order to limit the number of control lines necessary, and achieve better common mode noise rejection, for one device of each pair the upper plate has the higher voltage applied, while for the other the higher voltage is applied to the lower plate.

Figure 10 shows the effect of various substrate biases on the F_JS2W_Base design fabricated in the MUMPs 20 run. Similar data was collected for other devices. The pair of curves which slope upward from left to right are the pull-in and release voltages when the fixed actuation plates are held at 0V and the actuation voltage is applied to the top plate; in the other pair of lines the top plate is held at zero while the bottom plate has the actuation voltages applied. In order to control differential devices, the substrate bias is chosen so that distance between the lower of the upper pair of lines (the pull-in voltages) and the upper of the lower lines (the release voltages) is maximized. This region is

the range of voltages at which both devices will retain their state if one pair of control signals has the higher voltage on the moving actuation plates, while the other has the higher voltage on the fixed actuation plates. In order for X-Y addressing to be possible, this region must be larger than the sum of regions where one device actuates without the other device actuating. This constraint is necessary but not sufficient. The region must actually be larger than this, to account for the other factors discussed in section 3.2.5.

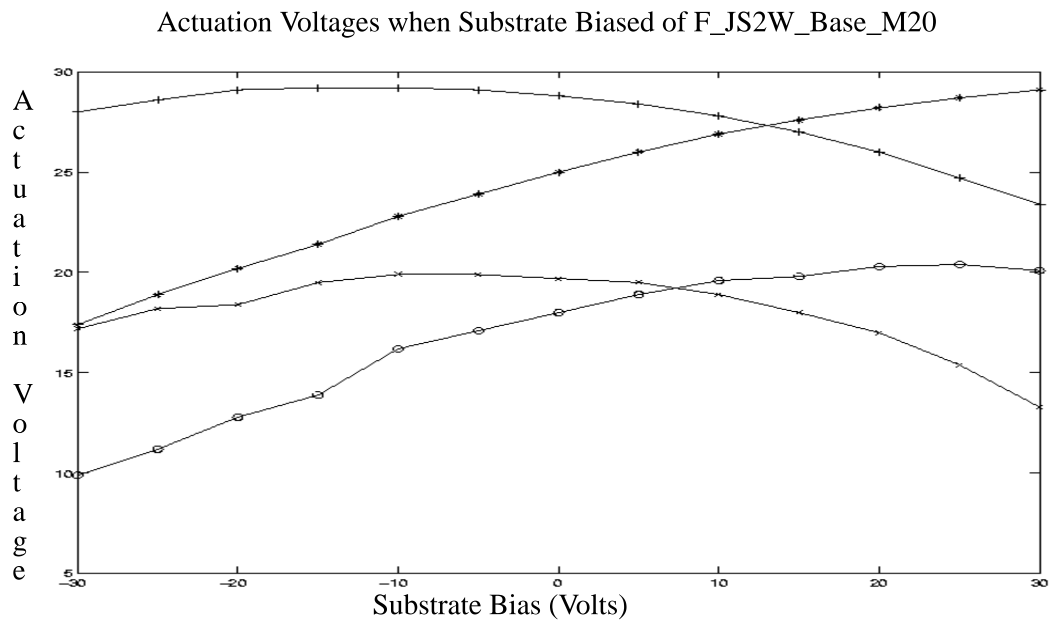

Figure 10 Substrate Effects on Pull-in and Release voltages

3.4 MUMPs Designs

The MUMPs process is a multilayer thin film MEMS process. It includes two releasable polysilicon layers, the top of which can have a liftoff deposited Au layer. MCNC developed the MUMPs process, and later spun of their MUMPs fabrication facility as part of CRONOS[39].

3.4.1 MUMPs 14 and 18

Designs utilizing the MUMPs process went through several stages. The initial designs were fabricated in the MUMPs 14 run. Figure 11 shows the basic devices made at that time.

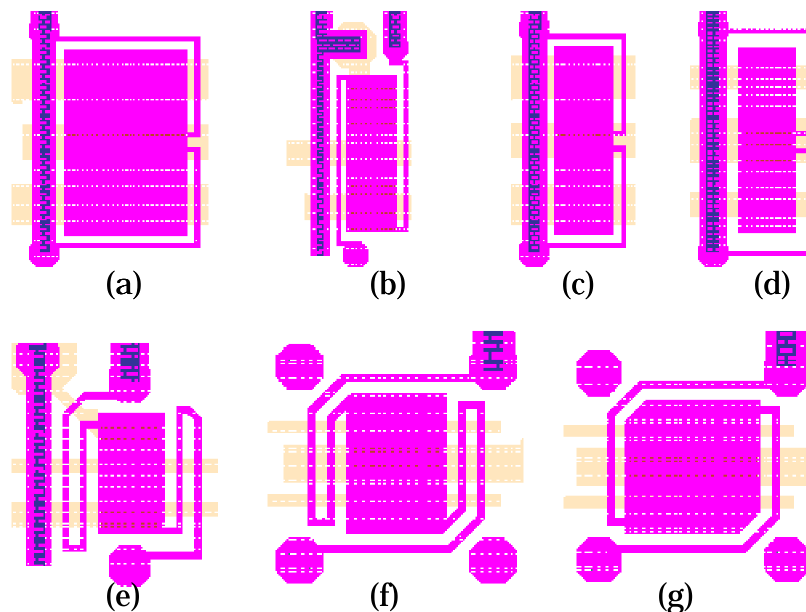

Figure 11 MUMPs 14 Devices (not same scale; springs 2um wide)

The basic approach used in all devices except (b), (e) is to put a single signal pad in the middle and bring in both signal and actuation through the springs.

In devices (b), (e) the actuation control is in the middle, and there are two signal plates, the signal passing through the device. Devices (b) and (e) never worked even with enough voltage applied that the signal line running adjacent to (b) flexed visibly without the device itself moving. Devices (a) (c) (d) are all basically the same except for the dimensions of the plate and clearances. (d) utilizes significantly larger clearances as a safety margin for process errors. This had no impact on device operation. (a) had problems with operation in which the far side of the device from the spring would actuate alone. While this occurred on occasion for most devices in MUMPs 14 it happened consistently with this device.

(e) (f) (g) utilized springs attached at opposite corners of the device. This tended to provide more consistent and proper operation especially of (g).

MUMPs 18 contained basically the same devices as MUMPs 14, with differences in the test structures. Instead of one line intended for DC probing from each direction, a ground line beside previous line was introduced to provide return paths. Also, several wire bondable arrays were included to allow for verification of the bistability property of array control.

One mistake made in all of the devices fabricated in MUMPs 14 and MUMPs 18 was to run POLY0 lines under the springs. While the springs were made more flexible as expected, significant variation in spring properties was introduced by thinning of the springs during the step-up and step-down at the edges of the POLY0. Also, the springs would sometimes touch actuation lines,

leading to a short and device destruction. This complicated wiring for later runs, since springs could no longer wrap all the way around devices.

In general the devices with shorter springs worked more consistently, even though higher voltages were required for control. No devices in these run had sufficiently good switching characteristics to measure using the test structures included in the run. With a signal generator and oscilloscope and change in a sine wave going through the device was barely noticeable visually for each of (c) (d) (e) using the DC probes for some frequencies in the hundreds of MHz.

In both runs, a capacitor was used in the test cell in order to couple the AC signal to the DC control line for the signal pad for all the devices with a single signal pad. There were no substrate pads in MUMPs 14 or MUMPs 18.

3.4.2 MUMPs 20

New devices (see Figure 12) were included in the MUMPs 20 run which incorporated several improvements.

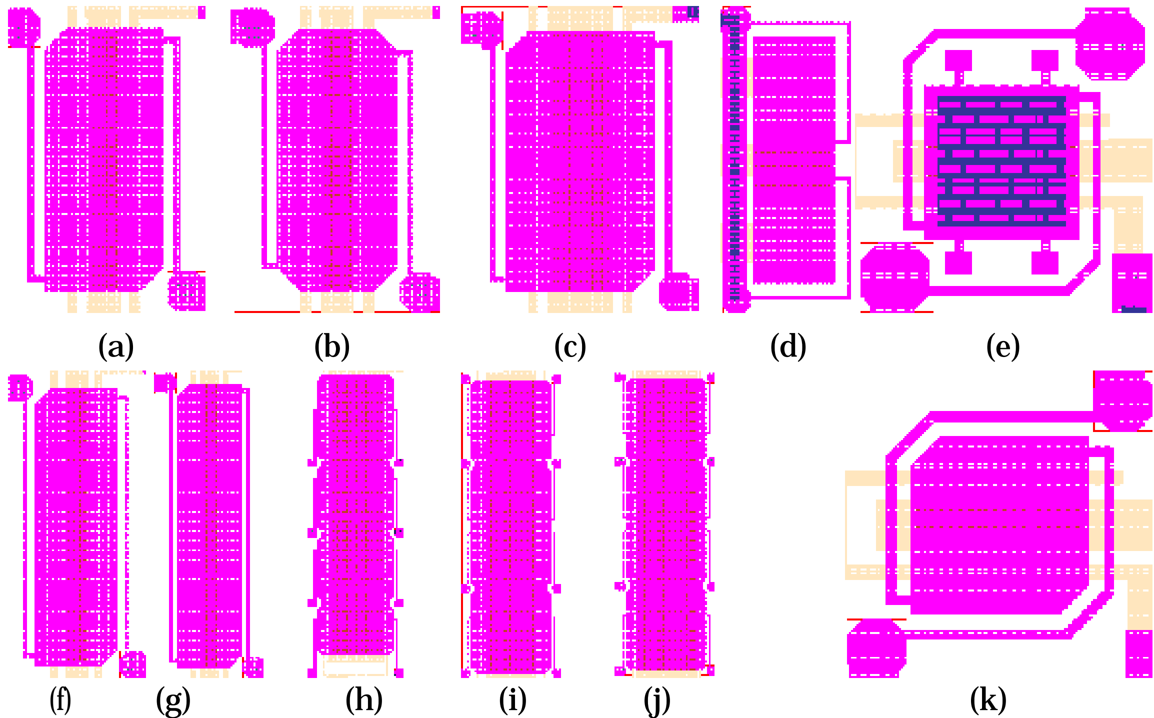

Figure 12 MUMPs 20 devices (each on different scale)

Some of the changes were minor, such as the increased anchor size in (k). Other such changes include the inclusion of metallized and nonmetallized versions of several devices ((e) shows a metallized version of one). (e) also utilizes landers as one approach to the problem of partial actuation.

The most important change, used in all devices except (e) and (k), was to redesign devices so that springs did not pass over POLY0. This led to better operation for device (d), as well as the creation of the new device family of

which (a) (b) (f) and (g) are a part with long narrow signal and actuation regions. These devices differ in details such as the h/w ratios, and spring attachment points; all except the extremely long devices (f) and (g) operated well mechanically. Device (a) is the F_JS2_Base_M20 design which the tests described in section 3.3 refer, (c) is the F_JS2W_Base_M20 version which was used in an array in MUMPs 20, and which was used to demonstrate array addressability as discussed in Section 3.2. These are the most robust devices developed with MUMPs.

Devices (h) (i) and (j) are much larger than all the others, utilizing multiple springs and very large signal plates. The goal of these devices was to provide a large amount of capacitance and serve as RF components. None of these devices actuated. The primary discovery of importance from these devices was the extreme degree of bending which occurs in large devices in the MUMPs process. This was a large factor in the decision to look more closely at the amount of plate curvature occurring in the smaller devices.

In addition to changes in devices, the test cells were significantly improved in MUMPs 20 and later runs. Ground-Signal-Ground pads were used for signal lines, and pads were moved closer to the device. Substrate pads were added to each test cell. The grounds were all connected, and provided the DC actuation level. Significant travel through non-metallized sections was avoided where possible, though the existence of only one metal level made it impossible to avoid polysilicon-only sections entirely. The arrays included in

MUMPs 20 demonstrated this problem most effectively, since with only one metal layer it is impossible for two lines to cross and remain metallized.

3.4.3 Later MUMPs runs

MUMPs 22 included no significant device changes, but did include test cells with pairs of devices hooked up differentially, and even some arrays using a differential signaling scheme.

MUMPs 23 included multiple devices in parallel in an attempt to make s-parameter measurement easier, as well as provide an alternative method of creating a large RF-suitable capacitance since larger devices are impractical. Too much non-metal routing was necessary to get good results from these structures.

MUMPs 25 included a better parallelization scheme, as well as the introduction of a variation on the Figure 12 (a) device which was more resilient in earlier runs. An extra signal pad was added, widening the device in the direction which increased the distance between the springs. The idea was to avoid the necessity of the signal passing through the springs, which add a large amount of resistance to the signal path in all the MUMPs devices which had operated properly. These devices did not exhibit the stability of their predecessors.

MUMPs 29 included a similar variation in device design as MUMPs25, except that the signal plates instead of running parallel each were in series, the path of both together in parallel with the actuation lines.

3.5 Modeling

In order to predict device operation, a variety of modeling approaches were used. The simplest approach was simple pencil and calculator estimates using the basic equations of the on and off capacitances. For better models, a variety of tools were utilized. The traditional formulas for estimating fringing capacitance did not provide any better estimates than the simple flat plate approximations, so for all models short of finite element approaches a flat plate approximation was used. Indeed, while fringing models usually add additional capacitance, devices usually act as if they have less capacitance than predicted by analytical models using the flat plate assumption.

3.5.1 Matlab

Matlab was used to create actuation curves assuming a flat plate model and estimate parasitic and signal capacitances. The basic equations given in Section 3.2 were used for actuation computations.

After discovering the substrate effects, the equations used in Matlab were modified, with a force from the substrate-plate attraction added to the system. This force is considered another flat-plate capacitance, this time between the

substrate and the actuation area of the moving plate. It is assumed for simplicity that because the signal pad is larger than the signal area of the plate, the effects of the signal area on the plate-substrate capacitance are not important in the curve region below actuation. The unreliable operation of devices during high (|Bias| >=25V) substrate bias was predictable when this was added to the model. Despite the flat plate capacitance simplification, this model was very useful in projecting regions of safe device operation, and areas to avoid.

If one adds the capacitive force of the substrate to the equation (repeated variables gain subscript s if dealing with the substrate, a for actuation plate; all voltages are relative to the plate voltage), then one can solve for the actuation voltage to generate a given displacement.

$$\frac{d_a}{d_s} \cdot \pm \sqrt{\frac{d_s^2 \cdot y \cdot 2 \cdot E \cdot b \cdot h^3 - A_s \cdot V_s^2 \cdot L^3 \cdot \varepsilon}{A_a \cdot L^3 \cdot \varepsilon}} = V_a$$

A fuller derivation of this equation, and the assumptions involved, is given in Appendix 1. The main problem with the Matlab simulations relative to the field solvers is from one of the assumptions used in the derivation. It is assumed that the force generated by the attraction between the substrate and moving plate, and the force generated by the attraction between the actuation controls and the moving plates, can each be computed separately based only on the voltages of the plate and the involved area. The effect on the charge dis-

tribution by the voltage on the third surface is neglected. The error introduced by this is illustrated by the upper left actuation curve in Figure 10. In this case, the moving plate voltage is held constant, and for various substrate bias values the voltage that must be applied to the actuation control area to actuate the device is measured. If the assumption used in deriving an expression for substrate controls were valid, this curve would be symmetric about the zero bias.

3.5.2 FastCap

FastCap was used to try to obtain better estimates of capacitances. In order to model a device in FastCap it is necessary to manually create the models of the surfaces of the devices using quadrilaterals. In order to simplify this process, the C preprocessor was used to create basic device shapes. The general use of FastCap is described further in section 4.3. MEMCAD incorporates the FastCap tool, so FastCap was not used separately once MEMCAD was acquired.

3.5.3 MEMCAD

MEMCAD is a finite element solver package produced by Coventor. The MUMPs device F_JS2_Base, including the substrate bias, was simulated using this solver. Pull-in analysis was done at each of the substrate biases for which we had lab data. Release analysis was also attempted, but while most of

the pull-in simulations converged to +/- 0.25V, the release simulations failed to get closer than a 4V range, and even so failed to produce release numbers for about half the cases. All releases were predicted to be in the range of 8 to 12 volts. Figure 13 is a comparison of the predicted actuation values from the Matlab equations, the MEMCAD simulations, and the data measured in the lab.

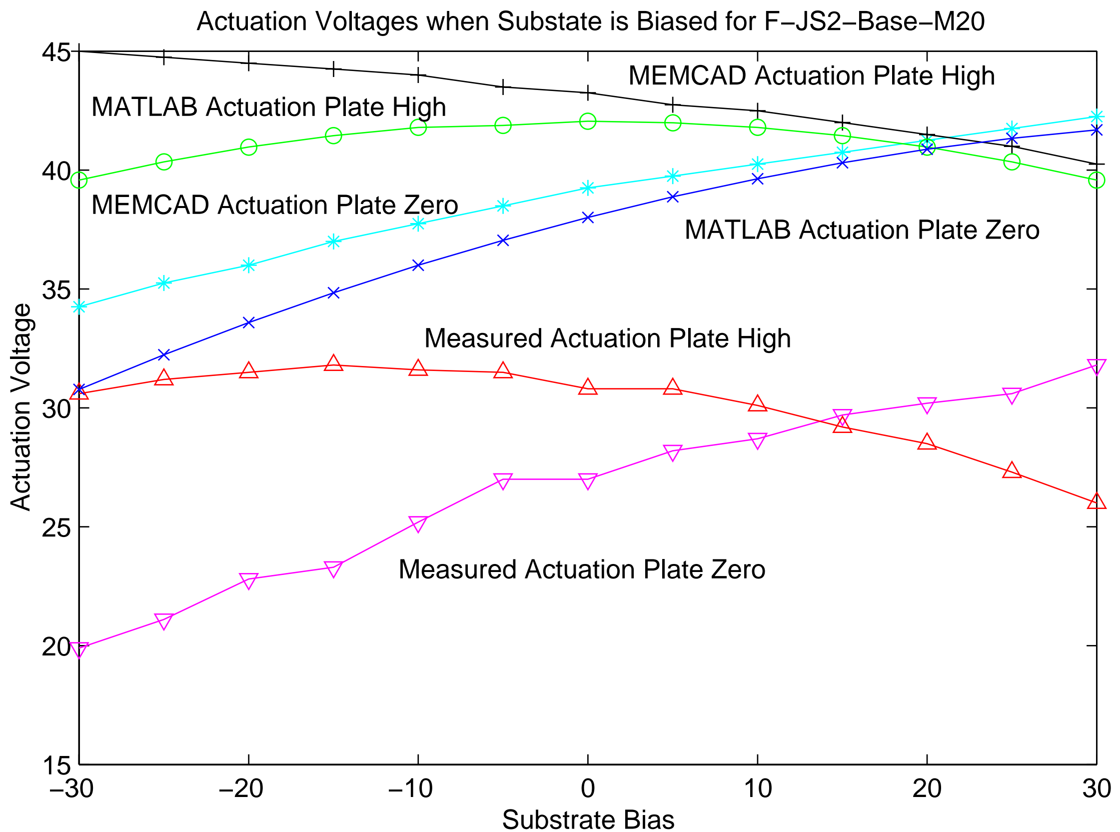

Figure 13 Simulated vs. Measured Actuation

It is interesting to note that while the general shape of the curves is the same, the simulated values are off by more than 10V for most of the range, and the MEMCAD results are farther from the measured results than the

simpler Matlab simulation. The MEMCAD curves do however indicate the same non-symmetry that the experimental results do, which the derived equations do not capture.

One possible reason that MEMCAD is providing results that differ more from experimental results than simple calculations is that the structure is too complex for the software to handle. While the tightest mesh the software was able to complete and simulate was used, it may not have been sufficient to allow accurate modeling of device operation. Much simpler situations (such as bending a single beam) provide significantly more accurate results when tighter meshes are applied than was feasible here, than when meshes of similar density are applied. Yet in order to get a good finite element simulation, usually more complex structures require more detailed meshes than simple structures.

Another possible reason for the offset of both curves is simply that there are differences in details such as material properties and layer thicknesses of the actual devices from the values used for simulation.

3.6 S-parameter results for example device

While the control characteristics of the MUMPs devices validated the bistability and control scheme, their characteristics as switches were too poor for use in a larger system. This is clearly visible from s-parameter measurements of the through characteristics of one of the MUMPs switches, as shown in Fig-

ure 14. The given switch is part of the JS2 family, and is sized to fit in 50um by 130um (excluding routing). The gain is on the low side, and the shape of the S21 curves greatly complicates the task of receiver design. Factors contributing to this include poor matching, large attenuation from polysilicon routing, lack of planarity, poor step coverage of the one metal layer, and the possibility of the polysilicon layers making up the device acting partially as dielectrics and partially as conductors. Rather than attempting to fit a model to this individual device, a later test setups involved banks of a device repeated several times.

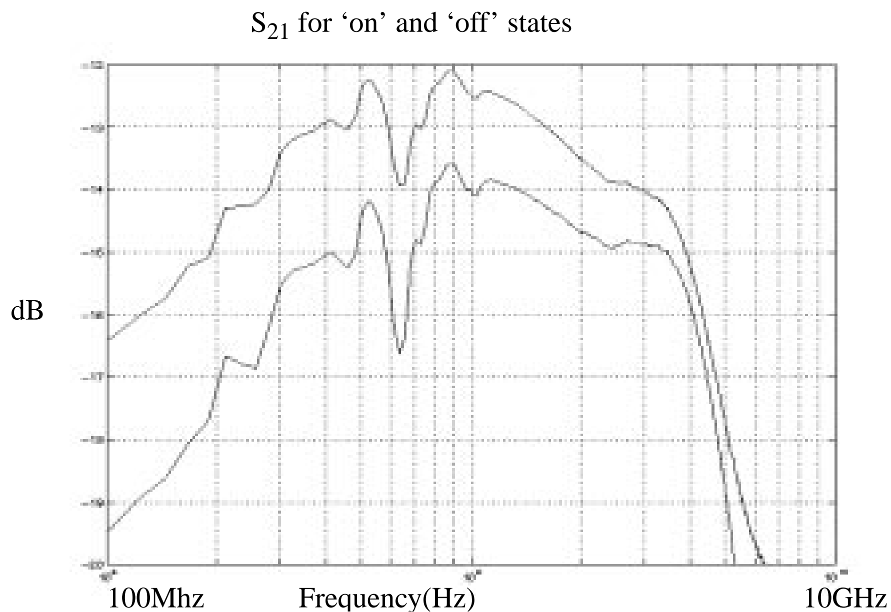

Figure 14 S-parameters of an early switch design

3.7 S-parameter results for a bank of devices

In MUMPs 29, there were a number of 16 device banks tested to explore the viability of such an approach for creating a variable capacitor for RF uses. While the banks take up too much area for using in building crossbars, and the capacitance ratio is too low, this approach does make modeling easier. Figure 15 shows the S21 of the 'on' state of such a bank with its fit.

