Silicon Micromachined Magnetic Actuators for Aerodynamic Flow Control Applications

Thesis by

Thomas Rocco Tsao

In Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

1998

Exam Date: May 8, 1998

To my parents

Approved:

Acknowledgements

I believe that I have grown tremendously as a person during graduate school. For this, I have to thank those with whom I have had the closest contact. I hope that I have absorbed the fine qualities I have seen in others and that I can live by these principles. I have been extremely lucky to have been associated with all of those mentioned below, and I would like to thank them for helping me grow.

First of all, I would like to thank my parents and sister for the support they have provided to me, not just during my graduate education, but in everything I have ever attempted in my life. Any success I have had or will have is a direct result of their love and encouragement.

Words cannot adequately describe what my girlfriend Eugenia who, along with Spanky, have meant to me over the years. Their love, honesty, humor, kindness, and thoughtfulness have meant the world to me, and I cannot describe myself without including them.

Professor Yu-Chong Tai has been an advisor in every sense of the word. I believe that as valuable as his technical advice has been, his philosophical advice will be of even greater value in everything - technical and non-technical - that I do in the future. In addition to being an advisor, he has been a mentor and a manager of the highest caliber. If I am ever recognized for being a compassionate mentor who helps those under me to achieve their potential, it will be a reflection of the man that Professor Tai is. To have your advisor place his trust and confidence in you is one of life's most rewarding feelings.

Professor Chang Liu was my first research partner, and it was an honor to work with him. He is one of the hardest working researchers I have ever known. He taught me a great deal about research, but I believe that the most important thing I learned from him is selflessness and sacrifice. I cannot count the number of times Chang would delay his own work to help another, especially when a machine would break. For a lab such as ours to run smoothly, it is crucial to have people like Chang.

Dr. Fukang Jiang has been working on the same projects as I have been, and through actions, he has shown me what it means to be a complete researcher. His scientific curiosity and work ethic are models for any incoming graduate student. I also had the pleasure of working with Ken Walsh, a true Techer with both a broad and deep knowledge of many fields. In addition to his insight into many problems, his sense of humor makes any environment he's in much more pleasurable.

The "P" in Ph.D. stands for philosophy, and I believe that any graduate student's life would be incomplete without many long, late-night conversations on obscure topics. Through his insightful thoughts, Xing Yang has provided me with countless hours of deep thinking. He is truly a scholar and an inspiration to those who want to learn. John Wright is one of the most creative people I know. I am grateful for all of his advice, help, and creative ideas, and his sense of humor is always welcome.

Wen Hsieh is a constant source of energy. I can always count on him to provide stimulating conversation and liven up any day. His humor and graciousness have always been very appreciated. Wen and Charles Grosjeans have also shown me the value in meticulousness. If I ever need a job to be done perfectly, I would certainly seek their assistance. Charles is an extremely fair and forthright person, and I know that I can

always trust him for an honest opinion. From Amish Desai, I have learned to look for humor in stressful situations and that it is not the end of the world if an LTO tube explodes. Shuyun Wu has always impressed me with his willingness to volunteer for work. I believe that this sort of attitude will be rewarded in the end. X.Q. Wang follows the mold of many others in this lab by being extremely hard-working and kind. He has always responded immediately and in an enthusiastic manner whenever I have asked for his assistance, and I hope to work with him in the future. I admire Dr. Sang-Wook Lee for his providing a level of professional dedication to his work, and I hope that I will be able to emulate this in the future.

I would also like to thank those who came before me in the lab, specifically Dr. Svetlana (Ceca) Tatic-Lucic, for her encouragement and for providing a smile to a new graduate student, and Dr. Raanan Miller, for his always forthcoming technical advice.

I would like to thank those others who have helped me including Dr. Weilong Tang, Dr. Yang Hsu, Ellis Meng, Marc Unger, Dr. Chuan-Cheng Cheng, Dr. Bhusan Gupta, Vincent Koosh, Jeff Dickson, and Zeke Kruglick. In addition to the guidance provided by Professor Chih-Ming Ho of UCLA, I owe a deeply grateful to Dr. Steve Tung and Dr. Vincent Lee for their fluid mechanics expertise.

Of course, nothing could have been done without the help of Trevor Roper, who maintains all of our equipment. I am truly grateful for all of the times he has helped me by fixing the equipment needed for my process. Janice Tucker has provided excellent support in helping me wade through all of the paper-work that has come my way.

Finally, I would like to thank DARPA and AFOSR for their financial support of my research.

Silicon Micromachined Magnetic Actuators for Aerodynamic Flow Control Applications

By

Thomas Tsao

In Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

Abstract

Active micro-scale flow control has been a long sought-after goal of fluid mechanists. One of the primary desires within this field is to control macroscopic events with microscopic forces. Recent research has provided shear stress sensors capable of detecting micro-scale structures present in aerodynamic flows. In this work, the results (fabrication, device testing, fluidic testing) of developing magnetic actuators to interact with and affect the flow are reported. A first attempt at M^3 (microsensor, microactuator, and microelectronic) integration is also presented.

One of the primary aerodynamic motivations is the control of the rolling moment of delta wing aircraft. For the demonstration of such an application, a line of flaps on a leading edge of the wing should be activated simultaneously. Therefore, arrays of passive magnetic flaps were fabricated and successfully used to create significant rolling moments in wind tunnel testing of delta wings.

The other motivation in the development of magnetic flaps is the reduction of drag in turbulent boundary layers. To succeed in this effort, the actuators must interact

with, and negate the effects of, millimeter sized turbulent structures. Due to the unpredictable nature of such structures, each flap across a control surface should be individually addressable. Many designs of actively controlled magnetic flaps were fabricated and tested. The various generations were designed not only with device performance in mind, but also eventual integration with sensors and electronics. To this end, several fabrication issues, most notably dealing with the sacrificial layer, are presented.

Finally, a first attempt at integrating sensors, actuators, and electronics on one substrate is presented. Such an effort is crucial both for the eventual use in flow control applications as well as in other generic MEMS/electronics integration efforts.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 MEMS and Microfluidics

Elements of micromachining have been around for over 30 years [1,2,3]. Interest in the field, however, waned soon after – quite possibly, it was a technology far ahead of its time. Then, approximately 10-15 years ago, there was a resurgent interest, spawned primarily from schools such as UC Berkeley and MIT. Almost all of the research conducted in the beginning of this renaissance related to the creation of individual devices or the study of the basic sciences and fundamental principles behind the device operations. Many of these devices were wondrous and quite novel. Unfortunately, however, many of these same devices were also not very useful and were coined "gadgets." The building of these gadgets and the study of the basic sciences laid the foundation for a more recent (approximately 5 years old) interest in micromachined systems that combine electronics and MEMS in a practical system.

One of the most interesting areas in which micromachined systems could make a large impact is the field of microfluidics, a scientific field that has never been adequately studied before due to a lack of experimental methods - the sensors simply were not mature enough. Interest in microfluidics extends beyond a pure theoretical interest. Many practical problems have their root in microfluidics. Much as the computing power once contained in large, room-sized arrays of vacuum tubes can now be held in the palm of a hand, there may be enormous benefit in shrinking macro-fluidic systems. These benefits may lay in size, weight, speed, economic, or distribution.

In addition, all macroscopic flows have microscopic components, which may or may not be relevant to the overall flow pattern. If the microscopic events are, indeed, relevant, the possibility exists that microscopic control could lead to macroscopic effects, in other words, fluidic amplification. Fluidic amplification already exists on a meso-scale. Companies such as Allied Signal produce analog fluidic computing elements that only use flow (no moving or electrical parts) to achieve the desired computing function.

Initial work has been done on understanding microfluidics, and many new phenomena have been observed, which reinforces the fact that our intuition about fluid flow on a person-sized scale does not necessarily translate to a micro-scale level. [4] One of the beautiful results of using micromachined devices is that scientists can now both study and affect physical phenomena on a level previously ignored due to lack of experimental apparatus.

One step beyond simply observing physical phenomena is to try and actively control it. This focus of the work in this thesis relates to the active control of fluid flow, in particular trying to control turbulent boundary layer flow and attempting to control the rolling moment of delta wing aircraft. Generically speaking, the problem of active fluid control entails integrating sensors, actuators, and control electronics repetitively across the control surface. Many issues need to be addressed, including discrete device performance, system performance, control algorithms, and fabrication compatibility.

1.2 Microfluidics on Airplanes

Before describing two of the specific projects targeted in this work, it is necessary to describe the impetus behind all of the studies of microfluidics in aerodynamic

applications [5]. The primary idea is the thought that an extremely small control force can have a macroscopic effect:

$$O(\varepsilon)\ \text{cause} \longrightarrow O(1)\ \text{result}$$

If proven successful, the results will have powerful implications – most importantly that a small expenditure in energy at the right location at the right time either can have the same effect as a large expenditure of energy or can produce a savings in energy far greater than that spent. These are indeed highly worthwhile goals!

1.2.1 Drag Control in Turbulent Boundary Layers

One of the main thrusts of this thesis is to present an attempt to solve a part of a problem that has been the subject of attention as long as man has moved in vehicles - the reduction of drag. From a practical point of view, any object moving in a fluid experiences drag. When the objects in question are man-made vehicles, and the drag experienced by such vehicles results in increased fuel costs and decreased operating efficiencies, the effort spent to understand and reduce drag becomes an extremely worthwhile endeavor. Indeed, for as long as man has spent moving in vehicles, he has been interested in moving faster with less effort.

There are many causes of drag on an object moving in a fluid. Of these, viscous drag refers to drag which occurs within a thin (boundary) layer next to the object. If the flow in this boundary layer is turbulent, as the air flows across the wing of a flying plane, many millimeter-sized fluidic structures form randomly in time and space across the surface. These streaks are comprised of pairs of counter-rotating vortices that bring high momentum fluid down to the surface. This causes an increase in the gradient of velocity with respect to vertical distance, which is proportional to the surface shear stress. This

shear stress, when integrated over an area, contributes significantly to surface drag. In 1985, it was estimated that if viscous drag on the fuselage of airplanes were to be reduced by 10%, the cost savings would be on the order of $350 million [6]. By extrapolating to 1998 using a 3% annual rate of inflation and a more than doubling of air travel [7], those numbers translate today into a savings of over $1.1 billion.

The fact that these streaks are randomly distributed in time and space implies that any system attempting to reduce the drag caused by these structures must be distributed across the entire surface. Also, the size of these structures at typical airplane speeds is on the order of a millimeter. Therefore, the size of the micromachined elements of such a system should also be of the same order. Finally, the control of such a system should be distributed and localized rather than centralized. The fact that vast numbers of sensors and actuators (thousands of each) need to be addressed in a timely manner implies that centralized processing and control probably cannot be used. A similar analogy can be made comparing such a system with the human nervous system. If a person touches a hot object, the control signal to pull back comes from nerves in the spine (localized control); the signal only reaches the brain (central computing unit) later. It is neither necessary (there is no need to know the condition of the foot when the hand is touching a hot object) nor prudent (the time saved by not sending the information to the brain minimizes the damage to the hand) for the brain to make this decision.

In fact, the entire concept of reducing drag has several relevant nature-inspired analogies, and much effort has been spent studying drag reduction in animals. In particular to viscous drag reduction, many examples have been found in marine animals. Some involve the coverage of surfaces with various "slimes" [8]. Many fast swimming

sharks utilize a passive drag reduction scheme with small riblets [9]. Passive here refers to the fact that the riblets are no more than simple attachments to the scales. No movement or control scheme is involved. Dolphins and porpoises have been the focus of many studies. In 1936, Gray [10] made some simple observations and performed some calculations which resulted in his belief that the drag associated with the flow around a dolphin must be laminar, if not "better". He believed that the flow could not be turbulent because the dolphins simply could not generate the power required to move through the water at the speeds he measured. Gray therefore postulated that dolphins have some means of turning turbulent flow laminar. In later works [11], others have speculated that this may be the result of compliant skins or perhaps even active control, where the dolphins would be able to sense local areas of turbulent flow and then vibrate their skin locally to cancel the effects of the turbulence. Others [12,13], still, have postulated that many of Gray's assumptions and measurement techniques were in error and that dolphins do not have any such means of reducing drag. To this day, no overwhelming evidence has been obtained proving either side correct. It is very difficult to obtain accurate experimental evidence, as the very act of observing can often affect the dolphins (i.e. the very presence of a boat may spook the dolphins into going underwater or the presence of a moving boat may generate wakes which affect the flow around the dolphins). One thing is certain, though, that nature has created animals that use many means of moving faster. Active control of viscous drag may be one of the means.

The development of an integrated MEMS system for turbulent boundary layer control is important for reasons other than reducing drag on airplane wings. Many sub-processes that are crucial for the overall system operation are very important and

significant in their own right. The MEMS sensor and actuator used in this project can easily be ported to other applications. Some of the basic processing technologies that were necessary either on the device level or the system level can also be applied to completely orthogonal MEMS problems. The integration issues that need to be addressed are problems of general interest that arise in many other integrated (not necessarily microfluidic) MEMS/electronics systems.

1.2.2 Rolling Moment Control on Delta Wing Aircraft

Another project described herein involves the control of the rolling moment of a delta wing aircraft. The basic concept behind this project (Fluid mechanical details will be presented in Chapter 2.) is the fact that a pair of counter rotating vortices are generated above a delta wing aircraft. These vortices are responsible for producing a significant portion of the lift experienced by such an aircraft. If these normally symmetrical vortices are somehow made asymmetric, a strong rolling moment could be produced, thus resulting in a highly agile aircraft. These vortices are created at a point upon the delta wing known as a separation point. After creation, the vortices grow in size as they travel the length down the plane. If an extremely small disturbance can affect the placement of the separation point, the vortices could be asymmetric, creating a lift imbalance.

To create such a system, it is necessary to have sensors spread around the area in which separation is expected to occur. Alongside these sensors, actuators are needed to produce the required epsilon-sized force. Some simple control electronics are also necessary to process the signals from the sensor array, determining both the initial separation point and also the separation point adjustment resulting from the actuator motion. From this information, some feedback control can be given to the actuators

depending on the desired rolling motion of the aircraft. Once again, it is necessary to have a distributed MEMS system integrating sensors, actuators, and electronics. However, the overall requirement is simpler than in the drag reduction project due to the facts that (1) the separation point is fairly well known for a given geometry; there isn't a random variation in time and space of the fluid mechanical objects under study, and (2) the control electronics are simpler because the control algorithm does not need neural net control or learning capabilities. This process can be more difficult than the control of turbulent structures, however, due to the fact that the eventual integrated system will need to be fabricated on a flexible skin.

References

[1] Nathanson, H., Newell, W., Wickstrom, R., and Davis, J., "Resonant Gate Transistors," *IEEE Transactions on Electron Devices*, Vol. 14, pp. 117, 1967.

[2] Petersen, K., "Silicon as a Mechanical Material," *Proc. IEEE*, Vol. 70 (5), pp. 420-457, 1982.

[3] Nathanson, H. and Guldberg, J., "Topologically structured thin films in semiconductor device operation," **Physics of Thin Films Volume 8**. Academic Press, New York, 1975.

[4] Liu, J., Tai, Y.-C., Pong, K., and Ho, C.-M., "Micromachined Channel/Pressure Sensor Systems for Microflow Study," *Technical Digest 7th International Conference Solid State Sensors and Actuators,* Yokohama, Japan, pp. 995-997, 1993.

[5] Ho, C.-M. and Tai, Y.-C., "Review: MEMS and Its Applications for Flow Control," *Journal of Fluids Engineering*, Vol. 118, pp. 437-447, 1996.

[6] Walsh, M., "Riblets as a Viscous Drag Reduction Technique," *AIAA Journal*, Vol. 21 (4), pp. 485-486, 1983.

[7] International Civil Aviation Organization (ICAO), "World Airline Passenger Traffic Growth Rate to Continue Through to 1998", *News Release*, 1996.

[8] Bushnell, D. and Moore, K., "Drag Reduction in Nature," *Annual Review Fluid Mechanics*, Vol. 23, pp. 65-79, 1991.

[9] Bechert, D., Bartenwerfer, M., and Hoppe, G., "Drag Reduction Mechanisms Derived from Shark Skin," Paper 86-1.8.3, 15th Congress of the International Council of the Aeronautical Sciences, 1986.

[10] Gray, J., "Studies in Animal Locomotion. VI. The Propulsive Powers of the Dolphin," *Journal of Experimental Biology*, Vol. 13, pp. 192-199, 1936.

[11] Kramer, M., "The Dolphins' Secret," *Journal American Society Naval Engineering*, Vol. 73, pp. 103-107, 1961.

[12] Au, D. and Weihs, D., "At High Speeds Dolphins Save Energy by Leaping," *Nature*, Vol. 284, pp. 548-550, 1980.

[13] Lang, T. and Pryor, K., "Hydrodynamic Performance of Porpoises (Stenalla atenuata)," *Science*, Vol. 152, pp. 531-533.

Chapter 2

Background

2.1 Fluids

2.1.1 Basic Fluid Concepts

In order to appreciate the primary problem being studied for this thesis, it is necessary to have a rudimentary understanding of the fluid mechanics phenomena that govern the relevant regimes of operation. First of all, there are four primary restrictions that are assumed to be met in all of the fluid mechanics covered in this work [17,18]:

(1) The laws of classical mechanics always apply.

(2) The length scale of the flow is large compared with the molecular mean free path. We are not considering rarified gas flow.

(3) The flow is considered incompressible. In gas flow, this condition implies that the flow velocity is low compared to the speed of sound.

(4) Fluids are considered to be Newtonian. This assumption essentially restricts the flow to be comprised of small molecules that interact in a simple way (i.e. no long-chain polymers). One result of this assumption is that the viscosity is considered to be independent of the velocity field and that the shear stress is directly proportional to the velocity gradient.

For steady (i.e. time invariant) incompressible flow with no body forces (i.e. gravity), the equations of fluid motion can be represented by the Navier-Stokes equation

$$\mathbf{u}\cdot\nabla\mathbf{u} = -\frac{1}{\rho}\nabla p + \upsilon\nabla^2\mathbf{u} \qquad (2.1)$$

where **u** is the velocity vector, p is pressure, ρ is density, and υ is the kinematic viscosity (defined as μ/ρ where μ is the viscosity). The Navier-Stokes equation needs to be combined with boundary conditions to obtain flow solutions. Two of the most important boundary conditions that apply at a surface are the facts that (a) there is no normal flow at the surface and that (b) tangential flow is reduced to zero velocity at the surface. The tangential boundary condition is known as the no-slip boundary condition and is a result of the viscosity of the fluid slowing down the fluid near the boundary to match the boundary speed. While there are many physical descriptions as to why this should be, for our purposes, the best justification, perhaps, is the fact that the no-slip boundary condition has shown itself to be valid experimentally.

One of the most important quantities in fluid mechanics is the non-dimensional quantity called the Reynolds number (Re). Note that in fluid mechanics, non-dimensional quantities are often abbreviated using two letters with the first being capitalized. Reynolds number is defined as the following

$$\mathrm{Re} = \frac{\rho \mathrm{UL}}{\mu} \tag{2.2}$$

where U is a characteristic velocity and L is a characteristic length. Reynolds number can be thought of as a ratio of inertial forces (force present simply due to the fact that fluid is moving) to viscous forces, or, in terms of the Navier-Stokes equation, as a ratio between the first and third terms.

When $\mathrm{Re} \ll 1$, the flow is called creeping motion. This sort of flow can be thought similar to tree sap flowing in winter. One interesting feature of creeping flow is

that the viscous interactions extend far into the flow. In other words, viscous effects can diffuse far into the flow before being convected downstream.

When Re>>1, the viscous forces become negligible compared to inertial forces. Therefore, the viscous term in the Navier-Stokes equations can be neglected and is reduced to the following.

$$\rho \mathbf{u} \cdot \nabla \mathbf{u} = -\nabla p \tag{2.3}$$

This equation is known at Euler's equation of inviscid motion. One interesting aspect of this equation is that each fluid particle is subject to an acceleration directly proportional to the pressure gradient.

For reasons of mathematical consistency, Euler's equation (which neglects viscosity) must be solved without the use of the no-slip boundary condition, which is a direct result of the presence of viscosity. This leads to an interesting paradox – boundary conditions either apply or don't apply. They cannot be thought of as existing, but being considered "small" relative to a different situation. Experimentally, Euler's equation has shown itself to be valid in high Reynolds number flow regimes. However, for high Re flow, solving Euler's equation for the flow around a cylinder, the solution implies that there is no drag force! This is known as D'Alembert's paradox.

The solution to this apparent paradox is the presence of a very thin layer next to the surface called a boundary layer. Inside this boundary layer, viscous effects are important and the no-slip boundary condition applies. Once outside the boundary layer, Euler's equation of inviscid flow holds. Therefore, to accurately model high Reynolds number flow, it is necessary to first apply Euler's equation to the majority of the flow around the object. This solution is then combined with the boundary layer flow solution

in a thin layer adjacent to the object (and in an area behind the object called the wake) to form the complete solution.

2.1.2 Boundary Layers

The boundary layer can be defined as the area in which the velocity differs from the inviscid solution by more than one percent (i.e. velocity is less than 99% the inviscid solution velocity – see Figure 2.1). For a steady, 2-D boundary layer, the velocity profile can be numerically calculated and shown as a Blasius profile. For flow across a flat plate (or an airplane wing), a thin laminar boundary layer develops at the leading edge and grows as approximately $x^{1/2}$. A short distance later, the laminar boundary layer becomes unstable, and disturbances within are amplified rather than attenuated. During this time, the laminar boundary layer undergoes transition and eventually turns into a turbulent boundary layer. The turbulent boundary layer is characterized by being much thicker than the laminar boundary layer (increasing as approximately $x^{4/5}$) and having a velocity profile that randomly fluctuates.

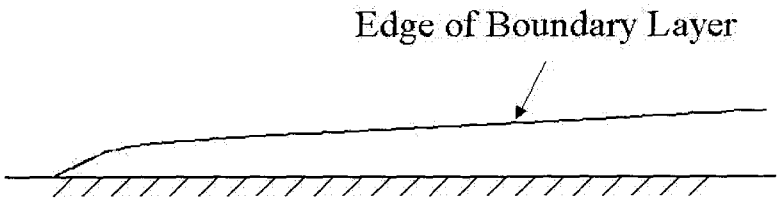


Figure 2.1 Drawing of a boundary layer across a flat plate. Note that the edge of the boundary layer corresponds to a flow velocity of 99% of the free stream velocity.

2.1.3 Separation

In low Re (Re<1) flow around an object, the flow follows the object shape very well. As the Reynolds number is increased, the fluid can no longer flow around corners

in a conformal manner. One way of thinking of this is to imagine a person walking around a sharp corner making a sharp 90° turn. If this person were to run around the corner, he would have to take a slightly wider turn and his path wouldn't follow the corner exactly. Separation is the term that refers to the condition in fluid mechanics when the flow around an object fails to follow the shape of the object. (See Figure 2.2) Separation can occur both in laminar and turbulent flows. At a high enough Reynolds number, the flow separating behind an object results in periodic shedding of Karman vortex streets which form two rows on opposite sides of the wake. Each row is comprised of vortices rotating in the same direction, and the vortices on opposite sides rotate in directions opposite of each other.

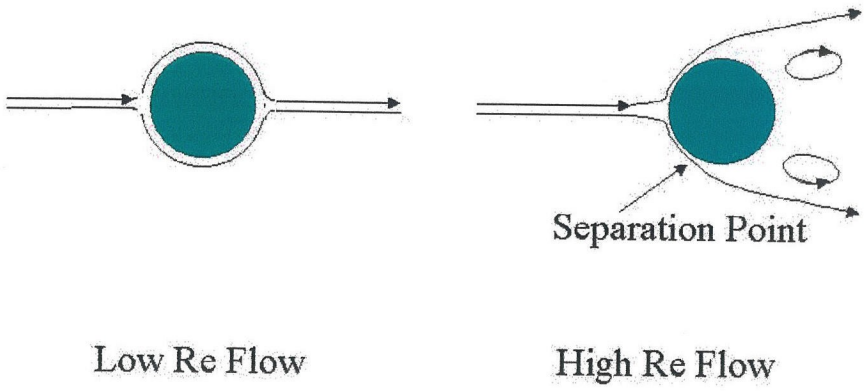


Figure 2.2 Conceptualization of separation

2.1.4 Drag

There are many sources of drag on any object moving through a fluid. Here, we will only consider two types: (1) pressure, or form, drag which is due to flow separation, and (2) and viscous, or skin-friction, drag.

First of all, there is form drag, which is a result of a negative pressure differential between the front and the back of the object. A fluid moving past an object will separate

from the object and cause a wake to be produced behind the object. The downstream velocity in this wake is less than the upstream velocity. By applying the laws of conservation of momentum, the velocity lost in by the flow is transferred to a retarding force (which can also be thought of as a pressure differential) on the body. Race car drivers often take advantage of these pressure differentials by "drafting" or closely following the driver ahead of them and taking advantage of the fact that the leading driver is "pushing" most of the air out of the way. To reduce form drag, it is necessary to delay the onset of separation for as long as possible.

Secondly, there is viscous (or skin-friction) drag, which results from the integration of the viscous shear stresses in the boundary layer, as described in the following equations:

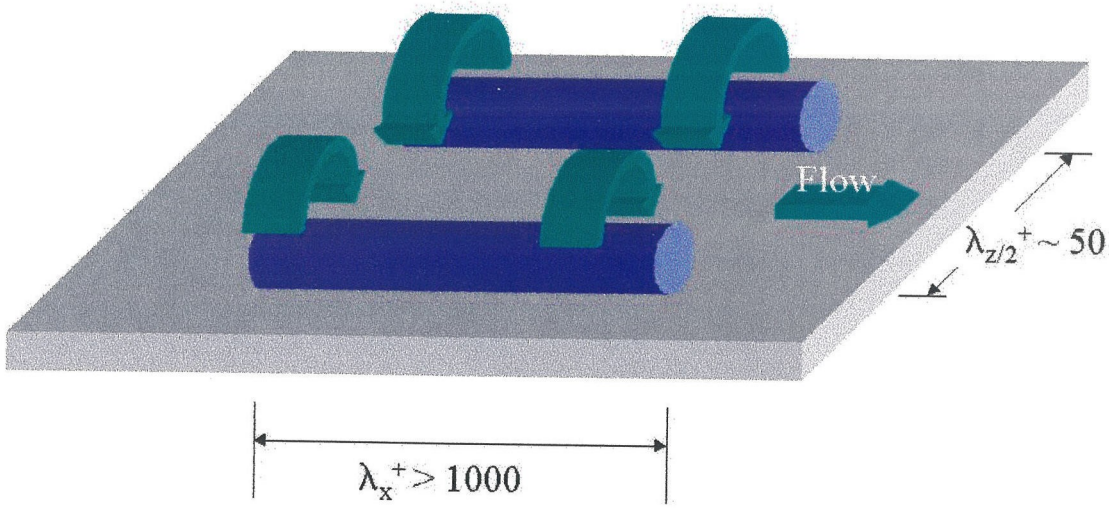
$$\tau = \mu \frac{du}{dy} \tag{2.4}$$

$$F_d = \int_A \tau dA \tag{2.5}$$

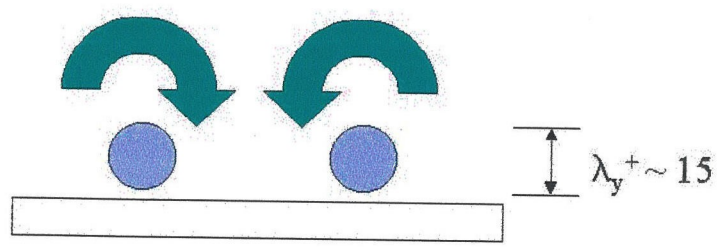
τ=shear stress
μ=viscosity
u=streamwise velocity
y=vertical direction
A=surface area
F_d=drag force

For a jet transport, [1] skin friction drag can account for approximately 50% of the drag, and it would make sense to spend some effort reducing it. Within a turbulent boundary layer, there exist many different types of structures [2]. Each type typically exists within a certain distance away from the wall. Closest to the wall are pairs of

counter-rotating vortices (see Figure 2.3). In between these vortices is a structure known as a streak, which is an area of increased shear stress caused by the vortices transporting high momentum fluid down to the surface. In dimensionless units, length scales are typically on the order of $x^+=1000$. Height scales are typically on the order of $y^+=30$. Spanwise scales are typically on the order of $z^+=100$.



(a)



(b)

Figure 2.3 (a) Perspective view of counter-rotating vortices bringing down high momentum fluid, and (b) Cross section view.

To convert between the above dimensionless units to lengths, other flow information such as Reynolds number and flow velocity must be known. Inside the

UCLA wind tunnel,the Reynolds number is calculated using the values of L=1.27 cm (half height of the tunnel) and υ=14.5 x 10^{-6} m^2/s. Therefore, Re can be caculated using the following equation:

$$\mathrm{Re}=876U \tag{2.6}$$

At the wall, there is a quantity called the wall friction velocity, which is defined as follows:

$$u_\tau = 0.1079(\mathrm{Re})^{-0.089}U \tag{2.7}$$

Then, the length scales (y^+, x^+, z^+) are defined as follows:

$$y^+ = \frac{\upsilon}{u_\tau} \tag{2.8}$$

The time scale, τ^*, is defined as follows:

$$\tau^* = \frac{y^+}{u_\tau} \tag{2.9}$$

The following table sumarizes the characteristic values for typical flow velocities inside the UCLA wind tunnel.

U (m/s)	Re	y^+	τ^*
10	8760	30 μm	62 μs
20	17520	16 μm	17 μs
30	26280	11 μm	8.4 μs
40	35040	8.4 μm	4.95 μs

Table 2.1 Non-dimensionalized structure lengths and times for experimentally achievable flow velocities.

Using these values and the estimates shown in Figure 2.3, we can see that typical lengths range from 8.4 mm - 30 mm, typical heights range from 240 μm – 9 mm, and typical spacings range from 840 μm – 3 mm. All of these numbers, particularly the numbers at the higher Re, fall well within the size range where micromachining can be used to create devices to interact with the flow structures.

One note must be mentioned here regarding the validity of the work described within this thesis as it applies to real-life situations on airplane wings. All of the fluid mechanics experiments were done in a wind-tunnel with velocities on the order of 25 m/s. The length scale of this wind tunnel is the half height, which is on the order of a few centimeters. If we now look at an airplane flying at cruising altitude, the Reynolds number is much larger (typically $Re=10^6$) because the velocity is an order of magnitude larger, which more than compensates for a reduction in air density (at cruising altitudes) by a factor of approximately three. Therefore, to ensure a flow similarity, the length scale must be smaller than a centimeter. In other words, the work done here is similar to the flow very close to the leading edge of an airplane wing.

Form drag and viscous drag are not entirely independent of one another. To reduce form drag, the flow about an object should separate as late as possible. It is known that turbulent flow about an object separates later than laminar flow does. To delay the onset of separation, turbulent flow is often purposely created on airplane wings by the use of vortex generators. Obviously, the skin friction drag due to the turbulent flow results in an increased viscous drag.

When attempting to reduce drag, then, it is necessary to take into consideration the object moving through the flow. For example, a golf ball has a dimpled surface. This

surface induces turbulence which moves the separation point further back, reducing pressure drag. The golf ball also has a very small surface area. Therefore, any increase in viscous drag forces is minimal compared to the savings in pressure drag reduction. A different example is an airplane wing, which is already nominally designed so that separation occurs as late as possible. Therefore, the drag component which can be most affected is the surface drag.

There are many other sources of drag on an airplane that are mentioned here primarily for completeness' sake. Induced drag can be highly significant and results from the generation of trailing vortex systems downstream of the airplane. Cooling drag accounts for the drag due to loss of air through cooling systems of the airplane. Interference drag is a result of the fact that the drag of two bodies in close proximity to each other can be greater than if the two bodies were separate.

2.2 Drag Reduction

2.2.1 Passive Boundary Layer Control

Passive boundary layer control has been the subject of numerous scientific works. Although passive control can be effective, the primary reason for studying passive, rather than active, control is one of simplicity - it is far easier to construct passive elements and test them in a flow than it is to devise an active control system.

One of the primary methods for passive control involves the use of small riblets on a surface [3-8]. At the expense of increasing the effective surface area (or "wetted" area), they are used primarily to prevent the streamwise vortices from bringing high momentum fluid down to the surface. These riblets are generally placed in a direction parallel to the flow and can have various cross sections, with the most popular one being

a V-groove. In the case of the V-groove, it is thought that the high momentum fluid being brought down cannot penetrate into the grooves, provided the groove size is on the order of the vortices' size. If the integration of this lower shear stress over the larger wetted area is less than the integration of the standard shear stress over a flat plate, drag reduction is the result. In fact, drag reductions of 4% [5] to 8% [7] have been achieved in a laboratory environment.

As previously alluded to, there exists an analogy in nature to the use of riblets for drag reduction. Biologists have known for over a century that the scales of certain sharks have (and have had for over 100 million years, as evidenced by fossils) tiny longitudinal ribs on their surface. However, it was not until the late 1960's that scientists began to speculate as to the fluid dynamical significance of these riblets. It was then noticed that the different shark species that have these ribs all are fast swimmers, while the shark species that do not have these small ribs are slow swimmers [3].

2.2.2 Active Boundary Layer Control

While passive drag reduction appears to be promising, it also has some drawbacks. Because the vortices occur randomly in space and time, any passive system with no sensors or feedback must necessarily be turned "on" at all times. As an example, riblets prevent high momentum fluid from being transferred down to the surface when counter-rotating vortices pass over them. However, riblets also result in a larger effective area over which drag can occur. The presence of riblets, then, actually increases drag when not reducing it. Also, should flow parameters, and hence the structure size change, fixed size riblets may not be effective. Such concerns lead one to consider an active

control system. Two questions exist, then, about such a system: (1) how effective would such it be, and (2) what would be the requirements?

The question of effectiveness is partially answered by Choi [9], who has shown, through numerical simulation, the effectiveness of various methods of active control on the reduction of drag in turbulent boundary layers. Some of the techniques described within that work showed drag reductions as large as 25%. Perhaps even more astonishing is the result that by controlling only 5% of the surface, drag reductions of 15% can still be obtained. This incredible result is a direct consequence of using an active control scheme to determine the locations where actuation would be most effective and the subsequent control of the most significant structures.

One important point must be noted about Choi's work. Due to the fact that the work was numerical simulation, the assumptions were that "sensors" could be placed anywhere in the flow, and actuation methods included the ability to exactly control all three velocity components (i.e. by blowing or suction at distinct locations). Most of these assumptions are not experimentally implementable. However, there was one practical set of simulations done with the sensors being placed only at the wall. Regardless of the methods of simulation, the important point to be drawn here is the fact that the structures contributing to high shear stress can be controlled in an active manner with positive results.

For testing the ideas related to an active control system, all fluid mechanics experiments for this project were conducted at UCLA. As was shown in Table 1.1, under typical operating conditions, the lifetime of the streaks is on the order of milliseconds, and the size is on the order of millimeters. In addition, these structures appear randomly

in time and space. Therefore, a fully integrated real-time MEMS would require, over a standard 100 mm wafer, thousands of millimeter sized actuators and tens of thousands of sub-millimeter sized sensors all connected by integrated circuitry.

2.2.3 Blowing and Suction

Finally, the concept of blowing and suction must be discussed. Blowing and suction refer to small, discrete holes in a control surface through which air is passed. These direct jets of air serve to affect the boundary layer either by bringing high momentum fluid to the surface or forcing high momentum fluid away. The idea of bringing high momentum fluid close to the surface seems counterintuitive because of the increase in shear stress, but in this case, the purpose is to stabilize the turbulent structures and prevent them from lifting off of the surface. The absence of this lifting prevents new structures from forming. [9]

Often, blowing and suction experiments are not considered active control because the system does not include sensors and a feedback system nor is it set-up for individual hole, rather than array, control. Theoretically speaking, however, blowing and suction could be a part of an active control scheme, as described in Choi's work.

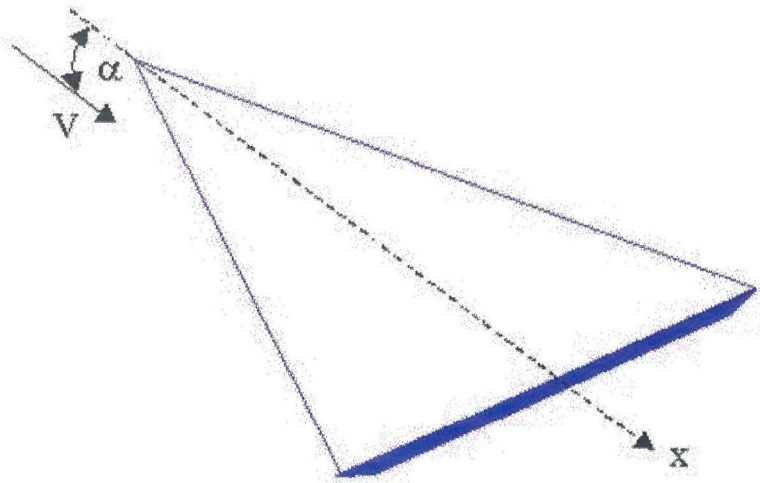
Blowing and suction apparatus on an airplane would most likely divert on-rushing air for use in control. It remains to be seen whether this diversion would lead to a drag increase far greater than any reduction resulting from the active control.

