ARO 27885.1-EC-CF

TECHNICAL DIGEST

IEEE

# SOLID-STATE SENSOR
# AND ACTUATOR

## Workshop

Sponsored by the IEEE Electron Devices Society

# 1990

Hilton Head Island, South Carolina
June 4-7

# DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

# REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

| 1. AGENCY USE ONLY (Leave blank) | 2. REPORT DATE 1990 | 3. REPORT TYPE AND DATES COVERED Final 1 Mar 90 – 30 Jun 91 |
|---|---|---|

**4. TITLE AND SUBTITLE**

Solid-State Sensor and Actuator Workshop

**5. FUNDING NUMBERS**

**6. AUTHOR(S)**

Richard S. Muller, (principal investigator)

DAAL03-90-G-0117

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Univ. of California
Berkeley, CA 94720

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

ARO 27885.1-EL-CF

**11. SUPPLEMENTARY NOTES**

The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

**12a. DISTRIBUTION/AVAILABILITY STATEMENT**

Submitted for announcement only.

**12b. DISTRIBUTION CODE**

**13. ABSTRACT (Maximum 200 words)**

The objectives of the Workshop were to provide researchers in the field of solid state sensors and microactuators an opportunity for exchange of results, ideas and plans and for in-depth discussions. A "workshop" format was established to facilitate this by scheduling single technical sessions with significant time alloted for discussion of each paper, through an evening discussion session, and by providing for unscheduled time to encourage informal interactions.

**14. SUBJECT TERMS**

Solid-State Sensors, Actuators, Workshop, Microactuators

**15. NUMBER OF PAGES**
191

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| UNCLASSIFIED | UNCLASSIFIED | UNCLASSIFIED | UL |

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

# 1990 SOLID-STATE SENSOR and ACTUATOR WORKSHOP
## Hilton Head Island, SC
### June 4-7, 1990

General Chairman      Ben Hocker
Honeywell Corp.

Program Chairman      Stephen D. Senturia
Massachusetts Institute of Technology

Treasurer      Joseph Giachino
Ford Motor Company

Publications      David S. Eddy
General Motors Research Laboratories

Local Arrangements      Tom Poteat
Sandia National Laboratories

Program Meeting Arrangements      Robert Huber
University of Utah

### Program Committee

Cliff Fung
The Foxboro Company

Wen H. Ko
Case Western Reserve University

Jed Harrison
University of Alberta

Antonio J. Ricco
Sandia National Laboratories

Roger Howe
University of California, Berkeley

Kensall D. Wise
University of Michigan

IEEE Catalog Number 90CH2783-9

Library of Congress Number 89-85516

i

## 1990 IEEE Solid State Sensor & Actuator Workshop
## Hilton Head Island, SC    June 4-7,1990

On behalf of the entire organizing committee, I am pleased to welcome you to Hilton Head Island for the fourth Solid State Sensor and Actuator Workshop.

The objectives of the Workshop are to provide researchers in the field of solid state sensors and microactuators an opportunity for exchange of results, ideas and plans, for in-depth discussions, and for the formation of new friendships and renewal of old. We have established a "workshop" format to facilitate this by scheduling single technical sessions with significant time allotted for discussion of each paper, through an evening discussion session, and by providing for unscheduled time to encourage informal interactions. Please take advantage of these opportunities, participate actively, and I believe we will all find this to be a most stimulating event in an exciting and growing field.

This Workshop is one of the regional meetings held in North America, Europe and Japan during even-numbered years alternating with TRANSDUCERS, the international conference series held in odd-numbered years. TRANSDUCERS '89 was held in Montreux, Switzerland, in June 1989, and TRANSDUCERS '91 will take place June 24-28, 1991, at the Hyatt Regency Hotel in San Francisco.

The Workshop would not be possible without the support of a number of organizations and individuals: The Workshop is sponsored by the IEEE Electron Devices Society which provides financial guarantees. We offer special thanks to the National Science Foundation and George Hazelrigg for the generous support provided for travel assistance funds which allow many graduate students to attend the Workshop. We are greatful to the US Army Research Office and Mike Stroscio for supporting the publication of the technical digest. Richard Muller of the University of California is due our special thanks for the work of arranging this outside funding. We also thank the General Motors Corporation for its continued contribution to these Workshops by offering publication services at attractive rates, and Dave Eddy of GM who has made this come about.

Finally, I want to thank the committee for their work in arranging the technical program and the authors for providing the exciting contributions which will serve to stimulate our discussions and future efforts.

Enjoy your Workshop!

G. Benjamin Hocker
General Chairman

# 1990 SOLID-STATE SENSOR AND ACTUATOR WORKSHOP
June 3-7, 1990
Mariner's Inn, Hilton Head, SC

## PROGRAM

### Sunday, June 3

7 - 9 PM   WELCOME RECEPTION with no host bar

### Monday, June 4   Session 1 - Acoustic and Resonant Devices
Chairman:  Stephen D. Senturia

7:30 AM   BREAKFAST

8:00 AM   Welcome and Introductions

8:15 AM   Quartz microbalance studies of solvent and ion transport in thin polymer films in sensor applications of the QCM
Steven J. Lasky, Howard R. Meyer, and Daniel A. Buttry, University of Wyoming (Invited)

10:15   COFFEE BREAK

### Monday, June 4   Session 2 - Microactuators - 1
Chairman:  David S. Eddy

11:55   LUNCH

### Monday, June 4   Session 3 - CAD
Chairman:  Wen H. Ko

2:05 - 4:00   POSTER SESSION with refreshments provided

### Tuesday, June 5   Session 4 - Chemistry and Physics of Materials and Processes
### Chairman: Jed Harrison

### Tuesday, June 5   Session 5 - Integrated Sensor Systems; Magnetic Sensors
### Chairman: Roger T. Howe

### Wednesday, June 6   Session 6 - Fabrication Methods, Structures, and Materials
### Chairman: Kensall D. Wise

*QUARTZ MICROBALANCE STUDIES OF SOLVENT AND ION TRANSPORT*
*IN THIN POLYMER FILMS IN SENSOR APPLICATIONS OF THE QCM*

Steven J. Lasky, Howard R. Meyer, and Daniel A. Buttry*

Department of Chemistry
University of Wyoming
Laramie, Wyoming 82071-3838

## ABSTRACT

Recently, the quartz crystal microbalance (QCM) has been applied to the study of mass changes within thin polymer films immobilized on the surface of the QCM. These studies have been in two contexts: 1) the determination of compositional changes within such thin films which have been subjected to electrochemical oxidation or reduction (i.e. injection or removal of charge) which necessarily causes ion transport to maintain electroneutrality and frequently causes solvent transport as well, and 2) the development of sensor elements capable of transducing chemical events, such as binding of an analyte to a host molecule immobilized within the thin film, into electrical signals. This talk will focus on the fundamental chemical and physical processes which can occur in such situations, and their influence on the resonant frequency and conductance spectrum of the QCM/thin film composite resonator. In particular, we will discuss the use of the conductance spectrum as a diagnostic tool for the evaluation of changes in the viscoelastic properties of the thin film which might be induced, for example, by solvent incorporation into the film. As examples, the operation of a real-time glucose sensor based on the enzyme hexokinase immobilized within a poly(acrylamide) gel on the QCM surface and a potassium ion sensor based on a crown ether host immobilized within a poly(vinylchloride) film will be discussed.

## INTRODUCTION

While the use of piezoelectric devices in chemical sensor applications is not new (1), activity in the area was quite limited until recently. The resurgence of interest in the area has occurred for a variety of reasons, the most important of which are their potentially low cost and excellent sensitivity. Several recent reviews have been published describing the use of surface acoustic wave (SAW) (2-4) and QCM devices (1,4-6) in a fairly wide variety of sensor applications. In the present contribution, we describe the use of the QCM in sensor applications involving thin polymer films.

Most of the past studies of QCM applications to sensors have relied upon a straightforward relationship between the resonant frequency of the QCM and the mass of a thin film on its surface (see below). In such studies, the thin film contains (or is composed entirely of) a so-called host molecule whose purpose is to bind to the analyte of interest, thereby generating a mass change of the thin film. The mass change causes a change in the resonant frequency of the oscillating piezoelectric device. Transduction of the chemical information to an electrical signal generally occurs by virtue of a broadband oscillator circuit whose function is to track the resonant frequency of the piezoelectric device. (In our QCM sensors, the working definition of the resonant frequency is that frequency at which the conductance of the QCM/thin film composite resonator is at its maximum value.) Such instrumentation provides a real-time measure of the mass of the thin film in the form of the resonant frequency of the device.

The host species used in such sensors are typically capable of some degree of selective binding to the analyte. Examples of such host species are synthetic ionophores such as crown ethers (which bind relatively selectively to a variety of mono- and divalent cations), enzymes (which generally bind with great selectivity to their substrates), immunochemical species (such as antibodies or antigens), and nucleic acid derivatives (such as DNA or RNA fragments). In all of these cases, the sensitivity and selectivity of the sensor are dependent on the binding characteristics of the host. The selectivity is mostly inherent to the host species itself, depending on its binding constant for the analyte, although the nature of the sample to be tested (e.g. the presence of possible interferents) also determines this to some extent. Thus, the use of highly selective host species is usually quite desirable. Biomolecules are especially attractive due to the remarkable selectivity of biorecognition processes. On the other hand, the sensitivity depends on a great many factors including the operating frequency of the QCM, the binding constant of the host for the analyte, the analyte concentration in the sample of interest, the noise level of the instrumentation, etc.

For cases in which the thin film is rigid (i.e. perfectly elastic, and with zero viscosity) the frequency is related to the mass change by the Sauerbrey equation:

$$\Delta f = - C_f \Delta m = -(f_o^2 / N \rho_q) \Delta m \qquad (1)$$

in which $\Delta f$ is the frequency change caused by the mass change per unit area on the surface of the QCM, $\Delta m$, $C_f$ is the sensitivity constant, $f_o$ is the operating frequency of the QCM prior to the mass change (typically between 1 and 10 MHz), N is the frequency constant for quartz (=0.167 MHz cm), and $\rho_q$ is the density of quartz (=2.648 g cm$^{-3}$) (5). Note that the negative sign indicates a frequency decrease will result from a mass increase. The crystals used in this contribution resonate at 5 MHz, so $C_f$ is 56.6 Hz cm$^2$ $\mu$g$^{-1}$, meaning that a mass change of 1 $\mu$g distributed across a one square centimeter area will cause a change in the oscillation frequency of 56.6 Hz. This behavior is called rigid layer behavior. This quantitative relationship between frequency and mass forms the basis for the widespread commercial use of QCM's as thin film thickness monitors for vacuum deposition systems.

Even in the absence of changes in rigid film mass, large frequency changes occur when QCM crystals are immersed into liquids. This is due to the energy loss from viscous loading by the liquid. Physically, this energy loss is caused by the strong attenuation of the shear wave which is launched into the viscous medium by the transducer. Kanazawa and Gordon (7) have presented a quantitative relationship between the resonant frequency of the QCM and the density and viscosity of the adjacent medium for QCM's which is given in Equation 2:

$$\Delta f = - f_o^{3/2} (\rho_1 \eta_1 / \pi \rho_q \mu_q)^{1/2} \qquad (2)$$

in which $\rho_1$ and $\eta_1$ are the density and viscosity of the liquid, respectively, and $\rho_q$ and $\mu_q$ are the density and shear modulus (=2.947 x 10$^{11}$ g cm$^{-1}$ s$^{-2}$)

of quartz, respectively. The decay length of this shear wave in pure water at room temperature for a QCM operating at 5 MHz is ca. 2500 Å. This dependence of resonant frequency on viscosity of resonating piezoelectric devices has been used as the basis of viscosity sensors.

When polymer films are used as a means to immobilize the host species onto the QCM surface, it is possible that viscoelastic behavior (8) on the part of the polymer film will render invalid the linear mass-frequency relationship given in Equation 1. This may happen, for example, if the entrance of the analyte into the film to bind to the host species causes a significant change in the extent of solvent swelling of the film, which, under the right conditions, could translate to a change in the viscosity of the film. This type of situation has already been observed in the development of univalent ion sensors based on crown ether hosts in poly(vinylchloride) films (9). Thus, it is crucial that both theoretical models and experimental tests of deviations from rigid layer behavior be available for those cases in which viscoelastic films are used in QCM sensors.

Kanazawa and coworkers have recently provided a detailed, analytical solution to the problem of the frequency change expected for deposition of a thin, viscoelastic film onto a QCM (10). The important film variables which are used to describe the frequency change are the thickness (1), the density ($\rho$), the shear modulus ($\mu$), and the viscosity ($\eta$). Note that the latter two terms are frequency dependent. Their solution provides a method for calculating the frequency change expected for deposition of such a film or, more importantly in the sensor context, for changes in the materials properties of the film caused by changes in its composition. However, the system is under-determined, in the sense that four film variables conspire to determine a single experimental observable, namely the resonant frequency. Thus, it is not possible to use the resonant frequency alone to arrive at a determination of whether or not the Sauerbrey equation is valid.

It is possible to gain a semi-quantitative estimation of the changing influence of film viscosity on the QCM resonant frequency through the measurement of the conductance spectrum of the QCM/thin film composite resonator. Such n  urements have been used for many years to study the viscoelasticity of polymer solutions, gels, and solid samples, and the classic text by Ferry gives very lucid discussions of the considerations involved in such measurements (8). We describe below the use of frequency dependent conductance measurements to probe changes in thin film viscoelasticity during QCM sensor operation for two different sensors, a glucose sensor and a potassium ion sensor. Note that the changes in the conductance spectrum described below are superimposed on those caused by the viscosity of the solution, which generally cause an increase in the full width at half height of the conductance peak from ca. 50 Hz (in air) to ca. 1500 Hz (in water), although these values depend somewhat on the crystal type and the solution properties (12).

EXPERIMENTAL

The experimental design, apparatus, and procedures for the measurement of QCM resonant frequencies (11) and conductance spectra (12) have been previously described. The preparation of the thin poly(acrylamide) (PAC) films containing hexokinase for the glucose sensors has also been described (11,13). The details of the preparation of the poly(vinylchloride) (PVC) films containing plasticizer and crown ether for the potassium ion sensor will be reported in a future contribution (9). Briefly, thin films of 43% PVC, 56% dioctylsebacate, and 1%

dibenzo-18-crown-6 (DB18C6) were prepared by casting a tetrahydrofuran solution of the appropriate concentrations onto the QCM electrode surface and rapidly drying. The solution compositions and casting aliquots (ca. 1-10 microliters) were chosen so as to cast films of ca. 2 micron thickness.

Conductance spectra were obtained with an HP 4192A low frequency impedance analyzer using an IEEE 488 interface to an IBM PC-AT within the ASYST programming environment and custom software.

RESULTS AND DISCUSSION

In previous publications, we have described the details of operation of the glucose sensor (11,13). In principle, its mode of detection is straightforward. The enzyme hexokinase, the biological function of which is to phosphorylate glucose to form glucose-6-phosphate (G6P) at the expense of ATP (14), is physically entrapped within a PAC gel. This is possible because the pore size of the gel can be easily manipulated by control of the extent of crosslinking (15). In the absence of ATP, glucose will reversibly bind to hexokinase without reaction to produce G6P, meaning that binding will occur as the glucose concentration in the gel increases and vice versa. The implication of this is that, within a certain range (i.e. the dynamic range of the sensor, which is determined by the binding constant of the enzyme-substrate pair (13)), the amount of glucose bound within the gel will provide a measure of the external glucose concentration, to the extent that the glucose concentration within the gel is proportional to the external concentration. This binding changes the mass of the gel, which should result in a change in the resonant frequency of the QCM. Under rigid layer conditions, the frequency will change linearly with the mass.

Figure 1 shows the frequency response of a QCM glucose biosensor toward changes in external glucose concentration. Note that the frequency changes are large, and vary in a roughly linear fashion with the external glucose concentration. Use of the Sauerbrey equation to calculate the expected mass change (assuming one bound glucose per enzyme) reveals that the mass change is much larger than expected.

We suspected the cause of this to be viscoelastic behavior on the part of the gel, so the conductance spectrum of the QCM/thin film composite resonator was measured. It is shown in Figure 2. The salient feature of the curves is that the width of the conductance spectrum, which provides a measure of the viscous loading due to the gel (12), increases dramatically upon glucose binding. Thus, the binding event causes a marked change in the viscosity of the gel. A possible source of this is the large conformational change which hexokinase undergoes when glucose binds to it (14). We are currently investigating this in more detail, but it is clear, both from theroetical treatments (10) and these experimental data that the mechanical properties of the thin film can have a profound influence on the frequency response of the QCM sensor. It is especially interesting that these mechanical effects can amplify the sensor signal, as in the present case.

Figure 3 shows the frequency response of a QCM potassium ion sensor towards changes in external potassium ion concentration. The signal from the reference crystal is also shown. This reference crystal also has a PVC film on it, but this film contains no crown ether. As in the case of the glucose sensor described above, the frequency changes are large and roughly linear in the external potassium ion concentration. Again, they are much larger than predicted by the Sauerbrey equation. Figure 4 shows the conductance spectrum of the QCM/thin film composite resonator, again revealing a very broad resonance indicative of extreme viscous loading by the

2

film. (Compare the large full width at half maximum of this curve with that for the a bare crystal in the same solution (ca. 1500 Hz). This is undoubtedly due to the large fraction of the film which is composed of plasticizer. In this case, we believe that the majority of the frequency response is due to an influx of water which accompanies the insertion of potassium ion into the film. This solvent would serve to further plasticize the film, thereby increasing its viscosity and causing increasing deviations from rigid layer behavior. We are currently conducting experiments using $D_2O$ (which weighs 10% more per mole) in place of $H_2O$ to test this hypothesis. The preliminary results of these experiments indicate that much of the mass change is, indeed, due to solvent accompanying the potassium ions into the film. However, there is no question about the presence of viscous loading by the film, even based on the data presented here.

## CONCLUSIONS

These results reveal that viscoelastic behavior on the part of the thin film can have a profound influence on the nature of the frequency response of a QCM/thin film sensor. It is especially interesting that amplification of the chemical event can occur by virtue of these effects. The disadvantage of such effects is that the quantitative relationship between mass and frequency can no longer be assumed valid. However, even in such cases, calibration of the device is possible, allowing for a relatively wide variety of possible applications. The theory by Reed, Kaufman, and Kanazawa (10) provides an extremely useful framework in which to analyze the behavior of viscoelastic films, however, it is limited in that it does not take into account the presence of a viscous medium adjacent to the film (i.e. the solution). Its great advantage is that it is an intuitive, physically based theory which allows for qualitative predictions to be made given reasonable knowledge of the mechanical properties of the film.

The use of QCM devices is shown here to be a potentially very useful strategy for the development of chemical sensors based on binding. They possess the necessary sensitivity, are extremely versatile in the derivatization chemistry which can be used to place host species within thin films at their surfaces, and the instrumentation required for their operation is simple and inexpensive. Perhaps their most important advantage over SAW devices is the possibility of their use in *in situ* applications, where SAW devices might experience too much viscous loading to allow for stable oscillation.

## REFERENCES

1. W.H. King, Anal. Chem., 36 (1964) 1735-9.
2. Chemical Sensors and Microinstrumentation, R. Murray, R.E. Dessy, W.R. Heineman, J. Janata, and W.R. Seitz, Eds., ACS Symposium Series No. 403; American Chemical Society, Washington, D.C., 1989.
3. D. Ballantyne and H. Wohltjen, Anal. Chem., 61 (1989) 704A.
4. M.D. Ward and D.A. Buttry, Science, in press.
5. M.R. Deakin and D.A. Buttry, Anal. Chem., 61 (1989) 1154A.
6. J.F. Alder and J.J. McCallum, Analyst, 108 (1983) 1169.
7. K.K. Kanazawa and J.G. Gordon, Anal. Chim. Acta, 175 (1985) 99.
8. J.D. Ferry, Viscoelastic Properties of Polymers; Wiley Interscience, New York, 1980.
9. H.R. Meyer and D.A. Buttry, manuscript in preparation.
10. C.E. Reed, K.K. Kanazawa, and J.F. Kaufman, J. Appl. Phys., in press.
11. S.J. Lasky and D.A. Buttry, Chapter 16, pp. 237-46, in reference 2.
12. R. Borjas, and D.A. Buttry, J. Electroanal. Chem., in press.
13. S.J. Lasky and D.A. Buttry, American Biotechnology Laboratory, 8 (1990) 8-16.
14. W.F. Bennett and T.A. Steitz, Proc. Natl. Acad. Sci., 75 (1978) 4848.
15. A.T. Andrews, Electrophoresis: theory, techniques, and biochemical and clinical applications; Clarendon Press, Oxford, 1981.
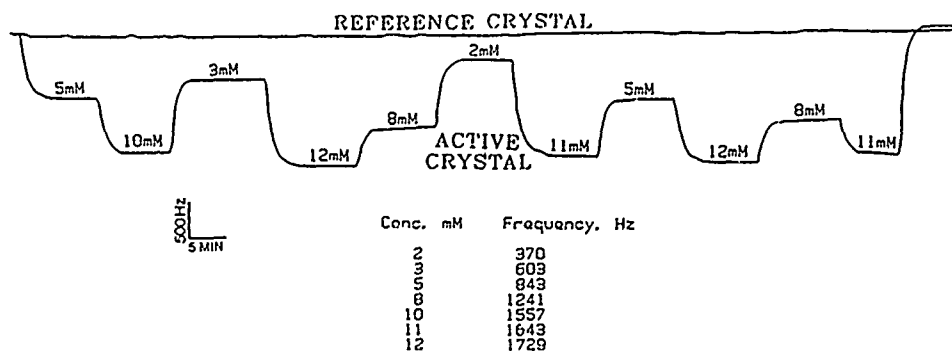
Figure 1. The frequency response to randomly selected glucose concentrations in the 2-12 mM range.

| Conc. mM | Frequency, Hz |
|---|---|
| 2 | 370 |
| 3 | 603 |
| 5 | 843 |
| 8 | 1241 |
| 10 | 1557 |
| 11 | 1643 |
| 12 | 1729 |

3
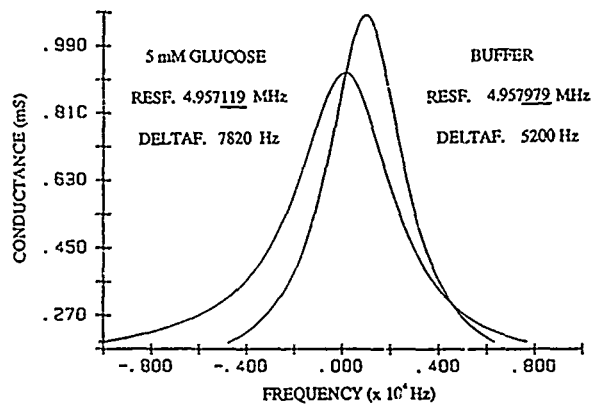
Figure 2. Conductance spectra for a glucose sensor film in buffer and 5 mM glucose. Note the large increase in the full width at half height of the conductance peak caused by glucose binding.
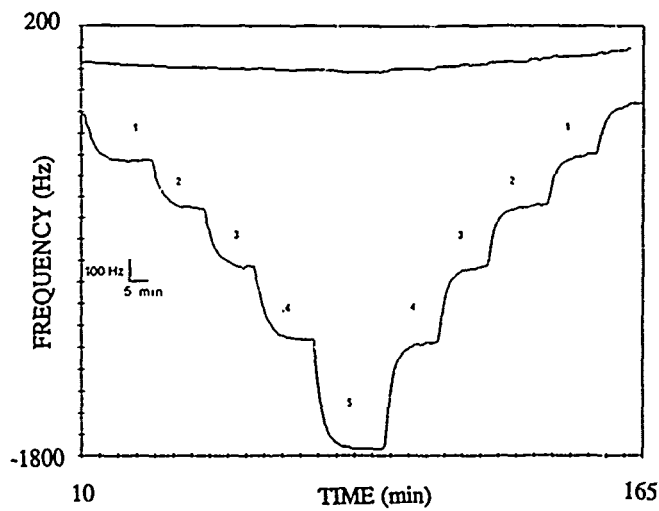


Figure 3. Concentration profile for a typical set of PVC films. Top trace: reference film response. Bottom trace: sensor film response. KCl concentrations are in millimolar (mM) units.
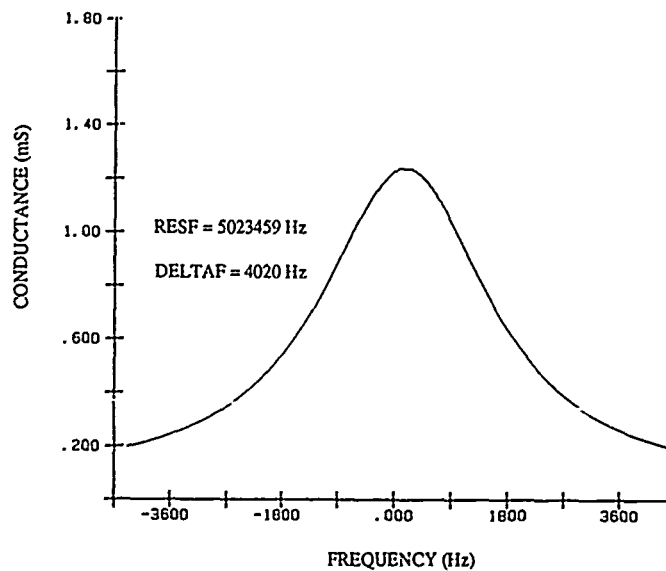


Figure 4. Conductance spectrum for a PVC sensor film in 1 mM KCl.

4

# MULTIPLE-FREQUENCY SURFACE ACOUSTIC WAVE DEVICES AS SENSORS

*Antonio J. Ricco and Stephen J. Martin*

Microsensor Division 1163
Sandia National Laboratories
Albuquerque, New Mexico 87185    USA

## ABSTRACT

We have designed, fabricated, and tested a multiple-frequency acoustic wave (MUFAW) device on ST-cut quartz with nominal SAW (surface acoustic wave) center frequencies of 16, 40, 100, and 250 MHz. The four frequencies are obtained by patterning four sets of input and output interdigital transducers of differing periodicities on a single substrate. Such a device allows the frequency dependence of AW sensor perturbations to be examined, aiding in the elucidation of the operative interaction mechanism(s). Initial measurements of the SAW response to the vacuum deposition of a thin nickel film show the expected frequency dependence of mass sensitivity in addition to the expected frequency independence of the magnitude of the acoustoelectric effect. By measuring changes in both wave velocity and attenuation at multiple frequencies, extrinsic perturbations such as temperature and pressure changes are readily differentiated from one another and from changes in surface mass.

## INTRODUCTION

As the utility of AW (acoustic wave) devices for chemical sensing and materials characterization is explored in more detail by an increasing number of workers, it has become apparent that monitoring only the frequency change resulting from a sensor perturbation can yield less than all the available information. For example, we recently pointed out the additional insight provided by simultaneously recording changes in both acoustic wave velocity (v) and attenuation ($\alpha$) when monitoring photopolymerization, acoustoelectric coupling, or liquid viscosity (1,2).

In general, the various perturbations that can affect AW propagation characteristics depend differently upon frequency (f). Furthermore, for a given perturbation, the frequency dependence of the fractional change in wave velocity, $\Delta v/v_0$, often differs in its frequency dependence from the change in wave attenuation. Thus, to further supplement the information available from acoustic wave sensors, we designed and fabricated the multifrequency acoustic wave device shown schematically in Figure 1. This MUFAW device enables the frequency dependence of perturbations to wave velocity and attenuation to be recorded. Application of such devices to problems in chemical sensing and materials characterization should make it possible to unambiguously separate from one another some of the many and varied perturbations which affect acoustic wave propagation.

In this paper, we report the design details and preliminary results from the MUFAW device. We have utilized the vacuum evaporation of a thin nickel film as a means to confirm the frequency dependence of mass sensitivity. In addition, the Ni film deposition provides an opportunity to examine a perturbation which should have no dependence on frequency, namely acoustoelectric coupling (3). Measurements of temperature and pressure effects at multiple frequencies are reported as well.
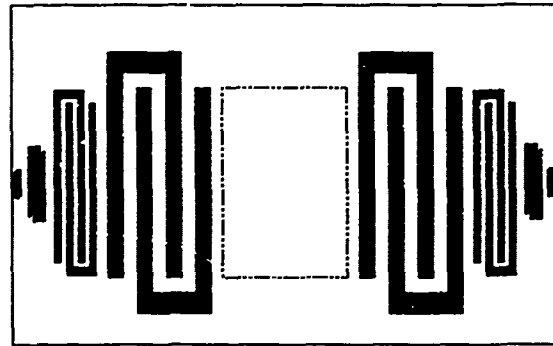


*Figure 1.   Schematic diagram of multifrequency acoustic wave device. Actual input and output transducers consist of 25 pairs of interdigitated electrodes for the 16 and 40 MHz transducers and 50 pairs for 100 and 250 MHz transducers. The interaction region is defined by the dashed box.*

### Background: Acoustoelectric Coupling

The coupling of an acoustic wave propagating in a piezoelectric substrate to charge carriers in an adjacent medium provides a novel mechanism for the study of changes in electrical conductivity in thin solid films and in solution (3,4). This acoustoelectric effect can be utilized to construct new chemical sensors and to study conductivity effects in general. For a SAW device, a surface film having sheet conductivity $\sigma_s$ perturbs the wave velocity by an amount $\Delta v/v_0$ and changes the attenuation (normalized by the wavenumber, k) by an amount $\Delta\alpha/k$, according to (3):

$$\frac{\Delta v}{v_0} = -\frac{K^2}{2}\ \frac{\sigma_s^2}{\sigma_s^2 + v_0^2 C_s^2} \tag{1}$$

$$\frac{\Delta\alpha}{k} = \frac{K^2}{2}\ \frac{v_0 C_s \sigma_s}{\sigma_s^2 + v_0^2 C_s^2} \tag{2}$$

where $K^2$ is the electromechanical coupling coefficient and $C_s$ is the capacitance/length of the substrate ($C_s = \epsilon_0 + \epsilon_s$, the sum of the permittivities of the substrate and the region above it).

## EXPERIMENTAL METHODS

Multiple-frequency acoustic wave devices were designed and fabricated in the Microsensor Division at Sandia National Laboratories. Each device has interdigital transducers of four different periodicities to excite and detect Rayleigh waves (SAWs) on ST-cut quartz at 16, 40, 100, and 250 MHz. The same transducers can excite shear-horizontal acoustic plate modes (SH-APMs) on ST-quartz at roughly 1.6 times the SAW frequencies, i.e. 25.6, 64, 160, and 400 MHz. The transducer periodicity defines the acoustic wavelength, $\lambda_0$, at the center frequency; the width of the fingers and spaces comprising each transducer is ¼ of its periodicity. Table I lists relevant design parameters for each of the four transducers.

Devices were fabricated on 0.5 mm-thick ST-cut quartz wafers. Transducers were defined photolithographically, using an etching process, from 100 to 200 nm-thick Au-on-Cr metallization. Wafers were diamond sawed into 3.6 x 1.3 cm devices, each of which was then attached to a custom-fabricated

## TABLE I

Design Parameters for Multifrequency Acoustic Wave Device

| SAW Center Frequency, MHz | 16 | 40 | 100 | 250 |
|---|---|---|---|---|
| Acoustic Wavelength, $\mu$m | 200 | 80 | 32 | 12.8 |
| No. of Finger Pairs | 25 | 25 | 50 | 50 |
| Acoustic beam width,[1] $N_{\Lambda_o,1}$ | 50 | 100 | 100 | 100 |
| Acoustic path length,[1] $N_{\Lambda_o,2}$ | 75 | 280 | 850 | 2400 |

[1]$N_{\Lambda_o,1}$ = length of transducer fingers in acoustic wavelengths; $N_{\Lambda_o,2}$ = center-to-center spacing between input and output transducers in acoustic wavelengths.

alumina PC board having a number of 50 Ω, lithographically defined coplanar-waveguides running from the device to contact areas along the board's periphery. 25 $\mu$m Au wire bonds were made between the transducer bonding pads and the waveguides. In some cases, several input transducers (of differing periodicity) were bonded in parallel to a single guide; corresponding output transducers were then similarly connected in parallel. PC board-mounted devices were installed in a custom-fabricated brass test case utilizing Au-plated Cu/Be spring contacts to connect the board's contact pads to SMA jacks, which were connected by coaxial cable to the external circuitry.

The electronic measurement circuitry can be configured in a number of different ways. The familiar oscillator loop can be constructed in the manner described elsewhere (5) for each input/output transducer pair. To measure oscillation frequency and device insertion loss, directionally coupled outputs from each loop are connected to a frequency counter and power meter (or vector voltmeter) via a computer-controlled rf multiplexer (described below), economizing on instrumentation.

The majority of the measurements reported below were made using the computer-controlled phase-locked loop configuration shown schematically in Figure 2. A 10 dB directional coupler directs a fraction of the power from a synthesized source (HP 8656B) to the reference channel of the vector voltmeter (HP 8508A). The balance of the power enters one of the MUFAW input transducers via the VHF multiplex card of an HP 3488A Switching & Control Unit. The $M_i$ denote impedance matching networks (these have not yet been incorporated). The output from the device is directed to the second channel of the vector voltmeter by a second multiplexer card.
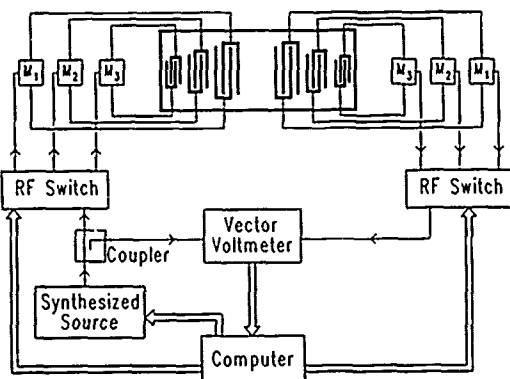


*Figure 2. Schematic representation of computer-controlled phase-locked loop used to simultaneously monitor acoustic wave velocity and attenuation. Vector voltmeter records phase shift and differential power level; as sensor is perturbed, computer adjusts source frequency to keep the phase shift constant for a given transducer pair. Computer-controlled rf multiplexers (denoted "RF Switch") allow several different transducer sets having different characteristic frequencies to be probed in rapid succession. The $M_i$ denote impedance-matching networks*

A Hewlett-Packard 9816 computer acquires differential phase and insertion loss data from the vector voltmeter in addition to setting the multiplexers to the appropriate channel and the source to the appropriate frequency. During each cycle (approximately every 3 s), the computer adjusts the source frequency sent to each input transducer in order to maintain the phase shift at a constant value. Thus, as wave velocity changes due to external perturbations, the source frequency tracks it according to:

$$\frac{\Delta f}{f_o} = \kappa \frac{\Delta v}{v_o},\qquad [3]$$

where $\kappa$ is the fraction of the acoustic path length being perturbed. Insertion

loss and source frequency, along with any other varying parameter (film thickness, temperature, etc.), are then recorded as a function of time by the computer.

Thermal deposition of Ni was carried out in a cryogenically pumped vacuum system with a base pressure near $10^{-7}$ Torr. Radio-frequency feed-throughs allow the various acoustic wave propagation parameters to be monitored *in situ* during the process of film deposition. Metal film thicknesses are measured using a commercial (Inficon) quartz crystal microbalance (QCM).

## RESULTS AND DISCUSSION

### Frequency Response

Frequency response data for the MUFAW device, Figure 3, reveal Rayleigh peaks (R) at approximately 15.5, 39.3, and 97.3 MHz. Various acoustic plate modes are apparent as well. This set
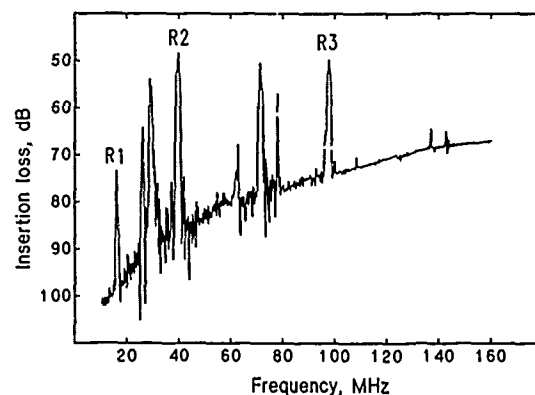


*Figure 3. Frequency response of the multifrequency acoustic wave device. Impedance matching networks have not yet been incorporated, accounting for the relatively large insertion loss of the acoustic peaks. Rayleigh peaks at 15.5, 39.3, and 97.3 MHz are indicated R1, R2, and R3, respectively. A number of acoustic plate modes are visible as well, including shear-horizontal acoustic plate modes at 1.6 times the frequency of R1 and R2.*

of data was obtained with the transducers for these three frequencies connected in parallel. Shear-horizontal acoustic plate modes are visible near 25 and 63 MHz. The SH-APM associated with the 97 MHz SAW would normally appear near 158 MHz (6); this mode is either entirely absent or may be represented by one of the highly attenuated features near 140 MHz. This is expected: the SH-APM has displacement components on both faces of the quartz plate and, in this particular case, the back side of the plate is unpolished and is secured using adhesive to a PC board, conditions which should attenuate the wave. It is also reasonable that the two lower frequency SH-APMs are less affected by the attachment and surface finish of the quartz plate, since the propagation paths at these frequencies contain many fewer acoustic wavelengths (75 and 280 $\Lambda_o$ for the 25 and 63 MHz SH-APMs, respectively, vs. 850 for the 158 MHz APM). The acoustic propagation loss, L (in dB), is proportional to the number of wavelengths in the wave path:

$$\Delta L = 54.6 n_{\Lambda_o} \frac{\Delta \alpha}{k},\qquad [4]$$

where $n_{\Lambda_o} = \kappa N_{\Lambda_o,2}$. Furthermore, attenuation due to coupling to a viscous medium (the adhesive in this case) often increases with frequency (6,7), further obliterating the 158 MHz mode.

Not shown in Figure 3 is the region near 250 MHz. This is because there is nothing to see. For one of several devices, a feature in the vicinity of 250 MHz which might be attributed to the highest frequency SAW was recorded. In general, however, we

believe that the acoustic path between the 250 MHz transducers is so long (2400 $\lambda_0$) that diffraction of the acoustic beam (8) results in significant signal loss. In addition, any acoustic loss due to attenuation will be largest at this frequency due to the large number of wavelengths between transducers. Finally, scattering off the other transducers has the greatest effect at this frequency, both because the wave must traverse more of the other transducers and because the thickness of the metallization is a larger fraction of the acoustic wavelength.

These considerations could lead one to question our sanity in choosing this design. The advantages of the sort of design represented by Figure 1 are that a single interaction region (the dashed box) or some part thereof is probed by all the waves of various frequencies. Reversing the order of the transducers would alleviate a number of the problems mentioned above, but creates others. In a "reversed" design, a portion of the wavefronts from the low frequency transducers would pass through the various high frequency transducers, causing parts of the wavefront to arrive at the relevant output transducer out of phase, i.e. the waves would not be straight-crested. We are therefore currently considering designs which give each pair of transducers an independent path having no intervening transducers. Although this puts more stringent requirements on the point-to-point uniformity of a thin film used for chemical detection or film characterization (different regions of the film will be probed at different frequencies), it may be the best way to make the higher frequency measurements.

## Nickel Deposition

Acoustoelectric coupling. The effects on SAW propagation characteristics of acoustoelectric coupling, as outlined by Eqs. 1 and 2, are revealed by the results of the experiment depicted in Figure 4, which is a plot of $\Delta v/v_0$ and $\Delta\alpha/k$ as a function of the
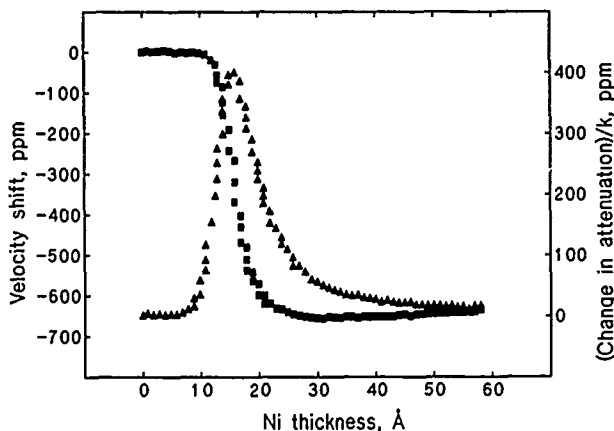


Figure 4. Experimental data showing the acoustoelectric effect in the form of fractional velocity shift (■) and normalized attenuation change (▲), both in parts per million, as a function of vacuum-evaporated nickel film thickness. Data are for a 97 MHz ST-cut quartz SAW device. The $\Delta v/v_0$ data have been corrected for the effect of mass loading.

thickness of a Ni film vacuum-evaporated onto the bare surface of a 97 MHz (single frequency) SAW device (2). It should be noted that the fractional velocity shift data of Figure 4 have been corrected for the effect of mass loading utilizing the relation (3):

$$\frac{\Delta v}{v_0} = -c_m f_0 \rho_s, \qquad [5]$$

in which $c_m$ is the SAW mass sensitivity coefficient and $\rho_s$ is the

mass/area added to the surface. The Ni thickness recorded by the QCM was used to compute the mass/area, and this linear perturbation was subtracted from the data according to Eq. 5. The attenuation data were calculated directly from changes in the measured insertion loss using Eq. 4.

Figure 5 shows the results of the vacuum deposition of a Ni film onto the area between the transducers (as defined by the dashed box of Figure 1) of the MUFAW device. The two curves shown were obtained at 39.3 (■) and 97.3 MHz (▲) using the computer-controlled phase-locked loop configuration of Figure 2.
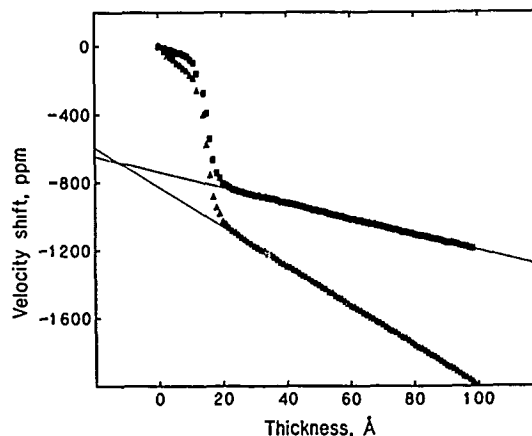


Figure 5. Effect of the evaporation of a thin Ni film on the acoustic wave velocity for the device represented by Figure 1. Data are for 39 (■) and 97 (▲) MHz SAWs. The acoustoelectric effect (represented by the sigmoidal portion of the curves) is frequency-independent in magnitude, while the ratio of the slopes in the linear, mass-loading regions of the data is 2.5, precisely the ratio of the frequencies. Straight lines are linear least-squares fits to the data between 30 and 100 Å of Ni.

Fractional velocity shift was calculated from the measured frequency change according to Eq. 3, with $\kappa = 0.25$ and 0.19 for the 39 and 97 MHz SAWs, respectively.

Eqs. 1 and 2 indicate that the magnitude of the acoustoelectric effect should be independent of frequency. Figure 5 indeed reveals that the acoustoelectric effect, which is manifested in the sigmoidal portion of the curves near 15 Å Ni thickness, is about the same for the two frequencies monitored. Measurements of propagation loss, carried out concurrently with the velocity measurements shown in Figure 5, give peak values of 190 and 220 ppm for $\Delta\alpha/k$ at 39 and 97 MHz, respectively. The difference in these values is insignificant within our estimates of their uncertainty.

Mass sensitivity. The linear regions in the data of Figure 5 are the result of mass loading. Linear least-squares fits to the data in the 30 - 100 Å thickness range, shown by the solid lines, give a measure of the coefficient of mass sensitivity. The results of these fits are slopes of -4.58 and -11.6 ppm/Å, consistent with the ratio of SAW frequencies being 2.5. The slopes yield values for $c_m$ in Eq. 4 of 1.32 and 1.34 cm$^2$/g-MHz for the 39 and 97 MHz SAWs, respectively. The $c_m$ value for the 97 MHz SAW is in excellent agreement with the previously reported value of 1.34 cm$^2$/g-MHz (9); the sensitivities at the two frequencies also agree quite well.

The oft-quoted $f^2$ dependence of mass sensitivity for SAW devices is based on reporting $\Delta f$ alone, incorporating both the $f_0$ from the denominator of $\Delta f/f_0$ and the factor of $f_0$ shown explicitly

7

in Eq. 5 into the value of $c_m$. In fact, our measurements at multiple frequencies reveal that, for the particular test arrangement used, the short-term noise levels (in Hz) are quite comparable at 16, 39, and 97 MHz. Thus, the short-term noise-limited mass sensitivity for the 97 MHz SAW is indeed about 6.25 times greater than that of the 39 MHz SAW. (We choose to express our results in terms of $\Delta f/f_0$ due to the ease of comparison with the various models for perturbations to the wave velocity.)

Consistent and reproducible mass sensitivity results were not obtained for the 16 MHz SAW. In fact, the mass sensitivity of this mode, measured in the same manner as described above, was inconsistent from one measurement to the next, sometimes nonlinear, and in some cases exceeded that of the 39 MHz SAW. We attribute this to the fact that the acoustic wavelength at 16 MHz is a significant fraction (about 40%) of the thickness of the quartz substrate: when the thickness of the medium along which the SAW propagates becomes comparable to the acoustic wavelength, the SAW degenerates into a Lamb wave (9,10) (a type of acoustic plate mode). We are currently investigating the effect of quartz substrate thickness on this phenomenon.

Temperature & Pressure Effects

A further demonstration of the utility of multiple frequencies is provided by Figure 6. The effects of changes in temperature and pressure are displayed by plotting $\Delta\alpha/k$ vs. $\Delta v/v_0$. Room temperature at 1 atm pressure is taken as the arbitrary reference point at which $\Delta v/v_0 = 0$ and $\Delta\alpha/k = 0$. Each set of points surrounded by a solid boundary represents a single temperature/pressure combination. Closed symbols represent temperature changes, while open symbols are used for pressure change. The three smooth curves are approximate isobars (temperature response at constant pressure) at one atmosphere for the three frequencies examined.
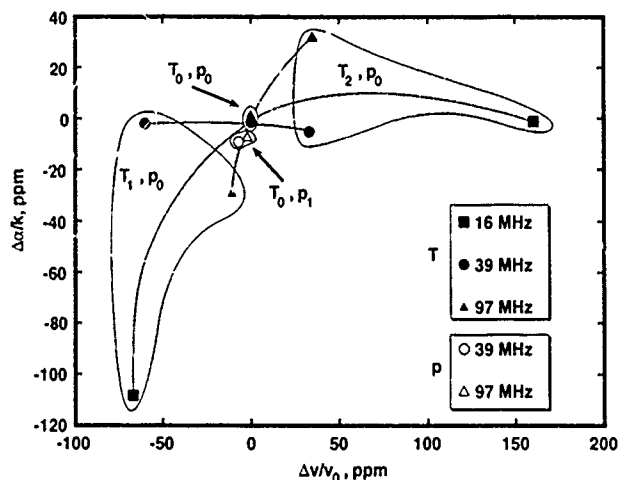


*Figure 6. Change in normalized attenuation plotted against fractional velocity shift at 16 (■), 39 (●), and 97 (▲) MHz for the device of Figure 1: temperature and pressure are the variable parameters. $T_0 = 24\ C$ and $p_0 = 1$ atm are arbitrarily chosen as zero points. Each encircled group of points represents the response to a given perturbation indicated by the T and p values: $T_1 = 10\ C$, $T_2 = 40\ C$, and $p_1 = 10^6$ Torr. Filled points result from temperature perturbations; open points represent a pressure change. Because none of the three temperature responses (smooth curves) intersects the pressure change points, these two perturbations cannot be mistaken for one another. On this plot, mass changes cause purely horizontal excursions at all three frequencies.*

It should be pointed out that the temperature effects in particular and the pressure effects to a lesser extent are not a result of the properties of the ST-quartz device alone, but also a consequence of the manner in which the device is fastened to the PC board. Different thermal expansion coefficients for quartz, the adhesive, and the PC board mean that changing any of the packaging details will alter the shape of the curves of Figure 6. Nevertheless, temperature changes are readily distinguishable from pressure changes or, more importantly, mass changes: variations in surface mass cause very little change in attenuation, and the relationship between velocity shifts at various frequencies is well documented by data like those of Figure 5.

## CONCLUSIONS

SAW and APM devices have demonstrated their potential as sensitive detectors of changes in surface mass and other properties in a number of laboratories. While their exquisite mass sensitivity alone will undoubtedly result in a number of specialized applications, the ability to distinguish changes in mass from variations in temperature, pressure, conductivity, permittivity, and viscoelastic properties holds the potential for much more general chemical and physical sensors. Simultaneously measuring the orthogonal acoustic wave propagation factors, velocity and attenuation, at a number of frequencies endows a single MUFAW device with the characteristics of a multisensor using several device technologies.

## REFERENCES

1. S. J. Martin and A. J. Ricco, "Monitoring Photopolymerization of Thin Films Using SH Acoustic Plate Mode Sensors." *Sensors & Actuators, A (22)*, 712-18 (1990).
2. S. J. Martin and A. J. Ricco, "Effective Utilization of Acoustic Wave Sensor Responses: Simultaneous Measurement of Velocity and Attenuation," *Proc. 1989 IEEE Ultrasonics Symp.*, 1989, pp. 621-25.
3. A. J. Ricco, S. J. Martin, and T. E. Zipperian, "Surface Acoustic Wave Gas Sensor Based on Film Conductivity Changes", *Sensors and Actuators, 8*, 319-333 (1985).
4. T. M. Niemczyk, S. J. Martin, G. C. Frye, and A. J. Ricco, "Acoustoelectric Interaction of Plate Modes with Solutions", *J. Appl. Phys., 64*, 5002-8 (1988).
5. Liquid Phase Sensors Based on Acoustic Plate Mode Devices, A. J. Ricco, S. J. Martin, T. M. Niemczyk, and G. C. Frye, in *Chemical Sensors and Microinstrumentation*, ACS Symposium Series No. 403. Washington, DC: American Chemical Society, 1989, ch. 13.
6. S. J. Martin, A. J. Ricco, T. M. Niemczyk, and G. C. Frye, "Characterization of SH Acoustic Plate Mode Liquid Sensors," *Sensors & Actuators, 20*, 253-68 (1990).
7. A. J. Ricco and S. J. Martin, "Acoustic Wave Viscosity Sensor", *Appl. Phys. Lett., 50*, 1474-76, 1987.
8. D. P. Morgan, *Surface Wave Devices for Signal Processing*. New York: Elsevier, 1985, § 6.2.
9. B. A. Auld, *Acoustic Fields and Waves in Solids*. New York: Wiley, 1973.
10. R. M. White, P. J. Wicher, S W. Wenzel, and E. T. Zellers, "Plate-Mode Ultrasonic Oscillator Sensors", *IEEE Trans. on Ultrasonics, Ferroelectrics, and Freq. Contr., UFFC-34*, 162-171 (1987).

# PERFORMANCE CHARACTERISTICS OF SECOND GENERATION POLYSILICON RESONATING BEAM FORCE TRANSDUCERS

J. J. Sniegowski, H. Guckel, T. R. Christenson

Wisconsin Center for Applied Microelectronics
University of Wisconsin, Madison, WI 53706

## ABSTRACT

Doubly clamped beams in vacuum display a shift in resonant frequency with applied axial load. Functionality, miniaturization, and batch fabrication has been accomplished using surface micromaching techniques with tensile, fine grained polysilicon as the construction material. The device could be a very accurate, cost-effective alternative to presently available force transducers if it displays sharply defined and stable resonant frequencies.

The devices are driven into resonance via electro-thermo-mechanical or capacitive forces. Analytical expressions are derived to illustrate the detrimental effect the drive mechanisms have on the resonant frequency. Material and process sensitivities are calculated from a closed-form expression for the fundamental resonant frequency.

Experimental beams have typical dimensions of 200 micron length, 45 micron width, and 2 micron thickness. The fundamental unloaded resonant frequency is near 650 kHz and can be adjusted by processing. It will shift to nearly 900 kHz with an applied strain level of +0.1%. A temperature coefficient of the frequency of -75 ppm/°C for the finished batch-fabricated device was demonstrated. Theory and experiment corroborated and yields the conclusion that the resonant frequency stability and not the ability to measure frequency limits the force resolution.

## INTRODUCTION

First generation devices demonstrated the viability of miniaturized clamped-clamped beams inside a vacuum cavity as a force transducer. The basic configuration of first and second generation devices has the beam located within a cavity formed from an overlying cover of polysilicon attached to the underlying silicon substrate. The interior vacuum, necessary to produce high quality (high Q) resonances, is produced by reactive sealing. Surface adhesion induced by surface tension is avoided by using freeze-sublimation techniques. The construction material is tensile, fine-grained polysilicon. The resonance is excited via capacitive forces. The deflection is sensed piezoresistively or optically. Typical beam dimensions are 200 micron length, 45 micron width, and 2 micron thickness.

Second generation fabrication techniques include an improved freeze/sublimation process which utilizes cyclohexane, optimized drive and sense configurations, and modified sacrificial etch/clean cycles that increased functional device yield to nearly 100% on a three inch wafer. Details on the fine-grained polysilicon construction material and the resonator process sequence are given elsewhere [1]. Figure 1 shows a completed second generation device.

## EXCITATION ANALYSIS

Two excitation schemes are readily available in the polysilicon construction technique. One, is the electro-thermo-mechanical excitation based on the thermal expansion of the beam material [2]. The periodic resistive heating results in a periodic thermal gradient, which produces a driving mechanical moment. The other is electrostatic excitation by formation of a capacitor by implantation of one electrode into the beam and with the other electrode either in the substrate or cover polysilicon.
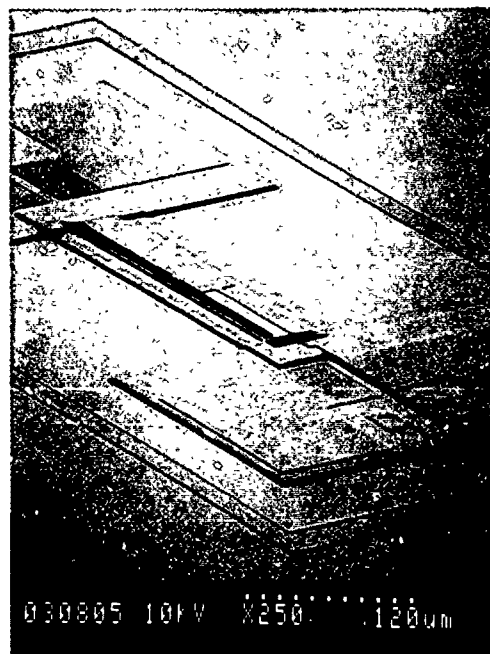


Figure 1. Vacuum encapsulated beam with dimensions. length=200μm, width = 40μm, and thickness=2μm. This beam illustrates a capacitive center drive. The metal electrode on the cover polysilicon with a lead running along the beam centerline out to a contact pad serves as the drive electrode. The bricklike patterns in the cover poly outside the beam edges indicate the location of the now sealed etch channels to the interior cavity.

### Electro-thermo-mechanical excitation

Since resistive heating is proportional to the square of the applied voltage, the frequency of the AC driver signal must either be at half the resonant frequency of the structure or by superimposing the AC signal on a DC bias, the AC signal can be at the resonant frequency. The detrimental effect of using a dissipative element in the beam structure can be seen from a single one-dimensional analysis of the thermal problem. Figure 2 illustrates the analysis.
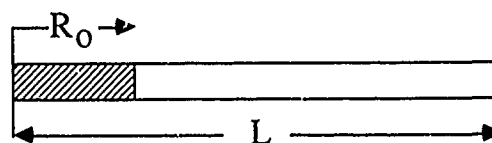


Figure 2. One dimensional analysis of heat transfer along a beam for the energy dissipated in a resistive element in the shaded area of the beam. $R_O$ is the extent of the resistor element onto the beam from one end.

The approximations are one dimensional analysis along the beam axis, the resistive element is a uniform heat source in the shaded region, heat conduction is through the beam ends only, and the substrate is a perfect heat sink, that is, the beam ends are at ambient temperature. By solving the heat equation, the thermal expansion of the beam due to a DC bias on the resistor element R is given by equation 1

$$\varepsilon = \frac{V_{DC}^2}{R} \frac{\alpha_T R_o^2}{Kbh} \left( 0.25 - \frac{R_o}{6L} \right) \qquad (1)$$

where,

| | |
|---|---|
| $V_{DC}$ is the applied voltage, | |
| R is the resistor value in ohms, | 7kΩ |
| $\alpha_T$ is the thermal exp. coef. of polysilicon | 4.0(10)$^{-6}$/°C |
| κ is the thermal conductivity | 1.412 W/cm K |
| b is the beam width | 45μm |
| h is the beam thickness | 2μm |
| $R_o$ is the length of the resistor | 40μm |
| L is the length of the beam | 400μm. |

This simple equation predicts resonant frequency shifts due to thermal heating to within a factor of 2 of that observed experimentally.

A second aspect of thermal excitation is the time response of the system. An estimate of the thermal time constant was obtained from a lumped element model of the beam depicted in figure 1. The resultant thermal time constant is given by equation 2

$$\tau^{-1} = \frac{D_{th}}{R_o (L - R_o)} \qquad (2)$$

where $D_{th}$=0.87 cm$^2$/sec is the thermal diffusivity of silicon. This indicates a time constant on the order of a tenth of a millisecond for a beam length of 200 micron. Since the resonant frequency of a 200 micron beam with built-in tensile strain is typically 650 kHz, the period of the resonant frequency is much less than the thermal time constant. Therefore the thermal drive becomes analogous to the AC ripple on a large capacitor of a simple AC to DC power supply. The net effects at the resonant frequency are a severely decreased drive efficiency and an effective DC thermal level. The consequential large drive signal required, capacitively feeds through to the sense electrode to only compound the problem. The large DC term significantly shifts the resonant frequency as cited earlier. Optical detection of the beam motion within the cavity verified the decrease in thermal drive efficiency at approximately 20 kHz. This corresponds to roughly 1/τ as expected. The calculations have recently been verified using the finite element analysis package ANSYS.

Capacitive excitation

The optimum design for the capacitive drive locates the beam electrode midway between the beam ends. This allows for efficient drive while effectively suppressing second and other even harmonics. The capacitive pressure, P, for this electrode configuration is

$$P = \frac{1}{2} \varepsilon \left( \frac{V_a}{G - w(x)} \right)^2 \qquad (3)$$

where,
 ε = permittivity of free space
 $V_a$ = applied voltage signal both AC and DC components
 G = equilibrium gap distance between beam and cover electrodes
 w(x) = displacement of the beam neutral axis from equilibrium.

Analogous to the thermal drive, the squaring to the input voltage results in the necessity of a DC bias if the resonant output frequency is to be fed back to the input. The resonator behavior can be adequately represented by a lumped mass and spring having the same resonant frequency [3]. Equation 4 is solved to illustrate the fundamental behavior of the system.

$$-kw - \eta \frac{dw}{dt} + \frac{1}{2} \varepsilon \frac{(V_{DC} + v_o \sin \omega t)^2}{G^2 (1 + \frac{w}{G})^2} bl = M \frac{d^2w}{dt^2} \qquad (4)$$

where, k = effective spring constant
 M = total beam mass
 η = damping coefficient
 b = beam width
 l = capacitor length on beam
 $v_o$ = magnitude of AC signal
 ω = drive frequency.

By suitable linearization of the forcing function, the following results are obtained,

$$w_{DC} = \frac{1}{2} \varepsilon \frac{V_{DC}^2}{G^2 \left[ k - \varepsilon \frac{bl}{G} \left( \frac{V_{DC}}{G} \right)^2 \right]} G \qquad (5)$$

$$w_{AC} = \frac{\alpha^2 \frac{v_o}{V_{DC}} G}{\left[ (\omega_R^2 - \omega^2)^2 + \left( \frac{\omega \omega_R}{Q} \right)^2 \right]^{\frac{1}{2}}} \sin(\omega t + \Theta) \qquad (6)$$

where, $\alpha^2 = \varepsilon \frac{bl}{G} \left( \frac{V_{DC}}{G} \right)^2 \frac{1}{M}$ $\qquad (7)$

$$\omega_R^2 = \frac{k}{M} - \alpha^2 = \text{resonant frequency} \qquad (8)$$

Q = quality factor of the resonator [4]

$$= \frac{\eta}{2M} \left( \frac{M}{k} \right)^{\frac{1}{2}}$$

Θ = phase angle.

The essential behaviors are the decrease in resonant frequency with applied bias and the fact that the amplitude becomes directly proportional to the quality factor Q near the resonant frequency. Experimental measurements are in agreement with this model. The model also illustrates the necessity of a high Q resonator. High Q devices require very small excitation signals thereby introducing small shifts in the resonant frequency. This eliminates the feedthrough from drive to sense electrodes.

SENSE SCHEME

The present sensing scheme is piezoresistive. This requires a constant voltage or current source for the measurement. From the thermal drive analysis, this sense scheme can obviously affect the resonance as well. The implication is that a drift in the measuring voltage over time would shift the effective reference resonant frequency $f_{Ro}$. This is detrimental to the accuracy of the device. The use of a one-port capacitive detection scheme or an optical detection scheme would eliminate this problem.

10

## RESONANT FREQUENCY - THEORY AND EXPERIMENT

The behavior of the free oscillation of a beam with applied axial load can be computed from the differential equation 9

$$EI \frac{\partial^4 w}{\partial x^4} + \rho A \frac{\partial^2 w}{\partial t^2} - P \frac{\partial^2 w}{\partial x^2} = 0 \qquad (9)$$

where $w(x)$ represents displacement of the neutral axis from equilibrium, $E=1.65(10)^{12}$ dynes/cm$^2$ is Young's modulus for the polysilicon beam of cross sectional area A and moment of inertia I. The mass density is $\rho$ which has a value of 2.33 g/cm$^3$. P is the total axial load in dynes. Equation 9 when combined with the boundary conditions for a clamped-clamped beam generates a transcendental equation from which the resonant frequencies can be determined. An alternative to this computation is the use of the Rayleigh-Ritz procedure. This technique when based on the deflection assumption

$$w(x) = Cx^2(x-L)^2 \quad \text{for } 0 \leq x \leq L \qquad (10)$$

yields,

$$\omega_r^2 = \frac{42Eh^2}{\rho L^4}\left(1 + \frac{2L^2}{7h^2}\varepsilon\right) \qquad (11)$$

where L is the beam length and h is its thickness. The total strain, $\varepsilon$, is simply the sum of the built-in strain and the applied strain, $\varepsilon_A$.

From the closed-form expression, various sensitivities can easily be calculated. Table 1 contains a summary of sensitivities covering the tensile range of built-in strain.

$$S_x^{\omega_R} = \frac{\partial \ln \omega_R}{\partial \ln x}$$

| X | $S_x^{\omega_R}$ ($\varepsilon_{bi} \to 0$) | $S_x^{\omega_R}$ ($\varepsilon_{bi} \to \infty$) |
|---|---|---|
| $\rho$ | $-\frac{1}{2}$ | $-\frac{1}{2}$ |
| E | $\frac{1}{2}$ | $\frac{1}{2}$ |
| h | 1 | 0 |
| L | $-2$ | $-1$ |

Table 1. Logarithmic sensitivities of $\omega_R$ are tabulated for limiting cases of the built-in strain. For near zero built-in strain the beam stiffness dominates. For large built-in strain the beam is dominated by the axial load and behaves like a stretched wire.

From equation 11, the dominant factor for temperature sensitivity stems from the temperature behavior of Young's modulus. The resonant frequency is expected to show approximately one half the temperature sensitivity of Young's modulus, that is, -37.3 ppm/°C. The following experimental data are representative of resonators with built-in tensile strain. Their dimensions are 200 micron length, 45 micron width, and 2.2 micron thickness. They display high Q fundamental modes near 650 kHz. Quality values as high as 35,000 have been obtained. The measured temperature behavior is

$$f_{ro} = AT + B \qquad (12)$$

with      A = 48.43 Hz/°C
and       B = 646,650 Hz.

This corresponds to a temperature coefficient of -75ppm/°C for the quantity $\Delta f_r/f_{ro}$. A significant contribution to this is from Young's modulus. Additional work is required to clarify the remaining behavior.

The strain sensitivity was measured by cleaving the substrate on which the devices are fabricated into the shape of a large beam. This beam is mounted as a cantilever in a micrometer driven deflection jig. The measured data has the form

$$\left(\frac{f_r}{f_{ro}}\right)^2 = 1 + a\,\varepsilon_A \qquad (13)$$

where $f_r$ is the resonant frequency which occurs when the strain $\varepsilon_A$ is applied via cantilever deflection. $f_{ro}$ is the unloaded resonant frequency. The applied strain, $\varepsilon_A$, includes both compressive and tensile excursions by applying both negative and positive cantilever deflections. Figure 3 illustrates the linear behavior of frequency squared to applied strain.

$$(f_r / f_{ro})^2 = 1.000 + 795.60\,\varepsilon$$

$$f_{ro} = 650.110 \text{ kHz}$$
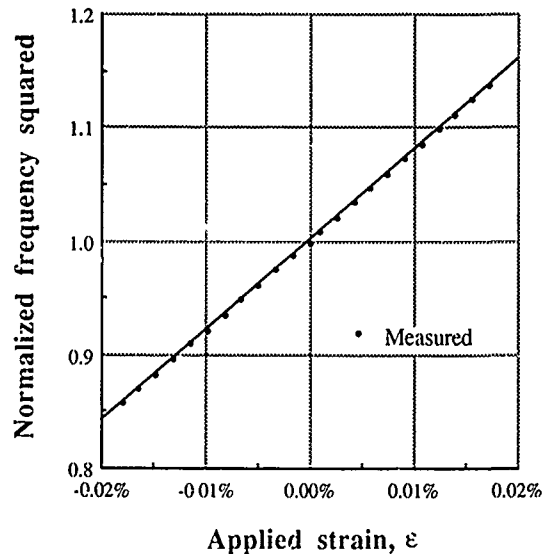


Figure 3. Normalized resonant frequency squared versus applied strain. The frequency is normalized to the resonant frequency without applied strain. The equation for the straight line fit is given at the top of the plot.

For a 200 micron beam, the application of an external strain of $8.82(10)^{-4}$ shifted the resonance from 644,950 cps, unloaded, to 853,530 cps. This applied strain is equal to an axial load of $1.4(10)^3$ dynes. An approximate force sensitivity is therefore near 150 cycles per dyne. A more accurate sensitivity calculation uses equation 13 with the experimentally determined coefficient a=795.6.

The accuracy and resolution of frequency measurements are far greater than those possible for piezoresistance measurement. Therefore, even the same relative change in frequency, $\Delta f_r/f_{ro}$, can be determined with many more figures of resolution than a relative change in resistance, $\Delta R/R_0$. This invalidates a comparison of "gauge factors" between resonant and piezoresistive strain sensors. The accuracy potential of resonant force sensors is inherently very high because of the basis for the measurement system, time. Whether this data acquisition sensitivity can be used depends on the stability of the resonator. One detrimental source of frequency drift is induced by temperature fluctuations. As an example, near $f_{ro}$ equation 13 indicates that $\Delta f_r \sim 174$ Hz/dyne. The temperature stability is 47 Hz/°C. If the temperature is held to within ±0.5°C, this implies a resolution of 0.27 dyne, or equivalently 275 micrograms of force. Clearly, temperature resolution or compensation is a far more dominant issue than frequency measurement limitations.

## CONCLUSIONS

The experimental device performance is very good, if not excellent. Agreement with theory is very sound. Frequency measurement does not present a limitation to force resolution. Temperature control or compensation is necessary to fully utilize the resolution capability of the resonant force sensor.

## ACKNOWLEDGEMENTS

We would like to thank Prof. E. Lovell and B. Choi for verification of the analytical theory using finite element programs.

## REFERENCES

[1]   a) H. Guckel, J. J. Sniegowski, T. R. Christenson, "Construction   and Performance Characteristics of Polysilicon Resonating Beam Force Transducers," in: Proceedings of the 3rd Toyota Conference: Integrated Micro Motion Systems ... Micromachining, Control and Application, Nissin, Aichi, Japan, Oct. 22-25, 1989, pp. (23-1) - (23-10).

b)   H. Guckel, J. J. Sniegowski, T. R. Christenson, F. Raissi, "The Application of Fine-Grained Tensile Polysilicon to Mechanically Resonant Transducers, in: Proceedings of the 5th International Conference on Solid-State Sensors and Actuators, in press, Montreaux, Switzerland, June 25-30, 1989.

c)   H. Guckel, J. J. Sniegowski, T. R. Christenson, "Advances in processing Techniques for Silicon Micromechanical Devices with Smooth Surfaces," Proc. of IEEE Workshop on Micro Electro Mechanical Systems, Salt Lake City, UT, Feb. 1989, pp. 71-75.

[2]   M. Elwenspoek, F. R. Blom, S. Bouwstra, T. S. J. Lammerink, F. C. M. van de Pol, H. A. C. Tilmans, Th. J. H. Popma, and J. H. J. Fluitman, "Transduction Mechanisms and Their Applications in Micromechanical Devices," Proc. of IEEE Workshop on Micro Electro Mechanical Systems, Salt Lake City, UT, Feb. 1989, pp. 126-132.

[3]  H. C. Nathanson, W. E. Newell, R. A. Wickstrom, and J. R. Davis, Jr., "The Resonant Gate Transistor," IEEE Trans. Electr. Devices, vol. ED-14, No. 3, March 1967, pp. 117-133.

[4]  W. T. Thompson, "Vibration Theory and Applications," Prentice-Hall, Inc., 1965, p. 74.

# THE EFFECT OF
## THERMOELASTIC INTERNAL FRICTION
## ON THE Q OF
## MICROMACHINED SILICON RESONATORS

Terry V. Roszhart
Research Scientist

Kearfott Guidance and Navigation Corporation
Little Falls, NJ

## ABSTRACT

Measurements of Q taken on chemically etched silicon beams show that thermoelastic internal friction can be a significant contributor to the damping found in miniature, solid-state resonators. Single crystal silicon beams were tested under vacuum conditions over a range of frequencies from 80 kHz to 1.6 MHz and over a range of temperatures from 300°K to 400°K. Measured values of Q varied from 10,000 to more than 70,000 and showed good agreement with theoretical estimates of thermoelastic internal friction. Data reported by other authors is also analyzed and shown to agree with theory at frequencies as low as 2.0 kHz. It is concluded that thermoelastic internal friction is a measurable phenomena in micromachined flexures and should be considered when designing, testing, and evaluating miniature resonators for solid-state sensors.

## INTRODUCTION

Thermoelastic internal friction is present in all structural materials and has been found experimentally in miniature silicon resonators. This form of damping is dependent on material properties such as thermal expansion, thermal conductivity, heat capacity, density, and elastic modulus. It is also related to mechanical design variables such as resonator dimensions and temperature. This paper describes a series of resonator damping measurements made on flexural beams micromachined from single crystal silicon. It also reviews a theoretical analysis of thermoelastic internal friction that accurately predicts the measurements that were taken. The results of these tests show that thermoelastic internal friction is a fundamental damping mechanism that can determine the quality of high Q resonators over a range of operating conditions.

These measurements were made because the Q of silicon resonators can be critical to the development of solid-state sensors. Miniature silicon resonators are often proposed as the basic sensing elements in micromachined, solid-state sensors. These sensors are frequency-modulated devices that exhibit a change in output frequency that is related to the measurement of a physical variable. The precision of this measurement is inherently dependent upon the frequency stability of the sensor's output. This, in turn, is dependent upon damping in the sensor's resonator. It is shown below that thermoelastic internal friction becomes a major part of resonator damping as the dimensions of a resonator become small. Since miniature structural dimensions are highly desired features in micromachined solid-state sensors, an understanding of damping caused by thermoelastic internal friction can be significant to the successful design and evaluation of these devices.

## THEORY

Thermoelastic internal friction was first analyzed by Zener (1) for thin metal reeds and, again by Zener (2) for a variety of polycrystalline metals. Measurements by Randall et. al. (3) showed that this analysis could accurately predict the functional relationship between internal friction and metallurgical grain size in polycrystalline brass. Thermoelastic internal friction can also explain why the attenuation of high-frequency, longitudinal, acoustic waves in elastic media is inversely proportional to wavelength (4).

Thermoelastic internal friction can occur in any material that is subjected to cyclic stress. It is most pronounced when the period of the cyclic stress is approximately equal to a structure's thermal time constant. The reason for this is that thermoelastic internal friction depends upon the rate at which mechanical energy is converted into nonrecoverable thermal energy. In particular, when a material is subjected to a mechanical deformation, most of the work needed to produce the deformation goes into elastic potential energy. This energy can always be recovered when the mechanical load causing the deformation is removed. However, some of the work of deformation also goes into thermal energy. It is well known that a material subjected to a compressive stress will increase its temperature. Conversely, a tensile stress causes a decrease in temperature. Since many types of deformations can create both tensile and compressive stresses in a structure, this produces adjacent regions of high and low temperatures. If the load that is causing the deformation is applied for a long enough period of time, thermal conductivity in the material will allow these hot and cold regions to equilibrate. When this happens, the energy that created the thermal gradients is lost to entrophy and cannot be recovered elastically. It is this lost thermal energy that causes the damping mechanism called thermoelastic internal friction.

For example, when an elastic beam undergoes bending, one of its surfaces experiences a tensile stress and the other a compressive stress. As a result, one side of the beam increases its temperature while the opposite side becomes cooler. As the force that produces the bending is applied to the beam, the beam's thermal conductivity will permit heat to flow from the warm side of the beam to the cool side. When the load is then removed, the energy lost to thermal conductivity is not recovered and the beam experiences thermoelastic internal friction. The magnitude of this damping in a flexural beam is given by the following equations (ref. 2):

$$\delta = \Gamma(T)\,\Omega(F) = \frac{1}{2Q} \qquad \text{eq. 1}$$

$$\Gamma(T) = \frac{\alpha^2\;T\;E}{4\rho\;Cp} \qquad \text{eq. 2}$$

$$\Omega(F) = 2\left[\frac{Fo\;F}{Fo^2 + F^2}\right] \qquad \text{eq. 3}$$

WHERE

| | | |
|---|---|---|
| $\delta$ | = | THERMOELASTIC CRITICAL DAMPING FRACTION |
| $\alpha$ | = | THERMAL EXPANSION |
| T | = | BEAM TEMPERATURE |
| E | = | ELASTIC MODULUS |
| $\rho$ | = | MATERIAL DENSITY |
| Cp | = | HEAT CAPACITY AT CONSTANT PRESSURE |
| K | = | THERMAL CONDUCTIVITY |
| F | = | BEAM FREQUENCY |
| t | = | BEAM THICKNESS |

AND

$$Fo = \frac{\pi\;K}{2\,\rho\;Cp\;t^2} = \text{CHARACTERISTIC DAMPING FREQUENCY} \qquad \text{eq. 4}$$

Thermoelastic internal friction has been expressed in the form of the critical damping ratio $\delta$ (5). The reason for this is that the total damping of a resonator can be found by adding the thermoelastic damping ratio to all other damping mechanisms such as gas damping and mounting losses. Q is then found by inverting the sum of the damping ratios in accordance with the right-hand side of Equation 1.

Equation 1 shows that thermoelastic internal friction can be expressed as the product of two different functions, $\Gamma(T)$ and $\Omega(F)$. $\Gamma(T)$ (Equation 2) calculates the peak magnitude of damping that is determined by material properties and temperature. The relevant material properties for single crystal silicon and quartz are listed in Table 1 along with the value of $\Gamma(T)$ at 300°K. It can be seen that the thermoelastic internal friction of quartz is more than an order of magnitude higher than silicon. This is primarily due to silicon's low thermal expansion, which appears as a squared term in Equation 2. (It should be noted that a rigorous calculation of thermoelastic internal friction should take into account the dependence of thermal expansion and elastic modulus on crystallographic directions. This is beyond the scope of this paper and the data shown in Table 1 should be used for general comparison purposes only.)

## TABLE 1. MATERIAL PROPERTIES

| Property | Silicon | Quartz | Units |
|---|---|---|---|
| Thermal expansion | 2.60 | 13.70 | ppm/°K |
| Elastic modulus | 1.70 | 0.78 | $10^{12}$ dyne/cm$^2$ |
| Material density | 2.33 | 2.60 | g/cm$^3$ |
| Heat capacity | 0.70 | 0.75 | J/g/°K |
| Thermal conductivity | 1.50 | 0.10 | $10^7$ dyne/°K/s |
| Peak damping @ 300°k | 1.06 | 11.34 | $10^{-4}$ |

The second function, $\Omega(F)$, calculates the frequency response of the damping in relationship to a characteristic damping frequency $F_0$ (Equation 4). The period of $F_0$ is proportional to the square of the beam's thickness and diffusivity, and is equal to the time during which heat from the warm side of the beam can flow to the cold side. The function, $\Omega(F)$, is symmetrical in frequency about $F_0$. When the frequency of oscillation F equals $F_0$, $\Omega(F)$ equals unity and damping is at a maximum. When F is significantly larger or smaller than $F_0$ the function $\Omega(F)$ is less than unity and therefore attenuates the magnitude of thermoelastic internal friction. As a result, a resonator's operating frequency should be designed away from its characteristic damping frequency $F_0$ to minimize damping and maximize resonant Q. The product of both $\Gamma(T)$ and $\Omega(F)$ is shown in Figure 1 for both materials over a range of frequencies.
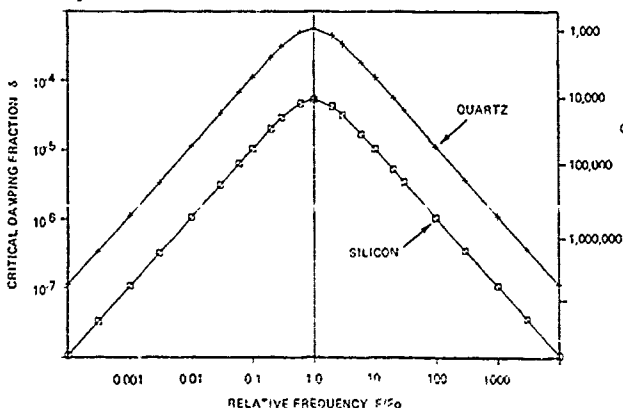


Fig. 1 The magnitude of thermoelastic internal friction versus frequency for single crystal silicon and quartz.

Although thermoelastic internal friction occurs in quartz it is not usually a major consideration in the design and performance of conventional quartz resonators. There are several reasons for this. Many quartz resonators operate in modes of vibration that are limited to shear deformations. This type of deformation does not exhibit a change in volume as a function of strain and, therefore, does not produce thermoelastic damping. Other quartz resonators operating in other modes of vibration have been limited by conventional fabrication techniques to mechanical dimensions that are relatively large in size. According to Equation 4, this makes the characteristic damping frequency $F_0$ small in relationship to the resonator's natural frequency. This, in turn, leads to values of thermoelastic internal friction that are small in comparison to other forms of damping.

However, as the size of a resonator becomes smaller, the effects of thermoelastic internal friction become larger. The advent of micromachining in materials like quartz and silicon is making possible the fabrication of mechanical resonators that have physical dimensions much smaller than could be achieved with conventional manufacturing methods. As the dimensions of micromachined resonators approach values usually used to describe metallurgical grain structure, the effects of thermoelastic internal friction become more important to high-quality resonator performance.

## TEST SETUP

All of the resonators tested below were fabricated from single crystal silicon using solid-state micromachining techniques. A Scanning Electron Micrograph of a silicon substrate supporting four of these beams is shown in Figure 2. Each resonator consists of a long flexural beam attached at one or two ends. The beam is driven by electrostatic electrodes that were plated onto a glass substrate spaced within 2 microns of the beam's surface.

The resonators were fabricated to the dimensions listed in Table 2. The first step in fabrication consisted of plasma etching the lithographic outline of each beam on one side of a four-inch, lightly doped (100) wafer. The depth of each beam was determined by the depth of the plasma etch. The backside of the beam was exposed by chemically etching the remainder of the substrate through a rectangular lithographic window on the opposite side of the wafer. The wafer was then anodically bonded to a glass electrode pair prior to diamond sawing. All beams were aligned parallel to one of the crystal's (111) axes. The resonators were evaluated in the test setup shown in Figure 3. The resonators were connected to a frequency synthesizer, a load impedance Z1, and a lock-in amplifier that was referenced to the synthesizer's output frequency. The output of the lock-in amplifier was proportional to current through the resonator and was plotted on an X-Y chart recorder as a function of frequency. Measurements of resonator frequency were made by locating the peak of the resonance curve on the frequency axis and measurements of Q were made from the half-power frequency bandwidth.



Fig. 2 Electrostatic cantilever beam fabricated from chemically etched, single crystal silicon.

14

Fig. 3 Vacuum test setup for resonator testing.

## TABLE 2. RESONATOR PARAMETERS

| Resonator | Length (microns) | Width (microns) | Thickness (microns) | $F_0$ (kHz) |
|-----------|------------------|-----------------|---------------------|-------------|
| SC4-10(B) | 395 | 54 | 10 | 1450 |
| SC2-10(A) | 192 | 54 | 10 | 1450 |
| SC1-10(B) | 190 | 27 | 10 | 1450 |
| SC1-10(A) | 192 | 27 | 10 | 1450 |
| SC1-20(A) | 191 | 27 | 15 | 644 |
| SC1-20(B) | 183 | 27 | 15 | 644 |
| SC2-15(A) | 190 | 54 | 17.50 | 473 |
| SC1-15(A) | 204 | 27 | 17.50 | 473 |
| SC1-15(B) | 182 | 27 | 17.50 | 473 |
| DC6-20 | 443 | 54 | 15 | 644 |

The resonators were placed in a vacuum chamber to eliminate the effects of gas damping. This effect usually disappears at about $10^{-4}$ Torr and all tests were made in the $10^{-6}$ Torr range. In general, it was found that the application of low drive levels produced the largest values of Q. At large amplitudes, the resonator begins to develop significant nonlinearities that produce excessive losses and an unsymmetrical resonance peak. All measurements were made at drive levels low enough to eliminate this form of energy loss but high enough to produce a good signal-to-noise ratio at the output. As a result, energy losses due to the measurement method were minimized and the values of Q measured with this setup were determined primarily by resonator damping.

## TABLE 3. RESONATOR TEST DATA

| Resonator | Frequency F (kHz) | Quality Q | Critical Damping Fraction $\delta$ |
|-----------|-------------------|-----------|------------------------------------|
| SC4-10(B) | 86.40 | 70,200 | $7.12 \times 10^{-6}$ |
| SC2-10(A) | 323.50 | 17,500 | $2.86 \times 10^{-5}$ |
| SC1-10(B) | 348.40 | 14,200 | $3.52 \times 10^{-5}$ |
| SC1-10(A) | 315.20 | 11,600 | $4.31 \times 10^{-5}$ |
| SC1-20(A) | 561.00 | 11,400 | $4.39 \times 10^{-5}$ |
| SC1-20(B) | 589.50 | 10,083 | $4.96 \times 10^{-5}$ |
| SC2-15(A) | 612.00 | 9,710 | $5.15 \times 10^{-5}$ |
| SC1-15(A) | 483.40 | 10,100 | $4.95 \times 10^{-5}$ |
| SC1-15(B) | 645.00 | 10,800 | $4.69 \times 10^{-5}$ |
| DC6-20 | 605.00 | 9,200 | $3.43 \times 10^{-5}$ |
| * | 1,620.00 | 12,500 | $4.00 \times 10^{-5}$ |
| ** | 2.47 | 35,000 | $1.13 \times 10^{-6}$ |

* Resonator DC6-20 operated in second vibration mode
** Data from Buser et.al. (ref 6)

The measurements of frequency and Q are listed in Table 3 and plotted in Figures 4 and 5. A total of 10 resonators were tested with frequencies ranging from 86 kHz to 1.6 MHz. Most of these resonators were single cantilever beams, but one resonator (DC6-20) was a double cantilever beam that was driven in the first and second modes of vibration. The values of Q ranged from 70,000 for the lowest frequency resonator to a minimum of about 10,000 for the resonators operating near their critical damping frequency.
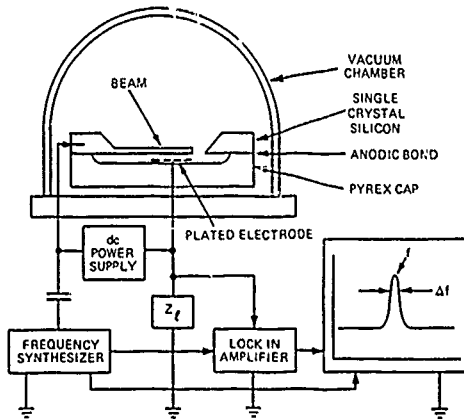
Figure 4 is a plot of the critical damping fraction $\delta$ as a function of frequency. The data shows good agreement with theory. Both the magnitude and shape of the frequency response agrees with calculations based on Equations 1 to 4. When $F/F_0$ is close to unity the damping is high (and the Q is low). The damping then decreases when $F/F_0$ is greater or smaller than unity. This supports the earlier observation that thermoelastic internal friction can be reduced by designing the resonator to operate at a frequency significantly different than the critical damping frequency $F_0$.
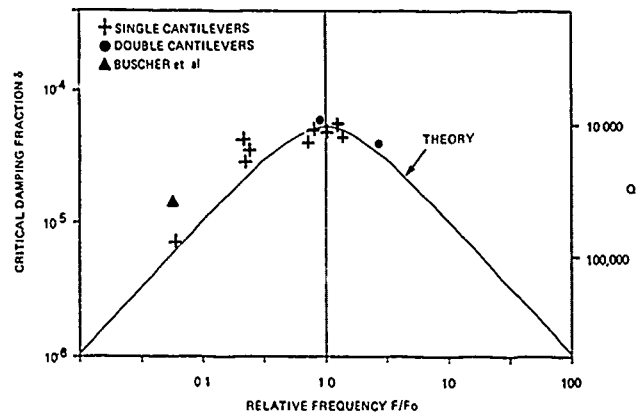


Fig. 4 Measured values of damping versus frequency (compared to theory).

Figure 4 also shows that thermoelastic internal friction in bending beams is relatively insensitive to the mode of flexural vibration. Both single-cantilever and double-cantilever beams exhibited the same level of damping when operated in the first (or fundamental) mode of vibration. When the double cantilever beam was operated in its next highest mode, its resonant Q increased by an amount that agrees with the corresponding increase in $F/F_0$. This shows that thermoelastic internal friction is more dependent on thermal conductivity through the thickness of the beam than it is on heat flow along the beam's length.

Comparisons of data can also be made with the results reported by Buser et.al. (6) on micromachined resonators of similar design. These resonators had thicknesses of 380 microns with resonant frequencies of 2.0 kHz to 2.5 kHz and Q's of 35,000. Plotting this data on Figure 4 shows that the damping in these resonators was a little higher than that predicted by thermoelastic internal friction alone. However, these resonators are larger than the resonators tested in this paper and the loss of energy to the surrounding substrate is likely to be larger. This is consistent with Buser's statement that the Q of his resonator was sensitive to the clamping pressure of his mounting fixture.

As a further check on theory, measurements of Q were made on one resonator over a range of temperatures from 295°K to 400°K. The results are shown in Figure 5 along with theoretical calculations. Equation 2 shows that damping is proportional to the temperature T. The experimental data plotted in Figure 5 shows this behavior. Buser also found a decrease in Q (and an increase in damping) as a function of increasing temperature.

15
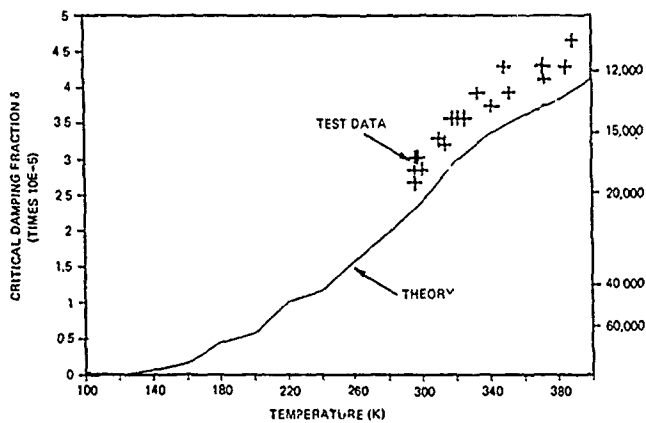
Fig. 5 Measured values of damping versus temperature (compared to theory).

It should be noted that the theoretical relationship in Figure 5 is not exactly linear since the properties of thermal expansion, modulus, and heat capacity have temperature dependances that are measureable. The dependance of these properties on temperature was included in the calculations plotted in this figure. In addition, if the calculations are continued to low temperatures, thermoelastic internal friction can be virtually eliminated. This occurs for two reasons. The first is that $\Gamma(T)$ is explicitly proportional to T. At $0^\circ K$ thermoelastic internal friction should vanish. However like other materials with a diamond like lattice structure, silicon exhibits zero thermal expansion at a temperature greater than $0^\circ K$. As a result, Figure 5 shows that $\Gamma(T)$ becomes small at temperatures near $120^\circ K$. This suggests the possibility of eliminating thermoelastic friction in silicon resonators operated at $LN_2$ temperatures.

## CONCLUSIONS

The results presented above show that thermoelastic internal friction can be a significant damping mechanism in miniature silicon beams:

- Analysis shows that thermoelastic internal friction can be calculated prior to resonator testing if thermal expansion, thermal conductivity, heat capacity, density, elastic modulus, beam thickness, and beam temperature are known.

- Measurements of Q taken on micromachined, single crystal, silicon beams with resonances from 2 kHz to 1.2 MHz show good agreement with the magnitude and frequency dependance predicted by analytical methods.

- Measurements of Q taken over temperatures from $295^\circ K$ to $400^\circ K$ show good agreement with analysis.

- Analysis also shows that thermoelastic internal friction can be reduced or eliminated by proper choice of resonant frequencies or by resonator cooling.

These factors should be considered for any solid-state sensor that incorporates micromachined solid-state resonators.

## REFERENCES

1) Zener, "Internal Friction in Solids (Theory of Internal Friction in Reeds)," Physical Review 52, Aug. 1, 1973; p 230.

2) Zener, "Internal Friction in Solids (II General Theory of Thermoelastic Internal Friction)," Physical Review 53, Jan. 1, 1938; p 90.

3) Randall, Rose, and Zener; "Intercrystalline Thermal Currents as a Source of Internal Friction," Physical Review 56, Aug. 15, 1939; p 343.

4) Mason, Physical Acoustics and the Properties of Solids, D. Van Nostrand Company, Inc.; 1958.

5) Harris and Crede; Shock and Vibration Handbook, McGraw-Hill Book Company, 1976.

6) Buser and De Rooij, "Resonant Silicon Structures," Sensors and Actuators, 17, (1989), p 145-154.

# FRICTION AND WEAR IN MICROFABRICATED HARMONIC SIDE-DRIVE MOTORS

Mehran Mehregany, Stephen D. Senturia, and Jeffrey H. Lang

*Microsystems Technology Laboratories*
*Laboratory for Electromagnetic and Electronic Systems*
*Department of Electrical Engineering and Computer Science*
*Massachusetts Institute of Technology*
*Cambridge, Ma 02139*

## ABSTRACT

This paper describes polysilicon variable-capacitance rotary harmonic side-drive micromotors, presents results from operational and frictional studies of these motors, and for the first time, reports *in situ* quantitative studies of wear under electric excitation. Voltages as low as 26V across 1.5 $\mu$m gaps are sufficient for operating these motors. Frictional force estimates at the bushings, 0.15 $\mu$N, and in the bearing, 0.04 $\mu$N, are obtained from measurements of stopping voltages. Extended operation of these motors to near 100 million wobble cycles at excitation frequencies of 10,000 rpm and 25,000 rpm, for operational durations of 150 hours and 71 hours, respectively, are studied. The results indicate that bearing wear is significant and results in changes in the gear ratio of the motors by as much as 20%. Typical gear ratios are near 90 at the start of motor operation and decrease to near 70 as the bearings wear out.

## 1. INTRODUCTION

Operation of variable-capacitance ordinary side-drive micromotors was first reported by Fan, Tai, and Muller [1,2]. Operation of similar micromotors using an air-levitation assist was reported by us shortly thereafter in [3]. However, in [3], failure in motor operation persisted without the air-levitation assist. At MEMS 1990 [4], we identified the native oxide of polysilicon as the cause of this motor operation failure, asserted that micromotor release and testing techniques are the determining factors for successful electrical operation of the motors, reported the operation of ordinary side-drive motors without air levitation, and presented results from frictional studies of the ordinary side-drive motors.

Additionally, at MEMS 1990 [4], we reported the development and demonstration of novel rotary variable-capacitance harmonic side-drive (hereafter referred to as wobble) micromotors. This paper further describes the wobble micromotors reported in [4], presents results from operational and frictional studies of these motors, and for the first time, reports *in situ* quantitative studies of wear under electric excitation.

## 2. MICROFABRICATED WOBBLE MOTORS

Figure 1 is a SEM photograph of a typical released wobble micromotor studied in this paper. Figure 2 is a schematic drawing of the cross-sectional view of the motor prior to release, which is when the low-temperature oxide (LTO) layers are dissolved in HF. Motor structural components are fabricated from heavily-phosphorus-doped polysilicon. After release, the rotor is supported on the bushings and is free to rotate about a center-pin bearing contacting an electric shield under the rotor. During motor operation, the rotor is intended to be in electrical contact with the shield positioned beneath it through mechanical contact at the bearing or at the bushing supports. As reported in [4], native oxide formation on the polysilicon surfaces can disrupt the intended electrical contacts and lead to the clamping of the rotor to the shield beneath it via electrostatic attraction, preventing motor operation. Proper release and testing techniques aimed at minimizing the native oxide formation are required for successful motor operation, as discussed in detail in [4], and are briefly review in Section 3.

The central feature of the design presented here is that the rotor wobbles around the center bearing post rather than the outer stator for conventional wobble motors [5-8]. A schematic
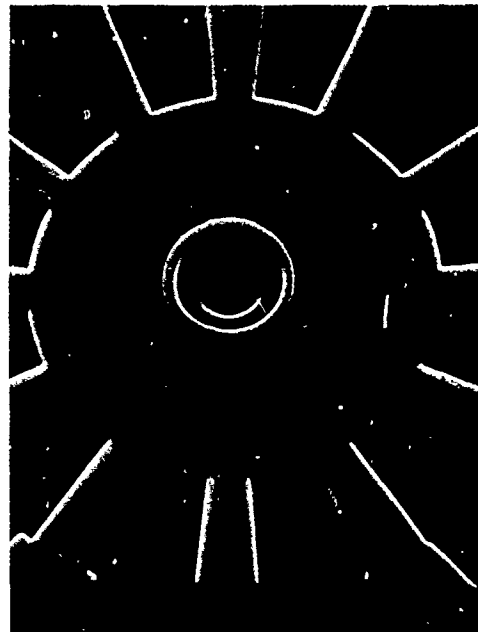


Fig. 1. A 12 stator-pole, 2.5 $\mu$m-gap, 100 $\mu$m-diameter, wobble micromotor with bushing 2 (i.e., the three small rotor indentations near the bearing).
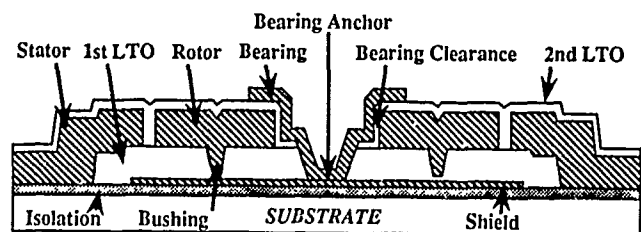


Fig. 2. Motor cross-section after fabrication completion and prior to release.

drawing showing the plan view of the design used here is shown in Fig. 3. In this design, the rotor lateral movement (i.e., wobble distance) is equal to the clearance in the bearing which is the difference between the bearing and inner rotor radii. This clearance, which we estimate to be near 0.3 $\mu$m, is specified by the thickness of the second sacrificial oxide. Furthermore, by ensuring a bearing clearance smaller than the nominal air-gap size, which is greater than 1.5 $\mu$m in our motors, the need for insulation between the rotor and the stator is eliminated. In this design, the rotor rotates in the same direction as the excitation signal and the point of contact.

A significant advantage of the wobble motor is that its drive torque is proportional to the motor gear ratio, $n$. The gear ratio relates the angle of rotation of the stator excitation to the angle of rotation of the rotor. For the wobble motor design presented here, $n$ is given by

$$n = \frac{r_b}{\delta}.\tag{1}$$

where $r_b$ is the radius of the bearing and $\delta$ is the bearing clearance. Let $\theta$ designate the angle between the diameter containing the contact point and the energized electrode, then the wobble motor drive torque, $T_w(\theta)$, due to a single excited stator pole at angular position $\theta$, estimated from in-plane two-dimensional simulations when the rotor is electrically grounded, is given by

$$T_w(\theta) = \frac{\epsilon_o t V^2}{2} \, \tau_w(\theta) \, n, \qquad (2)$$

where $\epsilon_o$ is the permitivity of air, $t$ is the rotor thickness, and $V$ is the applied excitation [8]. In (2), $\tau_w(\theta)$ is a normalized torque calculated by the simulations and specified by the motor design which includes effective air-gap size, pole width, and pole pitch [8]. The effective air-gap size is the actual distance of the rotor to the excited stator pole and is a function of the nominal air-gap size and the bearing clearance Note that the wobble motor drive torque is multiplied by the gear ratio, $n$ The torque obtained from multiply-excited stators can be derived by superposition.
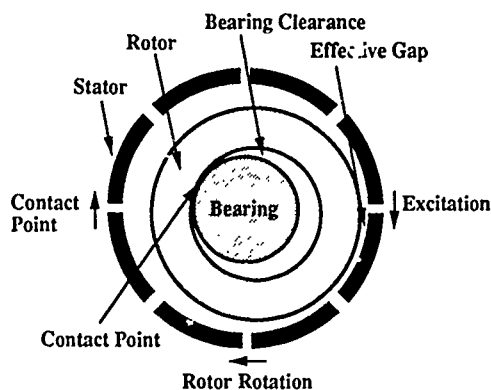


**Fig. 3.** The wobble micromotor using a center-pin bearing to eliminate the need for stator insulation.

A small bearing clearance in the wobble motor design used here is desirable because it reduces slip during motor operation, reduces rotor wobble, and reduces the risk of breakdown in the air gap. Recognizing that smaller bearing clearances decrease drive torque by increasing the effective air-gap size (see Fig. 3), we have increased the bearing radius to enhance the drive torque through further increase of the gear ratio.

## 3. FABRICATION AND TESTING

The wobble motor presented here, using a small bearing clearance and eliminating the need for rotor-stator insulation, simplifies fabrication and allows for a single fabrication process suitable for both wobble and ordinary side-drive motors. Details of this fabrication process, the polysilicon native oxide issue, as well as proper release and testing techniques are documented in detail elsewhere [1,9] and summarized here.

The motor fabrication process is discussed here in reference to the cross-sectional schematic of Fig. 2. Initially, substrate isolation is established using a 1 $\mu$m-thick LPCVD silicon-rich nitride layer over a 1 $\mu$m-thick thermally grown $SiO_2$ film [i 3]. A thin (3500Å) LPCVD polysilicon film is deposited, heavily doped with phosphorus, and patterned to form the shield. Stator anchors and bushing molds are patterned in the first sacrificial LTO layer which is 2.3 $\mu$m thick. The rotor, stator, and air gaps are patterned into a 2.5 $\mu$m-thick, heavily phosphorus-doped, polysilicon layer using RIE. The final rotor-stator polysilicon thickness is 2.2 $\mu$m since a patterned thermally grown oxide mask (not shown in Fig. 2) is used for the RIE etch of the polysilicon. Note that the inside radius of the rotor patterned at this point would correspond to the bearing radius in the final device. A bearing radius of 18 $\mu$m is used

for the wobble motors. The second sacrificial LTO layer is deposited, providing an estimated 0 3 $\mu$m coverage on the rotor inside radius side-walls, and patterned to open the bearing anchor. A 1 $\mu$m-thick LPCVD polysilicon film is deposited, heavily doped with phosphorus, and patterned to form the bearing The motor is released by dissolving the sacrificial LTO in HF During motor operation, the bushings are intended to provide electrical contact between the rotor and the shield as they slide over the shield.

The shield which is positioned under the rotor was suggested in [2] to be effective in the elimination of the rotor clamping forces by electrostatic shielding of the rotor from the substrate. However, even with electric shields incorporated in our devices, air levitation was found to be necessary in our previous work [3] to overcome frictional forces associated with the electrostatic clamping of the rotor to the shield beneath it. The rotor clamping was attributed to a lack of proper electrical contact between the rotor, the shield, and the bearing. This electrical contact is necessary to ensure that the rotor and the shield are at the same electric potential, thereby eliminating the clamping forces otherwise caused by the electric field between them. However, we believe that native oxide formation on the bushings, the bearing, and the shield surfaces results in a loss of contact, leading to an electric field between the rotor and the shield. This results in a clamping of the rotor to the shield (which was compensated for by air levitation in our previous work [3]). The reader is referred to [4] for a detailed discussion of these issues, including supporting studies, and to [10] for a possible model for calculating the electrostatic rotor clamping forces. With proper release and testing techniques which are summarized below and are directed at minimizing the formation of this oxide, we can readily operate the rotors without air levitation for extended periods of time.

Our release method consists of 15 minutes in a 49% (by weight) commercially-available HF solution, 2 minutes in a DI rinse, 5 minutes in a 1:1 $H_2SO_4$:$H_2O_2$ (pirana) clean, 10 minutes in a DI rinse, 30 minutes in the HF solution, 90 seconds DI rinse, and a nitrogen dry. The initial 15 minute oxide etch in HF is required to release the various motors on the die. The 5 minute pirana clean is effective in cleaning the exposed polysilicon surfaces from organic as well as ionic contaminants. The second HF etch removes the thin oxide formed on the polysilicon surfaces during the pirana clean. The final DI rinse time is critical. Rinse times under 60 seconds do not produce reliable results. Long rinse times may be detrimental since the native oxide growth rate is greater in DI water than in air. A 90 second rinse time has been typically used in this work with good results.

For all experimental studies reported here, the above release process is used. Immediately after release, the motors are stored and tested in nitrogen. In nitrogen, the rotor clamping failure mechanism is eliminated. In this case, the operation of wobble motors is repeatable (see measurement results in the next section) and has been studied for extended times up to six days (see Section 5).

## 4. OPERATIONAL AND FRICTIONAL STUDIES

Design parameter permutations including bushing style and air-gap size are incorporated into the fabrication of wobble motors. Electrical actuation for all of these design permutations has been confirmed. The wobble motor measurements reported in this section were carried out on 12 stator-pole, 100 $\mu$m-diameter motors (similar to that in Fig. 1) with 1.5 $\mu$m and 2.5 $\mu$m gaps. For these motors, the stator pole width and pitch are 27 and 30 degrees, respectively.

The wobble motors are studied using a six-phase, unipolar, square-wave excitation. Therefore, six stator poles are excited (independent of the total number of stator poles available) with a center-to-center angular separation of 60 degrees (e.g., every other stator pole for the motor in Fig. 1). This stator pole arrangement results from a limitation in our power supply which can excite only six independent phases. Two signal excitations were studied which differ in the number of the stator poles excited at one time. In the first case, excitation A,

18

the excitation followed the pattern '000001', '000011', '000010', '000110', and so on, where a '0' indicates a grounded stator pole and a '1' indicates an excited one. In the second case, excitation B, the excitation followed the pattern '000001', '000010', '000100', and so on. Unless otherwise specified, the rotor is electrically grounded through the shield. Note that excitation A approximately simulates the condition in which all twelve stator poles are excited with a signal excitation similar to that of excitation B. Excitation voltages as high as 150V across 1.5 μm-wide air gaps (i.e., electric field intensities of $1 \times 10^8$ V/m) are routinely used without electric field breakdown in the gaps. Using excitation B, the wobble motors have been operated at speeds estimated up to 700 rpm.

Measurement of starting and stopping voltages is performed on the wobble motors when operating in nitrogen. These measurements are performed by operating the motor at near 6 rpm independent of the excitation pattern. This requires that excitation A be switched twice as fast as excitation B, or approximately 10 ms and 20 ms, respectively. We measure the minimum voltage required for sustaining motor operation (stopping voltage) as well as the minimum voltage require to restart the motor (starting voltage). We have found that the starting and stopping voltages are nearly the same when the excitation is applied within a few seconds of previous rotor movement. Therefore, we insert a 30 second delay for the starting voltage measurements. At the slow operational speeds used for the stopping and starting voltage measurements, the rotor motion is step-wise since the settling time associated with the rotor transient response is much shorter than the switching times above. Therefore, the stopping and starting voltage measurements are only different in the length of time during which the rotor is at rest.

Freshly released motors which have not been previously operated are used in the measurements. The motors are typically operated for two minutes before taking measurements. The measured stopping voltages are reproducible to within 3% in the same testing session. The scatter becomes 10% when the same motor is tested in different sessions with a re-release prior to each session, or for devices on different dies. The starting voltage measurements are far more scattered. This scatter can be as high as 30% in the same session and from session to session, which indicates the complexity of static friction characteristics in the micromotors even after only 30 seconds at rest.

We have used two bushing designs in the wobble motors which are at an equal radius (lever arm), 29 μm, but have different apparent areas of contact. The apparent area of contact is an estimated 180 μm² for bushing 1 (which is of a continuous ring shape) and 6 μm² for bushing 2 (Fig. 1).

Table I summarizes the experimentally measured stopping and starting voltages as well as the corresponding estimated torques. All values listed in Table I are averages of the two bushing designs since no dependence was found on the bushing apparent area of contact, in agreement with the results reported in [4,11]. Each entry in Table I is an average computed from data on three motors (per bushing type) from different dies of the same wafer tested at different times. The wobble motors operate even when the rotor is not electrically grounded Therefore, it is not necessary to ground the shield (and therefore the rotor) as long as the rotor is in electrical contact with the shield. This is to be expected since as long as the shield and the rotor are in electrical contact, regardless of their electric potential, there is no electric field and hence no clamping force between them. However, as seen from lines 5 and 6 in Table I, higher voltages are required for initiating and sustaining motor operation. The reason is that without grounding, the rotor floats to a potential in the direction of the excited stator voltage so that a reduced voltage appears across the gap. Torque is thereby reduced for a given voltage. Since the rotor potential is not known in this case, corresponding torque values cannot be estimated from this type of measurement.

The measured voltages are used in conjunction with two-dimensional field simulations in the plane of the motors to estimate torque. Specifically, single phase drive-torque curves for

| ID | Gap (μm) | Excit. Pattern | Rotor Bias | Stopping | | Starting | |
|---|---|---|---|---|---|---|---|
| | | | | Voltage (V) | Torque (pN-m) | Voltage (V) | Torque (pN-r) |
| 1 | 1.5 | A | GND | 26 | 5 | 35 | 9 |
| 2 | 1.5 | B | GND | 33 | 6 | 52 | 14 |
| 3 | 2.5 | A | GND | 37 | 4 | 50 | 7 |
| 4 | 2.5 | B | GND | 49 | 5 | 70 | 10 |
| 5 | 2.5 | A | FLT | 61 | . | 81 | . |
| 6 | 2.5 | B | FLT | 80 | . | 105 | . |

Table I. Measured stopping and starting voltages as well as estimated torques for the wobble micromotors.

the motors studied here are calculated from the wobble motor analysis in [8]. A bearing clearance of 0.25 μm is assumed, accounting for the roughness of the rotor inside radius (see Section 5 for details). Since the air gap of the micromotors is comparable to the rotor thickness, the contribution of the fringing fields in the axial direction must be accounted for in interpreting the experimental results. A correction factor calculated from field simulations in the axial plane of the motor (see [9] for details) is used to approximately account for the contribution of the fringing fields in the axial direction. These correction factors are applied to the single-phase drive-torque curves since the system may be considered approximately invariant in the angular direction over a stator pole span. For the 1 5 μm air gap motors here, the corresponding correction factor is 1 95 while for the 2.5 μm air-gap motors this value is 2.27 [9].

Furthermore, in estimating torque values from the voltage measurements, an experimentally measured value for the gear ratio, $n$, is used rather than the value predicted by Eqn. (1). In the low excitation voltage range, corresponding to the stopping and starting voltage measurements, the rotor slips frequently, resulting in an apparent gear ratio higher than that predicted by Eqn. (1). The experimentally measured gear ratio in this voltage range is 85 for excitation A and 100 for Excitation B. Note that excitation B produces twice as much rotor angular rotation per excitation step than excitation A and therefore more slip. As a side point, it is worth mentioning that for motor speeds in the range of 5 rpm to 360 rpm, the measured gear ratio is constant.

In our torque analysis, the stopping rotor position has been assumed to be at the zero torque position with respect to the excited stator pole (or poles for excitation A). Therefore, the rotor starting position with respect to the next excited stator pole (or poles for excitation A) is known. In practice, it is difficult to experimentally determine the rotor position since the wobble distance (or rotor lateral motion) is very small. The assumed rotor position and the measured excitation voltage are used in conjunction with the field analysis described above to estimate the torques in Table I. Note that the wobble motor model used here predicts consistent torques from the stopping voltage measurements with variations in air-gap size and excitation pattern. In all recorded cases, the starting torques are higher than the stopping torques. Furthermore, the required starting torques are not as consistent as those of stopping torques, in agreement with results in [4]. The estimated averages for the stopping torque, 5±1 pN-m, and starting torque, 10±3 pN-m, are independent of the apparent bushing contact area. The average starting torque is twice that of the average stopping torque, in agreement with similar studies from the ordinary side-drive motors reported in [4].

In [4], we used stopping voltage measurements on ordinary side-drive motors to study frictional behavior of varying bushing styles and sizes. The estimated bushing frictional force in that study was 0.15 μN. For both ordinary and harmonic side-drive motors, the bushing operations produce sliding friction. Therefore, we have used a value of 0.15 μN for the bushing frictional force to estimate the bearing frictional force for the wobble motors. The lever arm of the bushings in the wobble motors is 29 μm while the bearing radius (or the lever arm of the bearing

frictional force) is 18 $\mu$m. Since the frictional torques at the bearing and the bushings should add up to 5 pN-m, which is the average stopping torque, the bearing frictional force can be estimated at 0.04 $\mu$N. Due to its larger lever arm, the bushing frictional force accounts for 87% of the total frictional torque of 5 pN-m. Note that the bearing frictional force estimate reported here is consistent with that obtained for the side-drive motors in [4]. In other words, as would be expected, the rolling action in wobble motor bearing produces less friction than that of the sliding action in the side-drive motor bearing.

## 5. WEAR STUDIES

Because the gear ratio of the wobble micromotor is equal to the bearing radius divided by the bearing clearance (see Eqn. (1)), under extended operation, changes in the gear ratio (reflected in changes in motor speed) can be a direct measure of wear in the bearing. We have used this concept to measure wear in the bearing of the wobble motors. Our experimental procedure is as follows. Freshly released motors are used and motor speed is measured periodically from frame by frame examination of a recorded video. A standard 8mm video system is used which exposes 30 frames/sec. The motor speed is measured by counting the number of frames required for the motor to make five turns. Note that the speed measurement accuracy increases by increasing the number of motor turns for which frames are counted.

The following definitions are needed before proceeding. Excitation frequency designates the number of cycles per unit time that the electrical signal travels on the stator and is related to the motor speed by the gear ratio. A wobble cycle is when the point of contact in the bearing has made one complete turn. Since the stator excitation and the point of contact rotate at the same rate, the number of wobble cycles can be calculated by multiplying the excitation frequency by the duration of operation. For a given excitation frequency, $\Omega$, for any two points in time during motor operation (designated by subscripts 1 and 2),

$$\Omega = n_1\omega_1 = n_2\omega_2, \qquad (3)$$

where $\omega$ is the motor speed. Therefore, the gear ratio as a function of wobble cycles can be calculated from the speed measured as a function of time. Note that as the bearing wears out, the motor speeds up which corresponds to a reduction in the gear ratio. In theory, the gear ratio is related to the bearing clearance by Eqn. (1). However, in practice, we have found that the gear ratio, $n$, is given by:

$$n = G(V) + \frac{r_b}{\delta}, \qquad (4)$$

where $G(V)$ is an excitation dependent offset in the gear ratio, accounting for rotor slip. For any two points in time during motor operation, the following expression can be derived:

$$\frac{\delta_2 - \delta_1}{\delta_1} = \frac{\delta_2}{r_b}(n_1 - n_2). \qquad (5)$$

An implicit assumption in deriving Eqn. (5) is that the slip term remains the same.

We have studied the extended operation of four identical wobble motors (in nitrogen) under three variations of excitation B involving differing voltage amplitude and frequency. A 90V excitation at 10,000 rpm frequency, corresponding to 10,000 wobble cycles per minutes, was used to test the first two motors. Figure 4 shows the data for the gear ratio as a function of wobble cycles for these two motors, Motor 1 and Motor 2. Motor 1 was continuously operated for nearly 30 million wobble cycles over 50 hours at which time it was dismantled for SEM inspection. Motor 2 was continuously operated for nearly 90 million wobble cycles over 150 hours at which time it was dismantled for SEM inspection. Figure 5 shows similar data for the gear ratio as a function of wobble cycles for Motor 3 and Motor 4. Both motors were continuously operated for nearly 100 million wobble cycles using an excitation frequency of 25,000 rpm, which

corresponds to a 2.5 times increase in the operational speed as compared to the tests of Fig. 4. A 90V excitation was used for Motor 3 and a 120V excitation for Motor 4.

Our measurements indicate three regions in the bearing wear characteristics of the wobble motors. The first region corresponds to the initial 0.3-0.5 million wobble cycles and is the burn-in period. At the start of the burn-in period, the gear ratio is typically near 90. As Figs. 4 and 5 indicate, except for Motor 2, the gear ratio increased (or the motor slowed down) in this region for the remaining three motors. It is likely that we missed detecting an increase in the gear ratio for Motor 2 by not taking enough data points in the initial stage. Furthermore, three of the four motors stopped once and the fourth (Motor 3) stopped several times in this region. After a stop, a momentary jet of nitrogen directed at the rotor was used to restart the motors. It is possible that the relatively larger number or size of the particles, produced during the initial (as compared to later) wear of the asperities on the rotor inner radius, is responsible for the observed behavior during the burn-in period. By filling in the bearing clearance or increasing slip in the bearing, these wear particles can lead to an apparent increase in the gear ratio and cease motor operation.

After the burn-in period, the motors continuously operate until the experiment is ended at a desired point. In the second region, which we call the break-in period, the general behavior for all four motors is similar, even though the detailed shape of the response varies. During the break-in period, the gear ratio decreases (i.e., the motor speeds up) significantly as the rotor inner radius asperities wear out, smoothing that surface and increasing the bearing clearance. By the end of the break-in period, the gear ratio is typically 70 to 76, reaching a plateau. This corresponds to nearly a 20% change in the gear ratio from the start of motor operation. In terms of motor speed, for example, Motor 1 operates at near 109 rpm at the start of operation and at near 137 by the end of the break-in period. Motor 3 operates at near 286 rpm at the start of operation and at near 357 rpm by the end of the break-in period. Note that the break-in period is under 30 million wobble cycles for three of the motors and up to 70 million wobble cycles for Motor 3.

After the break-in period, the gear ratio reaches a plateau at which point we enter the third region in the wear characteristics. Note that in this region, for Motor 2 and Motor 4, the gear ratio eventually increases (the motor speed decreases) with further operation. This gear ratio increase is attributed to an increase in rotor slip. This is corroborated by the fact that increasing the excitation voltage in this region of operation returns the gear ratio to the plateau value for the corresponding motor. We have not been able to identify the cause of rotor slip in this region of operation. It may be attributed to increased friction at the bushings or increased slip in the bearing. SEM inspection of the bushing surfaces after the wear tests have not given additional information. Note that Eqn. (5) is no longer accurate under these conditions since the slip term is now increasing significantly with motor operation.

As seen from Fig. 4, for two identical motors under identical excitations, the detailed wear characteristics during the break-in period may be significantly different. However, note that Motor 4 (Fig. 5) exhibits the most rapid wear during the break-in period as compared to the remaining three motors. This is to be expected, since the higher excitation voltage used in that test as compared to the remaining three, results in higher contact forces in the bearing. Furthermore, comparing Figs. 4 and 5 suggests that the higher operational speed results in increased wear.

As seen above, by the end of the break-in period, the overall gear ratio decreases (indicating a speed increase), corresponding to an increase of several hundred angstroms in the bearing clearance. The bearing clearance which is specified by the second sacrificial oxide thickness is estimated at 0.3 $\mu$m. However, the effective bearing clearance may be different due to the roughness of the RIE patterned rotor inside radius sidewalls (which in turn results in roughness of the bearing post walls). SEM inspection of unoperated motors reveal that the rotor inner radius roughness is as high as 600 Å. SEM inspection of the inner
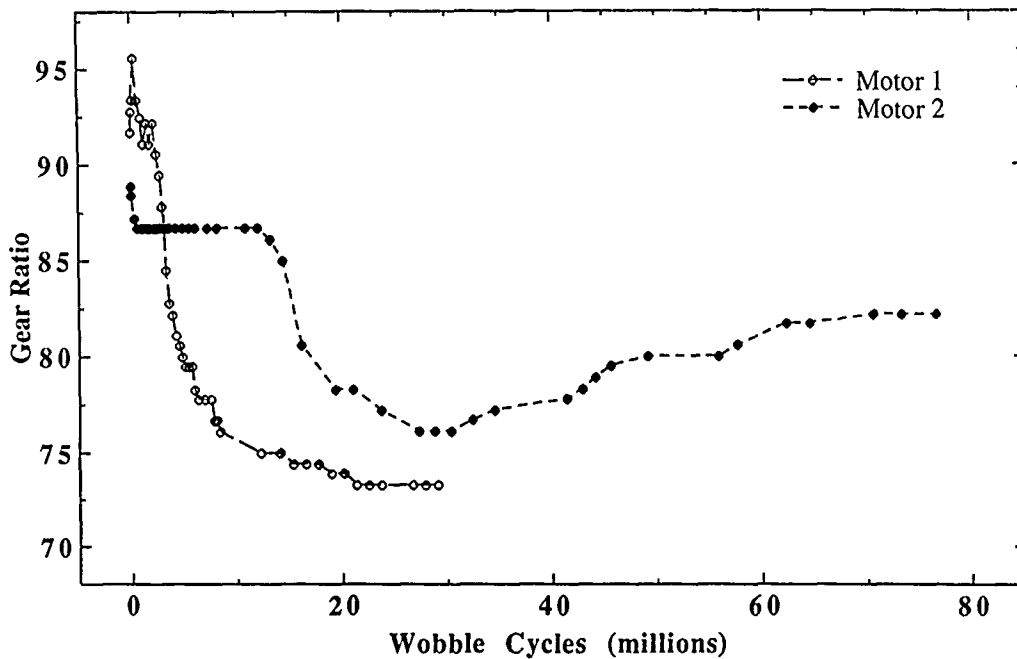
Fig. 4. Wear data for Motor 1 and Motor 2 operated at an excitation voltage of 90V and frequency of 10,000 rpm. Note that a decrease in gear ratio indicates an increase in speed.
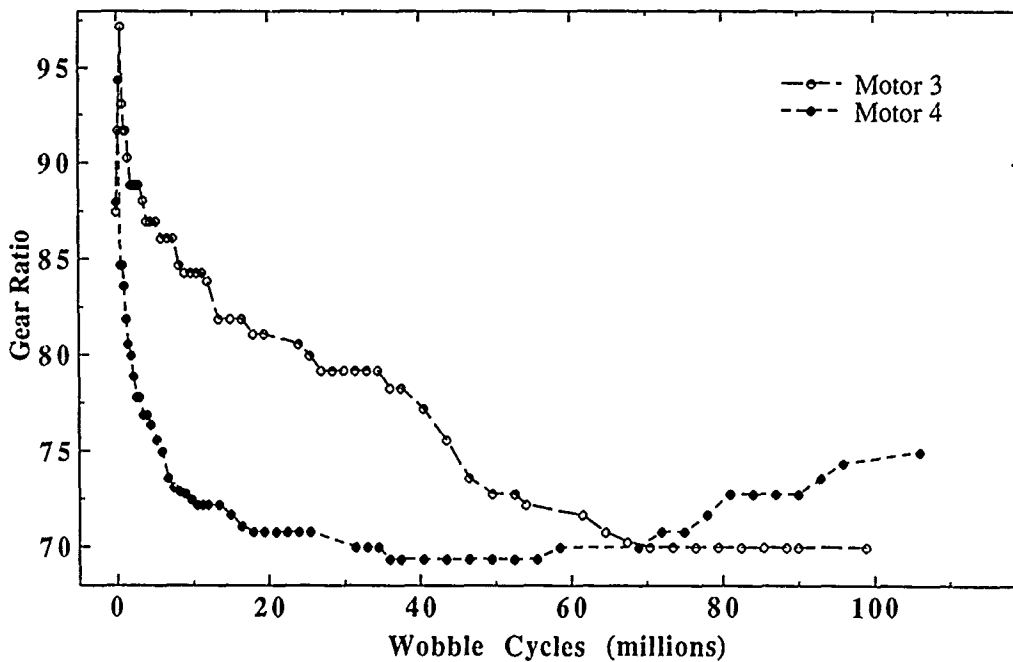


Fig. 5. Wear data for Motor 3 and Motor 4 operated at excitation voltages of 90V and 120V, respectively, and an excitation frequency of 25,000 rpm.

rotor radius of Motor 2 after 30 million wobble cycles shows the inner rotor radius to have been smoothed at least by a factor of three. The results are similar for the remaining three motors.

