SANDIA REPORT

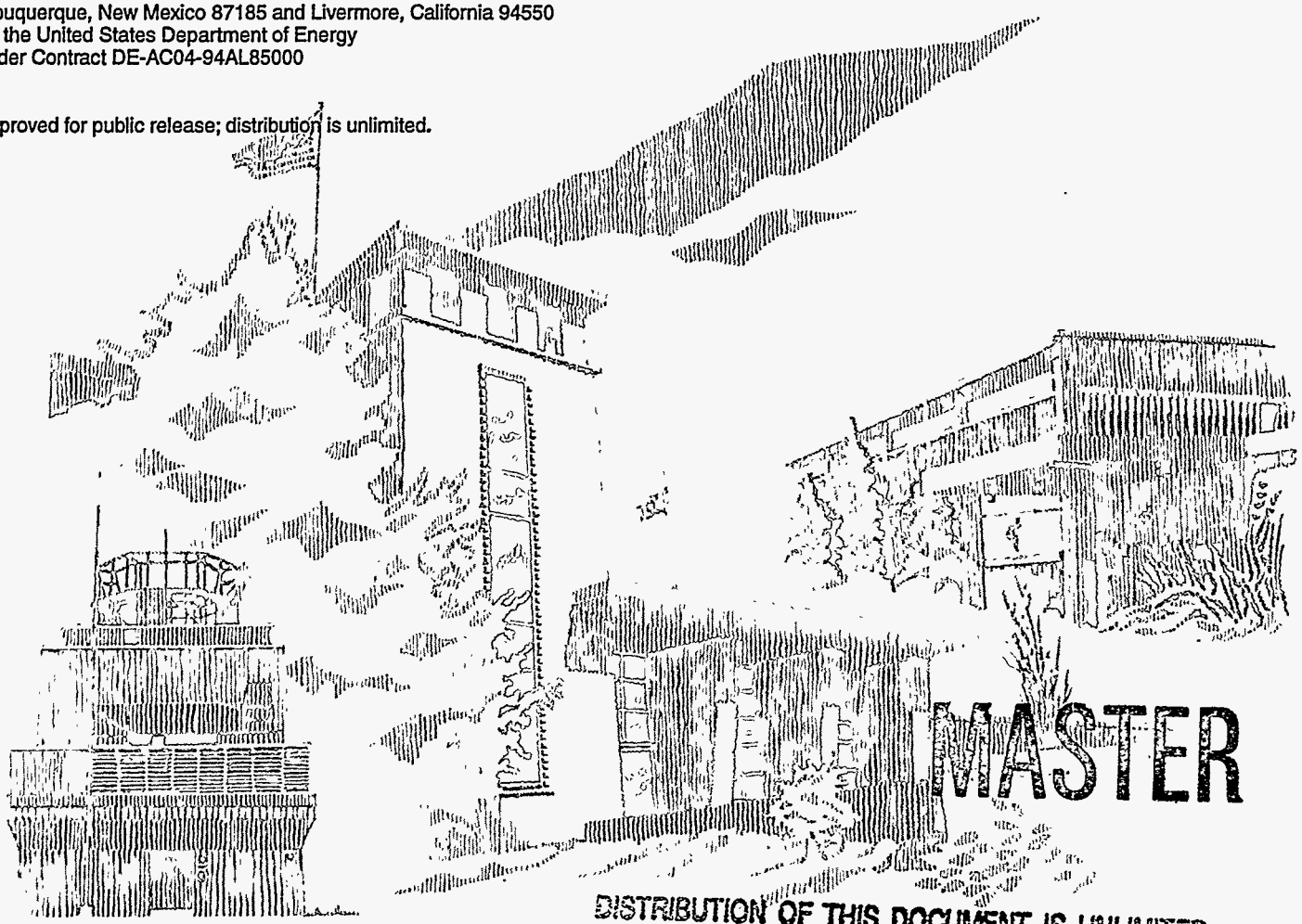
SAND96–2801 • UC–704
Unlimited Release
Printed November 1996

3D Packaging for Integrated Circuit Systems

Editors: Dahwey Chu, David W. Palmer

Author Contributors: Bruce Bainbridge, Richard Gassman, Pam Seigal, Bruce Draper, Jim Sweet, David Peterson, Simone Smith, Melaine Tuck, Jim Fleming, Glenn Laguna, Cathy Reber, Anthony Thornton, David Shen

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550
for the United States Department of Energy
under Contract DE-AC04-94AL85000



Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
Office of Scientific and Technical Information
PO Box 62
Oak Ridge, TN 37831

Prices available from (615) 576-8401, FTS 626-8401

Available to the public from
National Technical Information Service
US Department of Commerce
5285 Port Royal Rd
Springfield, VA 22161

NTIS price codes
Printed copy: A05
Microfiche copy: A01

SAND96-2801
Unlimited Release
Printed November 1996

Distribution
Category UC-704

3D Packaging For Integrated Circuit Systems

Editors:

Dahwey Chu
David W Palmer
Advanced Packaging Department 1333
Sandia National Laboratories, 87185 - 1082

Author Contributors:
Bruce Bainbridge, Richard Gassman, Pam Seigal, Bruce Draper, Jim Sweet, David Peterson, Simone Smith, Melanie Tuck, Jim Fleming, Glenn Laguna, Cathy Reber, Anthony Thornton, and David Shen

Abstract

A goal was set for high density, high performance microelectronics pursued through a dense 3D packing of integrated circuits. A "tool set" of assembly processes have been developed that enable 3D system designs: 3D thermal analysis, silicon electrical through vias, IC thinning, mounting wells in silicon, adhesives for silicon stacking, pretesting of IC chips before commitment to stacks, and bond pad bumping. Validation of these process developments occurred through both Sandia prototypes and subsequent commercial examples.

Contents

Introduction.....iii - iv
Chapter 1 – 3D IC Module – The Whole Enchilada.....1
Chapter 2 – Thermal Analysis of the 3D Module.....31
Chapter 3 – Thinning of IC chips.....39
Chapter 4 – Shallow Well in Silicon.....44
Chapter 5 – Screening ICs on the Bare Chip Level: Temporary Packaging.....58
Chapter 6 – Maskless Flip-Chip Solder Bumping Technique.....66
Conclusion 74

3D PACKAGING FOR INTEGRATED CIRCUIT SYSTEMS

INTRODUCTION:
A 30 month project was started in FY91 attempting to develop a set of design and assembly processes that would allow high density packing of ICs forming a high performance system. One continual theme for nuclear weapons is to shrink the volume used by all components including electronic components. For many of the integrated circuit components packaging volume is more than 100 times that of the IC volume. The goal of 3D packaging is to reduce the ratio toward 1 to 4 or even 1 to 2.

At the completion of this research, several customers continued support of these processes to achieve uniquely dense systems. The latest manifestation of 3D involves the stacking of image processing ICs for use in a detector array in a satellite.

The original enabling capability was the development of conductive vias through integrated circuits. That is, for each bond pad on the IC it was desired to form another electrical contact directly through the silicon on the back surface. Several processes involving laser drilling followed by oxidation and filling with conductors were developed giving initial encouragement to the effort to go 3D. Details of the silicon via research was documented in Sandia Report SAND93 - 1772. This report is complementary to the technology report in this report.

Chapter 1 will present an example of a 3D stack of multichip modules. This prototype was fabricated using silicon thinning, silicon MCM substrate processing, and silicon stacking adhesives. The MCMs were made with Sandia sensor chips so that temperature readings could be made throughout the 3D stack. The chips also have resistance heaters so that thermal loads could be applied. Extensive 3D thermal simulation was performed before and after the prototype tests.

Chapter 2 discusses the improvements in analysis brought about by the 3D MCM challenge. In particular, automatic gridding of complex packaging geometries and results using commercial thermal computer codes will be presented.

Chapter 3 will detail the processes developed for thinning the IC chips to fractions of their wafer thickness. For example, 6" wafers are typically 27 mils thick (.027"). However, early in this investigation methods were shown that created uniform 5 mil thick chips. Ordinary handling tools (tweezers, vacuum collets) were found able to handle chips with minimum breakage.

Chapter 4 will discuss the etching of shallow wells in silicon substrates. Often it proved advantageous to use recessed rectangles for the placement of ICs rather than placing the chips on a flat surface. Wells and groves were also used for spacing and alignment purposes.

Chapter 5 will discuss methods to evaluate IC chips before they are committed to 3D assemblies. Since in a 3D assembly many chips would be buried deep inside these systems. Thus it was important to make sure the chip could work over the frequency and temperature range expected before assembly. In addition, a week of high temperature burn-in is advised so that infant mortality on one chip does not bring the whole system down.

Chapter 6 discusses stud bumping. This is the ability to place a bump of gold or solder directly on the IC chip bond pads. Then the chips could be attached to the system by bumps rather than wire bonds. This ability saves about 30% on the substrate area needed for each chip. The final goal of high density would be reached by bumps, however, for Sandia's current needs wire bonds seem appropriate.

In 1991 the call for 3D packing was very faint. However, time has shown the call was real. Many computer makers now use memory chip stacking and a few high performance weapon systems sport stacked MCMs or complex chips. The processes described in this report were the result of initial research which was intended to prove feasibility. Once proved, specific applications have pushed continual development. This development often become proprietary to the development partner and is not detailed in other documents. This is similar to the 3D commercial techniques developed by Irvine Sensors and Chip Stack Inc.

CHAPTER 1 : 3D IC MODULE -- THE WHOLE ENCHILADA

PREFACE:
This chapter consists of a report presented during the IEEE's SEMITHERM 1993 meeting and then published in the IEEE Components, Hybrids, and Manufacturing Technology Society Transactions: **Analysis and Measurement of Thermal Resistance in a 3-Dimensional Silicon Multichip Module Populated with Assembly Test Chips,** James N. Sweet, David W. Peterson, Dahwey Chu, Bruce L. Bainbridge, Richard A. Gassman, and Cathy Reber.

The work combines many of the tools developed by this LDRD effort as well as some for related product oriented work: silicon MCM fabrication, use of Sandia test chips to measure internal 3D properties, computer simulation of 3D IC structures, thinning of silicon, adhesives for stacking silicon and demonstration of importance of thermal management in 3D high density microelectronics. We start this report with this complete, complex 3D example so the individual technologies discussed in subsequent chapters can be better appreciated.

Unexpected significance of this work was that we found we could gather high density thermal data on a complex 3D configuration faster than we could perform detailed computer analysis of the design. This led to further work at speeding up the computation as well as attracting business to Sandia for fast experimental thermal measurements in complex microelectronics packaging.

In general throughout this 2 1/2 year LDRD, each lesson learned in one task would modify our future efforts on other 3D packaging technologies. Sandia was learning by hands-on experience, not by top-down reasoning. Many techniques were unexpectedly found that allowed existing equipment to produce 3D assemblies, and many equally reasonable ideas were found not to work at all.

I. Introduction

As multichip module (MCM) circuit complexity and density increase, the module area also increases. This increase in module area is accompanied by longer conductor runs with a consequent increase in propagation delay. At some degree of complexity it becomes natural to think of making a transition from a two dimensional (2D) to a three dimensional (3D) architecture to reduce module volume and propagation delay. However, there are many problems in designing and constructing a 3D MCM with adequate thermal and structural properties. Although analysis is helpful in the design phase, it is highly desirable to be able to make experimental measurements to confirm theoretical predictions. The purpose of this work was to evaluate the use of special purpose test chips, designated Assembly Test Chips or ATCs, for determining the module heat transfer properties. This is similar in nature to the work of Wesling, Shiao, Chung, Pan, and de Simone in evaluation of the properties of a 2D TAB module [1]. The use of a somewhat similar special purpose thermal test chip with resistive heaters and an array of temperature measuring diodes has recently been described by Geeraerts, van Petegem, and Sansen[2]

The module described in this paper was not intended for actual system use. Rather, it is a test vehicle for studying both the problems associated with the 3D architecture and the problems of making detailed thermal resistance measurements for systems of this complexity. This paper is intended to provide some guidance and examples to other investigators who are working on thermal problems associated with complex MCMs.

II. Assembly Test Chips

The ATC03 die used to fabricate the module have been described previously [3] and are described more fully in a user's guide [4]. A drawing of the die is shown in Fig. 1. The CMOS die is square, 0.250 in. on a side, and operates at 5 V. There are four rectangular multiline polysilicon heaters located under Al triple track corrosion test structures. For the experiments described here, the heaters were either powered uniformly or with only a single heater energized. The single heater arrangement was useful for establishing large temperature gradients on the die under power. The presence of these gradients facilitates detailed comparison between experiment and theory.

The die also contains an array of 48 addressable piezoresistive stress sensing cells, each of which contains a diode for temperature measurement, as shown in Fig. 1. Most of the thermometry was done with the p^+-n diodes in the array. When a cell is addressed, its resistors and diode are connected to an output bus for measurement. In addition, a directly bondable n^+-p edge diode was used for thermometry in regions where the thermal gradients were small, such as the lower planes of the module. In such regions, an unpowered die was essentially isothermal and it was thus not necessary to resolve temperatures on a distance scale on the order of the spacing of the stress sensor cells, $\Delta x \approx 0.025$ in.

III. Substrate Fabrication

The module Si substrates were fabricated using the Sandia IC metallization and interconnect technology, as described by Chu, Reber, Draper, Sweet, and Palmer [5]. In this process, a 1.9 μm thick field oxide is grown on the Si wafer and then a 1.0 μm Al conductor pattern is formed. Usually, a PETEOS interlevel dielectric layer is then deposited and a second metal level is defined. However, for the 3D module described here, only one level of Al/Si/Cu metallization was used. Finally, P-glass passivation is deposited and bondpad windows are opened.

IV. Module Assembly

Each module plane or slice was assembled individually and electrically tested to assure functionality. A preprocessed silver filled thermoplastic adhesive film, STAYFORM 501, was used to attach the ATCs to the Si substrate [6]. This material has a published thermal conductivity of 3 W/m-°C. The 0.003 in. thick film was cut into preforms the approximate X- Y dimensions of the die, 0.250 in. square. The die attachment process was then performed on a heated work stage within a temperature range of 275-380 °C. Electrical connections between appropriate die bonding pads and the Si substrate were thermosonically attached using 0.001 in. diameter Au wire. The slice was then electrically tested prior to being assembled further.

After individual slices had been assembled and found to be electrically functional, the bottom slice was joined to the middle slice which was then joined to the top slice. A diagram of the module layout is shown in Fig. 2. Each slice is joined using two Si spacers 0.880 in. long, 0.050 in. wide and 0.050" thick. The spacers are attached in a similar manner as the die, however a thermoplastic film with a temperature range of 150-250 °C was used. Electrical connections

between slices were thermosonically attached using a 0.0005 in. thick x 0.003 in. wide Au ribbon.

The die were located in the same x-y locations on each plane, while the middle plane substrate was rotated 90° with respect to the bottom and top planes to enable the ribbon interconnect between planes, as shown in Fig. 2. A via technology, under development, is intended to replace the ribbon interconnect in actual applications. The vias would interconnect the slices with conductors running through the spacers.

A scanning electron micrograph of the top corner of a module is shown in Fig. 3. The array of light colored squares over the polysilicon heaters are actually holes cut in the P-glass passivation over the triple track corrosion sensors to make them more sensitive to external corrosive agents.

V. Diode Calibration

One of the goals of our experiments was to determine how accurately we could map a temperature distribution over the heated die surface. We performed a calibration of the diode thermometers in a Delta model 9023 oven with a type K thermocouple monitor. A thermocouple was placed inside a glass lid over the die to minimize temperature fluctuations produced by convective air currents in the oven. The forward bias *base-emitter* voltage of a diode, V_{BE}, was measured at a diode current of 10 μA using a measurement circuit described below. Data were obtained at four temperatures, T ≈ 25, 50, 75, and 100°C. When the oven was stabilized at the setpoint temperature, the 48 V_{BE} values for the diodes in the array were measured.

The details of the calibration are given in [3]. Here we summarize by noting that a quadratic fit of V_{BE} to T did not give a statistically better fit than a linear one. When T was treated as the dependent variable and V_{BE} the independent variable, the average slope relating the variables was $c_1 = \partial T/\partial V_{BE} =$ -0.488 °C/mV, with the standard error in that variable being $\sigma_{c_1} = 0.006$ °C/mV. A four sigma range corresponds to ≈ 5% overall spread. The edge n^+-p diodes had a different slope, -0.761 °C/mV with a standard error, 0.042 °C/mV.

VI.

VII. Thermal Measurements

The assembled MCM was mounted on a Cu heat sink with the same adhesive used for the spacer

attach. The wiring from the bottom slice was brought out to edge connectors on a special printed circuit board. The Cu heat sink was attached to a methanol cooled thermal heat sink with thermal grease and the MCM was protected with a plastic cover to minimize convective heat transfer effects. Thus, all data reported here relate to conduction heat transfer through the substrates and spacers of the MCM. Due to the symmetry of the module, as shown in Fig. 2, only one die was powered. Some measurements were made with all four die heaters shown in Fig. 1 powered, while others were made with only one heater energized. The one heater case magnified the thermal gradients on the chip which facilitated detailed comparison between experiment and theory. In the data reported below, the powered die was on the top plane.

A steady state power was established and the V_{BE} data were obtained after thermal equilibrium was established. The diode array was measured for the heated die and the edge diodes were measured for all die in the MCM. Data were obtained at powers of 1, 5, and 10 W to look for systematic power dependence in the thermal resistance values. For data reduction, *universal* calibration relations for the array and edge diodes, as given above, were used rather than the individual diode calibration data.

The diodes in the array could be addressed with a six-bit address word, as shown in Fig. 4. The addressed cell was then connected to the output or measurement bus through a CMOS transmission gate. Due to the circuit design, the four piezoresistors in the addressed and other cells were also connected to the diode in a series-parallel arrangement. This is shown schematically in the circuit diagram in Fig. 5. Assuming cell 1 is selected, the desired current path is to the V_{BE} bus, through the diode, and returning through the lower or common bus, as indicated by the heavy line through the cell 1 diode. However, there are many parallel current paths through components in other cells. One such path through the diode in cell 2, the n-vertical resistor in cell 2 and the n vertical resistor in cell 1 is shown by the dashed line. An equivalent circuit diagram of the measurement circuit is shown in Fig. 6. .The diode under test, D_1, is in series with R_1, the parallel combination of the piezoresistors in the cell under test. In parallel with D_1 and R_1 are the diode D_{eq}, the parallel combination of the remaining 47 diodes and R_2, parallel combination of the resistors in the other 47 cells.

The source measurement unit (SMU) serves as an additional current source which is programmed

to drive the error voltage, V_e, across R_1 to zero. The total current supplied by the SMU is denoted by I_t, It, is related to the diode current, I_d, and bypass current, I_b, by the relation,

$$I_t = I_d + I_b$$

Id is given by,

$$I_d = I_s + I_{R1}$$

where I_{R1} is the current through R_1. When the error voltage satisfies the relation, V_e=0, the current through R_1 satisfies the relation, I_{R1}=0, and thus by Eq. (2), I_d=I_s. The diode voltage, V_{BE}, is related to the SMU voltage and error voltage by,

$$V_{SMU} = V_{BE} + V_e$$

When V_e=0, V_{BE}=V_{SMU}. Thus, the diode voltage is given by the SMU voltage and the diode current by the current source setting, as shown in Fig. 4. For all measurements reported here, the current source was set at 10 μA. The measured voltage includes the voltage drop across the transmission gate which had a resistance of about 1 kΩ, corresponding to about a 10 mV drop.

As a further check on the accuracy of measurements on the top surface, we made additional measurements with a Luxtron fluoroptic temperature sensor, using a 0.050 in. diameter probe [5], [7].

VIII.Experimental Results

The experimental results reported here are for a total single chip power of 5 W. This power was applied to either a single die heater or with 1.25 W applied to each of the four heaters on a die. The 5 W power level yielded convenient temperature rises for measurement. No systematic deviation from linear dependence on chip power was observed. Figs. 7 and 8 show curve-fits to the 6x8 array of diode measured die surface temperatures for the single heater and four heater cases, respectively. The spatial relation of the die and heaters used are indicated in Fig. 2. In the single heater experiments, the die heater nearest the top slice corner was used. By symmetry, any die on the top slice should give an identical thermal response. Figs. 7 and 8 are oriented in the same way as Fig. 2. The chip heater indicated in Fig. 2 was powered in the single heater experiments, while all four heaters indicated by the white rectangles were powered in the four

heater experiments.

Fig. 7(a) shows a heated die map of the primary data (solid squares) together with two dimensional temperature surface curvefit to these 48 temperature data points. Fig. 7(b) shows the curvefit isotherms from Fig. 7(a) in the plane of the die. Fig. 8(a) shows the data and two dimensional curvefit results for the four heater case. In this case, the contour map in Fig. 8(b) shows two distinct temperature maxima near the maximum y coordinate of the die. The maximum measured ΔT was 78.1 and 66.9 °C for the single and four heater cases, respectively, where ΔT is the temperature rise above the heatsink temperature.

The finite element method (FEM) calculations discussed below predict that the chip temperature minimum will occur at the lower corner of the heated die nearest the spacer for the four heater case, while this minimum will be at the lower corner toward the module centerline for the single heater case. These predictions are confirmed by the measurements.

On the middle slice, the temperatures on the four unheated die were measured with the edge diodes. This ΔT was in the range 10-12 °C for the die not directly under the heated die and about 16 °C for the die under the heated die.

For determination of the location of excess thermal resistance it is useful to look at the temperature drops on the heated die surface, between the heated die and the edge of the top slice, and between the edge of the top slice and the heatsink. In Table I these drops are tabulated and compared to those calculated theoretically with the FEM method, as described below. The experimental ΔT die surface values were determined from the maximum and minimum temperatures derived from the diode array measurements. The FEM die surface ΔTs were determined from isotherm contour plots. The top slice experimental ΔT is the difference between the maximum of the diode array temperatures for the heated die and the minimum of the edge diode temperatures for the unheated die on the slice. The FEM top slice ΔTs were the calculated temperature differences between the maximum on the heated die and the temperature at the edge diode location on the unheated die. The top slice→heatsink experimental ΔT is the difference between the minimum edge diode temperature on the top slice and the heatsink temperature. The die surface→heatsink ΔT is the total module temperature drop. It should be

noted that the experimental maximum temperatures should be somewhat higher because none of the diodes in the array is located directly under a heater.

The Luxtron temperature measurements generally were about 2-3 °C below the diode temperature measurements due to contact resistance. Difficulties in probe positioning limit the utility of the Luxtron for measuring a heated die temperature distribution, although use of a smaller diameter probe might improve this. In regions where the thermal gradients are small, the Luxtron is a useful tool and complements the diode measurements.

IX. Analysis

Theoretical temperature distributions for the whole module were calculated using a full FEM analysis. The FEM analysis of the 3D MCM configuration involved the development of a solid model representation that was subsequently meshed and solved computationally. Pro/Engineer version 9 (Parametric Technology Corp.) served as the solid modeling tool that was used to develop the geometry of the substrates, spacers, and die [5]. For simplicity, the chip wiring and inter-slice ribbon interconnect were not included in the model. Estimates of the thermal conductance of these elements indicate that the errors associated with these approximations were negligible for our module. The resulting file was then imported by PDA Engineering's PATRAN [5] modeling program for the application of boundary conditions and the conversion of the solid geometry to a mesh of quadrilateral and hexagonal linear elements. The numerical model was then submitted to PDA's P/Thermal (QTRAN) version 2.5 solver for calculating the steady-state temperature field.

Boundary conditions for the model included a fixed-temperature lower surface of the bottom slice to simulate a heat sink. All other boundaries were adiabatic. Material properties were constant for adhesives and temperature dependent for Si. All materials were assumed to be isotropic thermal conductors.

