NRL/FR/5560—92-9329

# A Survey of High Density Packaging for High Performance Computing System

NHI-ANH CHU AND HENRY DARDY

*Advanced Concepts Section*
*Signal Processing Branch*
*Information Technology Division*

April 10, 1992

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. AGENCY USE ONLY *(Leave blank)* | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | April 10, 1992 | |

**4. TITLE AND SUBTITLE**

A Survey of High Density Packaging Technologies for High Performance Computing Systems

**5. FUNDING NUMBERS**

PR-513579-0-1
WU-DN150-155

**6. AUTHOR(S)**

Nhi-Anh Chu and Henry D. Dardy

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

Naval Research Laboratory
Washington, DC 20375-5000

**8. PERFORMING ORGANIZATION REPORT NUMBER**

NRL/FR/5560—92-9329

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

Office of Naval Technology
Arlington, VA 22203

Naval Ocean Systems Center
San Diego, CA 92152-5000

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

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION/AVAILABILITY STATEMENT**

Approved for public release; distribution unlimited.

**12b. DISTRIBUTION CODE**

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

In this age of modern warfare when precision-smart weapons and avionics are in high demand, there is a need for a dramatic increase of throughput density, defined loosely as the computation throughput per unit volume/weight of the electronic system. To date, the main approach has been the monolithic wafer scale integration (MWSI) approach. Over the last decades, monolithic integration circuit (IC) density has increased by several orders of magnitudes, and the monolithic IC area by 10 times to approach the dimension of the semiconductor wafer. The growth trends for both circuit density and IC area have leveled off significantly.

Hybrid wafer scale integration (HWSI) overcomes many of these limitations. This approach includes techniques to put monolithic chips very close together while maintaining or improving inter-chip signal quality. Major portions of the report survey techniques, performance and availability of HWSI technologies. The survey finds that HWSI can provide from one to two orders of magnitude of increase to the throughput density of current systems. There is a major trend in the industry to invest in this technology, and it is expected to be available to the industrial as well as military systems at competitive prices within a few years.

| 14. SUBJECT TERMS | | | | 15. NUMBER OF PAGES |
|---|---|---|---|---|
| High density packaging | Monolithic | Hybrid | | 39 |
| High performance computing | WSI | | | 16. PRICE CODE |
| MCM | Waferscale | | | |

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

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std 239-18
298-102

## CONTENTS

# A SURVEY OF HIGH DENSITY PACKAGING FOR HIGH PERFORMANCE COMPUTING SYSTEM

## 1. INTRODUCTION

The Office of Naval Technology (ONT) is interested in the development of high-density packaging for high-performance computing. This report reviews advanced high-density packaging of electronic systems, introduces the key techniques and developments, and discusses the technological and industrial status.

The fundamental question that the report addresses is how much size reduction (miniaturization) and performance increase may be achieved for a given level of reliability and cost. The review of advanced packaging technologies will be applicable to all electronics systems. The special high performance computing systems will be digital computers that can deliver a floating point calculation rate in the giga ($10^9$) operations per second (OPS) today and tera ($10^{12}$) OPS in the future.

Section 2 identifies the broad structure of the elements of high density packaging technology. The fundamental limits of the existing interconnection technology are explained, and a quick review of the optical interconnection is given.

Sections 3, 4, and 5 review the monolithic wafer scale integration (WSI), two-dimensional (2-D) hybrid WSI, and three-dimensional (3-D) hybrid WSI respectively.

Section 6 discusses the process of incorporating high density packaging technologies into Navy systems and summarizes the major findings of the report.

## 2. OVERVIEW: TRADITIONAL PACKAGING VS ADVANCED PACKAGING

Traditionally, as Fig. 1 shows, an electronic system is structured as a hierarchy in which the electrical and functional connection are inseparable from the mechanical structure of each interconnection level [1]. In this traditional system packaging concept, the active circuitry exists in a 2-D form of a "chip." Chips are interconnected by metallic wires that are fabricated on boards at multiple levels. Boards are then connected into a backplane; multiple backplanes form a rack, and so on.

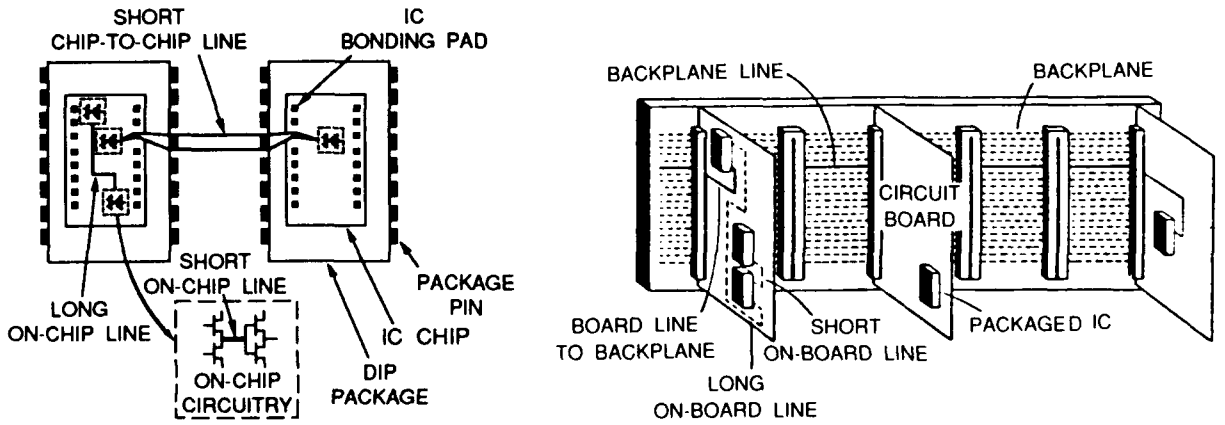


Fig. 1 — Traditional system packaging uses a hierarchy of interconnection structures that perform both the electrical conduction and mechanical support function. Courtesy of Tewksbury and Hornak [1].