The wear data in Figs. 4 and 5 corroborate the SEM inspection results. The bearing clearance is very nearly 0.3 $\mu$m at the end of the break-in period, when the rotor inner radius asperities are worn out. Therefore, using Eqn. (5) in conjunction with the wear data for each motor, the bearing clearance at the start of motor operation can be calculated. For the motors tested here, the average bearing clearance at the start of motor operation is 2400±100 Å. The overall increase in the bearing clearance from the start of motor operation can be calculated (Eqn. (5)) at 500-700 Å for the different motors. On the average, these numbers correspond to a 25% increase in the bearing clearance by the end of the break-in period.

These results are further corroborated by wear studies of the ordinary side-drive motors. We have operated these motors at two rotor speeds of 160 and 2500 rpm for a total of 40,000 rotor revolutions. SEM inspection of the bushing surfaces and the rotor inner radius reveal scattered wear particles several hundred angstroms in size. Figure 6 is a SEM photo showing a magnified view of the bushing surface (which slides on the shield). Note the wear particles which are adhered to the bushing surface. We have not detected wear on the polysilicon shield which the bushings slide over during motor operation. Figure 7 is a SEM photo showing a portion of the rotor inner radius from the backside. Again, wear particles are observed adhering to the area adjacent to the bearing surface. Unoperated motors which have been dismantled for SEM inspection exhibit clean surfaces with no particles.

Fig. 6. SEM photo showing magnified view of the bushing surface of an ordinary side-drive motor after 40,000 cycles of operation at 2,500 rpm.



Fig. 7. SEM photo showing a portion of the rotor inner radius from the backside for an ordinary side-drive micromotor after 40,000 cycles of operation at 2,500 rpm.

## 6. CONCLUSION

A novel, polysilicon, rotary, variable-capacitance, wobble micromotor was described. Voltages as low as 26V across 1.5 $\mu$m gaps are sufficient for operating the wobble motors. Frictional force estimates at the bushings, 0.15 $\mu$N, and in the bearing, 0.04 $\mu$N, were obtained from measurements of stopping voltages. Extended operation of wobble motors up to and over 100 million wobble cycles was studied. Since the gear ratio of the wobble micromotor depends on the bearing clearance, under extended operation, changes in the gear ratio can be a direct measure of wear in the bearing. The results indicate that bearing wear is significant and can result in changes in the gear ratio by as much as 20%. Typical gear ratios are near 90 at the start of motor operation and decrease to near 70 as the bearing wears out.

## REFERENCES

[1] L.S. Fan, Y.C. Tai and R.S. Muller, Proc. of 1988 IEEE Int. Electr. Devices Meeting, San Francisco, CA, Dec. 1988, p. 666.

[2] Y.C. Tai, Ph.D. Thesis, Department of Electrical Engineering and Computer Science, University of California-Berkeley, June 1989.

[3] M. Mehregany, S.F. Bart, L.S. Tavrow, J.H. Lang, S.D. Senturia and M.F. Schlecht, Proc. of The 4th Int. Conf. on Solid State Sensors and Actuators (Transducers'89), Montreux, Switzerland, June 1989, p. 106. Complete paper to appear in Sensors & Actuators A21, 173 (1989).

[4] M. Mehregany, P. Nagarkar, S.D. Senturia and J.H. Lang, Proc. of The 3rd Micro Electro Mechanical Systems, Napa Valley, CA, Feb. 1990, p. 1.

[5] W.S.N. Trimmer and R. Jebens, Proc. of Micro Electro Mechanical Systems, Salt Lake City, Utah, Feb. 1989, p. 13.

[6] S.C. Jacobsen, R.H. Price, J.E. Wood, T.H. Rytting and M. Rafaelof, Proc. of Micro Electro Mechanical Systems, Salt Lake City, Utah, Feb. 1989, p. 17.

[7] H. Fujita and A. Omodaka, IEEE Trans. Electron Devices 35, 731 (1988).

[8] R. Mahadevan, Proc. of The 3rd IEEE Workshop on Micro Electro Mechanical Systems, Napa Valley, CA, Feb. 1990, p. 120.

[9] M. Mehregany, S.F. Bart, L.S. Tavrow, J.H. Lang and S.D. Senturia, 36th National Symposium of the American Vacuum Society, Boston, MA, Oct. 1989, p. 108. Complete paper to appear in J. Vac. Sci. Tech. A., July-August 1990.

[10] J.H. Lang, Proc. of The 3rd Toyota Conference, Aichi-ken, Japan, Oct. 1989, p. 9.1.

[11] M.G. Lim, J.C. Chang, D.P. Schultz, R.T. Howe and R.M. White, Proc. of The 3rd IEEE Workshop on Micro Electro Mechanical Systems, Napa Valley, CA, Feb. 1990, p. 82.

# ELECTROSTATICALLY BALANCED COMB DRIVE
# FOR CONTROLLED LEVITATION

*William C. Tang, Martin G. Lim,* and Roger T. Howe*

University of California at Berkeley
Department of Electrical Engineering and Computer Sciences
and the Electronics Research Laboratory
Berkeley Sensor & Actuator Center
Berkeley, California 94720

## Abstract

This paper is an experimental study of the levitating force (*normal* to the substrate) associated with interdigitated capacitor (electrostatic comb) lateral actuators. For compliant suspensions, normal displacements of over 2 μm for a comb bias of 30 V are observed. This phenomenon is due to electrostatic repulsion by image charges mirrored in the ground plane beneath the suspended structure. By electrically isolating alternating drive-comb fingers and applying voltages of equal magnitude and opposite sign, levitation can be reduced by an order of magnitude, while reducing the lateral drive force by less than a factor of two.

## INTRODUCTION

Surface-micromachined polysilicon resonators which are driven by interdigitated capacitors (electrostatic combs) have several attractive properties [1-3]. Vibrational amplitudes of over 10 μm are possible with relatively high quality-factors at atmospheric pressure, in contrast to structures which move normal to the surface of the substrate. The comb-drive capacitance is linear with displacement, resulting in an electrostatic drive force which is independent of vibrational amplitude. Electrostatic combs have recently been used for the static actuation of friction test-structures [4] and microgrippers [5].

Potential applications of lateral resonators include resonant accelerometers and rate gyroscopes, as well as resonant microactuators [6]. For efficient mechanical coupling between a vibrating pawl and a toothed wheel, it is essential that both structures remain co-planar. However, 2 μm-thick polysilicon resonators with compliant folded-beam suspensions have been observed to levitate over 2 μm when driven by an electrostatic comb biased with a DC voltage of 30 V. This effect must be understood in order to design functioning resonant microactuators, with the possibility that levitation by interdigitated combs may offer a convenient means for selective pawl engagement. In this paper, the electrostatic forces responsible for levitation are described, along with a modified comb design with independently biased fingers. For appropriate drive voltages, the levitation effect can be nearly eliminated. Experimental measurements of this effect are reported for a variety of comb structures.

## VERTICAL LEVITATION AND CONTROL METHODS

### Vertical Levitation

Successful electrostatic actuation of micromechanical structures requires a ground plane under the structure in order to shield it from relatively large vertical fields [7,8]. In previous studies of the electrostatic-comb drive, a heavily doped polysilicon film underlies the resonator and the comb structure. However, this ground plane contributes to an unbalanced electrostatic field distribution, as shown in Fig. 1 [9]. The imbalance in the field distribution results in a net vertical force induced on the movable comb finger. The positively biased drive comb fingers induce negative charge on both the ground plane and the movable comb fingers. These like charges yield a vertical force which repels or levitates the structure away from the substrate. The net vertical force, $F_z$, can be evaluated using energy methods:

$$E = q\Phi \qquad (1)$$

where E is the stored electrostatic energy, q is the charge induced on the movable finger, and $\Phi$ is the potential. Differentiating with respect to the normal direction z yields

$$F_z = \frac{\partial E}{\partial z} = q\frac{\partial \Phi}{\partial z} + \Phi\frac{\partial q}{\partial z}. \qquad (2)$$
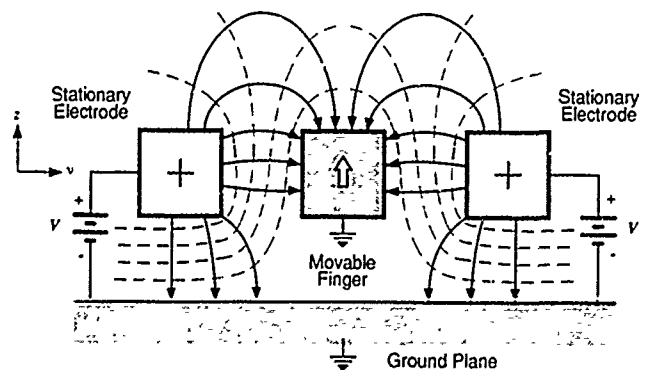


Fig. 1. Simplified cross section of a comb drive with the same voltage applied to all drive comb fingers, including potential contours (dashed lines) and the electric field distribution (solid lines).

However, we have that

$$q \neq 0 \text{ and } \frac{\partial \Phi}{\partial z} \neq 0, \tag{3}$$

and thus,

$$F_z \neq 0. \tag{4}$$

Whether this force causes significant static displacement or excites a vibrational mode of the structure depends on the compliance of the suspension and the quality factor for vertical displacements.

*Modified Comb Design*

There are several means to reduce the levitation force. By eliminating the ground plane and removing the substrate beneath the structures, the field distribution becomes balanced. Alternatively, a top ground plane suspended above the comb drive will achieve a balanced vertical force on the comb. Both of these approaches require much more complicated fabrication sequences. A simpler solution is to modify the comb drive itself. Reversing the polarity on alternating drive fingers results in an altered field distribution, as shown in Fig. 2. Continuing with the energy analysis from (2) and noting that we now have

$$\frac{\partial \Phi}{\partial z} = 0 \text{ and } \Phi \approx 0, \tag{5}$$

it follows that

$$F_z = 0. \tag{6}$$

Controlling levitation may be achieved by changing the comb drive or by modifying the ground plane. Various structures have been developed to alternate the polarity at every stationary drive finger (Fig. 3 (a)), every other finger, every fourth finger (Fig. 3 (b)) and even every sixth drive finger. Three different designs addressing the role of the ground plane are implemented. The conventional ground plane (Fig. 3 (a)) is compared to comb drives with a recessed ground plane and a ground plane which extends only beneath the comb fingers (Fig. 3(c)).

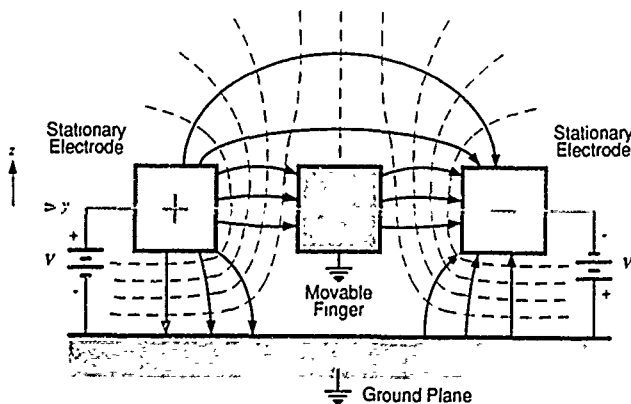

(a)



(b)



(c)
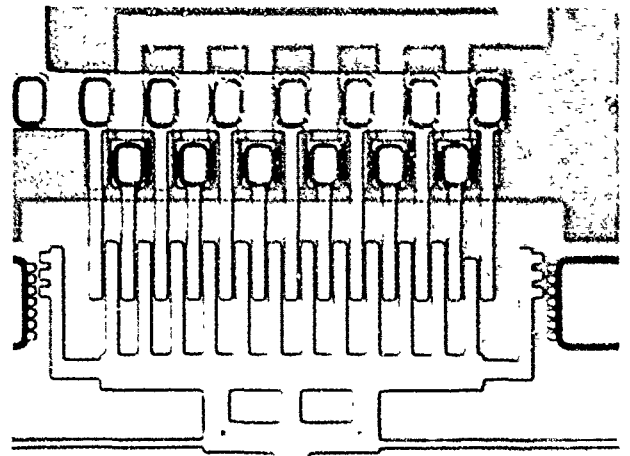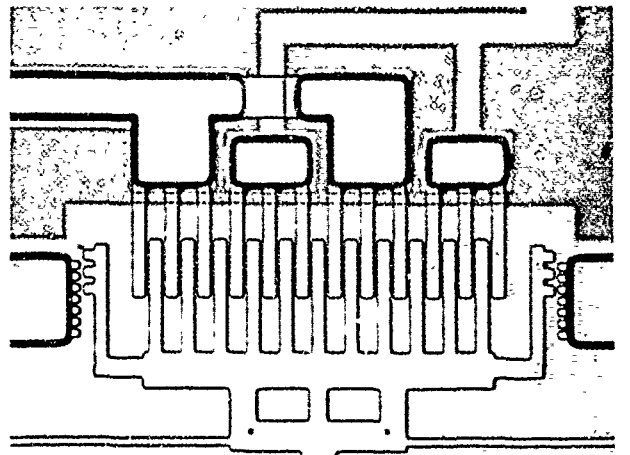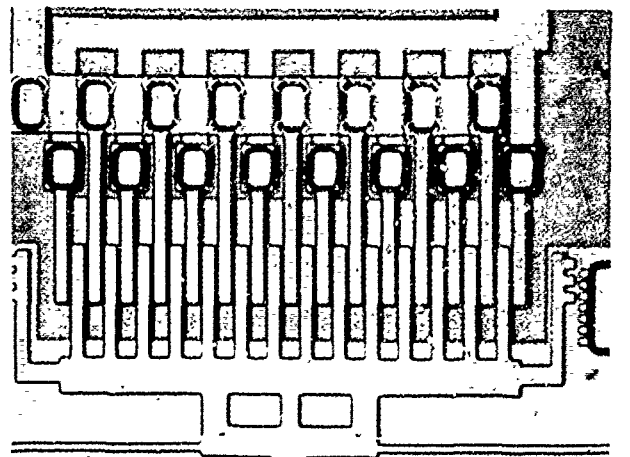


Fig. 2. Simplified cross section of a comb drive with equal and opposite voltages applied to alternating drive comb fingers.

Fig. 3. Optical micrograph of modified comb drives having two drive electrodes consisting of (a) alternating drive fingers and (b) sets of three drive fingers. (c) Modified comb drive with striped ground plane extending only under comb fingers.

24

## EXPERIMENTAL TECHNIQUES

Levitation amplitudes are recorded from low-voltage SEM pictures at various DC biases on the combs. All structures are wired together to make possible the measurement of a number of structures in a single SEM session. The angle of tilt and magnification are fixed for comparison between different structures. Vertical displacements are evaluated by accurately measuring the SEM images with a set of standard linewidths. An SEM of a levitated structure is shown in Fig. 4.

Although measuring the vertical displacement as a function of the comb voltages initially demonstrates levitation control, it is more useful to calculate the vertical force for a particular comb geometry (gap between fingers, film thickness, and offset from the substrate.) In order to do this, the Young's modulus of polysilicon is found by fitting the measured lateral resonant frequency of the structure to the formula from Rayleigh's method [1,2]. Due to difficulties with the polysilicon plasma etching process, the cross section of the suspension is trapezoidal, with the width at the top of the beam $a = 2.2$ $\mu$m and the width at the bottom of the beam $b = 2.8$ $\mu$m, for a film thickness $h = 1.94$ $\mu$m. The expression for lateral resonant frequency of the test resonator in terms of structural dimensions and Young's modulus is [10]:

$$f_r = \frac{1}{2\pi}\left[\frac{2Eh[a^3 + (b^4-4ba^3+3a^4)/4(b-a)]}{L^3(M_p+0.25M_t+0.343M_b)}\right]^{\frac{1}{2}} \quad (7)$$

where L is the length of the folded beam (400 $\mu$m), $M_p$ is the plate mass, $M_t$ is the mass of the outer connecting trusses, and $M_b$ is beam mass. Therefore, E can be expressed as

$$E = \frac{2\pi^2 f_r^2 L^3 (M_p+0.25M_t+0.343M_b)}{h[a^3 + (b^4-4ba^3+3a^4)/4(b-a)]} \quad (8)$$

Using a polysilicon density of $2.3\times10^3$ kg·m$^{-3}$, the Young's modulus is found to be E = 150 GPa for this process run, which is consistent with earlier results [1,2]. The vertical spring constant is found from the Young's modulus and the geometry of the suspension:
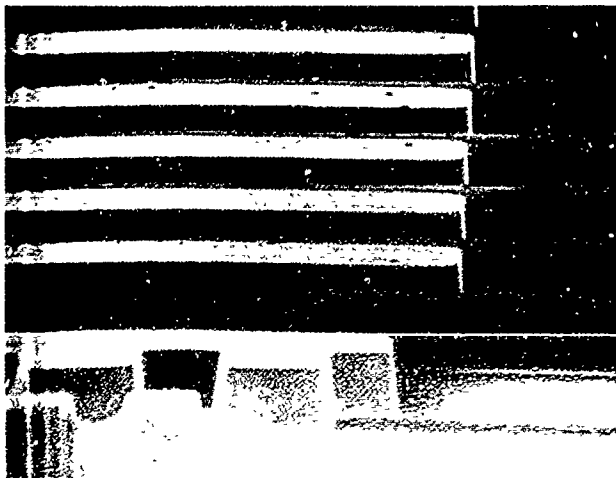


Fig. 4. SEM micrograph of a comb drive with 10V DC bias on the drive. The levitation is 510 ± 20 nm by direct linewidth measurement. Note that the drive fingers, because of the positive bias, appear darkened in the SEM.

$$k_z = \frac{2h^3(a^2+4ab+b^2)E}{3(a+b)L^3} = 86 \text{ nN·}\mu\text{m}^{-1}. \quad (9)$$

Finally, the levitation force is given by $F_z = k_z \Delta z$ where $\Delta z$ is the vertical displacement.

Similarly, the lateral (drive) force of the comb is found by observing the lateral displacement versus voltage, using a vernier with a 0.5 $\mu$m scale. The lateral force, $F_l$, is then evaluated with $F_l = k_x\Delta x$, where $\Delta x$ is the lateral displacement, with the lateral spring constant given by

$$k_x = \frac{2h}{L^3}\left[a^3 + \frac{(b^4-4ba^3+3a^4)}{4(b-a)}\right]E = 140 \text{ nN }\mu\text{m}^{-1}. \quad (10)$$

## EXPERIMENTAL RESULTS

### Levitation for unbalanced comb drives

Levitation is observed by applying a voltage of 0 to 25 V to all drive fingers and is plotted for the test structure in Fig. 5. The vertical displacement increases with applied voltage and reaches an equilibrium near 20 V where the attractive forces between the displaced interdigitated fingers offset the repulsive electrostatic forces between the ground plane and movable fingers. The suspension restoring force also retards levitation. The initial negative deflection for a grounded comb, shown in Fig. 5, cannot be attributed to gravity. Charging effects in the dielectric films are a possible source of this offset displacement. We define the vertical drive capacity, $\gamma_v$, as the levitation force per square of the applied voltage: $\gamma_v = F_v V^{-2}$ [N·V$^{-2}$]. Using the technique described in the
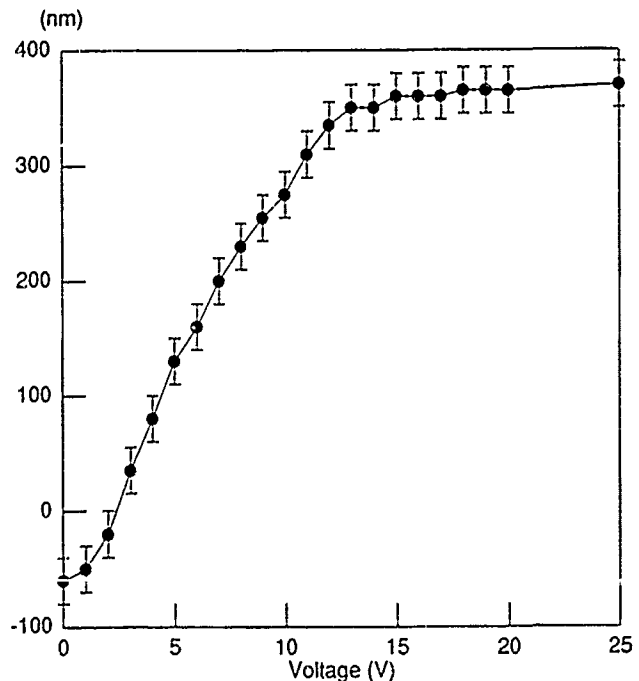


Fig. 5. Vertical displacement for a common voltage applied to both electrodes of a modified comb-drive structure with 18 movable comb fingers and 19 fixed drive fingers, each of which is 40 $\mu$m long, 4 $\mu$m wide, and 2 $\mu$m thick. The gap between fingers is about 2 $\mu$m and nominal distance from substrate is also 2 $\mu$m.

previous section, the levitation force is estimated to be 900 pN $V^{-2}$ of DC bias at $z = 0$ on this particular 19 drive-finger comb.

### Levitation for modified comb drives

Figure 6 is a plot of the levitation resulting from holding one set of an alternating drive fingers at 15V and varying the other set of electrodes from -15V to 15V. The structure is the same as that tested in Fig. 5. As expected, negative voltages in the range of -10 V to -15 V suppress the lifting behavior. As the disparity between the magnitudes of the voltages increases, more lifting occurs, with the limiting case of +15 V applied to all drive fingers yielding the same vertical displacement as found in Fig. 5.

### Drive capacity vs levitation control

We define the lateral drive capacity, $\gamma_l$, of an electrostatic comb drive as the lateral force per square of the applied voltage: $\gamma_l = F_l V^{-2} [N \cdot V^{-2}]$. Experimental results show that the drive capacity of the balanced comb drive is less than that of the unbalanced one with same geometry and number of fingers. The suppression ranges from 50% to 90% of the unbalanced drive, depending on the interdigitation method and ground plane design.

Both $\gamma_l$ and $\gamma_v$ for various balanced comb designs normalized to each drive finger are tabulated in Table I. The normalized value of $\gamma_v$ at $z = 0$ for the unbalanced combs is $47 \pm 1$ pN·$V^{-2}$ per drive finger, and the normalized $\gamma_l$ being $16 \pm 1$ pN·$V^{-2}$ per drive finger. None of the comb designs completely eliminates the levitation as predicted by idealized theory, which assumes evenly spaced comb fingers with vertical sidewalls and neglects native oxide films, which may serve as a charge-trapping dielectric layer.

**Table I**

Normalized vertical and lateral drive capacities per drive finger

| Type | $\gamma_v$ at $z = 0$ $\pm 1$ pN·$V^{-2}$ | $\gamma_l$ $\pm 1$ pN·$V^{-2}$ |
|------|------|------|
| A | 3 | 10 |
| B | 26 | 8 |
| C | 30 | 11 |
| D | 35 | 12 |
| E | 36 | 14 |
| F | 38 | 14 |

A is a structure with striped ground plane and the others are with conventional ground plane. The electrodes of B alternate every drive comb finger; C, every other finger; D, every third finger; E, every fourth finger; and F, every sixth finger. Comb fingers are 40 μm long, 4 μm wide, and 2 μm thick. The gap is nominally 2 μm and the comb is offset from the substrate by 2 μm.

Several qualitative observations were noted during testing of these structures. First, although vertical levitation can be induced with DC bias, none of the structures responds to vertical AC excitation even in vacuum. Therefore, we are unable to obtain the vertical resonant frequencies. Second, those structures with recessed ground planes behave unpredictably, especially when a step change is applied to the DC bias. They have a tendency to be pulled down and become stuck to the substrate, which may be due to charging effects in the underlying dielectric passivation layers.

## CONCLUSIONS

We have successfully evaluated and quantified levitation induced by electrostatic comb drives by direct tests in an SEM, which provides insights into designing structures for controlled out-of-plane motions. Among the various designs studied in this experiment, the best levitation suppression is obtained by alternating the two electrodes at every drive finger with a striped ground plane underneath the comb structure. Levitation can be further reduced by designing structures with vertically stiff suspensions. If controlled levitation is desired, soft suspensions can be used together with ratioed differential and common mode voltages applied to the two electrodes. Vertical AC excitation, charging effects, and etching processes for achieving vertical sidewalls are topics for further research.
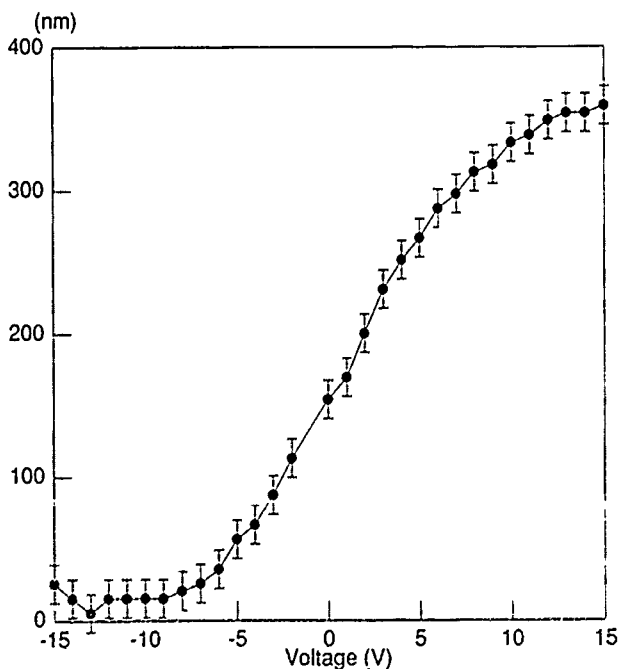
Fig. 6. Vertical displacement of the same structure used in Fig. 5 for varying voltage on one electrode of a modified comb structure, with the other electrode fixed at +15 V.

26

# REFERENCES

[1] W. C. Tang, T.-C. H. Nguyen, and R. T. Howe, "Laterally driven polysilicon resonant microstructures," *Technical Digest,* IEEE Micro Electromechanical Systems Workshop, Salt Lake City, Utah, February 20-22, 1989, 53-59.

[2] W. C. Tang, T.-C. H. Nguyen, M. W. Judy, and R. T. Howe, "Electrostatic-comb drive for lateral polysilicon resonators," *Technical Digest,* 5th International Conference on Solid-State Sensors and Actuators, Montreux, Switzerland, June 25-30, 1989, 138-140.

[3] R. A. Brennen, A. P. Pisano, and W. C. Tang, "Multiple Mode Micromechanical Resonators," *Technical Digest,* IEEE Micro Electromechanical Systems Workshop, Napa Valley, Calif., February 12-14, 1990, 9-14.

[4] M. G. Lim, J. C. Chang, D. P. Schultz, R. T. Howe, and R. M. White, "Polysilicon Microstructures to Characterize Static Friction," *Technical Digest,* IEEE Micro Electromechanical Systems Workshop, Napa Valley, Calif., February 12-14, 1990, 82-88.

[5] C. J. Kim, A. P. Pisano, R. S. Muller, and M. G. Lim, "Polysilicon Microgripper," *Technical Digest,* IEEE Solid-State Sensor and Actuator Workshop, Hilton Head Island, S.C., June 4-7, 1990.

[6] A. P. Pisano, "Resonant-structure micromotors," *Technical Digest,* IEEE Micro Electro Mechanical Systems Workshop, Salt Lake City, Utah, February 20-22, 1989, 44-48.

[7] Y.-C. Tai, L. S. Fan, and R. S. Muller, "IC-processed micromotors: design, technology, and testing," *Technical Digest,* IEEE Micro Electro Mechanical Systems Workshop, Salt Lake City, Utah, February 20-22, 1989, 1-6.

[8] M. Mehregany, P. Nagarkar, S. D. Senturia, and J. H. Lang, "Operation of microfabricated harmonic and ordinary side-drive motors," *Technical Digest,* IEEE Micro Electromechanical Systems Workshop, Napa Valley, Calif., February 12-14, 1990, 1-8.

[9] Maxwell Solver, v. 4.20. Ansoft Corp., 4 Station Square, 660 Commerce Court Building, Pittsburgh, Pa. 15219.

[10] J. Gere and S. Timoshinko, *Mechanics of Materials.* 2nd ed., Belmont: Wadsworth, Inc., 1984.

# A MICROMACHINED SILICON SCAN TIP FOR AN ATOMIC FORCE MICROSCOPE

L. C. Kong, B. G. Orr[†], and K. D. Wise[*]

*Center for Integrated Sensors and Circuits*

[†]*Physics Department*
University of Michigan
Ann Arbor, Michigan 48109-2122

## ABSTRACT

This paper describes the development of single and multichannel scan tips for use in atomic force microscopy. A combination of bulk and surface micromachining techniques has been utilized to fabricate a heavily-doped single-crystal silicon microprobe with integrated polysilicon scan tips. Polysilicon beams, configured both as cantilevers and as double-end-clamped structures 75-200µm long, and ~30µm wide with a thickness of 1-1.5µm have been fabricated. The fabrication process utilizes an undoped polysilicon sacrificial layer and a boron doped polysilicon release structure. By careful choice of the passivation layers, the strain in the structure can be reproducibly controlled. The beams are driven electrostatically and the response amplitude is sensed using interferometric techniques.

## INTRODUCTION

Since the development of the scanning tunneling microscope' (STM) [1] and the demonstration of its capabilities for the atomic-scale resolution of surface features, a variety of additional tools for surface analysis have been proposed based on atomic force, magnetic, and thermal coupling effects between a scanned stylus tip and an opposing surface [2,3,4]. These SXMs appear to hold great promise for material studies at the atomic level and for surface profilometry in the sub-100nm range. Such tools are particularly important for applications in semiconductor manufacturing, where features are expected to reach 0.15µm during the 1990s and where no satisfactory production tools now exist for measuring such surface topographies. The scanning force microscope is of particular interest here since it is viable on both insulating and conducting surfaces.

Atomic force microscope (AFM) employs the same generic principles as the STM. A microprobe, the force sensor, is brought near to the sample surface. As the cantilever beam on the force sensor nears the surface its resonance properties are effected. The interaction of the probe tip with the surface via van der Waals forces alters the effective spring constant of the system, shifting the resonant frequency. To produce an image of the surface, the cantilever is electrostatically excited near its resonance and the probe raster scanned across the surface. Variations in surface topography shift the resonant frequency thereby altering the amplitude and phase response of the cantilever. A feedback loop is used to adjust the height of the tip as it is rastered over the surface in such a manner as to maintain a constant amplitude response of the probe. By recording the feedback signal necessary to produce this trajectory a map of the surface is generated.

Figure 1 shows the initial approach to the realization of the scan tip. A number of surface-micromachined polysilicon cantilevers are formed on a silicon substrate, which is itself formed using bulk micromachining and a diffused boron etch-stop. The cantilevers are resonated using electrostatic drive circuitry and are read out using independent sense capacitors also located under the beams. Table 1 lists the simulated performance for a cantilever having the dimensions noted. The vertical and lateral resolution predicted for the system appear more than adequate for most applications in semiconductor surface profilometry; however, the capacitance variations to be sensed offer a significant challenge, requiring careful integration of on-chip detection and drive circuits and careful shielding of lead runs down to the drive/detect elements.
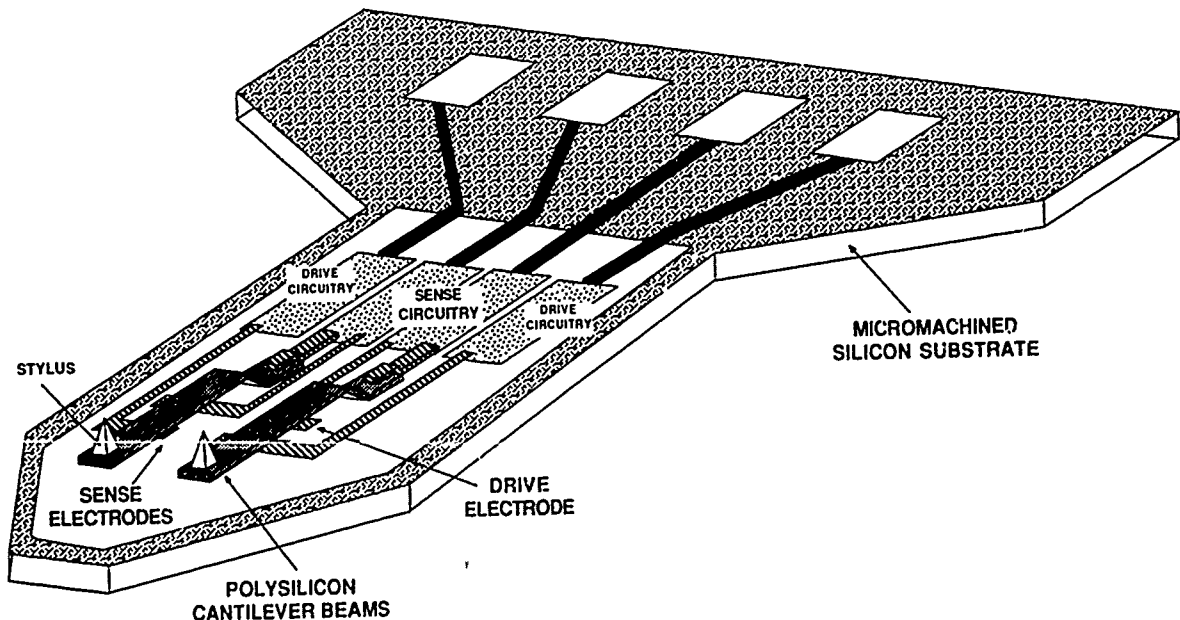


Figure 1: A Micromachined Silicon Multichannel Scanning Atomic Force Microscope with Electronic Drive and Readout.

## CANTILEVER SPECIFICATIONS

| | |
|---|---|
| Tip Height Above Surface | 5 nm to 20 nm |
| . Dimensions | 40 μm x 100 μm |
| . Thickness | 1 μm |
| . Capacitance Gap | 1.3 μm |
| . Capacitance | 7.5 fF |
| . Spring Constant | 2.0 Nt/m |
| . Generalized Mass | $4.0 \times 10^{-12}$ Kg |
| . Mechanical Angular Frequency | $5.0 \times 10^5$ rad/sec |
| . Vibration Amplitude | 1nm |
| . Quality Factor | 50 to 1000 |

## PERFORMANCE

| | |
|---|---|
| . Detectable Force Range | $10^{-8}$ to $10^{-11}$ Nt |
| . Resovable Force Gradient | $10^{-4}$ Nt/m |
| . Vertical Resolution | < 1nm |
| . Lateral Resolution | < 10 nm |
| . Resovable Amplitude | < $10^{-3}$ Å |
| . Resovable Capacitance | 100 aF |
| Scan Range | 10 μm x 100 μm |

Table 1: Simulated Performance of the Silicon Force Microscope

Because of the challenges inherent in capacitive detection of beam deflection, our current structure employs optical readout, as shown in Fig. 2, avoiding the initial need for on-chip circuitry and careful electrical shielding. Here, a hole is formed at the tip of the substrate behind the scan tip by masking the boron diffusion from this area during substrate formation. An optical fiber is positioned behind the beam as shown. Interference between the light internally reflected from the face of the fiber and light reflected from the cantilever beam allows very sensitive measurement of the beam vibration amplitude. A resolution of better than 0.01nm has been demonstrated in such applications [5]
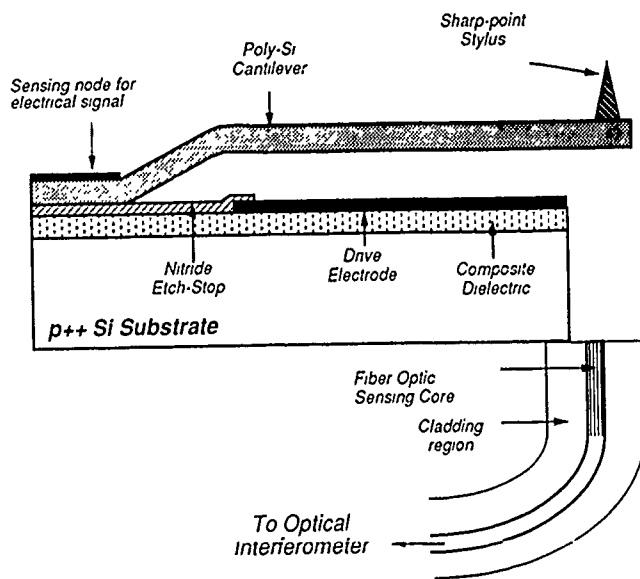


Figure 2. Detail of a Polysilicon Cantilevered Stylus with Electrostatic Drive and Fiberoptic Readout

## FABRICATION

The fabrication of the microprobe for the AFM involves both bulk and surface micromachining techniques. Deep boron diffusion and boron etch-stop techniques are used to form the thick support substrate for the force sensor; while surface micromachining is used to form the force sensor. The majority of surface micromachined structures and technologies have typically used a phosphorus-doped silicon dioxide (PSG) film as the sacrificial layer and a polysilicon layer as the microstructure material. Although the use of PSG has been widely accepted, a number of undesirable characteristics are generally thought to limit its universal applications. These include : 1) long etch time in concentrated HF to undercut the polysilicon microstructure which can cause attack of other materials on the wafer; and 2) undesirable interaction between the PSG sacrificial layer and the top polysilicon layer during the stress-relief annealing process that is often required in surface micromachining technologies [6]. In order to avoid these shortcomings, we have developed a new surface micromachining process that utilizes polysilicon for both the sacrificial layer and the microstructure material, as discussed below.

Silicon microprobes with silicon scan tips were fabricated using a seven mask process as shown in Figure 3. Fabrication starts with a (100) silicon wafer of standard thickness and resistivity. First a 1μm thermal oxide layer is grown and patterned to define the microprobe substrate area. A deep boron diffusion is performed at 1175°C for 15 hours and the thermal oxide is etched away. A composite dielectric layer consisting of silicon dioxide, silicon nitride and silicon dioxide is next deposited using low pressure chemical vapor deposition (LPCVD). This layer is patterned to form an insulator on the probe area and a contact window to the boron diffusion area. A heavily boron-doped poly-Si film is then deposited and patterned to form the drive electrode. Next, a 50nm layer of LPCVD silicon nitride is deposited and patterned. It protects against both out-diffusion of boron into the polysilicon sacrificial layer as well as attack by the final EDP etch. An undoped poly -Si sacrificial spacer is now deposited and patterned to define the undercut area. A second diffusion barrier composed of LPCVD silicon dioxide and silicon nitride is now deposited over the sacrificial layer. A 1.1μm layer of poly-Si is deposited and heavily doped with boron. It is passivated with a 50nm layer of LPCVD silicon nitride and the nitride/poly-Si layer is patterned to define the scan tip in the form of a cantilever or bridge. During this step holes are also etched in these structures and play a dual role. The first is to facilitate undercut etching of the sacrificial layer, and the second is to increase the quality factor of the resonant system by minimizing air damping. Lastly, metal conductors composed of Cr/Au are deposited and patterned using a lift-off process to form the electrical contacts. The field areas of the chip, including polysilicon interconnects, are protected by LPCVD silicon nitride. The final step is an EDP etch which releases both the individual dies from the wafer and removes the sacrificial layer under the cantilever/bridge. The silicon microprobes with integrated scan tips are then mounted on small tapered printed-circuit boards which plug directly into conventional DIP sockets.

Figure 4 shows a micrograph of a completely finished microprobe with the polysilicon beam located at the tip. The overall probe length is 2mm. Figures 5 and 6 show several SEM micrographs of polysilicon cantilever and bridge structures. The polysilicon is 1.1μm thick.

It was found that careful control of the residual stresses in the microstructure was needed to produce planar beams. For AFM applications a slight curvature away from the substrate is desirable. Figure 7 shows a comparison between a beam with a top layer of silicon dioxide and one with a top layer of silicon nitride. The oxide appears to be in compression bending the beam toward the substrate whereas the silicon nitride warps the beam away from the substrate. By varying the thickness and composition of the dielectric layers passivating the upper and lower sides of the released structure, good control of the curvature could be obtained [7].

Mask oxide

Composite dielectric

Deep boron diffusion area
Depth: 15 μm

Undoped area

Poly-Si for drive electrode

Nitride etch stop layer

Composite sacrificial spacer

Metalization    Poly-Si cantilever

EDP etching to removal substrate and sacrificial spacer

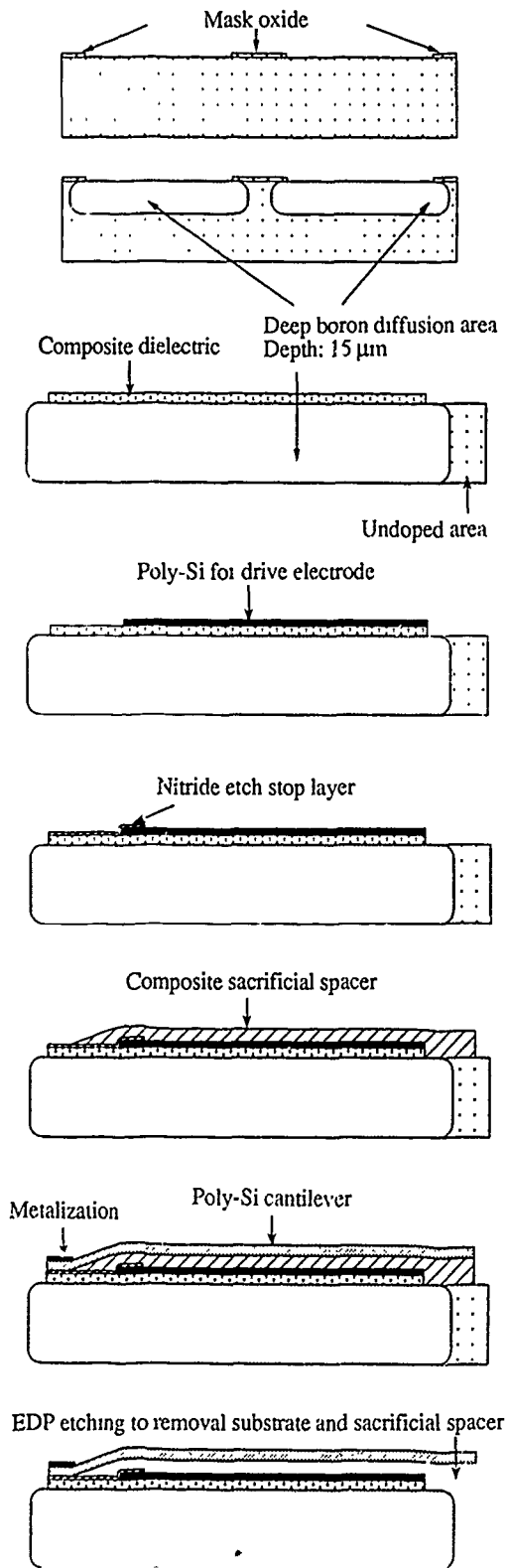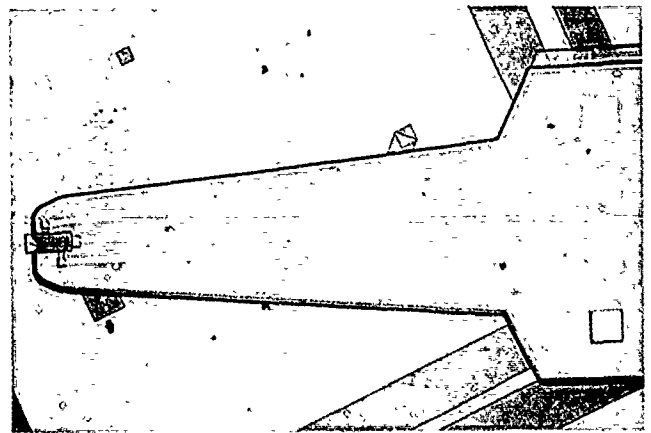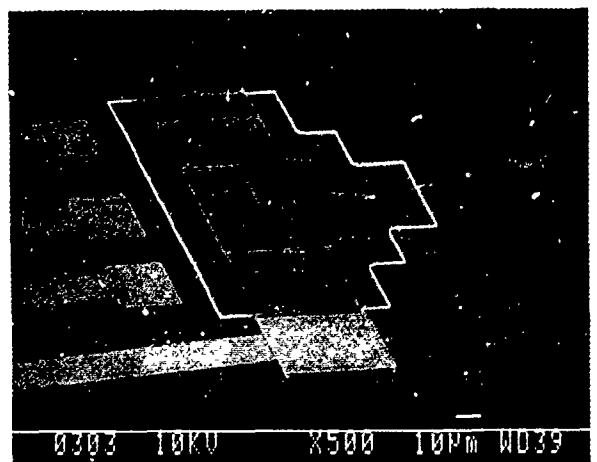Figure 3: Fabrication process for silicon microprobes with polysilicon scan tips.



Figure 4: Micrograph of a completely finished microprobe with the polysilicon beam located at the tip. The overall probe length is 2mm.



0391  10KV  X2,000  10μm  WD17



0303  10KV  X500  10μm  WD39
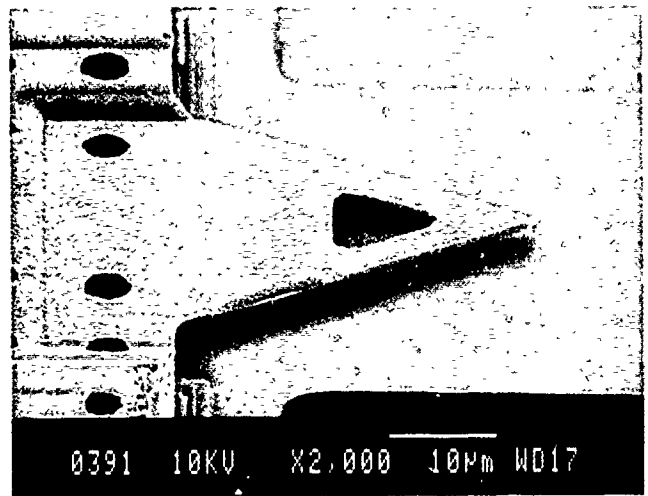
Figures 5:   SEM views of a released doubly-end-clamped polysilicon bridge structure. The polysilicon is 1.1μm thick.
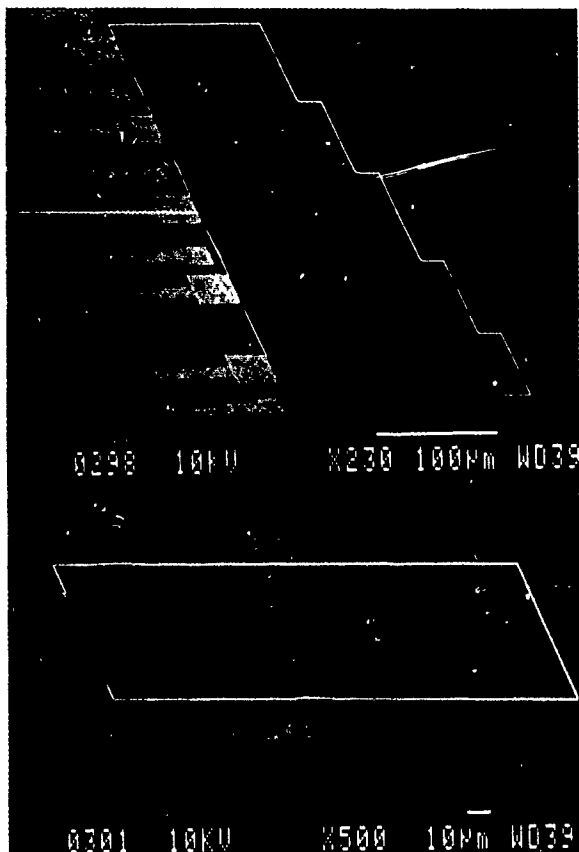
Figure 6: SEM views of released polysilicon cantilever structures having different lengths. The polysilicon is 1.1μm thick.
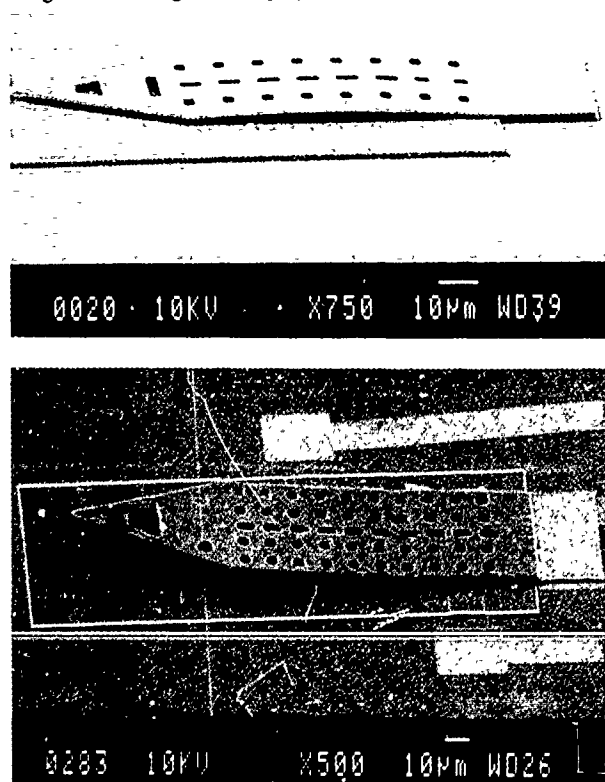




Figure 7: SEM views of a beam with a top layer of silicon dioxide bending the beam toward the substrate (top) and one with a top layer of silicon nitride bending the beam away from the substrate (bottom).

The above fabrication process is compatible with the integration of on-chip circuitry on the boron-doped silicon substrate. In addition, the process is simple, single-sided, and has a very high yield. Since an undoped polysilicon sacrificial layer is used which can be etched in EDP, chemical attack of other materials on the microprobe is avoided.

## CONCLUSION

A fabrication technology for a micromachined silicon microprobe with integrated polysilicon scan tips has been presented. A heavily boron-doped silicon substrate is formed using deep boron diffusion. Polysilicon beams, configured both as cantilevers and as double-end-clamped structures, having a variety of dimensions have been realized. The fabrication process utilizes an undoped poly Si sacrificial layer and a boron doped poly Si release structure. By careful choice of the passivation layers, the strain in the structure can be reproducibly controlled. On-chip circuitry for interfacing with integrated scan tips can be fabricated on the silicon substrate. A variety of techniques are being explored to form the stylus tips, where required, including ion-beam plating.

## ACKNOWLEDGEMENTS

## REFERENCES

1. G. Binning and H. Rohrer, *IBM J. Res. Dev.*, *30*, pg. 231 (1986).
2. G. Binning, C. F. Quate, and Ch. Gerber, *Phys. Rev. Lett.*, *56*, pg. 930 (1986).
3. G. M. McClelland, R. Erlandsson, and S. Chiang, *Rev. of Progress in Quantitative Non-Destructive Evaluation*, D. O. Thompson and D. E. Chimenti, eds., *6B*, pg. 1307. New York: Plenum Press, 1987.
4. C. C. Williams and H. K. Wickramasinghe, *Appl. Phys. Lett.*, *49*, pg. 1587 (1986).
5. D. Rugar, H. Mamin, R. Erlandsson, J. Stern, and B. Terris, *Surface Science*, June 6, 1988.
6. M.W. Putty, S.C. Chang, R.T. Howe, A.L. Robinson, and K.D. Wise, "Process Integration for Active Polysilicon Resonant Microstructures," *Sensors and Actuators*, 20 (1989) pp. 143-151
7. S.T. Cho, K. Najafi, and K.D. Wise, "Scaling and Dielectric Stress Compensation of Ultrasensitive Boron-Doped Silicon Microstructures," *Tech. Digest, IEEE Micro Electro Mechanical Systems Workshop*, pp. 50-55, Napa, CA, Feb. 1990

31

# PRESSURE SENSOR DESIGN AND SIMULATION USING THE CAEMEMS-D MODULE

Y. Zhang, S. B. Crary, and K. D. Wise

*Center for Integrated Sensors and Circuits*
University of Michigan
Ann Arbor, MI 48109-2122

## ABSTRACT

This paper reports a workstation-based simulation module for micromachined silicon diaphragm structures and its application to capacitive pressure transducers. The module offers a menu-driven user interface, access to a growing database of material parameters and performance data, and an interface to commercial finite-element software. It provides the user with three-dimensional stress, deflection, resistance, and capacitance information and allows three-dimensional visualization of complex diaphragm structures. Finite-element simulations of boron-diffused, thin-diaphragm capacitive pressure transducers, produced using the dissolved-wafer process, allow effects such as diaphragm stress, dielectric-based stress compensation, corrugations, and bosses to be explored. Detailed comparisons between the simulations and experimental data generally agree to within a few percent.

## INTRODUCTION

As the number and complexity of silicon microstructures continue to expand, the need for workstation-based software for their simulation and optimization is becoming critical. Such software is the logical medium for organizing existing knowledge regarding such structures and their associated materials, as well as for rapid design verification and, perhaps, synthesis. However, while powerful simulators exist for microelectronics at the process, device, and circuit levels, very few such packages exist for sensors, where mixed mechanical, thermal, and electrical modes offer particularly interesting challenges and opportunities.

CAEMEMS (Computer-Aided Engineering of Micro-Electro-Mechanical Systems) is a workstation-based system of interlinked design-analysis tools for micro-electro-mechanical systems [1]. The CAEMEMS architecture includes four principal components: a database of material properties and process-model parameters, a process modeler that takes as input mask and process information and produces a solid model of the simulated structure, a solid modeler that allows for visualization and design verification, and a device modeler that performs finite-element simulations of components and systems. The CAEMEMS-D (Diaphragm) module is the first realization of a software building block for the device-modeling component of CAEMEMS. The primary role of this module is to provide a high-level design tool for predicting the responses of silicon-based diaphragms to various loading conditions. The module contains the capabilities of SENSIM [2] as a subset of its capabilities. This paper describes the general organization of the module and its uses in exploring the performance of thin boron-doped silicon diaphragms, with and without dielectric-based stress compensation and non-planar features in the form of bosses and corrugations.

## GENERAL MODULE ORGANIZATION

The MEMS designer interacts with the CAEMEMS-D module via a user interface that allows the user to specify the structure of interest using a menu-driven format and also allows selection of a number of output options, see Figure 1. The simulator handles piezo-resistive or capacitive transducing modes on square, rectangular, or circular diaphragm geometries, allowing bossed and corrugated features if desired. An interpretation/control block controls the overall operation of the simulator, examines the structure defined, and selects from among stored performance data, analytical solutions, and finite-element simulations for determining the corresponding performances. Pressure and force act as forcing functions for these static linear and nonlinear simulations. Finite-element simulations can be run on local workstations, networked workstations, or external computers (e.g., a supercomputer). For much of the work reported here, execution has been on a Cray Y-MP computer with an Apollo DN-3500 acting as the local workstation. The module prepares data from the input specification for interface to commercial finite-element and visualization code, in our case ANSYS.
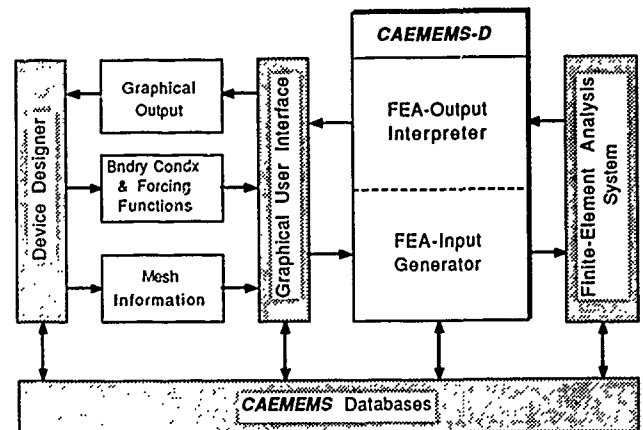


Figure 1. Block diagram of the CAEMEMS-D module

## THE USER INTERFACE

Figure 2 shows a prototype of the user interface for CAEMEMS-D as it would appear at one point in the specification of a square, corrugated silicon diaphragm. The user first specifies the geometrical features and materials of the structure. At present, the materials database is limited to 7740 glass, polysilicon, bulk silicon, silicon nitride, and silicon dioxide, with material parameters limited to those derived from nominal processing. As parameter dependencies on process variables are defined, this information is also being entered into the database so that when completed, process information will also be prompted by the interface. Eventually, the simulator will be linked to a full process simulator so that appropriate material

parameters can be generated (or processes optimized) automatically. All lateral and vertical (thickness) dimensions are entered using appropriate visualization figures. These include data on the overall diaphragm dimensions; boss dimensions; corrugation profiles, numbers, dimensions, and placements on the diaphragm; resistor placements, locations, and sensitivities (for piezoresistive devices); and capacitive electrode placements and gap spacings (for capacitive devices).

At the diaphragm edges, boundary conditions can be user-specified as fixed (simple), clamped (built-in), or micromachined (<111> bevel); or they can be specified as some combination of these by an edge-factor as in [2]. Where the appropriate edge conditions are unknown, the edge geometries can be input and the appropriate edge conditions can be determined by the simulator, or the entire structure, including the rim, can be simulated using ANSYS. The control software also allows the user to select outputs in the form of plots or tables of deflection, stress, resistance, or capacitance data. Where a solid model or contour plot is desired, the user may specify the desired viewing angle.
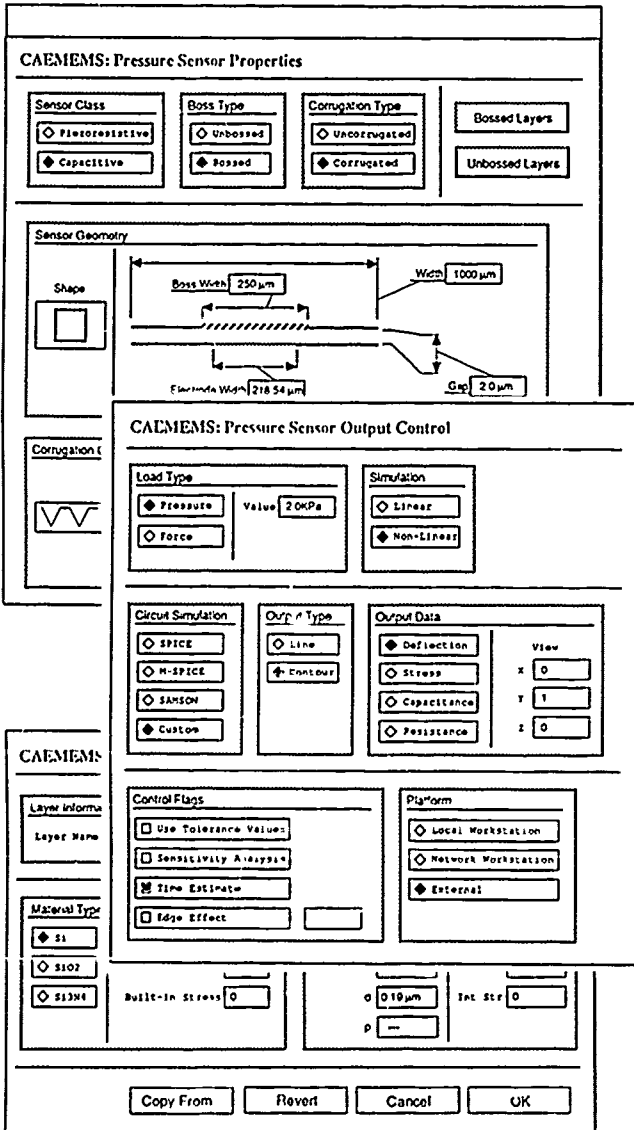


Figure 2. Sample viewscreens in the CAEMEMS-D user interface

## SENSOR SIMULATIONS

### Small-Signal Response of Planar Diaphragm Structures

As an initial application of this simulation module, we have focused on the capacitive pressure sensor structure first implemented by Chau [3] but with the geometries designed for an ultrasensitive flowmeter by Cho [4]. Figure 3 shows a cross section of the basic device. The transducer is produced using a four-mask dissolved-wafer process, with a deep-boron diffusion defining the chip rim and the boss and a shallow diffusion defining the thin portion of the diaphragm structure. An additional mask and etch would be used to define the corrugations when desired. The diaphragm size is 2mm x 2mm with nominal thickness and gap dimensions of 2.5µm and 3.5µm, respectively. The lower electrode is 1.5mm x 1.5mm.
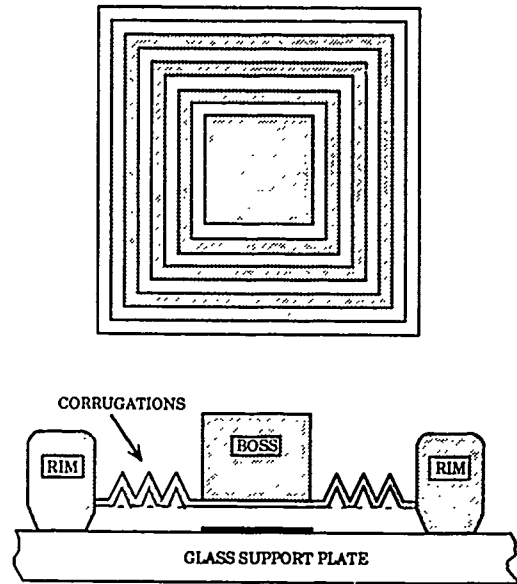


Figure 3. Basic device produced with the dissolved-wafer process

The heavily boron-diffused diaphragm contains considerable internal stress [5], which can be accentuated or compensated by coating the diaphragm with LPCVD silicon nitride or silicon dioxide dielectrics, respectively. The Young's moduli and internal stresses of these various films have been independently measured and reported by Cho [5]. Because the sensitivity of these thin diaphragms is a sensitive function of both geometrical and material parameters, and because detailed experimental data on these structures exist, they were chosen as a vehicle for initial use of the simulator.

Figure 4 shows the capacitance versus pressure characteristics for four planar-diaphragm transducers having the dimensions noted. Figure 4a shows the response of a bare silicon diaphragm over a 300mTorr range along with measured data. The sensitivity is 0.38fF/mTorr, with excellent agreement between the simulations and the experimental data. Figures 4b through 4d show responses of similar transducers with silicon thicknesses of 2.80, 2.60, and 2.45µm and with diaphragms coated with 1200Å of silicon dioxide, with 1000Å of silicon nitride, and with 1000Å of silicon dioxide over 1000Å of silicon nitride, respectively. Due to the known stress levels in these deposited dielectrics, the resulting
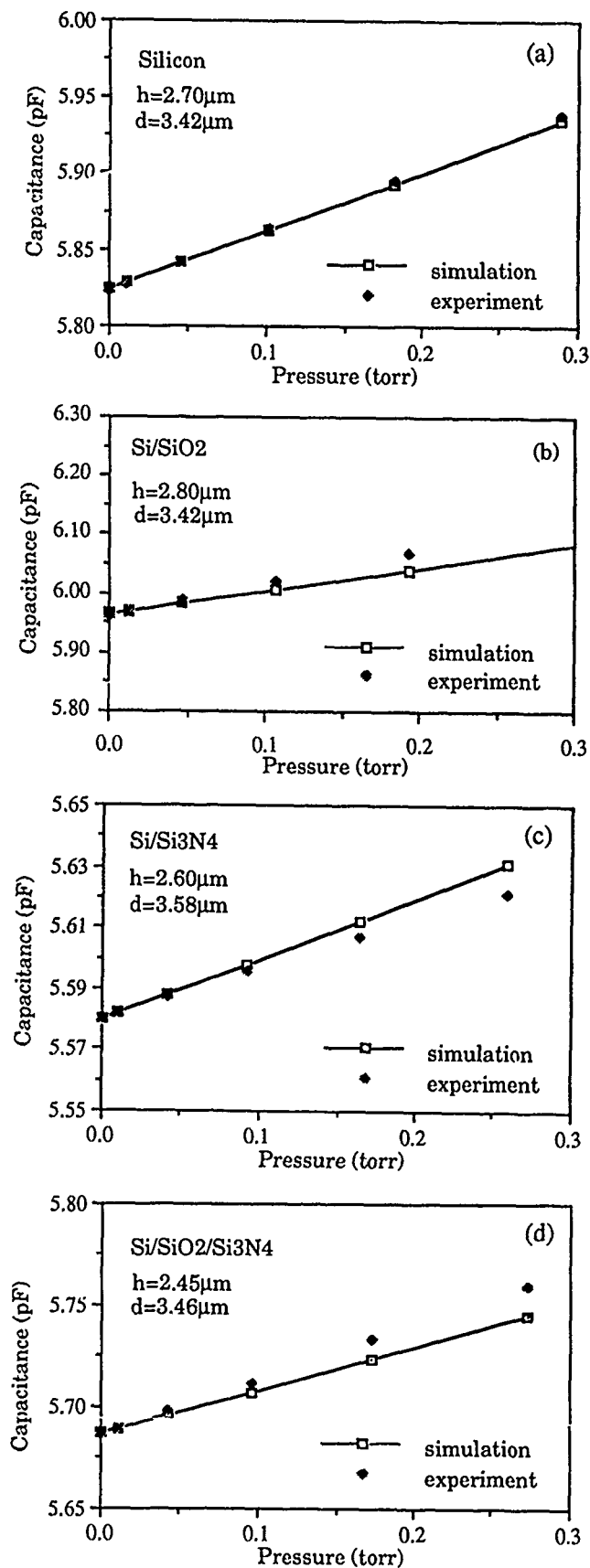
Figure 4. Capacitance vs. pressure characteristics for four planar-diaphragm transducers. All are 2mm x 2mm, but have different silicon thicknesses (h) and gaps (d).

composite diaphragms should have effective stress levels which are shifted to less tensile or more tensile values by the oxide and nitride layers, respectively. The resulting pressure sensitivities should be increased or decreased by the presence of the dielectric layers, according to the analytical theory described by Chau and Wise [6]. The simulated sensitivities are 0.39fF/mTorr, 0.20fF/mTorr, and 0.21fF/mTorr for Figs. 4b through 4d, respectively. Given that the values of the intrinsic stress used as inputs to the simulator are known only within error bounds of ±20%, the agreement between simulations and experiments is good and provides evidence that the simulator is correctly modeling the composite structures at this level of accuracy.

Another capability of the module is demonstrated in Figure 5, where the simulated effect of intrinsic stress on the center deflection of square, clamped diaphragms is compared with the analytic theory for round diaphragms [6]. The applied pressure for this demonstration is in linear-response regime, but can be chosen arbitrarily.
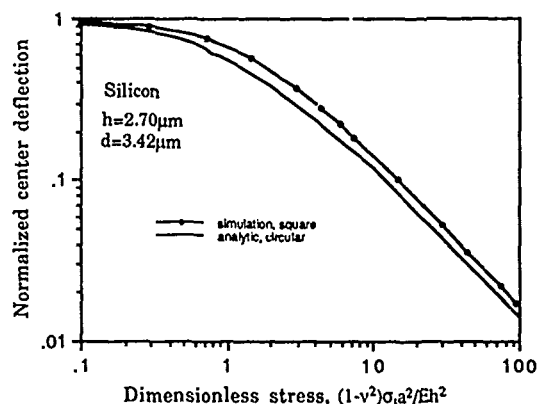


Figure 5. Comparison of normalized response curves for a planar-diaphragm transducer

### Non-Planar Diaphragm Simulation

As another example of CAEMEMS-D use, a square corrugated and bossed diaphragm is simulated. The diaphragm profiles are similar to the generic diaphragm of Figure 3. The diaphragm size is 1mm × 1mm, the boss occupies the center 0.25mm × 0.25mm, while the corrugation consists of eight V-grooves with a depth of 22.1µm and a center-to-center spacing 48.9µm. The non-ideal etching formation of corrugations along the <111> direction is included. Figures 6 and 7 show the simulated deflection and the capacitance vs. pressure of this non-planar diaphragm, respectively.

In addition, Figure 8 shows the normalized center deflection response to normalized (dimensionless) pressure for planar and non-planar diaphragms. The solid lines are based on known handbook curves [7] incorporated into the module database, whereas the dashed curves for various diaphragm thicknesses have been added to the database as the result of ANSYS simulations. The corrugation and boss reduce sensitivity in the low and middle pressure ranges. However, both approaches dramatically improve linearity in the midrange response and considerably smooth the overall response characteristics. Furthermore, if suitable software for organizing the performance library can be added to the module, it should be possible to automatically expand the performance database as it is used for the simulation of specific structures, thus

34

saving subsequent effort and expense for the user. The present simulations of bossed and corrugated structures run on the Cray Y-MP computer at the San Diego Supercomputer Center (SDSC) take approximately 20 cpu-seconds, whereas they take nearly 1200 cpu-seconds on an Apollo DN-3500.
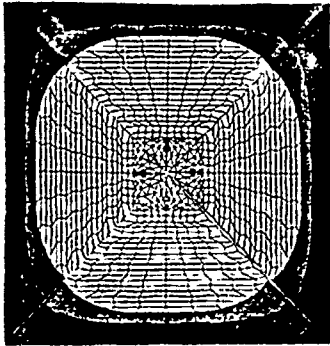


Figure 6. Contour plot of displacement of a non-planar diaphragm
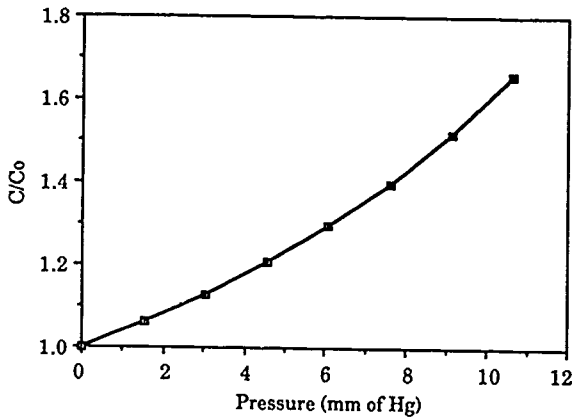


Figure 7. Capacitance vs. pressure of the non-planar diaphragm transducer
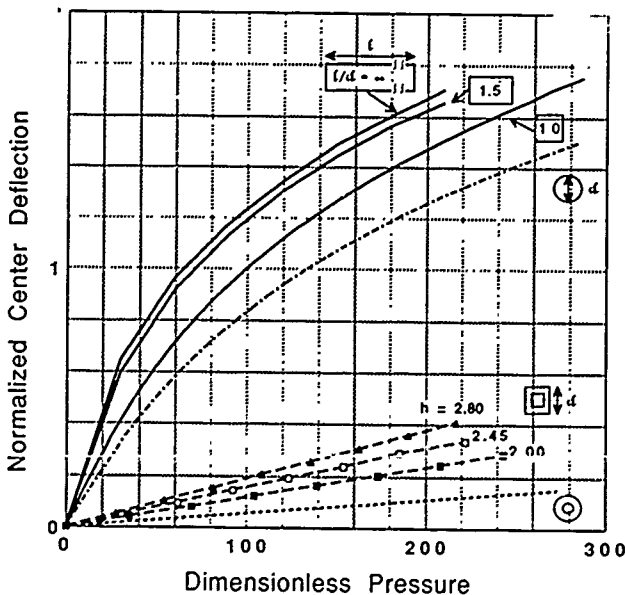


Figure 8. Normalized center-deflection response to normalized pressure ($Pa^4/Eh^4$) for a non-planar diaphragm of side-length a and Young's modulus E

## CONCLUSIONS

This paper has described the organization of a workstation-based simulator for diaphragm micro-structures. The module is capable of finding both analytic and finite-element solutions to distributed three-dimensional deflection/stress problems, incorporating internal stress effects and nonlinearities into the static solutions. Applications to the simulation of a dissolved-wafer capacitive pressure sensor has produced good agreement with measured results.

While useful at the present time, the software for CAEMEMS module is still evolving, with significant enhancements still to be added in the areas of user interface and visualization of three-dimensional structures. In addition, the incorporation of thermal forcing functions, combined thermal/mechanical/electrical effects, linear and nonlinear dynamics, and links with other components of CAEMEMS, provide important challenges for the future.

## REFERENCES

1. S. Crary and Y. Zhang, "CAEMEMS: An Integrated Computer-Aided Engineering Workbench for Micro-Electro-Mechanical Systems," *IEEE MEMS-90 Workshop, Napa Valley, CA,* Feb. 11-14, 1990.

2. K. W. Lee and K. D. Wise, "SENSIM: A Simulation Program for Solid-State Pressure Sensors," *IEEE Transactions on Electron Devices,* Vol. ED-29, No.1, January 1982.

3. H. L. Chau and K. D. Wise, "An Ultraminiature Solid-State Pressure Sensor for a Cardiovascular Catheter," *IEEE Transactions on Electron Devices,* Vol. 35, No. 12, December 1988.

4. S. T. Cho, K. Najafi, C. L. Lowman, and K. D. Wise, "An Ultrasensitive Silicon Pressure-Based Flowmeter," *IEDM Tech. Digest,* pp. 499-502, December, 1989.

5. S. T. Cho, K. Najafi, and K. D. Wise, "Scaling and Dielectric Stress Compensation of Ultrasensitive Boron-Doped Silicon Microstructures," *IEEE MEMS-90 Workshop, Napa Valley, CA,* Feb. 11-14, 1990.

6. H. L. Chau and K. D. Wise, "Scaling Limits in Batch-Fabricated Silicon Pressure Sensors,"*IEEE Transactions on Electron Devices,* Vol. ED-34, Apr. 1987.

7. M. D. Giovanni, *"Flat and Corrugated Diaphragm Design Handbook,"* Marcel Dekker, New York, 1982.

# Automatic Generation of a 3-D Solid Model of a Microfabricated Structure

Robert M. Harris, Fariborz Maseeh, and Stephen D. Senturia
Microsytems Technology Laboratories,
Massachusetts Institute of Technology,
Cambridge, MA, USA 02139

## 1  Introduction

As microfabricated sensors and actuators become increasingly sophisticated and complex, there is an increasing need fc CAD tools to permit rational design of these devices. Most critically needed are CAD tools for simulating the mechanical behavior of a sensor or actuator. Simulation presents two fundamental problems[1, 2]: (A) construction of a three-dimensional solid model from a description of the mask layout and process sequence to be used in fabricating the device; and (B) prediction of the material properties of each of the constituent components in the device, including possible process dependence of these properties. We recently presented an architecture that coherently addresses these issues[3]. (See Fig. 1.) In this architecture problem (A) is addressed by a solid modeling tool (the "Structure Simulator") and problem (B) is addressed by a "Material Property Simulator".

Also as part of that work[3], we presented the construction of a 2-D (axisymmetric) solid model for a process sequence of unmasked thermal oxidation optionally followed by removal of the back oxide. This structure was then analyzed using FEM to find the stress distribution in the structure, and, for the single-sided oxide structure, the wafer curvature Here, we present the automatic construction of a 3-D solid model from a simple process sequence and mask layout along with the simulation and analysis of a simple example. Koppelman[4] has developed a program ("OYSTER") which permits construction of a 3-D polyhedral-based solid model from a mask layout and primitive process description. The differences between Koppelman's approach and the one taken here are discussed in Section 3.

We first give an overview of our CAD system, then the Structure Simulator is described in some detail. Finally, we present the use of the Structure Simulator to simulate wafer bending due to a simple process sequence with a single mask.

## 2  CAD Architecture

The architecture of our CAD system is shown in Fig. 1. It consists of three basic subsystems (outlined by dashed lines in the figure). Microelectronic CAD, Material Property Simulator, and Mechanical CAD. The interactions among the subsystems and their various components are shown by arrows whose direction specify the direction of information transfer. The "User Interfaces" denote direct user access to architecture components.

This architecture is part of a lager Computer Aided Fabrication (CAF) project which is intended to integrated all phases of design and manufacture of microfabricated devices. Central to this integration effort is a uniform process description, the Process Flow Representation (PFR), to be used for both design and manufacturing[5]. This uniform description allows the designer to simulate the process from the *same* process description that would be used to actually fabricate the device.

The primary interface for mechanical modeling is through PATRAN[6], a mechanical CAD package which provides interactive construction of 3-D solid models, graphical display, and interfaces to FEM packages (we are using ABAQUS[7]). The 3-D geometry resides in the PATRAN Neutral File with additional model information stored separately. We have used the material-property format of the Neutral File in our initial version of the Material Property Simulator.

All of the commercially available code in Fig. 1 is installed and operating. The Structure Simulator has been implemented at an elementary level, and interfaced with the Mechanical CAD subsystem.

In the Microelectronic CAD subsystem, mask layout is created in CIF[8] format using KIC[9], and the process sequence is created in the PFR using a standard file editor. SUPREM-III[10] and SAMPLE[11, 12] are installed to provide depth and cross-sectional process simulation capability. The critical component of this subsystem is the Structure Simulator, which must merge the mask layout and process information to construct a three-dimensional solid model. Two kinds of information must be tracked, the *geometry* of the structure (position, shape, and connectivity of each component), and the *material type* and associated process conditions used to create each component of the structure. The geometry is passed directly to the Mechanical CAD subsection using the PATRAN Neutral File format. The material type and associated process conditions for each component are passed to the interface portion of the Material Property Simulator through what we call the Process History File. A more detailed description of the Structure Simulator is given in the next section.

The Material Property Simulator reads the process sequence for each component of the solid model from the Process History File and generates a set of material property data. The material properties are passed to the Mechanical CAD subsystem, either via the PATRAN Neutral File, the PATRAN interactive (graphics), or directly into the FEM input file The direct FEM input allows use of an FEM simulator independent of PATRAN,
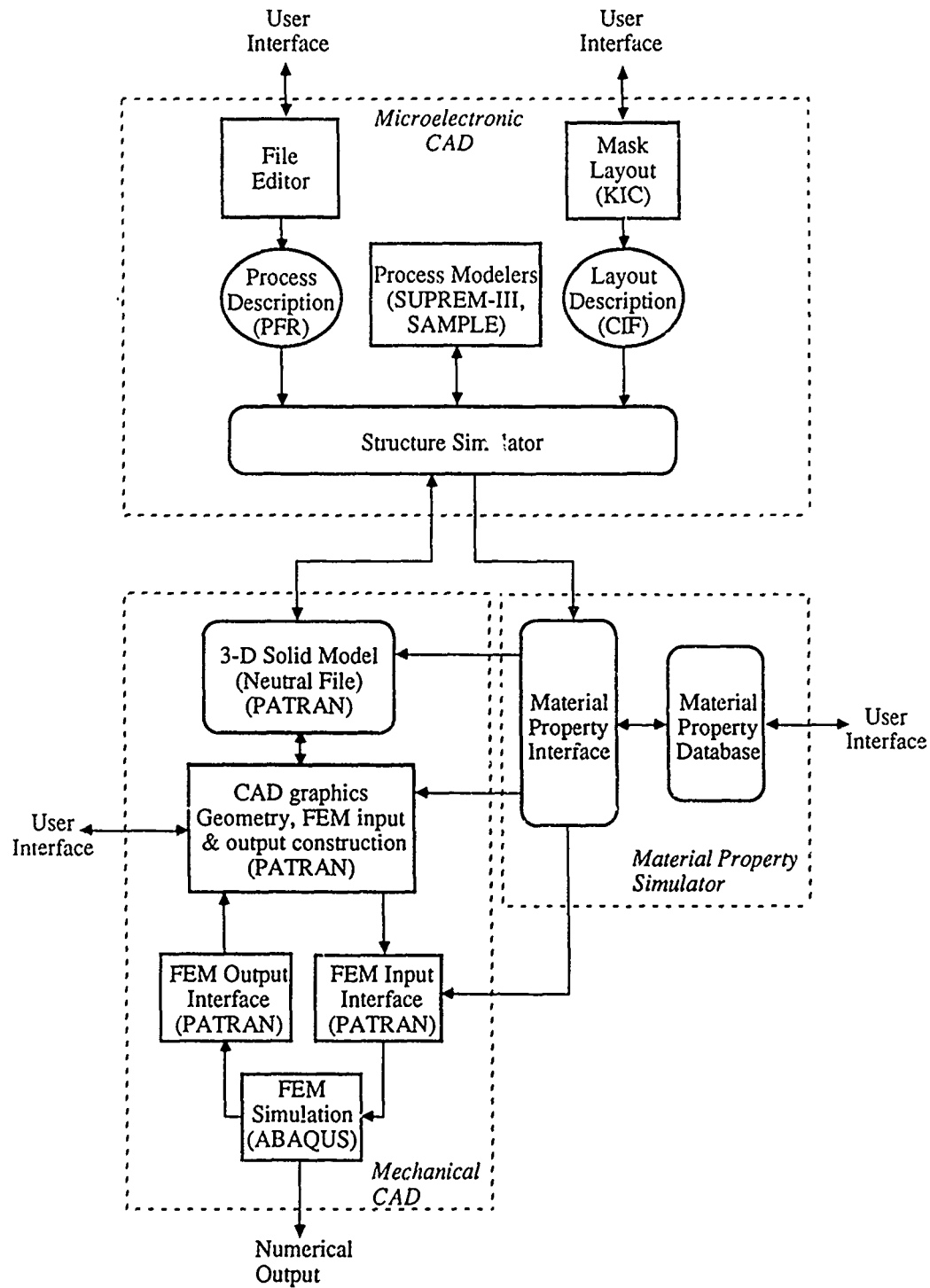
Figure 1: CAD architecture for microelectromechanical design.
Arrows denote information flow.

furthermore, it provides for a way to introduce *intrinsic stress* into the mechanical model. (Because of the detailed organization of PATRAN, intrinsic stress must be treated differently than *thermal mismatch stress*.)

Initially, we have elected to use pre-constructed PATRAN-readable data files to manually enter material property data for each type of material which might be a constituent layer of the device. For our example, only data for silicon and silicon dioxide are needed. This method, however, is impractical in the general case, since it requires manual construction of a data set for each possible permutation of material and process conditions. To address this problem, a more sophisticated object-oriented material property simulator is under development and has been described elsewhere[3].

In the Mechanical CAD subsystem, the geometric information from the Structure Simulator is discretized into finite elements, material properties from the Material Property Simulator are associated with their proper geometries, and the appropriate loads and boundary conditions specified. The solid model geometry and material properties are read and manipulated in PATRAN graphics, and a complete finite element input model is generated and optimized interactively. The FEM model is then translated into an ABAQUS input file for FEM analysis. (Intrinsic stress would also be entered into the input file at this point.) The results of the FEM analysis are then translated back into a PATRAN-readable form for display. PATRAN can then be used to examine simulated mechanical behavior of the device.

## 3 Structure Simulator

The Structure Simulator works on a process-step-by-process-step basis, modifying the solid model after each process step. This mimics the physical fabrication sequence, in which each process step causes a change in the wafer. The final structure is therefore the result of a sequence of such changes. The step-by-step operation of the Structure Simulator also permits simulation of intermediate results of the process sequence. This can be used, for example, to simulate whether the structure will maintain mechanical integrity throughout the entire process sequence.

For each process step, the Structure Simulator begins by reading the current step from the PFR, and the effect of this step is determined. (For example, the process *step* "oxidize in wet $O_2$ for 4hrs. at 1000°C" is converted into the process *effect* "grow 1 $\mu$m silicon dioxide on bare (100) silicon".) When necessary, the PFR information is passed to process simulators (e.g., SUPREM-III, SAMPLE) and the simulation results are used to determine the process effect. (In Koppelman's program, OYSTER[4], the process effect is entered directly using its Process Description Language (PDL). OYSTER, therefore, bypasses this step.) Layout information is consulted using the OCT database[13]. Although OCT is a general purpose database for microelectronic design, we use it here solely as a programmatic interface to the CIF layout description.

Modification of the solid model based on the process effect is done in two steps. First, the process effect is decomposed into a combination of primitive construction operators. These primitive operators are then used to modify the solid model from the previous process step, both the geometry (in the Neutral File), and the material-type information (in the Process History File). The updated model is output for use with the next process step. The above sequence is then repeated for the next process step. The operation of the Structure Simulator is summarized

Table 1: Summary of Structure Simulator Operation

1. Determine effect of process step.
   Consult layout information and process modelers, as necessary.

2. Decompose process effect into primitive construction operators.

3. Modify solid model using primitive operators.

4. Output results
   Geometry information $\Rightarrow$ Neutral File
   Material information $\Rightarrow$ History File

5. Repeat steps 1–4 for next process step.

in Table 1.

Modification of the solid model must be implemented to ensure that the resulting model is physically valid (i.e., describing a reasonable approximation to the actual structure). Without careful attention to the robustness of modification algorithms, invalid solid models could result (e.g. an unphysical topology such as a Klein bottle, or two objects occupying the same place). Implementation of a robust modification algorithm is considerably simplified if the modification is done using a small set of robust primitive construction operators. Hence, the two step modification procedure discussed above. For micromechanical design, the following primitive operators constitute a useful minimal set: film deposition and growth, film etching, impurity introduction and movement, and wafer joining. These primitives are combined using selection operators which restrict the operation of primitive operator to a certain region of the solid model. Selection may be done on the basis of layout (masking) or material type. For initial implementation, we have selected a restricted subset of operators: conformal deposition, and masked etching. These two primitives provide significant geometry flexibility, and permit the simulation of many interesting micromechanical systems.

The Structure Simulator must output the final structure to the Mechanical CAD subsystem for use in mechanical simulation. Therefore, the solid model representation used by Structure Simulator must be compatible with the one used by our Mechanical CAD, in this case PATRAN. Rather than convert from one solid model representation to another, the Structure Simulator uses the same solid model representation as PATRAN, which is analytic solid modeling (ASM). In ASM a solid is represented as an assemblage of non-overlapping hyperpatches[14]. A hyperpatch is essentially a distorted cube and is generated by mapping a unit cube in parametric $(\xi_1, \xi_2, \xi_3)$ space into the model $(x, y, z)$ space using a tricubic function. In general, a hyperpatch is defined by,

$$Z(\xi_1, \xi_2, \xi_3) = \sum_{k=1}^{4} \sum_{j=1}^{4} \sum_{i=1}^{4} S_{ijk} \xi_1^{4-i} \xi_2^{4-j} \xi_3^{4-k} \qquad (1)$$

where $Z$ is a vector in the model space and the $\{S_{ijk}\}$ are arbitrary vector coefficients. The advantages of ASM are that it allows modeling of the solid interior as well as the boundary and, due to the cubic mapping, can model curved surfaces. This is at the cost, however, of increased model complexity. In a boundary representation solid model, a continuous layer can

be represented as a single entity with a complicated boundary. In ASM, this layer must be broken up into an assemblage of hyperpatches.

OYSTER[4] functions like our Structure Simulator, in that it combines layout and process information to produce a 3-D solid model. There are two basic differences in design between it and the Structure Simulator presented here. As mentioned above, OYSTER's process description, PDL, describes the process *effects*, not the process *steps* which would be used to actually fabricate the device. Secondly, OYSTER is built upon a polyhedral solid modeler and cannot model curved surfaces directly Instead, a curved surface must be broken into facets. This could pose later difficulties in mechanical analysis due to artificial stress concentrations at facet corners.

Despite these differences, OYSTER is similar in many ways to the Structure Simulator presented a here. It works on step-by-step basis following procedure similar to that given in Table 1 (except for omitting Step 1 as described above). OYSTER also modifies the solid model using combination of robust primitive operators. In this case, the primitive operations are solid boolean operations (provided by the underlying solid modeler). The are combined into "cumulative translation sweep" operators to do the actual modifications[15].

Because of the current incomplete and elementary implementation of primitive construction operators in our Structure Simulator, OYSTER, at present, has a far superior modeling capability.

## 4   Example

The specific process considered in this example is. a 4-inch (100) silicon wafer is thermally oxidized at 1030°C in wet ambient for four hours. This grows just over 1 $\mu$m of silicon dioxide on both sides of the wafer. The oxide is removed from the back and a rectangular hole is optionally patterned and etched in the front oxide of each 10mm-square chip. The wafer is then diced into individual chips. Due to the compressive residual stress in the oxide layer the chip bends. This bending is less pronounced as larger holes are etched in the oxide.

The Structure Simulator is used to create a solid model of the chip in PATRAN Neutral File format. Currently, the Structure Simulator does not read the process sequence directly from the PFR Instead, the program queries the user directly for the necessary information. The process history information is maintained manually by the user.

The program begins by asking for information about the starting wafer state (thickness, orientation, and chip dimensions). The program then constructs a linear hyperpatch (see Fig. 2(a)) whose thickness is the wafer thickness and whose length and width is given by the chip dimensions. The oxidation conditions are specified by temperature, time, and ambient. Using this information and the starting wafer state the program constructs a SUPREM-III input file, and runs SUPREM-III to calculate the oxide thickness that would be grown Two linear hyperpatches of this thickness are attached to the front and back sides of the starting wafer to represent the oxide. The thickness of the starting wafer is also thinned at this time based on the SUPREM-III simulation results. This implements the planar growth operator. The result after this process step is shown in Fig. 2(b). Finally, the masked oxide etch is specified by the OCT cell name and layer to be used. (Complete etching of all unmasked oxide is assumed.) The program queries the OCT database and extracts the mask geometry of the desired layer. Currently, only a single layer with a single rectangular
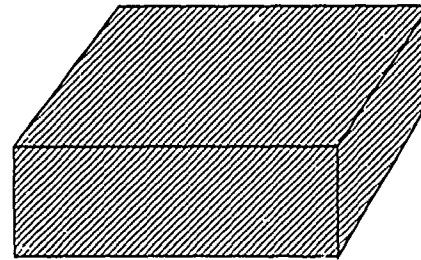


(a)
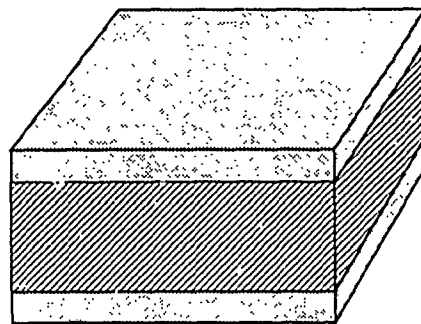


(b)
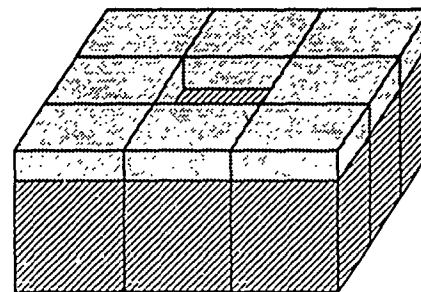


(c)

Figure 2   Three-dimensional solid models of the example process.
(a) starting wafer  (b) after application of planar growth operator (oxidation) (c) after application of masked etch operator
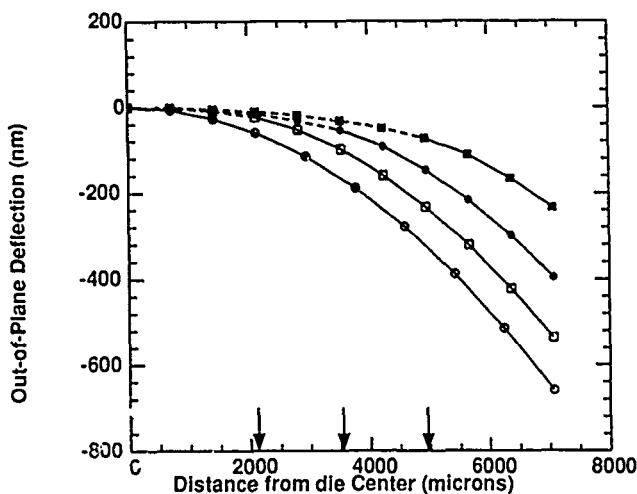
Figure 3. FEM results for intact oxide, 3mm-, 5mm-, 7mm-square holes ($\circ$,$\square$,$\bullet$,$\blacksquare$ respectively) along the diagonal from chip center to chip corner. Solid lines (—) are used for the area under the unetched oxide; dashed lines(- -) for the etched oxide. Arrows ($\downarrow$) show distance from center of oxide hole edges.



Figure 4. Comparison of Stoney Equation (—) with FEM results ($\circ$) for intact oxide

box is supported. To do the etch, the hyperpatch representing the back oxide layer is deleted, and the front oxide hyperpatch is divided into a 3x3 grid. The center hyperpatch of this grid is then deleted, thus implementing a masked removal operator. The final structure is shown in Fig. 2(c). In this example undercutting during the etch has been ignored — the sidewalls are vertical and the corners square. This is sufficient for deformation analysis but not for stress analysis. Curved sidewalls could be included in the solid model by replacing the linear hyperpatches with more elaborate ones.

The resulting structure (stored in PATRAN Neutral File format) is used directly by PATRAN. Using the PATRAN the structure is discretized (800 20-noded elements are used for the silicon and 400 20-sided elements are used for the unetched oxide layer) and appropriate boundary conditions are applied. Material properties are associated with each layer, in this case, by reading from a pre-constructed data file containing information on silicon and silicon dioxide. A thermal cooling load of $-1000°C$ was applied to the entire model, generating a residual stress due to the mismatched coefficients of thermal expansion. *Intrinsic* residual stress was ignored in this analysis.

PATRAN is then used to optimize the FEM model and create an ABAQUS input file. ABAQUS is then run to do the actual analysis. The analysis results are translated into a PATRAN compatible form through an ABAQUS-to-PATRAN translator. PATRAN is then used to graphically display and examine the final results.

Calculated deformations along a diagonal form chip center to chip corner for the unetched oxide layer, and layers with 3mm-, 5mm-, and 7mm-square holes are shown in Fig. 3. As can be seen in the figure, as the hole size increases the amount of deflection decreases. Also, the chip curvature where the top oxide has been removed is markedly shallower than underneath the unetched oxide.

For the case of unetched oxide, film stress is related to out-of-plane bending of the substrate by the Stoney equation[16].

$$\sigma_f = \frac{4}{3} \frac{E_s}{1 - \nu_s} \frac{t_s^2}{t_f} \frac{z}{L^2},$$ (2)

where $\sigma_f$, $E_s$, $\nu_s$ are the film stress, substrate Young's modulus, and substrate Poisson's ratio; $t_s$ and $t_f$ are the substrate and
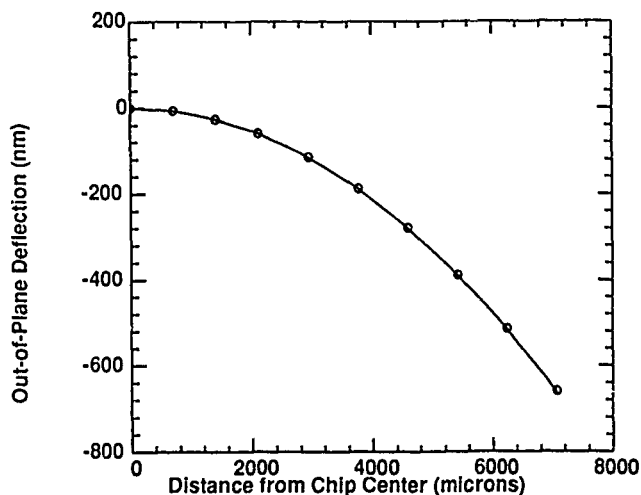
film thicknesses; and $z$ and $L$ are the out-of-plane deflection and length over which the deflection is measured. As can be seen in Fig. 4, analytic results given by the Stoney Equation are almost indistinguishable from the simulated deformation of the unetched oxide layer.

It should be noted that as the structure complexity increases, computer resources are rapidly exhausted during the mechanical analysis. For the relatively coarse discretization used here, 7 hrs of CPU time on a Sun-4/200 and 160 MBytes of scratch disk space were required.

## 5  Conclusion

We have demonstrated the automatic construction of a 3-D solid model from a specification of mask layout and process sequence, the process sequence, the process sequence consisting of planar deposition and masked etch. The generated 3-D solid model is directly usable by our mechanical CAD software and was used to do a deformation analysis of the structure.

## Acknowledgments

## References

[1] S. D. Senturia, "Microfabricated structures for the measurement of mechanical properties and adhesion of thin films," in *Transducers '87*, (Tokyo), pp. 11-16, 1987.

[2] S. D. Senturia, "Can we design microrobotic devices without knowing the mechanical properties of materials?," in *IEEE MicroRobotics and Teleoperators Workshop*, (Hyannis, MA), 1987.

[3] F. Maseeh, R. M. Harris, and S. D. Senturia, "A CAD architecture for microelectromechanical systems," in *Proc. IEEE Micro Electro Mechanical Systems*, (Napa Valley, CA), pp. 44–49, 1990.

[4] G. M. Koppelman, "OYSTER, a three-dimensional structural simulator for microelectromechanical design," *Sensors and Actuators*, vol. 20, pp. 179-185, Nov. 1989.

[5] M. B. McIlrath and D. S. Boning, "Integrating semiconductor process design and manufacture using a unified process flow representation," in *Proc. 1990 Intl. Conf. Computer Integrated Manufacturing*, May 1990.

[6] PDA Engineering, Costa Mesa, CA.

[7] Hibbit, Karlsson, and Sorensen, Inc., Providence, R.I.

[8] L. Conway and C. Mead, *Introduction to VLSI Systems*, pp. 115–127. Reading, MA: Addison-Wesley, 1980.

[9] G. C. Billingsley, "Program reference for KIC," Tech. Rep. UCB/ERL M83/62, U. C. Berkeley, Oct. 1983.

[10] C. P. Ho, J. D. Plummer, S. E. Hansen, and R. W. Dutton, "VLSI process modeling SUPREM-III," *IEEE Trans. Electron Dev.*, vol. ED-30, pp. 1438–1452, Nov. 1983.

[11] W. G. Oldham, S. N. Nandgaonkar, A. R. Neureurather, and M. M. O'Toole, "A general simulator for VLSI lithography and etching process. Part I — application to projection lithography," *IEEE Trans. Electron Dev.*, vol. ED-26, pp. 717–722, Apr. 1979.

[12] W. G. Oldham, A. R. Neureurather, C. Sung, J. L. Reynolds, S. N. Nandgaonkar, and M. M. O'Toole, "A general simulator for VLSI lithography and etching process: Part II — application to deposition and etching," *IEEE Trans. Electron Dev.*, vol. ED-27, pp. 1455–1559, Aug. 1980.

[13] D. S. Harrison, P. Moore, R. L. Spickelmeyer, and A. R. Newton, "Data management and graphics editing in the berkeley design environment," in *Technical Digest of ICCAD-86*, (Santa Clara, CA), pp. 20–24, Nov. 1986.

[14] M. S. Casale and E. L. Stanton, "An overview of analytic solid modeling," *IEEE Computer Graphics and Applications*, vol. 5, no. 2, pp. 45–56, 1985.

[15] R. C. Evans, G. Koppelman, and V. T. Rajan, "Shaping geometric objects by cumulative translational sweeps," *IBM J. Res. Dev.*, vol. 31, pp. 347–360, 1987.

[16] G. G. Stoney, "The tension of metallic films deposited by electroiysis," *Proc. R. Soc. London, Ser. A*, vol. 82, p. 172, 1909.

# STABILITY OF IRIDIUM OXIDE FILMS IN
# HIGH TEMPERATURE, 200-250°, SOLUTIONS

Kenneth Kreider and Michael Tarlov

Center for Chemical Technology
National Institute of Standards and Technology
Gaithersburg, MD 20899

## ABSTRACT

Iridium oxide films have been investigated as pH sensing electrodes. They have a Nernstian response, approximately 58 mV per pH at room temperature, under a wide range of conditions. The advantages of these solid state electrodes include their ruggedness, small size, high voltage/low impedance output, and the low cost of fabrication. They also have high temperature capability and have been considered for geothermal applications.

This study investigated the stability of reactively sputtered iridium oxide films on alumina substrates on exposure to pH 4, 7, and 10 solutions at 200°C (15 bar) and 250°C (40 bar). The exposures were made in a teflon lined bomb. The results of over 50 tests, cycling pH between 2 to 10 before and after exposure, are discussed. Changes of the E°, the formed potential, and the slope were most severe under acidic conditions at 250°C. The most severe conditions also caused film breakdown and loss of adhesion to the substrate. The incorporation of ion assisted deposition in the fabrication of the films lead to improved adherence under the most severe conditions of the exposure. The results indicated similar behavior with films deposited at 500K (crystalline) and those deposited at room temperature (amorphous).

## INTRODUCTION

Iridium oxide as a thin sputtered film is being considered as a pH electrode for high temperature (above 100°C) solutions. The desirability of the sputtered iridium oxide film (SIROF) for measurements in harsh environments has been described by Papeschi, et al. [1], Lauks, et al. [2], and by Dietz, et al. [3]. The glass electrode, which is the standard pH measurement instrument below 100°C becomes less stable at higher temperatures and the alkali interference problem is also aggravated. We at NIST have reported previously on SIROF sensors [4-6] fabricated in-house and are currently investigating their use for pH measurement in nuclear repositories. In this paper, we describe the results of testing the pH response of SIROF electrodes after exposure in pH 4, 7, and 10 solutions in a pressurized bomb at 200° and 250°C.

## EXPERIMENTAL

Iridium was sputtered from a pure (99.99%) iridium 5 cm diameter target with a 13.5 MHz rf planar magnetron source using approximately 0.40 Pa total pressure and a 10 cm target to substrate distance. Argon and oxygen were mixed in a 1:1 ratio to oxidize the growing film. The $1 \mu$ m thick depositions were made on alumina circuit board at 30-40°C and at 240°C.

The plan of testing was to compare the electrochemical potential of the films before and after exposure to deionized water, pH 4, 7, and 10 at 200°C (15.5 bar or 225 psi) and 250°C (40 bar or 600 psi); actually the higher temperature was 245°$\pm$5°C. This was accomplished simply by placing the bomb in a temperature controlled furnace for 20 (or 70) hours. High temperature exposure was made in the teflon lined bomb using potassium hydrogen phthalate (pH 4.01), disodium hydrogen phosphate-potassium dihydrogen phosphate (pH 7), sodium hydrogen carbonate-sodium carbonate (pH 10.01) and deionized water as test solutions for the iridium oxide.

The pH response testing of the SIROF electrodes consisted of a measurement of electrochemical potential versus a single junction Ag/AgCl reference electrode (+0.175 vs NHE) and compared to the pH as measured with a standard glass electrode. The electrochemical tests were performed in a stirred, air saturated, buffer solution (21-22°C) consisting of 0.01M acetic, phosphoric, and boric acids and 0.1M potassium nitrate. Titration was performed by first adding sodium hydroxide to the pH 2-3 buffer solution following by nitric acid additions to lower the pH.

## RESULTS

The twenty-one sample electrodes from three fabrication runs were tested before and after exposure in the teflon lined bomb. The results of the pH testing were plotted as in FIG 1 and 2 with the electrode potential versus pH. The first series of three electrodes were exposed to deionized water for 24, 500, and 1400 hours. The results of these tests did not indicate a consistent pattern of change and the spread in values of slope and formal potential was similar for exposed samples to the fresh samples. Figure 1 shows typical results of these tests after exposure. Ion assisted deposition improved the adherence for the long exposures

The next series of tests also included sample electrodes from three batches. Each sample was tested for pH response from approximately 2-12 and returned to pH 2 with $HNO_3$ when fresh and after subsequent exposure to pH 4, 7, and 10 at 200°C. All response curves were linear (correlation coefficient 0.999-1.000) indicating a Nernstian response and the slope of the E vs pH was not systematically affected by the 200°C exposure for 20 hours. A typical response curve after exposure to pH 4, 7, and 10 at 200°C is given in FIG 2. The open squares are the return from pH 11.5 to 2.5. Each electrode appeared to have a higher slope and Eo if it was exposed to air or heated in air and a lower potential after longer exposures. Exposure to the pH 7 solution for 70 hours did, however, lead to results with lower slopes and formal potentials. The solution contained $Na_2HPO_4$ and $KH_2PO_4$.

A third series of tests included exposures to pH 10 at 200°C for 20 hours followed by exposure at 240°C for 20 hours. Although no degradation was evident at 200°C, a significant decrease in slope (from 54 mV/pH to 45 mV/pH) was measured in all three batches. The electrodes exposed to 245°C at pH 7 and in deionized water were lost as the film came off the $Al_2O_3$ substrate as a powder and the liquid was a dark indigo blue indicating $Ir(OH)_4$. Techniques for improved adherence including ion assisted deposition did not solve the problem and the use of the SIROF at 250°C does not look promising.

To compare crystalline SIROF with the above discussed amorphous films [4] electrodes were deposited at 240°C under the same conditions. Their rutile structure was confirmed by x-ray diffraction. Exposure to 200°C at pH 4, 7, and 10 in the pressurized bomb yielded similar results to the amorphous films; the slopes were not changed substantially but the formal potential, Eo, was diminished.

## CONCLUSIONS

In summary, we have examined the sensitivity of iridium oxide thin films on alumina substrates to high temperature (200, 245°C), high pressure (up to 40 bar, 600 psi) acidic and basic solutions. Twenty-one sample coupons were electrochemically tested through titrations (pH 2-11) before and after several exposures. Under the most severe conditions at pH 7 and below and at 245°C the film is lost from the $Al_2O_3$. This may be a substrate or a film problem but the films were less stable on a silicon wafer substrate. The electrochemical potential was not strongly affected by exposures of 20 hours at pH 4, 7, or 10 at 200°C. However, both the pre-exponential factor and the electrochemical-pH slope appeared to drop after 70 hours (200°C) at pH 7 in $Na_2HPO_4$ and $KH_2PO_4$. It should be also noted that the scatter in the data for the electrochemical potential as previously reported [5] was a problem compared to a glass electrode.

## ACKNOWLEDGMENTS

## REFERENCES

1. Papeschi, G., S. Bordi, C. Beni and L. Ventura, Biochim. Biophys. Acta, 453, 192 (1976).

2. Lauks, I., M. F. Yueu, and T. Dietz, Sensors and Actuators 4, 375 (1983).

3. Dietz, T., and K. G. Kreider, Review of Materials for pH Sensing for Nuclear Waste Containment, National Bureau of Standards Interagency Report 85-3237 (U.S. GPO Washington, DC, 1985).

4. Kreider, K. G., J. Vac. Sci. Tech. A4 (3), 606, (1986).

5. Kreider, K. G., S. Semancik, and J. W. Erickson, Transducers '87 International Conference on Solid State Sensors and Actuators, Tokyo 1987, IEEE, New York, NY.

6. Tarlov, M. J., K. G. Kreider, S. Semancik, and P. Huang, "pH Sensors Based on Iridium Oxide". NUREG/CR-5484 U.S. Nuclear Regulatory Commission, Washington, DC, 1990.
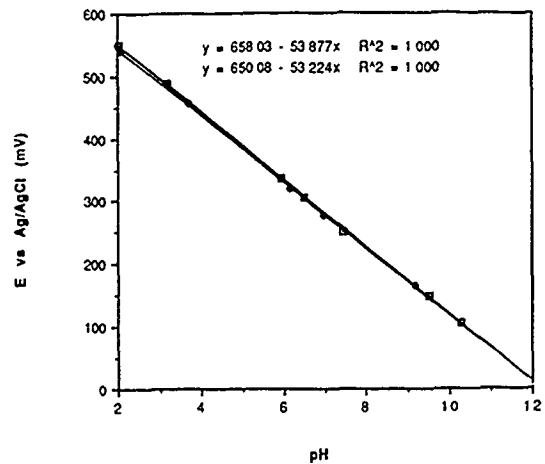
FIG 1  Electrode potential versus pH for SIROF after 1400 hr at 200°C (15.5 bar) in deionized water. Open squares are for increasing pH.
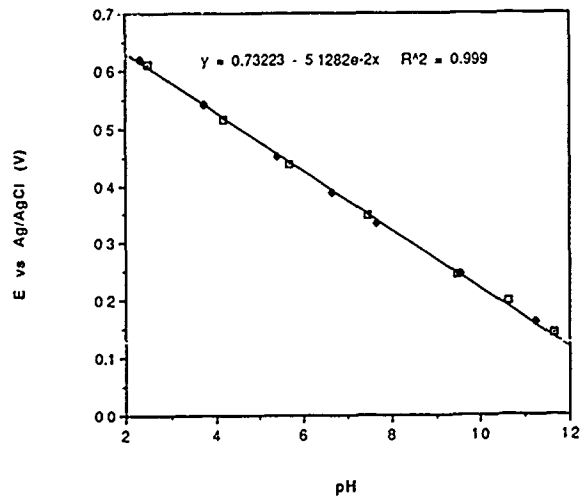


FIG 2  Electrode potential versus pH for SIROF after 20 hr at 200°C (15.5 bar) pH 4 plus 20 hrs at pH 7 plus 20 hrs at pH 10. Open squares are for increasing pH.

# SQUEEZE-FILM DAMPING IN SOLID-STATE ACCELEROMETERS

James B. Starr

Solid State Electronics Center
Honeywell, Inc., Plymouth, MN.

## ABSTRACT

Damping of proof mass oscillations is a fundamental consideration in the design of accelerometers. The introduction of silicon micro-machining into accelerometer technology facilitates very thin fluid layers that provide sufficient damping through viscous dissipation in a gas. Thicker layers require a liquid for sufficient damping. Because the viscosity in liquids is in general more temperature dependent than in gases, the smaller micromachined devices can exhibit less dependence of damping on temperature.

In the midst of these encouraging developments, the accelerometer designer is faced with three fundamental challenges. First, one must be aware of both the nonlinear and the compressible behavior of gas films that can lead to undesirable accelerometer performance, not to mention inadequate damping. Second, one must contend with other than relatively simple geometries in predicting damping coefficients. Third, one may approach the limits of continuum flow in very thin gas layers. Fortunately, there exist in the literature solutions for squeeze-film behavior that can be consulted when designing accelerometers. In addition, finite element techniques are readily available that facilitate damping predictions for other than simple geometries. In this paper criteria are set forth that are aimed at alerting the designer to the onset of nonlinear and compressible effects in accelerometers. A finite element technique is presented that essentially solves Reynold's equation for small displacements and "squeeze" numbers by analogy with heat conduction in a solid with internal heat generation. Examples are presented that show film pressure profiles generated for complex geometries with non-uniform film thicknesses and squeeze velocities. A technique for mitigation of compressibility effects through film pressurization is suggested.

## INTRODUCTION

Sensors that accurately replicate accelerations of mechanical systems require some mechanism for energy dissipation, i.e. a damper. An accelerometer is the mechanical equivalent of a spring-mass system. As such, its response is frequency dependent. In particular, without damping, the output signal will increase significantly with frequency. Indeed, at its resonant frequency, very large amplitudes may produce the electrical consequence of signal saturation or the mechanical consequence of structural failure. Silicon accelerometers are especially susceptible to these consequences by virtue of the highly elastic behavior of silicon at typical operating temperatures.

The extensive use of silicon technology in accelerometers in recent years has introduced accelerometer products that offer significantly improved performance and size for a given cost. The inherent mechanical stability of silicon along with stable force transduction technologies has resulted in batch-producibility of precision accelerometers. Piezoresistive, capacitive and beam-resonant approaches have all been shown to produce excellent accelerometer products.

The goodness of an accelerometer product does depend on considerations that are not an inherent part of either silicon material or transduction technologies. As previously mentioned, damping is a key consideration. It is needed to achieve a flat response over a wide range of frequencies. Fortunately, micromachining has been one area in which silicon sensor technology has grown significantly so as to enhance provisions for damping. It has facilitated the forming of very thin gas layers that serve as an excellent means of viscous dissipation, even when the fluid medium is a gas. This means that the accelerometer designer often does not have to resort to the use of liquid fill fluids to achieve adequate damping. The extent of damping becomes less temperature dependent because of the smaller relative change in viscosity of a gas as compared with a liquid. A simple damping configuration is illustrated in Figure 1. Damping forces result from the buildup of pressure that is needed to produce lateral motion of the viscous fluid.
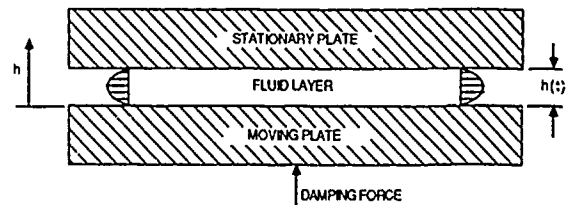


Figure 1. Squeeze-Film Damper Schematic

Full realization of the advantages of gas-film damping does require consideration of several factors affecting performances. Indeed, the designer must consider the following.

-The fluid behavior must be governed by viscous forces that are larger relative to momentum changes.
-Dimensions must be large enough to ensure the gas film exhibits continuum behavior.
-For certain conditions and geometries, damping may be adversely affected by compressibility effects.
-Displacements of the gas film may be large enough to introduce nonlinear behavior and signal distortions into the accelerometer
-Geometries of the damper may not necessarily be so simple as to utilize existing solutions for gas film pressures.

The following discussion addresses each of these factors. Results are presented in terms of concise formulations that should serve the accelerometer designer in his attempt to produce an excellent product.

## REYNOLDS NUMBER CONSIDERATIONS

The behavior of squeeze films is in general governed by both viscous and inertial effects within the fluid. However, for the very small geometries encountered in silicon micromachined devices, inertial effects can be ignored. In such a case the behavior of the fluid is governed by

$$\partial[(\rho h^3/\mu)\partial P/\partial x]/\partial x + \partial[(\rho h^3/\mu)\partial P/\partial y]/\partial y = 12\partial(\rho h)/\partial t \quad (1)$$

Where P is the film pressure, $\rho$ is density, $\mu$ is viscosity, h is

film thickness, and x and y are spatial coordinates. This expression is derived from the much more complicated Navier Stokes equation under the assumption of small Reynolds numbers. According to Langlois in Reference (1), the specific condition for validity of equation (1) is

$$\omega h^2 \rho / \mu \ll 1.0 \qquad (2)$$

where $\omega$ is the oscillation frequency of the damper plate.This condition is readily satisfied in typical silicon accelerometers. For example, an air-filled accelerometer with a squeeze-film thickness of 25 microns, operating at a frequency of 1.0 kHz, would have a Reynolds number defined by equation (2) of only 0.25.

## CONTINUUM LIMITS

The small film geometries producible by micromachining of silicon may result in a condition where the fluid film cannot be treated as a continuum. As film thickness is reduced, eventually a condition is reached where the mean-free-path of molecules within the gas becomes significant with respect to the thickness of the film. In such a case, the effectiveness of damping is reduced due to a slip-flow condition existing at the boundaries of the film. For example, the mean-free-path for air at 25°C and 1.0 atm is about .09 microns. Slip-conditions may be encountered when the mean-free-path is about one percent of the film thickness. Thus if a film thickness of under 9 microns is used, the possibility of slip-flow should be considered. Such effects may be mitigated by using a high-viscosity gas of larger molecular weight, such as neon, or by pressurizing the gas.

## SIMPLIFIED ANALYSES

The task of predicting damping coefficients for accelerometers is eased under certain conditions. First of all, thin gas films may be realistically assumed to be isothermal. This is because of the usually small conducting paths within the gas to bounding walls that have a relatively high heat capacity. For isothermal films, density is proportional to absolute pressure. Hence, all densities appearing in equation (1) can be replaced by pressure. This condition is in reality a very modest limitation on the analysis. Three additional simplifying assumptions are that pressure variations within the film are small relative to the average absolute pressure, the film thickness is uniform, and displacements of the moving surface are small relative to the film thickness. Subsequent sections of this paper will discuss conditions under which two of these assumptions are unrealistic. However, where they are realistic, there results a damping characteristic that is linear and characteristic of what is needed for precision accelerometers The isothermal, small-pressure-variation and small-displacement assumptions applied to equation (1) result in the expression

$$\partial^2 P_*/\partial x^2 + \partial^2 P_*/\partial y^2 = 12\mu(\partial h/\partial t)/h_o^3 \qquad (3)$$

where $P_*$ is the pressure departure within the film from the ambient value, and $h_o$ is the nominal film thickness. Equation (3) is simply a Poisson's equation that can be readily solved for simple geometries. The solution provides the pressure distribution within the film which can usually be integrated to provide an expression for the damping force. Two expressions for damping force are provided herewith for plates that move in a uniform manner transverse to the film plane. For a circular disk:

$$F_D = [3\pi\mu R^4 (\partial h/\partial t)]/(2h_o^3) \qquad (4)$$

where R is the radius of the disk. For a rectangular plate of width 2W and length 2L,

$$F_D = 16f(W/L)W^3 L (\partial h/\partial t)/h_o^3 \qquad (5)$$

where f(W/L) is given in Figure 2. The limiting case of W/L→0 would, of course, correspond to a very long narrow strip.
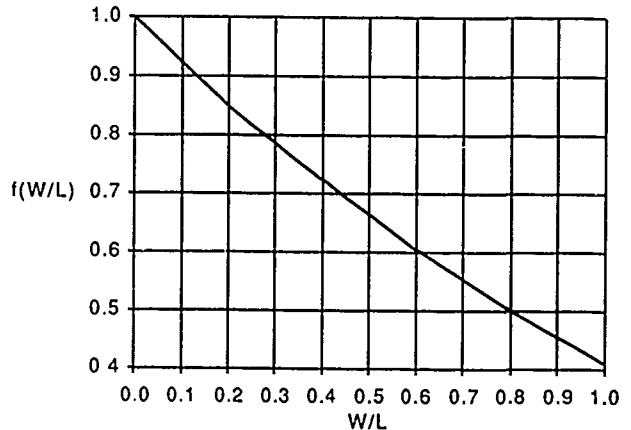


Figure 2   Effect of Aspect Ratio on Damping Produced by a Rectangular Plate

## LARGE DISPLACEMENTS

The characteristics of squeeze film damping will be altered when displacements become significant with respect to the nominal film thickness. Such a phenomenon is suggested by equation (3) even though it is derived for small displacements. The term on the right-hand side includes the inverse of film thickness cubed. One should therefore expect that as amplitude increases, there will be a strong tendency for the damping coefficient to increase. The extent of this increase can be determined from equations derived by Sadd and Stiffler in Reference 2. Damping under the limitation of small density changes but with larger displacements can be calculated by simply multiplying expressions such as those in equations (4) and (5) by a displacement function defined as

$$f_D(\varepsilon) = (1 - \varepsilon)^{-3/2} \qquad (6)$$

where $\varepsilon$ is the ratio of plate displacement to nominal film thickness. A plot of this expression is shown in Figure 3. Note the very significant increase in damping with displacement, it only takes a relative displacement of 0.25 to cause a 10 percent increase in damping. It is also important to note that large displacements will result in a distorted accelerometer response to a sinusoidal input acceleration.
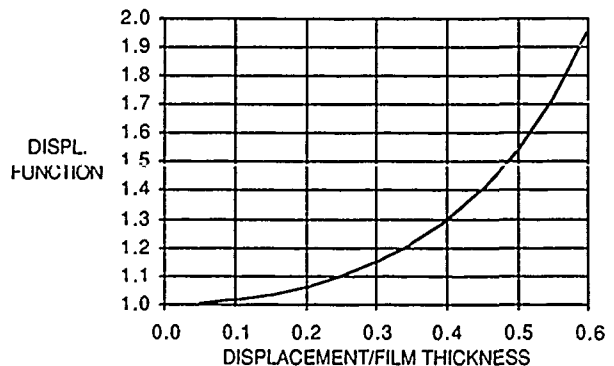


Figure 3.   Effect of Displacement on Damping

## COMPRESSIBILITY EFFECTS

Compressibility effects in squeeze-films have been addressed in both References (1) and (2). Use is made of a squeeze number, $\sigma$, defined by

$$\sigma = 12\mu\omega b^2/(h_o^2 P_a) \qquad (7)$$

where $\mu$ is viscosity, $\omega$ is frequency, b is a characteristic length, $h_o$ is the nominal film thickness, and $P_a$ is the ambient gas pressure. A power series in $\sigma$ is generated so as to satisfy Equation (1) for a uniform film thickness. Langlois showed that for small values of $\sigma$ the film behaved as though the fluids were incompressible. His results show that for a long narrow strip with b defined as its half-width, $\sigma$ values of about 0.2 or less exhibit essentially incompressible behavior. At the other extreme, where $\sigma$ is quite large, the film essentially acts as an air spring and exhibits little behavior desired of an energy dissipator. From these data one concludes that compressibility effects can be neglected by designing the accelerometer so that the squeeze number is much less than 0.2. Sadd and Stiffler in reference (2) solved the Reynolds equation also in terms of a power series in $\sigma$ but with the admission of displacements that are significant relative to the nominal film thickness. Their expressions are useful for quantify effects of both the squeeze number and displacement on air-spring like behavior and on signal rectification. In general, their expressions are concluded to be useful for $\sigma$ values up to about 0.3.