Two meshes were constructed: the first consisted of a fairly uniform mesh for a total of 3040 elements, while the second used a much higher mesh density in the region surrounding the heated chip, for a total of 4226 elements. The remeshing operation was performed as a check on the accuracy of the solution after an examination of the results from a test case indicated that some

refinement might be necessary. Steady-state convergence to $\leq 10^{-3}$ °C was attained after 500 to 1500 iterations with CPU times varying from 15 to 45 min. on a Sun Sparc-10.

The calculated thermal profile for the single heater case is shown in Figs. 9 and 10, for the top slice and whole module, respectively. Also shown on Fig. 9 are the heater and diode sensor locations for the heated die superimposed on the temperature map. Comparison of Figs. 7 and 9 shows that measured maximum temperature change, 78.1 °C, exceeds the calculated maximum, $\Delta T = 65.5$ °C, indicating the presence of additional thermal resistance not included in the model. In the four heater case, the maxima are 66.3 and 47.2 °C for the experiment and FEM model, respectively. An FEM isotherm plot for the top surface is shown in Fig. 11. The temperature minimum in the upper right hand corner of the heated die is clearly evident.

To complement the FEM whole module analysis, an analysis of the temperature distribution over the die top surface was done using the analytical rectangular parallelepiped solver TAMS [8],[9]. For the TAMS analysis, the boundary condition on the die bottom surface must be specified. In the calculations reported here, we specified the bottom surface as a convective boundary with an equivalent resistance, R=1/hA, characteristic of that associated with the die attach and Si resistance in the actual module, where h=heat transfer coefficient and A=die area. The result of a calculation for the four heater case is shown in Fig. 11. The temperature difference, ΔT, shown in Fig. 11 is the difference in temperature between the the die surface and a die corner, x=0, y=0. This ΔT is only weekly dependent on the actual value of h used. We used an h equivalent to R= 8 °C/W for this calculation.

It can be seen that the heaters are essentially thermally isolated from each other and, as a result, the maximum chip ΔT for the single heater case would be expected to be about four times that for the four heater case for equal total power.

X. Use of diamond as a substrate material

One potential way of reducing the thermal resistance seen by a die on the upper slice is to replace the Si substrate with a synthetic CVD diamond substrate. We are presently building a module with a diamond upper slice to test this concept. FEM calculations for a module with an 0.015 in. diamond upper substrate with a thermal conductivity of 1000 W/m-K indicate that, for the single

heater case, the predicted maximum temperature is reduced from 61 to 45 °C or a thermal resistance reduction from 12.2 to 9.0 °C/W. Further reduction could be achieved by replacing the middle slice with diamond also.

XI. Discussion

The experiments and calculations described above indicate that the module has a difference in thermal resistance of about 3 °C/W between the experimental and FEM calculated values. Most of this difference appears to be a result of high die attach resistance. About 0.4 °C/W of this excess can be accounted for by the thicker die attach region used in the actual module as compared to that used in the FEM. Table I shows that the temperature distribution over the die surface is predicted accurately by the FEM but that the experimental drop over the top slice surface, between the heated die maximum temperature and the slice edge, is under predicted. On the other hand, the slice to heatsink drop appears to be predicted accurately. This last result implies that thermal properties and geometric data used for the Si slices, Si spacers, and spacer adhesive in the FEM were reasonably accurate.

If the difference between experimental and FEM thermal resistance is attributed to an excess die attach thermal resistance, this resistance, $\Delta R \approx 2.6$ °C/W, corresponds to a specific interfacial resistance of $\delta R_i \approx 1$ cm^2-°C/W. The 0.003 in. die attach layer with a k = 3 W/m-°C has a predicted specific resistance $r_{da} = 0.25$ cm^2-°C/W and thus the excess is significant. An excess of this amount is probably due to voids in the die attach region, or possibly poor bonding between the adhesive and the Si. At this writing, microscopic examination of the module was in progress to investigate the die attach region.

XII. Conclusions

The measurements reported here on the temperature distribution of a heated 3D MCM demonstrate the utility of using the ATC03 test chip with a thermometer array for deriving detailed information about the module thermal performance. Through analysis of the measurements, the thermal model may be validated and sources of excess thermal resistance identified.

In the MCM discussed in this report, the measurements suggest that themajor source of excess

resistance was in the die attach region. The behavior of the spacers, spacer adhesives, and Si substrates appears to be well described by the FEM. The predication for the reduction in thermal resistance obtained by using a diamond substrate thus appears to be on a sound basis.

Acknowledgments

We would like to thank Nao Moore and Cathy Myers for their expertise and patience in assembling the MCMs used to make the measurements reported in this paper. Thanks are also due to Simone Smith and Jim Rife for their help in making high quality SEM pictures of the MCMs.

* This work was performed at Sandia National Laboratories, supported by the DE-AC04-76DP00789.

References

[1] P. Wesling, J. Shiao, T. Chung, T. Pan, D. de Simone, "CMOS Multichip Module Test Vehicle with TAB'd Components", *Proc. 41st Electronic Components & Technology Conference*, IEEE, 1991, pp. 712-718.

[2] J. N. Sweet, M. R. Tuck, D. W. Peterson, D. W. Palmer, "Short and Long Loop Manufacturing Feedback Using a Multi-Sensor Assembly Test Chip", *IEEE Trans. Comp., Hybrids, Manuf. Technol.*, Vol. CHMT-14, p. 529, 1991.

[3] J. N. Sweet, D. W. Peterson, M. R. Tuck, *Assembly test Chip 03 (ATC03) Description & Users Manual*, Sandia National Laboratories Report (draft edition available).

[4] D. Chu, C. A. Reber, B. L. Draper, J. N. Sweet, D. W. Palmer, "Multichip Module Enablers for High Reliability Applications", *Proc. 1992 IEEE Multi-chip Module Conference, MCMC-92*, IEEE, 1992, pp. 102-105.

[5] Reference to a particular product or company implies neither a recommendation nor an endorsement by Sandia National Laboratories, nor a lack of suitable substitutes.

[6] Luxtron model 755 manufactured by Luxtron Corp., Mountain View, CA.

[7] G. N. Ellison, Thermal Computations in Electronic Equipment, Van Nostrand, NY, 1984, chaps. 7, 9.

[8] G. N. Ellison, "Methodologies for Thermal Analysis of Electronic Components and Systems", in A. Bar-Cohen, A. D. Kraus, Advances in Thermal Modeling of Electronic Components and Systems, Vol. III, ASME Press, NY, 1993, pp. 153-206.

Figure Captions

Fig. 1. ATC03 layout showing major features on the die. The four polysilicon heaters are symmetrically located with respect to the die center. Temperature sensing diodes are located in the stress sensing cells.

Fig. 2. Layout of the 3D MCM showing the placement of the ATC03 die. Note that the Si spacers are offset.

Fig. 3. SEM picture of the top corner of the 3D MCM showing one of the ATC03 die. This would correspond to the die in the lower right hand corner of Fig. 2.

Fig. 4. Measurement system used to obtain diode voltage readings from the ATC03 die.

Fig. 5. Diagram of measurement setup to measure diode characteristic in cell 1 of the ATC03 die. The desired current path is shown by the heavy line. The dashed line shows one possible bypass current path. In this case, $V_s=V_{BE}$ but Is is greater than the diode current, I_d.

Fig. 6. Equivalent circuit for the diode measurement using the Source Measurement Unit. The long dashed line shows the bypass current, I_b, while the short dashed line shows the diode current, $I_d=I_s$ when V_c and $I_{R1} = 0$.

Fig. 7. Measured temperature distribution on heated die with one heater powered at 5 W. The surface is fit to the array of 48 measured temperature points.

Fig. 8. Measured temperature distribution on heated die with four heaters powered at 1.25 W each. Note that the temperature minimum is at a corner near the spacer in this case. Note that the vertical axis range is different from that in Fig. 5.

Fig. 9. FEM calculated temperature distribution for the single heater case. Even though the heating is highly asymmetric, the temperature distribution on the middle slice is approximately the same as it would be if the top slice was heated

Fig.10. FEM temperature profile on the top slice surface for the single die heater case

Fig.11. Temperature difference, ΔT, across the top surface of an ATC03 die with four heaters powered at 1.25 W each as calculated by TAMS.

Table Captions

Measured and calculated ΔT values for the single heater (ΔT_1) and four heater (ΔT_4) cases. See text for ΔT definitions in 2nd column.

Tables

Table I

Source	ΔT Definition	ΔT_1(°C)	ΔT_4(°C)
Expt.	Die Surface	14.2	5.9
FEM	Die Surface	26.6	6.3
TAMS	Die Surface	19.2	4.8
Expt.	Top Slice	58.3	46.5
FEM	Top Slice	50.3	31.7
Expt.	Top Slice→Heatsink	19.8	19.8
FEM	Top Slice→Heatsink	15.2	15.5
Expt	Die Surf.→Heatsink	78.1	66.9
FEM	Die Surf.→Heatsink	65.5	47.2

[1] P. Wesling, J. Shiao, T. Chung, T. Pan, D. de Simone, "CMOS Multichip Module Test Vehicle with TAB'd Components", *Proc. 41st Electronic Components & Technology Conference*, IEEE, 1991, pp. 712-718.

[2] B. Geeraerts, W. Van Petegem, W. Sansen, *Proc. 9th Semiconductor Thermal Measurement and Management Symposium*, IEEE, 1993, pp. 108-111.

[3] J. N. Sweet, M. R. Tuck, D. W. Peterson, D. W. Palmer, "Short and Long Loop Manufacturing Feedback Using a Multi-Sensor Assembly Test Chip", *IEEE Trans. Comp., Hybrids, Manuf. Technol.,* Vol. CHMT-14, p. 529, 1991.

[4] J. N. Sweet, D. W. Peterson, M. R. Tuck, *Assembly test Chip 03 (ATC03) Description & Users Manual*, Sandia National Laboratories Report (draft edition available).

[5] D. Chu, C. A. Reber, B. L. Draper, J. N. Sweet, D. W. Palmer, "Multichip Module Enablers for High Reliability Applications", *Proc. 1992 IEEE Multi-chip Module Conference, MCMC-92*, IEEE, 1992, pp. 102-105.

[6] Reference to a particular product or company implies neither a recommendation nor an endorsement by Sandia National Laboratories, nor a lack of suitable substitutes.

[7] Luxtron model 755 manufactured by Luxtron Corp., Mountain View, CA.

[8] G. N. Ellison, Thermal Computations in Electronic Equipment, Van Nostrand, NY, 1984, chaps. 7, 9.

[9] G. N. Ellison, "Methodologies for Thermal Analysis of Electronic Components and Systems", in A. Bar-Cohen, A. D. Kraus, Advances in Thermal Modeling of Electronic Components and Systems, Vol. III, ASME Press, NY, 1993, pp. 153-206.

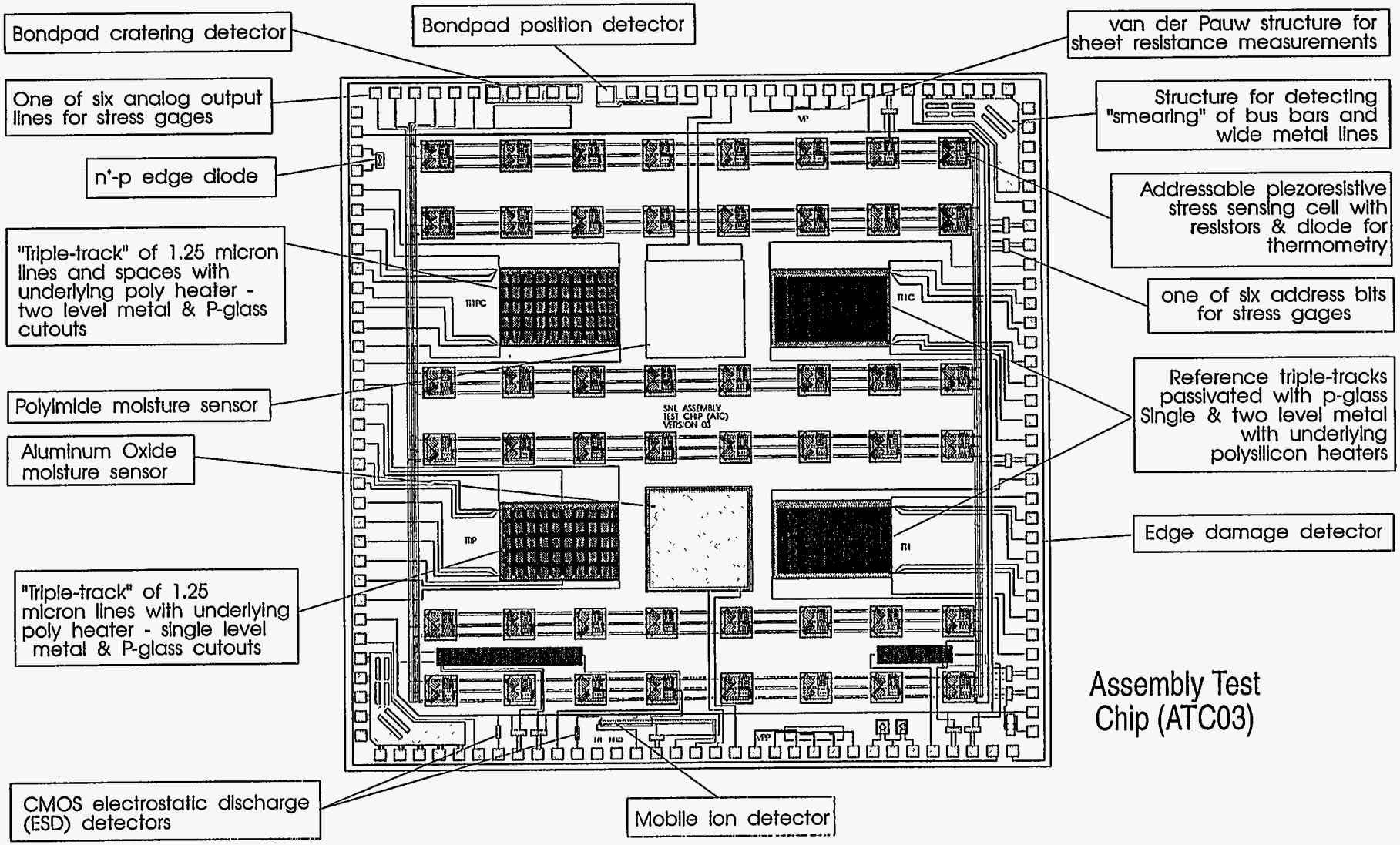

Bondpad cratering detector
Bondpad position detector
van der Pauw structure for sheet resistance measurements
One of six analog output lines for stress gages
Structure for detecting "smearing" of bus bars and wide metal lines
n+-p edge diode
Addressable piezoresistive stress sensing cell with resistors & diode for thermometry
"Triple-track" of 1.25 micron lines and spaces with underlying poly heater - two level metal & P-glass cutouts
one of six address bits for stress gages
Polyimide moisture sensor
Reference triple-tracks passivated with p-glass Single & two level metal with underlying polysilicon heaters
Aluminum Oxide moisture sensor
Edge damage detector
"Triple-track" of 1.25 micron lines with underlying poly heater - single level metal & P-glass cutouts
CMOS electrostatic discharge (ESD) detectors
Mobile ion detector
Assembly Test Chip (ATC03)

Figure 1

Ribbon
Top Si Spacer
Bottom Si Spacer
Top Si Spacer
Chip heater used for single heater experiments
Si Substrate
0.940"
1.060"
Top Si Spacer
Bottom Slice
Middle Slice
Top Slice
ATC03
ATC03
ATC03
ATC03
X
Y

Figure 2

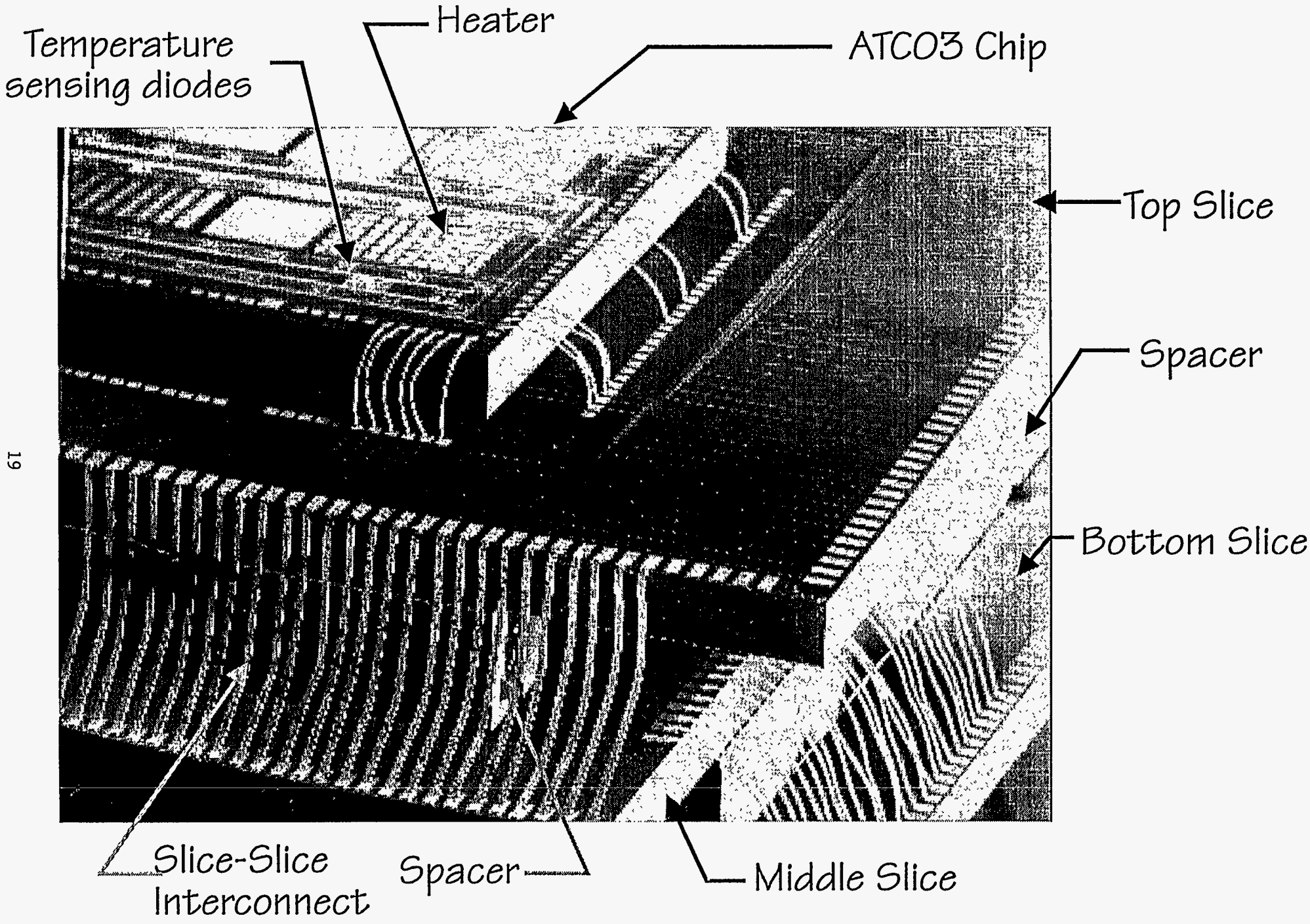
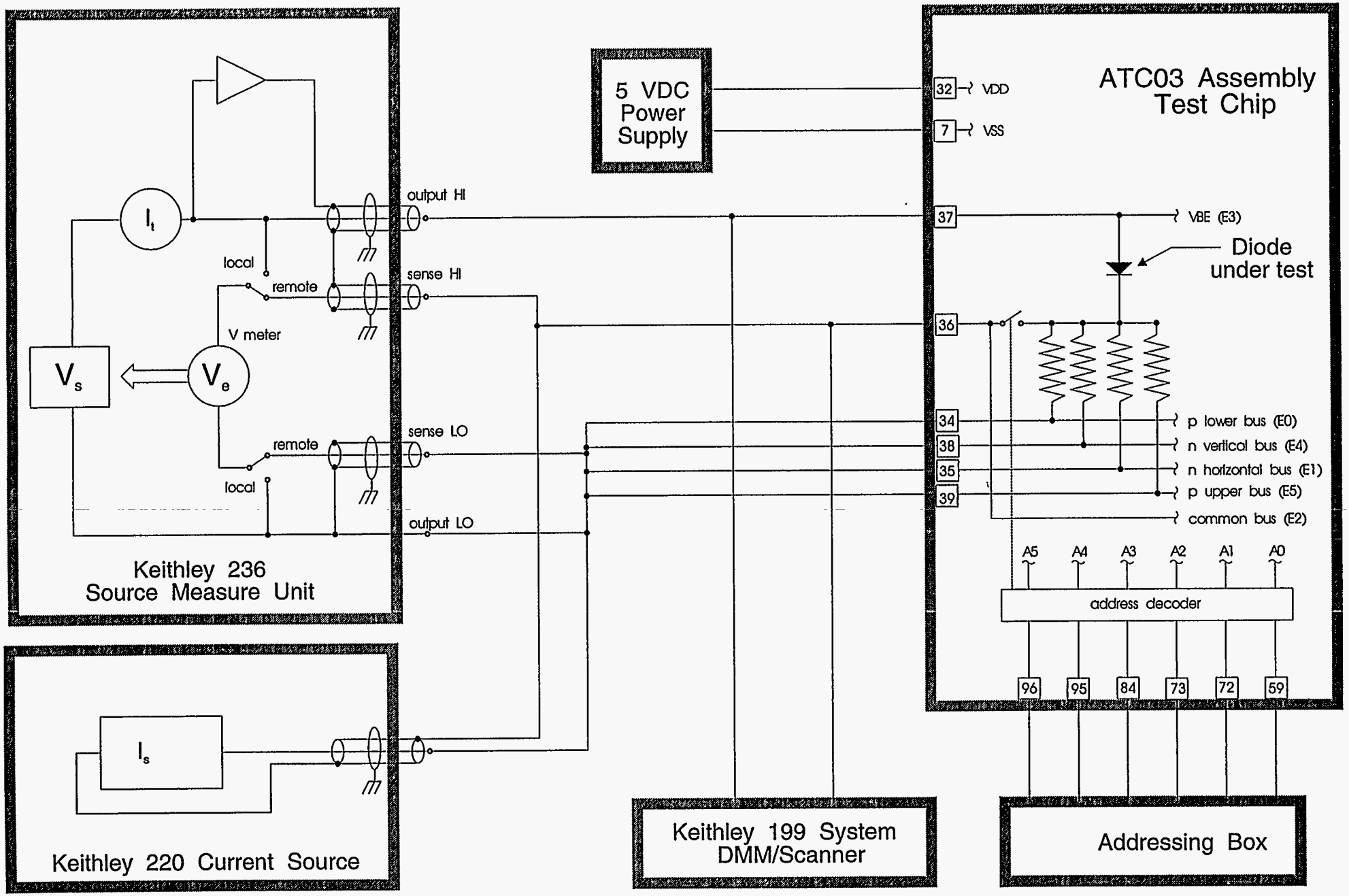

Temperature sensing diodes
Heater
ATCO3 Chip
Top Slice
Spacer
Bottom Slice
Slice-Slice Interconnect
Spacer
Middle Slice

Figure 3


5 VDC Power Supply
ATC03 Assembly Test Chip
32
VDD
7
VSS
37
VBE (E3)
Diode under test
36
34
p lower bus (E0)
38
n vertical bus (E4)
35
n horizontal bus (E1)
39
p upper bus (E5)
common bus (E2)
A5
A4
A3
A2
A1
A0
address decoder
96
95
84
73
72
59
output HI
sense HI
sense LO
output LO
local
remote
V meter
remote
local
I_t
V_s
V_e
Keithley 236 Source Measure Unit
I_s
Keithley 220 Current Source
Keithley 199 System DMM/Scanner
Addressing Box