As a final note, blowing and suction can also be used for other types of aircraft control. As an example, experiments have been conducted using a combination of both blowing and suction along the leading edge of a delta wing aircraft (described below) for controlling the breakdown of vortices [10].

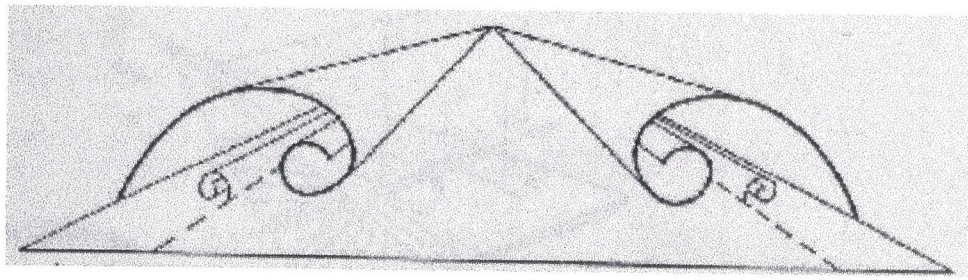
2.3 Rolling Moment Control of Delta Wing Aircraft

A delta wing aircraft is roughly shaped like a triangle and does not have traditional wings sticking out from the fuselage. (See Figure 2.4) Military aircraft are required to be able to maneuver over large ranges of angles of attack. Therefore, delta wing aircraft designs are often used because they can generate significant lift despite a high angle of attack. For example, a 70^{o} swept delta wing can increase lift until an attack angle of approximately 40^{o}, compared with a two-dimensional steady NACA 0012 airfoil, which will reach its maximum lift at approximately a 15^{o} angle of attack [11].

In a traditional aircraft, lift is generated by the fact that air flows faster over the top of the wings, resulting in a lower pressure on top of the wing than on the bottom. In a delta wing, a significant amount of lift, approximately 40% [12] is generated by the creation of large counter-rotating vortices above the aircraft. These vortices are generated due to separation, which occurs near the nose of the delta wing, and are convected downstream by the velocity component along the leading edges. If all geometric conditions are symmetric, the vortices generated on opposite sides of the wing rotate in opposite directions, but all other characteristics such as size are the exact same. If the geometric conditions can be made slightly asymmetric, the separation locations can change or the vorticity of the vortices can change, creating an imbalance in the lifting forces on each side of the wing. This imbalance creates a rolling moment in the wing, and, if large enough, can cause the plane to roll completely.



(a)



(b)

Figure 2.4 Drawings showing (a) a schematic of a delta wing where V is the velocity of the air flow and α is the angle of attack, and (b) representation of the leading edge vortices when looked at from the rear of the delta wing.

At the separation point, the vortices are extremely small and, being amplified in the flow, only grow larger as they progress down the length of the wing. Theoretically, an epsilon sized disturbance, placed at the correct location, could cause the separation point to move slightly.

2.4 Shear Stress Sensor

For reference sake, a short description of the hot-film shear stress sensor is included here. For a complete treatment, the reader is referred to the following works [13-16]. In essence the hot-film sensor consists of a polysilicon resistor (the heating element) sitting atop a thin nitride diaphragm. The nitride diaphragm is distanced from the silicon substrate by a near vacuum (on the order of tenths of a Torr) cavity. The resistor heats up when a current is applied, and the heat transfer from the resistor to the air flow can be related to the surface shear stress. The approximately 1 μm deep vacuum cavity is essential for thermal isolation, to ensure that the majority of the heat transfer is to the flow, not to the substrate. The presence of this vacuum cavity results in a device sensitivity an order of magnitude higher than for hot-film sensors sitting directly on a substrate. A scanning electron micrograph of a typical shear stress sensor is shown in Figure 2.5.

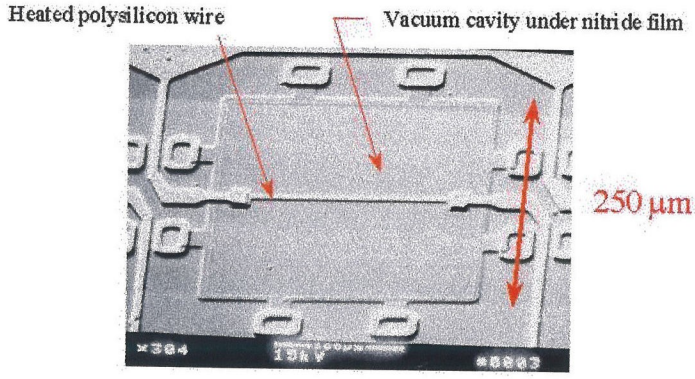


Figure 2.5 SEM of a shear stress sensor.

For reasons that will become apparent in Chapter 5, a different design of shear stress sensor was also designed and fabricated by Fukang Jiang and Xuan-Qi Wang. The operating principle is essentially the same, but the device structure is slightly different. A

picture of such a device is shown in Figure 2.6. This IC-compatible structure uses the gate polysilicon as the heating element. The diaphragm material comprises of gate oxide, interlayer dielectric oxide, and a thick layer of Parylene. Instead of sitting atop a vacuum cavity, this sensor sits atop a deep (approximately 20-30 μm) air cavity that was created by the BrF_3 etching of silicon.

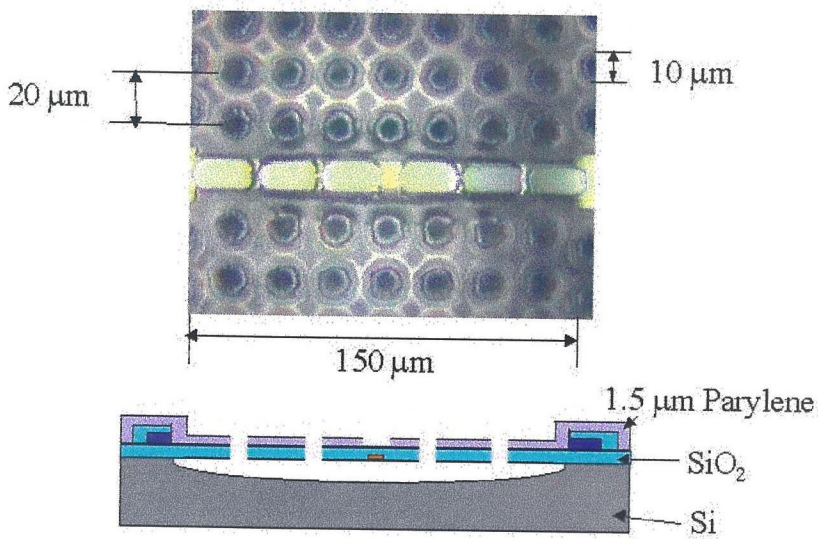


Figure 2.6 Picture (top) and cross section (bottom) of the shear stress sensor developed for use in an IC-compatible process.

References

[1] Bushnell, D., Hefner, J., and Ash, R., "Effects of Compliant Wall Motion on Turbulent Boundary Layers," *Physical Fluids*, Vol. 10, S31-48, 1977.

[2] Cantwell, B., "Organized Motion in Turbulent Flow," *Annual Review Fluid Mechanics*, Vol. 13, pp. 457-515, 1981.

[3] Bechert, D. and Bartenwerfer, M., "The Viscous Flow on Surfaces with Longitudinal Ribs," *Journal of Fluid Mechanics, Vol. 206*, pp. 105-129, 1989.

[4] Lazos, B. and Wilkinson, S., "Turbulent Viscous Drag Reduction with Thin-Element Riblets," *AIAA Journal*, Vol. 26 (4), pp. 496-498, 1988.

[5] Park, S.-R. and Wallace, J., "Flow Alteration and Drag Reduction by Riblets in a Turbulent Boundary Layer," *AIAA Journal*, Vol. 32 (1), pp. 31-38, 1994.

[6] Vukoslavcevic, P., Wallace, J., and Balint, J.-L., "Viscous Drag Reduction Using Streamwise-Aligned Riblets," *AIAA Journal*, Vol. 30 (4), pp. 1119-1122, 1992.

[7] Walsh, M., "Riblets as a Viscous Drag Reduction Technique," *AIAA Journal*, Vol. 21 (4), pp. 485-486, 1983.

[8] Sirovich, L. and Karisson, S., "Turbulent Drag Reduction by Passive Mechanisms," *Nature*, vol. 388, pp. 753-755, 1997.

[9] Choi, H., Moin, P., and Kim, J., "Active turbulence conrol for drag reduction in wall-bounded flows," *Journal of Fluid Mechanics*, Vol. 262, pp. 75-110, 1994.

[10] Gu, W., Robinson, O., and Rockwell, D., "Control of Vortices on a Delta Wing by Leading-Edge Injection," *AIAA Journal*, Vol. 31 (7), pp. 1177-1186, 1993.

[11] Lee, M. and Ho, C.-M., "Life Force of Delta Wings," *Applied Mechanics Review*, Vol. 43 (9), pp. 209-221, 1990.

[12] Polhamus E. C., " A Concept of the Vortex Lift of Sharp-edge Delta Wings Based on a Leading-edge-suction Analogy," NASA TN D-3767, 1966.

[13] Jiang, F, *Silicon-Micromachined Flow Sensors*, Ph.D. Thesis, California Institute of Technology, Pasadena, 1998.

[14] Jiang, F. and Tai, Y.-C., "A Micromachined Polysilicon Hot-Wire Anemometer," Technical Digest, IEEE Solid-State Sensor and Actuator Workshop, Hilton Head, SC, USA, pp. 264-267, 1994.

[15] Jiang, F., Tai, Y.-C., Walsh, K., Tsao, T., Lee, G.-B., and Ho, C.-M., "A Flexible MEMS Technology and Its First Application to Shear Stress Sensor Skin," Proceedings from IEEE Micro Electro Mechanical Systems, Nagoya Castle, Japan, pp. 465-470, 1997.

[16] Liu, C., Tai, Y.C. Huang, J. and Ho, C. M., 1994, "Surface-Micromachined Thermal Shear Stress Sensor," Application of Microfabrication to Fluid Mechanics, FED-Vol. 197, pp. 9-16, ASME.

[17] McCormick, Barnes W., **Aerodynamics Aeronautics and Flight Mechanics**, John Wiley and Sons, Inc., New York, 1995.

[18] Tritton, D.J., **Physical Fluid Dynamics**, Van Nostrand Reinhold Company, New York, 1977.

Chapter 3

Magnetic Actuator

The magnetic flap actuator was proposed as a means to interact with fluid flowing above a control surface. Three distinct flaps are described here. The first is a passive surface micromachined flap. The term "passive" refers to the fact that the actuation is soley controlled by an external magnetic field. In addition, flaps in an array cannot be individually addressed. These flaps are useful in cases where a large array of flaps needs to be activated at one time, so they are not useful for the turbulent boundary layer control project. The second flap described is an active surface micromachined flap. Like the surface micromachined passive flap, it requires an external magnetic field for operation, but unlike the passive flap, it can be individually addressed by means of an on-flap coil. These flaps are more complex than the passive flaps. The final flap described is an active bulk micromachined flap. Similar to the surface micromachined flap in operation, the main difference is in the geometry. With thicker support beams, the bulk flap can be designed to be very stiff and to have a high resonant frequency which, as will be shown later, is sometimes desired.

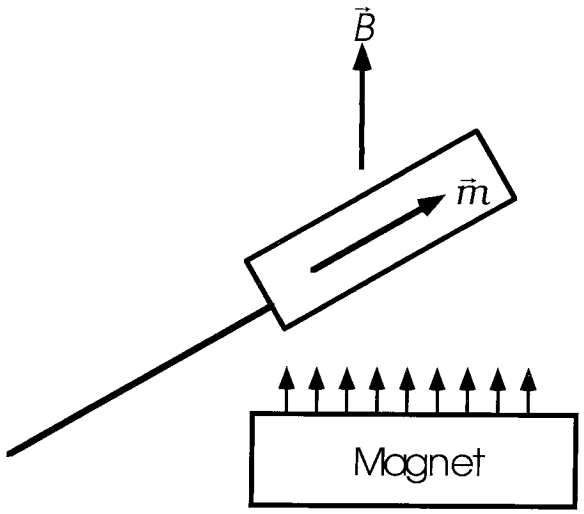
3.1 Passive Flap

For the rolling moment control project, a linear array of flaps needed to be activated simultaneously. Therefore, the less complicated passive flaps were used. Many different variations were fabricated, all in an attempt to increase device robustness.

3.1.1 Basic Operation Principles

The principles behind the operation of the permalloy flap have been well explained [1,2,17,18]. Essentially, the external field sets up a magnetization within a

permalloy thin film, which then rotates in an attempt to align itself with the magnetic field. The operating principles in both uniform and non-uniform fields are well explained in [1,15].



$$\vec{\tau} = \vec{m} \times \vec{B}$$

$$|\vec{\tau}| = M_s v H Cos\theta$$

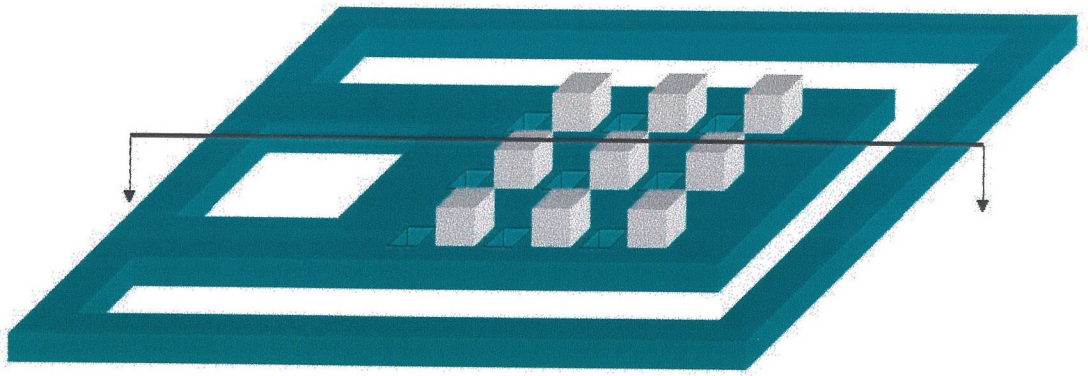
Figure 3.1 Operational principle behind the permalloy flap.

Permalloy is a soft magnetic material consisting of 80% nickel and 20% iron ($Ni_{80}Fe_{20}$). The fact that it is considered soft versus hard indicates that large magnetizations can be produced by relatively small external magnetic fields. Also, when the external field is removed, a soft magnetic material does not retain much magnetization. One result of these two properties is the fact that, during the electroplating process, it is not crucial to align the flaps in a preferred direction. Any external field present during usage will align the fields. If a hard magnetic material were being plated, the final magnetic properties would be highly dependant upon the wafer orientation during plating.

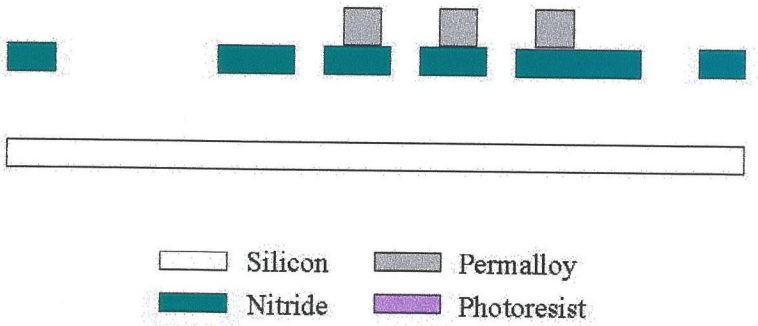
3.1.2 Design

Initial permalloy flaps used bending beams and were fabricated in a surface micromachined process using PSG as the sacrificial layer and polysilicon as the structural layer. Polysilicon was chosen due to the extremely high etching selectivity (infinite for all practical purposes) that HF acid has for etching oxide over polysilicon. Therefore, the actual device dimensions should be exactly as designed.

However, when tested in the wind tunnel these flaps were easily fractured by wind loading. Work has been done that showed the fracture strain of LPCVD polysilicon films to be approximately 0.93% [3]. More recent work [4] has shown that the fracture strain of silicon nitride can be as high as 6%. Therefore, in at attempt to improve device robustness, it was decided to use silicon nitride instead of polysilicon as the structural layer due to nitride's greater fracture strength.



(a)



(b)

Figure 3.2 (a) Perspective and (b) cross section views of a permalloy flap.

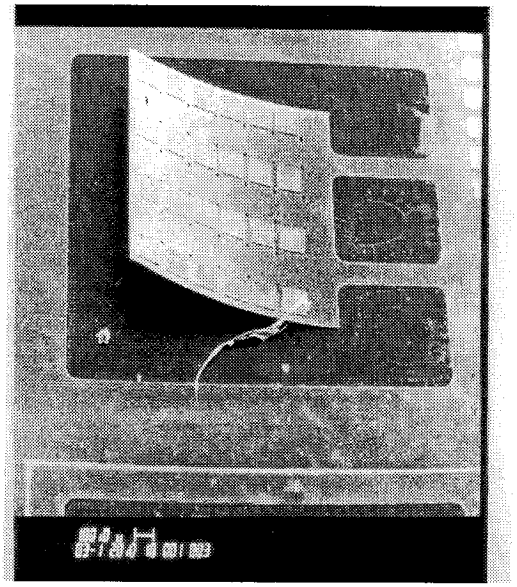


Figure 3.3 SEM photo of passive flap with bending beams.

When considering the extreme case of a flap aligning itself perfectly vertical, the support of the bending beams is placed under incredible stress. By the nature of the process, the support should be fixed, which implies that the boundary conditions are zero displacement and zero slope. When the flap is rotated to the vertical position, the beams are necessarily forced to also rotate to the perpendicular and the bottom side of the beam is strained beyond the fracture limit and breaks. However, if the flap is supported by torsional beams, rotation of the flap to the vertical position only implies that the end of the beams must rotate by 90^{o}. By distributing this rotation evenly over an appropriately long beam, the support boundary condition can still hold. Therefore, in addition to using nitride as the structural layer, a different geometry using torsional beams was also fabricated to take advantage of this fact.

Etch holes were spaced equidistantly to ensure that all portions of the flap (with the obvious exception of the beams) would be released in as simultaneous and symmetric

a manner as possible. The edges are released first due to the fact that the reactants can enter and the products can leave as a line of solution as opposed to a point source.

The permalloy was placed in discrete pieces upon the flap rather than as one large continuous piece. This was done for stress release purposes. Many individual pieces can cause a local stress in the underlying film, but one large piece can cause the stress to be distributed over the entire flap, perhaps causing global bending. In addition, should an individual piece be subject to a shear (for whatever reason) large enough to cause it to delaminate from the surface, it wouldn't result in other pieces delaminating as well (as would be the case for one large piece of permalloy).

3.1.3 Fabrication

The process for creating the passive permalloy flaps (nitride structural layer) is shown in Figure 3.4 and described in the following steps.

(1) Deposit 5.0 μm PSG for use as the sacrificial layer.

(2) Deposit and pattern 1.2 μm silicon nitride for use as the structural layer.

(3) Evaporate chrome/copper seed layer (0.01 μm and 0.1 μm respectively).

(4) Spin and pattern AZ 4400 photoresist for use as the electroplating mold. Also pattern an electrical contact ring surrounding the wafer.

(5) Electroplate permalloy to the desired height (usually 5.0 μm).

(6) Dissolve photoresist and etch the seed layer.

(7) Dice the wafer into strips of flaps and submerse in high concentrated (49%) HF acid for release.

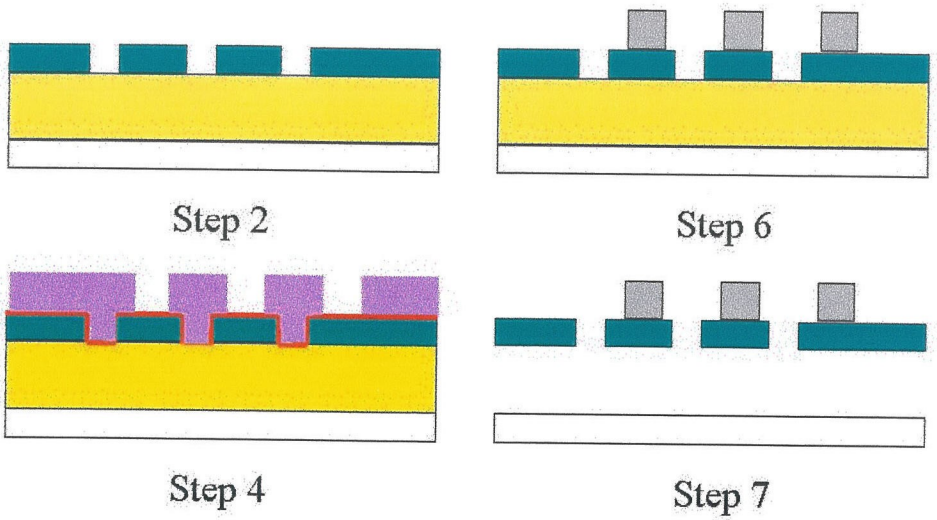


Figure 3.4 Process flow for permalloy actuator. Legend is as shown in Figure 3.2.

Typically, thin chrome layers can be easily etched in less than a minute using solutions such as dilute (10%) HCl or commercially available chrome mask etchants (such as Cyantek's Cr-7 Mask Etchant and Transene's Chrome Mask Etchant). Care must be taken when using the mask etchants, as they can undercut the chrome very quickly and can also attack the permalloy film.

After the electroplating process, the chrome adhesion layer often could not be etched. The electroplating is believed to affect the adhesion layer, perhaps creating a passivation layer between the chrome and copper which is highly resistant to etching. Sometimes, using a more highly concentrated (33%) HCl solution would eventually etch the chrome (in approximately 5 minutes), but this is not always the case.

One potential solution is to simply touch some of the exposed chrome (while in the HCl solution) with a piece of aluminum. The chrome in the touched area would then immediately etch, and an etch front would quickly spread across the wafer. A variation of this method involves the use of an external potential applied between the wafer and the

solution, using an aluminum rod as one electrode and the wafer as the other. This often became a frustrating process due to the fact that wafers would not etch consistently, and etch fronts would often cease to spread. An even worse occurrence would be the complete etching of the chrome except in recesses, where, by definition, the sacrificial layer is exposed. It is speculated that the thin chrome (on the order of 0.01 μm) deposited during the evaporation could not adequately cover the approximately 1.0 μm step height. Therefore, the chrome at the bottom of the steps is electrically isolated and is not affected by the electrochemical etching.

There is reason to believe that the chemically altered chrome is not a continuous layer. However, during the release process, if the quasi-chrome layer exists on the surface, the acid will eventually cause pieces of the chrome to lift-off of the surface, thus allowing the undercutting to begin. The small lumps of chrome which lift-off are not dissolved away in the solution, further implying that the chrome surface is porous enough to allow acid to undercut it.

The simple solution is to sputter etch the chrome. However, this option was not available at the time. Another option involves the use of titanium as the adhesion layer rather than chrome. Titanium is easily etched using an extremely dilute (0.5% or 1%) HF acid solution. Therein lies the main difficulty in using titanium as the adhesion layer when an oxide is used as the sacrificial layer. During the HF acid sacrificial layer etch, the titanium is easily undercut, thereby lifting off the permalloy. To circumvent this problem, a different variation of the flap was fabricated using polysilicon as the sacrificial layer, silicon nitride as the structural layer, and silicon etchants as the release agents. This technique is attractive because the silicon etchants used (TMAH and BrF_3)

do not attack titanium, and the dilute HF acid used to etch the titanium does not affect any other material. For specifics regarding the polysilicon sacrificial layer, and for more commentary on choice of sacrificial layer, see Chapter 5.

Typically, high concentrated HF acid can free the structure by undercutting approximately 100 μm of 5 μm thick PSG is roughly 10 minutes. If there still exists the un-etchable chrome layer, this time may be increased to approximately 13-14 minutes. During this time, approximately 0.06 μm of silicon nitride is also etched away from all sides of the flap for a total thickness decrease in the silicon nitride thickness of approximately 0.12 μm, or roughly 10% of the original thickness. This loss also affects the footprint of the flap, but to a much smaller percentage. Perhaps more detrimental to overall robustness of the flap is the fact that HF acid, when etching can either slightly roughen the surface or even create small notches. These surface imperfections lead to an increase in the possibility of fracture (See Appendix B).

Stiction [5-9] is a big problem associated with the release of surface micromachined structures. If, after the flaps are extracted from the HF acid, they are rinsed in water and air dried, they will all stick to the substrate and be unusable. A better solution is to rinse the flaps in alcohol before drying. In principle, this should improve the yield due to the fact that alcohol has a lower surface tension force than water does. However, the result is still a very low yield.

One solution developed to counter this problem uses the fact that the flap has permalloy on its surface and can be activated by an external magnetic field, lifting it up. After rinsing in water and alcohol, the strips of flaps are placed on an active electromagnet, thereby raising the flaps. By creating a force that resists the surface

tension forces, the alcohol evaporates without bringing the flap down to the surface. Therefore the stiction that is created by the release process is prevented.

It should be noted that in addition to release-inducted stiction, there also exists in-use stiction, which occurs when, during usage of the device, the structure comes close to or contacts the surface and sticks due to electrostatic, van der Waals, or other forces. Sometimes, during the magnetic release process, the flaps would be bent back far enough to slightly deform the beams plastically, leaving the flaps at a slight angle above horizontal. While this deflection may be undesirable in terms of affecting flows, it does serve one useful purpose in the fact that it greatly reduces the chances of in-use stiction occuring due to the fact that there is some finite energy stored in the beam which stiction forces need to overcome.

3.1.4 Testing

As previously described, these flaps were used in a rolling moment control project on a delta wing aircraft. Experiments were conducted in the UCLA wind tunnel to investigate the aerodynamic performance of the delta-wing model that was used for flow control. The wing model has a sweep angle of 56.5° and rounded leading edges. The delta wing, at different angles of attack, was mounted on a six-component force/moment transducer that recorded forces and moments in all three axes.

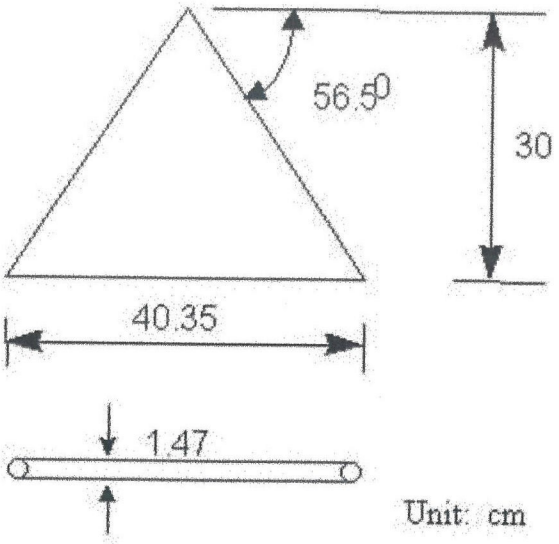


Figure 3.5 Schematic representation of the test delta wing. [10]

A linear array of actuators was placed at one of the rounded leading edges (Figure 3.6). The maximum torque (normalized by the torque generated by a single vortex) can be more than 15% for 1 mm actuators and 40 % for 2 mm actuators at $\alpha = 25°$ [10].

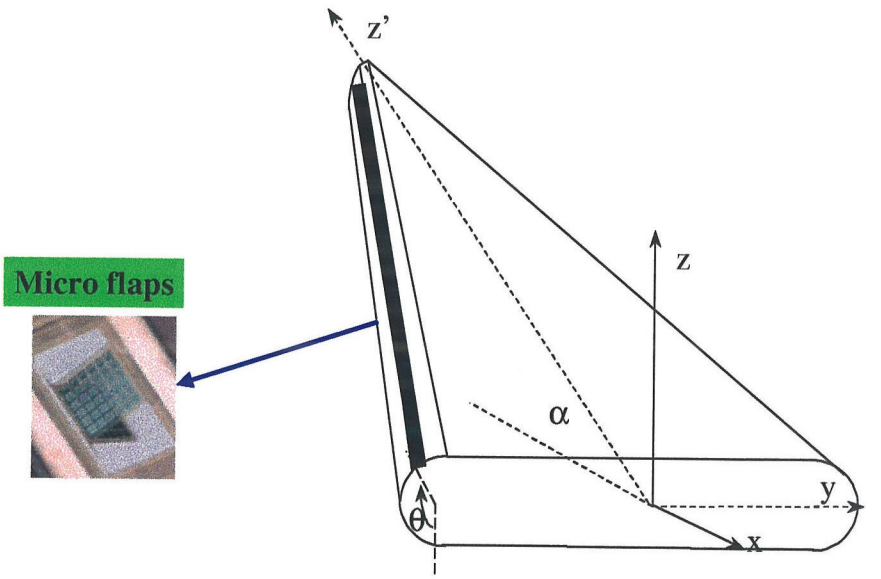


Figure 3.6 Schematic showing how the actuators were mounted on the delta wing. [10]

For large values of θ, the normalized rolling moment becomes negative in most cases. At high α, the negative value is fairly small. At $\alpha = 25^{\circ}$, the negative torque can be as large as –30 % for 2 mm actuators. This negative rolling moment can be employed to obtain an additional rolling moment if the actuators are mounted at the proper position at the opposite leading edge. For example, if the actuators are placed at $\theta = 50^{\circ}$ along one leading edge and at $\theta = 80^{\circ}$ along the other leading edge, the maximum torque increase reaches a staggering level of 70% (Figure 3.7). Since the linear summation of the individual incremental torques is equal to the value obtained by the two-sided actuation, the responses of the two vortices to the actuators seem to be independent.

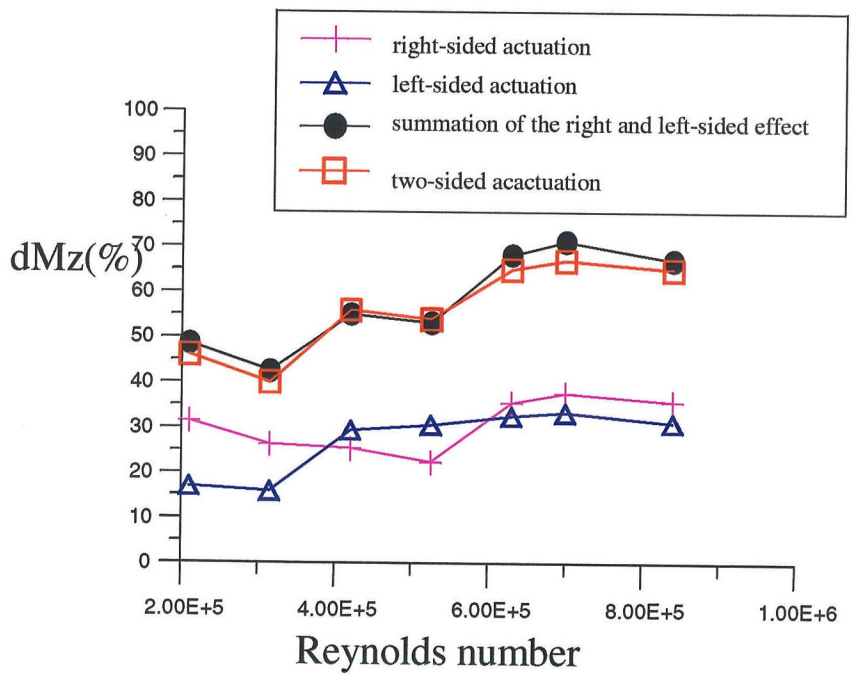
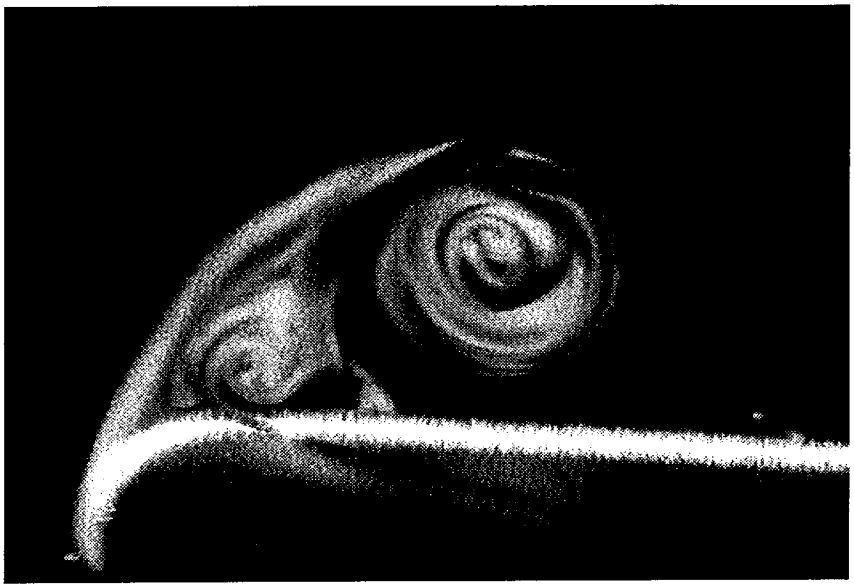
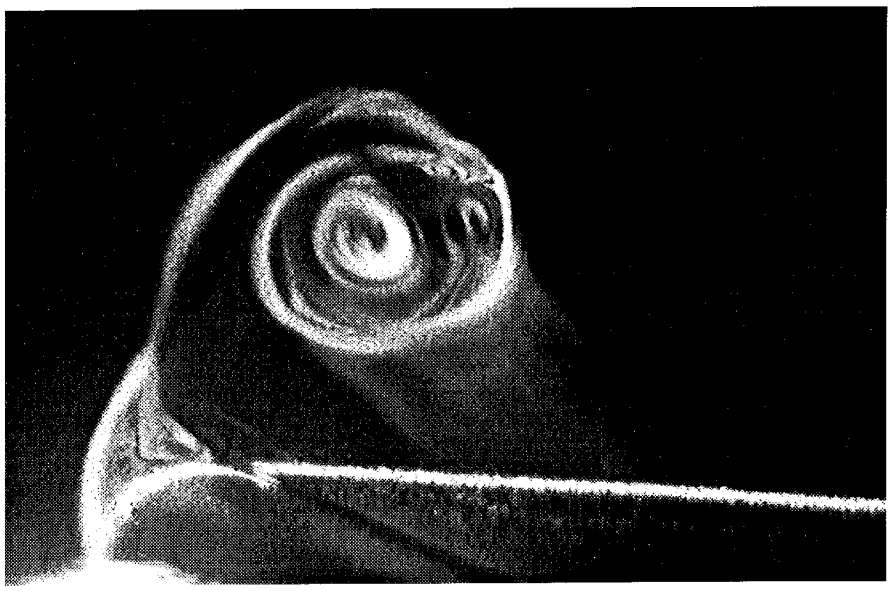


Figure 3.7 Graph showing results of combining torques from both sides of the delta wing. Note that the maximum change in rolling moment is equal in value to 70% of the value generated by one single vortex. [10]

Some excellent flow visualization work was done by Lee [10-13] and portions of it are presented here as an aid to help understand the effect of the actuators on the flow.



(a)



(b)

Figure 3.8 Vortical action (a) before and (b) after actuation at $\theta=50^{o}$ [10].

Here, for completeness sake, we mention an experiment done with a $1/7^{th}$ scaled model of a Mirage III radio controlled aircraft to further validate the concept of using micro devices to create macro effects. Due to the larger wind loading experienced during flight tests, a small strip-type mechanical actuator was used in place of the

micromachined actuators that were used in the wind tunnel tests. The maximum speed of the aircraft was 150 knots and the test speed for the maneuvering cases was about 50 knots. Many fighter aircraft maneuvering modes were achieved. It was found that swift rolls and small-radius turns can be accomplished at high angles of attack. In a high-G barrel roll mode, the aircraft was able to achieve an impressive 180° roll in 0.8 seconds.

3.1.5 Processing Variations

As previously mentioned, two of the primary problems with the flap process were the etching of the chrome adhesion layer and the stiction involved with the wet release process. The combined use of a titanium adhesion layer with a polysilicon sacrificial layer can potentially solve both problems.

As previously mentioned, titanium is easily etched using HF acid but cannot be used in an oxide sacrificial layer system for the same reason. In a polysilicon sacrificial layer system, titanium can be used. By using polysilicon as the sacrificial layer, the possibility of using a dry release agent, BrF_3, suddenly becomes viable. By using a dry release agent, there exists no possibility of release stiction due to surface tension forces.

One drawback of using BrF_3 is the fact that it attacks silicon nitride at a non-negligible rate. Although bulk tests have shown an etch selectivity as high as several hundred [14], for undercutting purposes, this selectivity can drop significantly. One potential solution to this problem is to use a wet etchant such as TMAH (more on this process will be discussed in Chapter 5), which has an extremely high selectivity, to etch a portion of the polysilicon and then follow with the dry etch. Another possibility is to heavily dope the polysilicon n-type to enhance the etch rate (and hence reduce the time the silicon nitride is exposed to the BrF_3). The principle behind this is opposite to that of

a boron etch stop layer, in which the high concentration of holes impedes the chemical etching process.