Manuscript approved February 14, 1992.

As the microelectronic technology progresses, artificial limitations imposed by the traditional system packaging model begin to impact system performance. Therefore, a new advanced packaging concept is needed to further the trend of speed increase and of system miniaturization.

The circuit density has been doubling every year in the last two decades through the scale-down of semiconductor circuits. However, this trend will be coming to a saturation point within five years, if not sooner, because the device size is becoming smaller than fabrication equipment can deal with.

The growing trend in the electronics industry is to examine the scale-down at the system level; space and weight will be substantially reduced by eliminating material other than the active components. Within a chip package, the semiconductor active area is typically 25% of a chip footprint. The chip-footprint-to-board area ratio is typically 10% to 30%. Furthermore, space is also underutilized in the vertical direction. The silicon thickness is no more than 20 mils or so in a typical chip, and the boards that contain them are spaced at a typical pitch of 1 in.—only 10% of the vertical space is used.

Speed increase is realized by using faster transistors, but an increase in the speed of the system also requires improving interconnections throughout the required levels. In common electronics that operate at less than tens of megahertz (MHz), the logic gate delay is an order of magnitude larger than the transmission line propagation delay, which is typically 2 ns/ft. In high-speed systems that use the fastest logic, the aggregate gate delay is comparable to the interconnection propagation delay, which is proportional to the line length. This places a limit on the maximum line length and thus dictates the mechanical volume of the systems package. For example, in a traditional supercomputer such as the NEC SX-2, the clock cycle is 6 ns. The delay through ten 0.25-ns emitter coupled logic (ECL) gates amounts to a few nanoseconds, requiring the longest line to be less than a foot. The entire process logic is contained on a 21 by 18 in. board.

Increasing the clock speed toward the gigahertz (GHz) range, however, is not the only way to increase throughput in a processing system. An alternative is to assemble a large number of processors, each working at some relatively slow rate. In fact, in the next decade we will certainly see the flowering of massively parallel architectures that in theory can scale up linearly with each processing element added.

However, in this architecture, the amount of information exchanged among processing elements grows with the number of elements $N$ at least linearly in the simplest case, and possibly exponentially if each new element has to communicate with the others. Fundamental limits exist in the current packaging approach: the bandwidth of signalling on metallic conductors is limited by line dispersion and driver/receiver problems long before the theoretical limit of the speed of light, and the conductor density is also bounded at various physical connection levels. For example, the line densities at the chip, multichip module, and printed wiring board levels are 4000, 500, and 40 lines per cm respectively, based on realizable line pitch.

The multilevel packaging model with drastically different line densities has to be replaced by one where the required channel density does not decrease as $N$ increases. Metal has to be replaced by a different medium that carries signal at the speed close to the speed of light, independent of line length. Optical interconnection has the potential to fulfill these requirements. Because of the limited scope of the present work, this topic could only be reviewed superficially here.

Electro-optical communication for computers currently concentrates on three basic approaches.

• In computer free-space electro-optical communication, optical signals are transmitted by an array of transmitters or spatial light modulators through the air and intervening material to a receiver array. The optical signals may be redirected through micro-optics—tiny mirrors milled into a planar surface,

or rerouted by a holographic plane. The loss through a typical silicon wafer is 0.5 dB. [2]. Deformable mirror devices have been reported with drive voltages of 28 V and sizes of 15 by 15 $\mu$m that modulate at tens of MHz [3]. The main disadvantage of free-space optical communication is precise alignment. There are inherent reasons that limit this technique to very close range communication. For a decade, the U.S. Defense Advanced Research Program Agency (DARPA) has sponsored many research efforts on this topic at a level exceeding $5 million per year.

• Integrated wave-guides have the potential to replace metallic substrate interconnections by grooves fabricated on the surface of a semiconductor substrate or insulator such as polyimide. The principal advantages come from the noninterfering nature of optical signals. Crucial mechanisms for signal routing such as 40° micro-mirrors facing up or sideways have been demonstrated [4]. Like the free space, the main difficulty in using this technology is alignment. Waveguides beyond the surface of a small module have yet to be demonstrated.

• Fiber optics technology has matured markedly for long-distance communication applications. During the last 10 years we have seen fiber optics migrate into applications of progressively smaller physical radius as the costs of the fiber and interfacing logic have been reduced: metropolitan area network (MAN) and local area network (LAN). The bandwidth of fibers, in excess of a terabits per second (1e12 bps) is underutilized at the present time, while the gigahertz semiconductor technology, gallium arsenide (GaAs) is being developed to address interface functions such as conversion, and multiplex-demultiplex. Despite the relatively high bending radius, the fiber optics technology can be an ideal optical interconnection mechanism for massively parallel computers whose largest distance between pipelines is not scaled down. At present, a pair of 1-Gbps GaAs transmitter/receiver modules with 8:1 multiplexer-demultiplexer exist commercially [5] in 0.5 by 2.5-in. packages. For less than $5,000 per module, fiber optics is very close to being ready for computer communications. A Gbps fiber-optics link for the massively parallel Connection Machine has been demonstrated [6].