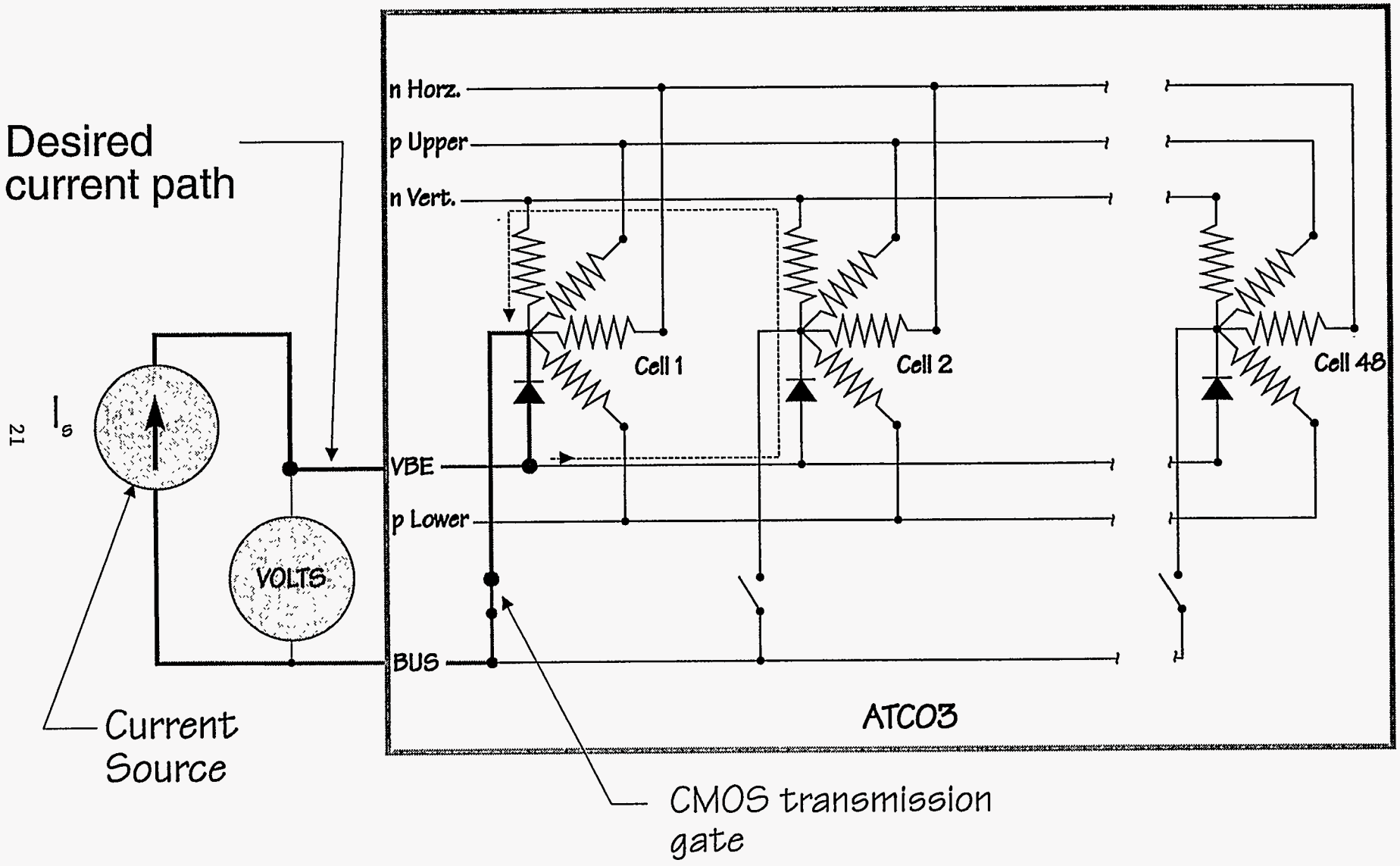
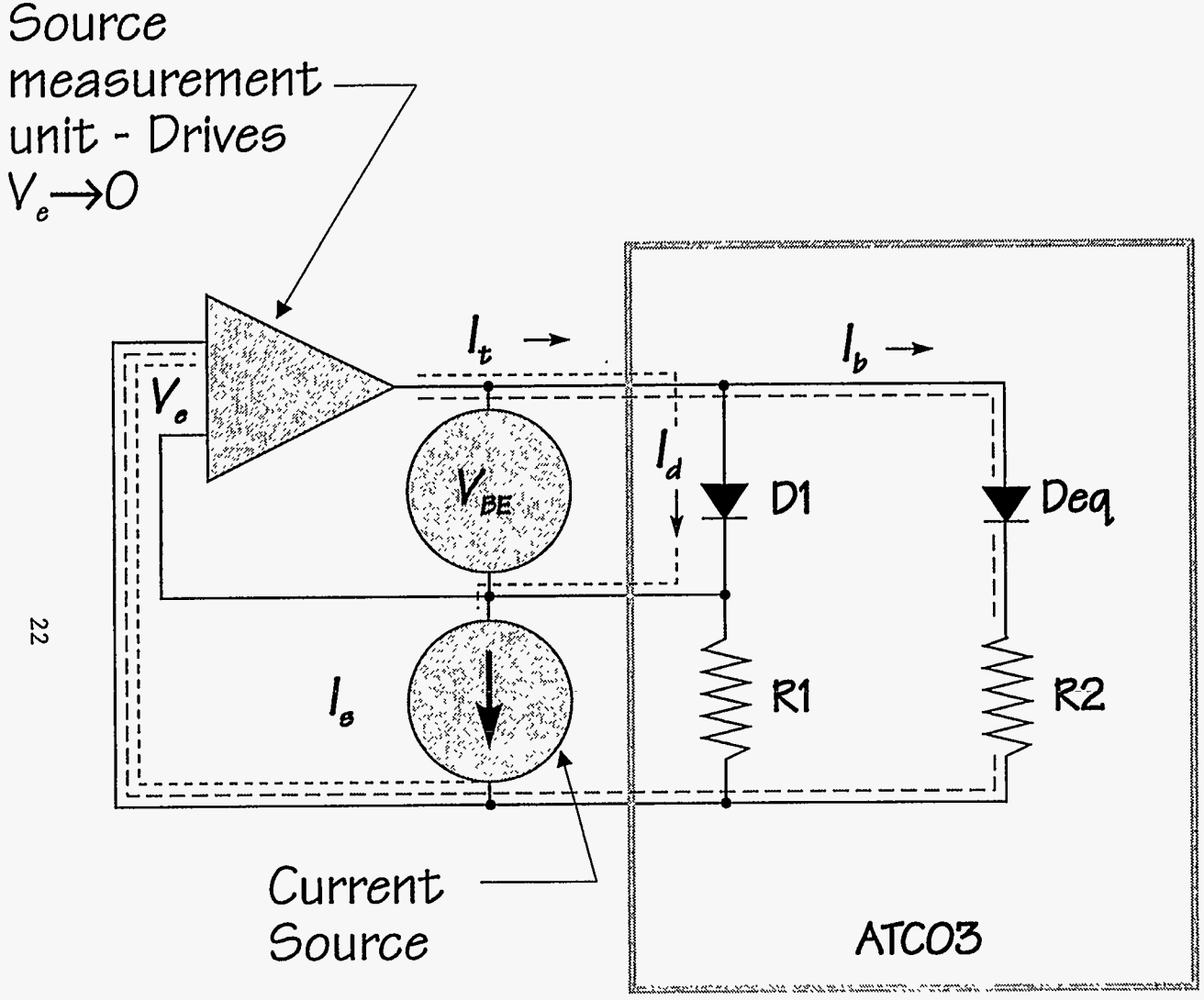

Desired
current path
n Horz.
p Upper
n Vert.
Cell 1
Cell 2
Cell 48
VBE
p Lower
BUS
VOLTS
I_s
21
ATCO3
Current
Source
CMOS transmission
gate

Figure 5



Deq: Parallel combination of remaining 47 diodes

R1: Parallel combination of piezoresistors in cell under test

R2: Parallel combination of piezoresistors in remaining 47 cells

Figure 6

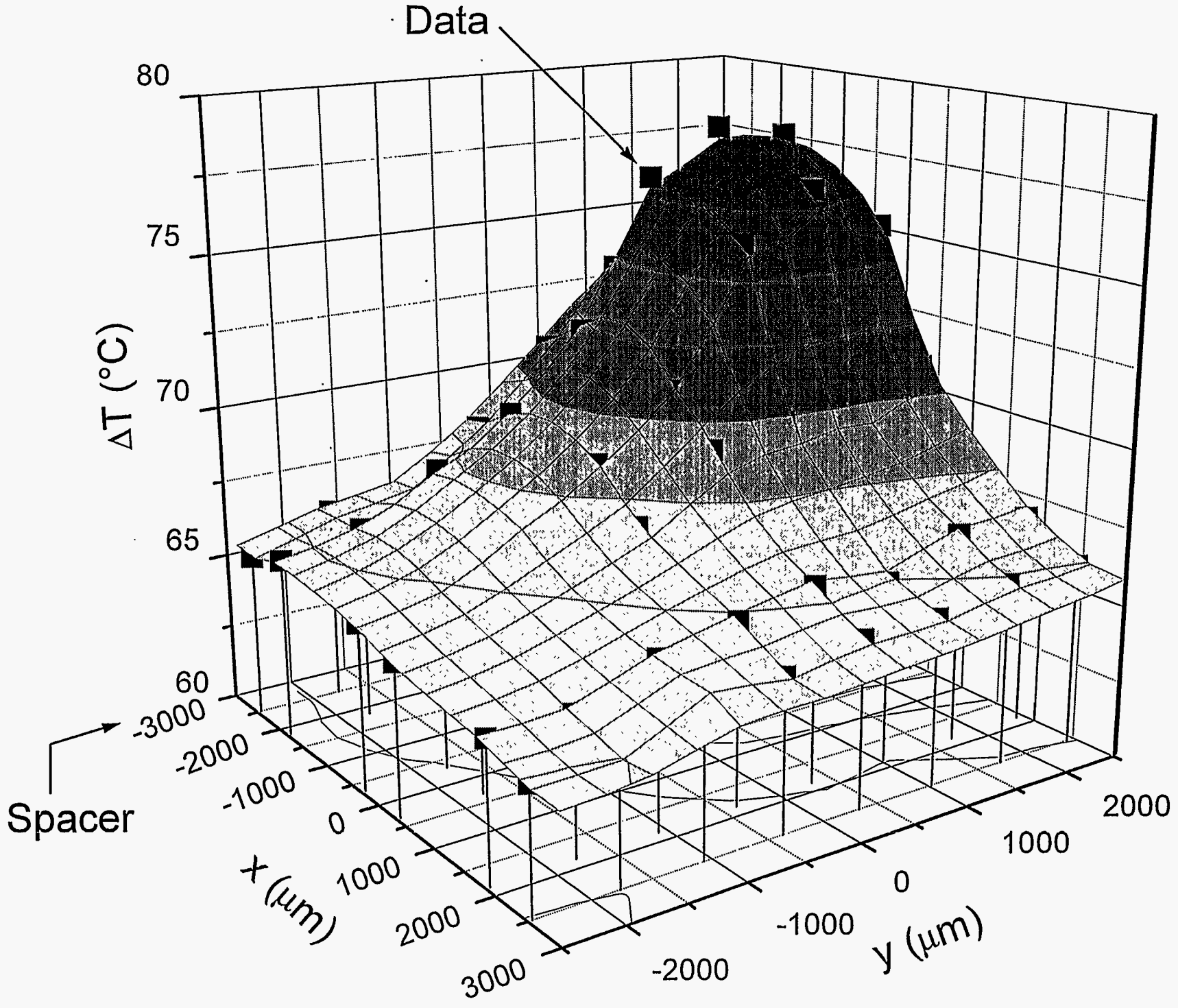

Data
80
75
70
65
60
ΔT (°C)
Spacer
-3000
-2000
-1000
0
1000
2000
3000
x (μm)
2000
1000
0
-1000
-2000
y (μm)

Figure 7
(A)

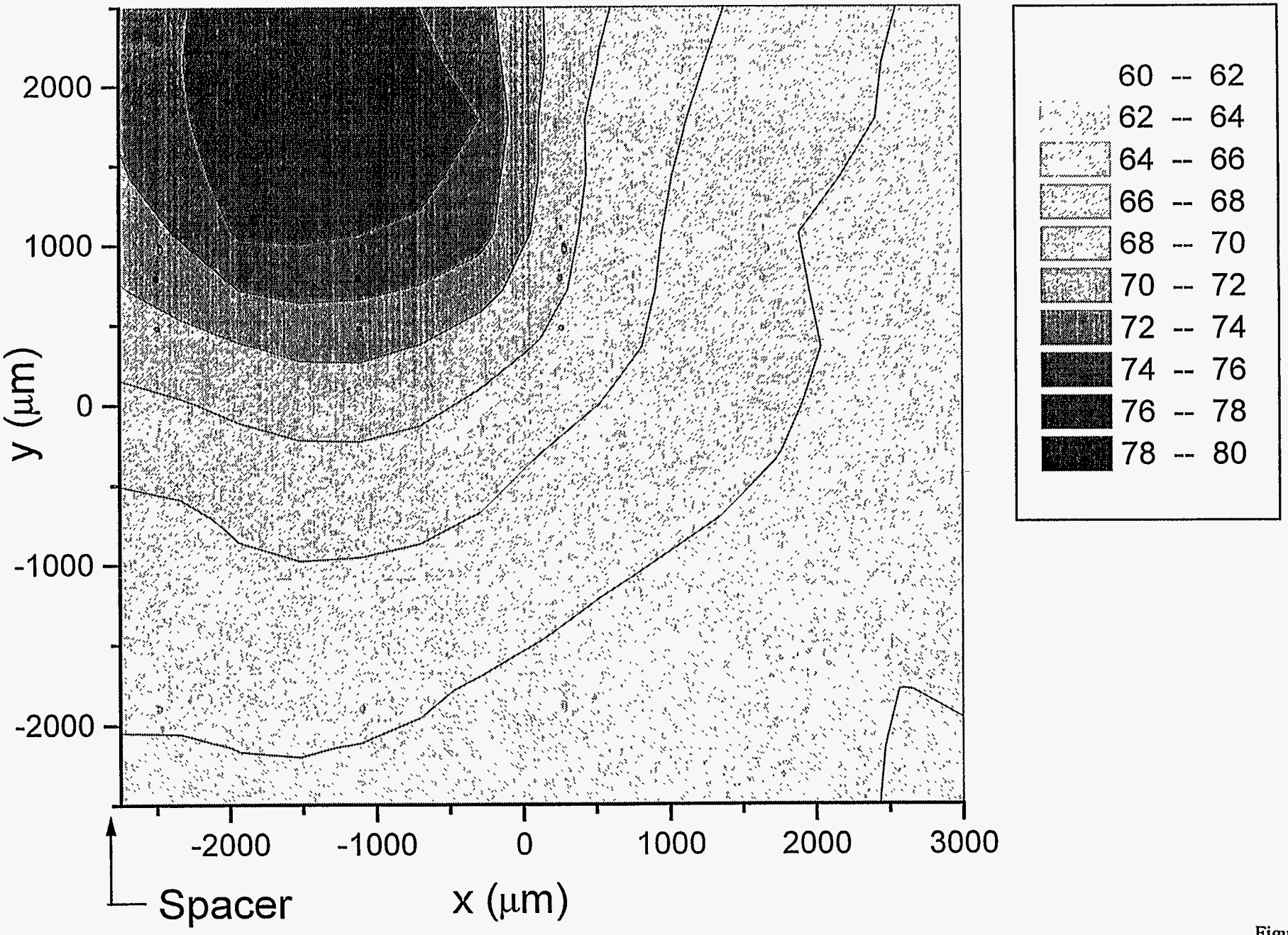

60 -- 62
62 -- 64
64 -- 66
66 -- 68
68 -- 70
70 -- 72
72 -- 74
74 -- 76
76 -- 78
78 -- 80
2000
1000
0
-1000
-2000
y (μm)
-2000
-1000
0
1000
2000
3000
Spacer
x (μm)

Figure 7
(b)

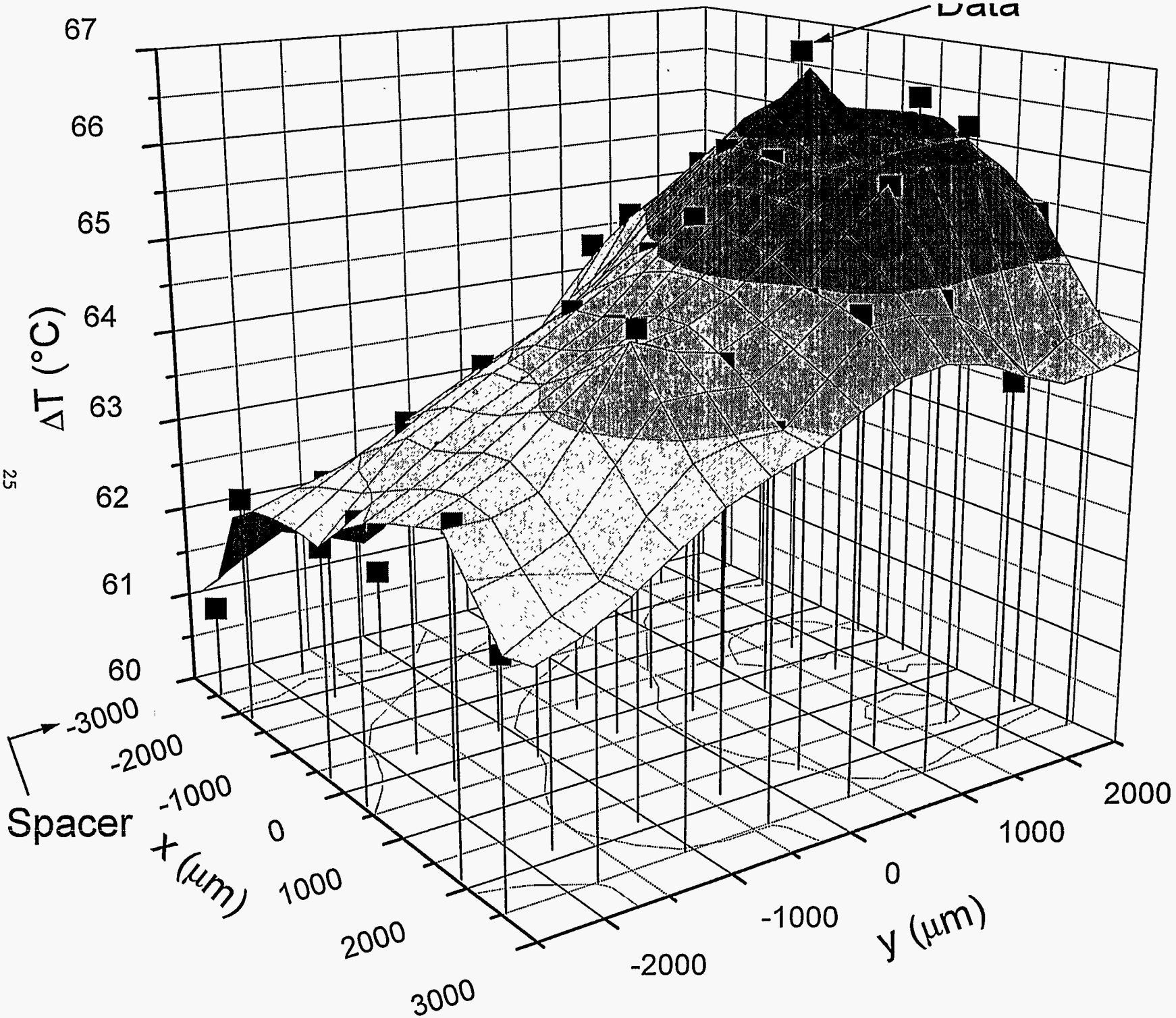

Data
67
66
65
64
63
62
61
60
ΔT (°C)
-3000
-2000
-1000
0
1000
2000
3000
x (μm)
2000
1000
0
-1000
-2000
y (μm)
Spacer

Figure 8
(A)

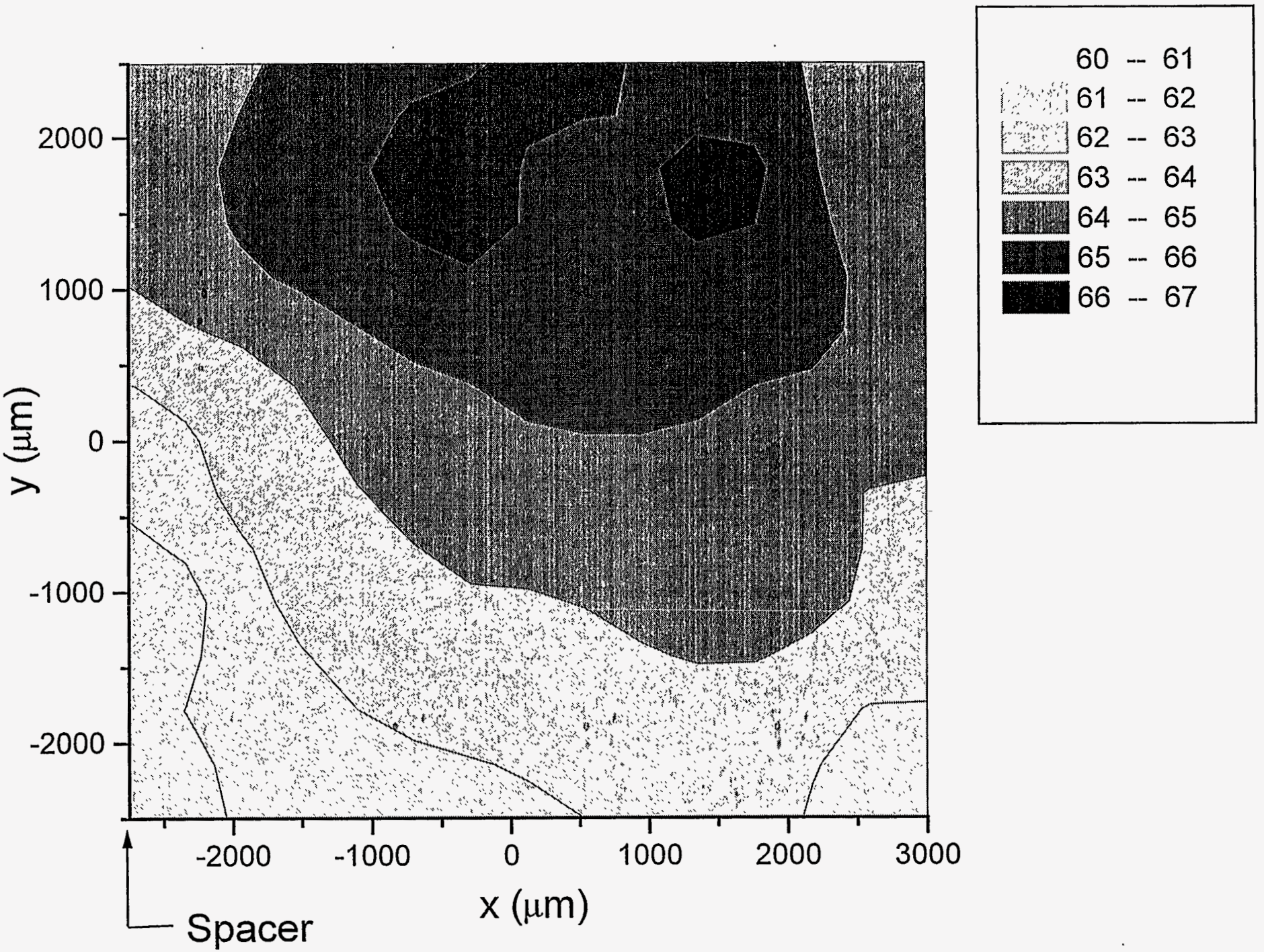

60 -- 61
61 -- 62
62 -- 63
63 -- 64
64 -- 65
65 -- 66
66 -- 67
2000
1000
0
-1000
-2000
y (μm)
-2000
-1000
0
1000
2000
3000
x (μm)
Spacer