3.2 Active Flaps

Many generations of active flaps were designed and fabricated for use in the drag reduction project. The first generation was fabricated as a proof of concept and to validate our ideas that the use of magnetic forces was a viable means to actuate micromachined devices. Later generations were designed and fabricated to provide more robust devices (here, the term robust implies both actual device robustness as well as process robustness) with the eventual goal of integration in mind. Sometimes, a new generation would be fabricated because our fluid mechanics partners believed that a particular geometry would prove to be advantageous. As with passive flaps, both bending and torsional active flaps were designed and fabricated in many generations.

3.2.1 Basic Operation Principles

Once again, excellent precise descriptions of the magnetic performance of the active type flaps is presented by Liu [1] and Miller [2]. The flaps have a series of planar coils patterned on top of them. When a current flows in these coils, a magnetic moment is generated.

$$\vec{m} = \sum_{n} i\vec{A}_{n} \tag{3.1}$$

This moment is the summation of the product of the current passing through each coil and the area encompassed by the coil. The vector nature of the magnetic moment reflects the use of the electromagnet "right-hand rule".

For flaps with torsional beams, it is assumed that the torque generated on the flap is directly transferred to the beams. In other words, the mechanical energy produced is 100% transferred into strain energy of the beams, with none of it serving to deform the plate. Assuming a uniform magnetic field, the torque experienced is simply a cross product of the on-flap magnetic moment and the external magnetic field.

$$\vec{\tau} = \vec{m} \times \vec{B} \tag{3.2}$$

In a non-uniform magnetic field, a force is also generated.

$$\vec{F} = (\vec{m} \cdot \vec{\nabla})\vec{B} \tag{3.3}$$

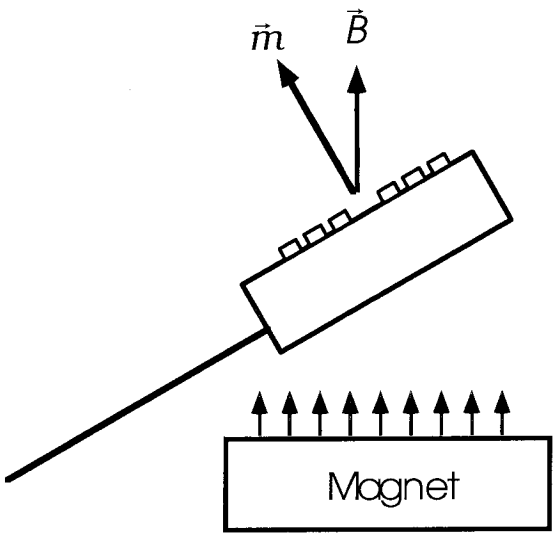


Figure 3.9 Drawing showing operational principles behind the active flap.

3.2.2 Overview of Different Generations

The first generation [16] flap consisted of three layers. A structural layer of silicon nitride was sandwiched between a top layer of metal coils and a bottom layer of

heavily phosphorous doped polysilicon. The polysilicon was used as a return path for the current. Several different geometries were fabricated. Plates of different sizes and shapes were designed. Some flaps had four corner supports instead of two bending beams. Different etch hole sizes and placements were also designed.

Unfortunately, the use of polysilicon as the return path for current was the root of two major problems: bending due to stress and excessive plate heating. The mismatch of the stresses of the polysilicon and silicon nitride films caused the flaps to curl up considerably. In general, this would cause problems in the sense that there would be a flow disturbance at all times. For the first generation, however, this was actually a serendipitous event because the stress served to provide a force to oppose the stiction forces. Without the natural bending, the flaps would have all stuck to the substrate.

Polysilicon has a resistivity that is orders of magnitude higher than that of metals. In fact, in IC processes, polysilicon is used as a resistor material. For our purposes, having a high resistance is detrimental. First of all, keeping in mind that the eventual goal is a distributed system with thousands of devices, the power consumption for each actuator must be minimized. Second, the heat generated induces a thermal bending of the plate that may not be desirable. Finally, the generation of heat may, in fact, affect the flow in a negative way by adding energy to an already turbulent flow.

Therefore, the second generation [19] was designed and fabricated with the intent of reducing the power consumption and also reducing the bending due to intrinsic stresses. As previously mentioned, the use of polysilicon is the cause of both problems. All second generation devices are characterized by the use of PSG as the sacrificial layer while using two layers of metal for the coils.

Typically, the IC industry uses some sort of LPCVD LTO as metal insulation. Immediately, one problem with using LTO in this process becomes apparent. Oxide is also used as the sacrificial layer. Therefore, the metal insulation would also be etched away, leaving behind air bridges. Other LPCVD options were not available to us due to temperature considerations. The low deposition temperature of LTO is important because the metal must not be melted or evaporated during the deposition. Even though the melting points of chrome and gold are significantly higher than that of aluminum, neither metal can stand LPCVD silicon nitride temperatures.

One simple alternative to using LTO as the insulation is to use photoresist, which can be spun on in a room temperature process. Photoresist is attacked by high concentrated HF acid, but is tends to break off in chunks as opposed to being dissolved away (as oxide is). Therefore, the possibility exists that while the resist integrity would be compromised, enough may remain to support the top layer of metal. Another alternative is to etch using buffered HF acid, which attacks oxide much slower than high concentrated HF acid does, but does not appreciably attack resist. Results of attempting these options are discussed below.

An alternative to LPCVD films for insulation would be PECVD films. Our lab has the capability of depositing PECVD silicon nitride at a temperature of 300^{o} C. Due to the nature of their deposition, PECVD films generally have poorer chemical resistance than LPCVD films. Therefore, PECVD nitride is appreciably etched in HF acid. However, if the flap can be freed before the insulating nitride is completely etched away, this process could also provide a viable means to insulate the coils.

The third generation of actuators was conceived with the idea of eventual process integration in mind. Up to this point, all of the actuators used PSG as the sacrificial layer. Unfortunately, as previously mentioned, IC processes typically use oxide both as inter-metal dielectrics, but also for overall circuit encapsulation (often called an overglass layer). Since the release step is usually the final step in any process involving the flap, the HF acid would horribly attack all of the oxide required for the electronics portion of the chip. Therefore, a different sacrificial layer process needed to be developed.

For reasons that will be detailed later (Chapter 4), metal was one option explored for use as a sacrificial layer. Metal can be deposited in thick layers relatively easily and at low temperatures. When using a metal sacrificial layer, care must be taken to choose a structural layer that can also be deposited at a low temperature. Therefore, using a metal sacrificial layer in conjunction with an LPCVD oxide structural layer was investigated. Two choices were considered for the sacrificial layer: aluminum and copper. Copper ended up being the most promising, and more detailed results are discussed in Chapter 4.

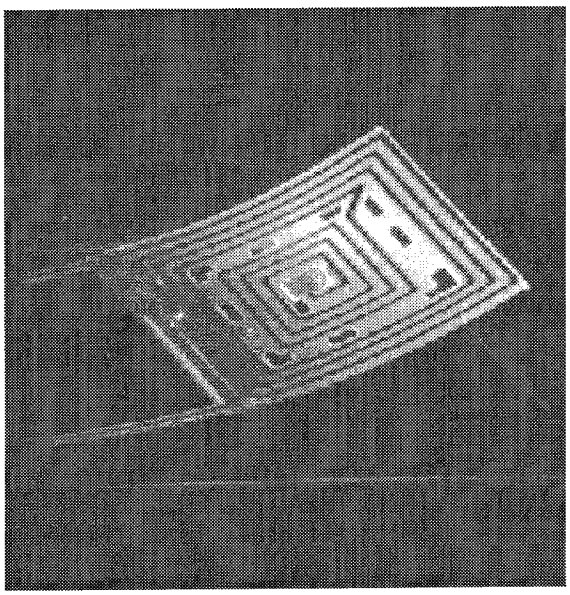
The fourth generation of flaps was designed with the ideas in mind that (a) a combined passive/active flap may prove beneficial, and (b) the passive component, permalloy, could also be used to aid in a magnetic release step. This generation of flap was designed with the idea in mind of IC process compatibility. Therefore, some versions of these flaps used polysilicon as the sacrificial layer. One particularly nice aspect of using polysilicon as the sacrificial layer is the possibility of releasing structures using a dry, gas-phase etchant. Details will be discussed below and in Chapter 5.

Finally, a fifth generation was designed. This generation is completely different from the previous four in the sense that it uses a bulk, rather than surface, micromachined

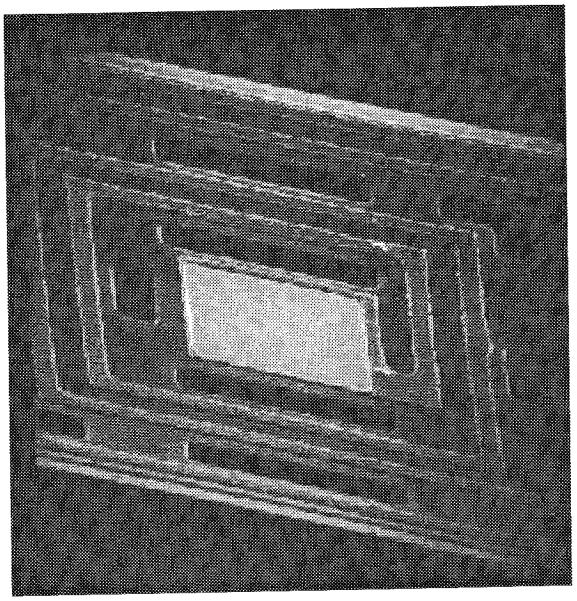
process. The primary reason for this is the desire for thicker beams, which result in a higher resonant frequency, than possible with a surface micromachined process.

3.2.3 Generation 1

A typical first generation actuator is shown in Figure 3.10. Flap sizes vary from 250 μm on to 900 μm on a side, the beam lengths vary from 100 μm to 360 μm, and the beam widths vary from 14 μm to 50 μm. Etch holes were typically 15x15 μm^2.



(a)



(b)

Figure 3.10 Picture of (a) a first generation active flap. The curvature is due to the stress levels in the silicon nitride and polysilicon films. (b) Close-up of the plate showing contact and etch hole detail.

3.2.3.1 Fabrication

The fabrication steps are well described by Liu [16]. A drawing is shown in Figure 3.11(a), and a brief process flow is shown in Figure 3.11(b) where the steps are listed as follows.

(1) Deposit 2.5 μm PSG for use as the sacrificial layer.

(2) Deposit and pattern 0.6 μm polysilicon for use as the return path.

(3) Deposit 0.5 μm PSG for use in doping.

(4) Anneal wafer at 950° C for one hour to dope polysilicon from both sides. This results in a sheet resistance of approximately 50Ω/square. Strip top PSG in BHF.

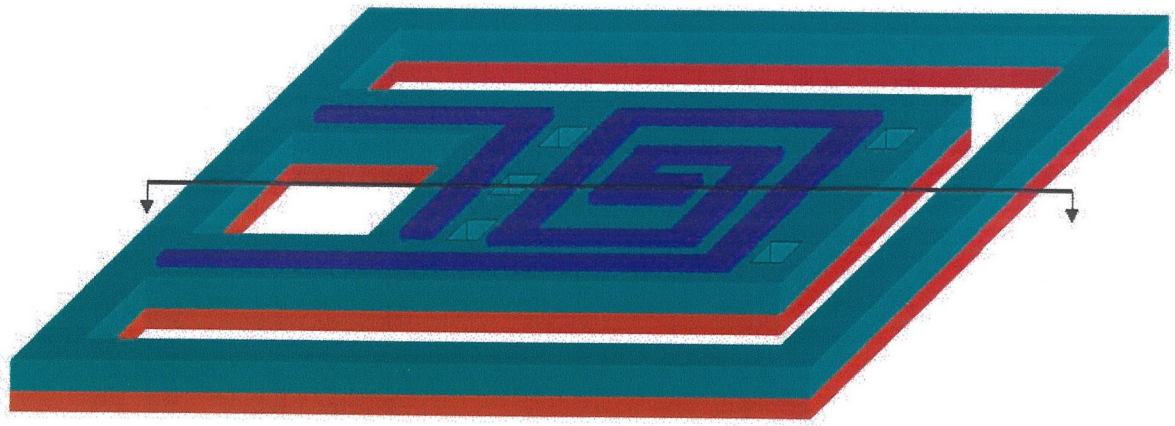
(5) Deposit and pattern 0.3 μm silicon nitride for use as the structural layer.

(6) Deposit and pattern Cr/Au (0.01 μm and 0.5 μm respectively) for use as the coils. Note that neither chrome nor gold can be etched by HF acid.

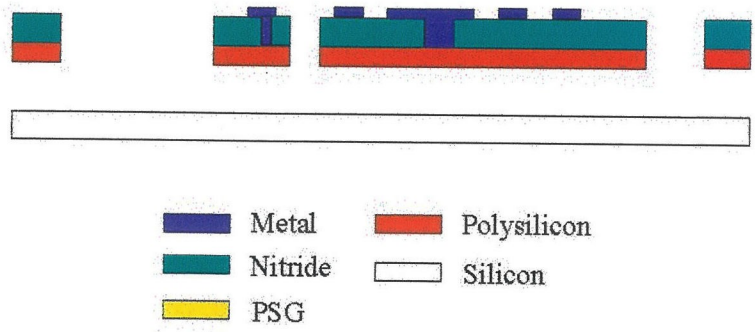
(7) Dice wafer and submerse dice into high concentrated HF acid for release.

The first generation fabricated used aluminum as the metalization, and both BHF and pad etchant were used as release agents. BHF would completely undercut a 200x200 μm^2 plate in approximately 30 minutes, while pad etchant would require approximately 3 hours to accomplish the same result. During both of these etches, the aluminum would be severely damaged, even when protected by photoresist, which, unfortunately, is designed to be resistant to these chemicals for durations typically encountered in IC processing (order of a few minutes).

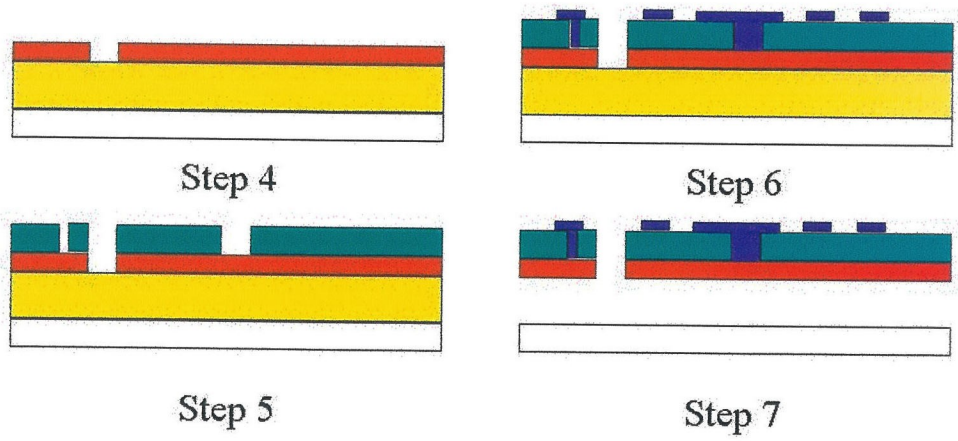
After switching from aluminum to chrome/gold metalization, the yield improved from approximately 10% to over 90%. High concentrated HF acid could release the above-described structure in approximately 4 minutes. After release, the structures needed to be dried. Several authors [5-9] have reported different drying techniques, but at the time, our standard technique included rinsing the dice in de-ionized water for 20 minutes, followed by rinsing in both acetone and isopropyl alcohol for 1 minute each. After a final 10 minute DI water rinse, the dice would be baked dry under an infrared lamp. Using this method, we were able to achieve an almost 100% yield, although stiction was occasionally observed.



(a)



(b)



(c)

Figure 3.11 (a) Perspective view of first generation active flap. (b) Cross section view. (c) Abbreviated process flow.

3.2.3.2 Testing and Results

Some of the prime concerns during testing included deflection (intrinsic, thermal, and magnetic), forces achieved, temperature, and frequency response. Some of these results have been well described by Liu [16] and Tsao [19] but are reproduced here for completeness' sake. A typical testing set-up is shown in Figure 3.12.

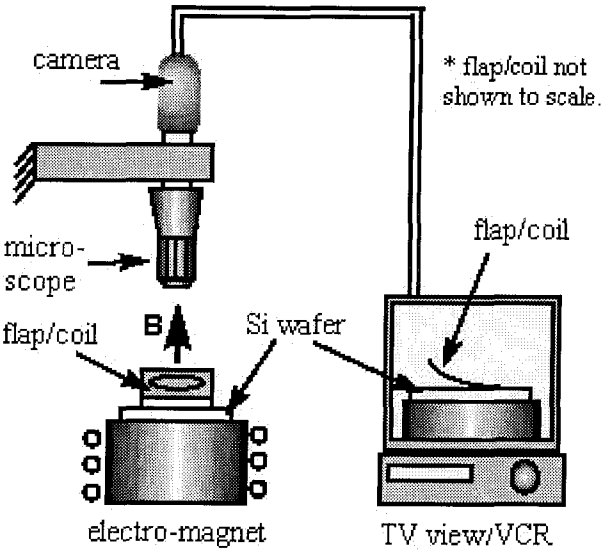


Figure 3.12 Drawing showing generic experimental set-up for actuator testing.

Most of our magnetic testing used an electromagnet with a maximum magnetic flux density in the vertical direction of 1.76 kGauss at a current of approximately 2.5 A. A calibration curve showing the magnetic flux density as a function of vertical distance is shown in Figure 3.13. Note that the only component of the gradient of the B field that can be derived from this graph is the z-component.

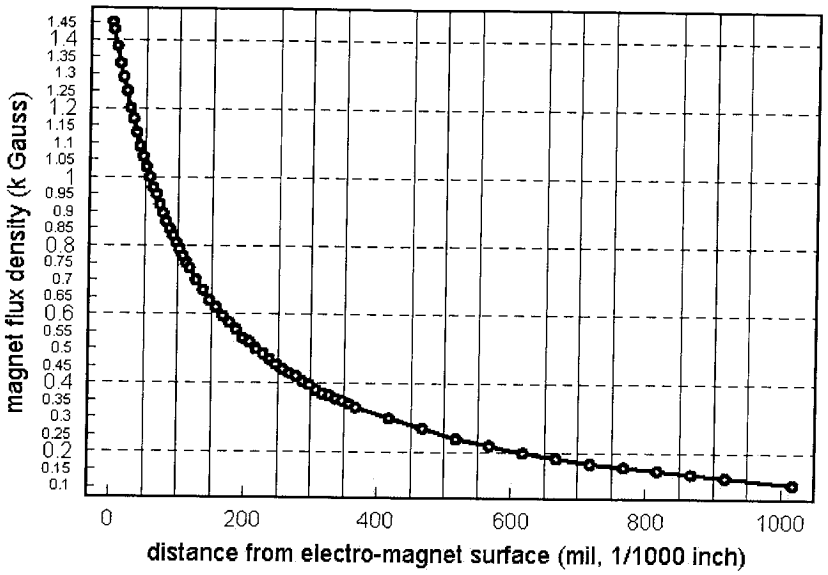


Figure 3.13 Calibrated magnetic flux density distribution along the axis of the electro-magnet core. The coil is located approximately 20 mil, or 0.5 mm, away from the magnet's surface, where the gradient of the B-field is linear

A typical first generation actuator used in the following descriptions and calculations has the following dimensions: 350 μm wide by 450 μm long, 8 turns of metal, and 20 μm wide by 300 μm long beams. Its radius of curvature at rest is 260 μm. When 30 mA is passed through its coils (80 Ω resistance), the radius of curvature increases to 460 μm, and the end of the flap deflects downward by 30 μm.

Intrinsic bending is due to the fact that two layers with different moments of intertia and different moduli of elasticity are placed together as if in a thermostat. If the combined structure bends, resulting in a non-infinite radius of curvature, the internal stress can be derived using [21]

$$\frac{Fh}{2} = \frac{E_1 I_1 + E_2 I_2}{\rho} \tag{3.4}$$

where F is the lateral force due to the intrinsic stress, h is the total height of the bi-material, E_1 and I_1 are the Young's Modulus and moment of inertia of the top layer, E_2 and I_2 are the Young's Modulus and moment of intertia of the bottom layer, and ρ is the radius of curvature. Using this formula with the above-mentioned flap, the bending moment due to internal stresses is 8.8×10^{-9} N-m and that the moment due to the thermal effect is -3.9×10^{-9} N-m.

With the first generation flap, there exists a hot spot centered around the contact hole, where the metal contacts the polysilicon. The temperature of the beams is relatively constant compared with the change of temperature of the plate. Therefore, when the actuator is heated, it is assumed that the vast majority of the bending is experienced by the flap, rather than the beams. The change in the radius of curvature due to heating can also be roughly estimated (or conversely, knowing the change in radius of curvature, the temperature can be estimated) using the following formula [21]

$$\rho = \frac{2}{3}\frac{\frac{7(t_1+t_2)^2}{4} - 2t_1t_2 + \frac{E_1b_1t_1^3}{E_2b_2t_2} + \frac{E_2b_2t_2^3}{E_1b_1t_1}}{\Delta\alpha\Delta T(t_1+t_2)} \tag{3.5}$$

This formula must be used carefully, however, because it assumes a uniform temperature distribution. In addition, this equation assumes that the composite beam is initially flat when $\Delta T=0$. A better method for correlating temperature with deflection is done by simple experiment where the flap is placed upon a hot plate, and the temperature is allowed to reach an equilibrium. Results of such an experiment are shown in Figure 3.14, which plots tip deflection (not radius of curvature) vs. temperature. It is interesting

to note that the heat-induced bending serves to counteract the bending caused by thin-film stress.

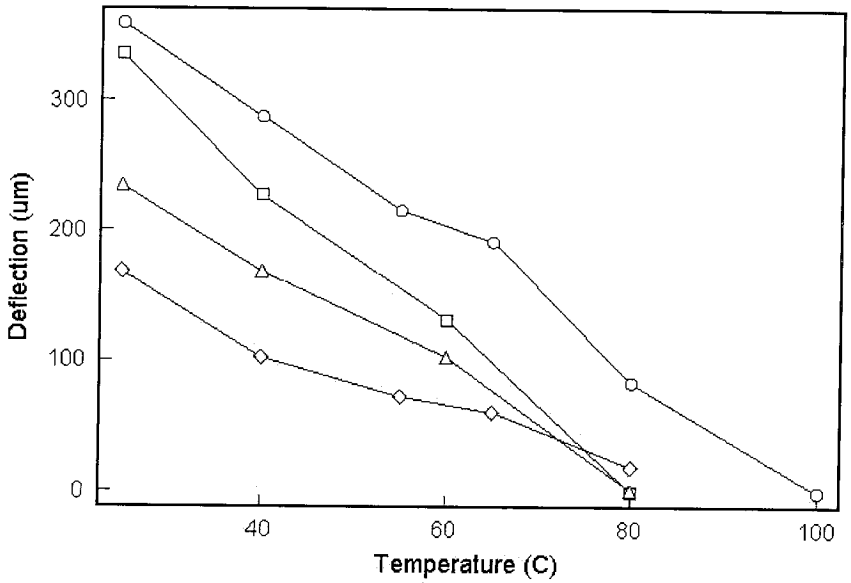


Figure 3.14 Deflection vs. Temperature for various first generation devices. The circular data points refer to a device 420x470 μm^2 with beams 280x20 μm^2. The square data points refer to a device 370x340 μm^2 with beams 360x50 μm^2. The triangular data points refer to a device 360x330 μm^2 with beams which are 360x50 μm^2. The diamond points refer to a device 310x370 μm^2 with beams 200x20 μm^2.

After determining deflection as a result of environmental temperature changes caused by a hot plate, we measured thermal deflection as a function of power applied to the actuator, i.e. localized heating. A typical curve showing thermal deflection versus power is shown in Figure 3.15.

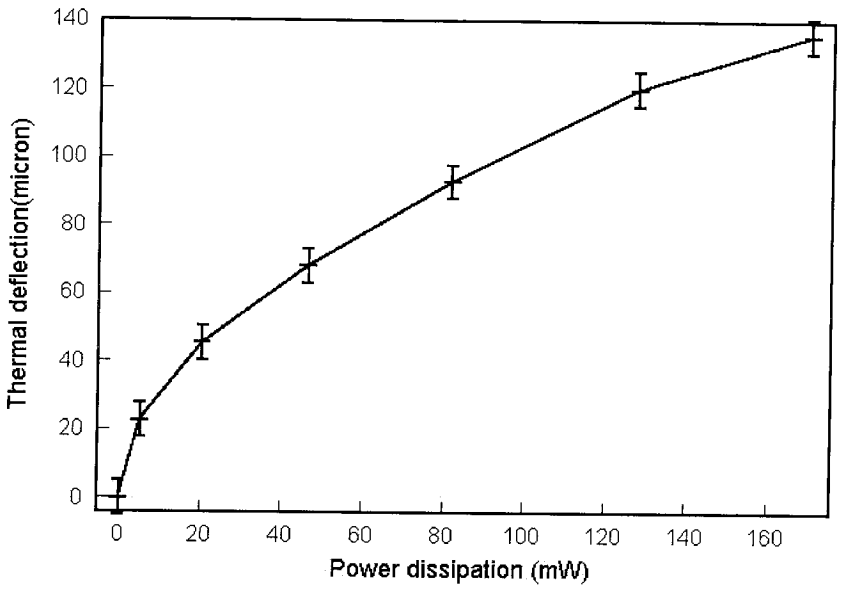


Figure 3.15 Typical Deflection vs. Power curve.

In addition, an infrared microscope (FSI 325) was used to measure temperature as a function of current.

For calculating the magnetic forces, it is assumed that a point force is generated at the center of the plate. (Minimal torque is assumed to be generated due to the fact that the magnetic moment and magnetic field are approximately parallel in this case. Therefore, the cross product between the two is extremely small.) Assuming that this force does not alter the shape of the plate, this point force can be translated to the interface between the plate and the beams and results in a point force and a moment. Once again taking the example flap described above, to obtain a beam deflection of 50 μm a force of approximately 1.2 μN is required at the center of the plate.

Figure 3.16 shows a plot of deflection vs. magnetic field gradient. The gradient was changed by varying the current in the electromagnet. The sign of the gradient was changed by reversing the direction of the current in the electromagnet. As expected, the deflection varies linearly with the gradient of the B-field. This further validates our

assumption that minimal torque is generated by the interaction of the magnetic moment with the B field.

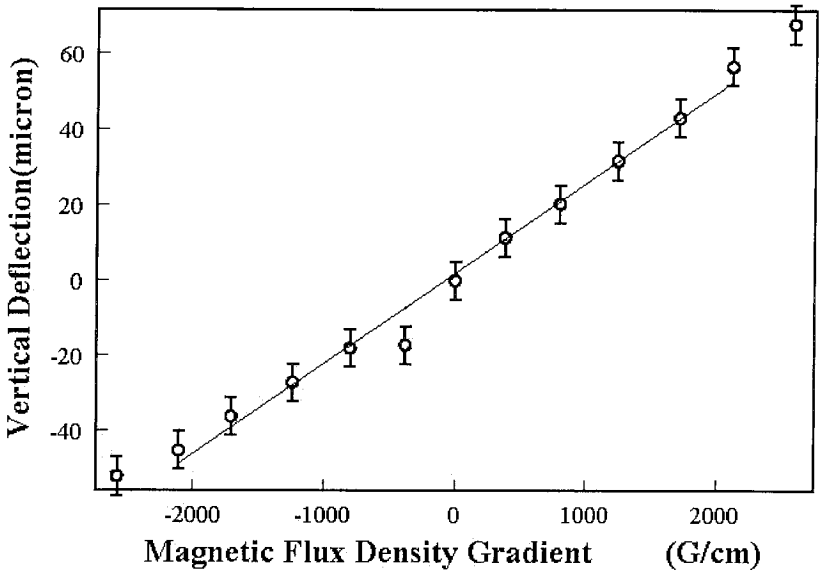


Figure 3.16 Magnetic deflection vs. B-field gradient for a first generation actuator. Note the linear dependence as predicted by theory.

A thermal frequency response curve is shown in Figure 3.17. The flap used in this testing is 300x300 μm^2 with two cantilever beams (200 μm long by 18 μm wide) as supports. Deflections were measured using a stroboscope. The bandwidth is approximately 1 kHz, with two distinct smaller peaks at 180 Hz and 360 Hz being observed. An ANSYS simulation of the same structure also shows the first mode resonant frequency to be approximately 1kHz. Differing time constants were observed for heating (0.48 ms) and cooling (0.26 ms). It is believed that the forced convection associated with the movement of the flap aids in cooling the flap and retards the heating.

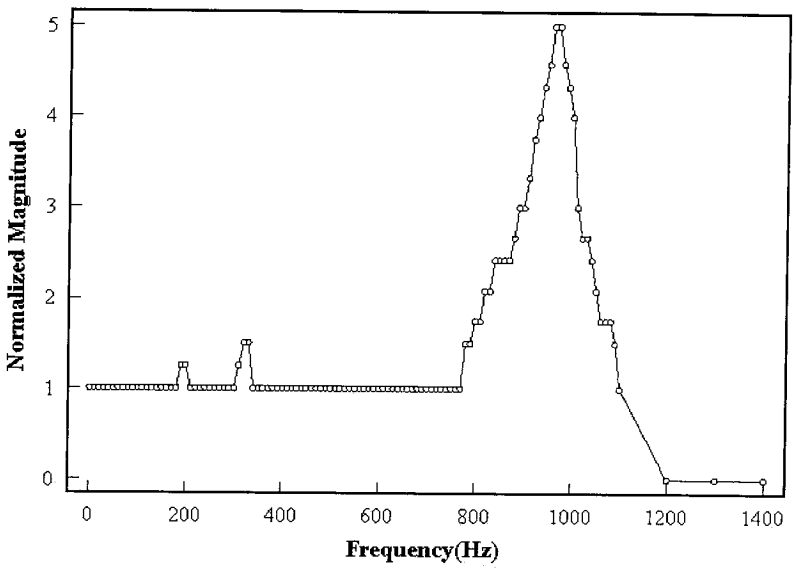


Figure 3.17 Example of freqency response of a first generation actuator.

3.2.4 Generation 2 – Two layers of Coils

Figure 3.18 shows a top view of a typical second generation magnetic flap. Note that in addition to on-flap coils, there are coils lying outside the perimeter as well. The purpose of the outside coils was to generate an on-chip magnetic field, thereby obviating the need for an external magnet.

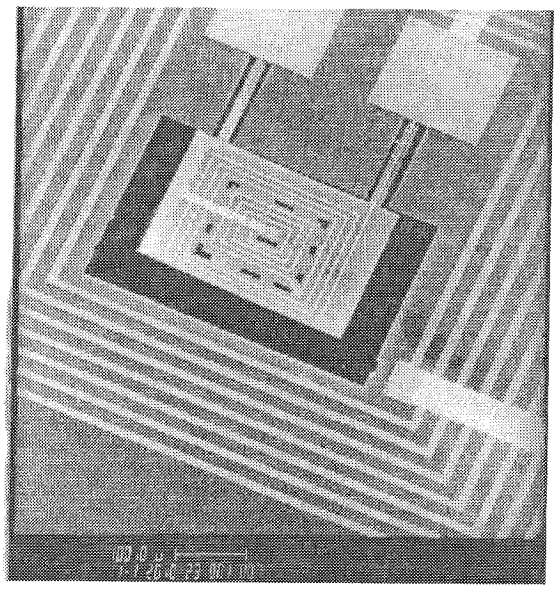


Figure 3.18 Top view of a second generation active flap.

3.2.4.1 Fabrication

The fabrication steps for the two-metal air bridge process is shown in Figure 3.19(b) and the steps are listed as follows. Note that step 4 could be changed to try different insulators such as photoresist or PECVD nitride.

(1) Deposit 5.0 μm PSG for use as the sacrificial layer.

(2) Deposit and pattern 1.2 μm silicon nitride for use as the structural layer.

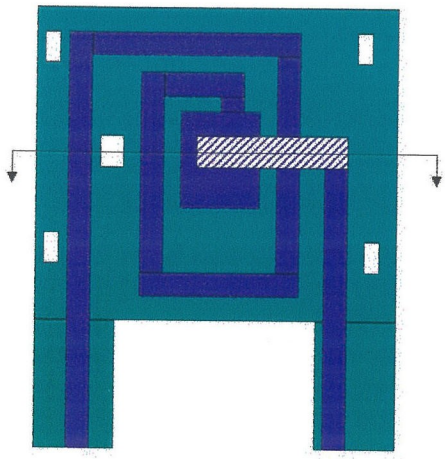
(3) Deposit and pattern Cr/Au (0.01 μm and 0.5 μm respectively) for use as the first metal layer.

(4) Spin photoresist for use as insulator. Resist (KTI 100cs) will be removed during the release process, leaving behind air bridges.

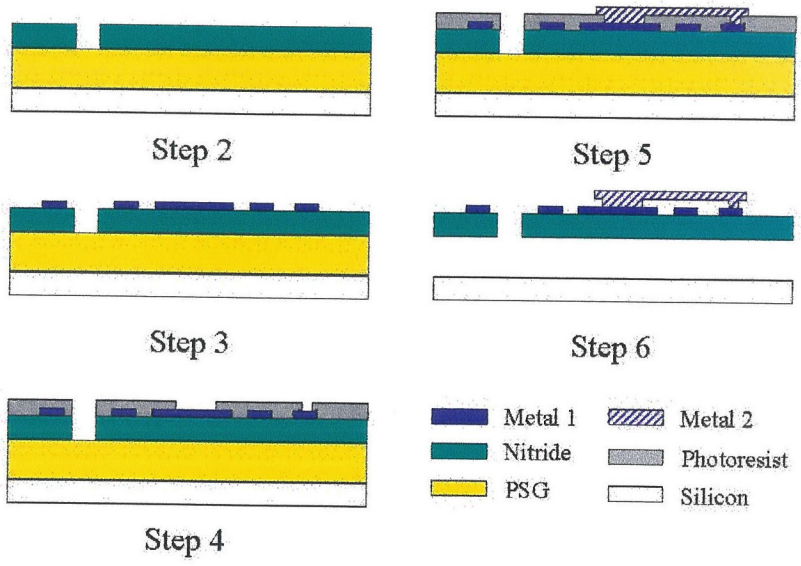
(5) Deposit and pattern Cr/Au (0.01 μm and 0.5 μm respectively) for use as the second metal layer.

(6) Dice wafer and submerse dice into high concentrated HF acid for release. Note that the acid also etches the insulation layer, thus leaving an air bridge.

Considering that stiction was a problem encountered during the release step of the first generation actuator, it was assumed that it would be an even larger problem if the intrinsic stress issue were solved. Therefore, several of the second generation actuators also used silicon nitride or photoresist tethers to connect the flaps with the surroundings during etching. The idea was to remove the tethers after etching.



(a)



(b)

Figure 3.19 (a) Schematic and (b) Process flow for second generation actuator.

3.2.4.2 Testing and Results

The silicon nitride tethers (typically 100 μm long by 6 μm wide) often were too narrow to support the large plates. Larger nitride tethers proved to be very difficult to break and altered the geometry considerably. Even if an acceptable geometry were constructed, manually destroying the tethers is a very time-consuming and yield-reducing process.

Photoresist tethers were thought to be ideal, because they could be made arbitrarily large, therefore supporting the structure, and could later be released by oxygen plasma etching, a dry process. Unfortunately, the experience with the resist tethers was the same as with other sorts of resist in high concentrated HF – they lifted off before the flaps were released.

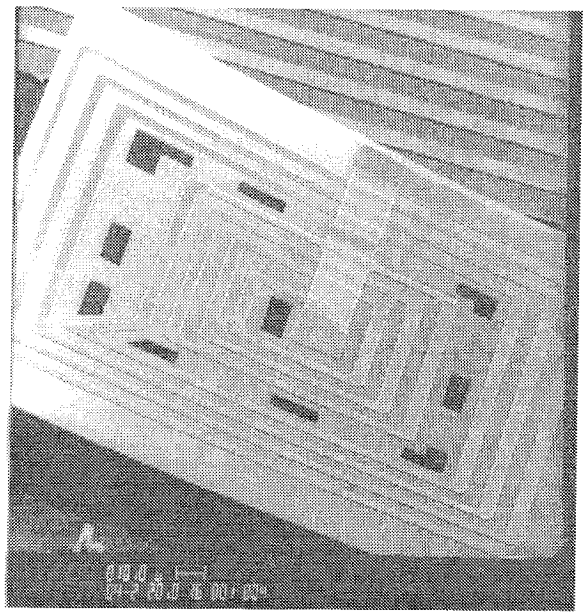
By using microscope probe manipulators to pry underneath the plates, the actuators could be released from the sticking position caused by in-use stiction. However, once the electrical leads to the plate would come into contact with any off-chip electronics (including power supply leads, people's hands, etc.), the plate would immediately stick down again and remain stuck even after electrical contact was broken. This suggests an electrosotatic attraction between the bottom of the plate and the substrate. It is believed that since the bottom of the plate (silicon nitride) is an insulator, it can trap charges. When a potential is applied between the charged plate and the substrate, an electrostatic force draws the two together. Once close enough, perhaps Van der Waals or other small distance forces keep the plate stuck to the substrate.

One method to prevent the plate from sticking to the substrate during use involved the placing of micron sized alumina (aluminum oxide) particles underneath the plate using probe tips. Plates generally required several conglomerations of particles in several locations before becoming permanently unstuck. However, this method is difficult and can be very destructive to the plate. If too much force is accidentally applied, the plate can be fractured.