Figure 2 shows an advanced packaging concept that attempts to increase the use of space for active components and to overcome metallic wiring speed limitations. The traditional wires on a hierarchy of boards are replaced by a fiber-optics network. The structure of the interconnection closely matches the logical structure required by the architecture of the computer system. The 3-D grid structure shown here could very well be a hypercube or torus, for example. Chips are replaced by large, e.g., 3 × 3 in., 3-D blocks of active circuitry called wafer-scale modules. Conceptually, real estate in each 3-D module is dedicated to the network-access function. The wafer-scale module is the main topic of discussion of this report.

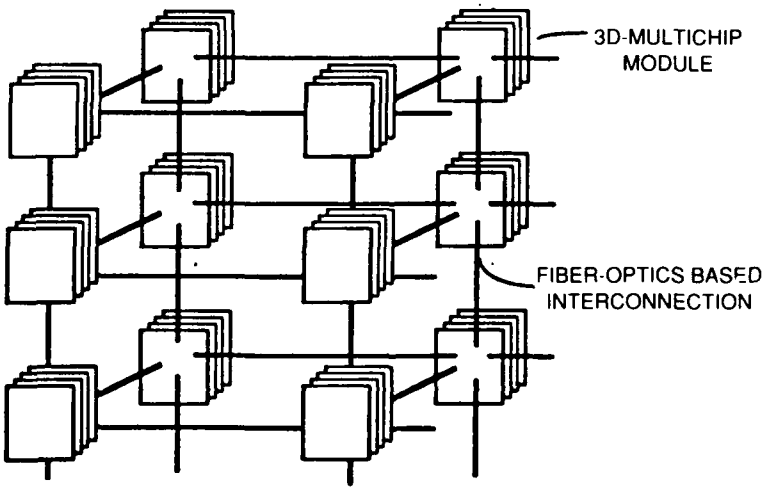


Fig. 2 — Advanced packaging concept based on 3-D active circuit modules and electro-optical interconnection

## 3. MONOLITHIC WAFER SCALE INTEGRATION

The term "wafer scale" relates to the size of a typical piece of material from which semiconductor devices are fabricated. Traditionally, silicon is fabricated from sand into a container called the ingot. The finished silicon ingot is sliced into thin, round pieces called wafers. Currently, most microelectronics manufacturers are set up for 4-in.-diameter wafers and are moving into 6-in diameters. The term is also used to mean monolithic WSI as opposed to hybrid WSI. Monolithic WSI is the fabrication of circuits from a whole wafer; hybrid WSI is the integration of prefabricated circuits onto a substrate as large as a wafer.

In theory, monolithic WSI is a simple extension of the very large scale integration (VLSI) technology, from the traditional size of a few square mils to wafer scale. We review the status of VLSI to lead to discussions of special considerations when extending this technology to wafer scale.

### Review of VLSI Technologies

Electronic devices are commonly fabricated from silicon-related material on a silicon substrate. Recently gallium arsenide (GaAs) has found its way into the semiconductor industry as a new material for very high speed logic and a unique substrate that can integrate electro-optical devices.

The key factor for the popularity of silicon as semiconductor material is its ability to grow a chemically stable protective layer of silicon dioxide ($SiO_2$) on the silicon substrate. Despite its widespread use, the cost of a silicon wafer is rather high. This is because of the difficulty in producing a wafer to a specification of resistivity, conductivity type, dopant, and a general description of orientation and quality of surface finish. The wafer's size dictates a specific thickness to avoid warpage—e.g., 0.550 mm for a 5-in. wafer. The wafer is polished to a fine degree then cleansed thoroughly through mechanical and chemical processes to rid of contaminants such as organic or inorganic debris (including water and salt).

Silicon-based transistors and capacitors form the basis for the analog and digital technologies that include bipolar, n-channel metal oxide semiconductor (nMOS), p-channel MOS, complementary MOS (CMOS) and ECL. Each of these technologies has its own relative advantages and disadvantages, even though speed generally influences the selection of one over another. Figure 3 shows the gate delay vs load capacitance for 3-input, 3-output gates in different technologies. Each gate is of a representative size found in the newest and densest gate array of the respective technology. The propagation delay are derived by using Barna's formulas [7].

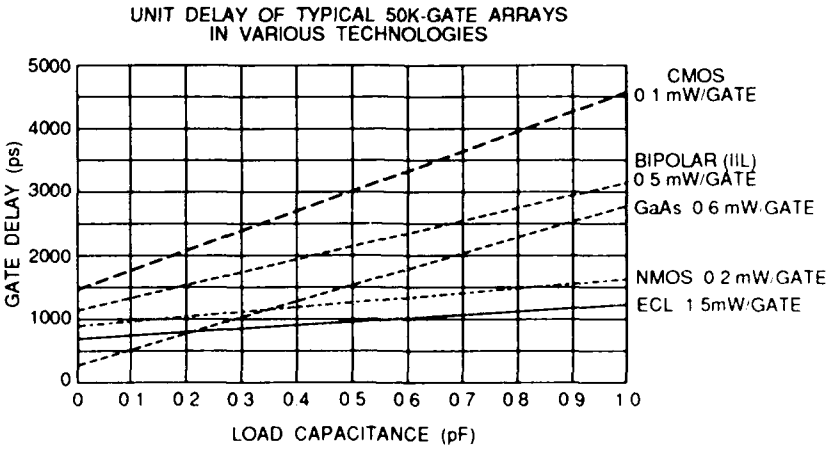


Fig. 3 — Gate delay vs load capacitance for high-performance and high-density gate-array chips in various technologies. Each gate is assumed to have a fan-in of 3 and a fan-out of 3. For CMOS, a clock duty ratio of 10 is assumed.

