Future WSI Technology: Stacked Monolithic WSI

Ronald Williams and Ogden Marsh

***Abstract*— This paper describes a methodology for stacking chips vertically and interconnecting them through the chips to achieve a three-dimensional (3-D) circuit. This methodology involves the integration of two technologies: indium bump interconnect technology, historically used to fabricate hybrid focal plane arrays, and the precision thinning of bonded silicon wafers by a process called Åcuthin. Substantial improvements in computing density, power dissipation, and signal propagation time can be realized. This paper describes some of the techniques and benefits of this 3-D interconnect methodology.**

I. Introduction

WAFER scale technology is advancing via the development of larger dice, stacking with chip-edge interconnections, and redundant, or fuse-link, architecture to improve yield. Routing and interconnection requirements for wafer scale integration are established in the literature [1], [2]. J. F. McDonald described a methodology for substantially lowering $x-y$ propagation resistance while retaining the features of negligible z-axis propagation inhibitors [3]. Previous work at the Hughes Research Laboratories (HRL) [4], [5] has provided a method of stacking wafers utilizing microbridge wafer interconnects.

A methodology to implement three-dimensional (3-D) wafer scale technology is described which combines two existing technologies: 1) indium bump interconnect technology (developed at the Hughes Technology Center) used to fabricate hybrid focal plane arrays, and 2) the precision silicon thinning technology called Åcuthin (developed at Hughes Danbury Optical Systems), which utilizes bonded silicon wafers. Circuits are fabricated on bonded silicon wafers and then thinned using the Åcuthin process to allow through-chip interconnects to be made to either side of the IC. Indium bumps allow these thin device layers to be stacked vertically and electrically interconnected to achieve a 3-D circuit. This methodology offers the potential for power savings of orders of magnitude by providing lower capacitance and inductance for drive lines as well as fully depleted transistor junctions. The chip and wafer stacking techniques described in this paper are higher density complementary technologies to the microbridge interconnects developed at the Hughes Research Laboratories.

Manuscript received April 21, 1993; revised August 6, 1993. The Work on the Wafer-Scale Associative String Processor (WASP) is supported by the U. S. Office of Naval Research. This paper was presented at the International Conference on Wafer Scale Integration, San Francisco, CA, January 20–22, 1993.

The authors are with the Technology Center, Hughes Aircraft Company, Carlsbad, CA 92009–1699.

IEEE Log Number 9212409.

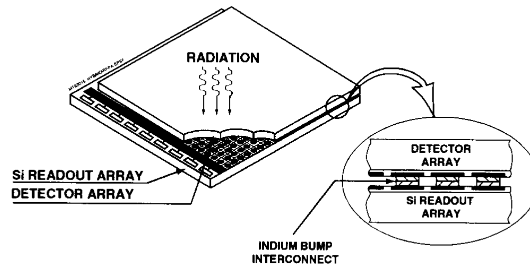


Fig. 1. Hybrid FPA. A hybrid FPA is composed of an IR detector chip connected to a read-out chip via vertical indium bump interconnects at each pixel (diode).

II. Indium Bump Interconnect Technology

Indium bump interconnect technology was originally developed to provide a means of mating an array of IR detectors to an array of amplifier circuits called a read-out (see Fig. 1). As shown in Fig. 1, the readout chip is placed below the detector (relative to the incident light), and each detector pixel is connected to its corresponding read-out circuit through an indium bump interconnect. The resulting two-chip vertical stack is referred to as a hybrid focal plane array (FPA). This technology has advanced to the stage where hybrid FPA's as large as 640×488 pixels, requiring nearly a third of a million interconnects, are being successfully fabricated. Yields of perfectly connected assemblies having as many as 97 000 interconnects per chip pair exceed 94%.