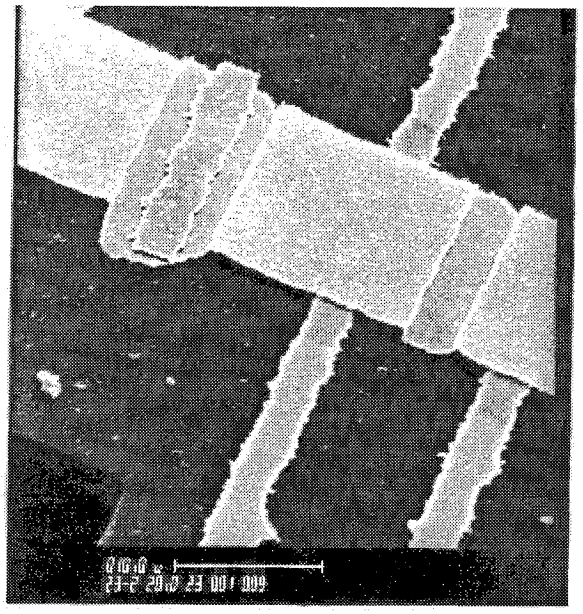
Stiction was not the only problem in the second generation. The air bridges would often fail. Interestingly enough, the bridges with anchors failed much more often than those without. SEM pictures [Figure 3.20] show that the anchors seem to actually pull the second layer of metal down at points away from the anchors.

The planar outer coils were not effective in creating an on-chip magnetic field strong enough to interact significantly with the on-flap magnetic moment. There is simply too much volume for the magnetic flux to escape into.

When PECVD nitride was used as the insulating layer, the flaps could often be released before the insulator could be etched away. However, the stress in this film was extremely high, resulting in structures that bent even more severely than our worst first generation actuators [Figure 3.21].

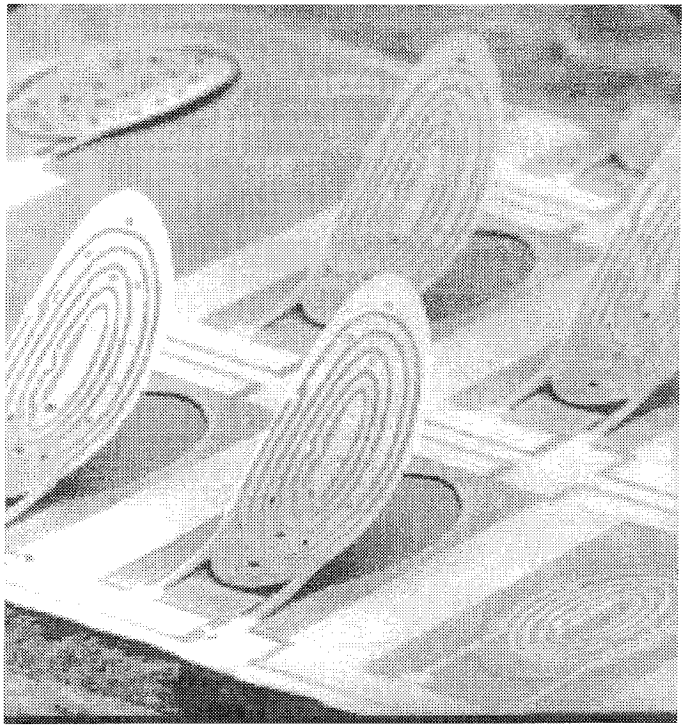


(a)

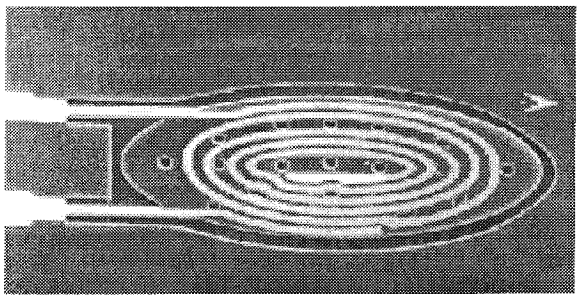


(b)

Figure 3.20 (a) Closer view of a second generation actuator. (b) SEM picture showing close-up of metal 2 air bridge over metal 1. These bridges would often fail around the support points.



(a)



(b)

Figure 3.21 SEM showing (a) array of flaps that used PECVD nitride as the insulating layer. The curvature is due to stress in the thin films. (b) Top view of an unreleased flap. The elliptical shape was used so that the flap would interact with the flow in a less abrupt manner.

Mechanical testing of the dynamic performance of the flaps showed that, unlike the first generation actuator, the frequency response was flat until it rolled off [Fig 3.22]. The physical structures were not too different from the first generation actuators. Therefore, it is believed that the lack of a resonant frequency is due to air damping. The

first generation actuator was free to vibrate in the air. The second generation flaps are very flat and lie within a recess that is filled with the sacrificial layer before freeing and with a thin film of air afterwards.

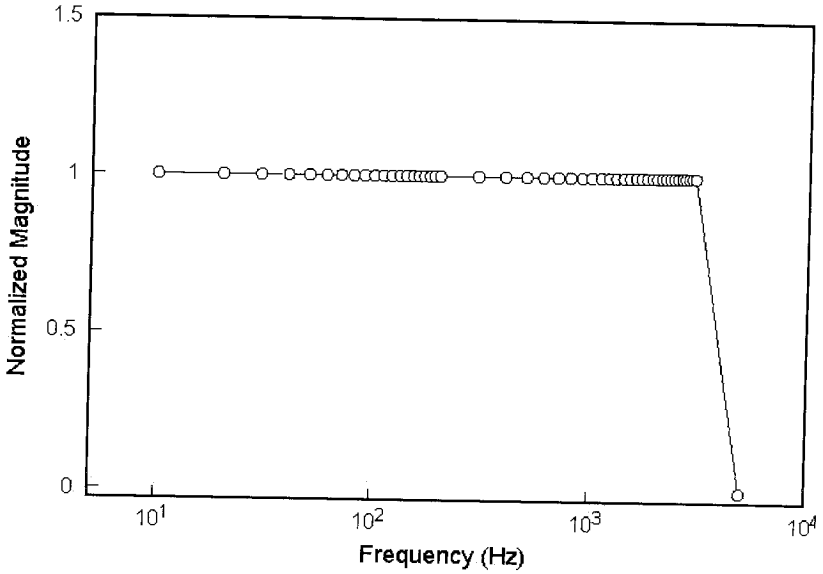


Figure 3.22 Figure showing typical frequency response of a second generation actuator.

3.2.5 Generation 3 – Metal Sacrificial Layer

With the completion of the first two generations of actuators, the next logical step was to design a magnetic actuator that could be used in an integrated process with electronics and shear stress sensors. Integration concerns will be presented in Chapters 4 and 5, but in brief, we decided to design a flexible actuator process that afforded a wide degree of latitude in sensor and actuator integration processing. If the entire actuator process could consist of low temperature steps, it would then place minimal constraints on the timing of the other processes. Most importantly, it would allow for the electronics to be completely fabricated first, which is often desirable for practical reasons to be discussed in Chapter 5.

The first thoughts, then, were to redesign the process to use low temperature materials for both the sacrificial layer as well as the structural layer. In addition, it was deemed desirable to not use an oxide sacrificial layer, because oxide is also used in standard electronics processing as an insulation layer. Therefore, the HF acid release etch would also severely attack the electronics insulation/protection.

3.2.5.1 Aluminum vs. Copper

The first consideration was to choose an appropriate sacrificial layer. Metal was an attractive choice primarily for the ease of depositing thick low temperature films through methods such as e-beam evaporation or electroplating. Thermal evaporation was not chosen because it can be difficult to obtain several microns of thermally evaporated aluminum. Two metals were chosen for preliminary studies – evaporated aluminum and electroplated copper.

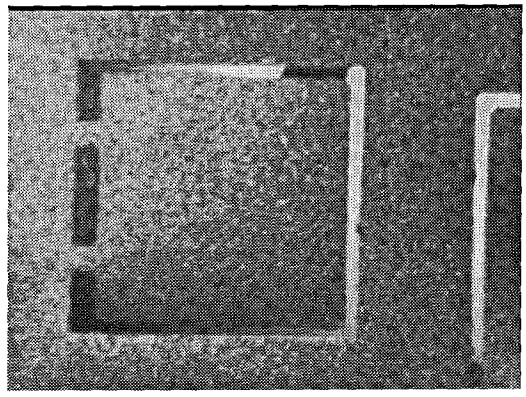
Some simple structures using aluminum as the sacrificial layer, LTO as the structural and insulation layers, and evaporated copper as the two layers of coils were designed and fabricated. Release was done using a HCl solution, which attack+6s aluminum, but not LTO or copper. Unfortunately, these structures experienced severe problems with the chrome adhesion layer. Ultimately, it was quickly decided to abandon this approach due to the fact that aluminum is used as the primary metalization for both the electronics and the sensor processing. Thus, the release etch could potentially destroy those parts.

Electroplated copper proves to be an almost ideal sacrificial layer for use with IC processing for reasons that will be explained in Chapter 4. The basis for most of the low temperature flaps, then, was to use thick copper as the sacrificial layer, gold as metal 1,

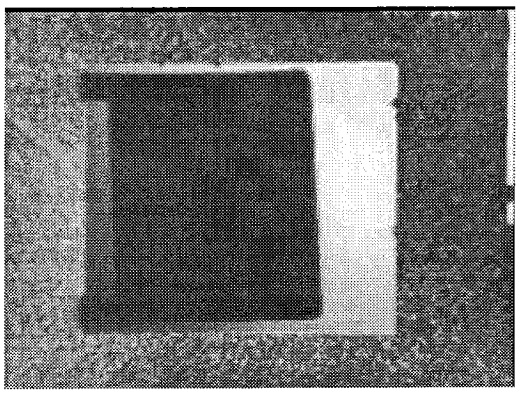
LTO as both plate material as well as insulator, and then gold again as metal 2. Since the solution for etching copper is very mild and doesn't attack LTO, there is no need to depend on air bridges with these actuators.

3.2.5.2 Different Generations

Within the low temperature actuator designs, there exist several sub-generations. The first attempt was completed to test the viability of using copper for the sacrificial layer. The easiest process, then, was to simply electroplate copper globally across a wafer, and then selectively plate permalloy for use as a structural layer. These structures, which consisted only of plates, were then released with a standard copper etch solution of acetic acid and hydrogen peroxide, which does not attack permalloy. The permalloy used in these tests was rather thick, on the order of 5.0 μm. All of the released permalloy structures were very robust. An example of such a structure, both in the resting and actuating positions, is shown in Figure 3.23.



(a)



(b)

Figure 3.23 Top view of permalloy flap in (a) resting, and (b) actuated positions. The permalloy was electroplated directly atop an electroplated copper sacrificial layer.

Having validated the performance of copper as a sacrificial layer, the next runs were all aimed at creating active flaps. The planned process flow for the first attempt is shown here and shown in Figure 3.24:

(1) Deposit a Cr/Cu (0.01 μm and 0.1 μm, respectively) Seed Layer

(2) Electroplate copper globally (sacrificial layer – 4.5 μm typically).

(3) Deposit/Pattern Cr/Au (metal 1 – 0.01 μm and 0.3 μm, respectively)

(4) Deposit/Pattern LTO (both structural and insulating layer – 0.4 μm)

(5) Deposit/Pattern Cr/Au (metal 2 – 0.01 μm and 0.3 μm, respectively)

(6) Release structures in copper etch solution.

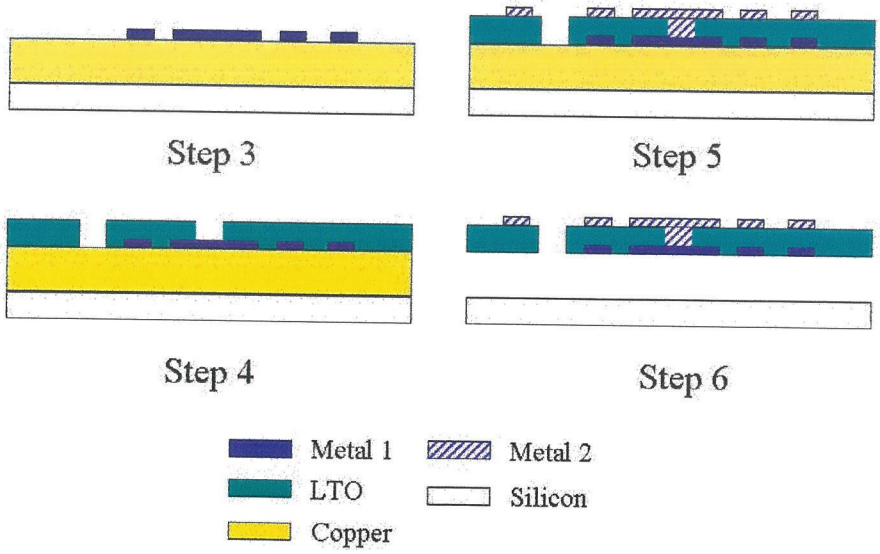


Figure 3.24 Ideal Process flow for first attempt at fabricating the low temperature actuator.

Unfortunately, none of the structures created using these steps could be released. The gold etch solution attacks copper at a reasonably fast rate. Therefore, during step 3 (gold etch), the copper is undercut. Due to the conformal nature of the subsequent LTO deposition, the undercut areas are also coated. The patterning of the LTO does not account for this coating and leaves it unetched. During the sacrificial layer etch, the copper actually has a protection layer of silicon dioxide [Figure 3.25].

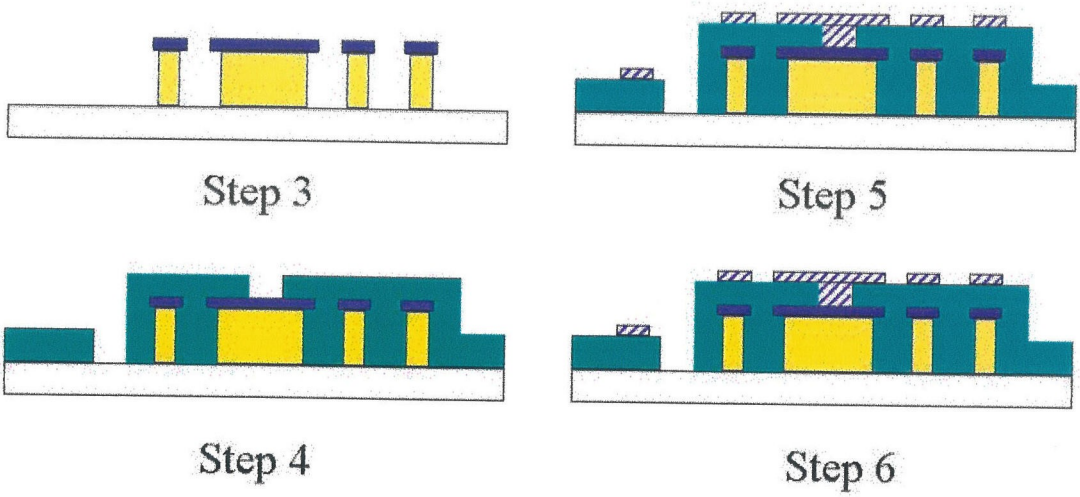


Figure 3.25 Actual process flow and cross section of the first attempt at fabricating a low temperature actuator. Note that Step 6 and Step 5 are identical. This is a result of the etching solution not being able to have contact with the copper. Note that as in all drawings, the vertical dimension is highly exaggerated.

The next attempt at a low temperature actuator used two additional steps to counter the possible inadvertent undercutting of the copper. The first metalization used a lift-off process, rather than a standard deposition and patterning. As a secondary precaution, a thin permalloy layer, which is not attacked by gold etchant, was electroplated on top of the copper sacrificial layer on certain wafers. An example of such a process is given below:

(1) Deposit a Cr/Cu seed layer (0.01 μm and 0.1 μm, respectively)

(2) Electroplate copper globally (sacrificial layer, 4.5 μm typically) on all wafers.

(3) Electroplate permalloy (0.5 μm, typically) on selected wafers.

(4) Perform lift-off of Cr/Au (metal 1 – 0.01 μm and 0.3 μm, respectively)

(5) Deposit/Pattern LTO (insulator and structural layer – 0.4 μm typically)

(6) Deposit/Pattern Cr/Au (metal 2 - 0.01 μm and 0.3 μm, respectively)

(7) Release structures in an acetic acid/hydrogen peroxide solution.

The use of lift-off greatly reduced the uncertainty associated with etching chrome and certainly solved the problem of the gold etch undercutting the copper. Therefore, it was not truly necessary to electroplate permalloy as a thin secondary protection layer. However, structures from wafers which used permalloy released with a much higher yield rate. This was almost certainly due to the extra structural support provided by the permalloy film during the bubble-generating release process.

One additional note should be mentioned here with regards to the presence of permalloy. During the sacrificial layer etching, no attempt was made to completely remove the permalloy, which covers the entire underside of the flaps. If the first layer of metal consisted of coils, this would not be an adequate solution, as the permalloy would short all of the lines. As it were, the design used in these actuators was the exact same as used in a first generation of actuator, in which a large plate of polysilicon was used as a current path, rather than as a second layer of coils. Hence, electrically speaking, the permalloy only serves to lower an already low plate resistance. If the first layer of metal was to be coils, then they would either have to be separated from the permalloy by an insulator, or the permalloy would have to be entirely undercut using a dilute sulfuric acid solution. This could be done quickly after the copper etch.

One problem with all of the above runs was the fact that globally plating copper resulted in a poor quality, highly non-uniform film that often contains streaks, pinholes, or other undesirable structures. Therefore, subsequent generations all had selective electroplating of the sacrificial layer. Such a process flow would be as follows:

(1) Deposit a Cr/Cu seed layer (0.01 μm and 0.1 μm, respectively) on wafers with thermal oxidation

(2) Pattern seed layer.

(3) Electroplate copper locally (sacrificial layer – typically 4.5 μm) on all wafers.

(4) Electroplate permalloy on selected wafers (typically 0.5 μm).

(5) Perform lift-off of Cr/Au (metal 1 - 0.01 μm and 0.3 μm, respectively)

(6) Deposit/Pattern LTO (insulator and structural layer – 0.4 μm)

(7) Perform lift-off of Cr/Au (metal 2 - 0.01 μm and 0.3 μm, respectively)

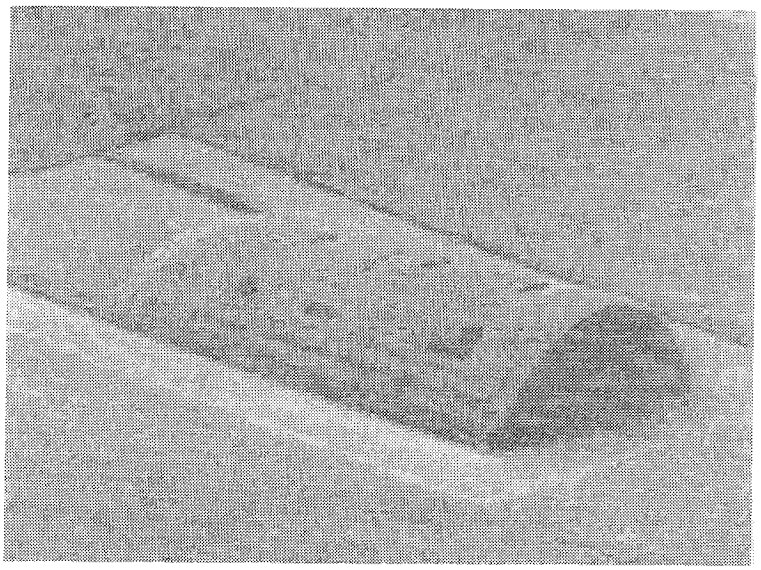
(8) Release structures in an acetic acid/hydrogen peroxide solution.

A few interesting notes are worth mentioning with regards to this process. First, lift-off is the process for both layers of metal. This is primarily due to the high level of irregularity associated with the etching of chrome. This problem seems to be only exacerbated by the presence of other metals, which may combine with the chrome (by diffusion, for example) and form unetchable alloys.

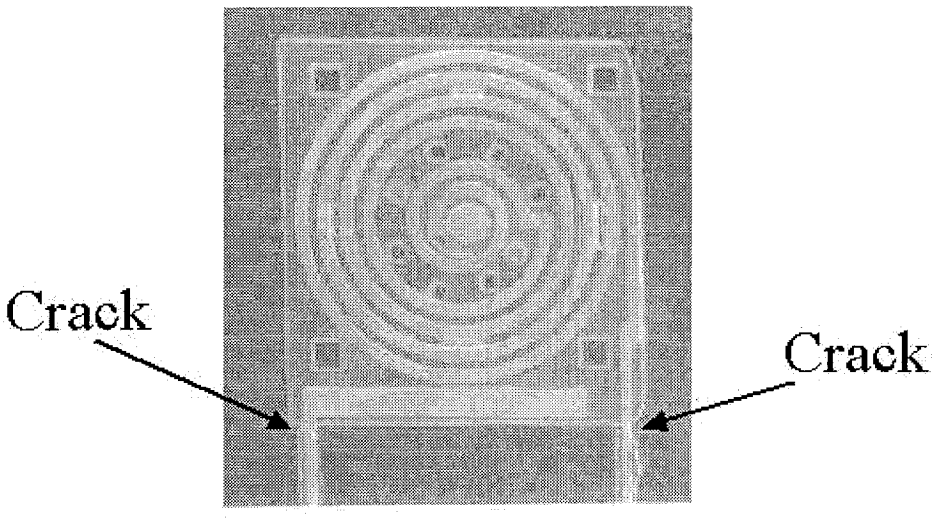
Second, the process starts with oxidized wafers. The reason for this is electrical isolation of the islands. Assuming the seed layer between islands could be etched away (not always a good assumption), there would still exist an electrical path between islands through the bulk silicon. For reasons of either poor design rules or non-complete etching of metals, the first layer of metal would often be shorted to the remaining copper sacrificial layer. This happenstance might not be disastrous if it were to occur on an individual island. In fact, this may be a beneficial structure, as the first layer of metal would be at the same potential as the substrate (shorted through the copper), and

therefore, no electrostatic forces would exist to cause in-use stiction. However, if islands were to be shorted together, then the devices on these islands would also be shorted, which generally speaking is not desirable.

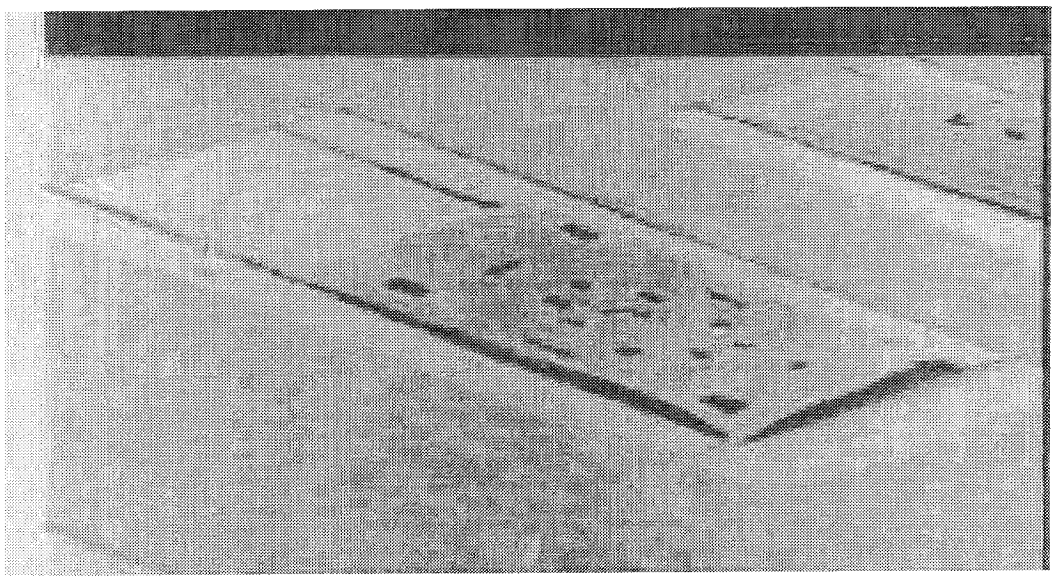
Ultimately, the low temperature actuator was abandoned. The reason for this is that while copper proved to be a usable sacrificial layer, LTO definitely is not a good structural layer for free-standing devices. The internal stresses cause an incredible amount of warpage, and LTO is a very brittle material, easily cracked during processing. [Figure 3.26]



(a)



(b)



(c)

Figure 3.26 (a) Perspective SEM of a low temperature actuator, (b) Top view of a actuator. Notice the cracking at the connection points between the beams and the flap. (c) released actuator held down by two LTO tethers. Even with the tethers, evidence of warpage is obvious.

3.2.6 Generation 4 – Hybrid

With the low temperature process abandoned, the debate began again to choose a sacrificial layer that would be compatible with a 2 metal process. Essentially, we decided to pursue two options: standard PSG and polysilicon. For all intents and purposes, an oxide is the only choice for a stable insulator. Therefore, in the PSG sacrificial layer process, it was decided to use thick electroplated permalloy as the top metal. Buoyed by our observations with the initial low temperature run, we believed that a thick, stiff layer would mechanically survive the release etching process, unlike the relatively thin metal layers used in the 2nd generation.

The pursuit of the second option was motivated by the fact that the entire problem of having a potentially fragile air bridge is made moot if polysilicon is used as the sacrificial layer and it etched using a selective etchant which does not attack oxide. Therefore, thin metals (as provided by standard CMOS processes) could potentially be used as the metalization in an integration run.

One may wonder as to the usefulness of combining permalloy with an active flap. One idea is that this would create a "normally-on" flap that has a "DC" position that is already deflected. By applying a current, the flap can then either be brought down to the horizontal position or deflected even further. Also, the presence of permalloy allows for the use of the magnetic release process first developed for use with the passive actuators. Therefore, a higher yield should be achievable.

3.2.6.1 Fabrication

The first version of the hybrid actuators had many different patterns of flaps and many different etch hole configurations (similar to the first generation of active flaps).

Permalloy was used as the current return path. The process flow is shown as follows and in Figure 3.27:

(1) Deposit 5.0 μm PSG sacrificial layer.

(2) Deposit and pattern 1.2 μm nitride structural layer.

(3) Deposit and pattern Cr/Au (Metal 1 - 0.01 μm and 0.35 μm, respectively)

(4) Deposit and pattern LTO (Insulator – 0.7 μm)

(5) Deposit and pattern Cr/Cu (Seed Layer – 0.02 and 0.2 μm, respectively)

(6) Pattern Electroplating Mold

(7) Electroplate NiFe (Metal 2 and Passive Element – 5.0 μm)

(8) Etch Seed Layer and Free Structures in HF.

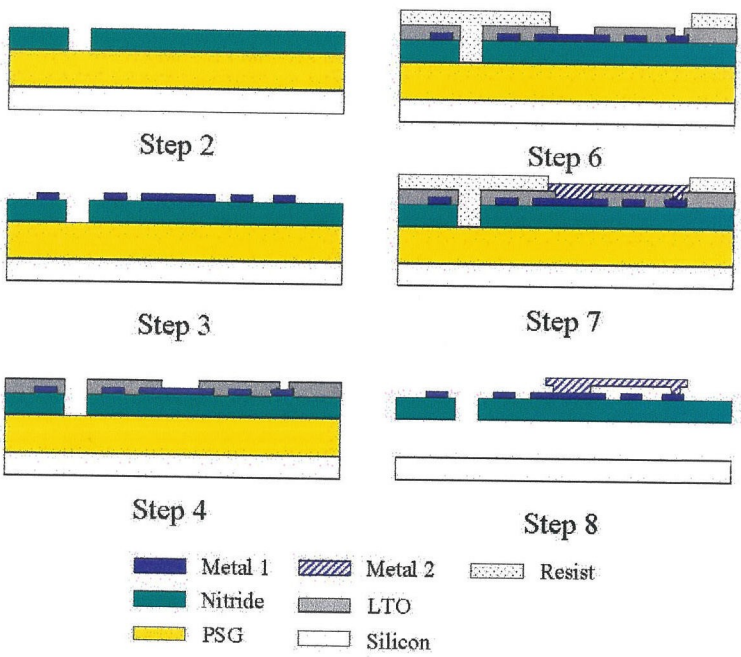
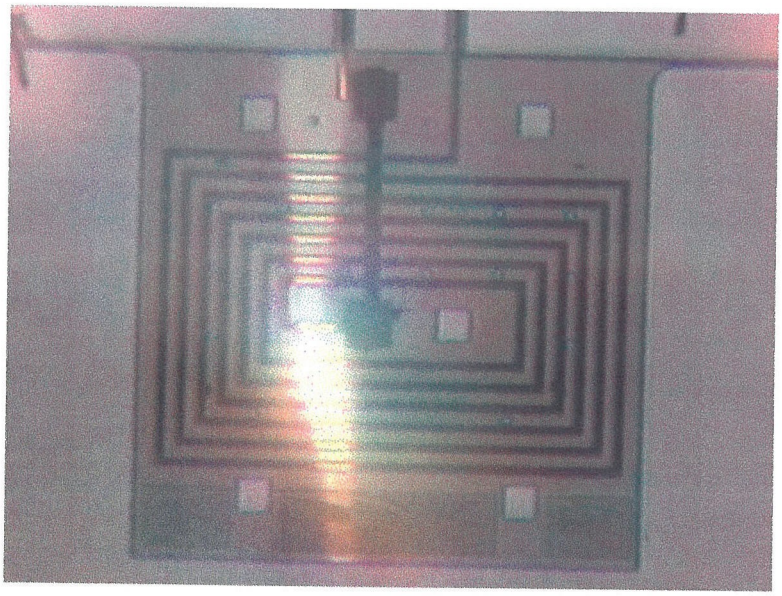
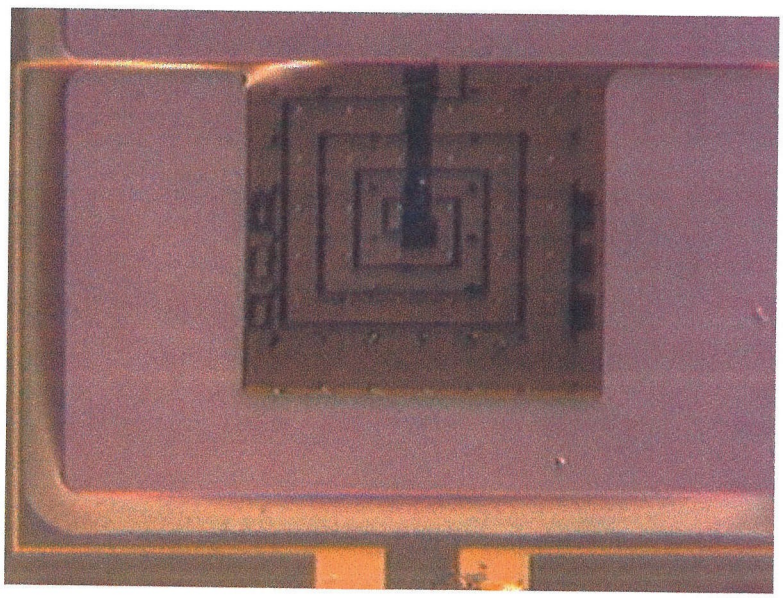


Figure 3.27 Process flow for the first iteration of the hybrid actuator.



(a)



(b)

Figure 3.28 Picture showing released hybrid actuators. (a) Note that the 2nd layer of metal is actually a gold wire-bond. (b) Standard 2nd layer of metal.

The second version used polysilicon as the sacrificial layer and chrome/gold as both metal layers. During the release etching in TMAH, many of the structures would crack. Most of these cracks would have one end at the edges or corners of etch holes. It

was believed that the uneven release of the flaps would cause unusual bending in the solution. Therefore, a re-design using equidistant and same size etch holes was fabricated, and the propensity for cracking decreased significantly. The process for this version of the hybrid actuators is shown as follows and in Figure 3.29:

(1) Deposit 1.0 μm polysilicon sacrificial layer on insulating (either 0.5 μm silicon nitride or 1.5 μm thermal silicon dioxide).

(2) Deposit and pattern 1.2 μm silicon nitride structural layer.

(3) Deposit and pattern Cr/Au (Metal 1 - 0.01 μm and 0.35 μm, respectively)

(4) Deposit and pattern LTO (Insulator – 0.5 μm)

(5) Deposit and pattern Cr/Au (Metal 2 – 0.01 and 0.4 μm, respectively)

(6) Deposit and pattern Cr/Cu (Seed Layer – 0.01 and 0.1 μm, respectively)

(7) Pattern Electoplating Mold

(8) Electroplate NiFe (5.0 μm)

(9) Etch seed layer and free structures using various methods.

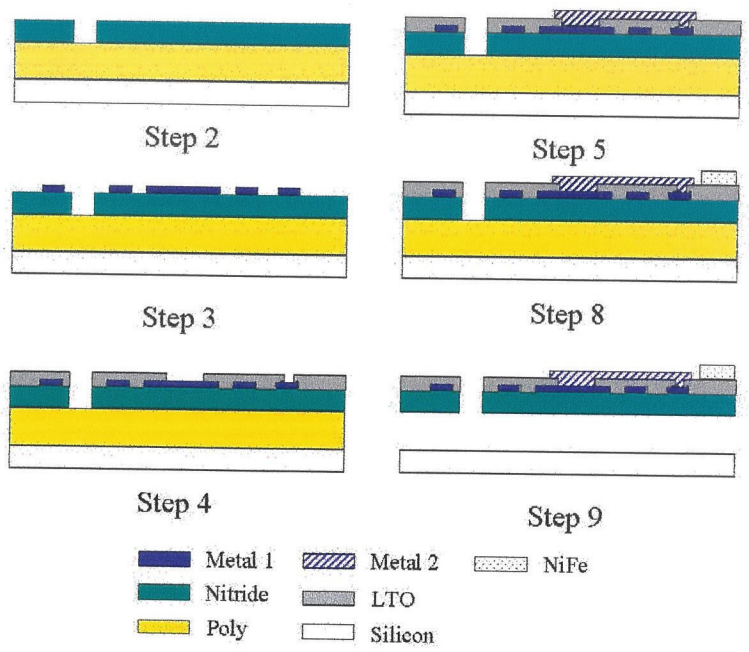


Figure 3.29 Process flow for hybrid actuators using polysilicon sacrificial layer, two layers of metal, and permalloy.

3.2.6.2 Testing and Results

One of the biggest problems associated with the testing of the hybrid flaps was the unexpectedly high resistance of the coils. At first, this was thought to be a result of high contact resistance. This belief was made stronger by the fact that several measurements of resistance seemed to indicated that the resistance of the first layer of would often seem to be close to predicted, but later measurements of both layers of coils would lead to abnormally high resistance values. However, by using an SEM, it was determined that the second layer of metal often did not cover the first layer steps in an adequate fashion and would crack at the steps as shown in Figure 3.30.

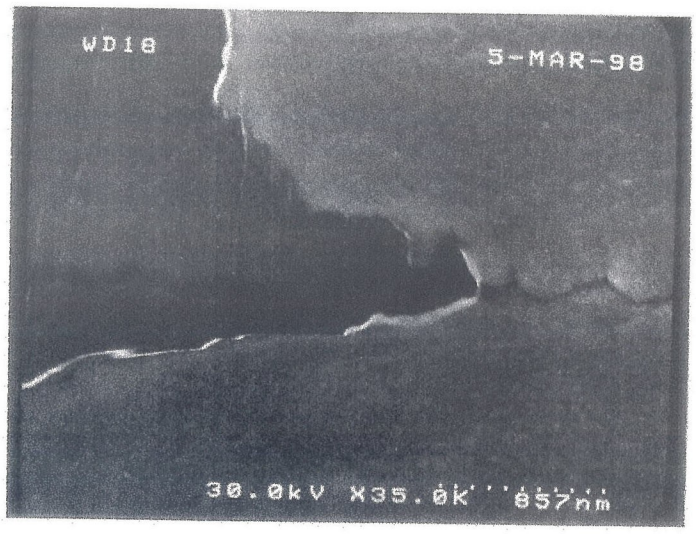
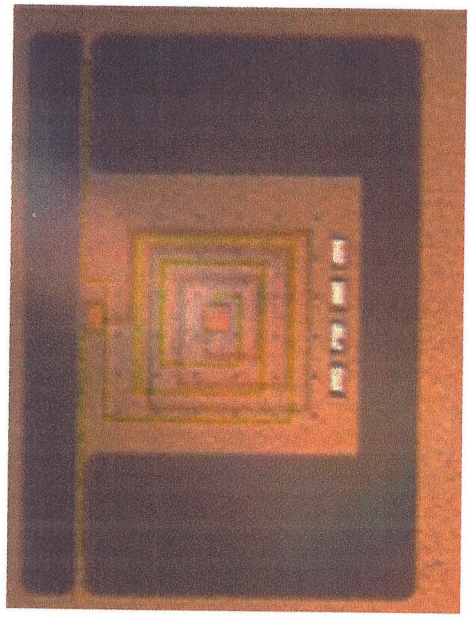
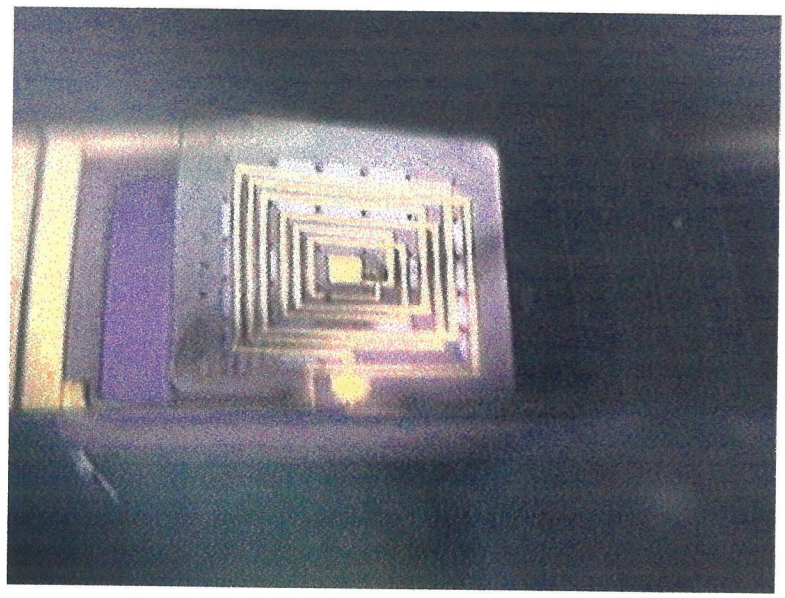


Figure 3.30 SEM showing Metal 2 running over Metal 1 and cracking at the step.