Air-spring behavior increases with $\sigma$ because the fluid film in effect forms a restriction against free outflow from the edges of the film. This effect can be quantified in terms of the relative increase in stiffness in a simple spring-mass representation of the accelerometer. By applying expressions in Reference (2) to a spring-mass system using a spring of stiffness k, there results

$$\Delta k/k = C_k f_k(\varepsilon)(\omega/\omega_n)\zeta\sigma \qquad (8)$$

where $\qquad f_k(\varepsilon) = (1 + 3\varepsilon^2 + 3\varepsilon^4/8)/(1-\varepsilon^2)^3,$

$\omega$ is the resonant frequency and $\zeta$ is the damping relative to the critical level for the nominal spring-mass system. Values of $C_k$ are 0.8 for a long, narrow strip and 0.33 for a circular disk. A plot of $f_k$ is presented in Figure 4. From Equation (8) and Figure 4 we conclude that increases in stiffness due to compressibility are significant functions of both the squeeze parameter $\sigma$ and relative plate displacement $\varepsilon$. In the limit of small displacements, assuming a critically damped accelerometer operating at resonant frequency, the relative increase in stiffness with a disk-shaped damper would be about $\sigma/3$.
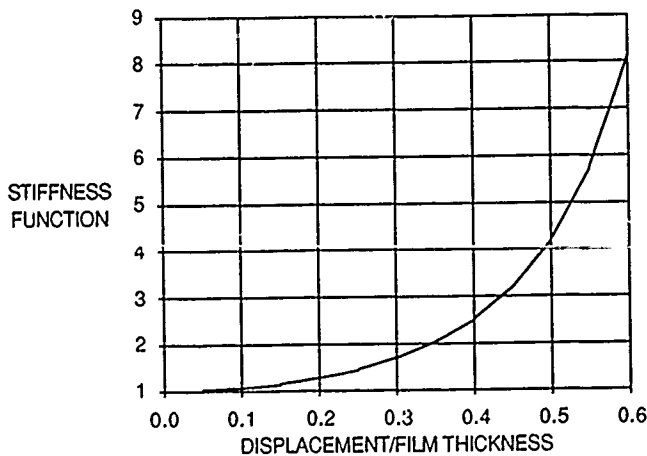


Figure 4. Fluid Stiffness Function

The inherent nonlinear characteristic of a damper may also introduce a rectifying effect into the accelerometer response. This effect exists only in connection with compressibility effects; it disappears when $\sigma$ is zero. Again using expressions from Reference (2), one can derive a relationship between the time-averaged fluid force generated by the damper, designated $F_R$ and the inertial force associated with the accelerometer's proof mass, designated $F_I$. The resulting expression, valid for $(\omega/\omega_n) \ll 1.0$, is

$$F_R/F_I = f_R(\varepsilon)(\omega/\omega_n)\zeta\sigma \qquad (9)$$

where for a disk

$$f_R(\varepsilon) = 5\varepsilon(4 + 3\varepsilon^2)/[48(1 - \varepsilon^2)^3]$$

and for a narrow strip

$$f_R(\varepsilon) = \varepsilon(4 + 3\varepsilon^2)/[4(1 - \varepsilon^2)^3]$$

Plots of these two functions are shown in Figure 5. Note that the rectification effect vanishes for zero relative displacement. This characteristic is very important in the design of precision accelerometers such as might be used in inertial navigational systems. Indeed, for such accelerometers provisions for limiting mass displacement would be required, e.g. through the use of a force rebalance servo system.
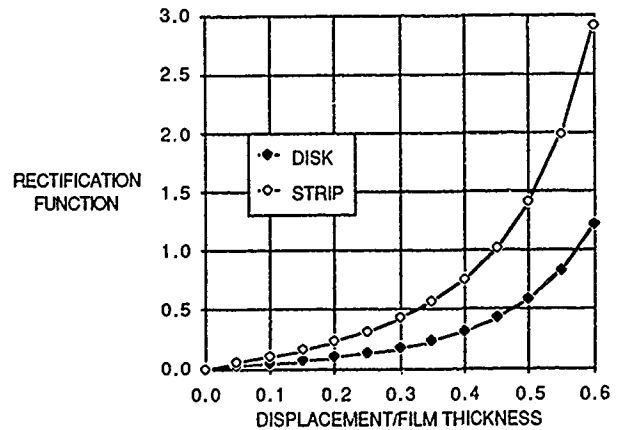


Figure 5. Rectification Function

When faced with the minimization of compressibility effects, it is important note the makeup of the squeeze number. Note that in its denominator is $P_a$, the ambient pressure. This suggests that one could reduce the value of $\sigma$ and therewith the effects of compressibility by increasing the absolute pressure of the fill gas in the accelerometer.

## COMPLEX GEOMETRIES

As previously discussed, the assumptions of incompressible behavior and uniform film thickness result in a governing equation that is in reality a Poissons Equation. As such, one may use existing finite element thermal analysis to generate solutions for film pressure. To that end, the Reynolds Equation may be written as

$$\partial^2\psi/\partial\eta^2 + \partial^2\psi/\partial\xi^2 = \lambda \qquad (10)$$

46

where $\eta$ and $\xi$ are spatial coordinates normalized by a reference length L, and

$$\psi = P \cdot h_0^3 / [12 \, (\partial h / \partial t)_{max} \, \mu L^2]$$

The term $\lambda$ facilitates the introduction of non-uniform motion of the damper as would occur with a hinged device; it is the ratio of the local plate velocity to the maximum plate velocity. Note that $\lambda$ is analogous to internal heat generation per unit volume and $\psi$ is analogous to temperature.

As an example, Figure 6 illustrates a pressure contour plot generated for a damping element. The damper is hinged at the lower edge. A hole within the device was required for routing current to a force rebalance device. With these data available, one can carry out a numerical integration of the pressure field to determine the damping force. It is also possible to determine damping action corresponding to higher order modes of oscillation that may occur. This is accomplished by evaluating damping with the plate hinged at the appropriate nodal axis.

## CONCLUSIONS

In designing an accelerometer that incorporates a squeeze-film damping device, care must be taken to avoid the deleterious effects of fluid compressibility and large displacements. In general, ideal damping will result if both the squeeze number and the relative displacement are kept much less than 1.0. If those conditions are satisfied, one may then use finite element techniques to estimate damping coefficients for other than simple geometries.

## REFERENCES

(1) Langlois, W. E., "Isothermal Squeeze Films", *Quarterly of Applied Mathematics*, vol.XX, no. 2, July, 1962, pp. 131-150.

(2) Sadd, M. H. and A. K. Stiffler, "Squeeze Film Dampers: Amplitude Effects at Low Squeeze Numbers", *Journal of Engineering for Industry*, Transactions of the ASME, vol. 97, Series B, Nov. 4, 1975, pp1366-1370.
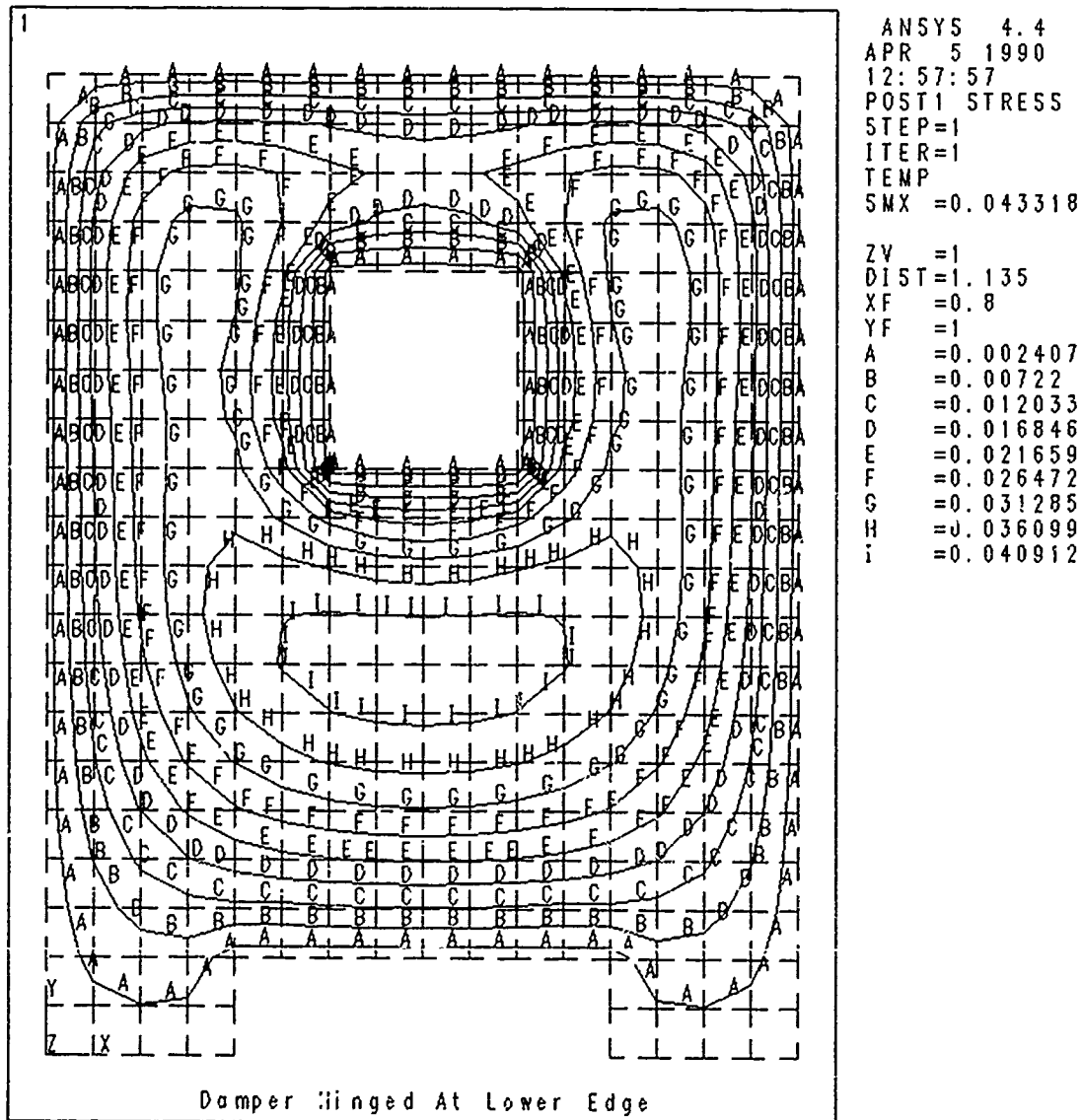
Figure 6. Contour Plots of Dimensionless Pressure ($\psi$) for a Hinged Damper

# POLYSILICON MICROGRIPPER

*Chang-Jin Kim, Albert P. Pisano, Richard S. Muller, and Martin G. Lim**

Berkeley Sensor & Actuator Center
*An NSF/Industry/University Cooperative Research Center*
University of California, Berkeley, California 94720

## ABSTRACT

A polysilicon, electrostatic, comb-drive microgripper has been designed, fabricated, and tested. Its main features are a flexible, cantilever comb-drive arm with a bidirectional actuation scheme and an over-range protector. Three different electromechanical models are developed and, along with fabrication constraints, are employed to design the microgripper and to simulate its performance. Experiments have demonstrated that a gripping range of 10μm can be accommodated with an applied potential of 20V. The motion dependence on drive voltage has been measured and compared with model prediction. The gripper motion is observed to be smooth, stable, and controllable. Measurements were carried up to the maximum of the voltage source (50V).

## 1. INTRODUCTION

Recent developments in micromechanics have led to success in building new types of microactuators [1, 2], elements which are essential building blocks for many micromechanical systems. Microgrippers capable of handling tiny objects have direct applications as end-effectors for currently available "macro"-sized micromanipulators. Mehregany, Gabriel, and Trimmer [3] have fabricated silicon tongs to be actuated by external means. Chen and co-workers [4] demonstrated tungsten microtweezers actuated by a voltage applied across two cantilevers about 2μm apart when fully open. These tweezers had a closure voltage of roughly 150V.

We report a prototype surface micromachined polysilicon microgripper. To avoid frictional effects and the relatively large clearance of micro-joints, the present microgripper is based on flexure structures. An electrostatic comb driver [2] has been modified for the gripper to obtain a reasonably large static movement. The microgripper has been modeled in three different ways. A simple, lumped-beam model has been used to predict the gripping range and force. Testing shows the required driving voltage of about 20V to accommodate a gripper range of 10μm, which is adequate to handle objects of the size range of micrometers, including most bacteria and animal cells. The measured gripper motion as a function of drive voltage is reported along with model predictions.

* Present address: Xerox Palo Alto Research Center
Palo Alto, California

## 2. DESIGN

The microgripper consists of two movable gripper tines driven by three electrostatic comb drivers shown in Fig. 1. The central gripper tines, labeled *drive arm* on Fig. 1, extend to sections labeled *extension arm* and *gripper tip*. The two electrode drivers, labeled *open driver* and *close driver* are attached to the substrate and actuate the gripper by bending the drive arm through the electrostatic fringe-field force of the comb-actuation mechanism. Lengths of drive and extension arms, $L_{dr}$ and $L_{ext}$, are 400 and 100μm, respectively.
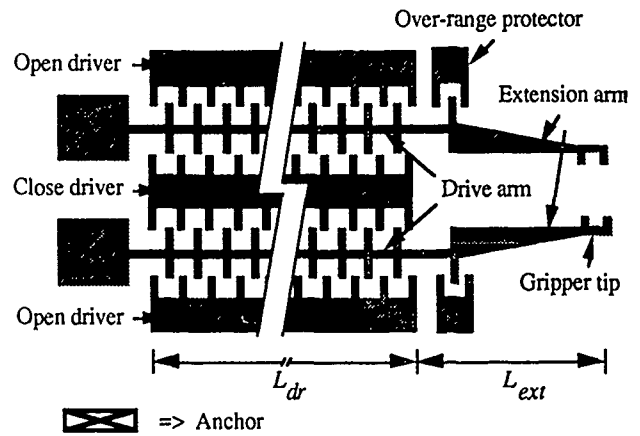


Fig. 1 Schematic of the microgripper

Electrostatic comb-drive actuation has been selected because of its stable drive over relatively long distances. It stands in contrast to actuation by fields applied across a variable-gap capacitor which results in unstable drive available over only very short distances (about 2μm, in practice). The generated fringe-field force by comb actuators, however, is generally smaller than the direct force across parallel-plate capacitors. To increase force and obtain the maximum displacement with this comb-drive mechanism, a cantilever with a large number of comb teeth has been designed.

The provision of separate open- and close-drivers doubles the effective gripping range for a given maximum voltage. The extension arm separates manipulated objects from the electric field in the comb area and also increases the gripping range. To reduce deformation while gripping objects, the extension arm is designed to be rigid compared to the flexible-drive arm. The over-range protector is at the same potential as the gripper tine. If a high voltage is applied during operation, this protector prevents contact between the drive arm and open drivers and thus enables repeatable operation even with electrical overdrive.

Configuration of the interdigitated combs used for the microgripper is shown in Fig. 2. The width of the comb teeth and beam $b$ is 2µm, the gap between the drive-arm comb teeth and driver-comb teeth $d$ is 3µm, and the length of the comb teeth is 10µm. There is no initial tooth overlap.
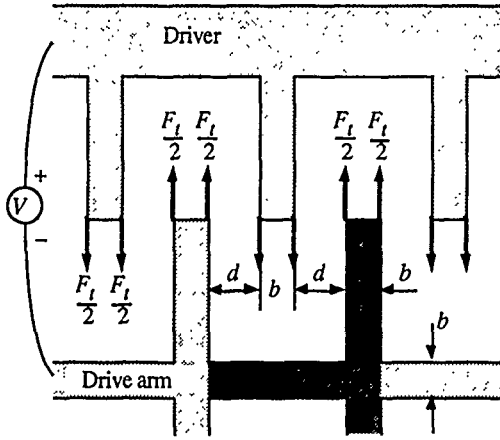


Fig. 2 Comb configuration

## 3. MODELS

Three electromechanical models of the microgripper have been investigated in which (a) deflections of beams are assumed to be very small compared to the beam lengths so that the rotation effect of the beam can be neglected, and (b) lengths of teeth are assumed to be much larger than the beam width which implies rigid-body behavior of the teeth as parts of the beams.

### 3.1 Concentrated-force model

Actuation is due to the fringe-field electrostatic force between the driver and drive-arm teeth. The drive arm is assumed to be a combination of $m$ successive comb elements (Fig. 3). The actuation force along each side of an individual comb tooth is [2]

$$\frac{F_t}{2} = \frac{\varepsilon t}{2d} V^2 \tag{1}$$

where $\varepsilon$ is the permittivity of air, $d$ is the gap between the drive-arm comb teeth and driver-comb teeth, and $t$ is the thickness of the gripper. The pitch of the comb is $p$, and $p'$ is the length of the flexible part of each comb element shown darker in Fig. 2. The gripper-tip displacement $\delta$ is calculated by equations (1) - (6), which are based on the free-body diagram of Fig. 3.
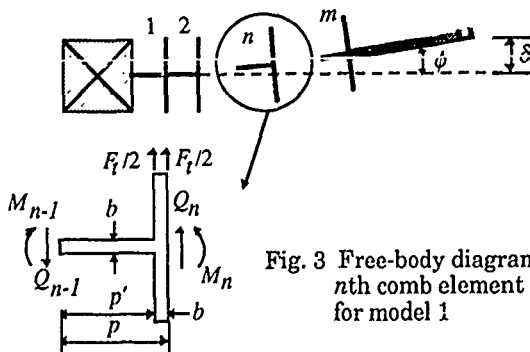


Fig. 3 Free-body diagram of nth comb element for model 1

$$\delta = \sum_{n=1}^{m} \delta_n + \phi_m L_{ext} \tag{2}$$

$$\delta_n = p\phi_{n-1} + b(\phi_n - \phi_{n-1}) + \frac{3bp'^2 + 2p'^3}{6E_Y I} Q_{n-1} + \frac{p'^2}{2E_Y I} M_n - \frac{bp'^2}{4E_Y I} F_t \tag{3}$$

$$\phi_n = \phi_{n-1} + \frac{2bp' + p'^2}{2E_Y I} Q_{n-1} + \frac{p'}{E_Y I} M_n - \frac{bp'}{2E_Y I} F_t \tag{4}$$

$$Q_n = Q_{n-1} - F_t \tag{5}$$

$$M_n = M_{n-1} + \frac{b}{2} F_t - p Q_{n-1} \tag{6}$$

where $n = 1, 2 \ldots m$, and $\delta_0 = \phi_0 = Q_m = M_m = 0$. $E_Y$ is modulus of elasticity, and $I$ is the second moment of inertia of the cross section of the flexible portion of the comb element.

### 3.2 Uniformly distributed-force model

Another model assumes a uniformly distributed load by averaging the fringe-field electrostatic force over the entire drive arm. This assumption should be admissible as long as the drive arm has a dense population of comb teeth. This model is similar to the concentrated-force model but employs a distributed load $w = F_t/p$ as shown in Fig. 4. The gripper-tip displacement $\delta$ is calculated by equations (1), (2), (7)-(10).
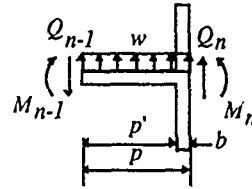


Fig. 4 Free body diagram of nth comb element for model 2

$$\delta_n = p\phi_{n-1} + b(\phi_n - \phi_{n-1}) + \frac{3bp'^2 + 2p'^3}{6E_Y I} Q_{n-1} + \frac{p'^2}{2E_Y I} M_n$$
$$- \frac{6b^2 p'^2 + 8bp'^3 + 3p'^4}{24E_Y I} w \tag{7}$$

$$\phi_n = \phi_{n-1} + \frac{2bp' + p'^2}{2E_Y I} Q_{n-1} + \frac{p'}{E_Y I} M_n - \frac{3b^2 p' + 3bp'^2 + p'^3}{6E_Y I} w \tag{8}$$

$$Q_n = Q_{n-1} - pw \tag{9}$$

$$M_n = M_{n-1} + \frac{p^2}{2} w - p Q_{n-1} \tag{10}$$

### 3.3 Lumped-beam model

A third model averages the compliance of the drive arm. We consider that $p$ is very small so that an infinitesimal analysis can be applied. A comb element under pure bending moment $M$ is shown in Fig. 5.

49
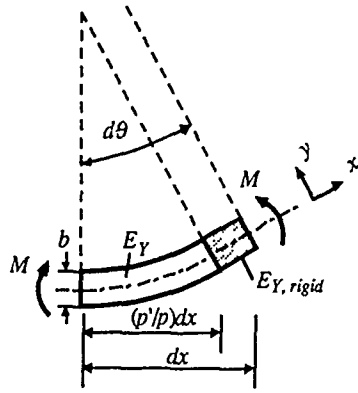
Fig. 5 A comb element under pure bending moment

The average strain along the comb element, $dx$, is (assuming $E_{Y,rigid} \longrightarrow \infty$)

$$\varepsilon_x = \frac{\sigma_x}{E_Y}(\frac{p'}{p})$$

(11)

From geometry, however, the strain is also

$$\varepsilon_x = \frac{d\theta}{dx} y$$

(12)

Moment equilibrium with eqns. (11) and (12) gives

$$M = \int_A y\sigma_x dA = \frac{E_Y}{(p'/p)}\frac{d\theta}{dx}\int_A y^2 dA = \frac{E_Y I}{(p'/p)}\frac{d\theta}{dx}$$

(13)

where $I = tb^3/12$. Eqn. (13) implies that the whole drive arm can be treated as a homogeneous flexible beam with flexural rigidity, $E_Y I/(p/p')$. Using this lumped-beam model, the deflection along the gripper tine is

$$\delta(x) = \frac{wx^2}{24E_Y I(p/p')}(x^2 - 4L_{dr}x + 6L_{dr}^2), \text{ for } 0 \le x \le L_{dr}$$

$$\delta(x) = \frac{wL_{dr}^4}{8E_Y I(p/p')} + \frac{wL_{dr}^3}{6E_Y I(p/p')}(x - L_{dr}), \text{ for } L_{dr} \le x \le L_{dr}+L_{ext}$$

(14)

The comparison of the predictions of these three models in Fig. 6 shows that the lumped-beam model, as simple as it is, generates numerical results that are very similar to the two more complicated models. The results shown in Fig. 6 are for an applied potential of 20V ($F_t = 2.5$nN).
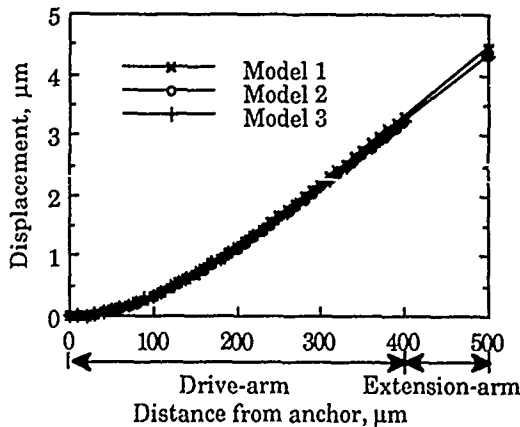


Fig. 6 Comparison of three different models

### 3.4 Gripping force

The gripping force $F_{gr}$ can be estimated by assuming a small gripper deflection and a rigid object. The initial displacement of the gripper tip before reaching the object is $\delta_0$. Using eqn. (14) of the previous lumped-beam model in Fig. 7, we can obtain gripping force as in eqn. (15). We can calculate, for example, that when $\delta_0$ is 2μm a drive voltage of 30V generates 44nN of gripping force. This force is almost 200 times the weight of a silicon cube 10μm on a side, and would be very effective in restraining it.
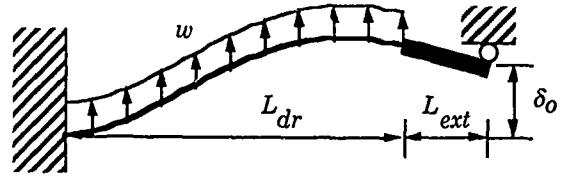


Fig. 7 Model for gripping-force estimation

$$F_{gr} = \frac{\frac{w}{24E_Y I(p/p')}L_{dr}^3(3L_{dr} + 4L_{ext}) - \delta_0}{\frac{L_{dr}}{E_Y I(p/p')}(\frac{L_{dr}^2}{3} + L_{dr}L_{ext} + L_{ext}^2)}$$

(15)

The predicted deformation of the neutral axis of the gripper arm before and after the gripper tip reaches the object is shown in Fig. 8 as a function of the drive voltage.
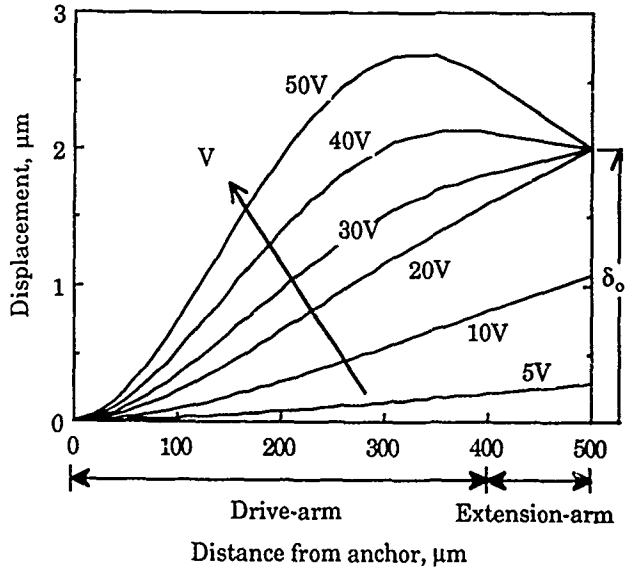


Fig. 8 Simulated neutral-axis deformation of a gripper arm during gripping

### 4. FABRICATION AND TESTING

Fabrication has been based on a four-mask process as described in Lim et al. [5]. A picture of a completed microgripper, including temporary breakaway support beams attached to the necks of the gripper tips, built-in scale next to the gripper tips, and over-range protectors at the left ends of the extension arms, is shown in Fig. 9. Dimples along each gripper

50

tine prevent the gripper from sticking to the substrate during the final step of freeing from the sacrificial PSG layer [1]. A closeup view of the gripper tips along with the built-in scale is shown in Fig. 10. The fabricated microgrippers are 2.2µm thick and have 2.1µm-wide beams and comb teeth.





500 µm
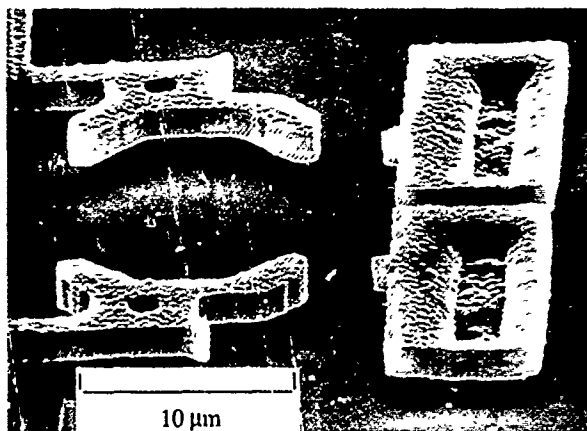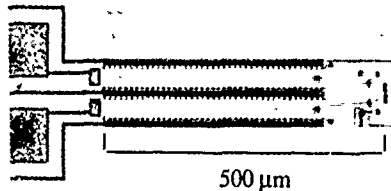
Fig. 10 SEM picture of gripper tip and built-in scale

After freeing the gripper tip from temporary support beams using a probe tip under the microscope, a dc bias was applied between the drive arm and open- or close-driver. The displacements of gripper tips have been measured with the built-in scales to give the results shown in Fig. 11. Also given on Fig. 11 are predictions of the lumped-beam model. The effect of finite beam thickness on the actuation force has been considered through a FEM analysis. The movement of the microgripper is observed to be smooth and stable, and the displacement is repeatable and controllable.
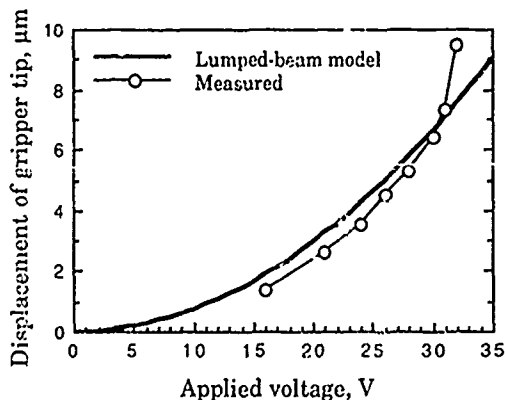


Fig. 11 Displacements of gripper tip

An applied potential of 20V resulted in 2-3µm movements of each gripper tip in both the open and close direction. Defining gripping range as the difference between the maximum-tip separation and minimum-tip separation for a given maximum applied voltage, we obtained about 10µm of gripping range with 20V as the maximum applied potential. In Fig. 11, voltages higher than 20V are to show roughly quadratic dependence of displacement on applied voltage. The assumptions for the models are not valid for large displacements, and the measured displacements deviate from the predicted values at high voltages. For gripper opening, the over-range protector worked as intended, resulting in repeatable gripper motion even after excessive driving voltages. Up to 50V (maximum of power source) were applied, but the over-range protector prevented further displacements at potentials higher than 32V. For gripper closing, higher voltages than that required for complete gripper closure resulted in tine bending comparable to that shown in Fig. 8.

## 5. CONCLUSIONS

In this research, we have demonstrated a polysilicon, electrostatic comb-drive microgripper. The electrostatic comb driver, developed originally for resonators [2], has been modified and used to provide static drive for a microgripper. The microgripper features a flexible, cantilever comb-drive arm with a bidirectional actuation and an over-range protector. Performance of the microgripper under gripping force has been simulated using three different models. Experiments showed that the microgripper has met its design goal of providing gripper-tip movement of several micrometers at drive voltages less than 20V. The gripper motion was observed to be smooth, stable, and controllable.

### REFERENCES

[1] L.-S. Fan, Y.-C. Tai, and R. S. Muller, "IC-processed Electrostatic Micro-motors." *IEEE Int. Electron Devices Meeting,* pp. 666-669, San Francisco, CA, Dec. 1988.

[2] W. C. Tang, T. H. Nguyen, and R. T. Howe, "Laterally Driven Polysilicon Resonant Structures," *Proc. IEEE Micro Electro Mechanical Systems Workshop,* pp. 53-59, Salt Lake City, UT, Feb. 1989.

[3] M. Mehregany, K.J. Gabriel, and W. S. Trimmer, "Integrated Fabrication of Polysilicon Mechanisms," *IEEE Trans. on Electron Devices,* Vol. 35, No. 6, pp. 719-723, June 1988.

[4] L. Y. Chen, Z. L. Zhang, J. J. Yao, D. C. Thomas, and N. C. MacDonald, "Selective Chemical Vapor Deposition of Tungsten for Microdynamic Structures," *Proc. IEEE Micro Electro Mechanical Systems Workshop,* pp. 82-87, Salt Lake City, UT, Feb. 1989.

[5] M. G. Lim, J. C. Chang, D. P. Schultz, R. T. Howe, and R. M. White, "Polysilicon Microstructures for Characterize Static Friction." *Proc. IEEE Micro Electro Mechanical Systems Workshop,* pp. 82-88, Napa Valley, CA, Feb. 1989.

# ISSUES REGARDING THE HIGH RATE AFTERGLOW AND DIRECT PLASMA ETCHING OF SILICON

Andrew M. Hoff

Center for Microelectronics Research

College of Engineering

The University of South Florida

Tampa, Florida 33620

## ABSTRACT

High rate chemical etching of silicon is an important technology to the microfabrication of electronic components. Examples of technologies which could benefit from this technique are via formation, trench or moat etching, and cavity formation. Fluorine containing compounds are the typical precursors in such processes. In purely chemical etching, the process may be optimized by judicial selection of the fluorine precursor species, the method of fluorine species production, appropriate substrate temperature selection, and mass flux of reactant to the chemical reaction zone. This paper presents results of silicon etching processes employing a variety of conditions. Direct plasma treatment was carried out in a parallel plate etch system with 13.56 MHz excitation and substrate temperature in the range of 23°C to 350°C. Afterglow processing was performed using species generated in a microwave-excited plasma tube. These species were allowed to react with the silicon in a remote thermal process zone at temperatures similar to those used in the parallel plate system. The gases utilized in this study consisted of $CF_4$, $CF_4 \cdot O_2$, and $NF_3$. Silicon etch rates as high as 47 $\mu$m min.$^{-1}$ have been achieved with selectivity to $SiO_2$ of 25, using $NF_3$ as the fluorine precursor. The implications of the results with regard to the precursors and the ambient conditions will be discussed for each case. In addition, the use of masking materials such as photoresist, $SiO_2$, and aluminum will be discussed. The high rate etching of silicon presented in this work provides a productive fabrication option in cases where device structures require large amounts of silicon to be removed.

## INTRODUCTION

Silicon, many silicon compounds, as well as many metals may easily be removed via a heterogeneous chemical reaction between the solid and a gas phase containing atomic fluorine [1]. Typically, an electric discharge is used to produce the atomic species. Silicon etch rates, at substrate temperatures near room temperature, vary from a few hundred angstroms per minute to ten thousand angstroms per minute depending on the fluorine precursor used, the concentration of species, and the method of reactant production [1,2]. Although such rates are acceptable for use in the removal of thin films, more productive methods of material removal will be required in the fabrication of structures such as deep trenches or moats, cavities, and through wafer vias.

In purely chemical processes, three factors must be addressed in the optimization of the process. The first is the concentration of reactant exposed to the substrate, second is the temperature of the substrate, and third is the mass flux of reactant to the substrate surface. The first two factors follow from accepted chemical practice. The third must be addressed whenever the chemical reaction at the interface occurs faster than the reactant can reach this surface. These relationships have been previously shown in an examination and optimization of photoresist stripping processes [3]. Proper attention to these factors resulted in an order of magnitude improvement in the productivity of the stripping process.

In this work, chemical etch rates of silicon have been determined in a flowing afterglow of reactant and in a direct plasma etching system. Although the results of these experiments are promising, other factors will require further examination in order to apply high rate etching to device structures.

## EXPERIMENT

In these studies, the silicon substrate material was p-type, <100> orientation, 6 to 10 $\Omega$-cm. Masking materials consisted of either $SiO_2$ or aluminum. These hard masks were used in deference to polymeric materials due to their temperature stability. The oxide films were thermally grown and the aluminum was sputtered. Both films were masked with photoresist and etched by conventional wet chemical methods, hydrofluoric acid for the oxide and phosphoric-nitric-acetic acid solution for the aluminum. The resist was removed in acetone and the samples were rinsed thoroughly and dried prior to any processing. In the case of the oxide films, thickness measurements were performed before and after processing using an ellipsometer. The aluminum films used were 0.5$\mu$m thick and no measurements were performed before and after the etch process as this material is not removed by reaction with fluorine [1].

Schematics of the two reactor systems utilized in this work are shown in Fig. 1. The afterglow/thermal apparatus consists of fused silica tube material surrounded by a microwave cavity and downstream of the cavity by a conventional furnace. This system has been described elsewhere [4]. The system is similar to a low pressure chemical vapor deposition system to which a microwave source has been added. Radicals and atoms generated in the microwave source by the 2.45 GHz excitation at 400 Watts are allowed to flow into the thermal process zone within the furnace where chemical reaction with the silicon substrate occurs. Typical process pressure was 1 Torr at a flow of 1 liter min$^{-1}$ and temperature was varied from 340°C to 586°C. In this manner, the silicon wafer is exposed only to chemical species created by non-thermal means and the reaction is driven by the temperature of the furnace. Hence, the production of reactive species is decoupled from the environment in which the reaction takes place. The gas species used in afterglow studies were $CF_4$, $CF_4$/10% $O_2$, $CF_4$/10% $N_2O$, and $NF_3$. Direct plasma studies were carried out in a commercial plasma enhanced chemical vapor deposition system [5]. In this system the wafer substrate temperature could be controlled from 35°C to 350°C. The excitation power was set at 450 Watts and the pressure maintained at 500 mTorr. Two frequencies, 50 kHz and 13.56 MHz, were used for the gas excitation. Only $NF_3$ was used in direct plasma studies at a flow of 50 sccm. In each case,
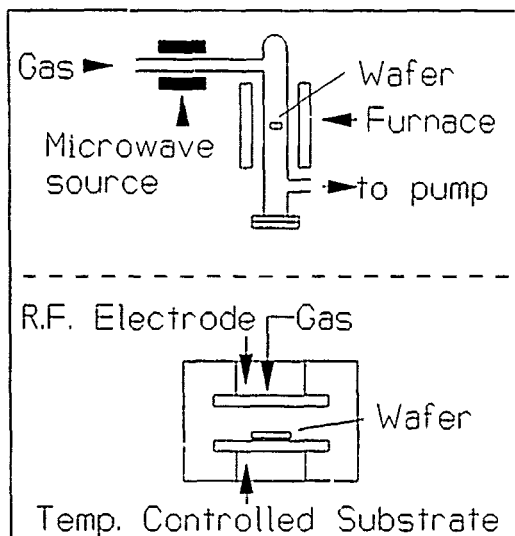
Fig. 1. Schematic of afterglow reactor (top) and direct plasma reactor (bottom).



Fig. 2. Afterglow/Thermal Silicon etch rates and selectivity vs. temperature. Etch, open structures: ( )$CF_4$, ( )$CF_4/O_2$, ( )$CF_4/N_2O$, ( )$NF_3$. Solid structures, selectivity, same patterns.

following the etch process and measurements of the remaining mask film thickness in the case of the oxide mask, the mask film was removed and the resultant step height created in the silicon was determined with a stylus profile instrument [6].

## RESULTS AND DISCUSSION

The afterglow/thermal system yielded quite interesting results. These data are summarized in Fig. 2. Keeping all parameters identical with the exception of gas mixture, $CF_4$ and mixtures of $CF_4$ produced silicon etch rates roughly 20% of those obtained in $NF_3$. The selectivity obtained to $SiO_2$, the ratio of silicon to oxide etch rate, averaged between 5 and 15 for the majority of $CF_4$ mixtures. The exception to this occurred at higher temperatures in pure $CF_4$. These latter data are inconsistent with models hypothesized describing etch rates of Si and $SiO_2$ [1] in fluorine atom chemistries, as a function of temperature. In other words, if F atoms alone are responsible for etching then the data should be uniform across temperature. The selectivities of the oxygen containing mixtures are depressed with respect to pure $CF_4$. This fact could be attributed to a slight oxidation of the silicon surface by atomic oxygen. Atomic oxygen has previously been shown to oxidize silicon at low temperature [4]. With the exception of the high temperature $CF_4$ case, however, the selectivities of the other mixtures are consistent with selectivities obtained in direct plasma processes [1,7] near room temperature. Therefore, oxidation is not expected to play a significant role in this process.

Approximating a wafer thickness at 650$\mu$m, the highest temperature $NF_3$ afterglow etch would require 14 minutes to clear a via, etching from only one side. Two sided masking is possible, especially with infrared alignment, but the process complexity necessarily suffers. Also the necessary oxide mask would be nearly 30$\mu$m thick. This point should direct potential users of such a process toward the use of masks with essentially infinite selectivity. Such films would form stable fluorides and not etch to any extent. Aluminum is one such material that could be used. However, the highest etch rates were obtained near 600°C. At this temperature the integrity of the Al° would be in question. Even with its negative aspects, $NF_3$ appears to be an excellent
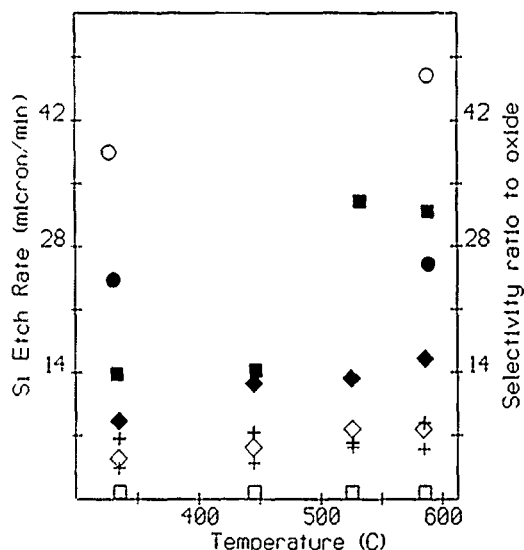
candidate for processes in which high silicon removal rates are required.

Fig. 3. depicts the results of silicon etch rate vs. temperature for the case of direct plasma treatment. These results were obtained with aluminum as the etch mask. These results again attest to the efficiency of $NF_3$ processes in producing large amounts of free fluorine. The most interesting feature of this data is the slight difference in rate obtained with two quite dissimilar frequencies. The high rate obtained at low frequency might allow some of the measurable values of the plasma to enhance the control of the process. RF current is one of these variables which could be monitored to control the etch rate or endpoints. Current measurement in high frequency plasma excitation is not nearly as straight forward.

The most significant challenge posed by high rate silicon etch processes is the surface texture problem which results when silicon is allowed to react with a fluorine atom environment. Although this damage could be removed by such processes as oxidation, many of the desired process fabrication sequences which could use high rate etch techniques occur at the end of device fabrication. Therefore, high temperature processes to remove damage are precluded. Accordingly, appropriate chemistries must be developed which will allow high rate silicon removal without roughening the surface. However, the basic mechanism of this surface texturing problem is not understood at this time. The afterglow process technique should be well suited to address this problem. Afterglow chemistry is very adaptable and by its nature may be used to create novel chemical species.

In conclusion, high rate chemical removal of silicon has been demonstrated using a variety of techniques. The response of each system indicates that the basic mechanism is chemical in nature and could be related to the concentration of atomic fluorine. Additional effort must be expended to eliminate surface texturing. Only then will this technique be of general use to individuals fabricating electrical devices in silicon.
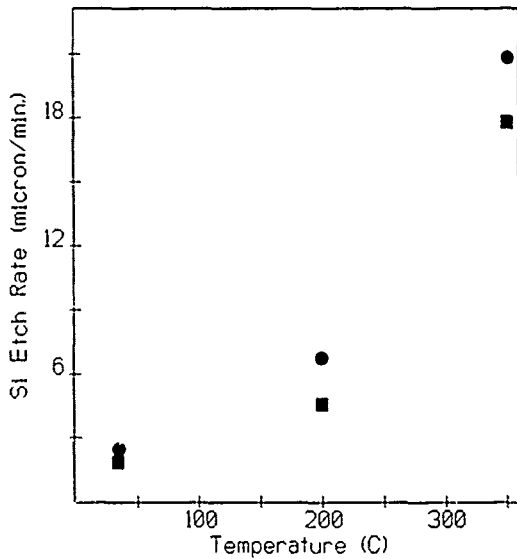
Fig. 3. Silicon etch rate vs. temperature for NF$_3$ direct plasma etching. ( ) 13.56 MHz, ( ) 50 kHz.

### REFERENCES

(1)    D.L. Flamm and V.M. Donnelly, "The design of plasma etchants," *Plasma Chem. and Proc.*, **1**, pp. 317-363, 1981.

(2)    B. Golja, J. Barkanic, A. Hoff, and J. Stach, "Plasma etching characteristics of Si and SiO$_2$ in NF$_3$/Ar and NF$_3$/He plasmas," *The Electrochemical Society Extended Abstracts*, Washington, D.C., pp. 207-208, October 9-14, 1983.

(3)    J.E. Spencer, R.A. Borel, and A. Hoff, "High rate photoresist stripping in a oxygen afterglow," *J. Electrochem. Soc.*, **133**, pp. 1922-1925, 1986.

(4)    A.M. Hoff and J. Ruzyllo, "Atomic oxygen and the thermal oxidation of silicon," *Appl. Phys. Lett.*, **52**, pp. 1264-1265, 1988.

(5)    Plasma-Therm, Inc. model 700. Plasma-Therm, Inc., Kresson, NJ..

(6)    Tencor Alpha-Step model 100 and 200, Tencor Instruments, Mountain View, CA..

(7)    B. Golja, J.A. Barkanic, and A. Hoff, "A review of nitrogen trifluoride for dry etching in microelectronics processing," *Microelectronics Journal*, **16**, pp. 5-21, 1985.

# Viscoelasticity and Creep Recovery of Polyimide Thin Films

Fariborz Maseeh and Stephen D. Senturia
Microsystems Technology Laboratories
Massachusetts Institute of Technology, Cambridge, MA 02139

## ABSTRACT

Creep and recovery of a polyimide thin film well below its glass transition temperature is demonstrated through use of circular membrane bulge test. Extensive use is made of a recently developed mechanical CAD system linked to the fabrication process to model the structure. A creep power law is used in a nonlinear finite element (FEM) analysis to fit the experimental results, thereby measuring the viscoelastic properties.

## INTRODUCTION

Recently, polyimide films have been used as structural components of mechanical microsensors and microactuators [1-3]. Polyimides, like most polymers, exhibit a time dependent mechanical response (viscoelasticity). This is potentially an important factor in the design of mechanical structures in which polyimide is subjected to sustained loads.

Viscoelastic properties of materials are traditionally measured by uniaxial tests [4]. Creep, stress-relaxation, and dynamic-loading tests are the commonly used techniques to measure the time-dependent analogs to elastic constants. The uniaxial creep test consists of measuring the time-dependent stretch resulting from the application of a steady axial load, and can be directly related to the viscoelastic properties. Uniaxial experiments, though effective for large samples, can be difficult to implement for thin films. We employ circular membranes to perform a biaxial creep experiment in which a time-dependent deformation is measured under a state of biaxial stress. This technique eliminates the sample handling problems and offers a way to detect and measure the viscoelastic behavior of the material. The membrane data, in the form of deflection as a function of applied pressure, need further modeling to extract the time dependent stress vs strain results. In this work, we report on creep and creep recovery of Dupont's commercial polyimide PI2525, and present a non-linear viscoelastic model for the interpretation of the creep behavior.

## SAMPLE FABRICATION AND MEASUREMENTS

The PI2525 polyimide is the imidized form of the polyamic acid precursor that results from the reaction of benzophenone-tetracarboxylic dianhydride (BTDA) with a blend of meta-phenylenediamine (MPDA-40%) and oxydianiline (ODA-60%) [5]. The polyamic acid is supplied commercially dissolved in N-methylpyrollidone and is spin-cast, then cured to produce the polyimide. Circular polyimide membranes one inch in diameter and 5 μm in thickness are microfabricated using the technique described in [6]. The PI2525 precursor is spin deposited on (100) silicon substrates in multiple layers, prebaked at 130°C for 15 minutes after each layer for solvent removal and partial imidization, and then cured (imidized) at 400°C in nitrogen for one hour. A VESPEL ring, one inch in inner diameter, 1-1/8 inch in outside diameter, and 1/16 inch thick, is mounted on the polyimide side with epoxy, and the substrate is removed by chemical etching with an nitric-hydroflouric acid solution. The ring, which is sufficiently large to be considered rigid, captures the state of biaxial stress in the film at the support. The sample is mounted in our bulge test apparatus and the center deflection as a function of the applied pressure is measured by tracking the focus point using a digital micrometer attached to a microscope stage. The applied air pressure is measured by a Kulite LQ-5-516 pressure sensor calibrated for the working pressure ranges (0 - 0.1 MPa).

Fig. 1 shows the center deflection vs pressure for a PI2525 membrane. The portion of the curve below the point indicated by the arrow is elastic. Data from this portion of the curve are used to extract the residual stress and biaxial modulus of the film, as explained in [6]. For pressures above this point, the center displacement increases with time at fixed pressure. In this particular figure, the pressure is gradually increased until the deflection reached 2200 μm, after which the increase in deflection is monitored as a function of time. In actual creep experiments, the pressure load is increased at small intervals and the deflection is observed over a time scale of 2 minutes. If a time-dependent deformation is observed, indicating the threshold of observable creep (on the time scale of tens of seconds), the pressure is then increased to a fixed value well above the creep threshold, and the center deformation as a function of time is measured. Fig. 2 shows typical creep data, plotted as creep deflection $d_{cr}$ vs time (the model is explained later). The creep deflection is defined as $d - d_S$ in Figure 1. The creep measurement continues until the center deflection stops changing on a time scale of 10 minutes. The sample is then unloaded and a plastically deformed dome shape surrounded by an elastic ring is observed shown schematically in Fig. 3. The width of the undeformed rim of the film is measured immediately after the unloading. This width is used in conjunction with elastic models of membrane deformation to estimate the strain at which creep initiates. After unloading, the polyimide sample gradually recovers its original flat shape. Fig. 4 shows a sequence of photographs for a PI2525 sample after unloading. The apparent total recovery time is 130 hours. After the recovery, a second load-deflection measurement is made in the elastic region. The biaxial modulus extracted from the second run agrees with the first run, however, the residual stress shows a 40% drop in value as a result of having cycled through the creep and creep-recovery sequence.
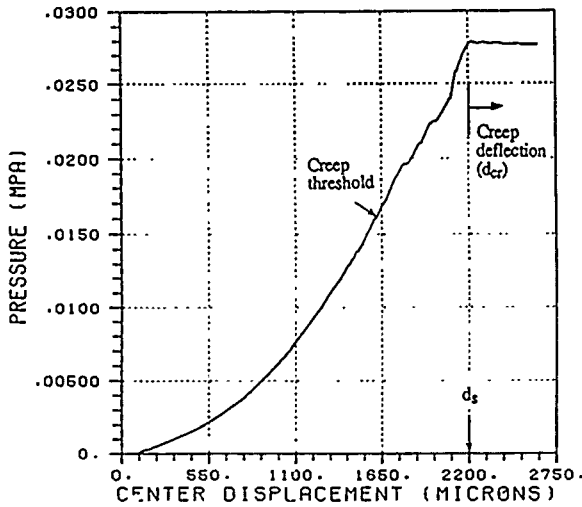
Fig. 1. Measured center deflection of a PI2525 circular membrane, one inch in diameter. The flat part is the viscoelastic deformation under constant pressure as a function of time.
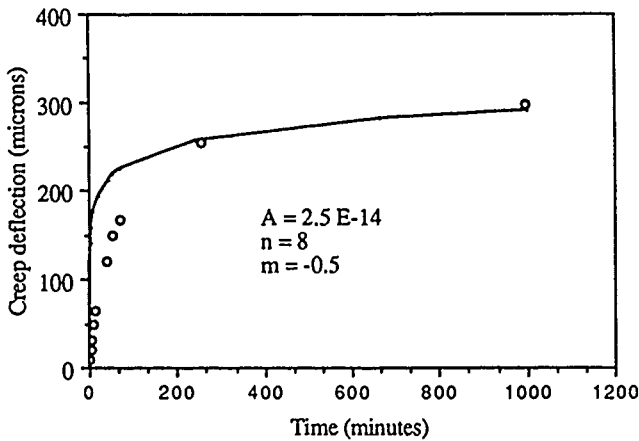


Fig. 2. FEM model of creep deflection vs time. Hollow circles are the measured data
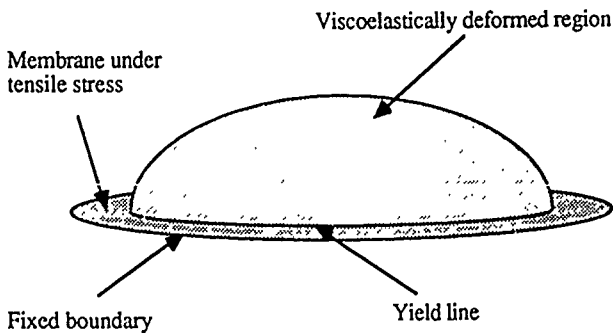


Fig. 3. PI2525 circular membrane viscoelastically deformed when unloaded. The deformed region recovers in time well below glass transition temperature.

## A. Modeling of the structure

The membrane fabrication and analysis is implemented in the environment of a previously reported CAD architecture [7,8], which uses a Structure Simulator supported by process modeling programs to create the solid model. The process modeling for spinning polyimide on a substrate uses data on film thickness vs spin speed for this particular polyimide [9] to generate a model for thickness ($t_f$ in μm) as a function of spin speed ($V_s$ in krpm) and the number of coats n. In the case of PI2525, a good representation of the data is provided by the expression $t_f = n [1/6 + 40/3V_s]$. A PFR (process flow representation [10]) of the polyimide deposition (including spinning, prebake, and cure steps) is developed. The Structure Simulator interprets the PFR process step description, determines the film thickness, and generates the solid model in a format which is directly readable by the mechanical modeling graphics pre/post processor PATRAN. Because of the radial symmetry of the samples, only two-dimensional models are generated. Within the PATRAN preprocessor graphics, the ring supporting the film is modeled by enforcing built-in conditions on the face of the film at the ring-film contact surface. The elastic properties (measured Young's modulus and Poisson's ratio from [6]) of polyimide (assumed isotropic) are loaded from a PATRAN-readable material property database file. Loads and boundary conditions and element types are specified. An ABAQUS input file is then generated assuming zero residual stress, and is modified manually to account for the residual stress in the polyimide. The viscoelastic properties are also entered by manual modification of the ABAQUS input file. The resulting model is sent to ABAQUS version 4-7 [11,12]; the analysis results are returned to PATRAN and are displayed graphically. The process modeling and FEM model generation are done on a SUN-4. The FEM simulations are done on both SUN-4 and the MIT·CRAY2 facility; in the latter case the ABAQUS output is transported back to SUN-4 for evaluation.

Since both the geometry and applied loads are axially symmetric, a two dimensional analysis is sufficient by taking a section through the thickness and a radius. 8-noded axisymmetric elements are used throughout the analysis. The suspended part of the film within the ring is modeled with 50 elements with element sizes linearly decreasing toward the edge to one-tenth of the one at the center, shown schematically in Fig. 5. The portion of the film adhering to the ring is modeled with 10 uniform elements fixed at the bottom edge as shown.

## B. Analysis of viscoelastic behavior

For deformations in the elastic regime, the state of the stress in the loaded membrane is shown in Fig. 6. As shown, an unequal state of biaxial stress develops when the membrane is subjected to uniform lateral pressure. The nonuniform stress distribution in the membrane requires a more complex analysis to determine the viscoelastic properties of the film than when the uniaxial measurement is used. Under the condition of the uniaxial loading, the commonly adopted creep law is of the form:

$$v_{cr} = A q^n t^m \qquad (1)$$

where $v_{cr}$ is the creep strain rate, q is the von Mises stress (a condition among the stress components at a point that must be satisfied for the onset of plastic deformation; in the case of simple tension: $2 q^2 = [\sigma_1 - \sigma_2]^2 + [\sigma_2 - \sigma_3]^2 + [\sigma_1 - \sigma_3]^2$ where $\sigma_i$'s are the principal stresses), A is a constant, n is a material dependent integer normally between 3 and 8 [13], and m is a (time hardening) constant between -1 and 0. All constants are to be determined by fitting to the creep measurement data. A modified creep law is used in this work by replacing q with q-$q_y$ where $q_y$ is the creep threshold, the stress below which the material creep is negligible. The creep strain $\varepsilon_{cr}$ is then determined by integrating eqn. (1) in time which results in:

$$\varepsilon_{cr} = [A / (m+1)] (q-q_y)^n t^{m+1} \qquad q > q_y$$
$$\varepsilon_{cr} = 0 \qquad q < q_y \qquad (2)$$

56

This results in creep when the stress is above the threshold stress (assumed to be 95 MPa for PI2525 from experimental observations). Eqn. (2) is incorporated in the FEM model. Investigation of this model reveals a linear relationship between the creep strain $\varepsilon_{cr}$ and the creep deflection $d_{cr}$ for different choices of the constants A, n, and m in the ranges of the measured displacements. Fig. 7 shows this relationship as obtained from FEM analysis. This result is used to determine the time-hardening exponent m directly from a fit to the measured $d_{cr}$ vs t. This value is then utilized to extract values of A and n by a fit to the experimental results.

To extract the creep compliance from the experimental results, an iterative FEM procedure is utilized. The initial (order of magnitude) estimate of A is based on assuming linear viscoelasticity (n=1) and calculating the creep strains from an elastic deflection model [14,15]:

$$\varepsilon_{cr} = .67\,(d/a)^2 - .67\,(d_s/a)^2 \tag{3}$$

where d is the total deflection, $d_s$ the static deflection and a the membrane radius. Using this approximation, a value for A is extracted from a logarithmic fit to the $\varepsilon_{cr}$ vs t curve. Then, with estimates of A and n (assumed unity in the first iteration) inserted into the FEM model, calculate the time-dependent deformation at the experimentally determined pressure (which at zero time yields the elastic deflection $d_s$), compare with experiment, and revise A and n until satisfactory convergence is achieved. In this case, no rigorous best fit is obtained, and error bounds on the resulting parameters can only be estimated. However, this procedure yields a reasonable approximation to the experimental data.
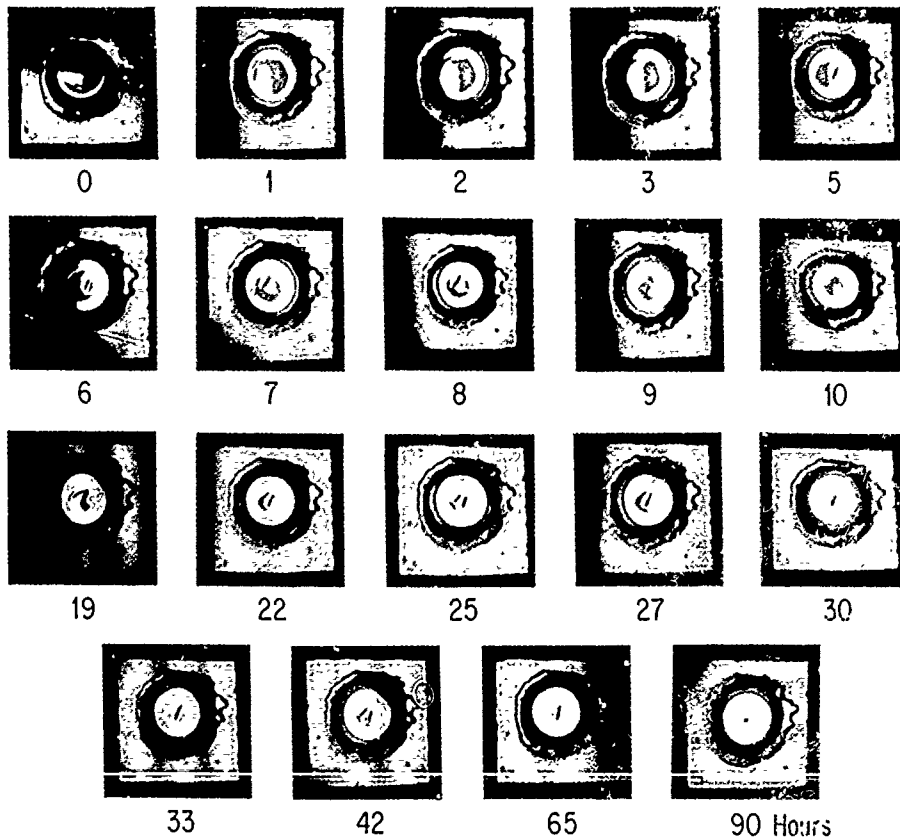


Fig. 4. Creep recovery of BTDA-MPDA/ODA polyimide (PI2525) at room temperature, when sample is unloaded. The glass transition temperature for this polyimide is above 300°C. The numbers below each photograph indicate the elapsed time in hours.

57

## C. Results

The membrane pressure vs center deflection data consists of an elastic and a viscoelastic region. The elastic portion is used to determine the elastic properties of the thin film with the procedure in [6].

Fig. 6 shows the distribution of strains and stresses from the FEM in the meridian and the circumferential principal directions. The state of strain along the meridian (principal) direction in the deformed shape remains fairly constant along the radius; however the circumferential (principal) strain distribution is equal to the meridian one at the membrane center and decreases to zero at the support. These distributions are in agreement with a generalized membrane model [14,15], but deviate significantly from the Beam's spherical cap assumption [16]. The stress distribution along the meridian and circumferential directions behave as shown in Fig. 6, in agreement with the plane stress constitutive relation. The shear stresses are negligible. The maximum Mises stresses (and therefore plastic flow) occurs at the center of the membrane and gradually decrease toward the support as in the experiment. The center element is under the state of equal biaxial stress; therefore, this element is used throughout the analysis to characterize the viscoelastic properties.

Fig. 2 shows the FEM creep deflection vs time (for the center element) compared with the experimental results. The agreement is not perfect, but key features are well represented. Fig. 8, shows the resulting center element biaxial creep strain as a function of time. The biaxial creep strain $\varepsilon_{bcr}$ is related to the uniaxial creep strain $\varepsilon_{ucr}$ by the relation [17]:

$$\varepsilon_{bcr} = \varepsilon_{ucr} / 2 + \sigma / 6K \qquad (4)$$

where $\sigma$ is the stress and K is the bulk modulus. Fig. 8 also shows the resulting uniaxial creep strain vs time.

The strain energy loss due to stress reduction in the film after recovery is related to the deviatoric (distortional) portion of the strain energy and is estimated form the expression [18]:

$$\Delta U_d = (1/6G) [ \sigma_i^2 - \sigma_f^2 ] \qquad (5)$$

where G is the shear modulus, and $\sigma_i$ and $\sigma_f$ are the in-plane stresses, in the unloaded state, before and after the recovery. This energy loss ( 0.098 MJ/m$^3$) may be attributed to dissipation during the molecular distortions of the polymer chains during the creep. The effects of elevated temperature anneals on this energy loss has not been investigated, but would be interesting.



Fig. 6a. FEM meridian and circumferential strains for one inch circular PI2525 membrane
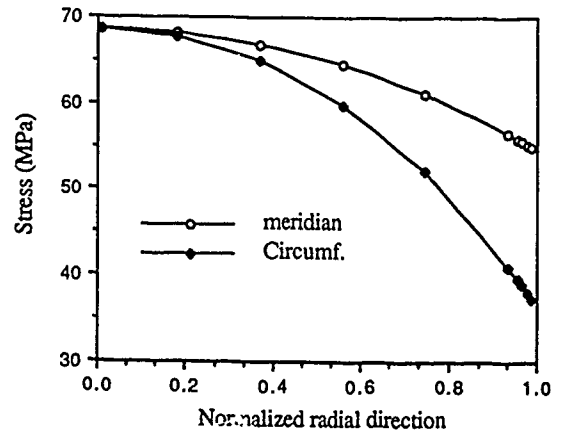


Fig. 6b. FEM Meridian and circumferential principal stresses for a circular PI2525 membrane
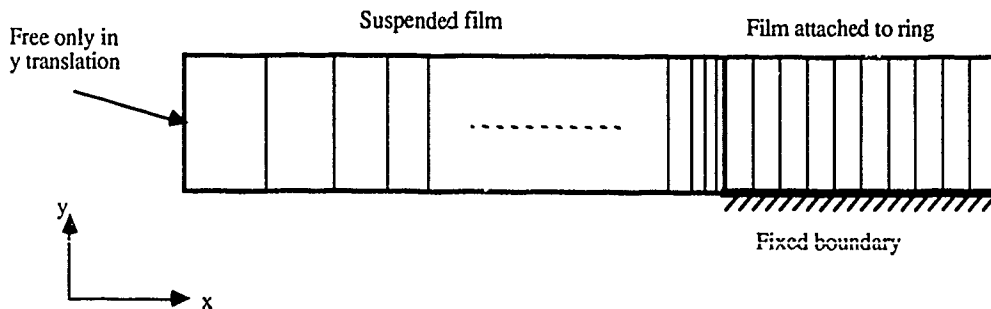


Fig. 5. Schematic finite element mesh for polyimide circular membranes. Axisymmetric elements (50 for suspended part and 10 for the fixed portion of the film) are used. Elements in the suspended film decrease uniformly from center to edge with a 10 to 1 ratio as shown.
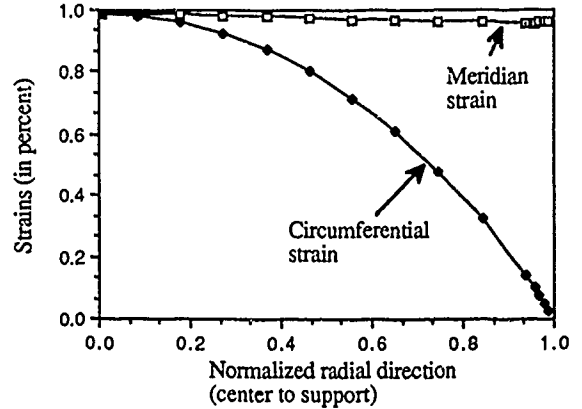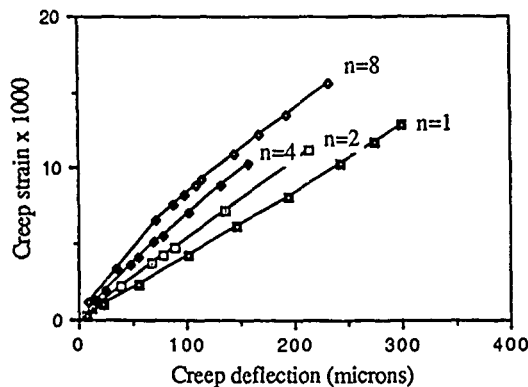
58

Fig. 7. Creep deflection vs biaxial strain in the center element
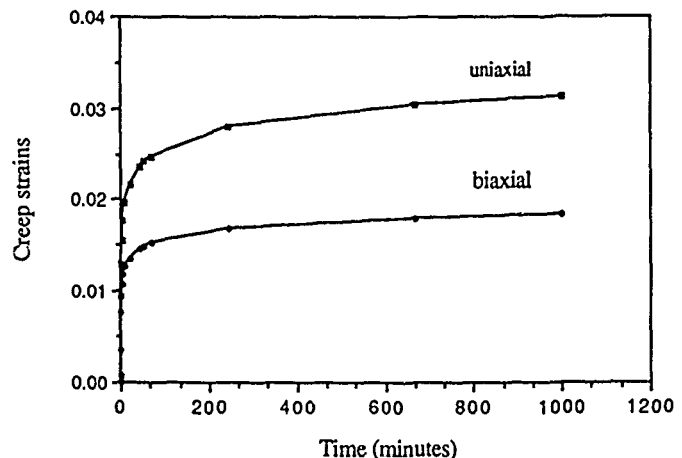(FEM model of PI2525 membrane)



Fig. 8. Creep strains from a FEM vs time; the uniaxial strain
is obtained from eqn (4)

## CREEP RECOVERY

The creep recovery phenomenon illustrated in Figure 4 (shape memory effect) has been observed in polymers that are plastically deformed below their $T_g$ when they are thermally loaded near or above their $T_g$; see e.g. [19,20]. In the experiment reported here, the polyimide is exhibiting creep recovery about 300 °C below its glass transition. Below the glass transition temperature, a polymer material must overcome two distinct sources of resistance before inelastic flow may occur [21,22]: (1) the material must be stressed above its intermolecular resistance to segment rotation (restrictions imposed on molecular chain motion by the neighboring chains), and (2) once the material begins to flow, molecular alignment occurs, altering the configurational entropy of the material. This entropy change, by analogy to the rubbery regime, may be related to the evolution of back stresses which can result in the recovery of the polymer [23]. The energy loss after shape recovery may be attributed to dissipating effects during reorientation of the polyimide chains.

Quantitative modeling of the recovery phenomenon will be reported elsewhere. The yield line for the recovering portion does not remain circular, and the contour for the deformed region is not easily measured without film disturbance. Qualitatively, it is observed that the recovery is temperature dependent. The sample recovers faster when the temperature is elevated. Humidity effects have not been investigated.

## CONCLUSION

The use of circular membranes in measuring the viscoelastic and plastic material properties of thin films has some advantages over the uniaxial techniques. The membranes are edge free and therefore less susceptible to cracks which can prohibit observation of yield in glassy materials such as PI2525. The maximum stress region (center point) is also identified in advance; hence the viscoelastic properties of thin films can be observed without having to yield or creep the entire sample. However, the analysis to determine the viscoelastic properties is more cumbersome. This paper has presented a first attempt at such data analysis, and has determined that the creep compliance of PI2525 is a very nonlinear function of the stress (n = 8), and also requires a threshold strain before creep begins.

## REFERENCES

[1] M. A. Schmidt, R. T. Howe, S. D. Senturia, and J. H. Haritonidis, "Design and calibration of a microfabricated floating element shear-stress sensor", IEEE Trans. Elec. Dev., v. 35, June 1988, p. 750.

[2] M. G. Allen, M. Scheidl, R. L. Smith," Design and fabrication of movable silicon plates suspended by flexible supports", Proceedings of IEEE Micro Electro Mechanical Systems, Salt Lake City , Utah, Feb. 1989, p.76.

[3] M Mehregany, R. Mahadevan, and K. J. Gabriel, "Application of electric microactuators to silicon micromechanics", Proceedings of Transducers'89, June 1989, Montreux, Switzerland.

[4] R Mohan and D F. Adams, "Nonlinear creep-recovery response of a polymer matrix and its composites", Exp. Mech., Sept. 1985, p. 262.

[5] P V Nagarkar, percentages from an XPS study, to be published.

[6] F. Maseeh, and S. D. Senturia,"Elastic properties of polyimide thin films", Polyimides : Materials, Chemistry and Characterization, ed. by C. Feger, M. M. Khojasteh and J. E. Grath, Elsevier, Amsterdam, 1989, p. 575.

[7] F. Maseeh, R. M. Harris and S. D. Senturia,"A CAD architecture for microclec.omechanical design", Proceedings of IEEE Micro Electro Mechanical Systems, Napa Valley, CA, Feb. 1990, p. 44.

[8] R. M. Harris, F. Maseeh, and S. D. Senturia, "Automatic generation of a 3-D solid model of a microfabricated structure", Proceedings of this conference.

[9] D. Volfson, private communications.

[10] D. S. Boning, private communication.

[11] H. D. Hibbitt, "ABAQUS/EPGEN, a general purpose finite element code with emphasis on nonlinear applications", Nuc. Eng. Des., vol. 77, 1984, p. 271.

[12] ABAQUS User Manual, Release 4-7, HKS, Providence, RI.

[13] A.R. S. Ponter, and F. A. Leckie, "The application of energy theorems to bodies which creep in the plastic range", J. Appl. Mech. Sept. 1970, p. 753.

[14] A. E. Green, and J. E. Adkins, Large elastic deformations, Clarendon press, Oxford, 1970, p.133.

[15] T. Tsakalakos, "A comparison of theory and experiment for isotropic and anisotropic films", Thin solid films, v. 75, 1981, p. 293.

[16] J. W. Beams, "Mechanical properties of thin films of gold and silver", in C. A. Neugebauer, J. B. Newkirk, and D. A. Vermilyea, ed., Structure and Properties of Thin Films, Wiley & Sons, New York, 1959, p. 183.

[17] L. E. Nielsen, Mechanical properties of polymers and composites, v. 1, Marcel Dekker, Inc., NY, 1974, p.121.

[18] E. F. Byars, R. D. Snyder, and H. L. Plants, Engineering mechanics of deformable bodies, Fourth ed., Harper & Row, 1983, p. 417.

[19] Y. P. Khanna, R. Kumar, and A. C. Reimschuessel, "Memory effects in polymers III. Process history vs crystallization rate of Nylon 6-comments on the origin of memory effect", Pol. Eng. & Sc., v. 28, No. 24, 1988, p. 1607.

[20] A. Charlesby, and B. Kidric, "Memory effect in irradiated polymers", Radiat. Phys. Chem., v. 30, No. 1, 1987, p. 67.

[21] A. S. Argon, "A theory for the low-temperature plastic deformation of glassy polymers", Phil. Mag., vol. 28, 1973, p. 39.

[22] R. N. Howard, and G. Thackray, "The use of a mathematical model to describe isothermal stress-strain curves in glassy thermoplastics", Proc. Royal Soc. vol. 302, 1968, p. 453.

[23] M. C. Boyce, D. M. Parks, and A. S. Argon, "Large inelastic deformation of glassy polymers. Part I: rate dependent constitutive model", Mech. Mat., vol. 7, 1988, p. 15.

# CONTROLLED MICROSTRUCTURE OXIDE COATINGS FOR CHEMICAL SENSORS

G. C. Frye, C. J. Brinker, T. Bein*, C. S. Ashley and S. L. Martinez

Inorganic Materials Chemistry Division 1846
Sandia National Laboratories
Albuquerque, NM  87185

## ABSTRACT

We have investigated the use of porous oxide coatings, formed using sol-gel chemistry routes, as the discriminating elements of acoustic wave (AW) chemical sensors. These coatings provide several unique advantages: durability, high adsorption capacity based on large surface areas, and chemical selectivity based on both molecular size and chemical interactions. The porosity of these coatings is determined by performing nitrogen adsorption isotherms using the AW device response to monitor the uptake of nitrogen at 77 K. The chemical sensitivity and selectivity obtained with this class of coatings will be demonstrated using several examples: hydrous titanate ion exchange coatings, zeolite/silicate micro-composite coatings, and surface modified silicate films.

## INTRODUCTION

The development of coatings that provide both chemical selectivity and increased sensitivity is critical for the development of effective chemical sensors (1,2). Chemical selectivity can be obtained based on selective sorption of a single species or a class of species. Increased sensitivity can be obtained based on high sorption capacities. Various types of coatings have been used for gas phase sensors utilizing surface acoustic wave (SAW) devices (1,3) and for liquid phase sensors utilizing either acoustic plate mode (APM) devices (4) or Lamb wave devices (5). Some examples of the types of coatings employed are organic polymers (1,3), phthalocyanines (1) and biological agents (5).

Sol-gel chemistry involves the hydrolysis and condensation of metal alkoxides to form inorganic polymers in solution. By varying the reaction conditions (e.g., reaction protocol, concentration of catalyst, water or alkoxide), the structure of these polymers can be varied from weakly branched chains to highly ramified structures (i.e., resembling a tumbleweed) to dense colloidal particles (6). Prior to gelation, films can be prepared from these solutions by either spin- or dip-coating. The final film structure is dictated by the film forming procedure and the polymer structure: dense films are formed from weakly branched polymers, while high surface area porous films are formed from highly ramified polymers or dense colloidal particles (7). The pore size distribution in the porous films can be tailored by varying the size and shape of the precursor polymers as well as by varying the coating procedure or thermal treatment (8).

These porous oxide coatings can provide several unique advantages as the discriminating elements of AW chemical sensors: (1) resistance to thermal or chemical degradation, (2) high surface area and pore volume resulting in increased sensitivity based on large sorption capacities, (3) controlled microstructure (i.e., pore size distribution) providing chemical discrimination based on molecular size, (4) capability to easily modify the chemical nature of the surfaces to provide discrimination based on chemical

interactions and (5) minimal changes in film properties (e.g., viscoelastic properties) during sorption resulting in simplified analysis of sensor response. Specific examples to be discussed are: (1) the use of SAW devices coated with zeolite microcrystals embedded in a dense sol-gel matrix to provide chemical selectivity based on molecular size, (2) the use of APM devices coated with hydrous oxides with large ion exchange capacities for sensing ionic species in solution, and (3) the addition of chemically reactive surface groups to porous oxide films using silane coupling agents.

## EXPERIMENTAL

### Characterization of Film Porosity

In developing oxide coatings for chemical sensors, the microstructural properties of potential films have been characterized. The frequency response of a coated SAW device is used to monitor the uptake of nitrogen (at 77 K) as a function of the partial pressure (p) of nitrogen in the gas phase over the device. This type of measurement, called a nitrogen adsorption isotherm, can be used to determine: (1) sample surface area using the BET analysis, (2) total pore volume based on the maximum uptake at partial pressures close to the saturation vapor pressure ($p_o$) at the temperature of the run and (3) pore size distributions based on an analysis considering capillary condensation into the pores (9).

The experimental system used for these measurements is shown in Fig. 1. The normalized nitrogen partial pressure ($p/p_o$) is varied from zero to about 0.95 and back to zero. Computer-operated mass flow controllers are used to vary the relative flow rates of a nitrogen stream and a nonadsorbing helium mix-down stream. To maintain the device at 77 K, the test case and a coil on the gas inlet side are immersed in a dewar of liquid nitrogen.
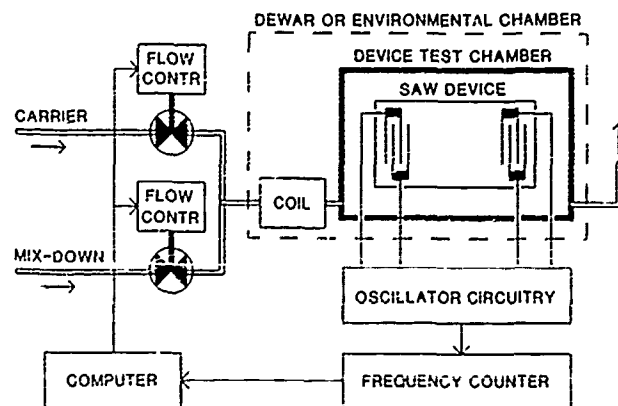


Fig. 1: Schematic of experimental system used to monitor nitrogen adsorption on the surfaces of porous thin films formed on SAW devices. The device is maintained at 77 K by immersion in a dewar filled with liquid N₂.

*Department of Chemistry
University of New Mexico
Albuquerque, NM  87131.

Nitrogen adsorption isotherms obtained in this manner for two films are shown in Fig. 2. The sol-gel system used contains $SiO_2$, $B_2O_3$, $Al_2O_3$, and BaO in ratios (by weight) of 71.18.7.4 (details of the sol-gel preparation are presented elsewhere (10)).

The first film (Fig. 2a) was formed by dip-coating using the initially prepared sol-gel solution which contains very small polymer precursors. The total amount of adsorption is small and the surface area, calculated using the BET analysis (9), is 1.3 $cm^2/cm^2$ of film. This value is very close to the value of 1.0 $cm^2/cm^2$ expected for a flat nonporous film (the slightly higher value may be indicative of surface roughness). This result agrees with ellipsometric (Rudolph AutoEL IV) analysis which gives a film thickness of .20 nm and a refractive index of 1.46, very close to the value expected for a dense film of this composition.

The second film (Fig. 2b) was formed by dip-coating after aging the same solution for three weeks at 50 C and pH 3. This aging results in larger, highly ramified polymeric species which, due to their inability to interpenetrate, create porosity in the film (7). This porosity is verified by ellipsometry which gives a refractive index of 1.18, indicating a volume percent porosity of 56%. This agrees with a value of 52% calculated based on the amount of adsorption at the high partial pressure end of the isotherm and a film thickness from ellipsometry of 148 nm.

The pore size distribution obtained from the desorption branch of the isotherm for the porous film (Fig. 2a) is shown in Fig. 3. A unimodal distribution with a median pore diameter of 6.2 nm is obtained. Smaller pore sizes are obtained with this system using
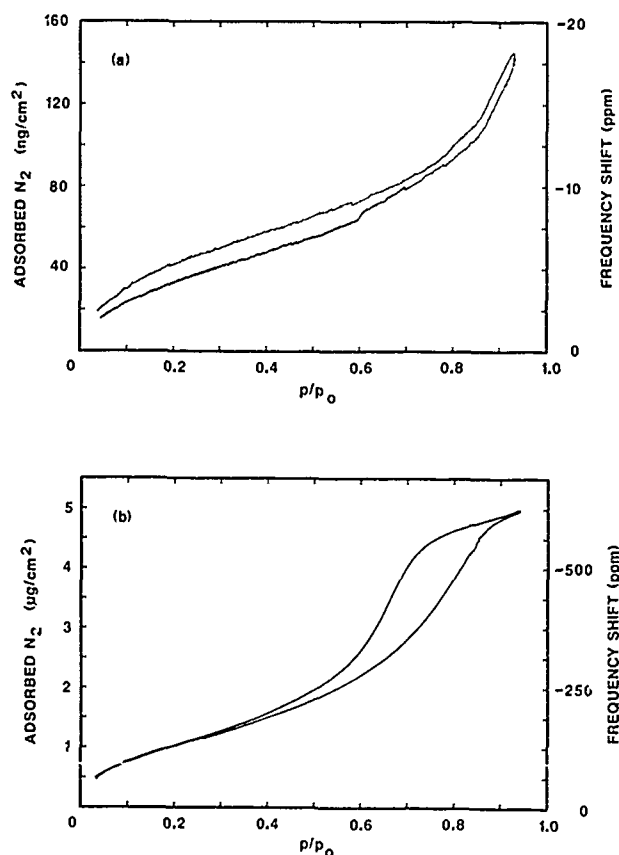
shorter aging times (8,10). Results of additional studies regarding the effect of sol-gel chemistry, reaction protocol, and film-forming procedure on film microstructure are presented elsewhere (8,11). These studies illustrate how sol-gel chemistry can be used to tailor surface area and pore size to fit a sensing application.
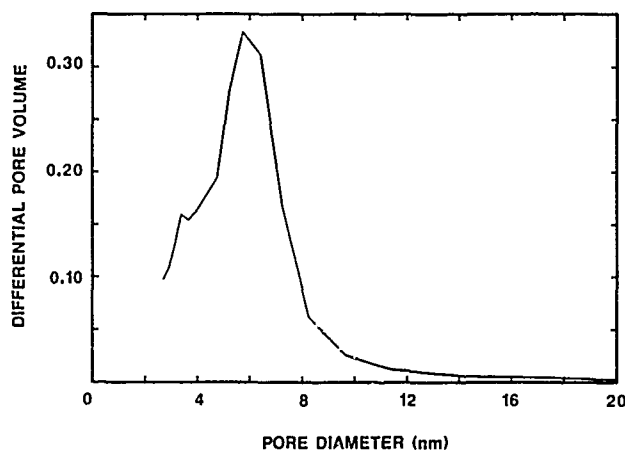


Fig. 3. Pore size distribution obtained from the desorption branch of the nitrogen adsorption isotherm for the film formed from the aged sol-gel solution (see Fig. 2b). A median pore diameter of 6.2 nm is obtained.

## Zeolite/Silicate Microcomposite Coatings

The most dramatic demonstration of chemical selectivity based on size exclusion is obtained with a film with a single characteristic pore size. Sol-gel chemical routes can provide a narrow range of pore sizes; however, the random nature of this process always results in a pore size distribution (see Fig. 3). Model inorganic porous systems with unimodal pore sizes are zeolites (aka molecular sieves). The most common zeolites are aluminosilicates which, due to their cagelike crystalline structures, have a single pore size that dictates accessibility to molecular adsorbate species. For example, the ZSM-5 zeolite has an opening with dimensions of 0.55 by 0.60 nm.

Zeolites can be synthesized using hydrothermal techniques. By rapid quenching at the early stages of the reaction, small (< 1 $\mu$m crystallite size) ZSM-5 crystals are formed (12). To create an AW device coating that utilizes the unimodal pore size of zeolites, the small crystals are embedded in a dense thin-film sol-gel matrix (13). This matrix forces adsorption to occur exclusively within the zeolite channel system. SEM analysis showed that the zeolites protruded from the sol-gel matrix. Quantitative titration of intrazeolite acid sites with organic bases such as pyridine indicated that the zeolite channels remained accessible (14).

To test the selectivity of a SAW device coated with this ZSM-5 zeolite/silicate microcomposite film, molecular probes with kinetic diameters ranging from 0.38 to 0.62 nm were employed. The observed SAW frequency shifts and the calculated mass changes are listed in Table I. At the relatively low vapor concentration used (0.1% of saturation for each species), large frequency shifts, indicating significant adsorption, were observed for methanol and isopropanol which have kinetic diameters less than the zeolite pore size (0.6 nm). A very small response was seen for iso-octane (kinetic diameter = 0.62 nm). These results demonstrate dramatic discrimination based on size exclusion using controlled microstructure coatings.



Fig. 2. Nitrogen adsorption isotherms using coated SAW devices for films formed from (a) unaged and (b) aged sol-gel solutions A BET surface area of 1.3 $cm^2/cm^2$ of film for the unaged indicates this film is nonporous while a value of 33 $cm^2/cm^2$ obtained for the aged indicates this film is highly porous.

62

Table I: Chemical discrimination based on molecular size using SAW devices coated with zeolite/sol-gel microcomposite films.

| Species | Kinetic Diameter (nm) | Frequency Shift (Hz) | Mass Change (ng/cm²) |
|---|---|---|---|
| Methanol | 0.38 | -6350 | 540 |
| Isopropanol | 0.47 | -10200 | 840 |
| Iso-octane | 0.62 | +74 | -6 |

## Hydrous Oxide Ion Exchange Coatings

It is well known that hydrous oxides, such as sodium titanates or niobates, act as ion exchange materials with large ion exchange capacities. This capability can be used to advantage in nuclear waste applications (15) or in the preparation of catalysts which, due to the highly dispersed nature of the ion exchanged metal atoms, show high activity even at low metal loading (16). By preparing coatings from these materials on APM devices, these large ion exchange capacities should result in large sensor responses during the exchange of ions from solution. Chemical selectivity can be varied by the choice of cations used (e.g., niobates form strongly acidic supports, titanates form amphoteric supports, and lanthanides form basic supports) and by controlling pore sizes.

To investigate the utility of these coatings, a sodium titanate system with a sodium:titanium ratio of 1:2 was chosen (15,16). The standard preparation technique for this material results in a solution which thickens too quickly to be usable in forming homogeneous coatings. However, by diluting (to 25% by volume) in hexane, reasonably homogeneous coatings are formed by spin-coating at 4000 rpm. Film thickness per coat is 300 nm. Initial experiments investigated the APM response during the uptake of $Ni^{2+}$ at pH 6. The APM response indicates the ion exchange is far from complete and is only partially reversible (significant amounts of exchanged Ni remain in the film after treatment with pH 3.0 $HNO_3$ and 1.0 M $NaNO_3$ solutions). This irreversibility contrasts results obtained for bulk samples (16).

To verify the APM results, similar coatings were formed on silicon and analyzed using a Tracor 5000 energy dispersive X-ray fluorescence (XRF) spectrometer. Based on the amount of titanium in the film, $Ni^{2+}$ exchange is incomplete (only 6% after 10 min. in a 0.05 M $Ni(NO_3)_2$ solution at pH 5.8). This incomplete exchange is verified by the observation that only 35% of the $Na^+$ initially present in the film was removed. The partial irreversibility of the $Ni^{2+}$ exchange is consistent with the lack of an experimentally significant change in the Ni/Ti ratio after a pH 3.0 $HNO_3$ rinse, a subsequent 1.0 M $NaNO_3$ rinse and a subsequent 0.1 M EDTA (ethylenediaminetetraacetic acid) rinse ˙6.5.

In contrast, XRF analysis of      nd film for the ion exchange of $Ag^+$ indicates that exch      rom a 0.05 M solution at pH 8 is effectively complete (97% of theoretical with only 3% of the initial $Na^+$ remaining in the film). In addition, the $Ag^+$ can be effectively removed from the film: only 23% remained after a pH 3.0 $HNO_3$ rinse, 15% after a subsequent 1.0 M $NaNO_3$ rinse and 9% after a subsequent 0.1 M EDTA rinse. Based on these results, $Ag^+$ exchange experiments were performed on a second coated APM device. Peristaltic pumps were used to pump solutions to and from a cell holding the liquid in contact with the unelectroded side of the APM device.

The results, shown in Fig. 4, are in good qualitative agreement with the XRF data. When a pH 8.0 solution of NaOH is replaced with a 0.05 M $AgNO_3$ solution at the same pH, a frequency decrease of about 190 ppm is observed after an initial frequency increase. Assuming the frequency response is due to changes in
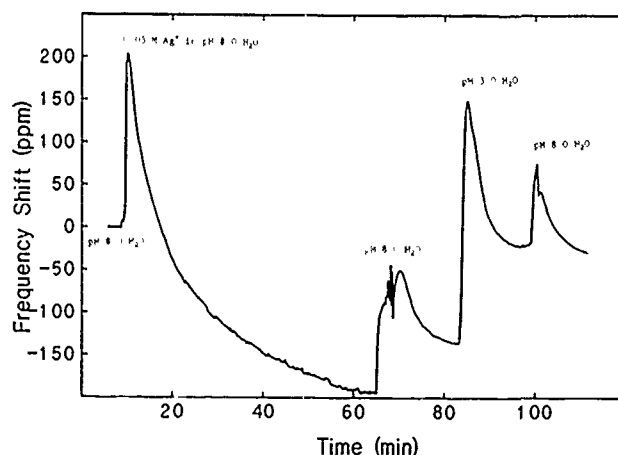


Fig. 4: Frequency response of hydrous titanate-coated APM device during ion exchange of $Ag^+$. A mass increase of 4.4 µg/cm² is detected for this exchange. Reversibility is indicated by the frequency returning to near its original value after a pH 3.0 rinse.

the mass loading of the film, this shift corresponds to a change in the film mass density of 4.4 µg/cm² (calculated using $\Delta f/f_o = c_t \Delta m$ where $\Delta m$ is the change in the mass/area and $c_t = 43$ cm²/g is the mass sensitivity coefficient of the device). This value is significantly lower than the value of 70 µg/cm² calculated for complete exchange of $Ag^+$ for $Na^+$ based on the film thickness of 600 nm (two coats were used), a sodium titanate skeletal density of 2.4 g/cm³, 40% porosity and a mass per ion exchange site of 176 g/mole (assumes site has form $Ti_2O_5$). This difference may indicate that many of the ion exchange sites are inaccessible. Thus, altering the film forming process to prepare films with fully accessible sites should result in a dramatic increase in the sensitivity of these sensors (full exchange of $Ag^+$ should give a frequency decrease of about 3000 ppm). The initial upward spike appears to be due to the loss of liquid flow which occurs during the change of solutions and was observed every time the solutions were changed. It should be possible to remove this erroneous signal using a solvent delivery system that can change solutions without loss of flow.

When the pH 8.0 $AgNO_3$ solution is replaced with a NaOH solution at the same pH, a 60 ppm frequency increase is observed indicating that about 30% of the $Ag^+$ is removed from the film. When a pH 3.0 $HNO_3$ solution is used, an additional frequency increase of over 110 ppm is observed. The total frequency increase in these two steps indicates that 90% of the exchanged $Ag^+$ has been removed, verifying the reversibility of this exchange process (this value may be a slight overestimate since the ion exchange sites will contain $H^+$ at this pH rather than $Na^+$ after the $Ag^+$ is removed). Except for the upward spike, a relatively small frequency shift is observed when the pH 3.0 $HNO_3$ solution is replaced with the pH 8.0 NaOH, indicating that the frequency increase seen with the pH 3.0 solution is due to mass loss during $Ag^+$ removal rather than due to some other effect related to the pH change. These results demonstrate that the large ion exchange capacities of hydrous oxide materials can be used to provide large sensor responses. Future work with these coatings will focus on optimizing the film forming process, determining the selectivity of the ion exchange process, and investigating the variability in ion exchange properties obtainable with various hydrous oxides.

63

## Surface Modification Using Silane Coupling Agents

Another technique being investigated which takes advantage of the high surface area and controlled pore structure of sol-gel coatings is the derivatization of the surfaces of these coatings with ligands that selectively bind chemical species of interest. With this technique, the desired pore structure (allowing for the decrease in the pore dimensions due to the attached surface layer) is first formed using appropriate sol-gel protocols. Then, using reagents such as silane coupling agents (e.g., XSi(OR)$_3$ where R is an alkyl group and X is the ligand), the reactive ligands are bound to the film surfaces.

For example, the enhanced binding of ionic species from solution is demonstrated with particulate silicate films derivatized with diethylenetriamine. Particulate silicate films were deposited by spin-coating a solution containing 22 nm diameter silicate particles formed by the Stöber process (13). XRF analysis indicated that the uptake of $Cu^{2+}$ in the as-deposited film after a 5 minute treatment in 1 M $Cu(NO_3)_2$ followed by rinsing with deionized water was less than 0.1 $\mu g/cm^2$ of film. However, after diethylenetriamine ligands were silane coupled to the surfaces, the $Cu^{2+}$ uptake increase to 3.0 $\mu g/cm^2$. Since a wide variety of ligands with silane coupling groups are commercially available, this generic technique should provide high surface area coatings with tailored chemical selectivities.

## CONCLUSIONS

Sol-gel chemistry can be used to prepare thin oxide coatings for sensor surfaces that provide both high sensitivity due to high surface areas, and chemical selectivity, based on both steric and chemical factors. Coating microstructure is evaluated by monitoring nitrogen adsorption-desorption at 77 K using SAW devices as sensitive microbalances. Gas and liquid phase AW sensor coatings based on zeolite/sol-gel, hydrous oxide ion exchange and surface derivatized sol-gel coatings have been demonstrated.

## REFERENCES

1. M. S. Nieuwenhuizen and A. Venema, "Surface Acoustic Wave Chemical Sensors," *Sensors and Materials*, 5 (1989) 261-300.
2. H. Wohltjen, "Chemical Microsensors and Microinstrumentation," *Anal. Chem.*, 56 (1984) 87A-103A.
3. S. L. Rose-Pehrsson, J. W. Grate, D. S. Ballantine, Jr. and P. C. Jurs, "Detection of Hazardous Vapors Including Mixtures Using Pattern Recognition Analysis of Responses from Surface Acoustic Wave Devices," *Anal. Chem.*, 60 (1988) 2801-2811.
4. S. J. Martin, A. J. Ricco, T. M. Niemczyk and G. C. Frye, "Characterization of SH Acoustic Plate Mode Liquid Sensors," *Sensors and Actuators*, 20 (1989) 253-562.
5. B. J. Costello, B. A. Martin and R. M. White, "Ultrasonic Plate Waves for Biochemical Measurements," *Proc. 1989 IEEE Ultrasonics Symp.*, 977-981.
6. C. J. Brinker, "Hydrolysis and Condensation of Silicates: Effects on Structure," *J. Non-Cryst. Solids*, 100 (1988) 31-50.
7. C. J. Brinker, A. J. Hurd and K. J. Ward, "Fundamentals of Sol-Gel Film Formation," *Ultrastructure Processing of Advanced Ceramics* (Wiley, New York, 1988) p. 223.
8. C. J. Brinker, G. C. Frye, A. J. Hurd, K. J. Ward and C. S. Ashley, "Sol-Gel Processing of Controlled Pore Films," *Proc. IVth Int. Conf. on Ultrastructure Processing of Glasses, Ceramics and Composites*, in press.
9. S. J. Gregg and K. S. W. Sing, *Adsorption, Surface Area and Porosity*, Second Ed., Academic, New York, 1982.
10. G. C. Frye, S. J. Martin, A. J. Ricco and C. J. Brinker, "Monitoring Thin-Film Properties with Surface Acoustic Wave Devices," *Chemical Sensors and Microinstrumentation* (ACS, Washington, 1989) Ch. 14.
11. C. L. Glaves, G. C. Frye, D. M. Smith, C. J. Brinker, A. Datye, A. J. Ricco and S. J. Martin, "Pore Structure Characterization of Porous Films," *Langmuir*, 5 (1989) 459-466.
12. T. Bein, K. Brown, C. J. Brinker, "Molecular Sieve Films From Zeolite-Silica Microcomposites," *Studies in Surface Science and Catalysis*, 49 (1989) 887.
13. G. C. Frye, A. J. Ricco, S. J. Martin and C. J. Brinker, "Characterization of the Surface Area and Porosity of Sol-Gel Films Using SAW Devices," *Mat. Res. Soc. Symp. Proc.*, 121 (1988) 349-354.
14. T. Bein, K. Brown, G. C. Frye and C. J. Brinker, "Molecular Sieve Sensors for Selective Detection at the Nanogram Level," *J. Amer. Chem. Soc.*, 111 (1989) 7640-7641.
15. R. G. Dosch, T. J. Headley and P. Hlave, "Crystalline Titanate Ceramic Nuclear Waste Forms. Processing and Microstructure," *J. Amer. Cer. Soc.*, 67 (1984) 354.
16. B. C. Bunker, C. H. F. Peden, S. L. Martinez, E. J. Braunschweig and A. K. Datye, "Hydrous Sodium Titanate Ion-Exchange Materials for Use as Catalyst Supports," *Characterization and Catalyst Development: An Interactive Approach* (ACS, Washington, 1989) Ch. 8.

# Electrically-Activated, Micromachined Diaphragm Valves

Hal Jerman
IC SENSORS
Milpitas, CA 95035

## Abstract

Electrically-activated diaphragm valves have been fabricated using heated bimetallic structures to provide the operating force. These valves can be designed to provide fully proportional control of flows in the range of 0-300 cc/min at input pressures from near zero to over 100 PSIG. The valves are batch fabricated using silicon micromachining techniques. By combining these valves with pressure or flow sensing elements, closed loop pressure or flow control is easily accomplished.

## Introduction

By using advanced silicon micromachining techniques, high-performance valves have been developed with integral actuators to control gas pressure or flow by the application of an electrical signal  These devices provide the electrical-to-pressure function which complements the pressure-to-electrical function offered by the millions of monolithic silicon pressure sensors fabricated yearly.

Conventional valves for pressure and flow applications have typically used magnetic actuation in the form of solenoids or motors to drive diaphragm or needle-type valves. In a miniature, monolithic silicon-based valve, however, magnetic actuation is not attractive due to the difficulties involved with providing sufficient force  Electrostatic valve actuation has been reported recently,[1] but the forces involved are large only for small electrode gaps. Additionally, only attractive force is feasible, and the $1/x^2$ nature of the force makes fully proportional control difficult  This particular device allowed only low pressure operation, with pressures over about 2 PSIG causing the valve to open even with voltage applied.

Electrostatic actuation is possible using piezoelectric drivers, but the force and deflection properties of materials such as PZT produce either high forces with small deflection for button type actuators, or relatively low forces with high deflection for bimorph type drivers. A button-type driver has been used as a valve actuator by Esashi at Tohoku University, but the complicated assembly and high voltage operation are unattractive for most commercial applications.

For very small structures, thermal actuation becomes more attractive due to the geometrical advantages of scaling. The amount of thermal mass decreases as the volume of material decreases, the thermal loss decreases as the thicknesses decrease, while the forces per unit area remains high.  A thermally activated valve using the vapor pressure generated in a heated fluid has been made by Zdeblick at Stanford University.[2] This valve is thermally inefficient due to the large volume of heated fluid, has difficult assembly problems involved with sealing the working fluid in the device, and is restricted to normally-open configurations.

Bimetallic beams[3] have been shown to provide significant force and displacement as integrated actuation elements, however the previously described devices have shown insufficient force to use in most valve applications. The use of bimetallically-driven diaphragms increases considerably the force available from an integrated structure while simultaneously achieving large displacements and symmetric, vertical travel. Although a large variety of combinations of materials can be used, the use of a silicon diaphragm and an aluminum metallic layer is one of the most attractive.

## Bimetallic Diaphragm Valve Structure

These valves, such as shown in Fig. 1, consist of a diaphragm actuator with a central boss which mates to an etched silicon valve body. The actuator consists of a circular silicon diaphragm with integral diffused resistors, which acts as one element of the bimetallic structure. An annular aluminum region on that diaphragm forms the other element. By varying the electrical power dissipated in the resistors and thus the temperature of the silicon diaphragm, the displacement of the central boss can be controlled. The actuator chip and the valve body chip make a very compact valve structure without the large, external actuators common in previously developed valves.

In one design a nominally circular silicon diaphragm has been used, and this valve closes with increasing temperature of the diaphragm, making a normally-open valve. A photograph of such a diaphragm structure is shown in Fig. 2.
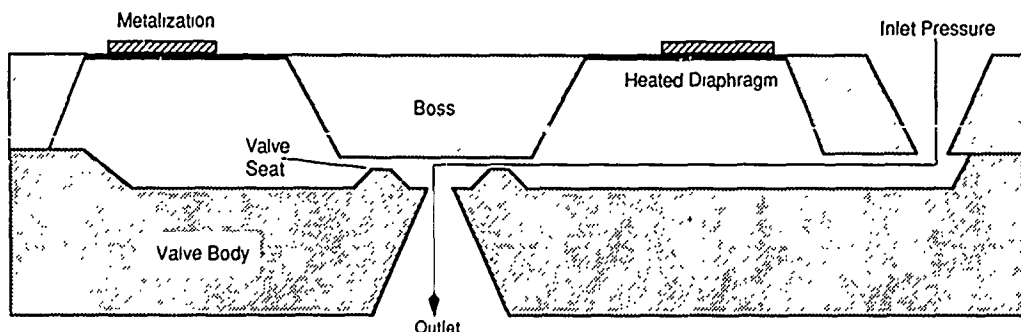


Fig. 1 Cross Section of a Micromachined, Bimetallically-Actuated Diaphragm Valve.
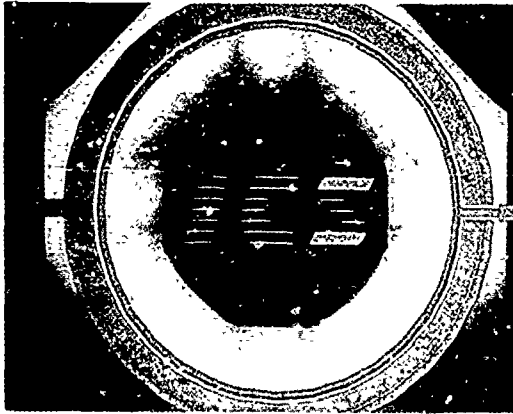
Figure 2. Photograph of typical bimetallic diaphragm with partial backside illumination.

An anisotropic silicon etch has been used to form the diaphragm, the extent of which can be seen from the light transmitted through the diaphragm. The corner compensation has been designed to provide an octagonal approximation to the outer diameter, while the inner diameter achieves a very close approximation to the desired circular shape. The star pattern visible from the center boss is the remnant of the original corner compensation in this non-etch-stopped diaphragm. The thick aluminum layer extends over the outer portion of the diaphragm. A diffused resistor is provided under the aluminum layer, and the aluminum is also used to provide electrical connection to the heater. The valve seat is on a mating silicon wafer, and is very similar to the original valve structure developed by Terry at Stanford, now almost 18 years ago.[4]

The temperature of the bimetallic structure on the diaphragm is used to control the force applied to the central diaphragm boss. Thus, the gap between the valve seat and the bottom surface of the diaphragm boss can be varied, controlling the flow of gas through the valve. There is a gas feedthrough hole in the bimetallic diaphragm chip to allow gas flow through the structure and to equalize the pressure drop across the diaphragm. This minimizes the force necessary to operate the valve.

The flow characteristics of this valve are controlled by the diaphragm and metal thicknesses, the material properties of the metals, the outside diameter, boss ratio, and metal coverage on the diaphragm, and the geometry of the valve seat. The temperature of the bimetallic diaphragm is controlled by varying the power dissipated in heaters either on or in the diaphragm. Gas flows of more than 100 cc/min at pressures over 30 PSIG have been controlled with such a valve, with on/off flow ratios greater than 1000. Prototype valves have been cycled millions of times with no observable change in performance. These valves can be operated as either on-off valves or as proportional control valves by changing the flow geometry and dimensions of the valve elements. Other aspects of the valve operation, such as speed, power dissipation, and magnitude and direction of deflection can be controlled by adjusting the suspension characteristics of the diaphragms.

It is of crucial importance in the design and fabrication of these devices to adequately model the overall flow as a function of applied power to allow the design of this type of valve for a particular application. The analysis of the valves is complicated by the interaction between the gas flow, the diaphragm deflection, and the bimetallic force. These complicated, linked analytical models can now be easily handled using computer programs which numerically solve the set of simultaneous non-linear equations. The individual models used in combination to predict the valve characteristics are described below.

## Bimetallic Diaphragm Model

A bimetallic structure consists of a pair of materials, not necessarily metals, with different coefficients of thermal expansion, which are bonded together. As the temperature of the composite is changed, stresses are generated in the structure which can cause useful forces and deflections to be generated by the element. The original analytical model was developed by Timoshenko in a classic paper in 1925.[5] Schematically the structure is shown below:
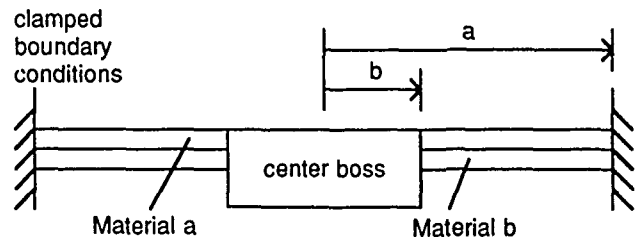


The deflection of this structure at zero force is given by:[6]

$$y(Q=0) = \frac{K_y}{(1+v)} \Theta a^2$$

where a is the radius of the diaphragm and $K_y$ is a constant which depends on the bossed ratio, b/a, and $\Theta$ is the temperature term:

$$\Theta = \frac{6(\gamma_b - \gamma_a)\Delta T(t_a + t_b)(1+v_e)}{t_b^2 K_1}$$

where $\gamma_a$ and $\gamma_b$ are the thermal expansion coefficients of the two materials, $t_a$ and $t_b$ are the respective thicknesses, $v_e$ is the effective Poisson ratio of the composite beam, and $K_1$ depends on the relative stiffness of the two components:

$$K_1 = 4 + 6\frac{t_a}{t_b} + 4\left(\frac{t_a}{t_b}\right)^2 + \frac{E_a t_a^3(1-v_b)}{E_b t_b^3(1-v_a)} + \frac{E_b t_b(1-v_a)}{E_a t_a(1-v_b)}$$

where $E_a$ and $E_b$ are the moduli of elasticity of the two components. The line load at the inner radius, b, at zero displacement is given by:

$$Q_{(y=0)} = \frac{K_q \Theta D_e}{(1+v)\, a}$$

where $K_q$ is a similar geometrical constant and $D_e$ is given by:

$$D_e = \frac{E_a t_a^3}{12(1-v_a^2)} K_{2p}$$

with:

$$K_{2p} \approx 6 + \frac{E_b t_b^3 (1-v_a^2)}{E_a t_a^3 (1-v_b^2)}$$

and the total force is just the line load times the inner circumference:

$$F = Q\, 2\,\pi\, b$$

The factors $K_y$ and $K_q$ are complicated functions of the bossed ratio (b/a) and $R_0$, the radius beyond which the diaphragm is heated. In this analysis the diaphragm is assumed to be flat at a diaphragm temperature equal to the temperature of the device during metal deposition. Thus by using the appropriate initial metal tension, the initial displacement of the diaphragm can be controlled.

There are a number of different bimetal combinations possible for the structure, but one pair that is attractive is aluminum and silicon. The force and deflection are maximized for pairs which have a large difference in thermal expansion coefficients. It is often desirable to use materials with low thermal conductivities to limit power loss through conduction. Unfortunately both silicon and aluminum have high thermal conductivities, resulting in relatively large power dissipation when these materials are used with thick diaphragms.

For example, for a clamped diaphragm, 2.5 mm in diameter with a 9 μm silicon thickness, a 6 μm aluminum thickness, and a boss ratio (b/a) equal to 0.4, the following deflection characteristics are obtained:
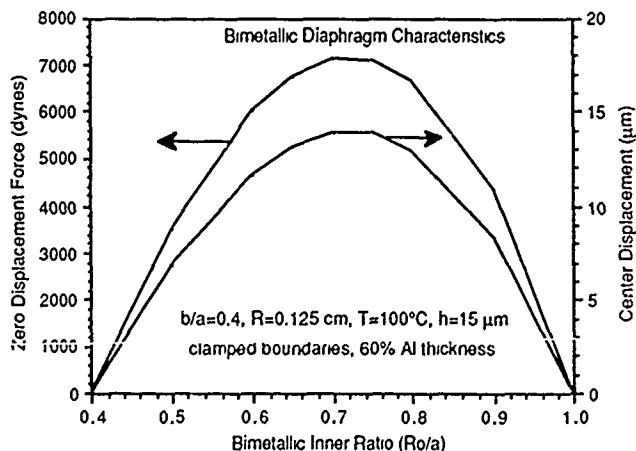


Figure 3. Force and displacement characteristics of a bimetallic diaphragm structure.

Here the deflection and force are maximized for aluminum which covers only about the outer half of the annular diaphragm. Note that for aluminum covering the entire diaphragm there is no force or deflection. The maximum force is about 7100 dynes and the maximum deflection is 13.3 μm. This deflection is downward for the aluminum on the top surface of the device.

Thinner diaphragms provide more deflection at lower force, while thicker diaphragms provide more force at smaller deflections. A silicon thickness between about 6 and 12 μm is optimum for most valve applications.

These first order bimetallic models do not fully take into account the difference in thermal expansion between the heated diaphragm and the supporting frame. In the clamped case the actual boundary condition is described as clamped but moveable. With true clamped boundary conditions, there is a possibility that thin diaphragms will buckle at high temperatures.

## Diaphragm Deflection

Flat silicon diaphragms are generally considered to have linear deflection vs. force characteristics. While this is true for the limited deflection of, for example, pressure transducers, in actuator structures the deflection is often a substantial fraction of the thickness of the diaphragm, where the non-linearities become significant. The deflection of any circular diaphragm can be described by an equation with the form:

$$\frac{Q a^2}{\pi E h^4} = A \frac{y}{h} + B \frac{y^3}{h^3}$$

where Q is the applied force, a is the diaphragm radius, E is Young's modulus, h is the diaphragm thickness, and A and B are coefficients which depend on the boss ratio. For non-bossed, flat, clamped diaphragms the value of A is generally taken to be 1.38 and B, 0.41. For bossed structures the diaphragm becomes much stiffer for a given thickness; for a boss ratio of 0.6 each coefficient is increased by a factor of about 20.

## Flow Model

The gas flow through the valve is determined by summing the appropriate forces on the valve diaphragm, which determines the deflection of the center of the diaphragm from the relation above. The flow through the valve is then assumed to be due to the restriction formed by the channel between the seating ring and the diaphragm boss. The volumetric, laminar flow through that channel is given by:

$$flow = \frac{gap^3\; w\,(P_{in}^2 - P_{out}^2)}{24\,\eta\,l\,P_{out}}$$

where w is the effective width of the seating ring, $\eta$ is the gas viscosity, $P_{in}$ and $P_{out}$ are the absolute input and output pressures, and l the effective length of the seating ring.

This equation is only valid for laminar flow through the channel. The Reynold's number of the flow can be examined

to determine if the flow is likely to be laminar. For a nitrogen flow velocity of 30,000 cm/s, nearly at the speed of sound, in a 5 μm channel, at a pressure of one atmosphere, the Reynold's number is only 79, well under the value of 2000 usually taken to be the onset of turbulent flow. Since the channel is 200 μm long, the flow should become well developed not far down the channel.

The average velocity at the output of the channel is found by dividing the volumetric flow given above by the area of the channel. The peak velocity is twice the average velocity for the parabolic flow profile found in laminar flow. For many combinations of high input pressure, low output pressure, and high flow, the average output flow velocity is found to exceed the speed of sound, using the laminar flow equation shown above. It is not possible, however, to obtain supersonic flow in a rectangular channel; that is only possible in a convergent-divergent nozzle.[7] A recent, computational method for the analysis of subsonic, compressible channel flows is available, but it does not extend to the sonic case.[8] Flow through a sonic nozzle becomes only a function of the upstream pressure and the area of the nozzle, at constant temperature. Thus the flow characteristic of the valve changes from a cubic dependence on gap to a linear dependence on gap when sonic flow conditions are achieved.

### Experimental Results

A normally open valve has been fabricated using the 2.5 mm diameter actuator shown in Fig. 2. It is mated to a valve body with an annular seat about 400 μm in diameter and 80 μm wide. The gas flow through this structure is typically from the topside of the structure, through a hole connecting the topside of the actuator diaphragm and the backside, across the valve seat, and through a hole in the valve body to the bottom of the valve assembly. By providing the inlet pressure to the topside of the actuator, the pressure tends to help close the valve, reducing the off-flow through the structure at high pressures. When closed, the valve will not suddenly open with inlet pressure pulses, as is the case with the electrostatic valve described by Ohnstein et. al

The valve flow at a variety of inlet pressures is shown in Fig. 4. This valve has a 10 μm thick diaphragm, a 5 μm thick
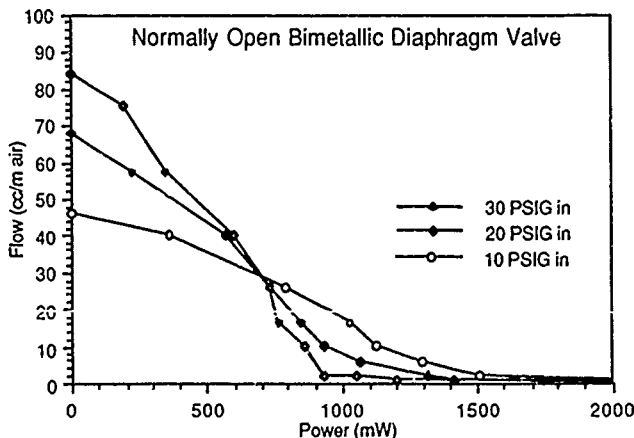


Figure 4. Experimentally determined flow characteristics of a valve with 2.5 mm diameter and 10 μm thick diaphragm.

aluminum layer, and a nominal 4 μm initial gap when the actuator is at room temperature.

Note that with 30 PSIG inputs, the valve actually closes earlier than with lower inlet pressures due to the action of the inlet pressure in helping to close the valve. The valves are fully proportional, with no obvious hysterysis. The leakage flow through this device at 20 PSIG input was about 0.045 cc/min, about a factor of 1,600 less than the full flow.

By using a combination of the models described earlier, the theoretical flow characteristics of such a valve structure can be calculated. The results are shown in Fig 5.
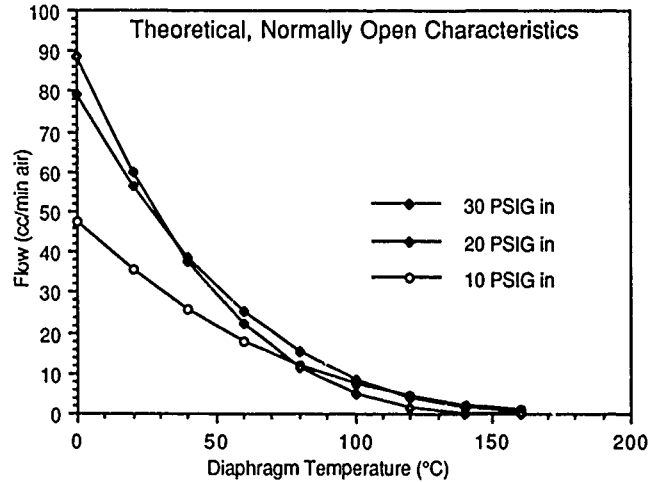


Figure 5. Theoretical flow characteristic of normally open valve structure.

The agreement with the experimental results is very encouraging. Here the initial gap has been selected to roughly match the room temperature 30 PSIG flow. The relatively thick diaphragm and small initial gap of this design results in a low flow, high pressure valve with relatively high power dissipation. By adjusting these parameters and the size of the diaphragms, a family of valves can be made with more optimum low pressure and high flow characteristics.

### Pressure Regulator

A similar valve has been operated in a closed loop pressure regulator system to show the ease in using these valves as a computer controlled pressure regulator. The valve was placed between a 30 PSIG gas source and a variable downstream restrictor. The difference between an electrical setpoint and the output of a silicon pressure transducer was used as the error signal to drive the valve open or shut. The restriction was changed to vary the flow through the valve while the electronics system maintained a constant pressure at the head of the restrictor. The results of one such test are shown in Fig. 6.
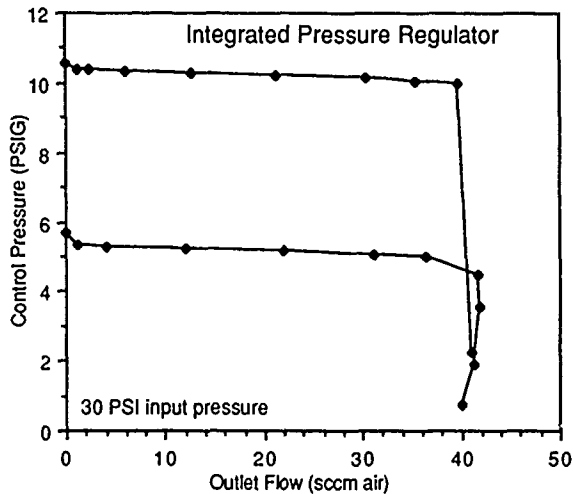
Figure 6. Characteristic of a closed-loop pressure controller using micromachined valve.

Here the electrical setpoint was set to about 10.2 and 5.1 PSI with a 30 PSI inlet pressure. The device provided accurate pressure regulation over a flow range from less than 1 cc/min to about 35 cc/min. Flows greater than 35 cc/min were above the maximum flow through this particular valve at this inlet pressure and thus the output pressure fell to a low value as the outlet restriction was reduced further.

## Conclusions

The use of bimetallic driven diaphragms has been shown to provide valve structures with flow and pressure ranges which are useful in a wide variety of applications. The characteristics of the valves can be greatly varied by changing the geometry of the structure, which can be used to optimize a valve for a particular application. The valves are batch fabricated and many hundreds of valves can be made on the same wafer, offering the possibility of a very low cost device. The theory, materials, and fabrication technologies are well understood, which shortens the development time for these devices. It is expected that these valves will greatly influence the types of integrated flow control systems available in the '90's.

## References

1  T. Ohnstein, T. Fukiura, J. Ridley and U. Bonne, Micromachined Silicon Microvalve, Proc. IEEE Workshop on Micro-Electro Mechanical Systems (MEMS), pp. 95-98, Feb. 1990.

2  M. J. Zdeblick and J.B. Angell, A Microminiature Electric-to-Fluidic Valve, Proc. 4th Int. Conf. on Solid-State Transducers and Actuators, June 1987, pp 827-829.

3  W. Reithmuller and W. Benecke, Thermally Excited Silicon Microactuators, IEEE Trans. Elec. Dev., v 35 n 6, 1988

4  S.C. Terry, A Gas Chromatography System Fabricated on a Silicon Wafer Using Integrated Circuit Technology, Stanford ICL Technical Report #4603-1, May 1975, p. 41.

5  S.P. Timoshenko, Analysis of Bi-Metal Thermostats, J. Optical Soc. Am., V. 11, 1925, pp. 233-255.

6  R. Roark and W. Young, Formulas for Stress and Strain, McGraw-Hill, 1975, p.165.

7  See, for example, Owczarek, J.A., Fundamentals of Gas Dynamics, Int. Textbook, 1964.

8  Schwartz, L.W., A Pertubation Solution for Compressible Viscous Channel Flows, J. Eng. Math., vol. 21, p. 69-86, 1987.

# VERIFICATION OF FEM ANALYSIS OF LOAD-DEFLECTION METHODS FOR MEASURING MECHANICAL PROPERTIES OF THIN FILMS

Jeffrey Y. Pan, Pinyen Lin, Fariborz Maseeh,
and Stephen D. Senturia

Microsystems Technology Laboratories

Massachusetts Institute of Technology, Cambridge, MA

## ABSTRACT

Accurate models are an essential element for determining the mechanical properties of thin films from load-deflection experiments. Analytical models are desirable because of their simplicity. Finite element method (FEM) models have the potential to be more accurate. Through an extensive FEM analysis of the load-deflection methods, we have confirmed that while the functional form of the analytical result is correct, three constants in the model must be corrected by as much as 30%. Experimental measurements of the deformed membrane shape have been made and they match the FEM results, verifying the accuracy of the FEM models. Experimental values, extracted from load-deflection analysis, for the biaxial modulus and the residual stress of thin films of Dupont PI2525 and Hitachi PIQ13 are presented.

## INTRODUCTION

The load-deflection method has previously been developed for the measurement of the mechanical properties of thin films [1-6]. In this technique, the deflection of a suspended film is measured as a function of applied pressure (Fig. 1). The biaxial modulus and the residual stress of the film can then be extracted from the data using various mathematical models. The models for both square and circular membranes have been developed using both analytical and finite-element methods (FEM).
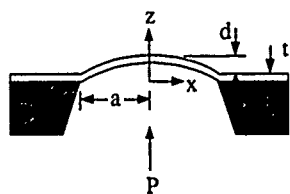


Figure 1: Deflection of a suspended membrane in
response to an applied pressure.

In the case of the analytical model, a functional form of the deflected shape is assumed, and the total strain energy minimized to find the load-deflection behavior. A hemispherical cap is assumed as the deflected shape of the circular membranes [2], whereas a cos(kx)cos(ky) form is assumed for the square membranes [7-8]. These choices are often made to simplify the mathematics at the expense of modeling accuracy.

The FEM models are generated by inputting a set of geometric and material parameters into a FEM program and simulating a load-deflection experiment. By noting the effect of variations of each material parameter on the simulation, a model can be derived.

In all cases the load-deflection behavior is of the form:

$$P = \frac{C_1 t}{a^2} \sigma_o d + \frac{C_2 f(v) t}{a^4} \frac{E}{1-v} d^3 \qquad (1)$$

where P is the applied pressure, d the center deflection, a the radius/half-edge length of the membrane, t the thickness, E the Young's Modulus, v the in-plane Poisson Ratio, $\sigma_0$ the residual stress, and E/(1-v) the biaxial modulus. The dimensionless constants $C_1$ and $C_2$ and the dimensionless function f(v) are geometry and model dependent.

It can be seen from (1) that the biaxial modulus and residual stress can be evaluated from the experimental pressure-deflection data and the membrane dimensions provided that $C_1$, $C_2$, and f(v) are known. Accurate values for these constants are therefore essential for the accurate determination of the mechanical properties. The choice of modeling method affects the values of these constants. Our goal in this study is to show that in contrast to the analytical models, which have to assume a shape function, the FEM model yields the shape function, and this shape function is in good agreement with experiment.

## ANALYTICAL MODELS

In the analytical model, a functional form of the deflected shape must be assumed and the total strain energy minimized to find the load-deflection behavior. For square membranes, a one-term Fourier approximation of the actual deflected shape can be assumed. The form is from Allen [7-8]:

$$z = d \cos\left(\frac{\pi x}{2a}\right) \cos\left(\frac{\pi y}{2a}\right) \qquad (2)$$

where x and y are the distances in the x- and y-axes away from the center, z is the deflected height at (x,y), a is the half-edge length, and d is the center deflection.

The analytical shape for the circular case is assumed to be a hemispherical cap [2]. The equation is:

$$z = d - R + (R^2 - r^2)^{1/2} \qquad (3)$$

where r is the distance from the center, R is radius of curvature of the deflected membrane, and d is the center deflection. The value of R can be calculated from

$$R = \frac{a^2 + d^2}{2d} \qquad (4)$$

where a is the radius of the membrane.

With these assumptions for the deflected shape, the minimization of the strain energy in the presence of a residual tensile stress is straightforward [2,7-8]. The resulting values for $C_1$, $C_2$, and f(v) are shown in Table II. Note that a simplified linear fit to the actual f(v) for square membranes is given in the table. The actual equation can be found in the references [4,7-8].

## FINITE ELEMENT METHOD

Finite element modeling of square and circular suspended membranes was carried out using both ABAQUS and ADINA. Agreement between the two programs was excellent. In order to check the consistency of the FEM models, several types of elements were used to solve this problem, and their results were compared.

The types of elements and the number of nodes selected for this work are shown in Table I. Three different types of elements were used for square membranes: 4-node and 16-node shell elements, and 20-node 3D solid elements. For circular films, 4-node and 16-node shell elements and 8-node 2D solid axisymmetric elements were used. Fig. 2 shows the schematic geometry of the different elements. Nonuniform grids were also used in some cases to put more nodes near the edges of the membranes. These elements are labeled as ratioed elements in Fig. 2.