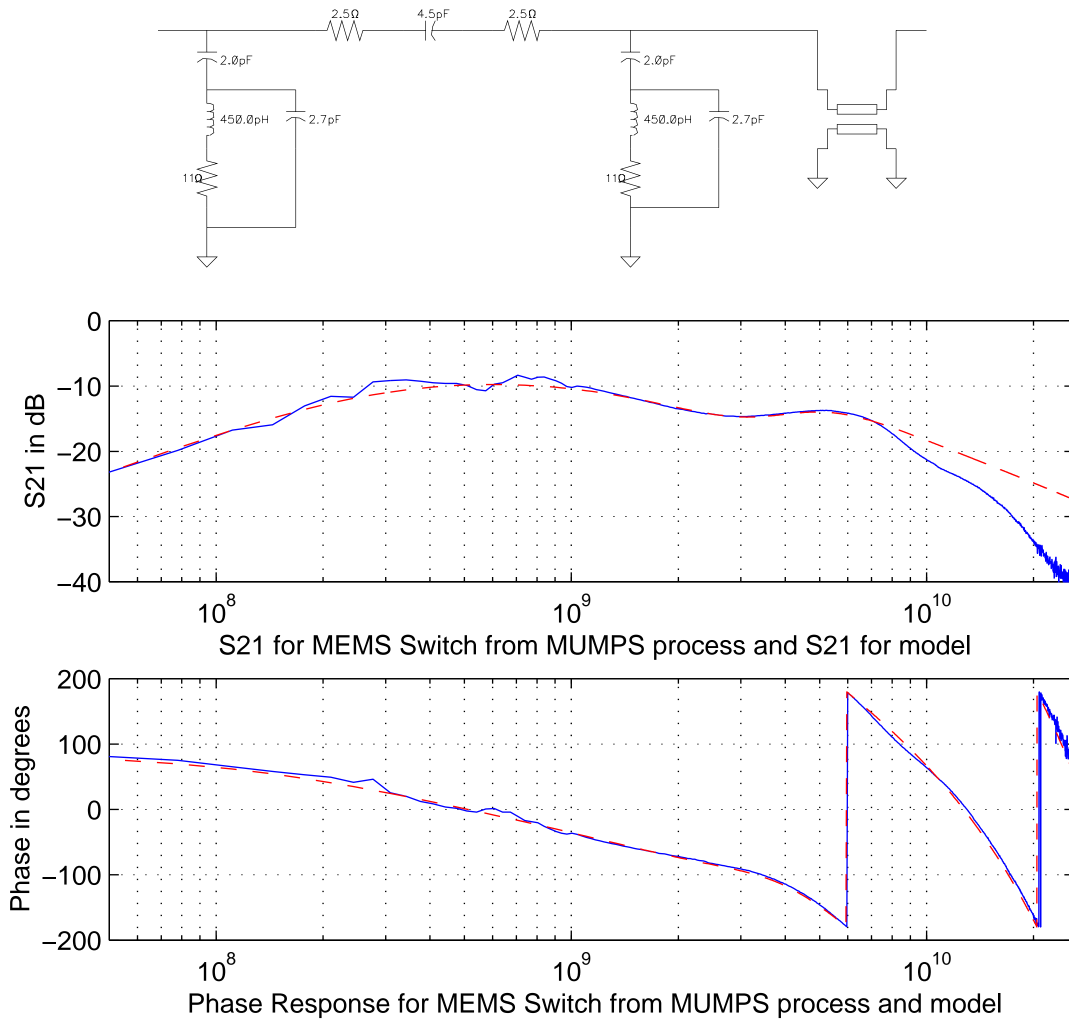

Figure 15 Bank of MUMPs devices S21 and circuit model

Based on a similar fit for the data from devices in the ‘off’ position, values for the through capacitance can be estimated. The capacitance ratio achieved is only 1.36, based on fits using values of 4.5pF ‘on’, and 3.3pF ‘off’. It is important to understand that although the fits are fairly good, and the circuit topology is reasonable for the physical situation, they do not of necessity represent the actual situation. Any model with enough variables to tweak can be made to match a given set of measured data, so there is always significant danger in reading to much into empirical fits. That said however, the model is sufficient to show clearly that these devices do not achieve the ‘on’ to ‘off’ performance desired for crosspoints. Working back to an individual through capacitance, the ‘on’ value determined is close enough to what a parallel plate assumption would predict that the main problem in this case is that the ‘off’ capacitance is too high.

Chapter 4 Design of MEMS Structures for Packaging and Test

4.1 Introduction

This chapter takes a look at designing surface thin film MEMS (MicroElectroMechanical Systems) structures, with an emphasis on facilitating simulation and measurement of devices as part of an integrated system. After a short discussion of some important things to remember in MEMS design, modeling of MEMS structures and transmissions lines on the MEMS chip are discussed. Then the simulation and testing involved with connection to a MEMS chip using probing, wirebonding, and flip-chip connections are discussed. Finally some physical requirements of the package are discussed. Reading this chapter is not strictly necessary to understanding the other chapters. It is rather presented to aid someone new to MEMS design in understanding some of the major issues involved.

4.2 Basic Thin Film MEMS Design Concerns

While this is not a complete manual on how to design a MEMS device, some general information and “gotchas” are discussed. The focus will be on those aspects of design relating to measurement and integration.

One of the more important aspects of design is that a design be testable. This usually involves probing, so it recommended that the section on probing

be read and understood early in the design process, and abilities of the testing facilities available be understood. Alternates to probing are to wirebond or flip-chip from the beginning; the section on wirebonding is also worth understanding early in the design process- in any case since they are the most likely method for achieving integration.

The MUMPs process [39] will be used for illustrations here. Whatever process will be used, it is critical to understand the layer stackup and design rules. The design rules are helpful for understanding how mechanical structures are made, as they often flow not only from lithographic concerns, but the possible interactions of the stacks and etches. Figure 16 illustrates the MUMPs layer stack.

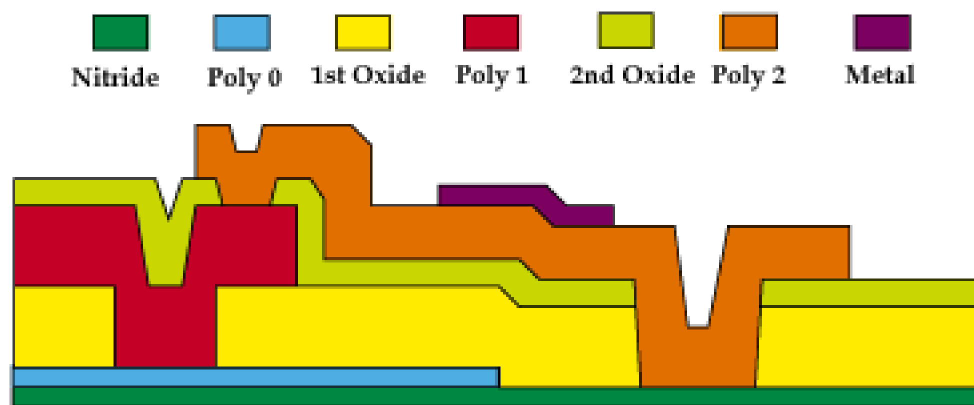

Figure 16 The MUMPs Layers. Source: [39]

One of the most important things to understand in making a MEMS device is what is happening vertically. When designing a device, it is a good idea to draw what happens as each layer is laid down, and each etch occurs. In partic-

ular, pay attention to rippling on upper layers from what happens below, and what will happen to layers below when above layers get etched. For example, if multiple polysilicon layers are stacked together, trying to then etch away one without etching the one below it is generally a bad idea. The layers below will also be cut into, perhaps etched all the way through. Also understand that structural properties will change when steps (changes in altitude due to lower layers) occur. For example, the Metal layer in MUMPs doesn't have good step coverage, so every step a metal line will cross has its cost in resistance. Once one has an understanding of how to stack things up the *reasons* behind the rules start to become apparent. Sometimes a rule is tighter than it is strictly necessary to follow because in that part of the device some material that lower down required that rule to be as large as it is isn't employed. Other times one needs to not only follow a rule, but expand it, because not only does the material need to be continuous, it is necessary to completely avoid any ripples from the layers below causing a tilt or lack of planarity. This being the case, another good idea when designing a device may be to contact some people with experience with the process when one has done a couple practice designs. Then go through each of the critical rules in that design with them to make sure that one is using it properly and understand why and how it applies. If a design is going to controlled by a CMOS chip, process variations become very important. This is because a change in a spring constant, or like thing affecting stability, can affect control voltages and dynamic properties significantly. In

particular, places like steps and small features are more affected by process variation than large flat areas. Therefore a design which utilizes lots of steps and small features will be more affected by process variations than a design that does not. The control chip has to be flexible enough to still control the device with the variations; this will sometime require some sort of feedback to the control chip that allows it to know what is going on in the MEMS chip.

When designing a device, care is needed to check out what is going mechanically in detail. If necessary make a physical macro-model of the device. Make sure that when parts move around, they will move as expected, and that when moved beyond the part of the range of motion intentionally being used nothing bad happens, and they can get back. Also be very careful about shorts when using electrostatically controlled devices. Avoid parts that could float away or tangle during release. In other words everything should be connected to something else directly, or should somehow constrained. Remember to leave sufficient etch-holes for the release step to be successful and remember that a device may be warped by residual stresses if care is not taken to control them.

4.3 Electrical Modeling of MEMS Structures

Most MEMS structures are small enough that for modeling them electrically transmission line modeling is not necessary in the device itself, though any transmission lines which connect to the device can be important and are discussed in section 4.4. Instead, bulk resistance, capacitance, and perhaps

inductance values will work for some devices if these are computed for the primary states of the devices. After solving for these with packages like FastCap for capacitance they can be put directly into an HSPICE model. For situations where this is not enough, such as capturing the dynamic and mechanical properties of devices, an integrated MEMS analysis package may be needed; for this, the documentation appropriate to the packages should be consulted.

A proper resistance estimate is very important. Resistance is usually much higher than one would expect by just looking at a design, especially if one has a CMOS background. In most thin film MEMS processes there is either no metal, or one metal with poor step coverage; usually most of the routing will be done in polysilicon layers with sheet resistances in the range of 9-40 Ohms in the best of conditions. Resistance is increased when going over bumps from lower layers, and vias between polysilicon layers are often high resistance (on the order of 220 Ohms)[39]. For estimating resistance of non-rectangular sections, [41] provides useful approximations for a variety of shapes.

Estimating capacitance is very important in electrostatic-based structures, and it is often important to take fringing into account. The standard formulas to handle this provide only rough approximations and are usually based on wiring rather than complex structures. One program that can be used to model capacitance in more detail is FastCap, which can be obtained from ftp://rle-vlsi.mit.edu/pub/fastcap/. One barrier to using this program is that entering shapes can be somewhat tedious. In order to facilitate modeling of objects

using FastCap, the practice of using macros that can be evaluated by the C pre-processor to prepare my FastCap inputs has been developed. Appendix 2.3 contains a short example of this technique. Keeping a design in the first octant (no negative coordinates, no zero coordinates) seems to be prudent. The fast-cap documentation should be read carefully- there are several gotchas, such as the fact that one must remember not to include interior surfaces. Also note that sufficiently large areas need to be subdivided to get good results; FastCap is a finite element solver and treats input as the surface meshing. The tightest parts of the mesh should generally be near edges.

4.4 Transmission lines in MUMPs

For sufficiently high frequency signals, transmission lines may be necessary to achieve good performance. For a transmission line to look distributed instead of like lumped components, the round-trip time for the signal must be greater than the rate at which the signal changes[42]. Thus, a given line may appear to have distributed effects at some frequencies, and lumped effects at others[43]. Transmission lines are important to keep the capacitance/inductance ratio of the line controlled and thus prevent reflections from causing trouble, as well as providing good return paths and thus keep phenomenon like crosstalk better controlled. One rule of thumb is that “if a segment of interconnect is longer than lambda/8 and significant signal distortion is intolerable, it must be built as a transmission line” [44].

The choices for transmission lines in the MUMPs process are quite limited, since there is only one metal in the process, and if a transmission line is needed, a polysilicon line is probably unsuitable. If the design rules are followed, instead of getting a fairly simple structure to analyze, stacks of materials result. The metal must be on poly2, either on a stack of more poly, over air, or on the nitride layer. Figure 17 shows the basic shape options other than simple strips of metal on poly2 on nitride.

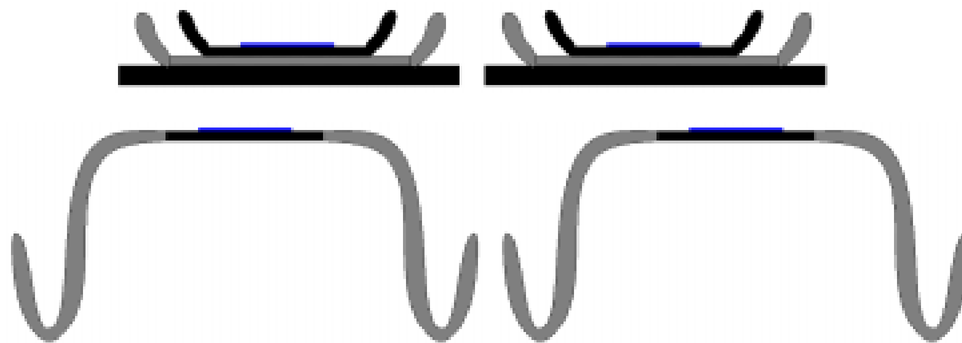

Figure 17 Basic MUMPs Transmission Line Shapes

In analyzing this structure, it is easy enough to find the resistance, but determining the characteristic impedance is much harder. Complicating the situation beyond the unusual shapes at the edges of the lines is the fact that many transmission line solvers assume that there is a non-lossy groundplane, and the resistance on the lines involved is also relatively low. Some more further issues include:

1. Should be the Poly be treated as a conductor or dielectric, or some combination?

2. If using a line in the air (which is necessary for crossing other signals on poly0), what effect do the periodic anchors have on the characteristic impedance?
3. If not using a line in the air, what effect does the substrate have? It is quite lossy, but is conductive, and can be made something of a ground plane using frequent substrate contacts.

Important design decisions include:

1. Whether to use the substrate as a ground plane, and how often to make ground contacts.
2. Whether to use lines primarily suspended or sitting flat on top of the nitride.
3. How much space to put between the lines.
4. If using air-lines, how often to place anchors, and what type of anchor design to use. Anything from a simple minimum size staple design, to much stronger structures can be used.

Besides simply building various transmission lines and testing their characteristics, the only method that would seem to provide a reasonable solution for modeling these lines would seem to be to utilize a good 3-D field solver that can handle the materials involved. This may provide a reasonable estimate of what the line looks like. Normal concerns in using transmission lines, such as selecting proper termination, continue to apply.

4.5 Testing Considerations: Probing

The first major step in testing occurs when the layout is being planned. Probing is a frequent test mechanism for MEMS devices. The selection of the type of probing used has a major impact on planning. Some general layout for probing concerns, and specifics about DC probing, and AC probing using the 50 Ohm and 500 Ohm probes available in the ERL lab.

4.5.1 Design for Probing

Probe pads must be made to proper specifications to accept the intended probes, and must be placed so as to allow all the necessary probes to be contacting at the same time. This will often involve placing the probe pads far enough from the structure being tested that the transmission line characteristics as well as the DC resistance of connections between the pads and the device become important. Another concern when planning for probing is that some types of AC probes will often bend out of their normal spacing. To compensate for this, probe pads significantly larger than the minimum size can be utilized there is space to allow this. Of course, the bent probes will no longer have the same characteristics as specified, particularly in terms of insertion and return loss.

The more probes to be landed at once, the harder it is to properly set up a probe station to get all the probes into place without accidental collisions. This becomes even more difficult using AC probes that require landing on both sig-

nal and ground pads, since the probe much come in at the proper angle. This is another reason to use larger pads than minimum, as well as why one should pay careful attention to angles and locations of probe pads. Finally, consider what type of microscope to be used. For example, in the labs used for this project, the basic needle type DC probes are out of the question for microscopes with lenses that must be close to the device. While some AC probes will fit, it's often close, to the point of being unable to rotate between lenses once the probes are in use.

4.5.2 Probe Specifics

Some of the probes used for measurements in the course of this work are discussed here. This section discusses how to read the probe sheets with some specifics about a couple of the probes as examples. The probes described here are from the Picoprobe line of probes [45]. Viewing the data sheets on the web site while reading this section may be beneficial.

For DC probes, such as the T-4 series, important things to know are the size of the landing point and the length of the probe (to know whether it will fit under the microscope with the available probe setup). The tip size and material basically tell how the tip will affect the landing surface: both how much damage it will do landing, and how good a contact it will make. Not much else will be specified since DC probes are not designed to carry high frequency signals.

For AC probes, there is a lot more to know. The impedance of the probe is important to know for matching the probe with the lines to the structure being tested. Frequency range and the relative strength of the top frequency (usually -3dB, the half power point) are important to know the probe is useful at the frequencies to be utilized. Which side of the probe the “ground” and “signal” lines are on is a simple, but important, attribute. One of the most important things to know before layout is the spacing, or pitch, of the probe tips. A pitch of 100um was used for most of the layout described here.

The Attenuation is related to a series resistance added near the signal probe tip. By adding the series resistance, the amount of signal which goes into the probe rather than remaining in the circuit is reduced by a factor of the ratio between the added resistance and the impedance of the rest of the probe system. So a 500 Ohm resistor on a (otherwise) 50 Ohm impedance probe would lead to a 10:1 attenuation.

The insertion and return loss are specified in dB. Insertion loss is the attenuation of the probing, so the smaller this number is, the better the probe. 3dB is the specified insertion loss for the model 10. The Return loss is the amount of reflections and such caused by deviation of the probe from its specified impedance, so the higher the better. Our probes are usually 10dB; a very good probe might be rated at 25dB return loss.

4.5.3 Two example probe stations

At ERL there are two probe stations. One has a microscope with a lot of clearance between itself and the sample, but not as much magnification. The sample attached to a moveable puck via vacuum, while the probes sit on a platform around it. The microscope is moveable relative to the sample, so once the puck is properly placed to look at a given device it's a good idea to leave it there and not move it. The puck should never be moved while probes are down. It will drag the probes across the chip, damaging both the chip and the probes. This station was intended for probing CMOS chips, rather than observing movement of MEMS structures, which is why the lower magnification with larger clearance was chosen.

The other station was designed in order to get data about optical MEMS devices. The magnification is significantly greater, and diffractive lenses are available so that interferometer measurements can be used to measure vertical distances and changes in them while devices are in use. The effectiveness of this depends on whether the building is shaking much: interferometer measurements require there be very little vibration. A heavy table (originally designed for use as an optical test bench) resting on compressed air is used to decrease vibration. Even when the interferometer is not available, the focal depth is such that when focused on a given height a rough idea of the height of other features can be determined by how much they are out of focus. When testing MEMS piston devices, this allowed observation of whether a device

was completely up, down, or somewhere in the middle. This is in contrast with the other probe station, where movement was only barely visible with large devices and careful lighting. Another feature of the second station is that the platform the probes and the chip sit on can be moved together and tilted together, allowing to be moved while the probes are down to look at other parts of the chip. Both probe stations have cameras attached to allow pictures. There are a number of drawbacks with the second probe station also. There is less space for probe manipulators on the platform, an there is much less clearance between the lenses and the chip. The 100x lens can only be used with the 500 Ohms AC probes due to clearance restrictions, moved into place counterclockwise when the probes are completely down.

4.5.4 Modeling of probing

When designing experiments and testing it is necessary to make a model of how the probes and probing setup will affect the operation of devices, even if the final system will not utilize probes. Some discussion of the issues involved in modeling follow.

In the case of a design where only DC is involved, so that a DC probe is sufficient, all one has to worry about is resistances. Of course, there are still pitfalls. The contact resistance of the DC probe to a polysilicon probe pad will often dominate the extra-sample resistance, and can vary a lot depending on how good the contact is, the condition of the probe tip, and some currently

unknown factors. Measuring the resistance of some simple polysilicon lines on one sample whose resistance should have been around 120 Ohms, the measured resistance varied from not far above the expected value all the way up to 230 Ohms. Simulations should take into account the various corners created by variations in the contact resistance while probing, as well as the process variations in the design as discussed elsewhere. Of course, the resistances of the lines from the DC source to the probe, the probe itself, and the supply model should be checked as well.

For AC probing, the DC concerns still apply, but the AC characteristics of the probes come into play as well. These can usually be determined from the data sheet for the given probe. The transmission line to the probes can be important; this can often be modeled in HSPICE using a transmission line model. The paths to the pads should be evaluated for possible transmission line characteristics, as well as possible impedance mismatches with the probes, should also be considered. Remember the insertion and return losses are specified for a matched system.

4.5.5 Doing the probing

Whenever using a probe station, one of the most important things is to think about every move one makes before doing it. Not only to ask “Is this what I want to do?”, but also “Will I bump any of the probes or the wires hooked to them?” and “Will this shake the probe station?”. If probes are down

and there is any shaking of either the probe station, or anything attached to a probe manipulator (i.e., the signal wires), the probe tips WILL move, probably removing them from their pads, and quite possibly damaging the probe tips and the device being tested. A few guidelines when setting up for probing follow:

1. Try to maximize the accessibility of all the controls on the manipulators.
2. Try to minimize the number of wires and connectors AC signals must flow through.
3. Secure everything. If a vacuum pump is available to hold the chip down, use it. If there are wires attached to manipulators, attach them to the probe table, and run them as out of the way of where hands will be moving as possible. But see rule 6 regarding placement of the pump.
4. Many common manipulators use magnetic bases to attach to the probe station. These provide a small measure of stability in the sense that they help prevent the manipulator from lifting from the table. It will not help keep them from rotating along the plane of the table, so don't expect them to.
5. Pay attention to the surface of that unsecured probe manipulators rest on. If it's not flat, the probe manipulator will be able to wobble, even if the surface is something a magnet will stick to. Keep in mind that if the manipulator tilts toward the chip, it could wipe out half the devices on the chip and destroy the probe as is dragged across the chip.
6. Don't create a situation in which something will have to be done later that will cause the probe station to vibrate. For example, avoid situations where turning on and off a power supply will shake the table. Instead, put the power supply somewhere else, such as a shelf next to the table.
7. Don't get out the sample for test, or the probe tips, until confident everything else than can be prepared in advance is set up correctly.