Many different types of detector arrays have been connected to read-outs in this 3-D arrangement, and many such hybrid assemblies have been used successfully in space, missile, astronomy, and airborne applications, having stringent reliability requirements. The various detectors used operate at temperatures ranging from 4K to greater than 77K. Since the read-out arrays are usually fabricated in silicon and the detector arrays are usually fabricated in materials other than silicon (e.g., HgCdTe, InSb), the mismatch in the coefficients of thermal expansion of these dissimilar materials causes mechanical stress on the indium interconnects during thermal cycling. The indium interconnects must, and do, remain intact and fully functional through repeated thermal cycles and the associated mechanical stress.

Indium bumps are produced using wafer processing techniques at the wafer level. Indium is a rather unique metal, softer than most plastics, annealing at room temperature, and able to melt without flowing. Indium bump interconnect technology has exhibited one failure mode: cohesive failure in the bump induced by severe mismatch of the thermal expansion characteristics of the different read-out and detector

0148–6411/93$03.00 © 1993 IEEE

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

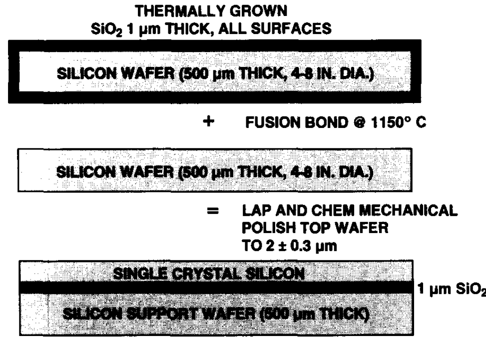


Fig. 2. Bonded silicon wafers.

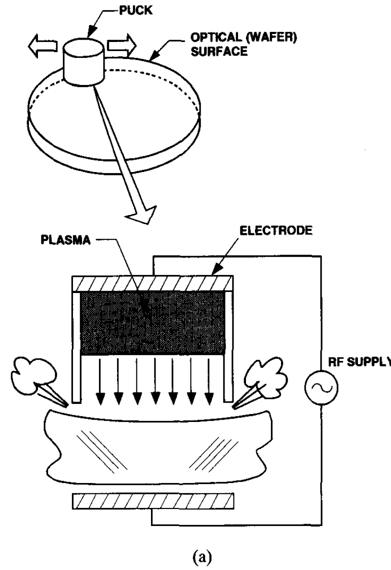
materials over many cool-down cycles (greater than 100) from room temperature to 77K or colder. However, since indium reanneals at room temperature, there is little concern about interconnect fatigue or cohesive failure occurring for chips operated in the normal military temperature range (-55–125°C). Also, for chips of similar materials (e.g., silicon and silicon) interconnected with indium bumps, this failure mode is eliminated regardless of thermal cycling because both chips expand and contract at the same rate. In addition, while indium has a tensile strength of only 9% of the more common solder bumps, its creep strength for sustained loading is twice as high.

These qualities make indium bump interconnect technology a viable process for use in the construction of 3-D silicon IC's. Preliminary testing at 150°C for thousands of hours shows that the indium bump interconnects remain reliable. The melting temperature of indium is 157°C, and testing has confirmed that bumps loose all strength above this temperature and chips remain in contact only due to surface tension effects.

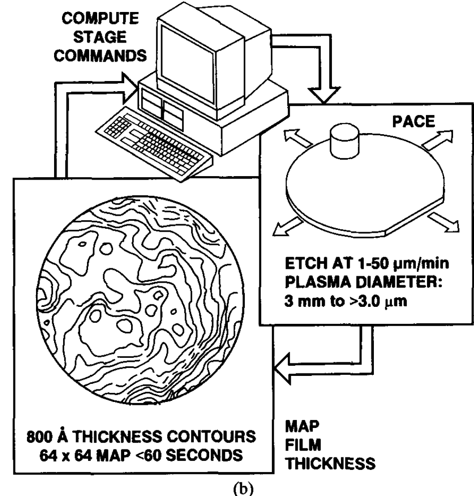
III. Bonded Silicon Wafer Thinning

To effectively stack and interconnect wafers and chips into a 3-D circuit arrangement, two requirements are imposed on the device layers: 1) the 2-D silicon chip or wafer must be thin (<10 μm) to allow electrical contact to both sides of the wafer/chip through trench vias provided in the IC design; and 2) the 2-D silicon chip or wafer must be very flat (<1 μm total thickness variation) to permit indium interconnecting over large circuit areas. These requirements are met by bonded silicon wafer thinning.