Figure 8

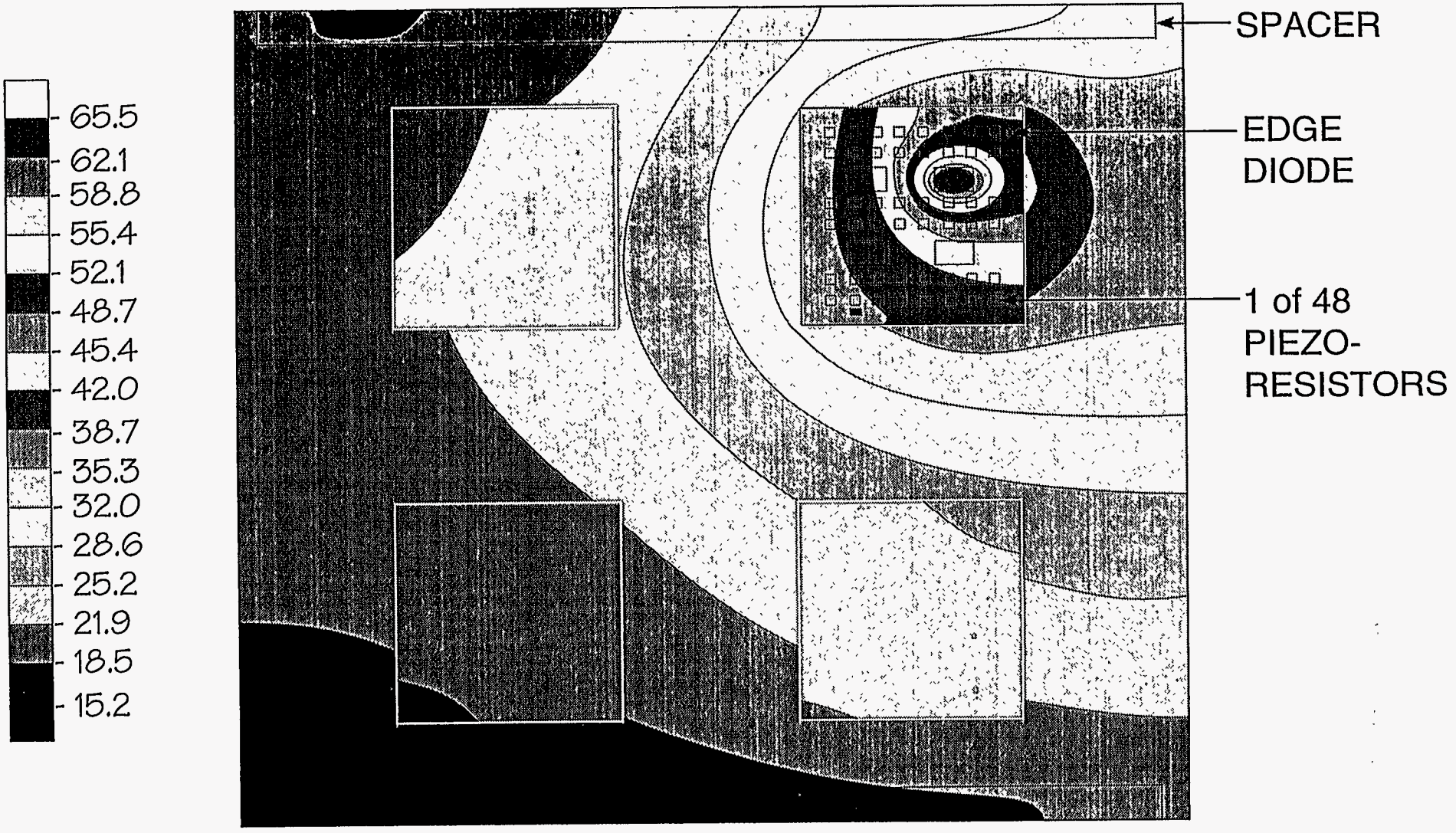

65.5
62.1
58.8
55.4
52.1
48.7
45.4
42.0
38.7
35.3
32.0
28.6
25.2
21.9
18.5
15.2
SPACER
EDGE DIODE
1 of 48 PIEZO-RESISTORS

Figure 9

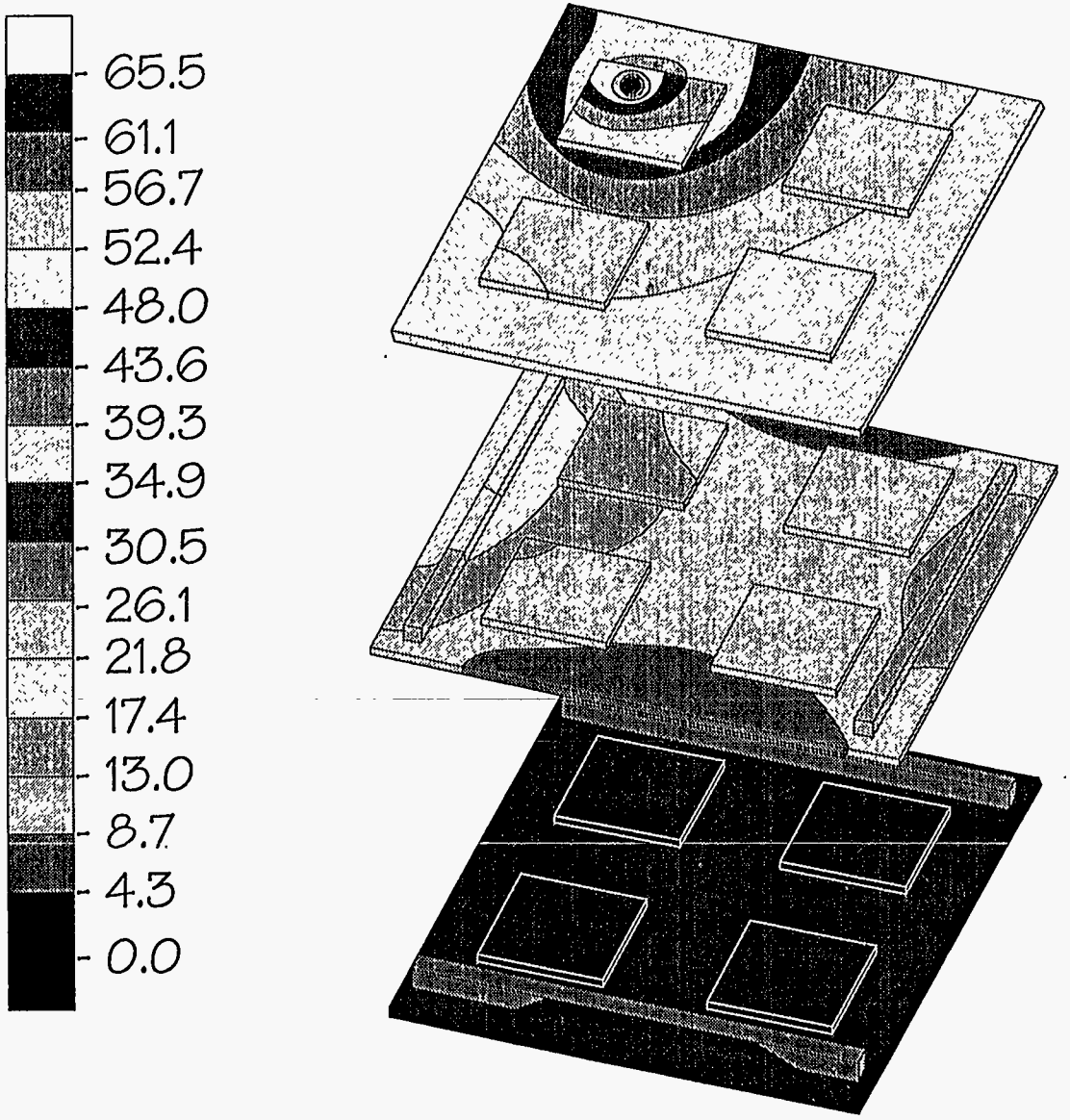


Figure 10

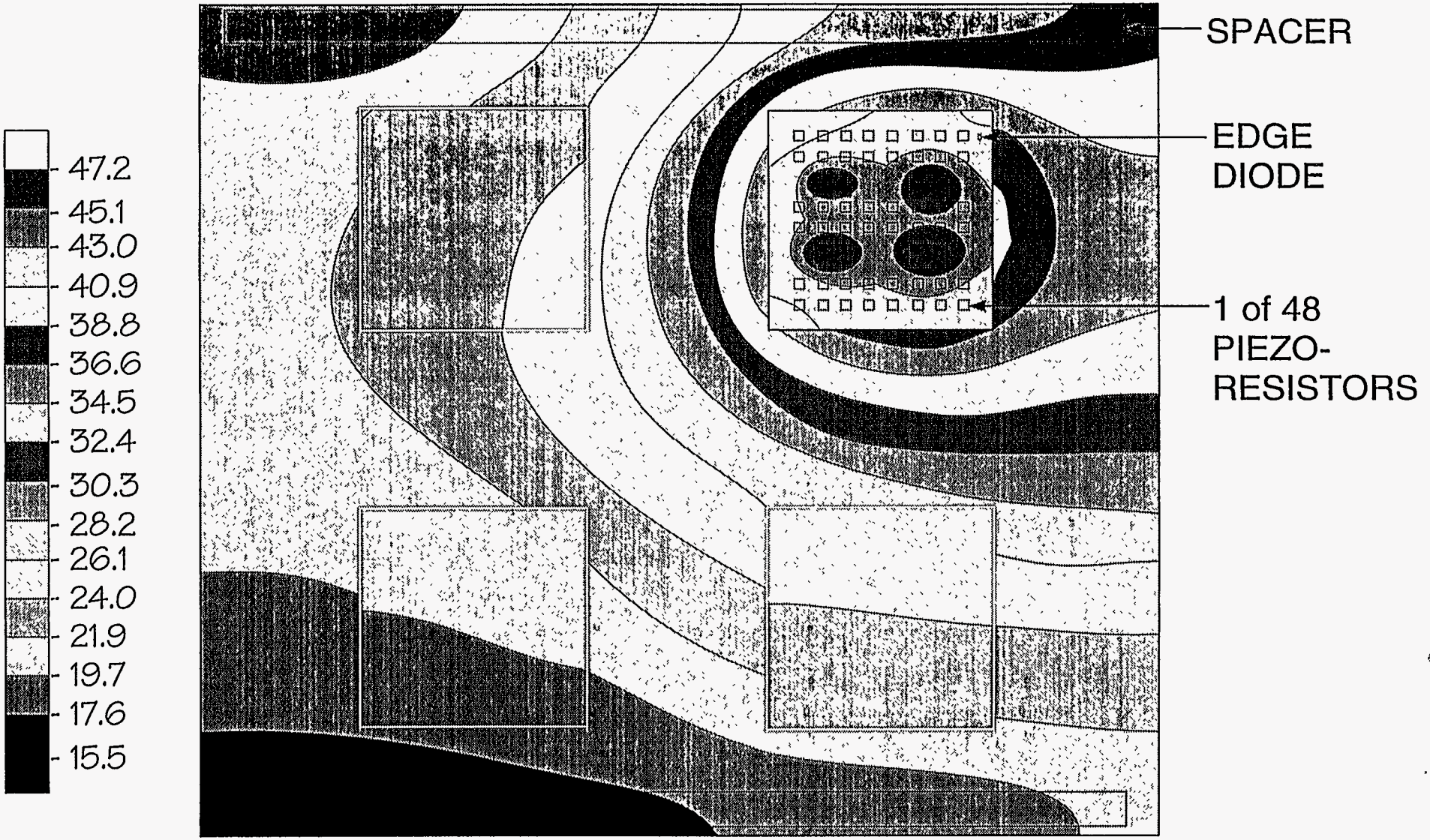

47.2
45.1
43.0
40.9
38.8
36.6
34.5
32.4
30.3
28.2
26.1
24.0
21.9
19.7
17.6
15.5
SPACER
EDGE DIODE
1 of 48 PIEZO-RESISTORS

Figure 11

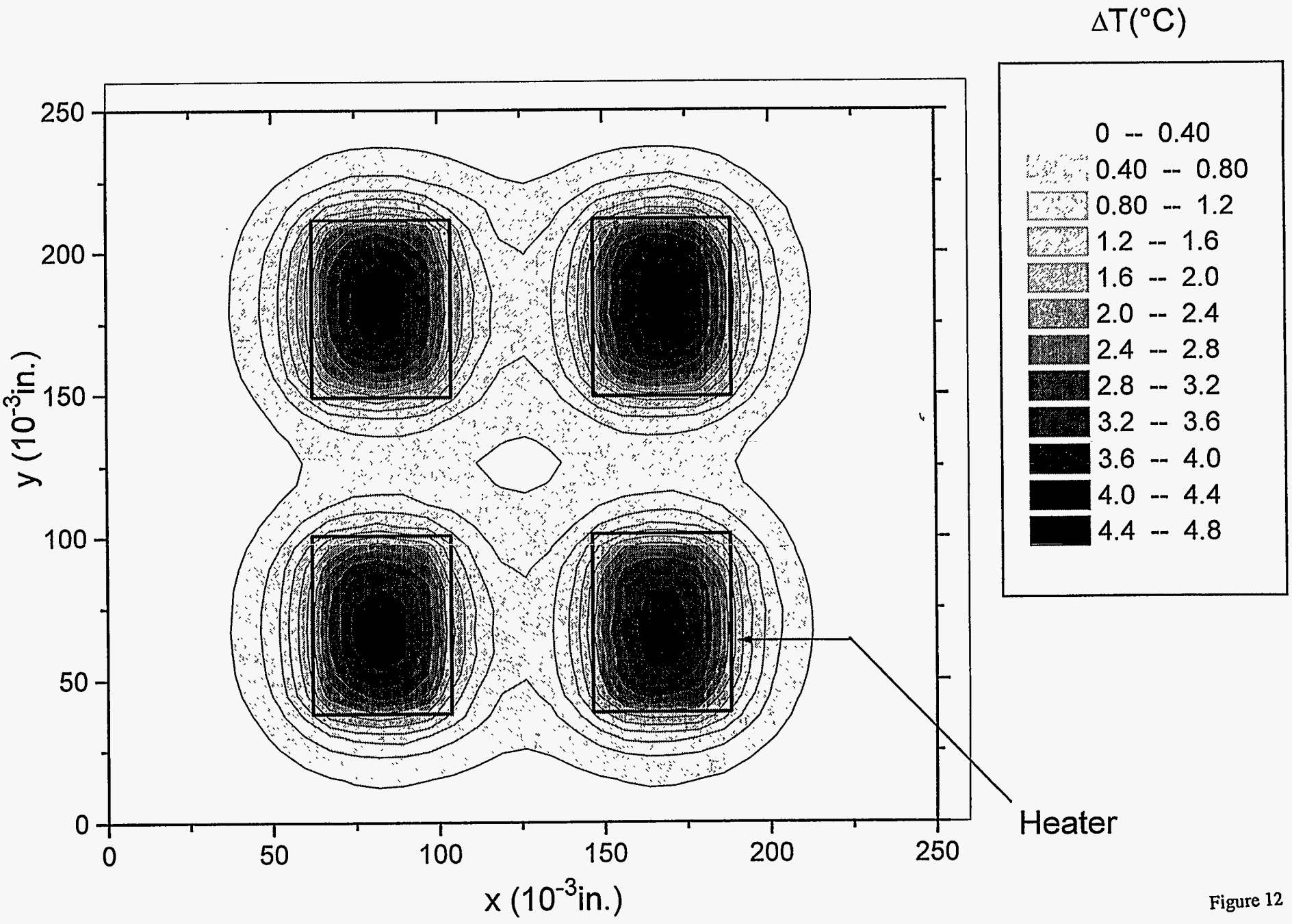

ΔT(°C)
0 -- 0.40
0.40 -- 0.80
0.80 -- 1.2
1.2 -- 1.6
1.6 -- 2.0
2.0 -- 2.4
2.4 -- 2.8
2.8 -- 3.2
3.2 -- 3.6
3.6 -- 4.0
4.0 -- 4.4
4.4 -- 4.8
Heater
y (10^-3 in.)
x (10^-3 in.)
0
50
100
150
200
250

Figure 12

CHAPTER 2: THERMAL ANALYSIS OF THE 3D IC MODULE

In many system cases the limit to density is set by the ability to extract waste heat from the system. A standard joke in the instrumentation business is "if you made a design mistake just buy a bigger fan for your equipment chassis to blow the extra heat away". With stacked integrated circuit chips or stacked multichip modules the thermal limit is easily reached. This limit is by definition the density above which some regions of the design exceeds the highest temperature for reliability in specification performance. For continuous operation this is often around 80°C, for 30 minute operation about 125°C.

Extending the limit can be accomplished in several ways:

1. Develop circuit design that dissipates less power. For example, run chips at lower voltage or have circuit segments that "go to sleep" if not being used.
2. Use high thermal conductivity substrates or blanks within the stack to conduct the heat to the stack edge and toward the system heat sinks.
3. Increase active cooling (air or fluid flow rate).
4. Reposition components to avoid hot regions in system.

To make the necessary trade-offs between thermal management and increased density, computer simulations of thermal flow are performed. In this way the design materials and design can be quickly changed in the computer model and the termal ramifications immediately determined. Many design changes can be tried on the computer much faster than one prototype can be built and tested. At least that is the simple vision. In fact, accurate 3D analysis of complex geometries takes considerable expertise and computer power.

This chapter contains the preliminary results of research in developing efficient automatic grid generation for complex packaging shapes. In the past the grid generation of a 3D problem often took months. This work was performed by Glenn Laguna and Anthony Thornton of Organization 1553, computational Aerodynamics. For comparison, previous specific thermal modeling of a complex 3D Multichip Module is presented. This simulation was performed by Bruce L. Bainbridge of Organization 1513, using the time-proven grid system of symmetry and grouping to minimize computation time, but still using more than most engineers can afford in early trade-off analyses.

Further improvements in codes and problem presentation were looming on the commercial horizon when the research study ended. Given the computational improvements shown possible in this work and the projected improvements from many others in this business, it appears that design trade-off from 3D analysis will become a common tool.

Grid Generation for Complex IC Packages

(excepts from technical memo to David W. Palmer on November 12 1992 from Glenn Laguna and Anthony Thornton, "Grid Generation for Heat Transfer in Semiconductors")

Until recently, computational aerodynamics has been done with structured grids. The grid points are arranged in rectangles or blocks, in an ordered ijk array. These grids have several advantages. The difference equations are fairly intuitive, and graphics are comparatively easy. Structured grid formulations also vectorize very well, so they can take maximum advantage of the current generation of vector supercomputers. It is possible to cluster the grid points close to a body in order to get good resolution in the boundary layer. Structured grids look very pretty, and appeal to our sense of order. The one major problem with structured grids is that generating the grid for a three-dimensional shape is fiendishly difficult. NASA has taken two months to generate a grid for an inlet-diffuser, and two years for the space shuttle. For complex shapes, developing a structured grid can be nearly impossible (see Figure 1).

Recently, unstructured grids have been getting a lot of attention in the computational aerodynamics community. In unstructured grid formulations, the volume is filled with triangles or with tetrahedra. The triangles can be any shape or size, and thus can fill an almost arbitrary volume. The advantage is that gridding is comparatively quick and easy. In fact, only a surface grid is really required since the volume grid generation is virtually automatic. The disadvantages are that many more tetrahedra are needed to capture the flow physics, and these formulations don't vectorize well, so the calculations can be costly in terms of computer time.

In the last 18 months we designed and implemented an unstructured grid capability for compressible aerodynamics computations. It has been extremely successful for calculating lift and drag when these are dominated by pressure as opposed to viscosity (for example, an airfoil at angle of attack would be pressure dominated, but would be viscous dominated at zero angle of attack). We used the LDRD to investigate this same approach as applied to heat transfer. The results are encouraging, but not yet as good as for aerodynamics.

One of the early I-DEAS grids we generated was for a 14 pin dual in-line package shown in Figure 3. Although this is a simple shape, it would be faily difficult to grid using structured techniques. The unstructured grid took only an afternoon. A RAMPANT solution for a generic automobile is shown in Figure 4. This took less than 24 hours from concept to solution.

Applications to Heat Transfer

Our primary concern in the area of heat transfer is the ability of the code to accurately capture the boundary layer, since this is where heat transfer takes place. It is critical because heat transfer depends on gradients of the flow quantities. This is still an active area of research in computational aerodymanics. Many people believe that an unstructured formulation is inherently unable to capture the boundary layer. Others believe there is no inherent difficulty, but that it is too expensive. It takes roughly six times the number of elements in the boundary layer as a structured code. We tested RAMPANT on the simplest possible boundary layer solution, flow over a flat plate. This is one of the few problems in fluid mechanics that can be solved analytically, resulting in

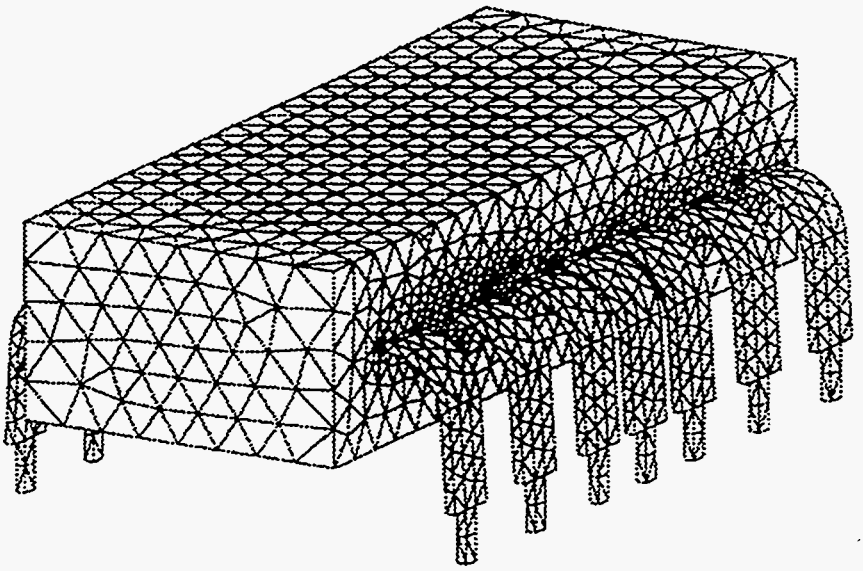


Figure 1 : Unstructured Surface Grid for Dual In-Line Package

the Blasius profile. The results of the calculation are shown in Figure 5. The variable in this plot are similarity variables, i.e. the curves plotted in this manner should fall on top of each other. The RAMPANT solutions are obviously not similar, and do not match the Blasius solution. (We reacently learned that Rick Matus at Fluent Inc. has done the same flat plate problem with the incompressible version of RAMPANT and got a nearly exact match to the Blasius solution. It could be that compressibility, even at low Mach number, accounts for the discrepency.)

Finite Element Analysis (FEM) of 3D IC Module

(excepts from July 30, 1991 Memo from B. L. Bainbridge to R. A. Gassman, "Thermal Analysis of the 3D Module)

A detailed numerical model of the 3-dimensional, integrated circuit (IC) module has been generated to examine its internal temperatures. The model incorporates the salient features of the design and should provide fairly accurate predictions of the temperatures.

Numerical Model

The model of the 3-D IC module (see Figure 2) was generated using PDA's PATRAN code. Information provided in reference 2 was used to define geometry and material properties. Basically, the module is divided into 4 slices, each one representing either a processor slice or a memory slice. Each slice is 3.17cm (1.25" by .95") and 0.147 cm (0.058") thick with a stacking order of (from the bottom up) processor-memory-processor-memory. The idividual slices are further divided into 4 layers; a via layer, a layer of solid silicon, a layer containing the ICs, and layer of Kapton. The various regions on each layer are defined by the material properties (see Figure 3).