TABLE I
The element types and the number of nodes per element used in FEM analysis.

| membrane shape | element type | nodes per element | number of elements | ratioed elements |
|---|---|---|---|---|
| circular | shell | 4 | 1200 | no |
| | shell | 16 | 105 | yes |
| | 2D solid | 8 | 960 | no |
| square | shell | 4 | 324 | no |
| | shell | 16 | 64 | yes |
| | 3D solid | 20 | 256 | yes |

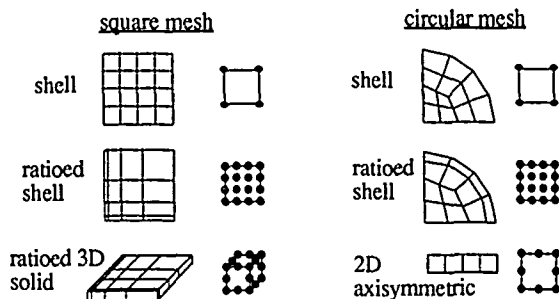square mesh                    circular mesh



Figure 2: The schematic geometry of FEM elements.

Ratioed elements were used for the 16-node shell and the 20-node 3D solid elements in order to accurately model the behavior of the membrane near its boundaries. The smallest size for the shell elements was 0.047mm or about 1.2% of the half edge length; for the 3D solid elements it was about 0.058mm or 1.5%. The deflected shape for a square membrane as a function of element type is shown in Fig. 3. The variation in both the deflected shape and the center deflection for the different elements is less than 0.3%.

The results for a circular membrane are shown in Fig. 4. Again the variation in the results is less than 0.3%. This indicates that our FEM analysis converges to the same answer regardless of the element type. We found the best balance between accuracy and computation time to be a grid of 576 4-node shell elements for square membranes and 250 8-node 2D solid elements for circular membranes.
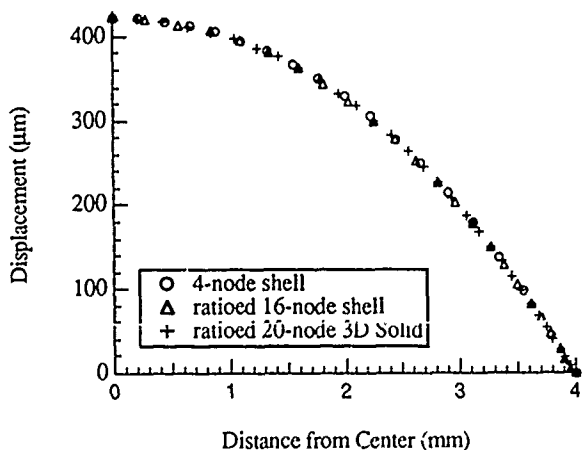


Figure 3. FEM-calculated deflected shape along center-to-midside for a square membrane as a function of element type.
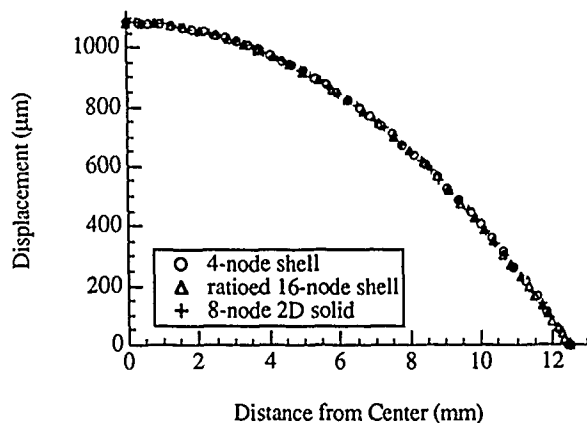


Figure 4: FEM-calculated deflected shape along radius of a circular membrane as a function of element type.

By computing the deflection vs. pressure for assumed values of $\sigma_o$, E, and $\nu$ for a sequence of applied loads, and fitting the results to Equation (1), it is possible to determine $C_1$ and the product $C_2 f(\nu)$. For convenience, we assign $C_2$ as the value assigned to the analytical model when $\nu=0.25$; i.e. $f(\nu)$ is scaled such that $f(\nu=0.25) = 1$. The results of extensive fittings are shown in Table II. The values are accurate to within 1%.

For the circular case, the FEM value for $f(\nu)$ differs from the analytical value by 12% at $\nu=0.4$. $C_1$ and $C_2$ are the same for both models. For square membranes, the FEM and analytical values for $C_1$ differ by 12% while $f(\nu=0.4)$ varies by 33%.

TABLE II
Modeled values of $C_1$, $C_2$, and $f(\nu)$ for both square and circular membranes.

| Membrane Shape | Model | $C_1$ | $C_2$ | $f(\nu)$ |
|---|---|---|---|---|
| Circular | Analytical | 4.0 | 2.67 | 1.0 |
| | FEM | 4.0 | 2.67 | $(1.026 + 0.233\nu)^{-1}$ |
| Square | Analytical | 3.04 | 1.37 | $1.075 - 0.292\nu$ |
| | FEM | 3.41 | 1.37 | $1.446 - 0.427\nu$ |

## EXPERIMENTAL PROCEDURE

### Fabrication of Samples

Square membranes of BTDA-MPDA/ODA (Dupont PI2525) were fabricated using micromachining techniques [5]. A combination of oxide mask, $p^+$ etch stop, and KOH etchant was used to define a thin silicon membrane on an <100> n-type silicon wafer. The polyimide was then spun cast onto the wafer and the silicon membrane etched in an $SF_6$ plasma. The final structure was cemented to an aluminum plate for pressure testing.

Circular membranes were made from Hitachi PIQ13 films using a variation of the technique outlined in [6]. PIQ13 was spun cast onto a bare silicon wafer. Using 6:1:1 $HF:HNO_3:CH_3COOH$, a hole was then etched through the center of the wafer, leaving a suspended polyimide membrane. The membrane was then transferred to a machined Vespel ring by epoxying the ring to the membrane and cutting it free from the silicon support   Finally, the ring was cemented to an aluminum plate to facilitate testing.

## Measurement of the Mechanical Properties

The thicknesses of the films were taken from Dektak measurements. The dimensions of the membranes were measured with a Nikon UM-2 measuring microscope equipped with two Boeckeler model 9598 digital micrometers connected to a Metronics Quadra-Chek II digital readout box. The maximum error in each axis for the calibrated xy stage was less than 3 microns.

The samples were clamped into a custom-designed pressurizing jig for testing. Computational routines inside the Quadra-Chek II were used to locate the center of the membrane under test. The deflection resulting from varying pressure loads for each membrane was then measured with the microscope [5-8]. A 40X objective with a numerical aperture of 0.5 provided the sub-micron depth of focus needed to locate the surface of the membranes. A 543-series Mitutoyo digimatic indicator was used to track the z-axis movement of the microscope head. The total measurement error in the deflection was less than 2 microns. Pressure readings accurate to 0.02 psi were made with a MICRO SWITCH 142PC05G pressure sensor.

All measurements were done at room temperature in dry air (dew point < -46° C). The strain in the sample was never allowed to exceed 1%. Ten to twenty pressure-deflection measurements were taken, and the points were fit to equation (1) using the FEM constants to determine the residual stress and biaxial modulus.

A ν of 0.4 was assumed for all calculations. We can check this assumption by comparing the Young's modulus predicted by this assumption against published values. The Young's modulus of PI2525 has been measured from uniaxial tests to be 3.2 ± 0.16 GPa [6]. The predicted biaxial modulus of 5.22 GPa combined with the assumed Poisson ratio of 0.4 yields a value of 3.13 GPa for E, so our computations are consistent. The results for two membranes are shown in Table III. The accuracy of the values is estimated to be ±5%.

TABLE III
Measured residual stress and biaxial modulus for
two membranes.

| Membrane Shape and Dimensions | Residual Stress (MPa) | Biaxial Modulus (GPa) |
|---|---|---|
| Square t = 5.2 μm, a = 4826 μm (Dupont PI2525) | 32.2 | 5.22 |
| Circular t = 11.4 μm, R = 12610 μm (Hitachi PIQ13) | 35.2 | 5.37 |

## Measurement of Deflected Shapes

The same apparatus as above was used to measure the deflected profiles. The calibrated xy stage allowed us to locate the center of the membranes and accurately translate the sample along the appropriate path. The results are plotted against the modeled results in Figures 5-6. The residual stress and biaxial modulus derived using the indicated model and the measured thickness, edge length/radius, and pressure were used to generate the modeled results.
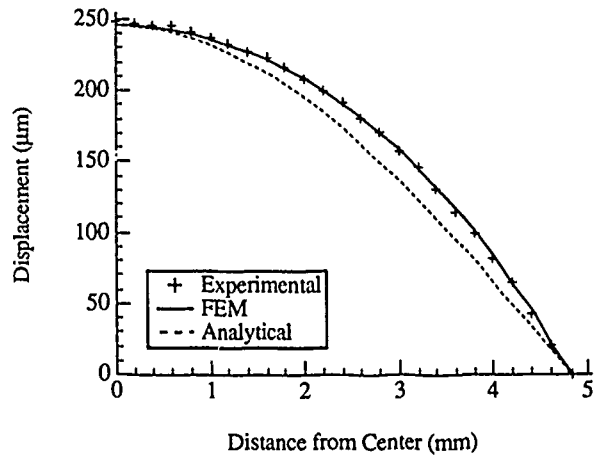


Figure 5a: Experimental deflected shape vs. FEM results and analytical model for center-to-midside of square membrane.
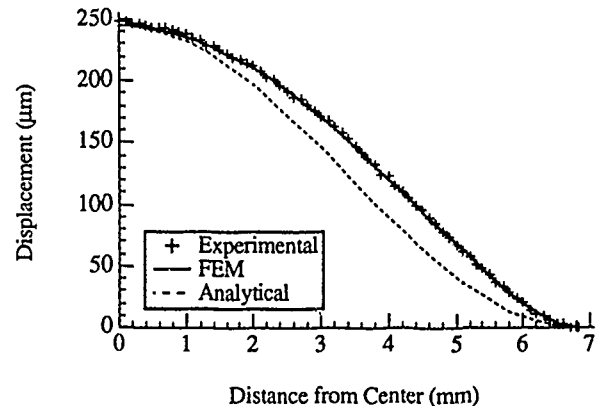


Figure 5b: Experimental deflected shape vs FEM results and analytical model for diagonal of square membrane.
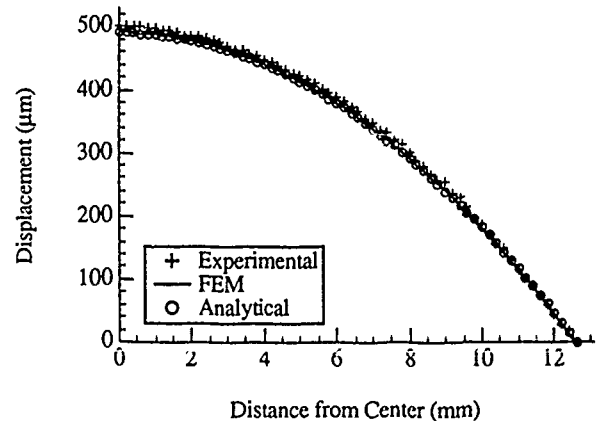


Figure 6: Experimental deflected shape vs FEM result and analytical model for radius of circular membrane.

## DISCUSSION

The center-to-midside deflected shape for a square membrane is shown in Fig. 5a; the deflected shape along the diagonal is given in Fig. 5b. The agreement between the experimental data and the FEM results is within 0.5% in both cases. The difference between the experimental data and the analytical solution is over 10% at the x-position of 4000µm.

Figure 6 shows the predicted and measured deflected shapes for circular membranes. Note that both the FEM result and the analytical result are within 3% of the experimental data. This indicates that the actual deflected shape is very close to a hemispheric function.

On the basis of this experimental confirmation of the FEM-predicted deflected shapes, we can conclude that the FEM values of $C_1$, $C_2$, and $f(v)$ given in Table II are the correct values to use for load-deflection analysis.

## CONCLUSION

Through a careful study of FEM modeling accurate values for $C_1$, $C_2$, and $f(v)$ in Eqn. (1) have been found. These values are substantially different from those predicted by analytical models. By direct experimental observation, we have verified that FEM modeling yields the correct deformed shape, and, hence the values of the constants obtained from FEM provide a reliable basis for using load-deflection data to measure the mechanical properties of materials.

## ACKNOWLEDGEMENT

## REFERENCES

[1] E. I. Bromley, J. N. Randall, D. C. Flanders, and R. W. Mountain, "A technique for the determination of stress in thin films," J. Vac. Sci. Technol. B, vol. 1, no. 4, pp. 1364-1366, Oct.-Dec. 1983.

[2] J. W. Beams, "Mechanical Properties of Thin Films of Gold and Silver," in Structure and Properties of Thin Films, C. A. Neugebauer, Ed. New York: Wiley, 1959, pp. 183-192.

[3] J. A. Hinkley, "A Blister Test for Adhesion of Polymer Films to SiO$_2$," J. Adhesion, vol. 16, pp. 115, 1983.

[4] O. Tabata, K. Kawahata, S. Sugiyama, and I. Igarashi, "Mechanical Property Measurements of Thin Films Using Load-Deflection of Composite Rectangular Membrane," Micro Electro Mechanical Systems: An Investigation of Micro Structures, Sensors, Actuators, Machines, and Robots, IEEE,1989, pp.152-156.

[5] M. G. Allen, M. Mehregany, R. T. Howe, and S. D. Senturia, "Microfabricated structures for the in situ measurement of residual stress, Young's modulus, and ultimate strain of thin films," Appl. Phys. Lett, vol. 51, no. 4, pp. 241-243, July 1987.

[6] F. Maseeh and S. D. Senturia, "Elastic Properties of Thin Polyimide Films," in Polyimides: Materials, Chemistry and Characterization, C. Feger, M. M. Khojasteh and J. E. McGrath, Eds. Amsterdam: Elsevier Science Publishers B. V., 1989, pp. 575-584.

[7] M. Mehregany, M. G. Allen, and S. D. Senturia, "The Use of Micromachined Structures for the Measurement of Mechanical Properties and Adhesion of Thin Films," Technical Digest, IEEE 1986 Solid-State Sensors Workshop, Hilton Head, S. C., June 1986.

[8] M. G. Allen, M. S. thesis, Department of Chemical Engineering, Massachusetts Institute of Technology, May 1986.

[9] M. G. Allen and S. D. Senturia, "Analysis of Critical Debonding Pressures of Stressed Thin Films in the Blister Test," J. Adhesion, vol. 25, pp. 303-315, 1988.

# BUBBLE FORMATION DURING SILICON WAFER BONDING:

## CAUSES AND REMEDIES

K.Mitani, V.Lehmann, and U.Gösele

School of Engineering, Duke University, Durham, NC., 27706,
USA

## ABSTRACT

Silicon wafer bonding is a new promising technique in silicon micromechanics. Unbonded areas, called voids or "bubbles", at the interface between silicon wafers limit the applicability of wafer bonding. We have investigated bubble generation and annihilation mechanisms occuring during silicon wafer bonding.

## INTRODUCTION

Silicon wafer bonding has recently been developed for silicon based electronic and micromechanical devices. SOI (Silicon-On-Insulator) appears to be the ideal structure for future integrated submicron circuit fabrications. Wafer bonding is one of the most promising technologies to fabricate SOI wafers [1,2]. In silicon micromechanics, the mechanical and electronic advantages of single crystal silicon are combined. For many micromechanical devices, two pieces of silicon have to be connected, which is usually accomplished by an intermediate glass layer having the same thermal expansion coefficient as silicon. This intermediate glass layer can be eliminated if silicon wafer bonding is used. Silicon wafer bonding also allows to fabricate more complicated or miniaturized micromechanical devices [3]. For all the above mentioned applications, the bonding has to be homogeneous over the entire interface without the presence of unbonded areas, called voids or "bubbles". Existence of bubbles does not only decrease the average bonding strength but also limits the device yield. This paper reports several causes of bubble formation and will especially concentrate on the role of surface contamination and on methods of reducing bubble formation. For our experiments, we have used unstructured polished single crystal silicon wafers although we are aware that for actual micromechanical applications, frequently wafers with grooved surfaces have to be bonded.

## CAUSES FOR BUBBLE FORMATION

There are four causes for bubble formation.

i) Insufficient wafer flatness: As we reported previously [4], the closing of gaps between wafers may be calculated based on elastomechanics. Let us assume that both wafers have flatness variations on a length scale r at the same location leading to a gap of height h between the wafers. Then the condition for closing the gap by bonding is:

$$h/r^2 < \sqrt{\sigma/(Ed^3)}$$

where $\sigma$ is the interface energy per area due to bonding, E Young's modulus and d the thickness of the individual wafers. It turns out that the flatness requirements for the best commercially available wafers are stringent enough to fulfill the above condition. T. Abe et al. [5] investigated the relation between surface roughness grade and bubble formation and came to the same conclusion.

ii) Trapped air: When the rim area of wafers is bonded first, air may be trapped between the wafers. Therefore, bonding should be initiated at some point on the wafers by pressing them together so that bonding can be propagated radially and no air is trapped.

iii) Particulates: Particles prevent locally the bonding of silicon wafers. In order to avoid particles, generally an ultra-clean room (class 1 or better) is required. However, as reported previously [6], at Duke University we developed a micro cleanroom set-up which does not require a cleanroom facility. Fig. 1 shows this micro cleanroom set-up schematically. After two wafers are placed face to face with 0.5 - 1.0 mm gap ensured by appropriate spacers, the wafers are flushed by deionized water. Then the wafers are covered with a transparent cover and
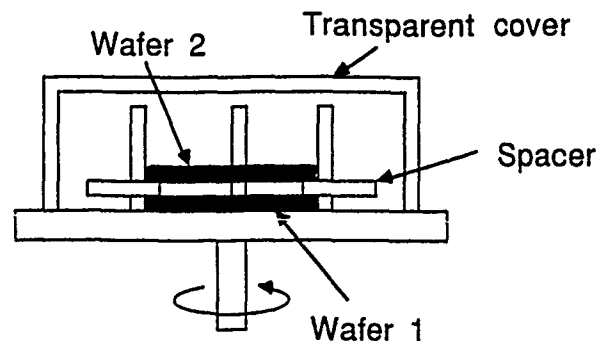


Fig. 1    Front view of a micro cleanroom. After flushing wafers by DI water, the cover is closed and the wafers are dried by spinning.

spin-dried with 2500 - 3000 rpm while illuminating the wafers with an infrared lamp. The whole process takes about 5 minutes. After drying, the spacers are removed without opening the cover so that the wafers can be bonded before particles can enter the space between the wafers. So far with this machine, bubble-free bonding of from 3 inch to 8 inch diameter silicon wafers has been accomplished successfully. In addition, GaAs and InP wafers were also bonded on silicon wafers [7].

iv) Surface contamination. In order to increase the initial bonding strength, wafers have to be annealed after the bonding process. As Ohashi et al. [8] had reported, when bonded wafers, which do not show bubbles induced by one of the above three causes, are annealed in the 200 °C - 800 °C temperature range, bubbles appear regardless of annealing ambient. Examples are shown in Fig. 2-1 and Fig. 2-2. We have studied the generation and annihilation mechanisms of these secondary bubbles and arrived at the conclusion that these bubbles are caused by hydrocarbon contamination of the surface. The following section will deal with our investigations of these bubbles in more detail.

# BUBBLE FORMATION IN 200 °C - 800 °C TEMPERATURE RANGE

We prepared 4 inch diameter (100) single crystal silicon wafers supplied by several silicon wafer manufacturers. By using the micro cleanroom set-up, wafers were bonded as described before and annealed in a box furnace, which allows wafer annealing up to 1100 °C . Bubbles are routinely inspected by an infrared (IR) camera which allows to detect a gap of 0.25 micron height or larger between bonded silicon wafers. X-ray topography is also used for inspecting smaller bubbles when necessary. We found bubble generation and growth to be dependent on annealing temperature and silicon wafer manufacturers but independent of annealing time. At each temperature, 15 -18 minute annealing of bonded wafers is used as a standard procedure to inspect bubbles.

Two different typical sizes of bubbles were generated when as received wafers from five manufacturers were bonded individually and annealed at 250 °C - 500 °C for 18 minutes. Therefore, we have classified the bubble generation behavior of silicon wafers from five wafer manufactures into two groups according to the size of the generated bubbles. In the "IR" group wafers, bubbles are generated which are large enough to be detected by an IR camera as shown in Fig. 2-2. "X" group wafers generate tiny bubbles which can only be detected by X-ray topography as shown in Fig. 3-1. IR-group wafer bubbles appear during warm-up to 250 °C , grow both in size and in number when increasing the temperature up to 600 °C - 800 °C with the maximum size of 15 mm diameter, and disappear at temperatures higher than 800 °C . The size and the spatial distribution of bubbles are random. These bubbles look quite similar to the bubbles reported by Ohashi et al. [8]. Tiny bubbles as can be seen in the X-group (Fig. 3-1) do not appear in the IR-group as shown in Fig. 3-2. X-group tiny bubbles show a similar growth dependence on annealing temperature but never grow to more than 1 mm diameter in size. They are randomly distributed over the entire interface with an areal density of 20 - 50 /cm$^2$. Our experiments clearly show that tiny bubbles may be present with a fairly high density even if no bubbles can be detected by an IR camera and that bubble configuration and distribution are exclusively different between the two groups.

The dependence of bubble type on the specific silicon wafer manufacturers indicates that bubble generation is affected by the surface condition of silicon wafers possibly due to the final wafer cleaning and/or contamination during transportation and storage. Stengl et al. [9] suggested that bubble formation is due to excess water adsorbed at the wafer surfaces. In our bonding process, deionized water is always used for flushing the silicon surface prior to bonding. In spite of this unchanging procedure, we always get two different bubble types, i.e.; the IR-group large bubbles and the tiny X-group bubbles. Therefore, we concluded that there must exist an additional or different cause for bubble formation, not related to adsorbed water J. Olsen et al. [10] suggested that after removing the native oxide by HF, the native oxide does not grow on the silicon surface for days but gas contaminations are piled up or adsorbed on the surface. Intuitively we assumed that hydrocarbons are the most probable contaminants. In order to test our assumption, the following experiment was performed.

Pairs of X-group wafers were exposed to different storage conditions for 15 hours at a temperature of 100 °C . These wafers were then bonded to each other and annealed at 500 °C for 10 minutes. Prior to the 500 °C anneal, wafers showed no bubbles when viewed by an IR camera. After the anneal, significant bubble formation appeared for the wafers that were stored in a conventional wafer box during the 100 °C anneal. The left part of Fig. 2-3 shows the IR image of this wafer pair. The right half of Fig. 2-3 shows a bonded wafer pair that was treated identically with the only difference that an inert glass container was used during the 100 °C anneal. This wafer pair is virtually free of bubbles after the 500 °C anneal. In addition, it was observed that the hydrophilic wafers stored in the wafer box (at 100 °C ) became hydrophobic.

The above described contamination results probably can be explained by the ESCA and HREELS investigations of hydrophilic and hydrophobic silicon surfaces by Grundner et al. [11]. The hydrophobic silicon surface was found to be covered mainly with $Si - H$ and $Si - CH_x$ groups. After annealing the samples at 400 °C , the HREELS signal of the $CH_x$ groups nearly disappeared These results support our assumption that hydrocarbons, adsorbed or bonded to the silicon surface, are the cause for the generation of large IR-group bubbles during annealing after bonding. Moreover, K. B. Kim et al. [12] recently reported that $SiC$ precipitation was detected by high resolution TEM at the interface between an epitaxial silicon layer grown by rapid thermal CVD and a silicon substrate, which also indicates that hydrocarbons may contaminate the silicon surface. Therefore, we have concluded that hydrocarbons adsorbed on silicon surfaces form bubbles by evaporation during annealing treatment required for increasing wafer bonding strength. The bubble size appears to depend on the specific type of hydrocarbon contamination.

# BUBBLE ANNIHILATION AT 800 °C AND HIGHER TEMPERATURES

The tiny bubbles of X-group wafers are usually annihilated during 800 °C - 900 °C annealing for 2 hours. The large bubbles of the IR-group wafers require 1000 °C - 1100 °C annealing for 2 hours for complete annihilation. In both groups, 1100 °C annealing treatment eliminates bubbles, which still exist after 18 minutes of 800 °C annealing. Basically two possible bubble

elimination mechanisms are likely. One is escaping of the whole bubbles to the outside by moving along the interface. The other is dissolution into the bulk silicon or the interface oxide layer. In order to investigate the annihilation process in the 900 °C - 1100 °C temperature range, intentionally hydrocarbon contaminated bonded wafers were used.

The bonded wafers were annealed at 900 °C , 1000 °C , and 1100 °C sequentially for 18 minutes each. As shown in Figs. 2-4, 2-5, and 2-6, increasing annealing temperature annihilates bubbles. In comparison with these three IR images, individual bubbles disappear without moving around. This result indicates that the gas in the bubbles is dissolved into the surrounding bulk silicon or the oxide layer.

## SUGGESTIONS FOR BUBBLE ELIMINATION

Once bubbles have been generated during annealing, a high temperature annealing at 1100 °C or higher for a few hours is necessary to eliminate the bubbles. This high temperature step may interfere with already fabricated devices or lead to the spread of etch-stop layers in the silicon wafers. Therefore, it is desirable to remove bubble sources prior to the bonding process. Based on our assumption that hydrocarbons are the source of bubbles, we may look at different ways of cleaning wafer surfaces from hydrocarbons. Wet cleaning is the usual approach to eliminate silicon surface contamination. A typical cleaning process consists of HF dipping and either $NH_4OH + H_2O_2$ or $H_2SO_4+H_2O_2$. This cleaning process can reduce bubbles from the IR-group wafer level to the X-group wafer level, but cannot completely eliminate them. Remaining hydrocarbons after cleaning still form tiny X-group bubbles in the 200 °C -800 °C temperature range.

Wafer preheating is another possibility to reduce hydrocarbon contents at silicon surfaces. Only the hydrocarbons evaporating at temperatures higher than the anneal temperature will then contribute to bubble formation. Resulting bubble configurations correspond to certain annealing temperatures. There is a bubble growth range between 200 °C and 600 °C , a bubble meta-stable equilibrium range between 600 °C and 800 °C , and a bubble annihilation range over 800 °C temperature. Therefore, in the 600 °C - 800 °C temperature range, all hydrocarbons contaminating silicon surfaces are likely to evaporate. Based on this idea, we investigated 600 °C preheating for 30 minutes using IR-group wafers. We verified that all bubbles visible by IR camera were eliminated by this process. When the bonded wafers preheated at 600 °C were annealed at 500 °C for 18 minutes, also no tiny X-group bubbles appear. However, similarly treated bonded wafers annealed at 800 °C showed tiny X-group bubbles. This result indicates that hydrocarbon contamination evaporation has a temperature distribution. When wafers are preheated at 600 °C , hydrocarbons with evaporation temperatures between 200 °C - 600 °C are evaporated and those with higher evaporation temperatures are still remaining on the silicon surface. Therefore, in spite of no bubble generation at 500 °C annealing, bubbles can form during 800 °C annealing. We presently investigate whether 800 °C preheating eliminates all hydrocarbons and therefore prevents any kind of bubble formation.

## CONCLUSIONS

We have reviewed the causes of bubbles which disturb development of wafer bonding applications to SOI, power devices, and micromechanical devices. Problems associated with wafer flatness, trapped air, and particles have been solved. The remaining problem of bubble formation in an intermediate temperature range has been discussed. We have proposed that the cause of wafer bonding bubbles formed during annealing is the presence of hydrocarbons contaminating silicon surfaces. Bubbles are generated when hydrocarbons are evaporated from the surface in the 200 °C - 800 °C temperature range. At higher temperatures, bubbles are dissolved into the bulk silicon or interface oxide layers. There are two different bubble sizes, the larger one (IR-group) may be detected by infrared, the smaller one (X-group) requires the use of X-ray topography. Preheating wafers at 600 °C - 800 °C prior to bonding can evaporate all the absorbed hydrocarbons on the surface and prevent bubble formation after bonding and annealing.

## REFERENCES

(1) J. Haisma, J. de Physique, Colloque C4, C4-3(1988)

(2) J. Haisma, G. A. C. M. Spierings, U. K. P. Biermann, J. A. Pals, Jpn. J. Appl. Phys. 28, 1426(1989)

(3) K. Petersen, P. Barth, J. Poydock, J. Brown, J. Mallon Jr, J. Bryzek, IEEE Solid-State Sensors Workshop, Tech. Digest, Hilton Head Island, SC (1988)

(4) R. Stengl, K. Mitani, V. Lehmann, U. Gösele, Proc. of the IEEE SOS/SOI Conf, Stateline NV, Sept, 123(1989)

(5) T. Abe, M. Nakano, Submitted to Proc. of 4th Intern. Symp. SOI Tech. and Devices, Montreal, May (1990)

(6) R. Stengl, K. Y. Ahn, U. Gösele, Jpn. J. Appl. Phys. 27, L2364(1988)

(7) V Lehmann, K. Mitani, T. Mii, R. Stengl, U. Gösele, Jpn. J. Appl. Phys. 28, 12(1989)

(8) H. Ohashi, K. Furukawa, M. Atsuta, A. Nakagawa, and K. Imamura, Proc. IEDM, 678(1987)

(9) R. Stengl, T. Tan, U. Gösele, Jpn. J. Appl. Phys. 28, 1735(1989)

(10) J Olsen, F. Shimura, J. Vac. Sci. Tech., Oct (1989)

(11) M. Grundner, H. Jacob, Appl. Phys. A39, 73(1986)

(12) K. Y. Kim, P. Maillot, A. E. Morgan, A. Kermani, and Y. H. Ku, J. Appl. Phys. 67, 2176(1990)
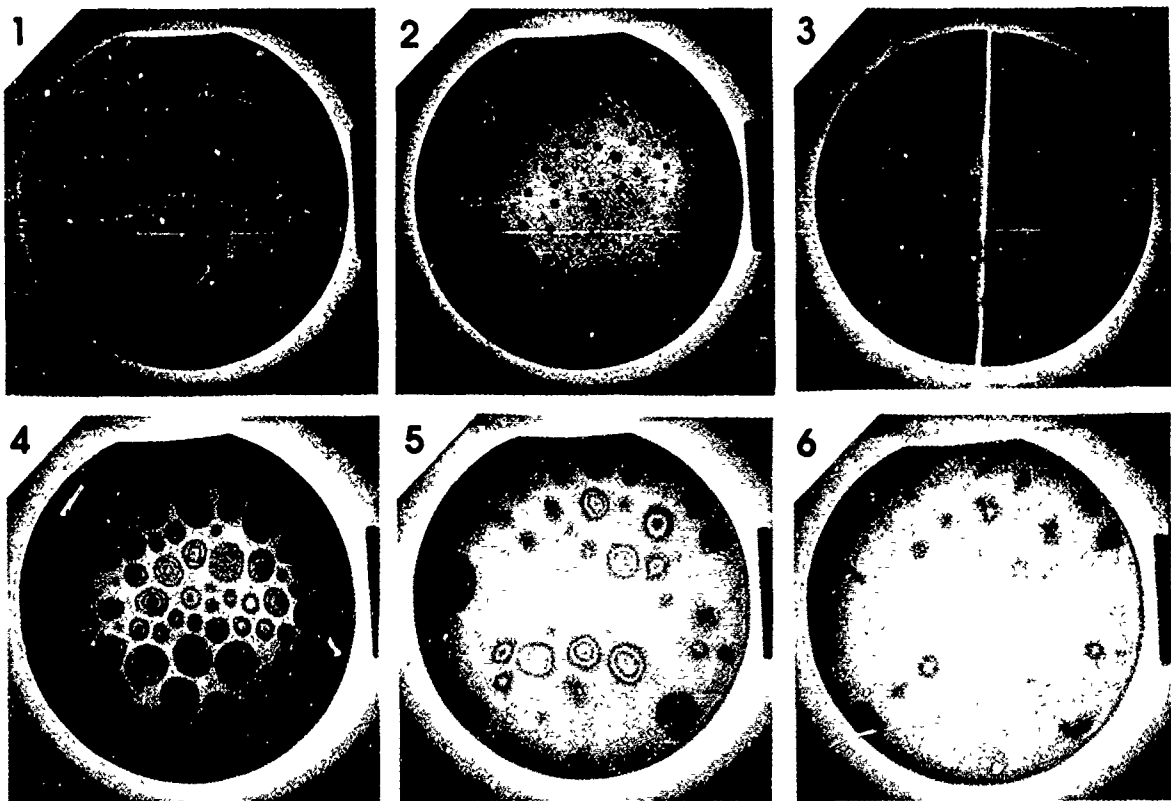
Fig. 2    Infrared images of bubbles in bonded silicon wafers.

(1): As bonded wafers with one particle related bubble. (2): Many bubbles appear after 250 °C annealing. (3): Two bonded wafer pairs after 500 °C annealing. The left pair was stored in a wafer box (100 °C , 15 hrs) before bonding, while the right pair was stored in a glass container under the same condition. (4)(5)(6): Bubble annihilation process. Bonded wafers were annealed at 900 °C , 1000 °C , and 1100 °C sequentially for 18 minutes each.



"X"-group wafer
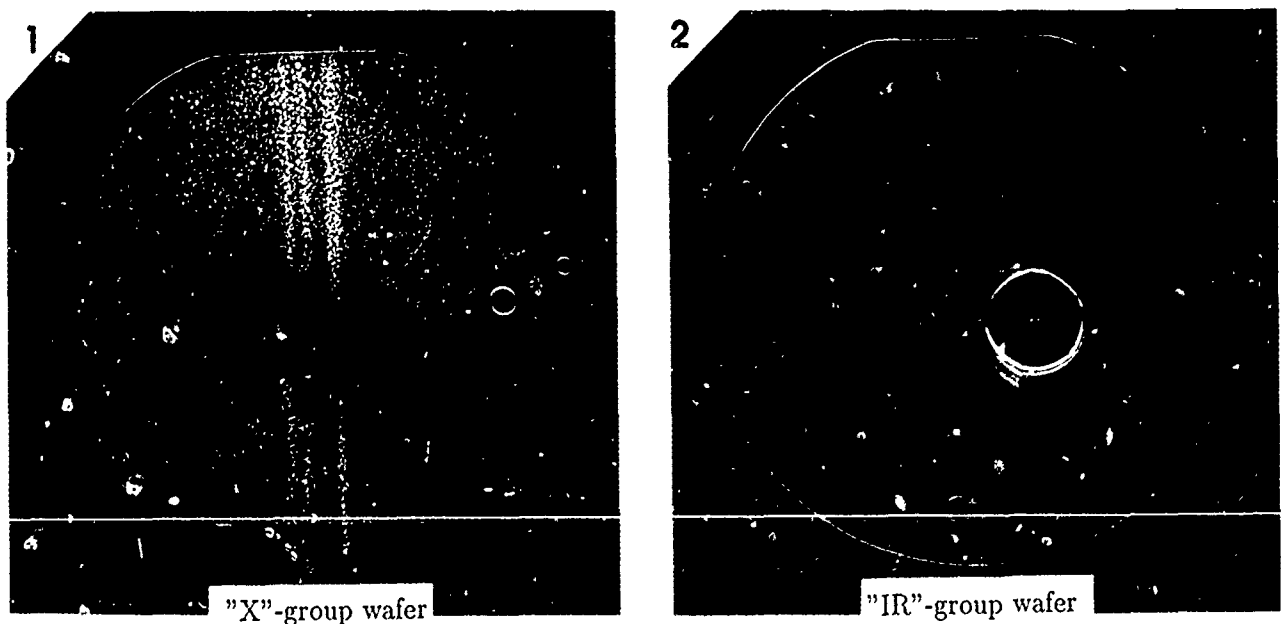
"IR"-group wafer

Fig. 3    X-ray topography of bubbles in bonded silicon wafers, which are annealed at 500 °C for 2 hours. (1): A typical "X"-group wafer. Many tiny bubbles are randomly distributed over the entire wafer. (2): A typical "IR"-group wafer. A bubble with the maximum size of 15 mm diameter appears.

# VARIABLE-FLOW MICRO-VALVE STRUCTURE
# FABRICATED WITH SILICON FUSION BONDING

Farzad Pourahmadi, Lee Christel, Kurt Petersen,
Joseph Mallon, Janusz Bryzek

NovaSensor
Fremont, CA

## ABSTRACT

A new concept in silicon micro-valves has been demonstrated for a specific commercial application, current-to-pressure (I/P) converters. Valves for this application demand high gas flow rates and large deflections of the active valve element (+/- 50 microns) which are not readily attainable with silicon microstructures. However, they do not require a tightly closed condition. This micro-actuator has been fabricated using silicon fusion bonding techniques together with conventional silicon anisotropic etching. The design of the device was optimized and its operation was analysed with the aid of finite element modelling.

## INTRODUCTION

Numerous types of micro-valves have been demonstrated using silicon micromechanical fabrication technology. Diaphragm-based valves can be effective for applications which require low leakage rates [1]. The maximum amount of deflection attainable with a silicon diaphragm is small, however, typically less than 20 microns. Such limited deflection magnitudes restrict the use of these valves (and other actuators) to low flow (and small motion) applications. Larger deflections can only be realized by making the diaphragms thinner, thereby decreasing the ruggedness and the operating pressures of the micro-valves.

Another difficulty with micro-valves is associated with the actuation mechanism. The most common actuation mechanisms employed in the literature are thermal [1], electrostatic [2], and piezoelectric [3]. All these mechanisms apply very large forces for small deflections, but are not effective for diaphragm or actuator deflections above about 20 microns. In other implementations, such as the fluid diode [4], the pressure of the supply fluid itself is used to actuate the valve.

The silicon micro valve described here is designed for very large deflections (over 50 microns) and large gas flow rates. The electromagnetic actuation mechanism is also designed for such large deflections.

Micro-valves have many possible applications. The exact application of such a device is the most important consideration in the design of a practical silicon micro-valve. The device described here is specifically designed for potential use in a current-to-pressure (I/P) converter. An I/P converter accepts a dc current signal (4-20 mA) as input and produces a pneumatic pressure as output. The pneumatic pressure is used to activate and regulate a "macro"-mechanical actuator such as a process flow valve. A typical pneumatic circuit consists of a 20 psi regulated source, flowing through an orifice into an expansion chamber which mechanically operates a macro-actuator. The pressure inside the expansion chamber is controlled by adjusting the flow rate in a leakage path between the orifice and the chamber. Flow rate is regulated by a small valve, typically electro-magnetically or piezo-electrically actuated, at flow rates of about 20 cc/sec. When the valve is fully open, the pressure in the chamber is minimized (3 psi); when the valve is closed, the pressure in the chamber is maximized (15 psi). In a well-designed instrument, the output pressure over this range is linear with input current. In this application, it is unnecessary (and undesirable) to maintain a leak-tight closed condition. These small, delicate industrial instruments are pains-takingly fabricated, adjusted, and calibrated at a rate of about 200,000 per year world-wide.

## FABRICATION

The fabrication sequence for the micro-valve is illustrated in Figure 1. Two silicon wafers are used in the fabrication of each micro-valve wafer. Wafer No. 1 is machined to become the valve exit port (backside) and recessed valve seat (frontside). Wafer No. 2 is bonded to wafer No. 1 and ultimately becomes the valve flexure (flapper) which closes the valve when forced against the seat by the actuator force.

The fabrication begins when wafer No. 1 is double-side polished to a thickness of about 400 um. A dielectric resistant to silicon etching is deposited on the wafer, and photolithography is used to define patterns on both sides of the
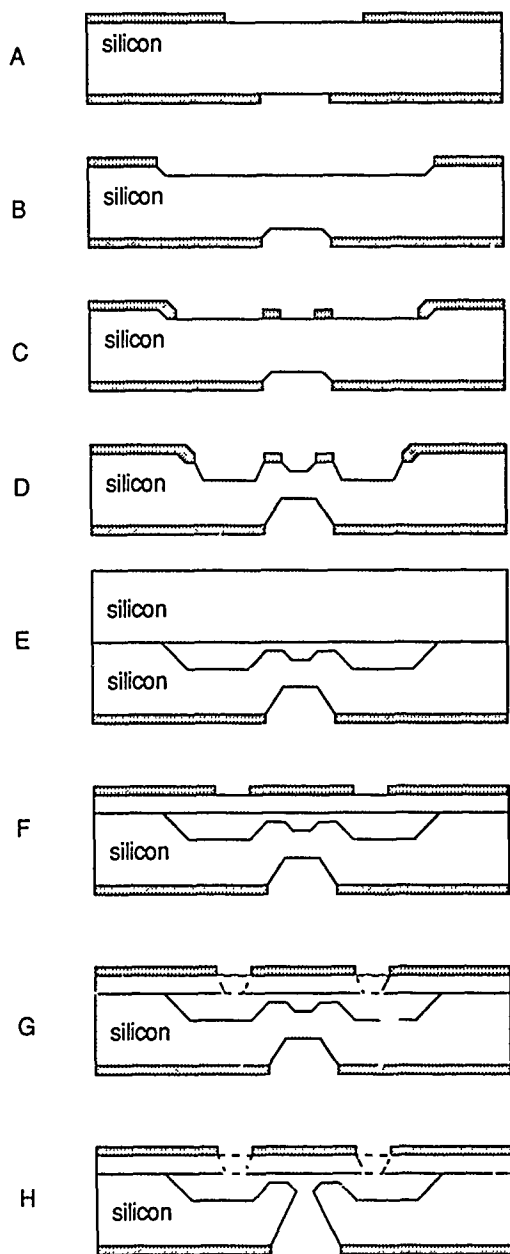
Figure 1    Fabrication sequence for the fusion bonded
micro-valve. A) Deposit dilectric and pattern
both sides; B) first silicon etch - 50 microns;
C) redeposit dielectric on front and pattern;
D) second silicon etch - 125 microns; E) strip
frontside and bond second wafer to first; F) thin
the top wafer, deposit dielectric, and pattern for
flapper etch; G) etch through top silicon to form
valve flapper; H) etch from backside to open
valve port.

wafer. On the back, a simple square pattern is defined. This
will ultimately be etched through the wafer to become the valve
backside port. The area defined on the frontside comprises
both the valve seat and an area surrounding the seat into which
the fluid will flow from the front of the device when the valve
is open.

After the first lithography step, the front of the wafer is etched
in KOH or another suitable anisotropic etchant to a depth of
about 50 um. This step recesses the surface that will eventually
be the top of the valve seat. An additional dielectric layer is
then deposited on the front of wafer No.1 and a second
photolithography step is used to define the valve seat. The
wafer is then etched in KOH again. During this etch step, the
top of the valve seat is protected, but the area surrounding it is
further etched to a depth about 75 um below the top of the
valve seat. During these first two etch processes, the backside
of the wafer is simultaneously etched, partially forming the
valve port.

After the front of wafer No. 1 is stripped of dielectrics, wafer
No. 2 is bonded to the frontside of wafer No. 1 using Silicon
Fusion Bonding (SFB) [5]. The wafer sandwich is then thinned
using grinding, polishing and etching until wafer No. 2 is
thinned to the desired flexure thickness of about 60 um.
Additional dielectric is deposited on the frontside and patterned
to define the valve flexures. Additional anisotropic etching is
used to etch through the thickness of wafer No. 2, forming the
beams and opening the top surface to the area surrounding the
valve seat. Finally, the backside of the wafer is etched until
the port meets the center of the valve seat, thus completing the
micro-valve fabrication. A scanning electron microscope photo
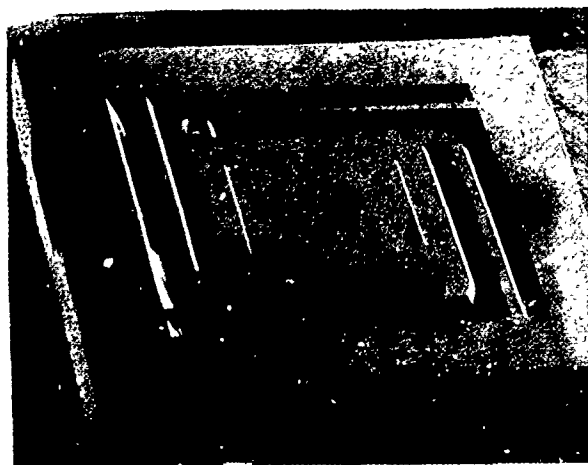of an early version of the microvalve is shown in Figure 2.



Figure 2    Scanning microscope photo of an early version
of the micro-valve chip. The chip dimensions
are 4 x 4 mm.

79

## DESIGN AND MODELLING

The flapper part of the micro-valve is of particular design importance because its function is to travel up and down, thereby adjusting the gap, and increasing the pressure drop across the valve. The motion of the flapper, its compliance with respect to the fluid pressure and actuator force, and its maximum stress levels are of great importance to a successful design.

Extensive Finite Element Analysis was performed to optimize the geometry and the strength of the device for the specified operating conditions and expected performance. ANSYS was used for the modeling of the micro-valve. The Finite Element Model is comprised of 3288 stif45 elements with 5795 nodes.

The structure of the micro-valve consists of a central region which is supported by four folded symmetrical beams each connected to a shoulder. The lateral motion of the flapper under the force exerted by the actuator induces considerable stress in the beams, especially in the regions where the beams are connected to the central flapper and also at the shoulder connection. The stress buildup is primarily due to bending moments generated by the flapper displacement. In addition, however, a torsional torque is superimposed on the beams, due to the particular design of the flexures. The combination of these two loads results in the buildup of highly tensile stress at the corner regions [6]. The existence of sharp corners at the intersection of <111> and <100> planes (due to anisotropic etching) causes stress concentration at those locations and magnifies the level of stress which can cause breakage during loading. Therefore, a stress relieved design is deemed necessary to reach the desired performance.



Figure 4    Schematic diagram of completed assembly for testing the performance of the micro-valve.

The approach here for relieving the stress at the beams' corner points has been to add multiple steps at the terminations of the beams, thus generating a gradual transition region preventing the stress from going through an abrupt change [7]. This design modification has proven successful in stress relief, resulting in an improved performance of the flexure for a given displacement. The dimensions for the step and beams have been optimized using FEM modeling.

Figure 3 shows the results of the Finite Element Model in which for 1 psi pressure applied to the top surface of the flapper element, 2928 psi maximum stress is generated. The corresponding central deflection of the flapper is calculated to be 3.4 micron/psi. This close-up shows the distribution of stress at the end of a folded beam design (different from that shown in Figure 2).

## OPERATION

The operation of the micro-valve is based on flow control through pressure drop in the valve flow channel by adjusting the flapper/valve-seat gap.

The complete micro-valve assembly is shown in Figure 4. The operation and performance of this assembly was characterized over its expected operating range. High pressure air is regulated at the inlet and the flow rate is measured with a precision rotameter. The pressure drop in the micro-valve is adjusted externally by the displacement of the core of a solenoid which is proportional to the electrical power input. A
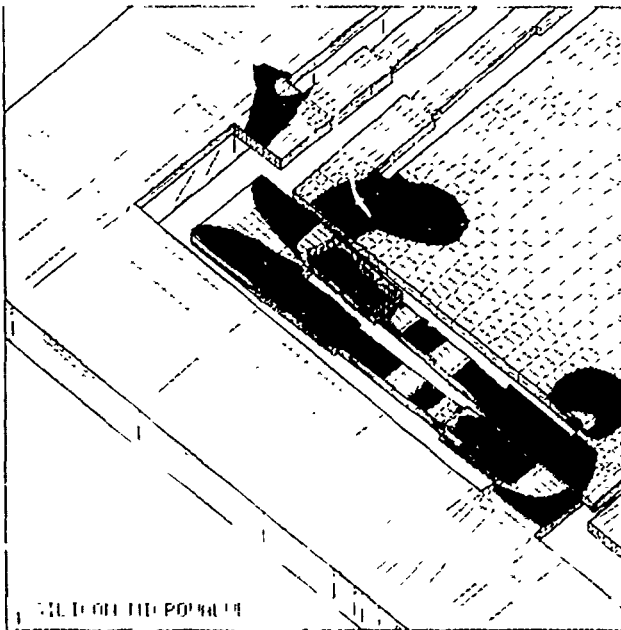


Figure 3    Stress distribution at the end of a folded cantilever beam determined by Finite Element Modelling. This beam shape was optimized for maximum vertical deflection, +/- 75 microns.
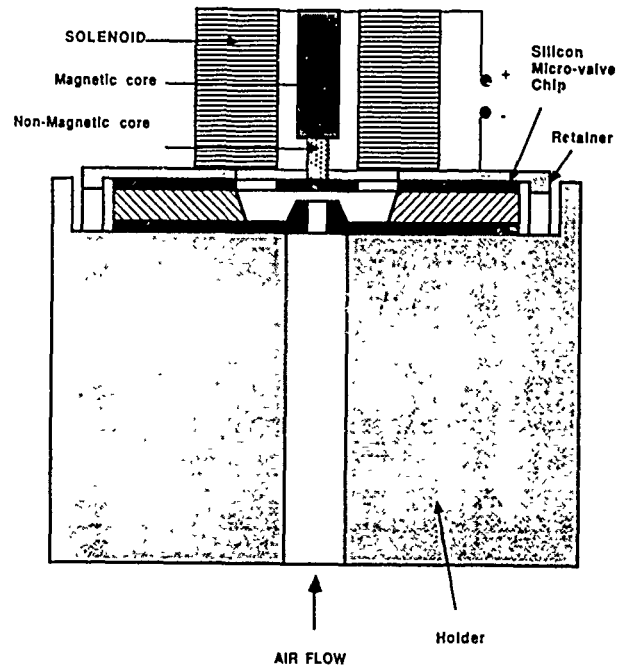
80

pressure sensor mounted on the back of the micro-valve measures the pressure buildup which opens the valve as it acts against the actuator's exerted force. The micro-valve outlet opens up to atmospheric pressure.

Figure 5 shows typical results for flow rate versus input power to the solenoid in the low pressure region of the operating range.

## CONCLUSION

The combination of a well-defined application, silicon fusion bonding fabrication technology, and detailed finite element modelling has provided a framework for implementing a new concept in silicon micro-valves. In addition, it has been shown that silicon fusion bonding is an important process, not only for sensor fabrication [5], but also for the fabrication of silicon micro-actuators. Optimization of design features with FEM increased deflection capability of the valve flapper by a least a factor of 5, over +/- 75 microns.

## ACKNOWLEDGEMENTS

Figure 5    Typical results for flow rate versus input power to the solenoid in the low pressure region of the operating range.

## REFERENCES

1]    M. Zdeblick and J. Angell, "A Microminiature Electric-to-Fluidic Valve," Proceedings of The Fourth International Conference on Solid-State Sensors and Actuators, Tokyo, 1987, pg. 827.

2]    T. Ohnstein, T. Fukiura, J. Ridley and U. Bonne, "Micromachined Silicon Microvalve", in the Proceedings of the IEEE Micro Electro Mechanical Systems Workshop, Napa Valley, CA, 1990, pg. 95.

3]    M. Esashi, S. Eoh, T. Matsuo and S. Choi, "The Fabrication of Integrated Mass Flow Controllers," Proceedings of The Fourth International Conference on Solid-State Sensors and Actuators, Tokyo, 1987, pg. 830.

4]    S. Park, W. Ko and J. Prahl, "A Constant Flow-Rate Microvalve Actuator Based on Silicon and Micromachining Technology", Proceedings of the IEEE Solid-State Sensor and Actuator Workshop, Hilton Head, SC, 1988, pg. 136.

5]    K. Petersen, P. Barth, J. Poydock, J. Mallon and J. Bryzek, "Silicon Fusion Bonding for Pressure Sensors", Solid State Sensor and actuator Workshop, Hilton Head Island, S.C., pp. 144-147, (1988)

6]    "Theory of Elasticity", S.P.Timoshenko, J.N. Goodier, 3rd edition, Mc-Graw Hill (1970).

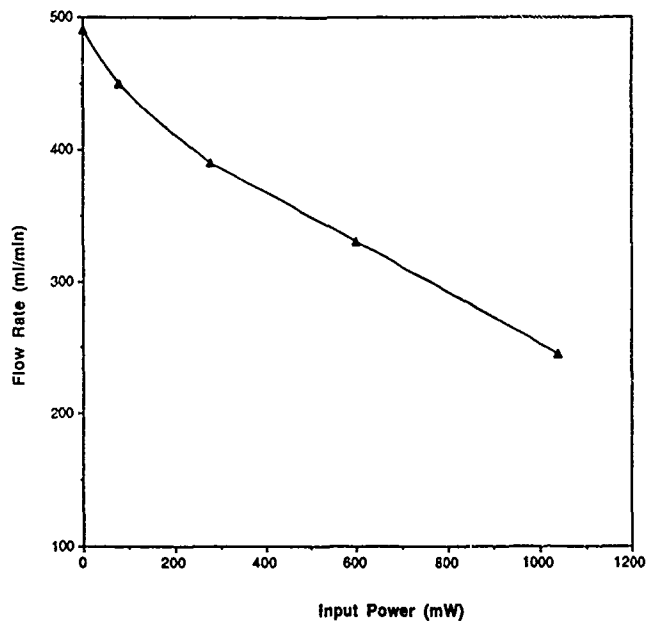7]    Mark's Handbook of Mechanical Engineering, American Society of Mechanical Engineers (ASME), Mc-Graw Hill (1980).

# AN ACOUSTIC PLATE MODE IMMUNOSENSOR*

J.C. Andle**, J.F. Vetelino and R. Lec
Department of Electrical Engineering and Laboratory for Surface
Science and Technology
University of Maine, Orono, Maine 04469


D.J. McAllister, BIODE, Inc.,
2 Oakwood Road, Cape Elizabeth, Maine 04107

## ABSTRACT

A novel acoustic plate mode (APM) sensor for sensing antigen-antibody reactions has been developed. The APM sensor utilizes a Z-cut X-propagating lithium niobate (LiNbO3) piezoelectric plate in which various types of acoustic waves are excited and received by means of a photolithographically deposited aluminum interdigital transducer (IDT). A judicious choice of the sensor operating frequency is made which results in the selective excitation of an acoustic mode which is highly sensitive to fluid loading of the LiNbO3 surface. This mode is then utilized as the sensing element in the sensor. Experiments have been performed using polyclonal antibodies for which the sensor has shown a strong response to the associated antigen. The biokinetics of the antigen-antibody reaction have also been studied. Experimental data compare favorably with theoretical results predicted by an affinity-purified human immunoglobulin antibody-antigen model. This novel immunosensor also has potential application in detecting various types of viruses such as AIDS (HIV), herpes simplex and hepatitis.

## INTRODUCTION

A critical need exists for fast, accurate microsensors for biosensing applications. The rapid detection of human viruses such as AIDS (HIV) or hepatitis and viruses or other organisms associated with plants, fish and animals is extremely important. A wide range of biosensors which include optical, calorimetric, conductimetric, potentiometric, amperometric and piezoelectric have been proposed (1). The piezoelectric sensors are particularly attractive since they offer such potential advantages as small size, portability, low cost and fast response time.

When piezoelectric sensors are used in biosensing applications, bulk acoustic waves are preferred as the sensing mechanism since biological materials are normally in solutions which severely attenuate surface acoustic waves. King (2) in 1972 first suggested that bulk acoustic wave (BAW) devices might have biomedical applications. Other reports (3) relating to the operation of BAW devices directly in a liquid or biological environment followed. Roederer and Bastians (4) reported the detection of biological molecules in solution on a surface acoustic wave (SAW) sensor. It was, however, shown later by Calabrese et al (5) that the acoustic plate mode (APM) was most likely the signal that was being monitored. Recently it has been proposed that piezoelectric devices utilizing the APM can be designed to sense gases (6), viscosity in liquids (7) and ions in solution (8).

It is the purpose of the present paper to study the application of an APM device as an immunosensor. Initially the theory of the APM device is presented and followed by the design of a prototype APM immunosensor. This sensor is tested by exposing it to affinity purified immunoglobulin G (IgG) from goats inoculated with human IgG and then measuring the response to human IgG and a mock antigen such as goat IgG.

## THEORY OF THE ACOUSTIC PLATE MODE DEVICE

The prototype APM device shown in Figure 1 is used as the sensing element in the immunosensor. The (IDTs) are deposited on the bottom face of the crystal while the biological layer is located on the top face. Both faces of the piezoelectric crystal are parallel and optically polished. Also shown in Figure 1 are the possible acoustic waves which may be excited by the IDT. These include the SAW, the surface skimming bulk wave (SSBW) and an APM. Obviously in order to probe the biological layer the APM is used as the sensing acoustic mode. The substrate is Z cut X propagating (ZX) LiNbO3. The coupling to APMs in this orientation is known to be very high (9) and exceeds the coupling to the SAW.
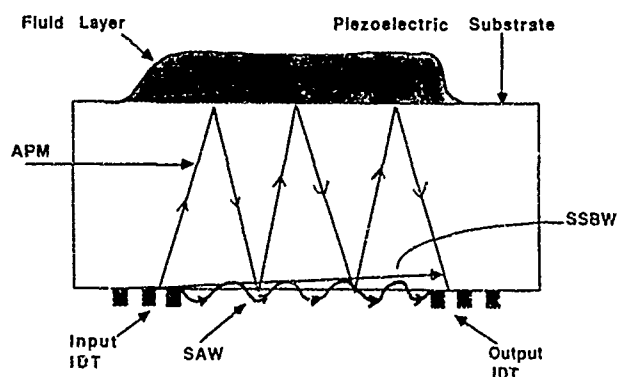


Figure 1. Prototype acoustic plate mode device

The frequency, $f_e$, associated with an acoustic wave excited by the IDT may be expressed as follows,

$$f_e = \frac{nV}{p\cos\theta},$$

(1)

where

| | |
|---|---|
| | 1 = fundamental mode |
| $n$ = | |
| | 2,3.. = higher order modes |
| $V$ = | acoustic wave phase velocity |
| $p$ = | spacing between electrodes of like polarity |
| and | |
| $\theta$ = | angle of acoustic wave with respect to bottom crystal face (see figure 1). |

In order to determine the BAWs excited one must know the variation of the three bulk wave velocities as a function of θ. The bulk wave velocity surface component, V/cosθ, for the three bulk acoustic modes is plotted as a function of θ in Figure 2.
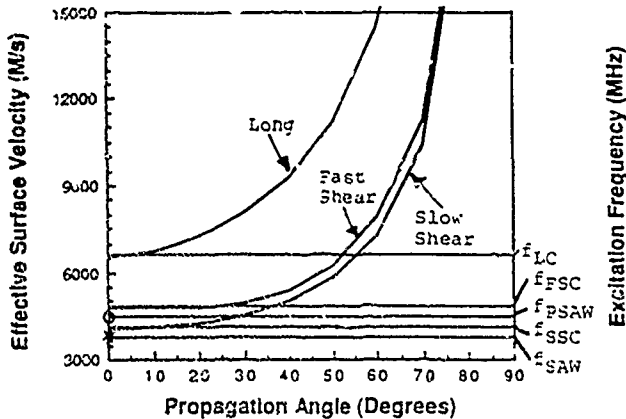


Figure 2. Variation of surface component of bulk wave phase velocities or effective surface phase velocity, V/cosθ, in ZX LiNbO3 as a function θ The cross and circle indicate the location of the SAW and pseudo SAW respectively. Also plotted on the right side of the graph are the associated excitation and cutoff frequencies for the SAW, PSAW, slow shear ($f_{SSC}$), fast shear ($f_{FSC}$) and longitudinal ($f_{LC}$) modes.

In order to calculate the delay time associated with an acoustic wave excited by the IDT, one must know the energy velocity of the wave. To determine the energy velocity it is useful to first consider a bulk wave of velocity, V, in an isotropic plate. When the wave reflects from the top surface part of the energy is converted to a wave with velocity, V, but at an angle -θ. The remaining energy is mode converted to another wave. For the reflected wave at an angle, -θ, the velocity components transverse to the propagation direction cancel and the surface component of the energy velocity is Vcosθ. Therefore, in an isotropic plate, the excitation frequency for a wave excited at an angle, θ, is related to V/cosθ while the delay time is related to Vcosθ. The resulting acoustic plate mode is seen to be dispersive with the ratio of the surface components of energy velocity, $V_E^S$, to phase velocity, $V_S$ as follows,

$$\frac{V_E^S}{V_S} = \cos^2\theta.$$

(2)

The variation of $V_E^S$ as a function of θ for the shear and longitudinal modes is presented in Figure 3. If the APM phase velocities are known, it is possible to predict the time response of the APM device Initially the spectrum of allowed phase velocities (9) is calculated. Since the APM phase velocity is exactly equal to the surface phase velocity, the angles at which the various bulk waves propagate are obtained from the phase velocity plot in Figure 2 These angles are then used to obtain energy velocities of each mode from Figure 3.
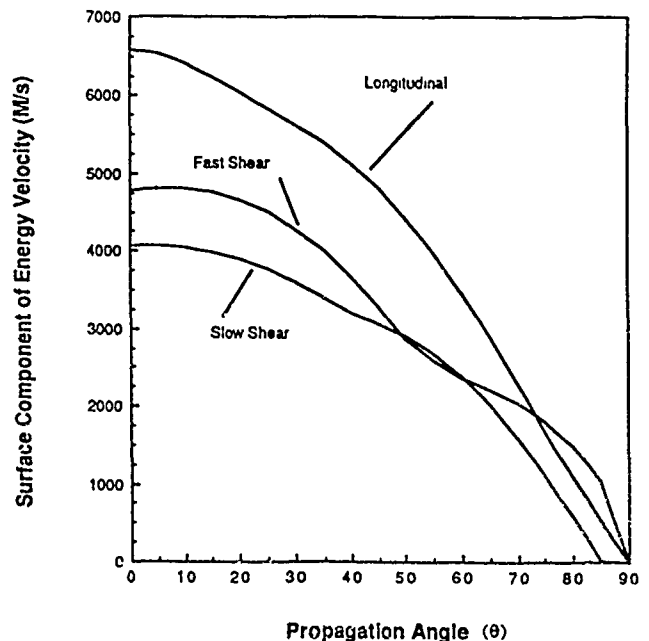


Figure 3. Variation of the surface component of energy velocity, $V_E^S$, as a function of θ for the shear and longitudinal modes in ZX LiNbO3.

Examining the allowed APM spectrum for a particular thickness, one can uniquely specify the excitation frequency and propagation angle, θ, for the bulk waves associated with a particular APM. Further, utilizing the theory developed, one can understand the nature of the different signals received at the output IDT in the time domain. One, however, must realize that the theoretical APM spectra are calculated for a plate with infinite dimensions in the propagation direction.

The positioning and spatial extent of the input and output transducers must be chosen such that the output IDT intercepts the acoustic mode of interest. Knowledge of the angle, θ, allows the correct positioning of the output IDT. It is then possible to determine an optimum spacing between the input and output IDTs.

The crystal transducer design parameters for an operating APM device may then be summarized as follows,

(1)    Choice of the IDT finger spacing and plate thickness.
(2)    Choice of the APM.
(3)    Proper choice of the distance between the input and output IDTs.

By varying the IDT finger spacing, the excitation frequency of the SAW and BAW is also varied. Therefore, for a plate of thickness, h, expressed in millimeters (mm), one can change the plate thickness when expressed in wavelengths by simply modifying the IDT finger spacing. It is then possible to study plates of many thicknesses (in wavelengths) by using only plates of a specified thickness (in mm). Generally, the higher the excitation frequency or the smaller the IDT finger spacing the thicker the plate becomes when expressed in wavelengths.

## ACOUSTIC PLATE MODE DEVICE

The ZX LiNbO3 crystals used in the present work were 1mm thick. A split finger IDT with four electrodes per wavelength, λ, was used to suppress acoustic reflections. The transducer finger spacing, p, was chosen in order to optimize the APM spectrum in terms of spurious mode rejection and mode spacing. Several APM spectra were calculated and the spectrum associated with h=3λ was selected. This spectra exhibited a strong APM near the pseudo SAW velocity with high spurious mode rejection and sufficient mode spacing to avoid interference. Thus, p was chosen to be .334 mm and h=3λ.

A number of frequencies corresponding to different APMs were given a cursory experimental examination for both a dry and water loaded surface. Of these, the mode corresponding to an excitation frequency of 43.53MHz was selected for the in-depth studies. The delay time for this mode experienced measurable shifts upon fluid loading, which is a clear indication that this mode is sensitive to surface effects.

Examination of Figure 2 shows that the longitudinal, fast shear and slow shear modes are excited. The propagation angles associated with these modes are 63°, 74° and 75° respectively. The corresponding surface components of energy velocity from Figure 3 are 3500 m/sec, 1800 m/sec and 1300 m/sec respectively.

Under fluid loading the two fast modes are observed but are heavily damped. This is due to the fact that these modes support compressional stresses in the fluid. The slow mode is only slightly damped and appears to have a slightly higher delay time and therefore was used as the sensing APM mode.

## ACOUSTIC PLATE MODE IMMUNOSENSOR

Several SAW and bulk measurement techniques can monitor either phase or energy velocities very accurately. One promising technique measures the transit time of an acoustic pulse by regenerating the pulse at the input when the output pulse is detected and then monitoring the repetition rate. If the output pulse is electrically detected and used to trigger a new input pulse, the rate at which the pulses are repeated is easily measured and gives a precise measure of the delay time. Interactions in the selective film will cause subtle changes in this delay time. This system is called a sing around measurement system (10).

In addition to the APM device, the immunosensor system consists of a pulsed-RF source, which may be externally triggered, an APM device with a selective coating, a gain block, a pulse detection module, a frequency counter, and a data logging system.

## RESULTS AND DISCUSSION

The candidate biological system used is based on an affinity purified immunoglobulin-G (IgG). IgG from one mammal (e.g. human) is introduced into another mammal (e.g. goat). Human IgG is thus an antigen and goat anti-human IgG is the corresponding antibody.

The antibody is covalently attached to the substrate using standard immobilization procedure. (11). This covalent attachment of the antibody to the surface ensures that the antibody-antigen reaction will occur at the surface. The binding density of antibody is estimated at 22 ng/mm$^2$.

A static fluid cell was created using rubber cement. The device was connected to the sing around measurement system and the repetition rate was monitored by a

computer. The fluid cell was initially filled with 200 µL of .15M NaCl, phosphate-buffered saline (PBS) at pH 7.4.

The system was allowed to stabilize for 640 seconds before the introduction of 25 µL .1mg/mL mock antigen (goat IgG). The system was then allowed to stabilize for an additional 6 minutes and no change was noticed indicating that the sensor did not respond to the mock antigen. 25 µL of .1mg/mL human IgG (the target antigen) were then added. The resulting frequency shift was .3%, stabilizing in about 15 minutes In order to ascertain the reusability of the device, the saline solution was removed and 150 µL of glycine were added. The glycine is supposed to reverse the antigen antibody reaction. The device response remained essentially unchanged until the glycine was removed and replaced with 250 µL of PBS. At this point, the response returned to baseline. The complete response of the immunosensor is shown in Figure 4.
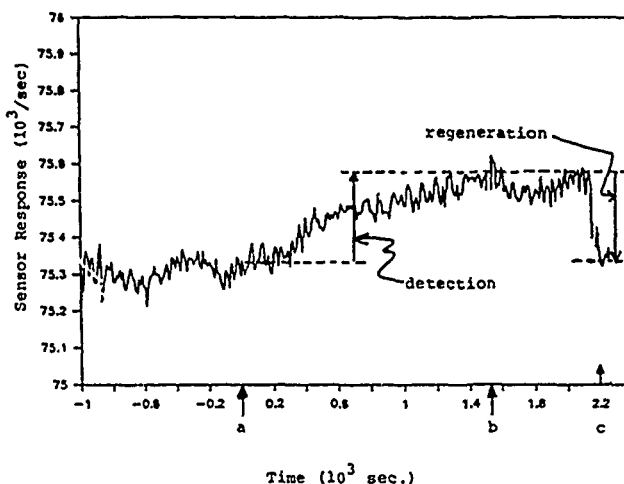


Figure 4. Total response of the immunosensor as a function of time. The time associated with a) introduction of the target antigen, b) removal of saline and replacement with glycine, and c) removal of glycine and replacement with saline are shown. The absolute sensor response is shown and both the changes due to detection and regeneration are indicated.

The device was then allowed to stabilize for 500 seconds and a second addition of 25 µL of antigen solution was added. The device responded with a positive shift of about .25% in a time frame similar to the previous measurements. This demonstrates the reversability of the antigen detection and the reusability of the sensor.

## SUMMARY AND CONCLUSIONS

An immunosensor which utilizes an APM device as the sensing element has been designed, fabricated and tested. The immunosensor basically consists of an APM device and external RF electronics capable of monitoring the APM group velocity. The immunosensor was tested by binding an antibody consisting of affinity purified goat IgG which had been innoculated with human IgG to the APM device surface. The sensor was shown to detect the target antigen, human IgG, reproducibly while not responding to a mock antigen. Although more work is needed to establish the APM immunosensor as a commercial sensor, the present work indicates that this type of sensor has promise.

84

# REFERENCES

1.  See A.P.F. Turner, Sensors and Actuators 17, 433
    (1989) and references therein.

2.  W.H. King, Bull. N.Y. Acad. Med. 48, 459  (1972).

3.  J.C. Andle, J.F. Vetelino, R. Lec and D.J. McAllister,
    1989 IEEE Ultrasonics Symp., pp. 579-584 and
    references contained therein.

4.  J.E. Roederer and G.J. Bastiaans, Anal. Chem. 55, 2333
    (1983).

5.  G.S. Calabrese, H. Wohltjen and M.K. Roy, Anal.
    Chem. 59, 833 (1987).

6.  C.T. Chuang, R.M. White and J.J. Bernstein, IEEE
    Electron. Device Lett. EDL-3, 145 (1982).

7.  A.J. Ricco and S.J. Martin, Appl. Phys. Lett. 50, 1474
    (1987).

8.  S.J. Martin, A.J. Ricco and G.C. Frye, Proc. 1988
    Ultrasonics Symposium, pp. 607-611.

9.  P.H. Carr, J. Jhunjhunwala, L.A. Veilleux, J.F.
    Vetelino and J.C. Field, Proc. 1977 IEEE Ultrason.
    Symp. p. 679.

10. R.L. Forgacs, IRE Trans. Instrumentation 9, 359
    (1960).

11. S.K. Bhatia, L.C. Shriver-Lake, K.J. Prior, J.H.
    Georger, J.M. Calvert, R. Bredehorst and F.S. Ligler,
    Anal. Biochem. 178, 408 (1989).

# THE MECHANISM OF ANISOTROPIC, ELECTROCHEMICAL SILICON
# ETCHING IN ALKALINE SOLUTIONS

H. Seidel

Messerschmitt-Bölkow-Blohm GmbH, P.O. Box 801109,
8000 München 80, Federal Republic of Germany

## ABSTRACT

In this paper, some experimental data on the anisotropy, selectivity, and voltage dependence of anisotropic silicon etchants are given. Furthermore, an attempt is undertaken to provide a general unifying model describing the reaction mechanism and the key features of all alkaline anisotropic etchants of silicon. It is shown that the reaction is electrochemical, comprising the transfer of four electrons between the electrolyte and the solid for the dissolution of one silicon atom. The crystallographic anisotropy can be attributed to small differences of the binding energy of surface atoms depending on their respective surface orientation. High boron concentrations induce the shrinking of a space charge layer on the silicon surface, which in turn leads to the fast recombination of electrons injected into the conduction band making them no longer available for the reduction of water. The electrochemical etch stop at positive potentials is due to the anodic oxidation of silicon. The finite etch rate observed at etch stop potentials was found to correspond well to the etch rate of $SiO_2$.

## INTRODUCTION

Anisotropic silicon etching is a key technology for the fabrication of micromechanical devices. It allows for the precise three-dimensional structuring of miniature sensors and actuators in an IC compatible way. The three main properties that make this technique so widely applicable are the dependence of the etch rate on crystal direction, dopant concentration and on an applied electric potential. The crystallographic anisotropy provides the possibility for a very precise lateral machining of a device by proper alignment of structural contours with either fast or slow etching crystal planes. The dependence on dopant concentration and on electric potential allow for the incorporation of well defined etch stop layers by either using a high boron concentration [1,2], or else exploiting the potential drop across a p-n junction [3,4].

Any alkaline solution of sufficiently high pH value (larger than 12) is known to exhibit the same key features listed above. That includes purely inorganic aqueous solutions of KOH, NaOH, LiOH, CsOH, and $NH_4OH$, with the possible addition of an alcohol [5-7], as well as organic aqueous solutions containing ethylenediamine [8,9], hydrazine [10,11], or choline [12], where additives like pyrocatechol and pyrazine can be present.

The main differences among these etchants are their degree of selectivity with respect to boron doped layers, the etch rate ratio of silicon to silicon dioxide, the influence of diffusion effects which is related to the obtainable surface morphology, as well as the tendency for residue formation.

So far, several papers have been published providing experimental data and models for describing specific aspects of the etching mechanism [8,13-20]. In this article, an attempt is undertaken to provide a general unifying model for all alkaline anisotropic silicon etchants, explaining the key features listed above.

## EXPERIMENTAL RESULTS

### Anisotropy

The central feature of all anisotropic etchants is the slow etch rate of the (111) crystal planes, being approximately one to two orders of magnitude smaller than for other principal crystal orientations. As an example the silicon etch rates of the three principal crystal orientations obtained in an ethylenediamine based solution (EDP type S) are shown in Fig. 1. This diagram also shows, that the activation energies are smaller for fast etching crystal planes and vice versa. This effect has been observed for other anisotropic etchants, as well [18]. Therefore, the anisotropy ratio generally decreases with rising temperature.

### Influence of dopants

For boron concentrations above a critical value $C_0$ of approximately $3*10^{19}$ $cm^{-3}$ a drastic reduction of the etch rate can be observed on both <100> and <110> wafers. This effect is shown in Fig. 2 for an EDP solution type S, where the absolute etch rate was normalized by the etch rate for moderately doped silicon. The drop is inversely proportional to the fourth power of the boron concentration.

When using KOH solutions, very similar results to EDP are obtained for low concentrations up to 10 %. However, the relative etch stop effect becomes smaller when increasing the concentration of the solution.

For comparison, results obtained for germanium doped epitaxial silicon layers and for phosphorous doped layers have been included in Fig. 2. Even at very high germanium concentrations of $1*10^{21}$ $cm^{-3}$ only a relatively small

reduction of the etch rate can be observed. Similarly, the reduction of the etch rate obtained for phosphorus doped layers, as published by Palik et al. [13], is relatively small.
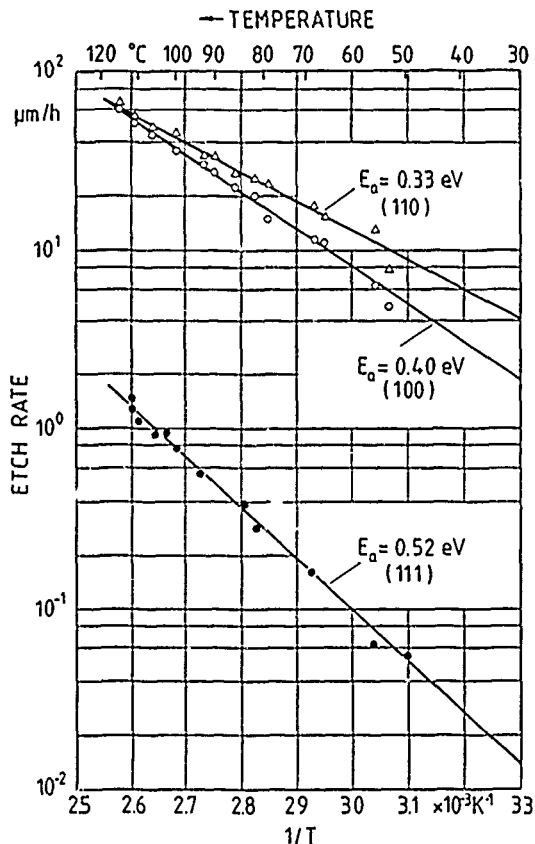


Fig. 1: Arrhenius diagram of the etch rates for the three main crystal planes, when using an EDP solution type S of the following composition: 1 l ethylenediamine, 133 ml water, 160 g pyrocatechol, and 6 g pyrazine.
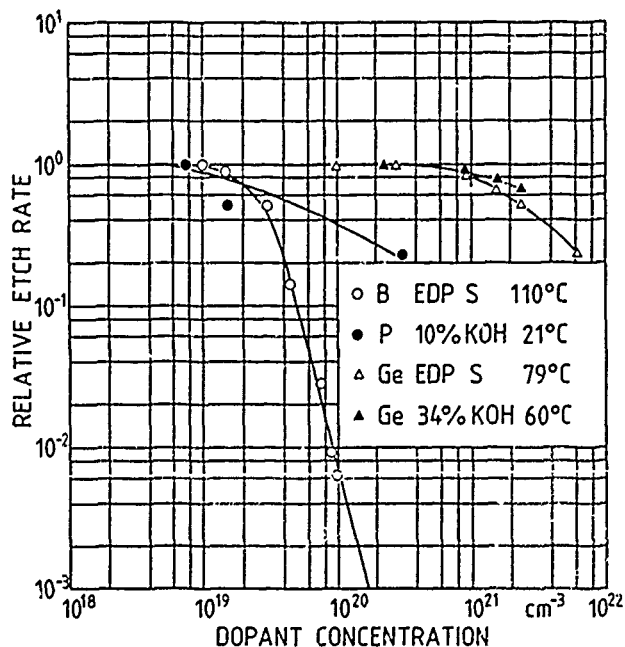


Fig. 2. Normalized etch rates for silicon layers doped with boron, phosphorus [13], and germanium, respectively.

## Electrochemical etching

When applying an external potential to a silicon wafer during the etching process, the etch rate can be influenced drastically. This is shown in Figs. 3 and 4 for p and n type silicon, respectively. Correlated voltammograms are also included in these diagrams. At anodic potentials the current density rises up to a certain point and then suddenly drops very sharply to a drastically reduced value, which is a behavior typical for the formation of a passivation layer. At this passivation potential a drastic drop of the etch rate can also be observed. The highest etch rate is obtained at a point slightly anodic from the open circuit potential (OCP). At cathodic potentials, the etch rate decreases gradually.



Fig. 3. Correlation between the voltammogram and the etch rate of p-type silicon in a 30 % KOH solution at a temperature of 65 °C. Potential is relative to a Hg/HgO reference electrode.

Whereas the etch rates obtained for n and p type silicon are very much the same, the currents flowing for cathodic potentials differ very much. P type silicon behaves like a backward biased diode for negative potentials, whereas n type silicon exhibits a large current similar to a forward biased diode.

The temperature dependence of the etch rate of n type silicon for three different potentials is shown in the Arrhenius diagram in Fig. 5. At the open circuit potential of about - 1.1 V the activation energy is 0.62 eV, which is in good agreement with previously observed values [18]. At a cathodic potential of - 2 V, the etch rate is somewhat lower with a slightly decreased activation energy. A drastically reduced etch rate can be observed for an anodic potential of 0 V. For comparison, a

curve for the etch rate of silicon dioxide has been included. Both the etch rate and the activation energy are found to be in good agreement. This is a strong indication, that the passivation layer that forms is silcon dioxide.

## DISCUSSION

In the following an electrochemical model for the etching mechanism is proposed, which is considered to be valid for all alkaline sili-con etchants. The model will first be descri-bed for moderately doped silicon with (100) orientation. Then the modifications arising for other crystal orientations and for high dopant concentrations, as well as when apply-ing an external potential will be discussed.

### Reaction mechanism

From the experimental results described above and from Raman spectroscopy measurements per-formed by Palik et al. [14] it can be conclu-ded, that hydroxide ions and water are the main reactants on the side of the electrolyte. Therefore, the redox couple $H_2O/OH^-$ is assumed to play a key role in the reaction. The Fermi level of this redox couple is initially higher than that of silicon, leading to a transfer of electrons into the solid after immersion into the electrolyte [18]. Thereby, a space charge layer is formed on the silicon surface, which corresponds to a downward bending of the ener-gy bands. This situation is shown in Fig. 6.
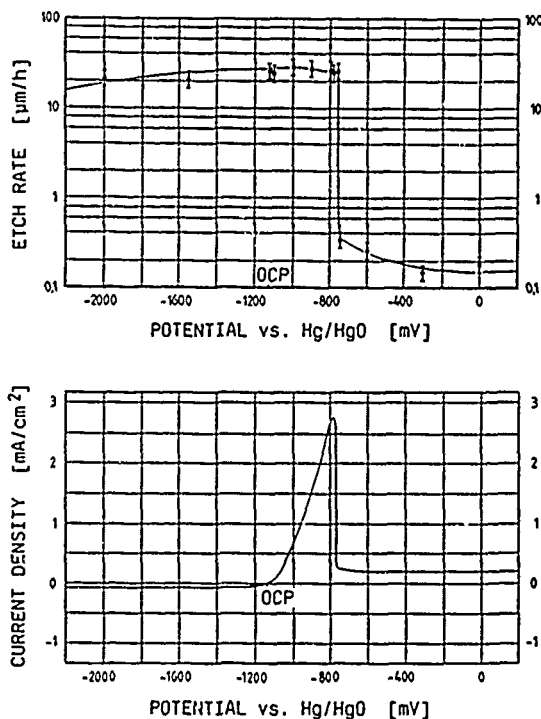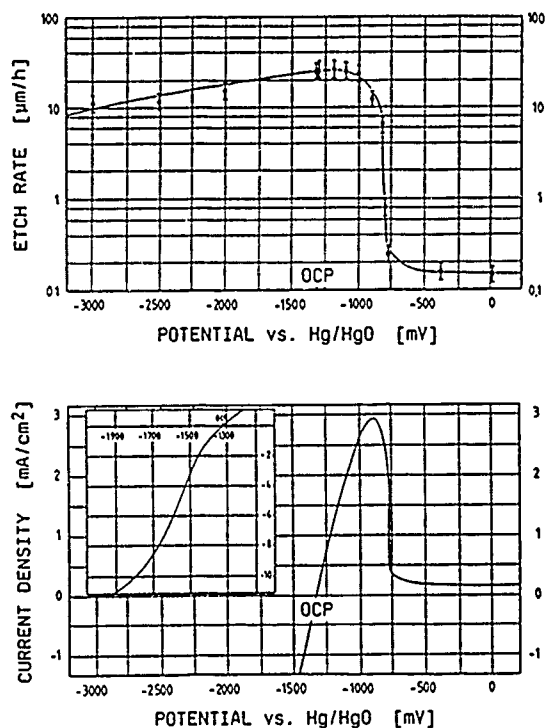
Fig. 4. Correlation between the voltammogram and the etch rate of n-type silicon in a 30 % KOH solution at a temperature of 65 °C.
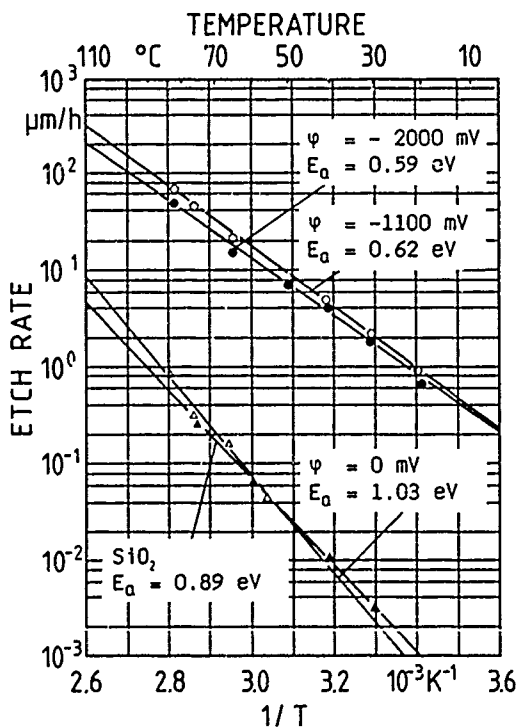


Fig. 5. Arrhenius diagram of the silicon etch rates obtained in a 30 % KOH solution for three different potentials. For comparison, the etch rate curve of silicon dioxide has been included.
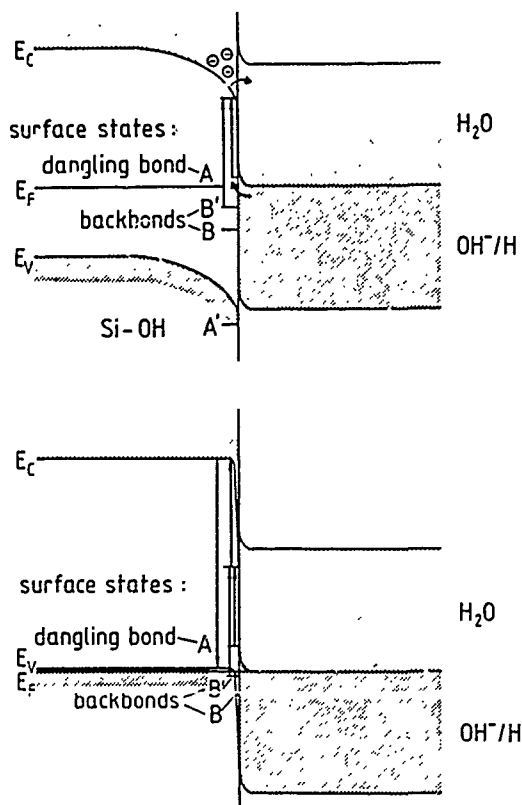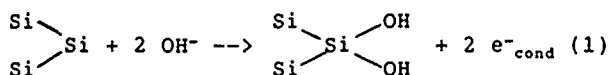


Fig. 6. Energy bands of moderately p doped silicon and the H2O/OH- redox couple. Surface states A and B in silicon correspond to dang-ling bonds and backbonds, respectively.
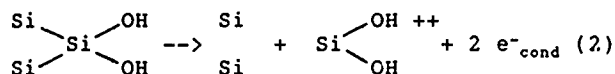
88

Due to the different bonding situation of atoms located near the surface of the silicon crystal, surface states arise which are located within the forbidden bandgap. In our model, two types of surface states are considered to be important. Type A, which is indicated in Fig. 6 is correlated to the dangling bonds. Type B with a lower energy level corresponds to the backward directed binding electrons of the surface atoms to the second layer of atoms, so-called backbonds.

In a first step, two hydroxide ions bind to the two dangling bonds available for a (100) surface silicon atom. They inject two electrons into surface state A, which are then lifted into the conduction band by thermal excitation:

$$\begin{array}{c} Si\diagdown \\ {\phantom{Si}}\diagup Si + 2\ OH^- \end{array} \longrightarrow \begin{array}{c} Si\diagdown \qquad \diagup OH \\ {\phantom{Si}}Si \\ Si\diagup \qquad \diagdown OH \end{array} + 2\ e^-_{cond} \quad (1)$$

Due to the presence of the bonded hydroxide groups on the crystal surface, the strength of the backbonds of the surface atoms will be reduced, leading to a shifting of the energy level of surface state B to a somewhat higher value.

In the next step, the backbonds of the $Si(OH)_2$ groups are broken by thermally lifting the corresponding surface state electrons (B) into the conduction band. Thus, a positively charged silicon-hydroxide complex is formed, which is detached from the lattice, but still attracted by electrostatic forces:

$$\begin{array}{c} Si\diagdown \quad \diagup OH \\ {\phantom{Si}}Si \\ Si\diagup \quad \diagdown OH \end{array} \longrightarrow \begin{array}{c} Si \\ {\phantom{}} \\ Si \end{array} + \begin{array}{c} \diagup OH\ ++ \\ Si \\ \diagdown OH \end{array} + 2\ e^-_{cond} \quad (2)$$

We assume that the energy difference between the surface state and the conduction band correlates to the measured activation energy in KOH solutions, being approximately 0.6 eV. Thus, this step can be considered to be rate limiting for the total reaction. The smaller activation energy of 0.40 eV for the same crystal plane obtained for an EDP solution can be attributed to the larger influence of diffusion effects due to the smaller pH value.

The silicon hydroxide complex reacts further with two more hydroxide ions producing orthosilicic acid [21]:

$$\begin{array}{c} \diagup OH\ ++ \\ Si \\ \diagdown OH \end{array} + 2\ OH^- \longrightarrow Si(OH)_4 \quad (3)$$
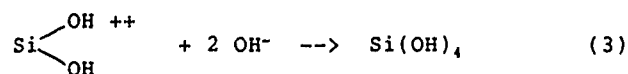
It can be assumed, that the breaking of the backbonds described by eqn. (2) and the bonding of hydroxide ions according to eqn. (3) happen more or less simultaneously. When the $Si(OH)_4$ molecule reaches the bulk electrolyte by diffusion, it will not stay stable. It is well known in silicate chemistry, that for pH values exceeding 12 the following complex will be formed by the separation of two protons [21], which was observed by Raman spectroscopy measurements [14]:

$$Si(OH)_4 \longrightarrow SiO_2(OH)_2^{--} + 2\ H^+ \quad (4)$$
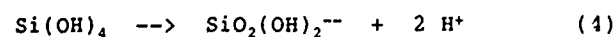
The excess electrons in the conduction band react with water molecules situated at the silicon surface, producing hydroxide ions and hydrogen atoms:

$$4\ H_2O + 4\ e^- \longrightarrow 4\ OH^- + 2\ H_2 \quad (5)$$
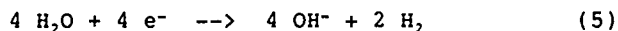
In the energy diagram shown in Fig. 6, this step corresponds to the transfer of conduction band electrons localized near the silicon surface into the unoccupied state of the $H_2O/OH^-$ redox couple, as indicated by the arrow. Therefore it is required, that an overlap between the conduction band and the upper state of the redox couple exists.

The four hydroxide ions consumed in the oxidation step are assumed to be generated by the reduction of water (eqn. (5)), rather than coming from the bulk electrolyte. This view is supported by the relatively small difference of the silicon etch rates between EDP and KOH solutions, although the concentration of hydroxide ions differs by a factor of more than 100. Furthermore, for large KOH concentrations the etch rate decreases with the fourth power of the water concentration [18].

The overall gross reaction is summarized as follows:

$$Si + 2\ OH^- + 2\ H_2O \longrightarrow SiO_2(OH)_2^{--} + 2\ H_2 \quad (6)$$
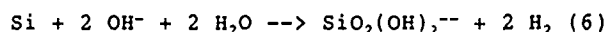
The formation of residues on the silicon surface is most severe for solutions with a relatively high water concentration. In this case the silicon dissolution rate is so high, that the transfer of the $Si(OH)_4$ complex away from the surface cannot keep up with its production. Thus, a polymerization by the separation of water leads to the formation of an $SiO_2$-like complex, which was actually observed by Wu et al. [22].

## Anisotropy

The central feature of the anisotropic behavior of silicon etchants is the very low etch rate of (111) planes. (111) surface atoms possess only one dangling bond, whereas there are two for all other main crystal surfaces. Thus in the initial reaction corresponding to eqn. (1), only one hydroxide can bind to a surface atom. Subsequently, three backbonds have to be broken in analogy to eqn. (2). This reaction step requires the transfer of the respective binding electrons into the conduction band by thermal excitation. The energy level of the corresponding surface state B must be lower than for a (100) surface. We assume, that this energy difference corresponds roughly to the observed differences in activation energy between (100) and (111) surfaces of approximately 0.12 eV (compare Fig. 1).

## Influence of dopants

For very high boron dopant concentrations, the silicon etch rate was found to decrease inversely proportional to the fourth power of the boron concentration. The critical boron concentration above which this decrease actually starts coincides well with the published value of $2.2*10^{19}$ cm$^{-3}$ for the onset of degeneracy [23]. For a degenerate p type semiconductor the Fermi level drops into the va-

lence band which is indicated in Fig. 6. As a consequence, the width of the space charge layer shrinks to a very small value on the order of one atomic layer, which is comparable to the situation observed on a metal electrode. Thus, the potential well given by the downward bending of the energy bands on the silicon surface becomes very small and, therefore, can no more confine the electrons injected into the conduction band by the oxidation step (eqns. (1) and (2)).

These injected electrons will therefore be forced to penetrate through the surface charge layer into deeper regions of the crystal. Due to the very large hole concentration, they will have a high probability to recombine with a hole from the valence band. As a consequence, these injected electrons are no more available for a subsequent reduction step, which is required for the generation of new hydroxide ions at the silicon surface (eqn. (5)).

The remaining etch rate observed for high boron concentrations is determined by the number of available electrons in the conduction band at the silicon surface. Under equilibrium conditions, this number is inversely proportional to the number of holes ($np$ = const.), i.e. to the boron concentration. We assume that this inverse proportionality is also valid at the crystal surface. Since four electrons are required for the dissolution of one silicon atom, the fourth power dependence of the etch rate on the boron concentration can be explained.

## Influence of electric potential

The key feature is the drastic reduction of the etch rate on both p and n type silicon electrodes for anodic potentials of approximately 0.5 V. The observed residual etch rate and its activation energy proofed to coincide well with the values for a silicon dioxide layer, as indicated in Fig. 5. Therefore, we assume that an $SiO_2$ passivation layer forms by the anodic oxidation of silicon, which was already suggested by several other authors [4,16]. Starting from the result of eqn. (3), this process can be described by the following reaction:

$$Si(OH)_4 \longrightarrow SiO_2 + 2\ H_2O \qquad (7)$$
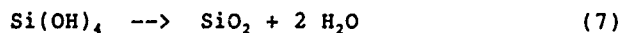
This oxide layer is assumed to start growing, as soon as the negative surface charge on the silicon electrode is cancelled by the externally applied positive potential. This corresponds to the flat band situation. Starting from this potential, hydroxide ions from the bulk of the solution are no more repelled and can easily approach the silicon surface in very large numbers. As soon as the oxide layer starts growing, the reduction of water according to eqn. (5) can no longer take place and the generation of hydroxide ions and free hydrogen at the surface stops. Interestingly, the measured current density at etch stop potential is well correlated to the silicon dissolution rate, according to the sum of eqns. (1-3) (compare with Figs. 3 and 4). This is a further proof for the transfer of four electrons required for the dissolution of one silicon atom.

The small increase of the etch rate at potentials slightly positive from open circuit conditions and, similarly, the gradual decrease of the etch rate at cathodic potentials indicates, that the electrostatic force on the $OH^-$ ions has an influence on their availability at the silicon surface. This view is supported by the observation that cathodically etched surfaces show typical signs of diffusion effects, like the formation of deeper trenches along the periphery of an etch cavity.

## CONCLUSIONS

The two key parameters for the etching behavior of alkaline solutions on silicon and on passivation layers are the molar water concentration and the pH value, which is a measure of the concentration of hydroxide ions. Anions do not play a significant role in the reaction. These effects are summarized in Tab. 1.

| | - H$_2$O + | - pH + |
|---|---|---|
| SiO$_2$ etch rate | no effect | - ◄——►► + |
| Si etch rate | - ◄——►► + | little effect |
| Solubility | no effect | - ◄——►► + |
| Si/SiO$_2$ ratio | - ◄——►► + | + ◄——►► - |
| Diffusion effects | - ◄——►► + | + ◄◄——►► - |
| Residue formation | - ◄——◄ ◄—► + | + ◄——►► - |
| p$^+$ etch stop | - ◄——►► + | + ◄——►► - |
| pn etch stop | - ◄——►► + | + ◄——►► - |

Tab. 1: Effects of water concentration and pH value on the characteristics of anisotropic silicon etchants.

The importance of organic etchants like of ethylenediamine or hydrazine based solutions does not derive from their complex anions, but rather from the possibility to adjust the pH value and the molar water concentration of water almost independently from each other. This is not the case in KOH solutions and related etchants. The only possibility to influence the water concentration there is by diluting with an alcohol.

## REFERENCES

[1] A. Bohg, "Ethylene Diamine-Pyrocatechol-Water Mixture Shows Etching Anomaly in Boron-Doped Silicon," J. Electrochem. Soc., 118 (1971) 401-402.

[2] H. Seidel and L. Csepregi, "Studies on The Anisotropy and Selectivity of Etchants Used for The Fabrication of Stress-free Structures," Electrochem. Soc. Ext. Abstr., No. 123, p. 194, Montreal, Canada, May 9-14, 1982.

[3] H.A. Waggener, "Electrochemically Controlled Thinning of Silicon," Bell System Tech. J., 50 (1970) 473-475.

[4] B. Kloeck and N.F. de Rooij, "A Novel Four Electrode Electrochemical Etch-Stop Method for Silicon Membrane Formation," in: Proc. 4th International Conference On Solid-State Sensors And Actuators — Transducers '87, Tokyo, Japan, June 2-5, 1987, pp. 116-119.

[5] J.B. Price, "Anisotropic Etching of Silicon with KOH-H2O-Isopropyl Alcohol," in: H.R. Huff and R.R. Burgess (Eds.), Semiconductor Silicon, Electrochemical Society Softbound Proceedings Series, Princeton, NJ, USA, 1973, pp. 339-353.

[6] L.D. Clark Jr., J.L. Lund, and D.J. Edell, "Cesium Hydroxide (CsOH): A Useful Etchant for Micromachining Silicon," in: Technical Digest, IEEE Solid-State Sensor And Actuator Workshop, Hilton Head Island, S.C., USA, June 6-9, 1988, pp 5-8.

[7] W. Kern, "Chemical Etching of Silicon, Germanium, Gallium Arsenide, and Gallium Phosphide," RCA Review, 39 (1978) 278-308.

[8] R.M Finne and D.L. Klein, "A Water-Amine-Complexing Agent System For Etching Silicon," J. Electrochem. Soc., 114 (1967) 965-970.

[9] A. Reisman, M. Berkenblit, S.A. Chan, F.B. Kaufman, and D.C. Green, "The Controlled Etching of Silicon in Catalyzed Ethylenediamine-Pyrocatechol-Water Solutions," J. Electrochem. Soc., 126 (1979) 1406-1415.

[10] M.J. Declercq, L. Gerzberg, and J.D. Meindl, "Optimization of the Hydrazine-Water Solution for Anisotropic Etching of Silicon in Integrated Circuit Technology," J. Electrochem. Soc., 122 (1975) 545-552.

[11] M. Mehregany and S.D. Senturia, "Anisotropic Etching of Silicon in Hydrazine," Sensors and Actuators, 13 (1988) 375-390.

[12] M. Asano, T. Cho, and H. Muraoka, "Application Of Choline In Semiconductor Technology," Electrochem. Soc. Ext. Abstr., No. 354, p. 911, 1976.

[13] E.D. Palik, J.W. Faust, H.F. Gray, and R.F. Greene, "Study of the Etch-Stop Mechanism in Silicon," J. Electrochem. Soc., 129 (1982) 2051-2059.

[14] E.D. Palik, H.F. Gray, and P.B. Klein, "A Raman Study Of Etching Silicon In Aqueous KOH," J. Electrochem. Soc., 130, (1983) 956-959.

[15] N.F. Raley, Y. Sugiyama, and T. van Duzer, "(100) Silicon Etch-Rate Dependence on Boron Concentration in Ethylenediamine-Pyrocatechol-Water Solutions, J. Electrochem. Soc., 131 (1984) 161-171.

[16] O.J. Glembocki, R.E. Stahlbush, and M. Tomkiewicz, "Bias-Dependent Etching of Silicon in Aqueous KOH," J. Electrochem. Soc., 132 (1985) 145-151.

[17] H. Seidel, "The Mechanism of Anisotropic Silicon Etching and its Relevance for Micromachining," in: Proc. 4th International Conference On Solid-State Sensors And Actuators — Transducers '87, Tokyo, Japan, June 2-5, 1987, pp. 120-125.

[18] H. Seidel, L. Csepregi, A. Heuberger, and H. Baumgärtel, "Anisotropic Etching of Crystalline Silicon in Alkaline Solutions — I. Orientation Dependence and Behavior of Passivation Layers," to be published in J. Electrochem. Soc.

[19] H. Seidel, L. Csepregi, A. Heuberger, and H. Baumgärtel, "Anisotropic Etching of Crystalline Silicon in Alkaline Solutions — II. Dopant Concentration Dependence," to be published in J. Electrochem. Soc.

[20] R. Voß, H. Seidel, and R.J. Behm, "Investigations on the Electrochemically Controlled Anisotropic Etching of Silicon in Aqueous KOH," Abstract of The Electrochemical Society European Workshop On Electrochemical Processing Of Semiconductors, Berlin, Germany, April 4-7, 1989, (full length paper to be published in J. Electrochem. Soc.).

[21] R.K. Iler, The Chemistry Of Silica, John Wiley, New York, 1979.

[22] X. Wu, Q. Wu, and W.H. Ko, "A Study on Deep Etching of Silicon Using EPW," in: Digest Of Technical Papers, The Third International Conference On Solid-State Sensors And Actuators, Philadelphia, USA, June 11-14, 1985, pp. 291-294.

[23] A.A. Vol'fson and V.K. Subashiev, "Fundamental Absorption Edge of Silicon Heavily Doped with Donor or Acceptor Impurities," Sov. Phys. Semiconductors, 1 (1967) 327-332.

# AN INVESTIGATION OF THE ELECTROCHEMICAL ETCHING
## OF (100) SILICON IN CsOH and KOH

Vincent M. McNeil, Simon S. Wang*, Kay-Yip Ng and Martin A. Schmidt

Microsystems Technology Laboratories
Massachusetts Institute of Technology
Cambridge, MA 02139

*Permanent Affiliation: General Motors Research Laboratories, Warren MI

## ABSTRACT

This paper presents results from an investigation of the electrochemical etching of silicon in $KOH:H_2O$ and $CsOH:H_2O$ solutions. Current versus voltage (I-V) scans were performed on both n- and p-type silicon as a function of etchant concentration (20%–60% by weight KOH and 25%–70% by weight CsOH) and temperature (25°C–80°C). Voltage scans were swept from potentials cathodic of the open-circuit potential (OCP) to potentials anodic of the passivation potential. The purpose of the I-V scans was to systematically investigate how varying etchant concentration and temperature affected the passivation potential and final passivation current density of both n- and p-type silicon. The results of the I-V scans are used to help investigate conditions for optimizing the performance of 3-electrode electrochemical etch-stop on $n^+$-p junction samples. A model is presented to describe the effect of reverse diode leakage on etch-stop performance which uses the the previously measured electrochemical etching parameters. Experimental measurements of the etch-stop are used to confirm the model.

## INTRODUCTION

The electrochemical etch-stop technique is a particularly attractive method for fabricating microsensors and microactuators since it has the potential for allowing one to reproducibly fabricate moderately-doped silicon microstructures and membranes with thickness control on the order of $\pm 0.1 \mu m$. A number of investigators have reported on the fabrication of silicon microstructures using electrochemical etch-stopping in various anisotropic etchants (KOH, ethylene diamine pyrochatecol, and hydrazine) [1,2,3,4,5,6,7,8]. A major challenge in using electrochemical etch-stop is in maintaining a potential sufficiently anodic to passivate the n-silicon over the entire wafer, while simultaneously keeping the p-silicon at a potential less than that required for passivation. This is particularly true for many of the newer microstructures fabricated using silicon wafer-bonding, where electrical contact to the entire n-region through the backside of the wafer is no longer possible, and currents needed to passivate the n-regions may have to flow along a thin, moderately-doped n-type layer. In addition, this problem could be further aggravated by the presence of large diode leakage currents. If $n^+$-p junction leakage currents are large enough, then the p-region may assume a potential sufficiently anodic to cause it to passivate. A four-electrode configuration has been proposed which can solve the problem of large diode leakage currents by making electrical contact to the p-region and forcing its potential to be near the OCP, thus supplying the necessary leakage current[2]. However, we have found that some microstructures do not readily provide electrical access to the p-silicon.

Therefore, we have decided to investigate etchant conditions for optimizing the performance of 3-electrode electrochemical etch-stop. We will divide this optimization into two parts. In the first part, we will characterize the electrochemical etching of n- and p-type silicon in two different etchants, KOH and CsOH. In the second part, we will use the results of the first part to model the behavior of the junction etch-stop.

The process of optimizing conditions for the etchant consists of two parts: i) characterizing the difference in passivation potentials between n- and p-type silicon and ii) characterizing the current density at passivation. A number of researchers have used current versus voltage (I-V) scans to characterize the electrochemical etching behavior of silicon [9,10,11,12,13], with a particular focus on using the technique for identifying the passivation potential and passivation current for n- and p-type silicon. The difference in passivation potentials for n- and p-type silicon becomes most important since under worst-case etch-stop conditions, a large diode leakage can cause the n- and p-type silicon to be at the same potential. It is therefore desireable to have the widest possible passivation 'window' between n- and p-type silicon. The second etch solution parameter which is of importance is the current density at passivation. After the p-type silicon has been completely etched away, the passivation current will cause ohmic drops along the thin n-type layer. If these ohmic drops are too large, some regions of the wafer will fail to etch-stop if they are at a potential less than the passivation potential for n-type silicon.

In this paper we will summarize the results of I-V data used to determine the passivation 'window' and passivation currents for n- and p-type silicon in $KOH:H_2O$ and $CsOH:H_2O$ solutions over a wide range of etchant concentrations and temperatures. We report for the first time the electrochemical etching behavior of silicon in $CsOH:H_2O$. CsOH was chosen as a promising etchant for optimizing the 3-electrode electrochemical etch-stop since it has been reported that it has a much lower silicon dioxide etch rate than KOH [14,15].

The optimization of the junction etch-stop will involve developing a model, using load-line analysis, to describe the effect of reverse diode leakage. This model will be verified by measuring the terminal characteristics before and during etch-stop. In particular, it will be possible to determine *in situ* if the etch-stop is operating correctly. We will first describe the experimental procedures used, and then present the results of both the etching studies and the junction etch-stop.

## EXPERIMENTAL PROCEDURE

All electrochemical measurements were carried out using a EG&G PARC Model 273 Potentiostat under computer control. Experiments were performed in darkness with the etch sample and solution immersed in a thermostatted-waterbath with $\pm 0.1°C$ control.

A Ag/AgCl double-junction electrode and 5 cm x 10 cm solid platinum metal foil were used as the reference electrode and counter electrode respectively. Electrical contact to the silicon etch samples was made via a gold-plated clip. A schematic sketch of this 3-electrode electrochemical etch system is shown in Figure 1.

Silicon etch samples for I-V scans and etch-rate versus potential experiments, as shown in Figure 2, were prepared using 10-20$\Omega$-cm (100) 4" Si wafers. A 1500$\mathring{A}$ layer of LPCVD silicon nitride is used as an etch mask over the entire silicon etch sample except for a 1 cm x 1 cm etch window patterned on the front of the sample and a 1 cm x 2 cm window on the back of the sample for electrical contact. Good ohmic contact to the silicon etch samples was maintained by depositing Al or Cr/Au onto the heavily-doped backside region of the sample. After metallization, the wafer is diced using a diamond saw, and the exposed edges of the 2 cm x 8 cm samples were coated with epoxy to passivate the exposed (110) and (111) planes from the etch solution.

Current versus voltage data was collected for both n-type and p-type silicon in KOH solutions of 20%, 30%, 40%, 50%, and 60% at 25°C, 60°C, and 80°C. I-V scans were also carried out in CsOH solutions of 25%, 50%, 60%, and 70% at 60°C and 80°C. During the I-V scans, the potential applied to the silicon sample was scanned from -1.6V to 0V, typically at a voltage scan rate of 5mV/sec.

Etch-rate versus potential data was collected by etching the same silicon samples as used above for 30 or 60 minutes at a number of fixed bias potentials and then using a surface profilometer to measure the depth of the etched surface. Etch-rate versus potential experiments were carried out on both n- and p-type silicon in 20% and 60% KOH and 50% CsOH solutions at 60°C.

Junction etch-stop samples were prepared by POCl$_3$ doping p-type wafers at 925°C for 90 minutes and then driving-in the n$^+$-diffusion at 950°C for 90 minutes, which produced junctions with a depth of 5-6$\mu$m. The wafers were then coated with 1500$\mathring{A}$ of LPCVD silicon nitride, patterned, and etched on the backside to form twenty 1 cm x 1 cm membrane etch windows. After the nitride in the windows was removed, the silicon in the window regions were pre-thinned down to a total thickness of 60-75$\mu$m in a 20% KOH:H$_2$O solution at 60°C before applying a bias for electrochemical etch-stop. Contact to the n$^+$ silicon was made by removing the silicon nitride passivating layer from a small region of the front of the wafer, and then contacting it with a gold-plated clip.

## EXPERIMENTAL RESULTS

### A. I-V Scans
Representative current-voltage results for n- and p-type silicon in KOH and CsOH are shown in Figures 3 and 7. We have found that the passivation peak for both n- and p-type silicon shifts cathodically with increasing etchant concentration at a fixed temperature for both KOH and CsOH. However,

the passivation peak of the n-type silicon shifts cathodically at a faster rate than does p-type silicon. Hence as the the concentration increases, a larger 'window' between the n-type and p-type passivation peaks occurs. In addition, the current at passivation decreases as the concentration of the etchant is increased.

The effect of etchant temperature on passivation potential was also observed. At a given concentration of either KOH or CsOH, as the temperature is lowered the passivation potential shifted cathodically for n-type silicon and shifted *anodically* for
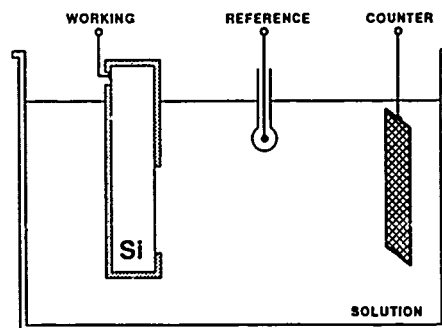


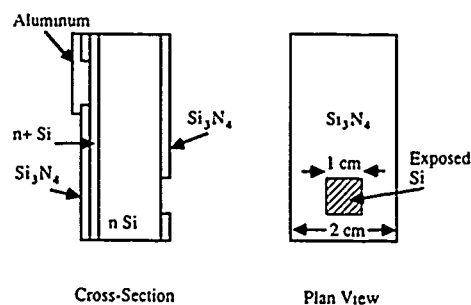Figure 1: Schematic of typical 3-electrode electrochemical etch system.



Figure 2: I-V and etch-rate vs. potential silicon etch sample.

p-type silicon. The result of this general trend is that as the etch temperature is lowered, the passivation potentials move in *opposite* directions, increasing the width of the passivation 'window'.

In addition to looking at the effect of concentration and temperature on the passivation potential, we were interested in studying how varying these parameters affected passivation current density. Variations in temperature had a pronounced effect on passivation current density. For example, for 20% KOH, the final passivation current for n-type silicon was 482$\mu$A/cm$^2$ at 80°C and 77$\mu$A/cm$^2$ at 60°C. Thus in order to to passivate the entire surface of a full 4" wafer, a current of 33.7mA is required for an etch at 80°C, as opposed to a current of 5.4mA for the same concentration etch performed at 60°C.

As the concentration of etchant was increased, there was generally only a small effect on passivation current density. For instance as the KOH concentration was increased from 20% to 60%, the passivation current density changed by less than 1%.

### B. Etch Rate vs. Potential
In order to clearly demonstrate whether or not the passivation 'window' as defined by looking at the difference in passivation potentials for n- and p-type silicon correlated to a n- and p-type etch 'window', etch-rate versus potential experiments were carried out in a chosen number of solutions. Figure 4 plots the relative etch-rate versus applied potential for both n- and p-type silicon in 20% KOH at 60°C. The etch-rate found at an applied potential was normalized to the etch-rate for a silicon sample etched under no applied bias (OCP). As can be seen, the etch-rate for both n- and p-type silicon fall off to less than 10% of their unbiased etch value as the applied potential becomes more anodic, with the etch-rate of n-type silicon falling off before the etch-rate for p-type silicon does. The n-p etch 'window' is about 100mV and corresponds fairly well to
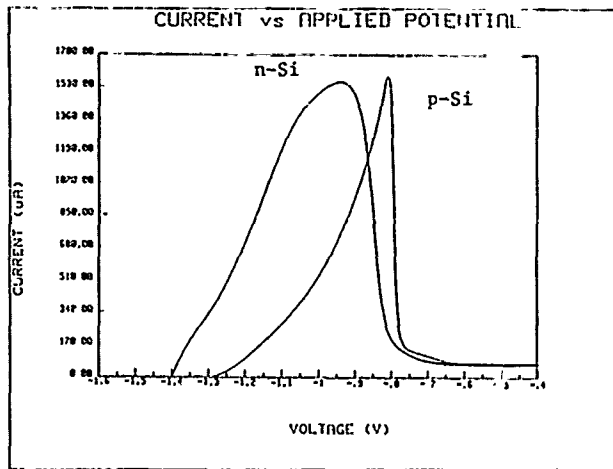
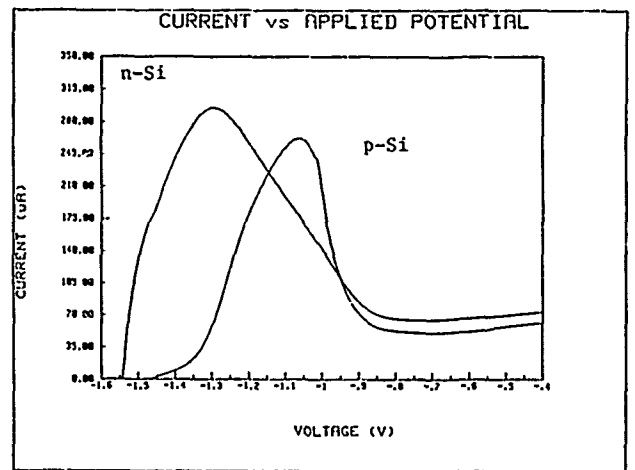Figure 3: I-V scan of n- and p-type silicon in 20% KOH at 60°C.



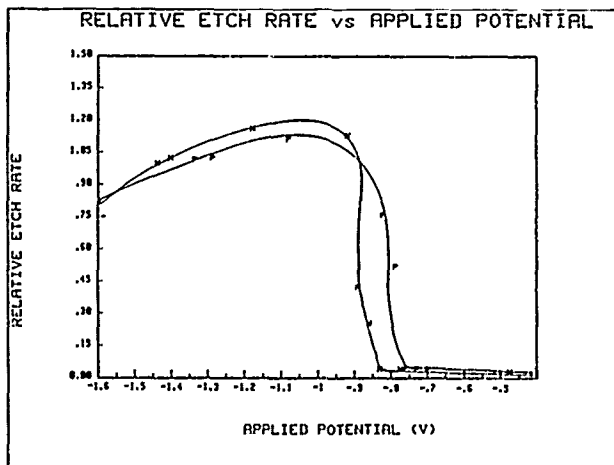Figure 5: I-V scan of n- and p-type silicon in 60% KOH at 60°C.



Figure 4: Plot of etch-rate vs. potential for silicon in 20% KOH at 60°C.
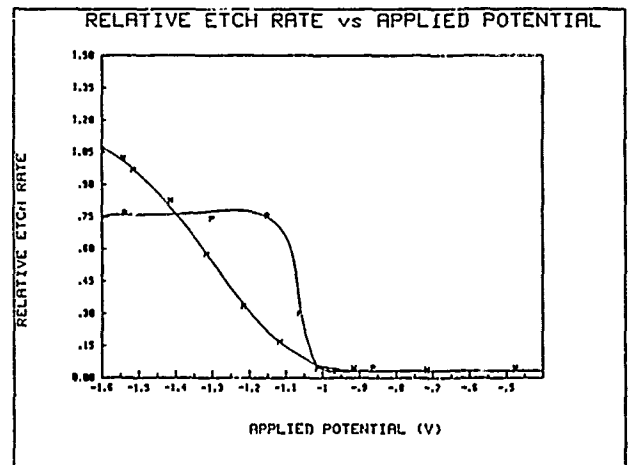


Figure 6: Plot of etch-rate vs. potential for silicon in 60% KOH at 60°C.

the passivation 'window' as predicted from I-V scans. Similar results were seen for etch-rate versus potential plots for n- and p-type silicon in 60% KOH and 50% CsOH at 60°C as shown in Figures 6 and 8 with the etch-rates for n- and p-type silicon falling off at the passivation potentials predicted by the I-V data. However for 60% KOH and 50% CsOH, the transition from high to low etch-rate for n-type silicon is less sharp than in the case for 20% KOH.

C. Junction Etch-stop Model

Figure 9 gives a schematic illustration of the junction etch-stop system where the potential of the n-type silicon is labeled $V_n$ and the potential of the p-type silicon is labeled $V_p$. Both potential are referenced to the solution. Since this system is a series combination of non-linear elements, with the characteristics of these elements shown in Figure 10, it is possible to use a load-line analysis to determine the potential of the p-type silicon ($V_p$). Figure 11 shows a graphical representation for the case of two diodes (ideal and leaky). As the applied potential is increased, the intercept labeled A will be the stable operating point for the ideal diode. If the diode is leaky, the potential of the p-type silicon will ultimately go beyond the passivation potential with increased applied bias and etching will cease (point B in Figure 11). From this analysis, we can therefore predict what the current-voltage characteristics

should look like in these cases, and this is shown in Figure 12. In the case of the leaky diode, a distinct transition should be visible from etching to passivation, making it possible to determine in-situ the upper bound on the potential which can be applied to the n-type silicon. This, in combination with the known passivation potential of n-type silicon, makes it possible to establish upper and lower bounds on the applied potential. Alternatively, we can specify a maximum tolerable leakage current given the I-V characteristics for p-type silicon and the total area being etched.

D. n+/p Junction Etch-Stop Measurement

Experiments to support the junction etch-stop model were performed. Figure 13 shows I-V diode characteristics curves for a $n^+$ p junction sample scanned at 5mV/sec in 20% KOH at 60°C. This particular sample had twenty 1 cm² etch windows exposed to the etch solution. Curve a in Figure 13 shows the 'ideal' diode characteristics for the sample scanned in darkness. In order to prove the model, we needed to create a sample which had a very leaky junction. This was accomplished by shining light on the sample during the I-V scan. Curve c is a plot of an I-V scan in which a high intensity of light was shone onto the junction. As a result of the increased leakage, the sample passivated when the applied potential was swept to a potential more anodic than the passivation potential for p-
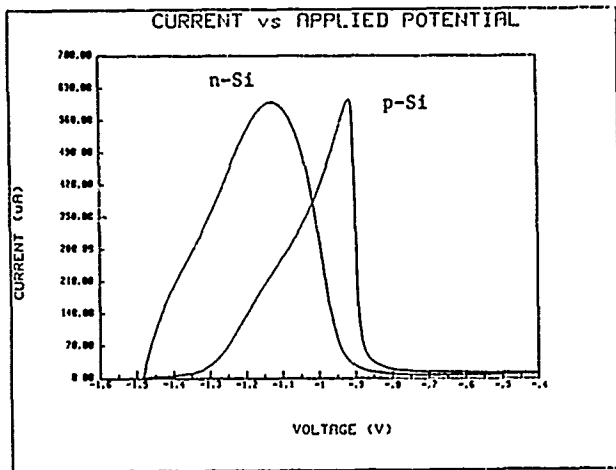
94

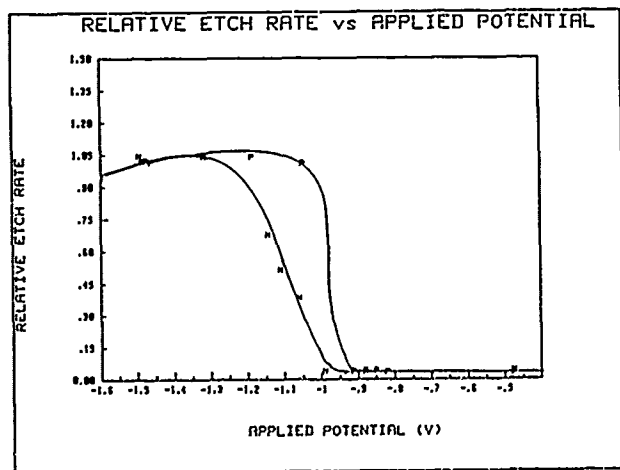Figure 7: I-V scan of n- and p-type silicon in 50% CsOH at 60°C.



Figure 8: Plot of etch-rate vs. potential for silicon in 50% CsOH at 60°C.