Recent developments in preparing silicon-on-insulator (SOI) material have utilized bonded silicon wafers (see Fig. 2). These wafers have a buried 1-μm thick silicon dioxide layer below a thin single-crystal silicon layer. Bonded silicon wafers are commercially available with thin silicon layer thicknesses down to 2 ±0.5 μm. Circuits are fabricated in the thin silicon layer, and the bulk supporting wafer is then etched away to begin the 3-D fabrication process. Production processes to thin a processed IC wafer to 10 μm or less from the backside were not available until recently with the development of the Åcuthin process, shown in Fig. 3.



(a)



(b)

Fig. 3. The Åcuthin process. Åcuthin allows high-precision thinning of silicon.

The Åcuthin process was initially developed for final optical figuring; however, its application to silicon is expected to become significant to the electronics industry. Fig. 4 shows the measured thickness profile for the thin silicon layer of a 100-mm bonded silicon wafer processed with the Åcuthin process. The Åcuthin process removes precise, precalculated amounts of material from the wafer surface by plasma-assisted chemical etching. The surface is first profiled, and a computer program determines the electrode movement profile to remove the desired amount of material. The Åcuthin process can produce wafers which have thickness uniformity within a few angstroms. Similarly, surface flatness of bonded silicon or bulk silicon wafers can be controlled with the same degree of

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

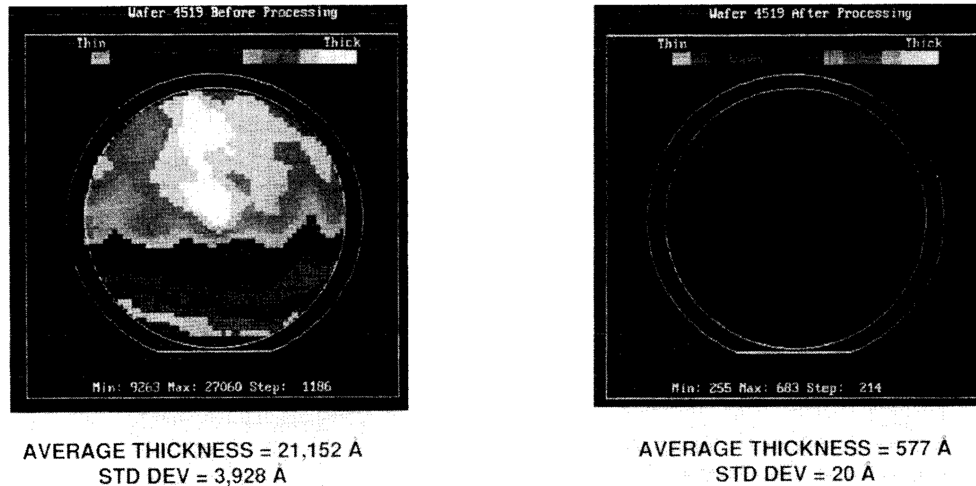
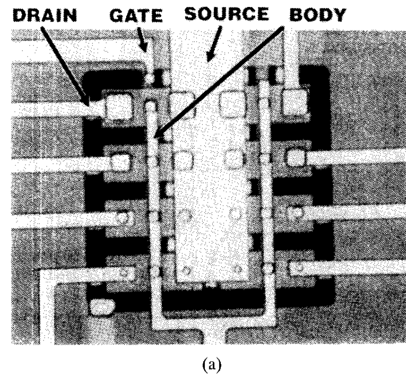
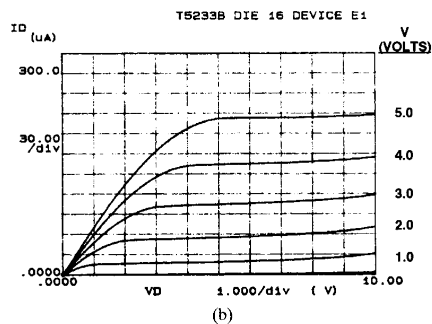


Fig. 4. Åcuthin thinning of 100-mm bonded silicon wafers.