Based on p-n-p and n-p-n transistors, bipolar technology has the highest current drive capability, and consequently it is used to implement high power drivers and receivers. The metal-oxide semiconductor (MOS) technologies are currently more popular because of the smaller transistor size, versatility in circuit design, and self-alignment capability. Because of the higher mobility of electrons relative to holes, nMOS is faster and more area efficient than p-channel, and for the same reason faster and more compact than complementary MOS. On the other hand, CMOS is the most power-efficient because each n- and p-transistor pair forms a very high resistance path from power to ground when not switching. This efficiency becomes less significant at higher speeds where dynamic power consumption dominates. Recently, new circuit techniques such as Domino logic and single-phase (NORA) clocking methods [8] were invented. These technologies take advantage of the complementary nature of CMOS to allow efficient synchronous circuits that operate at very high speed with a single-phase clock, rather than multiple-phase clocks that require extremely sharp and precise clocks. Another reason for CMOS' continued dominance in digital logic below 50 MHz is the recent trend to modify the CMOS process to allow the integration of bipolar drivers and receivers. This is the so-called BiCMOS process. As an indicator of the technology, we must consider the Oki CMOS gate array that contains 92 thousand useable gates with gate delay between 200 and 400 ps.

For speeds in the hundreds of MHz to 1 GHz, ECL has been dominant. High speed operation is possible mainly because the n-p-n and p-n-p transistors are deployed in the linear gain region. In addition, the voltage swing is an order of magnitude smaller than in saturation-mode logic. The Motorola macro cell array MCA3 is one of the most advanced high speed ECL gate arrays; it has 50 thousand useable gates, with 90-ps delay. National Semiconductor's newest entry contains 120 thousand ECL gates (including embedded RAMs) [9].

*Gallium Arsenide (GaAs)*

A transistor based on this semiconductor compound is inherently faster than the silicon counterpart because of the higher electron mobility. For a depletion-mode GaAs MOSFET, the process characteristic constant that depends strongly on mobility is more than three times the constant for a silicon depletion mode MOSFET. In addition the drain-gate capacitances are lower than those in silicon.

Because of the gigahertz operating frequency capability, GaAs is appropriate for radio frequency (RF) monolithic microwave integrated circuits (MMIC). Even more significantly, ternary or quaternary compounds based on GaAs are being used to construct laser sources, photodetectors, and solar cells [10]. These capabilities have been identified by workers for 40 years but only recent breakthroughs have brought the technology to the marketplace. The maturity of GaAs for digital application may be indicated by a GaAs gate array made by Vitesse, which incorporates 30 thousand gates. The unloaded gate delay is 90 ps at 0.3 mW per gate.

**Special Considerations to Extend VLSI to MWSI**

Yield — the extension from the dimensions of the largest VLSI chips ($1 \times 1$ cm) to wafer-scale devices-has proved to be more than a simple extension of the same design and fabrication process. In practice, success has been limited; it remains very difficult to fabricate a large circuit because of contaminations that occur during the wafer processing and handling. Pinholes in the photoresist, mask defects, contamination from various sources (including wafer breakage), and defects in the grown epitaxial layers contribute to circuit failures, especially if failure is associated with the active region of the device. The yield trend for MWSI may be examined statistically by using the well known Stapper's yield model. The yield $Y$ is a function of the defect density $D$, the chip area $A$, and a parameter $\alpha$ between 0.5 and 5 fitted to empirical yield data:

$$Y = 1/(1 + DA/\alpha)^{\alpha} \quad (1)$$

To illustrate the yield trend when the area increases from that of a VLSI chip to a MWSI module, Fig. 4 shows the optimistic case of $D = 1$ defect/cm$^2$ and a wafer area of 20 cm$^2$. The indication is clear: for a variety of yield parameters $\alpha$, the chip yield goes down sharply as the chip area approaches the wafer area (shown by arrows) because of the inclusion of the necessary clusters.

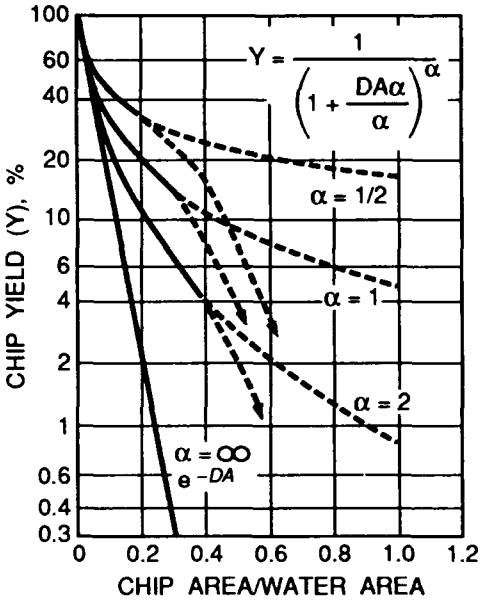


Fig. 4 — Expected drop in yield as chip area approaches wafer area (shown by arrows)

*Redundancy to Improve Effective Yield*

While fabrication technology continues to make marginal progress, workers have adapted various yield management strategies to make best use of the imperfect fabrication process. The general approach is to partition the circuit on a 2-D wafer into small areas (cells) such that any one cell with detectable defects may be isolated from the rest. The isolation of the defective cell may be performed physically by cutting links from the cell to the neighbors or by destroying prefabricated electronic fuses or antifuses that make up the links. Alternatively, the links may be controlled by switches that are opened or closed under program control.