Figure 9: $n^+/p$ electrochemical etch-stop system model.
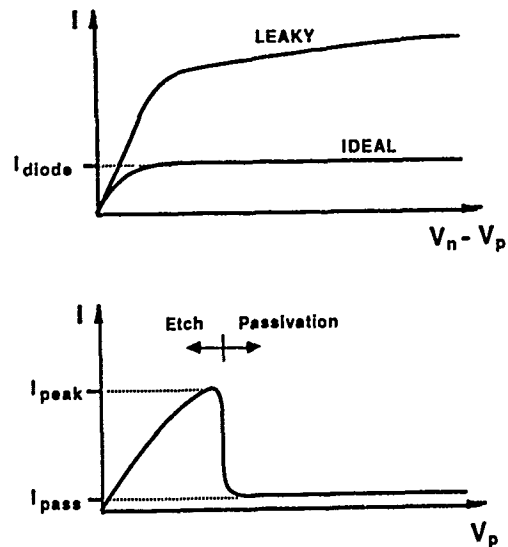


Figure 10: I-V characteristics of a p-n junction and a p-type silicon to solution junction.
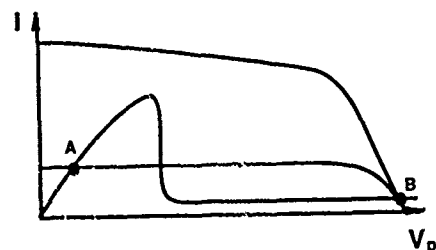


Figure 11: Load-line analysis for the case of an ideal diode and a leaky diode.
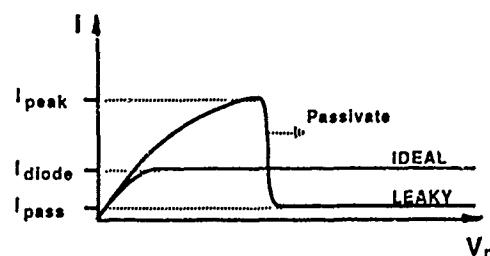


Figure 12: Predicted I-V behavior for $n^+/p$ etch-stop .

type silicon, as predicted by I-V experiments done previously. The final passivation current level of 1.6mA also correlates well with the expected passivation current as predicted by our previous measurements. Curve b is a plot of an I-V scan in which a moderate level of light created higher junction leakage than in curve a, but not sufficiently high enough to lead to passivation of the p-type silicon. In Figure 14 we show the pre- and post-etch-stop I-V scan characteristics for a 20 window sample on which a 3-electrode junction etch-stop was performed. The p-region in the etch windows were pre-thinned down to $60\mu$m before applying the etch-stop bias. In the lower part of Figure 14 one can clearly see the diode characteristic expected for an I-V scan done on the sample before initiating the etch-stop. After completion of the etch-stop, another I-V scan at 20mV/sec was done to demonstrate that the etch had terminated on the $n^+$ region in the etch windows. Figure 15 shows the corresponding current versus time plot for this sample which shows the current peak during formation of the anodic oxide and the steady-state passivation current afterward.
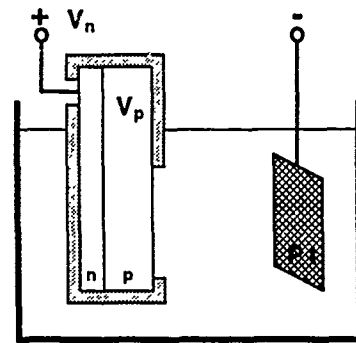
## DISCUSSION OF RESULTS

There are a number of important observations which may be drawn from the experimental results. Based on a desire to have a minimum passivation current density and a large window between n-type and p-type silicon passivation potentials we can conclude that the optimal etch conditions would be low temperature and high concentration. In particular, the CsOH etchant system is very attractive because of its very low passivation current densities. Since CsOH has a higher silicon/silicon dioxide etch rate ratio (ERR) than KOH, this optimized passivation current does not require a substantial reduction in silicon etch rate.

The junction etch-stop model imposes a different set of constraints for an optimized etch-stop system. First, we would like the lowest possible diode leakage. This is partly a technology constraint, but since the diode leakage is a function of temperature, it suggests that we would like a low temperature etch system. Additionally, we would like a system which provides us with a low silicon-to-solution impedance in the pre-passivation anodic region. In this way, the potential of the p-type silicon will be close to the potential of the solution, and most of the potential drop will be across the diode. In general, the lowest impedance is obtained for high temperature and lower concentration solutions. For example, the curves in Figures 5 and 3 clearly indicate that as the concentration of KOH is reduced from 60% to 20%, the effective impedance decreases by a factor of 6, while the diode leakage remains the same, because temperature is held constant. The advantage of going to higher temperature must be decided based on the temperature dependence of the impedance and diode leakage. If the decrease in impedance is a stronger function of temperature than the diode leakage, we would prefer to operate at a higher temperature. The answer to this question will depend on the details of the diode fabrication. It is also important to note that the relative areas of the etched surface and the diode will be important in determining the success of the electrochemical etch-stop. In particular, if the diode area is very large, and the etched surface consists of just a few small openings, then all of the diode leakage must flow through this small opening which could cause the surface to passivate.

Based on these observations we arrive at a set of conflicting requirements for optimized etch-stop behavior. The etchant optimization suggests lower temperatures and higher concentrations, while the diode model suggests higher temperatures (assuming diode leakage is a weak function of temperature in comparison to the silicon-solution impedance) and lower concentrations. We believe the answer will therefore be critically dependent upon the particular application. In the case of diaphram formation where electrical contact to the wafer is made at many points over the surface of the wafer, the resistive losses in the n-type layer may not be a concern. In this case, it would be advisable to use conditions which are optimum from the diode model arguments. Provided the electrical contact is good, the passivation current density will not matter. When thinning a bonded wafer with an $SiO_2$ layer between wafers, electrical contact can be made at only a few points near the edge of the wafer. In this case, the ohmic losses through the thin n-type layer can be of paramount concern, in which case the etchant optimization conditions will be important. In general, however, the CsOH has advantages in comparison to KOH because of its higher $Si/SiO_2$ ERR.
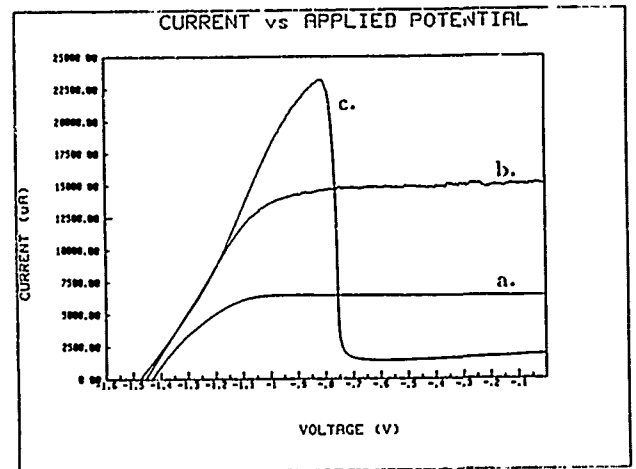


Figure 13: I-V scan for a $n^+/p$ sample with a.) minimum leakage, b.) moderate leakage, c.) high leakage.
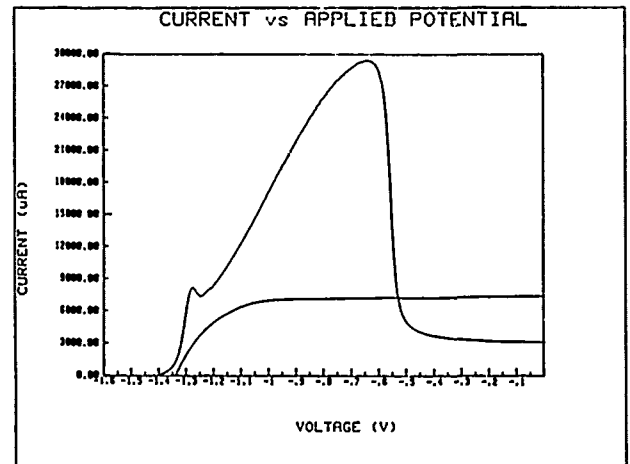


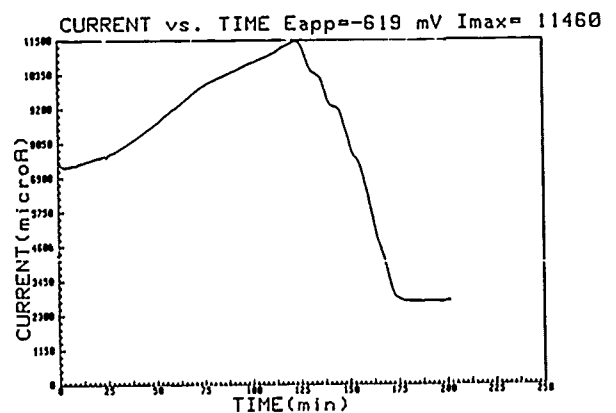Figure 14: Pre- and post-etch-stop I-V scan of $n^+/p$ sample.



Figure 15: Current vs. Time (I-T) plot for junction sample with 20 windows.

## CONCLUSIONS

We have demonstrated a method for optimizing etch conditions for the 3-electrode electrochemical junction etch-stop. Etching solution conditions in KOH and CsOH were identified that lead to the largest process latitude in terms of passivation potentials and passivation current densities. A model for the effect of junction leakage on the etch-stop, which permits in-situ analysis of optimized biasing conditions was presented. Additionally, experiments were performed to demonstrate the validity of the model.

## ACKNOWLEDGEMENTS

# References

[1] H. A. Waggner, "Electrochemically controlled thinning of silicon," *Bell Systems Technical Journal*, pp. 473–475, March 1970.

[2] B. Kloeck, S. D. Collins, N. F. de Rooij, and R. L. Smith, "Study of electrochemical etch-stop for high-precision thickness control of silicon membranes," *IEEE Transactions on Electron Devices*, vol. 36, pp. 663–669, April 1989.

[3] P. M. Sarro and A. W. van Herwaarden, "Silicon cantilever beams fabricated by electrochemically controlled etching for sensor applications," *Journal of the Electrochemical Society*, vol. 133, pp. 1724–1729, August 1986.

[4] Y. Linden, L. Tenerz, and B. Hok, "Fabrication of three-dimensional silicon structures by means of doping-selective etching," *Sensors and Actuators*, vol. 16, pp. 67–82, 1989.

[5] M. Hirata, S. Suwazono, and H. Tanigawa, "Diaphram thickness control in silicon pressure sensors using an anodic oxidation etch-stop," *Journal of the Electrochemical Society*, vol. 134, pp. 2037–2041, August 1987.

[6] Y. P. Xu and R. S. Huang, "Anodic dissolution and passivation of silicon in hydrazine," *Journal of the Electrochemical Society*, vol. 137, pp. 948–953, March 1990.

[7] T. N. Jackson, M. A. Tischler, and K. D. Wise, "An electrochemical p-n junction etch-stop for the formation of silicon microstructures," *IEEE Electron Device Letters*, vol. EDL-2, pp. 44–45, February 1981.

[8] S. C. Kim and K. D. Wise, "Temperature sensitivity in silicon piezoresistive pressure transducers," *IEEE Transactions on Electron Devices*, vol. ED-30, pp. 802–810, July 1983.

[9] E. D. Palik, J. W. Faust, Jr., H. F. Gray, and R. F. Greene, "Study of the etch-stop mechanism in silicon," *Journal of the Electrochemical Society*, vol. 129, pp. 2051–2059, September 1982.

[10] J. W. Faust, Jr. and E. D. Palik, "Study of the orientation dependent etching and initial anodization of Si in aqueous KOH," *Journal of the Electrochemical Society*, vol. 130, pp. 1413–1420, June 1983.

[11] O. J. Glembocki, R. E. Stahlbush, and M. Tomkiewicz, "Bias-dependent etching of silicon in aqueous KOH," *Journal of the Electrochemical Society*, vol. 132, pp. 145–151, January 1985.

[12] R. L. Smith, B. Kloeck, N. F. de Rooij, and S. D. Collins, "The potential dependence of silicon anisotropic etching in KOH at 60°C," *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*, vol. 238, pp. 103–113, 1987.

[13] H. Seidel, "The mechanism of electrochemical and anisotropic silicon etching and its applications," in *The Third Toyota Conference: Integrated Micro Motion Systems*, (Aichi, Japan), October 1989.

[14] L. D. Clark, Jr. and D. J. Edell, "KOH:H$_2$O etching of (110) Si, (111) Si, SiO$_2$, and Ta: an experimental study," in *IEEE Micro Robots and Teleoperators Workshop*, (Hyannis, Massachusetts), November 1987.

[15] L. D. Clark, Jr., J. L. Lund, and D. J. Edell, "Cesium hydroxide (CsOH) : a useful etchant for micromachining silicon," in *IEEE Solid-State Sensor and Actuator Workshop*, (Hilton Head Island, South Carolina), pp. 5–8, June 1988.

# UTILIZATION OF POLYMER VISCOELASTIC PROPERTIES
## IN ACOUSTIC WAVE SENSOR APPLICATIONS

S. J. Martin, A. J. Ricco, and G. C. Frye

Sandia National Laboratories, Albuquerque, NM 87185

## Abstract
The changes which occur in polymer viscoelastic properties in response to cross-linking reactions and due to absorption of gas phase species have been used advantageously in several acoustic wave-based sensor applications. When a polymer film is present on the surface of an acoustic wave device, changes in the viscoelastic properties of the film induce changes in wave propagation velocity and attenuation, providing two sensor responses. Film changes which occur during polymer cross-linking allow photo-polymerization to be monitored in real time using acoustic devices. A photoaction spectrum of photoresist reveals the cross-linking wavelength with maximum quantum yield. Changes in the viscoelastic properties of a polysiloxane film induced by vapor absorption are found to be unique for each of several species, enabling differentiation of species with a single film. A Maxwell model for polymer viscoelasticity, in combination with mass loading effects, provides a sound theoretical basis for explaining observed results for both polysiloxane and polybutadiene/polystyrene copolymer films.

## Introduction
Surface acoustic wave (SAW) and acoustic plate mode (APM) devices have demonstrated utility in a number of gas and liquid phase sensing applications (1-3) as well as in the characterization of thin films (4-7). Propagating acoustic modes are excited and detected by interdigital input and output transducers separated by an interaction region in which the acoustic mode has significant displacement at the device surface. If the acoustic wave medium is coated with a film, propagation of an acoustic wave along the surface results in sinusoidal deformation of the film. Consequently, wave propagation is sensitive to the viscoelastic properties of an overlay. In general, perturbations which change the stored energy density of the wave result in velocity changes, while perturbations which alter power dissipation by the wave result in attenuation changes (8).

In this paper, we investigate acoustic wave responses arising from changing viscoelastic properties of polymer film overlays induced by (1) a photo-cross-linking reaction, and (2) vapor absorption. Various acoustic modes can be used to probe the properties of a film overlay. In the photo-cross-linking reaction, a shear-horizontal (SH) acoustic plate mode (APM) was used (3). The SH-APM, which is excited and detected in a thinned quartz plate, has displacement in the plane of the surface and normal to the direction of mode propagation. In the vapor absorption experiment, a surface acoustic wave (SAW) was used. The SAW has two components of displacement: a shear component normal to the device surface and a compressional component lying in the direction of SAW propagation.

## Monitoring Photo-Cross-Linking Reactions
Cross-linking significantly alters polymer viscoelastic properties. The free movement of one polymer chain relative to its neighbors, which is possible in the uncross-linked polymer, is severely inhibited by the formation of cross-links between chains

(9). The shear modulus of the film in particular provides an indication of the state of cross-linking.

To monitor a photo-cross-linking reaction, negative photoresist films (Waycoat HR-100) were spin-cast at 3000 rpm for 30 seconds directly onto the metallized side of APM devices (over both the IDTs and the wave path), giving a nominal thickness of 1.1 $\mu$m. Films were heated to 85 C in air for 10 minutes to remove solvent. HR-100 photoresist consists of partially cyclized polyisoprene, along with 4% (by weight) of a coupling photoinitiator. Absorption of the appropriate wavelength of light by a photoinitiator results in the formation of a cross-link between two polymer chains. To cross-link the resist films, the coated APM device was exposed to monochromatic light produced by a 150 W Xenon lamp coupled to a CVI Laser Digikrom 240 single-pass monochromator. Illumination intensity was approximately 1.2 mW/cm² at 370 nm. This wavelength is near the peak spectral sensitivity reported (14) and measured (*vide infra*) for photo-polymerization of HR-100.

Shown in Fig. 1 are the changes in APM propagation velocity v and attenuation $\alpha$ measured as the HR-100 film was cross-linked by UV illumination. During the reaction, APM velocity increased while attenuation decreased. After an exposure time of 140 minutes, velocity and attenuation reached stable values, indicating the completion of the cross-linking reaction. The optical energy density incident on the film over this time period was approximately 10.2 J/cm².

Each cross-link formed in the HR-100 film has two effects which give rise to an acoustic wave response: (1) a change in viscoelastic properties due to the formation of a cross-link between polymer chains, and (2) a decrease in surface mass density due to the liberation of two $N_2$ molecules in the cross-linking reaction (10). As discussed in more detail below, changes in mass density lead to changes in wave velocity, while changes in viscoelastic properties generally cause both wave velocity and attenuation changes.
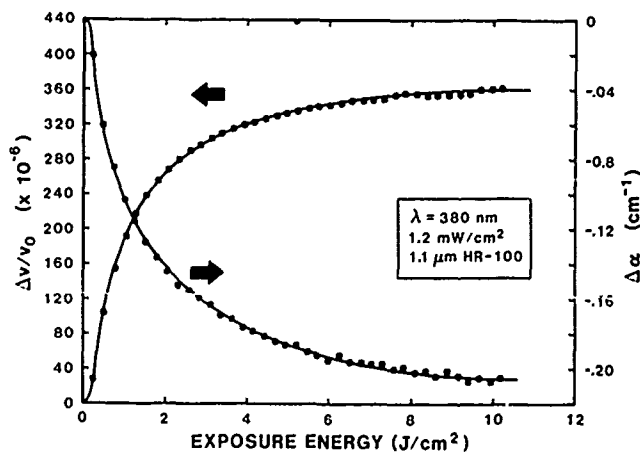


*Fig. 1. Changes in APM velocity and attenuation as a function of time measured during cross-linking of HR-100 negative photoresist film.*

## Spectral Dependence of the Cross-linking Reaction Rate

The wavelength dependence of the photosensitivity of the resist film was determined by measuring the rate at which APM velocity changed in response to cross-linking as a function of incident optical wavelength $\lambda$. The time rate of change of APM velocity at each wavelength is defined by:

$$R(\lambda) = \frac{1}{v_o} \frac{dv}{dt}\bigg|_\lambda . \qquad (1)$$

Wavelength was stepped under computer control from 200 to 650 nm in 10 nm increments while monitoring the frequency of the resist-coated APM device using the oscillator loop configuration (11). The 15-second dwell time at each wavelength was sufficient for a change in film properties to be measured without appreciably cross-linking the film: the entire spectral response experiment cross-linked the film by less than 20%. In a second wavelength scan, the spectral density $I(\lambda)$ of the source was determined using a pyroelectric detector. The relative spectral sensitivity for film cross-linking is then $R(\lambda)/I(\lambda)$.

The cross-linking rate of the HR-100 photoresist film vs. optical wavelength is shown in Fig. 2 (•). The peak in spectral sensitivity is at approximately 370 nm, in reasonable agreement with the sensitivity peak of 360 nm cited by the manufacturer (10). Also shown in Fig. 2 (x) is a UV-visible absorbance spectrum recorded for an identically prepared 1.1 $\mu$m-thick HR-100 film spin-coated onto a quartz substrate. Absorbance was measured over the 200 to 650 nm range using a Varian/Cary 2300 spectrophotometer. The absorbance spectrum has a peak at 355 nm, corresponding to the peak observed in the acoustic measurement of $R(\lambda)/I(\lambda)$. The film also exhibits very high absorbance at wavelengths below 300 nm, a result *not* observed in the acoustic measurement.
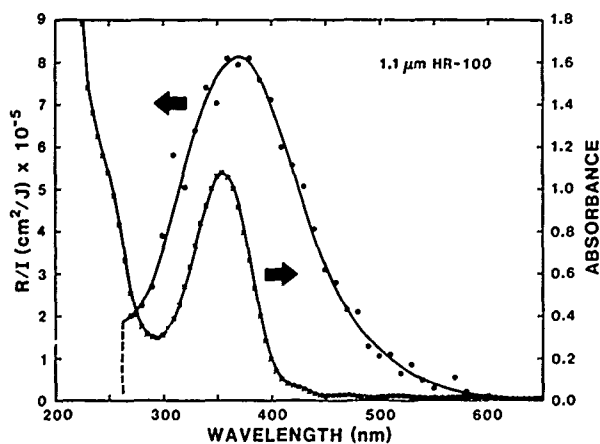
(cyclized polyisoprene), which is not photoactive; thus, its optical absorption affects neither mass density nor the elastic properties of the film.

## Viscoelastic Changes Induced by Vapor Absorption

The absorption of gas-phase species by a polymer film has several effects on SAW propagation. A decrease in SAW velocity due to the increased mass density has been well documented (1). In addition, small molecules which are molecularly dispersed in a polymer tend to plasticize the polymer, thereby changing the viscoelastic properties (9). In general, the relative contributions to velocity and attenuation changes made by different species are unique, enabling species to be discriminated on this basis.

Perturbations in SAW velocity and attenuation were measured with a polysiloxane-coated SAW device during exposure to organic vapors over a wide range of concentrations. The 1.0 $\mu$m-thick polysiloxane film was deposited by plasma-assisted chemical vapor deposition, using hexamethyldisiloxane, onto the surface of a 97 MHz SAW device. Shown in Fig. 3 is the manner in which velocity and attenuation vary in the polysiloxane-coated device as vapor concentration varies. Each data set is a parametric representation of an absorption isotherm—the locus of $(\Delta v/v_o, \Delta\alpha/k)$ points measured as solvent partial pressure varied from 0 to 97% of saturation (in nitrogen at 20 C) and back again. (For some species, the propagation loss became so large that the maximum concentration was set by the point at which insertion loss reached 65 dB.) The responses recorded during absorption and desorption track well over the two hour concentration ramp. Each vapor generates a unique locus of $(\Delta v/v_o, \Delta\alpha/k)$ points, enabling chemical differentiation on the basis of this two-parameter response. Conventional SAW gas/vapor sensors rely on velocity perturbations alone (12), necessitating the use of multiple films for discrimination.



*Fig. 2. Acoustically probed photoaction spectrum of HR-100 photoresist film (left axis); rate of change of APM velocity R(λ), normalized to spectral density I(λ), indicates rate of film cross-linking as a function of wavelength λ. Optical absorption spectrum of identical film (right axis).*
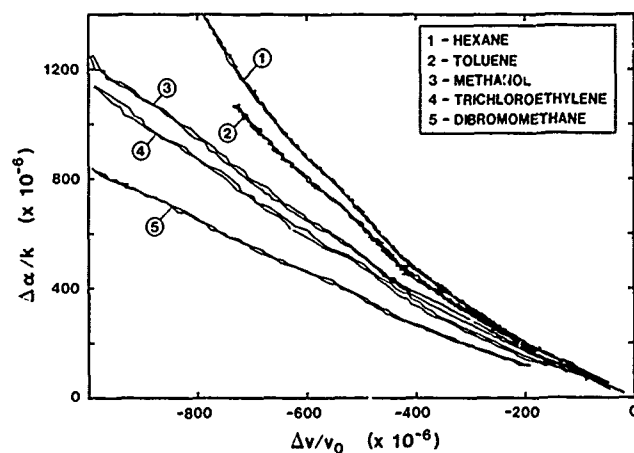


*Fig. 3. Attenuation per wavenumber vs. fractional velocity shift due to solvent vapor absorption by a polysiloxane-coated SAW device. Each vapor generates a unique locus of (Δv/v₀, Δα/k) points, enabling differentiation of species on the basis of these two responses.*

The spectral response measured using the acoustic wave device is an *action* spectrum in that only optical absorptions which lead to a chemical reaction and thereby contribute to changes in the elasticity or density of the film are registered. The peak in the absorbance spectrum (Fig. 4, x) at 355 nm is due to photoinitiator absorption, which leads to the formation of cross-links in the polymer and a corresponding peak in the acoustic $R(\lambda)/I(\lambda)$ response at 370 nm. The strong absorption at wavelengths less than 300 nm, however, is likely due to the polymer backbone

Velocity and attenuation changes measured during vapor sorption are due, to a large extent, to viscoelastic changes occurring in the film. The perturbation in SAW propagation caused by a thin-film viscoelastic overlay may be determined from a perturbation calculation. For several of the polymer films examined, the viscoelastic behavior can best be described by a simple Maxwell model, shown in Fig. 4. In this model, the viscous element (shear viscosity $\eta$) represents the frictional resistance of polymer chains to flowing past one another (9). The elastic element (shear stiffness
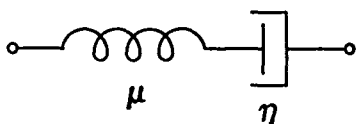
Fig. 4. The simple Maxwell model which best describes the shear viscoelastic behavior of the polymer films examined. Solvent absorption causes plasticization (reducing effective viscosity $\eta$) so that polymer changes from elastic to viscous behavior.

$\mu$) represents restoring forces arising from the tendency of polymer chains to seek their most probable configuration. The ratio of shear stress to shear strain in the polymer is given by the complex shear modulus $G = G' + jG''$. The real part, $G'$, is the storage modulus, while $G''$ is the loss modulus. A shear relaxation time $\tau = \eta/\mu$ gives the time scale for sinusoidal deformation (with angular frequency $\omega$) at which the material changes from elastic to viscous behavior: for $\omega\tau >> 1$, the material behaves elastically with $G = \mu$ (thus $\mu$ represents the glassy shear modulus of a polymer); for $\omega\tau << 1$ the material behaves viscously with $G = j\omega\eta$.

The changes in SAW velocity and attenuation due to a Maxwellian viscoelastic film are given by:

$$\frac{\Delta v}{v_o} = -\omega h \left[ c_1\rho + c_2 \frac{\mu(\omega\tau)^2}{1 + (\omega\tau)^2} \right] \qquad (2a)$$

$$\frac{\Delta\alpha}{k} = c_2\omega h \left[ \frac{\mu\omega\tau}{1 + (\omega\tau)^2} \right], \qquad (2b)$$

where $\omega$ and $k$ are the angular SAW frequency and wavenumber, respectively; $\rho$, $h$ and $\tau$ are the mass density, thickness and shear relaxation time, respectively, of the film; $c_1$ and $c_2$ are substrate-dependent constants (13). This result holds under the following assumptions: (1) film thickness is small compared with the acoustic wavelength, and (2) the magnitude of the film's bulk modulus is large compared with the shear modulus. The latter assumption is typically valid for polymers (14). The term involving $c_1$ in Eqs. 2 describes the effect of film mass density on SAW propagation; terms involving $c_2$ constitute the viscoelastic response. In general, changes in mass density affect only wave velocity, while viscoelastic changes affect both velocity and attenuation.

The viscoelastic velocity and attenuation changes given in Eqs. 2 take the form of a relaxation response. If the film is initially in the elastic (glassy) state (with $\omega\tau >> 1$) and goes through a transition to the viscous state ($\omega\tau << 1$), the SAW velocity increases monotonically, while attenuation goes through a peak at $\omega\tau = 1$. This behavior, which can be elicited by heating the film (15), is analogous to the zero-frequency glass transition.

An elastic-to-viscous transition can also be exhibited by a polymer film upon absorption of gas phase species. Small molecules which are molecularly dispersed in a polymer tend to plasticize the polymer, lowering the local friction between polymer chains and decreasing the effective viscosity $\eta$ (9). (We are neglecting here any chemical interactions between polymer and solvent, such as hydrogen bonding, which may also contribute to elastic changes.) In comparison to the large changes in viscosity which accompany this plasticization, $\mu$ is relatively unchanged, so that $\tau$ varies directly with $\eta$. The decrement in viscosity and relaxation time due to solvation of gas phase species is a physical effect found to vary exponentially with the volume fraction of the solvent present (9). Up to solvent volume fractions of approximately 30%, this may be expressed as:

$$\tau = \tau_e \exp\left[\frac{-\gamma NV}{1 + NV}\right] \qquad (3)$$

where $\tau_e$ is the relaxation time of the pure polymer, N is the number of solvent molecules per volume of pure polymer, V is the species volume, and $\gamma$ is a constant.

In order to observe the entire range of viscoelastic behavior predicted by Eqs. 2, SAW operating frequency and film properties must be such that $\omega\tau$ can be varied widely with respect to unity. With a SAW device having a center frequency of 97 MHz, a film whose relaxation time can easily be varied over this range is a copolymer of polybutadiene and polystyrene (PBPS). A 0.34 $\mu$m-thick PBPS film (composition: 80% polybutadiene, 20% polystyrene) was spin-cast onto a 97 MHz ST-cut quartz SAW device. (Very similar behavior to that reported below was observed with uncross-linked poly-1,2-butadiene, but without attaining as large a range in $\omega\tau$.)

In addition to causing plasticization, vapor absorption leads to an increase in surface mass density:

$$\Delta(\rho h) = Nmh \qquad (4)$$

where m is the mass of each vapor molecule absorbed.

The manner in which SAW velocity and attenuation vary during absorption and desorption of trichloroethylene by the PBPS film is shown in Fig. 5. The responses recorded during absorption and desorption track well over the two hour concentration ramp. When velocity and attenuation changes recorded during the isothermal measurement are plotted against each other, the data display a distinctive loop in the $\Delta v/v_o$-$\Delta\alpha/k$ plane. The shape of the isotherm is determined by the relative rates at which plasticization and mass loading occur during vapor absorption. The plasticizing action of absorbed species results in a semi-circle in the first quadrant of the $\Delta v/v_o$-$\Delta\alpha/k$ plane (15), while the mass contribution pulls these points toward lower velocity. Thus, the isothermal loop arises from a superposition of mass-loading and plasticization by the absorbed species.
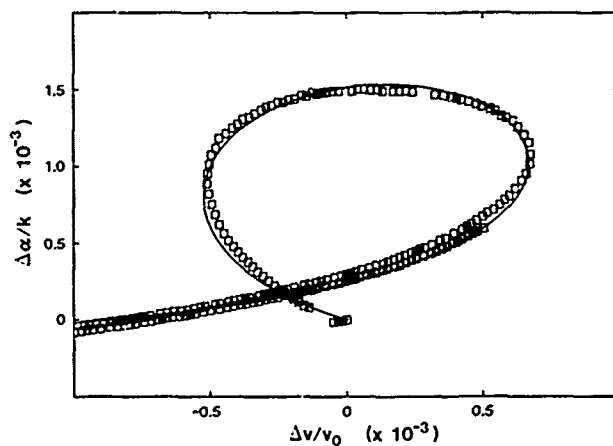


Fig. 5. Attenuation per wavenumber vs. velocity change due to absorption of trichloroethylene by a polybutadiene/polystyrene copolymer-coated SAW device. The data display an elastic-to-viscous relaxation caused by plasticization of the film by the vapor. The solid line, calculated from Eqs. 2 - 4, is based on a superposition of mass-loading and plasticization by the absorbed species.

The solid line in Fig. 3 shows how calculated values of $\Delta v/v_o$ and $\Delta\alpha/k$, based on Eqs. 2 - 4, vary with solvent concentration N.

In this calculation, N was varied from 0 to $4 \times 10^{20}$ /cm$^3$, with attenuation reaching a maximum for a volume fraction of approximately 20%. Best-fit model parameters were chosen as: $\mu = 6.9 \times 10^{10}$ dyne/cm$^2$, $\omega\tau_o = 15$, $\gamma = 0.013$. The calculated variation in $\Delta v/v$, and $\Delta\alpha/k$ follows experimentally measured values quite well. While the ability to fit the data with three parameters is not surprising, the same model parameters will also predict a diversity of data taken during absorption of pentane, methylene chloride, dibromomethane, as well as during a temperature ramp (15).

## Conclusions

Acoustic wave devices function effectively as real-time monitors of cross-linking reactions. The relative spectral sensitivity of a photoreaction can be determined from the rate of change of APM velocity as a function of optical wavelength.

Perturbations in SAW velocity and attenuation observed in some uncross-linked polymer films can be explained with a perturbational calculation based on a Maxwellian constitutive relationship for the film. This model accounts for responses observed during elastic-to-viscous transitions accompanying vapor absorption by films. The total response is a superposition of a viscoelastic response, arising from polymer plasticization, together with a mass-loading contribution.

With regard to gas/vapor sensing applications, SAW viscoelastic responses are significant when Maxwellian film properties are such that $\omega\tau$ is within an order of magnitude of unity. Since each vapor generates a unique locus of ($\Delta v/v$, $\Delta\alpha/k$) points, species can be discriminated by simultaneous measurement of these parameters during the vapor plasticization process (16). For $\omega\tau >> 1$ or $\omega\tau << 1$, the incremental changes in velocity should be dominated by mass changes. In this regime, velocity will decrease with solvent concentration without significant changes in attenuation.

## Acknowledgements

## References

1. H. Wohltjen and R. Dessy, "Surface Acoustic Wave Probe for Chemical Analysis", *Anal. Chem.*, 51 (1979) 1458-1475.

2. R. Lec, J. F. Vetelino, R. S. Falconer, and Z. Xu, "Prototype Microwave Acoustic Fluid Sensors", *Proc. 1988 IEEE Ultrasonics Symp.*, pp. 585-589.

3. S. J. Martin, A. J. Ricco, T. M. Niemczyk, and G. C. Frye, "Characterization of SH Acoustic Plate Mode Liquid Sensors, *Sensors and Actuators*, 20 (1989) 253-268.

4. D. S. Ballantine, Jr. and H. Wohltjen, "Use of SAW Devices to Monitor Visco-elastic Properties of Materials", *Proc. 1988 IEEE Ultrasonics Symp.*, pp. 559-562.

5. E. T. Zellers, R. M. White, and S. W. Wenzel, "Computer Modelling of Polymer-Coated ZnO/Si Surface-Acoustic-Wave and Lamb-Wave Chemical Sensors", *Sensors and Actuators*, 14 (1988) 35-45.

6. J. A. Groetsch III and R. E. Dessey, "A Surface Acoustic Wave (SAW) Probe for the Thermomechanical Characterization of Selected Polymers", *J. Appl. Polymer Science*, 28 (1983) 161-178.

7. G. C. Frye, S. J. Martin, and A. J. Ricco, "Monitoring Diffusion in Real Time in Thin Polymer Films Using SAW Devices", *Sensors and Materials*, 1 (1989) 335-357.

8. S. J. Martin and A. J. Ricco, "Effective Utilization of Acoustic Wave Sensor Responses: Simultaneous Measurement of Velocity and Attenuation, *Proc. 1989 IEEE Ultrasonics Symp.*, 621-625.

9. J. D. Ferry, *Viscoelastic Properties of Polymers*, Third Ed., Wiley, New York, 1980.

10. A. Stein, "The Chemistry and Technology of Negative Photoresists", Olin-Hunt Chemical Co., Palisades Park, N.J.

11. S. J. Martin and A. J. Ricco, "Monitoring Photo-polymerization of Thin Films Using SH Acoustic Plate Mode Sensors", *Sensors and Actuators*, Vol. A, No. 22 (1990) 712-718.

12. M. S. Nieuwenhuizen and A. Venema, "Surface Acoustic Wave Chemical Sensors", *Sensors and Materials*, 5 (1989) 261-300.

13. For SAWs propagating along the X-crystalline direction of ST-quartz: $c_1 = 2.03 \times 10^7$ cm$^2$-s/g, $c_2 = 2.47 \times 10^{14}$ s$^3$/g.

14. D. S. Ballantine, Jr. and H. Wohltjen, "Elastic Properties of Thin Polymer Films Investigated with Surface Acoustic Wave Devices", *Chemical Sensors and Microinstrumentation* (ACS, Washington, 1989) Ch. 15.

15. S. J. Martin and G. C. Frye, "SAW Response to Changes in Viscoelastic Film Properties", in preparation for *Appl. Phys. Lett.*

16. G. C. Frye and S. J. Martin, "Dual Output Acoustic Wave Sensors for Molecular Identification", in preparation for *Anal. Chem.*

# EFFECTS OF AGING ON POLYIMIDE:

## A STUDY OF BULK AND INTERFACE CHEMISTRY

Milan C. Buncick and Denice D. Denton

Department of Electrical and Computer Engineering and
Materials Science Program
University of Wisconsin-Madison
1415 Johnson Drive
Madison, WI 53706
(608) 263-2354

## ABSTRACT

Polyimides (PI)'s are used extensively in integrated circuit (IC) fabrication and in the fabrication of solid state sensing devices. While the hygroscopic nature of PI allows its use in relative humidity sensors, this property can also lead to long term stability and reliability problems. We are investigating the bulk and interfacial chemistry of PI exposed to heat and humidity stress over long periods. We have used electron spectroscopy for chemical analysis (ESCA) and Fourier transform infrared spectroscopy (FTIR) to examine degradation in PI when stressed.

The PI was formed from the polyamic acid precursor of pyromellitic dianhydride (PMDA) and oxydianaline (ODA). The acid was spun onto two inch Si wafers and cured at 300°C in nitrogen. The typical PI thickness was 1μm. The samples under test were placed in a high temperature (85°C) and a high humidity (85%RH) environment.

The evolution of the PI surface chemistry at the PI-air and the PI-Si interface was followed with ESCA. For exposure times greater than 550 hrs, the PI surface exhibits significant changes in emission structure and atomic concentration. These changes are, in part, attributed to breakage of the C-N bonds in the imide linkage and subsequent hydrolysis and a change in the band gap due to reduction in conjugation of the PI molecule. Angle dependent depth profiling indicates changes are consistent to the depth of the ESCA probe (about 13nm). Examination of PI at the PI-Si interface indicates no changes in the PI chemistry when compared to control samples. Bulk and gross surface chemistry were followed with FTIR. FTIR data were collected in transmission, ATR, and grazing incidence reflection configurations. In all cases there were no measurable differences in bulk PI chemistry between exposed and control samples.

The ESCA and FTIR data indicate that observed chemical changes in PI occur at the PI-air interface. Determination of the extent of this layer as a function of exposure time and a more detailed study of the chemistry is ongoing. This degradation has major reliability implications for sensors using interdigitated sensing elements, especially those with exposed PI surfaces. It also has reliability implications for use of PI as an inter-metal dielectric for sensors, IC's and packaging applications.

## INTRODUCTION

Polyimides (PI)'s have received considerable attention due to their wide industrial application. PI's are used extensively in integrated circuit (IC) fabrication as inter-metal dielectrics, passivation layers, sacrificial layers and as surface planarizers [1-5]. One goal of VLSI technology is to minimize the use of high temperature processing and wet etching chemistry. PI can be plasma etched and fully prepared at temperatures of 300-400°C.

PI has several advantages that make it attractive for microelectronics applications. In its fully cured form, PI is chemically inert in the presence of many solvents, mechanically tough and flexible, and thermally stable up to 450°C. PI is an excellent planarizer and is often used to planarize irregular surfaces. Electrically, PI is desirable for high speed applications because it has a low relative permittivity ($\varepsilon_r$ is 3.2 to 3.4 at audio frequencies [6]) and low loss over a wide frequency range. It also has a high breakdown voltage [7,8]. There are, however, some disadvantages associated with the use of PI in IC fabrication. These include sensitivity to moisture, sensitivity to variation in preparation procedures, and ionic contamination [6,9-13]. The hygroscopic nature of PI can lead to long term reliability problems in IC's such as increased insulator conductivity, loss of adhesion, and corrosion.

Very little work has been done to examine PI and PI-metal systems exposed to hostile environments over long periods. This is essential to the understanding of the long term reliability implications of introducing PI into IC fabrication schemes. We report a study of PI exposed to heat and humidity stress over long periods. We have used electron spectroscopy for chemical analysis (ESCA) and Fourier transform infrared (FTIR) spectroscopy to investigate the basic physical mechanisms that affect PI when stressed.

## EXPERIMENTAL DETAILS

The PI under investigation was formed from the polyamic acid precursor of pyromellitic dianhydride (PMDA) and oxydianaline (ODA) as shown in Fig. 1. The solvent is N-methyl pyrrolidone (NMP). The polyamic acid was spun onto two inch bare Si wafers. The spin speed was adjusted to give a PI film thickness of about 1 micron. The wafers were baked at 90°C for 30 minutes to drive off excess solvent and then cured at 300°C in nitrogen for five hours to imidize the PI. Wafers were sectioned and some pieces were exposed to heat and humidity while others remained in the laboratory environment to act as control samples.

The exposure environment was provided by a Parameter Generation and Controls environmental chamber. This chamber can provide a wide range of temperatures and relative humidities. Its operation is extremely stable. The water for the chamber is triply distilled, IC grade water. The chamber was set for 85°C/85%RH for our study.

ESCA experiments were performed using a Perkin-Elmer 5400 ESCA machine with a Mg K-alpha source (E=1253.7 eV). Data were typically acquired with the wafer normal at 45 degrees to the entrance of the analyzer. Operating pressure for the chamber was $3 \times 10^9$ Torr. The analyzer pass energy was 17.9 eV. The scan area was 1 x 3 mm. The data were smoothed and fitted using the data analysis software of the PE 5400. Charging was minimized by photoemission from the x-ray source. Emission peak shifts due to sample charging can be referenced to both the carbon 1s peak position and Auger peaks appearing in the data. Peaks shifts were less than 1.5eV.
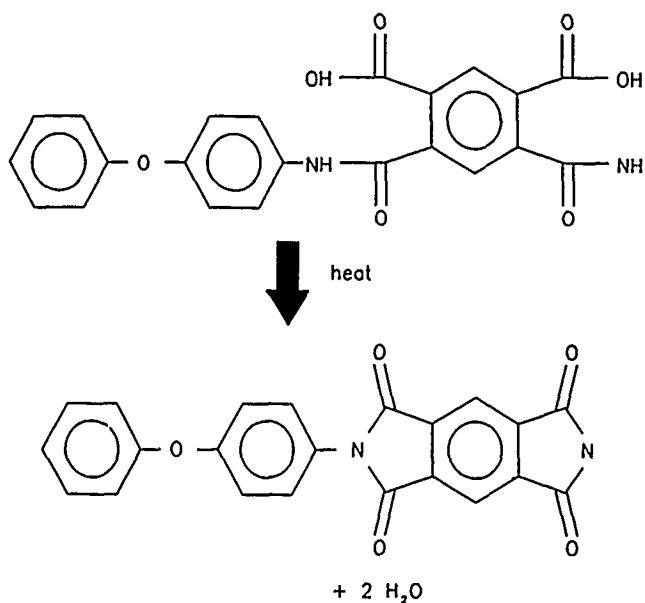
Fig. 1. Synthesis of PMDA-ODA polyimide from polyamic acid. As discussed in the text, curing was accomplished by heating at 300°C.

FTIR measurements were made with a Nicolet SX-70 spectrometer. All spectra were collected using a resolution of 4 wavenumbers and 400 scans. The sample compartment was purged for 10-15 minutes with dry nitrogen before beginning data collection.

## EXPERIMENTAL RESULTS

PI surface:

The PI surface chemistry was determined as a function of exposure time by monitoring the carbon, nitrogen, and oxygen 1s core levels with ESCA. Core level line shapes were decomposed by using summed Gaussian/Lorentzian curve fitting software on the PE 5400. Fig. 2 shows high resolution ESCA spectra for C1s, O1s, and N1s for PI control samples. The components from the curve fit for each type of bond are shown. The C-C peak in the C1s data is attributed to the two phenyl rings of the ODA part. The peak labeled PMDA/C-N is attributed to the phenyl ring and the carbon atoms bonded to nitrogen in the PMDA part. Our fit agrees with the results of previous studies of virgin films [14-18]. Area ratios are in approximate agreement with the ratios obtained by counting the number of atoms which should contribute to each peak if the intensity loss due to shake-up peaks is taken into account [15,16,18]. The intensity profiles show no significant changes between virgin PI and PI exposed for up to 500 hrs.

When the exposure time exceeds about 550 hrs we see significant changes in bond emission. The changes become more pronounced between 550 hrs and 750 hrs. For times longer than 750 hrs there is little variation in the peak shapes. Table I. shows mean atomic concentrations for several samples. The data show a slight increase in oxygen and a significant decrease in nitrogen after exposure. Fig. 3 shows the ESCA result for C1s, O1s, and N1s for a PI sample exposed 1100 hrs. The C1s data clearly show a decrease in the peak intensity for the PMDA/C-N bond indicating a change in the PMDA moiety. There is a reduction in carbonyl emission and broadening of all peaks. However, the O1s data also in Fig. 3, show an increase in the C=O/C-O-C ratio.

The most chemically unstable bond in the PI unit is the C-N bond of the imide structure [19,20]. When this bond is broken, the PI reverts to a state similar to the polyamic acid

Table I PI Surface Atomic Concentrations

| Sample | Concentration (%) | | |
| --- | --- | --- | --- |
| | C1s | O1s | N1s |
| PI unexposed | 75.5 | 18.06 | 6.44 |
| PI exposed | 73.9 | 22.11 | 3.99 |

state. The presence of water vapor hydrolyses the broken imide structure leading to a reduction in the intensity of the C-N bond a broadening and/or splitting of the carbonyl peak.
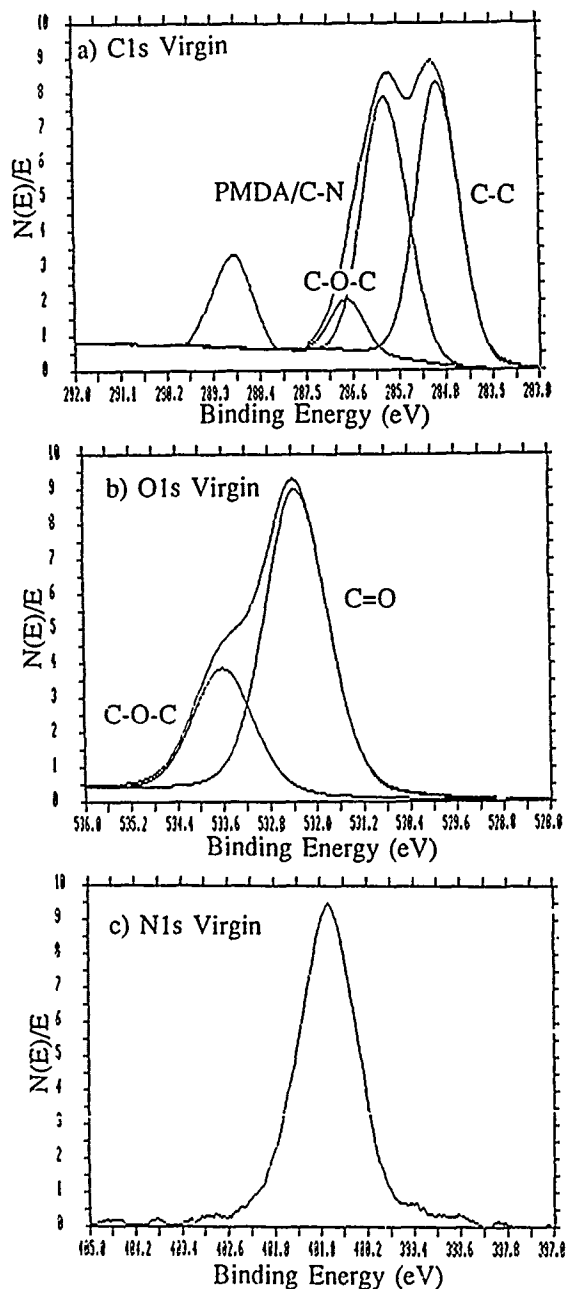


Fig. 2. PI ESCA data from the PEI Model 5400 ESCA system for an unexposed sample. a) carbon, b) oxygen, c) nitrogen.
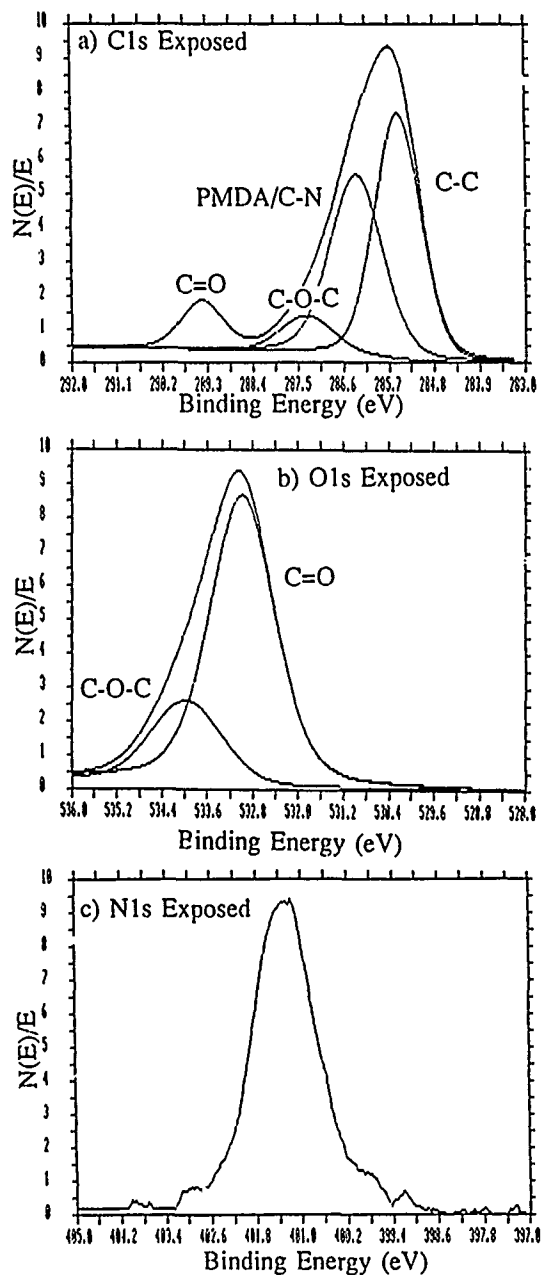
103

Fig. 3.   PI ESCA data from the PEI Model 5400 ESCA system for the PI surface of an exposed sample. a) carbon, b) oxygen, c) nitrogen.

consistent average shift to approximately 0.3 eV higher energy of all peaks for the exposed samples. While no attempts were made to correct for charging, the x-ray source tends to minimize charging. Also, exposed and unexposed samples were attached to the sample stage in pairs with each sample attached in an identical manner. Thus each sample would be expected to charge similarly. However, unexposed samples did not exhibit this shift. We interpret the energy shift in exposed samples as a change in the Fermi level, the energy reference. This can be the result of band gap shift caused by a reduction in conjugation in the polymer unit or an alteration in the number and distribution of trap states within the band gap.

## Buried PI-Si Interface:

PI-Si adhesion is typically poor and thus it is rather easy to peel the PI from the Si substrate and examine the PI at the Si-PI interface with ESCA. Fig. 4 shows the data for exposed PI at the interface after it has been separated from the Si.



There is a change in the peak intensity in the O1s due to the formation of hydroxyl groups. This emission structure has been observed in ESCA studies of polyamic acid (PAA) and during the cure of PAA into PI [21-24].

In contrast to the expected PAA emission structure, the O1s data of our study show no evidence of the formation of O-H bonds. O-H bond emission would overlap the ether bond emission and a larger peak would appear at 533-534 eV [22-24]. These data indicate that while the imide structure may be changing there is additional change in the PI surface chemistry.

By changing the collection angle of the electron analyzer ESCA can provide depth profile information to approximately 13 nm into the PI [25]. There is no change in the peak structure as a function of collection angle. Therefore to the depth of the ESCA probe the PI appears to be altered by exposure to heat and humidity.

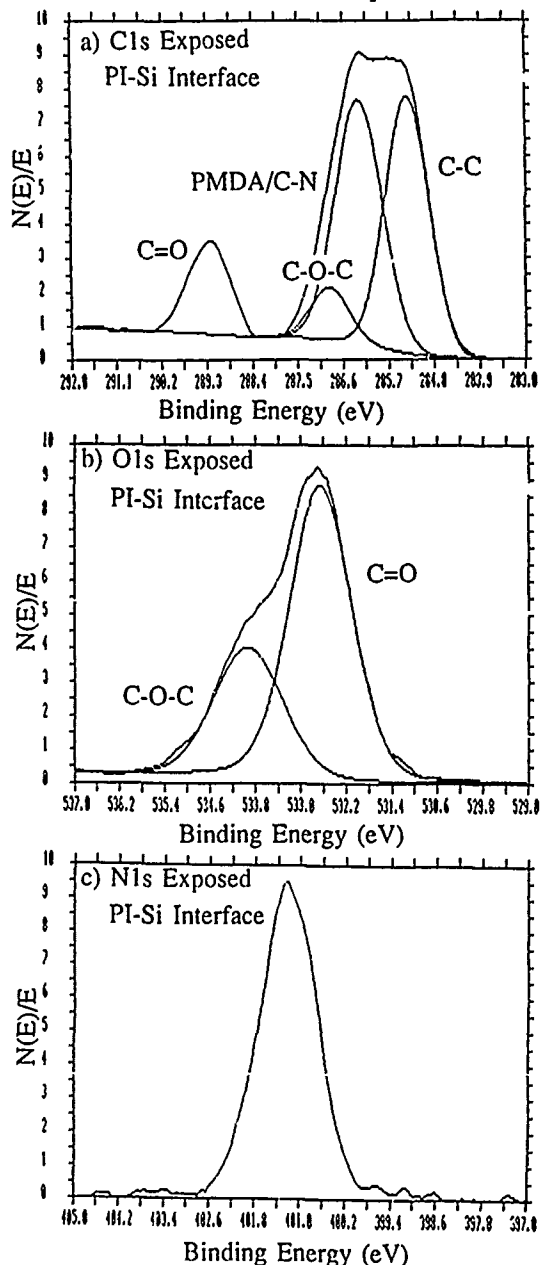In addition to the changes described above, there is a

Fig. 4.   PI ESCA data from the PEI Model 5400 ESCA system for the PI-Si interface of an exposed sample. a) carbon, b) oxygen, c) nitrogen.

The PI shows no changes from the virgin form even after an 1100 hr. exposure. Table II shows atomic concentrations for PI-Si interface scans of exposed and virgin samples. Angle of incidence depth profiling with ESCA indicates that the PI is unchanged to energy analyzer collection angles within 15° of

**Table II PI-Si Interface Atomic Concentrations**

| Sample | Concentration (%) | | |
|---|---|---|---|
| | C1s | O1s | N1s |
| PI unexposed | 73.31 | 19.78 | 6.91 |
| PI exposed | 71.87 | 20.96 | 7.17 |

the surface. At this collection angle the sampling depth is only a few Angstroms. Examination of the Si substrate shows that very little PI remains attached to the substrate. These data indicate that the PI chemistry at the interface is unaffected by exposure to heat and humidity.

**Bulk PI:**

Additional chemical characterization of PI was obtained using FTIR spectroscopy. Fig. 5 shows the absorbance spectrum for

unexposed PI. The data were taken in a transmission configuration. The peak at 1500 $cm^{-1}$ is the skeletal stretching mode for the aromatic ring in the amine moieties. This has been found to be a reliable internal standard for comparing sizes of relevant peaks [21,26]. The imide structure has absorbance peaks at 1780, 1370, and 730 $cm^{-1}$. Recently it has been found that both the 1780 and the 730 $cm^{-1}$ peaks overlap anhydride absorption bands [21]. Thus changes in the imide concentration can be reliably determined by following the 1370 $cm^{-1}$ peak as a function of the internal standard. The presence of hydroxyl groups gives rise to an absorbance band at approximately 3500 $cm^{-1}$. Fig. 6 is the absorbance spectrum for the PI sample exposed 1100 hrs. There is no measurable difference in the IR absorbance between the exposed sample and the unexposed sample when normalized using the 1500 $cm^{-1}$ band. Further, there is no measurable evidence of the formation of hydroxyl groups in the FTIR data. These data are not contrary to the ESCA results reported above. Transmission FTIR is primarily a measure of the bulk response of the PI while ESCA is a surface sensitive probe. Therefore the changes in the PI must be confined to a surface layer at the PI-air interface. This layer is thicker than the ESCA probe depth, but not of sufficient extent to be measurable by transmission FTIR.



Fig. 5. PI FTIR data for an unexposed sample. Imide bands in our data are at 725, 1377, 1725, and 1777 $cm^{-1}$. The aromatic skeletal stretching mode peak used as an internal standard appears at 1500 $cm^{-1}$.
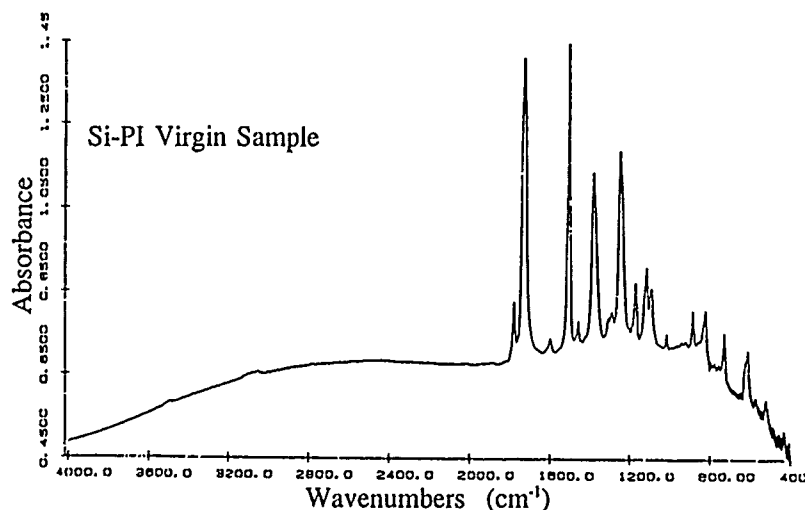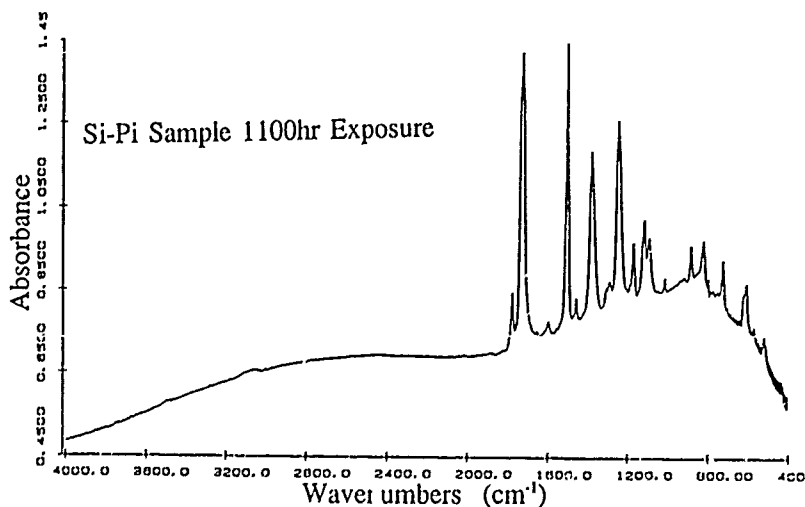


Fig. 6. PI FTIR data for an exposed sample. Imide bands in our data are at 725, 1377, 1725, and 1777 $cm^{-1}$. The aromatic skeletal stretching mode peak used as an internal standard appears at 1500 $cm^{-1}$.

105

## CONCLUSION

We have made ESCA and FTIR measurements to investigate the effects of long term exposure of PI to heat and humidity stress. The results show that changes are confined to a surface layer at the PI-air interface. ESCA data show significant changes at the PI-air interface and that the PI-Si interface remains unchanged. FTIR transmission data indicate that the bulk chemistry appears to be unaffected by such exposure. The surface chemistry is characterized by a significant reduction in the PMDA moiety and changes in the bonding of oxygen. There appear to be changes in the imide structure for the PI surface but the mechanisms of change require further study. This has reliability implications for microelectronic systems using PI as a dielectric.

## ACKNOWLEDGEMENTS

## References

1 P. Shah, D. Laks, and A. Wilson, "High Performance, High Density MOS Process Using Polyimide Interlevel Insulation", IEEE Proc. IEDM, paper 20.4, pp. 465-468, 1979.

2 P.G. Rickerl, J.G. Stephanie, and P. Slota, Jr., "Processing of Photosensitive Polyimides for Packaging Applications", IEEE, T-CHMT December 1987, pp. 690-694.

3 A.J. Gregoritsch, "Polyimide Passivation Reliability Study", IEEE Spring COMPCON, 1979, 228-233.

4 H. Umezaki, et. al., "Planar Process for 8-um Period Bubble Devices Using Polyimide Resin", IEEE Trans. on Magnetics, MAG-18, 1982, 753-758.

5 D.R. Γ.y, et. al., "Polyimide Planarization in Integrated Circuits", in K. L. Mittal (ed.), Polyimides, Plenum Press, New York, 1984, 767-782.

6 S.M. Zalar, "Dielectric characterization of polyimide thin films," Proc. First Tech. Conf. on Polyimides, Ellenville, NY, Nov. 10-12, 1982.

7 J.H. Nevin, and G.L. Summe, "DC Conduction Mechanisms in Thin Polyimide Films", Microelectronics and Reliability, 21, 1981, 699-705.

8 A.M. Wilson, D. Laks, and S.M. Davis, "Characterizing Polyimide Films for Semiconductor Applications", Organic Coatings and Plastics Chemistry, 43, 470.

9 D.D. Denton, and S.D. Senturia, "Fundamental issues in the design of polymeric capacitive moisture sensors," Proc. Int'l. Conf. on Solid st. Sensors and Actuators. Philadelphia. PA. 202-205, 1985.

10 D.D. Denton, J.B. Camou, and S.D. Senturia, "Effects of moisture uptake on the dielectric permittivity of polyimide films," Proc. Int'l Symposium on Moisture and Humidity, Washington, DC, 1985, 505-514.

11 D.D. Denton, D.R. Day, D.F. Priore, S.D. Senturia, E.S. Anolick, and D. Scheider, "Moisture diffusion in polyimide films in integrated circuits," J. Electronic Mater.,14(2) 119-136, 1985.

12 E. Sacher and J.R. Susko, "Water permeation of polymer films I Polyimide," J. Appl. Polymer Science, 23, 2355-2364, 1979.

13 E. Sacher and J.R. Susko, "Water permeation of polymer films III High-temperature polyimides," J. Appl. Polymer Science, 26, 679-686, 1981.

14 H.J. Leary and D.S. Campbell, "Surface analysis of aromatic polyimide films using ESCA," Surf. and Inter. Anal., 1(3) p. 75, 1975.

15 B.D. Silverman, P.N. Sandia, P.S. Ho, and A.R. Rossi, "Origin of the 1s-core level shifts in polyimide model compounds," J. Poly. Sci.: Poly. Ch. Ed. 23, 2857-2863, 1985.

16 B.D. Silverman, J.W. Bartha, J.G. Clabes, P.S. Ho, and A.R. Rossi, "Molecular Orbital analysis of the XPS spectra of PMDA-ODA polyimide and its polyamic acid precursor," J. Poly. Sci.: A 24, 3325-3333, 1986.

17 L.P. Buchwalter, B.D. Silverman, L. Witt, and A.R. Rossi, "X-ray photoelectron spectroscopy analysis of hXexafluorodianhydride-oxydianiline: substantiation for substituent effects on aromatic carbon 1s binding energies," J. Vac. Sci. Technol. A 5(2), 226-230 Mar/Apr 1987.

18 R. Haight, R.C. White, B.D. Silverman, and P.S. Ho, "Complex formation and growth at the Cr- and Cu-polyimide interface," J. Vac. Sci. Technol. A 6 (4), 2188-2199, Jul/Aug 1988.

19 H. Yu and W.J. Pietro, Dept. of Chemistry, University of Wisconsin, Private Communication.

20 C.A. Pryde, "Polyimide Hydrolysis: Measurement by FTIR Spectroscopy", ACS Symposium Series 407, J.H Lupinski and R.S. Moore, Eds. Chapter 4, p. 57 (1989).

21 C.A. Pryde, "IR Studies of Polyimides: I. Effects of Chemical and Physical Changes During Cure", J. Poly. Sci. Part A, Polymer Chem. 27, P. 711 (1989).

22 J.R. Salem, F.O. Sequeda, J. Duran, and W.Y. Lee, "Solventless polyimide films by vapor deposition,' J. Vac. Sci. Technol.A 4(3), May/June 1986.

23 S.G. Anderson, H.M. Meyer, III, Lj Atanasoska, and J.H. Weaver, "Dynamics of polyimide curing and degradation: An in situ x-ray photoemission study," J. Vac. Sci. Technol. A 6(1), Jan/Feb 1988.

24 M. Grunze and R.N. Lamb, "Preparation and adhesion of ultrathin polyimide films on polycrystalline silver," Chem. Phys. Let. 133(4) p. 283, January 1987.

25 T.A. Carlson, "Photoelectron and Auger Spectroscopy," Plenum, New York, 1975.

26 L.J. Bellamy, "Infrared Spectra of Complex Molecules," 3rd. Ed. Chapman and Hall, London, p. 79, 1975.

# An Implantable CMOS Analog Signal Processor for Multiplexed Microelectrode Recording Arrays

Jin Ji and Kensall D. Wise

*Center for Integrated Sensors and Circuits*
University of Michigan
Ann Arbor, MI 48109-2122

## ABSTRACT

This paper describes a second-generation multichannel probe designed for measuring single-unit activity in neural structures. The probe includes CMOS circuitry for amplifying and multiplexing the recorded signals and for electronically positioning the recording sites with respect to the active neurons. The probe offers a typical ac gain of 150 (50Hz to 10KHz), a dc gain of 0.3, and an equivalent input noise of 13$\mu$V-rms. Eight active recording sites are selected from among 32 on the probe shank using a static input channel selector. The probe implements channel selection, self-testing, data output, and initialization using a three-lead connection to the outside world. The probe is realized using 12 masks in a high-yield single-sided dissolved wafer process with a 3$\mu$m feature size for the circuitry and a 3$\mu$m pitch on the electrode shanks.

## INTRODUCTION

Over the past twenty years, there have been many efforts to adapt solid-state process technology to the formation of thin-film microelectrode arrays for the study of signal processing in the central nervous system. Using micromachined silicon substrates formed using boron etch-stops, such microprobes have been shown [1] to couple tightly to neural tissue and accurately sample spatially separated neural activity in short-term (acute) experiments. The major barriers for longer-term (chronic) use are associated with the output leads. Large-numbers of recording sites are desirable for sampling the total activity within even a restricted tissue volume; however, the tethering on the probe from such leads would be unacceptable even if there were room to attach them to the probe substrate. In addition, the 1-10 megohm impedance levels associated with typical recording sites make the signals extremely sensitive to leakage on the output leads and make these leads very difficult to encapsulate adequately for chronic use. Thus, one of the major advantages to the use of a silicon substrate is the ability to embed signal processing electronics in it, both for signal amplification and buffering (to reduce output lead encapsulation difficulties) and for multiplexing to reduce the number of leads required.

The ability to realize circuitry within the probe substrate has been previously demonstrated [2,3]. Using three-channel non-multiplexed and ten-channel multiplexed designs in E/D NMOS technology, active probes have been demonstrated, and single-unit cortical activity has been recorded. This paper represents a significant extension of these results, reporting an on-chip CMOS signal processor which offers higher gain, electronic positioning of the active recording sites, reduced noise, low- and high-frequency bandlimiting, bidirectional signaling over a single output lead, and functional upward compatibility with the previous circuit generation. The merging of high-performance CMOS circuitry with the deep-boron-diffused probe also illustrates a process having applicability to a variety of other devices as well.

## MERGED PROCESS CONSIDERATIONS

Figure 1 shows a view of the overall probe structure along with

a photograph of a first-generation ten-channel E/D NMOS probe. While the width of this probe shank is about 140$\mu$m, scaling studies [4] have shown the process compatible with the formation of shanks as narrow as 10-20$\mu$m, so that width becomes a function primarily of the number of shank conductors and the feature size of the lithography used. A focus of the second generation probe development was also to shrink the feature size on the shank from the 6$\mu$m of the first-generation probe to 1.5$\mu$m.
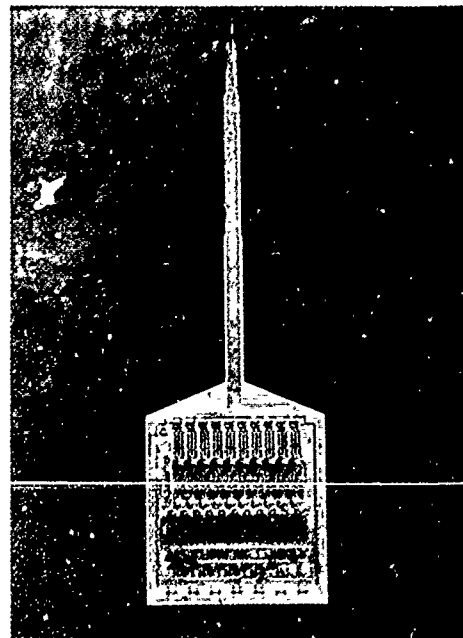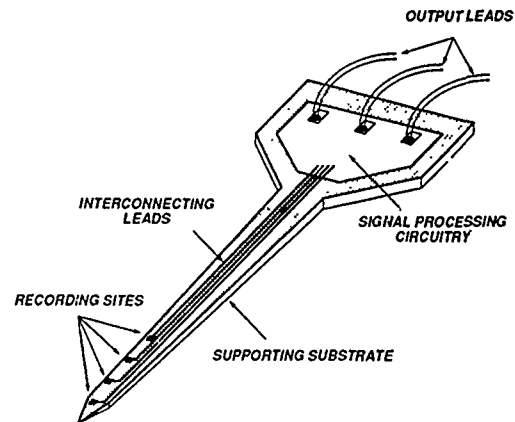




Fig. 1 Structure and Photograph of a Multichannel Microprobe. The overall length of the probe is 4.7mm, with recording sites separated by 100$\mu$m in depth along the shank.

Figure 2 shows the process flow for the second-generation device. There are three areas in which the merged process differs from a standard bulk p-well CMOS process. 1) the p-well driven in and the deep-boron shank diffusion are performed simultaneously; 2) the pad oxide and LOCOS nitride layers are also used for the lower dielectric under the shank conductor; and 3) the metal system is modified to allow the use of high-temperature LPCVD dielectrics over the entire structure prior to final etch. The resulting process is single sided on wafers of normal thickness and requires 12 masks (eight for the circuitry alone). Both the bonding pads and the recording sites are inlayed with a noble metal, typically gold, just prior to patterning the field dielectrics.



(a) SELECTIVELY DIFFUSE SUBSTRATE TO DEFINE PROBE SHAPE

(b) PROCESS CIRCUITRY AND PATTERN ELECTRODES
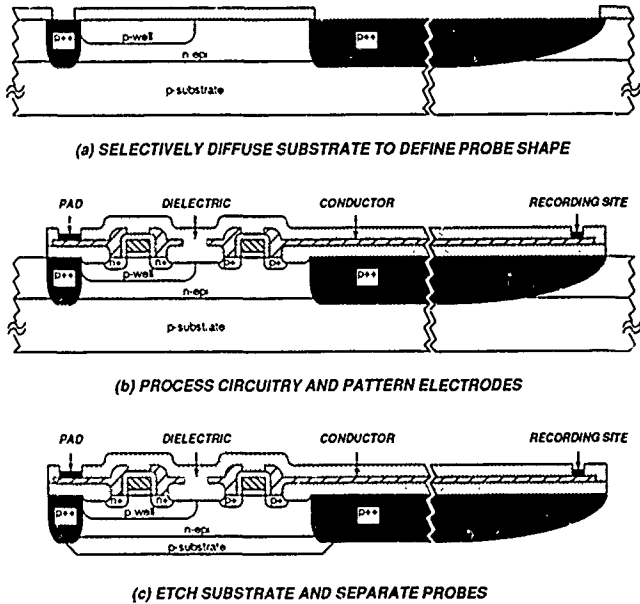
(c) ETCH SUBSTRATE AND SEPARATE PROBES

Fig. 2: General Process Flow for an Active CMOS Microprobe.

Figure 3 shows the process modification needed for the p-well drive step, which is the most significant of the circuit process modifications. In the usual process, the implant would be done at 100KeV using a dose of $1.5 \times 10^{13}\text{cm}^{-2}$, producing a measured surface concentration of about $1.4 \times 10^{16}\text{cm}^{-3}$ and a final junction depth after drive-in of $6.7\mu\text{m}$. In the merged process, a thick oxide must be used to cap the p-well (and the rest of the active circuit area) during the deep-boron diffusion. As a result of surface conversion to oxide and boron segregation effects, the implant is modified by increasing its dose to $2.5 \times 10^{13}\text{cm}^{-2}$ and its energy to 400KeV. This produces a deeper peak which flattens during drive-in to produce the same surface concentration. The modified p-well depth is $5.3\mu\text{m}$. The same drive-in cycle produces a deep-boron etch-stop depth of $15\mu\text{m}$. All other CMOS processing is unchanged from the standard process with the exception of final metal, which must be compatible with LPCVD process temperatures. The process produces n- and p-channel thresholds of +0.8V and -0.7V, respectively.

## CIRCUIT AND SYSTEM DESIGN

### System Organization

Figure 4 shows the block diagram of the signal processor circuitry. An input channel selector is used to select eight active recording sites from among 32 sites located on the probe. This approach allows the channels selected for processing to be optimized for maximum electrode-cell coupling and effectively implements electronic site positioning. The selector is configured as eight independent 4:1 multiplexers, allowing 65,536 different site

selections using input latches which are loaded serially using a two-phase non-overlapping 4MHz clock. The channel selection switches are designed to minimize on-state resistance and clock noise.

STANDARD CMOS PROCESS          MODIFIED CMOS PROCESS



Grow Ion Implantation Masking Oxide; p-Well Ion Implantation

Grow Ion Implantation Masking Oxide; p-Well Ion Implantation

Grow Boron Diffusion Masking Oxide

p-Well Drive-in

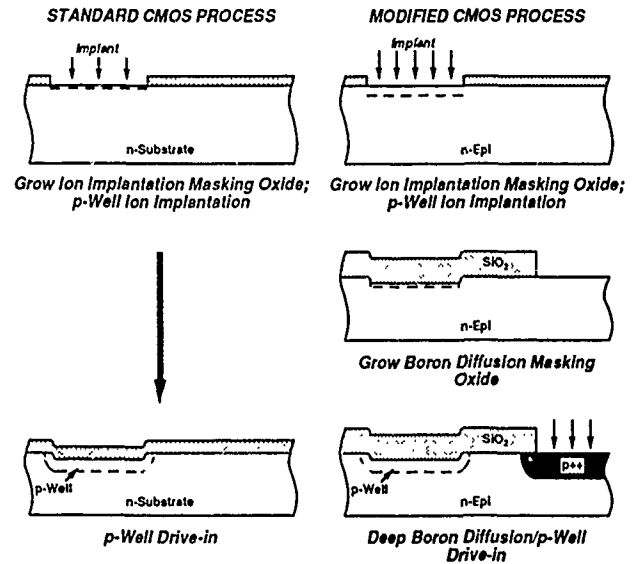Deep Boron Diffusion/p-Well Drive-in

Fig. 3: Process Modifications Required in Merging the p-well and Deep Boron Diffusion Steps in Probe Fabrication.
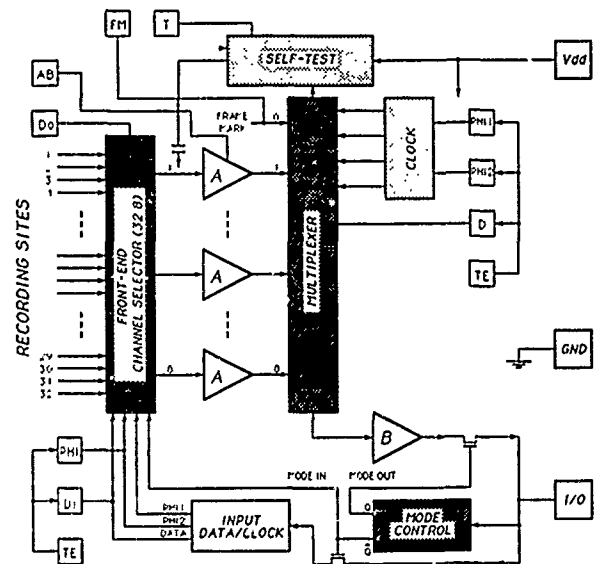


Fig. 4: Block Diagram of the Integrated Signal Processor. Only three leads are required for interface with the external world.

Channel selection data is obtained over the same lead normally used for multiplexed data output from the probe. Operation of this data lead is shown in Fig. 5. A 5V pulse applied externally to this lead selects its mode. In the normal output mode, the eight signal samples and frame marker are time-multiplexed to the external circuitry, while in the input mode, binary input data and clock signals are superimposed on the data lead. On-chip circuitry separates these signals and clocks the input channel selector to enter the desired selection code.

108

Fig. 5: Input and Output Signaling Levels on the I/O Data Line.

## Preamplifier Design

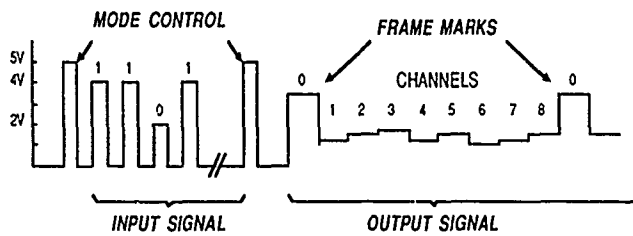Since extracellular neural signals typically have frequency components from about 100Hz to 6KHz with an amplitude of 0-500μV, the preamplifiers should offer a bandwidth of about 50Hz to 10KHz with a gain of 100-300. The circuitry must achieve this performance with no off-chip components and minimum die area. A high-frequency cutoff is necessary to avoid aliasing in the multiplexer, while a low-frequency cutoff is needed to avoid preamp saturation due to input offset voltages. The resistance of per-channel input clamp diodes is used to avoid DC input drift due to instability in the metal-electrolyte recording sites [3]. A new resistance-capacitance structure formed using a floating p-well diode is used to set the lower cutoff corner in the present amplifier design, realizing the very high resistance required without chopping or off-chip/hybrid components The preamplifier schematic is shown in Fig 6 The preamp offers an ac gain of 150 and a dc gain of 0.3. The measured passband is from 5Hz to 10KHz, with an ac input dynamic range of ±1mV.
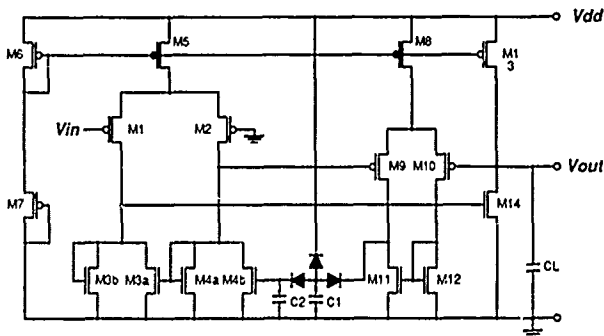


Fig. 6: Circuit Schematic of the Per-Channel Preamplifier. The preamplifier uses a diode-capacitance network formed using the resistance of a floating p-well to set the low-frequency cutoff of the circuit.

## Noise Sources

There are four important noise sources in these structures: thermal/flicker noise in the input electronics; multiplexer noise due to feedthrough of the multiplexer control signals into the analog signal path; ground bounce due to the finite resistance in the ground line of the probe and any differences in the current levels associated with different states of operation; and, optical noise induced by ambient lighting. Figure 7 shows the measured equivalent input noise voltage for the CMOS probe as a function of frequency. The noise level is about 150nV/√Hz at 1KHz and 500nV/√Hz at 100Hz, giving rise to an equivalent input noise of 13μV-rms over the signal bandwidth. This is comparable to the noise from the recording site itself, so that system noise is not substantially degraded by the on-chip electronics. Multiplexer noise in the CMOS implementation is reduced to negligible levels by differential clocking so that clock feedthrough is effectively cancelled. Multiplexer switching intervals are a' blanked by the external circuitry for additional noise suppression
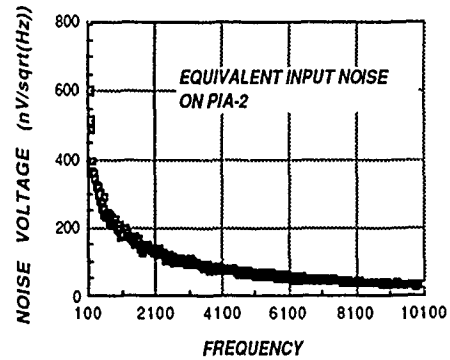


Fig. 7. Equivalent Input Noise on the CMOS Microprobe as a function of Frequency.

Since probe ground is established using an off-chip lead which must be very small, the presence of finite resistance in this lead is possible. For ribbon and discrete wire leads, a lower limit on this resistance is in the range of 2-10 ohms for a typical implant assembly. The total supply current to the probe (about 400μA) can generate a voltage of several millivolts across this resistance, however, since most of the current is consumed by the amplifiers (which do not switch) and since the multiplexer switches differentially (and at frequencies far out of band for the amplifiers), the resulting noise induced in the probe ground is negligible. Thus, the use of this ground as a reference for one side of the preamplifier is possible, in some situations, this reference can also be tied to a large-area electrode in order to cancel gross potentials resulting from aggregate neural activity.

Figure 8 shows the optical noise generation mechanism in the NMOS probe structure. Incident illumination on the back of the lightly-doped circuit area on the rear of the probe generates electron-hole pairs, which diffuse into the bulk of the silicon toward the device junctions. Electron charge is swept across these junctions into the n-type source-drain areas. Since the electrode impedance levels at the input to the preamplifier are capacitive and of the order of several megohms, charge collection on the wells of the input protection devices results in a very sensitive optical detector (closely resembling a floating-gate detection amplifier). For acute situations where light is present in the operating room, the resulting noise can swamp any neural signals present. In CMOS, however, the presence of the epitaxial-substrate junction effectively shields the input devices from any charge incident from below. Use of metal optical shield plates on top of the structure can also prevent any generation due to light incident on the upper surface of the probe. Such metal films can also form an additional layer of passivation against the saline environment.
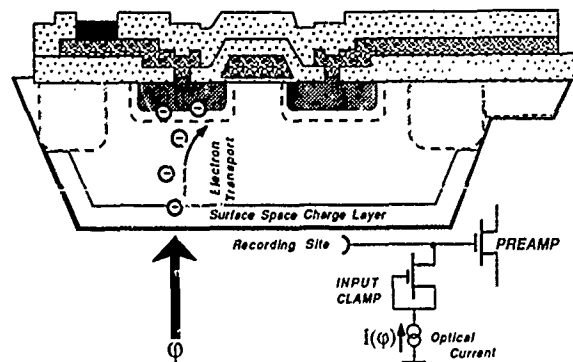


Fig. 8. Generation of Optical Noise in an Active Microprobe.

109

## Measured Performance

Figure 9 shows a photomicrograph of the signal processor portion of the second-generation probe, implemented here with all 32 input leads available externally for test purposes. The active die area is 2.5mm$^2$, with a power dissipation of 2mW. Measured performance and circuit waveforms are given in Fig. 10. The on-chip functions performed include channel selection, amplification, signal bandlimiting, multiplexing, clocking, power-on-reset (initialization), and self-testing. The feature size for this circuitry is 3$\mu$m, while to minimize the width of the shank the interconnect pitch there is 3$\mu$m (1.5$\mu$m line and space).
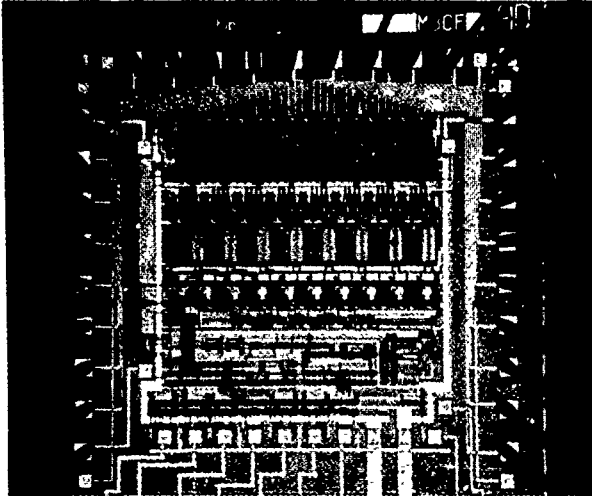


Fig. 9: Photomicrograph of a Completed Circuit Die. The chip is fabricated in a 3$\mu$m bulk p-well CMOS process with a die size of 1.58mm x 1.60mm.

## Acknowledgement

## References

[1] K. L. Drake, K. D. Wise, J. Farraye, D. J. Anderson, and S. L. BeMent, "Performance of Planar Multisite Microprobes in Recording Extracellular Single-Unit Intracortical Activity," *IEEE Trans. Biomed. Engr.*, 35, pp. 719-732, September 1988.

[2] K. Najafi, K. D. Wise, and T. Mochizuki, "A High-Yield IC-Compatible Multichannel Recording Array," *IEEE Trans. Electron Devices*, 32, pp. 1206-1211, July 1985.

[3] K. Najafi and K. D. Wise, "An Implantable Multielectrode Array with On-Chip Signal Processing," *Digest ISSCC*, pp. 98-99, February 1986. See also, *IEEE Journal of Solid-State Circuits*, pp. 1035-1044, December 1986.

[4] K. Najafi, J. Ji, and K. D. Wise, "Scaling Limitations of Silicon Multichannel Recording Probes," *IEEE Trans Biomed Engr*, 37, pp. 1-11, January 1990.

[5] J Ji, K Najafi, and K D Wise, "A Low-Noise Demultiplexing System for Active Multichannel Microelectrode Arrays," *IEEE Trans Biomed Engr*, to be published.

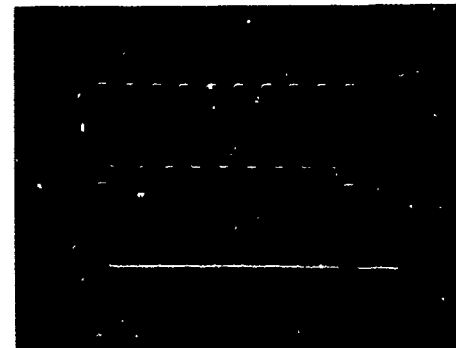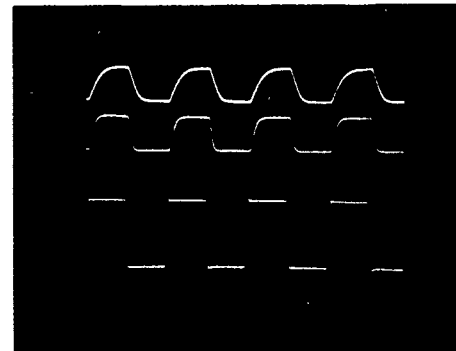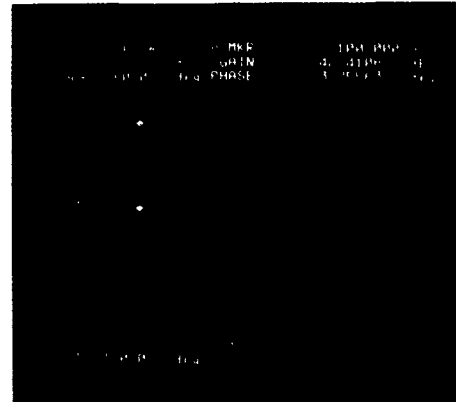| Function | Measured Performance |
| --- | --- |
| Mid-Band AC Amplifier GAin | 150 |
| Amplifier Bandwidth | 5 Hz to 10 kHz |
| DC Amplifier Gain | 0.3 |
| Input DC Range | ±100 mV |
| Input AC Range | ±1 mV |
| On-Chip Clock Frequency | 220 kHz |
| Output Strobe Level *(Frame Marker)* | 3.9 V |
| Buffer Output DC Level | 3.0 V |
| Data/Clock Separation Threshold | 3.1 V |
| I/O Switching/Control Level | 4.5 V |
| Self-Test Enable Level *(on VDD)* | 8.1 V |



Fig. 10. Measured Performance of the On-Chip Electronics. The photographs show (from top) the gain/phase response of the preamplifier as a function of frequency, separation of superimposed data and clock signals (with an all-1's signal at the bottom and the separated signals on the upper traces), and the on-chip multiplexer clock and frame reset signals.

# 2-D MAGNETIC FIELD SENSOR BASED ON VERTICAL HALL DEVICE

Lj. Ristic, M. Paranjape and M.T. Doan

Department of Electrical Engineering and Alberta Microelectronic Centre
University of Alberta, Edmonton, Alberta
Canada, T6G 2G7

## ABSTRACT

A silicon sensor capable of measuring two components of the magnetic field in the plane of the chip, based on Vertical Hall device, is presented. The device is designed and fabricated in 2 μm CMOS process. The device consists of five current contacts and four Hall contacts. It has a linear response to magnetic field on both channels.

## INTRODUCTION

Magnetic field sensors are usually designed to sense one component of magnetic field, either perpendicular [1-4] or parallel [5-11] to the chip surface. Only a few attempts can be found in literature for two-dimensional magnetic field measurements [7,12,13]. Zieren et al. [7] and Maenaka et al. [12] have reported results on 2-D magnetic field sensors based on vertical magnetotransistors fabricated in a bipolar process. We have reported recently [13] a 2-D magnetic field sensor based on lateral magnetotransistor fabricated using a CMOS process, as well as a lateral magnetotransistor with single collector capable of sensing magnetic field in two different directions [14].

In this paper, we present a novel silicon structure fabricated in 2 μm CMOS process, based on Vertical Hall device [8].

## DEVICE STRUCTURE AND PRINCIPLE OF OPERATION

A new device capable of simultaneously measuring two components of magnetic field, based on Vertical Hall device, is shown in Figure 1. It is comprised of five current contacts and four Hall contacts, all of which are made as $n^+$ within an n-substrate. $C_0$ serves as a central current contact, while $C_1$ to $C_4$ are outside current contacts. The four Hall contacts, labelled $V_{H1}$ to $V_{H4}$, are placed symmetrically with respect to the central current contact. They facilitate the measurement of the generated Hall potential. The central current contact is designed in a special way to create two channels of lateral carrier flow in the x and y directions. This is achieved by $p^+$-regions placed at the four corners of $C_0$.



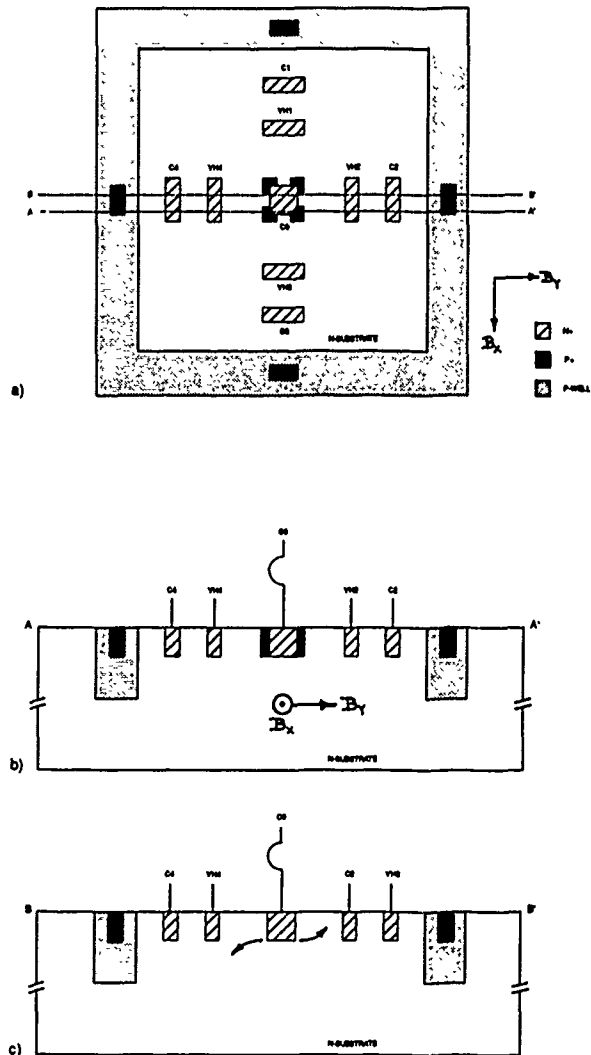Fig. 1. Vertical Hall device designed to sense two components of magnetic field. $C_0$, central current contact. $C_1$ to $C_4$, outside current contacts. $V_{H1}$ - $V_{H4}$, Hall contacts. a) Top view. b) Cross-section A-A', suppression of current flow in corner regions. c) Cross-section B-B', deflection of carriers by Lorentz force.

These p+-regions form p-n junctions and suppress current flow in the corner regions of $C_0$, as in Figure 1b. However, the middle edges of current contact $C_0$ represents an n+- n junction which allows flow of carriers in the x and y directions. The whole device is surrounded by a p-well ring, which serves as a device shield [8]. The n+-regions were realized using the standard doping process for the source and drain of n-channel MOS transistors. The p+-regions on the four corners were formed using the standard doping process for the source and drain of p-channel MOS transistors.

The device operates as follows. A constant current I is supplied to the central current contact $C_0$. This current is shared by the four outside contacts $C_1$ to $C_4$, which are grounded. The lateral flow of carriers is shaped in the x- and y- directions, forming two channels. In the absence of magnetic field, the currents of all four outside contacts are equal because of the device symmetry. Upon application of a magnetic field parallel to the chip surface and consisting of two components, $B_x$ and $B_y$, the currents of the outside contacts will change due to the action of the Lorentz force (deflection of carriers). For example, the $B_x$ component, which is parallel to the chip surface and oriented in the x-direction, will cause a deflection of carriers flowing in the y-directions, as shown in Figure 1c. As a result, the potential at $V_{H2}$ will increase, while $V_{H4}$ will decrease. Therefore, the difference $V_{H2}$ - $V_{H4}$ = $V_H(x)$ can be used to detect $B_x$. In a similar manner, $V_{H1}$ and $V_{H3}$ are affected by the magnetic field component, $B_y$.

## EXPERIMENTAL RESULTS

The characterization of the device has been done using equipment which includes the following: a variable four-inch electromagnet (Varian V-400S); Gauss meter using an axial InSb probe; and a programmable semiconductor parameter analyzer (HP 4145A).

The device is characterized in terms of sensitivity:

$$S_i = \frac{V_H}{I} \times \frac{1}{B_i} \qquad (1)$$

for i=x, y

where I is the driving current, $B_i$ the intensity of magnetic induction, and $V_H$ the Hall voltage.

Figure 2 shows the change of the Hall voltage for both the $V_H(x)$ and $V_H(y)$ channels as a function of magnetic induction. It can be noticed that the response of both channels is linear to magnetic field. The characterization of the sensor with respect to biasing current, I, is presented in Figure 3. As can be seen, $V_H$ increases linearly with current.

Finally, our investigation has shown that there is little or no cross-sensitivity between the two channels.



**Vh vs B FOR X- AND Y-AXES**
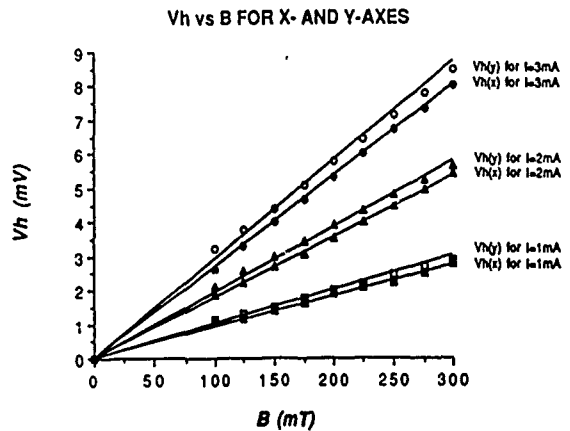
Fig. 2.    $V_H$ as a function of $B_i$, where i=x, y. $C_0$ is driven by I. All outside contacts, $C_1$ to $C_4$, are grounded. $V_H$ (x) = $V_{H2}$ - $V_{H4}$ and $V_H$ (y) = $V_{H1}$ - $V_{H3}$. $S_x$=9 V/T·A and $S_y$=10 V/T·A.



**Vh vs I FOR X- AND Y-AXES**
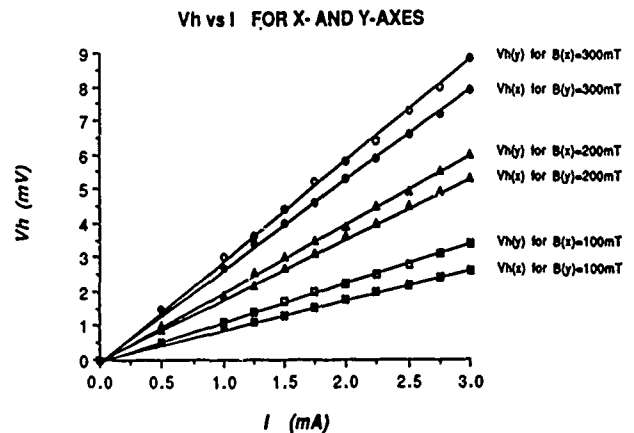
Fig. 3.    $V_H$ as a function of I. All outside contacts, $C_1$ to $C_4$, are grounded. $V_H$ (x) = $V_{H2}$ - $V_{H4}$ and $V_H$ (y) = $V_{H1}$ - $V_{H3}$.

## CONCLUSIONS

A novel magnetic field sensor based on Vertical Hall device has been designed and fabricated in a standard 2 μm CMOS process. The device is capable of simultaneous measurement of two components of magnetic field in the plane of the chip. The sensor exhibits a linear response to magnetic field along both channels, yet does not show cross-sensitivity between the two channels.

112

## REFERENCES

1. G.S. Randhawa, "Monolithic Integrated Hall Devices in Silicon Circuits", Microelectronics Journal, Vol. 12, pp. 24-29, 1981.

2. S. Takamiya and K. Fujikawa, "Differential Amplification Magnetic Sensor", IEEE Trans.

3. R. C. Gallagher and W. S. Corak, " A Metal-oxide Semiconductor (MOS) Hall Element", Solid-State Electronics, Vol. 9, pp. 571-580, 1966.

4. L. W. Davies and M. S. Wells, "Magnetotransistor Incorporated in an Integrated Circuit", Proc. IREE Australia, 1971, pp. 235-238.

5. O. S. Lutes, P. S. Nussbaum, and O. S. Aadland, "Sensitivity Limits in SOS Magnetodiodes", IEEE Trans. Electron Devices, Vol. ED-27, pp. 2156-2157, 1980.

6. I. M. Mitnikova, T. V. Persiyanov, G. I. Rekalova, and G. Shtyubner, "Investigation of the Characteristics of Silicon Lateral Magnetotransistors with Two Measuring Collectors", Soviet Phys. Semiconductors, Vol. 12, pp. 26-28, 1978.

7. V. Zieren and B. P. M. Duyndam, "Magnetic Field Sensitive Multicollector n-p-n Transistors", IEEE Trans. Electron Devices, Vol. ED-29, pp. 83-90, 1982.

8. R. S. Popovic, "The Vertical Hall-Effect Device", IEEE Electron Devices Lett., Vol. EDL-5, pp.357-358, 1984.

9. R. S. Popovic and R. Widmer, "Magnetotransistor in CMOS Technology", IEEE Trans. Electron Devices, Vol. ED-33, pp. 1334-1340, 1986.

10. Lj. Ristic, H. P. Baltes, T. Smy, and I. Filanovsky, "Suppressed Sidewall Injection Magnetotransistor with Focused Emitter Injection and Carrier Double-Deflection", IEEE Electron Devices Lett., Vol. EDL-8, pp. 395-397, 1987.

11. Lj. Ristic, T. Smy, and H. P. Baltes, "A Lateral Magnetotransistor Structure with Linear Response to the Magnetic Field", IEEE Electron Devices, Vol. ED-36, pp. 1076-1086, 1989.

12. K. Maenaka, H. Fujiwara, T. Ohsakama, M. Ishida, T. Nakamura, A. Yoshida, Y. Yasuda, "Integrated Magnetic Vector Sensor", Proc. of the 5th Sensor Symposium,1985, pp. 179-183.

13. Lj. Ristic, M.T. Doan, and M. Paranjape, "2-D Integrated Magnetic Field Sensor in CMOS Technology", Presented at the 32nd Midwest Symposium on Circuits and Systems, 1989, Urbana, Illinois, U.S.A.

14. Lj. Ristic and M.T. Doan, "Lateral Transistor Structure Sensitive to Magnetic Field Applied Either Parallel or Perpendicular to Chip Surface", Presented at MIEL '90, Ljubljana, 1990.

113

# A HIGH-SENSITIVITY MAGNETORESISTIVE SENSOR

J.E. Lenz, G.F. Rouse, L.K. Strandjord, B.B. Pant, A. Metze,
H.B. French, E.T. Benser, D.R. Krahn

Honeywell Systems and Research Center
Honeywell Solid State Electronics Center
Minneapolis, Minnesota

## ABSTRACT

A high-sensitivity magnetic field sensor has been fabricated and tested based on a permalloy (NiFe) magnetoresistive (MR) transducer and low-power electronics. The basics of the magnetoresistive transducer, and the concept, design, fabrication, and characterization of a single-axis, thin-film magnetometer are presented. Minimum sensor noise is balanced with minimum sensor power dissipation to provide a sensor with a minimum noise of 2 $\mu G/\sqrt{Hz}$ at a power dissipation of 800 $\mu W$. A discussion of the use of this sensor in magnetic anomaly detection applications is also included.

## INTRODUCTION

Magnetic sensing techniques exploit a broad range of physics and chemistry disciplines. For the past 40 years, magnetic sensor development has concentrated on two attributes—higher sensitivity and lower power.[1] In this paper a sensor technology that addresses both these attributes is described. The sensor is based on the magnetoresistive effect exhibited in a thin film of ferromagnetic material. Solid-state processing of the sensing element (i.e., transducer) is employed to offer uniformity and repeatability for producing the ferromagnetic thin film. An added benefit to basing the sensor design on solid-state technology is the low unit cost for high-volume production and the ability to integrate the sensor electronics with the transducer.

## THE MAGNETORESISTIVE EFFECT

Figure 1 shows the magnetoresistive effect where the resistance of a ferromagnetic thin film varies with the direction of magnetization in the film. The resistance is highest when the magnetization is parallel to the current and lowest when it is perpendicular to the current. The magnetization rotates in the direction of an applied magnetic field, however, it generally does not end up pointing in the same direction as the field because its direction is determined by several competing factors. One factor is the easy axis (the direction along which the magnetization in a circular film prefers to lie), which is determined by the magnetic field present during deposition of the film. Another is the shape of the film, which in the case of a long, thin film, keeps the magnetization in the plane of the film and tends to make it point along the length of the film.[2]

Permalloy is the most common material for magnetoresistive sensors because it has a relatively high magnetoresistive coefficient and also because its characteristics (e.g., a zero coefficient of magnetostriction) are compatible with the fabrication techniques employed to make silicon integrated circuits. Figure 2 shows a typical MR curve for permalloy. The resistance change is approximately 2% at saturation. Other materials or material geometries can offer a much higher MR effect. For example, multilayer stacks of magnetic thin films

exhibit an MR coefficient of nearly 25%;[3] however, for this sensor development a stable and repeatable readout of the MR effect is equally critical. Permalloy is preferred at this time because it has many years of processing history with its use in temperature sensors.
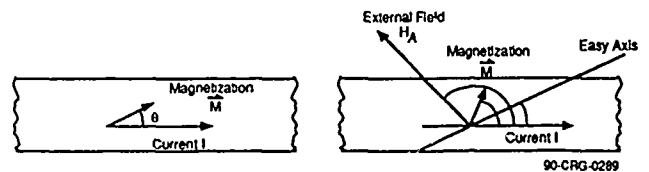


**Figure 1. The Magnetoresistive Effect in a Thin Current Strip**
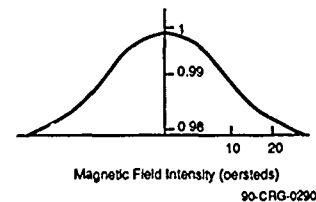


**Figure 2. The Nominal Magnetoresistive Effect for a Permalloy Thin Film**

## MAGNETORESISTIVE PATTERNS FOR SENSING

Magnetoresistive sensors incorporate various intricate patterns to exploit certain profiles of the magnetic field to be measured. Most of these patterns are variations on a familiar circuit element: the Wheatstone bridge (see Figure 3). The resistance value of the four legs is designed to be identical. Because of processing variations, laser trimming can be done to accurately balance the bridge in the null condition. By matching the legs, the temperature coefficient of the sensor can be reduced by six orders of magnitude. To make the bridge operate in the linear region of the magnetoresistance curve, a bias field is applied along a magnetic symmetric axis of the bridge. The bias field affects the resistance of the legs, but because the magnetic field and the current form the same angle at all four legs, all of the resistances change by the same amount. The bridge still has no output voltage. If the sensor is then placed in a signal field (i.e., the magnetic field to be sensed) directed along the other axis of the bridge, because of magnetic asymmetry in the pattern, the resultant field vector forms one angle with the current in two opposite legs and a different angle with the current in the remaining two legs. In two opposite legs the resistance decreases and in the remaining two it increases. The bridge develops an output voltage that is proportional to the signal field.

The output voltage is also proportional to the sensitivity of the bridge. For a given supply voltage, the bridge's sensitivity is the ratio of the output voltage to the signal field. The more sensitive

the bridge, the larger the output voltage it will develop in a signal field. Although the sensitivity of a bridge can be manipulated by the designer, the sensitivity of a permalloy bridge is typically 300 $\mu V/V/G$. That means a bridge with a power supply of 10V exposed to a field of 10 mG (0.01G) would generate an offset voltage of 30 $\mu V$. For our most recent designs we have achieved bridge sensitivities as high as 3000 $\mu V/V/G$.
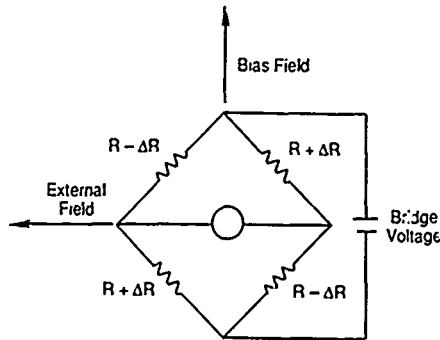


Figure 3. The Wheatstone Bridge

Figure 4 shows the three basic methods for biasing the MR sensors. For the first two biasing methods, the four legs of the bridge are laid out in herringbone patterns. Each resistor is a single trace that continuously doubles back to form a close packaging of the long, thin strip. Opposite legs of the bridge are oriented at 90 degrees to each other. The first bias method is to wrap a solenoid around the package containing the
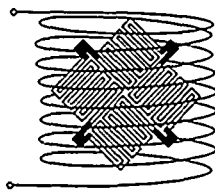


a. EM Coil
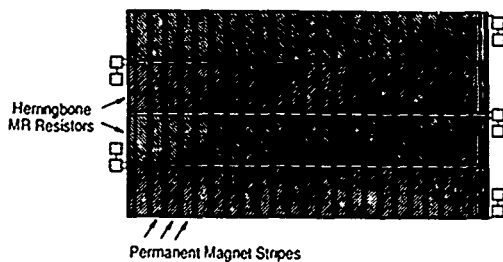


b. Permanent Magnet
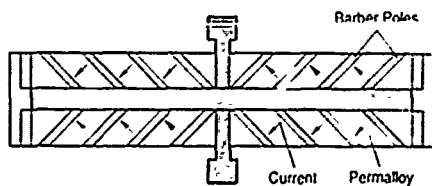


c. Barber Pole

Figure 4. Basic Methods for Biasing Magnetoresistive Sensors

magnetoresistive films. Current in the solenoid gives rise to a fairly uniform magnetic field along its axis. For some applications, however, it may be a disadvantage that power is needed to maintain the bias field. A second method is to deposit thin-film permanent magnets on top of the magnetoresistive layer. These thin films, which are made of materials such as cobalt or alloys of cobalt, can be magnetized after they are deposited by the application of a large magnetic field. This biasing scheme requires no power, but the bias magnets tend to be susceptible to temperature variations.

Rather than rotating the magnetization away from the current, the films can be biased through a third method by rotating the current away from the magnetization. This can be done by depositing angled strips of a highly conductive material such as gold on the magnetoresistive films. The current takes the shortest path through the magnetoresistive film between the conducting strips. If the angle between the magnetization and the current is suitably defined, the film operates in a linear region of the magnetoresistance curve. This method, called barber-pole biasing after the pattern the strips form, is quite elegant but has the drawback that only about half of the magnetometer's area is used for sensing.

## HIGH-SENSITIVITY SENSOR DESIGN

The high-sensitivity magnetic sensor uses the method of barber pole biasing so that no external magnetic bias is necessary. The sensitivity of the transducer is about 3000 $\mu V/V/G$. Two sensors using the MR transducer, each with a different readout technique, were designed and fabricated (see Figure 5). One uses a dc voltage applied to the bridge (see Figure 6) so that the differential voltage that develops is due to the magnetic field applied to bridge. The second magnetometer uses an ac voltage at frequency $f_m$ applied to the bridge (see Figure 7) so that the differential voltage must be demodulated to extract the output signal. (To avoid confusion with the magnetic bias, we refer to the electrical bridge bias as excitation.) The advantages of the dc-excited concept are simplicity and lower power consumption, since less circuitry is needed. The advantages of the ac-excited method are lower dc drift and potentially lower noise operation since the signal is now amplified in a frequency region where 1/f noise is at a lower level.

The sensors operate in a magnetically closed-loop configuration, so that the transducer output is nulled (bridge balanced) to the zero-output condition. This has the advantage of linearizing the output, therefore allowing the scale factor and dynamic range of the sensor to be tailored to the application. It also allows the frequency response to be designed to meet the application and results in a sensor with performance that is more consistent over the range of environmental changes. Figure 8 shows a block diagram of the magnetic feedback concept and another method of balancing the bridge; that is, by adding resistance to one element of the bridge. This latter method is not as desirable since it diminishes the advantage of using the differential output of the bridge.

Additional sensor capability was pursued by the inclusion of permeable material into the region around the MR bridge. This permeable material acts to concentrate the flux around the MR sensor and to increase the flux density within the MR transducer.

A series of finite element modeling (FEM) analyses were performed to determine the potential increase in flux density due to the inclusion of the permeable material. Initial results of the

FEM study showed that the shape-dependent demagnetization fields within the materials studied limited the effective permeability of the flux concentrators. Therefore, materials studied for the flux concentrators were selected to offer minimum remnant fields as opposed to maximum permeability.

The flux concentrators are positioned within the feedback coil (see Figure 8), and therefore, they magnify both the sensed field and the feedback field. Because the scale factor represents the feedback current that is needed to cancel a given applied field, it has the same value when the flux concentrators are present as when they are not. However, because the flux concentrators effectively increase the sensitivity of the chip, the magnetic gain of the circuit is increased. This allows the electrical gain to be decreased by the same amount, which decreases the noise level. The noise measurements presented elsewhere in this paper imply that the flux concentrators magnify the field by about a factor of 3 for the dc-excited sensor and by about a factor of 8 for the ac-excited sensor. The flux concentrators are more effective for the ac-excited than the dc-excited magnetometer because a smaller chip size is used for the ac-excited sensor, which allows for a smaller air gap between the two pieces of flux concentrators.

be varied from roughly 200 to 900 $\mu$W by controlling the current to the operation amplifiers.

Testing of the sensors demonstrated a linear output versus input field with a scale factor of about 10 V/G. The linearity measured better than 1% with a fifth-order polynomial fit. Figure 9 shows the output versus input field.

Sensor noise was characterized versus total sensor power (for the dc-excited sensor) and shown to be very close to the theoretical thermal noise limited sensitivity. Figure 10 shows the noise density values versus sensor power. Minimum noise was measured at 6 $\mu$G/$\sqrt{Hz}$ achieved at a sensor power of about 800 $\mu$W without flux concentrators, and 2 $\mu$G/$\sqrt{Hz}$ achieved at a sensor power of about 800 $\mu$W with flux concentrators. Figure 11 shows the noise output of the dc-excited magnetometer versus frequency.

Figure 8. Block Diagram of the dc-Excited Magnetoresistive Sensor

Figure 5. Two Completed Sensor Units
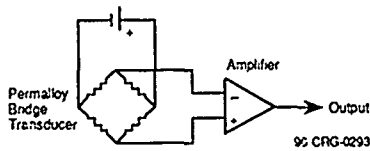
Figure 6. dc-Excited Permalloy Bridge Concept
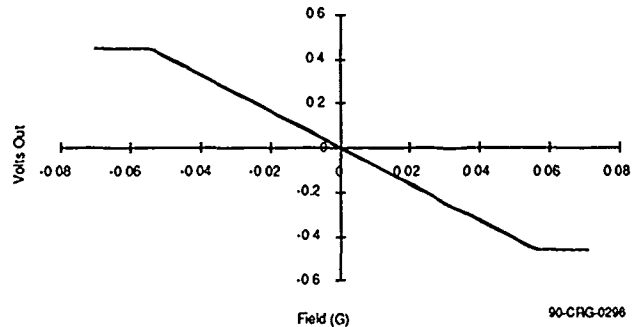
Figure 9. ac-Excited Sensor Transfer Function with Flux Concentrators

Figure 7. ac-Excited Permalloy Bridge Concept

## SENSOR TEST RESULTS

The sensors were designed with low-power electronics with a response bandwidth of dc to 10 Hz. The dc-excited sensor was designed so that the bridge bias voltage could be changed from $\pm$1.5 to $\pm$3.0V, and the electronics power consumption could also

Figure 10. Measured Sensor Noise for the dc-Excited Magnetoresistive Sensor vs. Sensor Power (without flux concentrators)

116
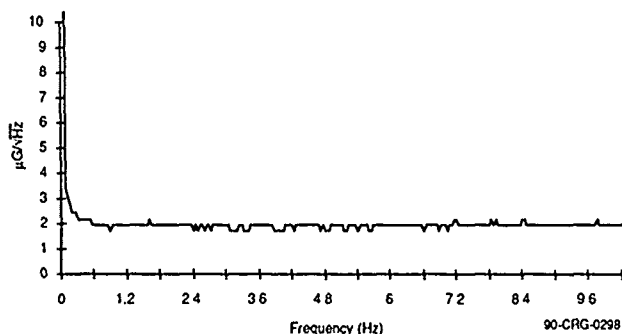
**Figure 11. dc-Excited Sensor Noise Density vs. Frequency (with flux concentrators)**

## APPLICATIONS

The net result of this sensor development has been a more than 5000 times sensitivity improvement over our previously reported results[4] and at the same time a reduction in the sensor power by a factor of more than 1000. With these improvements, the MR sensor meets two critical factors (high sensitivity with minimal power) for magnetic anomaly detection (MAD) applications.
Magnetic anomaly detection involves the detection and localization of magnetic disturbances in Earth's magnetic field. Two examples are detecting automobiles at traffic lights and locating submarines in the ocean. Figure 12 shows this effect. These distu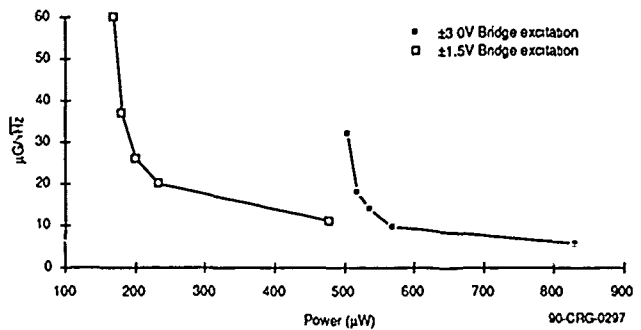rbances can be modeled in the far field as magnetic dipoles and thus the levels of the disturbance decays as the cube of the distance between the sensor and the disturbance.

The key parameter for detecting a disturbance is calculating the signal-to-noise ratio. The signal can be estimated based on the effective dipole moment representative of the ferromagnetic material in the disturbance and the geometry involving the dipole orientation in Earth's field with respect to the sensor location. The noise is comprised of many factors but for most MAD applications is dominated by the sensor noise and geomagnetic noise.

We have achieved a sensor noise of 2 $\mu G/\sqrt{Hz}$ with our MR sensor. With a bandpass covering 0.1 to 1 Hz, this sensor should be able to detect cars passing 100 ft away at speeds between 5 and 50 mph with a signal-to-noise ratio exceeding three.
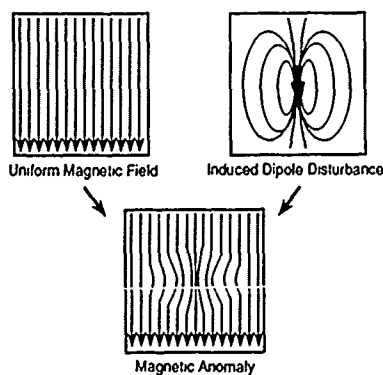


**Figure 12. Magnetic Anomaly Superposition**

Background noise must also be considered, of which a major component at low frequencies is due to charged particle motion in the ionosphere and magnetosphere. We derive the power spectrum

shown in Figure 13 for the horizontal component of this geomagnetic noise, based on measurements made in our laboratory over evenings and weekends. This noise has a roughly 1/f character, and is comparable to the sensor noise for a 0.1 to 1 Hz bandpass.

An approach for reducing geomagnetic noise is to use multiple sensors and signal processing to perform a common mode subtraction. The assumption built into this approach is that all these noise sources are coherent over the baseline of the sensors but the magnetic anomaly is not. This tends to push many MAD applications toward using several low-cost sensors in place of a single high-sensitivity sensor. Also, the ability to produce miniature sensors makes it more practical to get the sensors close to the disturbance, which greatly improves the $1/r^3$ range problem.

Another important consideration for MAD is often long-term remote operation. A three-axis MR magnetometer will consume roughly 2.5 mW of power. This sensor could operate continuously for more than one week from two AA cell Ni-Cd batteries. By duty cycling the sensor, because of its fast response time and the usual low frequency band of interest, the three-axis sensor could operate from these batteries for more than one year.
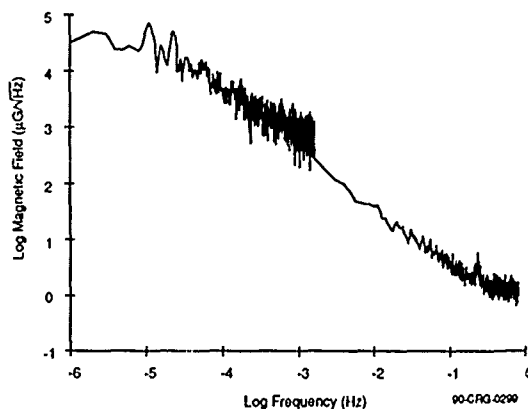


**Figure 13. Geometric Noise Power Spectrum (horizontal component)**

In summary, the improvements in increased sensitivity and lower power of this MR sensor technology open new possibilities for magnetic anomaly detection applications. The major issue now is to develop a system that maximizes the capability of the sensor and minimizes compromises due to background noise.

## REFERENCES

1. "A Review of Magnetic Sensors," J. Lenz, to be published in Proceedings of the IEEE, June 1990.

2. "Scaling in Thin Magnetoresistive Film," B. Pant, J. Applied Physics, Volume 67, No. 1, January 1990.

3. "Giant Magnetoresistance of 001 Fe/001 Cr Magnetic Superlattices," M.N. Baibich, J.M. Broto, A. Fert, F. Nguyen Van Dau, and F. Petroff, Physical Review Letters, Vol. 61, No. 21, November 1988.

4. "Comparison of Five Readout Techniques for an Integrated Magnetometer, J. Lenz, K. Kawai, R. Fryer, L. Strandjord, Solid State Sensor Workshop, Hilton Head, South Carolina, June 1986.

# DEEP X-RAY AND UV LITHOGRAPHIES FOR MICROMECHANICS

H. Guckel, T.R. Christenson, K.J. Skrobis, D.D. Denton, B. Choi,
E.G. Lovell, J.W. Lee, S.S. Bajikar and T.W. Chapman

College of Engineering
University of Wisconsin-Madison
1415 Johnson Drive
Madison, Wisconsin 53706

## ABSTRACT

Micromechanical devices are three-dimensional structures. They therefore require processing sequences which differ from standard IC-processing which is essentially two-dimensional. The use of very thick photoresist layers is but one method to achieve this. However, when thick photoresist processes are combined with electroplated metal technology, major advances in micromechanics can occur because the available material base expands greatly.

The photoresist for our work is poly(methyl methacrylate) or PMMA. It is sensitive to x-ray photons and deep ultraviolet radiation. It can be used for spin coating or can be applied by casting and in situ polymerization. This implies that there are at least four possible versions of the deep lithography process. All of them hinge on the availability of a developer with very high selectivity and low stress formation during liquid processing.

The work which is reported here deals first of all with a bi-level photoresist system which uses optical processing only. This process, DUVL, results in PMMA definitions with excellent resolution for photoresist thicknesses of up to 10 μm. It can be extended to 30 μm or so by a simple process change. The second approach uses thicker, in our case larger than 40 μm, photoresist layers. These films involve cross-linked PMMA and are exposed by synchrotron radiation. Both films are developed in an excellent wet developing system which was formulated by a German research group. The results show good pattern definition.

The exposed and developed substrates are electroplated with either gold or nickel films. Gold is used to produce x-ray masks on polysilicon mask blanks. The patterns are optically defined and produce good results above 0.6 μm. Synchrotron-exposed samples are plated with nickel. They produce LIGA-like structures with good results.

## INTRODUCTION

X-ray lithography has two major application areas: submicron VLSI and micromechanics. The XRL technology difference for these two fields arises primarily from the dramatically different photoresist thicknesses which are used. Thus, XRL for VLSI employs spun-on layers in the one micrometer range. XRL for micromechanics, on the other hand, becomes interesting at photoresist thicknesses of more than ten micrometers and may employ thicknesses as large as several hundred micrometers. The term. deep x-ray lithography or DXRL is therefore richly deserved.

The photoresist thickness issue causes a second major difference for the two application areas. VLSI x-ray lithography can be implemented with rather modest x-ray flux densities and is therefore practical with traditional point and plasma sources. DXRL requires the high brightness which is currently only available from a synchrotron source. This is somewhat unfortunate because synchrotron access is required for process research in this potentially important field. This difficulty is somewhat reduced by W. Ehrfeld's work in a process which he calls LIGA [1]. This procedure, which is the earliest example of DXRL, utilizes synchrotron exposure only once to produce a master pattern. It then copies the master by injection molding and therefore avoids further synchrotron use for DXRL device production. A somewhat different approach for modest photoresist

thicknesses has been used at Wisconsin. It is based on the fact that the most important photoresist for DXRL is poly(methyl methacrylate) or PMMA. This resist can be exposed with a deep UV source at 2300 Å where it has an absorption length of roughly 4 μm. Thus, if deep lithography or DL is defined to start at 10 μm or so, many aspects of DXRL can be studied and useful devices can be derived from ultraviolet deep lithography or DUVL which of course will not involve synchrotron access. The implications are then as follows: DUVL is accessible to many research groups. It is also a tool which leads to much more rapid process development for DXRL.