(a)



(b)

Fig. 5. Demonstration of NMOS transistor I–V characteristics after front- and backside Åcuthin wafer fabrication devices are on 0.3-μm transferred silicon with backside body contact.

precision. As shown in Figs. 5 and 6, transistor circuits have been fabricated on Åcuthin material. These transistors are 0.3 μm thick and illustrate the essential vertical-through-silicon contacts necessary for 3-D wafer scale integration.

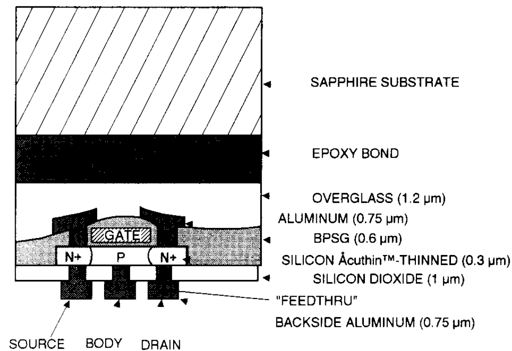


Fig. 6. Wafer incorporating Åcuthin single-transfer and backside processing.

Bonded silicon wafers can be processed in an identical fashion to standard silicon wafers using the same design rules and fabrication procedures. It is even possible to mix the bonded silicon wafers with standard bulk silicon wafers in a process lot with no process compensation needed. Bonded silicon wafers are somewhat expensive at this time, about four times the cost of a standard bulk silicon wafer, but market projections indicate that the cost of bonded silicon wafers will approach that of standard bulk silicon wafers as the bonded silicon market matures.

Recent developments in via technology allow thinned wafers to be electrically accessed from either side (top and bottom) for electrical interconnects. This via technology requires thinned silicon, usually less than 10 μm thick, so that the via holes can be fabricated and insulated during normal semiconductor processing using existing silicon dioxide and metal steps. The thinned wafers are electrically connected with indium bumps on either side of the thinned silicon circuitry. These operations are performed at the wafer level. Hughes has developed

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

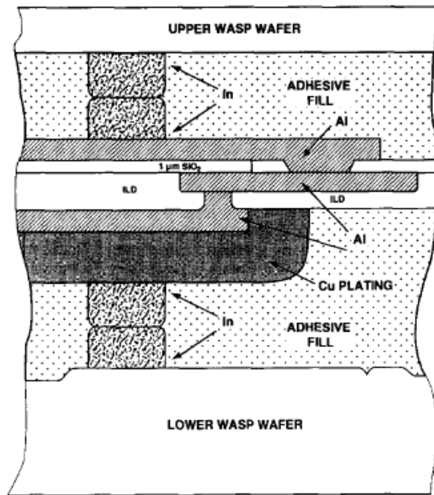


Fig. 7. High density interconnect for WASP 3 D.

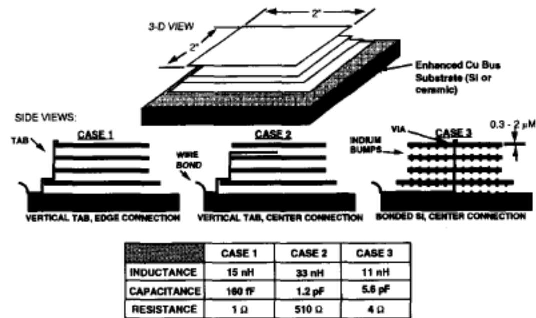
prototype hardware to demonstrate all key elements in this approach to wafer scale integration and stacking.