Illustrative of the physical configurability techniques are the vertical laser link and laser diffused link invented at MIT Lincoln Laboratory under the Restructurable VLSI program [11]. The vertical laser link uses a silicon-rich silicon nitride (Si Nx) as a link insulator rather than the traditional silicon oxide that proved to be too fragile. During the linking process, an argon laser pulse of durations ranging from 0.084 to 1 ms and power between 1.2 and 2.5 W is focused to a 4.2 $\mu$m full-width half-maximum spot. This results in melting the second metal, removing the nitrogen from the Si Nx while melting the remaining silicon, and melting part of the first metal, thus creating a silicon aluminum alloy. The typical parameters of the link are an on-state resistance in the 1 to 10 $\Omega$ range, an off-state of about $10^{14}$ $\Omega$, and a very high reliability rate for successful linking. The laser diffused link can be produced in a standard MOS process. It consists of two diodes formed by implantation into a well or substrate, isolated by a gap of 1 to 4 $\mu$m. A laser pulse is used to melt the silicon in the gap and on the edges of the diodes; the

dopant from the diodes undergoes liquid phase diffusion across the gap, forming an ohmic connection, with resistance on the order of 100 Ω.

The Xilinx programmable gate array is an example of dynamic configurability technique [12]. It consists of a transistor-based switch controlled by a configuration logic bit stored in an array dynamic memory which is in turn loaded on power-up from EEPROM on the chip.

*Special Considerations For Long Distance Interconnects and Power Distribution*

Another fundamental problem exists if the VLSI process is used without modification for WSI. Interconnect lines of the typical VLSI geometry exhibit excessive resistance across the large distance on a wafer. Thicker metal interconnection layers and thicker dielectric layers are required to make lossy transmission lines with low impedance.

The theoretical upper and lower limits of digital delays of a lossy transmission line (RLC) are estimated by Stopper in Ref. 2, and shown in Fig. 5.

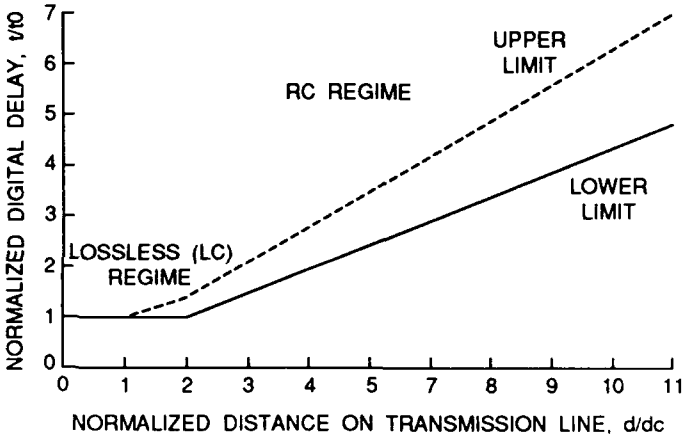


Fig. 5 — Estimated theoretical bounds of digital delay on a transmission line. Delay is normalized to the minimum delay achieved at the speed of light. Distance on the line is normalized to the critical distance $d_c$ of the transmission line. A line shorter than $d_c$ exhibits lossless transmission line delay—proper termination is required. Others operate in the lossy regime where the delay is that of an RC line.

The critical distance $d_c$ of a transmission line, as used in Fig. 5, is a characteristic of the line, which is determined by the geometry and material of the construction. Table 1 shows $d_c$ values of typical microstrip lines (a long conductor over a ground plane) used in VLSI chips, thin- and thick-film multichip modules, and printed wiring board. The formula derived in Ref. 13 for $d_c$ is

$$d_c = 2 \ln 2 \, Z_0/R_0 \tag{2}$$

where $Z_0$ is the characteristic impedance and $R_0$ is the resistance per unit length of the conductor. The characteristic impedance is calculated from the ideal (minimum) delay $t_0$ and line capacitance as follows:

$$Z_0 = t_0/C, \tag{3}$$

$$t_0 = (\sqrt{\epsilon_r}) \, d/c_0, \tag{4}$$

$$C = (\epsilon_r \epsilon_0/h) \, (d + 1.6h)(w + 1.6h) \tag{5}$$

where

$\epsilon_r$ is the dielectric constant of the insulator material,
$\epsilon_0$ is the dielectric constant of air,
$c_0$ is the speed of light,
$h$ is the insulator thickness,
$d$ is the conductor length, and
$w$ is the conductor width.

Note that the transmission line characteristic impedance and critical distance are very slightly dependent on the line length $d$. This dependence is introduced in the calculation of the microstrip capacitance, which is attributed to Barna [8].

Table 1 — Transmission Line Characteristics of Microstrip Lines in Some Typical Constructions

| | VLSI | Thin Film | | Thick Film | PWB |
|---|---|---|---|---|---|
| | | 1 | 2 | | |
| Conductor resistivity ($\mu$Ohm-cm) (aluminum/copper) | 2.6 | 2.6 | 1.7 | 1.7 | 1.7 |
| Dielectric relative constant er ($SiO_2$/polyimide) | 3.9 | 3.5 | 3.5 | 3.5 | 3.5 |
| Typical line length (cm) | 1.0 | 10.0 | 10.0 | 10.0 | 25.0 |
| Conductor width ($\mu$m) | 4.0 | 25.0 | 25.0 | 100.0 | 100.0 |
| Conductor thickness ($\mu$m) | 1.0 | 3.0 | 5.0 | 20.0 | 20.0 |
| Dielectric thickness $\mu$m | 1.0 | 2.5 | 10.0 | 20.0 | 20.0 |
| Conductor resistance ($\Omega$) | 65.0 | 34.7 | 13.6 | 0.9 | 2.1 |
| Capacitance (microstrip) $C$ (pF) | 1.9 | 35.9 | 12.7 | 20.5 | 51.1 |
| Minimum delay to = sqrt(er) 1/c (ns) | 0.1 | 0.6 | 0.6 | 0.6 | 1.6 |
| Characteristic impedance $Z_0 = t_0/C$ ($\Omega$) | 34.1 | 17.4 | 49.1 | 30.5 | 30.5 |
| Critical RLC line distance $d_c$ (cm) | 0.7 | 6.9 | 50.0 | 497.3 | 497.4 |

As line lengths increase beyond a 1 cm or so, it is clear from Table 1 and Fig. 5 that the construction of conductor lines in VLSI circuits is not sufficient. Table 5 suggests a minimum delay may be obtained for a large circuit at the cost of increased conductor and insulator geometries.