Micromechanics based on DXRL involves four major areas. The source, as stated earlier, is a synchrotron with a suitable beam line and fixturing for the mask and substrate. The photoresist process is the second area. Since very large photoresist thicknesses are involved, photoresist application via spinning may not be feasible. Furthermore, since developing of thick layers may involve long immersion times, the selectivity of the developer for exposed and unexposed photoresist films must be nearly infinite. Swelling and distortions must be avoided if mechanical damage and geometric distortions are not allowed. The third topic involves a suitable mask for DXRL. The mask blank should not absorb any photons. This requires a low atomic number membrane with a thickness in the micrometer range as a compromise solution. The absorber is formed by patterned high atomic number material, i.e., gold or tungsten. The desired contrast ratio sets the absorber thickness for a given mask blank. For very thick photoresist and therefore long exposures, absorber thicknesses of several micrometers are required. This normally implies that the absorber must be electroplated. Bath compatibility with the photoresist system, built-in strain and deposit uniformity are some but not all of the difficulties which must be addressed.

The previous discussion indicates that process development for DXRL involves a "Catch-22" condition: a mask is needed for DXRL. Its construction requires DXRL. This dilemma must therefore be resolved if a viable research program in DXRL is to proceed.

## MASKS FOR DEEP X-RAY LITHOGRAPHY

Mask blanks for x-ray lithography as stated earlier are fully supported membranes of low atomic number materials. In our case a suitable mask blank for VLSI x-ray lithography had already been developed by using silicon nitride and tensile polysilicon films before work on DX₁ was initiated [2,3,4]. Because DXRL exposures are long and radiation damage was of concern, a DXRL mask blank made from one micrometer thick tensile polysilicon was selected as the blank of choice. A diaphragm size of 2.5 cm × 0.8 cm acknowledges the fact that synchrotron beam scanning via x-ray mirrors which is used in VLSI x-ray lithography would not be used in order to reduce exposure time. Blanks of this type are now being produced quite routinely in a high yield process.

The absorber for a DXRL mask was chosen to be electroplated gold. The argument for this is found in the fact that absorber thicknesses for the polysilicon mask blank become quite large for good contrast ratios if long exposure times are anticipated. The plating occurs into photoresist recesses with high edge acuity on a properly prepared mask blank. The mask blank preparation involves a suitable

plating base which is currently produced by sputtering 50 Å of chromium followed by 150 Å of nickel.

The photoresist for mask fabrication has to be at least 7.5 μm thick to accommodate absorber thicknesses of up to 4 μm. The requirement on edge acuity is very high because any non-vertical flank on the plated absorber is in effect a contrast change and therefore causes geometric drift between the mask and the work piece. At this photoresist thickness the acuity issue can be satisfied by using PMMA as the photoresist of choice. This material has the capability to produce essentially vertical flanks as demonstrated by the LIGA process. However, in the present situation, x-ray exposure is not feasible because this would require a mask which does not exist. Either E-beam or optical patterning is needed. Since one of us, H.G., had worked on a high resolution bi-level optical approach, this procedure was adopted and led to the concept of an optically defined x-ray mask.

The actual process, the DUVL process, involves first of all a substrate with a plating base. The substrate material is normally silicon. However, there are many other possibilities because the maximum process temperature is only 180°C. The PMMA photoresist layer is formed from 496,000 M.W. PMMA in a 9% by weight chlorobenzene solvent. This material can be bought from KTI. The optical characteristics of the material can be improved by adding a dye which in the present case is Kodak Laser Grade Coumarin 6. The mixture is 50mg of dye per 25cc of PMMA. The material is spun at 2000 RPM for 45 seconds. It is then annealed very carefully in a precision bake-out oven by ramping to 180°C at 1°C/minute, holding at 180°C for 1 hour and ramping to room temperature at -1°C/minute. This results in a PMMA thickness of 1.6 μm. The process is repeated until the desired photoresist thickness is obtained. The prepared substrate is next covered with a compatible optical photoresist. This material must have the following properties: it must be a good optical receiver for mercury arc exposure, it must be opaque at deep UV wavelengths and it must not cause interface problems. The commercial KTI 809 system satisfies all of these requirements. The material is spun at 5000 RPM for 30 seconds. It is then pre-baked at 90°C for 15 minutes. Exposure in hard contact on a Canon 501 follows. Developing in a mixture of 1:1 KTI 809 developer and DI-water produces easy pattern transfer at 1 μm line width and optimized pattern transfer at 0.6 μm. Blanket exposure at 2300 Å at 15 mw/cm² for 31 minutes in an HTG system at 8 μm PMMA thickness completes the photoresist processing phase.

The next processing step is that of developing the PMMA. It was stated earlier that this field owes a lot to Dr. W. Ehrfeld. His progress in turn is based very much on a nearly perfect developing system [5]. The developer is

| | |
|---|---|
| 60 vol% | 2-(2-butoxyethoxy)ethanol |
| 20 vol% | tetrahydro-1, 4-oxazin (morpholine) |
| 5 vol% | 2-aminoethanol |
| 15 vol% | water |

typically used at 35°C in a recirculating spray developing system. The selectivity of the developer is nearly perfect and stress build-up due to the developing process is extremely low. The importance of this system cannot be overstated: without this type of performance, there would be no DXRL.

The developing cycle is followed by a 20 minute rinse at 35°C. The rinse consists of 80% 2-(2-butoxyethoxy)ethanol and 20% water. A second rinse for 5 minutes in DI-water completes the process. Figure 1 illustrates typical results.

Plating follows a mild oxygen plasma de-scum cycle. Typical non-optimized plating conditions for gold are 50 mA/cm² at 60°C with a plating rate of 0.3 μm/min. The plating composition is a commercially available solution which is manufactured by Technic Inc.

Photoresist removal is accomplished by blanket UV exposure and developing. Figure 2 illustrates the mask result.

The processing sequence which has been discussed here is stable and can be extended to thicker PMMA layers. Thus,



Fig. 1  Partially exposed PMMA test pattern. The pattern has been developed throughout the exposed PMMA. The pattern height is 8 μm.
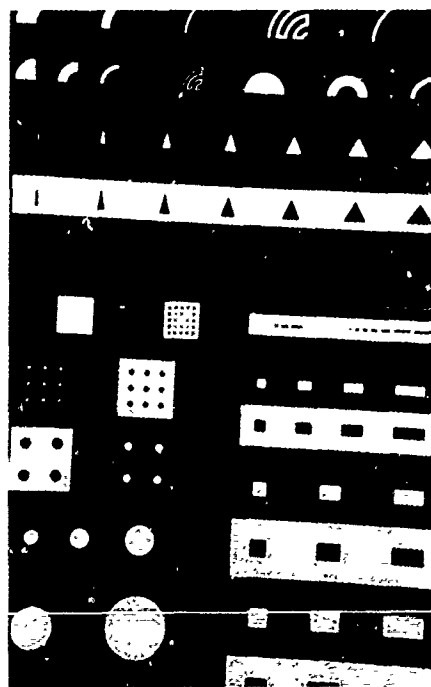


Fig. 2  Polysilicon x-ray mask. The figure contains an optical micrograph of a backlit x-ray mask. The light areas are x-ray transmissive; the dark areas are covered by 3.5 μm of plated gold. The polysilicon thickness is 1 μm.

119

spun-on films of 30 μm thickness have already been produced. They can be exposed with multiple deep UV exposures and can be processed with several developing cycles if a slightly different optical transfer system is used. This is necessary because the PMMA developer removes the KTI 809. The cycles are needed because the absorption length of the PMMA at 2300 Å is roughly 4 μm and therefore exposure to 8 μm is a practical limit at UV for a single cycle. It is therefore possible to claim that DUVL for micromechanics is readily implemented in-house for most research groups.

## THICK PHOTORESIST TECHNOLOGY FOR X-RAY EXPOSURE

It is possible to argue that DXRL can become significant with spun-on films only. However, there are many advantages for cast films. Perhaps the two most important issues are planarization and flexibility in polymer composition and, in particular, the fact that cross-linked films can be produced. The difficulty in this area is simply the fact that this type of processing has not been studied in detail and that therefore a learning curve is required. The learning cycle has been shortened considerably by the publications of the German LIGA group [6,7,8].

The procedures which have been developed here are based on a casting syrup which is made from pre-polymerized PMMA with a number-averaged molecular weight of 100,000. It is dissolved in MMA and a cross-linking agent, ethylene glycol dimethacrylate (EGDA) is added. Two initiators are used. Thus, solution I contains dimethyl aniline or DMA and solution II contains benzoyl peroxide or BPO. The two solutions are mixed at room temperature which causes free radical formation via a reaction between the two initiators. Polymerization starts and proceeds at a rapid pace. Since the two solutions have to be well mixed and degassed, special equipment was designed and constructed to do this.

Typically, 1 to 2 ml of the syrup are applied to a casting fixture which supports the wafer. The function of this device is that of forcing the syrup into a uniform film. This implies flat plane contact to the upper surface of the film which is accomplished by using a sheet of Mylar as a release film. Mechanical spacers and film shrinkage set the final film thickness. The cast films are next annealed at 110°C. Precision positive ramping at 20°C/hour and negative ramping at 5°C/hour together with a 1 hour hold at 110°C complete the process.

## X-RAY EXPOSURE

PMMA exposure via synchrotron radiation is not a simple issue. The reason for this is, in part, the spectrum of the source: it is broad band as demonstrated in Figure 3. The second problem is found in the absorption length of PMMA: it is very wavelength-sensitive as indicated by Figure 4. Direct exposure with this type of source and absorber will evidently be nonuniform in the direction normal to the substrate plane. This nonuniformity typically involves overexposure of the PMMA surface and underexposure at the substrate/PMMA interface. One can estimate proper exposing conditions by using data from reference [9]. The threshold of exposure for the developing system which has been mentioned here is 1.6 kJ/cm³. Reported resist damage occurs at a dose of 20 kJ/cm³. These numbers are based on the Bonn synchrotron which operates at higher energy and much lower beam current. ALADDIN, the synchrotron at Madison, is a different machine and test exposures indicate that the upper exposure limit may be quite a bit higher. There are also some indications that exposure levels on a higher current machine are rate dependent. In any case, since both an upper and lower exposure limit exist, one has no choice but to change the spectral output of the machine. This can be done quite easily by inserting a filter between the incoming beam and the mask-substrate combination. Figure 5 shows the result. Figure 6 shows the corrected spectrum after a 50 μm Be filter and the 1 μm silicon mask blank.

Initial exposures with the polysilicon mask were accomplished by using the beam line ES-O which was constructed by the Wisconsin Center for X-Ray Lithography
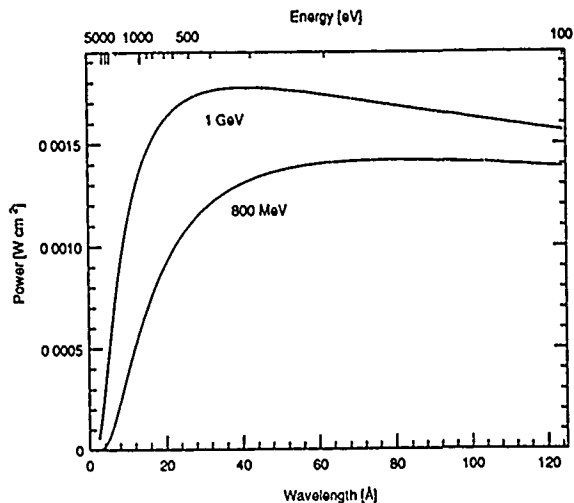


Fig. 3 Spectrum of ALADDIN at 800 MeV and 1 GeV, with a beam current of 100 mA and horizontal acceptance of 4.858 mrad.
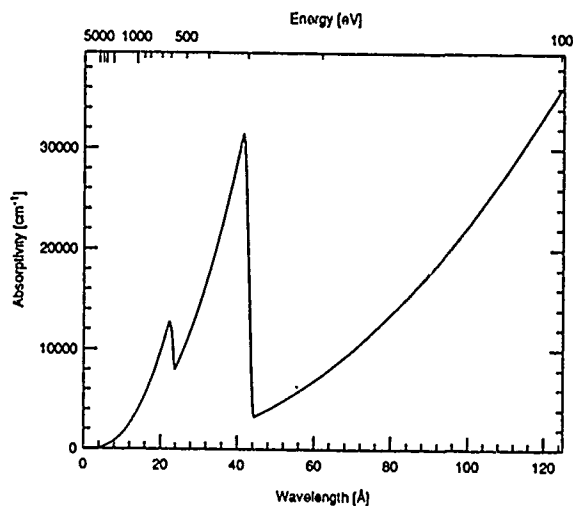


Fig. 4 Absorptivity vs. wavelength of poly(methyl methacrylate).
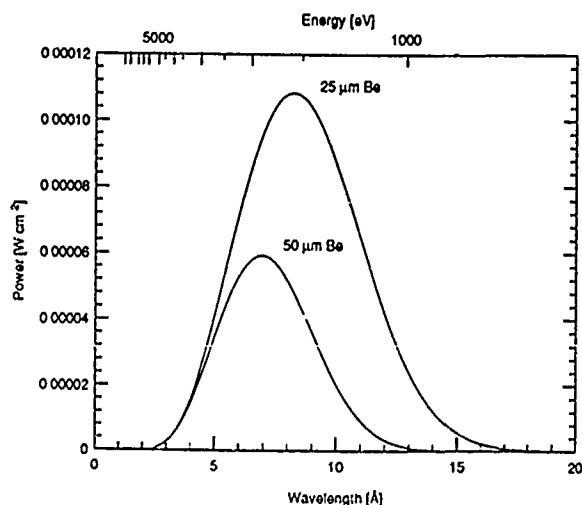


Fig 5 Spectrum of ALADDIN modified by beryllium filters, at 800 MeV, with a beam current of 100 mA and horizontal acceptance of 4.858 mrads.

120

group. Maximum PMMA thickness was 40 μm and exposures were done by mechanically scanning the wafer-mask combination at 2.7 scans/min. with a 2 cm scan amplitude. Mask to wafer separation was 50 μm. The wafer-mask assembly was exposed to 20 torr of helium for cooling purposes. Typical exposure levels were 50 J/cm$^2$ with exposure times always less than 1 hour.
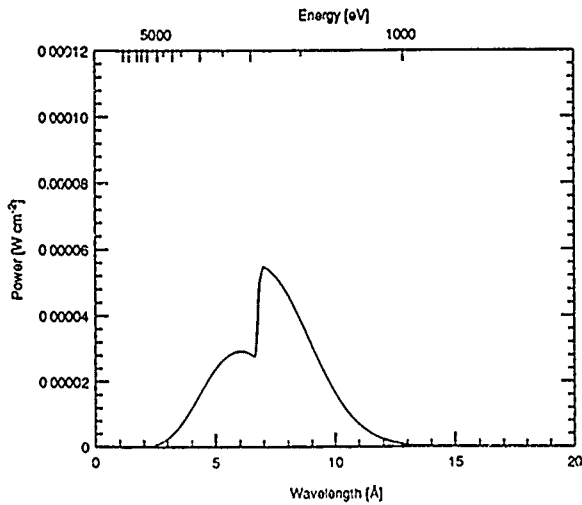


Fig. 6 Spectrum of ALADDIN after a 50 μm beryllium and a 1μm polysilicon mask at 800 MeV, with a beam current of 100 mA and horizontal acceptance of 4.858 mrads.
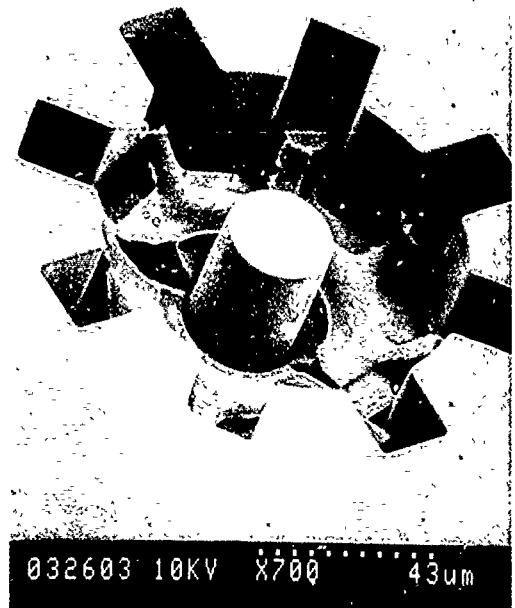


Fig. 8 X-ray exposed cast PMMA test pattern. The PMMA thickness in Fig. 8 is 40 μm. The x-ray exposure was too short for complete developing.



Fig. 7 X-ray exposed PMMA test pattern. This pattern v produced from an 18 μm thick spin-coated PMMA f  . and x-ray exposure. It was then fully developed.



Fig. 9 Nickel plated LIGA-like test structure. Fig. 9 illustrates plating resolution. The structures are nickel columns of 5 μm diameter and 18 μm height.

## DXRL RESULTS

The x-ray exposed wafers were developed as indicated earlier. Figure 7 indicates the resist performance. Particularly noteworthy are the smoothness of the unexposed sections and the edge acuity. The absence of crazing is recognized. Figure 8 involves a cast PMMA sample for which the exposure was too short. It demonstrates the selectivity of the developer and of course the edge acuity for a cross linked film.

Some of the samples were electroplated. For this purpose a nickel sulfamate plating system was constructed. The bath is operated at 50°C at a pH of 4.3 and uses a plating current density of 50 mA/cm². It resulted in the structures which are shown in Figure 9.

## CONCLUSIONS

The progress which has been reported here indicates that DUVL is a viable tool for micromechanics which is accessible without synchrotron access. This process can be used with many metals which can be plated and therefore extends the micromechanics material base considerably. This is good even though it is an admission that silicon does have limitations which one of us, H.G., finds somewhat disturbing.

The Wisconsin x-ray lithography effort has passed the stage of relying on LIGA information. It is also past the feasibility stage. It will be applied to new and interesting structures in the near future.

## ACKNOWLEDGEMENTS

## REFERENCES

(1) W. Ehrfeld, F. Götze, D. Münchmeyer, W. Schelb and D. Schmidt, "LIGA Process: Sensor Construction Techniques via X-Ray Lithography," Technical Digest, IEEE Solid-State Sensor and Actuator Workshop, pp. 1-4, 1988.

Reference (1) was an invited presentation at the 1988 Sensor and Actuator Workshop. The LIGA literature is quite extensive and cannot be fully cited here due to space limitations. The interested reader should note that a full list of LIGA references is available from the authors. Other key LIGA references are cited below.

(2) C.C.G. Visser, J.E. Uglow, D.W. Burns, G. Wells, R. Redaelli, F. Cerrina and H. Guckel, "A New Silicon Nitride Mask Technology for Synchrotron Radiation X-Ray Lithography: First Results," Microelectronic Engineering, Vol. 6, pp. 299-304, 1987.

(3) H. Guckel, D.W. Burns, T.R. Christenson and H.A.C. Tilmans, "Polysilicon X-Ray Masks," Microelectronic Engineering, Vol. 9, pp. 159-161, 1989.

(4) H. Guckel and D. Burns, "Polysilicon Thin Film Process," U.S. Patent No. 4,897,360, Jan. 30, 1990.

(5) V. Ghia and W. Glashauser, "Verfahren für die Spannungsfreie Entwicklung von bestrahlten Polymethylmethacrylat-Schichten," Offenlegungsschrift DE 3039110, Siemens AG, Munich.

(6) J. Mohr, "Analyse der Defektursachen und der Genauigkeit der Strukturübertragung bei der Röntgentiefenlithographie mit Synchrotronstrahlung," Diplomarbeit, Universität Karlsruhe, F.R.G., 1986.

(7) A. Stutz, "Untersuchungen zum Entwicklungsverhalten eines Röntgenresists aus vernetztem Polymethylmethacrylat" Diplomarbeit, Universität Karlsruhe, F.R.G., 1987.

(8) S. Harsch, "Untersuchungen zur Herstellung von Mikrostrukturen großer Strukturhöhe durch Galvanoformung in Nickelsulfamatelektrolyten," Diplomarbeit, Universität Karlsruhe, F.R.G., 1987.

(9) D. Münchmeyer, "Untersuchungen zur Abbildungsgenauigkeit der Röntgentiefenlithografie mit Synchrotronstrahlung bei der Herstellung technischer Trenndüsenelemente, "Diplomarbeit, Universität Karlsruhe, F.R.G., 1984.

# A Pressure-Balanced Electrostatically-Actuated Microvalve

Michael A. Huff, Michael S. Mettner*, Theresa A. Lober, and Martin A. Schmidt
Microsystems Technology Laboratories
Massachusetts Institute of Technology
Cambridge, MA 02139

*Robert Bosch Company
Stuttgart, Federal Republic of Germany

## Abstract

We report a new microvalve structure which is designed to enhance the limited actuation forces available in microfabricated devices by using a pressure-balancing scheme. The concept is to allow the fluid to provide a balancing force on the moving part of the device thereby reducing the force required to open the valve. Although various methods may be used to actuate the valve, we have chosen electrostatic actuation since it is readily integrated with the valve fabrication sequence. This paper will discuss the design and fabrication of this microvalve. Flow testing of the valve has not been completed and will not be presented in this paper. The process for implementing the valve concept uses multiple wafer bonding steps (three in the present prototype), and has yielded valves which have been successfully actuated in air using voltages below 350 Volts. While developing a process for the microvalve we investigated two important facts related to bonded wafers with sealed cavities. First, by bonding two silicon wafers together, one of which has a cavity etched into it and electrochemically etching back one of the wafers to a very thin layer, we have been able to measure the residual gas pressure in the cavity. Using the theory of large deflections of circular plates, we have determined that the residual pressure is approximately 0.8 atms. Second, we have found that high temperature processing of bonded silicon wafers which have a sealed cavity between the bonded layers results in a permanent set in the top layer of lightly doped silicon due to the expansion of the trapped gases within the cavity. We have determined the onset of plastic deformation to be within the temperature range of 800-850 C.

## Introduction

One of the first reported microfabricated valves was designed as an injection valve for use in a integrated gas chromatograph in 1978 [1]. This valve consisted of a silicon micromachined valve seat and a nickel diaphragm which was actuated by an external solenoid. Much of the more recent work is similar to this first design except the nickel diaphragm has been replaced by a silicon micromachined diaphragm[2]. In addition to these externally actuated valve designs, there has been a number of one-way check valves[3,4], a thermally actuated flow-channel type valve capable of controlling large fluid pressures[5], and an integrated electrostatically actuated valve fabricated by surface micromachining [6].

The use of microfabrication is attractive for valves because of small dead volumes, potentially high speed, and low-cost due to batch fabrication. However, the creation of an integrated microvalve is made difficult by the inability to generate enough mechanical force to actuate the valve using conventional actuation schemes.

The basic concept of the pressure-balanced valve, illustrated in Figure 1, is to have the pressure of the fluid provide a balancing force on the moving part of the valve. The fluid pressure produces an upward directed force which tends to close the valve and simultaneously a downward directed force tending to open the valve. Consequently, the total force required to actuate the valve can be designed to be only a small fraction of the total pressure force of the fluid by properly sizing the area of the top and bottom surfaces of the valve. Any number of methods may be used to actuate the valve including thermal, pneumatic, magnetic, and electrostatic. We have chosen to use electrostatic actuation in the prototype design due to the ease with which this type of actuation can be integrated into the fabrication sequence. The valve is actuated by an electrostatic force created by an applied voltage between the lower and upper conducting layers which are separated by an insulating layer. Even though it is typically only possible to electrostatically generate several psi of pressure force, the pressure-balancing design makes it possible to control fluids at much higher pressures. The limit of the



**VALVE CLOSED**



**VALVE OPEN**
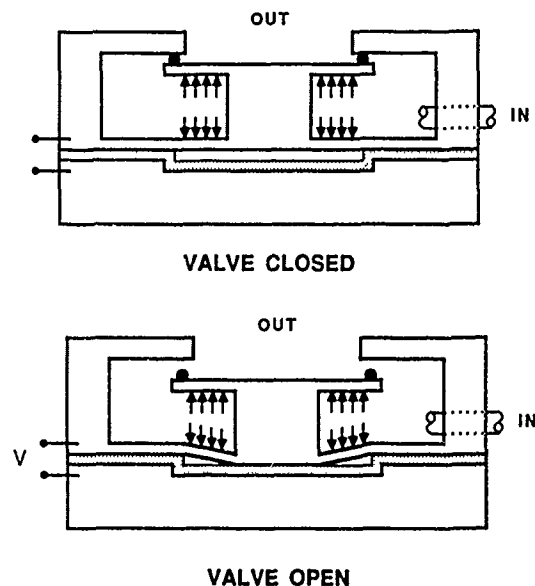
**Figure 1.** Valve Concept

balancing will be determined by the ability to match and control the important surface areas of the valve.

The valve is designed to be used in a "bi-stable" mode, in which the structure is either fully open or closed. In this case, what is of most interest in the operation of the valve is the voltage required to cause electrostatic "pull-in".

## Design

### Fluidic Behavior

Figure 2 shows the typical dimensions for a completed prototype microvalve. We can estimate the fluid flow rate through the valve by modelling the structure as an orifice. In laminar flow, the flow rate through an orifice is given by [7]:

$$(1) \qquad Q = \frac{2\delta^2 D_h A_o}{\mu} (P_{IN} - P_{OUT}),$$

where $\delta$ is the discharge coefficient. $D_h$ is the hydraulic diameter, $A_o$ is the area of the orifice opening, $\mu$ is the viscosity of the fluid, and $P_{IN}$ and $P_{OUT}$ are the pressures on either side of the orifice. For a sharp-edged slit orifice $\delta = 0.157$. Substituting into equation (1) values for, the hydraulic diameter $D = 2w$ (w is the deflection of the structure into the cavity), the orifice area $A_o = 2\pi cw$, $\Delta P = 70$ bar, and assuming the fluid is silicon oil where $\mu = 100$ mPa sec, the flow rate is estimated to be 78 ml/min.

To ensure that the flow is laminar through the orifice, the Reynold's number must be calculated and compared to the transition Reynold's number $R_t$. If $Re < R_t$ then the flow is laminar. The

$$A_1 = 2\pi a^2 \qquad A_2 = 2\pi b^2 \qquad A_3 = 2\pi c^2$$
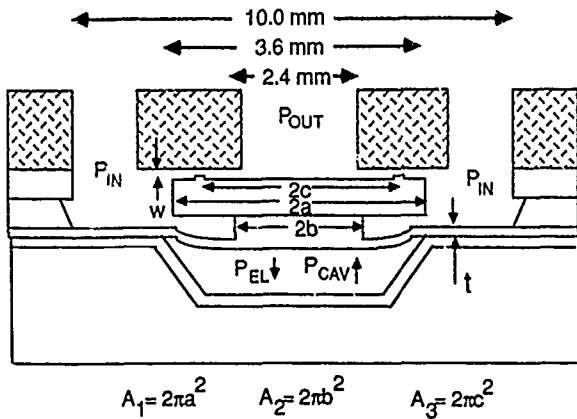
Figure 2. Valve Dimensions

Reynold's number for an orifice is defined by:

$$(2) \qquad Re = \frac{\rho (Q/A_o) D_h}{\mu},$$

where $\rho$ is the density of the fluid, $\rho = 0.8$ g/cm$^3$ for silicon oil. The transition Reynold's number is given by:

$$(3) \qquad Re_t = \left(\frac{0.611}{\delta}\right)^2,$$

and is found to be approximately 15 for a sharp-edged slit-type orifice. Substituting values for the variables into equation (2) above it is found that Re $\ll$ Re$_t$ and therefore the flow is laminar.

Mechanical Behavior

The mechanical behavior of the valve is examined by modelling the valve structure as shown in Figure 2. We assume that the circular plate is symmetrically loaded with respect to the axis perpendicular to the plate, the deflections are small in comparison with the thickness of the plate, and there are no residual stresses in the structure. The solution for the maximum deflection of this plate under the loading of a concentrated force, $Q_T$, acting on the thick center region or "plunger" and a uniform pressure force, $q_T$, acting on the thin annular region is given as follows[8]:

$$(4) \qquad w = K_1 \frac{Q_T a^2}{E t^3} + K_2 \frac{q_T a^4}{E t^3},$$

where w is the maximum deflection of the plate, a is the outer radius of the circular plate, E is Young's modulus of silicon, t is the thickness of the plate in the thinner region, and $K_1$ and $K_2$ are constants given in tabular form for various ratios of the outer and inner diameters. This equation assumes poisson's ratio is equal to 0.3.

Referring to Figure 2, the total concentrated force, $Q_T$, acting on the center plunger can be written as:

$$(5) \qquad Q_T = P_{EL} A_2 - P_{CAV} A_2 - P_{IN}(A_3 - A_2) + P_{OUT} A_3,$$

where $P_{IN}$ and $P_{OUT}$ are the pressures of the fluid inside the valve cavity and outside the valve, respectively, $P_{EL}$ is the electrostatic force generated by application of a voltage on the structure, and $P_{CAV}$ is the pressure inside the cavity which is dependent on the deflection of the plate into the cavity. $A_1$, $A_2$ and $A_3$ are the areas of the plunger cap, the plunger base, and the sealing ring, respectively, as shown in the Figure. The total uniform load on the plate can be written as:

$$(6) \qquad q_T = P_{EL} + P_{IN} - P_{CAV}.$$

The electrostatic force can be written as:

$$(7) \qquad P_{EL} = \frac{1}{2} \varepsilon_o \left(\frac{V}{h_{AIR} + \varepsilon_r^{-1} h_{OX}}\right)^2,$$

where $\varepsilon_o$ is the permittivity of air, $\varepsilon_r$ is the relative permittivity of silicon dioxide, V is the voltage applied, $h_{OX}$ is the thickness of the oxide layer and $h_{AIR}$ is the gap spacing between the cavity bottom and the deflecting plate. For ease of calculation, we assume that the electric pressure acting on the thin annular region is equal to the pressure acting on the plunger. This is an approximation, since in reality P depends on the electrode spacing and thus will vary over the thin annular region. The air gap spacing is related to the deflection of the plate as:

$$(8) \qquad h_{AIR} = h_o - w,$$

where $h_o$ is the initial cavity depth. The pressure inside the cavity is dependent on the deflection of the plunger into the cavity and can be written as:

$$(9) \qquad P_{CAV} = \frac{P_1 V_1}{V_1 - \Delta V},$$

where $P_1$ is the initial pressure inside the cavity, equal to 0.8 atm, and $V_1$ is the initial volume of the sealed cavity. We neglect the static deflection of the structure due to the sealing ring. The mode shape assumed for the structure as it deflects into the cavity is a frustum of a right circular cone. Therefore $\Delta V$ is given by:

$$(10) \qquad \Delta V = \frac{1}{3} \pi w (a^2 + b^2 + ab).$$

The variable a is the outer radius and b is the radius of the plunger. Combining the above equations and substituting in the values for the dimensions shown in Figure 2, a general equation relating the applied voltage to the deflection of the plate into the cavity is derived:

$$(11) \qquad w = 8.91 \times 10^{-6} + 4.25 \times 10^{-22} \left(\frac{V}{5.41 \times 10^{-6} - w}\right)^2$$
$$- \left(\frac{5.45 \times 10^{-11}}{7.11 \times 10^{-6} - w}\right),$$

where it is assumed t = 20 µm, $h_o$ = 5 µm, $h_{OX}$ = 1.6 µm, and $P_{IN}$ = $P_{OUT}$ = atmospheric pressure. Using equation 11, it is found that the structure has a "pull-in" voltage of 328.4 Volts.

Fabrication

The process for implementing the valve concept uses three wafer bonding steps and does not yet have the sealing ring as shown in Figure 2. The fabrication sequence (Figure 3) of the valve begins with a n-type <100> 0.5-2.0 ohm-cm 4-inch silicon wafer. The wafer is placed in a phosphorus diffusion furnace at 925 C for 1.5 hours in order to highly dope the surface. This step is done to ensure that good electrical contact can be made in order to actuate the valve. After a one hour drive-in at 950 C, the wafer is stripped of the phosphorous-doped glass and a 1000 Å thermal oxide is grown. After the masking oxide is patterned the wafer is placed in 20% KOH at 56 C for approximately 22 minutes thereby etching circular recessed electrodes 5 µm deep and 3.6 mm in diameter, Figure 2a. The masking oxide is then stripped and a thermal oxide, 1.6 µm-thick, is grown on the wafer. This thick layer of silicon dioxide acts as the dielectric isolation during electrostatic actuation and will also serve to protect the handle wafer from the subsequent silicon etching steps.

A second 4-inch wafer, <100> p-type 10-20 ohm-cm, is thermally bonded to the front side of the first wafer, Figure 2b. Prior to bonding, the two wafers are cleaned using a standard pre-oxidation clean and then hydrated by immersion into a 3.1, sulfuric acid.hydrogen peroxide solution. After a spin rinse and dry, the polished surfaces of the wafers are physically placed into intimate contact. Using an infrared inspection system, the wafers are examined for voids. Assuming the bond is void-free, the composite two-wafer structure is placed into a dry oxidation furnace for one hour at 1000 C to complete the bond. Once removed from the

(a) Etch Trench In Silicon Substrate

(b) Grow Oxide and Bond Wafer

(c) Etch-back and Pattern Wafer #2

(d) Bond Wafer #3

(e) Pattern Wafer #3

(f) Packaged Valve
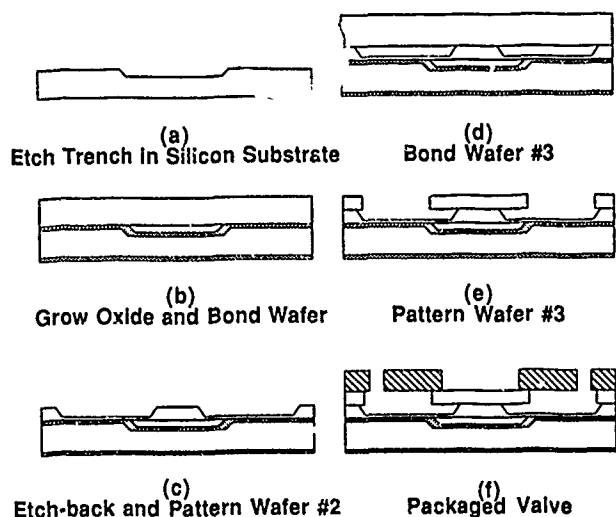
Figure 3. Valve Fabrication Sequence

furnace, the bonded wafers are inspected again using the infrared inspection system.

The bonded wafers are then placed in a 20% KOH solution at 56 C for approximately 23.5 hours and the second wafer is etched back to a resulting thickness of 75 $\mu$m. The surface of the etched back wafer is then mechanically polished to a mirror-smooth finish. The resulting thickness of the second silicon wafer is 50 $\mu$m. A masking oxide, (LTO), 5000 Å thick is deposited onto the just polished surface and subsequently patterned. The wafers is then etched using a 20 % KOH solution at 56 C for approximately 1.5 hours forming the base of the valve, Figure 2c. After the masking oxide is stripped a third 4-inch <100> silicon wafer, p-type 10-20 ohm-cm, is thermally bonded to the polished surface of the second wafer, Figure 2d. The bonding is done in exactly the same way as described above. After bonding, the composite structure is inspected using the infrared inspection system. Figure 4 is a thermal print-out of the infrared inspection system clearly indicating the plunger base bonded to the third wafer.

The third wafer is etched back in a 20% KOH solution at 56 C to a resultant thickness of 75 $\mu$m. A 5000 Å-thick oxide is deposited onto the third wafer surface using either LTO or an ACVD oxide. Subsequently, the wafer is patterned and etched in a 20% KOH solution at 56 C forming the top layer and releasing the valve, Figure 2e. The complete wafer is sawed into individual valves, which are then packaged using a capping glass plate which contains inlet and outlet ports, Figure 2f.

We have successfully fabricated the valve without the sealing ring shown in Figure 2. Figure 5 shows a top-view SEM of a valve that has been sawed in half and Figure 6 is a SEM of the valve structure displaying the quality of the bonds between the various layers.

The complete microvalve prototype, without sealing ring, was successfully actuated in air with voltages below 350 Volts. We are currently in the process of flow testing.
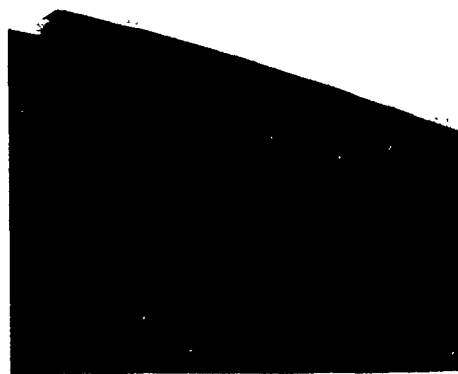


Figure 5. SEM of Valve Structure



Figure 4. Infrared Image of Valve Wafer
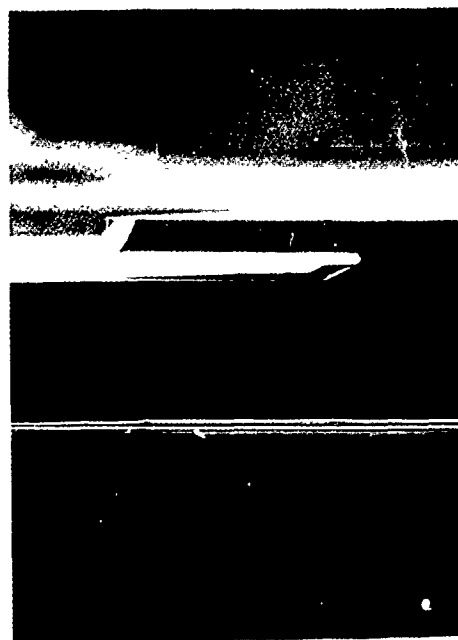


Figure 6. Close-up SEM of Valve

## Residual Pressure in Sealed Cavities

The bonding of a silicon wafer with cavities to another silicon wafer results in trapped gases within the sealed cavity. It was of interest to determine the residual pressure of this trapped gas in order to model the behavior of the valve. Starting with p-type <100> silicon wafers, we diffused phosphorous into the wafers to act as an electrochemical etch stop. These wafers where then bonded to other silicon wafers which had circular cavities of various depths ranging from 30 to 50 μm. After selectively removing the oxide from the backside of the phosphorous doped wafers, the wafers were then etched in KOH in order to thin the wafers to a resultant thickness of 50 μm. The wafers then were electrochemically etched in KOH, resulting in a smooth, $8.0 \pm 0.2$ μm-thick capping layer of silicon over the sealed cavity. The thickness of this layer was verified using an SEM. After etching, it is immediately noticed that the capping layer is deflected into the cavity, indicating that the residual pressure is below atmospheric. The deflections of the capping layer of silicon were measured with a calibrated microscope. Using the theory of large deflection of circular plates we can write an expression relating the maximum deflection of the circular plate to the pressure differential across the plate as [8]:

$$(12) \qquad w = .662 \, a^3 \left( \frac{q \, a}{E \, t} \right)^{1/2},$$

where E is Young's modulus of silicon, a is the radius of the circular cavity, t is the thickness of the capping layer. In this equation it is assumed that the residual stress in the silicon layer is negligible.

The differential pressure across the plate is a function of the deflection of the capping layer into the cavity and is given by.

$$(13) \qquad q = P_{ATM} - \frac{P_1 V_1}{V_1 - \Delta V},$$

where $P_{ATM}$ is atmospheric pressure, $P_1$ and $V_1$ are the initial pressure and initial volume of the cavity, respectively, and $\Delta V$ is the change in cavity volume. The deflection of the capping layer is assumed to have a spherical mode shape. Calculated values of the residual pressure inside the cavities measured are given in Table I. where it is seen the average residual pressure is 0.79 atm.

## Table I. Residual Pressure Inside Sealed Cavities

| Wafer # | Average Depth of Cavity (μm) | Average Deflection of Capping Layer (μm) | Residual Pressure Inside Cavity (Atm) |
|---------|------------------------------|------------------------------------------|----------------------------------------|
| 1 | 53 | 20.1 | .778 |
| 2 | 53 | 19.5 | .787 |
| 3 | 32 | 12.8 | .795 |
| 4 | 31 | 13.1 | .782 |

## Plastic Deformation of Lightly Doped Silicon

High temperature thermal processing of a bonded wafer with a sealed cavity results in the expansion of the trapped gas causing the capping layer to be loaded beyond the yield point. After removing the wafer from the furnace and allowing it to cool, it is observed that a residual strain is present in the capping layer, as is clearly seen in the SEM in Figure 7. The occurrence of plastic deformation of lightly doped silicon and heavily doped silicon has been reported elsewhere[9,10]. It was of interest to determine the onset temperature of plastic deformation in lightly doped silicon. Therefore, we prepared bonded wafer samples with cavities and placed the wafers in a furnace at various temperatures. Starting at 600 C the wafers were annealed in nitrogen for one hour and then



Figure 7. Plastically Deformed Silicon Over Cavity

removed. The change in deflection of the capping layer of the cavity was then observed using a calibrated microscope. The wafers were placed back in the furnace, this time with a temperature of 650 C and removed after one hour and re-measured. This sequence continued, incrementally increasing the temperature each time by 50 C until 1000 C was reached. We found the onset of plastic deformation to be between 800-850 C. It has been reported that the onset of plastic deformation is highly dependent on the amount of oxygen precipitates within the silicon and therefore the onset temperature we observed may vary considerably depending on the wafers and process[9].

As a result of this experiment, it was concluded that high temperature processing of the device wafers must be avoided whenever a sealed cavity exists inside the bonded wafers. Consequently, the masking oxide used to pattern bonded silicon layer 2 and layer 3 was a low-temperature deposited oxide. It is also noted that the bonding of layer 3 to the composite structure of layer 1 and layer 2 requires a high temperature anneal. However, in this case the pressure balancing of the valve helps to prevent the valve from being permanently deformed.

## Conclusion

We report on a new type of valve microstructure using the concept of pressure-balancing which allows the control of fluids at very high pressures. The process to implement the concept of the valve uses a series of wafer bonding steps to form the structure. It has been observed that gas is trapped within a sealed cavity between two bonded wafers and it was experimentally determined that this trapped gas had a residual pressure of 0.8 atm. This would indicate that the oxygen trapped within the cavity had reacted with the exposed silicon sidewalls and that the amount of inert gases left within the cavity are in direct proportion to their content in air. By exposing a bonded wafer having a sealed cavity and a relatively thin capping layer, we observed plastic deformation of the capping layer. We determined that the onset temperature of this plastic behavior to be between 800 and 850 C. We have successfully actuated prototype microvalves using voltages below 350 Volts. We are in the process of flow testing.

## Acknowledgements

## References

[1]  S.C. Terry, J.H. Jerman and J.B. Angell, "A Gas Chromatogragh Air Analyzer Fabricated on a Silicon Wafer," IEEE Transactions on Electron Devices, vol. ED-26, No. 12, p.1880, December, 1979.

[2]  S. Nakagawa, S. Shoji and M. Esashi, "A Micro Chemical Analyzing System Integrated on a Silicon Wafer," Proceedings, IEEE Micro Electro Mechanical Systems, Napa Valley, CA, p. 89, February 11-14, 1990.

[3]  H.T.G. Van Lintel, F.C.M. Van De Pol and S. Bouwstra, "A Piezoelectric Micropump Based On Micromachining of Silicon," Sensors and Actuators, Vol. 15, p. 153, 1988.

[4]  S. Park, W.H. Ko, and J.M. Prahl, "A Constant Flow-Rate Microvalve Actuator Based On Silicon And Micromachining Technology," Technical Digest, IEEE Solid-State Sensor and Actuator Workshop, Hilton Head, S.C., p. 136, June 6-9, 1988.

[5]  M.J. Zdeblick, "A Planar Process For An Electric-to-Fluidic Valve," PhD thesis, Stanford University, June, 1988.

[6]  T. Ohnstein, T. Fukiura, J Ridley and U. Boone, "Micromachined Silicon Microvalve," Proceedings, IEEE Micro Electro Mechanical Systems, Napa Valley, CA, p.95, February 11-14, 1990.

[7]  H.E. Merritt, Hydraulic Control Systems, New York: Wiley, p.44, 1967.

[8]  S.P. Timoshenko, Theory of Plates and Shells, New York: McGraw-Hill, 1940.

[9]  B. Leroy and C. Plougonven, "Warpage of Silicon Wafers," Journal of the Electrochemical Society, vol. 127, No. 4, p. 961, 1980.

[10]  F. Maseeh and S.D. Senturia, "Plastic Deformation Of Highly Doped Silicon," 5th International Conference on Solid-State Sensors and Actuators & Eurosensors III, Transducers '89, Montreux, Switzerland, p. 246, June 25-30, 1989.

# A Study on Silicon-Diaphragm Buckling

Xiaoyi Ding, Wen H. Ko, Yinli Niu, Weihua He
Elecronics Design Center
Case Western Reserve University
Cleveland, Ohio 44106

## ABSTRACT

The buckling of P$^+$ silicon diaphragm is studied. A series of experiments were designed to explore the effects of wafer doping, size and shape of rectangular diaphragm, sequence of diffusion process -(predeposition and drive-in), oxidation and annealing on buckling. It is found that all samples with predeposition as the last thermal process are flat.

Qualitative analysis is presented and the conclusions are: (1) there is no plastic deformation due to oxide layer in all the thermal processes; (2) the residual stress in P$^+$ is dependent upon the last thermal process; (3) the stress in P$^+$ layer is relieved by 50% by drive-in process and (4) the buckling of the diaphragm can be controlled by making the last process to be Boron diffusion predeposition. It is suggested that when the substrate is etched to form diaphragms, the redistribution and generation of stresses on the edges cause the buckling of the diaphragms.

## INTRODUCTION

In many cases, the silicon films used in microsensors are in a state of internal stresses. For example, the P$^+$ silicon diaphragms are seriously buckled under certain fabrication processing conditions. This was observed by the authors and others [1,2].

At present, the mechanism of silicon-diaphragm buckling is not agreed upon. The suggestion [2] that the thermal oxide layer on the diaphragm surface, created during boron diffusion, produced plastic deformation in highly doped silicon is not observed by the investigators of this study.

This article attempts to search for the mechanism of diaphragm buckling and to present a set of experiments with qualitative analyses on the buckling of p$^+$ silicon diaphragms.

## EXPERIMENTS

A diagnostic mask, shown in Fig.1, was made to define diaphragms with various shapes and sizes (0.8 to 6 mm),.to explore the influences of diaphragm shape and size on buckling. Both low-doping and high-doping diaphragms were fabricated with different high-temperature processes to examine the effects of high-temperature process, doping impurities and their distribution on diaphragm buckling.

After completion of the diaphragm etch, the diaphragms were first examined visually. The diaphragms without apparent deformation were further measured by the optical interferometer at 589 nm wavelength.

### Low-doping Diaphragms

Low-doping diaphragms were fabricated using p-n junction etch-stop technique[3], in which silicon etch occurs on p-side and leaves n-layer as the diaphragm at the completion of etch.

The (100)-oriented n on p epitaxial wafers with resistivities of 6-8 $\Omega$-cm and thicknesses of 4-6 $\mu$m were used to fabricate the diaphragms. The wafers were first thermally oxidized at 1150 °C for 2 hours in wet oxygen ambient, and then window opened on the p-side for the diaphragm etch. The oxide on the n-side was kept as the etch mask. The diaphragm etch was done at 75 °C in 20 wt % KOH etchant.

Upon completion of the diaphragm etch, all of the diaphragms were buckled due to the stresses induced by the oxide layer on the diaphragm surfaces. However, after the removal of the oxide using the buffer hydrofluoric acid (BHF), all of the diaphragms became flat, indicating that the observed buckling caused by the covering oxide was an elastic deformation.

Experiments were performed on four n/p epitaxial wafers with the same parameters as previously indicated except that two of them had thicker epitaxial layers (~10 $\mu$m). Same results as described above were observed.

In order to see the influence of high-temperature process on the stability of the low-doping diaphragms, some flat diaphragms etched in the above experiments were annealed at 1175 °C for two hours in N$_2$ and dry O$_2$ ambient. All of the diaphragms annealed in nitrogen ambient were still flat and those annealed in O$_2$ ambient were buckled but became flat again after the removal of the oxide formed during annealing in oxygen.

### Boron-doped P$^+$ Diaphragms

P-type, (100)-oriented, single-side polished silicon wafers with resistivities of 6-8 $\Omega$-cm were used to fabricate P$^+$ diaphragms. They were first thermally oxidized, then the oxide on the front side (the polished side) was removed for boron diffusion, but the oxide on the back side (the unpolished side) was kept as the diffusion mask. Solid-source boron diffusion was performed at 1100 °C for 1 to 6 hours with 2% oxygen and 98% nitrogen gas flow. After diffusion, the wafers were divided into three groups and were processed in three different ways, as shown in Fig. 2.

The wafers in group 1 were stripped of the boronsilicate glass (abbreviated BSG below) while opening the oxide windows on the back side for the diaphragm etch in EPW etchant. In groups 2 and 3, after stripping of the BSG the wafers were oxidized at 1175 °C for 1 to 4 hours in dry oxygen, referred to as "drive-in". The wafers in group 2 were then processed the same way as in group 1. After drive-in and removal of oxide on the front side, the wafers in group 3 were processed with another boron diffusion at 1125 °C for 1 hour, and then the same procedure used for group 1 was repeated.

After completion of the diaphragm etch, the diaphragms in groups 1 and 3 were all flat, but those in group 2 were all buckled. Since there was no oxide or BSG on the front surfaces before the silicon etch for the wafers in all three groups, the diaphragm buckling in group 2 was not related to oxide-induced stress.

Sixteen wafers were processed with repeated thermal cycles (processes of boron diffusion followed by drive-in are referred to as one cycle). They were divided into four groups with four wafers in each, referred to as groups 4-7 The performed processes and the corresponding diaphragm states are summarized in Table 1.

In group 5, four wafers were first diffused on the front side with boron at 1100 °C for three to four hours in 2% O$_2$ and 98% N$_2$ ambient. After stripping off the BSG, they were driven-in at 1175 °C for 2 hours in dry O$_2$ ambient . Following this, two cycles of process sequence were repeated When the last process, drive-in , was done, the oxide was removed from the wafer on the front side while patterning the oxide on the back for the diaphragm etch, and then the wafer were etched in EPW etchant at 115 °C. Upon completion of the diaphragm etch, the diaphragms on these four wafers were found to be buckled.

The processes for the wafers in group 7 were repeated using the same cycles as in group 5, but were diffused with boron one more time before the diaphragm etch. Upon completion of the silicon etch, all of the diaphragms in this group were flat.

The experimental results listed in Table 1 suggest that (1) the diaphragm states are only related to the last high-temperature

process but not the numbers of the repeated cycles; (2) the diaphragms etched from the wafers with boron diffusion as the last high-temperature process are flat, but those etched from the wafers with drive-in as the last high-temperature process are buckled.

The drive-in process was also performed at 1175 °C in nitrogen ambient instead of oxygen ambient. During this process, almost no oxide was formed, therefore, the influence of oxide could be eliminated. Two wafers, referred to as group 8, were first diffused with boron on the front side for two hours, and after stripping off BSG, were driven-in at 1175 °C for three hours with $N_2$ gas only. After the diaphragm etch the diaphragms were buckled, but the buckling was slightly less severe as compared with those in group 2 fabricated in the same way except with oxygen as the drive-in ambient.

In order to isolate the influences (if any) of the etched holes underneath neighbouring diaphragms on the same wafer, two wafers, referred to as group 9, were processed with only one rectangular diaphragm with an area of 1.5 x 2.5 mm² in the center of a two-inch silicon wafer. The same phenomena observed in samples with many diaphragms on each wafer were found. That is, the diaphragm etched from the wafer with diffusion as the last high temperature process was flat, but the one etched from the wafer with drive-in as the last high temperature process was buckled.

In another experiment, the wafers with flat diaphragms and those with buckled diaphragms were annealed at 1175 °C in nitrogen ambient for two hours. After annealing, it was found that the diaphragms which were already buckled before annealing were still buckled, and the diaphragms which were flat before annealing became buckled. In order to eliminate the possibility that this buckling might be caused by the very thin layer of oxide formed during annealing, the annealed wafers were etched in BHF acid for 3 minutes. No change in shape of the buckled diaphragms was observed.

It was also found that the buckled diaphragms could be flattened by additional boron diffusion process. Two wafers with buckled diaphragms, one from group 2 and one from group 5, were diffused with boron at 1125 °C for 1 hour. After this, it was found that all of the diaphragms became flat.

## ANALYSES

The experimental evidence--that diaphragms with diffrent shapes etched from wafers with boron-diffusion as the last high-temperature process were all flat but those etched from wafers with drive-in as the last high-temperature process were all buckled-- eliminates the possibility that the diaphragm buckling might result from the nonsymmetric stress distribution on the diaphragm plane, which may exist in the diaphragms with non-circular shapes. The experimental results for the samples in group 9 indicate that the buckling of a diaphragm is not due to the actions of neighboring diaphragms and holes under them.

Since only compressive or bending stresses in the diaphragms can cause them to buckle, all of the possible sources of the compressive or bending stresses will be considered.

### Critical Load of Diaphragm Buckling

The energy method can be used to investigate the critical load of buckling of diaphragms in the same way as it is used for the analysis of buckling of thin plates [4]. Assume that a rectangular diaphragm with length a width b with clamped edges is compressed in its middle plane by forces uniformly distributed along the edges. Let the magnitude of this compressive force per unit length of edge be denoted by $N_0$. When $N_0$ is increased to a critical value, $N_{ocr}$, the flat diaphragm becomes unstable and buckling occurs. $N_{ocr}$ can be found to be: [4]

$$N_{ocr} = \frac{4\pi^2 D}{3} \frac{(3\frac{b^2}{a^2} + 3\frac{a^2}{b^2} + 2)}{(a^2 + b^2)}$$

(1)

where D is the flexural rigidity of the films with thickness h,

$$D = \frac{Eh^3}{12(1 - v^2)}$$

(2)

In the special case of a square plate (a = b), Eq. (4-1) becomes

$$N_{ocr} = 5.33\frac{\pi^2 D}{a^2}$$

(3)

The smallest diaphragm defined by the mask is square and has an edge length of 0.8 mm. Using Eq. (3), the critical values of compressive stresses for the diaphragms with edge length of 0.8 mm and thicknesses from 1 μm to 5 μm, can be calculated to be about 1 MPa to 21 MPa ($10^7$ dyn/cm² to 2.1 x $10^8$ dyn/cm²).

Once the compressive stresses in the diaphragms are greater than these critical values, all of diaphragms will be buckled because diaphragms have larger dimensions will be buckled under even smaller stresses.

### Effect of $SiO_2$ Layer

It is known that the thermal oxide is in a state of compression on the $Si$ surface [5]. It is also observed by the experiment that the low-doping silicon diaphragms were buckled by the stresses due to the thermal oxidation layer covered the diaphragm. However, this buckling is not a plastic deformation because the buckled diaphragms became flat after the removal of the $SiO_2$ layer.

It is reported that for oxidations carried out at 1200 °C, the measured stress is 3.1 x $10^9$dyn/cm² [5], and that the yield stress in silicon whiskers at elevated temperature is about 5.5 x $10^9$dyn/cm² [6]. We also measured the stress in $SiO_2$ layer formed at 1175 °C, using the wafer curvature method [7], to be about 4.9 x $10^9$dyn/cm². If taking the thicknesses of $SiO_2$ layer and silicon substrate to be 1μm and 300 μm, respectively, the stress existing in the silicon surface under the oxide layer is only about 1.6 x $10^7$ dyn/cm², which is much smaller than the yield stress at the drive-in temperature.

### Stress Redistribution Due to Drive-in

From previous study [8],the plastic deformation and thus generated dislocations were observed in boron-doped P⁺ silicon in transmission electron micrographs and was caused by the tensile stresses developed during boron diffusion. The diaphragms etched from the wafers with boron-diffusion as the last high-temperature process had a very high density of dislocations (the order of $10^{13}$/cm²) but were flat. It is also known that: (1) the boron- diffusion induced stresses are tensile and nonuniform through the thickness of the diffused layer; (2) the generation of dislocations during boron diffusion greatly relieves the stresses; (3) the average residual stresses are measured to be about 60 MPa (6 x $10^8$ dyn/cm²).[8,9].

During drive-in, dislocations would not be further generated, because the residual stresses in the diffused layer have been less than the critical stress $\sigma_c$ for generation of dislocations. However, the movement of dislocations might happen, because the stress needed to move dislocations is smaller than $\sigma_c$ [6]. Therefore, dislocations may diffuse into the interior of silicon so that they will be more uniformly distributed in the diffused layer. The movement of the dislocations would further relieve some tensile stresses [6].

The SUPREM simulation curve for the wafer diffused with boron at 1125 °C for two hours, and the curve for the same wafer after additional drive-in at 1150 °C for two hours in dry $O_2$ and another two hours in $N_2$ were compared. It was found that the thickness of the diffused layer with boron concentration greater than 5 x $10^{18}$ /cm³ is increased from 4.5 μm to 8.0 μm and the maximum boron concentration is reduced from 5 x $10^{20}$/cm³ to 9.4 x $10^{19}$ /cm³. and moved from the surface to 3.5 μm below the surface. The decrease of boron concentration near the surface is due to the boron segregation at the oxide-silicon interface.

129

The boron atoms are nearly symmetrically distributed for the layer with a concentration larger than $5 \times 10^{19}/cm^3$ (The etch stop lever). This profile has been proven by our cantilever beam experiments and the observation of misfit dislocations at the surface [9]. The silicon wafer with boron diffusion and following drive-in were flat. The boron distribution in the cantilever beams are either uniform or nonuniform but symmetric with respect to the middle plane of the beam. According to the analysis for impurity diffusion [9], the symmetric distribution would be reasonable, for the diaphragms etched from the wafers without drive-in, the misfit dislocations were only observed from the etched side of the diaphragms. However, the misfit dislocations were observed at both surfaces of the diaphragms etched from the wafer with drive-in as the last high-temperature process, which indicates the steep boron gradient exists at both surfaces.

The wafer curvature measurement was also performed to see the influence of drive-in process on the residual stress. The sample wafer #1 was used in the measurement of tensile stresses due to boron diffusion, which was measured to be about 60 MPa. The wafer was driven-in at 1150 °C for two hours in dry $O_2$ followed by another two hours in $N_2$. Then the surface profile of the wafer was measured by DEKTAK after removing the $S_iO_2$ formed during drive-in. The same surface profile as the one before drive-in was obtained, indicating that the residual stresses of the diffused layer after drive-in are still tensile, and the total tensile forces are about the same. Because the diffused layer thickness is nearly doubled (from 4.5 to $8.0 \mu m$) The average residual stress, $S_2$, of the diffused layer (after drive-in) was determined, from the curvature of the wafer, to be 32 MPa, which is about 50% of the residual stresses measured before drive-in (60 MPa). The reduction of tensile stresses due to drive-in results from the decrease of the boron concentration and the relief of the stresses due to the movement of dislocations and generation of new misfit dislocations at the silicon surface.

From the above analyses, it would be expected that the residual stresses in a silicon wafer with drive-in as the last high-temperature process may have a distribution as shown in Fig. 3., which is symmetric with respect to the middle plane of the $P^+$ layer with thickness, d, and having boron concentration larger than $5 \times 10^{19}/cm^3$.

### Generation of Local Compressive Or Bending Force

The fact that etched diaphragms after drive-in were buckled. indicates that a local compressive state or a bending moment must exist after the diaphragm etch.

When the silicon wafer is in equilibrim, two basic conditions must be satisfied, which are: summation of stress forces equals to zero and the summation of moments about any axis equals to zero

As a cavity is etched from the silicon wafer to form a diaphragm, the equilibrium is broken in the local areas around the cavity. The edges represent the discontinuities where new stresses are created or enhanced. This stress enhancement arround the edges is similar to the phenomenon that occured on the edges of the cavity etched out into the oxide on silicon, as reported by Jaccodine et al. [5]. They indicated that at the point of the discontinuity stresses approach infinity, but in real material, plastic flow and other material adjustments occur limiting and reducing these stresses. They also reported that clusters of dislocations bordering a cavity in oxide after a heat treatment were observed, which proved that the edges of the hole represented areas of stress enhancement. In our case, the plastic deformation does not occur during the diaphragm etch because the temperature used for etching is low (115 °C), but the internal stresses and the internal bending moments will be adjusted in the nonequilibrium areas until a stable state is set up again It is due to these adjustments that the local compressive stresses and bending moment occur.

Fig.3 shows a schematic of the cross-sectional view of a diaphragm and the surrounding base, in which the stress distribution in the equilibrium area is also illustrated. After diaphragm etch, in any plane, BB', perpendicular to the surface in region AA'C'C where the quilibrium is broken, the tensile forces are not equal to the compressive forces. Therefore, an internal bending moment about the neutral axis is produced, which will enforce the plane to rotate with respect to the neutral axis until it is balanced by the bending moment of the new stresses produced by

the change of the curvature in the local area. The experimental evidence that the buckled beams with clamped ends all have convex shape indicates that this bending moment adjustment constitutes one source of the local stresses in the diaphragm.

In the flat diaphragms etched from the wafers with boron-diffusion as the last high-temperature process, the relatively high tensile stresses are only partially relieved during the adjustments of internal stresses when the cavities etched on silicon and the residual stresses in the diaphragms are still strong tensile. When these flat diaphragms are annealed, the tensile stresses are reduced, the same procedure of adjusting internal stresses and bending moments occurs. The residual tensile stresses in the diaphragms are not strong enough to maintain the flatness, therefore the diaphragms buckle.

The quantitative analysis on the local compressive stresses in the diaphragm needs to be carried out with finite element analysis on computer in the future.

## CONCLUSION

It can be concluded that the buckling of the boron-doped $P^+$ diaphragms with drive-in as the last process is due to the local compressive stresses generated when the cavity is etched on the back side of the silicon to form the free-standing diaphragms.

The key factors introducing buckling stresses are: (1) the residual stresses in heavily doped silicon wafers are very dependent upon high-temperature processes; (2) the tensile stresses in the $P^+$ layer are relieved about 50% by drive-in process so that they are smaller than the local stresses produced when the cavity is formed.

For low-doping silicon diaphragms, these two factors do not exist, therefore, they are flat if there is no other stressed films covering the surface.

The buckling problem of boron doped $P^+$ diaphragms can be controlled by selecting high-temperature sequence. The boron diffusion should be arranged as the last high-temperature process. If the drive-in process must be performed after boron diffusion, for instance, the oxide is needed as the mask, the buckled diaphragms can be flattened by an additional boron diffusion process.

## REFERENCES

[1]  Maseeh, F., and Senturia, S.D., "Plastic Deformation of Highly Doped Silicon", Abstract Digest, Transducers' 89, Montreux, Switzerland, p.246 (1989).

[2]  Huff, M.A., Senturia, S.D., and Howe, R.T., "A Thermally Isolated Microstructure Suitable for Gas Sensing Application" IEEE Solid State Sensor and Actuator Workshop, Hilton Head Island, SC, p. 47 (1988)

[3]  Kloeck, B., Collins, S.D., DeRooij, N.F., and Smith, R.L., "Study of Electrochemical Etch-Stop for High-Precision Thickness Control of Silicon Membranes", IEEE Transactions on Electron Devices, Vol. 36, p.663 (1989)

[4]  Timoshenko, S., "Theory of Elastic Stability," First Ed., McGraw Hill Book Co., Inc. p. 364 (1936)

[5]  Jaccodine, R.J., and Schlegel, W.A., "Measurement of Strains at Si-SiO2 Interface", J. Appl. Phys., Vol;. 37, p.2429 (1966)

[6]  Prussin, S , "Generation and Distribution of Dislocations by Solut Diffusion," J. Appl. Phys., Vol. 32, p. 1876 (1961)

[7]  Glang, R., Holmwood, R.A. and Rosenfeld, R.L., "Determination of Stress in Films on Single Crystalline Silicon Substrates", The Review of Scientific Instruments, Vol. 36. p 7 (1965)

[8]  Ding, Xiaoyi , "Mechanical Properties of Silicon Films and Capacitive Microsensors", Ph.D Thesis, Case Western Reserve University, May (1990)

[9]  Ding, X., Ko, W.H. and Mansour, J.M. "Residual Stress and Mechanical Properties of Boron-doped P+ -silicon films", Sensors and Actuators, Vol. A. p 866 (1990)
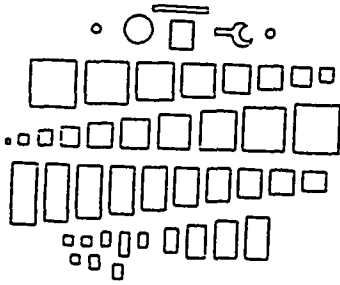
Fig. 1 Diagnostic mask for the study of diaphragm buckling (The edges vary from 0.8 to 6mm).
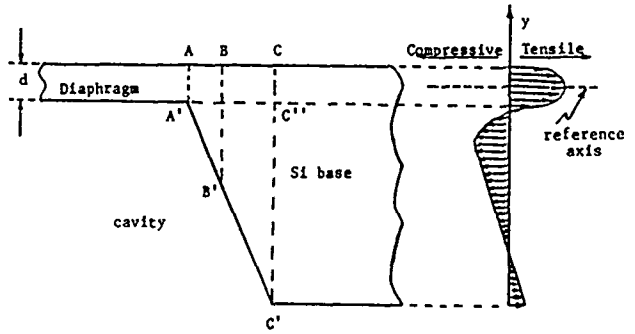


Fig.3   Schematic of diaphragm and surrounding Si base, and stress distribution in continuous regions.



(A) Oxidation and stripping of front side oxide

(B) Boron diffusion

Group 1    Group 2    Group 3

(C1) Stripping BSG on front side. Patterning oxide on back side

(C2) Stripping of BSG on front side and drive-in

(C3) Stripping of BSG on front side and drive-in

(D1) Diaphragm etch and oxide removal

(D2) Stripping oxide on front side Patterning oxide on back side

(D3) Removing oxide on front side

(E2) Diaphragm etch and oxide removal

(E3) Boron diffusion

(F3) Stripping BSC on front side.Patterning oxide on back side

(G3) Diaphragm etch and oxide removal

Fig.   2 Manufacturing sequences of p+-diaphragms in three groups.

Table 1   Fabrication process and diaphragm state for boron-doped P+ diaphragms.

| PROCESS & DIAPHRAGM STATE / GROUP # | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| WAFER QUANTITY | 9 | 10 | 8 | 4 | 4 | 4 | 4 |
| BORON DIFFUSION (1100°C, 2% O2 + 98% N2) | 1-6HRS | 1-6HRS | 1-6HRS | 3-4HRS | 3-4HRS | 3-4HRS | 3-4HRS |
| DRIVE - IN (1175°C, DRY O2) | NONE | 1-4HRS | 1-4HRS | 2HRS | 2HRS | 2HRS | 2HRS |
| FOLLOWING DIFFUSION - DRIVE IN CYCLES | NONE | NONE | NONE | 1 | 2 | 1 | 2 |
| LAST HIGH TEMPERATURE PROCESS | DIFFUSION | DRIVE IN | DIFFUSION | DRIVE-IN | DRIVE IN | DIFFUSION | DIFFUSION |
| DIAPHRAGM THICKNESS (μm) | 0.8-3.5 | 0.8-3.5 | 0.8-3.5 | 3-4 | 5-6 | 3-4 | 5-6 |
| DIAPHRAGM STATE | FLAT | BUCKLE | FLAT | BUCKLE | BUCKLE | FLAT | FLAT |

131

# POLYCRYSTALLINE DIAMOND FILM FLOW SENSOR

C.D. ELLIS, D.A. JAWORSKE*, R. RAMESHAM, T. ROPPEL

*ELECTRICAL ENGINEERING DEPARTMENT*
*ALABAMA MICROELECTRONICS SCIENCE AND TECHNOLOGY CENTER*
*AUBURN UNIVERSITY, ALABAMA 36849-5201*

*LEWIS RESEARCH CENTER*
*NATIONAL AERONAUTICS AND SPACE ADMINISTRATION*
*CLEVELAND, OHIO 44135-3127*

## ABSTRACT

A new batch fabricated micromachined silicon flow sensor has been developed incorporating a selectively deposited diamond film as a thermal element. It consists of a boron-doped silicon heater resistor thermally coupled to a boron-doped silicon temperature sensing resistor by a free-standing diamond film.

## INTRODUCTION

The recent development of a process to selectively deposit high quality polycrystalline diamond films on silicon realizes many possibilities for sensor structures [1]. Diamond films have a unique combination of material properties including: high thermal conductivity, high modulas of elasticity, and a very high electrical resistivity [2,3]. With the use of silicon, a batch fabrication process is possible, also it allows a small size (low mass) and thermally conductive base. The low mass of the diamond sensor should yield a fast response time.

Some of the other devices incorporating diamond films presently under investigation at the Alabama Microelectronics Science and Technology Center include microchannels, accelerometers, pressure and flow sensors.

## DIAMOND FLOW SENSOR FABRICATION

Three key steps were required in the fabrication of the diamond flow sensor used in this study: establishing the pattern for the boron-doped resistors, preparing the silicon substrate for the selective deposition of diamond, and depositing the diamond using microwave plasma chemical vapor deposition technology.

The essential features of the diamond flow sensor are summarized in figure 1. The resistors making up the heat source and the heat sink were derived from doping n-type silicon with boron. In this way, diamond can nucleate and grow on top of the resistor patterns, as well as in the region between the resistor patterns. Boron-doping was carried out in a two step process, with diffusion of the boron occurring at 1050°C and drive-in of the boron occurring at 1100°C. Metalized resistors were not considered for this application because nucleation and growth of diamond on metal was in question.

Dual four-point probes were chosen as the patterns for the heat source and temperature sensing resistors. They were placed symmetrically about the dia-

mond window. Hence, the functions of the resistors could be interchanged if needed. The four-point probe design was chosen over a serpentine resistor design so the exact temperature dependence of the resistivity could be determined for subsequent thermal conductivity measurements, based on the transient method developed by Van der Held [4]. In addition , the size of the diamond film window was scaled to 7x10 mm, so subsequent observations of the heat distribution could be made using an infrared camera.

The silicon substrate was prepared for the selective deposition of diamond according to the scheme proposed earlier [1]. Briefly, the surface of the silicon wafer was first scratched with 0.25 µm diamond paste. A layer of silicon nitride was deposited over the entire wafer using a mixture of dichlorosilane and ammonia gas at a temperature of 800°C to protect the scratches from the subsequent oxidation. After the silicon nitride is removed from all areas except those desired for diamond growth, a layer of silicon dioxide approximately one micron thick was grown on the wafer using pyrogenic steam at 1000°C. Removing the nitride layer exposes the original scratched area.

The diamond was grown in the exposed regions by microwave plasma chemical vapor deposition using a mixture of hydrogen and methane gas, and a substrate temperature of 900°C. The growth was only in the non oxidized regions. The act of depositing an oxide layer on top of the scratched silicon appears to protect the silicon from diamond growth, even after the oxide has been completely reduced by the atomic hydrogen in the plasma chamber. An SEM photograph showing the high selectivity of this process is shown in figure 2. The as-grown diamond films are continuous and between 10 and 20 microns thick.

The resistor values were recorded after each of the high temperature steps. After boron doping, silicon nitride deposition and thick oxide growth, the resistors measured 1000 ohms. After diamond growth, the values increased to 2500 ohms.

After the diamond growth, contact holes were opened in the oxide covering the resistor contact pads located outside of the hydrogen reducing area. Next an aluminum pad array was defined using a lift-off process to make electrical contact to the resistors. Finally, silicon was etched from the backside, in a KOH solution, to form the bridge structure. A completed device is shown in figure 3.

## THEORY OF OPERATION

The free-standing diamond film flow sensor described above is similar, in kind, to a hot-wire thermo-anemometer. The relationship between the heater power and the flow rate is given by [5]:

$$P = \Delta T (A + BU^{1/2}) \qquad (1)$$

where P is the power supplied to the device, $\Delta T$ is the temperature difference between the flowing medium and the sensor, A is a function of the thermal conductivity of the fluid, B is a collection of terms describing sensor geometry, fluid density, and fluid specific heat, and U is the velocity of the flowing medium.

The flow sensor works by forcing a sufficient current through the heater resistor to maintain a constant temperature at the temperature sensing resistor. According to equation 1, the power required to keep the sensor at a constant temperature can be used to obtain the velocity of the flowing medium. A possible circuit which accomplishes this task is shown in figure 4. As flow is increased, heat is removed from the diamond film lowering the temperature of the sensor which would cause the circuit to increase the current to the heater. Generating power vs. flow curves for different gases would allow the constants in equation 1 to be determined analytically.

## DISCUSSION

There are many parameters which need to be evaluated to determine the effectiveness of a flow sensor (i.e. response time, stability, sensitivity, chemical resistance, and temperature sensitivity), these tests are presently being performed by the authors. Advantages offered by the diamond film flow sensor are derived from the chemical inertness and hardness of diamond, making the diamond sensor particularly suited for corrosive or abrasive environments. Other flow sensors may also benefit from the addition of thin diamond films, such as the silicon sensor described by Van Oudheusden [5] which could use the diamond film as a thermally conductive and chemically resistant layer over the silicon heater and temperature sensors.

## SUMMARY

The fabrication of a unique flow sensor was described which uses a recently developed process of selective diamond deposition on a micromachined silicon substrate. It appears to be a promising structure for use in corrosive or abrasive environments as well as in general flow measurements.

## ACKNOWLEDGEMENTS

## REFERENCES

1. J.L. Davidson, C. Ellis, R. Ramesham, J. Electronic Materials, 18(6): pp. 695 - 699, 1989.
2. J.C. Angus, F.A. Buck, M. Sunkara, T.F. Groth, C.C. Hayman, and R. Gat, MRS Bulletin, October: pp. 38 - 47, 1989.
3. R.N. Boggs, Design News, April: pp. 70 - 74, 1989.
4. E.F.M. Van der Held, J. Hardebol, and J. Kalshoven, Physica, 14: pp. 208 - 216, 1953.
5. B.W. Van Oudheusden, IEEE Proceedings, 135(5): pp. 373 - 380, 1988.
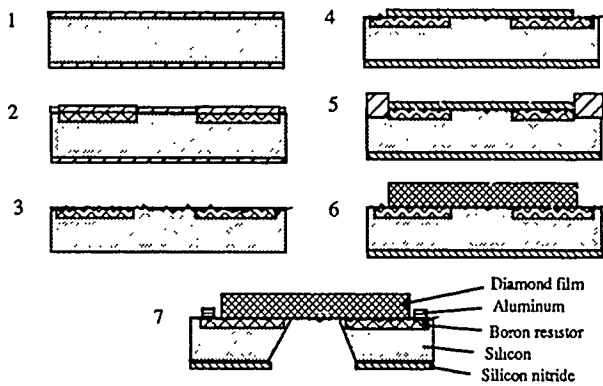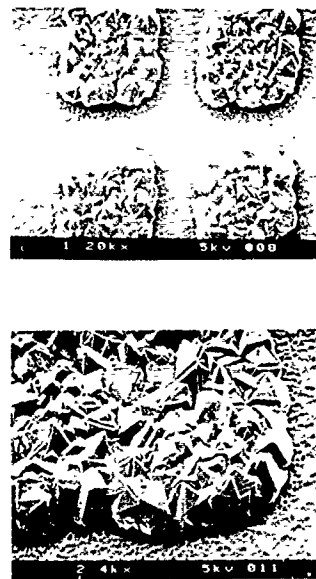
FIGURE 1

DIAMOND FLOW SENSOR PROCESS



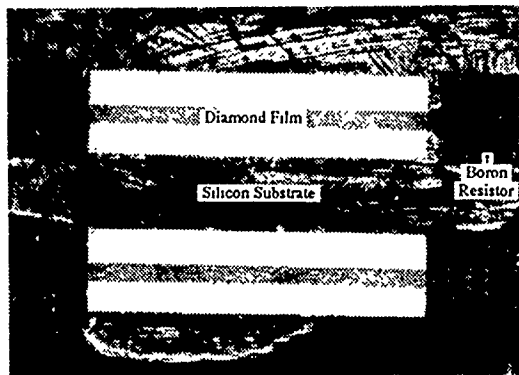FIGURE 2. Selective Deposition of Diamond



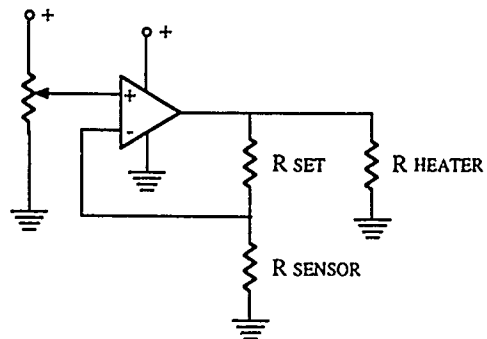FIGURE 3.

COMPLETED DEVICE



Figure 4. Flow Sensor Circuit Diagram

# LARGE DISPLACEMENT LINEAR ACTUATOR

Reid A. Brennen, Martin G. Lim, Albert P. Pisano, and Alan T. Chou

Berkeley Sensor & Actuator Center
Electronics Research Laboratory
Department of Electrical Engineering and Computer Science
University of California, Berkeley CA 94720

## ABSTRACT

In this paper, a tangential drive (T-drive) polysilicon, linear actuator is presented which produces large magnitude tangential motion by flexing a microstructure in an essentially straight line with moderate input voltage. These devices have been designed, fabricated, and successfully tested to have substantial displacements even for static positioning. The working principle of the T-drive is that the strong electrostatic forces of attraction between a fixed bar and a free bar are converted to large amplitude tangential motion by the parallelogram flexural suspension of the free bar. Operating devices are capable of tangential displacements as great as 32 μm. This displacement is stable and easily varied, since the tangential displacement can be controlled by adjusting the potential between the fixed electrode and the traversing bar. Static displacements are detectable for voltages as low as 15 VDC. Typical T-drives have free bars 200 μm long, 12 μm wide, with flexural suspensions 450 μm long and 2 μm wide All component thicknesses are 2 μm. Theoretical models for both the flexural suspension and electrostatic forces have been derived and they predict the relation between tangential displacement and excitation voltage.

## INTRODUCTION

Within the last two years, several designs of micro-actuators have emerged for specific applications. Among these, one may find shape-memory allow diaphragms for micropumps [1], cantilevered, electrostatically actuated structures to guide probe tips for scanning tunneling microscopes [2], bi-stable mechanical actuators for non-volitile logic elements [3], electrostatic linear actuators for micro friction evaluation [4], a linear micromotor for magnetic disk drives [5], electrostatically-actuated tweezers [6] as well as grippers [7]. A common design theme among these and other micro-actuators are the use of electrostatic, piezoelectric, or shape-memory properties of materials to actuate the micro structures through relatively small displacements (0.01 to 10 μm) with respect to the microactuator size. Although actuation amplitudes and forces for these devices seem to be adequate, the design of higher performance microactuators will require the development of larger amplitude motion with no compromise of actuation force.

Among the electrostatic actuators, one may find two basic designs. The first basic design utilizes parallel plate capacitors with one moving plate that is allowed to displace in the direction of the major field lines, yielding a large-force, small-amplitude actuator. The second basic design utilizes the fringe field of capacitors to drive the moving plate parallel to the fixed plate and perpendicular to the major field lines [8]. This results in a low-force, large amplitude actuator.

In this paper is described the design and performance of a large-force, large-amplitude electrostatic actuator that exhibits the best features of previous electrostatic actuator designs. Thus, the moving capacitor plate is allowed to move parallel to the electric field lines (generating large force) as well as displace parallel to the fixed capacitor plate (generating large amplitude motion). This new electrostatic actuator design utilizes tangential motion of the moving capacitor plate, and thus, is called the T-drive.

Unoptimized designs of T-drive actuators show that low voltage, large force, and large displacement linear actuation is possible.

## DESIGN

The force-generating components of the T-drive are made up of two parallel bars which are separated by a small gap. When a voltage is applied across the gap, an electrostatic force is created that acts perpendicular to, and between, the bars. For convenience, this direction is called the normal direction. By fixing one of the bars and attaching the other bar (the free bar) to one side of a parallelogram flexure suspension, Fig. 1, this force can be utilized for motion both parallel and perpendicular to the fixed bar by adjusting the geometry of the suspension. The parallelogram suspension is used in order to constrain the two bars to remain parallel. If the suspension beams of the flexure are not parallel to the normal direction, the total perpendicular force generated, $F_{total}$, can be decomposed into two force components: one component parallel to and the other component perpendicular to the suspension beams. The tangential force component acts at the end of the flexures and deflects them, resulting in lateral motion of the free bar. This lateral movement can only occur when the initial orientation of the beams is not parallel to $F_{total}$, i.e. the normal direction. Gravity acting on a simple pendulum serves as an analogy to the concept of the T-drive.

The two critical design features of the T-drive are the flexure geometry, which controls the gap distance as the structure displaces, and the free bar length, which affects the magnitude of the electrostatic force (Fig. 1). The flexure geometry parameters include the offset angle of the beams, the beam dimensions (length, width, and thickness), and the initial gap distance between the free and fixed bars. Larger initial angles of the beams result in a more rapid crease in gap width and a more rapid increase in applied force as the free bar moves laterally. However, if the initial angle is too large, the free bar will contact the fixed bar as it displaces, and thus short the circuit preventing further movement. The rate of approach of the free bar toward the fixed bar is a function of beam length. The initial gap between the bars determines whether the bars will contact and, if so, what maximum actuator displacement is possible before they meet. The width, thickness, and length of the beams determine the stiffness of the flexure suspension and the magnitude of $F_{total}$ is proportional to the free bar length.
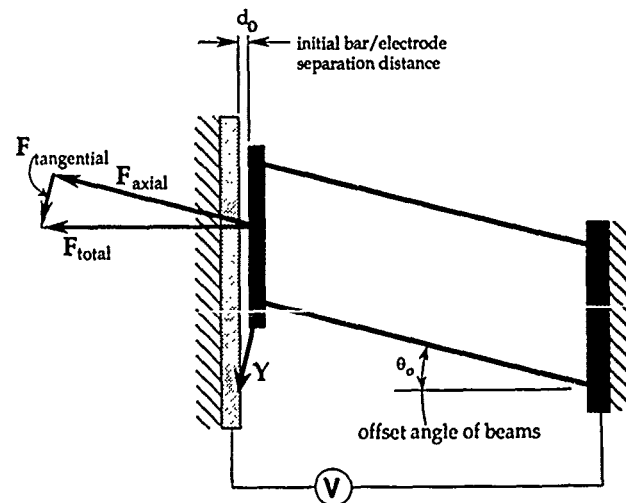


### Figure 1

Schematic of Tangential Drive Linear Actuator. Note that it is $F_{tangential}$ that deflects the structure in the Y-direction.
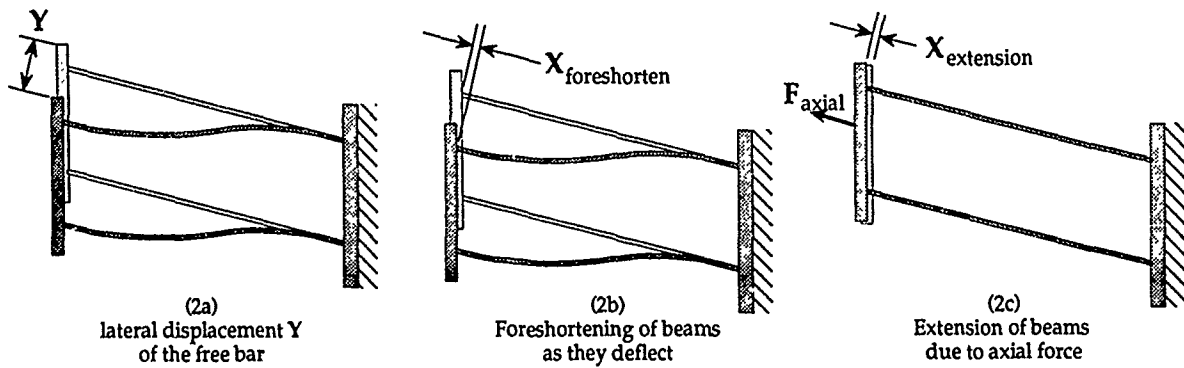
**Figure 2**
**Types of beam displacements**
(Shaded structures are in displaced position)

| (2a) | (2b) | (2c) |
|---|---|---|
| lateral displacement Y of the free bar | Foreshortening of beams as they deflect | Extension of beams due to axial force |

## THEORY

Using a modified parallel-plate capacitance model, the total electrostatic force $F_{total}$, generated between the facing sides of the bars is estimated by:

$$F_{total} = C_1 \frac{\varepsilon_o A}{2} \left(\frac{V}{d}\right)^2 \qquad (1)$$

where A is the capacitor plate area, V is the applied potential, d is the gap between the bars, $\varepsilon_o$ is the permittivity of air, and $C_1$ is a constant factor. The factor $C_1$ is based on results from a electrostatic finite element package, Maxwell [9] and improves the estimate for $F_{total}$ by including fringe fields in the capacitance analysis. For our model, $C_1$ is set to 1.15. Clearly the factor, $C_1$ must be recalculated if the design of the device is modified.

As the free bar moves laterally, it approaches the fixed bar. The rate of approach is governed by three separate phenomena ( Fig. 2). The first, due to the initial angle of the support beams, is the approach of the free bar as it moves laterally a distance Y (Fig. 2a). The second is the foreshortening, $X_f$, in the *axial* direction, of the support beams as they deflect (Fig. 2b). The third is the axial extension, $X_e$, of the support beams as the axial component of the total force is applied (Fig. 2c). If $\theta_o$ is the angle of the beams with respect to the normal direction, then the gap, d, can be expressed as:

$$d = d_o - Y\sin\theta_o + (X_f - X_e)\cos\theta_o \qquad (2)$$

where $d_o$ is the initial gap. Using small displacement beam theory, expressions for Y, $X_f$, and $X_e$ can be derived [10], and, expanding each of the terms in Eq. 2:

$$d = d_o - \frac{F_{total} L \cos^2\theta_o}{b t E} \qquad (3)$$

$$+ \frac{\tan\theta_o \sin\theta_o}{4\lambda}\left\{-\lambda L\left(1+\gamma^2\right)+\sinh(\lambda L)\left[\cosh(\lambda L)\left(1+\gamma^2\right)-2\gamma\sinh(\lambda L)\right]\right\}$$

$$\text{where} \quad \gamma = \frac{\cosh(\lambda L) - 1}{\sinh(\lambda L)} \qquad (4)$$

$$\lambda = \left[\frac{F_{total}\cos\theta_o}{EI}\right]^{\frac{1}{2}} \qquad (5)$$

and L is the length of the flexure beams, b is the width of the beams, t is the thickness of the beams, E is Young's modulus of elasticity for polysilicon, and I is the area moment of inertia of the beams.
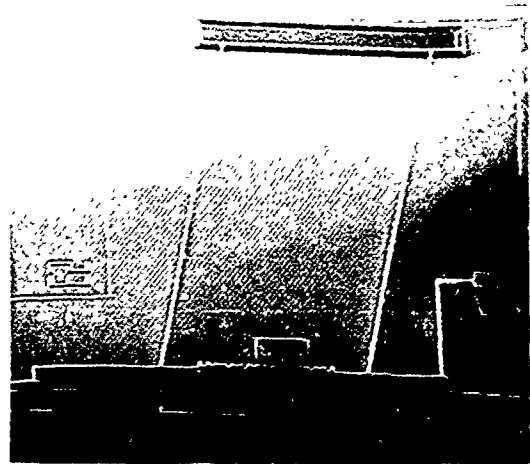


**Figure 3**
SEM of a single bar T-drive structure. The L-shaped breakaway support is removed before testing.
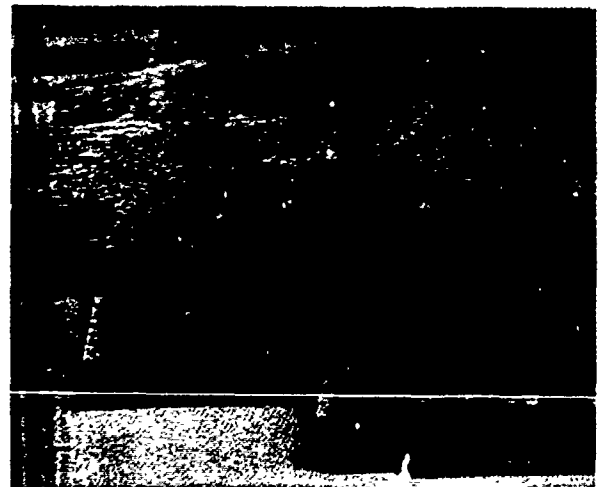


**Figure 4**
SEM of the free bar of a single bar T-drive.
The scale and pointer can be seen between the two suspension beams on the left. The dimples can be seen in the center of the free bar. Note the vertical side walls of the polysilicon in the gap.
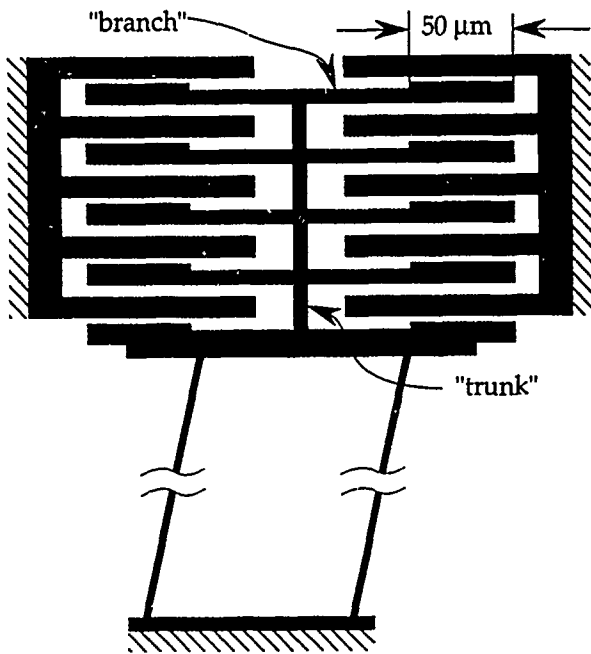
136

"branch"

50 μm

"trunk"

**Figure 5**
**Schematic of Multiple bar T-drive Structure**
Note the 50 mm long sections on each branch which
act as the parallel plate attractors.

By solving equations (1) and (3) simultaneously, values for $F_{total}$ and d can be found for given values of the applied voltage V. Once $F_{total}$ is known, the lateral displacement Y of the free bar can be determined, by:

$$Y(F_{total}) = \frac{\tan\theta_o}{\lambda}\left[\lambda L - \sinh(\lambda L) + \frac{(\cosh(\lambda L) - 1)^2}{\sinh(\lambda L)}\right] \quad (6)$$

Two designs were implemented to demonstrate the T-drive concept. The first design was very similar to the schematic in Fig. 1. This design used a single free bar attached to the parallelogram suspension (Figs. 3 and 4). In this design, the electrostatic attracting force is linearly proportional to the length of the free bar. The second design increased the effective bar length by using multiple free bar/fixed bar pairs which form a "tree" (Figs. 5 and 6). The odd-shaped bars or "branches" attached to the central "trunk" of the tree of the suspension each have a wider 50 μm long section which has a 2 μm distance from the free bar. All other parts of the branch, including the back side of the 50 μm section were separated by a distance of at least 6 μm gap from the fixed bars. In this manner, the 50 μm section acted as a constant attraction surface area; as the structure displaced, the effective area of the parallel "plates" did not change.

Both of the structure types used a scale and pointer method to measure displacements. Due to the long overhang lengths of the structures, sometimes as great as 600 μm long, breakaway supports were used to provide added stiffness to the structure during fabrication (seem in Fig. 3). The breakaway supports were mechanically removed by physically rupturing them with test station probes. Small dimples were introduced into the free bars to prevent large area contact between the free bars and the substrate, thereby keeping the released structure free from surface tension adhesion during rinsing and drying.

### FABRICATION and EXPERIMENTAL METHODS

The structures were fabricated using the process developed by Lim [4]. The structural layer defining the device was 2 μm thick phosphorus-doped polycrystalline silicon. The T-drive structures were fabricated with 1.7 μm wide, 400 and 446 μm length beams and with beam angles ranging from 4.6° to 5.2°. The initial gap between the free and fixed bars was 2.1 μm. A ground plane was used beneath the free bar and its suspension to prevent electrostatic interference from the substrate.



**Figure 6**
**SEM of the free bar component of a multiple bar T-drive.**
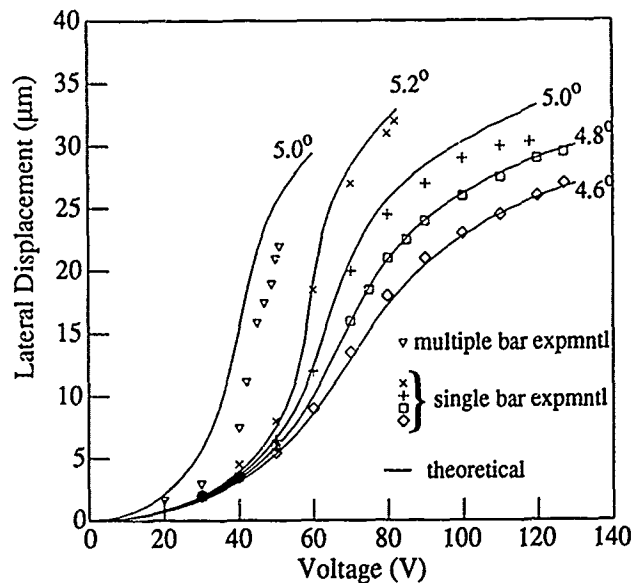


**Figure 7**
**Lateral displacement Y versus applied voltage for various T-drive structures. The corresponding angle of the beams with respect to the primary direction is shown next to each set of data.**

Static displacement measurements as a function of applied voltage were performed for several structures by use of the scale and pointer. This method allowed a ± 0.5 μm accuracy. The applied voltages ranged from 0 to 127 V. Out-of-plane deflections of the free bar were made using a microscope focusing method with an accuracy of about ±2 μm for small deflections. All testing was performed in air observed under under 1000X magnification both directly through a microscope and on a video monitor. The beam widths of the structures were measured using SEM images.

### RESULTS and DISCUSSION

The theoretical determination of static displacement was dependent on the accuracy of both Young's modulus and the suspension beam widths. After measuring the beam widths, the value of Young's modulus used in the theoretical model was varied until theoretical displacements were found that corresponded to the experimental data. Using this method, the Young's modulus was found to be 105 ±15 GPa.

Static displacement amplitudes as a function of voltage were measured for several single and multiple bar structures and are plotted in Fig 7. The displacements of the structures were smoothly controlled by varying the voltage. The T-drive designs with larger suspension beam angles deflected more for a given voltage. The larger beam angle resulted in a larger tangential component of the total electrostatic force

137

**Table 1** Maximum Displacements attained for Single and Multiple-Bar T-drive structures each with its beams at different angles with respect to the normal direction.

| | Maximum displacement (μm) (voltage at maximum displacement) | | | |
|---|---|---|---|---|
| | Angle of beams | | | |
| | 4.6° | 4.8° | 5.0° | 5.2° |
| single bar T-drive (446 μm beam) | 27.0 (127V) | 29.5 (127V) | 30.5 (118)* | 32.0 (82V)* |
| Multiple bar T-drive (446 μm beam) | 22.0 (57V)† | 20.0 (53V)† | 22.0 (50)† | 21.0 (45)† |

\* At higher voltages, the free bar contacted the fixed bar.

† At higher voltages, one of the branches deflected and contacted the fixed bar.



**Figure 8**
Theoretical force exerted by the T-drive versus applied voltage at fixed lateral displacements. The fixed displacement is noted by each curve.

since $F_{tangential} = F_{total} \sin\theta$. This larger $F_{tangential}$ reduced the gap more for a given voltage than that of the lesser-angle beam designs. This increased $F_{total}$ and correspondingly, $F_{tangential}$. Consequently, the displacements of the larger angle beam designs increased faster than the lower voltage designs as the voltage was increased.

For a given voltage, the single bar structures deflected less than the multiple bar structures. The experimental results bear out the expected proportionality between the length of the attracting surfaces (i.e. the bar lengths) and $F_{total}$.
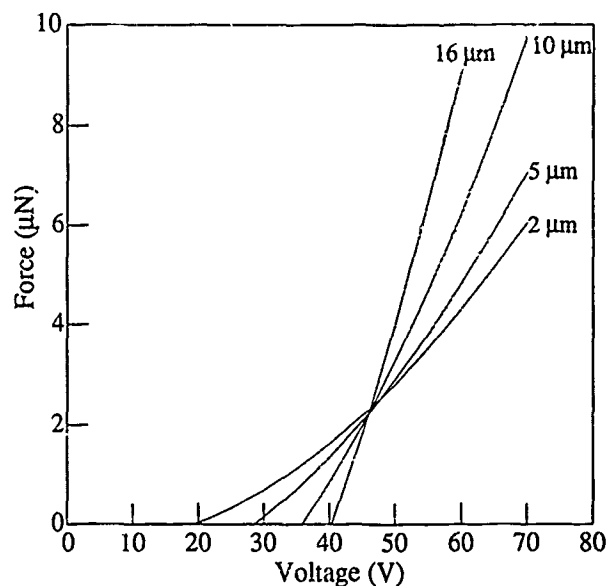
The maximum static displacement amplitudes attained for the structures are shown in Table 1. For an applied voltage of 127V, a static displacement of 32 μm was measured for a single bar T-drive with 446 μm beams at an angle of 5.2°. The 32 μm displacement was the greatest displacement measured for any of the T-drive structures. The maximum displacement increased with an increase in $\theta_0$. Geometrical constraints limited the maximum displacement of the structures with large $\theta_0$. In these cases the bars contacted before the voltage reached 127V.

There was a significant difference between the maximum attainable displacements of the single and multiple bar structures. The multiple bar design of the tree end was inherently more compliant than the single bar structures. It was noted during testing that the branches deflected towards their respective fixed bars due to the large forces generated as the gap became smaller. When the gap decreased to a certain amount, the force balance between the electrostatic force and the resisting force in one of the branches would become unstable and the branch contact the fixed bar. The complexity of the "tree" may have contributed to the less than predicted displacements.

During testing of the single and multiple bar T-drives, it was noted that the free bar section of the devices levitated out of plane. The amplitude of the levitation varied from device to device depending on the physical dimensions of each device. For most of the devices, the out-of-plane levitation increased as the applied voltage increased until about 30 V was reached, after which the levitation decreased and was not perceptible above about 50 V This levitation is due to the interaction of the ground plane and the free and fixed bars [11].

Theoretical calculations reveal that the forces that the T-drive would be able to generate are dependent on how much the suspension has been displaced before the force is exerted This effect is due to the elastic forces required to deflect the suspension For example, for one multiple bar design, an initial displacement of 10 μm would require approximately 36V (Fig. 8). Assuming the structure at this point contacts the the structure upon which it acts and the T-drive does not displace further, an increase of 8V would provide a force of 2 μN (1 μN is the force exerted by a 350 μm cube of silicon in gravity). Lower initial displacements reduce the rate at which the applied force would increase with voltage.

The differences in displacements between the theoretical model and the experimental data can be attributed to several causes. First, the structures not only have compliant suspension beams, but the free bars could also deform. As noted above, this was of especial concern for the multiple bar T-drive. Further, in the multiple bar free bar/fixed bar design, only part of each branch was near to the fixed bar, the nominal gap being 2 μm. However, the branch was surrounded by other features that were about 6 μm distant that were at the opposite potential. Since an electrostatic force was generated by this gap, the total force was actually less than the theoretical force.

The model had several limitations. As noted above, the beam theory used to describe the T-drive suspension was valid only for small displacements, or more accurately, for small changes of slope of the beams. These T-drive suspension beams did have significant changes of slope when they deflected. The effect of this increased slope was to effectively increase the stiffness of the system [12]. A second limitation of the model was that the model assumed perfect geometric dimensions whereas the actual structures had dimensions that varied slightly with the fabrication process. Since large displacements only occurred when the gap had become very small, the uniformity of the gap was very important.

The motion generated by the T-drive is not perfectly linear, but since the free bar travels in a very large radius arc as the beams deflect, but this non-linearity is very small.

## CONCLUSION

Electrostatic large displacement structures have been developed, fabricated, and tested. The T-drive concept demonstrates that direct electrostatic attraction can be used to generate large displacements. These structures attained static displacements of up to 32 μm with an applied voltage of 82V and 22 μm with an applied voltage of 50V. Structures with longer beams show promise for even greater displacements. A theoretical model has been developed which closely describes the experimental data. Using this model, a value has been calculated for Young's modulus of 105 GPa. Further, it has been shown that relatively large forces can be applied with small increases in voltage once the free bar undergoes a large displacement and the gap between the free and fixed bars is reduced.

## REFERENCES

[1] J.D. Busch and A.D. Anderson, "Prototype Micro-Valve Actuator", *Proc. IEEE Micro Electro Mechanical Systems Workshop*, Napa Valley, CA, Feb. 1989.

[2]     S. Akamine, T.R. Albrecht, M.J. Zdeblick, and C.F. Quate, "Microfabrication of Integrated Scanning Tunneling Microscopes", *Transducers '89, 5th International Conference on Solid-State Sensors and Actuators*, Montreux, Switzerland, 25-30 June 1989.

[3]     P.L. Bergstrom, T. Tamagawa, and D.L. Polla, "Design and Fabrication of Micromechanical Logic Elements",*Proc. IEEE Micro Electro Mechanical Systems Workshop*, pp. 15-20, Napa Valley, CA, Feb. 1989.

[4]     M. G. Lim, J. C. Chang, D. P. Schultz, R. T. Howe, and R. M. White, "Polysilicon Microstructures for Characterize Static Friction." *Proc. IEEE Micro Electro Mechanical Systems Workshop*, pp. 82-88, Napa Valley, CA, Feb. 1989.

[5]     Martin G. Lim, Roger T. Howe, and Roberto Horowitz, "Design and Fabrication of a Linear Micromotor for Magnetic Disk File Systems", ASME Winter Annual Meeting, San Francisco, December 10-15, 1989

[6]     L. Y. Chen, Z. L. Zhang, J. J. Yao, D. C. Thomas, and N. C. MacDonald, "Selective Chemical Vapor Deposition of Tungsten for Microdynamic Structures," *Proc. IEEE Micro Electro Mechanical Systems Workshop*, pp. 82-87, Salt Lake City, UT, Feb. 1989.

[7]     Chang-Jin Kim, Albert P. Pisano, Richard S. Muller, and Martin G. Lim, "Polysilicon Microgripper", *to be presented at IEEE Solid-State Sensor and Actuator Workshop*, Hilton Head Island, South Carolina, June 4-7, 1990.

[8]     William C. Tang, Tu-Cuong H. Nguyen, Michael W. Judy, and Roger T. Howe, "Electrostatic-Comb Drive of Lateral Polysilicon Resonators", *Transducers '89, 5th International Conference on Solid-State Sensors and Actuators*, Montreux, Switzerland, 25-30 June 1989.

[9]     Ansoft Corporation, Pittsburgh, PA 15123 *Maxwell Users Guide*, January, 1989

[10]    Raymond J. Roark and Warren C. Young, Formulas for Stress and Strain, 5th Ed., McGraw-Hill, New York, 1975

[11]    William C Tang, Martin G. Lim, and Roger T. Howe, "Electrostatically Balanced Comb Drive",*to be presented at IEEE Solid-State Sensor and Actuator Workshop*, Hilton Head Island, South Carolina, June 4-7, 1990

[12]    Egor P. Popov, Introduction to Mechanics of Solids, Prentice-Hall, Inc., Englewood Cliff, New Jersey, 1968

# A Miniature Fabry-Perot Interferometer
## with a Corrugated Silicon Diaphragm Support

by
J. H. Jerman
IC Sensors, Milpitas, CA

D. J. Clift and S. R. Mallinson
British Telecom Research Laboratories, Ipswich, England

## Abstract

The techniques of silicon micromachining have been used to fabricate a second-generation Fabry-Perot interferometer for use in the near infra-red spectral region. The device consists of two silicon wafers with deposited, highly reflective dielectric mirrors. The wafers are bonded together with a small gap between the mirrors, and sawn into individual devices. The wavelength tuning and parallelism control of the mirror elements are achieved electrostatically, by varying the voltage between control electrodes. This second-generation device includes a thin, etch-stopped corrugated diaphragm as the suspension element, which allows increased travel in a device occupying only 13% of the previous die area.

## Introduction

A Fabry-Perot interferometer (FPI) is an optical element consisting of two partially reflecting, low loss, parallel mirrors separated by a gap. The optical transmission characteristic of such an element consists of a series of sharp resonant transmission peaks when the gap equals multiples of a half wavelength of the incident light. These transmission peaks are caused by multiple reflections of the light in the cavity, and by using highly reflective mirrors, small changes in gap can produce large changes in transmission response. Even though two reflective mirrors are used, at the peak wavelengths, the transmission through the element approaches unity. In these silicon devices the mirrors are deposited on the interior surfaces of a bonded pair of silicon parts. Such a device can operate in the near infrared, at wavelengths greater than 1.15 µm, where the silicon substrates are highly transparent. A device previously reported by the authors[1] showed the feasibility of producing a high performance silicon-based FPI, although a very large, flat diaphragm was used as the suspension element for the movable optical mirror, which resulted in a large die size. The present structure uses an etch-stopped, corrugated diaphragm support in place of a flat diaphragm. These corrugated structures offer many significant advantages over conventional diaphragms,[2] including increased linear travel, isolation from case stresses, and resistance to off-axis bending. They are particularly useful in providing large vertical travel in a bossed structure with a relatively small suspension area. These corrugations are formed by etching concentric grooves in the front surface of a wafer and diffusing an etch-stop layer into that non-planar surface. After etching from the back surface, the diaphragm that is formed follows the etched contour of the front surface. By using an isotropic etch for the groove formation, the corrugations can follow any desired pattern, including circular, spiral, and serpentine. In this case, the grooves have straight sides with rounded corners, surrounding a

square center boss. The overall die size is 5 mm square, with a central optical cavity 1.4 mm square.

The increasing pressure on transmission capacity of optical fibers has led to increasing interest in having several optical channels operating in the same transmission window using wavelength division multiplexing. An essential component of such a system is a low loss, narrow linewidth optical filter capable of being tuned to select a particular channel and to provide isolation from other channels. The silicon micromachined FPI represents such a device and has high potential for future use in this application. A first generation part has been successfully used in such a system and described in a recent paper,[3] and some aspects of this trial are discussed later in this report.

## Device Fabrication

This FPI, shown in Fig. 1, consists of two silicon wafers, bonded together at the wafer level, and sawn into individual devices. Both wafers have highly reflective, multi-layer dielectric mirrors and arrays of metallic electrodes deposited on the interior surfaces. One wafer consists of silicon mesas surrounded by corrugated diaphragm suspensions. The other wafer has a matching set of electrodes which are used to electrostatically vary the gap in the devices.

This new device uses a novel technique for forming the initial optical gap between the wafers. The silicon surfaces on which the optical coatings are deposited must be equivalent in surface finish to un-processed silicon wafers. This precludes the use of an etched silicon surface as the mirror surface. Instead, the cap wafer is oxidized and a sacrificial wafer is fusion bonded to the eventual mirror surface. The sacrificial wafer is then etched down to the desired gap thickness, in this case about 25 µm, and that etched surface is then patterned and etched to form the optical cavity. Since the original cap wafer surface is protected by the original oxide layer during this process, it remains acceptably flat and smooth when re-exposed. The optical coatings and metalizations are then deposited and patterned. The cap and mesa wafers are bonded together using a metallic bonding system, which allows electrical crossovers between the interior surfaces, providing connections to electrodes on both the cap and the movable silicon mesa from bonding pads on one of the surfaces.

The quality of the optical coatings is critical to the operation of the device. The coatings must have high reflectivity to provide narrow optical bandwidth, but with very low loss to provide high peak transmittance. The reflective coatings used are multi-layer dielectric mirrors with alternating layers of high and low index material, a common optical coating technique.
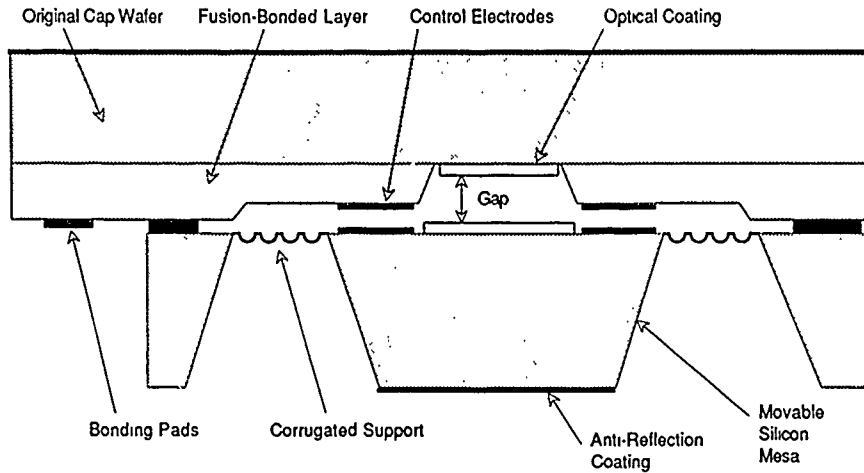
Figure 1. Cross-sectional, schematic view of second generation Fabry-Perot interferometer. Vertical scale exaggerated for clarity.

They have a reflectivity which ranges from about 95% at 1.3 μm to 97.5% at 1.55 μm. Since silicon has a refractive index of about 3.5, antireflection coatings are required on the external surfaces to avoid substantial Fresnel reflection loss in the devices.

The devices are controlled by the application of voltages between pairs of control electrodes on the mesa and cap. The electrostatic force tends to draw the mesa towards the opposite surface, thus reducing the cavity spacing. This motion is balanced by the elastic restoring force of the etched diaphragm support. There are four pairs of drive electrodes on the device, one pair for each side of the square cavity. This allows control of not only the spacing of the cavity, but also the parallelism of the two mirrors, maximizing the finesse of the system. Additionally, the spacings of the electrodes can be monitored capacitively and active feedback applied to maintain parallelism and environmental stability.[4]

**Optical Transmission**

The transmission of a FPI is described by the Airy function:

$$ I = \left(1 - \frac{A}{(1-R)}\right)^2 \frac{1}{1 + \frac{4R \sin^2(kd\cos\theta)}{(1-R)^2}} \qquad (1) $$

where
A = mirror absorptance,
R = mirror reflectance,
d = cavity gap,
θ = angle of incidence of the beam, and
k = $2\pi/\lambda$

The wavelength response described by the Airy function consists of a series of resonant peaks. It can be seen from (1) that with kd=nπ, and with n, the order of the interference, equal to a positive integer, the transmission for normal incidence is just:

$$ T_{max} = \left(1 - \frac{A}{(1-R)}\right)^2 $$

and the transmission intensity minimum between these peaks is given by:

$$ T_{min} = T_{max} \left(\frac{(1-R)}{(1+R)}\right)^2 $$

The transmission peaks are separated in wavelength by $\Delta\lambda = \lambda^2/(2d)$, and this spacing is refered to as the free spectral range (FSR). Each peak has a full width at half maximum (FWHM) bandwidth given by:

$$ FWHM = \frac{\lambda(1-R)}{n\pi\sqrt{R}} $$

The instrument finesse, F, is the ratio of the spacing of the transmission maxima, the free spectral range, to the FWHM linewidth. Defects are characterized by the defects finesse $F_D$ which is related to the instrument finesse by:

$$ \frac{1}{F^2} = \frac{1}{F_R^2} + \frac{1}{F_D^2} $$

where $F_R$ is the reflection finesse given by:

$$ F_R = \frac{\pi\sqrt{R}}{(1-R)} $$

An important feature of a FPI is that the free spectral range and the FWHM bandwidth can be independently controlled. For a given wavelength, the cavity gap sets the free spectral range and the mirror reflectivity controls the bandwidth. Thus the properties of the filter can be tailored to a particular application.

For example, for the current devices with about a 24 μm gap and a mirror reflectance of 0.95 the theoretical response is shown in Fig. 2. Note that in this theoretical case the free spectral range and the FWHM bandwidth (both measured at λ=1.3 μm) are around 37 nm and 0.6 nm respectively. This filter could effectively separate up to about 15 optical channels that are spaced in a window less than 37 nm wide.
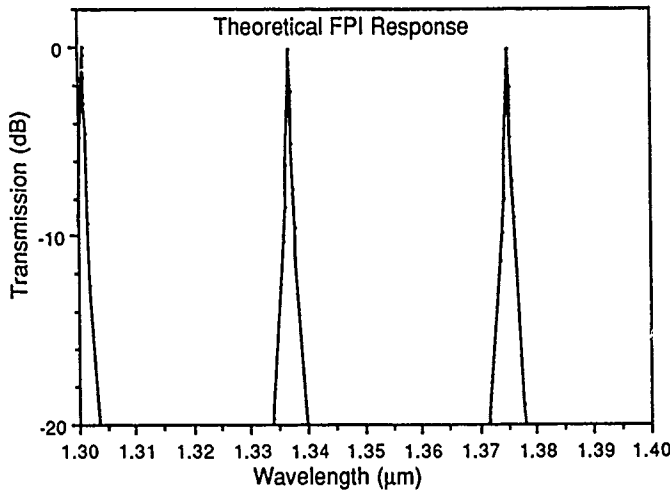
141

Figure 2. Theoretical response of a FPI interferometer with 95% mirror reflectance and 24 μm gap.
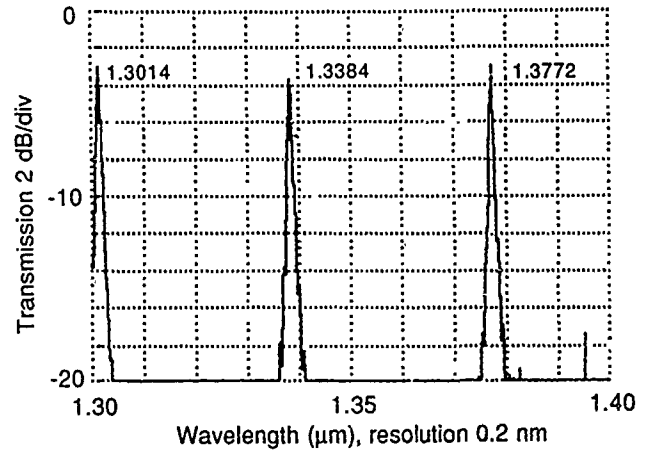


Figure 3. Experimental FPI transmission as a function of wavelength for device J11 when electrostatically adjusted to parallelism. The FWHM optical bandwidth is 0.9 nm.

## Experimental Results

It is possible to observe the parallelism of two surfaces of the device by using a monochromatic infra-red source and TV camera. After fabrication, there is typically a misalignment of less than two fringes, corresponding to an angular misalignment on the order of $10^{-2}$ degrees. This result testifies to the excellent uniformity and stress-free nature of the suspension membrane after processing. By the application of appropriate bias voltages to the drive electrodes, the parallelism could be improved to better than $3 \times 10^{-4}$ degrees.

When the devices are tuned to parallelism the optimum performance is obtained. The optical performance of the device was assessed by measuring the transmission response. A broadband infrared source was launched into a standard single-mode fiber attached to a proprietary beam collimation component. The transmitted light coupled into a second single-mode fiber was then measured with an optical spectrum analyzer. The filter was inserted into the collimated beam and the transmission, at a given wavelength, was recorded by measuring the light present at that wavelength when the device was inserted compared to when it was not. The results from one such measurement are shown in Fig. 3.

The second generation devices have an optical gap of around 24 μm and an instrumental finesse of 40. At a wavelength of 1.3 μm, this results in a FWHM optical bandwidth of 0.9 nm and a free spectral range of 38 nm. The peak wavelength transmission of the devices is around -3.5 dB.

The relationship between the peak transmission wavelength and the applied voltage can be obtained by solving for the equilibrium position of the center boss by modeling the non-linear stiffness of the suspension and the non-linear force characteristics of the control electrodes, and comparing that result to the measured wavelength tuning as a function of applied voltage. The electrostatic force is given by:

$$F = \frac{V^2 A_e \varepsilon}{2 x^2}$$

where V is the applied voltage, $A_e$ the total electrode area, $\varepsilon$ is the permittivity of air, and x is the electrode gap. For these devices the electrode gap is set by a different etch than the optical gap, and is about 6 μm. The total electrode area is about 4 square mm.

The stiffness characteristics of the diaphragm can be determined from measuring the deflection vs. applied differential pressure on an actual diaphragm. These results have been compared to a theoretical, circular corrugated diaphragm with the measured thickness and corrugation depth of the actual devices, a boss ratio of 80%, and an effective diameter 10% greater than the length of the square diaphragm, and are shown in Fig. 4.
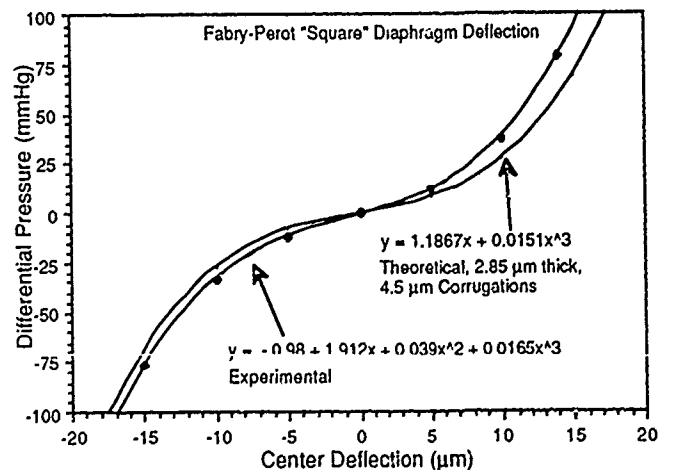


Figure 4. Experimental and theoretical deflection characteristics of the center boss supported by a shallow, corrugated diaphragm.

Note that the linear deflection term is roughly 60% greater than the theoretical value, reflecting the increased stiffness of the diaphragm due to the residual stress of the p+ etch stop used to form the diaphragm. The shallow corrugations used in this device serve to minimize this effect but not eliminate it completely. The cubic terms are quite similar, which is commonly achieved with these structures. The deflection as a function of force can be determined from the product of the diaphragm area, the applied pressure, and the so-called effective area of the diaphragm.
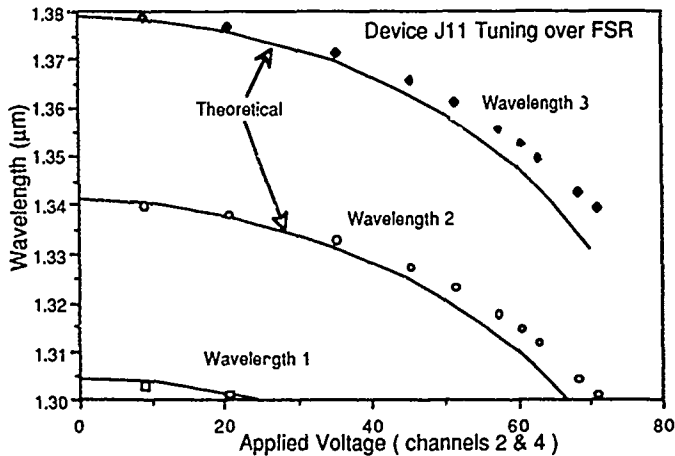


Figure 5  Theoretical and experimental tuning characteristics of device J11.

These relations can be numerically solved and compared to the actual tuning of a device. These results are shown in Fig. 5. There are 3 transmission peaks in the range of 1.3 to 1.38 since the FSR of this device is about 37 nm. With the application of 0-70 V, on a pair of control electrodes, the device can be tuned over a full free spectral range, which would tune any optical channel in that FSR.

The experimental results imply that either the initial electrode gap was slightly greater than 6 µm or the particular diaphragm suspension of this part was slightly stiffer than the suspension used in the pressure vs. deflection measurements.

### Application to Optical Fiber WDM Systems

Passive optical networks have been proposed to link large numbers of customers to a central exchange. A first generation silicon FPI was used in an experiment incorporating a 32-way passive optical split between a head end and a receiver  The FPI provided isolation between two trial optical systems operating over the same fiber in the same, 1.3 µm, window. The first system operated bidirectionally at 1.33 µm and at a data rate of 20 Mbit/s, simulating up to 128 voice channels. The second system offered broadband services at 1.31 µm using subcarrier multiplexing (SCM) techniques to provide 32 wide-band FM video channels (950-1750 MHz) to each customer. The test system is shown in Fig. 6.
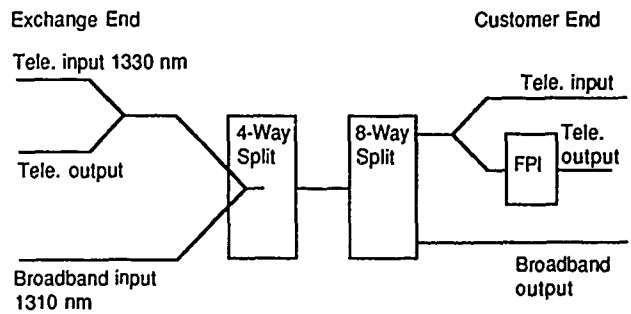


Figure 6.  Test system for FPI evaluation.

The systems were tested separately and the telephony channel was found to provide error-free (BER < $10^{-9}$) operation with a received optical power of -44 dBm. However, the addition of the broadband system reduced the receiver sensitivity by 7 dB without the use of the FPI. With the FPI inserted and the optical power adjusted to make up for the insertion loss of the filter, the telephony system was again found to operate error free without the SCM system. With the filter present, however, the broadband SCM system could also be operated with no observable degradation of the telephony system. The first-generation FPI filter provided over 20 dB of isolation between the two channels. The second-generation part has substantially narrower FWHM bandwidth compared to the previous part and should provide greater isolation between adjacent channels than was shown in this test.

### Application as Sensors

As noted in the 1988 paper, these devices can be used in a number of ways as a sensor. Since the transmission of the device is a function of both the gap spacing and the wavelength, the device can be used as a sensor of either parameter. By varying the gap and monitoring the intensity, the device can be used as a monochromator, within the free spectral range of the device. By maintaining the incident wavelength constant, the device can be used as a very sensitive position sensor, when the gap is near a transmission maxima.