The via layer (vias are copper-filled through-holes used for electrical connections between layers) is 0.0254 mm (0.001") thick and is used to represent the compliant contacts that feed signals from one slice to the next. The layer is mostly air except for the region that contains the vias. The via region is modelled using elements composed of 95% air and 5% copper by volume. The percentage was obtained by assuming a 0.102mm (0.004") diameter for each of the 198 vias.

The silicon layer forms the substrate for the dies. It is 0.609mm (0.024") thick and is composed almost entirely of elemental silicon. The exception is the via region which is composed of 95% silicon and 5% copper.

The next layer contains the ICs. As indicated in Figure 3, the 0.660mm (0.026") thick layer is silicon except for the via region and air-filled gaps that isolate each die from the rest of the layer. The via region has the same composition as the previous layer. The placement of the air gaps depends on whether the layer is part of a processor slice or a memory slice. In fact, the air gaps, in conjunction with the heat flux boundary condition, *define* the slice.

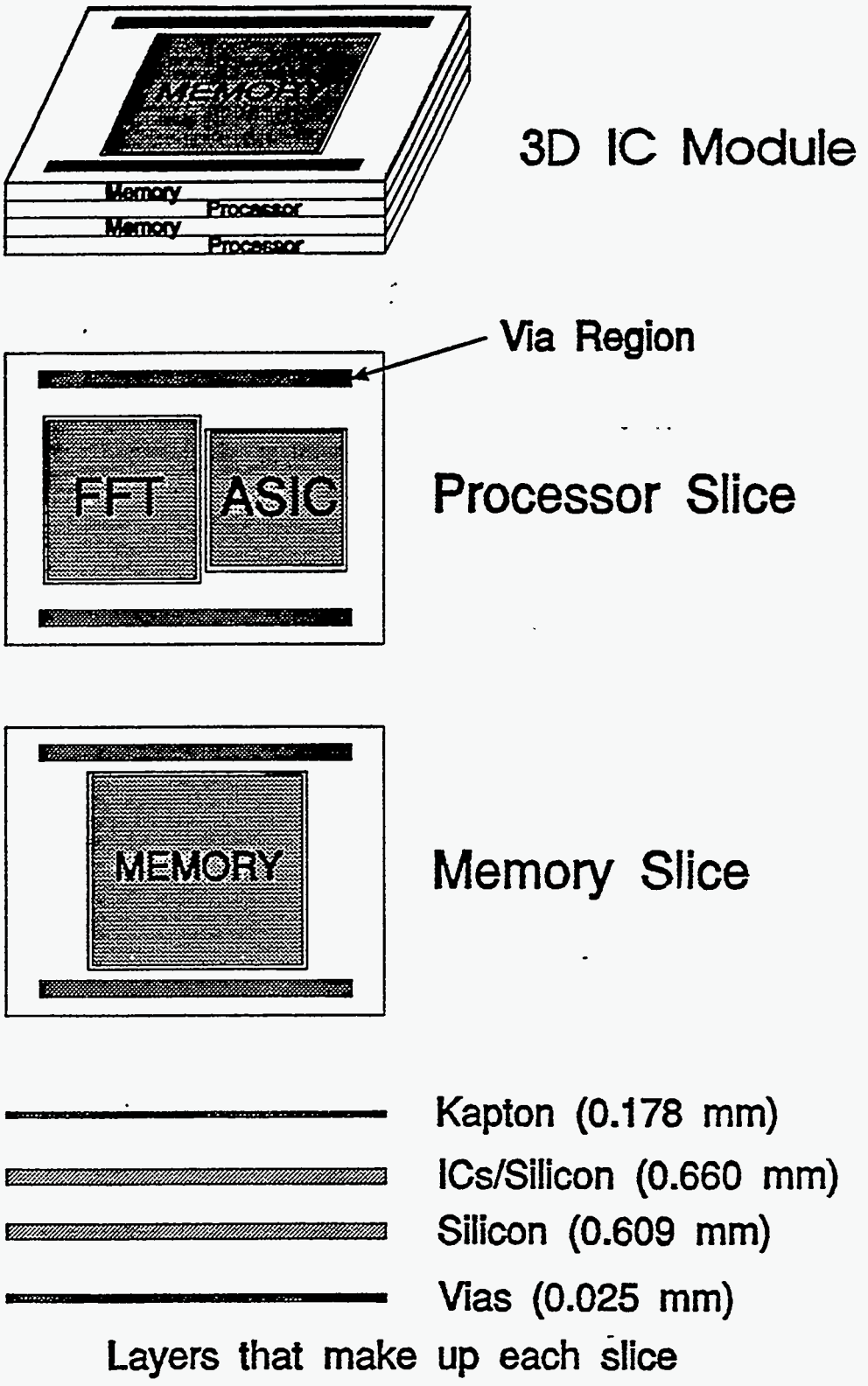
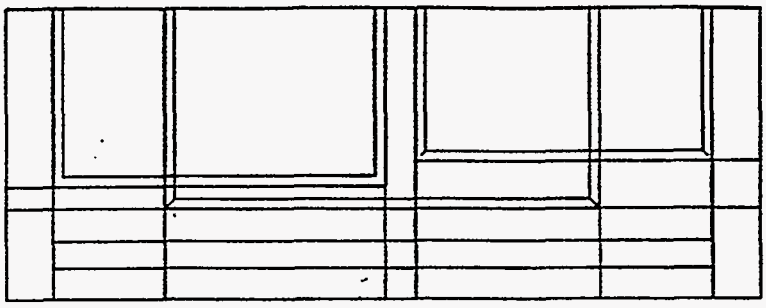
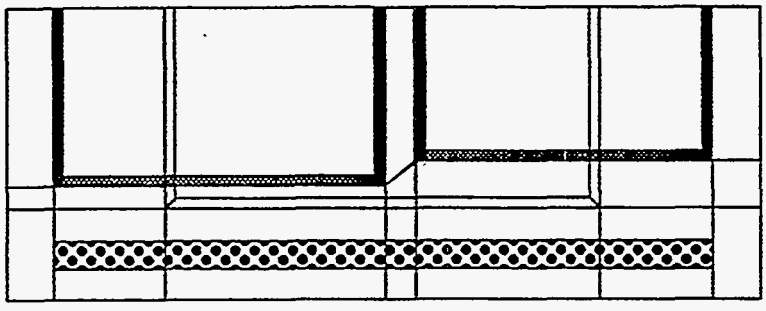


Figure 2: Schematic of 3D IC module

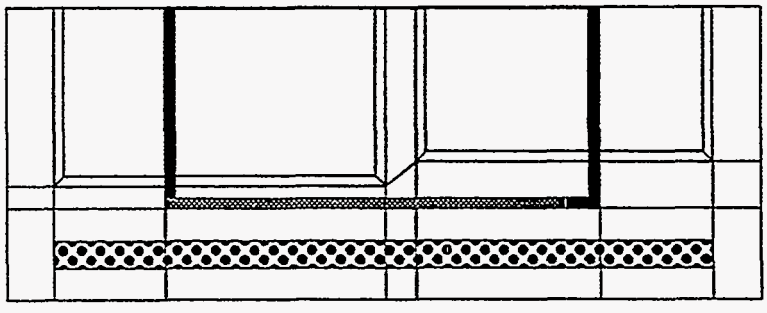
The processor and memory slices differ in the number of ICs and their location. Each slice is modelled using 4 layers: a layer of Kapton, and IC/silicon layer, a layer of silicon, and a via layer.



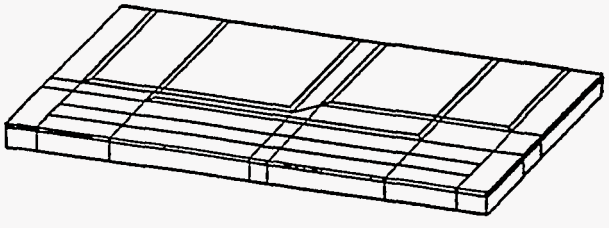
Basic Layer Showing Hyperpatch Outlines



Processor Slice



Memory Slice



Isometric

Silicon

Air

Silicon+Copper

Figure 3: Hyperpatch entities on MCM layers

The model uses hyperpatch entities to outline 3D regions of similar properties. The hyperpatches are then meshed with elements. The material properties used for the elements in each hyperpatch depends on which layer they are in.

The last layer represents the Kapton Type-H film that support the electrical connections between the dies and the vias. The via region has dissimilar properties (95% Kapton, 5% copper) from the rest of the layer. In reality the Kapton layer containes a large number of thin copper traces. As their number can vary from one case to the next, it was decided to examine the most conservative case and include none at all. This results in the lowest thermal conductivity for the Kapton layer.

The numerical model takes advantage of the module's symmetry by splitting the module lengthwise (defined by the side that is 3/17 cm long) and discretizing only one half. The resulting model is smaller and computationally less expensive. Figure 4 show a solid-filled representation of one slice of the model.

Power dissipation was defined[2] as 7W, 2W, and 1W for the FFT, ASIC, and memory ICs, respectively. Thus, the processor slices dissipated 9 W and the memory slices 1 W for a total of 20 W for the entire module.

The material properties used for the analysis are listed in the appendix. Values for copper, air, elemental silicon and Kapton Type-H film were obtained from the department 1510 material property data base. The copper vias are modelled by modifying the thermal properties of the via region. Assuming the copper through-holes were 0.102mm in diameter (0.004"), approximately 5% of the via region was copper and the remainder either air, Kapton, or silicon. The material properties were altered to reflect the presence of the copper based on that percentage.

The initial temperature of the model was 20°C (68°F). A heat flux was applied to the upper surface of each chip region according to the previously described power dissipation values. Boundary conditions were assumed to be adiabatic except for the bottom surface of the module. At that surface, a 17.1mm (0.675") wide area running the length of the module between the via regions was in contact with a copper heat sink. Two cases were examined: the first included an air gap of 0.0254mm (0.001") between the module surface and the heat sink (modelled as a constant temperature boundary fixed at 20°C), while the second case did not. The boundary condition applied over the via regions also differed for the 2 cases. The first case assumed that the surface in those areas was adiabatic while the second case assumed it was also in contact with a heat sink.

The numerical model was run on the Cray YMP using PDA's P/Thermal code. The 15,000 node model required approximately 45 minutes of cpu time to simulate 2 minutes of real time.

(editors' note: traditional computer simulation of thermal flow in a typically complex electronic package takes lots of cpu time on powerful computers.)

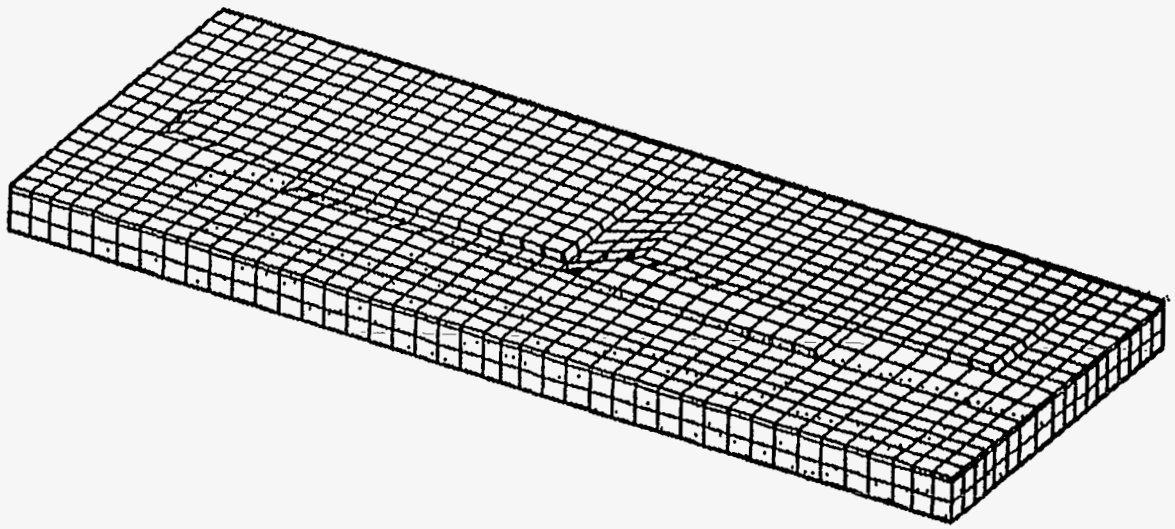


Figure 4: Mesh of elements

One slice of the numerical model that shows a solid-filled representation of the element mesh. Details of the various regions can be seen in figure 3.

CHAPTER 3 - THINNING OF IC CHIPS

One method to increase density in integrated circuits (IC) is to stack die to create a 3-D multichip module (MCM). special post wafer processing can be done to bring interconnects out to the edge of the die. The die were sawed, glued, and stacked. Special processing was done to create interconnects on the edge to provide for interconnects to each of the die.. Die thinning techniques were developed to allow for die thickness as thin as 50 μm. This would allow for high 3-D density when the die are stacked. To increase the density for 3-D MCM in the vertical direction, thinning the die or MCM substrate down to as small as one quarter of their original thickness needed to be demonstrated.

Two methods for mechanical thinning were initially investigated. They are the "free floating" and "wax mount" techniques. The active side of the die were protected using a coating of polyimde that was cured for 30 minutes at 90 ° C or with photoresist. The lapping equipment used was made by Strasburg.

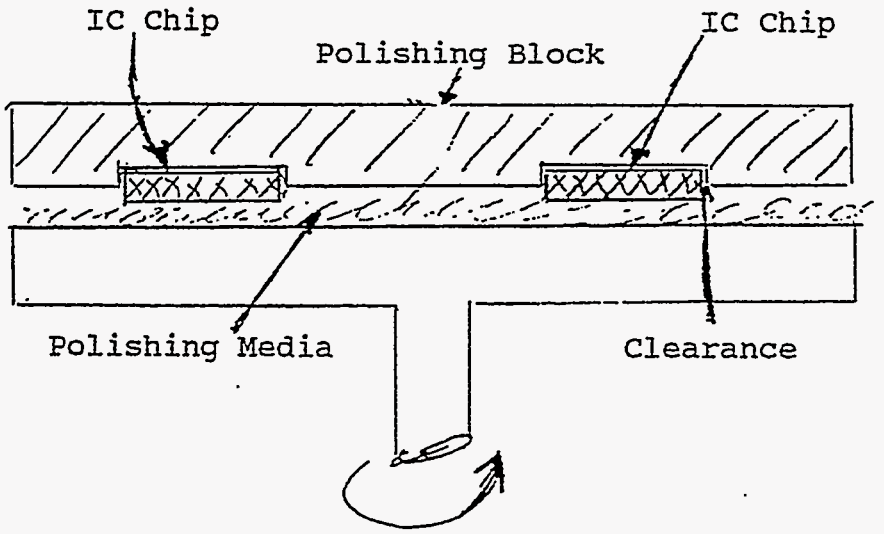


Figure 1. Schematic Representation of the "Free Float" Thinning Process

Free Floating: The original attempt at die thinning was the free floating technique. In this technique the chips are confined to a recessed area in the polishing block, and are allowed to move during the lapping process, see Figure 1. A set of die were lapped down to 0.006". Handling of the thinned die prove to be very difficult due to breakage. A problem was identified

on the thinned die. The problem was created by a design and/or processing deflects, which allowed the die some motion during the lapping process, and the clearance around the die edge in the recess of the polishing block. the motion of the die during the lapping and the clearance around the die edge allowed polishing media to work its way between the top of the chip and the polyimde protective coating, the bond pads and circuitry near the edge of the chip. Detail optical inspection confirmed this, see Figure 2.

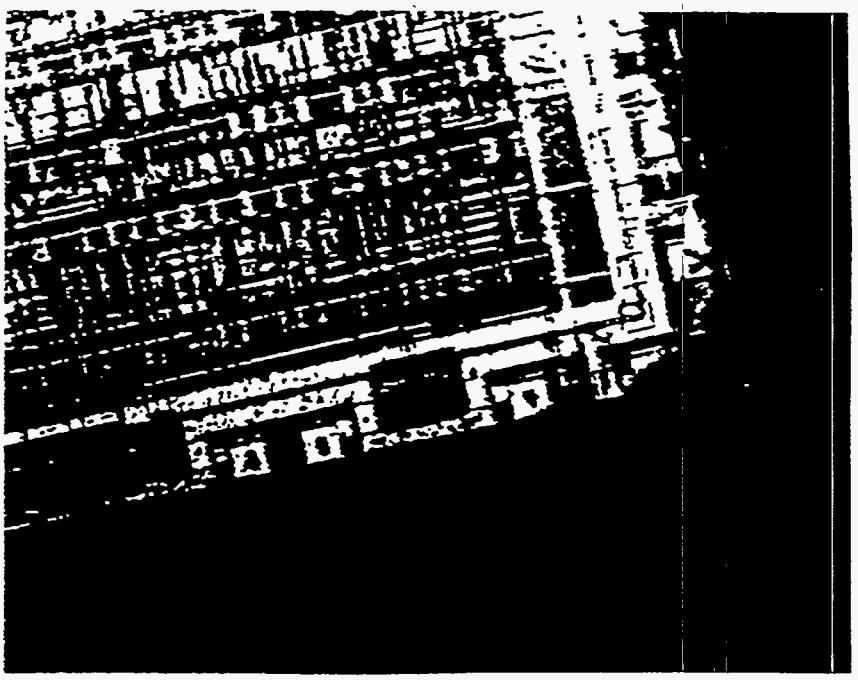


Figure 2. Effects of the Free Float Lapping Process

Wax Mount: This technique uses an organic polymeric materiel to bond the die to a flat polishing block. extreme care must be taken to ensure the die are mounted flat, otherwise planarity to the front surface can not be achieved. Prior to lapping the active side of the die were protected using a coating of polyimde that was cured for 30 minutes at 90 ° C. After lapping, the residual mount wax was removed with acetone. The polymide was stripped off. Inspection of the thinned die, top and backside indicated all die were in good condition with no damage or cracks. All bond pads and circuitry near the edge of the die was in good condition. Figure 3 shows the photomicrograph of a 0.006" thinned die prior to clean up. The photomicrograph, Figure 4 shows the die after wax and polymide removal. There was little or no edge damage for die thinned to 0.006" +/-0.0005".

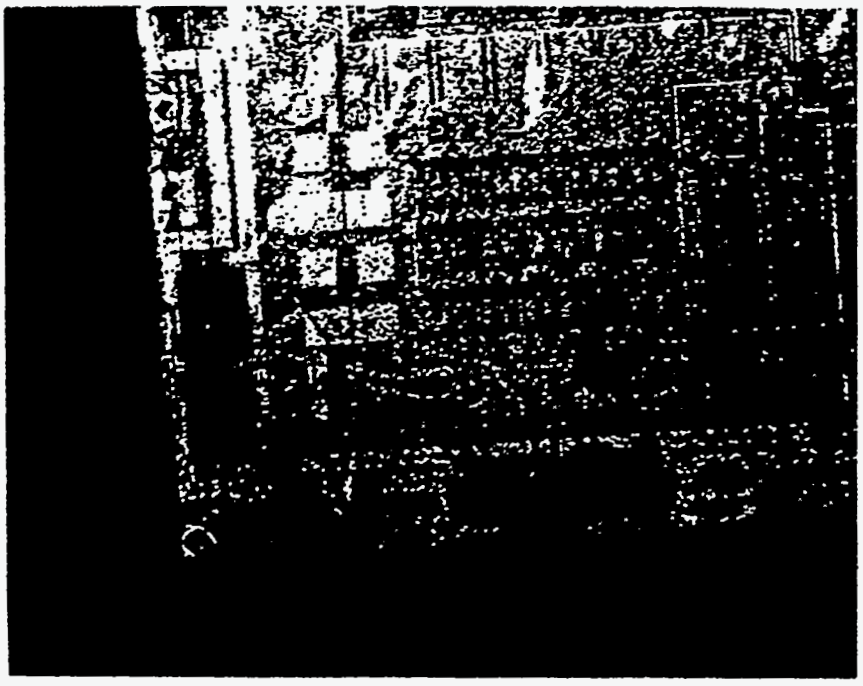


Figure 3. An 0.006" Thick Die after the Wax Mount Lapping, No Cleanup

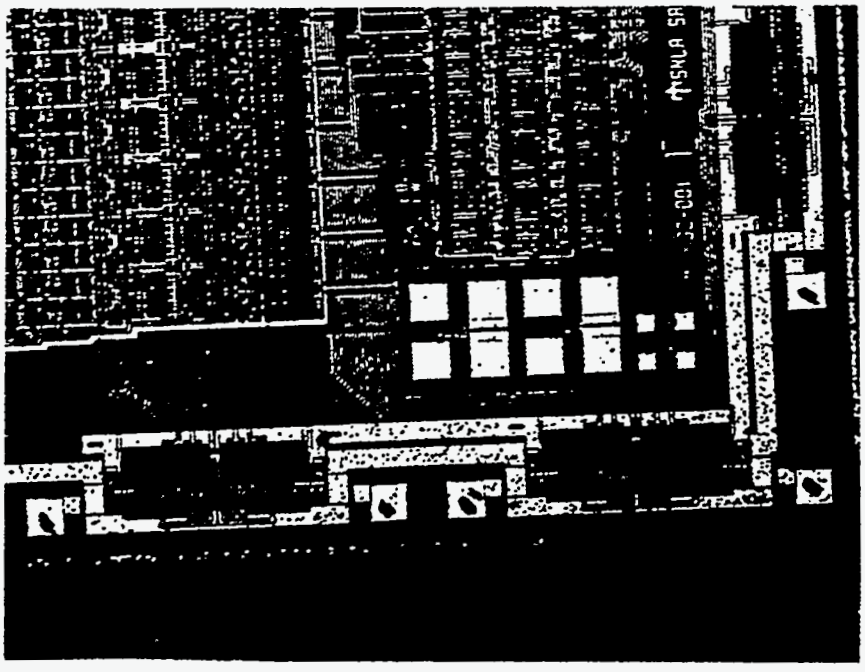


Figure 4. An 0.006" Thick Die Thinning by the Wax Mount Process after Complete Cleanup

To thin die down to less than 0.006" another piece of thinning equipment and technique was used. The equipment used for this type of die thinning is a Logitech PM4A precision lapping and polishing machine. Two other pieces of equipment used in conjunction with the Logitech PM4A are the Logitech VPB1 vacuum pressure bonding jig and the Logitech PP5GT precision polishing jig mounted with the Logitech PSM1 programmable sample monitor. Figure 5 shows the thinning equipment setup. The die are mounted onto 83 mm glass discs using quartz wax. The Logitech VPB1 system is then used to achieve a uniform bond between the sample and the glass disc. This glass disc is held on to the Logitech PP5GT jig by vacuum. The Logitech PSM1 which is mounted to the jig consists of two components: the sample monitor itself and the digital linear gauge. The Logitech PSM1 allows one to program in the amount of material to be removed and automatically stops the procedure when completed. The thickness of the die is measured by an Ono Sokki digital linear gauge. After thinning has been completed, the die are carefully removed from the glass disc by dissolving the wax in hot limonene and cleaning in isopropyl alcohol.

Using the Logitech PM4A, the die or MCM substrate as large as 4" in diameter could be thinned down to 50 μm. Figure 6 shows the cross section of die at regular thickness of 625 μm, 150 μm , and 50 μm. After thinning has been completed, the die are carefully removed from the glass disc by dissolving the wax in hot limonene and cleaning in isopropyl alcohol. By conducting regular assembly packaging experiments on the thinned die, it was determined that the minimal thickness of a die should be no less than 150 μm. Die thinned to less than 150 μm have yield loss due to assembly handling of more than 25%.