(a)



(b)

Figure 3.31 (a) Top view of a hybrid flap, before release, using polysilicon as the sacrificial layer. (b) Perspective view of a different hybrid flap, after release. Note the warpage.

3.2.7 Generation 5 - Bulk

3.2.7.1 Design

As described in Chapter 2, the lifetime of the structures being studied is on the order of milliseconds. Therefore, the minimum speed with which the actuator should operate is on the order of 1 kHz. In actuality, it may be desirable to pump each structure with several cycles of actuation – therefore requiring a frequency of several kHz. Therefore, as a starting point, it was decided to fabricate actuators with a resonant frequency of 3 kHz, with the thought that it would be able to beat three times during the lifetime of a 1 ms streak.

Also, it was hypothesized that the important actuator parameter for controlling the flow in a turbulent boundary layer was not simply the displacement of the actuator but rather the tip velocity, which is the product of angular frequency and displacement. If

true, this hypothesis implies that an actuator moving quickly but moving over small distances would not have any lesser effect as a slowly moving actuator that has a larger displacement. Given a choice, the higher frequency actuator has the added advantage of not contributing as much to form drag. Therefore, the combination of all these reasons led to an effort to create and study high frequency, low displacement actuators. Figure 3.32 shows some work done [22] at low frequencies and high displacement and is presented primarily for showing motivation to pursue high tip velocities.

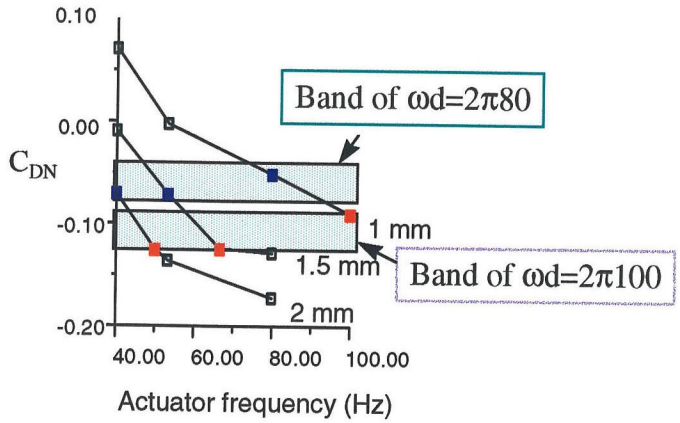


Figure 3.32 Graph showing result of localized drag reduction when using actuators with different frequencies and amplitudes. Note that the bands of constant tip velocity appear to give similar drag reduction results. Also note that the actuator used for this testing was a bulk micromachined flap fabricated by Miller [1].

The surface micromachined actuators typically had resonant frequencies on the order of up to, but not exceeding, 1000 Hz. At higher frequency operations, the amplitude of such a structure is highly attenuated. Therefore, it was decided to use an actuator with a much higher resonant frequency. In addition, if such an actuator were to be used at its resonant frequency, the resulting amplitude could be much larger than otherwise possible. In addition, prior experience gained from the surface micromachined actuators indicated that torsional, rather than bending, beams should be used.

In general, the resonant frequency of a torsional structure is described as in the following equation where ω is the angular frequency, k_θ is known as the torsional spring constant, and I is the moment of intertia of the structure.

$$\omega = \sqrt{\frac{k_\theta}{I}} \tag{3.6}$$

The torsional spring constant is calculated as follows, where G is the modulus of rigidity and k is a geometric factor which has been determined experimentally and represents the fact that a non-circular cross section will experience warping (extension into the third dimension) under an applied torque. E is the Young's modulus (or modulus of elasticity), and ν is Poisson's ratio (not to be confused with kinematic viscosity in fluid mechanics). The dimensions of the beam (a,b, and L) are shown in Figure 3.33.

$$k_\theta = \frac{kG}{L} \tag{3.7}$$

$$G = \frac{E}{2(1+\upsilon)} \tag{3.8}$$

$$k = ab^3[\frac{16}{3} - 3.36\frac{b}{a}(1 - \frac{b^4}{12a^4})] \tag{3.9}$$

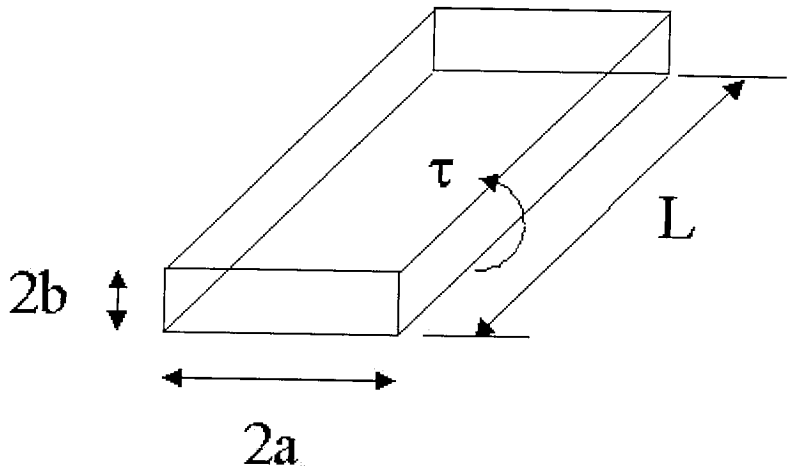


Figure 3.33 Drawing showing dimensions used to calculate the resonant torsionsal characteristics of a beam.

Note that in determining the value of k, *2b* is used to represent the smaller of the two sides of the cross section. The torsional resonant frequency of a thin beam therefore depends much more heavily upon the height of the beam rather than the width.

The moment of inertia, *I*, is calculated from the following general formula where r is the distance from the axis of rotation to each point and dM is the infinitesimal mass associated with that point.

$$I = \int r^2 dM \tag{3.10}$$

If we take the case of the actuator described as in Figure 3.34, we can use symmetry to simplify the calculation significantly.

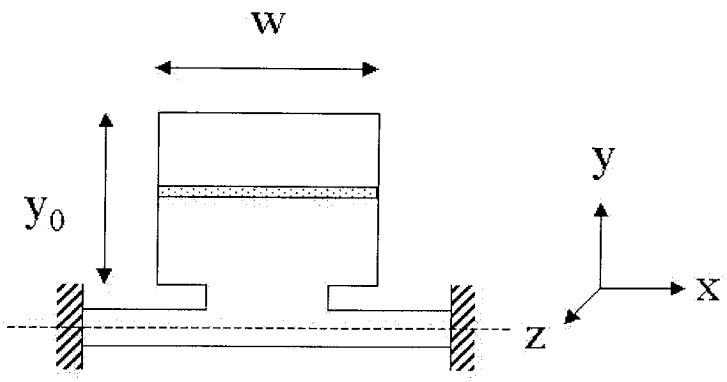


Figure 3.34 Drawing showing an idealized torsional bulk micromachined flap. Note that the picture is not to scale. The beam width is exaggerated for clarity sake.

We can assume that the contribution from the beam itself is negligible compared to the contribution from the plate. The reasoning behind this is the fact that the distances from the axis of rotation to points on the plate is on the order of a millimeter while the distances to points on the beam are on the order of tens of microns. In Figure 3.34, the dotted line is the axis of rotation while the shaded rectangle represents an infinitesimal width bar. In this case, the moment of inertia can be approximately calculated using the following formula where ρ represents the density of silicon (2.33 g/cm^3).

$$I = \int r^2 \rho dV = \int y^2 \rho(2b)(w)dy \tag{3.11}$$

which can be integrated to yield

$$I = \frac{2}{3} \rho b w y_0^{\,3} \tag{3.12}$$

When this result is substituted into the resonant frequency equation, the following formula results.

$$\omega = \sqrt{\frac{3ab^2[\frac{16}{3} - 3.36\frac{b}{a}(1 - \frac{b^4}{12a^4})]G}{\rho w y_0{}^3 L}} \tag{3.13}$$

The question then becomes which of the 3 parameters (a, b, or L – the width and length of the plate are not included here because they are assumed to be fixed by area constraints for magnetic actuation purposes) can be adjusted to result in the desired resonant frequency. Beam widths are constrained by the fact that the metal lines running atop must have a minimum width to provide decent current carrying capabilities (to prevent burning or melting of the metal) with reasonably low resistance. Typical beam widths are on the order of 20 μm. In surface micromachining, typical structural layer thicknesses are on the order of 1 μm or less, while in bulk micromachining, thicknesses are on the order of 10's of microns. From simple calculations, it is apparent that in order to achieve the minimum desired resonant frequency of 3 kHz, a surface micromachined beam 20 μm wide by 2 μm thick should have a length of approximately 30 μm. A bulk structure having a similar width but a thickness of 20 μm should have a length of approximately 500 μm. Both cases assume a flap size of 1 mm X 1 mm, which is a reasonable size requirement for generating a significant magnetic moment. If resonant frequency were the only or primary constraint, it would appear that both surface and bulk flaps are viable options.

3.2.7.2 Fabrication

The process flow for the standard bulk flap is fairly straightforward (See Figure 3.34) and the final structure is similar – in shape, not size - to all of the flaps previously described.

(1) Etch double sided alignment marks

(2) Grow 2.0 μm thick thermal silicon dioxide (Bulk etch mask)

(3) Pattern backside windows.

(4) KOH etch backside cavities (approximately 19-20 hrs).

(5) Strip thermal oxide.

(6) Regrow thin thermal oxide (Insulator – 0.3 μm).

(7) Deposit and pattern aluminum (Metal 1 – 0.6 μm).

(8) Deposit and pattern LTO (Insulator – 0.6 μm)

(9) Deposit and pattern aluminum (Metal 2 – 0.6 μm).

(10) Pattern silicon for RIE etch.

(11) Release structures in RIE.

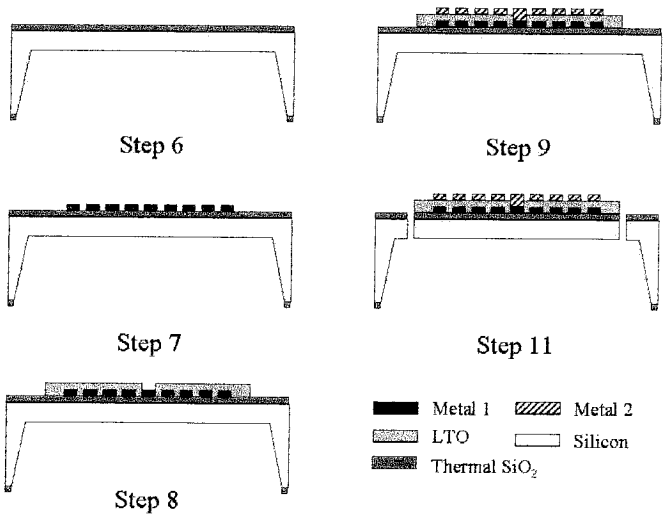


Figure 3.35 Process flow for bulk flaps.

The first thermal oxide (grown in step 2) is not used as the electrical insulation because it is slowly etched in a slightly non-uniform manner during the KOH etch. This

oxide layer cannot be easily planarized. Therefore, a better process is to strip all of the oxide and re-grow a uniform layer (steps 5 and 6).

It is conceivable that a thinner LTO layer could be used in step 8 and still provide a conformal covering. However, the problem of stress-induced curvature is far less in bulk structures, which are an order of magnitude thicker than surface structures, than it is in the surface structures. Therefore, it was deemed that a slightly thicker, more conservative layer would be used to provide short-proof insulation.

This process is elegant in the fact that if used in an integration process, only two steps need to be completed in a MEMS environment – the bulk etching and the final release. Every other step can be completed by an electronics foundry. Further descriptions as to the usefulness of this property will be included in Chapter 5. A picture of a bulk micromachined active flap is shown in Figure 3.36.

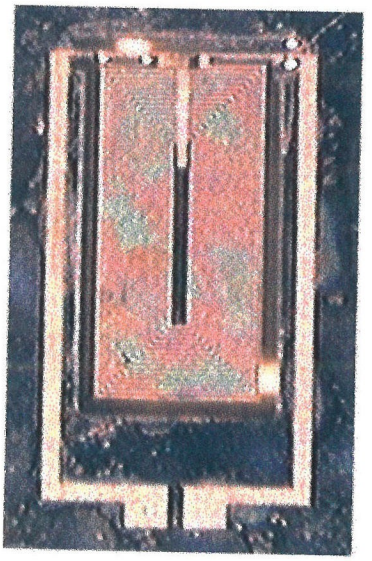


Figure 3.36 Top view of a bulk flap.

For eventual integration purposes, it is conceivable (see Chapter 5) that only after electronics processing will the MEMS processing be attempted. In such a case, the bulk silicon etching must be done carefully so as to ensure the survival of the front side

electronics, in particular the metal. One option is to use a one-sided etch apparatus. This approach was not chosen due to the fact that it is difficult to obtain such an apparatus that is 100% leak proof. In addition, such a system requires a clamping mechanism that induces stress on the wafer. It is conceivable that this stress could cause the weakened wafer to fracture. Finally, such a device typically leaves an air gap on the sealed side of the wafer. Pressure differentials between this air and the liquid on the other side of a thin membrane could cause individual membranes to rupture, thereby allowing the etchant to attack the front side.

A different option from using a one-sided etch apparatus is to cover the front side with a protective film. Such a deposition process must be low temperature so as to not melt the aluminum. Typically, silicon nitride and silicon dioxide are used as masking materials. However, the temperature requirement implies that the highest quality processes - LPCVD nitride and thermal oxidation - cannot be used. LPCVD oxide and PECVD nitride satisfy the temperature requirement, but their quality is not as good and their etch rates in silicon etchants is reasonably high. In addition, the process of removing this film can be tricky due to the fact that IC processes often use the same films as protection layers. It is nearly impossibly to remove the silicon-etch-protecting dielectric films while not attacking the electronic-protection dielectric films, especially due to the fact that standard operating procedure includes an overetch to ensure wafer-wide clearing of the film. This could lead to circuit contamination and eventual failure.

Wax is a material that can be easily deposited as a general protection layer onto a wafer by melting. Wax is also easily removed using highly selective organic solvents. There are many different types of wax available. Experiments were conducted with two

types of wax for protection during KOH etching. Etch resistance is not the only important attribute, however. Adhesion is also very important. The ideal wax cannot peel off of the wafer, or if it does peel after a reasonably long period of time, it should be able to be re-melted and reacquire the same adhesion characteristics.

Two types of wax were studied for this purpose – KPW-A Protect Wax, which is produced in Japan and specifically designed for use as protection in caustic environments, and Kerr Sticky Wax, which is used by dentists to make molds of teeth. KPW-A is black and color and has a softening point of 120° C. Kerr Sticky Wax is orange and has a melting point of approximately 58° C. The best method for applying either wax is simply to place a solid piece of wax atop the wafer and set them upon a hot plate or into a convection oven. Upon melting, the wax flows evenly across the surface and typically stops once it reaches the edges. If left to flow for a long enough time, this results in a fairly evenly coated wafer.

Neither wax could be left unattended for a 24-hour period in KOH. However, the Kerr Sticky Wax seemed to be a better product. It has a lower melting temperature and better adhesion characteristics than the KPW-A film, which would bubble, become brittle, and lift-off at the edges after approximately 4 hours of etching in a standard KOH bath. The Kerr wax would also lose some elasticity after extended etching but would never lift-off from the edges. It would, however, sometimes form pinholes or small cracks after a long (12 hr) unattended etching period. While such cracks are obviously undesirable and care should be taken to remelt the wax at frequent intervals (typically 4 hours between meltings), they are often not catastrophic. Minimal KOH seeps through, as is evidenced by the lack of silicon etching after the removal of the wax.

Both waxes could be remelted after exposure to KOH. However, the physical qualities of the wax after each re-melt were noticeably different. These differences included slight color changes, the inability to form a very smooth surface, and a loss of softness or pliability. Also, after each subsequent re-melting, the physical properties of the waxes would change quicker when exposed to KOH (i.e. KPW wax would become brittle in a shorter and shorter time period).

An alternative to re-melting the wax is to completely remove the wax and reapply a second coat. In any case, after the etching is complete, the wax needs to be removed. Therefore, some consideration must also be given to ease of dissolution of the wax. KPW-A wax proved to be quite resistant to many organic solvents including acetone, alcohol, and photoresist stripper. Boiling TCE seemed to be the only available solvent which would dissolve the wax, and slowly at that. It takes several hours to remove a coat, and during those hours, the solution needs to be constantly replenished, as it quickly becomes black. By contrast, Kerr Sticky Wax is slowly dissolved with acetone and is attacked relatively quickly with TCE. Soaking in a TCE solution for an hour at room temperature causes the wax to soften and dissolve at an appreciable rate. If desired, at this point, the wax can also be easily scraped off. If immersed in heated TCE (a temperature as low as 30^{o} C is adequate), this process can be accelerated. In any case, the removal of Kerr Sticky Wax is orders of magnitude easier than the removal of KPW-A.

The process of creating the backside cavity involves a timed KOH etch. Due to processing variations, it is difficult to obtain uniform etching across one wafer, let alone across many wafers. There may exist slight wafer thickness variations. During the etch, many bubbles are produced which may stick to the surface, preventing etching in certain

locations which can also lead to a roughening of the surface. Towards the end of an etch, close attention must be paid to the thickness to ensure that cavities are not over or under etched. The use of wafers with an epitaxially grown silicon layer sitting atop a buried boron doped layer solves many of these problems (See Appendix B). The overall process would be identical to the one described above except for the fact that silicon nitride would be used as the protection layer during the KOH etch.

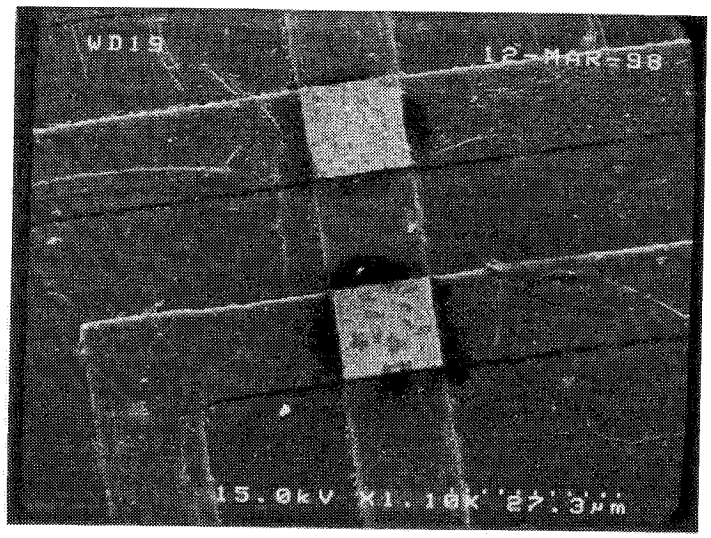
3.2.7.3 Testing and Results

The resistance of the flaps is an important parameter. Miller [1] used a similar process to fabricate flaps, but he was able to use electroplated copper on the order of ten of microns thick as the metal layers. In our fabrication runs, in an attempt to simulate an integration process, the thickness is limited to be an order of magnitude thinner. Hence the resistance for a similar footprint is an order of magnitude higher. This limits the current that can be passed and accordingly, the deflection distance. Therefore, every effort was made to reduce the resistance caused by non-idealities in processing.

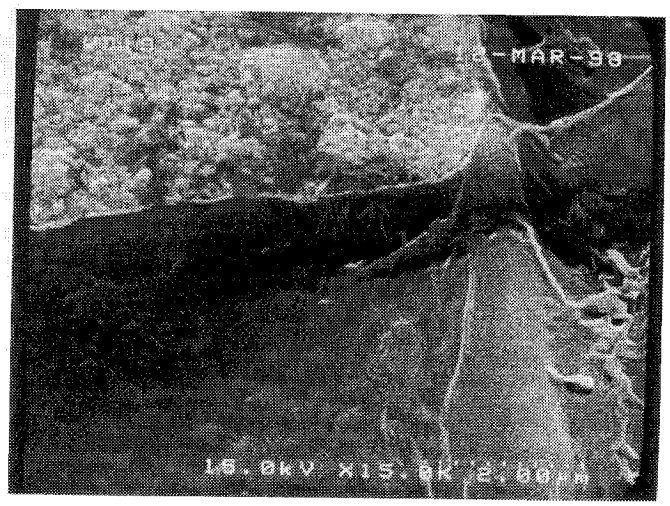
One such non-ideality that caused a significant increase in resistance is the contact resistance between metal 1 and metal 2. Despite overetching the contact holes by as much as 50%, there would often be a thin barrier between the two layers. The result would be a high device resistance upon first testing of the devices with an ohm-meter. After passing a typical operating current (30 mA) through the coils, though, the resistance would immediately drop to more expected levels. This current is believed to burn through any thin insulating barrier.

Even with the burn-in, however, the resistance was often a factor of 3-4 higher than calculated. At first, this was thought to result from a higher resistivity (i.e. poor

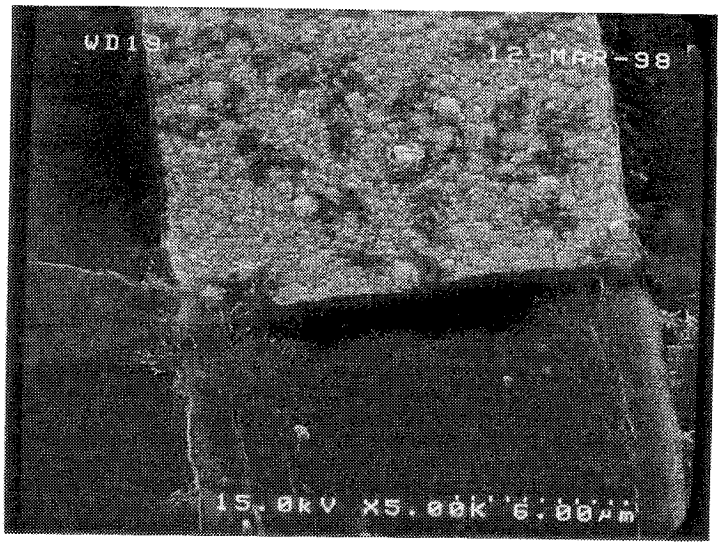
quality) aluminum. However, further inspection showed that the first layer of coils would have a resistance very close to the calculated value. Therefore, it was deemed unlikely that the second metalization would always produce a poor quality film. Instead, upon closer examination with an SEM, it was found that several cracks would occur when the metal 2 would cross over metal 1.



(a)



(b)



(c)

Figure 3.37 Photos showing (a) overview of metal 2 crossing metal 1 (b) close-up of cross-over, and (c) close-up of opposite side of cross-over. Note that in (b) and (c), there are obvious gaps in metal coverage.

As previously mentioned, the flaps were designed to have a resonant frequency on the order of 3 kHz. In the equations, the thickness of the beams is the most critical dimension. Due to the nature of the process, this thickness, which is determined by the bulk silicon etching, is also the parameter over which we have the least control. Therefore, the flaps had a wide range of resonant frequencies, which were measured using a laser vibrometer.

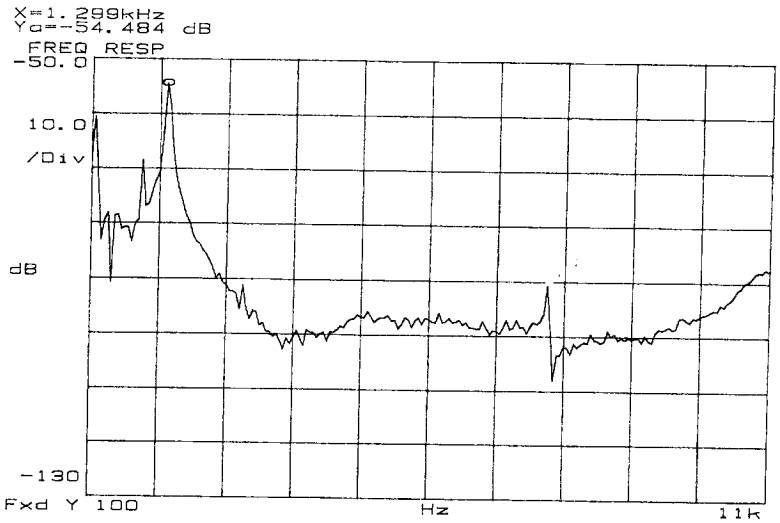


Figure 3.38 Frequency response of a typical bulk actuator. The first order resonant frequency is 1.3 kHz. The higher resonance (7.45 kHz) corresponds to a twisting mode. Other actuators have resonant frequencies ranging as high as 7 kHz.

Deflection measurements and general observations about the flap behavior were made optically. For the surface micromachined flaps, dynamic information could be observed by using a stroboscope to capture fast action. However, our lab's stroboscope does not operate fast enough to capture kilohertz dynamic information. Therefore, a high power LED was used to illuminate the flap.

The actuator was also tested in a UCLA wind tunnel. First order experiments were run to determine if the smaller deflections (approximately 100 μm vs. 2 mm)

achieved by the bulk flap could affect the flow. In order to do this, a vortex generator was placed in the flow to guarantee turbulent flow downstream. The flap was placed after the vortex generator, and a shear stress sensor was placed behind the actuator. The idea behind this experiment was to use the sensor to monitor the flow and determine the difference, if any, in the flow when the flap is in action. A schematic of the experiment is shown in Figure 3.39, and a picture of the actuator/sensor combination is shown in Figure 3.40.

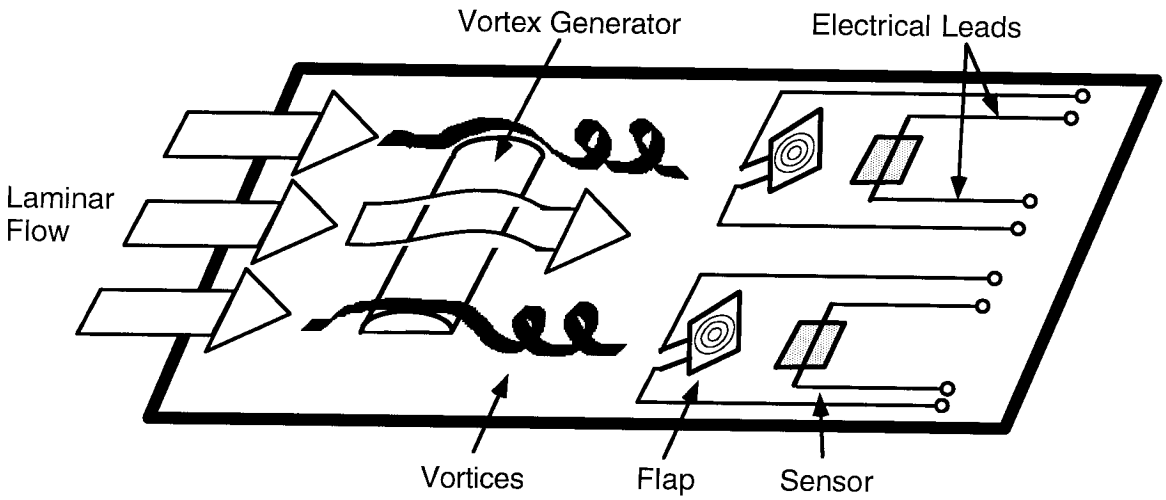


Figure 3.39 Schematic representation of the actuator/flow experiment.

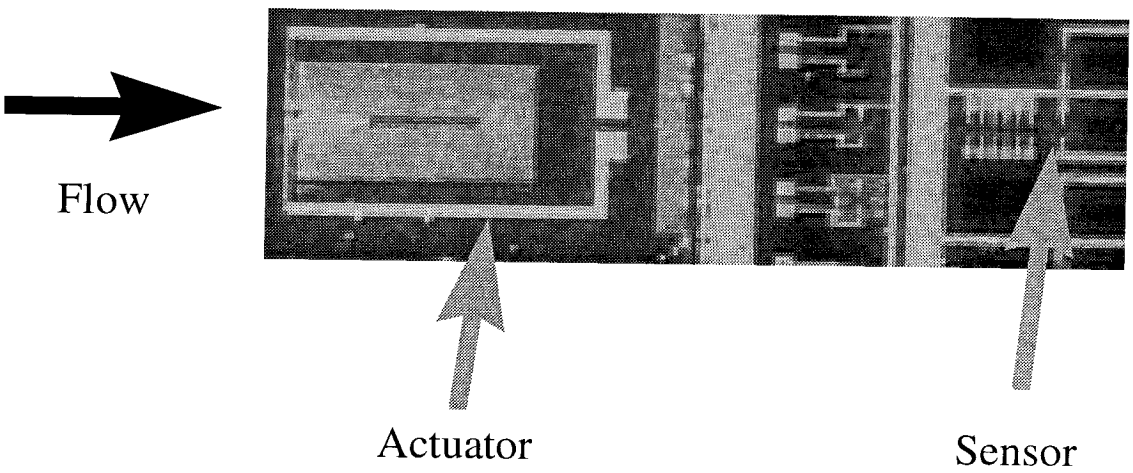
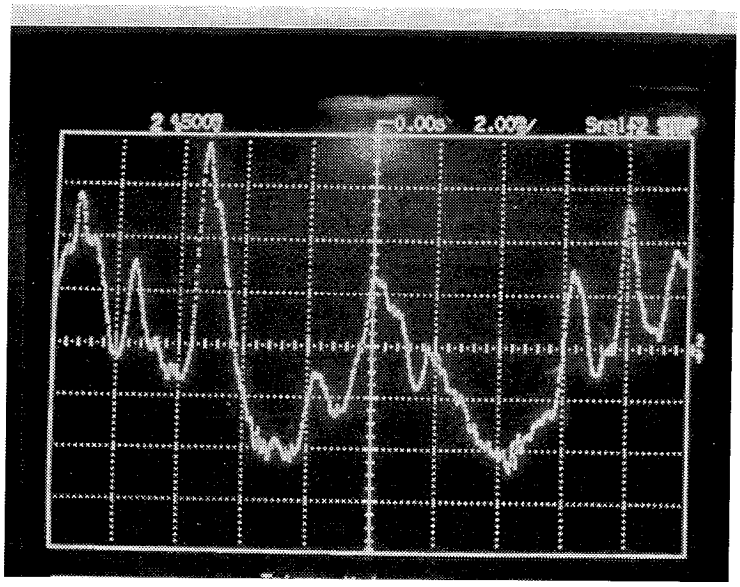
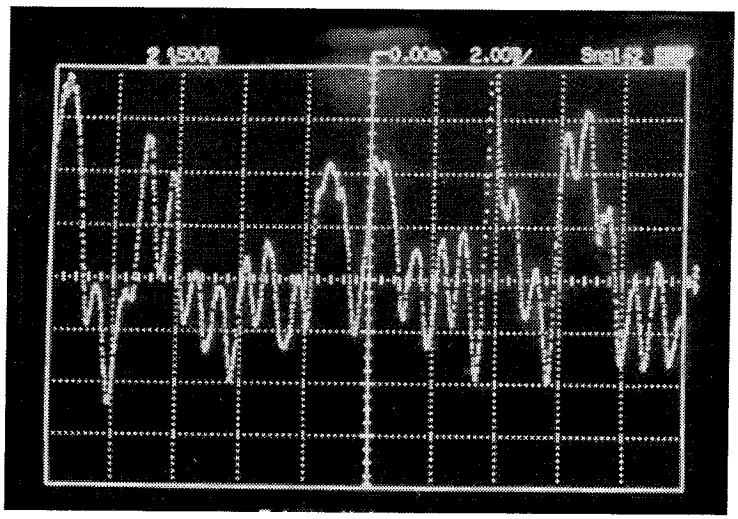


Figure 3.40 Picture showing the actuator/sensor experimental set-up.

The results of the initial experiments were promising, as shown in Figure 3.41. Figure 3.41(a) shows the sensor output when the actuator is not operating. It is a typical response from a shear stress sensor in turbulent flow. When the actuator is used at 3 kHz, a distinct 3 kHz signal then appears on the sensor output, as shown in Figure 3.41(b). Care was taken to ensure that the signal was not a result of any sort of coupling (either thermal or electrical). Further testing needs to be done to determine if the tip velocity effect is still valid using this actuator as is was in Figure 3.32.



(a)



(b)

Figure 3.41 Oscilliscope output of the shear stress sensor. (a) without and (b) with a 3 kHz actuator movement.

References

[1] Miller, R., *Microfabricated Electromagnetic Flap Actuators and Their Applications*, Ph.D. Thesis, The California Institute of Technology, 1997.

[2] Liu, C., *Silicon Micromachined Sensors and Actuators for Fluid Mechanics Applications*, Ph.D. Thesis, The California Institute of Technology, 1995.

[3] Tai, Y.-C. and Muller, R., "Fracture Strain of LPCVD Polysilicon," *Technical Digest, IEEE Solid-State Sensors and Actuators Workshop*, Hilton Head Island, SC, USA, pp. 88-91, 1988.

[4] Xing Yang, personal communication.

[5] Alley, R., Cuan, G., Howe, R., and Komvopoulos, K., "The Effect of Release-Etch Processing on Surface Microstructure Stiction," *Technical Digest, IEEE Solid-State Sensor and Actuator Workshop*, Hilton Head, SC, USA, pp. 202-207, 1992.

[6] Liu, C., Tsao, T., and Tai, Y.-C., "A High-Yield Drying Process for Surface Microstructures using Active Levitation," *1997 International Conference on Solid-State Sensors and Actuators*, Chicago, IL, USA, June 1997.

[7] Mastrangelo, C. and Hsu, C., "A Simple Experimental Technique for the Measurement of the Work of Adhesion of Microstructures," *Technical Digest, IEEE Solid-State Sensor and Actuator Workshop*, Hilton Head, SC, USA, pp. 208-212, 1994.

[8] Mastrangelo, C. and Saloka, G., "A Dry-Release Method Based on Polymer Columns for Microstructure Fabrication," *Proceedings from IEEE Micro Electro Mechanical Systems*, Fort Lauderdale, Fl, USA, pp. 77-81, 1993.

[9] Mulhern, G., Soane, D., and Howe, R., "Supercritical Carbon Dioxide Drying of Microstructures," *Proceedings, 1993 International Conference on Solid-State Sensors and Actuators,* Yokohama, Japan pp. 296, 1993.

[10] Lee, G., *Control of a Delta-Wing Aircraft by Micromachined Sensors and Actuators*, Ph.D. Thesis, University of California, Los Angeles, 1998.

[11] Lee, G. B., Ho, C. M., Jiang, F., Liu, C., Tsao, T., and Tai, Y. C., "Distributed Flow Control by MEMS", *ASME 1996 International Mechanical Engineering Congress and Exposition*, Atlanta, Georgia, USA, Nov. 17-22, 1996.

[12] Lee, G. B., Ho, C. M., Jiang, F., Tsao, T., and Tai, Y. C., " Control of an Aircraft by M^3 systems", *Bulletin of the American Physical Society, 50th Annual Meeting of the Division of Fluid Dynamics*, Nov. 23-25, 1997, San Francisco, CA.

[13] Lee, G., Jiang, F., Tsao, T., Tai, Y.-C., Ho, C.-M., "Macro Aerodynamics Devices Controlled by Micro Systems", *1997 IEEE Aerospace Conference*, Feb. 1-8, 1997, Snowmass, Colorado.

[14] Wang, X.Q., Yang, X., Walsh, K., and Tai, Y.-C., "Gas Phase Silicon Etching with Bromine Trifluoride," *1997 International Conference on Solid-State Sensors and Actuators*, Chicago, IL, USA, June 1997.

[15] Judy, J., Muller, R., and Zappe, H., "Magnetic Microactuation of Polysilicon Flexure Structures," *Technical Digest, IEEE Solid-State Sensor and Actuator Workshop*, Hilton Head, SC, USA, pp. 43-48, 1994.

[16] Liu, C., Tsao, T., Tai, Y.-C., and Ho, C.-M., "Surface Micromachined Magnetic Actuators," *Proceedings from IEEE Micro Electro Mechanical Systems*, Oiso, Japan, pp. 57-56, 1994.