The maximum slope of the transmission curve occurs at a transmitted intensity of about 0.75, where the theoretical slope is 0.005 nm per percent of intensity change for these second generation devices. Given the ability to detect a change in intensity of 0.1%, a change in wavelength of 0.0005 nm could be detected. For the current devices the slope of the tuning curve is on the order of 1 nm/V, thus a change in wavelength of 0.0005 nm represents a difference in drive voltage of 0.5 mV on a 50 V drive.

FPI's with relatively large gaps do not generally make the most sensitive position sensors, since $\partial x/\partial \lambda = n/2$, where n is the order of interference of the cavity. For this device n=36, and thus the ability to detect a wavelength change of 0.0005 nm represents the ability to detect a change in gap of about 0.009 nm. When operated as a sensor, the FPI is generally operated in a servo loop, where the gap is maintained such that the transmitted amplitude remains constant. A very stable laser source is required so that the wavelength of the incident light is constant. The drive signal needed to

143

maintain the gap is the output signal. Similar performance has been reported for a confocal FPI, where a minimum sinusoidal displacement of about $2.5 \times 10^{-3}$ nm was detected between 10 and 200 Hz.[5]

This position servo signal can be used to measure either differential pressure across the diaphragm support or acceleration of the center mass. Given the spring constant of the diaphragm support of the current device, a change in gap of 0.009 nm corresponds to a differential pressure of $3.4 \times 10^{-7}$ PSI.

## Conclusions

The use of silicon micromachining has resulted in a high performance Fabry-Perot interferometer. This fabrication technique allows the inclusion of electrodes to tune and adjust the parallelism, along with the ability to closely specify the initial gap of the device. The ability to batch fabricate these devices should result in a substantially less expensive device than is currently available. The use of an etch-stopped, corrugated support in these second generation devices has resulted in a device with increased performance and much smaller die size. The devices have the significant benefits of being considerably smaller than other types of FPI and requiring low drive voltages. They will undoubtedly prove to be of use as wavelength selection components in optical fiber communication systems, and offer significant promise in use as sensors.

## Acknowledgements

## References

1. J.H. Jerman and S.R. Mallinson, A Miniature Fabry-Perot Interferometer Fabricated Using Silicon Micromachining Techniques, Technical Digest 1988 Hilton Head Workshop, p. 16

2. Jerman, J.H., The Fabrication and Use of Micromachined Corrugated Silicon Diaphragms, Abstracts, 5th Int. Conf. on Solid-State Sensors and Actuators, Montreux, June, 1989, p. 274.

3. Harper, J.S., Rosher, P.A., Fenning, S., and Mallinson, S.R., Application of Miniature Micromachined Fabry-Perot Interferometer to Optical Fibre WDM System, *Eletronics Letters,* Vol. 25, No. 16, August, 1989, p. 1065.

4. Hicks, T.R., Reay, N.K., and Atherton, P.D., "The application of capacitance micrometry to the control of Fabry-Perot etalons", *J. Phys. E.*, 1984, 17, p 49.

5. Jackson, D.A., Kersey, A.D., and Corke, M., Confocal Fabry-Perot Sensor, *Electronics Letters,* 4 March 1982, Vol. 18 No. 5, p. 227.

# RF Telemetry Powering and Control of Hermetically Sealed Integrated Sensors and Actuators

*Tayfun Akin, Babak Ziaie, and Khalil Najafi*

## Center for Integrated Sensors and Circuits
Department of Electrical Engineering and Computer Science
University of Michigan
Ann Arbor, Michigan 48109-2122

## Abstract

A RF telemetry system for powering and control of micro-miniature integrated transducers packaged in a hermetically sealed custom made glass capsule is described. A class-E power amplifier is utilized to energize and control a single-channel microstimulator for neuromuscular applications. The transmitter can deliver 1.5W of power through a solenoid antenna using a 12V rechargeable battery, while achieving efficiencies greater than 90%. The microstimulator antenna is a 1.5mm diameter, 4-5mm long ferrite-cored solenoid coil which can generate peak voltages of up to 50V, and up to 50mW of power. A custom-made glass capsule electrostatically bonded to the microstimulator's silicon carrier provides a hermetic package for the implant. Lead transfer into the sealed chamber is achieved using thin films of iridium insulated with LPCVD thin films. Helium leak tests on these structures show leak rates below $10^{-9}$ cm$^3$/sec.

## Introduction

The application of silicon micromachining techniques to sensors and actuators for use in health care has become widespread as the structures developed offer smaller size, better performance, lower cost, and improved reliability. One of the most challenging areas in many implantable transducers has been the transfer of data and power into/out of the body. Although hard-wired systems have been generally used, they impose many problems on the overall structure in terms of: 1) hermetic encapsulation at the entrance and exit points of the output leads; 2) tethering effects due to the output lead(s); 3) infection at the site of the wires breaking the skin barrier; and 4) low yield and labor intensive assembly and packaging. In many applications, transmission of power and data using radio-frequency (RF) telemetry has been employed with success [1]. However, many of these previous systems utilized implantable antennae that were typically very large (in the few centimeter range) in order to generate enough power. It is believed that before many of the transducers developed for biomedical applications can be practically utilized, they need to be miniaturized while eliminating all external leads to the implant. Therefore, micro-miniaure transducers that can be powered and controlled using RF telemetry and combine the sensor/actuator, electronics, antenna, and power storage units into a very small area (a few square millimeters) are needed.

One such device being developed by our group is a single-channel implantable microstimulator for neuromuscular stimulation shown in Figure 1. It consists of: 1) a silicon substrate which incorporates the stimulating electrode(s), as well as CMOS control and bipolar power regulation circuitry; 2) a custom-made glass capsule electrostatically bonded to the silicon carrier providing a hermetic package for all of the receiver-stimulator circuitry and hybrid elements; and 3) hybrid-connected chip capacitor(s) and receiving coil. The microstimulator should measure 1.8x1.8x9mm³ and should be powered and controlled using RF telemetry. It should be capable of delivering constant current pulses of 10mA amplitude for durations as long as 200μsec into loads of 800Ω or less. This requires the microstimulator to store at least 2μC of charge in a storage capacitor, and the on-chip power supply to be at least 8V in
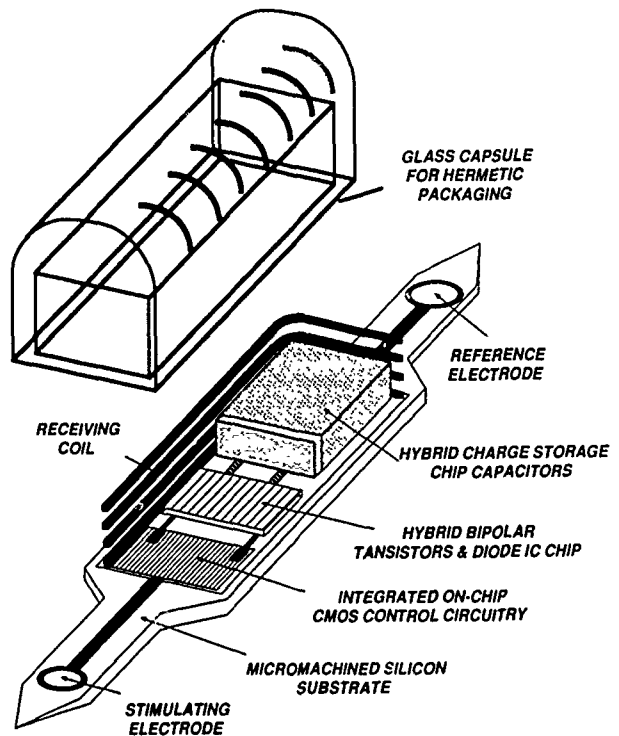


Figure 1: Structure of a single-channel implantable microstimulator for neuromuscular stimulation.

order to sustain the constant 10mA current across the equivalent 800Ω resistance offered by the tissue. Therefore, a large amount of power and high voltage amplitudes need to be induced at the receiver coil. In addition, because of the high signal levels involved, and because of the need for the long-term operation of this implant in corrosive biological environment, it is imperative that the entire stimulator electronics, as well as the receiver coil and capacitor be packaged in a hermetically-sealed chamber while allowing access to the stimulating electrode(s).

This paper reports the design and development of the transmitter and miniature receiver units for transmission of power and data (control) signals to the microstimulator, and the development of the hermetic package for its long-term encapsulation. The application of microtelemetry techniques to other sensors and actuators will become more realistic and attractive as receiving antennae become smaller. In addition, hermetic packaging of telemetry powered and controlled devices is very attractive for a variety of transducers that need to operate in vacuum, and/or need to operate in environments where physical access is difficult. Therefore, the results and techniques reported below should be applicable to a wide variety of microsensors and microactuators.

## Telemetry Powering and Control

### RF Telemetry Link

The RF telemetry link developed for the present application is based on two coaxial solenoid coils, as shown in Figure 2. The transmitter coil surrounds the receiver coil and sets up an electromagnetic field that induces sufficient power in the receiver coil. The most important requirement in this, and other similar applications, is the very small size of the receiver coil. Because of this restriction, the inductive link provides for a very small mutual inductance between the coils, i.e. the coils are weakly coupled and the mutual inductance is very low. The open-circuit voltage induced at the receiver, $V_r$, for such a system is obtained from [2]:

$$\frac{V_r}{V_t} = \frac{M}{L_t} Q_t Q_r \qquad (1)$$

where $V_t$ is the transmitter voltage, $M$ the mutual inductance, and $Q_t$ and $Q_r$ are the quality factors of the transmitter and receiver respectively. The mutual inductance $M$ is in turn given by [2]:

$$M = \frac{\pi}{2} N_t N_r \mu \frac{Y_r^2}{Y_t} \qquad (2)$$

where $N_t$ and $N_r$ are the number of turns in the transmitter and receiver coils, $\mu$ the permeability of the receiver coil, and $Y_r$ and $Y_t$ are the diameters of the transmitter and receiver coils respectively.

These analytical equations clearly indicate that for a given transmitter voltage and coil parameters, it is desirable that the receiver coil have: 1) a large number of turns; 2) as large of a diameter as possible; and 3) as high a $\mu$ as possible in order to maximize the induced voltage across it. In microtelemetry systems, however, the receiver coil diameter is forced to be very small. Therefore, a coil with large number of turns and high permeability (achieved by using ferrite cores) is needed. Eq. (1) also indicates that a large transmitter voltage is desired if a high voltage amplitude is to be induced at the receiver. This may be needed in some microactuators that operate using electrostatic forces generated by applying large dc voltages across air-gap capacitors. The design of a simple, high-efficiency transmitter, therefore, becomes an important goal for this and other applications.

### Transmitter Design

Since in most prostheses applications the devices involved usually operate from batteries, the efficient use of the battery energy is an important goal in the design of the transmitter. Therefore, the transmitter should have a high efficiency in converting dc battery energy into RF (ac) signal for powering and control of the transducer. Figure 2 shows the circuit diagram of a class E power amplifier [3] which has been selected for the transmitter. The active transistor acts as a switch when driven by an appropriate signal at the fundamental resonant frequency of the load network defined by the capacitor-inductor combination. As this switch is operated at this frequency, dc energy from the power supply through the RF choke inductor is converted to ac energy at the switching frequency. High efficiency is obtained by avoiding the imposition of simultaneous substantial voltage and current on the switch. This will result in very low power dissipation in the active switch, ensuring that all the power is transmitted to the output coil. A step-by-step procedure for the optimum design and implementation of the transmitter is given in [4].

The transmitter designed for this application employes elements with the following values: $L_1$=1.5mH with a Q of 43 at 1MHz; $L_2$=136$\mu$H, with an unloaded Q of 105 (loaded $Q_{tL}$ of ~20) measuring 9cm in diameter, 7cm long, with 41 turns; $C_1$=2nF, $C_2$=200pF, and a Motorola bipolar power transistor 2N5320. This transmitter operates from a single 12V rechargeable battery and generates 1.5W of power at a center frequency of 1MHz. The measured transmitter efficiency is ~90%. The transmitter antenna voltage reaches 150V peak, while the peak current through the transmitter is 0.3A. Rechargeable Gel-Cel batteries are used to operate the transmitter, and are rated at 1.9A-Hours. The battery lifetime at this power level is approximately six hours. Figure 3
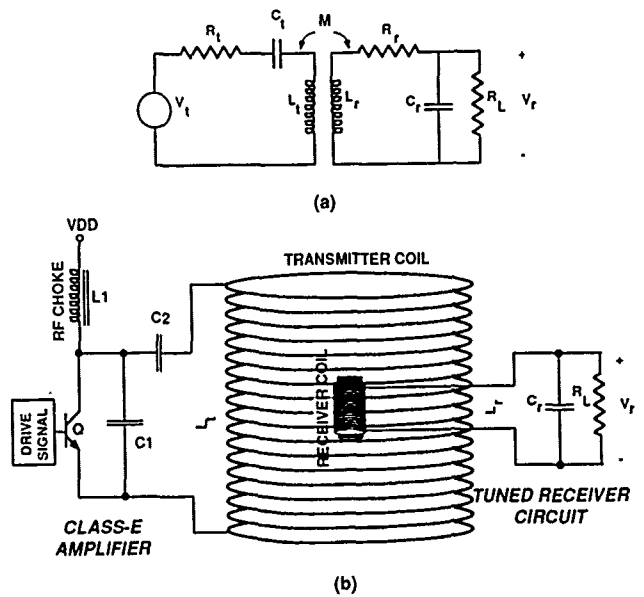


**(a)**



**(b)**

Figure 2. The RF telemetry link for the microstimulator, and circuit diagrams of the class-E transmitter and the receiver circuit.
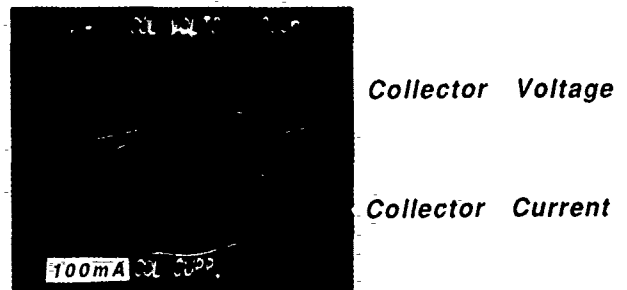


Figure 3: Collector voltage and collector current waveforms of the switch transistor.

shows the collector voltage and collector current of the transistor. It is seen that at peak collector voltage, the current is nearly zero, while at peak collector current, the voltage is equal to the saturation voltage of the transistor.

The above design for the transmitter results in a minimum number of components, delivers the high voltage and power levels required in this application, and provides the high efficiency that is so crucial in many battery-operated applications.

### Receiver Design and Tuning

As discussed above, the receiver coil has to be very small, while allowing sufficient power to be generated for the on-chip electronics and stimulator. A ferrite-cored solenoid coil has been selected for the receiver antenna as shown in Figure 4. The coil is ~5mm long, with a diameter of 1.5mm and includes 180 turns of 40 gauge wire resulting in a total inductance of ~75$\mu$H and a Q of ~60 at 1MHz. The receiver coil can pick up a peak-peak voltage of as high as 100 volts which is rectified and regulated to generate power for the implant. Figure 5(a) shows the unregulated receiver voltage, while Figure 5(b) shows the output voltage of the regulator as a function of output resistance on the regulator. In this application, a regulated supply of ~8V is provided resulting in a delivered power of ~30mW. Higher supply voltages have been produced by using a different design for the voltage regulator. The received power is stored in a miniature hybrid capacitor and is sufficient to drive the electronics and deliver current into tissue for stimulation. An important requirement for any RF powered device is minimum sensitivity to misalignments between the receiver and transmitter coils. The receiver voltage is proportional to $\cos\theta$, with $\theta$ being the misalignment angle between the two coils [5]. The telemetry link

146

described here is operational for misalignment angles as large as 50°. It should be noted, however, that misalignment sensitivity is a function of the received RF voltage and the desired regulated supply voltage. Table 1 summarizes the characteristics of the transmitter-receiver combination.

Table 1: Measured performance of transmitter and receiver circuits.

| PARAMETER | PERFORMANCE |
|---|---|
| Transmission Frequency | 1MHz |
| Transmitter Supply | 12V Battery |
| Transmitter Antenna Size | 7cmø x 9cm Long |
| Receiver Antenna Size | 1.5mmø x 5mm Long |
| Transmitted Power | 1.5W |
| Received Regulated Power | 30mW across 2.3kΩ |
| Peak-Peak Received Voltage | 50-60V across 8kΩ |
| Battery Lifetime | 6 Hours |
| Transmitter Range | 4-5cm |
| Max. Misalignment | 50° |

Since the telemetry link described here is based upon a double-tuned inductively-coupled system, provisions for tuning of the receiver coil should be made. The capacitance required to resonate the above coil at 1MHz is approximately 320pF. If slightly higher resonant frequencies are used, this capacitor can be reduced further in the range of tens of picofarad. This capacitor value can be easily achieved using on-chip MOS capacitors which will avoid the use of discrete capacitors and will allow tuning the receiver tank to the transmitter frequency thus ensuring maximum power transfer.

The above design for RF telemetry powering of miniature implanted units satisfies all the requirements for the present application. Most other applications require either lower power levels or lower voltage amplitudes. Therefore, implementation of other telemetry links for these applications is certainly possible, and the design parameters can be easily selected for a particular application.

RF Telemetry Control
An important aspect of the design of any telemetry link is the ability to transmit control (data) information to/from the implant in addition to power. The implant should be controlled using the same antenna in order to save space. Control signals can be superimposed on the RF carrier by properly modulating the carrier. In most applications, a serial stream of digital code is transmitted to either program the device, or to initiate proper actions. Digital control signals are encoded onto the carrier in the form of long and short pulses imposed on the RF carrier, corresponding to logical "1"s and "0"s. The on-chip electronics demodulates the carrier and generates a digital bit stream. Figure 6 shows a representative transmitted and decoded digital control code. The transmitter switch is driven with a frequency modulated drive signal. The load network consisting of the capacitor-inductor transmits only the signal at its resonant frequency, while inhibiting transmission of the signal away from its resonant frequency. The resultant transmitted signal is shown in the top trace of Figure 6. Driving the transmitter with a FM signal is crucial in obtaining high-speed modulation of the carrier. The lower trace in Fig 6 shows the demodulated signal which consists of long and short pulses corresponding to "1"s and "0"s. A number of remarks and observations should be made with regard to the data transfer technique presented here.

First, the receiver electronics designed to perform the demodulation is not optimized and is implemented using discrete elements which produce high capacitive loading thus effectively reducing the useful bandwidth of the telemetry link for information transfer. However, even with this preliminary design, it is seen that a bandwidth of 50kHz is easily obtained which is sufficient for most applications. Second, for many microactuator applications that operate using electrostatic voltages, a separate power supply is not needed since high-amplitude signals similar to those shown in Figure 5(a) can be directly used to operate the actuator. Third, we believe that with optimum design of the demodulation electronics, especially when integrated circuits are used, the data transfer rate can be improved by up to an order of magnitude. The design of such an IC for the microstimulator is underway. Fourth, for transducers that
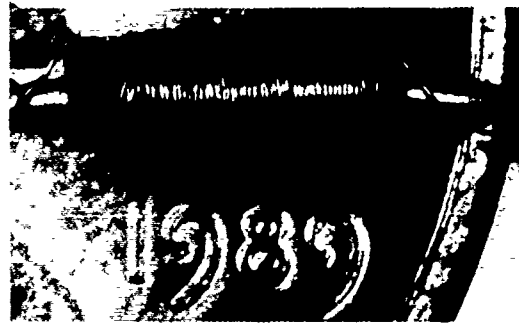


Figure 4: Photomicrograph of the receiver coil. The ferrite core measures 5mm long with a 1.5mm diameter, and includes a 3mm-long winding of 180 turns of 40 gauge wire.
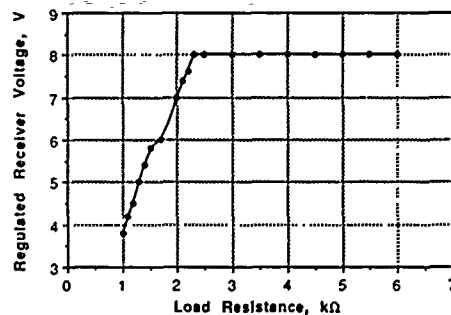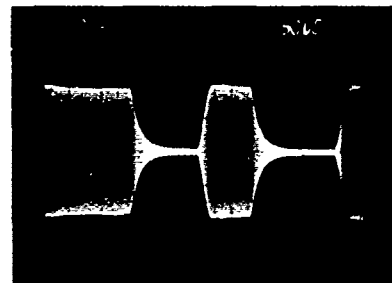


Figure 5: (a) Unregulated receiver signal across the coil, and (b) output voltage of the regulator as a function of output resistance.
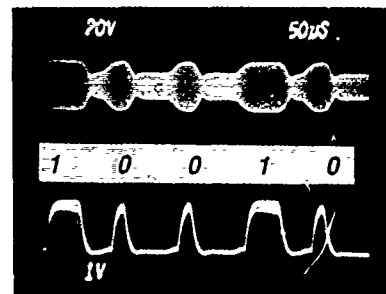


Figure 6. Transmitted modulated carrier signal and demodulated received signal consisting of long and short pulses corresponding to "1"s and "0"s.

need to transmit information back to the outside world, both active transmission using an on-chip transmitter, and passive telemetry can be used [6].

Hermetic Encapsulation

In order to ensure long-term operation of the implantable microstimulator and to avoid damage to the surrounding tissue, the implant should be encapsulated in a hermetically-sealed package. Conventional packaging techniques cannot be readily applied to this device since the package should not only be biocompatible but should also provide the necessary space for housing the receiver coil and storage capacitor. This has been achieved by utilizing a custom-made glass capsule that is electrostatically bonded to the silicon substrate.

147

## Package Fabrication

Figure 7 shows a SEM view of a custom-made glass capsule (#7740 Pyrex glass). The capsule measures $1.8 \times 1.8 \times 8.5 \text{mm}^3$ and is fabricated such that the bonding surface between the glass and the silicon substrate is polished to result in a high-quality seal Electrostatic bonding is used to hermetically seal the glass to the silicon substrate and is performed at temperatures of ~420°C with an applied voltage of ~1000V across the glass-silicon sandwich Bonds to both bare silicon and silicon covered with deposited LPCVD dielectrics have been made [7]. Figure 7 shows a capsule bonded to a silicon substrate, and the cross-section of a cleaved glass-silicon structure. Helium leak tests of glass capsules bonded to both bare silicon and silicon insulated with 1μm of LPCVD silicon dioxide and silicon nitride show a leak rate lower than $5 \times 10^{-9} \text{cm}^3/\text{sec}$ at a pressure of 1 atmoshphere. Since the inside volume of the capsule is approximately $10 \text{mm}^{-3}$, it will take about 2.5 years to completely penetrate the capsule at this rate. A nice feature of this technique is that the glass capsule is bonded only around the perimeter leaving enough room for the circuitry and the hybrid components

## Lead Transfer

An important requirement for any hermetically-sealed device is the ability to transfer leads into the sealed cavity without compromising the integrity of the seal. In this application, provisions for the interconnection of the stimulator electronics and the stimulating electrode(s) should be made. This is achieved by insulating a thin (<500Å) iridium conductor using LPCVD dielectrics, and by bonding the glass capsule directly over the conductor-dielectric combination, as shown in Figure 8. It has been shown that electrostatic glass-silicon bonds can cover over steps less than 500Å [7]. In order to reduce the total resistance of the interconnect, the iridium film need be thin only under the bond regions. Lead transfers into the glass chamber using this technique can be achieved without increasing the leak rate above levels mentioned before. In addition, since the interconnect is sandwiched between LPCVD dielectrics, excessive electrical leakage that has prevented the use of such techniques as diffused interconnects can be avoided. This technique of providing hermetic encapsulation is also very attractive for many sensors and micromechanical systems and actuators which need to be isolated from external contamination.

### Acknowledgements

The authors wish to thank Drs. F.T. Hambrecht and W.J. Heetderks of the National Institutes of Health for their encouragement and support throughout this project. This work is supported by the Neural Prosthesis Program, National Institutes of Health, under contract NIH-NINDS-N01-NS-8-2312.

### References

[1] B. Smith, P.H. Peckham, M.W. Keith, D.D. Roscoe, "An Externally Powered, Multichannel, Implantable Stimulator for Versatile Control of Paralyzed Muscle," *IEEE Trans. Biomed. Eng.,* Vol. BME-34, No. 7, pp. 499-508, July 1987

[2] W.J. Heetderks, "RF Powering of Millimeter and Submillimeter-Sized Neural Prosthetic Implants," *IEEE Trans. on Biomedical Eng.,* Bol. BME-35, pp. 323-327, May 1988

[3] N.O. Sokal, A.D. Sokal, "Class E- A New Class of High-Efficiency Tuned Single-Ended Switching Power Amplifiers," *IEEE J. Solid-State Circuits,* Vol. SC-10. No. 3, pp. 168-176, June 1975

[4] F.H. Raab, "Idealized Operation of the Class E Tuned Power Amplifier," *IEEE Trans. Circuits and Systems,* Vol. CAS-24, No. 12, pp. 725-735, December 1977

[5] R. Jacob, "Aerial Design in Inductive Loop Telemetry," *IEEE Trans. Biomed. Eng.,* Vol. BME-23, No. 2, pp. 118-123, March 1976

[6] B.C. Towe, "Passive Biotelemetry by Frequency Keying," *IEEE Trans. on Biomed. Eng.,* Vol. BME-33, No. 10, pp. 905-909, October 1986

[7] L.J. Spangler, K.D. Wise, "A Technology for High-Performance Single-Crystal Silicon-On-Insulator Transistors," *IEEE Electron Device Letters,* Vol. EDL-8, No. 4, pp. 137-139, April 1987
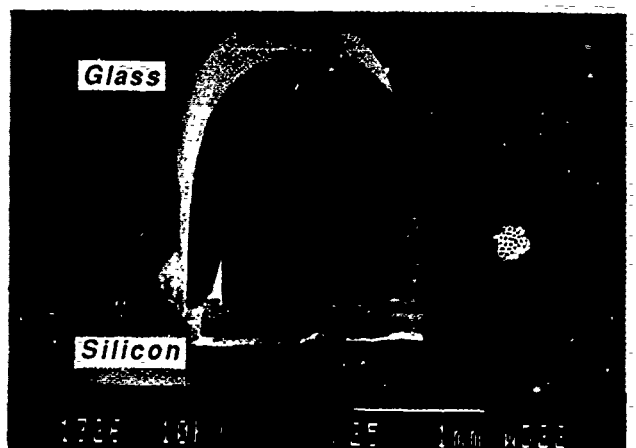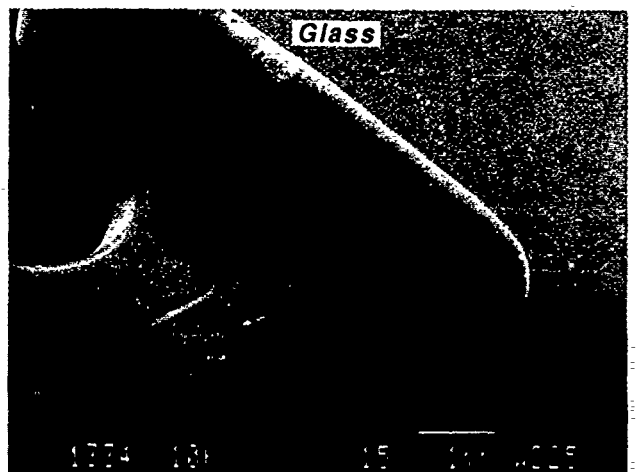
Figure 7. SEM views of the custom-made glass capsule and silicon-glass bonded structure. The capsule measures $1.8 \times 1.8 \times 8.5 \text{mm}^3$.
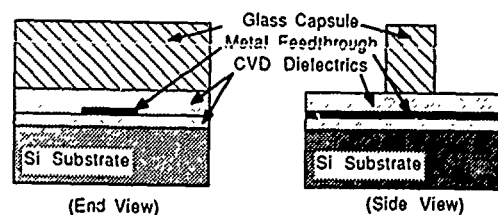


Figure 8: Cross-sectional views for lead transfer into the hermetically-sealed package for the microstimulator. The 500Å thick iridium conductor is insulated on top and bottom with LPCVD dielectrics and runs under the glass capsule which is electrostatically bonded to the top dielectric.

# Second-order Effects in Self-testable Accelerometers

Diederik W. de Bruin, Henry V. Allen, Stephen C. Terry
IC SENSORS
Milpitas, California 95035

## Abstract

Self-testability in accelerometers has application in measuring their sensitivity and frequency response during fabrication, without the need for applying mechanical vibration, and in verifying their operation and calibrating the sensitivity over temperature in the field. The self-test force in devices that have been reported to date has only been able to deflect the seismic mass to about 15 % of maximum range. The most effective way to increase the electrostatic force is to narrow the gap between the electrode on the mass and the over-force stops on the silicon top cap. However, this affects the operation of the sensor, both during regular use as well as in self-test mode. The device becomes more overdamped and the output signal may exhibit more non-linearity. These effects can be minimized by a careful design of the structure. Overall, minimizing the electrode gap in order to maximize the force is only recommended when the self-test feature is used for qualitative measurements where a large full scale deflection is necessary. For quantitative measurements, a spacing which produces a 10 % deflection seems a good compromise between reasonable electrostatic force and accuracy.

## Introduction

Self-testable sensors incorporate built-in actuators, which allows testing of a part over frequency and temperature during assembly or during operation. The concept of a self-testable accelerometer has previously been presented.[1,2,3] However, issues related to optimizing the electrostatic force and the influence of the self-test effect on the performance of the accelerometer have not yet been presented. This paper discusses some phenomena associated with increased self-test force.

The self-testable accelerometer is a modification of a piezoresistive, doubly supported cantilever structure with top and bottom caps as reported at this conference[4]. In order to incorporate the self-test function, the over-force stops which limit travel have been enlarged to form electrodes. The cross-section of the device is depicted in figure 1. A metal plate is deposited on the seismic mass and connected across the flexures to an outside bond pad. The top and bottom caps that provide protection against over-force and contamination are processed such that they are electrically connected to the substrate, thereby surrounding the self-test electrode in an electrically neutral environment.

The mass is deflected toward the top cap when a voltage difference is applied between the electrode and the substrate. The electrode is connected to the substrate through a reversed-biased diode in order to avoid damage by electrostatic discharge. In this device, p-type resistors are used, and the substrate must be biased to the most positive voltage on the chip. Consequently, the electrode voltage must be negative

with respect to the substrate. A unique feature of this improved structure over the previously reported devices is that no electrode is required on the cap.
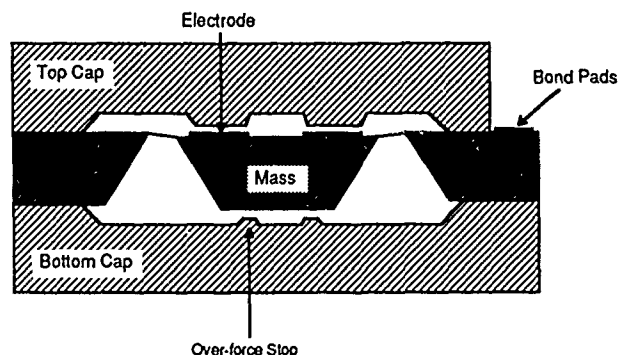


Fig. 1. Cross section of the self-testable accelerometer, showing an electrically isolated electrode on the mass and air-handling channels between electrode areas

## Increasing the electrostatic force

The devices reported to date have a 1 mV/g output for a 5 V drive, while the electrostatic deflection for a 20 V voltage between the electrode and substrate amounts to a force equivalent to 7.5 g. For certain applications, such as verification of maximum travel of the seismic mass, a larger force is desirable. It will be shown that modifying the structure to increase the self-test force is not without trade-offs.

The electrostatic force (in g's) is a function of the applied voltage V, the electrode area A and the gap spacing between the electrodes:

$$F_{el} = \frac{\varepsilon_0 \, A \, V^2}{19.6 \, m \, (x_0 + x)^2} \tag{1}$$

where $x_0$ is the gap at zero deflection and x is the deflection of the mass. Referring to figure 2, x > 0 for displacements away from the top cap and towards the bottom cap.
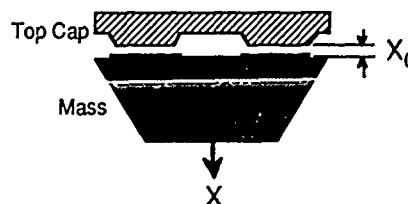


Fig. 2. Initial gap spacing $x_0$ and deflection x.

The electrostatic force is maximized if the electrode area occupies the entire surface of the mass with the exception of the area needed for the resistor metal interconnections.

Therefore, the area cannot be enlarged without increasing the size of the accelerometer, which has many disadvantages. Additionally, increasing the electrode voltage has more effect because the force increases with the square of the voltage. The maximum voltage that can be applied is, however, limited by the breakdown voltage of the dielectric layer under the electrode and of the diode which is used to protect the electrode from ESD. Also, it is often not convenient to generate a large (negative) voltage.

The force is inversely proportional to the electrode gap spacing. Thus, if the spacing is decreased, a larger electrostatic force is generated at the same applied electrode voltage. However, this reduction in spacing between the seismic mass and the over-force stops limits the maximum travel of the mass, so the gap must not be chosen smaller than the deflection of the mass at the maximum range. If the gap is decreased from 5 μm to 2 μm, the electrostatic force will be 6.25 times larger. However, the maximum deflection before the mass hits the over-force stops decreases from 100 g to 40 g. This will have the following implications.

*Regenerative feedback of the electrostatic force*

The restoring force of the silicon springs is proportional to the deflection of the mass, while the electrostatic force is inversely proportional to the square of the gap between the electrodes. At some electrode voltage, the restoring force is no longer able to counter the electrostatic force, and the mass will move until it rests against the over-force stops. This theoretically occurs when the mass has traveled $1/3$ of the gap spacing, if the restoring force of the springs is linear over displacement. Measurements indicate that the real value is closer to half the gap spacing due to the increase of the spring constant with increasing deflection. After the mass rests against the stops, the output of the accelerometer will be equal to its maximum value, and verification of full travel is thus possible. At that point, the electrode gap is determined only by the thickness of the dielectric layer covering the over-force stops.
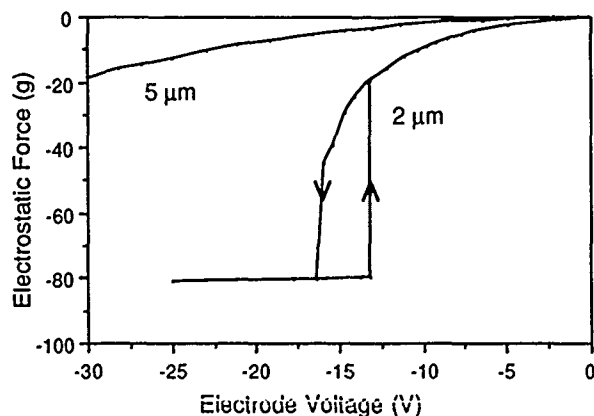


Fig. 3. Electrostatic force vs electrode voltage for two values of the initial gap spacing.

The holding force is consequently very large, and the mass will not release until the voltage is reduced to the point where the restoring force exceeds the electrostatic force again. Figure 3 shows the force as a function of electrode voltage for devices with two different electrode spacings. The 'release'

voltage is lower than the value needed to achieve maximum deflection, resulting in hysteresis in the force versus voltage curve. It would require a much larger voltage for a device with a standard 5 μm gap than for a device with a 2 μm gap to observe this effect.

*Non-linearity of the electrostatic force*

A deviation from an ideal parabola in the force versus voltage curve exists because the $(x_0 + x)^{-2}$ term in (1) cannot be simplified to $1/x_0$. This may effect the accuracy of the self-test function. The self-test function will generally be implemented by applying a voltage pulse to the electrode and measuring the change in output of the accelerometer. If the deflection of the seismic mass is small compared to the gap, the output of the sensor will be the sum of the signals due to the electrostatic force and the externally applied 'mechanical' g-force. Ideally, with an external g-force present, the output change upon applying the self-test voltage is proportional to the electrostatic force. For devices with narrower gaps, this is not necessarily true, because the gap can no longer be considered constant. If the deflection is not small compared to the initial gap spacing, the mechanical acceleration will modulate the electrostatic force through the $(x_0 + x)^{-2}$ term in (1). As a result, the sensitivity will increase with increasing electrode voltage. Figure 4 depicts the sensitivity to a signal with 1 g amplitude for different values of the electrode voltage. It can be seen that the sensitivity is modulated by the electrode voltage.
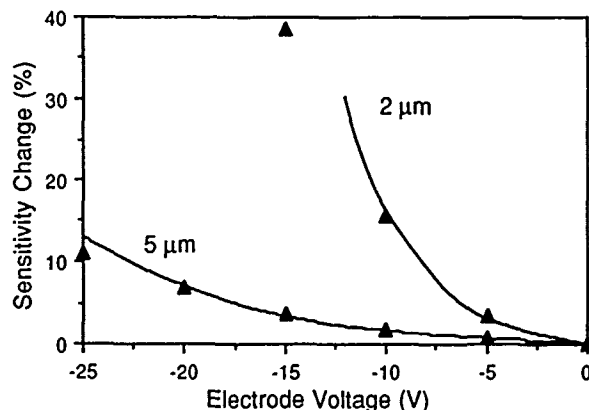


Fig. 4. Sensitivity increase vs electrode voltage for a 1 g small signal for two gap spacings.

The lines in figure 4 represent a simulation of this effect using a model with a linear spring constant, while the dots are measured values. The simple model is already in good agreement with the measured data.

*Increased Damping*

In the earlier versions of the accelerometer, cavities were etched in the top and bottom cap to obtain a critically damped system using squeeze-film damping. The damping ratio is inversely proportional to the cube of the depth of the cavities and is further determined by the stiffness of the silicon support of the seismic mass. For a critically damped sensor like the one being discussed, the gap is about 20 μm  Because the gap between the electrodes is smaller than the target cavity depth, the sensor will be overdamped  The damping

ratio increases with increasing electrode voltage because the gap becomes progressively more narrow. This is particularly apparent for devices with a 2 μm spacing (see figure 5). The damping can be improved by etching air-handling channels in the electrode. This way, the electrode looks like a plurality of stops rather than a single mesa. The damping is then set by a contributions of the electrode areas, where the gap is small, and of the air handling channels. The object in the design is to place multiple channels in the structure while maximizing the electrode area. The ideal pattern might look like a waffle.
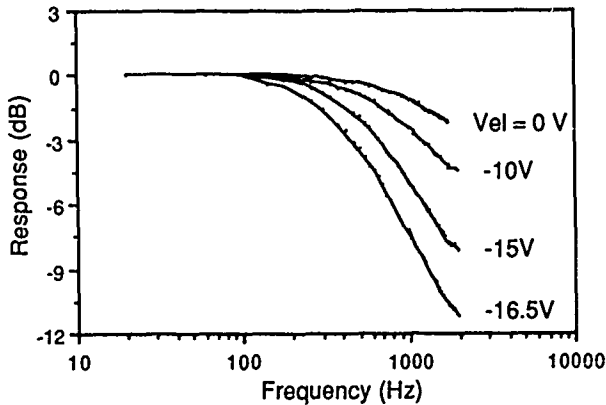


Fig. 5. Frequency response at different values of the electrode voltage.

Note that this effect allows for electronic tuning of the frequency response. As the magnitude of the electrode voltage is increased, the response rolls off at a lower frequency. However, there will be disadvantages in adapting this as a simple filter, such as an implied change in sensitivity and increased non-linearity.

If acceleration needs to be measured while a voltage is applied to the self-test electrode, the non-linearity of the response will be more accentuated with increasing electrode voltage. This is caused by the combination of modulation of the electrostatic force by the movement of the mass and non-linear air damping when the gap approaches zero. Figure 6 shows the observed total harmonic distortion at the output of the sensor when subjected to a sinusoidal vibration with an amplitude of 5 gRMS.
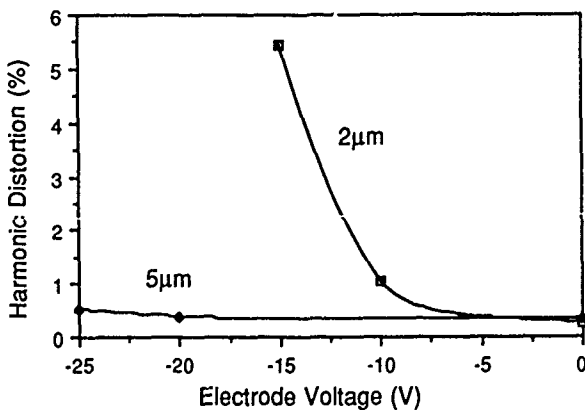


Fig. 6. Total harmonic distortion for a 1 g small signal as a function of electrode voltage.

The non-linearity increases rapidly with increasing electrode voltage in devices with a 2 μm gap, while it remains small in devices with 5 μm gap.

## Self-test during manufacturing

Besides verification of performance and calibration during operation, the self-test feature can also be use to screen a device during manufacturing. By applying a voltage at the electrode, the functionality of the sensor can be determined and a rough value for the sensitivity can be computed. Note that this allows mapping of the sensitivity across the wafer prior to dicing, a feature that is not available for piezo-resistive pressure sensors. In addition, it is possible to utilize sinusoidal control voltages of several frequencies and thus measure the frequency response without the need for mechanical vibration. If the electrode voltage is:

$$V_{el} = B + C \sin(\omega t) \qquad (2)$$

then the resulting electrostatic force follows from substituting $V_{el}$ in (1):

$$F_{el} = \frac{\varepsilon_0 \, A \, \alpha}{19.6 \, m \, (x_0 + x)^2} \qquad (3)$$

where $\alpha = B^2 + 0.5 \, C^2 + 2BC \sin(\omega t) - 0.5 \, C^2 \cos(2\omega t)$

If B=0, the output of the accelerometer contains only the second harmonic. The frequency response can be determined by measuring the output signals at the double frequency for several values of $\omega$. However, because of the diode used to protect the electrode from ESD, the sine wave cannot be centered around zero, and an offset B must be added to the electrode voltage. In that case, the response contains signals with both the fundamental and second harmonic frequency and it is necessary to use a frequency-discriminating AC voltmeter or spectrum analyzer to determine the frequency response.
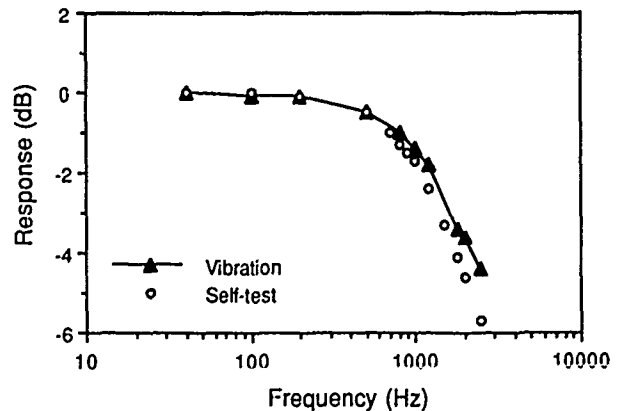


Fig. 7. Frequency response measured mechanically and using the self-test.

The accuracy of the measurement, shown in figure 7, is quite good. The electrode voltage is this case is:
$$V_{el} = -2.5 + 2.5 \sin(\omega t),$$

which is a good compromise between a minimum average force (which might influence the damping) and reasonable sine wave amplitude.

## Accuracy of electrostatic force

As reported before, the self-test feature can be used to calibrate the sensitivity of the device over temperature. This is done by periodically applying a voltage at the electrode and measuring the change in output voltage. Provided that the electrostatic force for that voltage is constant over temperature, the change in output voltage is proportional to the sensitivity. Using some electronics, the sensitivity can be temperature compensated without prior characterization or the need for individual compensation resistors. The residual sensitivity error depends on the variation of the electrostatic force over temperature, which is caused by temperature-dependent changes of the gap, mostly due to built-in stress. Using the self-test function for calibration at low values of the force is not optimum because of the low signal levels. At deflections that are not small compared to the initial gap, variations in the gap as a result of temperature-induced stresses have more influence on the force; thus the temperature dependence will increase again. For devices with a 5 μm gap, a sensitivity error of below 1 % over a temperature range of -25 °C to 75 °C is routinely achieved for an electrostatic force around 4 g (see figure 8).
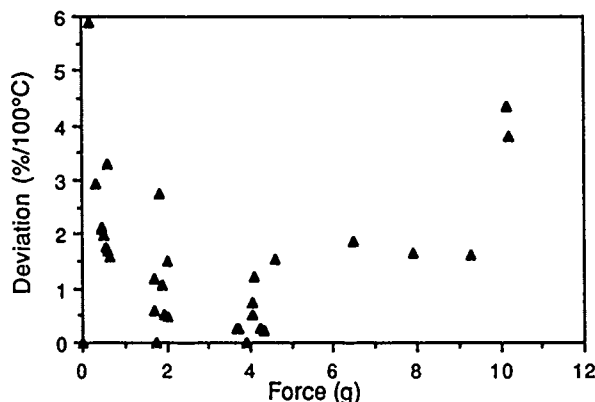


Fig. 8. Variation of the electrostatic force over a 100 °C temperature range when used for calibration of sensitivity.

## Conclusion

A number of conclusions can be drawn from the results presented above. First, breaking up the electrode into multiple sections separated by air handling channels is necessary to minimize the effect of viscous air damping between the plates. Second, when a self-test voltage is applied while acceleration is measured, the effective sensitivity increases with increasing electrode voltage as a result of modulation of the electrostatic force. Third, the self-test can only deflect the mass controllable to about half the gap, after which the mass is pulled against the stops when the restoring force of the flexures can no longer counter the electrostatic force. Fourth, the accuracy of the self-test effect, when used to calibrate sensitivity, deteriorates when the deflection is not a small fraction of the gap. When the self-test feature is used to calibrate the sensitivity over temperature, the electrostatic

deflection should not exceed about 10 % of the maximum range. This allows for the gap spacing to be chosen sufficiently large that the dynamic performance of the accelerometer is not compromised. Finally, only when full-scale electrostatic deflection is required at an electrode voltage below 20 V does the initial gap need to be reduced substantially.

## Acknowledgement

[1] Patent Pending

[2] Henry V. Allen, Stephen C. Terry, Diederik W. de Bruin, "Self-Testable Accelerometer Systems", Proceedings of IEEE Micro Electro Mechanical Systems, IEEE Catalog No. 89THO249-3, February 1989, pg 113-115.

[3] Henry V. Allen, Stephen C. Terry, Diederik W. de Bruin, "Accelerometer Systems with Built-in Testing", Abstracts of Transducers '89, Montreux, Switzerland, June 1989, pg 148-149. (to be published in Sensors and Actuators)

[4] Stephen C. Terry, "A Miniature Silicon Accelerometer with Built-in Damping", Technical Digest of IEEE Solid-State Sensor and Actuator Workshop, IEEE Catalog No. 88THO215-4, June 1988, pg 114-116.

# WIDE DYNAMIC RANGE DIRECT DIGITAL ACCELEROMETER

Widge Henrion, Len DiSanza, Matthew Ip*
Stephen Terry and Hal Jerman**

Triton Technolgies, Inc.* and IC Sensors, Inc.**

## ABSTRACT

Silicon micromachining techniques have been used to fabricate a high-precision, micro-g accelerometer. Operating in a closed loop configuration, the accelerometer utilizes electrostatic field sensing and electrostatic force feedback. The sensor assembly consists of an assembly of three silicon chips, bonded together at the wafer level. The center layer is comprised of the proof mass, springs and supporting structure. Electrochemical etching from both sides of the wafer is utilized to form a double-sided symmetrical structure which minimizes orthogonal sensitivity and harmonic distortion. The springs which support the mass are formed with a composite material to obtain near-zero net stress over the operating temperature range. The two outside silicon caps form a cavity for the mass and provide accurately spaced electrodes as well as over-force protection.

The micromachined sensor is operated in a vacuum to eliminate non-linear viscous damping and to provide a high-Q second-order mechanical resonant circuit. Near critical damping is provided by the closed loop control system. The control system is a highly over-sampled sigma-delta modulator, which produces a wide dynamic range and a direct digital output. The second-order spring-mass system with a high mechanical Q provides the integration for the sigma-delta modulator. Noise shaping of the modulator allows for a dynamic range from micro g's to the g-range, while producing extremely low total harmonic distortion. The single-bit output is decimated by an 8,000-gate, two-stage digital filter designed specifically for the accelerometer and fabricated using 1.5 micron CMOS technology.

The paper will describe the micromachined 3.5 x 4.0 mm sensor chip, the "acceleration input-digital output" sigma-delta modulator and the finite element analysis of the mechanical structure. The performance obtained from prototype units will be presented.

## INTRODUCTION

The design of a micro-g accelerometer with a full scale input of 0.1 g, a dynamic range of 120 dB and total harmonic distortion of less than 0.1%, required a different design approach. To achieve the 120 dB dynamic range, it was assumed that a digital output would be required. Initial attempts at converting the output of a capacitive accelerometer to an analog frequency, and then converting the frequency to a digital signal, did not yield results that would meet the above specifications. The capacitive sensor, when operated at atmospheric pressure with the required narrow gaps, has non-linear viscous damping, over damping, and electrostatic force problems. To solve the viscous damping problems, a sensor operating in a vacuum was considered and eventually selected. Operation of the sensor in a vacuum results in a high Q resonant peak, which creates its own set of problems. All attempts to passively damp the sensor, without introducing distortion, failed. It was then decided that the high Q second-order spring-mass system could be substituted for the second-order transfer function which is required in a second-order feedback system. The electrostatic forces on the mass, rather than being a problem, are used as the feedback force. The sensing, rather than being capacitive, is accomplished by electrostatic field sensing. A sigma-delta modulator system was selected because of its wide dynamic range possibilities. The use of a digital sigma-delta modulator results in an all digital closed loop, force balance sensor.

Numerous problems associated with an open loop

sensor are solved by the closed loop approach. The digital nature of the system begins at the sensor itself. The sigma-delta modulator's only concern is whether the position of the proof mass is above or below its at-rest position, and not by how much. Therefore, this closed loop system has a digital form from the mechanical spring mass on through all of the electronics. Using the sensor in a closed loop feedback configuration constrains the proof mass to a position very near its at-rest position. The proof mass's total excursion is reduced by a factor equal to the open loop gain (in the case of a sigma-delta control system, the gain is signal dependent). The amount the mass position differs from it's at-rest position at the end of a sample period is carried forward to the next sample period.

The output of the sigma-delta modulator is a high speed serial bit stream. This serial bit stream is then converted to a binarily weighted sampled word by use of a digital decimation filter.

Once the sigma-delta system configuration was selected, the next step was to select a sensor design to meet the modulator requirements and specifications. The decimation filter, as well as the sensor and the sigma-delta modulator will be described in the following sections.

## SENSOR DESIGN

The accelerometer consists of a 500 micron thick, <100> lightly doped single crystalline silicon (SCS) spring-mass layer, sandwiched between two identical material and thickness SCS layers. Except for the seal interlevel bond areas, a gap of 1.7 microns separates the top cover layer and the middle mass layer. The same gap also appears between the bottom layer and the middle mass layer. In order to allow for the sag of the proof mass due to the earth's gravity, a depression is etched into the bottom layer and a mesa on the bottom of the top cap is necessary. The small gap is essential for a low-g, high sensitivity electrostatic acceleration sensor. Because of the closed loop environment, the full scale travel of the proof mass is limited to only a small fraction of the gap. A cross sectional view of the sensor construction is illustrated in Figure 1.
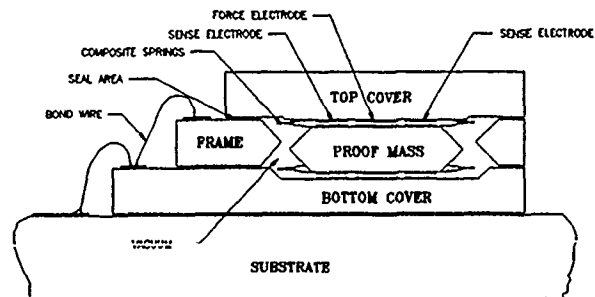


Figure 1: Sensor Cross Sectional View

The proof mass is suspended evenly by 8 spring sections; four of the spring sections are attached to the top four corners of the mass and the remainder to the matching bottom corners. A picture of the spring mass layout can be seen in Figure 2. Each spring section, as shown in Figure 3, is equivalent to having a pair of double cantilever beams, joined together by a stiffener. The stiffener is used to prevent
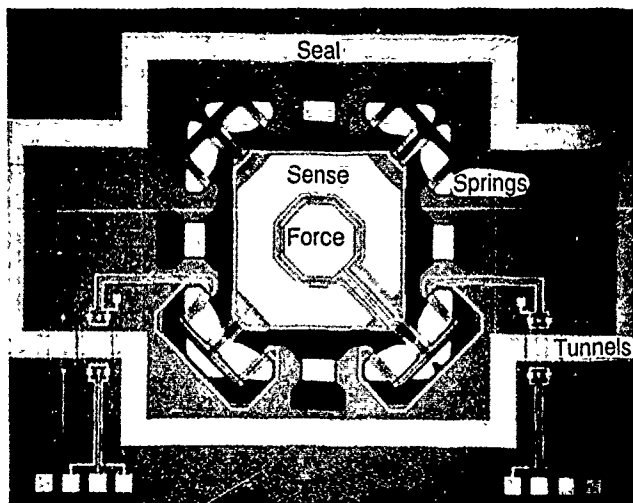
Figure 2: Accelerometer Spring Mass Layer

any torsional components of the spring from creating non-linearities. Each spring is composed of 200 micron long, 50 micron wide, and 1.3 micron thick undoped fine-grain polysilicon, silicon oxide, barrier metal, and gold. The present spring materials and their thickness and width ratios, were selected only after extensive research in processing techniques and finite element simulations. This is necessary in order to manufacture a flat spring using materials with different coefficients of thermal expansion. Flat springs are extremely important to ensure good open loop linearity and the correct stiffness. An SEM picture of the spring section can be seen in Figure 3. The proof mass is basically a 1000 micron square x 500 micron thick prism with the corners etched back slightly. The <111> plane etch slope from the surface to the middle of the mass increases the volume by 40%, producing a silicon mass of approximately 1.63 mg.
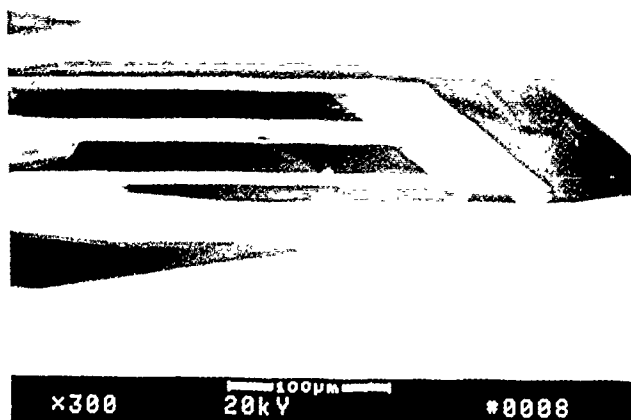


Figure 3: SEM of Accelerometer Spring

The displacement of the proof mass is controlled by a pair of concentric hexagonal-shape force and sense electrodes, located at the top and bottom surfaces of the mass. Identical patterns of the force and sense electrodes are located on the top and bottom covers, opposite their corresponding force and sense electrodes on the proof mass surfaces. The force and sense electrodes are thin metalized layers of gold, with leads connected to the gold of the composite springs and to the bonding pads outside the accelerometer. The areas of the sense and force electrodes are 548,000 and 127,000 square microns respectively. Surrounding the force and sense electrodes on the mass are guard rings and diffused guard plates.

The top and bottom covers are sealed to the middle spring-mass layer through a metal seal ring along the periphery of the sensor. Electrical conductors from the force and

sense electrodes, guards and substrates are brought out through diffused tunnels underneath the seal rings, and cn to the external bond pads. The tunnels are P-diffusions in N-tanks. Internal inter-layer electrical connections are made through bump bonds. All the bump bonds possess the same elevation as the seal such that the electrical connections are made at the same time as the seal bonds.

The open loop fundamental resonant frequency of the accelerometer design is 266 Hz. The closed loop cut-off frequency is in the kilohertz range. However, the low pass decimation filter has a cut-off of 200 Hz. Other vibratory modes of the sensor are much higher in frequency than the fundamental. Figure 4 shows the resonant frequencies of the first 3 modes of the sensor.
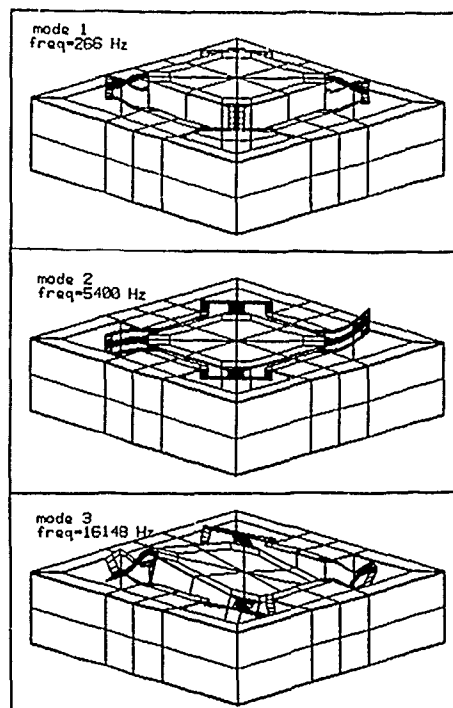


Figure 4: Sensor Modal Plot

The 8 spring sections on both the top and bottom of the proof mass not only provide the most balanced and linear spring-mass system, they also provide the rigidity to resist cross axis excitation and rotation from the horizontal axis. Finite element analysis showed that the cross axis motion is less than .001% of the sensing axis sensitivity. The accelerometer is also thermally stable, there is no air present inside the sensor to affect the damping due to temperature fluctuations. Both static and dynamic temperature analysis were performed to ensure no thermal mismatches would cause buckling in the springs. Since the composite cantilever springs themselves are free to change length, slight mismatch in their coefficients of thermal expansion with the surrounding material will not create any unwanted thermal stresses due to temperature fluctuations.

The overall sensor dimension is about 3.5 mm x 4.0 mm, it is mounted on to a ceramic substrate inside a 24-pin hybrid package. The units built for testing have the sensing and control loop circuitry and the decimation filter chip mounted externally.

## SENSOR PROCESSING

One of the unique features of this accelerometer structure is the complete front-to-back symmetry of the etched silicon proof mass and springs. This symmetry is achieved by performing each process step, including lithographies, implants, diffusions, metalizations, and silicon etches,

154

simultaneously on both surfaces. Including all of the cap and mass processing, there are 27 photolithography steps performed on 5 of 6 wafer surfaces in the 3-wafer assembly.

The center wafer is processed by first implanting (on both sides) and diffusing n-type tanks into the lightly doped p-type wafer. Into the tanks, p-type resistors are diffused to form tunnels and guard regions under the seals. These diffusions are followed by the deposition and patterning of low-stress polysilicon regions which serve as part of the springs. After the passivation of the poly, the metal electrodes and interconnects are deposited and patterned. Following the metalization steps, regions for the silicon etch are opened on both sides of the wafer. Note that since the etch proceeds from both sides of the wafer simultaneously, the metal must survive the entire length of the silicon etch. An anisotropic electrochemical etch is employed to undercut the springs and their supporting structure. The potential of the n-tanks and the p-substrate are controlled separately during the etch. The complex geometry of this device, as opposed to simple diaphragm structures employing electrochemical etch stops, required that the applied voltages be carefully controlled and optimized to result in the proper etched shapes. After etching half-way through the wafer, the etches meet, freeing the proof mass. After some residual passivation oxide is removed to insure flat and stress-free supports, the caps are aligned and bonded to the center wafer. The device is then sawn apart into individual die and the two levels of bonding pads exposed.

The fabrication of the cap wafers is somewhat simpler. The top cap is nominally flat, thermal grown silicon oxide is employed to adjust the gap between the top cap electrode and the proof mass with the 1-g sag. A vent hole is first etched through the cap to the central region of each device to provide a means by which the air surrounding the mass can be pumped out before use. The metal electrodes are then deposited and patterned. The bottom cap must allow for the 1-g sag of the mass, so a shallow depression, is formed before the electrode metal is deposited. A schematic cross section of the device is shown in Figure 1.

The accelerometer die are assembled onto ceramic substrates which are in turn mounted in metal packages. To reduce the number of connections to the outside, the sense and drive electrodes and the guards on the upper and lower surfaces of the mass are connected together on the ceramic substrate. The metal package is then welded to a metal cover in a high vacuum welder.

## SYSTEM

The major component blocks of the accelerometer system are shown in Figure 5. These components are: 1) the sensor, 2) the buffer amplifier and associated housekeeping circuitry, 3) the lead-lag network, 4) the quantizer, 5) the sampler, 6) the level shifter, and 7) the decimation filter.
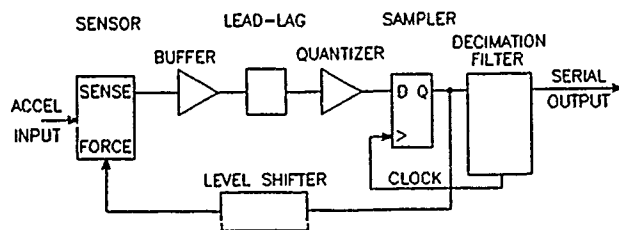


**Figure 5: Accelerometer Block Diagram**

The first block is the sensor and has been previously described. The second block is a buffer amplifier which provides the necessary loop gain and isolation between the sense electrodes and the rest of the circuitry. It also provides a low output impedance unity gain drive for the guard electrodes. The lead-lag network following the amplifier provides the compensation necessary to stabilize the loop. The next block, which is the quantizer, utilizes a comparator to determine if the input signal is above or below a reference level. It

provides an asynchronous digital output. The sampler converts the output of the quantizer to a synchronous digital signal. The level shifter in the feedback converts the digital pulses from the sampler to extremely accurate (in amplitude and time) pulses. These feedback pulses generate the force which is applied to the proof mass.

Finally, there is the decimation filter which is a custom digital signal processor. The purpose is to produce a 24 bit digital output word and to act as a brick wall low pass filter. This stage has many requirements, and some of them are: that it be able to run at the desired sample clock speed (up to 1MHz), that it have programmable decimation ratios, and that it produce output levels and format suitable for use with other processors.

## SENSOR

Force Feedback:
The closed loop sigma-delta converter requires a single bit digital to analog conversion of plus or minus one full scale feedback unit (shown as the level shifter in Figure 5). A unit in this case is a fixed amount of force which is equal to a multiple (generally about 2) of the full scale "g" range. The mechanical sensor converts either a fixed charge applied to the center mass electrode, or a fixed voltage applied between the top or bottom force electrodes and the center mass, to a plus or minus unit of force. The purpose of the feedback from the closed loop system is to always move the proof mass back to a position in the center of device. If a voltage feedback method is employed, then the center electrodes on the mass are grounded and a voltage is applied to either the top or bottom cap electrode while the other is grounded. The mass will always move toward the ungrounded electrode. The equation for the feedback force in terms of the physical parameters of the device and the terminal voltages is given by:

$$F = -\epsilon A v^2 / 2 x^2$$

where $e$ is the permittivity of free space, A is the area of the force electrodes, x is the distance of the mass to the fixed electrode on the cap and V is the applied feedback voltage. This is of course non-linear with respect to x for any given fixed V. The distortion produced by this non-linearity however can be made small in a closed loop feedback system where the change of x is very small compared to the gap.

An alternative where extremely low distortion is required is to use constant charge feedback. The expression for the force in this case is:

$$F = Q/(2\epsilon A)$$

where Q is a fixed packet of charge. The equation is linear and the force is not a function of the displacement x. The implementation of charge feedback is more difficult than voltage feedback, which has proven to be satisfactory for the required specifications.

Sense:
In order to sense the position of the proof mass, three sense electrodes are employed. Two of the electrodes are fixed and are on the top and bottom caps. The third electrode (which is actually two electrodes) is located on the mass and moves with it. A voltage V is applied between the cap electrodes. The mass or "center" electrode is connected through a switching arrangement to a high input impedance amplifier. The amplifier is essentially "floating" electrically for at least 50% of the time during a single control cycle (which is on the order of 2 microseconds for 512 kHz system clock). The total distance from the top cap electrode to the bottom cap electrode is fixed, and hence the electric field is fixed and uniform. If the proof mass is assumed to be conductive, then the equation for the electric field is given as:

$$E = V/(2 \cdot gap)$$

Then by definition, the voltage on the center electrode with

respect to ground at any position x (defined as zero at the bottom electrode and twice the gap at the top electrode) in the uniform field is given by:

$$V_{sense} = Ex$$

This simple relationship implies that the position of the mass can be found by monitoring the voltage on the sense electrode. In reality, the input impedance of the amplifier used to buffer the signal is finite, and hence, the signal will always be smaller than expected. This error is very small however and relatively static. A more serious problem with real devices, however, is that they will always accumulate charge over time. The concern here is that this charge build up will cause a time varying (non-static) error voltage. To address this issue, a "housekeeping" cycle is required. This cycle is performed as often as possible to prevent charge accumulation on the electrodes, switch, and amplifier. This step is also needed to prevent the center sense electrode from acquiring a charge and producing a force error. An added side benefit of this cycle however is that it actually helps to reduce sensor noise. The noise reduction is possible because most of the sensor noise is resistive in nature. This noise source resistance forms an RC time constant with the electrode capacitance, and thus the noise is reduced by the ratio of the housekeeping frequency to the RC product.

Guards:

As with any small signal level, high impedance transducer, good shielding and layout is mandatory. To this end great care has been exercised in the physical placement of all lines and electrostatic shields. Guard rings and diffused regions have been used throughout. The guards are actively driven wherever possible to help reduce stray capacitance and any adverse loading or coupling effects the stray capacitance might produce. In addition, charge injection compensated MOS switches have also been used to minimize errors.

## SIGMA-DELTA

The electrostatic spring-mass sensor is a mechanical low pass filter. The cut-off frequency (limit of the pass band) or resonance is given by:

$$\omega_n = \sqrt{(K/M)}$$

where K is the spring constant and M is the mass. $\omega_n$ is the natural resonant frequency in radians. The transfer function of the sensor is of the form:

$$H_{(s)} = C/[s^2 + s\omega_n/Q + \omega_n^2]$$

where C is a gain constant and Q is the quality or the reciprocal of the damping the mass encounters (quite small in a high vacuum).

This low pass characteristic of the sensor is used advantageously by the closed loop system. The sigma-delta converter produces a digital signal which is a representation of the analog acceleration input. The advantage of using a closed loop system is reduction of the non-linearities, resonant peaks, and unwanted electrostatic effects which would otherwise be produced by the device. The advantage of using a sigma-delta converter as the feedback method is that the output as well as internal signals are all digital.

In all analog to digital (A/D) converters, the analog signal (in this case acceleration) being represented has been "quantized", and can be no more accurate than one half the least significant bit (lsb) that the converter produces. Obviously then, the smaller the value of the lsb available, the finer is the resolution or accuracy (this determines the smallest analog signal which the converter can represent). Since it is also desirable to accurately represent large signals, these converters should work over as wide a dynamic range as possible. What this really means is that the converter should produce as large a digital word as is meaningful to the technique involved.

Another figure of merit of A/D's is their linearity, which, simply stated, means how the accuracy or scale factor of the A/D conversion varies throughout the dynamic range. Sigma-delta converters are extremely linear and can offer very large dynamic range with relatively simple circuitry. It is for all of these reasons that this method has been employed for use with the sensor.

A sigma-delta converter uses two techniques to reduce the quantization noise and thus decrease the minimum detectable signal. The first assumes that the total noise power is constant and that it has a flat spectral distribution. In this case, the signal plus the noise are sampled at a rate much higher than the signal frequency, and by doing so, the noise is spread over a larger bandwidth. Since the noise power is constant and the noise bandwidth is increased, the magnitude of the noise in the baseband is reduced. Each doubling of the sampling frequency further reduces the noise by 3 dB (this is equivalent to 1/2 of a binary bit). The quantization noise of the sigma-delta modulator is much more aggresively reduced by the noise shaping characteristics of the system. Here the noise power is once again assumed to be constant but now it is spectrally shaped. If the frequencies of interest are below a cutoff frequency $\omega_n$, then the noise shaping is used to push the baseband noise down for all frequencies below $\omega_n$, and it rises at frequencies above $\omega_n$. The converted output signal is digital, and so digital filter techniques are employed to strip off the unwanted high frequency components, and hence most of the noise. The noise shaping results from the signal and the quantization noise having different transfer functions (see Figure 6). A low pass filter is used to perform the noise shaping.
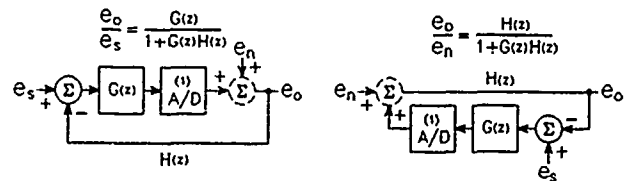


Figure 6: Signal and Noise Transfer Functions

The order of the filter (n) along with the sampling ratio determine the extent to which the in-band noise is reduced. The combined noise reduction in dB obtained by oversampling and noise shaping is $(3+6n)$ times the number of octaves of oversampling. (See Figure 7 for a typical sigma-delta output spectrum.)

Since the sensor is a second order low pass filter and is the noise shaping transfer function, the theoretical resolution of the converter will be $3+(6\cdot2)=15$ dB times the oversampling ratio. Therefore for eight octaves of oversampling the resolution will be 120 dB.
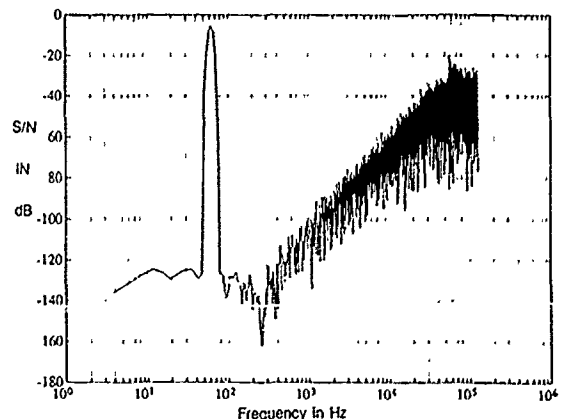


Figure 7: Sigma-Delta Spectrum Plot

## DECIMATION FILTER

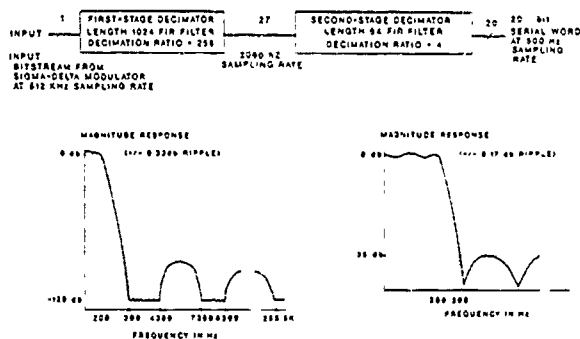The decimation filter block diagram is shown in Figure 8. It is a custom digital signal processor (DSP) which is

156

**Figure 8: Two Stage Decimation Filter**

used to reduce the long length, high rate serial bit stream that comes from the sigma-delta into a "decimated" 24 bit digital word. It is also used to strip off the high frequency noise by employing a brick wall low pass filter. The filter is realized in two programmable sections and allows for numerous decimation ratios of up to 1024. The output word is "burst" at a high rate along with a frame synchronization signal so that up to 32 accelerometers can be time division multiplexed on a single line. The output format and levels are directly compatible with most DSP chips and also easily interfaced to most computers including PC's. In a typical application, the computer or DSP chip would perform a Fast Fourier Transform (FFT), which would yield complete amplitude and frequency information of the acceleration being analyzed.

## RESULTS

Some experimental units were specially constructed to measure the electrostatic control of mass displacement by varying the amount of voltage on the bottom electrode. These units have a large etch hole on the top layer, so the vertical movement of the proof mass can be monitored through interference microscopy. In addition, the force and the sense electrodes were shorted together at the mass layer and the bottom layer, so a bigger electrostatic force can be produced when a voltage is applied. The top layer only served as a protective layer with no force or sense electrodes. When a DC voltage is applied at the bottom electrode, the mass will move according to the following relationship:

$$V = \sqrt{(2Kx^2/A\epsilon)(d-x)}$$

where K is spring constant, d is the original at-rest gap distance, x is the displaced distance between the two electrodes, and $\epsilon$ is permittivity of free space. A series of different voltages was applied to the bottom electrodes during the test and the displacement of the mass measured. The results of the experiment can be seen in Figure 9. The results show that the
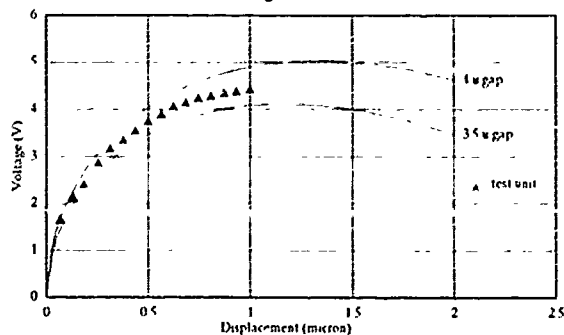


**Figure 9: Proof Mass Displacement vs. Voltage**

empirical data matches very well with the analytical formula with the gap approximately equal to 3.7 microns. The figure also shows that when the mass displacement reaches a point at which the rate of change of the electrostatic force exceeds the spring restoring force, the system becomes unstable. The

unstable point was experimentally observed when the mass displacement exceeded 1 micron.

A variety of different experimental sensor configurations have been examined. All were designed to be 0.1 g devices with a resonance of about 266 Hz. The sensors' characteristics were verified using a B&K 4809 shaker mounted on a double isolation table in an anechoic chamber (measured noise floor of less than $-160dBg/\sqrt{Hz}$). Electrical outputs were observed on an HP 3561 dynamic signal analyzer and an HP 3585A spectrum analyzer. The average resonance was about 350 Hz, the noise floor was well below $-100$ dBg-$/\sqrt{Hz}$ (no housekeeping or noise reduction techniques were used during those tests), the dynamic range was approaching 100 dBg, and all sensors had low pass characteristics for both mechanical and electrical drive. Sensitivity was hampered by stray wiring capacitance, but when corrected for the strays, the sensitivity approached 0 dBv/g.

## SUMMARY

A 0.1g full scale accelerometer designed to operate over a 120 dB dynamic range has been described. It is a direct digital sensor in the truest sense. The device was designed to withstand severe shock of over 700g's in any direction. New accelerometers are under development to complement the 0.1g unit. One of the new accelerometers will operate over a range of 10 micro-g's to 10g's. A low range unit is in development with a minimum detectable signals of $-160$ dBg (or 10 nano-g's).

For extremely low signal levels, quantization noise floors, circuit noises and mechanical noise (Browning noise) must all be considered in the design of accelerometers. Testing at these low levels of accelerations require anechoic chambers specially designed for low frequency noise rejection, with massive gas isolated tables supported on geologically stable bedrock.

Micromachined silicon accelerometers can be built to serve widely diverse markets at a reasonable cost. Frequency ranges from DC to thousands of hertz, accelerations from nano g's to hundred of g's, and dynamic ranges in excess of 120 dB can all be accommodated in small rugged silicon accelerometers.

## REFERENCES

[1] B.E. Boser and B.A. Wooley, "The Design of Sigma-Delta Modulation Analog-to-Digital Converters", IEEE Journal of Solid State Circuits, Vol. 23, No. 6, Dec. 1988.

[2] J.C. Candy, "A Use of Double Integration in Sigma-Delta Modulation", IEEE Trans. Commun., Vd. COM-33, pp. 249-258, Mar. 1985

[3] J.C. Candy, "Decimation for Sigma-Delta Modulation", IEEE Trans. Commun., Vol. Com-34, pp. 72-76, Jan. 1986.

[4] H. Inose, Y. Yasuda, and J. Marakami, "A Telemetering System by Code Modulation", IRE Trans. on Space Electronics and Telemetry, pp. 204-209, Sept. 1962.

[5] Martin A. Plonus, Applied Electro-Magnetics, McGraw-Hill, Inc., pp. 186-189, 1978.

[6] L.M. Roylance, J.B. Angell, "A Batch-Fabricated Silicon Accelerometer", IEEE Trans. on Electron Devices, Vol. Ed-26, No. 12, pp. 1911-1917, 1979.

[7] Felix Rudolf, Alain Jornod, Philip Bencze, "Silicon Microaccelerometer", Transducers '87, pp. 395-398, 1987.

# ENVIRONMENTALLY RUGGED, WIDE DYNAMIC RANGE MICROSTRUCTURE AIRFLOW SENSOR

T. R. Ohnstein, R. G. Johnson, R. E. Higashi, D. W. Burns,
J. O. Holmen, E. A. Satren and G. M. Johnson
Sensor and System Development Center
Honeywell, Inc.
Bloomington, Minnesota

R. E. Bicking and S. D. Johnson
MICRO SWITCH
Freeport, Illinois

## ABSTRACT

A silicon microstructure airflow sensor has been developed with a wide dynamic operating range and is rugged for long life operation in harsh environments. Platinum metallization is used for the airflow sensor because of its resistance to corrosion. Processing of thin film platinum has been developed to achieve a high, stable value of the platinum temperature coefficient of resistance (TCR). A first order coefficient for the platinum TCR of .003 ($\Omega/\Omega/°C$) has been achieved. The airflow sensor was designed with an exceptionally wide dynamic range, from less than 3 ft/min to over 30,000 ft/min (>10,000:1). The sensor design can be adjusted for sensitivity or range depending upon the application requirements. The airflow sensors were subjected to accelerated life testing to demonstrate the ability to maintain electrical stability and physical integrity in harsh environments. The life testing consisted of operation of the sensors with airflow in overpowered, high temperature and high humidity conditions for extended periods. The sensors were also subjected to extended periods of time during which dust particles were added to the flow to simulate a dusty environment. The sensors performed well throughout the accelerated life testing with little change in output characteristics.

Figure 1. Photomicrograph of the airflow sensor die.

## INTRODUCTION

A silicon microstructure airflow sensor has been developed for applications which require a wide dynamic response range and high volume flows. The sensor has also been designed for operation in dirty or corrosive environments where the gas flow environment cannot be filtered. This paper details the airflow sensor structure, performance and results of environmental testing.

## FLOW SENSOR STRUCTURE AND OPERATION

The airflow sensor consists of a thin film structure fabricated on a silicon substrate. Figure 1 shows a photograph of the airflow sensor chip. There are three temperature sensitive resistor elements on a dielectric diaphragm near the center of the chip. The diaphragm which can be seen as a dark square area on the chip is thermally isolated from the silicon substrate. There are two other resistors on the chip that are in thermal contact with the substrate for measuring the reference ambient temperature.

The photograph in Figure 2 is a closer view of the diaphragm area and the three resistor elements that make up the active part of the airflow sensor. A cross section of the airflow sensor through the diaphragm is shown in Figure 3. The diaphragm
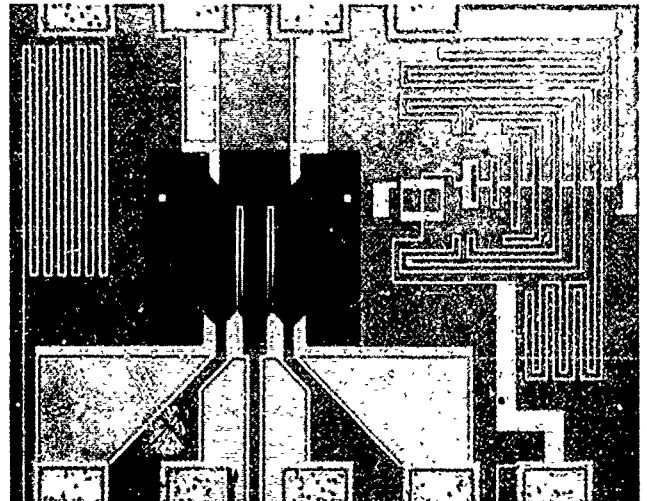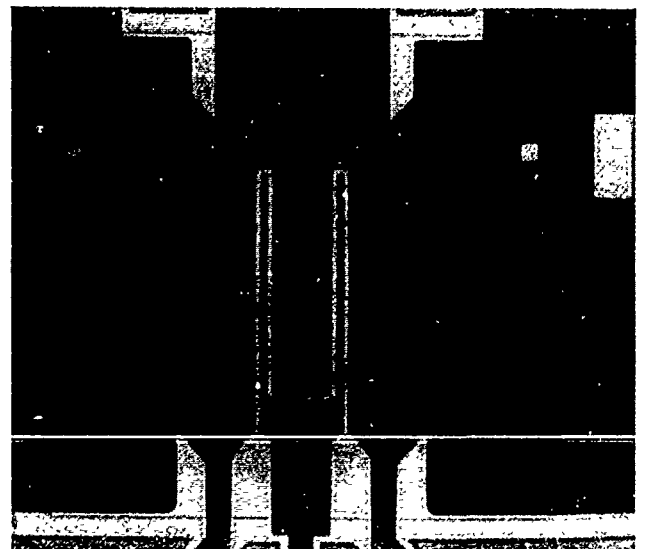


Figure 2. Photomicrograph of the airflow sensor diaphragm area showing the heater and sensor elements.

has a continuous closed surface with no large steps or exposed edges which make it resistant to impact damage or buildup of particles in dusty environments. Silicon nitride is used as the diaphragm material. The resistor metallization is completely passivated by the silicon nitride. The diaphragm is thermally isolated from the silicon substrate by anisotropically etching the silicon from the backside of the wafer.

A platinum metallization is used in the airflow sensor to provide stable resistors that are resistant to corrosion in harsh environments. A thin film platinum process has been developed for this sensor which has a first order temperature coefficient of resistance(TCR) of $3.0 \times 10^{-3}$ $(\Omega/\Omega/°C)$ and a second order TCR of $-5.0 \times 10^{-7}$ $(\Omega/\Omega/°C^2)$. The platinum metallization is passivated by silicon nitride and only gold is exposed at the bonding pads. However, in the event of scratches or pinholes in the nitride passivation, the platinum provides a metallization that is resistant to corrosion even if exposed.
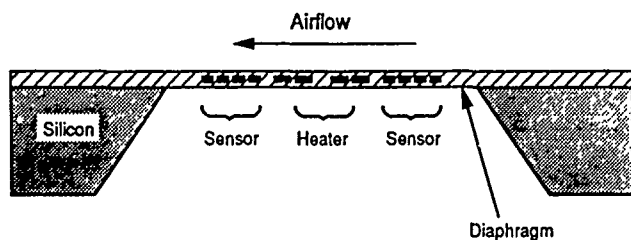


Figure 3. Cross section of the airflow sensor through the diaphragm area (not drawn to scale).

Referring again to Figures 2 and 3, the center resistor element is used as a heater to heat the diaphragm to a constant temperature above ambient during operation. The excellent thermal isolation of the diaphragm allows the operation of the heater at a temperature 160 °C above ambient with less than a 1 °C temperature rise on the sensor chip. Two sensing resistors are symmetrically positioned on either side of the heater along the axis of the airflow. Airflow is directed laterally across the chip as shown in Figure 3. The flow and direction of flow are sensed by the transfer of thermal energy primarily through the air from the upstream sensor to the downstream sensor. The airflow cools the upstream resistor and heats the downstream resistor. The flow detection circuit is a Wheatstone bridge where the upstream and downstream sensors make up two arms of the bridge. In a no-flow condition the sensors will be at the sa     ·perature and the bridge will be balanced with no outpu'     . A temperature difference between the sensors cause(     ı airflow unbalances the bridge circuit resulting in a vJltage output. The output voltage can be calibrated to flow velocity, volume flow or mass flow. A description of the operation of the airflow sensor in greater detail can be found in previously published work (1)-(3).

## PACKAGING AND TESTING

The response characteristics of the airflow sensor depend greatly upon the packaging of the sensor die. The design of the flow channel in which the sensor is mounted is critical in determining the characteristics of the flow in the channel. It will govern whether the flow that the sensor sees is laminar or turbulent in the desired flow range. For the applications of this airflow sensor it was desired that the flow is in the laminar

flow regime over a large range. The flow channel was designed for laminar flow over the range from 0 - > 30,000 ft/min.

The sensor die was mounted on a ceramic thick film circuit substrate. The laminar flow channel was then mounted over the sensor. The dimensions of the flow channel were 0.1" wide, 0.025" high and 3.0" long. The response of the airflow sensor was tested by applying a calibrated flow through the laminar flow channel.
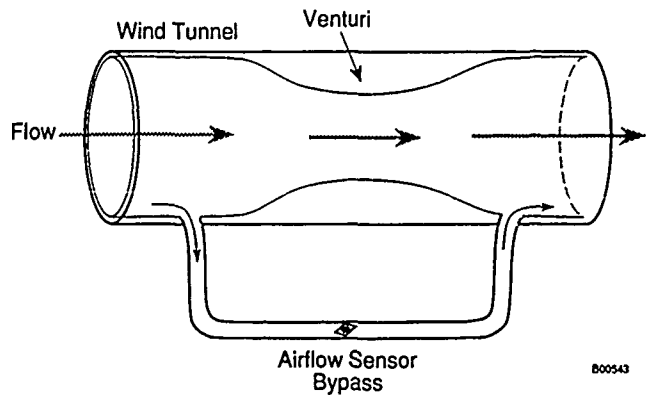


Figure 4. A section of the wind tunnel with the sensor connected as a bypass for measuring airflow.

To simulate larger volume flow applications, a wind tunnel was used for testing the airflow sensor. The airflow sensor in the laminar flow channel was mounted as a bypass element on the main duct of the wind tunnel similar to that illustrated in Figure 4. A venturi element in the main flow channel provides a pressure drop which drives the flow in the bypass. The response of the airflow sensor was calibrated to the flow in the main channel.

## EXPERIMENTAL RESULTS

The flow response of the airflow sensor was measured in the laminar flow tube and is shown in Figure 5. The curve is a plot of the output voltage of the Wheatstone bridge circuit versus flow. The sensor has a wide dynamic range of response from less than 3 ft/min to over 30,000 ft/min. It has been found that the sensitivity and the range of the sensor response can be tailored to the application by varying the sensor chip design. The sensor response shown in Figure 5 is from a device that was designed for a large operating range.

An accelerated life test was performed on the airflow sensors which simulated 10 years of sensor use and is an indication of the stability of the sensor over time. For this test airflow was through the laminar flow channel directly. The heater was powered to a level 50% higher than in normal operation so that it was running at a higher than normal temperature. The ambient temperature of the airflow past the sensor was heated to 85°C with 85% RH. The test under flow conditions continued for at least 100 hours. The criteria for passage of the test was that the no-flow output or null voltage of the bridge circuit could shift by no more than 1mV and the output voltage could shift by no more than 5% of full scale at full flow from before the test to after the test. The sensors were within these specifications after 400 hours of testing.

To simulate operation of the sensor in a dusty environment over a 10 year time period, an accelerated life dust test was performed. The airflow sensor was mounted as a bypass element on the wind tunnel as described above and illustrated in Figure 4. The sensor response was calibrated to the flow in the wind tunnel and the output voltage of the Wheatstone bridge circuit was amplified to give a maximum output of 5 volts full scale. Standard Arizona "coarse" road dust was used for this test. Arizona "coarse" road dust is a standard mixture of dust particles that is used in environmental testing. The "coarse" name denotes a mixture of particles ranging in size from <5 microns to 200 microns. The test was over a 15 hour time period where 100 grams of the dust was evenly mixed and added to the airflow in the wind tunnel. The airflow was periodically cycled between 10% and 100% of full flow throughout the test. Full flow in the wind tunnel was approximately 16,000 ft/min. The flow response of the airflow



Figure 7. A plot of the output voltage shift after the dust test. The data is plotted as the percentage change in voltage from before the test to after the test.
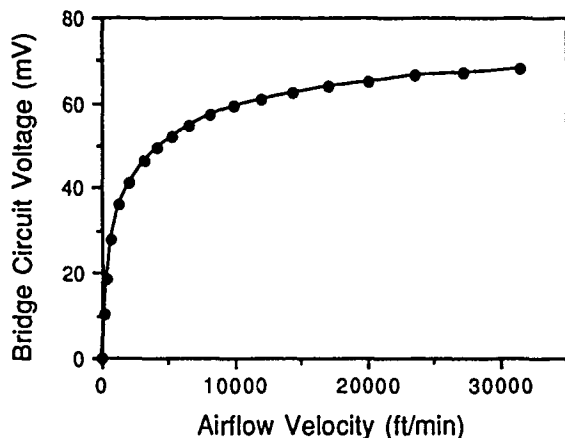


Figure 5. A plot of airflow velocity versus the bridge circuit output voltage measured in the laminar flow channel.

sensor was measured before and after the dust test. The results are shown in Figure 6. There was very little change in the curve after the dust test. The percentage change in the output voltage is plotted in Figure 7. Over most of the range the change was less than 1%. The bypass mounting of the sensor chip gives some protection to the sensor from the dust in the main flow but it was evident that dust did pass through the sensor flow tube. There was however no significant accumulation of dust either on the sensor chip or in the flow channel.

## SUMMARY

A silicon microstructure airflow sensor has been developed that has a wide dynamic range of flow response. The sensor has been shown to be environmentally rugged, passing accelerated life tests for electrical stability and physical integrity. A stable thin film platinum process has been developed for the sensor metallization that is resistant to corrosion in harsh operating environments.

## REFERENCES

(1) G. Benjamin Hocker, "Solid State Sensor Research At Honeywell: Active Thin Films Plus Microstructures," Tech. Digest IEEE Solid-State Sensor Conference, Hilton Head Island, South Carolina, June 6-8, 1984, pp. 18-19.

(2) R. G. Johnson, R. E. Higashi, P. J. Bohrer and R. W. Gehman, "Design And Packaging Of A Highly Sensitive Microtransducer For Air Flow And Differential Pressure Sensing Applications," Proc. 1985 Int. Conf. Solid-State Sensors And Actuators (Transducers '85), Philadelphia, Pa., U.S.A., June 11-14, 1985, pp 358-360.

(3) R. G. Johnson and R. E. Higashi, "A Highly Sensitive Silicon Chip Microtransducer For Air Flow And Differential Pressure Sensing Applications," Sensors And Actuators, Vol. 11, 1987, pp. 63-72.
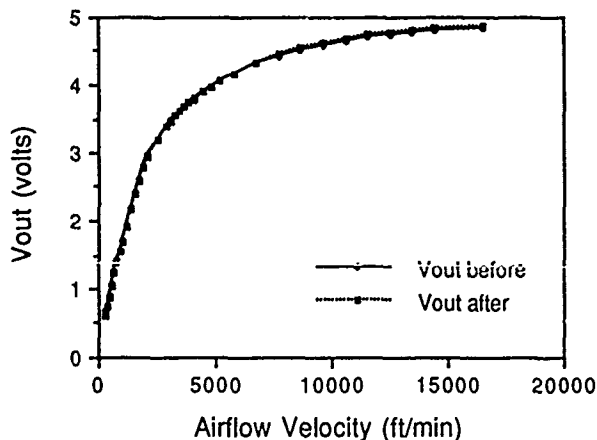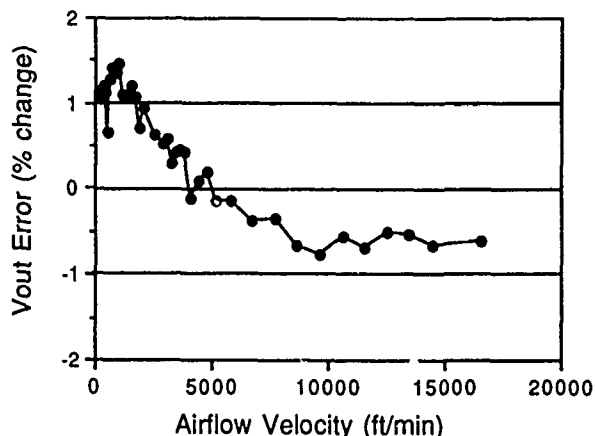
Figure 6. A plot of airflow velocity versus output voltage before and after dust testing. The data is for airflow in the wind tunnel. The output voltage of the bridge has been amplified.

# A MULTI-ELEMENT MONOLITHIC MASS FLOWMETER WITH ON-CHIP CMOS READOUT ELECTRONICS

Euisik Yoon and Kensall D. Wise

*Center for Integrated Sensors and Circuits*
Solid-State Electronics Laboratory
University of Michigan
Ann Arbor, MI 48109-2122

## ABSTRACT

This paper reports the implementation of control circuitry for a multi-element mass flow sensor capable of simultaneously measuring five different gas-related parameters necessary for mass flow. A process has been developed to merge the micromachined sensing structures and the CMOS readout circuitry on a single chip. The total process requires thirteen masks. eight for the double-poly single-metal CMOS process and five for additional processing of the transducer elements. The on-chip circuitry includes offset-free instrumentation amplifiers, an analog multiplexer, heater drive circuits, self-test circuitry, and a band-gap temperature sensor using substrate npn bipolar transistors. The resulting multi-element chip requires only ten external pins and delivers high-level buffered output signals describing gas velocity, type, direction, pressure, and temperature. The chip dissipates $120mW$ from $\pm 5V$ supplies and measures $3.5mm$ x $5mm$ in $3\mu m$ features.

## INTRODUCTION

During the last few years, one of the more important topics associated with the development of silicon sensors has been the possible integration of several different types of transducers along with appropriate interface circuitry on a common chip [1-4]. Recently, we reported a multi-element monolithic mass flow sensor capable of simultaneously measuring flow velocity, flow direction, gas type, pressure, and ambient temperature [5]. While each of these elements performs well individually, the resulting chip requires that ten separate transducers be sensed or controlled. Further optimization therefore requires on-board circuitry to boost and multiplex the transducer output signals and to reduce the number of external leads. In this paper, we report the realization of these on-chip interface electronics and discuss details of the circuit design and of the process modifications needed to allow on-chip circuitry. The resulting multi-element sensor illustrates a nearly generic structure, allowing the measurement of five independent variables with minimal cross-parameter sensitivity and with on-board signal conditioning. Our initial application for this device is in monitoring semiconductor process gases in the atmospheric pressure range, although a number of other applications are also possible. In monitoring process gases, the device, using its on-board electronics, functions as part of a VLSI sensing node, working over a standardized sensor bus with the system tool controller. Digital compensation of the device for offset, linearity, and slope errors is implemented at the controller level using coefficients uploaded from the node.

## MERGED PROCESS FLOW

Figure 1 shows a cross-section and top view of the multi-element flowmeter chip. The transducing structures are identical to those of the multi-element mass flow sensor chip previously reported [5]. Flow velocity, direction, and gas type sensors are supported on dielectric windows selectively etched in the silicon wafer. The four windows are located in a common well and are supported and thermally isolated using boron-diffused support beams. These beams are formed by the boron etch-stop when the silicon is selectively removed from the back of the wafer as the final step in wafer fabrication. The windows are $0.5mm$ x $0.5mm$ x $1.5\mu m$ in size and are formed using a layered silicon dioxide/silicon nitride/silicon dioxide sandwich which is
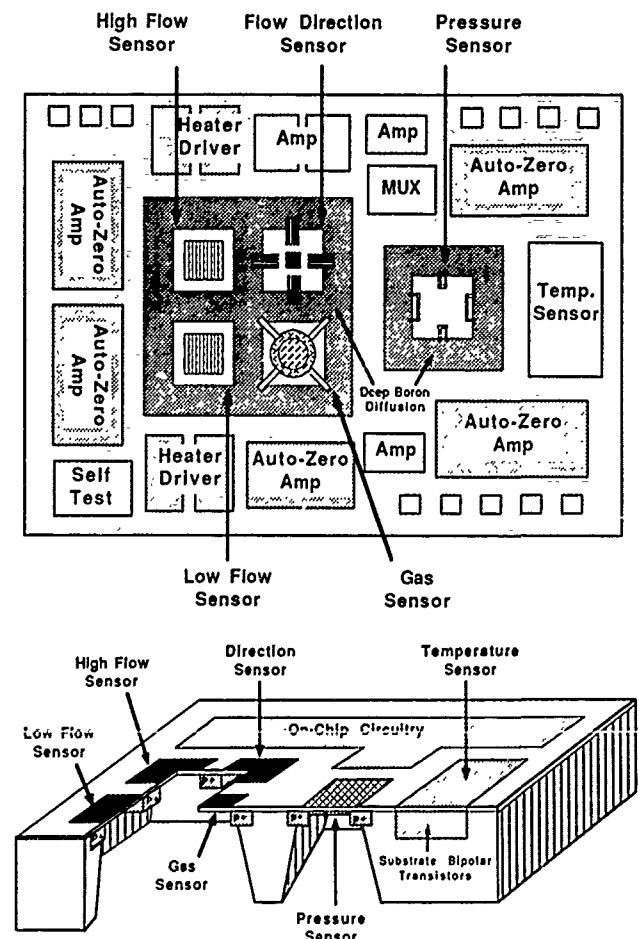


Fig. 1. Top View and Cross-Section of the Multi-Element Mass Flow Sensor with On-Chip Readout Electronics.

in mild tension. The two sensors for flow velocity use a thin gold/chromium ($260nm/40nm$) film as a temperature detector and polysilicon as a heater. Flow direction is measured using a third window, which contains two orthogonal pairs of differential polysilicon-gold thermopile detectors together with a polysilicon heater. The fourth window contains a conductivity cell for the measurement of gas type [6]. A piezoresistive polysilicon pressure sensor is located in a separate well for the measurement of ambient pressure.

On-chip sensing and control feedback circuitry is fabricated on the unetched bulk portion of the chip. A standard $3\mu m$ p-well CMOS process is used with modifications for compatibility with the sensor fabrication. Five additional masks are necessary in addition to the double-poly single-metal CMOS process itself (implemented in eight masks). These additional masks define the deep boron diffusion for front-side window definition, the shallow boron diffusion for the pressure sensor window, the thin film for the gas sensor, the dielectric cuts to remove the window dielectrics over the circuit area, and the backside silicon etch wells. The deep boron diffusion used for window definition is performed simultaneously with the p-well drive-in diffusion [7]. The depth of this diffusion is important in that it determines (along with the lateral dimension) both the strength of the window support structure and the effectiveness of the beams in preventing any thermal interactions (crosstalk) between windows. At a depth of $15\mu m$, the crosstalk between windows is less than 0.4 percent.

Figure 2 shows the merged process flow, which combines the standard CMOS and the micromachined diaphragm processes. Fabrication begins by implanting the p-well areas of the active devices at a typical dose of $2.0x10^{13}cm^{2}$ with boron at $150KeV$. A $1.5\mu m$-thick cap oxide is next deposited by LPCVD and opened to define
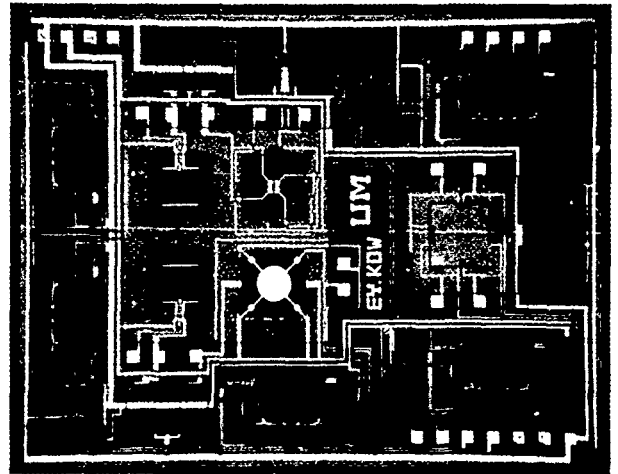


Fig. 3. Fabricated Multi-Element Flow Sensor Chip On-Chip control circuitry is included on the unetched bulk silicon region

the window rim areas. The deep boron diffusion is then performed at $1175°C$ for 16 hours, simultaneously driving the p-well and giving etch-stop and p-well depths of $15\mu m$ and $5\mu m$, respectively at the end of process. The pressure sensor diaphragm area is diffused next to produce a typical etch-stop thickness of $2\mu m$ This diffusion sets the sensitivity of the pressure sensor

At this point, a pad oxide and LPCVD nitride layer are formed, the field areas are implanted, and a $0.9\mu m$ field oxide is grown, after which the window dielectrics are deposited by LPCVD. Typical windows consist of silicon dioxide, silicon nitride, silicon dioxide composite films in thicknesses of $200nm$, $300nm$, and $800nm$, respectively. The temperatures used in the formation of these films ($820-920°C$) do not significantly alter the p-well or etch-stop profiles. The window dielectrics are removed over the circuit areas, and the rest of the process flow is identical with the standard CMOS process (including all channel implants). The silicon wafer is etched from the back side in ethylenediamine pyrocatechol to form the window/diaphragm wells (and die separation) as the final step in the process. Using this etchant, an alternative to aluminum metallization is needed; gold on chromium has been used here both for circuit interconnect and for thin-film temperature sensing ($3400ppm/°C$).

For this process, the sensor diffusions are performed first so that only the CMOS p-well diffusion needs to be adjusted as a result of the merged process. The more critical portions of the circuit process are unaffected up to final metal. Figure 3 shows a photograph of a fabricated flowmeter with on-chip readout electronics. The overall die size is $3.5mm$ x $5mm$ in $3\mu m$ features.

## DESIGN OF ON-CHIP CIRCUITRY

Figure 4 shows a block diagram for the on-chip readout circuitry. The circuitry requires as inputs $\pm5V$, GND, three multiplexer address bits (S0, S1, and S2), a self-test control signal ($\phi_{TEST}$), and a two-phase nonoverlapping clock ($\phi1$ and $\phi2$), and delivers a multiplexed high-level output. The nominal clock frequency is $14KHz$. Switched-capacitor offset-zeroing amplifiers are used where DC levels are important, while simple



**P-well Implant**

(100) n-type Substrate

**Deep Boron Diffusion and P-well Drive-In**

P+     P+     p-well
(100) n-type Substrate

**Dielectric Diaphragm**

P+     P+     p-well
(100) n-type Substrate

**CMOS Circuitry and Micromachining**
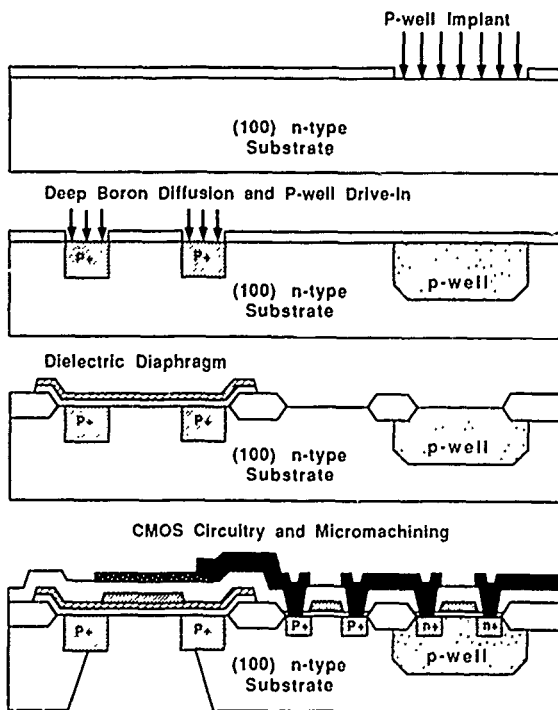
P+     P+     P+  P+  n+  n+  p-well
(100) n-type Substrate

Fig. 2: Detailed Flow Chart of the Merged Process which Combines the Standard CMOS and the Micromachined Diaphragm Process
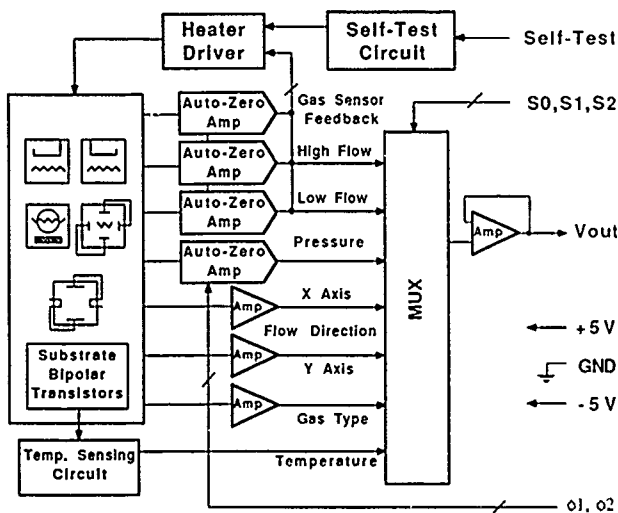
162

Fig. 4. Block Diagram of the On-Chip Readout Electronics

resistively-coupled amplifiers are used elsewhere (e.g., in the thermopile direction sensor). The thermally-based sensors (for flow velocity, flow direction, and gas type) operate in constant temperature mode, with temperatures set by on-chip feedback loops and heater drivers.

Figure 5 shows the self-test circuitry implemented in one of the velocity sensors to measure changes in the electrothermal characteristics of the diaphragm window *in situ*. With the self-test signal ($\phi_{TEST}$) off, the velocity sensor operates in the normal constant-temperature feedback loop for measuring flow. When $\phi_{TEST}$ is on, two additional circuit blocks are included in the feedback loop. These circuits form an electrothermal oscillator whose pulse duration is a function of the diaphragm time constant and the flow velocity (which continues to be monitored by the other velocity sensor during the self-testing mode). Figure 6 shows typical waveforms of the self-test outputs. The time durations $t_1$ and $t_2$ can be used to calculate the intrinsic thermal time constant ($\tau$) of the diaphragm sensor window given by:

$$\frac{L}{\tau} = \frac{\ln\xi}{t_2} \ [1 \ - \frac{P_{fc}}{P_e} \cdot \frac{(T_l - T_f)}{(T_v - T_f)} \cdot \frac{(\xi^{(1+t_r)} - 1)}{(\xi^{t_r} - 1)}]$$

where $\xi = \frac{T_h - T_f}{T_l - T_f}$, $t_r = \frac{t_1}{t_2}$,

$T_h$ and $T_l$ are the high and low switching temperature levels of the Schmitt Trigger, respectively, $T_f$ is the ambient flow temperature, and $P_e$ is the power supplied to the heater during the heating time ($t_1$) [5, 8]. $P_{fc}$ is the forced convection heat dissipation due to flow and is measured from the other velocity sensor which operates at the temperature $T_v$. This transient thermal response of the diaphragm can be used to detect the buildup of surface deposits that would lead to errors in the velocity measurements.

Substrate npn bipolar transistors are used to form a temperature sensor. Figure 7 shows the circuit schematic of the temperature sensor and the structure of the substrate transistor. The source/drain diffusion of the NMOS transistors is used as emitter, the p-well

region as base, and the n-substrate as collector. The Gummel plot of the substrate bipolar transistor is shown in Fig. 8. A typical common emitter current gain of 100 has been measured. The temperature sensor output response is given by:

$$\frac{dV_{out}}{dT} = \frac{R_2}{R_1} \cdot \frac{dV_l}{dT} \ lnA \ - \frac{dV_{BE}}{dT}$$

where $V_T$ is the thermal voltage ($kT/q$), $V_{BE}$ is the base-emitter voltage of bipolar transistor (typically, $dV_{BE}/dT \approx$ -2mV/°C), and $A$ is the ratio of emitter size between two bipolar transistors. The output response of the temperature sensor is designed to be +4mV/°C using the resistor ratio, which can be set precisely.

## RESULTS AND DISCUSSION

Table 1 summarizes the characteristics of the individual sensor elements including the temperature sensor realized with circuit components. Each of these transducers has been successfully implemented without any significant change in its characteristics due to the addition of the process steps necessary for the on-chip circuitry. The baseline threshold voltages are +0.8V for NMOS transistors and -0.8V for PMOS transistors, respectively. The total chip size is only 30 percent larger than the previously-reported multi-element chip [5] with
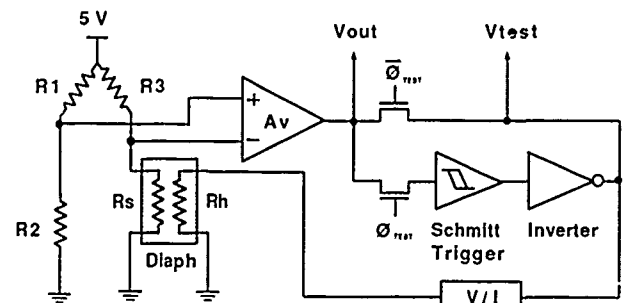


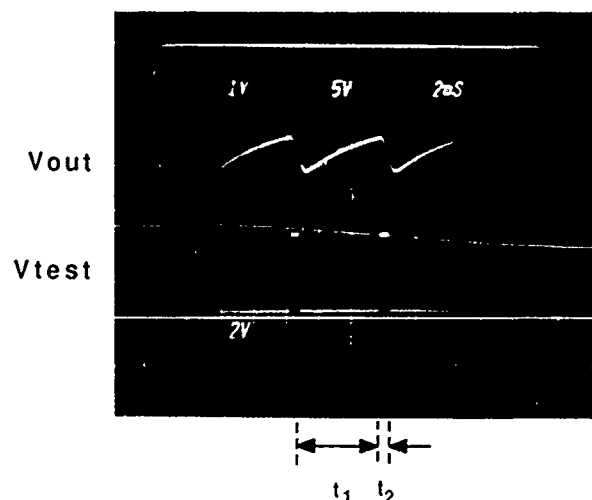Fig. 5. Schematic of the Self-Test Circuitry Implemented in One of the Flow Velocity Sensors.



Fig. 6 Typical Waveforms of the Self-Test Outputs. This transient response can be used to detect any changes in the electrothermal characteristics of the diaphragm sensor window.

163

no on-chip circuitry. The number of external leads required is reduced from 24 to 10 using the analog multiplexer, considerably simplifying the bonding and packaging of the device. The on-chip circuitry also improves noise immunity by amplifying the sensed signals by a factor of 20 before presenting them to the output pads. Application of the chip to the measurement for semiconductor process gas flows is underway.

## CONCLUSION

The thin diaphragm structure, responding to different pressure or to thermal events, is suitable for the measurement of a number of parameters including those involved in the determination of mass flow. This paper has reported the integration of sensing and control circuitry on the same chip with these transducers, preserving the performance of both the transducers and the circuitry and allowing functions such as amplification, multiplexing, and self-testing to be realized to improve system reliability. The resulting chip requires ten I/O pads, dissipates $120mW$ from dual $5V$ supplies and measures only $3.5mm$ x $5mm$ in $3\mu m$ features.

## ACKNOWLEDGMENTS

## REFERENCES

[1] D. L. Polla, R. S. Muller, and R. M. White, 'Integrated Multisensor Chip,' IEEE Electron Device Lett , vol. EDL-7, pp. 254-256, Apr. 1986

[2] O. Tabata, H. Inagaki, and I. Igarashi, 'Monolithic Pressure-Flow Sensor, IEEE Trans. Electron Devices, vol. ED-34, pp. 2456-2462, Dec 1987.

[3] D. L. Polla, H. Yoon, T. Tamagawa, and K. Voros, "Integration of Surface-Micromachined Zinc Oxide Sensors in n-well CMOS Technology,' in IEEE IEDM Tech. Digest, pp. 495-498, December 1989.

[4] E. S. Kim, R. S. Muller, and P. R. Gray, "Integrated Microphone with CMOS Circuits on a Single Chip," in IEEE IEDM Tech. Digest, pp. 880-883, Dec 1989.

[5] E. Yoon and K. D. Wise, 'A Dielectrically-Supported Multi-Element Mass Flow Sensor", in IEEE IEDM Tech Digest, pp. 670-673, Dec 1988

[6] C. L. Johnson K. D. Wise, and J. W. Schwank, 'A Thin-Film Gas Detector for Semiconductor Process Gases," in IEEE IEDM Tech Digest, pp 662-665, Dec. 1988.

[7] E. Yoon and K. D. Wise, 'A Monolithic RMS-DC Converter Using Planar Diaphragm Structures," in IEEE IEDM Tech Digest, pp 491-494, Dec 1989

[8] G. Stemme, 'A CMOS Integrated Silicon Gas-Flow Sensor with Pulse-Modulated Output, Sensors and Actuators, vol 14, pp 293-303, 1988
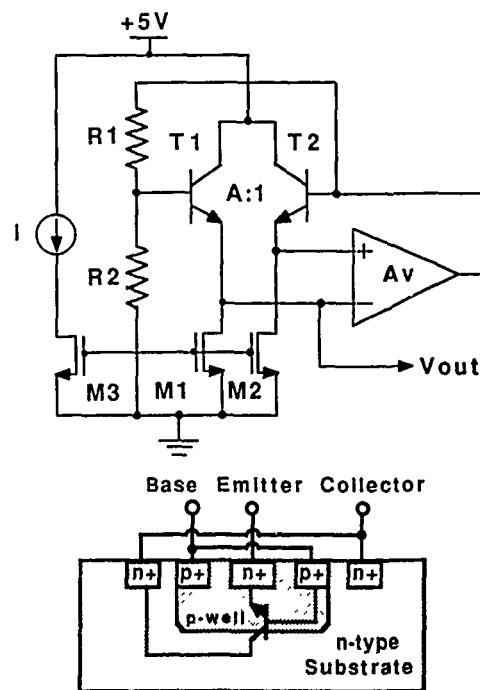
Fig. 7 Circuit Schematic of the Temperature Sensor and the Structure of the Substrate npn Bipolar Transistors Realized Using the p-well and the Source/Drain Diffused Layers in the Standard CMOS Process.
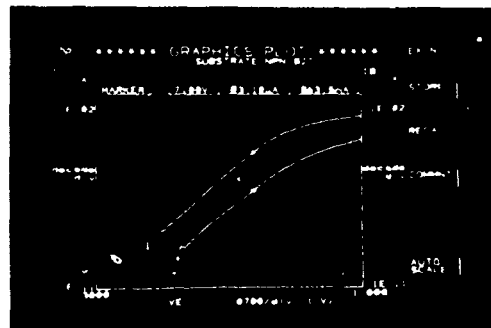


Fig. 8. Gummel Plot of the Substrate npn Bipolar Transistor.

Table 1

| Sensor | Range | Resolution | Sensitivity |
|---|---|---|---|
| Flow Velocity | 0 to 5 m/sec | 0.5 cm/sec | 34mV/V @5m/sec |
| Flow Direction | ± 180 ° | ± 5 ° | |
| Pressure | 0 to 800 mmHg | ± 0 5 mmHg | 12ppm/mmHg |
| Temperature | -55 to 125 °C | 0.1 °C | 4 mV/V²C |
| Gas Type | See reference [6] | | |

164

# INTEGRATED ION SENSORS: HOW MUCH MORE SHOULD BE DONE?

D. Jed Harrison,* Slobodan Petrović, Xizhong Li, Elisabeth M.J. Verpoorte, Alem Teclemariam, Alebachew Demoz

Department of Chemistry, University of Alberta

Edmonton, Alberta, Canada, T6G 2G2

The ion sensitive field effect transistor (ISFET) was first reported twenty years ago (1). Through the 1970's these devices were the subject of controversy regarding their mechanism of operation (2,3). During the 1980's they again became controversial, in regard to their ability to meet the early perceptions of their usefulness and practicality. The perceived advantages of integrating the chemical sensor have not been met in entirety, and the devices remain relatively difficult to fabricate. Many of the difficulties are due to practical impediments, and are not fundamental. However, there are questions of theoretical importance pertaining to stability that have not been addressed sufficiently, and several of the practical problems will likely be solved only by developing a deeper understanding of the devices. Commercialization of the devices will require creative approaches to the problems identified in the past ten years (2-4), and equally importantly may require improved interaction between semiconductor and electrochemical engineers within industry to bridge the chasm between these two fields.

This paper examines the endeavors of a number of academic labs to address the more serious difficulties associated with ISFETs. The focus is entirely on these devices, since most others, such as enzyme FETs, extended gate and diode devices, are all based on the same principles. A limited examination of industrial efforts will be presented verbally.

This discussion focuses on several of the more troublesome areas associated with ISFETs coated with ion sensitive membranes that have been identified and at least partially addressed. These include the problem of membrane adhesion, one of the problems identified relatively early on in ISFET development (5). The inherent interfacial stability of these devices is of obvious concern, but this point requires more attention. Finally, the problems associated with isolation and packaging of the devices will be reviewed.

## DISCUSSION

Chemical specificity for ions is often obtained by coating the ISFET gate with a polymeric layer. Typically, ion sensitive membranes used with conventional electrodes and based on poly-vinylchloride (PVC) or polysiloxane, a plasticizing agent, and an ion carrier have been used to coat the ISFET (2-5). Drift of these devices was addressed by Blackburn and Janata, who used a grid structure to mechanically adhere the membranes (5). Sibbald et al. later presented a simple lock and key arrangement at the edges that was claimed to improve adhesion (6). Photoresist materials with significant adhesion were evaluated by Wen et al. but these exhibited inferior electrochemical performance (7). The mechanical schemes make the assumption that the membranes are otherwise well suited to the membrane/solid interface.

In this laboratory we have used chemical modification of the membrane materials to improve their durability on solid surfaces. In Scheme I (8,9) the PVC matrix is made to bond to the surface and to itself, while Scheme II (10) results in cross-linking of the plasticizer without surface adhesion. Both of these methods were found to improve sensor lifetime and decrease drifting, implying that an increase in self-cohesion (strength) of the polymer matrix may be as beneficial as increased adhesion in enhancing performance. Further, in evaluating these materials we were led to postulate that stress develops in the polymers coated on solid surfaces in a manner different to, or less easily relieved than that when solution contacts both faces (10).

Recently, other approaches to membrane preparation have been reported that involve siloxane surface bond adhesion

enhancement, photopolymerizable polymers and reduced reliance on plasticizers (11-13). These can in some cases be used to photolithographically define the active sensor layers and indicate increased opportunities for simpler manufacturing. Complete evaluations of these systems do not appear to have been reported yet.
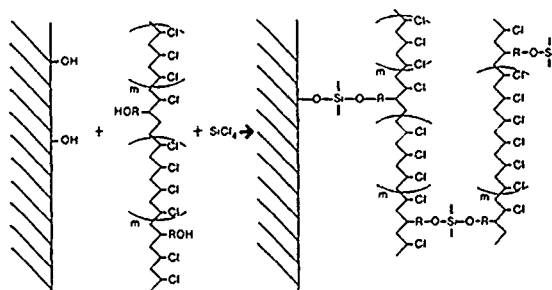


Scheme I
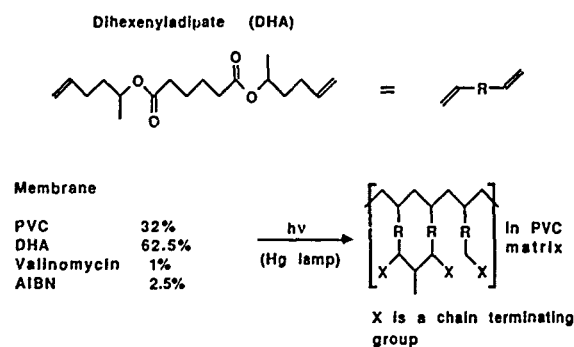


Scheme II

Recently, we developed an optical probe that is capable of measuring the concentration profile of $H_2O$ inside an ion-sensitive membrane (14). Figure 1 shows the absorbance due to
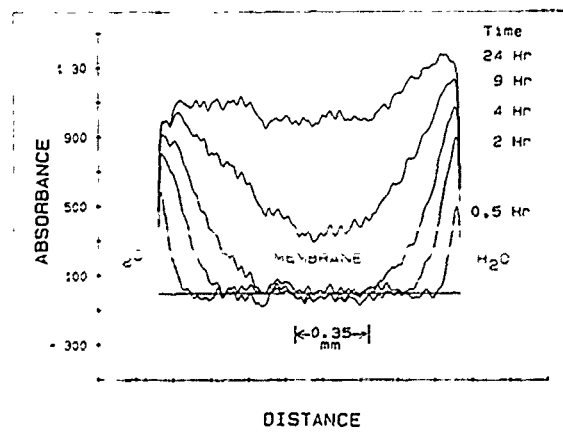


Figure 1

---

*Author to whom correspondence should be addressed.

H2O uptake in a PVC, diocatyladipate membrane when contacted at both faces with H2O. A uniform distribution is attained after ~24 h corresponding to a diffusion coefficient, D, of $3 \times 10^{-7}$ cm$^2$/s. When one of the solution phases is replaced by a solid material the uptake of H2O was considerably changed, as shown in Figure 2. Even after 111 h a non-uniform profile is observed, indicating a time dependent, spatially varying value of D. The result is provocative as it must arise from stress developed in the membrane. The non-uniform H2O distribution indicates a marked difference between solution and solid-contacted membranes, and provides some basis for understanding the increased rate of degradation of the latter structure. Further experimentation evaluating the extent and role of stress in the performance of these ISFET coatings is clearly warranted.
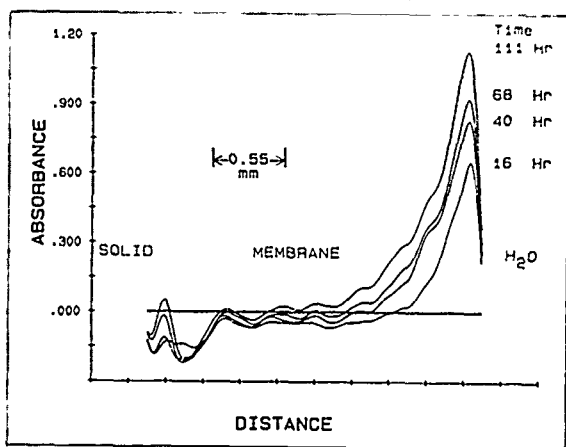


Figure 2

The theoretical basis of ISFET response has been carefully analyzed for the electrolyte/insulator/semiconductor junction, for which a site-binding model is appropriate (3,4). The model applies to pH sensitive inorganic coatings such as $Si_3N_4$, $Ta_2O_3$, etc. As an aside, efforts to prepare other site binding surfaces using a monolayer of a ligand such as a crown ether have seen only modest success, most likely due to the establishment of mixed potentials as has been discussed by Janata.