Figure 5. Logitech PM4A Thinning Equipment

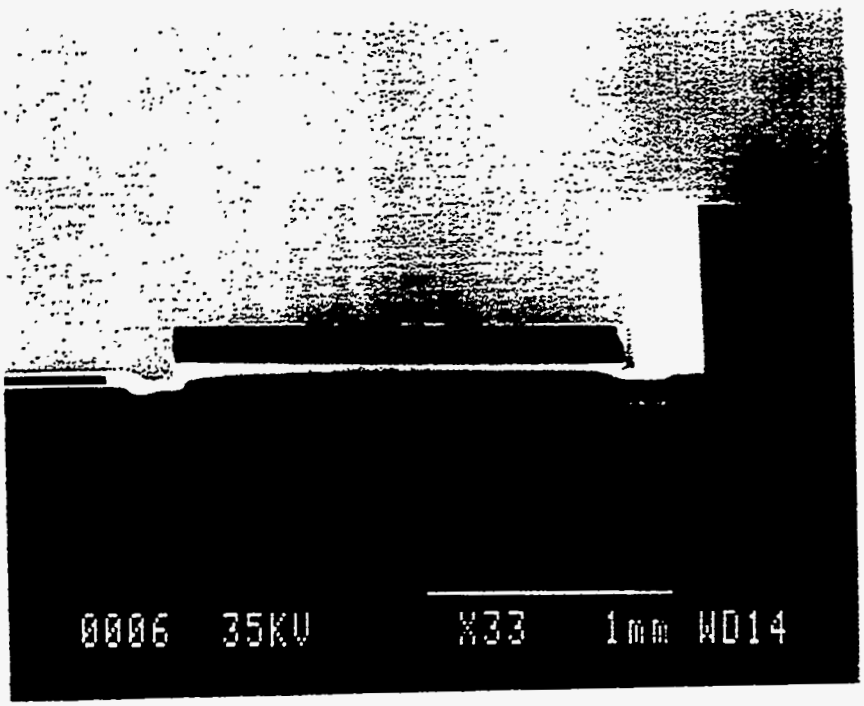


Figure 6. Cross Section of Thinned Die

Back side isolation: After the die are thinned, the backs must be electrically isolated. This is accomplished by "painting" on an insulative material. Initially both spin on glass (SOG) and polyimide were tried. It was decided that polyimide was more suitable both for its isolation properties (tested by using probes/digital multimeter) and ease of handling. The procedure includes: placing the die face down on blue tape, applying the polyimide to the backs with a small paintbrush, air drying for one hour while still on the tape, removing from tape and curing in nitrogen for 1 hour at 175° C and 1 hour at 300° C. The polyimide used was Micro-Si SPI-115.

CHAPTER 4 - SHALLOW WELL IN SILICON

The following discuss the progress made on etching of shallow wells in silicon substrates. It is often advantageous to use recessed rectangles for the placement of die rather than placing the chips on a flat surface. Wells and grooves were also used for spacing and alignment purposes.

Initial Experiments

A series of six wafers were micromachined. The wafers are of three types, the first set of two consists of oxidized 50 mil thick silicon wafers with wells; there is one 50 mil wafer which has not been oxidized; and three wafers consist of two wafers bonded together and are not oxidized, except for the etch stop oxide at the bottom of the wells. The etched wells were targeted to a depth of 24 mil in the 50 mil wafer case, and to the point where the etch "stopped" on the bonding oxide in the bonded wafer case.

Procedure for Well Etching

1. 5:1 piranha clean, 5 min., 95° C.
2. Standard megasonics clean, OP #1210.
3. Deposition of 1175A thermal silicon nitride, OP #2700.
4. Photopatterning with the first "3-D packaging" reticle set.
5. Etch SiN in the ONO using the "micromachining" recipe.
6. Strip photoresist with PRS 1000.
7. KOH etch the wells, 4 M solution 85° C stopping at 24 mil depth. Final well depth 24 mil +/-0.25 mil.
8. Strip SiN in HF.
9. KOH cleanup in 6:1:1 (H_2O:H_2O_2:HCL) 15 min. One wafer finish at this point.
10. Oxidize the surfaces of two of the 50 mil wafers, 950° C, 1 hour, steam, in VTR-10.

Procedure to Bond Wafers

1. Piranha clean of the 25 mil single side wafers, 5 min., 95° C, OP #1000.
2. Deposit 5500A BPSG (5%, 5%) (BPTEOS replacement process) on 3 single side polished wafers.
3. Reflow BPSG OP #7021.
4. Piranha clean three, 24 mil, n+, double polished wafers together with the three BPSG wafers listed above. OP #1000.
5. Join each of the double polished wafers to one of the BPSG coated wafers. The flats ere aligned at 45° to each other (+/-10°).

6. Deposition of 1175A thermal silicon nitride, OP #2700. This process both deposits the SiN and forms the bond between the wafers.
7. Photopatterning with the first "3-D packaging" reticle set.
8. Etch SiN in the ONO using the "micromachining" recipe
9. Strip photoresist with PRS 1000.
10. KOH etch the wells, 4 M solution 85° C. the etch was stopped after the well depth had uniformly reached the level of the bonding oxide.
11. KOH cleanup in 100:1 HF 15 sec. followed by 6:1:1 (H_2O:H_2O_2:HCL) 15 min.

Observation Made on the Processes

The unpolished backs of the 5 mil wafers appear to be more prone to failure of the thermal nitride film that the backs of the standard 25 mil wafers. These failures result in the formation of pits on the backs of the wafers. the backs of the bonded wafers had between three and four pits ~300 microns on a side which were probably related to a machine induced defect. The 50 mil wafers had tens of failures larger than 0.5 mm on a side and hundreds of smaller pits. Since the nitride coating was identical in each case, this suggests that the 50 mil wafer has a poorer than usual surface quality. This adversely effects the robustness of the protective nitride film on the wafer backs. In the case of the 50 mil backs, the pits sometimes seemed to have an order to them which can not be matched to the tools in the fab. There is evidence that one of the machines used to handle the wafers, probably during photo, has damaged the nitride backs. This results in a distinctive pattern of three pits and probably causes in the failures observed on the backs of the bonded wafers.

There are several possible approaches to the problem outlined above. The first is to send the 50 mil wafers out to be double polished. The second is to protect the front surface of the front polished wafers and use a KOH etch to smooth the back surface. This has been done earlier on the unpolished 50 mil wafers and resulted in a reduction in the pitting resistance of the material. However, isolated pits were still observed. The third approach would be to use bonded wafers. It should be a simple task to isolate the machine damaging the backside nitride since the three point pattern is quite distinctive. Once identified, it may be possible to reduce or eliminate this particular problem.

The nitride etch process appears to be giving rise to a ring defect on the backs of the wafers. The ring is ~1/4 inch in from the wafer edge and results in nitride removal or damage in this area. subsequent KOH exposure then attacks the silicon degrading the edge quality. This problem will probably not effect yield by may effect subsequent wafer handling.

There were several reasons for investigating the use of bonded wafers for a package. The first is that work on other projects has lead to significant advances in the bonding process. Secondly, bonding the two wafers with their flats rotated should result in a more robust package since the wafer cleavage planes are no longer aligned. Thirdly, if the die thickness is known in advance,

or if the dice can be thinned to the same thickness, then bonding will result in the formation of reproducible wells with very smooth bottoms. The process is also more controllable in that the etch will slow greatly on encountering the bonding oxide, significantly increasing process control. The thickness of the bonding oxide is only 5500A, and well depth can be readily controlled to within this value. The major source of error would probably exist in thickness of the top, double polished wafer and the dice thickness.

In one of the bonded pairs there appears to have been incomplete bonding. This is probably due to particles at the center of the wafer. This problem can be addressed by going to a higher bonding temperature than the 800° C used in this set of runs. 800° C is the nitride deposition temperature and was used to simultaneously deposit the nitride and bond the wafers. Use of a higher bonding temperature (1000° C) would result in BPSG reflow which would eliminate the effects of smaller particles.

Stress cracking was observed on one of the oxidized wafers. the crack occurred as the system was cooling and initiated at a v-groove to be used as a sawing guide. A similar problem has been observed in another project. The solution was to ramp the temperature down more gradually during the 2 micron oxide deposition and eliminate the v-groove.

Production of Silicon Substrates with a Differing Well Depths

The use of epoxy masking techniques to obtain wells of differing depths and a overlap process to produce differing depths within the same well, without the need for a ridge between the two regions of differing depth have be investigated.

A successful masking material must have the following properties:

1. Be able to withstand the KOH etch and have no effect on the etch rate.
2. Be able to be removed after the etch.
3. Adhere well to the silicon.
4. Maintain dimensional stability during cure.

It is desirable, but not necessary, that the material have the following properties of being transparent and have a cure that is below 300° C.

During experiments investigating the formation of wells of four differing depths on the same substrate, it was found that the etching of the silicon was inhibited at high surface coverage of "2-ton" epoxy. This was not evident at the lower surface coverage. It is interesting to note that the epoxy actually seems to deactivate the solution and not just form a protective film over the areas being etched. Increasing the cure time and temperature reduced, but did not eliminate the problem.

As a solution to the problem outlined above, Ecobond 104, cured at 100° C overnight, was tried on samples with high coverage of silicon. This material resisted attack and had no effect on the etch rate of the silicon. Adhesion of the material is excellent. When this material is cured at

200° C, it shrunk slightly and pulled chunks of silicon out of the substrate as it contracted. The problem with this material lies in it removal after masking. The ideal removal technique is the use of a piranha solution (5 parts H_2SO_4, one part H_2O_2 at ~95° C). This solution is commonly used in the fab to remove organic contaminants. Ecobond 104 is very resistant to piranha. Multiple exposures will remove most of the material but small stringers are left. Ecobond 104 can be stripped using commercially available paint strippers. However, there are ES&H contamination issues with the use of these strippers. There is a bench in the fab where chlorinated solvents may be used, but no KOH contamination can be tolerated in that bench.

Some experiments were attempted in which the Ecobond 104 was used to cover most of the masking layer of the "2-ton" epoxy. This approach seemed to work, but rejected as being too cumbersome.

It was found that a third epoxy "Epoxy Patch 1C", cured at 50° C overnight, was able to withstand the etch and did not effect the etch rate. The workable lifetime of the material is 30 minutes at room temperature. Shrinkage on curing does not appear to be excessive. It was also found that small amounts of the uncured material could be removed using acetone. This is important since it means that small errors may be reworked. Room temperature cure of the epoxy was not found to be sufficient.

The use of epoxy masking will allow one to produce wells of differing depth on the same die, as long as each of the regions of differing depths separated by a ridge. The tolerances of the bottoms of the well can be expected to be within +/-0.5 mil of the target, with +/-0.25 mil being typical. Starting with patterned nitrided masked wafers, the fabrication process have been defined as follows:

1. Etch to the shallowest well depth, standard 4M, 85° C KOH solution. If the wafer type is unknown, or if the etch rate is suspect, stop the etch halfway through and determine the actual rate and readjust the etch time accordingly. Etch rates are typically 100 microns per minute.
2. Once the final depth has been obtained, thoroughly rinse the wafer. Spin dry in the contaminated spin drier.
3. Mask the wells at depth one by squeezing out equal lengths of epoxy parts A and B; completely mixing the two parts; apply the epoxy to the wells at depth one; minor errors can be corrected by removing the uncured epoxy with acetone; cure at 50° C overnight.
4. Etch to well depth two.
5. Continue from Task 1 until all the different well depths have been defined. After each depth visually inspect the epoxy masks. If there is any sign of degradation or undercutting of the masks, strip the masks using a piranha etch and reapply epoxy.
6. Once all the well depths have been defined, strip the epoxy in 5:1 piranha solution. Note, the initial reaction of the epoxy to the piranha etch will be quite violent. Make sure that the table top is protected from the heating of the solution.

While at first glance this technique may not appear applicable to industry, we believed that it would probably lend itself to screen printing techniques. A similar, alternate approach would be SiN masking. Epoxy masking is advantageous compared to this approach in that it saves the KOH cleanup, nitride deposition and photo steps. Application of the epoxy is a relatively short process and curing can be done overnight. The technique is more accurate than the overlap approach described next in the that the accuracy (in depth) of each etch step is independent of the etch steps which have gone before. In the overlap approach errors in well depth add, thereby limiting the number of times the technique can be performed on a single wafer.

Using epoxy masking, it is not possible to produce a single well having two different depths. The overlap approach consist of fabrication sequences shown in Figure 1. This illustrates the process flow for the formation of two well depths within a single well. In this approach, the deepest well is defined first and etched t a depth equal to the difference in thickness between it and the deepest well. The system undergoes a KOH cleanup and the second depth region is defined photolithographically and the nitride etched. KOH etching is the continued until the second well depth is defined. This step also finishes the etching of the deepest wells.

Similar approaches were considered earlier and rejected for a number of reasons:

1. The process flow is longer and involves more possible routes for KOH contamination.
2. Errors in depth determination are cumulative.
3. It requires photolithography over etched wells.

However, if the technology is required the (1) can be discounted. The current well depths can be controlled to within ~0.5 mil, within 0.25 mil being typical. thus, at least two different levels would be possible while keeping within a 1 mil tolerance. Also, it appears that photolithography will be required for metal patterning so that a photo process compatible with well will needs to be developed. As the result of these consideration, an experiment to investigate this approach was undertaken. The experimental process flow is given below:

1. Standard monitor grade 25 mil wafer were coated with 1175A thermal SiN, OP #2700.
2. The nitride was photopatterned by adjusting the shutters on the Nikon steppers. This approach is less accurate than that using a mask and does not give as good a corner definition. However, it is fast and does not require a mask. The openings in the photoresist were ~ 1 cm by 1 cm.
3. The nitride was etched using standard processes, "ONO micromachining recipe".
4. 4M. 85° C, KOH was used to etch wells 100 micron deep, actual depth across the two wafers was 95-100 microns.
5. KOH contamination was removed using 100:1 HF and $HCl/H_2O_2/H_2O$.
6. Standard processes were used to spin on the photoresist. Significant problems were encountered on the outer wells.

Cross Sectional Views of "Overlap Process"

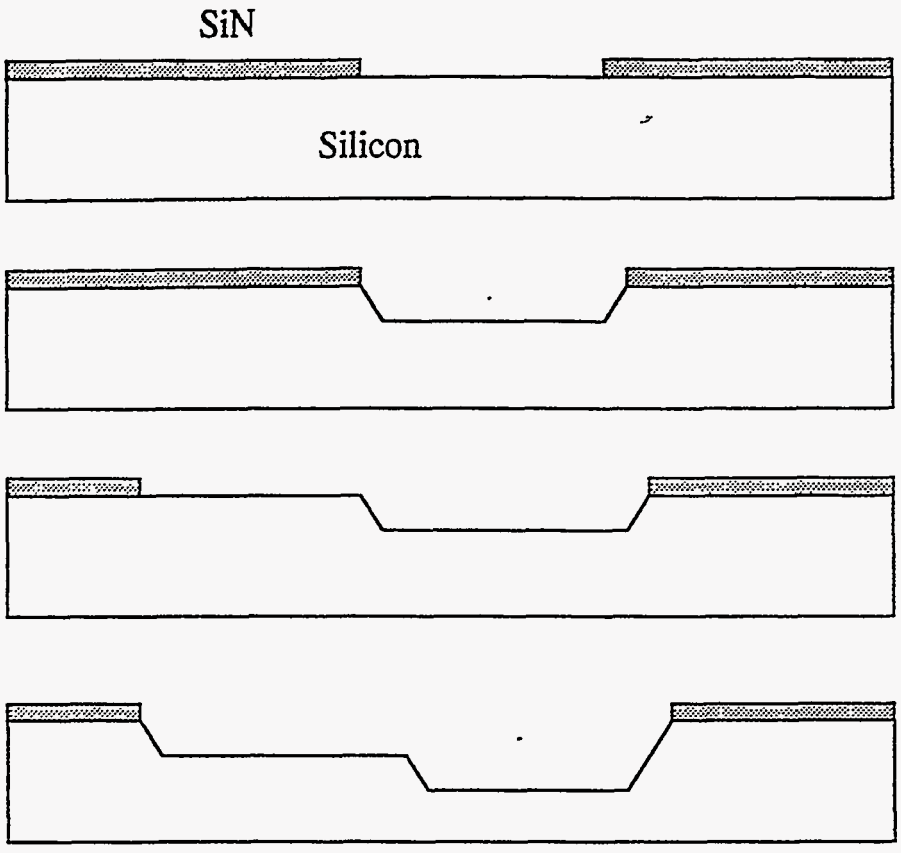


Figure 1. Processing Sequence Used for the Overlap Process

7. The Nikon shutters were used to define a pattern ~1 cm by 0.5 cm which overlapped the edge of the previous pattern by ~0.5 cm (see Figure 1).
8. The nitride was etched. Etching or the silicon in the exposed part of the previously etched wells was observed, indicating that it should be possible to pattern metal within the wells.
9. The samples were etched in KOH, 4M 85° C, a further 16 mil. This results in targets of 16 mils for the shallow portion of the well and 19.94 mil (100 micron + 16 mil) in the deep portion.
10. Observation were made of the actual depth of the wells, the interface between the two different depths and the state of the corners between the two well depths. The results of these observation are given below

Actual well depth measurements were taken on the two wafers.

Wafer #1	deep	shallow	difference
	20.75	16.6	4.15
	20.75	16.8	3.95
	20.4	16.55	3.85
	20.5	16.6	3.9
	20.3	16.55	3.75
	20.4	16.5	3.9
Wafer #2	20.35	16.4	3.95
	20.3	16.25	4.05
	20.3	16.4	3.9
	20.15	16.25	3.9
	19.95	16.2	4.05
	20.25	16.2	3.75

The difference between the two measurements, average, 3.93 mil, range 3.75-4.15 mil, is in excellent agreement with the initial depth etched 3.75-3.94 mil (target 3.94). The deep wells average depth, 20.35 mil, range 20.75-19.95 is in good agreement with the target, 19.94 mil with all values being within 1 mil of the target. The shallow well depth is somewhat high, 16.45 mil, range 16.2-16.8 mil (target 16 mil).

An alpha step scan of the transition between the two depths is shown in Figure 2. Tens scans were taken of different wells and all of them had the same general shape. There is initially a sharp fall off from the shallower to the deeper portion, followed by a tail. The distance from the start of the fall-off to the point where the depth is within 0.25 mil of the final depth is ~24-30 mil. the shape of this transition is somewhat unexpected, but appears to be consistent.

As expected, the outside corners are attacked and rounded by the KOH, Figure 3. all the corner had the same general morphology as that displayed in Figure 3. this effect could be eliminated in the case where the edges of the second well correspond with those of the first. There are reports in the literature of ways to preserve outside corners. These approaches were not tried and the geometries are somewhat different.

In order to implement the overly approach, it is necessary to spin on photoresist to pattern the second well. It is well known that indented features can lead to pooling of the photoresist and may also spin out in a tail. It is clear that the depth of the well will lead to focusing problems within the well. These effects for this approach created no serious problems. The second well was designed to significantly overlap the first and the photoresist in the bottom of the wells was patterned and the silicon was clearly etched during the nitride etch process. This illustrates that it is possible to pattern in the wells to some degree. Tail of resist were observed. However, these did not seem to adversely effect the photoresist patterning.

While patterning in the wells and comets tails do not appear to be serious problems, pulling away of the photoresist from the edges of the etched features during spinning was found to be problem. this problem is schematically shown in Figure 4. During spinning, photoresist seems to pile up on the lip of the well. Photoresist on the flat portion above lips being spun off. Since it is not being replenished, the photoresist either thins down or completely removed. In this case, standard photoresist processing was used and the pull back was clearly visible to the eye and was worse on the edges. This pullback clearly effects the dimensions of the well and during metal patterning will result in unintentional cutting of the metal lines. This problem could be minimize by using different photoresist and spinning procedures. Two levels within a single well are possible, while keeping the well depth specification range within 1 mil. Three levels within 1 mil may be possible, but, because of the additive nature of using this technique would require further study.

3-D Well Repeatability Experiments

A lot of 25 wafers were processed to investigate the repeatability of well formation. 25 n-type monitor wafers were coated with a ~1175A of VTR CVD SiN. These wafers were then coated with Hunt Photoresist 204, and baked at 105° C for 100 sec. The 3-D well's 2 inch mask was used on the OAI contact aligner to pattern the photoresist. Hunt 419 developer and a 105° C, 60 sec. post bake were used. The SiN was SF_6 plasma etched with a programmed 100% over etch, and the photoresist was stripped suing PRS 1000. The wafers were next cleaned in 120° C 10:1 sulfuric acid - hydrogen peroxide bath for 5 minutes, DI rinsed and then rinsed in 1% HF solution until the well areas were hydrophobic. The wafers were again DI rinsed until the rinse water resistivity was 14 MΩ.