[17] Liu, C., Tsao, T., Tai, Y.-C., Leu T.-S., Ho, C.-M., Tang, W., and Miu, D., "Micromachined Permalloy Magnetic Actuator for Delta-Wing Control," *Proceedings from IEEE Micro Electro Mechanical Systems*, Germany, pp. 332-335, 1995.

[18] Liu, C., Tsao, T., Tai, Y.-C., Liu, W., Will, P., and Ho, C.-M., "A Micromachined Permalloy Magnetic Actuator Array for Micro Robotics Assembly Systems," *1995 International Conference on Solid-State Sensors and Actuators (Transducers 95)*, pp. 328-331,1995.

[19] Tsao, T., Liu, C., Tai, Y.-C., and Ho, C.-M., "Micromachined Magnetic Actuator for Active Fluid Control," *ASME Application of Microfabrication to Fluid Mechanics 1994*, Chicago, IL, USA, pp. 31-38, 1994.

[20] Monk, D., Soane, D., and Howe, R., "Sacrificial Layer SiO_2 Wet Etching for Micromachining Applications," *Proceedings of the 6th International Conference on Solid State Sensors and Actuators*, pp. 647-650, San Francisco, 1991.

[21] Timoshenko, S., "Analysis of Bi-metal Thermostats," *Journal O.S.A. and R.S.I 11*, pp. 233-255, September 1925.

[22] Ho, C-M., Tung, S., Lee, G., Tai, Y.C., Jiang, F., and Tsao, T. "MEMS- A Technology for Advancements in Aerospace Engineering", *AIAA 35th Aerospace Sciences Meeting & Exhibit*, AIAA 97-0545, Jan. 6-10, 1997, Reno, NV.

Chapter 4

Sacrificial Layer Technology

4.1 Importance of Sacrificial Layer

The choice of a sacrificial layer is often the most important step in a surface micromachined process – or perhaps more specifically, the choice of the release agent is very critical. An ideal release agent should have a high degree of selectivity over all other materials being used and should etch the sacrificial layer at an acceptable rate. In this chapter, we describe some results from investigating the uses of copper and polysilicon as sacrificial layers.

In the search for post-IC micromachining techniques, it is often desirable to have a sacrificial layer that can be deposited at a low temperature. If MEMS processing is to be attempted on a completely finished IC wafer, all subsequent steps must be done at a temperature lower than approximately 450° C, above which the aluminum metalization, which is prevalent in IC fabrication, will degrade. The ideal low temperature sacrificial layer should be easy to deposit at a reasonably fast rate and should be etched by a chemical means (either wet or dry) with a high degree of selectivity over other semiconductor or MEMS materials. Electroplating of copper is a widely-available, simple, low cost, and room temperature process. Electroplated copper can be thick, be highly conductive, and have low stress. In addition, it can be undercut quickly and deeply with relatively mild solutions (such as acetic acid/hydrogen peroxide) that do not attack many other films. Therefore, using copper as the sacrificial layer results in many possibilities for post-IC MEMS processes.

Sometimes, the steps following the sacrificial layer deposition need to be conducted at high temperature. If this is the case, copper (and most other metals) cannot be used as the sacrificial layer. This sort of constraint is usually placed upon an interweaved integrated process where IC and MEMS steps are intertwined. The traditional sacrificial layer, oxide, cannot be used because it is also used as the interlayer dielectric material in the IC process. Therefore, the sacrificial layer etch would most certainly destroy the metalization as well. One potential choice for a sacrificial layer is polysilicon, which is readily available during standard IC processes, and therefore, by definition, is IC compatible. There are several etchants that can be used to release the structural layer (typically silicon nitride), but the most promising appears to be TMAH. One important point that must be considered when choosing the release agent is that it must have a high selectivity of etching polysilicon over oxide. If it does not, the protection oxide can be etched away, leaving the gate polysilicon vulnerable to being etched.

4.2 Copper Sacrificial Layer

4.2.1 Set-up and Devices

Initial results of using copper as a sacrificial layer have been promising [1], and Wagner [2] has developed devices using this technology. Copper electroplating is a well understood process. The copper electroplating was done using a copper sulfate/sulfuric acid solution (6 L H_2O:1500 g $CuSO_4 \cdot 5H_2O$: 300 g H_2SO_4). Typical plating current densities are on the order of 5 mA/cm^2, which results in a plating rate of approximately 6 μm/hr. This technique can be used to plate up to tens of microns of copper.

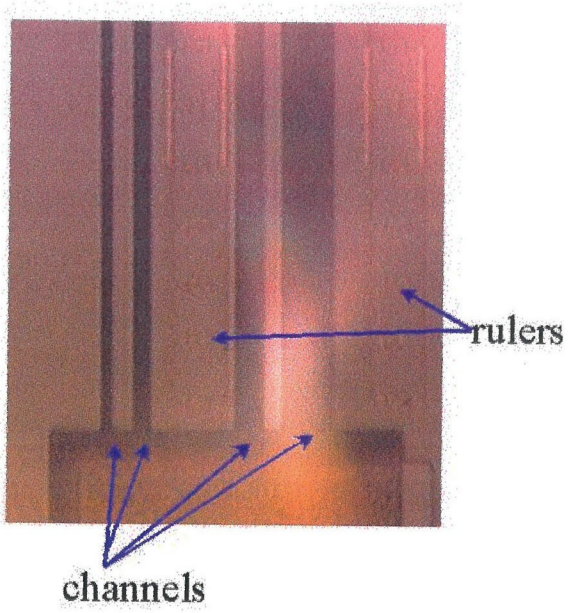
We have successfully used multiple solutions (including acetic acid/hydrogen peroxide, sulfuric acid, and nitric acid solutions) for etching the sacrificial layer. Many other acids can be used, as long as the etching chemical reaction is thermodynamically favorable (i.e. negative Gibbs free energy).

$$\Delta G = -nFV$$

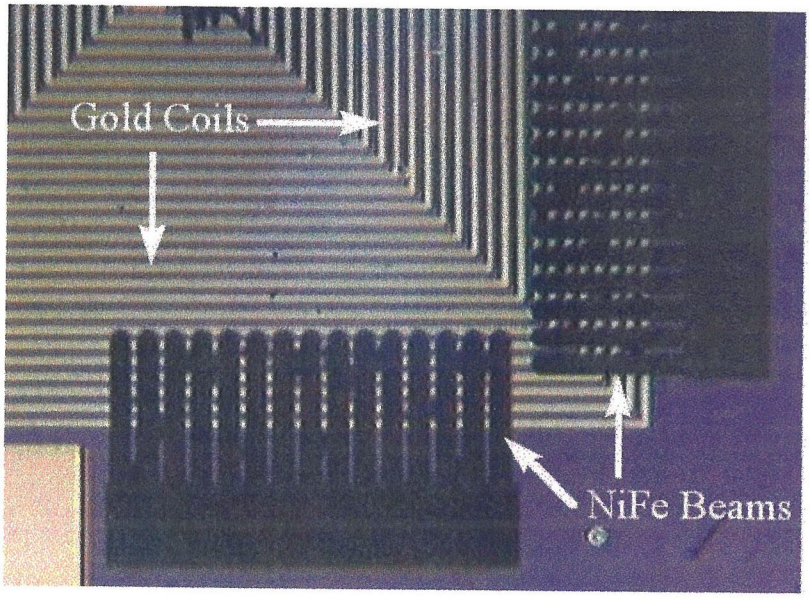
nF = coulombs of charge
V = overall cell voltage

Our test structures were fabricated by depositing a 1 μm LTO structural layer atop 2 μm of electroplated copper. Etching holes were opened atop the structural layer, and gold was deposited and patterned into rulers for use in measuring etch distances. The structures included channels of differing (10, 20, 40, and 80 μm) widths (Figure 4.1), diaphragms, plates, etc.

In addition to the test structures, we have freed permalloy ($Ni_{80}Fe_{20}$) cantilever beams suspended over insulated gold coils (Figure 4.1(b)). This process demonstrates the compatibility of electronics metallization with copper sacrificial layer technology. We have also fabricated LTO flaps with gold coils both the top and bottom (Figure 3.26). This entire process can be done after CMOS fabrication. Details on such structures are presented in Chapter 3.



(a)



(b)

Figure 4.1 Pictures showing (a) top view of etch channel structures, and (b) top view of permalloy coils over gold coils.

A video microscopy system was used to optically measure etch distance as a function of time. Some of the etching solutions produced heat in an exothermic reaction when the reactants were originally mixed. To filter out gross temperature effects, all solutions were allowed to cool to room temperature before use.

The etch front was rarely perfectly straight. In order to determine an etch rate, an equivalent etch front was used. At a given time, the real etch front represents the fact that a given mass of copper has been etched. The equivalent etch front, at the same time, represents the same mass of copper that would have been etched if the etch front were perfectly straight. As will be seen later, this is an essential tool, as many etch fronts have dramatically different forms.

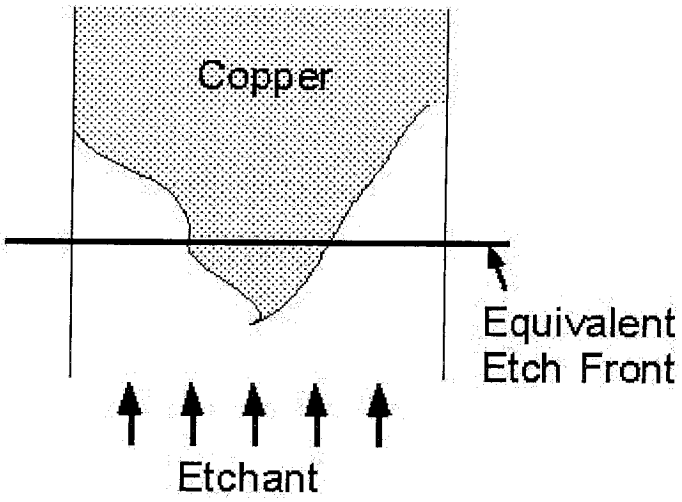


Figure 4.2 Drawing showing approximation of an equivalent etch front.

Hydrogen peroxide is often used in conjunction with acids to etch the sacrificial copper layer. Over time, hydrogen peroxide will decompose into water and oxygen gas and therefore become reduced in concentration and effectiveness. As with all chemical

reactions, this one is highly temperature dependent. For example, hydrogen peroxide is used in conjunction with sulfuric acid (the solution is referred to as "piranha") at an elevated temperature of 120° C for wafer cleaning in semiconductor processing. Fresh hydrogen peroxide must be used every few hours. In the copper etching experiments, however, the solution was placed atop a heat sink. Even in the cases that an exothermic chemical reaction occurred upon mixing of reactants, the etch solution was always cooled to room temperature within a manner of minutes. Also, the concentrations used were relatively low compared with the extreme example of piranha mentioned above. By Le Chatelier's principle, this keeps the kinetics favorable (i.e. keeps the dissociation of hydrogen peroxide into water at a relatively low rate). In addition, having large amounts of water dilutes the heat generated.

4.2.2 Experimental Results

The results show that channels with cross sections as small as 2x2 μm^2 can still be etched, that the etchants do not appreciably chemically damage the structural layer after 4 hours of etching and that diaphragms can easily be formed. Bubble generation (Figure 4.3) plays a large part in the etching reactions. Data shows that in certain cases the etch reaction does not become diffusion limited (Figure 4.5), even after long distances. This is due in large part to the bubble-induced forced convection. However, care must be taken to not choose a violently bubbling reaction that could result in fractured diaphragms. Another interesting effect is the edge effect (Figure 4.3), which results in the edges etching faster than the middle of the channels. It is believed that the strain energy released from the sacrificial layer/structure layer interface is transferred into the chemical reaction.

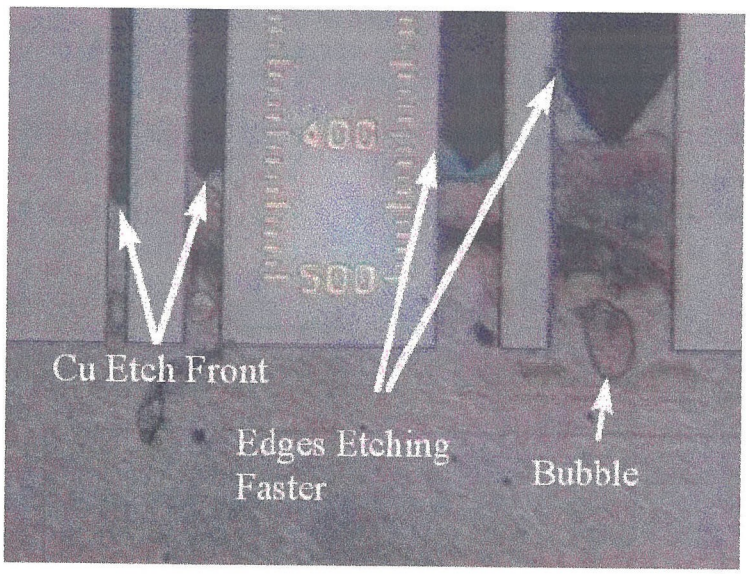


Figure 4.3 Etch Channels. Top view of four etch channels during etching. The widths are 10, 20, 40, and 80 μm. The height is 2 μm. Note the sharp tip shape of the etch front and the bubble generation. In the wider channels, bubbles enhance etching by forced convection.

Copper placed in pure acetic acid will etch very slowly. Hydrogen peroxide, when added to the solution, will eventually decompose into water. This decomposition process requires hydrogen ions, which are provided by dissociation of an acid, and electrons, which are provided by the dissolving of solid copper into Cu^{2+} ions. Thermodynamically, this is a favorable reaction, as seen by the two electrochemical half reactions:

$Cu(s) \rightarrow Cu^{2+} + 2e^-$ V = - 0.340 V

$H_2O_2 + 2H^+ + 2e^- \rightarrow 2H_2O$ V = +1.776 V

$Cu(s) + H_2O_2 + 2H^+ \rightarrow Cu^{2+} + 2H_2O$ V = +1.436 V

The effect of acetic acid/hydrogen peroxide concentration on etching is shown in Figure 4.4. It is worth noting that the lowest concentration (chemically milder) solution etches the fastest. Also, looking at the data for the 10% solution, the undercutting etch rate is approximately 2 μm/min in the initial stages. Compare this with a slower etch rate

of approximately 1.2-1.3 μm/min when etching bulk copper. One possible explanation for this phenomena is the fact that the standard enthalpies of formation, ΔH_f^0, for the dissociation of hydrogen peroxide into water and copper into Cu^{2+} are -191.17 and 64.77 and kJ/mol respectively. Therefore, the overall reaction is exothermic. During channel etching, this generated heat is kept more localized and is not diffused away as easily as during bulk etching, where there is an abundant quantity of water available to both convect and diffuse away the heat. Therefore, the chemical reaction rate at the etch front is increased.

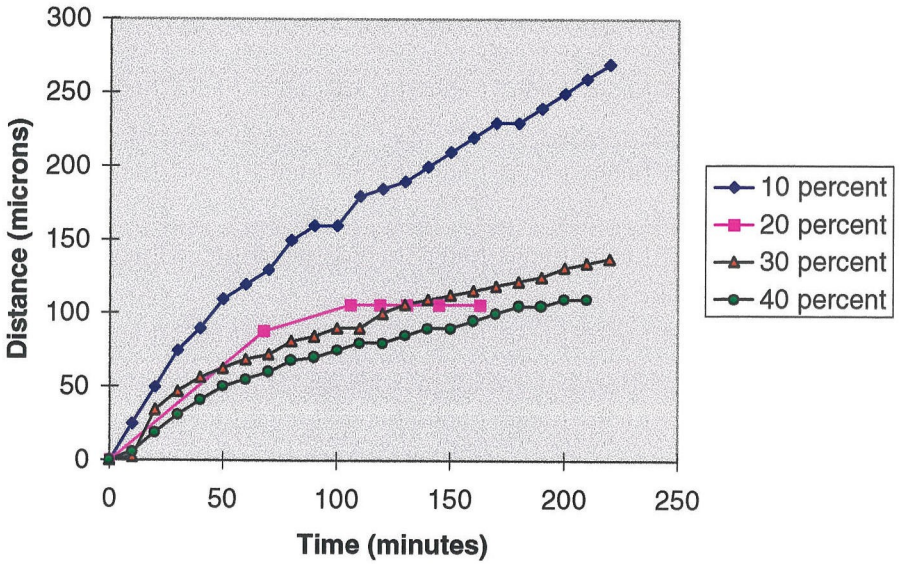


Figure 4.4 Concentration Effect. Graph showing the acectic acid/hydrogen peroxide concentration effect on etching for a 10 μm wide channel. Note that the lowest concentration etches the fastest, while the highest concentration etches the slowest. Compare with HF etching of PSG [8,9], in which case higher concentrations etch faster.

An example of the width effect is shown in Figure 4.5. Wider channels often do not exhibit diffusion limited etching. The two wider channels (80 μm and 40 μm) have an approximate etch rate of 1.5 μm/min when using a 30% solution, which corresponds to the etch rate shown in Figure 4.4 for the initial etching of a 10 μm wide channel using

a 10% solution. The linear nature of both of these etch curves implies that the system is chemical reaction rate limited. Also, the geometrical independence of the etch rates further validates this postulate.

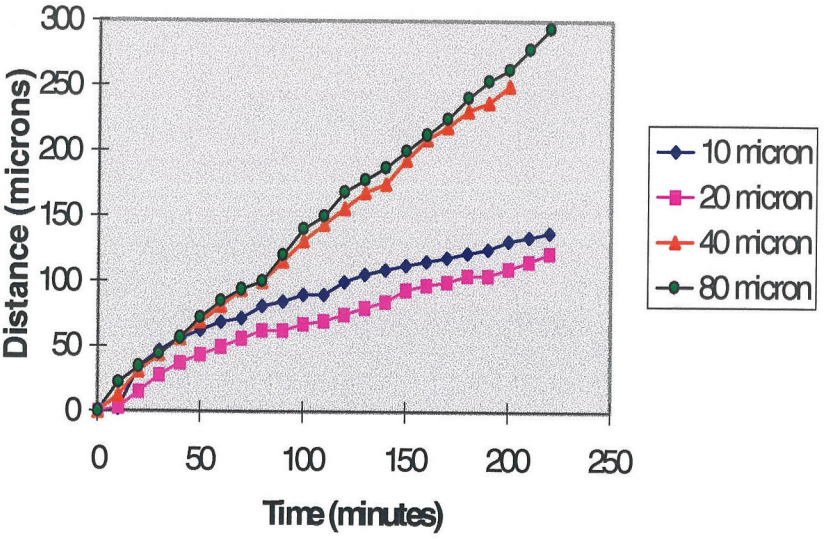


Figure 4.5 Width Effect. Graph showing etch distance vs. time for the channels shown in Fig. 3. Note the linear relationship for the wider channels. The etchant was acetic acid and hydrogen peroxide in water (30:30:100).

For extremely long channels, a sulfuric acid/hydrogen peroxide system etches much more quickly than an acetic acid/hydrogen peroxide system does [Figure 4.6]. As expected, the Gibbs free energy is also negative for such an etching set-up. It is interesting to note that sulfuric acid by itself will etch copper very slowly, and only to the extent that the water in the acid solution slowly produces electrons and hydrogen ions. This can be seen if only the first two chemical reactions below are considered.

$Cu(s) \rightarrow Cu^{2+} + 2e^-$ V = - 0.340 V
$SO_4^{2-} + 4H^+ + 2e^- \rightarrow H_2SO_3 + H_2O$ V = +0.200 V
$H_2O_2 + 2H^+ + 2e^- \rightarrow 2H_2O$ V = + 1.436 V

$Cu(s) + H_2O_2 + SO_4^{2-} + 6H^+ + 2e^- \rightarrow Cu^{2+} + H_2SO_3 + 3H_2O$ V = +1.296 V

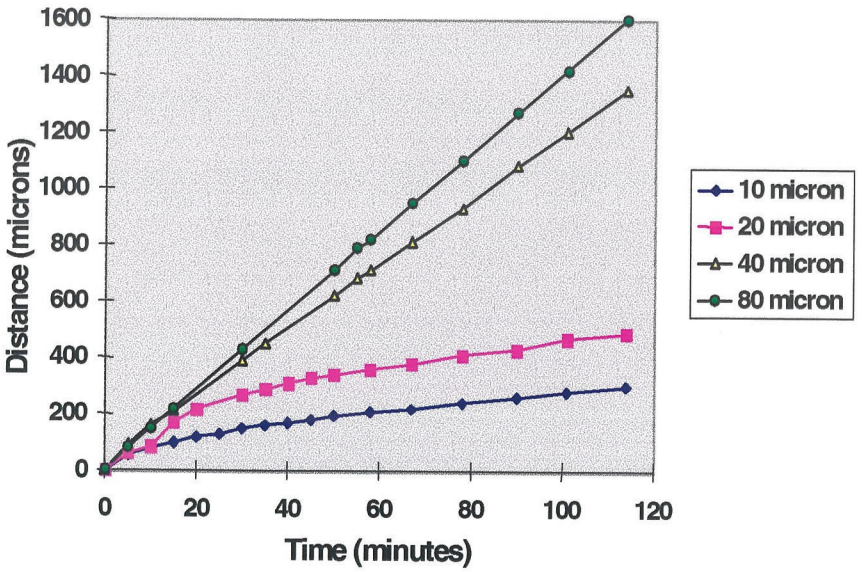


Figure 4.6 Long channel etching. Graph showing etch distance vs. time for channels similar to those shown in Fig 3. Dilute sulfuric acid (5%) was the etchant. Note the extremely fast etch rate (~ 1 mm/hr for the widest channel).

Nitric acid [Figure 4.7] was also investigated as a potential etch solution. Unlike acetic and sulfuric acids, nitric acid does not need the addition of hydrogen peroxide to etch copper. This can be seen in the following chemical reactions.

$Cu(s) \rightarrow Cu^{2+} + 2e^-$ V = -0.34 V

$3e^- + NO_3^- + 4H^+ \rightarrow NO + 2H_2O$ V = +0.96 V

$3Cu + 2NO_3^- + 8H^+ \rightarrow 3Cu^{2+} + 2NO + 4H_2O$ V = +0.62 V

The nitric acid solution etched the channels very fast. Even without the hydrogen peroxide, the solution created much more bubbling than any other solution. This excessive bubbling would often crack the structural layer. Often, data could not be taken for more than one hour, as the channel would crack, allowing solution to enter at those locations. In the case of diaphragms, entire structures could be broken and lifted off of the surface [Figure 4.8]

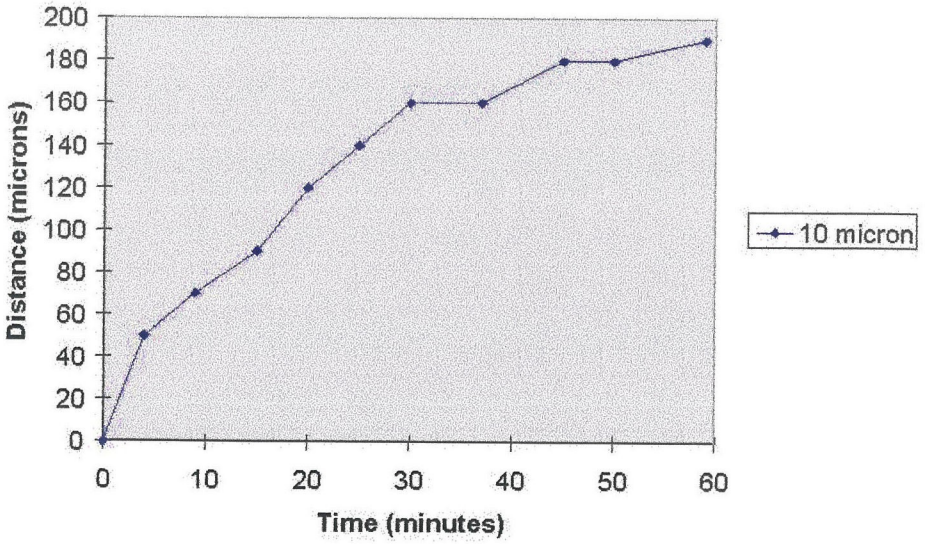


Figure 4.7 Nitric Acid Etching. Graph showing an example of 10% concentration nitric acid etching of a 10 micron wide channel.

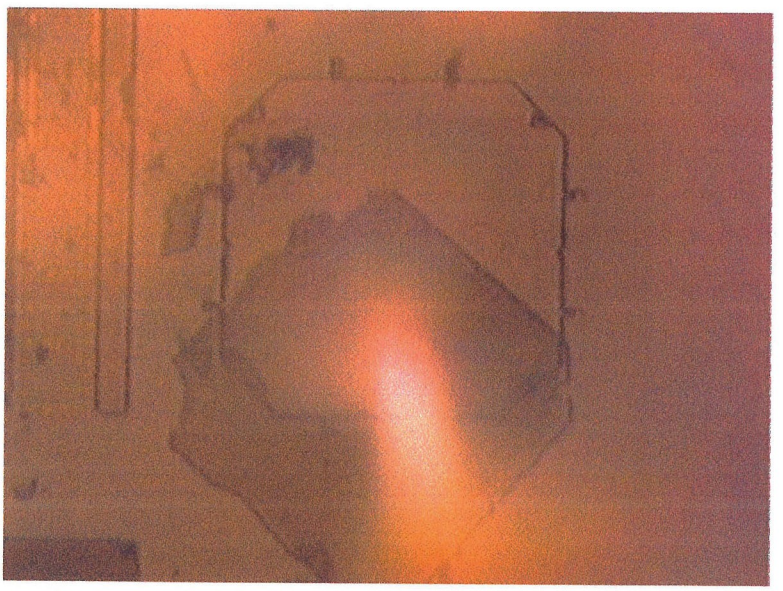


Figure 4.8 Cracked Diaphragm. Picture showing diaphragm lifted off by excessive bubbling, which cracked the edges.

There is one note that should be mentioned regarding the use of copper as a sacrificial layer in a post-IC MEMS process. Recent developments in the field of copper metalization by large IC research centers such as IBM, Motorola, and Texas Instruments have shown that copper will soon replace aluminum as the metal of choice in IC

fabrication. Copper has a much lower resistivity than aluminum and therefore copper lines have a much lower RC delay than comparably sized aluminum lines.

These developments could imply the use of the copper both as metalization as well as sacrificial layer – provided, of course, that the metal lines were properly protected during the release etch. If the etchant chosen were the acetic acid/hydrogen peroxide combination, this concern is easily alleviated since materials such as silicon dioxide or even photoresist could be used as protection. The combination of using copper for both metal lines as well as sacrificial layer could result in an overall process that is simpler and shorter – this results in a less expensive, higher yielding, more robust design.

As a final aside, if copper proves to be the dominant metal used in the IC industry, future work on developing sacrificial layer technology could refocus on using aluminum. Schmidt et al [3] have used aluminum as a sacrificial layer in a purely MEMS process, but it has not been attempted in an integrated process. The exact same arguments for using copper as a sacrificial layer in an aluminum metalized process can be applied for using aluminum in a copper metalized process.

4.3 Polysilicon Sacrificial Layer

Traditionally, people studying surface micromachining have used some form of oxide as the sacrificial layer and polysilicon as the structural layer. The primary reasons for this include availability of both processes in standard IC fabrication facilities and the infinite selectivity the HF acid has of etching oxide over polysilicon.

Recently, however, there has been an interest [4,5] in using polysilicon as the sacrificial layer. While it is not etched as quickly as oxide can be in HF acid, using polysilicon has the obvious advantage that oxide can potentially still be used in the other

processing steps. This opens up the door to on-chip integration of MEMS with electronics since VLSI processes depends on oxide as inter-metal insulation. Also, the potential exists that the etchant used to dissolve the polysilicon may not attack silicon nitride or other structural layers. This will be discussed below.

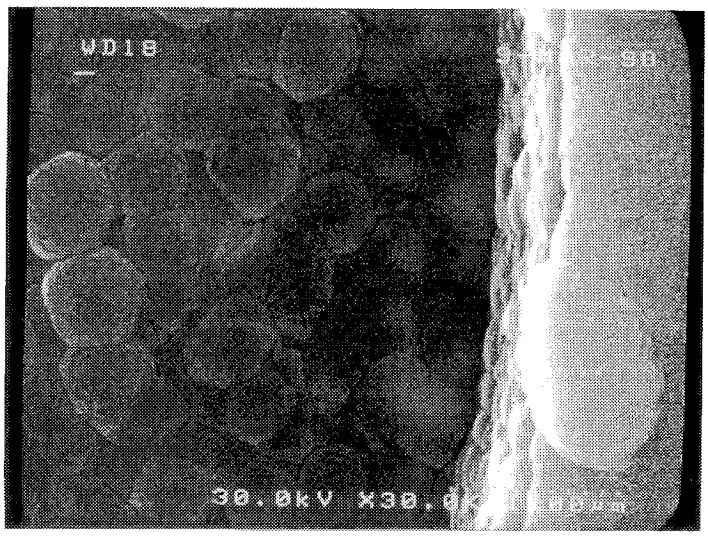


Figure 4.9 SEM showing close-up of the bottom and edge of a etch hole. Note that the polysilicon grain sizes are on the order of 300-500 nm in diameter.

All of the standard etchants used to etch bulk silicon can also be used to etch polysilicon. However, bulk etching is often done in the beginning of a process, when minimal processing, particularly metal deposition, has been done. Therefore, care must be taken to select an appropriate polysilicon etch when electronics integration issues are important. EDP and KOH attack most metals, especially aluminum, rather rapidly. Also, the few films that can stand such etchants, such as thermal silicon dioxide and LPCVD silicon nitride, can not be deposited on top of metal for obvious reasons. Due to the presence of HF acid, HNA will slowly attack both silicon nitride and silicon dioxide. Also, the presence of nitride acid will attack many metals. Therefore, neither HNA nor

EDP should be used to etch polysilicon on a wafer with electronics. However, some calibration was done using both etchants for completeness sake. TMAH attacks aluminum, but not other metals such as copper and gold. If aluminum is present, it can be protected by LTO for a limited time.

All of the etch tests done in this section were done on the hybrid-type structures described in Chapter 3. The polysilicon is 1 μm thick. Etch rates are sometimes different depending on whether the etch front is measured from an etch hole ("point entrance") or from a side ("line entrance").

Four different HNA recipes were used. In each of these, the HF concentration was kept low so that the actual sacrificial layer oxide etching would be the rate limiting process. The etching would then probably become a diffusion limited process. If the HNA concentration is kept low, the oxidation rate is the limiting process, and all the many factors that affect oxidation rates (such as electron concentration, silicon "orientation" direction and defects) become important. The initial basis for choosing these particular volumes comes from a standard recipe used at UC-Berkeley for an etchant that has both a reasonable etch rate and a decent selectivity.

Relative Volume HNO_3	Relative Volume H_2O	Relative Volume HF	Etch Rate – Edge (μm/min)	Etch Rate – Hole (μm/min)
25.2	11.7	1	0.10	0.16
25.2	11.7	2	0.47	0.44
25.2	11.7	3	0.64	0.57
25.2	11.7	4	0.91	0.69

Table 4.1 HNA recipes used in etch tests.

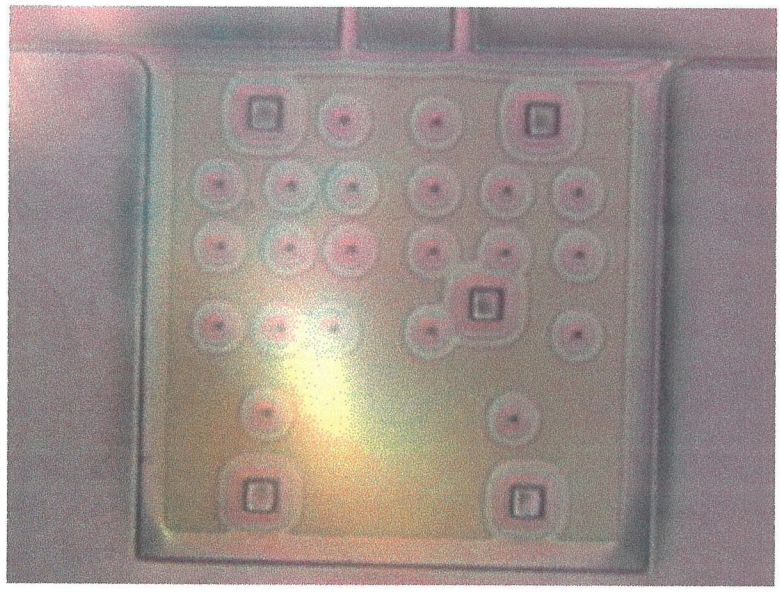


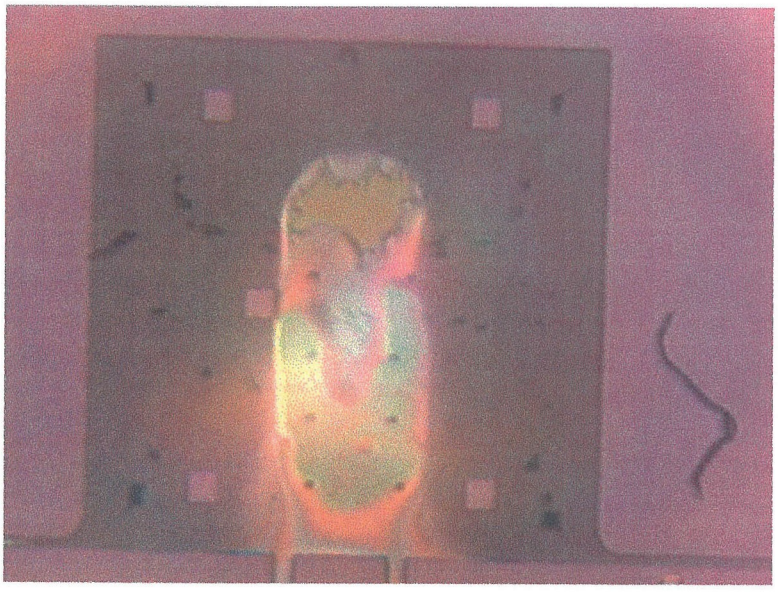
Figure 4.10 Result of HNA etching of polysilicon sacrificial layer.

As would be expected, the etch rate increases with increasing HF concentration. Another interesting effect is the fact that the etching along an edge is faster than from a point source. In other words, there is a geometry effect (either a width effect or an opening effect), which implies that this is a diffusion rate limited process, as hypothesized when choosing HF-poor etching solutions.

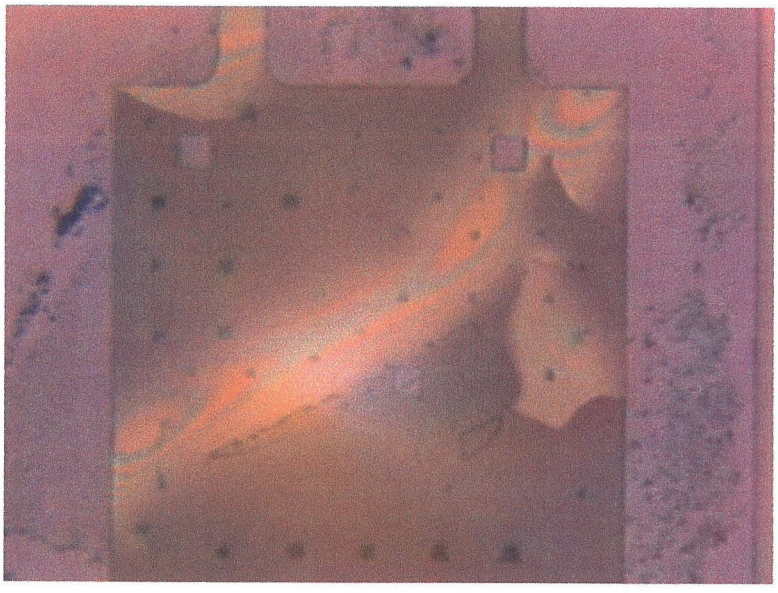
One drawback to using HNA is the fact that oxide cannot be used as the buffer layer between the silicon substrate and the polysilicon sacrificial layer. The HF acid will slowly attack the oxide, and when the oxide is completely etched away, the HNA will start etching the bulk silicon. New solution pouring into the etch hole has a tendency to diffuse straight down, rather than sideways. Therefore, much of the etchant is wasted and undercutting is minimized [Figure 4.10], even after an extremely long time etching.

EDP was also tested, although those tests were all negative. Most devices (>95%) cracked during EDP etching. None were freed. One interesting effect seen during the EDP etch was the fact that the etch fronts were often neither smooth nor symmetric.

(Figure 4.11) This is probably due to the highly anisotropic nature of EDP etching combined with the random orientation of polysilicon grains. It is worth noting that this effect is not seen during TMAH etching, which has a lower selectivity.



(a)



(b)

Figure 4.11 Pictures showing (a) uneven etch fronts and (b) cracking resulting from the use of EDP to etch the polysilicon sacrificial layer.

When etching bulk <100> silicon, the TMAH concentration affects the etch rate significantly. In fact, up to a point, lowering the concentration will actually increase the etch rate [6]. This effect is also seen in sacrificial layer etching. Another interesting effect is the fact that the etching rate remains constant over a time period of 5 hours, which was the limit of our testing. This implies that the etch process is not limited by diffusion. Also, the width of the channel does not appear to affect the etch rate, nor does a continuously increasing "channel width". For example, when the etch front is a circle (i.e. when the etchant is introduced through a single point), the rate at which the radius increases is constant with time. This is equivalent to a constantly increasing width channel. This further validates the chemical rate limitation of the reaction – enough etchant is always available at the etch front, even when the channel width is increasing with distance.