In contrast, for PVC membrane coated ISFETs there appears to have been be only two major attempts to analyze the potential developed by this interfacial structure. The analysis of Lauks unfortunately used an inappropriate space charge model for the membrane and so makes some dubious predictions (15). Hubert and Janata's analysis (2) is based on thermodynamic considerations, however, ultimately it is concluded that Nerstian response requires the internal membrane concentration of the target ion remain constant. This requirement is at odds with the known behavior of these membranes, which has been substantially elucidated since the latter analysis was presented (16).

The Donnan potential of a membrane will be rapidly established in the double layer region (~100 Å) inside the membrane, leading to the predicted potential change, however, transport within the membrane must also occur to establish the new equilibrium ion concentration. We have experimentally established the significance of transport for a Co-porphyrin based NO2$^-$ sensitive carrier (17). Figure 3 shows the distribution of NO2$^-$ across a membrane 24 h after a concentration change from 0 to 0.1 $M$ NO2$^-$ on the right hand side, as evidenced by the increase in absorbance of the Co-porphyrin complex at 544 nm. The concentration profile obeys Fick's laws of diffusion (smooth curve at 25 h), confirming the significance of diffusion based transport in the ion-sensitive membrane bulk. Clearly, over the short term there will be an internal region with a constant concentration (and so constant potential) but over time this must change. The standard cell potential of the sensor would consequently vary over time as a function of the electrode history.
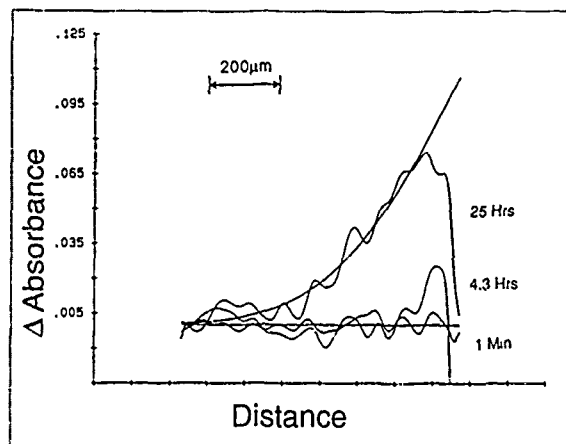


Figure 3

In addition to the source of drift discussed above Fogt et al. (18) have clearly identified the membrane-solid interface itself as a source of instability. Many solid surfaces bear acid-base functionality and it is possible for neutral species such as $CO_2$ or salicylic acid to permeate an ion-sensitive membrane, react with the surface acid-base sites, and change the charge state at this "blocked" interface, resulting in a spurious signal.

Both of the problems discussed point at the need for a properly defined internal reference material when making a solid membrane contact. For silver halide membranes this has been carefully studied by Buck (19). Rhodes (20) showed that a Ag/AgCl contact to K$^+$ sensitive PVC membranes showed less drift than a Si/SiO2 contact, but a buffered hydrogel was necessary as a contact to reduce the rift of H$^+$ sensitive PVC membranes. In our laboratory we have shown (21) that the chemical interferences of CO2 and other species can be eliminated via a Si/SiO2/Ag/AgCl contact, and the overall device drift is also reduced. Recent reports on ISFETs show there is now a much greater appreciation of the need to establish an internal potential (22). Several schemes involve the use of hydrogels, and the transport of H2O through PVC membranes means problems related to differential swelling and stress will need to be considered.

Understanding of the ion sensitive membrane-solid contact remains incomplete. One of the surprising effects is seen in the impedance behavior. A membrane contacted by two solutions shows impedance characteristics as a function of frequency that are independent of applied potential. In contrast, with a solid-state contact a fraction of the structures studied exhibit impedance that varies markedly with potential, as shown in Figure 4. These
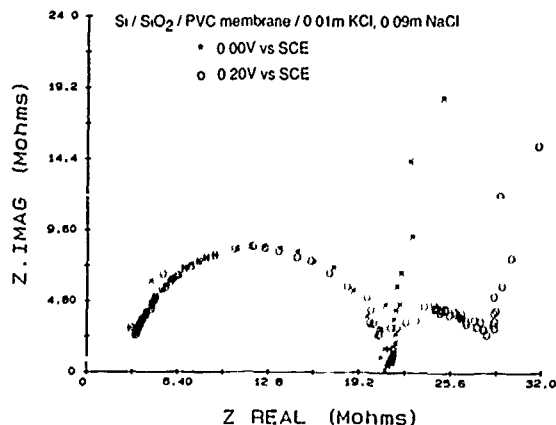


Figure 4

effects have not been examined in detail, but may arise from formation of a "surface film" akin to those Buck has discussed, or to some form of potential dependent shunt arising as an early sign of encapsulation failure. Possibly it is related to the potential dependent effects that Lauks predicted (15), although this seems less likely.

Bergveld has developed a unique application of the ISFET, integrating it with a coulostatic generating electrode for coulometric titration of the analytical reagent, and using the ISFET as the endpoint detector (23,24). This obviates problems with device drift, and means a pseudo-reference electrode (such as a metal wire) with a poorly defined potential is readily used. Coulometric titration is a well studied method, and a large range of electrochemically generated titrants have been developed, most of which could be utilized by this approach. The resulting devices represent an "end run" around many of the difficulties associated with the ISFET, and do not significantly change the fabrication methods.

The manufacture of ISFETs, in particular their encapsulation, appears to be one of the major difficulties in their large scale production and commercialization (2). It is ironic that low cost manufacturing by lithography could be one of the advantages of these devices, but has been a great hindrance. A number of approaches have been proposed to reduce the demands on the encapsulating material. Harame suggested the use of CMOS fabrication methods to isolate the FETs in a p-well from the n-substrate, preventing leakage currents through the Si bulk, and Smith (25) has presented data suggesting some improvement with this approach. Others have suggested silicon on sapphire technology (4) for the same reasons regarding isolation, while the concept of backside contacts to the device to eliminate lead wires on the side exposed to solution has also been pursued (4). Epoxies remain as the most common encapsulant for these ISFETs (26), although kaptan tape bonding has also been examined (2). Various epoxies are used with reported lifetimes ranging from days to several months. Several industrial labs have developed encapsulation methods that may be feasible, but data is not readily available on their efficacy. A number of concepts developed are appealing, but the net impact of utilizing the best, or most compatible of these ideas on the packaging problem is unclear. This is in part because comparisons of several technologies arising out of one lab, or several in combination, have not been reported. In fact a workshop focussing on this aspect would appear to be very useful.

## CONCLUSIONS

Research on ISFETs provides no lack of challenging problems. The basic concept of membrane coated devices is relatively easily demonstrated, however, practical devices have been more elusive. The increasing effort to address the more difficult issues is apparent, and has led to significant advances in improving device durability and ease of preparation. In essence the problem of potential drift comes down to how stable the device must be for a given application. Certainly the present ion-sensors are not drift-free. Similarly the acceptable cost of manufacture depends on the end use, and the present sensors are not cheap to prepare.

It is certainly the case that ISFETs can be made to work well when the proper care is taken. That level of care must be reduced to make the device more acceptable to those that presently hold the market-share in the medical applications of ion-sensors. The present state of the art suggests this can be accomplished. However, it must also be asked whether the extent of effort required would be better spent in this endeavor, or in exploring alternate, less developed integrated ion-sensing technologies. With so much effort invested the ISFET should decidedly not be abandoned. Further, any method that utilizes a solid/membrane contact will draw extensively on knowledge gained from the ISFET. Research efforts on integrated ion sensors do appear to need to broaden to some extent, to explore other approaches to planar, lithographically prepared sensors. These include diverse approaches such as the use of hybrid technologies with off-chip electronics (27), more extensive use of micromachining in conjunction with sensor design (25), and novel structures such as the chemical transistor (28).

## REFERENCES

(1) P. Bergveld, "Development of an Ion-Sensitive Solid-State Device for Neurophysiological Measurements," IEEE Trans. Biomed. Eng., 1970, BME-17, 70-71.

(2) J. Janata, "Solid State Chemical Sensors," J. Janata, R.J. Huber, Ed., Academic Press, London, 1985, Chapter 2.

(3) A. Sibbald, "Chemical-Sensitive Field-Effect Transistors," IEE Proc., 1983, 130, 233-244.

(4) P. Bergveld, A. Sibbald, "Comprehensive Analytical Chemistry, Vol. 23, Analytical and Biomedical Applications of Ion-Selective Field Effect Transistors," Elsevier, Amsterdam, 1988.

(5) G. Blackburn, J. Janata, "The Suspended Mesh Ion Selective Field Effect Transistor," J. Electrochem. Soc., 1982, 129, 2580-2584.

(6) A. Sibbald, P.D. Whalley, A.K. Covington, "A Miniature Flow-Through Cell with a Four-Function Chemfet Integrated Circuit for Simultaneous Measurements of Potassium, Hydrogen, Calcium and Sodium Ions," Anal. Chim. Acta, 1984, 159, 47-62.

(7) C.C. Wen, I. Lauks, J.N. Zemel, "Valinomycin-Doped Photoresist Layers for Potassium Ion Sensing," Thin Solid Films, 1980, 70, 333-340.

(8) T. Satchwill, D.J. Harrison, "Synthesis and Characterization of New Polyvinylchloride Membranes for Enhanced Adhesion on Electrode Surfaces," J. Electroanal. Chem., 1986, 202, 75-81.

(9) D.J. Harrison, L.L. Cunningham, X. Li, A. Teclemariam, D. Permann, "Enhanced Lifetime and Adhesion of $K^+$, $NH_4^+$ and $Ca^{2+}$ Sensitive Membranes on Solid Surfaces Using Hydroxyl Modified Polyvinylchloride Matrices," J. Electrochem. Soc., 1988, 135, 2473-2478.

(10) D.J. Harrison, A. Teclemariam, L.L. Cunningham, "Photopolymerization of Plasticizer in Ion-Sensitive Membranes on Solid-State Sensors," Anal. Chem., 1989, 61, 246-251.

(11) E.J.R. Sudhölter, P.D. Van Der Wal, M. Skowronska-Ptasinska, A. Van den Berg, D.N. Reinhoudt, "Ion-Sensing Using Chemically-Modified ISFET's," Sensors and Actuators, 1989, 17, 189-194.

(12) E.J.R. Battilotti, R. Colilli, I. Giannini, M. Giongo, "Ion-Selective Membrane Chemically Bound to a Field-Effect Transistor," Sensors and Actuators, 1989, 17, 209-215.

(13) R.W. Cattrall, P.J. Iles, I.C. Hamilton, "Photocured Polymers in Ion-Selective Electrode Membranes," Anal. Chim. Acta, 1985, 169, 403-406.

(14) X. Li, S. Petrović, D.J. Harrison, "A Novel Spectroscopic Method to Image $H_2O$ Distribution in Ion-Selective Membranes," Sensors and Actuators, 1990, B1, 275-280.

(15) I. Lauks, "Polarizable Electrodes," Sensors and Actuators, 1981, 1, 261-288.

(16) W.E. Morf, W. Simon, "Caton-Response Mechanism of Neutral Carrier Based Ion-Selective Electrode Membranes," Helv. Chim. Acta, 1986, 69, 1120-1131.

(17) X. Li, D.J. Harrison, manuscript in preparation, 1990.

(18) E.J. Fogt, D.F. Untereker, M.S. Norenberg, M.E. Meyerhoff, "Response of Ion-Selective Field Effect Transistors to Carbon Doxide and Organic Acids," Anal. Chem., 1985, 57, 1995-1998.

(19) R.P. Buck in "Ion Selective Electrode Methodology," Vol. I, A.K. Covington, Ed., CRC Press, Boca Raton, Florida, 1979, Chapter 9, pp 203-205.

(20) R.K. Rhodes, "Development of Low-Drift pH Electrodes Based on Neutral Carrier in PVC Overlayered Metallizations," IEEE Trans. Biomed. Eng., 1986, BME-33, 91-97.

(21) X. Li, E.M.J. Verpoorte, D.J. Harrison, "Elimination of Neutral Species Interference at the Ion-Sensitive

Membrane/Semiconductor Device Interface," Anal. Chem., **1988**, *60*, 493-498.

(22) P. Bergveld, "Sensors and Actuators, Twente," Sensors and Actuators, **1989**, *17*, 3-26.

(23) B. Van der Schoot, P. Bergveld, "An ISFET-Based Microlitre Titrator: Integration for a Chemical Sensor-Actuator System," Sensors and Actuators, **1985**, *8*, 11-22.

(24) W. Olthus, B.H. Van der Schoot, P. Bergveld, "A Dipstick Sensor for Coulometric Acid-Base Titrations," Sensors and Actuators, **1989**, *17*, 279-283.

(25) R.L. Smith, "An Integrated Sensor for Electrochemical Measurements," IEEE Trans. Biomed. Eng., **1986**, *BME-33*, 83-90.

(26) A. Grisel, C. Francis, E. Verney, G. Mondin, "Packaging Technologies for Integrated Electrochemical Sensors," Sensors and Actuators, **1989**, *17*, 285-295.

(27) R.E. Belford, R.G. Kelly, A.E. Owen in "Chemical Sensors," T.E. Edmonds, Ed., Blackie & Son, Glasgow, **1988**, Chapter 11, 236-255.

(28) D. Belanger, M.S. Wrighton, "Microelectrochemical Transistors Based on Electrostatic Binding of Electroactive Metal Complexes in Protonated Poly(4-vinylpyridine): Devices That Respond to Two Chemical Stimuli," Anal. Chem., **1987**, *59*, 1426-1432.

# Ion-Selective Sensors Incorporating Strongly Adhesive Polymeric Membranes

H. D. Goldberg, G. S. Cha, and R. B. Brown

Center for Integrated Sensors and Circuits
Solid-State Electronics Laboratory
University of Michigan, Ann Arbor, Michigan 48109-2122

## ABSTRACT

Chemical-selective membranes having matrices of PVC/(hydroxylated PVC), polyurethane/(hydroxylated PVC), and moisture-curable silicone rubber are compared to conventional PVC membranes in terms of adhesion to silicon nitride. A well-controlled peel test, developed for this evaluation, yields repeatable, quantitative results for both wet and dry membranes. Polyurethane and silicone membranes have much better adhesion to the sensor surface than do PVC or hydroxylated PVC. The hydroxylated PVC- and polyurethane-based membranes have electrochemical performance equivalent to that of PVC membranes in terms of slope, detection limit and selectivity. Though the electrochemical properties of the silicone-matrix membrane are degraded somewhat by its high resistance, it has extremely good adhesion.

## INTRODUCTION

Central to the design of longer-lifetime solid-state chemical sensors is development of polymeric membranes having better adhesion to the surface of the integrated sensor. Ionophore-doped polymeric membranes are used as transducers in these sensors because of their excellent selectivity toward the ion of interest, the wide range of ions for which ionophores are available, and because they can borrow from ongoing developments in conventional ion-selective electrode (ISE) technology. As in ISEs, poly(vinyl chloride) (PVC) is the most commonly used membrane matrix in solid-state ion-selective sensors.

The leading cause of failure in these microsensors has been electrolyte shunts around the membrane due to poor membrane adhesion [1, 2]. This problem has been addressed by others through modification of the PVC matrix for binding to hydroxyl-bearing surfaces, mechanical attachment of the membranes, and the use of other materials [3]. These methods usually achieve improved adhesion at the cost of more complex processing and/or degraded electrochemical performance as compared to conventional sensors. We have previously reported [3] improved adhesion of polyimide-matrix membranes; the electrochemical properties of these membranes were somewhat inferior to those of PVC-based membranes.

We have now optimized three other membrane compositions which incorporate:

- Poly(vinyl chloride) with the hydroxylated copolymer poly(vinyl chloride / vinyl acetate / vinyl alcohol) ((PVC)/(PVC/Ac/Al)),

- Polyurethane (PU) with the hydroxylated copolymer ((PU)/(PVC/Ac/Al)), and

- Moisture-curable silicone.

Though adhesion of membranes to the sensor surface is only one of the factors determining sensor lifetime, it has been the limiting factor in most sensors reported to date. Unfortunately, methods used for characterizing this important parameter have been imprecise and irreproducible.

The mechanical adhesion of a material to a substrate is a function of substrate surface morphology, temperature, interfacial electrostatic forces, and interfacial chemical bonding. Exact determination of the adhesive strength of a material to a given substrate is difficult because the material is usually deformed by the test procedure. Often, the best that can be achieved is to quantify the adhesion through a threshold test, or to measure a peel force which includes film thickness and elasticity, as well as adhesion. Common methods for adhesive evaluation of thin films are tape [4, 5], blister [5–7], scratch/scrape [4, 6], ultrasonic bath [2], and peel [4–7] tests.

The tape test involves cutting the film into a grid pattern with a knife, applying an adhesive tape over the film, peeling off the tape, and observing the percentage of squares that were removed from the substrate. This is a threshold test since the pulling force of the tape is assumed to be constant (roughly 6 pounds/inch$^2$ for PVC). This most popular test for chemical membrane adhesion has a number of drawbacks. Adhesion of the tape itself varies with membrane material; some polymers have very poor adhesion to the tape. This variation invalidates comparisons between types of membranes. Adhesion is also a function of several uncontrolled variables: preparation technique, angle of the tape, pulling force and pulling rate. The tape test cannot be used on wet membranes (the condition of greatest interest). And finally, results of the tape test are binary for all membranes except those having adhesion to the substrate which is similar to their adhesion to the tape. For example, tape test results for conventional PVC membranes and our three new membrane compositions were not very enlightening: all of the PVC and PVC/(PVC/Ac/Al) films pulled off of the substrate, while none of the PU or silicone films came off.

The blister test is implemented by forming a suspended membrane structure and applying an increasing pressure to the membrane backside until peeling from the substrate is initiated. Average work of adhesion can be calculated from spatial deflection, pressure, and area peeled [8]. The blister test yields quantitative results, but depends on operator judgment for accuracy. In addition, it requires rather involved processing to prepare samples for testing (back-side masking and micromachining of the silicon substrate). In sample preparation, membranes are exposed to the wafer etching process which can alter their characteristics before the test begins. Because membranes are deformed in this test, results are subject to membrane elasticity and thickness. Membranes having strong adhesion compared to cohesion cannot be tested with this method because they rupture before peeling begins. More complex structures have been developed to extend the range of measurements [8].

The scratch/scrape test is performed by contacting the film with a moving stylus and increasing the normal force until the substrate is seen. This test, too, gives quantitative results, but is dependent upon operator judgment. Observed adhesion will be a strong function of other membrane physical properties, and of membrane thickness. In some cases the stylus just crushes the film and the applied force is not correlated to adhesion [4].

The ultrasonic bath is also a threshold test since the forces induced on the membrane/substrate interface by ultrasonic stimulation fall within a narrow range. This method inherently convolves effects of soaking with forces to dislodge the membrane. When membranes have strong adhesion, the test is very time consuming, lasting days, during which time, the membranes must be checked at regular intervals for detachment. Unless its temperature is controlled, the solution is heated over time by the ultrasonic energy, increasing the effect that soaking has on results. In evaluating a variety of membrane materials in ultrasonic-bath accelerated lifetime tests, we have observed a correlation between adhesion time and compliance of membranes [3]. The ultrasonic vibrations are more effective at detaching rigid membranes which are unable to absorb the ultrasonic energy through elastic deformation. Comparisons of dissimilar membrane materials with this method suffer from a systematic error which favors softer materials.

Peel tests are used primarily to study adhesion of metal films to various surfaces. A machine similar to the tensilometers used for finding Young's modulus and other physical properties from stress/strain curves [9] is used to peel the film from the substrate while maintaining a constant pulling rate and angle. Physical attachment to metal film is usually made by soldering a metal strap to it. When the film is not metal, it is often deposited over a bracket at one end, which can be gripped by

the test equipment. The pulling force is monitored as the film is peeled. From this data and width of the strip removed, peeling force is calculated. Peel tests produce a more quantitative and reliable measure of interfacial adhesion than any of the tests described above.

# EXPERIMENTAL

We have modified the peel test in order to apply it to testing adhesion of thin polymeric membranes to surfaces of silicon substrates, such as the silicon nitride passivation layer used in our solid-state chemical sensors. The goal of this task was to develop a procedure for testing adhesion of both dry and wet membranes, which would yield repeatable, quantitative results quickly, and with simple sample preparation. In our work it is important to produce valid comparisons of adhesion both over variations of a given membrane composition, and between membranes made of dissimilar polymer matrices. The method described hereafter could be used by others who are concerned with adhesion of thin polymer layers to solid surfaces.

## Sample Preparation

Fig. 1 outlines the overall adhesion test; the first three steps in this figure show the process flow for sample preparation. The sample preparation for this test, though straightforward, is important in assuring reliable data. After a prefurnace clean, three-inch silicon wafers have 3600Å of silicon nitride deposited by chemical vapor deposition. A programmable diamond saw is then used to cut an 8-mil deep by 3-mil wide scribe lane across the center of the 15-mil thick wafers on the back side. Acetone, isopropyl alcohol, and nitrogen are used to clean and dry the wafers, which are then ready to have polymeric membranes cast on the front side.
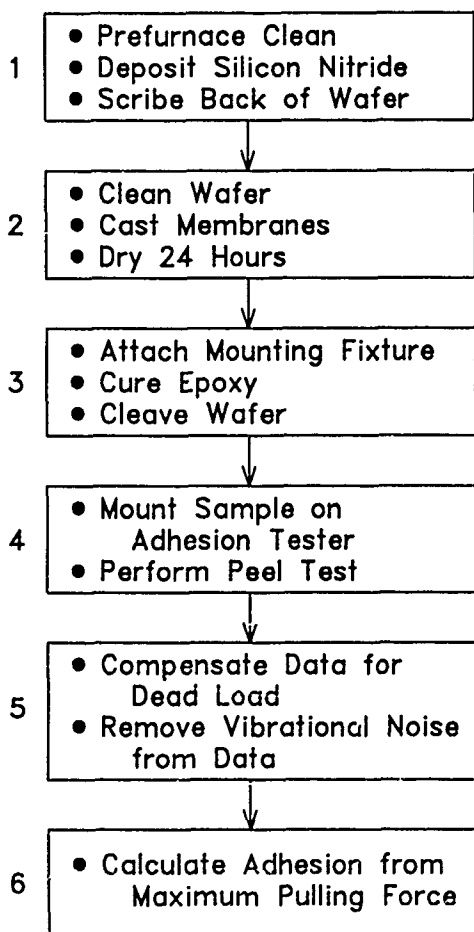
Membrane casting solutions are mixed by completely dissolving the polymer membrane components in 1.2 ml of anhydrous tetrahydrofuran (THF, water content < 0.005%). Glass rings, 22 mm in diameter, are placed on the front of the silicon wafer, centered over the back-side scribe lane. These glass rings are filled with casting solution, and the solvent is allowed to evaporate for 30 minutes until the membranes are partially cured. The glass rings are then removed, and the membranes are dried for 24 hours.

Mounting brackets are attached with epoxy to the back side of the wafer on either side of the scribe lane. When the epoxy is dry the wafer is ready for adhesion testing. The wafer is scribed down the precut lane and the wafer is cleaved by carefully folding the halves, held together by the membrane, forward, so that the membrane halves are touching.

## Adhesion Testing Apparatus

The cleaved wafer is mounted on an Instron 1131 mechanical tester (see Fig. 2) with one half attached to the load cell (Cardinal Scale SP-50L) the other to a stationary grip. This mounting method leaves the membrane free of deformation from mounting fixtures or grips since the membrane contacts only the substrate. The tester has a traveling crosshead which moves at 0.5 inches/minute, inducing a pulling force at both membrane/substrate interfaces. The membrane peels from both surfaces at approximately the same rate, maintaining a constant peel angle of 90°. The load cell converts the pulling force to a corresponding differential voltage.

The load-cell voltage is amplified by a precision instrumentation amplifier and then converted to digital representation at 135 samples/second by a 12-bit A/D converter, which is part of a programmable data acquisition and control adapter card (IBM part# 6451502) in an IBM-XT microcomputer. The microcomputer stores the force data and calculates the crosshead position from the known pulling speed.

## Signal Processing

Processing variations such as imprecise scribing of the wafer, variations in mounting-bracket weight, and differences in the amount of epoxy used on the brackets bias the force data. These effects are compensated by reading a dead-load weight after the membrane is completely peeled, and subtracting it from the force data. For consistency, the membrane weight is always included in the dead load.



1
- Prefurnace Clean
- Deposit Silicon Nitride
- Scribe Back of Wafer

2
- Clean Wafer
- Cast Membranes
- Dry 24 Hours

3
- Attach Mounting Fixture
- Cure Epoxy
- Cleave Wafer

4
- Mount Sample on Adhesion Tester
- Perform Peel Test

5
- Compensate Data for Dead Load
- Remove Vibrational Noise from Data

6
- Calculate Adhesion from Maximum Pulling Force

Figure 1: Process flow for membrane adhesion test, including sample preparation (1–3), peel test (4), signal processing (5), and computation (6).



Traveling Crosshead

Load Cell

Cable to Instrumentation

Mounting Bracket
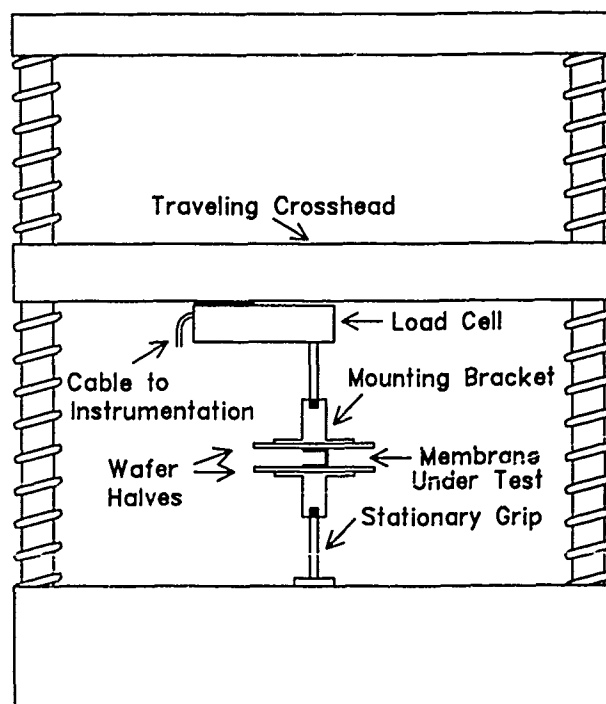
Wafer Halves

Membrane Under Test

Stationary Grip

Figure 2. Schematic drawing of adhesion tester.

Table 1. Compositions and electrochemical characteristics for membranes of four different matrices, selective to potassium and ammonium

| | Composition of Matrix (wt.%)[1] | | Slope (mV/decade) | | Detection Limit ($\mu M$) | | Selectivity $\log k_{i,j}^{pot}$ | | | $R$ | Composition Notes | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | | | | | | $i = K^+$ | $i = NH_4^+$ | | | | |
| | | | $K^+$ | $NH_4^+$ | $K^+$ | $NH_4^+$ | $j = Na^+$ | $j = Na^+$ | $j = K^+$ | | | |
| 1 | PVC | 33% | 57.1 | 55.8 | 0.58 | 0.98 | −4.28 | −2.82 | −0.85 | 1.0 | PVC | poly(vinyl chloride) |
| | | | | | | | | | | | PVC/Ac/Al | 80% vinyl chloride |
| 2 | PVC | 20% | 57.3 | 56.0 | 0.52 | 0.91 | −4.22 | −2.77 | −0.82 | 0.8 | | 5% vinyl acetate |
| | PVC/Ac/Al | 20% | | | | | | | | | | 15% vinyl alcohol |
| | | | | | | | | | | | PU | polyurethane |
| 3 | PU | 26.4% | 57.2 | 56.8 | 0.59 | 0.99 | −4.21 | −2.89 | −0.89 | 0.9 | P-SS | polydimethylsiloxane, silanol terminated |
| | PVC/Ac/Al | 6.6% | | | | | | | | | Si-CN | 10–12% (cyanopropyl)methyl |
| | | | | | | | | | | | | 88–90% dimethylsiloxane copolymer |
| 4 | P-SS | 78% | 56.5 | 52.9 | 1.0 | 1.4 | −4.16 | −2.64 | −0.66 | 33.5 | | |
| | Si-CN | 21% | | | | | | | | | $R$ | Normalized Resistance |

[1]Potassium- and ammonium-selective membranes are doped with 1% valinomycin and nonactin, respectively. The balance of the membrane is DOA (bis(2-ethylhexyl)adipate) plasticizer. (The silicone rubber membrane contains no plasticizer.)

Crosshead motion, being driven by mechanical gears and screws, imparts vibrational noise to the load cell. This random (white) noise becomes part of the data stream. We have used both averaging and median filters to remove noise from the signal, with virtually identical results. Since no advantage was seen using the median filter, a computationally simpler averaging algorithm is used. Averages are calculated on samples in a sliding data window. The window is sized large enough so that vibrational noise is removed, but small enough to avoid distorting data by excessive smoothing. Fig 3 shows the raw sampled data compared to the filtered data. The filtered signal has virtually no noise, yet follows all of the trends of the raw data. The small fluctuations in the filtered curve are due to mechanical imperfections of the interface. Adhesive failures propagate along these imperfections [7, 10].

The processed data is then scanned for the maximum force value, which is taken to represent the adhesion of the material. Use of the maximum value eliminates artifacts related to membrane elasticity, plastic or viscoelastic deformation, or membrane thickness, which shift the force curves and make their shapes peculiar to membranes of different matrices. In our experience, the maximum force the interfacial bond can withstand before yielding, is the best representation of the adhesive strength of the membrane material. Since all of our tests are comparative, rather than absolute, and membranes are all of the same size, there is no need to normalize for area.
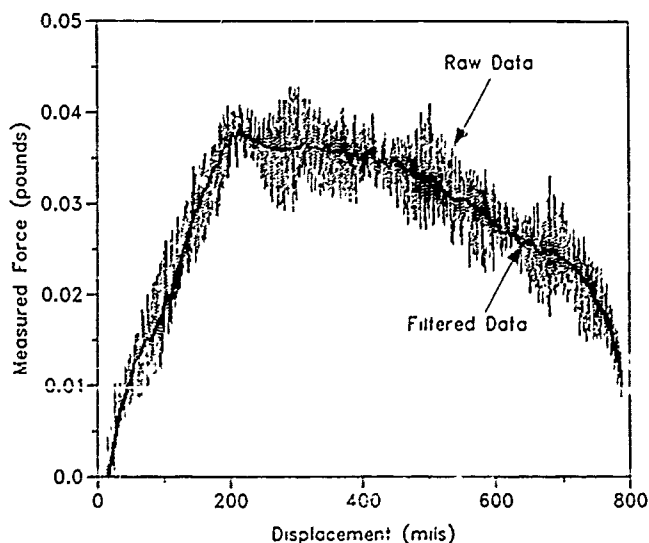
## Results and Discussion

A conventional PVC membrane and three new membrane compositions were compared for adhesion using the procedure described above. The new compositions (including polymer matrix, neutral carrier ionophore, and plasticizer as appropriate) had been optimized for response and mechanical strength. Our new peel testing capability will allow optimization for adhesion, as well. Table 1 lists the compositions and electrochemical properties of membranes based on these matrices, selective to potassium and ammonium. The hydroxylated PVC and polyurethane membranes have electrochemical performance equivalent to that of PVC membranes in terms of slope, detection limit and selectivity. Due to its high resistance, the silicone rubber membrane has somewhat inferior (though still useful) electrochemical properties.

Samples for three adhesion test runs were made: a) a dry adhesion test using three samples of each of the four membrane matrices, b) a dry adhesion test using three samples each of membranes 1–3 (see Table 1), with 10 wt.% SiCl4 mixed with the casting solution, and c) a wet adhesion test using samples identical to those in b). Fig. 4 plots typical data for the four membrane types. Peel test data from multiple samples is averaged and summarized in Table 2. The polyurethane and silicone membranes have much better adhesion to the sensor surface than do PVC or hydroxylated PVC.
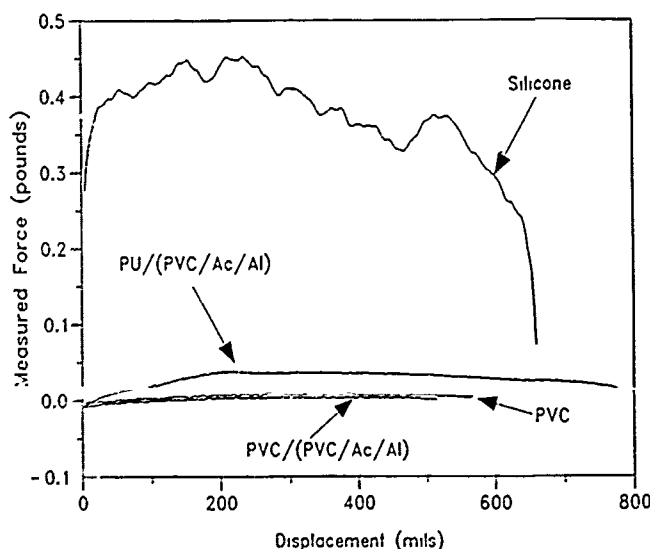


Figure 3. Raw and filtered force data versus displacement for a typical polymeric membrane.



Figure 4. Typical peel force curves for the four membrane matrices evaluated in this study.

5. By multiplexing different light sources in different spots, the device can be a multisensor without additional process complexity. Also, the zone covered by one light source can be split up into many individual small sensing areas without added complexity.

6. The absence of front-side contacts or metalization means the sensor surface can be very flat and smooth.
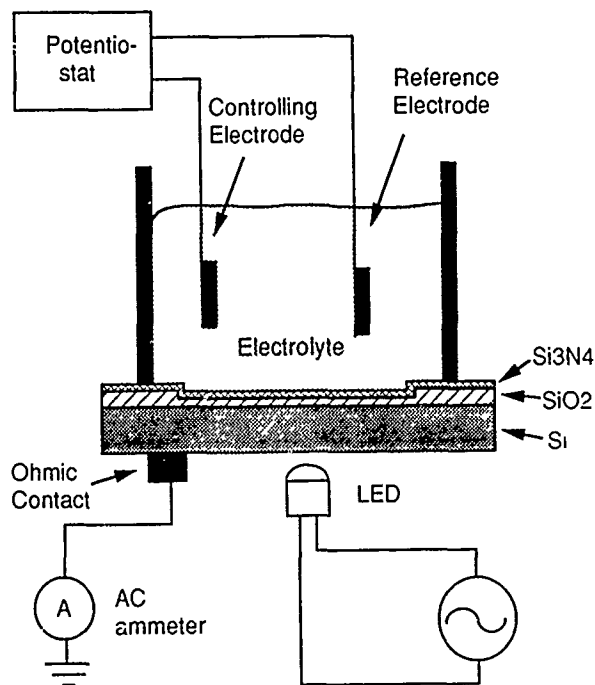


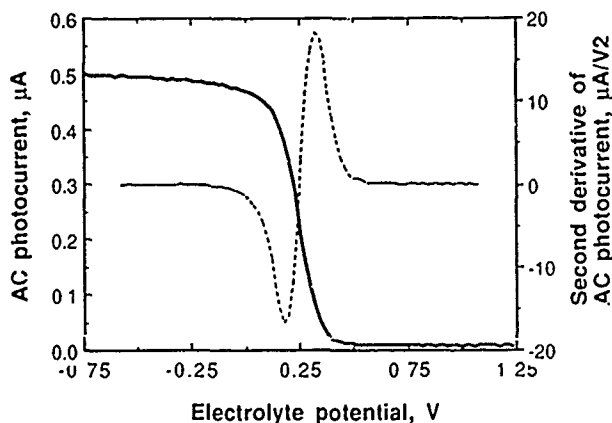Figure 1. Diagram of the setup to measure AC photocurrent/voltage curves with a LAPS.



Figure 2 Example of an AC photocurrent/voltage curve, together with the numerically calculated second derivative. The zero of the second derivative is used to find the inflexion point of the curve.

## MEASUREMENT OF METABOLIC RATES

In order to survive, the cells have to be bathed in a relatively fresh culture medium. When cells are placed in a microvolume chamber through which culture medium can flow, the chamber pH is determined by that of the medium. The flow rate is usually about 100 μL/min. To measure cell metabolism, the medium flow is periodically interrupted. As the cells consume glucose, they excrete acidic metabolic products (carbon dioxide and lactic acid, produced by aerobic respiration and glycolysis, respectively). The cells therefore act as a source of protons, which will change the pH at a rate depending on the cavity volume and the pH buffering capacity of the medium. This situation is similar to the one where an enzyme catalyzes a reaction that changes the pH in a small volume (7).

The flow on/flow off cycle is repeated with a period of 50 to 300 seconds. A typical chamber has dimensions approximately 100 μm high, 3mm wide and 12 mm long (volume 3.6 μL); the volume used for pH detection, however, is about 100 nL, and contains typically about 1000 cells. Figure 3 shows an example of metabolic rate detection, with a typical rate of 60 μV/s, or 0.06 pH units/min.

The method used for retaining the cells in the microvolume chamber depends on the cell type. If the cells are adherent, a confluent monolayer can be grown on a solid support. One example is glass cover slips, which are used as the chamber wall opposite the silicon pH detector. Non-adherent cell types, however, need to be retained in a mechanical structure, such as the one shown in Figure 4. This consists of a set of wells deep enough to hold the cells, even when medium is flowing over the wells.
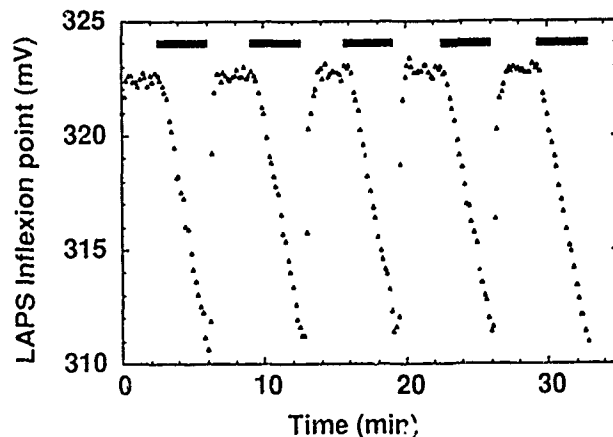


Figure 3. Example of data used for the determination of metabolic rates.

174

## WET ANISOTROPIC ETCHING FOR FLOW CHANNELS

A first application of silicon micromachining is the etching of channels in the silicon to serve as the fluidic pathways, and to accurately define the volume of the microvolume chambers. The chamber is made by placing a flat plate with entry/exit ports over the etched silicon; it is possible for instance to use a glass cover slip coated with adherent cells. The main requirement for this etch is the creation of a smooth bottom surface, which was achieved with an anisotropic wet etch. In the basic process sequence given above, this step is added between steps 2 and 3. The result is that the active area is the same as the area which is etched down. Figure 5 is a picture of a sensor chip with a fluidics pattern etched into it. The dimensions of the etched-down area are relatively large (about 13 × 7 mm), which means that the 55 degree angle at the edges has little impact.

## PLASMA ETCHED CAVITIES FOR CELL CAPTURE

A more critical application of micromachining is the etching of arrays of cavities for the retention of non-adherent cells. The requirements are that cells remain in the cavities even when a sufficient flow of nutrient medium goes through the chamber. This implies that the flow lines cannot penetrate very much into the cavity itself. Clearly, this will require deep cavities with walls close to vertical. The size of the cavities must be somewhat larger than typical cell dimensions, which are 10 to 20 µm. Our typical cavities are 50 µm squares, although we have experimented with various patterns. The cavity depth must be at least equal to the side of the square, so that the etched volume is roughly a cube.

To meet these requirements while still retaining the (100) orientation of the starting material, we chose anisotropic plasma etching. This method was described by Zdeblick et al. (8); the gases are $SF_6$ and $C_2ClF_5$ in a 1:1 ratio. The cavity depth was measured to be about 50 µm, and the angle of the walls was between 85 and 90 degrees. As in the case of the wet etch, the etch was done after the active area definition and before gate oxide growth. Therefore, only the sides and bottoms of the cavities are active pH-sensing areas. This means that the inside of each cavity, with a volume of 0.125 nl, acts as an individual pH sensor. The photocurrent generated in the LAPS device is the sum of the the contributions from a 1 $mm^2$ array of about 180 of these cavities. The backside illumination with modulated light thus acts as a virtual backside contact for potentiometric sensors, and enables all those in the area covered by the light beam to be accessed simultaneously.

The method for filling the cavities consists of filling the entire chamber with a suspension of cells, and allowing them to settle by gravity into the wells. A slow flow of liquid (linear velocities 20 to 100 µm/s) then sweeps away the cells that are not trapped in the bottom of a well. Figure 6 is a composite photograph showing a set of cavities before and after this filling process.

The pattern of laminar flow over a deep rectangular cavity is relatively well understood, both theoretically (9) and experimentally (10). Figure 7 shows the flow pattern. The most important aspect is the existence of separating flow line, which penetrates slightly into the cavity. Mass transport across this line is by diffusion only. Below this flow line, there are one or more vortices, depending on the depth of the cavity. For a 1:1 aspect ratio, there is room for only one vortex. At about a 2:1 depth to

width aspect ratio, a second, much slower, vortex appears. In this context, the important point is that cells packed in below the separating flow line will not be swept out of the cavities by a slow flow in the laminar range. The vortex will only agitate cells within a cavity, as we have observed.
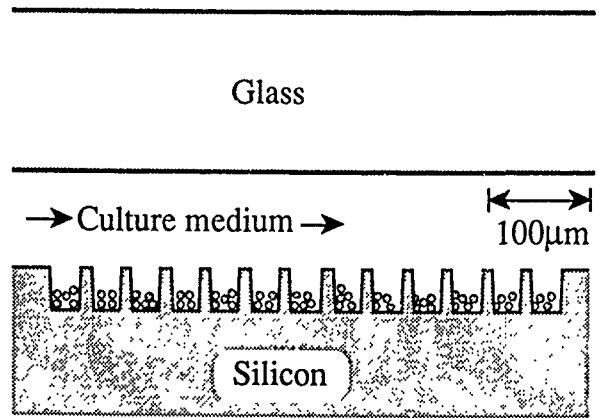


Figure 4. Cross-section of a flow chamber for holding nonadherent cells in etched wells.
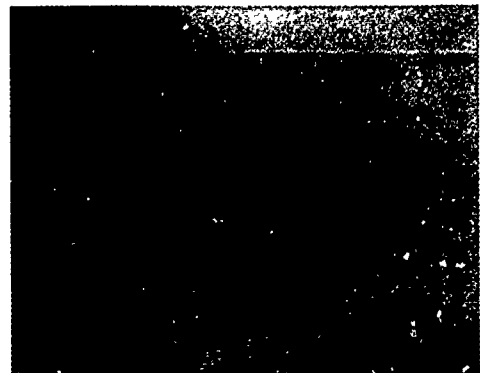


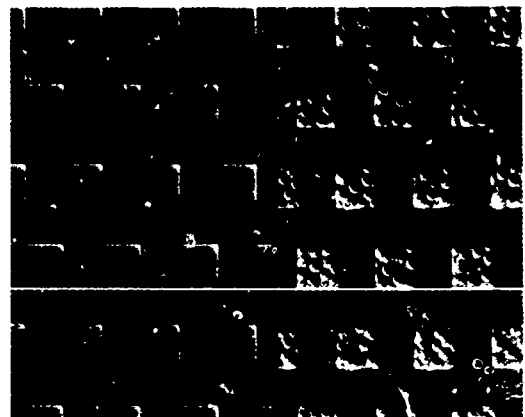Figure 5. Picture of a LAPS chip with two parallel fluidic chambers etched into it. Chip size is 13 × 17.5 mm.



Figure 6. 50 µm square and 50 µm deep wells etched into the surface of a LAPS chip. On the left, empty wells, and on the right, living nonadherent P388D1 cells trapped in the wells.
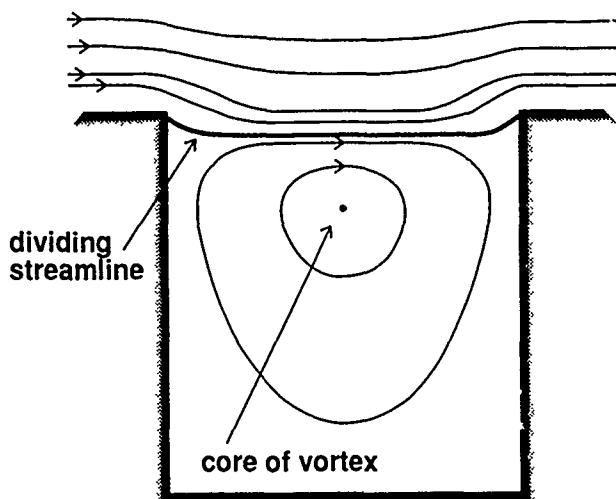
Figure 7.  Pattern of streamlines for laminar flow past a square cavity, adapted from Takematsu (9).

## CONCLUSIONS

The silicon chips described here perform dual roles as sensor (to measure metabolic rates) and structural component (defining flow chambers and cell traps to maintain nonadherent cells in the flow chamber). Using silicon technology to achieve these two functions is particularly valuable when compared to alternatives. This combination is common in physical or mechanical sensors, where for instance micromachining is used to make a thin diaphragm or a proof mass, and the piezoresistivity of the silicon measures the local strain. In chemical sensors, however. most attention has been directed towards the sensing functions themselves, and not towards the system context that requires these functions. We believe that it is more fruitful to design the whole measurement system rather than only the sensors in it. The greatest value of silicon sensor technology resides in its capability to accomplish several critical system functions simultaneously. There has been a considerable effort lately to implement fluidic functions such as valves and pumps in silicon; such components are also candidates for future integration into the silicon chip which lies at the heart of the microphysiometer.

## ACKNOWLEDGEMENTS

## REFERENCES

1. D.G. Hafeman, J.W. Parce, and H.M. McConnell, "Light-addressable potentiometric sensor for biochemical systems", Science, Vol. 240 (1988) pp. 1182–1185.

2. J.W. Parce, J.C. Owicki, K.M. Kercso, G.B. Sigal, H.G. Wada, V.C. Muir, L.J. Bousse, K.L. Ross, B.I. Sikic, and H.M. McConnell, "Detection of cell-affecting agents with a silicon biosensor", Science, Vol. 246 (1989), pp. 243–247.

3. J.C. Owicki and J.W. Parce, "Bioassays with a microphysiometer", Nature, Vol. 344, 1990, pp. 271–272.

4. J.C. Owicki, J.W. Parce, K.M. Kercso, G.B. Sigal, V.C. Muir, J.C. Venter, C.M. Fraser, and H.M. McConnell, "Continuous monitoring of receptor-mediated changes in the metabolic rates of living cells" (submitted to Proc. Natl. Acad. Sci.)

5. G.B. Sigal, D.G. Hafeman, J.W. Parce, and H.M. McConnell, "Electrical properties of phospholipid bilayer membranes measured with a light addressable potentiometric sensor", in Chemical Sensors and Microinstrumentation, R.W. Murray et al. Eds., ACS Symposium Series nr. 403, American Chemical Society, Washington DC, 1989 (pp. 46–64)

6. David L. Harame, Luc J. Bousse, John D. Shott, and James D. Meindl, "Ion-sensing devices with silicon nitride and borosilicate glass insulators", IEEE Trans. Electron Devices, Vol. ED-34 (1987) p. 1700.

7. Luc Bousse, Gregory Kirk, and George Sigal, Biosensors for detection of enzymes immobilized in microvolume reaction chambers, Sensors and Actuators, Vol. B1, 1990, pp. 555–560.

8. M.J. Zdeblick, P.P. Barth, and J.B Angell, "A microminiature fluidic amplifier", Sensors and Actuators, Vol. 15, 1988, pp. 427–433.

9. M. Takematsu, "Slow viscous flow past a cavity", J. Phys. Soc. Japan, Vol. 21, 1966, pp. 1816–1821.

10. Milton Van Dyke, "An album of fluid motion", The Parabolic Press, Stanford, 1982.

# A SURFACE GENERATED ACOUSTIC WAVE LIQUID MICROSENSOR*

P. Clarke, R. Lec and J.F. Vetelino
Dept. of Electrical Engineering and
Laboratory for Surface Science and Technology
University of Maine
Orono, Maine 04469

## ABSTRACT

Normally when a thin fluid layer comes in contact with the delay path in a surface acoustic wave (SAW) device the SAW is severely attenuated and no appreciable output signal is observed. However, if the fluid layer is sufficiently thick, it is possible for the SAW to excite acoustic waves in the fluid. In this work the properties of the acoustic waves in a fluid overlay on a piezoelectric substrate are investigated. Acoustic wave properties such as velocity and attenuation are obtained as a function of the thickness and physical properties of the fluid layer. It is shown that subtle changes in fluid properties modify the acoustic wave properties which can then be used as the sensing element in a fluid microsensor.

## INTRODUCTION

In the last ten years several papers have appeared (1) on the use of surface acoustic wave (SAW) devices as sensors. In applications such as gas sensing the sensor is configured as a dual SAW oscillator where one oscillator acts as the reference and the other oscillator has a gas sensitive film deposited between the input and output interdigital transducers (IDTs). The film selectively sorbs the gas of interest which in turn change the SAW properties. The corresponding relative change in the SAW oscillator frequency is then found to be directly proportional to gas concentration. If one replaces the film in the gas sensor by a fluid layer one might realize the possibility of using the dual oscillator configuration as a viscometer, immunosensor or simply as a sensor for a particular solute in solution. Unfortunately, once a fluid layer is placed in the SAW delay path, the SAW is severely attenuated and no appreciable signal is observed at the output IDT.

Recently work has appeared (2-18) demonstrating that a piezoelectric crystal such as quartz can oscillate in a bulk wave mode when in contact with a liquid. Although the losses are high, liquid bulk wave sensors have been able to detect viscosity and various solutes in solution. The bulk acoustic wave (BAW) propagates in the piezoelectric crystal and is modified only by the boundary conditions at the fluid-solid interface.

Another type of sensor suggested (19-21) for application as a fluid microsensor is the plate mode sensor. This device is similar to the SAW device in that the acoustic waves are excited by an IDT. However, instead of the energy propagating as a SAW it is directed into the crystal resulting in the excitation of the plate modes of the crystal. These modes then interact with the boundary between the liquid and the crystal and can be used to monitor liquid properties such as viscosity or density. A report (22) of a fluid microsensor which utilizes the SAW has appeared in the literature. However, it appears that in this case the energy is being transmitted as a plate mode and not a SAW (23).

Recently work has appeared (24,25) relating to the use of special types of surface guided acoustic waves for sensing fluid properties. In particular a leaky wave has been suggested (24) for use in a fluid microsensor. The leaky wave, however, attenuates in the direction of propagation hence the effective path length is small. Also theoretical work has suggested (25) that shear surface waves such as the Bluestein-Gulyaev wave and the surface skimming bulk waves might also be used in a fluid microsensor.

A sensor which allows an acoustic wave to propagate through a fluid instead of simply interacting with the fluid-crystal boundary as in the case of the fluid BAW and plate mode sensors might offer the possibility of being more sensitive to subtle changes in fluid properties. It is the purpose of this paper to investigate a sensor configuration which allows the acoustic wave to not only interact with the fluid-crystal boundary but also to propagate through the fluid layer.

## THEORY

In order to couple SAW energy into a fluid layer the geometry shown in Figure 1 which consists of a YZ-cut LiNbO3 SAW delay line upon which a small amount of fluid has been deposited will be studied. The fluid covers only a fraction of the delay path and is constrained at the top surface by a glass plate which is parallel to the LiNbO3 substrate. Normal capillary action holds the fluid between the two solids. When the IDT is excited a SAW of velocity, $v_r$, propagates along the crystal surface and is incident on
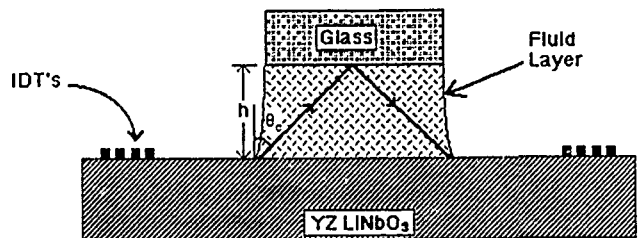


Figure 1. Prototype fluid microsensor geometry

the fluid layer. A significant portion of the SAW energy is converted into a bulk compressional acoustic wave of velocity $v_f$, which radiates into the fluid at an angle, $\theta_c$, with respect to the normal to the crystal surface which is defined as follows (26),

$$\theta_c = \sin^{-1}\left(\frac{v_f}{v_r}\right).$$

(1)

The acoustic wave in the fluid is incident at the top solid and radiates some of its energy to acoustic waves which are transmitted into the solid. The reflected acoustic wave propagates through the fluid and is incident at the piezoelectric substrate at an angle, $\theta_c$. At this point the wave is converted to a SAW and propagates to the output IDT.

Depending on the fluid layer thickness the acoustic wave may suffer one or more bounces off of the top solid. If the acoustic wave strikes the LiNbO3 substrate near or at the air-fluid boundary as shown in Figure 1, it is possible for a significant amount of the acoustic energy to reach the output IDT. Any changes in fluid properties such as density, viscosity, etc. will alter the velocity and attenuation of the acoustic wave in the fluid and will hence modify the device response.

## EXPERIMENT

The IDT consisted of 20 finger pairs, a separation of 25.4 mm between transducers and an excitation frequency of 59.5 MHz. The spacing between the IDT and the fluid was 10.6mm while the fluid covered 4.2mm of the delay path. The delay line was mounted on a microscopic stage which could be moved in the X,Y and Z direction and rotated about the vertical (Z) axis. The delay line and the upper glass solid were mounted on two different stages which had three degrees of freedom and allowed parallelism to be achieved between the two solids. The substrate and upper solid could be moved vertically with an accuracy of $\pm 1 \mu$, horizontally with an accuracy of $\pm 100 \mu$ and rotated with an accuracy of $\pm 0.1°$.

## RESULTS AND DISCUSSION

The velocity and attenuation of the acoustic waves were measured using RF and pulse response techniques. The acoustic wave properties were determined as a function of the fluid thickness for pure water and water with various dissolved solutes.

In Figure 2 photographs of the various RF pulse responses are shown. In Figure 2a the exciting pulse is shown while in Figure 2b the SAW response of the bare delay line is shown. In Figure 2c the delay line response associated with the single bounce configuration is shown for a 2.53 mm layer of water. The SAW response has disappeared and a new response appears, with a time delay longer than that associated with the SAW response. If one uses equation (1) which yields $\theta_c = 25.4$ and the velocity of sound in water it may be shown that the new response is the single bounce mode. The amplitude of the single bounce mode is roughly 24 db less than the SAW response but is clearly observable. Finally, in Figure 2d the device response for a 2.53 mm thick layer of 20% NaCl in water solution is presented. Close examination of Figures 2c and 2d reveals that the time delays and the amplitudes for the two responses were different. This is directly attributable to the NaCl which was added to the water.

In Figure 3 the experimental measurements of velocity (v) of the bounce mode normalized with respect to the compressional acoustic wave velocity in the fluid ($v_f$) are presented as a function of fluid thickness. It can be clearly seen that the bounce mode velocity does not vary with thickness and is hence nondispersive. Also shown in Figure 3 is the amplitude of the bounce mode as a function of thickness. It can be seen that the optimum single bounce mode, therefore the wave which strikes the piezoelectric substrate and travels closest to the air-fluid boundary, occurs at a thickness of 2.67 mm.

In Figure 4 the normalized bounce mode velocities for different NaCl concentrations in water are presented
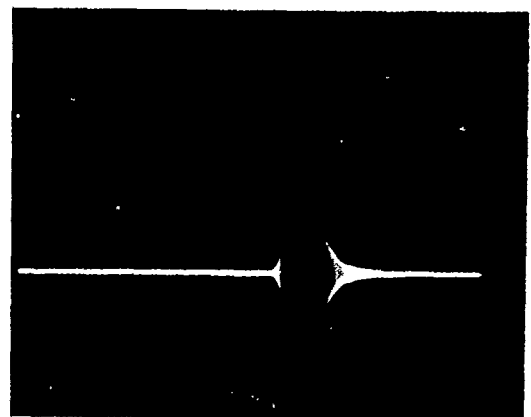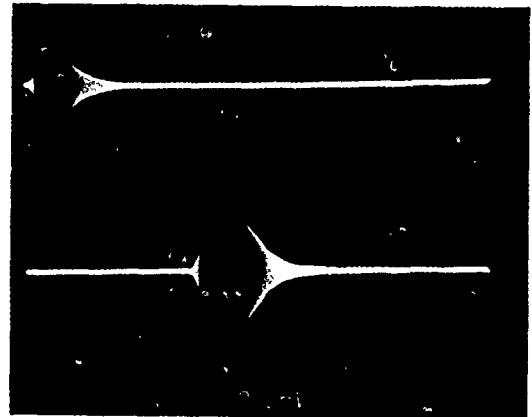


Figure 2    Photographs of the various RF pulse responses of the bounce mode geometry.    (a) Exciting pulse, (b) SAW response of the bare delay line, (c) bounce response within a 2.53 mm water layer, (d) bounce response with a 2.53 mm 20% NaCl solution in water.
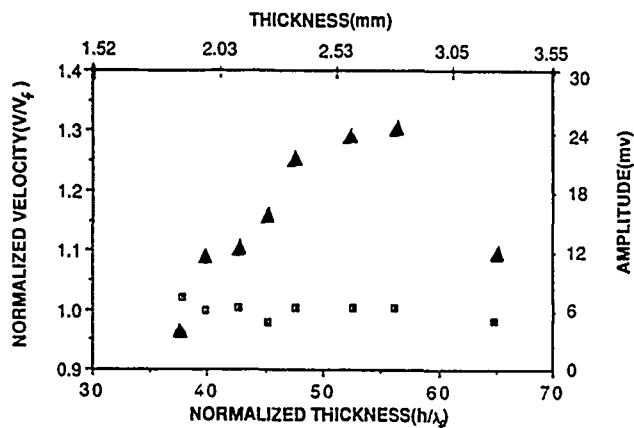
Figure 3. Velocity and amplitude of the bounce mode as a function of fluid thickness. Triangles - the velocity, v, of the bounce mode normalized with respect to the compressional wave velocity in pure water, vf and triangles - the amplitude of the bounce mode.
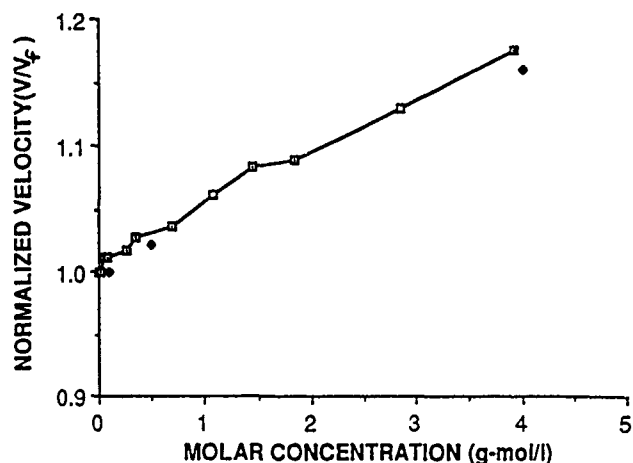


Figure 4. Normalized bounce mode velocity for various NaCl concentrations in water. Squares - present work, dots - ref 27.

and compare favorably with suitable literature data (27). Similar experiments were performed for various solutions of glycerin in water and are presented in Figure 5.

Examination of Figures 4,5 reveals that the velocity at a particular fluid thickness is a linear function of the concentration of solute in water. This is true for both low viscosity (NaCl) and high viscosity (glycerine) solutions.

## SUMMARY AND CONCLUSIONS

A new SAW liquid sensor configuration using BAWs generated by a SAW has been investigated. A theoretical model has been developed and verified experimentally by measuring the acoustic velocity and attenuation using the RF pulse technique. Various liquids containing different concentrations of solutes in solution have been studied. Both velocity and attenuation of the acoustic mode varied in a predictable fashion as the fluid properties changed.
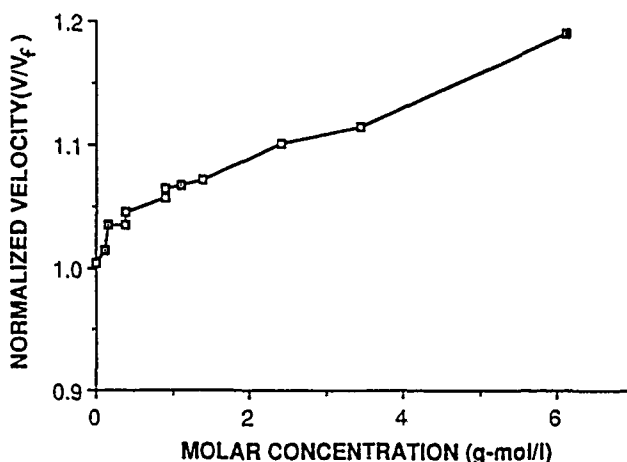


Figure 5. Normalized bounce mode velocity for various glycerin in water solutions. Squares - present work.

The sensor exhibits high sensitivity to minute variations in solute concentrations. The sensor response was not extremely sensitive to geometric factors such as the parallelism of the two solids. This type of acoustic sensor may have applications in such diverse areas as water pollution control, chemical sensors and biosensors.

## ACKNOWLEDGEMENTS

## REFERENCES

(1)    See for example IEEE Trans. Ultrason. Ferr. Freq. Cont. and references therein.

(2)    A.P.M. Glassford, "Response of a Quartz Crystal Microbalance to a Liquid Deposit," J. Vac. Sci. Technol. 15, pp. 1836-1843 (1978).

(3)    R.A. Crane and G. Fischer, "Analysis of a Quartz Crystal Microbalance with Coatings of Finite Viscosity," J. Phys. D. Appl. Phys. 12, pp. 2019-2026 (1979).

(4)    P.L. Konash and G.J. Bastiaans, "Piezoelectric crystals as detectors in liquid chromatography," Anal. Chem., 52, pp. 1929-1931 (1980).

(5)    T. Nomura, "Single-Drop Method for Determination of Cyanide in Solution with a Piezoelectric Quartz Crysta," Anal. Chim. Acta 124, pp. 81-84 (1981).

(6)    T. Nomura and M. Iijima, "Electrolytic determination of namomolar concentration of silver in solution with a piezoelectric quartz crystal," Anal. Chim. Acta. 131, pp. 97-102 (1981).

(7)    T. Nomura and M. Okuhara, "Frequency shifts of piezoelectric quartz crystals immersed in organic liquids," Anal. Chim. Acta. 142, pp. 281-284 (1982).

179

(8) T. Nomura and T. Mimatsu, "Electrolytic determination of traces of iodide in solution with a piezoelectric quartz crystal," Anal. Chim. Acta, 143, pp. 237-241 (1982).

(9) T. Nomura and M. Maruyama, "Effect of metal solution and the adsorptive determination of iron (III) as phosphate," Anal. Chim. Acta, 147, pp. 365-369 (1983).

(10) T. Nomura and T. Nagamune, "Internal electrolytic determination of silver in solution with a piezoelectric quartz crystal," Anal. Chim. Acta, 155, pp. 231-234 (1983).

(11) T. Nomura and K. Tsuge, "Determination of silver in solution with a piezoelectric detector after electrodeposition," Anal. Chim. Acta, 169, pp. 257-262 (1985).

(12) T. Nomura, M. Watanabe, and T.S. West, "Behavior of piezoelectric quartz crystals in solutions with application to the determination of the iode," Anal. Chim. Acta, 175, pp. 107-116 (1985).

(13) K.K. Kanazawa and J.G. Gordon, "Frequency of a quartz microbalance in contact with liquid," Anal. Chem. 57, pp. 1770-1771 (1985).

(14) K.K. Kanazawa, "The oscillation frequency of a quartz resonator in contact with a liquid," Anal. Chim. Acta, 175, pp. 99-106 (1985).

(15) S. Bruckenstein and M. Shay, "Experimental Aspects of Use of the Quartz Crystal Mirobalance in Solution," Electrochim. Acta, 30, pp. 1295-1300 (1985).

(16) H.E. Hager and P.D. Verge, "Multicomponent Gas Analysis via Transient Operation of Liquid Coated Piezoelectric Crystal Detectors," Sensors and Actuators 7, pp. 271-283 (1985).

(17) M. Thompson, C.L. Arthur and G.K. Dhaliwal, "Liquid-Phase Piezoelectric and Acoustic Transmission Studies of Interfacial Immumochemistry," Anal. Chem. 58, pp. 1206-1209 (1986).

(18) M. Thompson, G.K. Dhaliwal, C.L. Arthur and G.L. Calabrese, "The Potential of the Bulk Acoustic Wave Device as a Liquid-Phase Immunosensor," IEEE Trans. Ultrason. Ferr. Freq. Cont. 34, pp. 127-135 (1987).

(19) A.J. Ricco and S.J. Martin, "Acoustic Wave Viscosity Sensor," Appl. Phys. Lett. 50, pp. 1474-1476 (1987).

(20) R.M. White, P.J. Wicher, S.W. Wenzel and E.T. Zellers, "Plate-Mode Ultrasonic Oscillator Sensors," IEEE Trans. Ultrason. Ferr. Freq. Cont. 34, pp. 162-171.

(21) J. Hou and H. van de Vaan, "Mass Sensitivity of Plate Modes in Surface Acoustic Wave Devices and Their Potential as Chemical Sensors," Proc. 1987 IEEE Ultrason. Symp., pp. 573-578.

(22) J.E. Roederer and G.J. Bastiaans, "Microgravimetric Immunoassay with Piezoelectric Crystals," Anal. Chem. 55, pp. 2333-2336 (1983).

(23) G.S. Galabrese, H. Wohltjen and M.K. Roy, "Surface Acoustic Wave Devices as Chemical Sensors in Liquids. Evidence of Disputing the Importance of Rayleigh Wave Propagation," Anal. Chem. 59, pp. 833-837 (1987).

(24) T. Moriizumi, Y. Unno and S. Shiokana, "New Sensor in Liquid Using Leaky SAW," Proc. 1987 IEEE Ultrason. Symp., pp. 579-582.

(25) F. Josse and T. Shama, "Analysis of Shear Surface Waves at the Boundary between a Piezoelectric Crystal and a Viscous Fluid Medium," J. Am. Acous. Soc., Oct. 1988.

(26) B.A. Auld, "Acoustic Fields and Waves in Solids," John Wiley & Sons, N.Y., N.Y. (1973).

(27) C.J. Montrose, K. Fritsch, "Hypersonic velocity and absorbtion in aqueous electrolytic solutions," J. Acoust. Soc. Am. 47, pp. 786-790 (1970).

# HIGH-STRESS AND OVERRANGE BEHAVIOR OF
# SEALED-CAVITY POLYSILICON PRESSURE SENSORS

K. H.-L. Chau, C. D. Fung, P. R. Harris, and J. G. Panagou

The Foxboro Company
N01-2A, 38 Neponset Ave.
Foxboro, MA 02035

## ABSTRACT

This paper reports the overrange behavior of sealed-cavity polysilicon pressure sensors with full-scale pressures ranging from 35 to 300 psi (0.24 to 2.0 MPa). Because of a built-in overpressure stop, rupture pressure between 3500 to 6000 psi (24 to 41 MPa) was observed. After repetitive cycles of overpressure to 3000 psi, no degradation in sensor performance was apparent, and the zero repeatability was better than 0.1% of span. These results demonstrate the outstanding overrange capabilities of sealed-cavity polysilicon sensors and the exceptional mechanical strength and reproducibility of polysilicon diaphragms for pressure sensing.

## INTRODUCTION

There are many applications in industrial process control where pressure sensors with good overrange capabilities are required. The sealed-cavity polysilicon pressure sensor, as originally reported by Guckel et al [1], represents a new class of sensing structures which employs a polysilicon diaphragm formed by first removing an oxide spacer underneath a polysilicon layer, and then sealing off the remaining cavity in vacuum using a gaseous reaction process. Piezoresistive readout is provided by polysilicon sensing resistors located at the diaphragm edge. Overrange protection is built-in for this type of pressure sensors, since the diaphragm can only deflect by a distance set by the cavity depth before reaching and being stopped by the silicon substrate. In this paper, the high-pressure behavior and overrange capabilities of sealed-cavity pressure sensors will be investigated. Experimental results, including the nonlinear high-pressure response, the rupture pressure, and the output reproducibility after subjecting to overpressure, will be reported and correlated with stress simulations by finite element analysis.

## DEVICE FABRICATION

First, a thin pad oxide layer was grown on n-type (100) silicon substrate. This was followed by a silicon nitride film, deposited by low-pressure chemical vapor deposition (LPCVD). At the prospective diaphragm locations, a 0.75 μm-thick oxide layer was formed by means of local oxidation of silicon. The remaining nitride-oxide layer was then completely removed to bare substrate except at those regions where a network of etch channels was patterned, leading to the thick oxide at the diaphragm sites. Then, a 2 μm-thick LPCVD polysilicon diaphragm material was deposited by pyrolysis of 100% silane at 580 °C and 300 mTorr. The as-deposited polysilicon exhibited a high compressive stress, which was subsequently reduced to near zero by annealing at 1150 °C for three hours in a nitrogen environment. Etch wells were then made through the polysilicon layer to expose the nitride-oxide channel areas, and the wafers were immersed in hydrofluoric acid solution to remove the thick oxide and free the polysilicon layer at the diaphragm regions. The etch channels were then sealed via an oxide growth, followed by a silicon nitride LPCVD, leaving a vacuum-sealed cavity underneath each polysilicon diaphragm [1]. To form the piezoresistive readout for the pressure sensor, 440 nm of polysilicon was deposited by LPCVD and ion-implanted with boron at 80 keV to a dose of $5 \times 10^{15}$ cm$^{-2}$. This layer was patterned to form resistor sensing elements at the diaphragm edge. Finally, a thin LPCVD nitride was deposited and followed by contact opening and metallization. This completed the fabrication of the sealed-cavity sensor.

## TEST SETUP

The fabricated devices were mounted onto stainless steel headers, which were furnished with high pressure fittings and feedthroughs for the wire leads. An oil-filled, hydraulic dead-weight tester was used to set the applied gauge pressure. Since the sealed-cavity sensor has a vacuum reference, the actual applied pressure is the sum of the applied gauge pressure and the barometric pressure, which was also monitored. Testing was performed at 27 °C in a constant temperature chamber. In order to further eliminate the effects due to the temperature coefficient of resistance, the pressure sensing resistor was compared with a pressure-insensitive resistor of identical geometry located outside the diaphragm. Using equal excitation currents for both resistors, the voltage difference between them was largely independent of temperature.
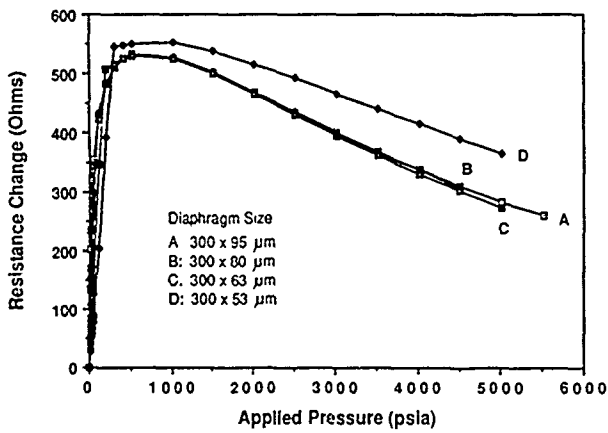
Fig. 1 Measured resistance change versus the applied pressure for a 9-kΩ sensing resistor located at the diaphragm edge.
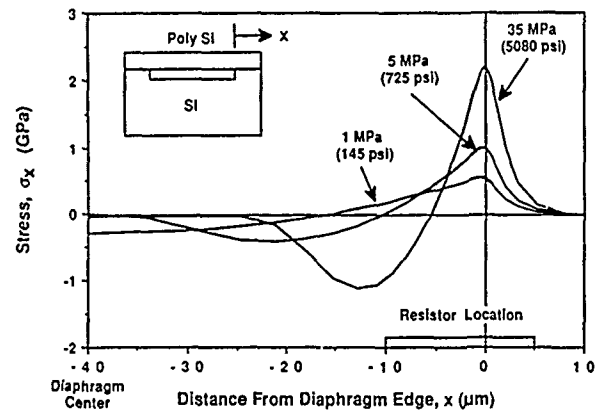


Fig. 2 Theoretical stress distributions on the top surface of a long rectangular sealed-cavity sensing diaphragm under various applied pressures. The diaphragm is 80 µm wide, 2.2 µm thick, and the cavity depth is 0.65 µm. The location of the sensing resistor is also shown.

Overrange testing was performed in steps of increasing pressure. Each time a new high pressure was set and the test results recorded, the sensor was rechecked for low-pressure response before proceeding to an even higher pressure. Any abnormal response or open resistor would indicate that the rupture point had been reached.

## RESULTS

Sealed-cavity sensors with rectangular diaphragms, ranging from 53 µm x 300 µm to 95 µm x 300 µm in size, were realized. The final diaphragm thickness and cavity depth were 2.2 µm and 0.65 µm, respectively. The full-scale pressure of these diaphragms, based on a linear response, ranges from 35 to 300 psi (0.24 to 2.0 MPa), with the smallest diaphragm yielding the highest range. A single 9-kΩ serpentine resistor, located along the longer edge of each diaphragm, provided a full-scale resistance change from 3% for a 35 psi device to 5% for a 300 psi device.

Consider first the overrange characteristics, i.e., the output response as the diaphragm is being stopped and supported by the silicon substrate. Figure 1 shows a typical set of measured resistance changes versus the applied pressures for four different diaphragm sizes. Resistance increases linearly with pressure until the diaphragm center comes into contact with the substrate. Since, at this time, the diaphragm area closer to the edges has not yet reached the stop, resistance continues to increase slowly and nonlinearly with further applied pressure. Eventually, at sufficiently high pressure (about 500 psi), the contact area has advanced to where the sensing resistor is located, thereby reversing the bending stresses from tensile to compressive. As a result, resistance finally reaches a maximum and then decreases with pressure throughout until rupture occurs. Finite element analyses of stresses using the ANSYS program further illustrate the overrange behavior of these sensors. Figure 2 shows the theoretical stress distribution along a path
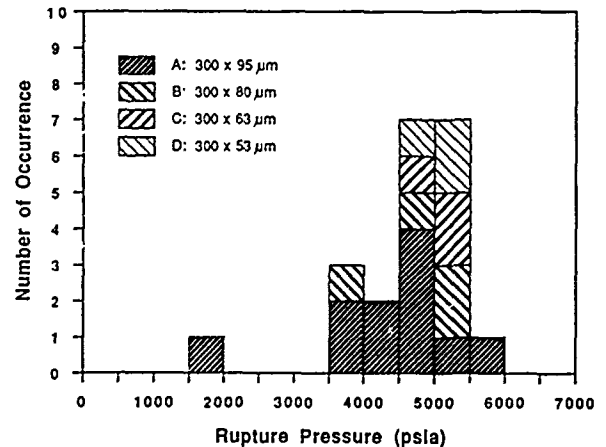


Fig. 3 A histogram for the measured rupture pressures of 21 sensors with full-scale pressures from 35 psi (Type A) to 300 psi (Type D).

perpendicular to the length on the top surface of a long rectangular diaphragm. Stresses increase linearly with pressure before the diaphragm reaches the stop. Afterwards, stresses are relieved over portions of the diaphragm that have been stopped while at the edges, the maximum stress continues to increase approximately with $P^{0.42}$, where $P$ is the applied pressure, until rupture stress is reached.

A histogram of measured rupture pressures for 21 sensors, with the previously mentioned sizes and pressure ranges, is shown in Figure 3. Rupture occurs between 3500 to 6000 psi (24 to 41 MPa), and appears to be relatively insensitive to the particular size or range of the sensors investigated. This is in agreement with finite element analyses, which show that while long rectangular diaphragms of different widths generally reach different stress
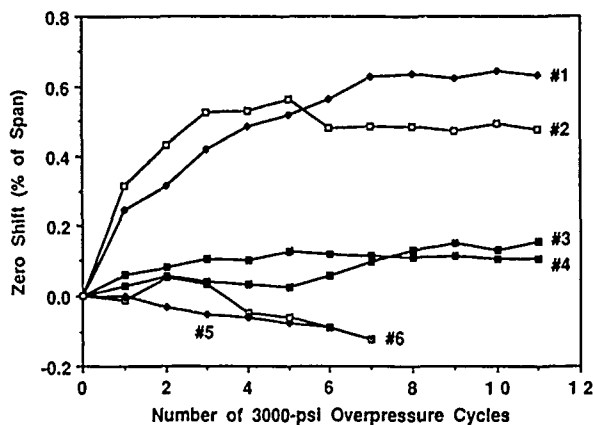
Fig. 4 Cumulative output shift at zero applied pressure versus the number of 3000-psi overpressure cycles for 6 sensors.

values before touching, their stress level after touching is roughly the same regardless of width, and is a function of the applied pressure only. This holds true provided that all the diaphragms have the same thickness and cavity depth, and hence they all tend to rupture at about the same pressure. The calculated diaphragm strains in the rupture range are on the same order of magnitude as previously reported for LPCVD polysilicon [2].

Below the rupture limit, the reproducibility of sensor response, after being subjected to repetitive overpressure, is also of importance. In this study, we overranged the 65 psi sensors to 3000 psi and measured their linear-range response every time after overpressure. Figure 4 shows the cumulative output shift at zero applied pressure versus the number of overpressure cycles. Sensors generally showed an initial trend of zero shift for the first few overpressure cycles, but all devices stabilized to within 0.1% of span for subsequent repetitive overpressures. These sensors were tested up to 90 overpressure cycles of 3000 psi. It should be pointed out that other than the above mentioned zero shift, no apparent degradation in span, hysteresis, or repeatability has been found.

## CONCLUSIONS

The overrange behavior of sealed-cavity polysilicon sensors has been studied, and the results demonstrated the outstanding mechanical strength and reproducibility of polysilicon diaphragms for pressure sensing. The superb built-in overrange protection represents a significant advantage of the sealed-cavity polysilicon sensors for industrial applications where high overrange pressure is common.

## REFERENCES

[1] H. Guckel *et al*, "Fine Grained Polysilicon and its Applications to Planar Pressure Transducers," *Dig. Tech. Papers*, Transducers'87, Tokyo, pp. 277-282, June 1987.

[2] Y. C. Tai and R. S. Muller, "Fracture Strain of LPCVD Polysilicon," *Dig. Tech. Papers*, IEEE Solid-State Sensor and Actuator Workshop, Hilton Head Island, SC, pp. 88-91, June 1988.

# SECONDARY SENSITIVITIES AND STABILITY OF ULTRASENSITIVE SILICON PRESSURE SENSORS

S. T. Cho, K. Najafi, and K. D. Wise

*Center for Integrated Sensors and Circuits*
Solid-State Electronics Laboratory
Department of Electrical Engineering and Computer Science
University of Michigan
Ann Arbor, Michigan 48109-2122

## ABSTRACT

The performance of boron-doped silicon pressure sensors as a function of variables such as plate thickness, membrane size and internal stress has been reported. However, the effect of temperature, creep, hysteresis and fatigue on device sensitivities is not well-known. This paper reports experimental results on the characterization of these effects in ultrasensitive boron-doped silicon membranes. Temperature coefficient of sensitivity (TCS) and zero pressure temperature coefficient of offset (TCO) were found to be as much as an order of magnitude higher than previously reported values; this is due to the strong dependence of pressure sensitivity on internal stress and the large mismatches in thermal expansion coefficients between silicon and dielectrics. Creep and fatigue affected the offset by <0.2% full scale and the change in pressure sensitivity was insignificant; hysteresis observed on all devices was also <0.2% full scale. The results reported here indicate that it may be possible to extend the pressure range by an order of magnitude, which would increase the resolution from 10 to 16 bits.

## INTRODUCTION

Ultrasensitive silicon micromachined pressure sensors are attractive for a wide range of emerging applications, and the pressure sensitivity of these structures as a function of variables such as plate size, membrane thickness and internal stress has been recently reported [1-5]. However, the temperature sensitivities and effects due to hysteresis, fatigue and creep in these thin diaphragm structures are not well known. Understanding these effects is critically important since they determine the ultimate resolution, accuracy, stability, and reliability of the device. This paper presents experimental results on the characterization of such secondary effects in boron-doped silicon membranes. Capacitive pressure sensors were fabricated with and without a variety of dielectrics on the diaphragms and were tested over pressure and temperature. The internal stresses and thermal expansion coefficients for LPCVD silicon dioxide, LPCVD silicon nitride and oxide/nitride composite films were also evaluated as a function of temperature. Thin silicon membranes were tested for creep and fatigue to determine if stretching or plastic deformation is an effect in diaphragm reliability.

## FABRICATION

Capacitive pressure sensors were fabricated using a versatile, single-sided dissolved wafer process. The fabrication sequence involves both silicon and glass processing[5]. A cross-section of a capacitive pressure sensor is shown in Fig. 1. Processing starts with a p-type (100) silicon wafer of moderate doping (>1$\Omega$-cm). A recess is etched into the silicon using KOH. This recess defines the cavity of the pressure sensor. Two high temperature (1175°C) boron diffusions follow. The first is a deep diffusion (12-15$\mu m$) which defines the rim of the membrane. The second is a shallow diffusion (2.6$\mu m$) which determines the thickness of the diaphragm. By varying the KOH etch depth and the shallow boron diffusion time, a wide variety of operating ranges and sensitivities for the capacitive pressure sensor can be obtained. LPCVD oxide or nitride films are then deposited on the membrane to compensate the stress and to provide protection against the metal plate and diaphragm

shorting together. The silicon dioxide was deposited at 920±10°C at a pressure of 450 *mTorr*, and a nitrous oxide: dichlorosilane gas ratio of 2:1. The nitride was deposited at 820±10°C at a pressure of 250 *mTorr* and an ammonia:dichlorosilane gas ratio of 4:1.

The glass processing involves depositing and patterning a multi-metal system (10$nm$ Ti/ 20$nm$ Pt/ 250$nm$ Au) on a #7740 Corning glass wafer. The metal forms the bottom plate of the capacitor and the output leads from the transducer. The silicon and glass are then electrostatically bonded together. The bonding takes place at 400°C with a potential of 1000V applied between the glass and silicon. In order to make electrical contact to the silicon, a metal lead on the glass is allowed to overlap the silicon rim over a small area. The gold is removed from the metal area here to reduce the step height between the glass and silicon. During the electrostatic bonding process, the silicon and glass are drawn tightly together, ensuring a low-resistance Pt/Ti contact to the silicon (40$\Omega$ for a 40$\mu m$ x 20$\mu m$ area). The final step in the process is a selective etch in ethylenediamine-pyrocatechol-water (EDP) [6]. EDP dissolves the silicon and stops on the heavily-doped (p+) diffused layers. The overall fabrication sequence requires only single-sided processing with four non-critical masking steps on silicon. Precise dimensional control is obtainable and the process is high yield as well as compatible with batch processing. A fabricated capacitive pressure sensor is shown in Fig. 2.
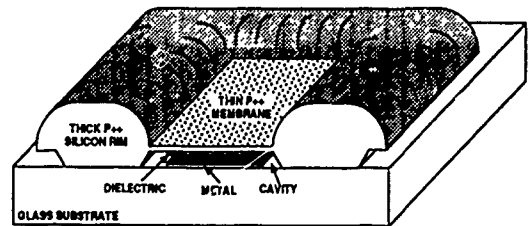
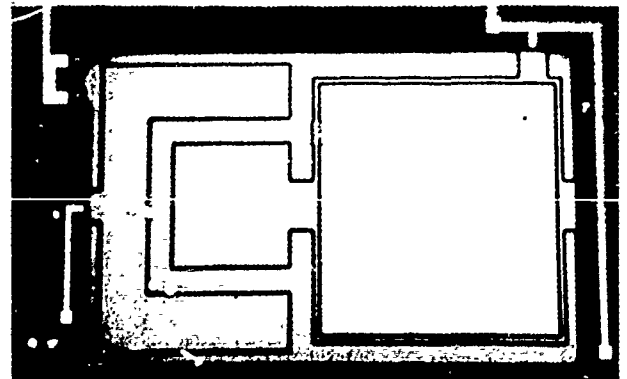Fig. 1: A Capacitive Silicon Pressure Sensor Fabricated Using the Dissolved Wafer Process

Fig. 2: A Fabricated Capacitive Pressure Sensor. The Membrane is 2mm x 2mm x 2.6$\mu m$.

## MATERIALS CHARACTERIZATION

Stress plays a large role in determining the pressure sensitivity of thin membranes [5]. In order to understand the temperature sensitivity of these structures, the variation of dielectric stress with temperature must be determined. Three different sets of dielectrics were deposited on silicon wafers: $100nm$ of CVD nitride, $200nm$ of CVD oxide and a $100nm/100nm$ oxide/nitride composite. These stress measurements were made using both electrostatic pull-in structures [5, 7] and a laser-based stress guage which measures wafer curvature. The room temperature stress results of the two techniques agree to within twenty percent.

The dielectrically-coated wafers were placed on a hot chuck and the laser-based stress guage monitored the stress as a function of temperature. The results are shown in Fig. 3. In general, as temperature increases, the stresses become more tensile and the change in stress is linear. Based on these observations, it is assumed that the change in stress is primarily due to the mismatch in thermal expansion coefficients between the silicon and the dielectric film. Since the thickness of the substrate is much greater than the film, the expansion coefficient of the film can be derived by [8]:

$$\Delta\sigma = (\alpha_{Si} - \alpha_f)\frac{E_f}{1 - v_f}\Delta T \qquad (1)$$

where $\Delta\sigma$ is the change in stress; $\Delta T$ is the change in temperature; $\alpha_{Si}$ and $\alpha_f$ are the thermal expansion coefficients of silicon and the film, respectively; $v_f$ and $E_f$ are the Poisson's ratio and Young's modulus for the film.

The measured expansion coefficient for silicon dioxide ($0.5 \times 10^{-6}/°C$) is in good agreement with the theoretical value ($0.4 \times 10^{-6}/°C$) [9]. However, the expansion coefficient for the nitride is a factor of two less than previously reported values ($1.1 \times 10^{-6}/°C$ versus $2.4 \times 10^{-6}/°C$).

## TEMPERATURE SENSITIVITY

Several sets of capacitive pressure sensors were fabricated for testing. The devices had a plate size of $2mm \times 2mm$ and an average thickness of $2.6\mu m$. In order to characterize the effect of dielectrics on diaphragm performance, the membranes were coated with several different combinations of dielectrics (no dielectric, $100nm$ nitride, $200nm$ oxide and $100nm/100nm$ oxide/nitride). All device testing was performed on devices that had unsealed cavities. The capacitance versus pressure response for these devices was measured over a temperature range of 0-100°C. The results are shown in Fig 4.

In evaluating the temperature sensitivity, there are two principal values to consider: the temperature coefficient of sensitivity (TCS) and the zero pressure temperature coefficient of offset (TCO). These variables are given by:

$$TCO = \frac{1}{C_o}\frac{\Delta C}{\Delta T}; \quad TCS = \frac{1}{S_o}\frac{\Delta S}{\Delta T}, \qquad (2)$$

where $C_o$ is the room temperature zero pressure capacitance, $S_o$ is the room temperature pressure sensitivity; $\Delta C/\Delta T$ and $\Delta S/\Delta T$ are the change in capacitance and pressure sensitivity per change in temperature. Pressure sensitivity (S) is defined by:

$$S = \frac{1}{C_o}\frac{\Delta C}{\Delta P} \qquad (3)$$

where $\Delta P$ is the change in pressure. The values for pressure sensitivity over temperature are given in Table 1 and the values for TCS and TCO are shown in Table 2.

Capacitive pressure sensors that have low temperature sensitivity [10] typically have very thick plates (>10μm) and are

Fig. 3: Dielectric Stress versus Temperature for Several Compositions of Dielectrics.

### Oxide and Nitride Stress vs. Temperature



### Oxide/Nitride Stress vs. Temperature
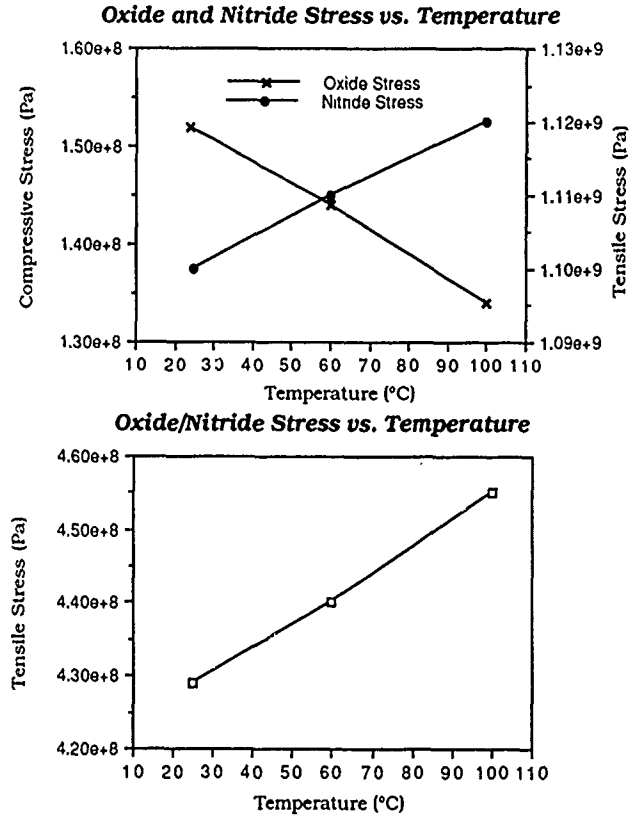


insensitive to internal stress. Previously reported values for TCS range from 0-300$ppm/°C$. The TCS values presented in Table 2 are as much as an order of magnitude higher than these values. The pressure sensitivity of devices with thin, large plates is a strong function of internal stress and the principle mechanism responsible for the high TCS is believed to be the dependence of stress on temperature.

This mechanism can be verified by comparing the theoretical sensitivity to the measured sensitivity. An exact solution of the effect of stress and temperature on the membrane is extremely difficult, expansion and contraction of the different materials creates membrane bowing which could change the boundary conditions at the plate edges. However, simple stress models can be used to predict the pressure response. The pressure sensitivity of ultrasensitive membranes is given by [5]:

$$S_{tension} = \frac{a^2}{8\sigma_i hd} \qquad (4)$$

where $a$ is the half plate length; $h$ is the diaphragm thickness; $d$ is the plate separation and $\sigma_i$ is the internal stress. By determining the internal stress as a function of temperature, the pressure sensitivity can be roughly calculated. The stress as a function of temperature is given by:

$$\sigma_i = \sigma_o + \alpha E \Delta T \qquad (5)$$

where $\sigma_i$ is the internal stress; $\sigma_o$ is the stress at 25°C; $\alpha$ is the expansion coefficient; $E$ is the Young's modulus and $\Delta T$ is the change in temperature from 25°C. The value for $\sigma_o$ was determined by using pull-in voltage structures; the derived $\alpha$ value from the materials characterization was also used. Once the stress of each material is known, the composite stress can be computed from [10]:

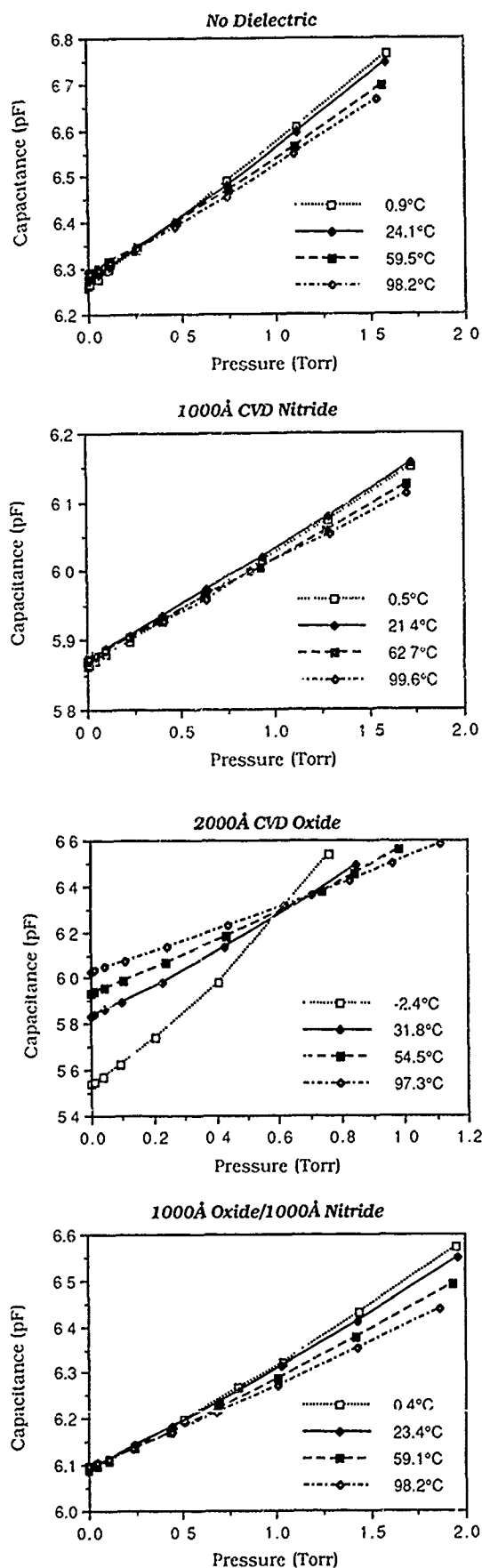$$\sigma_c t_c = \sigma_1 t_1 + \sigma_2 t_2 + ... \qquad (6)$$

185

## No Dielectric

Capacitance (pF) vs Pressure (Torr)

Legend:
- 0.9°C
- 24.1°C
- 59.5°C
- 98.2°C

## 1000Å CVD Nitride

Capacitance (pF) vs Pressure (Torr)

Legend:
- 0.5°C
- 21 4°C
- 62 7°C
- 99.6°C

## 2000Å CVD Oxide

Capacitance (pF) vs Pressure (Torr)

Legend:
- -2.4°C
- 31.8°C
- 54.5°C
- 97.3°C

## 1000Å Oxide/1000Å Nitride

Capacitance (pF) vs Pressure (Torr)

Legend:
- 0 4°C
- 23.4°C
- 59.1°C
- 98.2°C

Fig. 4: Capacitance versus Pressure of Thin p++ Silicon Membranes Coated with Various Deposited Dielectrics Over Temperature.

where $\sigma_c$ is the composite stress; $t_c$ is the total thickness and $\sigma_{1,2}$ and $t_{1,2...}$ are the corresponding stresses and thicknesses of each component material. The measured pressure sensitivity and the theoretical sensitivity are given in Table 3. The two values track each other well and the maximum error is about twenty percent.

Previously reported TCO values range from 0-100 $ppm/°C$ and this offset was attributed to the mismatch in thermal expansion coefficients for silicon (2.6 x $10^{-6}/°C$) and glass (3.2 x $10^{-6}/°C$). However, for thin boron-doped silicon membranes, the dielectric not only affects the internal stress of the diaphragm, but also the thermal expansion of the plate. If large mismatches exist in thermal expansion coefficients, the membrane could bow, which would create large zero pressure offsets. In general, the devices which have the lowest TCO's are the membranes coated with nitride, while a very large TCO exists for the membranes with oxide. This is due to the close match in expansion coefficient between the silicon and nitride (2.4 x $10^{-6}/°C$) and the large mismatch with oxide (0.5 x $10^{-6}/°C$).

If high TCO's are created by thermal expansion mismatches, it should be possible to reduce this TCO by compensating the oxide thermal expansion coefficient with a nitride. An ultrasensitive membrane (160 $ppm/mTorr$, 2.6$\mu m$ thick) was fabricated with a 320$nm$/56$nm$ oxide/nitride dielectric coating and was tested over temperature. Although this device was 50% more sensitive than the membrane with 200$nm$ of oxide, the TCO was reduced by 50% (408$ppm/°C$). However, due to the thicker dielectrics, the TCS was higher (9098 $ppm/°C$).

During all pressure testing, hysteresis was monitored. The hysteresis observed was <0.2% full scale (FS) on all devices, where the full scale pressure is defined to be the pressure when the diaphragm deflects half the plate separation distance ($\Delta C=C_0$).

Table 1: Pressure Sensitivity of Thin p++ Silicon Membranes with Deposited Dielectrics At Different Temperatures (ppm/mTorr)

| Material | 0°C | 25°C | 60°C | 100°C |
|---|---|---|---|---|
| No Dielectric | 46.7 | 43.0 | 37.5 | 35.1 |
| 2000Å SiO2 | 177 | 113 | 92.2 | 72.8 |
| 1000Å Si3N4 | 28.0 | 28.2 | 24.8 | 24.6 |
| 1000Å/1000Å SiO2/ Si3N4 | 33.7 | 32.2 | 29.4 | 26.9 |

Table 2: Temperature Coefficients of Offset and Sensitivity

| Material | TCO (ppm/°C) | TCS (ppm/°C) |
|---|---|---|
| No Dielectric | 41 | 2772 |
| 2000Å SiO2 | 841 | 7014 |
| 1000Å Si3N4 | 15.5 | 1225 |
| 1000Å/1000Å SiO2/ Si3N4 | 15.1 | 2159 |

Table 3: Measured and Theoretical Pressure Sensitivity Over Temperature ($ppm/mTorr$)

| Membrane Type | Temp.(°C) | $S_{measured}$ | $S_{theoretical}$ |
|---|---|---|---|
| No Dielectric | 25 | 43.0 | 47.4 |
| | 60 | 37.5 | 38.5 |
| | 100 | 35.1 | 31.6 |
| 2000Å Oxide | 25 | 113 | 137 |
| | 60 | 92.2 | 99.2 |
| | 100 | 72.8 | 75.1 |
| 1000Å Nitride | 25 | 28.2 | 32.4 |
| | 60 | 24.8 | 29.4 |
| | 100 | 24.6 | 28.1 |
| 1000Å/1000Å Oxide/Nitride | 25 | 32.2 | 37.2 |
| | 60 | 29.4 | 33.5 |
| | 100 | 26.9 | 31.7 |

186

## CREEP AND FATIGUE

A major concern in the development of thin membranes for sensor applications is the possibility of plate stretching or plastic deformation. In order to better understand the stability and robustness of boron-doped silicon diaphragms, tests for creep (constant stress at high temperature) and fatigue (repeated bursts of stress) were performed. Creep testing involved exposing dielectrically coated membranes (320nm/560nm oxide /nitride) and membranes without dielectrics to overpressures 10 times the full scale range at 100°C for 75 hours; fatigue testing involved pressure pulsing a membrane deposited with dielectrics (320nm/560nm oxide /nitride) with pressure bursts 10 times the full scale range at 0.5Hz. The diaphragm was stressed for a total of 10,000 pulses.

The zero pressure offset and pressure resolution ($\Delta C/\Delta P$) were measured to evaluate the effect of both creep and fatigue. For creep, the change in offset was <0.2% FS for both types of membranes. The membrane with dielectric had a change in pressure resolution of +0.028fF/mTorr and the membrane without dielectrics had an increase in pressure resolution of 0.007fF/mTorr. The measurement accuracy is estimated to be approximately 0.020fF/mTorr; thus, the change in pressure resolution is not considered to be signifi_ant.

Figure 5 presents the offset and pressure resolution as a function of pressure cycles. The zero pressure offset is <0.1% FS and the change in resolution is 0.016fF/mTorr. Since the dielectrics are much more brittle than the boron-doped silicon, it can be assumed that fatigue does not affect thin membranes without dielectrics. The hysteresis was <0.2% FS for both creep and fatigue testing.

## EXTENDED RANGE OPERATION

In designing capacitive pressure sensors, a major trade-off exists between pressure range and sensitivity. By increasing the sensitivity, the pressure range is decreased. A typical pressure range extends over three orders of magnitude and has a resolution of 10 bits. The creep and fatigue results indicate that the device performance does not significantly degrade at overpressures of 10 times the full scale range; thus, it may be possible to extend the range another order of magnitude.

Figure 6 shows the overpressure characteristic of a capacitive pressure sensor. At full scale range, the membrane deflects only half the separation distance of the two capacitor plates; at approximately twice the full scale pressure, the diaphragm touches the bottom metal plate and begins to spread out against this plate. The glass acts as an overpressure stop which prevents the membrane from rupturing and the dielectric prevents the capacitor plates from shorting together. The pressure range of the sensor can be extended into this overpressure region, which would increase the resolution to 16 bits.

## CONCLUSIONS

Experimental results involving the effects of temperature, hysteresis, creep and fatigue on capacitor pressure sensor performance have been reported. Temperature offset and pressure sensitivity coefficients can increase by more than an order of magnitude due to the mismatch in thermal expansion coefficients of silicon and dielectrics and the dependence of stress on temperature. Creep and fatigue show no effect on offset or pressure resolution; there is no evidence that plastic deformation takes place at high pressures. Based on these results, there exists the possibility of extending the pressure range by an order of magnitude.
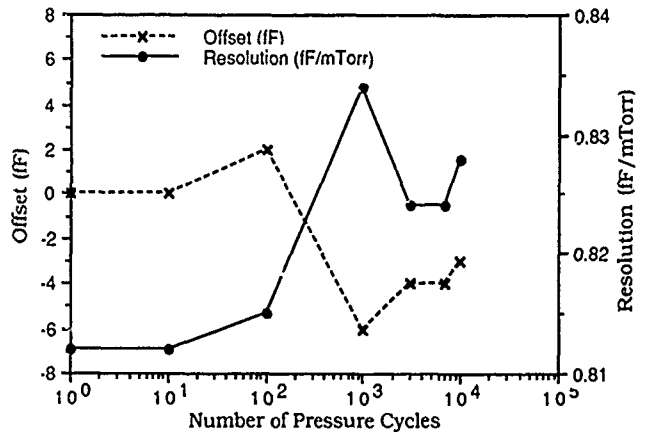
## ACKNOWLEDGEMENTS

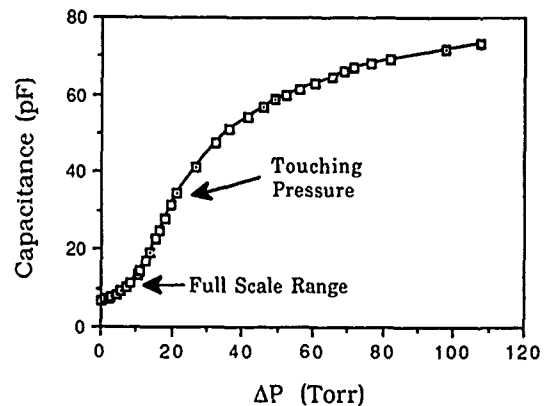Fig. 5: Offset and Resolution versus Number of Pressure Cycles.



Fig. 6. Capacitance versus Pressure. The Full Scale Range is Approximately 10 Torr.

## REFERENCES

[1] H.L. Chau and K.D. Wise, "An Ultraminiature Solid-State Pressure Sensor for a Cardiovascular Catheter," IEEE Trans. Electron Devices, pp. 2355-2362, December 1988.

[2] J.T. Suminto, G.-J. Yeh, T.M. Spear, and W. Ko, "Silicon Diaphragm Capacitive Sensor for Pressure, Flow, Acceleration and Attitude Measurements", Transducers '87, pp. 336-339, July, 1987.

[3] S.T. Cho, K. Najafi, C.L. Lowman, and K.D. Wise, "An Ultrasensitive Silicon Pressure-Based Flowmeter," IEDM Tech. Digest, pp. 499-502, Dec. 1989

[4] H.L. Chau and K.D. Wise, "Scaling Limits in Batch-Fabricated Silicon Pressure Sensors," IEEE Trans. Electron Devices, Vol. ED-34, No.4, pp. 850-858 , April 1987.

[5] S.T. Cho, K. Najafi and K.D. Wise, "Scaling and Dielectric Stress Compensation of Ultrasensitive Boron-Doped Silicon Microstructures," Proc. IEEE Workshop on Microelectromech. Sys., pp. 50-55, Feb. 1990.

[6] N.F. Raley, Y. Sugiyama, and T. Van Duzer, "(100) Silicon Etch-Rate Dependence on Boron Concentration in Ethylenediamine Pyrocatechol Water Solutions," J. Electrochem. Soc., Vol. 131, pp.161-171, January 1984.

[7] K. Najafi and K. Suzuki, "A Novel Technique and Structure for the Measurement of Intrinsic Stress and Young's Modulus of Thin Films," Proc. IEEE Workshop on Microelectromech. Sys, pp. 96-97, Feb. 1989.

[8] Retajcyk and Hill, Appl. Phys. Lett., Vol.36, p.161, 1980.

[9] Werner Riethmuller and Wolfgang Benecke,"Thermally Excited Silicon Microactuators", IEEE Trans. Electron Devices, Vol. ED-35, No. 6, pp.758-762, June 1988.

[10] Y.S. Lee and K.D. Wise, "A Batch-Fabricated Silicon Capacitive Pressure Transducer with Low Temperature Sensitivity", IEEE Trans. Electron Devices, Vol. ED-29, pp.42-48, January 1982.

[11] O. Tabata, K. Kawahata, S. Sugiyama, and I. Igarashi, "Mechanical Property Measurements of Thin Films Using Load-Deflection of Composite Rectangular Mer _oran " Proc. IEEE Workshop on Microelectromechanical Systems, pp.152 156, F. b. '989.

# Surface-Micromachined Piezoelectric Pressure Sensors

P. Schiller and D. L. Polla
University of Minnesota
Department of Electrical Engineering
Minneapolis, Minnesota 55455
(612)   625-4873

M. Ghezzo
General  Electric Company
Corporate Research and Development Center
Schenectady, New York 12301

Fig. 1. Cross-section schematic of the surface-micromachined piezoelectric pressure sensor.

## ABSTRACT

Piezoelectric pressure sensors based on zinc oxide (ZnO) thin films have been fabricated in a silicon *IC*-compatible process using surface micromachining techniques. The diagnostic pressure sensors fabricated in this work have dimensions ranging from 50x50 $\mu m^2$ to 250x250 $\mu m^2$ and make use of 0.95 $\mu m$-thick sputtered ZnO piezoelectric capacitors supported 0.6 $\mu m$ above the silicon substrate on 2.0 $\mu m$-thick polysilicon membranes. These sensors exhibit approximately 0.36 mV/$\mu bar$ sensitivity and 3.4 dB variation over the range 200 Hz to 40 kHz.

## INTRODUCTION

This paper describes surface-micromachined piezoelectric ZnO pressure sensors fabricated on deformable polysilicon membranes. The process is completely compatible with CMOS technology because it uses conventional fabrication techniques, materials, and chemicals[1]. Integrating microsensors and circuitry onto a single chip reduces power consumption and allows localization of information processing operations to be carried out over a small area. Previous reported pressure/acoustic microsensors have been fabricated using bulk-machining techniques [2-4] and/or hybrid sealing techniques [5-6]. Bulk machining severely limits device density and weakens the mechanical integrity of the wafer. Front-to-backside alignment unnecessarily complicates the process while silicon etchants such as KOH and EDP (or EPW) can potentially lower the yield of working circuits on the wafer. Hybrid bonding and sealing techniques often require prolonged high temperature processing steps after the formation of circuits.

## PROCESS APPROACH

This paper presents an alternate approach to solid-state pressure sensor fabrication and is based on two principle technologies.

1) Piezoelectric pressure sensors have been formed by a planar surface-micromachining process. This eliminates the need for front-to-back side alignment, does not weaken the mechanical integrity of the silicon wafer through bulk-anisotropic etching, and does not complicate the fabrication of on-chip circuitry.

2 ) Dielectric sputtering is used as a directional-sealing technique to form an enclosed pressure reference cavity. This eliminates the need for high temperature oxidation or anodic bonding as a cavity sealing technique.

## FABRICATION

Initially 1000 Å thermal oxide is grown to electrically isolate the sensor from the silicon substrate. A lower stationary electrode is formed by depositing 0.1 $\mu m$ LPCVD polysilicon and is electrically encapsulated by 0.1 $\mu m$ LPCVD $Si_3N_4$ (see Fig. 1). A 0.6 $\mu m$-thick spacer oxide is then deposited by LPCVD in two steps. The first 0.3 $\mu m$ of low-temperature oxide is deposited without doping and is followed by a second 0.3 $\mu m$ deposited with 7% phosphorus doping. Prior to patterning, the sacrificial layer is densified in nitrogen at 1000 °C for 20 minutes. Partial redistribution of impurities takes place during this step allowing phosphorus atoms to diffuse into the previously undoped LTO layer. A taper in the spacer sidewall is formed due the higher chemical
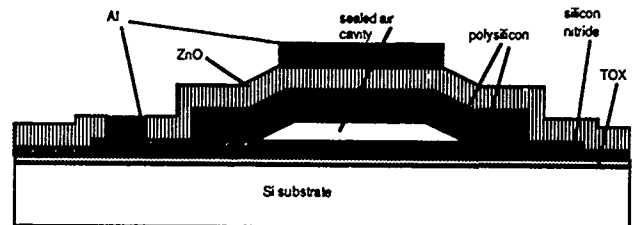
etching rate in phosphorus-doped oxide. Phosphorus-doped polysilicon is next deposited and defined in the form of a 2.0 $\mu m$-thick film. Thermal annealing at 1000°C for 30 minutes is then carried out to relieve stress in the polysilicon membrane. Lateral etching (or surface-micromachining) of the spacer oxide is used to form an electrically conductive polysilicon structural support and cavity for the sensor. The resulting microbridge structure formed is shown in Fig. 2.

The active ZnO thin film is next deposited by rf planar magnetron sputtering to a thickness of 0.95 $\mu m$. In addition to forming the active piezoelectric film, the directional nature of sputtering also seals the sidewalls of the pressure sensor cavity in an ambient of 50% $O_2$/50% Ar at 10 mT. Contact openings are cut and a top metallization is then deposited allowing for the simultaneous readout of both a piezoelectric signal induced across the ZnO film and change in capacitance between the polysilicon and lower stationary electrode.
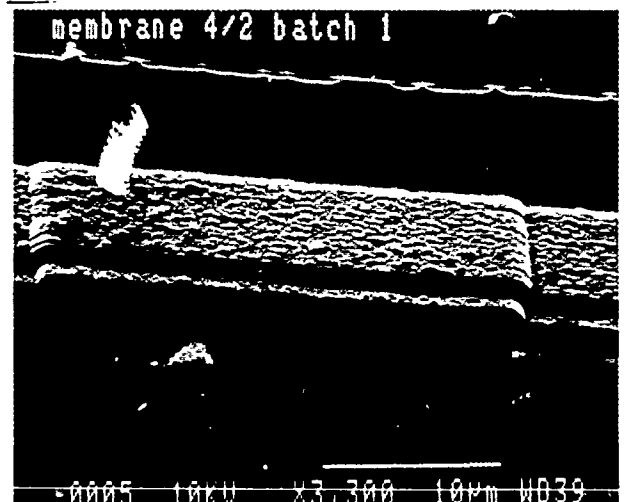


Fig. 2. Surface-micromachining  of  polysilicon  mechanical support structure.

## SENSITIVITY

The derivation of the theoretical pressure sensitivity is based on the mechanical theory of deformable plates [7]. The coordinate axes and dimensional variables used in this derivation are shown in Fig. 3. The deflection, *w*, of a rectangular plate with all edges built-in can be obtained by solving the following differential equation under appropriate boundary conditions
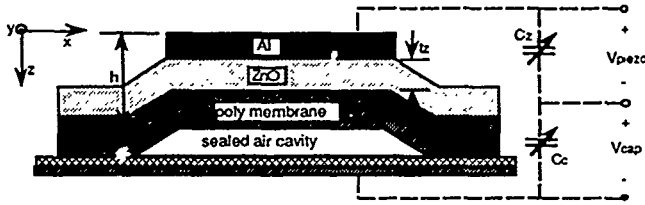
Fig. 3. Electrical equivalent circuit showing both piezoelectric and capacitive readout.

$$\frac{d^4w}{dx^4} + 2\frac{d^4w}{dx^2dy^2} + \frac{d^4w}{d^4y} = \frac{q}{D} \qquad (1)$$

where

$$D = \frac{Eh^3}{12(1-\upsilon^2)} \qquad (2)$$

and $q$ is the loading force intensity assumed to be uniform across the plate, $E$ is the modulus of elasticity, $\upsilon$ is Poisson's ratio, and $h$ is the thickness of the plate. The detailed derivation and solution is given in [7,8]. From the solution for deflection, the normal stress components, $\sigma_x$ and $\sigma_y$, are given by

$$\sigma_x = -\frac{Ez}{1-\upsilon^2}(\frac{d^2w}{dx^2} + \upsilon\frac{d^2w}{dy^2}) \qquad (3)$$

and

$$\sigma_y = -\frac{Ez}{1-\upsilon^2}(\frac{d^2w}{dy^2} + \upsilon\frac{d^2w}{dx^2}). \qquad (4)$$

Since the crystal structure of ZnO is classified as dihexagonal polar, the principal equations for the direct piezoelectric effect can be expressed as follows [8]

$$\begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{15} & 0 & 0 \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \tau_{yz} \\ \tau_{zx} \\ \tau_{xy} \end{bmatrix}$$

(5)

For ZnO the piezoelectric coefficients are $d_{15} = -8.3\times10^{-12}$ C/N, $d_{31} = -5\times10^{-12}$ C/N, and $d_{33} = 12.4\times10^{-12}$ C/N. The components of polarization are represented by $p_x$ - $p_z$, and $\sigma$ and $\tau$ represent the normal and shearing stress components, respectively. The average stress components are calculated by integrating Eqn. (3) and (4) over the area of the membrane and dividing by the total membrane area

$$\sigma_{xav} = \frac{4}{ab}\int_0^{a/2} \int_0^{b/2} \sigma_x \, dxdy \qquad (6)$$

and

$$\sigma_{yav} = \frac{4}{ab}\int_0^{a/2} \int_0^{b/2} \sigma_y \, dxdy . \qquad (7)$$

where $a$ and $b$ are the length and width of the membrane, respectively. The average stress components in the piezoelectric film, $\sigma'_{xav}$ and $\sigma'_{yav}$, are then calculated by integrating equations (5) and (6) over the thickness of the film.

$$\sigma'_{xav} = \frac{1}{t_z}\int_{-h/2}^{-(h/2 - t_z)} \sigma_{xav} \, dz \qquad (8)$$

and

$$\sigma'_{yav} = \frac{1}{t_z}\int_{-h/2}^{-(h/2 - t_z)} \sigma_{yav} \, dz . \qquad (9)$$

The average polarization is then obtained from

$$P_{zav} = d_{31}(\sigma'_{xav} + \sigma'_{yav}) \qquad (10)$$

Since the electrodes are formed normal to the z axis, the induced charge on the electrodes is nearly independent of $\sigma_z$. From Eqn. (10), the unamplified sensitivity of this device is

$$S = \frac{V_m}{q} = \frac{P_{zav}}{qC_z} = \frac{d_{31}(\sigma'_{xav} + \sigma'_{yav})}{qC_z} \qquad (11)$$

where $C_z$ is the capacitance per unit area between the upper and lower electrodes. The details of the solution reveal that the sensitivity is linearly dependent on the side length of the device and is maximized when the membrane is square. A complete derivation is given in Ref. 8.

For the geometries used in this work, the theoretical unamplified sensitivity of the piezoelectric pressure sensor is calculated to be 293 $\mu V/\mu bar$ for a 20x20 $\mu m^2$ device. Given a square membrane of side length $a$, the theoretical sensitivity is given by

$$S = \frac{a}{20}(293 \ \mu V/\mu bar) \qquad (12)$$

where the dimension $a$ is in units of micrometers.

## RESULTS

The sensitivity was measured using a calibrated acoustic source and an EG&G PAR 5209 lock-in-amplifier. A sensitivity of 0.36 mV/$\mu$bar at 1.4 kHz, T=23 °C was measured for a device with ZnO electrode dimensions of 145x145 $\mu m^2$ as shown in Fig. 4. The frequency response of a 50x50 $\mu m^2$ device is shown in Fig. 5.

The measured responsivities are in reasonable agreement (and somewhat lower) with that predicted by theory. The lower responsivity may be due in part to the dc atmospheric pressure differential across the membrane. Since the cavity is sealed at 10 mT, the membrane may be slightly deformed by atmospheric pressure. This deformation increases the capacitance between the deformable polysilicon electrode and the lower stationary electrode producing a reduction in sensitivity.

## CONCLUSIONS

Piezoelectric pressure sensors based on ZnO thin films have been formed in an IC-compatible fabrication process. The main features of this work include the use of surface-micromachining for cavity formation rather than bulk-micromachining and directional cavity sealing using sputtering rather than high temperature reactive sealing or hybrid bonding approaches. Measured responsivities of 0.36 mV/$\mu$bar for a sensor with active dimensions of 145x145 $\mu m^2$ suggest the use of ZnO-based pressure sensors for low-level acoustic detection applications.

189
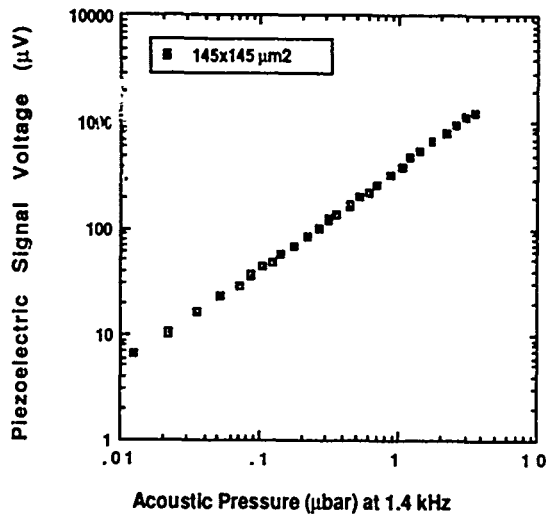
Fig. 4. Unamplified voltage responsivity versus acoustic pressure at 1.4 kHz of a piezoelectric pressure sensor with an active area of 145x145 μm².
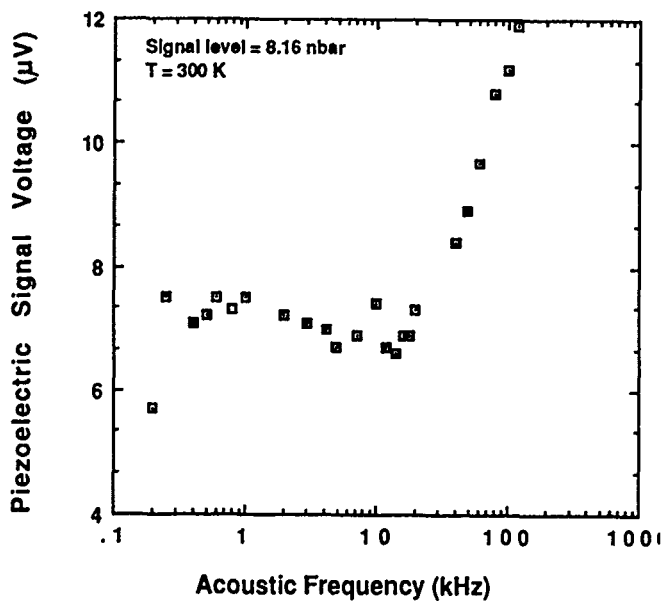


Fig. 5. Unamplified voltage responsivity versus frequency for a piezoelectric pressure sensor with an active area of 50x50 μm².

**References**

[1] D.L. Polla, H. Yoon, T. Tamagawa, and K. Voros, "Integration of Surface-Micromachined ZnO Sensors in n-Well CMOS Technology," IEEE Int. Electron Dev. Mtg., Washington, D.C., Abs. 19.2, pp. 495-498, Dec. 1989.

[2] W.H. Ko, "Solid-State Capacitive Pressure Transducers," Sensors and Actuators 10, pp. 303-320, 1986.

[3] S.K. Clark and K.D. Wise, "Pressure Sensitivity in Anisotropically Etched Thin-Diaphragm Pressure Sensors," IEEE Trans. on Electron Dev., ED-26, pp. 1887-1896, 1979.

[4] E.S. Kim and R.S. Muller, "IC-Processed Piezoelectric Microphone," IEEE Electron Dev. Lett., EDL-8, pp. 467-468, 1987.

[5] H. Guckel, D.W. Burns, H.H. Busta, and J.F. Detry, "Laser-recrystallized Piezoresistive Micro-Diaphragm Sensor," IEEE Int. Conf. Solid-State Sensors and Actuators, Philadelphia, pp. 182-185, 1985.

[6] T. Ishihara, K. Suzuki, S. Suwazono, M. Hirata, and H. Tanigawa, "CMOS Integrated Silicon Pressure Sensor," IEEE J. Solid-State Circuits, SC-22, pp. 151-156, 1987.

[7] S. Timenshenko and S. Woinowsky-Krieger, Theory of Plates and Shells, McGraw-Hill, 1959.

[8] K. Suda,"IC Compatible Pressure Sensor," M.S. Thesis, Dep't. of Elect. Eng., Univ. of Minnesota, Aug. 1988.

# AUTHOR INDEX