IV. Stacking Wafers

The first application of this 3-D technology is to be the wafer scale associative string processor (WASP), an exploratory computer application [6], [7]. The associative string processor architecture has the important advantage of being fault-tolerant and able to diagnose and electrically switch around defective cells. Fault-handling capability is necessary in wafer scale projects to compensate for contemporary wafer and die yields. Fig. 7 shows the make-up of the high density interconnect layer. This wafer scale integration requires only a few thousand bump interconnections. Another simplification incorporated into this application is the use of silicon for both wafers: because of chip heating during operation, it would be very difficult to stabilize dissimilar materials over a 2.1 in by 2.1 in area.

The design of a 3-D logic device is an inventive process, because vertical partitioning is not a common element in microcircuit design. The initial selection of the associative string processor simplifies this somewhat, since the processor is composed of many identical cells. Gross modeling of the processor, as shown in Fig. 8, illustrates the necessity of through-chip interconnects rather than the conventional edge interconnect structure. The more common method of stacking and connecting large dice is to metallize or TAB bond edge pads. This method does not allow large interconnect structures, which have very low capacitance and inductance, to supply power and signals to the center of the wafer. Low resistance through-chip vias substantially improve this situation. The completed 2.1-in square chip stack must operate at 20 MHz.

As a step in the direction of a full-up WASP, we have fabricated a multichip module (MCM). This parallel string processor module, shown in Fig. 9, has only the routing, indium bump interconnect circuitry, and polysilicon resistors



	CASE 1	CASE 2	CASE 3
INDUCTANCE	15 nH	33 nH	11 nH
CAPACITANCE	160 fF	1.2 pF	5.6 pF
RESISTANCE	1 Ω	510 Ω	4 Ω

Fig. 8. Comparison of gross electrical parameters for three wafer stacking methods.

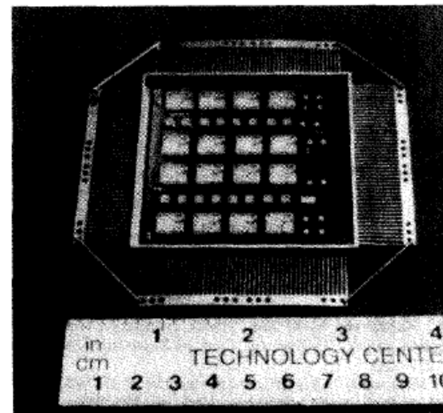


Fig. 9. Parallel string processor module.

of the full wafer scale integration device and replaces the upper and lower wafer with discrete dice. When the entire process is assembled later in this decade, the stack will be as illustrated as shown in Fig. 10. The WASP architecture is an ideal first attempt into this new technology application.

V. Conclusion

To summarize, the elements of a successful 3-D wafer scale integration product will be:

- fault-tolerant circuitry;
- total control of the silicon thickness and flatness;
- reliable multithousand bump interconnections.

The benefits of a successful 3-D wafer scale integration product will be:

- higher speed;
- lower power;
- lower production cost.

The principal market inhibitor will be the availability of proper design tools to establish a product which efficiently uses the shorter chip-to-chip interconnects to improve performance and lower costs.

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.

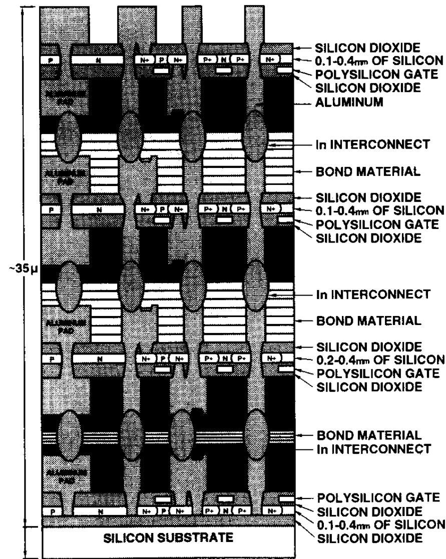


Fig. 10. 3-D SOI wafer scale IC's.