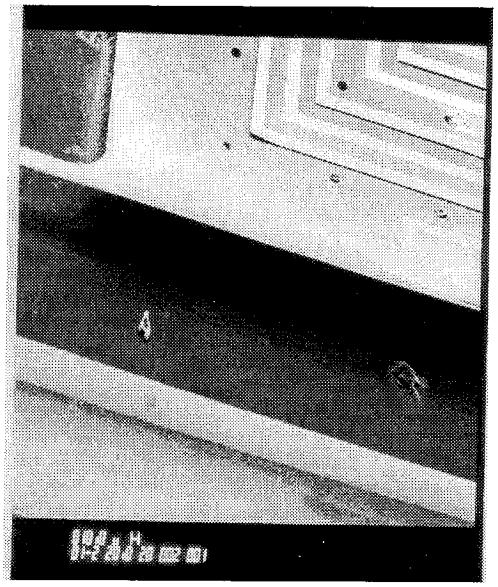


Figure 4.12 SEM showing close-up of an ideal case of TMAH etching.

First, testing was done to try an quantify the temperature effect. Results for three different temperatures are shown in Table 4.2. The concentration is kept at 25% for all of

these experiments. The etch distances are linear with time, thus indicating that this is not a diffusion limited process.

Temperature (°C)	Etch Rate (μm/min)
90	1.46
80	0.91
60	0.55

Table 4.2 Etch rate of 1 μm polysilicon sacrificial layer in various temperature TMAH solutions.

Several experiments were run to determine if the concentration effect for etching bulk silicon also held for sacrificial layer etching. The temperature was kept constant for these experiments at approximately 63-65 °C. There apparently is a local maximum for etch rate between 12.5% and 25% and also between 0% and 12.5%. More detailed results need to be obtained to fully quantify this result.

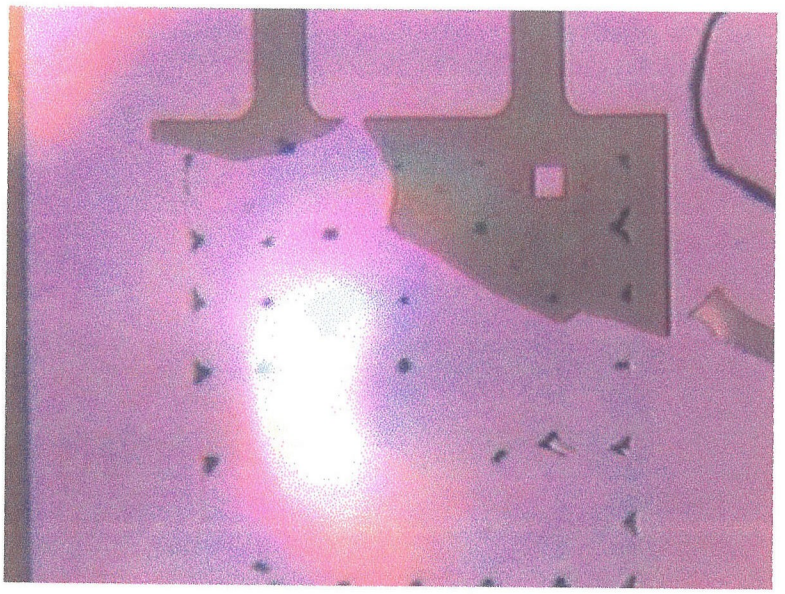
TMAH Concentration (%)	Etch Rate (μm/min)
25	0.45
15	0.57
12.5	0.48
5	0.52

Table 4.3 Etch rate of 1 μm polysilicon sacrificial layer in various concentration TMAH solutions.

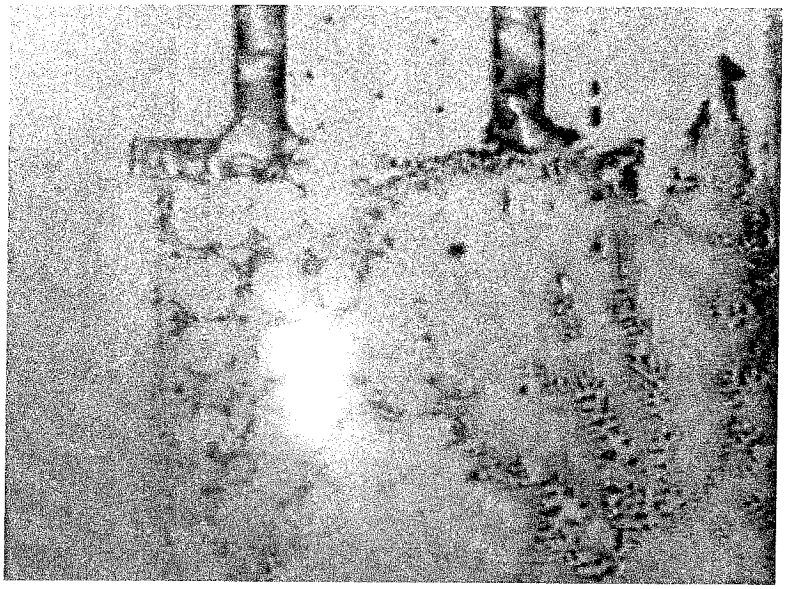
Etch tests were also run with heavily n-doped polysilicon in the hopes that it would etch faster in a TMAH solution than undoped polysilicon would (the opposite effect of heavily p-doped silicon etching at a much slower rate). Unfortunately, this did not prove to be the case, and the etch rates were more or less the same. The polysilicon was doped by diffusion of phosphorous from one side in the same way it was doped in

the first generation actuator. Therefore, the doping density should be the solid solubility of phosphorous in polysilicon. The sheet resistance of the film was approximately 60 Ω/ , which, for a polysilicon thickness of 1 μm, yields a concentration of approximately 10^{19} cm^{-3}.

Despite the promise that using TMAH has, care must be taken to appropriately design the etch hole spacing. Poor choices in spacing can lead to cracking or even complete destruction of the flap [Figure 4.13].



(a)



(b)

Figure 4.13 (a) A completely destroyed flap. (b) Dirty residue is sometimes left behind after etching.

In addition to TMAH, both XeF_2 and BrF_3 were also used to etch polysilicon due to their high selectivities over metals, and, just as importantly, due to the fact that they're gas-phase etchants and therefore don't cause release stiction. The results of the xenon difluoride etching, which was done at UCLA's microfabrication facility, were not positive. In addition to etching the polysilicon, the silicon nitride was also etched at an appreciable rate. A green film deposited upon the gold bonding pads. It did not appear to affect the resistance, however.

Unfortunately, it was difficult to repeat the etching parameters due to the manual controlling of all parts of the system. In the end, etching with xenon difluoride showed a high degree of potential, but it was decided to build a bromine trifloride/xenon difluoride system at Caltech. The main advantage of having such capabilities "in-house" is control – in terms of both contamination and ability to vary the system (and hence, etching) parameters.

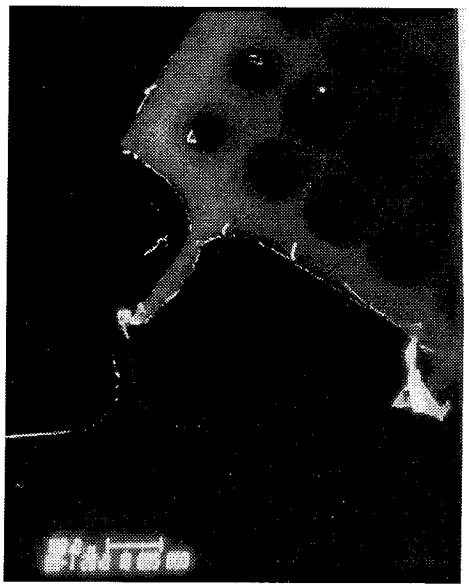
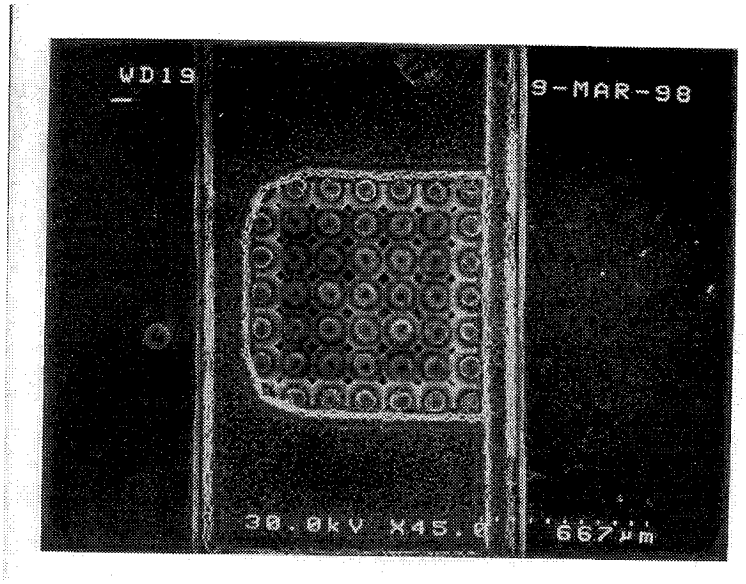
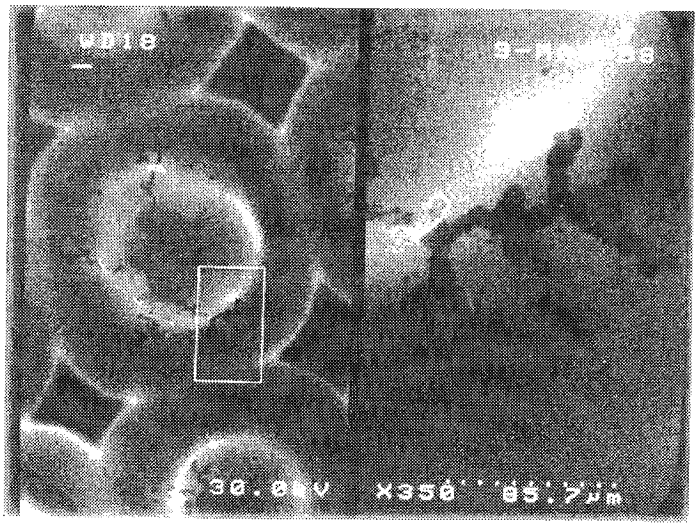


Figure 4.14 SEM showing remnants and residue of a flap completely etched away using XeF_2.

The BrF_3 etching system has been used to etch bulk amounts of silicon very efficiently [7] with high selectivity over other thin films such as silicon nitride and polysilicon. However, when used to undercut polysilicon, a distinct problem arises: in the standard procedure, the primary etch chamber is kept at approximately 1-2 Torr of BrF_3. The mean free path of the molecules is then approximately 25-50 μm. The typical polysilicon thickness is 1 μm. Therefore, the primary etch mechanism probably is not diffusion of gases, but rather a combination of surface adsorption and subsequent surface diffusion. Unfortunately, in the case of using low stress silicon nitride as the structural layer, this means that the BrF_3 molecules, rather than attack the polysilicon, all directly contact the silicon rich film and attack it at an appreciable rate.



(a)

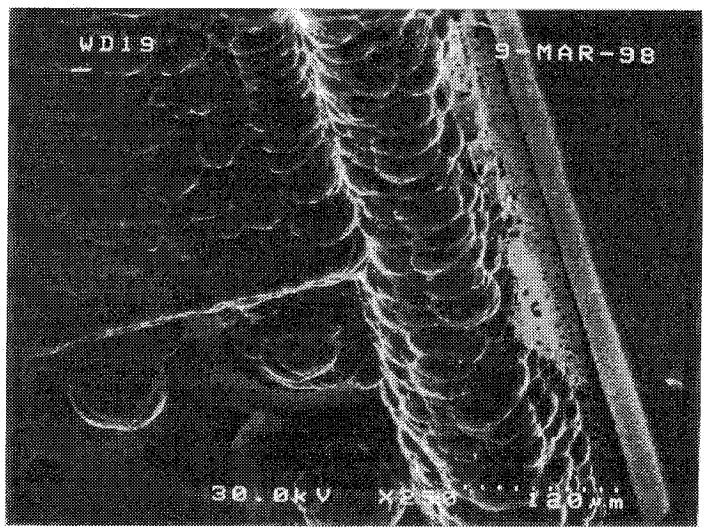


(b)

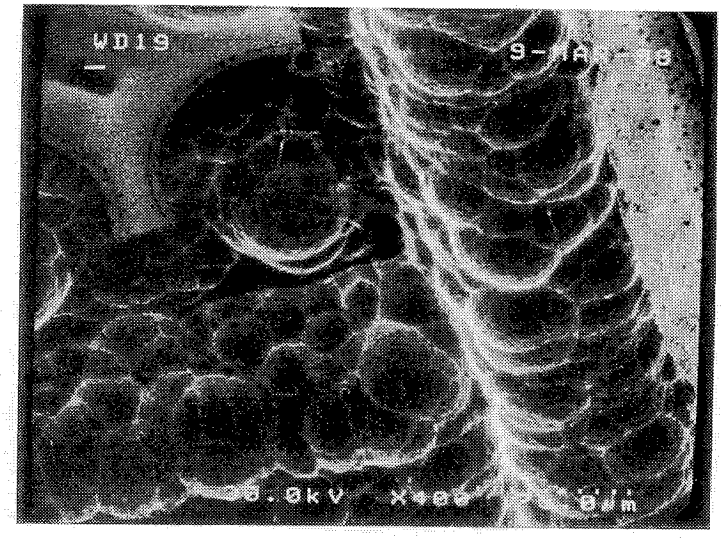
Figure 4.15 (a) SEM showing a flap that was subject to BrF_3 etching. Notice that the silicon nitride flap is obviously very damaged compared with previous examples. (b) Close-up of etch hole areas, showing obvious damage to the nitride.

Figure 4.15 shows an SEM of an obviously attacked flap. Close-up views review much more information. For example, it appears that the silicon nitride beams are still intact, when in reality, they have been completely etched away, and the visible artifact is

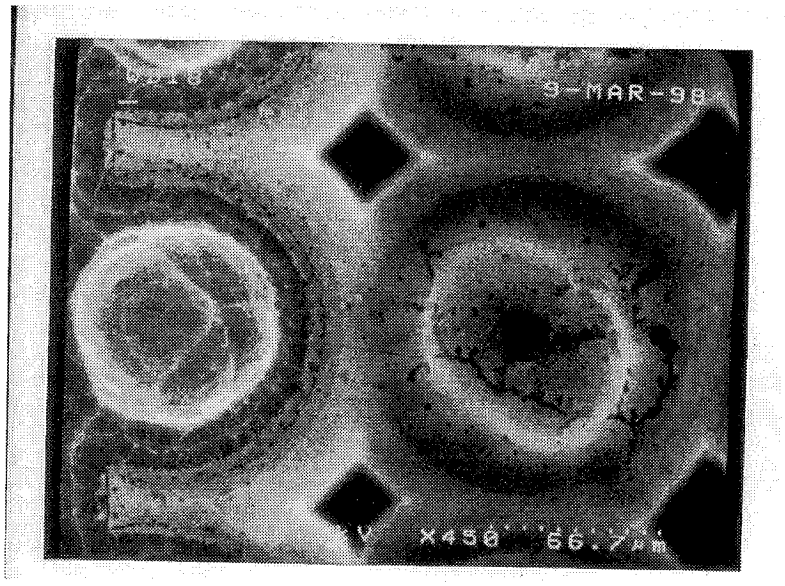
actually a knife edge that remains after the etching is completed [Figure 4.16(b)]. If the edges of the flap are examined, it is apparent that minimal undercutting occurs. Rather, any molecules that diffuse in through the etch holes necessarily have a minimal horizontal velocity component (due to the large mean free path) and have a tendency to etch straight down, rather than outwards. The edges of the flaps appear to fare a little better in terms of undercutting, due to the fact that an entire edge is open to the gas, rather than just one hole. However, the nitride in these areas is also destroyed, as can be seen in Figure 4.16(c).



(a)



(b)



(c)

Figure 4.16 (a) SEM showing silicon knife edge that used to be under a beam. (b) Close-up of a cover of a flap. Note that there appears to be more downward etching (as evidenced by the pit) than undercutting. (c) Edge of a flap. Notice that the left (edge) hole is completely destroyed because it has more exposure to the gas.

References

[1] Tsao, T., Hsu T.-Y., and Tai, Y.-C., "Copper Sacrificial Layer Technology for use in Surface Micromachining," SCCAVS Micromachining Workshop III, Anaheim, CA, 1996.

[2] Wagner, B., Reimer, K., Maciossek, A., and Hofmann, U., "Infrared Micromirror Array with Large Pixel Size and Large Deflection Angle," Transducers '97.

[3] Schmidt, M., Howe, R., and Senturia, S., "A Micromachined Floating-Element Shear Sensor," *Proceedings of the 4th Int. Conference on Solid State Sensors and Actuators*, pp. 383-386, Tokyo, Japan, 1987.

[4] Sugiyama, S., Suzuki, T., Kawahata, K., Shimaoka, K., Takigawa, M., and Igarashi, I., "Micro-Diaphragm Pressure Sensor," *Proceedings of IEDM*, pp. 184-187, Los Angeles, 1986.

[5] Tabata, O., Funabashi, H., Shimaoka, K., Asahi, R., and Sugiyama, S., "Surface Micromachining Using Polysilicon Sacrificial Layer," *Proceedings 2nd International Symposium on Micro Machine and Human Science,* pp. 163-172, Nagoya, Japan, 1991.

[6] Tabata, O., Asahi, R., Funabashi, H., and Sugiyama, S., "Anisotropic Etching of Silicon in $(CH_3)_4NOH$ Solutions," *Proceedings of the 6th Int. Conference on Solid State Sensors and Actuators*, pp. 811-814, San Francisco, 1991.

[7] Wang, X.Q., Yang, X., Walsh, K., and Tai, Y.-C., "Gas Phase Silicon Etching with Bromine Trifluoride," 1997 International Conference on Solid-State Sensors and Actuators, Chicago, IL, USA, 1997.

[8] Monk, D., Soane, D., and Howe, R., "Sacrificial Layer SiO_2 Wet Etching for Micromachining Applications," *Proceedings of the 6th International Conference on Solid State Sensors and Actuators*, pp. 647-650, San Francisco, 1991.

[9] Liu, J., Tai, Y.-C., Lee, J., Pong, K., Zohar, Y., and Ho, C.-M. "In Situ Monitoring and Universal Modelling of Sacrificial PSG Etching Using Hydrofluoric Acid," *Proceedings, IEEE Micro Electro Mechanical Systems Workshop (MEMS '93)*, Fort Lauderdale, FL, pp. 71-76, 1993.

Chapter 5

Integration

5.1 Introduction

Being able to fully integrate sensors, actuators and electronics on the same substrate is necessary for the success of the turbulent boundary layer control project. However, the importance of this integration extends beyond this application specific task. There are many generic concerns in MEMS-Electronics integration that have been addressed in this work. Elements of this MEMS-Electronics integration effort can be applied to other, more generic integration efforts.

The primary concern of all integration efforts revolves around the choice of when the MEMS processing steps are completed relative to the IC processing steps. There are three options: (1) an interweaved process, (2) electronics first, followed by MEMS, and (3) MEMS first, followed by electronics. Each alternative has its advantages and disadvantages.

One assumption being made while comparing these different options is that the MEMS designer does not have full and high priority access to an IC capable fabrication facility. This is a reasonable assumption for an academic environment where few laboratories have even basic MEMS capabilities, let alone IC fabrication facilities. In the few institutions (Berkeley and Stanford to name two) in the entire country which have CMOS capabilities, the electronics yield is usually unacceptably low compared with industry. The lack of dedicated access to a fabrication line is not limited to academia. Even in a corporation with large fabrication facilities, internal MEMS research facilities often are completely separate from VLSI lines and MEMS researchers often are not

allowed full access to such lines. If full access were available to MEMS researchers, the interweaved process would almost definitely be the choice for reasons that will be discussed below.

From a device-robustness point of view, the interweaved process is ideal. The order in which steps are completed reflects a process flow designed with optimum device performance in mind. This is in stark contrast to the other approaches, which often require steps that are not needed for the eventual devices (or electronics) per se, but rather are crucial for the overall survival of the process. For example, in an electronics-first process, during the MEMS processing, layers are often deposited, patterned, and later removed simply to protect the aluminum metalization on the electronics portion of the wafer. In an interweaved process, the metalization steps could occur after any harsh MEMS processing steps have been completed. From this, it can be seen that the interweaved process also is usually the shortest (both in time and in number of processing steps). General processing wisdom holds that the fewer the steps, the more robust the process. For commercial applications, fewer steps also reduces the cost significantly while increasing the yield.

Extremely few IC facilities will allow pre-processed wafers to enter their fabrication lines due to fear of contamination, especially the fear of unknown (by VLSI standards) MEMS materials. Therefore, finishing the IC processing first has one primary advantage: one has many choices of IC foundries that can provide wafer level electronics. After the foundry fabricates the IC electronics, the wafers are then subject to MEMS processing, which can be completed in a facility separate from the VLSI fabrication line.

Oftentimes, the choice of which approach to take for integration depends on metalization concerns. Typically, VLSI fabrication uses aluminum metalization. Aluminum is considered a low temperature material and cannot withstand processing temperatures above 450° C. Therefore, certain steps such as diffusion and thermal oxidation cannot be attempted after aluminum deposition. If electronics including metalization is completed before MEMS processing, the user then can only use low temperature steps to finish the process. Other considerations regarding metalization are that (1) ideally only one deposition/pattern/etch step should be used for each metal layer, and (2) fine line widths are often required for the electronics portion of the chip. These two concerns may imply that the intelligent approach is to complete all the metalization (for a given layer of metal) at one time and to do it at the VLSI fabrication facility, where fine line metal patterning and etching are well characterized steps. There may exist some facilities which will take pre-processed wafers on a one-time basis, in which case MEMS processing up to but not including metalization can be done before submitting the wafers to the foundry. Complete CMOS processing is then completed, with the metalization connecting both electronic and MEMS devices.

Finally, although less interesting from a research point of view, economic reasons can often dictate which option is chosen. When choosing an outside foundry as a vendor, it should be noted that the least expensive option involves requesting as standard a package as possible. This usually implies that, of the three above-mentioned options, the second (electronics first, then MEMS) option may be the most viable. Cost, however, is often not the only economic concern. The idea of interweaving a process is often the most appealing option from a technological point of view. From a foundry's point of

view, it is also the most costly. While most foundries will not turn down business that requires standard technology, they do require financial justification before attempting an expensive custom run which could divert their manpower and resources from other high-volume projects. Foundries such as Standard Microsystems Corporation (SMC) which have dedicated MEMS-Electronics fabrication facilities often also have long lines of corporate customers buying high volumes of devices. They often are not interested in low volume applications (such as university research or small company products).

Therefore, oftentimes, small volume applications are limited to pursuing the electronics first approach. If high temperature MEMS steps are required, it is possible to switch to an interweaved approach where the electronics processing up to, but not including, metalization is completed at the VLSI foundry. Next come the MEMS steps, and the process is completed with metalization at the MEMS facility (or another facility willing to take pre-processed wafers). This approach requires the ability to dry etch contact holes, pattern fine lines, and dry etch aluminum in the post-IC facility. This may or may not be a limitation.

Here, then, in this chapter, two attempts are presented – one interweaved and one metal-first.

5.2 Interweaved Process

5.2.1 Overview

The CMOS fabrication for our interweaved process [3] was completed at UC-Berkeley's Microfabrication Laboratory, using their standard 2-poly 1-metal run. Their laboratory policy allowed for the reintroduction of previously processed wafers. The shear stress sensor design is as described in Chapter 2. Both surface micromachined and

bulk micromachined actuators were designed, but due to processing errors, only the wafers designed with surface micromachined actuators in mind survived the full processing. A simplified process flow is shown in Figure 5.1. In the first section, Berkeley completed CMOS electronics fabrication up to, but not including, contact hole opening for metalization. The wafers were then returned to Caltech for high temperature MEMS processing, which includes most of the sensor and some of the actuator processing. The wafers were then returned to Berkeley for contact hole opening and metalization. Finally, the wafers were returned to Caltech for the low temperature MEMS processing, which included the majority of the actuator processing and the final actuator freeing.

The actuator in this Berkeley integration run used polysilicon as the sacrificial layer and BrF_3 and TMAH were both used as the sacrificial layer etchant. Silicon dioxide was not used as the sacrificial layer because it is also used in the CMOS process as the insulating and encapsulation layers. These layers would have been vigorously attacked during the sacrificial layer etch step. In addition, the exposed aluminum metalization would also have been stripped by the HF acid etch. Both BrF_3 and TMAH have high etching selectivities of silicon to silicon nitride, silicon dioxide, and in the case of BrF_3 but not TMAH, most metals (including aluminum). However, BrF_3 is often preferred to TMAH due simply to the fact that BrF_3 is a dry etching system, and therefore release stiction is not a problem (although in-use stiction may occur regardless of the release mechanism).

One major difference exists when choosing whether to use TMAH or BrF_3 as the sacrificial layer etchant. As mentioned above, TMAH attacks aluminum quickly, but it

attacks silicon dioxide relatively slowly but appreciably. Therefore, a thick layer of low temperature oxide (LTO) was deposited and patterned to protect the aluminum during the sacrificial layer etch. In addition to its protective attributes, LTO is a decent choice for the protection layer because it is deposited at a relatively low temperature (approximately 400-450° C) in our laboratory's low pressure chemical vapor deposition (LPCVD) system. Silicon nitride has an even higher etch selectivity over silicon in TMAH, but it is deposited at a much higher temperature (approximately 835° C), which would destroy the metal. Organic films are also etched by TMAH and therefore cannot be used as protection layers. If available, plasma enhanced chemical vapor deposition (PECVD) systems operate at a much lower temperature than their counterpart LPCVD systems and could be used to deposit a silicon nitride film, but generally speaking, PECVD films are of lower quality than LPCVD films.

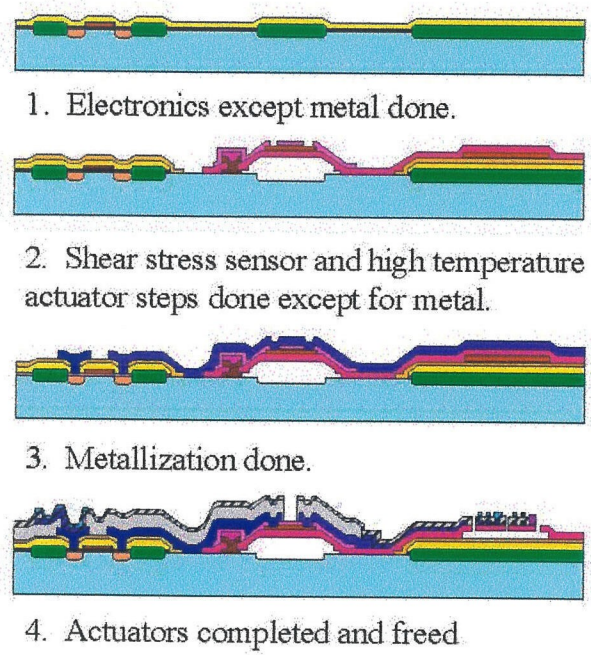


Figure 5.1 Simplified integration process flow using the Berkeley CMOS process.

5.2.2 Design Issues

The sensor used for this process is the standard shear stress sensor as described by Jiang [1]. There are two rows of sensors – one upstream and one downstream of the row of actuators. Each row of sensors consists of three units of three sensors each. Each unit is associated with one actuator. The idea behind this design was that each upstream sensor unit would be of an appropriate size to be able to detect a high shear stress streak by comparing the outputs of the three individual sensors. Then, once a high shear stress streak is identified, the processing circuitry would cause the actuator to move. The downstream sensor unit would then be used to determine what effect the actuation had.

The electronics that was designed for this run by Gupta [2] can be broken down into three main functions. First of all, there is the constant temperature anemometry circuitry used for driving the sensors. The sensor signals are then fed into a non-linear filtering network which can sense large differences between adjacent sensors while smoothing small differences or signals from distant sensors. If a large difference is sensed between adjacent sensors, a trigger signal is sent to the actuator, which is driven by a large transistor.

The actuator chosen for the Berkeley integration run is a surface micromachined flap. The design was chosen more for robustness' sake than performance's sake. The flap used a design (1x1 mm^2 flap with 1 mm long torsional beams) that had been proven to be relatively easily released with a high yield. In the end, however, a surface micromachined flap may not be used as there are good arguments for using either surface or bulk flaps.

As previously mentioned, the tip velocity, v, is the product of frequency, ω, and displacement, ysinθ, and may be used to better decide which, if either, choice is better. (Refer to Figure 3.34 for a drawing of the flap.) In the simplest form, tip velocity can be calculated as follows where the displacement is assumed to be small such that sin θ can be approximated by θ.

$$v = \omega y_0 \theta \tag{5.1}$$

$$\theta = \frac{\tau}{k_\theta} \tag{5.2}$$

τ=applied torque
k_θ=torsional spring constant

Torque is created by the cross product of the coil-generated magnetic moment and the external magnetic field. Increased current flowing through more coils would increase the magnetic moment. To increase the number of coils, the plate area needs to be increased, and both the resistance and power consumption also increase. The increase in area of the plate also decreases the resonant frequency. Therefore, with so many conflicting demands, it makes sense to define a figure of merit. Thinking long term, if these actuators are to be used in a large array, power consumption may become an issue. Therefore, a reasonable figure of merit may be the tip velocity divided by the power consumed.

The power consumed is simply i^2R. We will assume that the current is independent of everything else. In practical cases, the upper limit may be reached when the power dissipated in the metal contacts creates enough heat to vaporize the metal. Resistance for a coil if defined as

$$R = \frac{\rho L}{A} \tag{5.3}$$

where ρ is the resistivity, L is the length of the coil and A is the cross-sectional area of the coil. The length of the wire can be calculated by measuring all of the lengths of the individual turns, but a more efficient way is to assume that the length is equal to the total area occupied by metal, which is approximately one half of the plate area, divided by the width of the coil. If the metal widths and spacings are not the same, then the area occupied by the metal changes appropriately. Therefore, the resistance can be calculated using the following:

$$R = \frac{\rho w y_0}{2 w_m{}^2 t} \tag{5.4}$$

where w_m refers to the width of the metal and t refers to the thickness of the metal.

If the angle deflected is small, then this figure of merit can be represented by the following equation:

$$FM = \frac{\omega y_0 \theta}{P} \tag{5.5}$$

which, with some substitution becomes

$$FM = \frac{\sqrt{\frac{k_\theta}{I}} y \frac{\tau}{k_\theta}}{i^2 \frac{\rho w y_0}{2 w_m{}^2 t}} \tag{5.6}$$

which, with further substitution and simplification becomes

$$FM = \frac{2 w_m{}^2 t \sqrt{\frac{y_0{}^2}{k_\theta I}} \vec{m} \cdot \nabla \vec{B}}{i^2 \rho w y_0} \tag{5.7}$$

where

$$\vec{m} = \sum i\vec{A} \tag{5.8}$$

which can also be expressed as

$$\vec{m} = i\sum_{n=1}^{\frac{w}{4w_m}}(w-2nw_m)(y_0-2nw_m) \tag{5.9}$$

If this equation is simplified to look at the geometric relations, the following equation results (C is simply a constant that accounts for all of the non-geometric variables):

$$FM = C\sqrt{\frac{L}{ab^4w^3{y_0}^3}}{w_m}^2 t\sum_{n=1}^{\frac{w}{4w_m}}(w-2nw_m)(y_0-2nw_m) \tag{5.10}$$

If one takes this equation by itself, the figure of merit can be increased by reducing the spring constant (i.e. increasing L and decreasing a and b) and by reducing the plate area (i.e. reducing w and y_0, which is the same as reducing the moment of inertia). However, these rules must be applied with the constraints that the resonant frequency should still be on the order of several kHz.

To illustrate the example of calculating the figure of merit, we can take two cases: surface and bulk micromachined flaps. For purposes of clarity, we will assume that both flaps have identical plate dimensions, metal thicknesses, and metal pitch. We will also compare flaps that have the same resonant frequency. The figure of merit is normalized to that of the bulk actuator.

	a	b	L	ω	FM
Surface	10 μm	1 μm	30 μm	2.3 kHz	30
Bulk	10 μm	10 μm	1.36 mm	2.3 kHz	1

Table 5.1 Figure of merit for a surface and a bulk actuator.

From this table, it would appear that the surface actuator is the proper choice for eventual integration issues when power consumption becomes a large issue. However, for practical reasons, it may not be as simple as choosing a larger figure of merit. As evidenced by the frequency response of the 2nd generation active magnetic actuator, the expected gain in amplitude due to operation at resonance may be negated by the fact that the flap is sitting in a shallow cavity. Therefore, air damping becomes a critical part of the equation. As such, a bulk micromachined flap may be the best choice.

5.2.3 Fabrication Issues

The detailed process flow for fabricating the integrated sensor/surface micromachined actuator and electronics is presented here with commentary on the thought that went behind choosing the steps and the challenges encountered. The descriptions in square brackets [] represent mask names.

Berkeley (electronics up to but not including metal)

1. Start with p-type Si <100> 8-12 Ω/
2. Grow SiO_2 100nm
3. Define well [N Well]
4. Implant P 5e12/cm^2 @ 150KeV
5. Drive-in well to 3.4 μm
6. Grow SiO_2 30 nm (pad oxide)
7. Deposit Si_3N_4 100nm
8. Define active area [Active]
9. Etch Si_3N_4 (where LOCOS)
10. Define p-field

11. Implant B 1.5e13/cm^2 @ 70 KeV
12. Grow LOCOS 650-700nm
13. Remove nitride (from 7)
14. Remove oxide (from 6 - pad oxide)
15. Grow sacrificial oxide 20nm
16. Implant B 1.7e12/cm^2 @ 30 KeV
17. Remove sacrificial oxide (from 15)
18. Grow gate oxide 30nm
19. Deposit gate poly 450nm [Poly]
20. Define gate poly
21. Etch gate poly (from 19)
22. Deposit cap. oxide 80nm
23. Deposit cap. poly 450nm [Poly2]
24. Define cap. poly
25. Etch cap. poly
26. Define N+ source/drain
27. Implant As 5e15/cm^2 @ 160 KeV
28. Anneal N+ source/drain
29. Deposit PSG 700nm

Caltech (shear stress sensor and part of actuator)

30. Etch PSG (from 29) [fberk_psg]
31. Deposit PSG 400nm
32. Deposit poly 1um
33. Pattern poly [tpoly] {TNP}
34. Pattern PSG (from 31) [fcaltech_psg]
35. Deposit Si_3N_4 1.2 μm
36. Pattern Si_3N_4 for etch holes [fhole_etch]
37. Etch cavity w/HF
38. Deposit LTO 1μm
39. Deposit Si_3N_4 400nm
40. Pattern Si_3N_4 [fseal_pad_nitride] {NB}
41. Pattern LTO (from 37) [fseal_pad_lto]
42. Deposit poly 500nm
43. Dope poly B 1e16 @80kev
44. Pattern poly (from 41) [fpoly]
45. Deposit Si_3N_4 200nm
46. Pattern Si_3N_4 [fcontact]
47. Etch LTO (from 38)
48. Pattern Si_3N_4 (from 35) [fnitride]

Berkeley (finish metalization)

49. Open contact holes.
50. Sputter Al 600nm and pattern Al (Lam/plasma) [Metal1]

Caltech (finish sensor and actuator)

51. Pattern Al (wet etch) [fberk_metal]
52. Pattern Si_3N_4 (from 35) [tnitride_etch]
53. Deposit LTO 2μm
54. Pattern LTO [tlto_etch]
55. Deposit Cr/Au 400nm
56. Pattern Cr/Au [tmetal1+]
57. Deposit LTO 500nm
58. Pattern LTO [tlto2_etch] <add to over sensor like LTO1>
59. Deposit Cr/Au 400nm
60. Pattern Cr/Au [tmetal2+]
61. Deposit seed layer (PR) (optional)
62. Pattern seed layer [tplating] (optional)
63. Plate NiFe (optional)
64. Remove PR (optional)
65. Protect NiFe [tcr_protect] (optional)
66. Remove seed layer (optional)
67. Etch w/TMAH or BrF_3

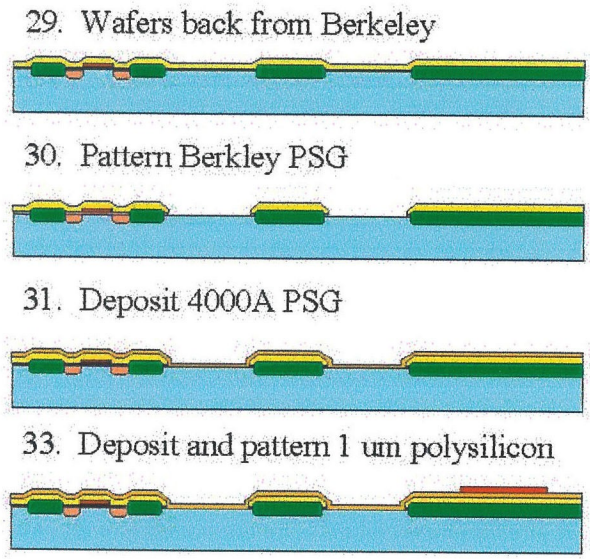


Figure 5.2 Berkeley integration process up to step 33.