Traditional VLSI techniques distribute power through a grid or H-tree structure by using the same two or three metal layers on which signals are distributed. This is not adequate for two reasons. For a high speed logic WSI circuit it is necessary to have a continuous ground plane for metal lines to behave as good transmission lines. A power plane and ground plane separated by a very thin insulator of high dielectric constant will also be required to form a large capacitor to filter out noise on the supply lines. This noise can be very strong because of the simultaneous drawing of current at clock edges in a synchronous circuit. For memory devices, the power plane requirement is not severe because of the small power consumption and exclusive-use nature of memory cells.

## Review of Some Recent Significant MWSI Efforts

*Memory*

The most prominent and effective implementation of monolithic wafer scale integration is seen in memory devices that are extremely regular and require very simple interconnection structures. In general, higher yield for memory has been noted because the fabrication processes for memory are sufficiently fine-tuned through incremental improvement iteration cycles. For wafer scale devices, even high yield memory processes make use of the special techniques described above.

An example is the INOVA 1Mbit SRAM which achieves 80% effective use of silicon through the use of a laser-fuse technique to disable bad or unnecessary blocks. A 1Mbit die is fabricated as a total of 40 32K-bit blocks, 32 of which are selected [14]. SGS-Thomson of Europe published reports on a 4.5 Mbit SRAM design, which is based on 64Kbit blocks, laser fuses, and antifuses. 20-$\mu$m thick metal is not used for signal interconnection but for power supply routing. The 6-in. wafer was fabricated in 1988, but no information is yet available regarding the outcome [15].

*Logic*

The Lincoln Laboratory RVLSI program has demonstrated the technology since 1983 with six working wafer scale digital signal processing circuits. On average, each generates about 1 to 3 W, contains about 390 thousand CMOS transistors, clocks at 6 to 10 MHz (except for the small digitizer that runs at 25 MHz). Copies of a single wafer type were restructured to make four wafers: the fast Fourier transform (FFT), constant false alarm rate filter, Hough transform and two-dimensional convolution. Several other larger RVLSI designs are under way: the single-instrumentation-multiple-data (SIMD) image processor and the MUSE linear systolic array for beamforming. Each takes up approximately 3 million transistors and are being designed for 5-in. wafers.

The penalty for silicon underutilization incurred by the restructuring approach may be observed by comparing some RVLSI wafers with the VLSI iWarp chip. Even after accounting for feature size, which is 3$\mu$m on the latest RVLSI vs 1$\mu$m for the iWarp, it is interesting to see that the RVLSI wafer area is 25 times larger but contains half as many transistors, while the RVLSI clock speed is four times less than the iWarp 40MHz. These low clock rates seem to confirm the lossy transmission line problem.

Another interesting work is the U.K.WSI Associative String Processor (WASP) 2b version developed by Aspex Microsystems and fabricated by Plessey in Roborough in the first quarter of 1990 [16]. This is a 6-in. wafer with device area of 9.1 by 9.8 cm. It contains 8.43 million devices. The architecture is fine-grained SIMD massively parallel; the WASP2b contains 6480 processing-elements. The goal of a 40 MHz internal clock appears ambitious considering that no special attention is indicated for wafer length transmission or ground planes.

Figure 6 shows these and other recent WSI circuits organized according to year of announcement and circuit size. The Intel 80386 and iWarp VLSI chips are also shown for comparison.

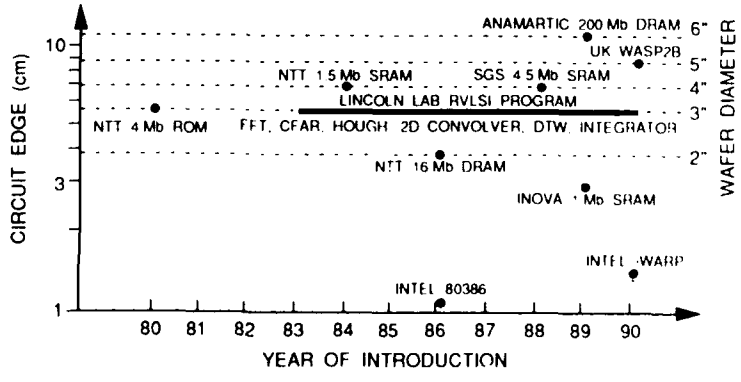


Fig. 6 — Circuit size of some monolithic WSI circuits

## CONCLUSION

VLSI technology is making incremental steps in circuit size, not the great leap forward to monolithic wafer scale integration. Only very regular structures such as memory devices seem to be successfully integrated at the wafer level.

It is interesting to observe that while special techniques to implement wafer-scale transmission lines and quiet ground planes have been identified and applied universally in hybrid wafer scale integration, none of the monolithic wafer scale integration (MWSI) projects reviewed have actually used these techniques. No doubt this is responsible, at least partially, for limiting success to low speed and low power circuits.

Another observation is that the market trend should be followed to take advantage of the availability of high density application-specific integrated circuits (ASIC). This should be a first step toward high-density packaging at the active component level. The combination of high-density mixed-logic ASIC and hybrid WSI promises to achieve the goals of advanced packaging.