4.6 Packaging

Wirebonding can often be done using similar pads as would be used with a DC probe.

Connecting two chips together directly, or connecting the face of the MEMS chip to an MCM are referred to here as flip-chip approaches. Solder-bumps are usually used for alignment and attachment when a MEMS chip is connected in this manner. The solder bumps also carry signals between the MEMS chip and what it is attached to. One advantage of this method is relatively low inductance: <1.0nH[46]. Another method of signaling which is inappropriate for low-frequency signals, but which may work for high-frequency signals, is to capacitively couple the MEMS chip to the other chip. This involves putting large plates on both sides and sending the signal across without the plates actually coming into contact.

Besides the electrical considerations, there are some mechanical considerations with using a flip-chip approach. Devices may need to be placed distant from the solder areas due to stresses caused by the solder-bumps. Also, except when a smaller chip is put on top of a MEMS chip access to any visual information about the devices will be lost. When making the connection, it is necessary to be careful not to crush or otherwise mangle the MEMS device with the other chip or substrate. Another thing to be careful with is the relative thermal expansion; a bad mismatch can cause large stresses in the MEMS chip

even when it does not lead directly to failed connections. These stresses then may interfere with proper device performance.

Any package used for a MEMS chip must avoid making physical contact with the devices (except where required by design, such as in a pressure sensor). In addition, it is usually necessary for the package to prevent the entry of moisture, since moisture leads to stiction in many parts. Finally, the orientation of the devices is important in such application as accelerometers, and any devices that are affected by gravity. Another case where alignment is important is that of optical devices. Depending on the intended application, it may be necessary to test the ability of devices to handle vibration and physical shock absorption due to their mechanical nature.

4.7 Conclusions

When designing MEMS devices it is important to consider the intended methods of connection and packaging early in the design process. Of particular importance are understanding the effects of parasitics, especially resistance, and the difficulty of making good transmission lines in most MEMS processes. Taking these things into account when preparing and modeling a design should increase significantly the chances of reaching a working design, or at least a testable one.

Chapter 5 MEMS Crossbar Architecture

In order to build a large crossbar using MEMS crosspoints, more than just the MEMS devices are necessary. This chapter explores the other aspects of such a system. A particular size crossbar is used as a target for illustration; other large crossbars could be built using the same principles and the same or similar components.

5.1 System Description

The target system is a high performance 192x192 non-blocking crossbar switch. This means a switch that can simultaneously route any of 192 inputs to any of 192 outputs. While such a component would be very useful, it is not usually made today because with CMOS technology it isn't really feasible due to speed, power, and area constraints. For this reason, it seems a good point to explore the capabilities of using MEMS capacitive switches instead of conventional methods.

The system would consist of three basic components: the chips with the MEMS devices, the chips with drivers and receivers for the capacitively coupled signal, and the chips which provide the controls to the MEMS devices. Other chapters describe the chips with the MEMS devices at length. Chapter 2 investigates methods for design of the drivers and receivers, but actually making them is beyond the scope of this dissertation. Some of these compo-

nents were built by John Wilson as part of the DARPA project funding this work. Section 5.6 will discuss the design and testing of the control chip. The rest of this chapter discusses how these components would be connected to make the overall system.

The major drawback of this system for many applications will be that since mechanical switches are used it will take longer to reconfigure the switch than a traditional CMOS switch might. Therefore this switch is targeted more at applications in which the switch will remain in one configuration for a long enough time that a 1ms reconfiguration time is unimportant. This significantly impacts the architecture chosen; for example the long reconfiguration time contributes to the need for a non-blocking approach.

Rather than comparing to CMOS crossbars, another possible comparison is to all-optical crossbars. In that area, the comparison would be against other types of MEMS switches used in a large system. Compared to these, the benefits of the capacitive switches is that they are much smaller, and can be used in arrays, thus having the potential to provide large physical volume savings. Also, optical switches are currently quite expensive, so significant cost savings are possible. When comparing to mechanical optical switches, the reconfiguration time due to the use of mechanical capacitive switches is no longer an issue.

5.2 Crossbar Architecture Selection

Consistent with the goal of building a non-blocking crossbar and the previous work discussed in Chapter 2, there are only two feasible options to explore. One is to build a single 192x192 array, and the other is to make a Clos network. Due to the current quality of the currently available processes and the maximum achievable ON/OFF ratio, it is not possible to fit an entire 192x192 array on a single MEMS chip in either available process. The routing necessary to achieve acceptable resistances, and the device sizes necessary to create sufficiently high capacitances in the signal areas while having large enough area actuation plates to operate at reasonable voltages simply require too much area; the additional area necessary to bring 192*2 signals and 192*2 control lines onto the MEMS chips (or even 192*3 lines, assuming the coupling is done beforehand) just make it even more impossible. Therefore an MCM would definitely be needed if the direct implementation were to be used. In the Clos case the interconnect between stages are sufficiently complicated to provide significant challenges even with post-process metallization and SUMMiT, and even the reduced area required for the crosspoints is still too great to put on a single MEMS chip. The move to the SUMMiT process which has only 1/4 the area per chip increases the number of MEMS chips needed even more. Since a single 32x32 array is nearing the upper limit of what can be fit on a single SUMMiT process chip, a 6x6 array of such chips would be

required to make a 192x192 array in the direct implementation with the chips tiled.

5.3 Characteristics Of A Direct Implementation

While the Clos network was the chosen architecture, it is worthwhile to consider the characteristics of the direct implementation. It has the advantages of being very simple to program, and by avoiding CMOS sub-arrays would probably achieve better power performance. Ignoring spacing between chips, a 6x6 array of SUMMiT chips would be about 3cm per side. The signal quality management issues would be more difficult, with six times as many “off” capacitors in shunt to other signals on each side of the “on” series capacitor. Given the difficulties in achieving sufficient switch and interconnect quality, it is just as well that this architecture was not chosen. If the whole array is programed as one large array, 24 of the high voltage driver chips described in Section 5.6 are needed to control all of the devices. One driver is needed for each input line; one receiver is needed for each output.

5.4 Size Limitations

Based on the size of the MEMS chips, and the likely sizes of devices, it is possible to determine how many devices can be fit on a single SUMMiT chip. Chips in the SUMMiT process have a usable area of 4660um by 4660um. The devices described in Chapter 7 are designed to fit into a 120um by 120um box,

including their local routing, when tiled to form an array. 4660/120 = 38, but area is needed for pads also; 32x32 would leave sufficient area for perimeter connections. If there area for the device were decreased (and the decreased performance were bearable) as large as an array as 39x39 might be squeezed into a chip, but it's probably not worth it.

If a 32x32 array is used, at 120um by 120um cell size, this leaves 4660-32*120 = 820um for I/O on each of length and width. Wirebond pads that are spaced 160um center to center would allow only 24 in line with each row or column, so three lines of wirebond pads would be needed to bring in 2 lines per device. This would fit, with a little space left over to ease things a little farther apart and allow substrate pads and taps on the corners of the chip.

5.5 Clos Array Sizes Selection

Traditional Clos theory would suggest the number of inputs to each first stage array be determined by utilizing the relation $N \sim 2n^2$ [7]. This is based on a minimization of the number of crosspoints needed. For a 192x192 array, this would lead to the case of 20 10x19 subarrays at the input, 19 20x20 arrays in the middle, and finally 20 19x10 arrays at the output. This would require approximately 25 MEMS chips using the SUMMiT process. The possibility of using other valid Clos configurations that might make an easier to manufacture array at the cost of an increased number of crosspoints was explored. A chart of the valid networks was generated, and is included in Appendix 2

along with C code to produce the set of Clos networks for an arbitrary sized array. Examining the possibilities, it can be noted that if some sections of the subarrays were sufficiently large, other portions could be small enough that they could be fabricated in CMOS. This would allow the advantages of a MEMS array and CMOS array to be combined, by making each section of the array using the technology most appropriate to its size. The design space in this area was explored, examining the area and latency trade-offs in the inner stage MEMS arrays and outer-stage CMOS arrays. For the 192x192 case specifically, if one pushed the amount one could fit on a SUMMiT chip to the limit, the sweet point is to use 5x9 and 9x5 CMOS crossbars (which could be made with acceptable throughput) for the input and output stages, and 39x39 MEMS subarrays in the center which could probably be fit into a single SUMMiT chip with sufficiently good interconnect and area to maintain a 1GHz data rate and low latency for a good crosspoint of reasonable size. Thus, only nine MEMS chips would be needed, thus potentially significantly decreasing the packaging cost compared to the 36 needed for the direct case while preserving data rate, and increasing the latency to approximately 1.4ns. Power consumption per crosspoint (and probably overall) in the CMOS sections will be significantly greater than in the MEMS sections. Using the more conservative 32x32 array, 11 rather than 9 MEMS chips are required. This leads to 6x11 and 11x6 CMOS arrays, which is still feasible.

Based on comments at the poster session of a DARPA PI meeting there is an interest in 256x256 crossbar switches. Noting the interest in making large crossbars of various sizes, this provided another reason to select a standard size for the center crossbar: 32x32. This will allow a higher quality of on-chip interconnect, allowing lower latency and a wider data bandwidth while still filling a SUMMiT chip than pushing to 39x39 would. In order to use this component to make a 256x256 array 8x15 and 15x8 outer stage crossbars would be required as well as 15 MEMS chips. Alternately 16 chips each containing two 16x31 arrays each could be used with 31 square 16x16 middle stages (requiring another 8 MEMS chips) if the 8x15 and 15x8 arrays are not feasible for the desired data rate and latency in CMOS.

Therefore, the chosen overall switch architecture for the 192x192 switch is to utilize 32 6x11 CMOS crossbars followed by signal drivers into the MEMS sections on the first stage utilizing a 0.25um technology so as to achieve required performance, followed by a center stage of 11 32x32 MEMS crossbars, with the higher voltage controls for these arrays on larger feature size (cheaper) CMOS chips. The final stage would again be CMOS, receivers for the signals from the MEMS arrays followed by 32 11x6 crossbars. Figure 18 illus-

trates the connectivity of such an array. Each crossbar connects to each crossbar in the next stage exactly once.

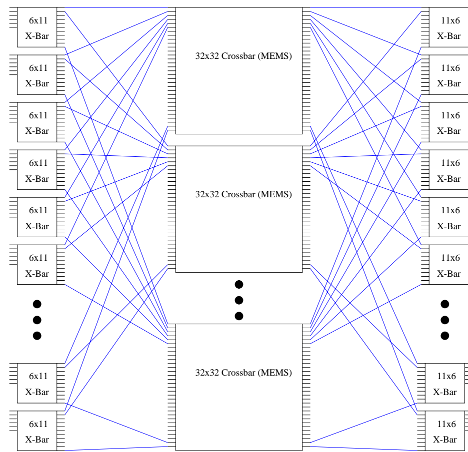

Figure 18 A 192x192 Clos Array using 6 32x32 centers

Information for programing the array can by a relatively simply program as shown in Appendix 2, or a control chip can be made to handle this. One advantage of the offboard control program is that it allows changes in programing strategy without adding nonvolatile memory and computing elements to the already crowded MCM. The physical reprogramming time for the MEMS (on the order of 1 ms) is already great enough that external computation of reprogramming arrangements should not have an important time impact.

5.6 High Voltage Drivers

In order to control the MEMS devices utilized, each MEMS chip requires high voltage drivers for each row and column. To perform this task, chips were fabricated using the Orbit 2um process. Each of these chips has 16 output

lines, so two for each axis of a 32x32 crossbar chip. The combinational logic of the chip of the chip is shown in Table 3. A<0:4> column is the output of the Address latches controlled by the clock.

Table 3 HighVolt Logic

Enable1	**Enable2**	**Disable**	**Row**	**A<0:4>**	**Out_A**	**$Out_{non\text{-}A}$**
0	X	X	1	X	V_H	V_H
0	X	X	0	X	V_L	V_L
X	0	X	1	X	V_H	V_H
X	0	X	0	X	V_L	V_L
X	X	1	1	X	V_H	V_H
X	X	1	0	X	V_L	V_L
1	1	0	1	X	V_L	V_H
1	1	0	0	X	V_H	V_L

The input and output have separate supplies, so that the inputs can be controlled by standard logic levels, while the outputs are appropriate drivers for MEMS devices. The output driver stage was done by John Wilson. In order to control a 32x32 array, four chips would be used, two to control the rows and two to control the columns. The ‘row’ chips would be at the low levels of V_H and V_L, while the ‘column’ chips would be at the higher levels; the difference between the ‘row’ V_H and high V_L should be a valid release voltage; the difference between the ‘row’ V_L and high V_H should be a valid set voltage; the other two voltage combinations should fall within the hold region for the devices controlled.

The logic has been verified to work correctly in chip testing. The pinout of the chip is given in Table 4; Figure 19 is the custom chip layout.

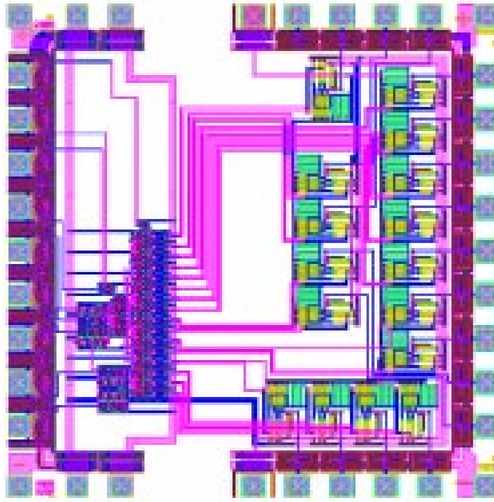

Figure 19 Layout Of Control Chip

Table 4 Pinout for HighVolt Chip (Pad supplies noted with P)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	O9	O5	O13	O15	V_HP	NC	O7	O11	O3	NC
10+	Vdd_H	Vdd	Gnd	VddP	NC	NC	Row	En1	En2	Dis
20+	CLK	A1	A0	A3	A2	GndP	Gnd	Vdd	V_L	O12
30+	O4	O8	O0	NC	V_LPd	O10	O2	O14	O6	O1

Power measurements of the chip indicate the power consumption of the high voltage driver stage to be fairly low; maximum current draw into the V_H supply at 10V was 0.61mA. The heavier power draw was on the control side, where the pad ring and circuitry can consume 2mA from its 5V supply; this could be improved with better pad design. As the output voltage is increased, power consumption starts increasing quickly, due to the protection diodes on

the output pads. 1.8mA is drawn when V_H is set to 20V, and only 14V is output. Thus, these chips are best used for hold regions of 10V or less, which is consistent with the MEMS devices developed. Assuming 10V is used, total power for each chip would be about 16mW.

When controlling multiple MEMS chips, it is possible to trade off reconfiguration speed for less control chips. If the MEMS chips are arranged in a logical grid for control purposes, with only one device in all the chips reconfigured at a time instead of reconfiguring all chips simultaneously, then only one chip for each 16 devices in each row, and again in each column, is needed. For example, if 11 MEMS chips are needed, and they are arranged in a 3x4 array (with one corner missing), then 6+8=14 high voltage control chips can be used instead of 11*4=44. Similarly, if a 192x192 array is built directly, requiring a 6x6 array of 32x32 MEMS crossbars, then 12*2 =24 high voltage control chips are needed.

5.7 The System

32x32 seems to be the largest sub-array that can be reasonably included on a single SUMMiT process chip. Of course, multiple chips could be combined in larger direct crossbars.

So, we have 32x32 subarrays. Four of the control chips are needed for each of these subarrays if separately controlled, or two per row and column if they are logically arranged in an array for control purposes. Using 6x11 CMOS

crossbars in the driver chips and 11x6 CMOS crossbars in the receiver stage allows a 192x192 Clos network to be made with 11 of these 32x32 subarrays. 32 each of the driver and receiver sections are needed. Two of these CMOS crossbars can be fit in a single chip, so the system would consist of 32 (CMOS) +11 (MEMS) + 14 (control) = 57 chips; if each MEMS array is controlled separately, and additional 30 control chips are needed. The low cost control chips made in a 2um feature size process. Except for the MEMS chips, the main reason for the large number of individual chips is the I/O limits of each chip. The drivers and receivers for the signaling through the MEMS switches are incorporated into the CMOS crossbars. The CMOS and MEMS crossbars would be mounted on an MCM or other substrate.

5.8 Signal Path

It would be nice if the signal path through the crossbar would simply look like a series capacitance, which would sufficiently dominate other components that they could be neglected. In practice, the signal path is considerably more complex. The crosspoint itself has a shunt capacitance to AC ground due to the actuation plates. If the signal is routed through the springs, there is a significant series resistance added here. If the signal goes through signal capacitances in series to avoid the resistance, the effective through capacitance is significantly decreased.

The path to get to the crosspoint passes by some number of ‘off’ devices, from zero to N (where N is the size of the array) and then passes another set of ‘off’ devices to reach the output. By making connections to each edge of the array, instead of one side only, this situation can be improved somewhat. Of course, the connections from the driver to the crossbar, and crossbar to receiver, need to be modeled as well. While certain smaller capacitances can be neglected on the device level (such as the shunt capacitance from the signal line to the actuation line, they may become significant in a long row of devices. Similarly, for a large enough array, inductance may come into play. Another factor is that the signal passing through ‘off’ crosspoints turns into noise on other lines.

A detailed model of a crosspoint without metallization based on inspection of the layout is presented in Figure 20. This device does utilize the springs as part of the signal path. Several of the components in this model can be

neglected or further combined in analysis, but this has not been done for the figure to leave each component's physical source discernible.

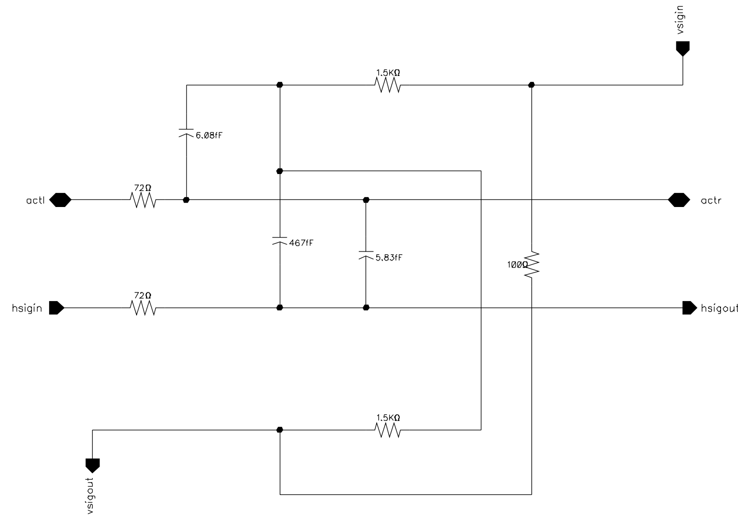

Figure 20 Unmetallized Model of SUMMiT device 'on' state

Most of these resistances are greatly decreased with metallization. The 1500 Ohm resistors represent the springs, which are shielded from the metallization. The capacitance between the hsigout line and these resistors is the through capacitance for this device.

As will be discussed in section 7.9, a SUMMiT devices with two capacitors in series was fit to a rough model with only the series and shunt capacitances, and a series resistance. This device achieved a 4:1 ratio between the worst 'on' capacitance and the best 'off' capacitance. A ratio of approximately 5:1 is exhibited for the nominal voltages. With the receivers built, this would only be sufficient for very small arrays. The capacitance ratio needs to be approxi-

mately 2*n, for an n*n array of devices to barely work. 3*n is preferred. Therefore a device with at least a ratio of 64 (preferably the originally planned 100).

5.9 Overall Power Consumption

Assuming a 32x32 MEMS array is used, and that the CMOS crossbars, receivers and high voltage driver chips developed are either suitable or indicative of overall power consumption, it is possible to estimate the total power that would be consumed by the crossbar illustrated in Figure 18. Each of the CMOS driver chips would consume about 48mW; each of the CMOS receiver chips would consume about 62mW. The 14 high voltage control chips would each consume about 16mW. This brings total power consumption to approximately 3.75 Watts. Even if 44 high voltage control chips are used, total power consumption is only approximately 4.25 Watts. This is much less than the power consumption of the crossbars discussed in section 2.2.1.

Chapter 6 Metallization

While polysilicon provides a good mechanical layer, it's resistance is much too high to provide a good chip-wide interconnect. In order to achieve low stress polysilicon, anneal steps at much higher temperatures than most metal layers could survive are utilized in both MUMPs and SUMMiT. This prevents them from including a metal layer before the final anneal. One of the major problems with the MUMPs designs is the poor metallization. The original plan involved modifying the MUMPs process to produce better devices, but sufficient process modifications to create an acceptable interconnect were infeasible. In order to fix other problems, movement to the SUMMiT process (described further in Chapter 7) became necessary; post-processing to improve the SUMMiT process by adding metal is needed to contain attenuation losses.

6.1 Metallization Options

In order to reduce resistance to reasonable levels for high speed electrical switching, some form of metallization is necessary. Since SUMMiT does not have a metal, some form of post processing is needed when using this process. A pre-release metallization method, such as the lift-off method used in MUMPs, would limit metal to the top level of the design. This would not provide the low resistance interconnect desired in any design where lines crossed.

Also, in the case of lift-off, some of the other problems of the MUMPs metallization would recur: poor step coverage and relatively thin metal.