REFERENCES

[1] K. C. Keezer and V. K. Jain, "Design and evaluation of wafer scale clock distribution," in *Proc. Int. Conf. on Wafer-Scale Integration*, 1992, pp. 168–175.

[2] D. C. Keezer and V. K. Jain, "Clock distribution strategies for WSI: A critical survey," in *Proc. Int. Conf. on Wafer-Scale Integration*, 1991, pp. 277–283.

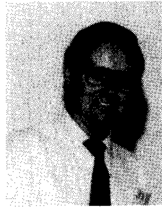
[3] J. F. McDonald, "An overview and analysis of 3D WSI," in *Proc. Int. Conf. on Wafer-Scale Integration*, 1991, pp. 223–235.

[4] M. W. Yung and M. J. Little, "A new approach to implement a defect tolerant power distribution network in WSI circuits," in *Proc. Int. Conf. on Wafer-Scale Integration*, 1991, pp. 215–222.

[5] M. J. Little and J. Grinberg, "The 3-D computer: An integrated stack of WSI wafers," in *Wafer Scale Integration.* E. Swartzlander, ed. Boston, MA: Kluwer Academic, 1989.

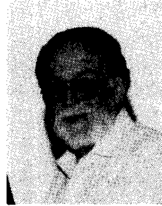
[6] I. P. Jalowiecki, S. J. Hedge, and R. Williams, "The development of the WASP 3 processor," in *Proc. Fifth Ann. IEEE Int. Conf. on Wafer Scale Integration*, 1993, pp. 20–29.

[7] R. M. Lea, "Comparing monolithic and hybrid WSI massively parallel processing modules for cost-effective real-time signal and data processing," in *Proc. Int. Conf. on Wafer Scale Integration*, 1991, pp. 199–206.



Ronald Williams received the B.S. degree in physics from the University of California at Los Angeles. He took part in graduate studies in metallurgical, electrical, and materials engineering.

He has several years experience in applied vacuum technology and crystal polishing and lapping. Currently, he is a Laboratory Scientist with the Engineering Laboratory, Hughes Aircraft Company Technology Center, where he has made major contributions to the Center's highly successful indium bump hybridization techniques. He has 22 years experience in vacuum deposition, component and process design, and the design and fabrication of specialized machinery, including six years experience as Manager of HTC's Assembly Laboratory. He currently provides technical leadership and management of groups responsible for production of precision quartz resonators, flexible printed wiring, and miniaturized electronic assemblies. He currently holds 5 patents.



Ogden Marsh received the B.S.E.E. degree from the University of Connecticut and the M.S.E.E. degree from the University of California at Los Angeles.

He is currently a Chief Scientist with the Hughes Aircraft Company Technology Center. He is presently investigating the applications of new materials and processes to device structures for flat-panel and projection displays, focal plane arrays, signal processing circuitry, high temperature electronic circuitry, and radiation hard nonvolatile memories. He is involved in developing bonded silicon wafer (BSW) technology for application to silicon-on-insulator (SOI) high speed and radiation hardened integrated circuits. He is also involved in the development of a new high frequency liquid crystal light valve (LCLV) and is heading a team developing this light valve for high density television projection systems. He is also involved in developing the silicon molecular beam epitaxy (MBE) process at the Hughes Research Laboratories and its applications to new device structures, such as Si/SiGe heterojunction bipolar transistors. In addition, he has managed infrared detector programs as well as several projects involving crystal growth and evaluation of semi-insulating GaAs and Si crystals for LWIR detectors and SOI materials for VHSIC. He has been involved in experimental research on space charge-limited current in semi-insulators and its applications. He was one of the first investigators of ion implantation in semiconductors and semi-insulators. He is currently a Visiting Associate with the Division of Engineering and Applied Science, California Institute of Technology. He currently holds 15 patents.

Mr. Marsh is a member of APS, AVS, the Electrochemcial Society, and MRS.

DOCKET ALARM Find authenticated court documents without watermarks at docketalarm.com.