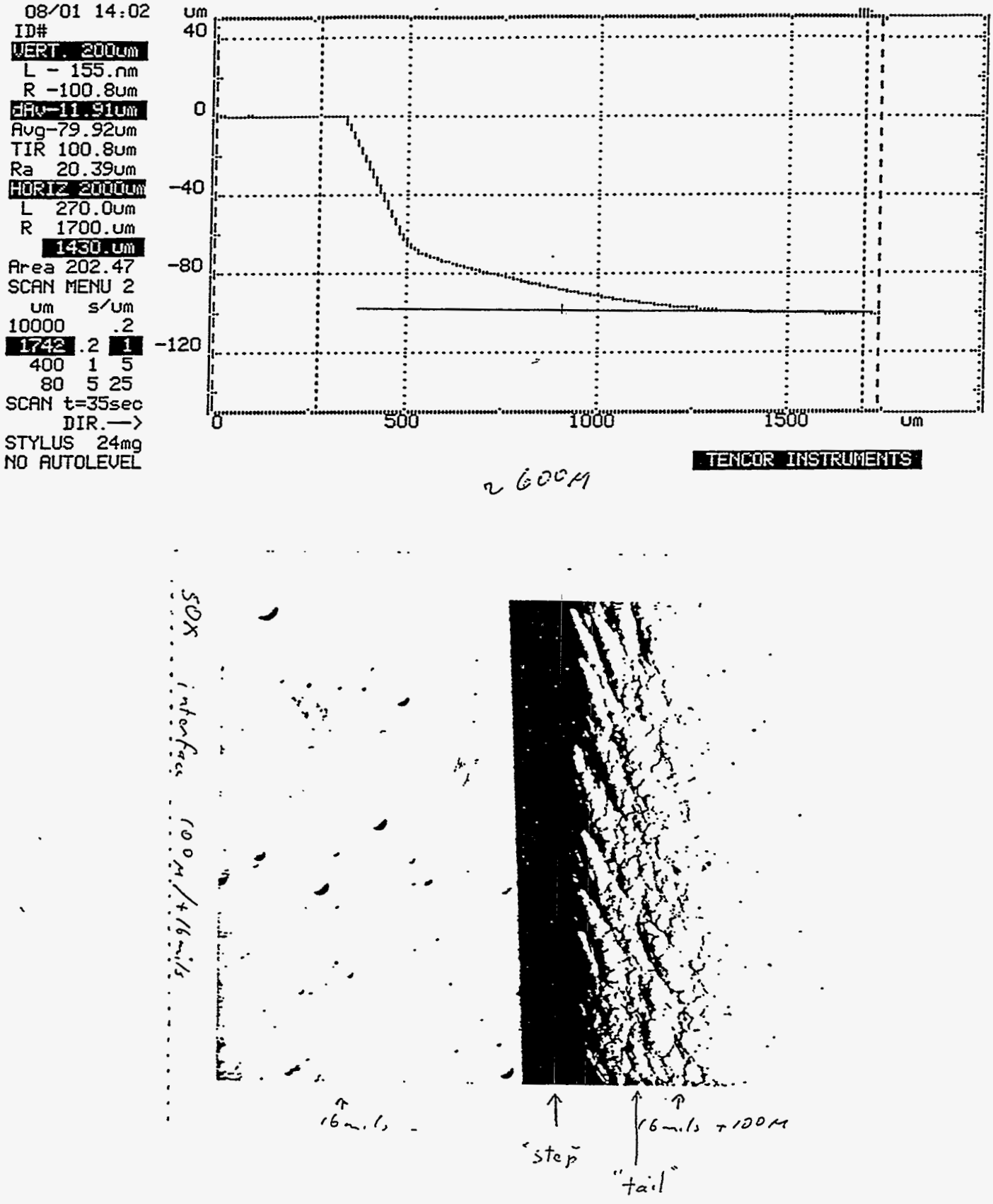


Figure 2. Alpha Step Trace of the Profile of the Step between the Two Well Depth and Optical Micrograph of the Top View of the Step

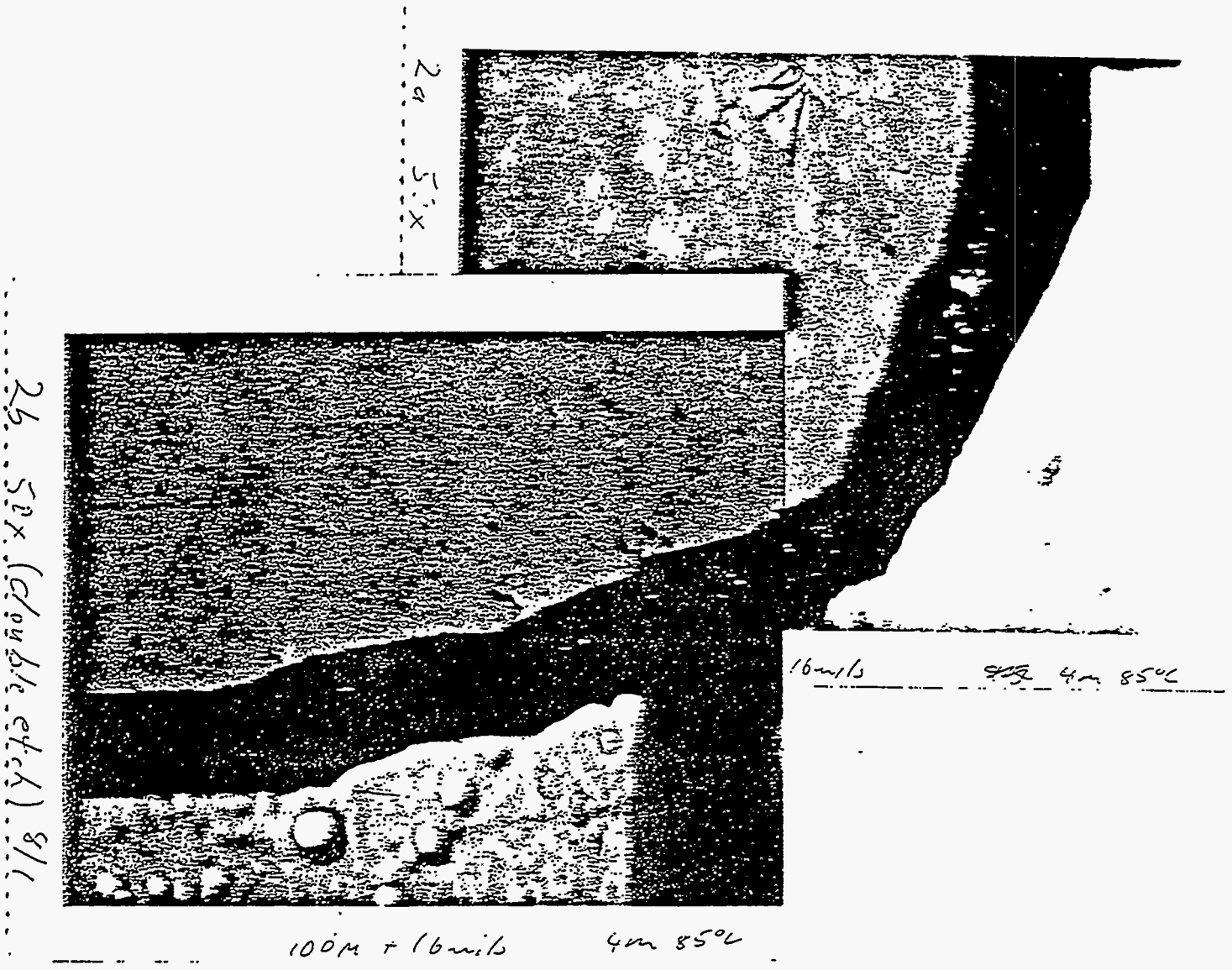


Figure 3. Composiste Micrograph Showing the Corner Morphology between the Two well Depth (50 Times Mag.)

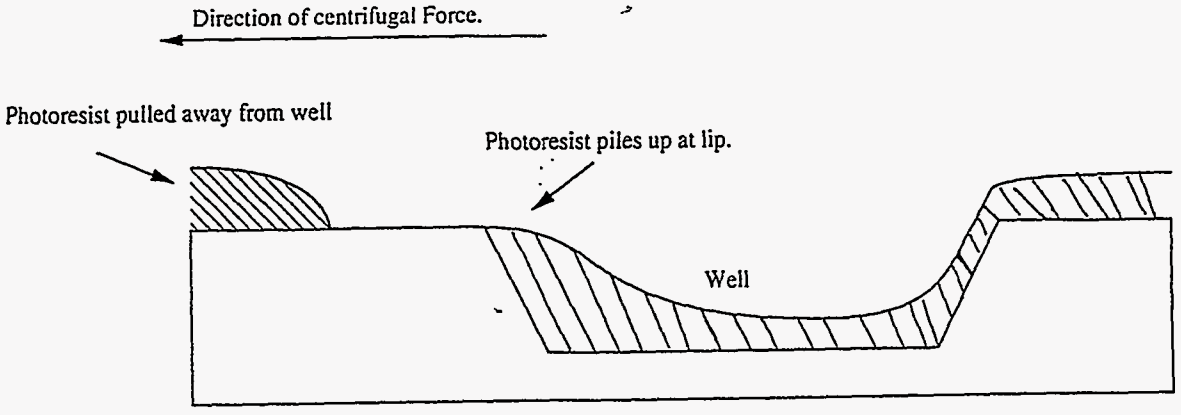


Figure 4. Schematic Representation of the Photoresist Pull Back Problem

The well were formed using a KOH etch. The KOH used is a pelletized semiconductor grade which contains ~14% water. This mixed with DI water to make an etch solution of 4M not correcting for the water of hydration; the corrected molarity is 3.43. This solution was heated in a quartz tank inside a heated recirculating water bath to 85° C before the wafers were immersed. The solution was not agitated during the etch except by the bubble action of the etch process. The targeted etch depth for this lot was 13 mil +/-0.5 mil. From past experience, the etch rate with the above conditions is ~ 1 mil per 15 minutes. One wafer from each batch of wafers was pulled to confirm the etch rate and to determine the exact etch duration. For this experiment the lot of 25 wafers was etched in ten runs. Each run had from one to eight wafers.

Of the 100 well produced, four per wafer in the experiment were inspected. Figure 5 shows the inspection from that was generated for each of the 25 wafers. 72 of the wells passed. The break out of the data is as follows:

1. Good. These wafers had none of the below mentioned defects, but were not necessarily perfect. allowed imperfections included orange peel texture, haze, etch depth tolerance of +/-0.5 mil for the target of 13 mil, pits in the well bottoms which did not go through the remaining silicon, nitride defects and nicks in the well walls which did not impinge on the outer well edges, and bumps rising less than 0.1 mil from the well floor. 60 of the well were classified as good.
2. These wafers had "finger ridges" less that 1 mil in heights as measured from the well floor and extending less than 1 mm into the well from the base of the walls. 12 of the wells were in this group.
3. Bad. These were broken down into the following sub-classes:
    - Fingers greater than 1 mm long or bumps greater than 0.1 mil in height. 15 of the 100 wells were in this group.
    - Breakage. Two of the wells were in this group.
    - Out of spec. wells - either too shallow or too deep. 7 of the wells were in this group.
    - Nicks around the exterior perimeter of the well. 5 of the wells were in this group.

In the course of this experiment, it became apparent that the wafer which was pulled during the etch and measured, to determine the etch rate, was generally hazier and had more defects than the wafers which remained in the etchant. A side experiment was done to confirm this observation. Four wafers of eight wells each were patterned and etched. Two wafers were removed from the etchant was measured, the other two were left in the KOH. Of the 16 uninterrupted etch wells only one had finger defects, and both of these measured wafers were much hazier than the unmeasured. This knowledge was incorporated into the later runs of the repeatability study, and the wafers which had to be removed to be measured, were removed as late as possible to minimize the size of the finger defects.

7/6/1

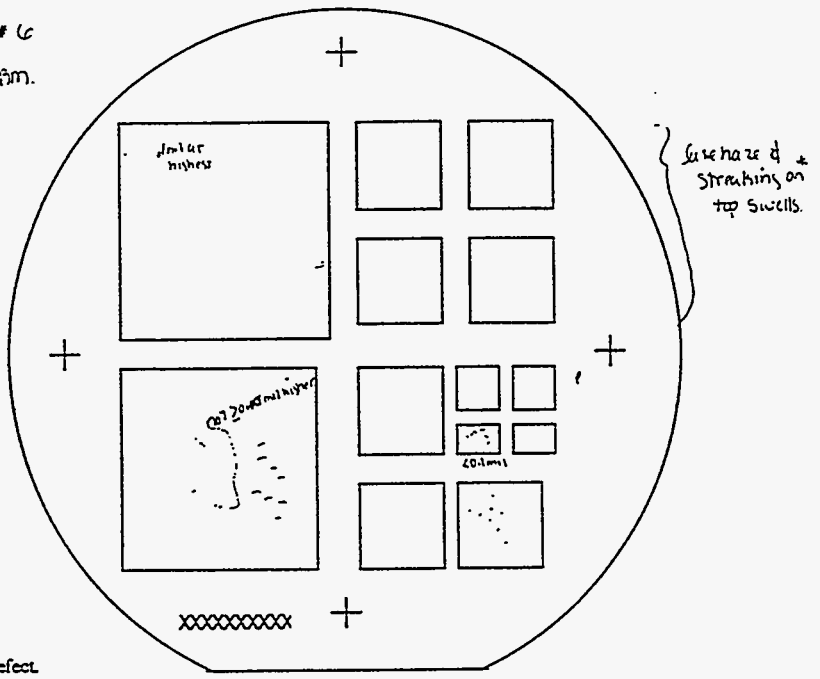
Lot 7

Wafer # 1 (Scribe on ~~back or edge~~) Front Flat 6/1

date:

KOH run # 6

Initials: PJM.



Key:

F= finger defect.

H= Haze

P= Pit in surface.

N= Nick in wall.

A= Apparent Pattern/TIR value

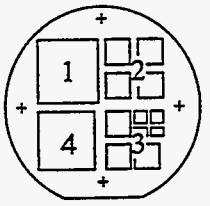
X= Scrap.

Overall depth: 13 mils

Range: 12.85-13.10

NOTES:

Well Ordering.



eg, 10-4-3 stands for KOH run 10, wafer 4, well 3

✓

† where streaking is (scratch-like) can see change in surface texture under scope - finer grained crystal material is often deeper than surrounding surface

Figure 5. Form for Inspection

During the course of the study, the wafer handling improved, as well as the photolithography techniques. If the same experiment were to be repeated, it would be reasonable to assume that the overall etch quality would be improved. The greatest quality improvement would be by maximizing the number of wafer per etch group, so that the ratio of measured to uninterrupted etched wafers be as small as possible. This was observed in the study. In one run performed near the end of the study, eight wafers were etched in a group. Of these eight wafers (32 wells total), there were no rejects, no finger defects, one wafer had raised bumps on the well floor. These bumps were less than the rejection criterion of 0.1 mil in height.

In summary, at least 72% yield could be obtained using an acceptance criteria of :

- Wells with a depth within 0.5 mil of specification.
    - No ridges or bumps on the floor of the well greater than 0.1 mil
    - No finger ridge defects longer than 1 mm extending away fro the wall at the base of the well.
- Acceptable wells would be free of nicks in the sidewalls greater than 0.5 mm .
- Have no nitride defects impinging on the wells.

CHAPTER 5 -SCREENING ICS ON THE BARE CHIP LEVEL: TEMPORARY PACKAGING

Abstract

Several different temporary packaging concepts of integrated circuits for pretest at speed and burn-in are introduced. Temporary packaging is achieved using standard labor and equipment resources already employed in permanent packaging. Experiments were carried out to validate the pretest process, and results are presented for the various materials used in the pretest process. The preferred method for temporary packaging along with the selected materials used is presented. Temporary packaging of integrated circuits for pretest with reasonable yield is demonstrated as feasible.

Introduction

Temporary packaging of integrated circuits (ICs) is required in order to perform chip screening prior to its use in a multichip module or in chip on board. In today's electronic industry, ICs typically receive extensive electrical and aging testing only *after* they have been individually placed in the final packages. It is generally not possible to test an IC chip at speed using wafer probe techniques[1]. The desire to increase system performance by putting 10 to 50 IC chips in one package for multichip module applications makes it necessary to perform extensive electrical and reliability testing at actual speed, temperature, and burn-in conditions before commitment of each chip to final assembly. If this is not done, extensive rework could be necessary.

A promising solution for pretest would be a method of temporary packaging in a robust package. The package needs to be compatible with today's military standard requirements. Temporary packaging would allow IC chips to be extensively tested as in the past. Only now, the chips are tested prior to placement. The chips could then be extracted undamaged for use in multichip modules.

It is imperative for the success of multichip modules with many chips that there be a near 100% chip reliability prior to assembly[2]. In most of the multichip module architecture existing today, rework is expensive, time consuming, and technically difficult. Anything that reduces the amount of rework required will have a major impact on multichip module cost.

Methods

For temporary packaging, there are three primary technical problems that must be addressed: die attachment, electrical interconnection, and package lid sealing. Figure 1 illustrates one concept.

Die attach, approaches include:

- a vacuum attachment of the die through the package during wire bonding
- a re-workable thermoplastic
- various types of solvent removable polymers

The latter two are the die attach techniques which were concentrated on.

For electrical interconnection to the package, the approaches include:

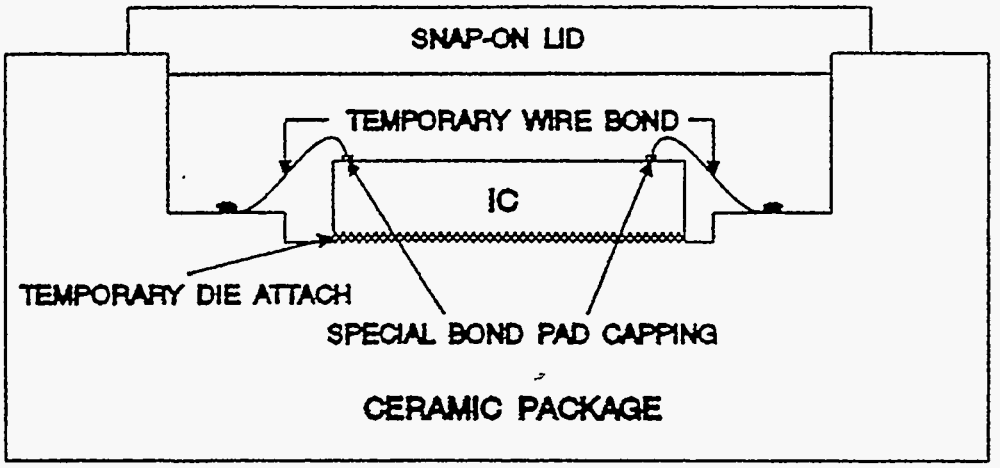


Figure 1 Temporary Packaging Concept

- the use of very low pressure wire bond attach to die pads, with removal after testing
- the use of copper wire for wire bonding and subsequent chemical etch off with nitric acid
- the use of plating to provide a temporary capping of bond pads on the die and wire bond with Aluminum, Copper or solder wire
- flipchip interconnect on temporary test substrate

The low pressure wire bond technique was the easiest to implement, and the most consistent with standard assembly practices.

For package lid sealing to protect the die, the approaches investigated include:

- snap-on lids
- taping the lid down with various adhesives
- sealing the lid with polyimide bead

Taping the lid down proves the simplest and most flexible in implementation. In pretest, there is no hermeticity requirement.

Extensive testing on extracted die after repackaging has been done to determine the effect of temporary packaging on performance and reliability.

Experiments

An experimental matrix was designed to validate the die pretest process. The intent of the experiment was to determine the electrical yield and success rates after dice were assembled, submitted for burn-in, disassembled and reassembled into a final test vehicle. Figure 2 shows the

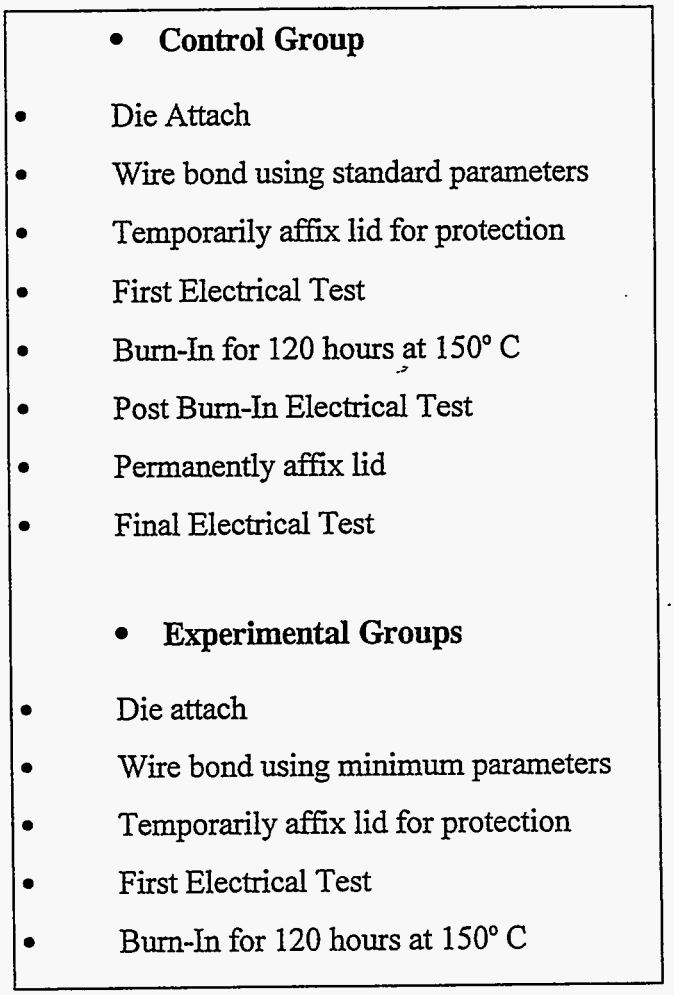


Figure 2 Assembly Process Flow

assembly process flow for both control groups and experimental groups. The main difference between the control group and the experimental groups is that (in the control group) there is no disassembly or reassembly of the IC. In addition, the package lid is permanently attached for the control group.

Discussion

A 16K SRAM fabricated by AT&T was selected as the device vehicle based on availability. A twenty pin dual-in-line multilayer ceramic package was used as the temporary and final package due to the ease of testing.

A test matrix was developed to include three experimental temporary attach mediums as well as a control group. Each of the die attach mediums were split into two sub-groups, one bonded with .001 in. aluminum wire and the second bonded with .001 in. gold wire. In the control group there was an additional subgroup bonded with gold ribbon (.001 in. x .003 in.) wire.

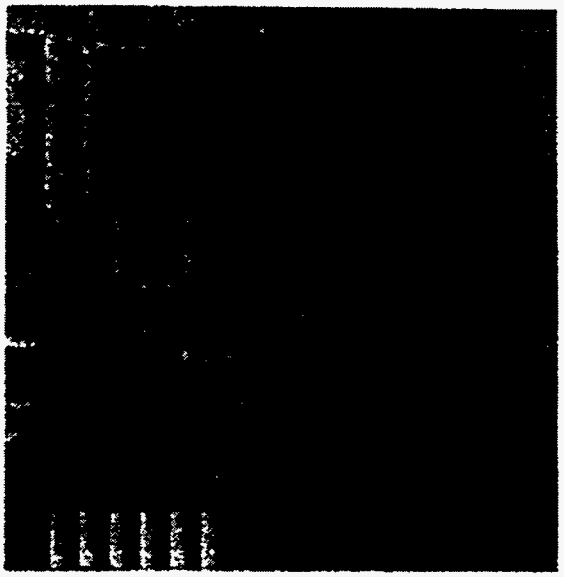


Figure 3 Bond Pad with Aluminum Wire

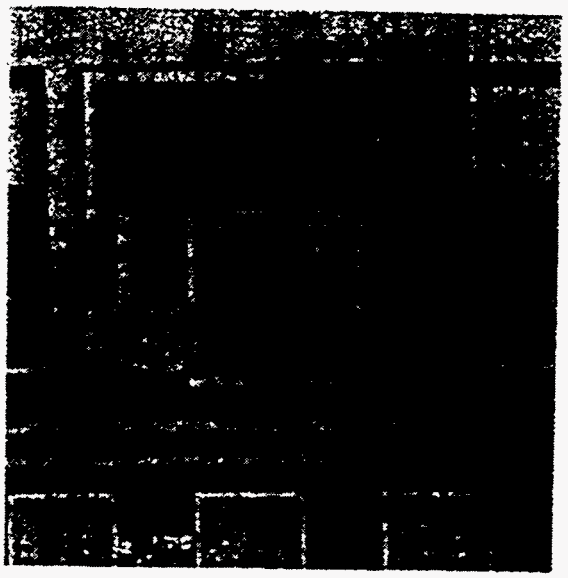


Figure 4 Bond pad with aluminum wire removed.

The control groups were eutectically die attached with 99.9% gold preform. The three experimental attach mediums included two thermoplastic adhesives and one positive photoresist.

The photoresist was manufactured by WAYCOAT, part number HPR-206. The first thermoplastic adhesive was a low dielectric constant preimidized Polyimide Siloxane manufactured by Altisil, part number SPI115. The second was an unfilled, insulative, fully polymerized thermoplastic adhesive manufactured by Staystik, part number 173. The evaluation die attach materials were manually dispensed and allowed to cure at room temperature for a period of 24 hours.

Aluminum wire bonding parameters were established to achieve sufficient electrical connection while still maintaining ease of removal. Although the intent was to completely remove all evidence of the temporary aluminum wire bonds, there was a fairly high percentage of bond residue remaining on the aluminum bonding pad. Of the 1760 bond pads inspected 1113 (63%) had some evidence of the original bond remaining on the pad. This high occurrence is most likely associated with the fact that the parts underwent burn-in at 150° C for an average of 140 hours causing annealing of the aluminum wire to the aluminum pad. For example, prior to burn-in, monitor devices were bonded with the wire removal yield with no residue in the 80-85% range. Fig. 3 shows a bond pad with aluminum wiring. Fig. 4 shows a bond pad with the aluminum wire removed. The evidence of remaining bond residue was not of great concern during the repackaging of the part as there was sufficient space left on the bond pad to attach new wire connections.

Gold wire bonding parameters were established to achieve minimum damage to the bonding site on the die. Since gold wire bonds are virtually impossible to completely remove from the bonding pad, a portion of the bond "foot" remained on every pad. Bonds were offset to allow sufficient space to bond directly next to or slightly over the original bond "foot".

Several problems arose during the experiment that significantly reduced the quantity of the IC parts used for testing. The individual group quantities were reduced from 250 each to 100 each based on burn-in board limitations. In addition, during the removal process of the first photoresist group, 29 of 43 dice were lost due to ground pad corrosion. Analysis of the problem revealed that the distilled water used in the cleaning process had a high conductivity reading. The problem occurred because our laboratory is equipped with a portable distilled water system which is only activated as needed. The conductivity of the water could be brought to an acceptable range by continuously running the distilled water for 15 to 30 minutes prior to use. After running the water, a sample of photoresist attached die were cleaned to verify that the corrosion problem had been alleviated entirely.