When metallizing after release, some form of masking is necessary to avoid shorting the entire chip together. This can take the form of a proximity mask, or building the mask into the design. Both solutions have been used in the past[50]. A self-masking design allows greater control of where the gold is deposited than a separate shadow mask. Also, in the case of a released chip, such a mask could not contact the surface of the chip without causing damage. On the other hand, design of a self-masking chip is much more complex. Nevertheless, this approach has the potential for much greater control of the final location of the metal, as the 'masks' are already well aligned, and very close to the location desired. Deposition without an additional external mask seemed the best option for the intended application and resources, so layouts were prepared with this in mind. As gold was selected for the metal to deposit, a thin Cr or other suitable metal precedes the gold to enhance adhesion. In addition to adapting the general self-masking technique described in [50], novel crossover techniques were developed improve wiring resistance.

6.2 Layout for Metallization

In order to avoid the deposition of metal on unintended areas, the mechanical layers of polysilicon are used to mask layers below them. The metal is assumed to be deposited from a source above the chip, and is assumed to

spread out at some maximum angle from the vertical. To allow use of sputtering, a 45 degree angle of spread during deposition was assumed in preparing the first chips. This implies that metal passing through a 1um square gap in one layer of polysilicon, and then traveling 1um, is assumed to spread to cover no more than a 3um by 3um area centered below the gap. In practice, a number of factors can lead to a smaller coverage. A complication arises for moving parts, in that one must account for all their possible positions at the time of metallization. For fixed parts however, one can generate a table of necessary overlaps for one layer to shield another layer for normal layer heights. The 45 degree assumption allows easy estimation of clearances for a given set of layer heights, but care must be taken to account for such features as dimples, anchors, and other features which change the distance between given layers. In practice, complex interactions should be checked using a cross sectioning CAD tool.

An important distinction between the approach to self-masking used here and the guttering approach described in [51] is that while in guttering, the metal is primarily intended to be deposited on the top layer, and the gutters between top level features are shielded to protect from metal passing through, here metal was specifically prevented from landing in certain areas to avoid shorting and mechanical problems, but otherwise allowed to form a large ground plane on the nitride. In contrast, in a guttering approach most of the metal not used for something else winds up on a polysilicon stack gutter.

6.2.1 Mechanical Concerns

It is important to put metal on the lines carrying signals, but metal should not be permitted to land on the springs, or cause the capacitor plates to be significantly more 'bumpy' than they otherwise would be. Nonuniformity in the capacitor plates would compromise the flatness and decrease the 'on' capacitance. Metal on the springs would change the spring constants, as well as raising issues of metal fatigue and delamination. Specific application of this to some of the device designs in SUMMiT can be found in the device design description in Section 7.5.1.

The other general mechanical concern is that the added stress from the metallization not compromise planarity. This limits the total thickness of metal that can be deposited unless important flat areas are shielded. While adding metal to the moving plate will change it's mass, and thus somewhat affect it's dynamic characteristics, this should not affect the actuation points, as the effects of gravity are much smaller than the other forces involved.

6.2.2 Low Resistance Line Crossings

A novel feature of this metallization scheme is a technique to decrease the resistance of line crossings. A usual aspect of a single layer of metallization is that when two lines cross (in particular the MMPOLY3 lines and the MMPOLY1/MMPOLY2 stack lines) there would normally be no metal on the lower level. Even when metallizing a released structure, there would be no

metal on the lower level where the stacking occurs. In order to allow the sputtered metal to reach the lower level, minimum sized for the time (2umx2um for MMPOLY3) holes in a rectangular grid with minimum spacing (2um edge-edge for MMPOLY3) in the crossing areas have been added so that the metal is able to reach the lower level MMPOLY layers. While this leads to higher resistance in each layer than if it were solid and completely gold-covered, it will lead to a lower total resistance to a signal that must pass through both directions. Since this is a crossbar, this is true of all important signals. It is also important to use planar wires without changes in levels, since vias consist of polysilicon and greatly increase resistance. Figure 21 illustrates the basic crossover plan.

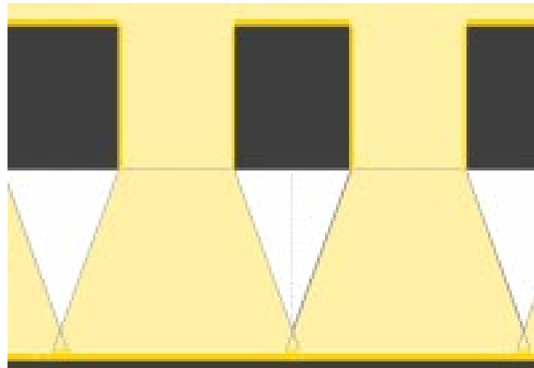

Figure 21 Basic crossover metallization

In preceding discussions, the important aspect of spread angle has been the maximum spread angle. In order to design a crossover, the minimum spread angle (i.e., the minimum the metallization spreads out after passing through an opening) becomes important in determining whether the technique is suc-

cessful. Note that a deposition passing through such a grid may not retain the full spread it had entering the grid. Consider the effect of the size of an opening in a layer on the maximum spread possible. Assume that the deposition travels only in straight paths, and does not bounce or diffract off the polysilicon. The thickness of the polysilicon layer is important. MMPOLY3 can be used as an example. MMPOLY3 is 2.5um thick. The cutoff for a 2um wide slit can be determined using simple trigonometry. Metal missing one of the upper edges by an infinitesimal, and the opposite lower edge by an infinitesimal, travels 2um across and 2.5um down. So the cut off angle for this situation would be the inverse tangent of 2/2.5, or 38.66 degrees. For a 3um wide hole, 50.2 degrees. For a 4um wide hole, 58 degrees. Similar calculations can be done for other layers and cut sizes.

If we wish for two lines to cross, the top line in MMPOLY3, we can determine for a given grid and minimum spread how far the destination layer must be from the MMPOLY3 in order for the metallization coming through the grid to form a continuous path along a lower level. For a 2umx2um grid with holes spaced 2um edge to edge, we need the metal to spread out by 1um from each hole for each adjacent hole to connect, so at the maximum 38.66 spread, 1um*2.5/2 = 1.25 um. Assuming no dimple is used, the oxide between MMPOLY2 and MMPOLY3 is thicker than this, so a 2um grid should be sufficient. On the other hand, a 4umx4um grid, with holes spaced 4um apart, would need 2.5um of travel distance, if the same spread applied. Of course, a

greater spread can actually pass through such an opening- assuming the metallization procedure supplied that spread. Indeed, controlling spread from the metallization technique is a key factor for success.

One possible method to increase the distance of travel between MMPOLY3 and MMPOLY2 or MMPOLY1 when needed for a given situation would be to put the line in a dimple to increase the gap.

6.3 Metallization (Sputtering): First SUMMiT Run

Samples of chips from the first SUMMiT run were metallized using sputtering. A Cr adhesion layer (20nm) was utilized, and the chips were on a planetary rotator during metallization. Both a thin and thick deposition of gold were used on a few samples each. During both depositions a base pressure of 1e-7 Torr, sputtering pressure of 2e-3 Torr Argon, and sputtering current of 0.3A were utilized. Approximately 20 nm of Cr was used as an adhesion layer in both cases. The thin deposition was 23 seconds (approximately 50nm); the thick deposition was 5 minutes 15 seconds (approximately 700nm). Eight chips were metallized; two at each setting for each of the two designs from the first SUMMiT run.

6.3.1 Visual Inspection

A visual inspection with partial dissection provided some information about the characteristics of the deposition. The first chip removed was from the thick

film metallization of the primarily electrical switching chip. In order to examine the spread, one of the devices was disassembled one. Some key observations can be made using the images in Figure 22.

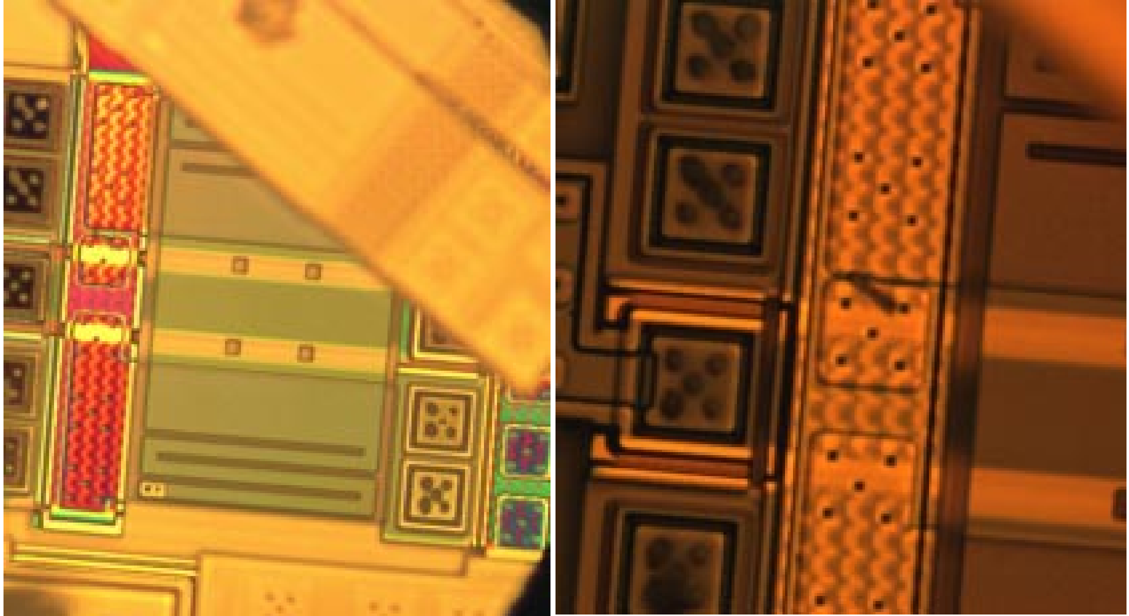

Figure 22 Dissected Device from first metallization

In the left image, the MMPOLY3 covering is being removed; the image on the right is a higher magnification view of the left center region of the first image. The wavy lines running vertically along the MMPOLY2 line are due to the crossover grid in the MMPOLY3. The grid was made of 2um square holes with 2um edge to edge spacing. These pictures were taken using a DI lens, therefore the differences in brightness represent differences in the thickness of the metal. The obvious theory as to the orientation of the wavy lines is that this is dependent on the position of the chip relative to the source throughout the disposition process, so that one can not afford to assume that the metal

will be deposited so optimally in the future (in this case, the thickest metal makes paths in the direction in which the signal travels). The slightly darker areas do however also have metal on them, just not as thick. The device itself only has metal on some places, as the metal has been intentionally shielded from much of it to protect the mechanical properties.

Another feature of interest is the parallel lines of metal running along the center of the device. These are due to slits in the MMPOLY3 between the signal and actuation plate. From the width of these lines it is possible to estimate the spread of the metal in the direction perpendicular to the slits. From the wavy lines one knows that the maximal spread is close to this axis. Note that the width of the lines is greater than the etch holes in the device. These holes are 4um square. Note that the amount the metal extends beyond the holes is greater toward the bottom of the picture than toward the top. The slit is 2um wide and centered over the hole. The distance the metal traveled from MMPOLY3 to MMPOLY1 is estimated at 4.8um by an in-house elevation tool. Note this tool does assume we know the thickness of OXIDE3, which in practice varies due the planarization step and was not specifically measured here. The extension of metal past the holes appears to be approximately 1um toward the top of the picture and 2um toward the bottom. This represents a maximum spread slope of 2/4.8, or a spread angle of 22.6 degrees. Note that this is less spread than the success of the grid would seem to imply. It is quite possible however that the device partially actuated during the metallization

process, which would change this distance as well. The width of the MMPOLY3 gap over the MMPOLY2 line is 3um, and here the spread would appear to be about 1um. The estimated vertical gap in this area is 3um. So a spread slope 1/3, or 18 degrees. These angles are fairly similar, given the roughness of the estimates, so the device probably did not significantly actuate if the assumption of OXIDE3 thickness is reasonable.

6.3.2 Electrical & SEM Tests

A variety of electrical tests were done to check for shorting and measure resistivity. It was discovered that signal pads were systematically shorted to the general metal coating on the nitride. For the thick metal this short was relatively low resistance; for the thin metal it was in the kiliohms. It was not determined with the electrical tests whether this was a result of the metallization scheme or a layout problem (i.e., a mistake violating the metallization plan). The contact resistance of the probes varied widely, making measuring the exact value of the resistance of the short difficult. Section 6.6.2 discusses the source of some of these probing difficulties.

Samples were viewed using a SEM to look for the error. One hypothesis at the time was that the shorts resulted from stringers. One of the more interesting SEMs illustrating the metallization surface is shown in Figure 23.

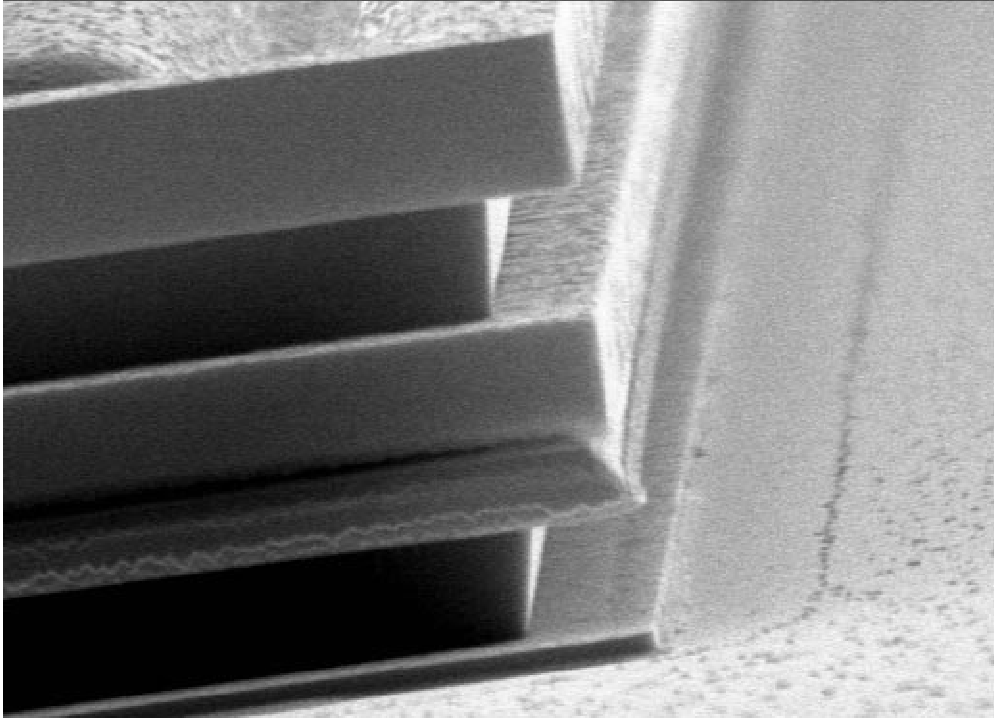

Figure 23 Metallization near back of pad

Some features visible here are a slant in MMPOLY1, and some interesting variability in the metal on the end of this slant. Further the texture of the metal on the nitride, and a clear, but irregular, line in the metal near the pad, but far enough away not to expect to be the actual edge of the metal, is of interest.

Eventually a layout error was discovered in the GSG pads used all over the chip which leads to a short in the MMPOLY1 layer. Throughout the tripad

design, the cutting of MMPOLY2 is used to cut MMPOLY1 as well, and there is no SACOX2. The problem is that at the edge, the SACOX2 cut stops before the MMPOLY1 stops. The MMPOLY1 is not separately etched. This causes there to be a short between the isolation ring around the pad and the rest of the pad. The isolation ring is electrically connected to the substrate when metallization occurs (by design). Figure 24 illustrates the location of the error in the layout:

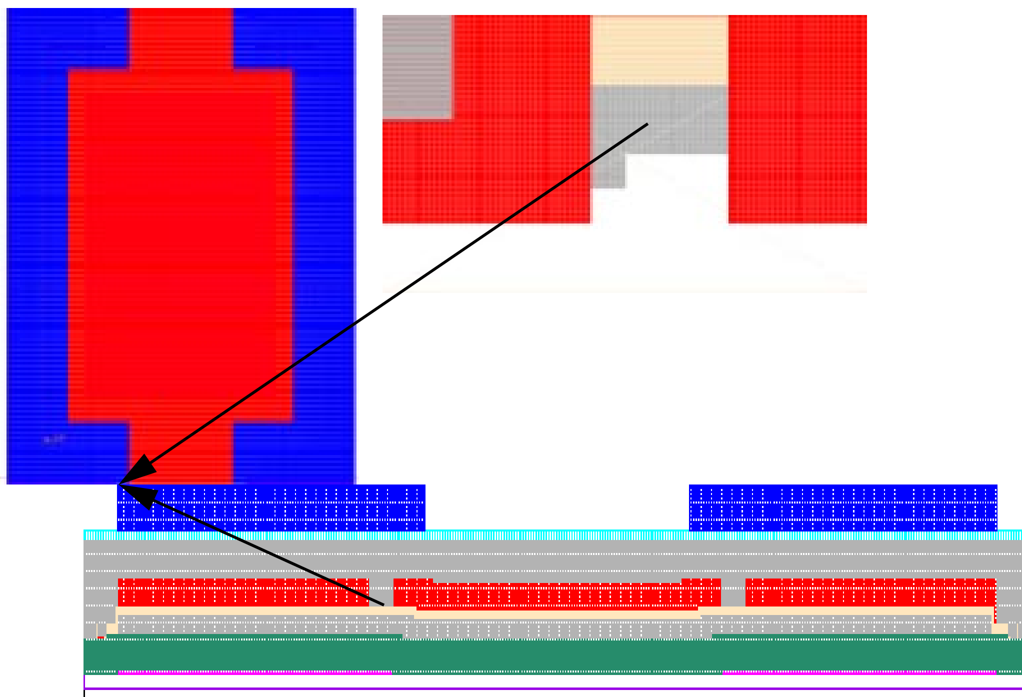

Figure 24 Pad Layout Error

The short is readily visible in an SEM once one knows what to look for. It was corrected for future runs. Also it was fixed for a device test in the current

run by FIB (Focused Ion Beam) so that device design could be improved for the next run using actual data as well as pictures.

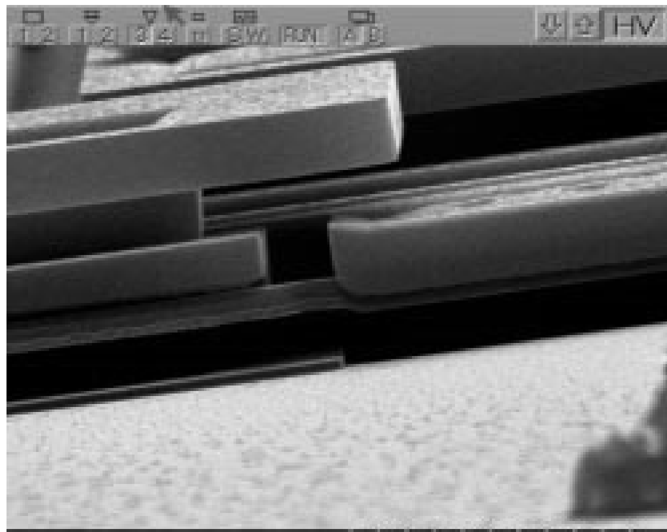

Figure 25 Built in pad short

6.4 New Metallization Test Structures

In addition to fixing pad shorts, adding new device types, and revising the metallization shielding a little, special structures were included in the next run to characterize the effectiveness of the metallization at reducing the resistance lines. Transmission lines with GSG pads were created, and a 802um long region of them was either uncovered (to allow full metallization), covered by 2um square holes with 2um edge-to-edge spacing (to test extended crossovers), covered with a similar grid of 4um holes with 4um edge-to-edge spacing (to show by comparison that the close spaced crossovers leverage the spread angle to reduce resistance further), and completely covered sections

that would be free of metal and allow measurement of the resistance of the MMPOLY lines.

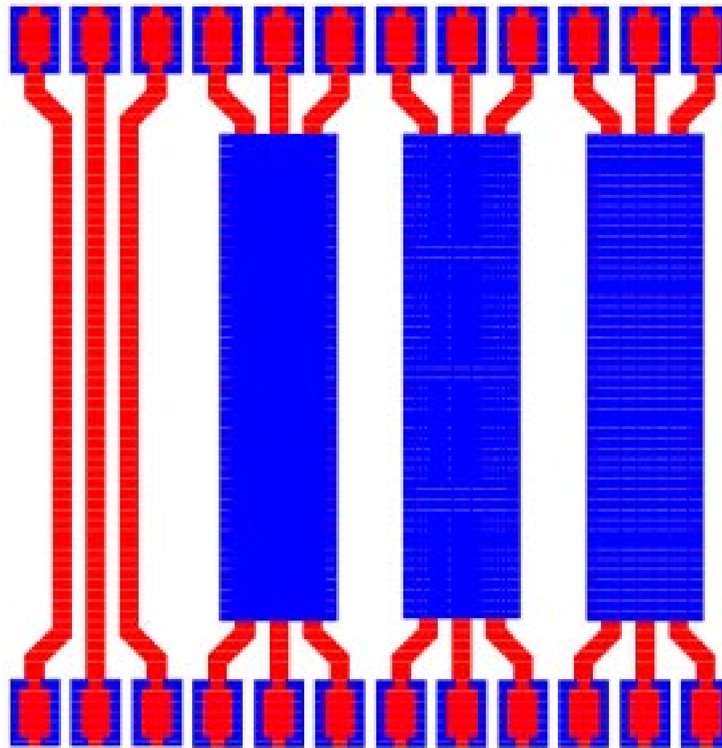

Figure 26 Metal Test Lines

6.5 Metallization(Sputtering): Second SUMMiT Run

The first metallization done on the chips from the second SUMMiT run was a deposition of 0.4um of gold by the same lab as did the metallization for the first run. Exact deposition parameters (i.e., temperature) were not recorded, but cleanroom personnel reported conditions were ‘standard’.