## 4. TWO-DIMENSIONAL HYBRID WAFER SCALE INTEGRATION (2-D HWSI)

The term HWSI is commonly used to describe the integration of prefabricated components into a substrate as large as a wafer. Two important characteristics differentiate HWSI from conventional printed wiring board: the components in die form are inserted directly to the substrate without the traditional chip package, and the interconnection density on the substrate is one order of magnitude higher. Figure 7 shows the multichip module (MCM) technology relative to printed circuit board (PCB) and VLSI technologies in terms of interconnection density [17]. A HWSI module is commonly identified with the MCM that has been around for more than 20 years; these two terms are used interchangeably in this report.

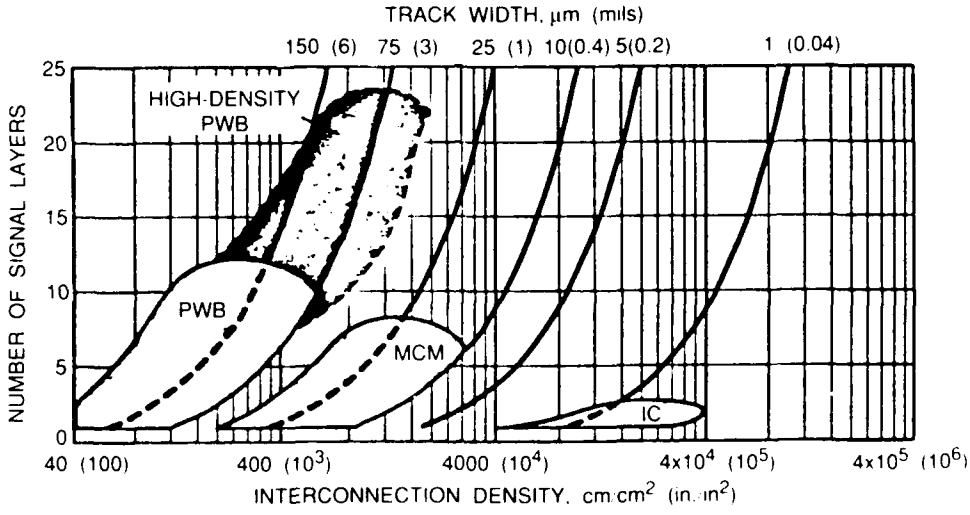


Fig. 7 — MCM technology relative to PCB and VLSI in terms of interconnection density. Courtesy of Sage et al. [17].

Depending on component sizes and MCM implementation techniques, the MCM may offer between 4 to 10 times silicon-to-board area ratio compared to a conventional printed wiring board. Furthermore, multiple MCMs may be stacked to form a 3-D module (3-D MCM). In comparison to the typical 1 to 2-in. board pitch in current systems, the thickness of each MCM layer is as little as a tenth of an inch

(the combined thickness of the support substrate, the circuit die, and the interconnecting layers). The 3-D MCM may offer 10 to 20 times more circuit density in the vertical direction. The combined volume packing is then between 40 to 1 and 200 to 1.

## Overview of 2-D HWSI Modules

Two main families of MCM are classified by the thickness of the metal traces that can be deposited on the substrate. The older thick-film family uses screen printing to make lines with a typical thickness of 20 $\mu$m. The current generation uses advanced techniques and metallurgically pure conductors, typically 5 $\mu$m thick.

## Thick-film MCM

Historically and technically, thick-film MCM technology may be considered as an extension of printed-wiring-board (PWB) technology to meet stringent mechanical and environment requirements. These requirements are beyond the capability of the plastic substrate commonly used in the PWBs. The thick-film process may be defined as the sequential printing and firing of conductor and dielectric paste formulation onto a substrate.

Fundamental reasons limit the high-speed signalling of thick film to below 50 MHz. At gigahertz frequencies, thick microstrip lines are severely affected by the skin-effect, which significantly increases the line resistance [18]. In addition, because of the method of construction of microstrip lines in thick film, the thickness of the conductor and of the dielectric planes are not uniform and therefore do not maintain a high quality transmission line.

Features as small as 4-mil line width and 8-mil line pitch are possible with a state-of-the-art process, while the typical line width and pitch are 8 mil and 16 mil respectively. Even at this resolution the pitch is two to four times that of typical I/O pads in a VLSI die. A fan-out pattern that is consequently required to adapt the chip to the thick film MCM takes up a significant footprint. For advanced complex modules that require a large number of interconnection nets, many layers are required. Each added layer gives a diminishing return because of the routing complexity problem compounded by the large area required to construct the staggered vias to interconnect the layers.

### *Construction*

The selective deposition of thick-film materials onto a substrate is done by forcing the material in ink or paste form through a fine metal screen. To form the pattern on the screen, an ultraviolet (UV)-sensitive emulsion fills the screen. Each aperture of the screen is then opened or closed when the screen is exposed to a UV source through a photographic mask containing the circuit pattern. Figure 8 illustrates the screen printing process [19]. A metallic powder, usually gold, is mixed with a binder generally composed of glass powders and oxides, and a "vehicle" typically containing a solvent and a resin. After the paste is screened on, the module is oven dried to remove solvents. Firing at high temperatures (500 to 1000°C) is done to burn off the solvent and other nonmetallic ingredients in the paste. As the temperature rises to the peak, metal particles begin to sinter, glass frits begin to soften and flow, and a film is gradually formed.

For each layer, the hour-long process of printing, drying, and firing is repeated until all layers are formed. Each layer requires as many as four firings in the most conservative process. Some steps may be eliminated by cofiring. This is a process in which two thick-film compositions are printed and dried, one after the other, and then fired at the same time.