Observations

A small amount of die attach residue, primarily from the photoresist and Staystik thermoplastic attachment, remained on the backside surface of the die immediately following the die removal process. Mechanical scrapping to clean the die backside was required prior to reattachment. In contrast, the Polyimide Siloxane was completely removed with N-MethylPyrrodlidone (NMP), thus this material is the suggested option.

X-Ray analysis of the backside surface was performed on a sample die from each of the temporary attach mediums. Samples were selected immediately following the die removal process to identify if any residual contaminants were present. None were found.

The tape used to temporarily affix the lid for protection during testing did not withstand the burn-in temperature of 150° C. Therefore, high temperature polyimide tape is recommended. Alternative temporary lid recommendations include the use of spot welding or the use of similar materials used in the temporary attach.

Table I shows the electrical yield results of the control groups. The column labeled "Burn-In" is the number of parts available after burn-in. The column labeled "Post Elec." is the number of

functional parts after electrical test. The column labeled "Final Elec." is the number of tested functional parts after permanent lid seal. There were two groups bonded with aluminum wire, one group bonded with gold wire, and one group bonded with gold ribbon. After burn-in the electrical yield of the parts ranged from 90.5% to 100% with a total weighted average yield of 96.5%. After permanent lid sealing the parts were tested again. Using the post electrical parts as the standard, the yield ranged from 94.7% to 100% resulting in a total weighted average yield of 97.3%.

Table I: Control Group

Wire	Burn-In	Post Elec.	Post Elec. % Yield	Final Elec.	Final Elec. % Yield
Al	50	49	98	47	95.9
Al	20	20	100	20	100
Au*	21	19	90.5	18	94.7
Au	25	24	96	24	100
Total	116	112	96.5	109	97.3
*ribbon					

Table II shows the results using photoresist as the die attach medium. The column labeled "Burn-In" is the number of parts available after burn-in. The column labeled "Post Elec." is the number of functional parts after electrical test. The column labeled "Re-Pkg." is the number of parts available after disassembly and reassembly. The column labeled "Final Elec." is the number of tested functional parts after reassembly. Yield after burn-in ranged from 89% to 100% with a total weighted average yield of 95.8%. Some parts were lost, due to handling, during extraction. Thus, resulting in a repackaging yield of 93.7 % to 100% and a total weighted average yield of 94.2%. Using the post electrical parts as the standard, these parts were retested resulting in a final electrical yield of 63% to 93% (the worst being gold wire bonded parts) resulting in a total weighted average yield of 83%. Handling loss could be driven to zero as these processes become more standardized.

Table II: Photoresist Die Attach

Wire	Burn-In	Post Elec. (P.E.)	P.E. % Yield	Re-Pkg. (R.P)	R.P. % Yield	Final Elec. (F.E.)	F.E. % Yield
Al	18	16	89	15	93.7	14	93
Al	35	34	97	31	91.2	28	90.3
Au	19	19	100	19	100	12	63
Total	72	69	95.8	65	94.2	54	83

Table III shows the results using Polyimide Siloxane as the die attach medium. Yield after burn-in ranged from 98% to 100% with a total weighted average yield of 98%. Some parts were lost after repackaging resulting in a repackaging yield of 92.3 % to 95.2% and a total weighted average yield of 91%. Using the post electrical parts as the standard, these parts were retested resulting in a final electrical yield of 83% to 100% (again the worst being gold wire bonded parts) and a total weighted average yield of 91.2%.

Table III: Polyimide Siloxane Die Attach

Wire	Burn-In	Post Elec. (P.E.)	P.E. % Yield	Re-Pkg. (R.P)	R.P. % Yield	Final Elec. (F.E.)	F.E. % Yield
Al	50	49	98	47	95.9	41	87.2
Al	39	38	97	32	84.2	32	100
Au	13	13	100	12	92.3	10	83
Total	102	100	98	91	91	83	91.2

Table IV shows the results using Staystik as the die attach medium. Yield after burn-in range was 100%. Some parts were lost after repackaging resulting in a repackaging yield of 90.7 % to 100% and a total weighted average yield of 94%. Using the post electrical parts as the standard, these parts were retested resulting in a final electrical yield of 93% to 100% and a total weighted average yield of 94%.

Using Polyimide Siloxane as the temporary die attach material and aluminum wire as the temporary interconnection results in the least residue with repackaging yields of up to 100%. Using photoresist as the temporary die attach and gold wire as the temporary interconnection results indicate the worst combination.

Table IV: Staystik Die Attach

Wire	Burn-In	Post Elec. (P.E.)	P.E. % Yield	Re-Pkg. (R.P)	R.P. % Yield	Final Elec. (F.E.)	F.E. % Yield
Al	75	75	100	68	90.7	63	93
Al	19	19	100	19	100	18	95
Au	20	20	100	20	100	20	100
Total	114	114	100	107	94	101	94

Summary

The ability to reliably pretest die by means of temporarily packaging has been demonstrated on a standard assembly line. The best wire bond method uses aluminum with minimal bonding parameters. Respectable yields were found with both the STAYSTIK and Polyimide Siloxane thermoplastic adhesive samples. The best lid attach method is to tape the lid down with high temperature polyimide tape. Additional experimentation using the photoresist is required prior to recommending it as a temporary attach medium. Continued efforts are underway on minimizing the bond residue remaining during assembly. The additional cost to Sandia for a temporary package is $4.00 per part.

[1] D.C. Keeser, "Bare Die Testing and MCM Probing Techniques", in the Proceedings 1992 IEEE Multi-Chip Module Conference, 1992, pp. 20-23.

[2] M. M. Salatino and R. C. Braken, "Die and MCM Test Strategy: The Key to MCM Manufacturability" in the Proceedings of Eleventh IEEE/CHMT International Electronics Manufacturing Technology Symposium, 1991, pp. 440-445.

This work was published in the IEEE Transactions on Components, Hybrids, and Manufacturing Technology, Vol. 16, No. 4, June 1993, pp. 394-395, by D. Chu, C. A. Reber, and D. W. Palmer.

CHAPTER 6 - MASKLESS FLIP-CHIP SOLDER BUMPING TECHNIQUE

Abstract

Conventional flip-chip solder bumping, utilizing masks, evaporation, and plating equipment, is ideally suited for the mass production of components. Large batches of wafers can be deposited with multilayer solder at a single time. However, this process can be excessively costly, time consuming, and unnecessary for small, specialized applications. A process has been developed to solder bump individual pads, one at a time, without the need for expensive masks, evaporation chambers, or die layout information. This information, critical for solder bump placement, is typically not available to sources outside the design house. Using a standard ball wirebonder and specialized bonding wire, solder bumps can be formed on both Aluminum and Gold bond pads of integrated circuits (IC) using thermosonic bonding in a forming gas (Hydrogen-Argon). The bonding effectively wets the solder bump to the pad without the use of special intermetallic adhesion layers. This process can cut prototyping time and cost by eliminating the need for elaborate equipment and complex solder compositions without sacrificing electrical or thermal performance.

Introduction

The advancement of integrated circuits has resulted in exponentially increasing chip densities. This has put great demands on the functionality and reliability of ever-increasing numbers of input / output (I/O) connections. [1] Wirebonding, although still the most common chip-level interconnection method, is unable to satisfy the large numbers of I/O's demanded by current and future devices. A search for a more robust interconnection scheme led to a technique referred to as Flip-Chip bonding or Controlled Collapse Chip Connections (C4) [2]. Flip-chip bonding allows for area arrays of interconnects. This in turn permits higher lead counts and thus denser designs.

The process employs solder bumps or balls to electrically, as well as mechanically, connect an integrated circuit device to a substrate. The solder bumps are first applied to the metallic bond pads on the device. The solder must be able to wet to the bond pad to form a stable junction. The device is then flipped over (hence flip-chip bonding) and attached via solder reflow to a substrate, which possesses a mirror image of the device's bond pad layout (Figure 1).

The central piece of equipment used in solder bump formation is an evaporation chamber. Typical evaporators possess numerous metal loads enabling sequential as well as simultaneous evaporation. Metals commonly used include Cu, Cr, Sn, Pb, Au, Ag, In, Pt, and Pd. State-of-the-art systems can cost hundreds of thousands of dollars. The solder is deposited through holes in a metal mask, which is clamped to the wafer. A different mask is required for each bond pad layout. Thus this process is suited for mass production, but has the following disadvantages:

- complicated process
- higher facility cost
- environmental impact
- not possible to bump at chip level

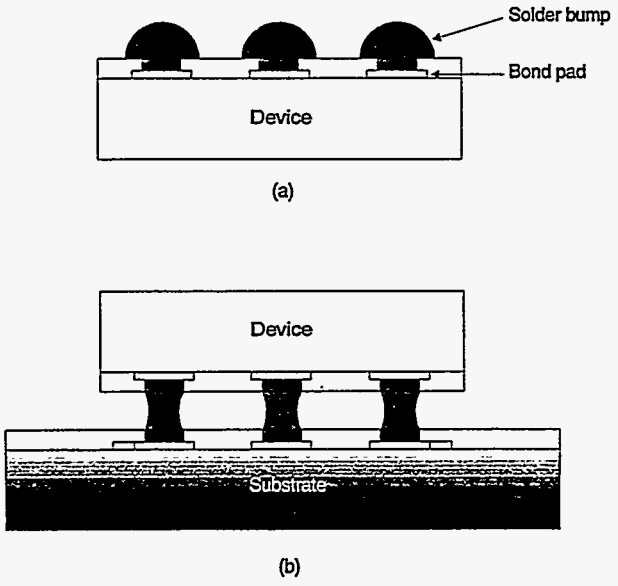


Figure 1. Illustration of flip-chip attach process. (a) Wetting of solder bump to bond pad. (b) Device to substrate attachment via solder reflow.

As will be discussed later, our process requires no evaporation chambers or masks.

Solder composition is a major issue in flip-chip interconnection technology. The choice of solder influences the electrical, mechanical, and thermal characteristics of the junction. One such composition is illustrated in Figure 2(a).

A layer of Cr is necessary for proper solder adhesion to the Al pad and to serve as a diffusion barrier between the Al and the solder. A phased Cr+Cu layer provides resistance to multiple reflows. [3] This complex array of metals can be replaced with a Sn-Ag alloy (Figure 2(b)). The solder effectively wets to Al or Au pads without the use of an adhesion layer. Moreover, the low temperature solder is easily reflowed.

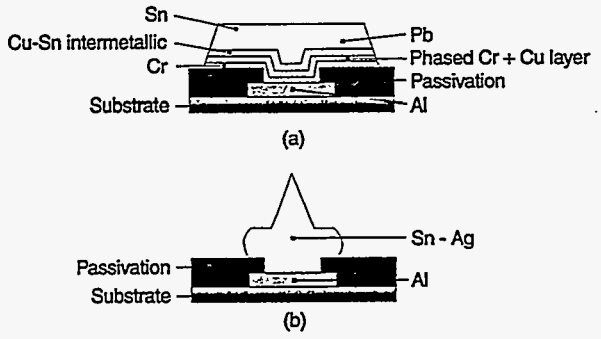


Figure 2. Illustration of the multilayer deposited solder. (a) Standard flip-chip solder composition prior to reflow. (P. A. Torta and R. P. Sopher, "STL Device Metallurgy and Its Monolithic Extension," *IBM J. Res. Dev., Vol.13. No. 3, May 1969, p.226.* (b) Sandia single layer solder bump.

Bump Formulation on Die Bond Pads

Instead of the methods discussed previously, we are using a conventional manual wire ball bonder to formulate different types of solder bumps on bond pads of die. Advantages to this method included: less complex method of bump formulation: dry and clean process;controllable bump size; maskless process; chip level bumping.

•The solder bump metalization that can be made directly on the die pads are: Pb/Sn bumps on Au pads; Pb/Sn bumps on Au bumped Al pads; Sn/Ag on Al pads.

A Kulicke and Soffa (K&S) Model 4124 Universal Thermosonic Ball bonder with forming gas wand and bumping software is shown in Figure 3 was used to achieve this type of bumping with different types of solder wire. The bonding method used is thermosonic under a forming gas atmosphere.

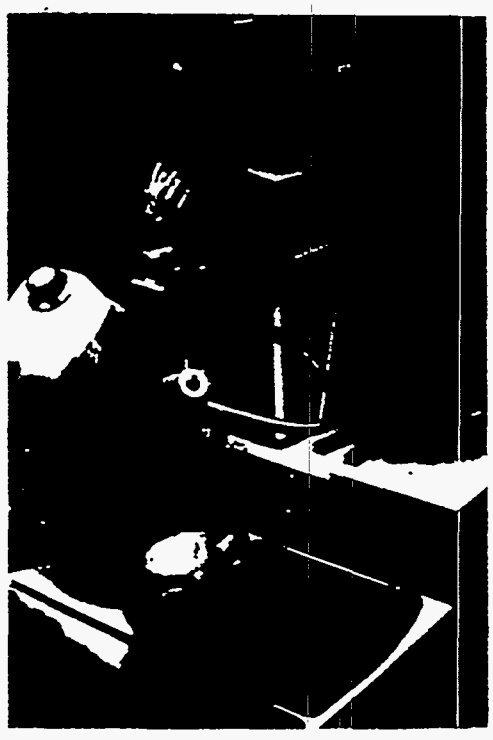


Figure 3. K&S Model 4124 Bonder

There are three methods of wire bonding to the pad metalization. They are thermocompression, ultasonic, and thermosonic. Thermocompression bonding is achieved by pressing the wire against the pad metalization a an elevated temperature. Ultrasonic bonding is achieved by pressing the wire against the pad metalization at the same time a burst of ultrasonic energy is being applied. Thermosonic bonding uses both themocompression and ultrasonic bonding to bond between the wire and the pad metalization. Thus, by using heat and a burst of ultrasonic energy when the wire is being pressed against the pad metalization. a good metallurgical bond between the wire and the pad metalization is achieved. Figure 4 illustrates the bumping method with this thermosonic bonder.

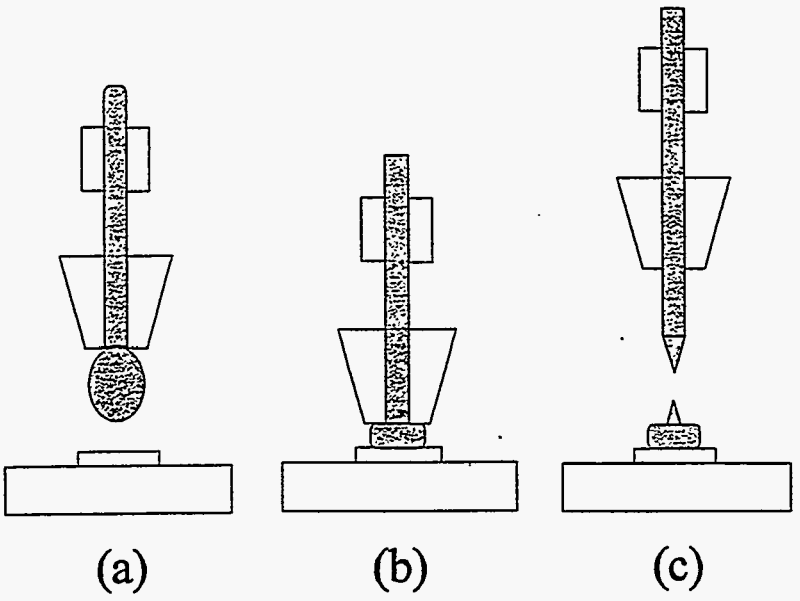


Figure 4. Bumping Method

Two special solder wire formation made by Tanaka Denshi Kogyo K. K. were used Pb/Sn (Tanaka Part # SB18) and Sn/Ag (Tanaka Part # SB22S). The Pb/Sn wire was composed of 60 wt. % Pb and 40 wt. % Sn with a wire diameter of 30 μm and a melting temperature of 183° C. The Sn/Ag wire was composed of 97 wt. % Sn and 3 wt. % Ag with a wire diameter of 30 μm and a melting temperature of 221° C. A solder ball is formed by high voltage arc discharge (1200 V) to the solder wire with the ground being on the wire side in a forming gas atmosphere of 95% Ar and 5% H_2. The ball size is controlled by the voltage, current, and time of the arc discharge.

The solder wire ball is then joined to the pad metalization by using the thermosonic method of bonding under the forming gas atmosphere. After joining the wire is then cut off by the clamp while the capillary tip is moved up in the vertical direction.

The SB18 Pb/Sn solder ball wire bonds only to Au pad metalization. In order to make solder bumps with this solder formulation, on the Al pad metalization, a separate Au bump is put on top of the Al pad metalization prior to solder bumping. The Au bump is then leveled with an flatten tool. the SB18 solder wire is then bumped on top of the Au bump. With the SB22S Tin/Silver solder wire, solder bumps can be formed directly on the Al pad metalization. Figure 5 shows a SB22S 100 μm solder bump with a small tail.

Since this is a manually operated bonder, a small tail is left after bonding. The tails are manually removed. A special flattening tool is used instead of the capillary to level the bumps. Figure 6 shows the SB22S solder bump after leveling.

There are three parameter that affect the thermosonic bonding. they are force, time, and energy. The force is the measurable static load on the capillary during the bonding sequence. the time is the duration of the capillary, wire, and the pad surface are in contact. The energy is the

Table 2. Temperature Runs

Run	Station 1 °C	Station 2 °C	Station 3 °C	Station 4 °C
1	150	175	250	275
2	350	350	350	350
3	250	350	400	25
4	325	350	375	400

The bumped die were placed into a 40 pin ceramic d.i.p. package, which was used as a carrier. The package was clipped to an aluminum block to facilitate heat transfer from the stage to the ceramic. A gold plated cover (typically used to seal a hermetic package) was then clipped over approximately three-quarter of the package cavity. The gap allowed the N_2 gas to completely fill the cavity but prevented the die from blowing out of the cavity from the gas pressure. A continuous flow of N_2 gas at 300 SCFH was pumped into an inert gas hood installed onto the oven. The samples remained in each station for five minutes. There were two additional cool-down stations giving a total run time of 30 minutes. Figure 8 show SB22S solder bump after reflow.

Figure 8. SB22S Solder Bump

In parallel experiments, we attempted to reflow the solder in a Delta Design 2300 environmental test chamber. The bumped die were situated on ceramic sheets and placed into the oven. Again N2 gas was used in the chamber. The oven was set to 300o C and the samples were left to reflow for 20 minutes. The results were mixed and generally unsatisfactory. Because we chose not to use flux to facilitate reflow, the SnO2 did not completely break down in the chamber, thereby preventing the solder from reflowing into a ball. However, we believe that the slight

vibrations caused by the moving belt in the Sikama oven aided in the breakdown of the SnO2 and permitted the solder to reflow correctly..

Flip Chip Bonding

After bump formulation, the bumped chip can be joined to a substrate by using a flip chip bonder. The metalization on the substrate for both solder alloy needs to be of Gold metalization. The flip chip bonder we use is the Semiconductor Equipment Corporation (SEC) Model 4150 manual flip-chip die bonder (Figure 9). this flip chip bonder has a heated motorized x-y-theta

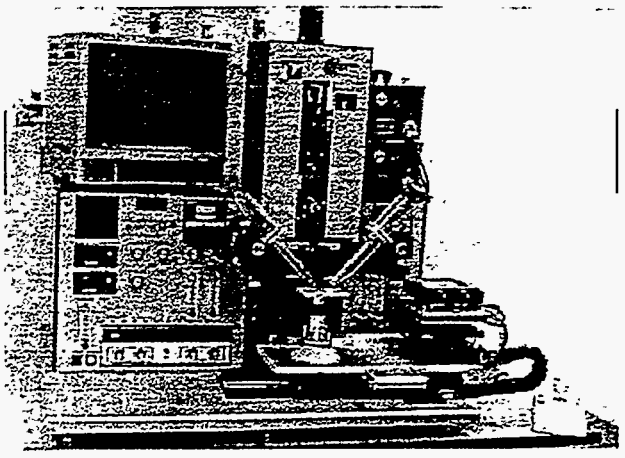


Figure 9. SEC Model 4150 Flip-chip Bonder

work stage and an ultrasonic vacuum pick up head with a programmable motorized z motion. Flip chip alignment is accomplished by a cube beam splitter vision system. The beam splitter will superimpose the die pattern over the substrate bond pattern on the video system for alignment. bonding temperature is 190° C for the SB18 solder bumped chip and 240° C for the SB22S bumped chip. Figure 10 show a bumped die flip -chip bonded to a test substrate.

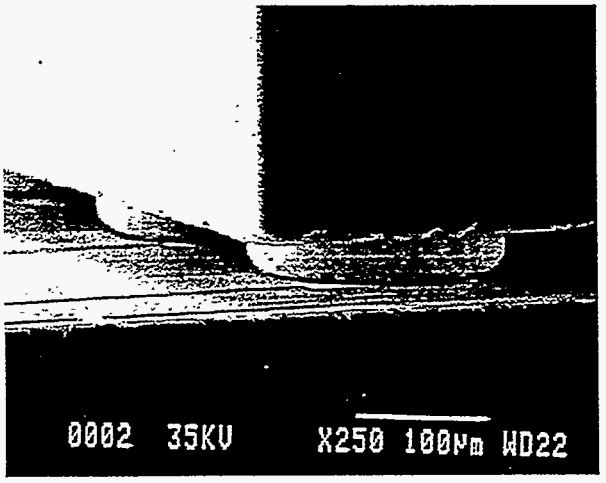


Figure 10. Flip Chip Bonding

Cost

To implement this bumping process, the following equipment is required:

- K&S Model 4124 Bonder with forming gas wand and bumping software - $25K
- Sikama Reflow Oven - $7K
- SEC Flip Chip Bonder - $65K
- Tanaka solder wire - $26 per meter

Thus for less that $100K solder bumping and flip chip bonding can be achieved.

Summary

Using a standard wire bonder, a low cost method of solder bumping method for prototype has been introduced. This process is valid for single die or whole wafer. Solder bumps can be made on Gold and Aluminum die bond pads. This bumping process could be automated. Both K&S and Shinkawa have suitable automatic ball wire bonder for use in this type of application.

[1] Seraphim, D. P., Lasky, R., Li, C. Y., Principles of Electronic Packaging, McGraw-Hill Series in Electrical Engineering, 1989.

[2] Miller, L. F., "Controlled Collapse Reflow Chip Joining", IBM Journal of Research and Development, Volume 13, May 1969, pp. 239-250.

[3] Tummala, R. R., Rymaszewski, E. J., Microelectronics Packing Handbook, Van Nostrand Reinhold, 1989.

[4] Torta, P. A., Sopher, R. P., "STL Device Metallugy and its Monlithic Extension", IBM Journal of Research and Development, Volume 13, May 1969, p. 226.

This work was published in the Proceedings of the 43rd IEEE Electronic Components and Technology Conference , June 1993, D. Chu and D. S. Shen.

Distribution:

Copy	Org.	Name	Mailstop
1	01300	Romig, Alton D.	1079
1	01301	Rohr, Donald F.	1077
10	01333	Chu, Dahwey	1082
10	01333	Palmer, David W.	1082
1	01333	Peterson, David W.	1082
1	01333	Reber, Cathleen A.	1082
1	01333	Seigal, Pamela K.	1082
1	01333	Smith, Simone R.	1082
1	01333	Sweet, James N.	1082
1	01333	Tuck, Melanie R.	1082
1	09113	Bainbridge, Bruce L.	0835
1	01277	Gassman, Richard A.	1073
1	01306	Draper, Bruce L.	1074
1	01323	Fleming, James G.	1084
1	09621	Laguna, Glenn A..	1008
1	03600	Thornton, Anthony L.	1356
5	04414	Technical Library	0899
1	04523	Meyers, Chuck	1436
2	12630	Review & Approval Desk for DOE/OSTI	0619
1	08523-2	Central Technical Files	9018