6.5.1 Metallization Test Structure Testing

The metallization test structures designed to test the crossover scheme were tested using an HP ohmmeter, and the HP 8510 network analyzer. Table 5 summarizes the DC resistance measurements. The convention here is that

Table 5 Resistances of Transmission Lines after Sputtering

(Ohms)	Uncovered	2um spacing	4um spacing	All Covered
“Horizontal”	3.1	62.5	49(partial short)	645
“Vertical”	2.9	68.4	247	539

“Horizontal” indicates the lines in the a corner of the chip, the “Vertical” structures are rotated so that the lines are at a right angle to the “Horizontal”. Table 5 reflects some deembedding by subtracting the 5.2 Ohms measured across a “short” on the test substrate, but does not reflect any deembedding for the pad contact resistance. Note that there was some difficulty landing on the “Horizontal” 4um spacing, and so the pad may have been bent to cause a partial short to the substrate. Another attempt at measurement using the model 10 probes determined that gentle use of probes is necessary to avoid pad damage leading to shorting. Resistance more in line with the “Vertical” measurement at 4um is indicated by these tests, but the measurements fluctuate significantly with minor table vibrations. A final important note is that later testing indicates partial contact often greatly increases resistance; further the “All Covered” case resistances are both much higher than the resistance of the polysilicon without metal should be, and similar to later measured bad con-

tacts. Based on the 9 Ohms per square resistance Sandia specifies for a MMPOLY1/MMPOLY2 stack such as the one used for the line, the resistance of the covered area should be on the order of 240 Ohms, much like the 4um case. Added contact resistance on the order of 200 Ohms per pad is not unusual for probing polysilicon pads with the setup used, so the additional resistance may be attributable to that.

Despite measuring to 0.1 Ohm precision in the lower resistance cases, the differences in resistance with multiple probe landings indicated insufficient accuracy to determine whether there was significant variation in the metallization angle with orientation. Part of the reason for this variation is discussed in Section 6.6.2. The s-parameter measurements told another story. In every case, the s12 and s21 for the “horizontal” case had much less loss than the “vertical” case- including the completely covered case, which should not be affected by metallization. This raises the possibility of some other phenomenon impacting the effectiveness of the transmission lines. A combination of the probing difficulties discussed in section 6.6.2 and some tilt in the chip mounting might be the culprit.

Another problem with the 4um case is that a layout error was discovered in the grid; there is one row of holes with a 5um instead of 4um gap. For the purposes of this test, that should not be very important, but for other tests (with wider metallization deposition angles where a continuous metallization is theoretically possible) this could cause a line break.

With all these problems however, the improvement of resistance with the use of the grid to allow the sputtering to pass through the upper layer is readily apparent, with the overall resistance much lower than the completely covered case, while much worse than the uncovered case. Due to the tight angle of the sputtering, the metal that passed through adjacent holes does not seem to have formed a continuous conductive path in either of the grid spacing cases; but the spread was sufficient so as to provide a significant benefit in the 2um case. Alternately, it is possible that the metal patches touch, but the areas of the contact are small enough to lead to the cumulative resistance measured. All in all, any confident verdict on the overall value of the metallization crossover should wait for the data collected later, when the metallization and probing practices were further refined.

6.5.2 "Bubbling" and the SAM

During the testing of devices, a problem occurred which manifested itself with the appearance of the surface of the chip bubbling. The bubbles occurred during device testing (device data was gathered, and is described in Chapter 7), but when the bubbles collided with a device it resulted in the device destruction. The bubbles seemed to form near ground pads of devices under test, and move randomly about afterwards. The conditions for bubble formation seemed to be voltage applied to the device, but no current draw. In fact, some devices drew current during some tests without bubble formation, and

later produced bubbles without current draw when higher voltages were applied later. Figure 27 shows a device destroyed by bubbling and surrounding bubbles.

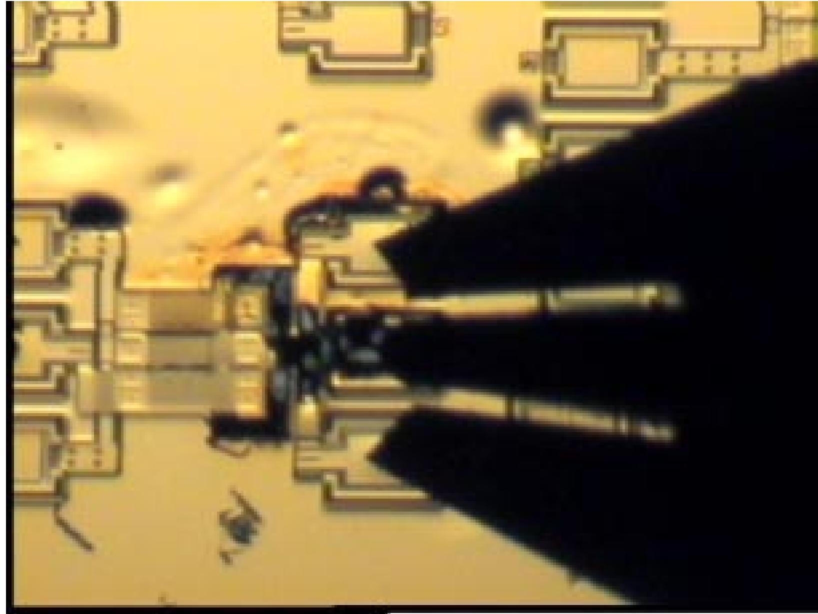

Figure 27 Bubbling Near Bubble Damaged Device

Several hypothesis were investigated. Due to the lack of current draw (the supply's internal meter for current limiting was supplemented by use of a more sensitive ammeter) accompanying the bubbling phenomenon, it seemed unlikely that the metal was actually boiling; and if actual boiling was occurring that energy must be stored somewhere on the chip during the earlier part of the testing, which was for some reason being released when the higher voltages were applied. Of course, localized heating from a short between the ground pad and the metallization on the nitride was the most likely culprit for the fact that the bubbles originated near the ground pads.

Inquiries to Sandia National Labs (the fabricators of the chip, and the ones who had done device release) as to whether there were additional layers or processing aspects that would account for this strange behavior were fruitful. During their release processing a Surface Adhesion Monolayer (SAM) of proprietary composition is added to reduce stiction. It is possible that this layer (which after metallization was trapped under the metal) was outgassing, which could lead to the visible phenomenon. If a voltage across the layer were a contributing factor to the outgassing, this would explain in part why the bubbling occurred when voltages were applied, rather earlier in testing when current was being drawn. That current was being drawn implies the existence of small shorts, that while they did not prevent device operation, did allow voltages (and perhaps some heating) to be transferred to the general ground plane. Sandia provided us with additional samples which had undergone a supercritical CO_2 release process without use of the SAM layer so we could investigate if this phenomenon recurred.

6.6 Metallization of chips from no-SAM release

The results of testing these chips motivated the switch from sputtering to evaporation. While the bubbling disappeared (presumably due to the lack of a SAM layer to outgas), other phenomenon indicating random shorts were observed.

6.6.1 Deposition Conditions

Chips were exposed to HF vapor to remove the native oxide before the deposition. 10mL of 49% HF (not diluted) was placed in a 50mL PTFE beaker, and all 6 chips were in series picked up with tweezers and held near the surface of the HF for 10 seconds. A color change was visible. Six samples were placed in the sputtering chamber with a variety of orientations. The chamber was pumped down to 10e-6 Torr, and 2mTorr of Ar was added. 0.3A was used for both targets. Each source was presputtered before using, then the chips were shifted into place. The planetary rotation was not used. The Cr deposition was for 60 seconds at 350 Volts for about 20nm. The Au deposition was for 46 seconds at 380 Volts for about 50nm.

Visual inspection after deposition did not reveal any errors; partial disassembly of some of the metallization test structures indicates a smaller spread angle than the designed maximum, though similar to the last metallization run. This implies that barring layout errors, large unwanted shorts should not have been introduced by the metallization. This does not preclude smaller shorts due to non-uniformity of spread angle by insufficient material to be visible.

6.6.2 Some Probing Abnormalities Explained

In the course of further testing, part of the reason for difficulties encountered thus far with the landing of the high frequency GSG probes was discovered. Figure 28 shows a close-up of the probe tips.

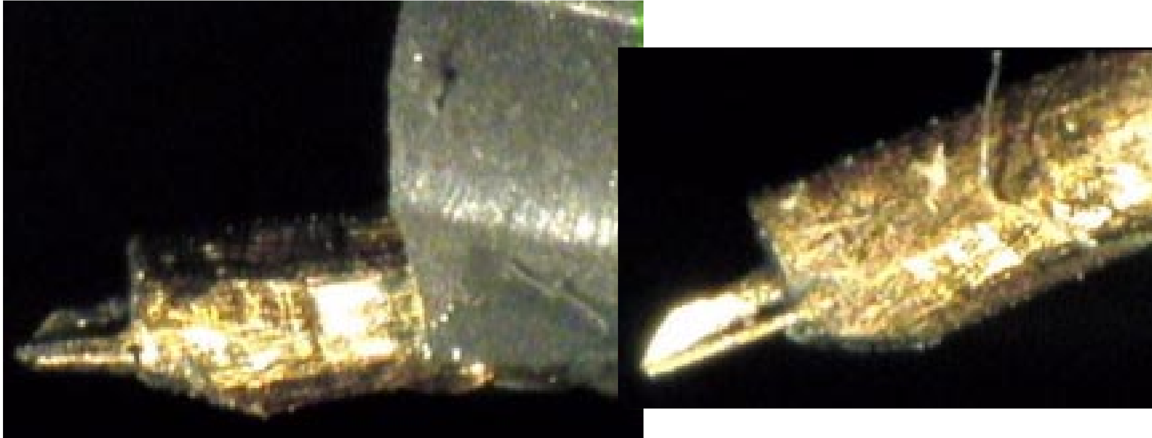

Figure 28 Probe Tip close-ups

The problem is that part of the probe hits the chip before the tip. As normally mounted, the thicker part attached to the tip hits the chip approximately when the tip does. This is not a problem for normal chips, which have a protective overglass layer. But it caused significant trouble on MEMS chips (whose surface is not even, since the glass has been removed). This explains both the damage to unrelated structures during testing, and the difficulty in getting a solid probe landing. We employed a couple shims on each probe mount to increase the angle, and avoid this problem. After adding the shims, probe landings became much easier and more reliable. It is necessary however to be careful not to introduce a tilt in another direction. This would cause the

three probes not to all come down at the same time, which also leads to unreliable probing.

6.6.3 Testing

These chips did not show as much resistance benefit as the metallization with planetary rotation; this is probably a function of the lower spread angle. The 2um, 4um and covered cases all had 247 Ohms of DC resistance, and similar s-parameters. No bubbling occurred, but probing for contacts between pads and the overall groundplane measured a number of shorts. Some devices were free of shorts, others had multiple pads shorted. The locations of shorts was not consistent from sample to sample. Attempting to recreate the bubbling effect by putting significant voltage across the shorts created significant local heating, which did affect the surface of the metallization. An apparent crinkling effect can be produced by the heating and cooling of the metal, which looks more like staining at normal probing magnifications.

Based on the continued existence of random shorts, it is concluded that some flaw in the metallization plan (probably the lack of directional uniformity of sputtering) exists. The next section describes how this problem was surmounted.

6.7 Evaporation

Due to the intermittent shorts after the sputtering deposition, a much more directional deposition technique was sought. E-beam evaporation has the potential for much more directionality. Sputtering is often used in situations where step coverage is important, taking advantage of the spread to achieve low resistances when a line changes height. Evaporation is preferred for lift-off, as it does not tend to cover such steps, so the metal on top of the photoresist is not strongly connected to the lines. One other possibility considered, but not attempted due to it's complexity, was to metallize by first using CVD to deposit oxide, then a blanket RIE to remove the portion of the oxide that is 'visible', then the sputtering deposition, and finally a normal release to remove the oxides and effectively lift off the excess metal. This would rely on RIE directionality.

6.7.1 Evaporator Geometry

The e-beam evaporator in the EGRC cleanroom does not make an incidental deposition due to its geometry. Items to be metallized are placed on one of four two inch square metal plates, facing downward. The plates are arranged around a circle. The source is located 10 inches below and 3 inches to the side of the center of the inner edge of each plate. Thus, material from the source

hitting the plates arrives at an angle between 16.7 and 26.6 degrees from the incident. At the center of the plate, the beam arrives at 21.8 degrees from the incident.

In order to partially compensate for this geometry, a 21.8 degree tilted ramp was fabricated in the center of a 2x2 inch steel plate. Placing the ramp so the sample faces inward allows metallization at the incident. Other ramps were also fabricated, as discussed in the next section.

6.7.2 Synthetic Spread Angle

As discussed in section 6.2.2, going through a hole of a given width in a given layer introduces a cutoff angle, for which any beam at greater angle from the incident will not make it through the hole. Of course, for metallization that all travels in a parallel path the metallization ceases to provide much practical benefit from the metal passing through the gap well before the cutoff angle, because most of the metal winds up on the sidewall of the slit rather than on the destination surface. This is one of the major challenges of changing from a sputtering to evaporation deposition. While sputtering is not highly directional, so that a certain amount of spread can be used to advantage, e-beam evaporation is highly directional. Very little spread seems to occur. In order to make a connected metal path on the lower metal level in the crossover scheme, it is necessary to create an artificial spread. One conceptually simple approach would be to tilt the sample through the range of desired

spread angles during the metallization process. This would have the advantage of creating a metal deposition pattern much closer to the original plan than what was finally done. This approach, while suitable for volume production, requires a lot of work for preparing a few samples. Creating a platform that will provide the desired tilt inside the e-beam chamber would be a complex task, probably involving clockwork to avoid the contamination difficulties of using some other form of energy source inside the chamber such as a battery.

A much simpler approach is to deposit some metal, cease deposition, manually adjust the incident angle of the sample to the source, deposit some more metal, and repeat until the desired amount of spread has been achieved. This is very expensive in lab time, as each adjustment requires the evaporator be opened, and pumped back to vacuum afterwards. Nevertheless, to test feasibility with a small number of samples it is preferred to creating a complex mounting mechanism. All wiring crossings were aligned with manhattan directions, so approximating a spread with only tilts in the manhattan directions is assumed sufficient for these tests.

The following deposition scenarios were developed to provide a continuous metal on the lower layer for the two grids fabricated. Note that here 'continuous metal' is used to mean only that adjacent patches connect, not that there are no areas without metal. These scenarios do not directly connect patches diagonally. To facilitate these experiments, mounts for the evaporator have

been machined from steel. The ramps were designed to provide a net 35 degree and net 45 degree angle of deposition. On these two ramps, a circular platform rotatable through use of a setscrew was fabricated. By reversing the direction of the ramps, one could also introduce additional angles. The complete set of producible angles using all three ramps and a flat plate in conjunction with the inherent 21.8 degree incident angle are 0, 1.4, 8.6, 21.8, 35, 43.6, and 45 degrees. Unlike the sputtering case, these are not general spread angles, but the angle of a beam from the incident.

For the 2um grid with 2um spacing: Use a flat plate (no ramp) from each direction to achieve 21.8 degrees, and the 13.2 ramp to generate 8.6 degrees in order to get a continuous metallization. This means a total of 8 depositions, 4 per manhattan direction to create a continuous field of metal. In order to deal with crossovers in only one direction, only 4 depositions would be needed. Figure 29 illustrates this technique showing the intended paths of two depositions.

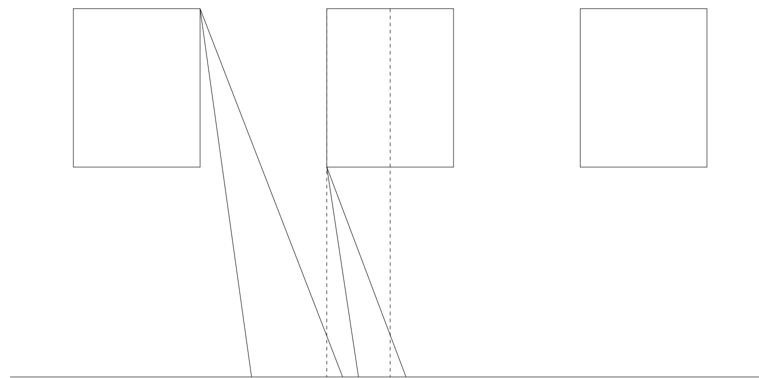

Figure 29 Evaporation Through 2um Grid Line Below

For the 4um grid with 4um spacing: A 35 degree deposition from each direction, then a single direct deposition using the normalizing ramp should be sufficient to get a continuous metallization. This means a total of 5 depositions for complete metallization, 3 to deal with one direction.

6.7.3 Initial Sanity Check

In order to verify that the evaporation method was highly directional and no shorts were created, a test metallization using the ramp designed to compensate for the inherent tilt was done. This deposition involved a few samples included with someone else's tests, so the HF wave was not done; the deposition thickness is approximately 180nm of Au.

Many pads and structures were tested for shorts, and structures were manually disassembled to visually check the spread through grids. By all measures the metallization was highly directional. A significant interface resistance was present between the metal and polysilicon; this is attributed to the lack of HF wave (the native oxide would introduce such an interface resistance). Usually, before metallization, the chips are waved over HF for a time, so that the vapor from the HF strips away the native oxide that builds up on the polysilicon. If the native oxide is not removed, it creates such an interface resistance between the polysilicon and metal.

6.7.4 First Four Depositions

Next, a total of four Cr/Au depositions were done in order to apply the recipe described in Section 6.7.2 to one direction on the metallization test structures. Each deposition involved 5nm Cr for adhesion followed by 70nm Au. Before the first deposition, the chips were held over HF for 10 seconds for removal of oxide buildup.

Both recipes were implemented, each using both SAM coated, and non-SAM coated chips. Also, one chip was metallized at a 45 degree angle from all four manhattan directions, and another similarly at 21.8 degrees. These depositions were to verify the effectiveness of the self-masking at preventing shorts, rather than to make a continuous metal covering.

Testing of these chips indicated that a continuous metal covering had not been achieved. The resistance of the lines with metallization grids was similar to that of the completely covered lines. This indicates also that there is sufficient interface resistance between the metal and polysilicon that the signal does not see a lower (or even similar) resistance path passing through the interface to the metal, traveling along the metal for a few microns, and returning to the polysilicon than simply traveling through the polysilicon.

Manually pealing back the metallization grids, and examining the surfaces using an SEM verifies that the metallization coating was not complete. Figure 30 shows both recipes, as well as the failure to achieve good overlap.

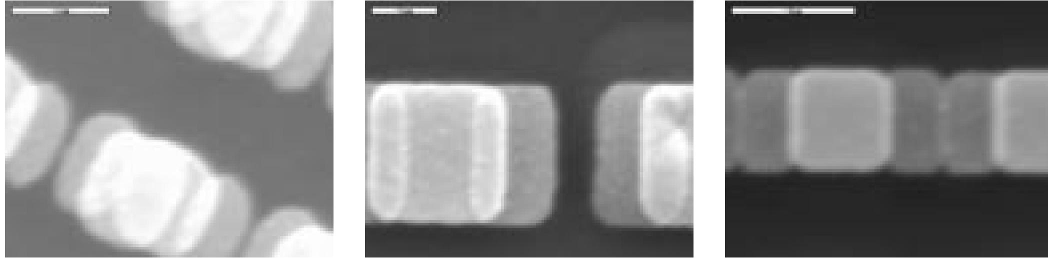

Figure 30 Evaporation 1 (2um recipe, 4um recipe on line, nitride)

On the leftmost SEM, we see the bright areas where the metal was deposited through each hole overlap each other; but there is no overlap between the patches of adjacent holes. Similarly with the 4um recipe, we see a large bright patch where two depositions almost completely overlap, and the sections to the sides from angled depositions, but these do not overlap with adjacent regions. Finally, the rightmost image shows that at the nitride, the adjacent patches do overlap, although their overlap area is not large, providing a continuous metal from patch to patch.

Despite not succeeding in providing full overlap on the middle polysilicon layers, this run of metallization achieved several things. Probing was much easier than past metallization, and as described in Section 7.7.3 device operation was much improved as a result. One major factor in the metallization not connecting was that as the thickness of the oxide between MMPOLY2 and

MMPOLY3 is not exactly known before fabrication, the recipe assumed a thickness greater than it turned out to be. The SUMMiT process uses a planarization step on this oxide, so it's thickness varies. Thus, the metallization did not spread out as much as anticipated before landing on MMPOLY2. In future runs, a metallization grid with larger holes and thinner MMPOLY3 lines would be utilized to compensate. At the time of the run, the 2um grid was minimally sized, but now MMPOLY3 of only 1um width is legal. If the holes were expanded to 3um per side and the center to center spacing was held constant, the recipe used for the 2um case should work well.

6.7.5 Another Evaporation Run

It seemed that it should be barely possible to fill in these gaps using additional depositions. Using a 35 degree deposition from each direction should fix the 2um deposition; similarly a 45 degree deposition should fix the 4um deposition. These deposition angles are close to the maximum possible angles to make it through each grid, so little additional material will make it to the lines, but based on line tracings using projected cross sections it should be enough to fill in the gaps. Figure 31 illustrates some of the various cumulative depositions through the 2um grid. The solid paths represent depositions from the original deposition. The horizontal dotted line indicates the actual level of the line below relative to the grid computed from spacings on the SEMs. The

slanted dotted lines indicate one of the new depositions, intended to fill the gap left from the original depositions.

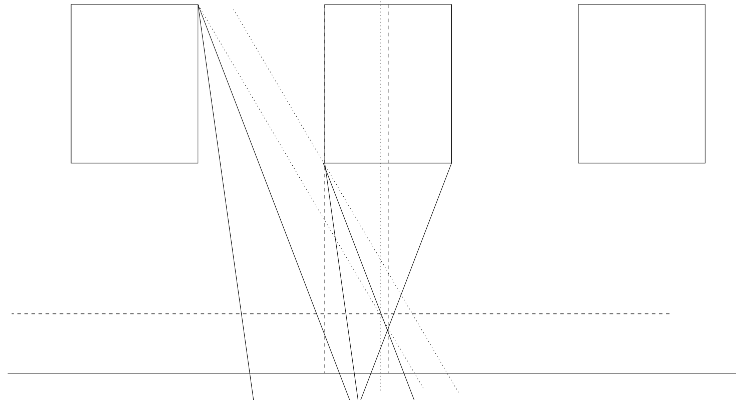

Figure 31 2um Grid Additional Evaporation

This was attempted in the cleanroom. The same procedure for evaporation as before was used, except for the different angles necessary, and the use of Ti rather than Cr for the adhesion layer. 70nm of Au, and 5nm of the adhesion layer were again used for each deposition. The three remaining samples of each recipe had the additional metal applied. Unfortunately, when the resistance of the transmission lines was measured, they did not show the marked improvement that would be expected had the metallization become continuous.