The IC (one poly, one metal) process is completed up to, but not including contact hole openings. (Step 29). Once back at Caltech, the first step (step 30) is to etch away the thick Berkeley deposited PSG in the areas where the etching channel will be for the

shear stress sensors. Thin PSG is then deposited (step 31) for use as the etching channels. Polysilicon is then deposited (step 33) for use as the sacrificial layer for the actuator.

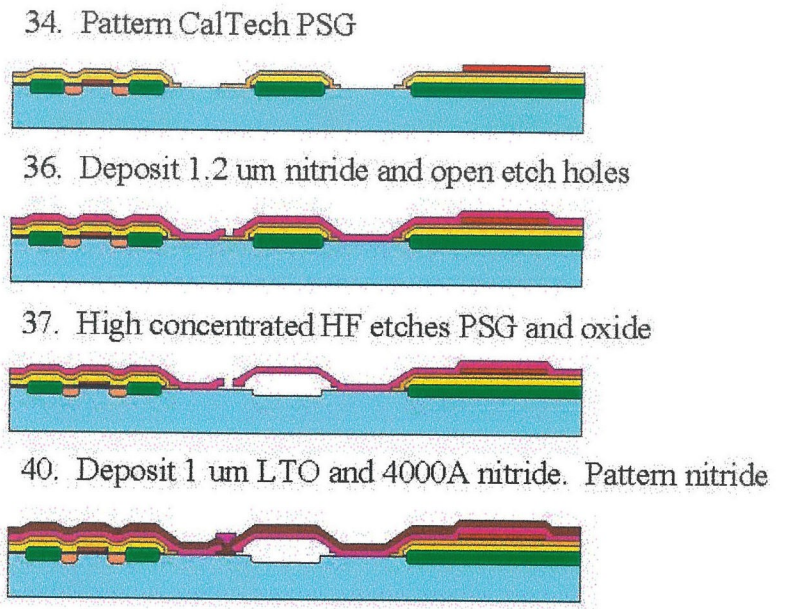


Figure 5.3 Berkeley integration processing up to step 40.

The thin PSG is then etched (step 34), and nitride is deposited (step 36) for use as the shear stress sensor diaphragm and the actuator flap material. This nitride is also crucial for protecting the electronics during the subsequent cavity sacrificial layer etch (step 37). The shear stress sensor cavity is then sealed (step 40) using LTO and nitride. The reason both are used to seal is that (a) nitride is a better sealant, but (b) LTO provides a good etch stop when plasma etching the nitride.

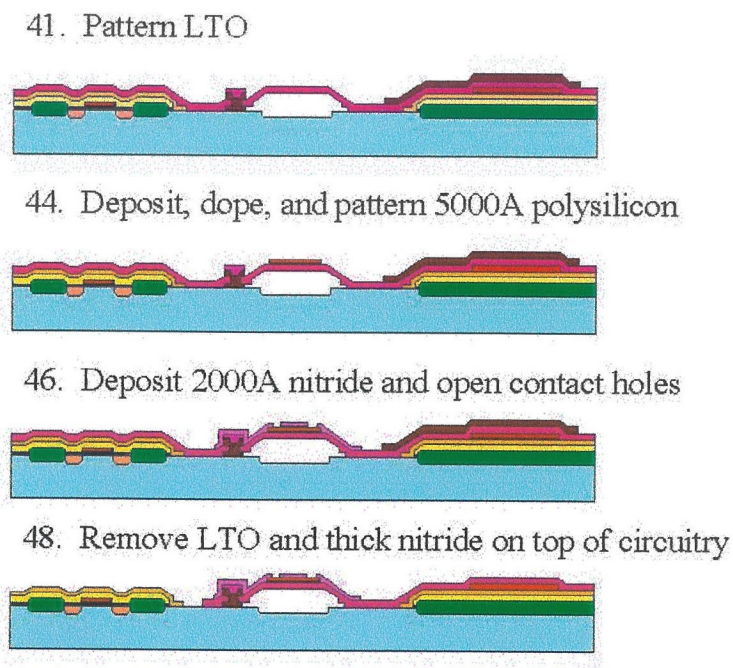


Figure 5.4 Berkeley integration process up to step 48.

After the nitride sealing holes are patterned, the LTO can be wet etched (step 41) away easily, without damaging anything underneath. This is another reason for choosing to seal the etch holes with both LTO and nitride. Next, polysilicon is deposited and implanted (step 44) for use as the resistor for the sensor. The poly is then encapsulated with nitride (step 46). Note that the nitride can easily be etched over the actuator area without worry of etching the flap nitride due to the presence of protection LTO over the flap. Next, all of the remaining LTO is wet etched away, and the protection nitride over the circuitry is removed using a dry etch (step 48). Once again, the nitride etch naturally stops once the PSG covering the electronics is reached.

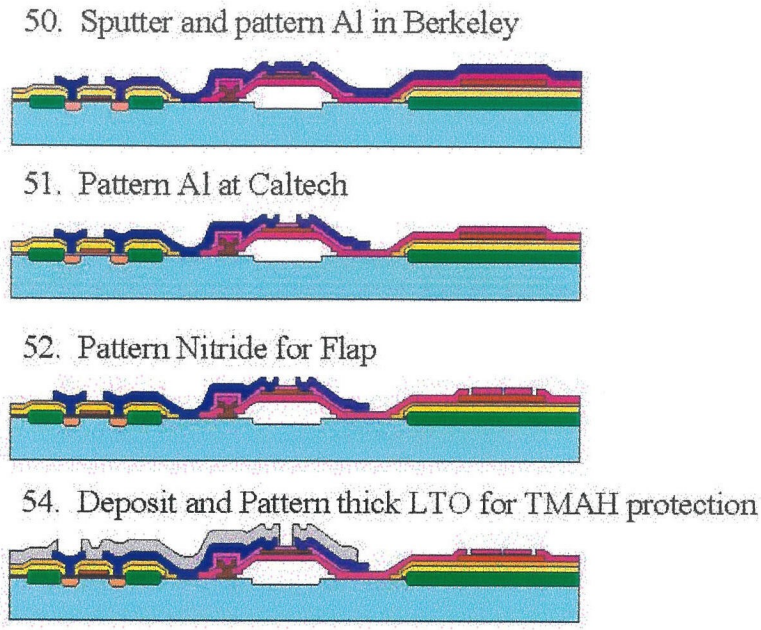


Figure 5.5 Berkeley integration process up to step 54.

At this point, the wafers are returned to Berkeley for contact hole opening and aluminum metalization (step 50). Due to the fact the aluminum is plasma etched and the fact that we were not sure of the effect the chlorine-based plasma has on silicon nitride, we decided not to etch the sensor leads at the same time, and instead wet-etched the remaining aluminum (sensor leads and protection for actuator nitride) at Caltech (step 51). At this point, the sensor is completed, and the nitride for the actuator is patterned (step 52).

For this first attempt at integration, every effort was made to be as conservative as possible. One of the biggest concerns was the protection of the electronics metalization during the sacrificial layer etch. As previously mentioned, due to the fact that TMAH attacks aluminum, a thick LTO layer was deposited (step 54) to protect the vast majority of the metal lines. However, after leaving Berkeley for the final time, the pad aluminum needed to be exposed. Therefore, we decided to pattern final contact holes in the thick

protection LTO and deposit chrome/gold as a true top layer of metal. Note that the sensor area, which is protected by nitride, and the bonding pad aluminum are not covered with LTO.

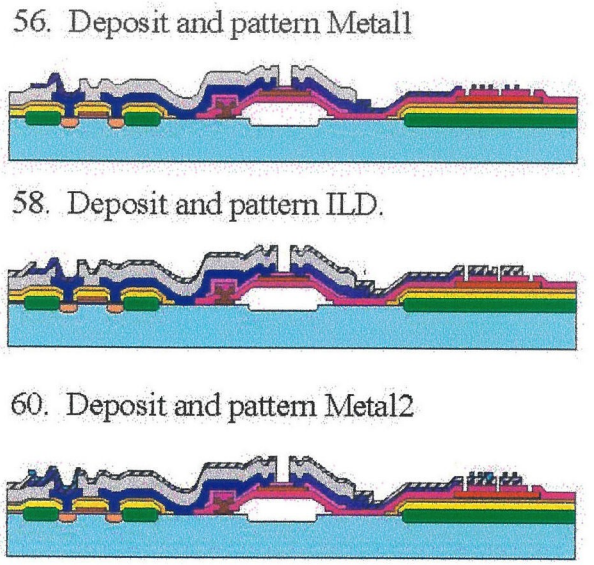


Figure 5.6 Berkeley integration process up to step 60.

The first layer of actuator metal (Cr/Au) is then deposited and patterned (step 56). This metal is also used as to contact the aluminum bonding pads. A thin LTO insulator is then deposited and patterned (step 58). Finally, the second layer of actuator metal is deposited and patterned (step 60).

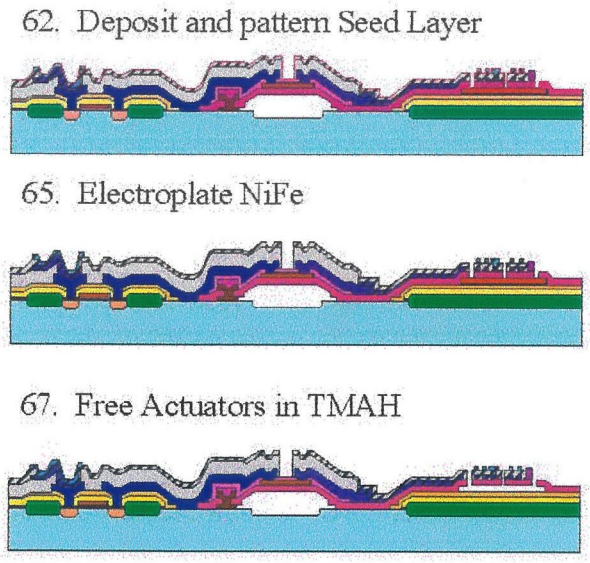


Figure 5.7 Berkeley integration process to completion

If so desired, this process also accomodates the electroplating of permalloy. The seed layer is deposited and patterned (step 62), the permalloy is plated to the desired height (step 65), and the seed layer is removed. Finally, the actuators can be freed in TMAH or BrF_3 (step 67) and the process is finished.

5.2.4 Results and Discussion

A completed die is shown below in Figure 5.8. Unfortunately, the overall device yield low. Therefore, no system level tests could be run.

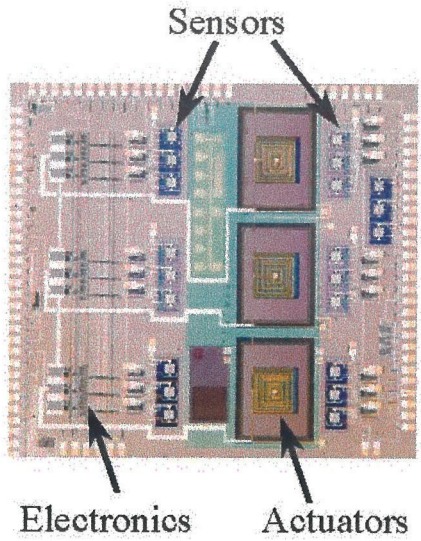


Figure 5.8 Picture showing a chip from an integration wafer.

From a beginning wafer lot of 8 wafers, we eventually ended up with one wafer. There were many reasons for the loss of wafers. One of the largest issues is that of spinning photoresist, especially once the wafers have been returned to Berkeley for contact hole opening and metalization. Contact hole opening is a very critical process and is often the most difficult step in an IC process. To facilitate the openings of small contact holes, thinner photoresist is preferred over thick. However, our wafers had very large (by IC standards) step heights. The Berkeley photoresist spinning process could not be adjusted. Once a standard recipe is developed, it is used exclusively and fabrication facilities are very hesitant to change process parameters. Therefore, on some wafers, problems existed with contact hole openings and metalization.

Another problem existed with the design rules applied to the opening of thick LTO protection and subsequent deposition and patterning of Cr/Au for actuator metal 1 (steps 54-56). The undercut of the LTO was far greater than expected (and appeared to be larger in magnitude than the thickness of the layer, indicating a non-isotropic etch).

Later, then the gold was etched, this resulted in the gold etch attacking the aluminum and destroying the vast majority of the bonding/contact pads.

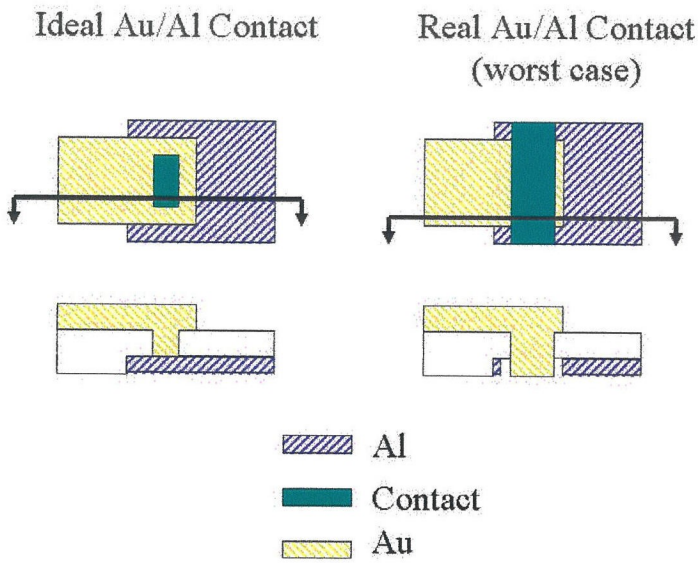


Figure 5.9 Drawing showing top view and cross section of gold/aluminum contact.

Other reasons for reduction in yield are mostly due to issues of miscommunication. For example, some wafers had contact holes opened at Berkeley before being shipped to Caltech for the first time. Due to all of the processing conducted afterwards, these contact holes were then filled with various materials, and it proved impossible to re-open them later.

Finally, with any process, it is necessary to have several test wafers in one's possession. These test wafers ideally should have the same wafer history as one's real wafers. Unfortunately, we did not have the luxury of having such test wafers, and many previously unproven steps had to be completed using the real wafers. Anytime this happens, a reduction in yield can be expected.

Despite all of above-mentioned problems, some transistor level electronics testing was completed. These tests showed that individual working transistors could be fabricated. For example, Figure 5.10 shows an NMOS transistor (W/L=50/10) drain current vs. drain voltage characteristics for various values of gate voltage. The overall shapes of the curves look correct. However, the current level (maximum of 10 μA appears quite low.

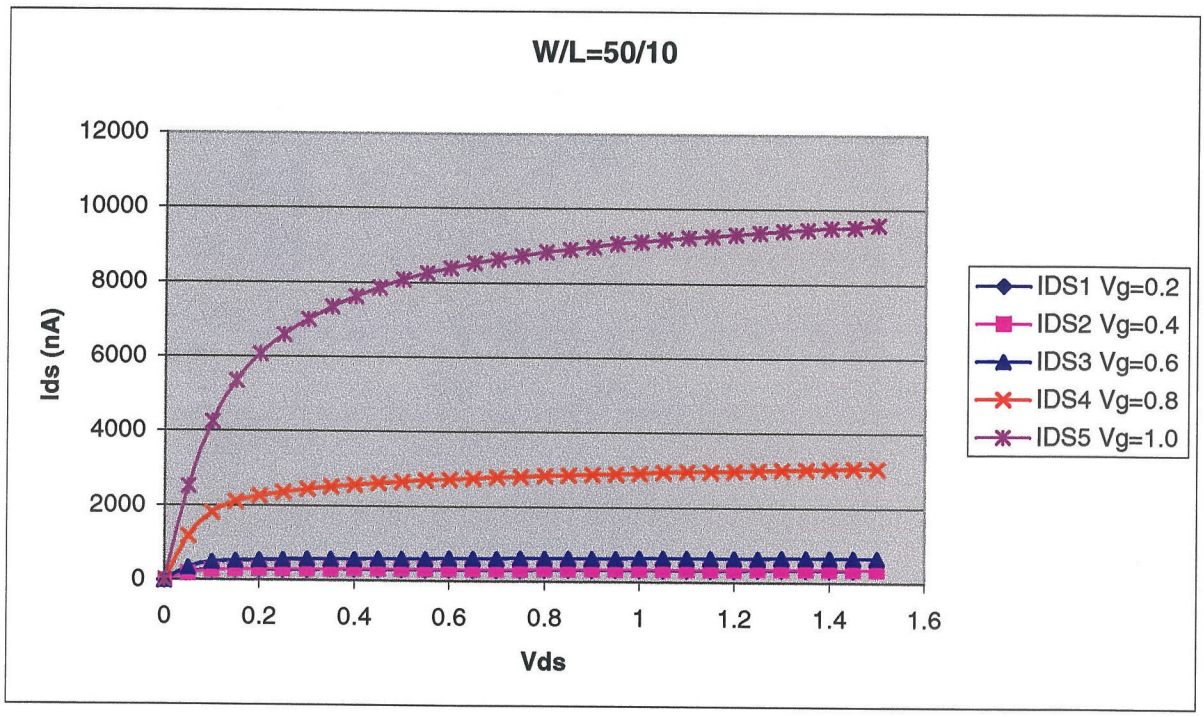


Figure 5.10 Drain current vs. drain source voltage for a W/L=50/10 NMOS transistor.

Figure 5.11 shows the Sqrt (I_{ds}) vs. V_{ds} curves for various NMOS transistors biased into saturation. By extrapolating the linear parts of these curves to the x-axis, one can determine the threshold voltage (i.e. when Vg=Vds, Ids=0). For each of the three curves, the value of threshold voltage appears to be in the range from 0.325 V – 0.375 V. Such values are lower than expected - values

tested by Berkeley before all of the MEMS processing are in the range of 0.84 – 1.00 V for the 50/3 transistors, 0.84 – 0.85 V for the 50/10 transistors, and 0.82 – 0.88 V for the 50/25 transistors. Care was taken to ground the appropriate terminals during circuit testing. Therefore it is assumed that the MEMS processing somehow affected the threshold voltages. Even if the threshold voltages didn't change much from the expected values, the spread may or may not be acceptable for use in larger circuits.

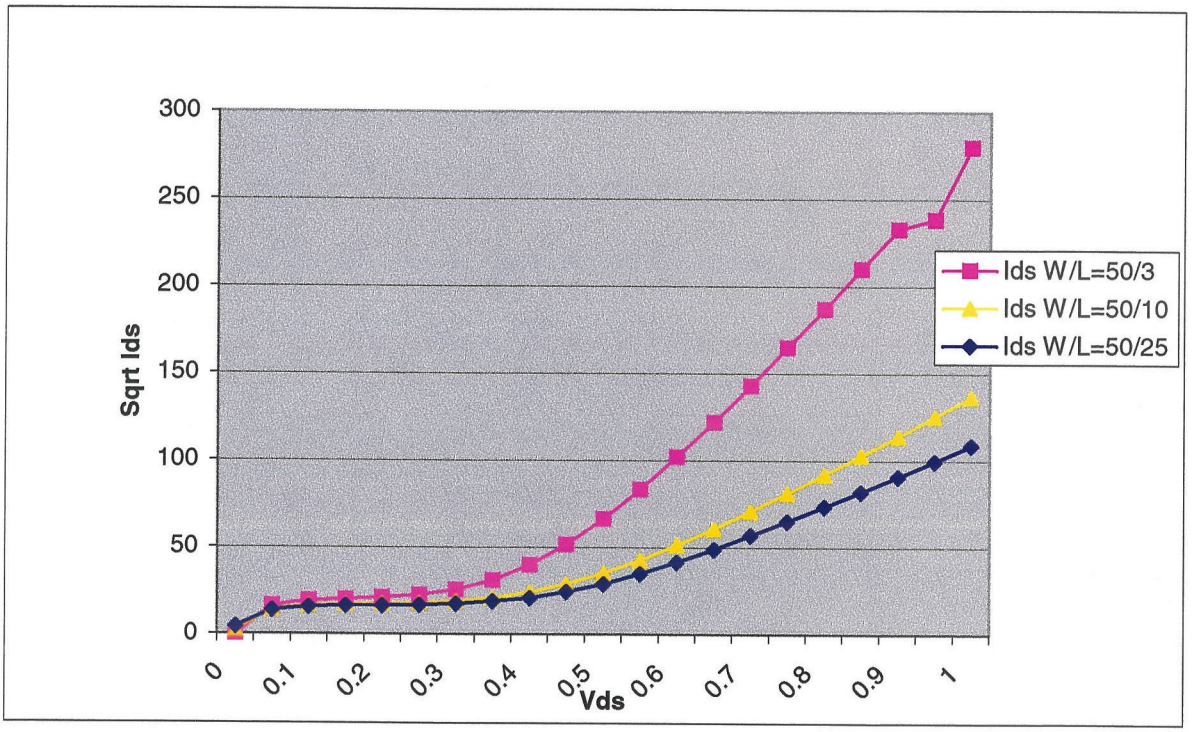


Figure 5.11 Sqrt(I_{ds}) vs. V_{ds} curves used to find threshold voltage.

5.3 Electronics First Process

5.3.1 Overview

The Berkeley run was ideal in the sense of the latitude in processing steps that were allowed. However, the circuit yield was not believed to be acceptable for the eventual need of wafer level integration. In order to produce a high enough yield, it was

necessary to pursue the foundry option. Orbit Semiconductor proved to be willing to explore the option of interweaved processes. However, for economic reasons, and more importantly, the fact that they intended to switch their fabrication facilities to a 6-inch wafer line which would be incompatible with the Caltech Micromachining Lab's 4-inch line, this option was soon abandoned. For economic reasons, another interweaved fabrication facility (Semiconductor Manufacturing Corp., SMC) was also investigated and then abandoned. At this point, it was realistically decided to pursue the electronics-first approach, and Mitel Semiconductor was chosen to fabricate the electronics portion of the wafer. At the time of this writing, the design for the Mitel fabrication run is being completed. One of the largest differences between the Mitel process and the Berkeley process is the fact that the Berkeley process is a 1-poly, 1-metal process while the Mitel process is a 2-poly, 2-metal process. The MEMS significance of these features will be described in the next section. It is anticipated that details of this run will be presented in future papers or theses. Here, in brief then, is the plan for the run.

5.3.2 Process

The proposed final structure is drawn below. Notes and details on the chosen shear stress sensor and actuator follow.

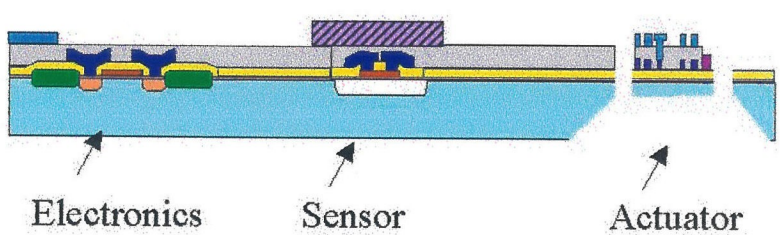


Figure 5.12 Proposed structure for Mitel (electronics first) integration run.

First of all, due to the fact that this process is completely electronics first, no high temperature processing may be completed after the wafers are returned from the fabrication facility. Therefore, the sensor and actuator requirements are significantly different from those of the Berkeley run. Most significantly, there can be no after-electronics silicon nitride deposition.

The standard shear stress sensor uses a silicon nitride diaphragm sitting atop a silicon dioxide sacrificial layer. For purposes of this run, a shear stress sensor was designed, fabricated, and tested using low temperature oxide as the structural layer and either polysilicon or single crystal silicon as the sacrificial layer. If polysilicon is used as the sacrificial layer, then a 2-poly process is necessary – poly-1 would be the sacrificial layer while poly-2 would be the structural layer. This structure uses BrF_3 as the etchant and vacuum deposited parylene both to seal the vacuum chamber and to provide additional structural support.

The standard surface micromachined actuator uses silicon nitride as the flap material. As was shown in Chapter 3, silicon dioxide flaps are not robust. Therefore, for purposes of the Mitel integration run, a bulk actuator was chosen for use. One great advantage of using a bulk actuator with the Mitel process (or any 2-metal process) is that the flap material (silicon) is obviously provided as are both metalization layers. Therefore, the only post-IC processing that needs to be completed is the creation of the backside cavity and the RIE release from the front side. As was described in chapter 3, the frontside can be easily protected during the cavity etch by the use of Kerr Sticky Wax.

Care must be taken, however, to properly open the backside cavities. First of all, backside alignment marks must be created. When created at the same time as front side alignment marks (by using a double-sided alignment jig, for example), alignment between front and backside marks can be quite good. However, when created at separate times, this process requires a very careful alignment between front side alignment marks and the marks available on the double side alignment machine you are using. If the machine used is a jig, this could be difficult. If the machine used is an infrared double side contact aligner, the problem is not so difficult. These problems can be made less severe, however, by the use of conservative design rules on the sizes of the backside cavities. If the cavities are made large enough, the process can be made to withstand several tens of microns of misalignment. Finally, care must be taken when etching the protective layers on the backside. Several layers, all of differing thicknesses, are deposited during the IC processing. Therefore, to ensure that the proper etching is done, the wafer may need to be taken between plasma and wet etching systems several times.

References

[1] Jiang, F, *Silicon-Micromachined Flow Sensors*, Ph.D. Thesis, California Institute of Technology, Pasadena, 1998.

[2] Gupta, B., *Analog VLSI for Active Drag Reduction*, Ph.D. Thesis, The California Institute of Technology, 1997.

[3] Tsao, T., Jiang, F., Miller, R., Tai, Y.-C., Gupta, B., Goodman, R., Tung, S., and Ho, C.-M., "An Integrated MEMS System for Turbulent Boundary Layer Control," *1997 International Conference on Solid-State Sensors and Actuators*, Chicago, IL, USA, June 1997.

Appendix A
Standard Processes

Here in Appendix A, standard recipes are presented and then followed with the detailed process flows for selected work.

A.1. Lithography

All photoresists were spun on, although methods such as spraying and dipping also exist. Several different types of photoresist were used. KTI 100cs (made by OCG) positive resist was the first photoresist chosen for use in the Caltech Micromachining Lab because of its qualities such as high contrast, good physical stability, and tolerance to overdeveloping of fine structures. However, it was phased out by the producer, and the AZ series of photoresists was used. AZ 1518 is a thin resist good for patterning thin lines but poor for covering large steps. AZ 4400 is a thicker resist good for step coverage, electroplating molds, and medium term dry-etch protection. AZ 4620 has very similar properties to AZ 4400 but is more viscous and therefore produces thicker films at the same spin speed.

Step Number	Parameters
1. HMDS	1.0 minute
2. Spin	KTI 100cs- 3.5 krpm, 40 seconds yields 3.0 μm film AZ 1518 - 3.0 krpm, 40 seconds yields 1.0 μm film AZ 4400 – 3.5 krpm, 40 seconds yields 4.5 μm film AZ 4400 – 3.0 krpm, 40 seconds yields 5.0 μm film AZ 4620 – 3.5 krpm, 40 seconds yields 7.0 μm film
3. Softbake	20 minutes at 90° C
4. Hardbake	20 minutes at 120° C

Table A.1 Standard resist spinning parameters. In addition, note that a longer spin time (60 seconds, say) will result in a thinner resist but better step coverage.

Exposures were done on a 10:1 GCA stepper. Standard exposure times are given below.

Resist	Exposure Time Given Standard Spinning Process
KTI 100cs	0.7 s
AZ 1518	0.3 s
AZ 4400	0.7 s
AZ 4620	0.9 s

Table A.2 Standard exposure times.

Dilution of photoresist developers is crucial for obtaining high contract exposures. With proper dilution, overdeveloping by as much as 100% does not significantly affect the lithography steps for the processes described in this work. With no dilution, AZ 351 developer acts as a photoresist stripper.

Developer	Dilution (water:developer)
OCG 934 1:1	1:1
AZ Developer	2:1
AZ 351	4:1 or 5:1

Table A.3 Standard dilution of developers.

A.2. Double-side alignment

Double-side alignment is done with a jig that sandwiches a wafer between two symmetrical mask plates.

Step Number	Parameters
1. Spin resist on backside	AZ 4400 3.5 krpm
2. Softbake	90° C 10 minutes
3. Spin resist on frontside	AZ 4400 3.5 krpm
4. Softbake	90° C 10 minutes
5. Expose	7 seconds frontside, 7 seconds back
6. Develop	2 minutes
7. Hardbake	120° C 20 minutes
8. Etch	RIE – silicon etch recipe – 10 minutes front, 20 min back PE II – silicon etch recipe – 10 minutes front, 15 min back
9. Remove resist and clean	Piranha

Table A.4 Standard double-side alignment steps.

When spinning resist on the frontside, care must be taken not to scratch the resist on the backside. During developing, the wafer should be flipped upside down several times (assuming developing is being done in a flat dish) to ensure the both sides are properly developed. If a given alignment mark (out of 4) is underdeveloped but the other ones are well developed, further spot-developing should be done by taking the wafer out of the solution, drying it, and then using a pipette to drop individual drops of developer onto the underdeveloped alignment mark. After developing, to prevent unwanted etching of the silicon, care must be taken to ensure that any scratches in the resist are painted over. In choosing whether to use the RIE or the PE II to etch the alignment marks, consideration must be given to the facts that the PE II (a) has a higher throughput, but (b) has a more isotropic etch (i.e. it will undercut the alignment marks more).

A.2. Thin Film Deposition

A.2.1. Wafer Cleaning

Piranha cleaning is required before any deposition in a clean tube and highly recommended for any other deposition assuming that the wafers are not metalized. The standard piranha clean is done at 120° C with sulfuric acid and hydrogen peroxide. If the

wafer is already metalized, TCE cleaning can be substituted. RCA1 (5:1:1 H_2O:NH_4OH:H_2O_2) and RCA2 (6:1:1 H_2O:HCl:H_2O_2) cleans are used at room temperature to remove metal contaminants from quartzware, tweezers, wafer holders, etc.

A.2.2. Thermal Oxidation

The two primary means of growing thermal oxide are wet, used to grow thick films relatively quickly, and dry, used to controllably grow thin films. In particular, a 0.1 μm thick film is needed for the optical determination of polysilicon thin film thicknesses.

Process	Typical Parameters
Wet Oxidation	1050° C, 10 Hrs, 7 sec/drop will yield a 1.5 μm film
Dry Oxidation	950° C, 30 minutes will yield a 0.1 μm film

Table A.5 Typical thermal oxidation parameters.

A.2.3. Low Pressure Chemical Vapor Deposition

Low pressure chemical vapor deposition (or LPCVD) is a an excellent method for depositing high quality conformal dielectric thin films.

Film	Temperature	Deposition Rate	Gas Flow Rates
Low Stress Silicon Nitride	835° C	60 A/min	DCS – 64.7 sccm NH_3 – 15.6 sccm
Polysilicon	550° C	136 A/min	SiH_4 – 80.0 sccm
PSG	450° C	85 A/min	O_2 – 61.8 sccm SiH_4 – 19.8 sccm PH_3 – 10.2 sccm
LTO	450° C	100 A/min	O_2 – 61.8 sccm SiH_4 – 42.8 sccm
LTO (for copper/aluminum coverage)	380° C	85 A/min	O_2 – 61.8 sccm SiH_4 – 42.8 sccm

Table A.6 Standard LPCVD recipes and approximate deposition rates.

A.2.4. Evaporation

Typical films used during the processes described in this thesis include aluminum, chrome, titanium, copper, and gold. Typical rates of thermal evaporation are approximately 1.0 A/sec for adhesion layers and 3.0 A/sec for other metals.

Chrome and titanium are used as adhesion layers for copper and gold. Chrome can be extremely difficult to etch at times. Therefore, if subsequent processing does not expose the wafer to HF acid, which attacks titanium extremely fast, titanium is preferable. During chrome evaporation, a significant amount of out-gassing occurs immediately preceding the actual evaporation. Therefore, close attention must be paid to avoid pinning the high-vacuum gauge. Titanium evaporation is done at a high temperature but care must be taken to avoid using too much power, which may disintegrate the titanium wire.

A.2.5. Electroplating

Both permalloy and copper plating are done at a rate of approximately 5 μm/hr. To achieve good quality films, as much wafer area as possible should be plated, with the undesired film later etched away.

Step Number	Comment
1. Standard Lithography	If thicker films are desired, multiple layers of resist may be spun.
2. Expose Plating Ring	1 minute for a 5 μm resist.
3. Develop Wafer	For thick resists, AZ 351 should be used.
4. Electroplate	Test wafers are important for determining the correct plating current.
5. Remove Resist	
6. Strip Seed Layer	See Table A.8 for recipes.
7. Pattern Unwanted Areas	Optional
8. Etch Unwanted Areas	Optional

Table A.7 Standard electroplating recipe.

A.3. Thin Film Etching

A.3.1. Wet Etching

Film	Chemicals	Rate
Copper	100 H_2O:10 HAc:10 H_2O_2	~ 500 A/min
Chrome	100 $H_2$0:20 HCl	~600 A/min if no passivation
Titanium	100 H2O:1 HF	~50 A/min
LTO	BHF	~3-4000 A/min
PSG	BHF	~6-7000 A/min
Silicon Nitride	49% HF	~ 5-60 A/min

Table A.8 Typical wet etch rates for various films.

A.3.2. Dry Etching

All plasma-based dry etching systems suffer to some degree from a loading effect. Therefore, the following etch rates are approximate. When using the PEII, the wafers should be rotated 4 times during the etch. Note that either CF_4 or SF_6 can be used for etching nitride and polysilicon. SF_6 etches both faster, but has a poor selectivity of polysilicon over nitride. Therefore, when etching a nitride film atop polysilicon, CF_4 should be used.

Film	Power	Gases	Etch Rate
Low Stress Silicon Nitride	300 W	SF_6 – 210 mT O_2 – 50 mT	700-1000 A/min
Low Stress Silicon Nitride	200 W	CF_4 – 200 mT	~800-1000 A/min
Polysilicon	300 W	SF_6 – 210 mT O_2 – 50 mT	> 2000 A/min
Polysilicon	200 W	CF_4 – 200 mT	~800-1000 A/min
Photoresist	300 W	O2 – 300 mT	> 1 μm/min
Silicon	400 W	SF_6 – 210 mT O_2 – 50 mT	> 5000 A/min (if etching alignment marks)

Table A.9 Standard PEII etch recipes.

The primary RIE etching done during the processing described herein is a silicon etch using a recipe developed by Xing Yang. The O_2 set point is 4.32 and the SF_6 set point is 5.35. Power is 650 W. The etch rate depends on the opening area and wafer placement, but a typical value is approximately 0.5 μm/min.

A.4. Silicon Etching

KOH bulk silicon etching was done using a reflux system to prevent the evaporation of solution and changing of chemical concentrations during extended etches. The easiest method of depositing an etch mask is thermal oxidation, which produces a thick, uniform, pinhole free film. Typical etching parameters are shown as follows:

Etchant	Concentration	Temperature	<100> Etch Rate
KOH	20	58° C	~ 20 μm/hr

For precise control of bulk etched structures, an epitaxial layer sitting atop a heavily doped boron layer is often used. The boron layer serves as an etch-stop for most bulk etchants (KOH, EDP, and TMAH). After the etching in the openings has reached the boron layer, a selective etchant (HNA 1:3:10 in 100 H_2O) that etches boron-doped silicon much faster than it etches n-doped silicon can then be used to remove the remaining boron layer without attacking the epitaxial layer. This process leaves behind a mirror-smooth surface of precise thickness.

When using this method, however, LPCVD silicon nitride must be used as the backside mask. The elevated temperature of thermal oxidation causes the boron to diffuse significantly. If the concentration within the boron layer drops to 10^{19} cm^{-3}, the etch selectivity essentially disappears.

List of Abbreviations Used

BHF	Buffer HF Acid - HF Acid with buffers to maintain constant pH and etching rate of oxide.
CMOS	Complementary Metal Oxide Silicon - The most common type of circuitry.
EDP	Ethylene Diamine Pyrocatechol – One of the first wet silicon etches developed. Originally used for wet etching polysilicon.
HF	Hydrofluoric Acid - An oxide etching acid.
HNA	Hydrofluoric, Nitride, Acetic Acid – An isotropic silicon wet etch.
IC	Integrated Circuit
KOH	Potassium Hydroxide – An anisotropic silicon etchant.
LPCVD	Low Pressure Chemical Vapor Deposition - A thin film deposition system, generally used for depositing highly conformal dielectric or polysilicon films.
LTO	Low Temperature Oxide - An LPCVD deposited silicon dioxide.
PECVD	Plasma Enhanced Chemical Vapor Deposition - A thin film deposition system which depends upon a plasma to deposit the film at a lower temperature than otherwise possible in an LPCVD system
PSG	Phosphosilicate Glass - LTO doped with phosphorous. PSG has a lower reflow temperature and is easier to etch than LTO.
RIE	Reactive Ion Etching
SEM	Scanning Electron Micrograph
TMAH	Tetramethylammonium Hydroxide – A relatively non-toxic anisotropic wet silicon etch.
VLSI	Very Large Scale Integration