Chapter 7 Devices Using SUMMiT

7.1 SUMMiT Process Features Relative to MUMPs

The move to the SUMMiT process was due to various shortcomings of the MUMPs process that impact yield and the achievable capacitance ratio, such as insufficient flatness. The SUMMiT layer spacings should also allow bigger ‘off’ gaps, which improves the SNR ratio, a key factor in designing these devices. The greater flatness of the SUMMiT devices should allow a better ‘on’, also improving the SNR. Since MCNC is unable to provide the process changes we wanted (in particular a 2nd metal) the only remaining advantage MUMPs has is the 1cm/side rather than 0.5cm/side chips. Because of this size difference, it is not possible to fit as many devices of a given size on a SUMMiT chip.

The SUMMiT process lacks a metal layer, but does have lower resistance routing in polysilicon, and gold can be added as a post-processing step. This postprocessing is discussed in Chapter 6. Polysilicon resistances in SUMMiT are generally lower than in MUMPs, at 9 Ohms/square for MMPOLY3 and also for a MMPOLY1/MMPOLY2 stack.

7.2 Initial Planning

Based on experiments in the MUMPs process, the most stable design utilizing two springs (2 springs preferred due to vast area savings relative to more springs) are those with springs connecting to opposite corners of the device, the signal plate in the middle, and actuation strips on the outside. Employing as much symmetry as possible in both spring and device arrangement also seems beneficial. Mechanical stability seems to be best served when the device has width:height ratios in the range of 0.5 to 2 when utilizing this basic design. Springs perform much more consistently when flat (nothing running under them in lower levels causing ripples in them).

In the first SUMMiT run, four micron wide springs are used to minimize the variance in spring constant due to process variation. While the SUMMiT process will hopefully provide more uniform springs anyway, the 2um springs used in the MUMPs devices varied too much and the number of runs to try things in SUMMiT is much less than MUMPs.

In order to maximize the ON:OFF capacitance ratio for the bistable switch, it is necessary to create the largest possible vertical throw for the moving plate without compromising flatness. The two possible choices in the SUMMiT process would be a MMPOLY1/MMPOLY3 gap with a MMPOLY1 plate moving upward, or the MMPOLY3/MMPOLY0 gap with the plate moving downward. Due to the need to shape the plate so that the signal area lands without danger of the actuation regions shorting, the moving region must be con-

trolled. Since MMPOLY3 is planarized, the way to shape it vertically is by using a dimple added after planarization. Due to the planarization, the depth of the DIMPLE3 (constructed by etching oxide, then back-filling the void) may vary sufficiently to make the difference in actuation and signal area heights vary enough to cause control problems. As discussed in section 3.2.3, the gap height is a critical design parameter. Also, since MMPOLY1 is much thinner than MMPOLY3 and can be anchored directly, as it is more suitable for building spring structures. The substrate effects discussed in Section 3.3 should be less if the device is further from the substrate when actuated. It was therefore decided to use a MMPOLY1 plate moving upward to MMPOLY3.

7.3 Rule Adjustments

The SUMMiT process is usually used for creating large gears and related mechanical structures. The design rules were not developed with a focus on vertical motion. In order to achieve maximum vertical throw and allow proper device operation, design rules relating to the use of DIMPLE1_CUT and its effect on MMPOLY1 and MMPOLY2 were updated to allow more design freedom while still avoiding the problems the original rules were designed to avoid.

There are two ways to shape MMPOLY1 vertically (without endangering flatness as etching the nitride would): DIMPLE1, and MMPOLY0. The thickness of MMPOLY0 is only 0.3um, which allows very little space for post-pro-

cess metallization and plate tilting without creating a short. Therefore, while some devices were made which utilized this method, it is the 1.5-um DIMPLE1 which has the best potential for favorable results. The design rules as written seem to prevent the use of DIMPLE1 for this purpose, however. Two rules present problems: ERR_D1C_W/O_P2 and ERR_D1C_W/P1C. ERR_D1C_W/O_P2 exists as part of the effort to prevent formation of stringers in MMPOLY2 due to DIMPLE1. By putting a solid sheet of MMPOLY2 over all instances of DIMPLE1, it effectively removes the possibility of MMPOLY2 and MMPOLY1 stringers in these regions. It also prevents the use of MMPOLY1 for making a plate that can move up to MMPOLY3, as the MMPOLY2 would block this motion. Examining the stacking layers and the MMPOLY2 etch depth, it appears that this rule is not always necessary. If OXIDE2 is left there to prevent the merging of MMPOLY1 and MMPOLY2, the MMPOLY2 etch, which is designed to cut through a MMPOLY1 and MMPOLY2 stack, should be sufficient to safely remove all of the MMPOLY2 over a DIMPLE1 edge, as long as other low level features (such as nitride or MMPOLY0 edges) are kept far enough from this edge to avoid further increasing the height of deposited MMPOLY2. At such an edge, the thickness of MMPOLY2 would be the normal 1.5um + 1.5um, or 3.0um from the dimple, while the etch of MMPOLY2 including the 25% overetch is equal to 1.5um for etching the MMPOLY2, plus 1.0um to etch the polysilicon, multiplied by 1.25, or 3.125um. The deposition of MMPOLY1 and Sacox2 after the dimple tends

to smooth sharp corners. The new rule used in this case pertains to using DIMPLE1 without an overlying MMPOLY2 layer. In this region, there can be no cuts in MMPOLY1; MMPOLY0 must either be completely removed or be a continuous sheet; and there can be no nitride etch.

The purpose of ERR_D1C_W/P1C is also to prevent stringers in MMPOLY1 and MMPOLY2, and if strictly adhered to would also prevent the desired layer combination. The DIMPLE1 regions are the ones farthest from MMPOLY3, and thus must be used for actuation areas. The actuation areas in MMPOLY3 must be anchored, and consequently must cross the edge of the plate at least once. At this location, the plate must be dimpled in order to avoid shorting. Hence it must be possible to dimple the edge of the plate. This cannot be done without cutting MMPOLY1 somewhere in a dimple. Breaking this rule cannot be done simultaneously with violating the ERR_D1C_W/O_P2 rule without the possibility of making MMPOLY2 too thick to be removed by the MMPOLY2 etch. A solution is to extend the dimple well beyond the moving platform, then cover the edge of the dimple with MMPOLY1. In this way, (despite the existence of regions where there is a MMPOLY1 cut where the is a dimple) there is no region where the edges of the dimple are uncovered by MMPOLY1, and can therefore cause stringers. Since in this border area the MMPOLY1 will not be moving, there is little reason not to additionally overlap this dimple edge with MMPOLY2, which also prevents MMPOLY2 stringers in this area. The new rule used in this case applies to cutting MMPOLY1

within a dimpled region, e.g., the edges of the dimple must still be covered by MMPOLY1 to prevent stringers.

7.4 Some Basic Design Trade-offs

Maximizing the ON:OFF capacitance ratio is done by getting as large a throw as possible, with flat plates. This means selecting layers that are as flat as possible and well separated. The rule changes already discussed aside, there is some trade-off involved in deciding how close the limit of various design rules to go in order to maintain flatness versus the area cost. Also, residual stresses versus area cost a is general issue to keep in mind. For example, the placement and frequency of anchors has great potential to interact with the effects of any residual stresses.

There are two basic spring design trade-offs. One is that the stiffer the spring is, the higher the fundamental frequency of the device. Thus, faster mechanical operating speeds are possible with stiffer springs. On the other hand, stiffer springs require a combination of higher voltages and greater area devoted to actuation to activate. Note that for reasonable choices of springs the fundamental frequencies (kHz) are well below the signaling frequencies, so there is no danger of signaling producing mechanical resonance.

The other spring trade-off regards spring shape and resistance. This is particularly important for devices where the signal must travel through the springs. A wider spring needs to be longer in order to avoid increasing stiff-

ness, so while the spring resistance can be decreased by making the spring wider, the spring must also be lengthened. While the total resistance still decreases since the length only needs be increased by a factor of sqrt(3) for a doubled width, the device size begins to increase greatly after a certain point, which can also increase the overall resistance.

7.5 Device Design

While with the MUMPs process, the moving portion of the device moved down (toward the substrate), the devices designed for the SUMMiT process all use MMPOLY1 as the spring layer, and move upward (away from the substrate) to plates made with MMPOLY3. This has some important effects on device characteristics. The substrate now has less of a pull on the device when actuated, rather than more. Off devices are still hanging in the air, but can now potentially be pulled downward as well as upward (using the substrate or a MMPOLY0 plate used to apply a potential). There are some drawbacks with upward actuation as well. Visibility is severely decreased by the MMPOLY3 overhang, which makes visually diagnosing problems with the design much more difficult. The MMPOLY3 must be firmly attached, and stress in the MMPOLY3 must not be sufficient to warp it as this would cause trouble with device operation.

One neat technique possible is that moving upside down may allow is to use the substrate to pull the device away from the landing area when it is ‘off’,

increasing the available capacitance ratio and making device mis-actuation due to resonance vibration of the device much less likely. Since there is more force exerted by the substrate on an 'off' device than an 'on' device, it is possible that after all the desired 'on' devices are set the remaining devices can be pulled to this more-off position. Some devices were made to test this. Most devices employed a MMPOLY0 plate to set the MMPOLY0 voltage to the same level as the device above it. In this way we completely isolate the device from substrate effects, removing the floating substrate problem that plagued early attempts at consistent device operation using the MUMPs process. Since the substrate is all connected, such a shielding method is not possible using the substrate, as different devices will be at different voltage levels.

Pisano's "5 of Hearts" anchor[49] was modified to utilize circular instead of square vias to reduce total anchor size, still providing a stable low stress anchor for wires, and signal and actuation regions. The anchor rests on a platform created by using MMPOLY1 and MMPOLY2 to trap the lower oxides. Some of the techniques for beam stepups discussed in Chapter 2 were used to maintain planar flexure beams for the springs in one family of devices.

Since the springs are made out of the thinnest available polysilicon layer, they are highly resistive. Despite the fact that simulations indicate we can still send a signal through them and the capacitive pad, in some device designs there are two moving capacitors. In this way, the signal goes through a path more purely capacitive, and the RC constant seen by the driver may be

smaller, depending on the resistance of the wiring. Of course, for a given area devoted to signal plate, only 1/4 the capacitance is available if it is split in half to create two capacitors in series that the signal must travel through.

7.5.1 Design to Survive metallization

In order to protect the springs from gold landing on them and modifying the spring constant, a covering of MMPOLY3 which shadows the spring from the gold is used. There are several nuances to this part of the layout. One is that it is not known during layout that the device will be in it's central 'off' state during the sputtering. Stray charges could actuate the device, or move it closer to the substrate, and the design must work in either case. This means that in no place can MMPOLY1 and MMPOLY3 touch at one of the gaps in the MMPOLY3, otherwise the metal may glue them together. Only the dimpled areas or holes in the MMPOLY1 can be under the openings in MMPOLY3 because of this problem. On the other hand it is possible that the device will be attracted to the substrate. Since getting gold on portion of the signal plate would lead to the rest of the plate not getting as close to the MMPOLY3 during actuation, the dimple therefore needs to extend far enough to catch all the metal which is deposited. Currently a 6.0um extension is being used, which effectively adds directly to the width of the device on each side of the signal plate.

7.6 The First SUMMiT Run

Unlike when preparing for the first MUMPs run, there was no designer experienced with the processes locally to check designs against. A certain amount of additional conservatism in local layout was employed, despite the modifications to some design rules. Probe pads, transmission lines, substrate contact pads, and other components were all made from scratch. Designs were done using Cadence, and then converted into the Autocad format required for submission. One advantage to this was that because the local DRC and the DRC available by ftp using Sandia's design kit were both in use, deficiencies of each were often compensated by the other. Two chips were prepared for this run; an electrical chip, and a primarily optical chip with some electrical devices as well. The first run actually occurred twice, as a mask had been left out in the first fabrication and the chip was rerun; in the rerun, some refinements were included.

7.6.1 Devices Included

Two basic device families were included in the first SUMMiT run. One family utilized the MMPOLY0 layer in order to create the signal/actuation height gap, resulting in a 0.3um gap. The other used the 1.5um DIMPLE1 cut to create the gap, utilizing the design rule modifications discussed in Section 7.3.

The 1.5um gap design utilized a continuous sheet of MMPOLY0 electrically connected to the moving plate to decrease the influence of the substrate on

device operation. This device's topology is shown in Figure 32. The actuation control surfaces are the two thin MMPOLY3 lines, with the central area providing the landing pad and signal pad area. The outer lines provide covering for the springs, protecting them from metallization. The frame around the device is necessary to meet the new design rules for the dimple cut, which covers all the area within the border, except a square in the center of the device for the signal area. The springs are connected to the border. The use of a single signal pad requires the signal to travel through the springs to reach the spring; the use of 4um wide springs helps some with the resistance, but this design still has a significant series resistance with the capacitor.

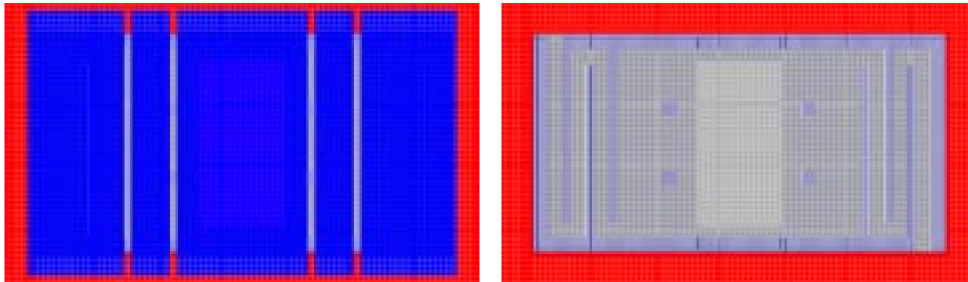

Figure 32 JSM Device Family (full layout on left, without P3 on right)

The device topology utilizing the 0.3um gap is shown in Figure 33. In this case, there are two signal lines on the outside (also serving as spring shields) while the actuation area is in the middle. The springs are anchored to MMPOLY0. This device can be hooked up in two configurations. Either the signal can come in to the plate through the springs, and leave through both signal areas, or the signal can enter through one signal area, travel through

the plate, and leave through the other plate. This provides a lower effective capacitance, but also provides a much lower resistance path through the device. Which configuration is superior for using the device in an array would depend on other characteristics such as the resistance of the wiring to the device.

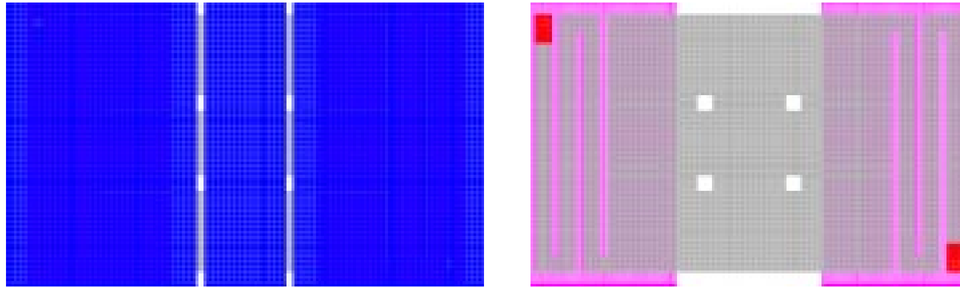

Figure 33 DubPo Device Family (full layout on left, w/o P3 on right)

An example array illustrating anchoring and local wires, utilizing a device of the DubPo topology, is shown in Figure 34. The horizontally traveling signals are routed on a MMPOLY1/MMPOLY2 stack, and are connected to the right signal plate of each device. The vertically routed signals travel down the left side. Five point anchors are used for all lines in MMPOLY3 to minimize stress; these anchors use round rather than square anchor points in order to

minimize anchor size. Where the horizontal and vertical routing cross, the 2um grid described in Chapter 6 is used.

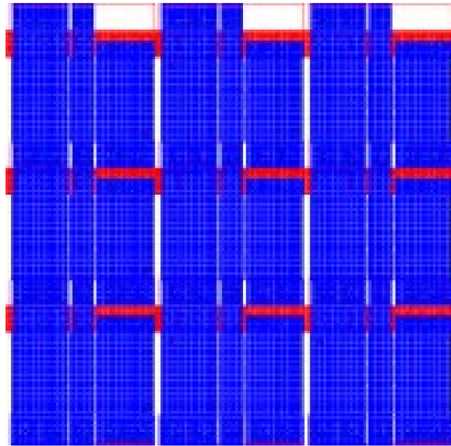

Figure 34 3*3 Array Segment using DubPo devices

7.6.2 Test Results

Due to the metallization problems, it was not possible to test a large number of devices. Once the problem with the pads was found, the pads around a single device were FIB'd in order to remove the shorts, and the device was tested. Figure 35 shows the S12 magnitude results measured for this device.

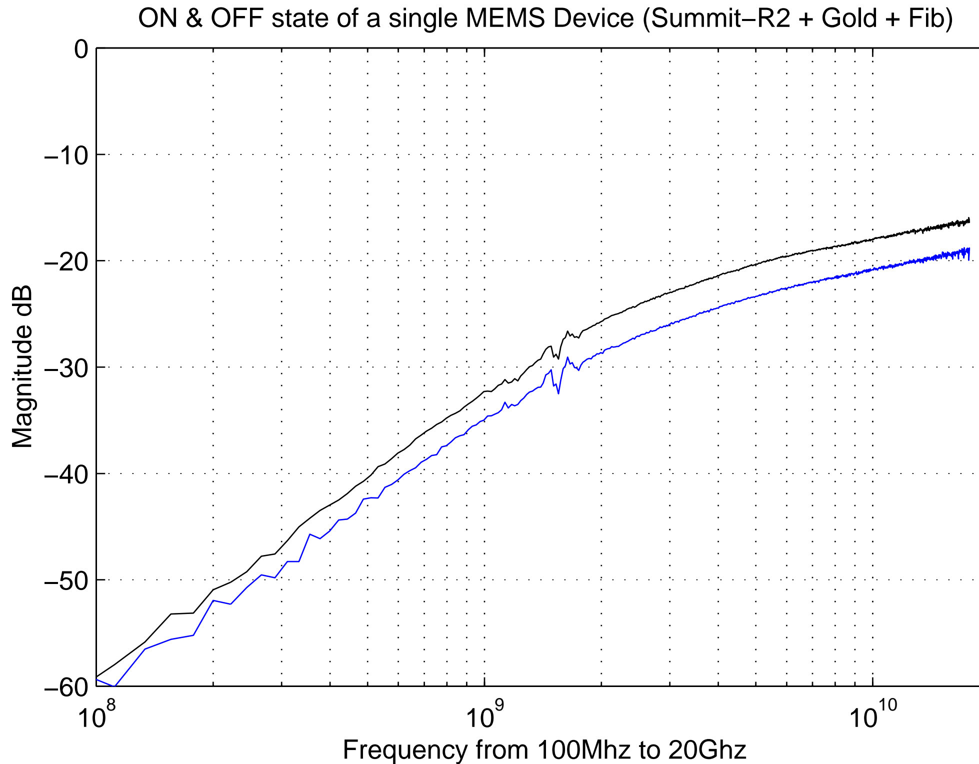

Figure 35 FIB'd Device First SUMMiT Run S12

It is possible that this device was not fully actuating; indeed based on later measurements this seems likely. The gap between the two curves varies between 0.5dB and 1dB from 200MHz through 20GHz. The primary value in this measurement is that whether it is an on or off curve, it is at least more

consistent with the series capacitor dominating than the MUMPs devices. Significant improvements to metallization and test setup were necessary to achieve further results.

7.7 Second SUMMiT Run

The second SUMMiT run had the main purpose of fixing difficulties with metallization, but some of what was learned from the first run was also incorporated, and a new device family was introduced. In addition, a more compact device test cell was introduced to minimize wiring, and make device testing easier. Finally, metallization test structures, discussed in Chapter 6, were included.

7.7.1 Devices Included

Most of the devices from the first run were included again, but an additional device type was included. In order to increase total capacitive area and decrease the spring constant, the moving plate was made using MMPOLY2, with the springs in MMPOLY1 below it. The success of the rules modifications used to create the dimpled devices led to a higher use of that family of devices, with variations utilizing less wide springs to allow lower actuation voltages. The dimple rule modifications are used in the new MMPOLY2 plate devices. While these devices have the disadvantage of less distance for the plate to

move (and thus should exhibit a lower ON:OFF ratio), locating the springs below the plate makes shielding them from metallization simpler, and allows more efficient use of area and larger plate sizes.

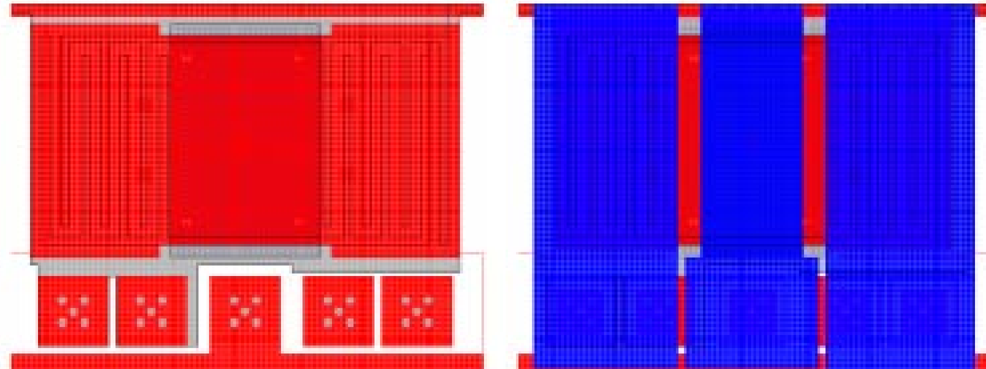

Figure 36 Poly2 Plate Device (with and without top panels)

7.7.2 Test Results: Sputtering

A pre-metallization visual inspection looks like everything was processed as expected. Using model 10 probes, a range of 0-86V was applied across the actuation gap of one of the new devices. The device was not visibly damaged, but it was impossible to tell visually whether actuation occurred.

In order to test for actuation, the 8510 was used to measure the s-parameters of the devices while voltages were applied. Many devices were tested from one of the samples that had 0.4um sputtered. Due to troubles with the metallization, many devices could not be tested. The s-parameters shown in Figure 37 are for a device with sufficiently low current draw that continuing with measurements seemed safe. Device operation continued during the bubbling discussed in Chapter 6 until a large bubble shifted into the device and caused

it's destruction. Due to the odd behavior, this same device were tested on an unmetallized sample for comparison. The s-parameters for it are shown on the left side of the figure. Despite the excellent 'on' to 'off' ratio at lower frequencies of the sputtered version, the metallization problems make the usefulness of this result, and the others before evaporation, questionable.

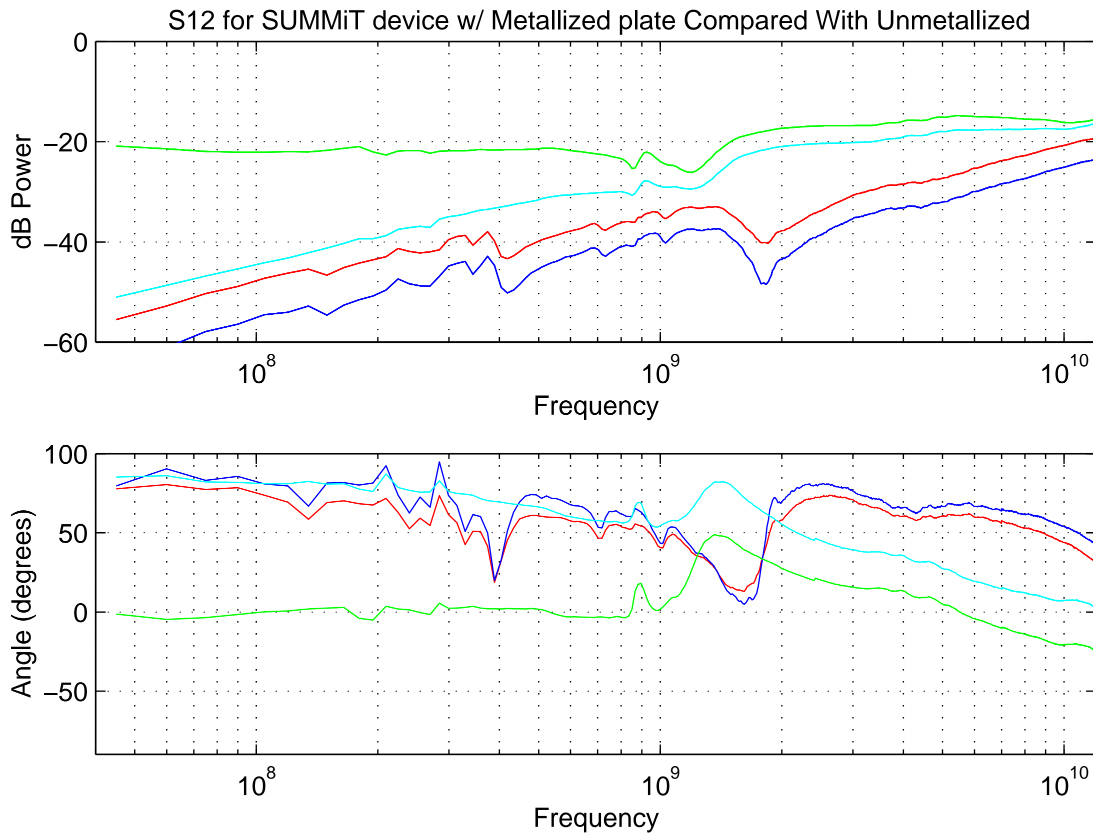

Figure 37 Comparison of Metallized(Sputtering) and Unmetallized

The green curve is the 'on' metallized, cyan is 'off' metallized, blue is 'on' unmetallized, and green is 'off' metallized.

7.7.3 Test Results: Evaporation

The new device described in section 7.7.1 was tested first as the most conservative device mechanically. After some initial ramping up and down (presumably to get the device moving, if it started slightly stuck down), the device exhibited very repeatable and consistent actuation. The device did not actuate before 18.5V and always actuated by 19V, even if very large voltage change steps were used. Release occurred at 16V for a step down from 19V, 16.2V if stepping down by 0.1V, providing a couple volts control window. This small control window is expected due to the relatively small available gap. The s-parameters are particularly smooth for this device, exhibiting classical high-pass behavior. The lowest “ON” curve and highest “OFF” curve have a separation ranging from 3dB (at 12Gz) to 5dB (by 500MHz). While the total loss is high in either case, it should be noted that the testing system is 50 Ohms, which is a severe mismatch relative to the device being tested.

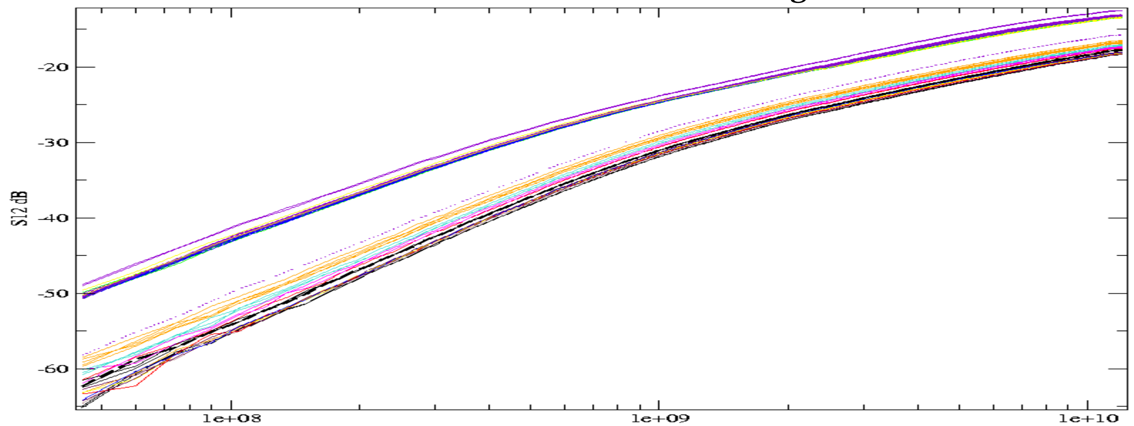

Figure 38 Magnitude (dB) of S21 for SUMMiT Device (many voltages)

Comparing the performance of the evaporation treated and unmetallized versions of one of the devices in this family, the metallized version exhibits lower loss, and a smoother frequency response, than the devices without metal. This is despite the incomplete crossover coverage of the evaporation for the device tested, the variant of the new device using 2um wide springs and the two micron squares version of the crossover grid. Figure 39 show the metallized device measurements in green 'on', and cyan 'off'; the unmetallized have blue 'on' and red 'off'.

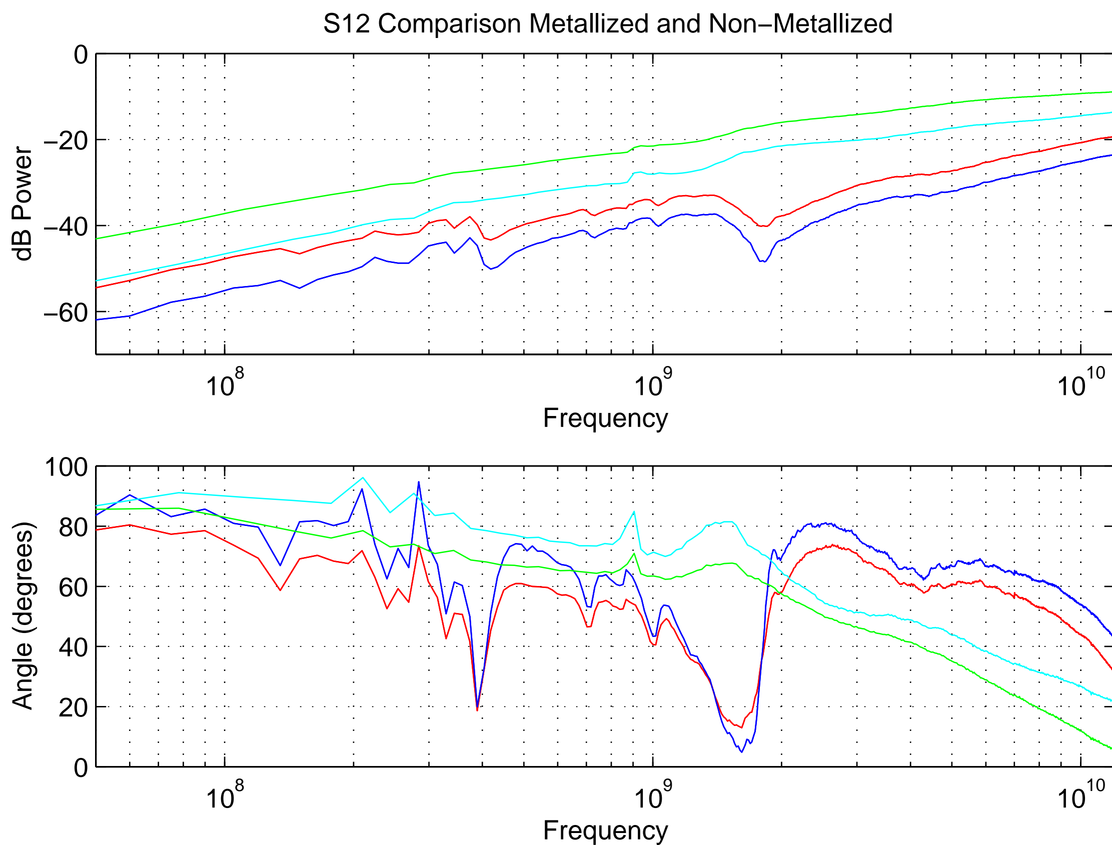

Figure 39 Comparison of Evaporation and Non-Metallized S12

7.8 Modeling

The same techniques were used to model SUMMiT devices as MUMPs devices. Since the MATLAB models provided the most complete potential predictive ability despite the flat-plate assumption, these were used most extensively in device design. Since MMPOLY0 shields to isolate the devices from the substrate were used in many devices, the substrate effect was much less important than for the MUMPs devices. While the models predicted values too high for the MUMPs devices, they predicted values too low for SUMMiT. For example, the actuation voltage for the device shown in figure 39 is predicted by the Matlab models to be on the order of 7V, less than half the actual actuation voltage. Some possible factors which might explain part of this include stiffer polysilicon in the SUMMiT process, stiffer springs due to the larger number of bends utilized, and variations in expected layer thickness (due to the planarization of OXIDE3 for example).

7.9 Capacitance Ratio

A rough fit using the a simplified topology model of the one used in section 3.7 was done using the data presented in figure 38, using the curves for the lowest 'on' capacitance and highest 'off' capacitance. Based on this fit, the

through capacitance of this device in the ‘off’ state is on the order of 60fF, and the ‘on’ capacitance is on the order of 240fF. This represents an ‘on’ to ‘off’ ratio on the order of 4. The model used is shown in figure 40.

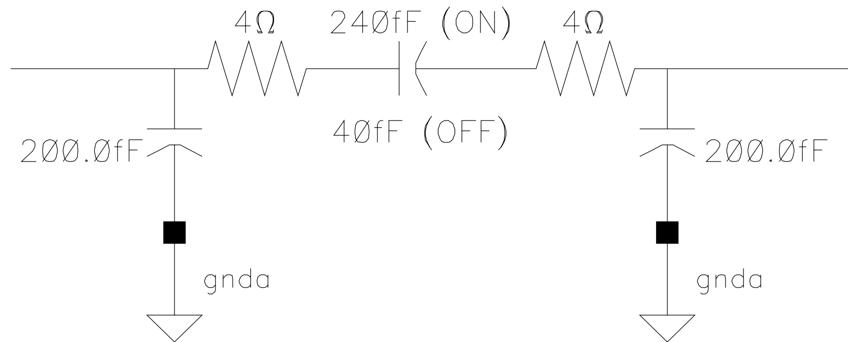

Figure 40 SUMMiT Model Fit (single device)

The reason a simpler model was possible is due to improvements in the interconnect and return paths. Note that this model is of both the device and the test structure around it. As such, the shunt capacitances include a lot of capacitance added by the test structure. In the case of a large array of devices, a portion of this shunt capacitance will be added by each device, but most of it will remain on the ends of the arrays. Depending on packaging technique, this portion may be replaced by other components.

If the model is fit using reasonable ‘hold’ voltages for when programming is not occurring a ratio of 5:1 is achievable.

These ratios should be enough for the switch to work as a single isolated switch, but is insufficient to form part of a crossbar of non-trivial size. They

may also suitable for building a phase shifter; decreasing the through resistance by improving metallization would help with that application.

Chapter 8 Conclusions & Future Work

8.1 Achievements

Several of the key prerequisites to building a large MEMS based crossbar have been met. A novel bistability and addressablity scheme has been demonstrated, and the effect of the substrate bias on controllability has been explored. For an n*n array of devices, this allows the number of control lines to be proportional to n, rather than n-squared, greatly increasing the utility of large arrays of MEMS devices. To facilitate control of such arrays, a high voltage control chip was produced. A promising method for post-release metallization of thin film MEMS processes has been adapted from the methods introduced by [50], and a novel method to decrease crossover penalties has been introduced. The space of possible Clos implementations was explored and the approach of optimizing array sizes for arrays utilizing CMOS and MEMS based crossbars for different stages rather than the traditional method of minimizing crosspoints was proposed. Design rule changes to improve the flexibility of the SUMMiT process were proposed and tested. A crosspoint with significantly improved highpass characteristics was developed, which may be suitable for broadband switching. Estimated overall power consumption for a 192x192 array using 32x32 MEMS crossbars (if sufficiently good crosspoints were developed) would be about 4.25 Watts, which compares favorably with the 17W consumed by the 140x140 crossbar examined in section 2.2.1.

8.2 Remainders

Despite many successes, not all goals were achieved. The crosspoint and interconnects developed are not yet sufficiently good for building large arrays, though small arrays may be possible; in any case large demonstration arrays have not been built. System tests have therefore not been possible, and the global interconnect substrate necessary to assemble the full system was never built.

8.3 Changes

If this project were begun again from the start, there would be several changes in approach. An early focus on processing, with hands on cleanroom experiences, would have allowed more time to develop the metallization scheme further. Knowing that changes in the MUMPs process are not an option, use of the SUMMiT process would have begun much earlier, allowing more design iterations, and qualified supporting components (such as pads) by the final stages. More custom process monitors would have been included in each chip to improve understanding of process variations in general, and the thickness of OXIDE3 in particular. It is possible that developing a custom process to allow larger gap heights and larger die sizes in parallel with using SUMMiT would have been beneficial, but full MEMS process development takes sufficient time and expertise that it could be the basis of a thesis in

itself; developing a process with sufficient planarity to allow comparable results to SUMMiT would be difficult for an individual.

8.4 Directions for Future Work

Future work to make MEMS crossbars suitable for applications such as digital and broadband switching using thin film polysilicon processes may be most fruitful in three paths. The metallization crossover scheme using evaporation merits further exploration, as it can potentially greatly improve the utility for interconnect on MEMS chips. Using the improved interconnect, and further refining device design to take advantage of the latest revision of the SUMMiT process, crossbars could be built and suitability for both digital and broadband switching assessed. Finally, a process optimized for production of these devices would the potential for greatly improved device operation.

Rotating variable capacitors currently in development hold promise for certain RF applications, but have insufficient 'on' to 'off' ratios for crossbar applications.

The use of new processes with alternate mechanical materials, such as copper, has the potential to solve the metallization problems. While work has been done to convert the copper metallization stack of a standard process to a MEMS process, this is still at an early stage. Achieving a sufficiently low residual stress level to allow successful release of devices (i.e., the devices don't curl up), and characterization of the mechanical properties of the result-

ing layers is the next step in that effort. Another important concern in that regard is the long term mechanical properties of the copper; for example, whether spring fatigue occurs affects device lifetime, and whether operating points remain consistent over time.

References

[1] Dharma P. Agrawal, Tutorial Advanced Computer Architecture. 1986 IEEE,pp129-132.

[2] Laxmi N. Bhuyan and Dharma P. Agrawal, “Design and Performance of Generalized Interconnection Networks,” IEEE Transactions on Computers, Vol C-32, No. 12, December 1983, pp. 1081-1089.

[3] Kai Hwang and Faye A. Briggs, Computer Architecture and Parallel Processing. pp. 325-392,481-508. McGraw-Hill (1984).

[4] Susan R. Dickey and Richard Kenner, “Combining Switches for the NYU Ultracomputer”, Frontiers of Massively Parallel Computation, 1992. pp 521-523.

[5] D. A. Padua, “The Cedar Parallel Processor: Machine Organization and Software”, International Specialist Seminar on the Design and Application of Parallel Digital Processors, 1988 pp. 62-66.

[6] Josep Torrellas and Zheng Zhang, “The Performance of the Cedar Multistage Switching Network”, IEEE Transactions on Parallel and Distributed Systems, vol. 8, no. 4, April 1997, pp. 321-336.

[7] C. Clos, “A Study of Non-Blocking Switching Networks”, Bell System Tech. J 32, 406-424 (1953).

[8] V.E. Beneš, Mathematical Theory of Connecting Networks and Telephone Traffic, 136-158 Academic Press (1965).

[9] Anujan Varma and Suresh Chalasani, “Asymmetrical Multiconnection Three-Stage Clos Networks,” Proceedings of Sixth International Parallel Processing Symposium, March 1992. pp. 411-414

[10] Yuanyuan Yang and Jianchao Wang, “Wide-Sense Nonblocking Clos Networks Under Packing Strategy,” IEEE Transactions On Computers, vol. 48, no. 3, March 1999.

[11] J. Lenfant, “Parallel Permutations of Data: A Benes Network Control Algorithm for Frequently Used Bijections,” IEEE Transactions on Computers, Vol. C-27, No. 7, July 1978, pp. 637-647.

[12] J. Robert Heath and Stephen Riley, “Modeling And Implementation Of An NxN Clos-Type Interconnect Network Employing a ‘Clashing’ Control Procedure”, Proceedings Southeastcon ‘89, vol. 3 pp. 1211-1215.

[13] J. Robert Heath and Eric Allen Disch, “A Methodology For The Control And Custom VLSI Implementation of Large-Scale Clos Networks”. Computer Design, 1988, pp. 472-477.

[14]Eric Allen Disch, “A Methodology for the Control and Implementation of Large-Scale Clos Networks”, *Master’s Thesis*, Dept. of Electrical Engineering, Univ of Kentucky, Lexington, KY., Nov 1981.

[15]Deepak Rana, “A Control algorithm for 3-Stage non-blocking networks”, GLOBECON ‘92 Conference Record, vol 3, pp 1477-1491.

[16]Thaddeus J. Gabara, “On-chip Terminating Resistors for High Speed ECL-CMOS Interfaces”, AT&T Bell Laboratories.

[17]Tadahiro Kurodo et al.,“Automated Bias Control (ABC) Circuit for High-Performance VLSI’s,” IEEE Journal of Solid-State Circuits, Vol.27, No. 4, April 1992.pp641-648.

[18]Makoto Nakamurea, “An Instantaneous Response CMOS Optical Receiver IC with Wide Dynamic Range and Extremely High Sensitivity Using Feed-Forward Auto-Bias Adjustment,” IEEE Journal of Solid-State Circuits, Vol.27, No. 4, April 1992.pp991-997.

[19]Thaddeus J. Gabara and Wilhelm C. Fischer, “Capacitive Coupling and Quantized Feedback Applied to Conventional CMOS Technology,” IEEE Journal of Solid-State Circuits, Vol. 32, No. 3, March 1997.

[20]Leonard A. Hayden and Vijai K. Tripathi, “Pulse Signaling Using Capacitively-Coupled CMOS”, Electronic Performance of Electronic Packaging 1996. pp.7-9.

[21]David Salzman, Thomas Knight Jr, and Paul Franzon, “Application of Capacitive Coupling to Switch Fabrics”, Proceedings of MCMC ‘95, pp 195-199

[22]Rajeshwar Naik and D.M.H. Walker, “Large Integrated Crossbar Switch,” Proceedings Seventh Annual IEEE International Conference on Wafer Scale Integration, Jan 1996, pp. 217-227.

[23]Kyusun Choi and William S. Adams, “VLSI Implementation of a 256x256 Crossbar Interconnection Network,” Proceedings Sixth International Parallel Processing Symposium, March 1992, pp. 289-293.

[24]Frank E. Barber et al, “A 64x14 Non-Blocking Crosspoint Switch,” IEEE International Solid State Circuits Conference, February 1988, pp 116, 117, 322.

[25]Device specification from Applied Micro Circuits Corporation, part S2025

[26]Preliminary Product Summary from Velio Communications, Inc, part VC3003. May 22, 2001. http://www.velio.com.

[27]J.F. Dickson, “On-chip high-voltage generation in MNOS integrated circuits using an improved voltage multiplier technique,” IEEE Journal of Solid-State Circuits, vol. SC-11, pp. 374-378, June 1976.

[28]Ming Zhang, Nicolas Llaser, and Francis Devos, “Integrated Multivalue Voltage-To-Voltage Converter”, The 6th IEEE International Conference on Electronics, Circuits and Systems 1999. Proceedings of ICES ‘99. Volume 1, 1999.

[29]Toru Tanzawa and Shigeru Atsumi, “Optimization of Word-Line Booster Circuits for Low-Voltage Flash Memories”, IEEE Journal of Solid State Circuits, vol. 34, Issue 8, pp.1091-1098. August 1999.

[30]Kurt E. Petersen, “Silicon as a Mechanical Material”, Proceedings of the IEEE, Vol. 70, No. 5, May 1982.

[31]Hiroshi Hosaka, Hiroki Kuwano and Keiichi Yanagisawa, “Electromagnetic Microrelays: Concepts and Fundamental Characteristics,” MEMS ‘93 Proceedings An Investigation of Micro Structures, Sensors, Actuators, Machines and Systems. IEEE. Feb 1993.

[32]M. A. Gretillat et al. “Electrostatic Polysilicon Microrelays Integrated with MOSFETs”, Proceedings Workshop on MEMS ‘94, Jan 1994, pp 97-101.

[33]H. F. Schlaak, F. Arndt, J. Schimkat, and M. Hanke, “Silicon-microrelay with electrostatic moving wedge actuator- New functions and miniaturization by micromechanics,” Proceedings 5th International Conference Micro Electromechanical Systems, 1996, pp. 515-520.

[34]William P. Taylor, Oliver Brand, and Mark G. Allen, “Fully Integrated Magnetically Actuated Micromachined Relays,” Journal of Microelectromechanical Systems, Vol. 7, No. 2, June 1998.

[35]Qingyuan Meng, Mehran Mehregany, and Robert L. Mullen, “Theoretical Modeling of Microfabricated Beams with Elastically Restrained Supports,” Journal of Microelectromechanical Systems, Vol. 2, No. 3, September 1993.

[36]John Jung-Yeul Gill et al, “Elimination of Extra Spring Effect at the Step-Up Anchor of Surface-Micromachined Structure,” Journal of Microelectromechanical Systems, Vol. 7, No. 1, March 1998.

[37]Hirotsugu Matoba et al., “A Bistable Snapping Microactuator,” Proceedings, IEEE Workshop on Micro Electro Mechanical Systems, 1994.

[38]M. F. Chang et al. “Micro-Electro-Mechanical Systems (MEMS) for RF and Microwave Applications”, 3rd Annual Wireless Communications Conference. November, 1998, pp. 116-120.

[39]http://www.memsrus.com/cronos/svcsmumps.html; http://www.mcnc.org/mumps.html

[40]http://www.mdl.sandia.gov/Micromachine/

[41]Neil H. E. Weste and Kamran Eshraghian. Principles of CMOS VLSI Design:A Systems Perspective. Addison-Wesley, 1993.

[42]Ron K. Poon. Computer Circuits Electrical Design. Prentice Hall, 1995.

[43]Howard W. Johnson and Martin Graham. High-Speed Digital Design: A Handbook of Black Magic. Prentice Hall, 1993.

[44]Paul Franzon. "Electrical Modeling, Simulation, and Design of Electronic Packages: Volume 1 - Electrical Design.". Course notes for ECE544 at NCSU.

[45]http://www.ggb.com/

[46]Daryl Ann Doane and Paul D. Franzon [edited] Multichip Module Technologies and Alternatives: The Basics. Van Nostrand Reinhold, 1993.

[47]D. A. Winick, B. E. Duewer, S. Palchaudhury, and P.D. Franzon, "Performance evaluation of micromechanical binary phase-only holographic optical elements," 47th Electronic Components and Technology Conference, pp. 419-424, Components, Hybrids, and Manufacturing Technology Society, IEEE, (San Jose, CA), May 1997.

[48]David Winick, Bruce Duewer, Som Chaudhury, John Wilson, John Tucker, Umut Eksi, Paul Franzon, "MEMS-Based Diffractive Optical Beam Steering Technology" Miniaturized Systems with Micro-Optics and Micromechanics III, Vol. 3276. January 1998, pp. 81-87.

[49]Sandia MEMS Short Course. February 17-19, 1998. Sandia National Laboratories. Albuquerque, NM.

[50]M. Adrian Michalicek, John H. Comtois, and Heather K. Schriner, "Design and fabrication of optical MEMS using a four-level, planarized surface-micromachined polycrystalline silicon process" Miniaturized Systems with Micro-Optics and Micromechanics III, Vol. 3276. January 1998, pp. 48-55.

[51]"Methodology for Design of Electrostatic MEMS Devices Using the SUMMiT Process" Bruce Duewer, David Winick, John Wilson, Jeremy Palmer, Paul Franzon. Proceedings of the 45th International Instrumentation Symposium. pp 511-520. May 2-6,1999.

Appendices

Apprendix 1 Equation Derivations

The basic equations are as follows:

The force between two capacitive plates, where C is the capacitance, V is the voltage differential, and d is that distance between the plates, is given by:

$$F_c = \frac{C \cdot V^2}{2 \cdot d}$$

The capacitance between parallel conductive plates is given by:

$$C = \frac{\varepsilon \cdot A}{d}$$

The displacement, y, for a long beam when a force is applied to create a guided movement is given by:

$$y = \frac{F_s \cdot L^3}{6 \cdot E \cdot I}$$

with the moment of inertia, I, of the beam given by:

$$I = \frac{b \cdot h^3}{12}$$

Now, these basic equations can be used to determine the operation of a capacitive plate held above another capacitive plate by two long beams. Noting that when that when the plate is stationary, the force generated by the capacitive plates is equal to the force necessary to hold the beam at a given

displacement, the forces are set equal. Since two beams are used, the spring force is twice what it would be with only one beam. A little algebra yields:

$$y = \frac{\varepsilon \cdot A \cdot V^2 \cdot L^3}{2 \cdot b \cdot E \cdot h^3 \cdot d^2}$$

Often, there is a force to the substrate as well; it can also affect movement of the plate. It adds and additional force to the system. This additional force can be treated as another plate to plate interaction, and the same equations can be used (with the distance d from the substrate different, as well as the voltage difference). This continues to ignore the important point that the physical situation and fields are more complicated, and makes the assumption that the substrate to moving plate capacitance is unaffected by the actuation plate.

The force equation is now not just a matter of setting two forces equal, but the relations between three forces. Using the convention that the substrate force acts in the same direction as the actuation force, and recalling that we are modeling the static position:

$$2 \cdot F_s = F_{Cactuation} + F_{Csubstrate}$$

So now we introduce some new variables to allow easier manipulation of the various values. Let V_p be the voltage on the moving plate, V_a be the voltage on the actuation plate, and V_s be the substrate voltage. We still assume the signal plate voltage is the same as V_p so no attraction exists there. Let A_a

be the area for the actuation capacitance, and A_s be the area for the substrate actuation capacitance. Let d_a be the distance of the plate from the actuation plate, and d_s be the distance of the plate from the substrate.

So the force equation can be expanded. As this situation is more complicated, intermediate equations are given to facilitate following along.

$$2\frac{y \cdot 6 \cdot E \cdot I}{L^3} = \frac{C_a \cdot (V_p - V_a)^2}{2 \cdot d_a} + \frac{C_s \cdot (V_p - V_s)^2}{2 \cdot d_s}$$

Now plugging in the Inertia and Capacitance formulas, and rearranging:

$$2\frac{y \cdot 6 \cdot E \cdot \frac{b \cdot h^3}{12}}{L^3} = \frac{\frac{\varepsilon \cdot A_a}{d_a} \cdot (V_p - V_a)^2}{2 \cdot d_a} + \frac{\frac{\varepsilon \cdot A_s}{d_s} \cdot (V_p - V_s)^2}{2 \cdot d_s}$$

$$\frac{y \cdot E \cdot b \cdot h^3}{L^3} = \frac{\varepsilon \cdot A_a \cdot (V_p - V_a)^2}{2 \cdot d_a^2} + \frac{\varepsilon \cdot A_s \cdot (V_p - V_s)^2}{2 \cdot d_s^2}$$

$$\frac{d_a^2 \cdot d_s^2 \cdot y \cdot 2 \cdot E \cdot b \cdot h^3}{L^3 \cdot \varepsilon} = d_s^2 \cdot A_a \cdot (V_p - V_a)^2 + d_a^2 \cdot A_s \cdot (V_p - V_s)^2$$

$$\frac{d_a^2 \cdot d_s^2 \cdot y \cdot 2 \cdot E \cdot b \cdot h^3}{L^3 \cdot \varepsilon} =$$

$$d_s^2 \cdot A_a \cdot (V_p^2 - 2 \cdot V_p \cdot V_a + V_a^2) + d_a^2 \cdot A_s \cdot (V_p^2 - 2 \cdot V_s \cdot V_p + V_s^2)$$

To proceed from here, and solve for anything, some further assumptions are made. Since Vp is a voltage, we arbitrarily choose it as the point in the system where the voltage is zero.

$$\frac{d_a^2 \cdot d_s^2 \cdot y \cdot 2 \cdot E \cdot b \cdot h^3}{L^3 \cdot \varepsilon} = d_s^2 \cdot A_a \cdot V_a^2 + d_a^2 \cdot A_s \cdot V_s^2$$

Then it must be decided if the device is moving "up" (away from the substrate) or "down". Devices were fabricated that do each, so both paths are derived.

If actuation is downward, then $d_s = d_a + d_t$ (d_t is a new variable). Then

$$\frac{d_a^2 \cdot d_s^2 \cdot y \cdot 2 \cdot E \cdot b \cdot h^3}{L^3 \cdot \varepsilon} = (d_a - d_t)^2 \cdot A_a \cdot V_a^2 + d_a^2 \cdot A_s \cdot V_s^2$$

$$\frac{d_a^2 \cdot d_s^2 \cdot y \cdot 2 \cdot E \cdot b \cdot h^3}{L^3 \cdot \varepsilon} = (d_a^2 - 2d_a d_t + d_t^2) \cdot A_a \cdot V_a^2 + d_a^2 \cdot A_s \cdot V_s^2$$

Alternately,

$$\frac{d_a^2 \cdot d_s^2 \cdot y \cdot 2 \cdot E \cdot b \cdot h^3}{L^3 \cdot \varepsilon} = d_s^2 \cdot A_a \cdot V_a^2 + d_a^2 \cdot A_s \cdot V_s^2$$

$$\frac{d_a^2 \cdot d_s^2 \cdot y \cdot 2 \cdot E \cdot b \cdot h^3}{L^3 \cdot \varepsilon} - (d_a^2 \cdot A_s \cdot V_s^2) = d_s^2 \cdot A_a \cdot V_a^2$$

$$\frac{\frac{d_a^2 \cdot d_s^2 \cdot y \cdot 2 \cdot E \cdot b \cdot h^3}{L^3 \cdot \varepsilon} - (d_a^2 \cdot A_s \cdot V_s^2)}{d_s^2 \cdot A_a} = V_a^2$$

$$\frac{\frac{d_a^2 \cdot d_s^2 \cdot y \cdot 2 \cdot E \cdot b \cdot h^3}{L^3 \cdot \varepsilon} - \frac{(d_a^2 \cdot A_s \cdot V_s^2)(L^3 \cdot \varepsilon)}{L^3 \cdot \varepsilon}}{d_s^2 \cdot A_a} = V_a^2$$

$$\frac{d_a^2 \cdot d_s^2 \cdot y \cdot 2 \cdot E \cdot b \cdot h^3 - (d_a^2 \cdot A_s \cdot V_s^2)(L^3 \cdot \varepsilon)}{d_s^2 \cdot A_a \cdot L^3 \cdot \varepsilon} = V_a^2$$

$$\frac{d_a^2 \cdot (d_s^2 \cdot y \cdot 2 \cdot E \cdot b \cdot h^3 - (d_a^2 \cdot A_s \cdot V_s^2)(L^3 \cdot \varepsilon))}{d_s^2 \cdot A_a \cdot L^3 \cdot \varepsilon} = V_a^2$$

$$\frac{d_a}{d_s} \cdot \pm \sqrt{\frac{d_s^2 \cdot y \cdot 2 \cdot E \cdot b \cdot h^3 - A_s \cdot V_s^2 \cdot L^3 \cdot \varepsilon}{A_a \cdot L^3 \cdot \varepsilon}} = V_a$$

This should provide a solution of the V_a necessary to hold the device in place for any given displacement. Note the simplification of the substrate capacitance however; to compute d_s properly the effective distance for an air gap is being used, despite that there is some distance through the nitride.

In order to use this equation on the situation where the control voltage is applies with the plate voltage changing, then a fixed substrate voltage can no longer be used in this equation.

The sign of the actuation voltage is an interesting concern. While according the parallel plate model the sign does not matter, in the real world the fields will act differently when the actuation and substrate voltages have different signs than when they are the same.

If we don't choose a voltage as zero, and use Maple some of the algebra, the situation can be reduced to the form:

$$\frac{d_a}{d_s} \cdot \pm \sqrt{\frac{d_s^2 \cdot y \cdot 2 \cdot E \cdot b \cdot h^3 - (A_s \cdot (V_s - V_p)^2 \cdot L^3 \cdot \varepsilon)}{A_a \cdot L^3 \cdot \varepsilon}} = V_a - V_p$$

Another version of the equation rearranged to illustrate the effect of the substrate:

$$(V_a - V_p)^2 = \frac{y \cdot 2 \cdot E \cdot b \cdot h^3 \cdot d_a^2}{A_a \cdot L^3 \cdot \varepsilon} + \frac{A_s \cdot d_p^2 \cdot (V_s - V_p)^2}{A_a \cdot d_s^2}$$

Here we see the change that the substrate has on standard operation segregated to the term on the right; if the substrate effect is removed (set Vs=Vp), the old (non-substrate effect) equation can be seen by inspection to be equivalent.

Apprendix 2 Program Listings & Outputs

Selected short programs written to in the course of this research are included. They are here to provide a starting point for others.

2.1 Clos Array Combination Generator

```
/* A simple program to generate Clos networks to create switches of a
certain minimum size */

#include<stdio.h>

int closgen(int size) {
        int loop, asize, last_asize=size;
        int asize2, resultsize;
        int crosspoints, powerconsumed;

        printf("\nInput StageInner StageFinal StageResult Size");
        printf("Crosspoints");

        for(loop=2; loop < size; loop++) {
                asize = (size-1) / loop + 1;
                if (asize == last_asize) continue;
                last_asize = asize;

                asize2 = 2*asize-1;
                resultsize = loop * asize;

                printf("\n%d %dx%d   %d %dx%d   %d %dx%d   %dx%d",
                        loop,asize,asize2,
                        asize2,loop,loop,
                        loop,asize2,asize,
                        resultsize, resultsize);

                crosspoints = 2*loop*asize*asize2 + loop*loop*asize2;
                printf("%d",crosspoints);
        }

}

void main() {
        closgen(192);
        printf("\n\n");
}
```

Here is the output for a 192x192 array.

Input Stage	Inner Stage	Final Stage	Result Size	Crosspoints
2 96x191	191 2x2	2 191x96	192x192	74108
3 64x127	127 3x3	3 127x64	192x192	49911
4 48x95	95 4x4	4 95x48	192x192	38000
5 39x77	77 5x5	5 77x39	195x195	31955
6 32x63	63 6x6	6 63x32	192x192	26460
7 28x55	55 7x7	7 55x28	196x196	24255
8 24x47	47 8x8	8 47x24	192x192	21056
9 22x43	43 9x9	9 43x22	198x198	20511
10 20x39	39 10x10	10 39x20	200x200	19500
11 18x35	35 11x11	11 35x18	198x198	18095
12 16x31	31 12x12	12 31x16	192x192	16368
13 15x29	29 13x13	13 29x15	195x195	16211
14 14x27	27 14x14	14 27x14	196x196	15876
15 13x25	25 15x15	15 25x13	195x195	15375
16 12x23	23 16x16	16 23x12	192x192	14720
18 11x21	21 18x18	18 21x11	198x198	15120
20 10x19	19 20x20	20 19x10	200x200	15200
22 9x17	17 22x22	22 17x9	198x198	14960
24 8x15	15 24x24	24 15x8	192x192	14400
28 7x13	13 28x28	28 13x7	196x196	15288
32 6x11	11 32x32	32 11x6	192x192	15488
39 5x9	9 39x39	39 9x5	195x195	17199
48 4x7	7 48x48	48 7x4	192x192	18816
64 3x5	5 64x64	64 5x3	192x192	22400
96 2x3	3 96x96	96 3x2	192x192	28800

2.2 Clos Programing Code

```
/* A function for programming a clos array, plus test functions to assure
   that it actually works for all combinations*/
/* Right now it assumes 3 stage clos, equal number of inputs and outputs
*/

#include<stdio.h>

/* These defines set up the size and composition of the crossbar. Based
on
   the overall size (SIZE) and the size of an available center section
(CENTER)
   of arrays in the center section, the rest of the crossbar composition
   can be determined */
#define SIZE 32
#define CENTER 8
#define ASIZE ((SIZE-1)/CENTER+1)
#define ASIZE2 (2*ASIZE-1)
```

```
#define RSIZE (CENTER*ASIZE)

#define NODEST -1

/* Array data structures */

/* The following structures would work for tracking the state of each
   crosspoint. This isn't really necessary yet, but would be if the
   program were actually controling switches, and were worried about
   things like testing each switch to check yield, etc. */
/*
typedef struct instage_t {
     int config[ASIZE][ASIZE2];
} instage_t;

typedef struct midstage_t {
     int config[CENTER][CENTER];
} midstage_t;

typedef struct outstage_t {
     int config[ASIZE2][ASIZE];
} outstage_t;
*/

/* The next ones just show the mapping which is sufficent to track the
routes chosen through the crossbar until yield and multicast sims are
needed */

typedef struct instage_t {
     int config[ASIZE];
} instage_t;

typedef struct midstage_t {
     int config[CENTER];
} midstage_t;

typedef struct outstage_t {
     int config[ASIZE2];
} outstage_t;

typedef struct fabric_t {
     instage_t instage[CENTER];
     midstage_t midstage[ASIZE2];
     outstage_t outstage[CENTER];
} fabric_t;

typedef struct equiv_t {
     int config[SIZE];
} equiv_t;
```

```
/* reset_fabric simply goes through and sets all devices off; physically
   this operation is equivalent to moving all control lines to release */
reset_fabric(fabric_t * fabric) {
     int loop, loop2;

     for(loop=0; loop < CENTER; loop++) {
          for(loop2=0; loop2<ASIZE;loop2++) {
               fabric->instage[loop].config[loop2] = NODEST;
          }
     }
     for(loop=0; loop < ASIZE2; loop++) {
          for(loop2=0; loop2<CENTER; loop2++) {
               fabric->midstage[loop].config[loop2] = NODEST;
          }
     }
     for(loop=0; loop < CENTER; loop++) {
          for(loop2=0; loop2<ASIZE2;loop2++) {
               fabric->outstage[loop].config[loop2] = NODEST;
          }
     }
}

/* This would display the state of the switch fabric in an appropriate
   manner. */
void show_fabric(fabric_t *fabric) {
printf("Write me in detail later\n");
}

/* Finds a way to get from line "from" to line "to" and set the array */
int make_path(fabric_t *fabric, int from, int to){
     int midtry;
     int from_block, to_block;

     from_block = from / ASIZE;
     to_block = to / ASIZE;

     for(midtry=0; midtry < ASIZE2; midtry++) {
          if ((fabric->midstage[midtry].config[from_block] == NODEST) &&
              (fabric->outstage[to_block].config[midtry] == NODEST)) {
               /* Now make connections */
               fabric->instage[from_block].config[from % ASIZE] = midtry;
               fabric->midstage[midtry].config[from_block] = to_block;
               fabric->outstage[to_block].config[midtry] = to % ASIZE;
/*             printf("Connected %d %d %d\n",from,midtry,to); /**/
               return 1; /* Using one for success here. Who cares about
tradition? */
          }
     }

return 0; /* No routes found, so failing */
```

```
}

/* A simple naive way to program the fabric- assign one line at a
   time with no concern for painting myself into a corner. Because the
   array is a Clos array, we can get away with doing this. In fact, that
   is one of the reasons Clos arrays are so useful. */
void program_fabric(fabric_t * fabric,equiv_t * equiv) {
     int loop;

     for(loop=0; loop < SIZE; loop++) {
          if(!make_path(fabric, loop, equiv->config[loop]))
               printf("Failed connecting %d %d\n",loop,equiv-
>config[loop]);
     }
}

/* Used in recursively testing all configurations */
void recurse (equiv_t * equiv, int place, int used[]) {
     int loop;

if(place) {
     --place;
     for(loop=0; loop<SIZE; loop++)
          if (!used[loop]) {
               equiv->config[place] = loop;
               used[loop] = 1;
               recurse(equiv,place,used);
               used[loop] = 0;
          }

} else {
     fabric_t fabric;

     reset_fabric(&fabric);
     program_fabric(&fabric,equiv);
}
}

/* Exhaustively test all possible routes. A less than efficient way to
check for some forms of mistakes (coding for now, failed crosspoints
if we were controling an actual switch) that is simple to code. For
a real application, use something more intelligent. */
void exhaust_test() {
     int place;
     int loop;
     int used[SIZE];
     equiv_t equiv;

     place = SIZE-1;
```

```
     for (loop=0; loop < SIZE; loop++) used[loop]=0;

     for(loop=0; loop<SIZE; loop++) {
          equiv.config[place] = loop;
          used[loop] = 1;
          recurse(&equiv,place,used);
          used[loop] = 0;
     }
}

void main() {
     fabric_t fabric;
     equiv_t equiv;
     int loop;

     printf("Input Stage\tInner Stage\tFinal Stage\tResult Size\n");
        printf("%d %dx%d   \t%d %dx%d   \t%d %dx%d   \t%dx%d\n",
          CENTER,ASIZE,ASIZE2,
          ASIZE2,CENTER,CENTER,
          CENTER,ASIZE2,ASIZE,
          RSIZE,RSIZE);

     reset_fabric(&fabric);
     show_fabric(&fabric);

     /* A simple test, route each input to corresonding output */
     for(loop=0; loop<SIZE; loop++) equiv.config[loop] = loop;
     program_fabric(&fabric,&equiv);
     show_fabric(&fabric);

     /* Now let's really waste some runtime. Exhaustively test all
           possible routes. Don't bother doing this for anything much
        larger than the provided setup. It's just in this example for
           an expensive sanity check */
     exhaust_test();
}
```

2.3 Macro Code for FastCap

I found that the macro preprocessor for a C compiler could greatly simplify the creation of a file for FastCap. The following code segment creates a simple quadralateral. Note that as written it does not subdivide faces. Choosing good points at which to subdivide faces is a key decision in preparing to simulate a device.

```
/* This makes a quick quadrilateral */
#define CNUM 2

/* The x/y/z boundaries */
#define x1 100e-6
#define x2 37.0e-6
#define y1 33.0e-6
#define y2 57.0e-6
#define z1 0.1e-6
#define z2 0.5e-6

/* The corners */
#define C1 x1 y1 z1
#define C2 x2 y1 z1
#define C3 x1 y2 z1
#define C4 x2 y2 z1
#define C5 x1 y1 z2
#define C6 x2 y1 z2
#define C7 x1 y2 z2
#define C8 x2 y2 z2

/* The planes */
Q CNUM C1       C2      C4      C3
Q CNUM C5       C6      C8      C7
Q CNUM C1       C2      C6      C5
Q CNUM C3       C4      C8      C7
Q CNUM C1       C3      C7      C5
Q CNUM C2       C4      C8      C6
```