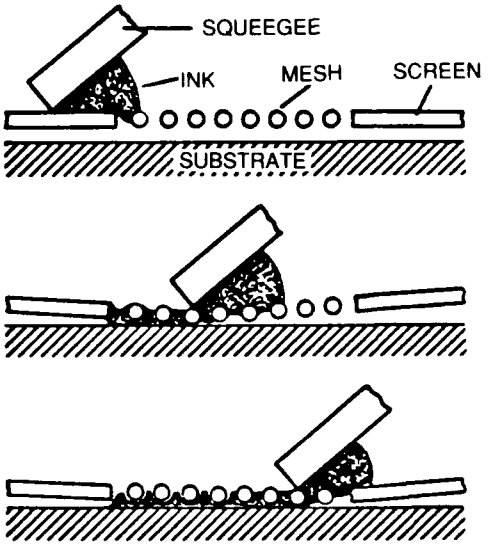


Fig. 8 — The screen printing process used in PWB and thick-film MCM. Courtesy of Borland [19].

The distinct disadvantages of such a high temperature and multiple firing and printing cycles have recently pushed the technology toward a new process called the low-temperature (about 800°C) cofireable ceramic process. In the most advanced variation of this process, a ceramic tape with punched vias is laminated onto an alumina substrate before firing. The ceramic dielectric tape is of a glass-based compound that allows a lower sintering temperature, permitting the use of high-conductivity air-fired precious metal or nitrogen-fired copper systems.

## Thin Film MCM

Currently two main approaches are used for thin-film MCMs; the so-called conventional and the overlay approaches. Many variations exist in each approach in terms of choice of material, material deposition techniques, via constructions, mask vs laser-based direct write, prefabricated/programmable interconnection, etc.

### *Conventional MCM.*

The basic construction includes a base substrate onto which multiple layers of metal conductors are pattern-deposited with intervening layers of an insulating dielectric material such as polyimide to make up the interconnections of the MCM. For large volume productions, a set of lithographic "hard" masks are fabricated first. A mask consists of a large black and white layout of the desired pattern of openings that is made and then reproduced photographically and finally transferred into a metal plate. For smaller quantity productions, a computer-driven laser beam system can be used to directly draw the desired patterns on photosensitive resist layers.

Although the construction of thin-film MCM is very similar to the VLSI fabrication process, a much better yield can be reasonably expected. First, the difficult steps involved in diffusion and ion implantation are not needed since no active transistors are involved, only metal interconnections. Second, calculation of the yield for wafer scale interconnections by using the Stapper models for particulate contaminations strongly suggests that high yields are possible.

The layers are electrically interconnected by vias that are typically etched at each sequential layer. Figure 9 shows the profile of the IBM VHSIC chips on silicon (VCOS) MCM with a solder bump at the top surface and four layers of metal interconnected by staggered vias. In this particular construction, two of the layers are for the power and ground distribution.

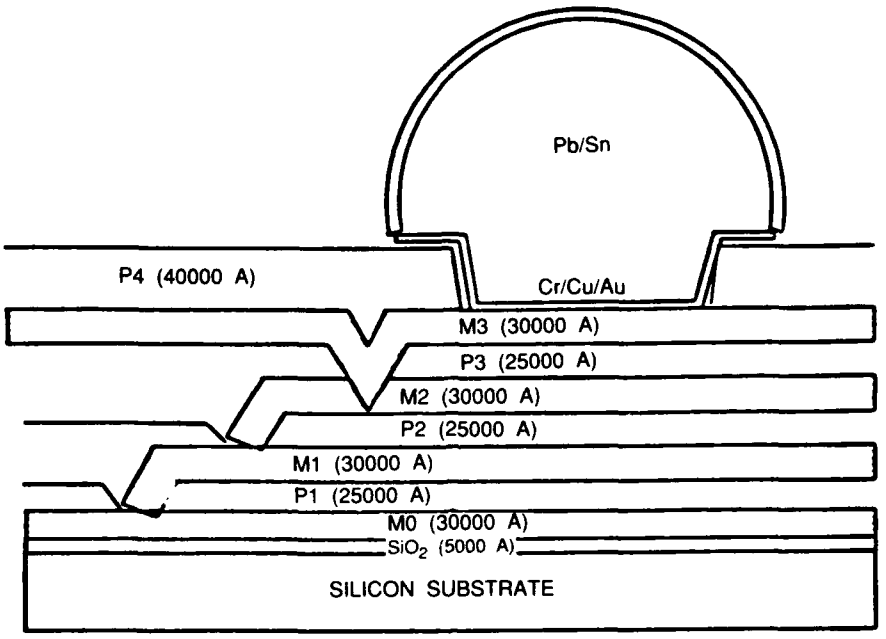


Fig. 9 — Profile of IBM VCOS MCM with flip-chip (C4) interconnection

The second column of Table 1 shows the critical distance for transmission line performance for the conductor of this construction to be 6.9 cm. This is a rather small number for a 3-in. wafer, but it is consistent with IBM's claim of performance in the 50 to 100 MHz range, rather than in the GHz range. Similar profiles are reported for the Hughes and AT&T MCM constructions.

The MCM designed by the Microelectronics and Computer Technology Corporation (MCC) uses 5-$\mu$m thick copper and a 15-$\mu$m polyimide with dielectric constant 3.5. The third column of Table 1 shows this critical distance to be 50 cm. The penalty for increased polyimide thickness is in thermal resistance because this material has thermal conductivity that is 2 orders of magnitude lower than metals that are used in the other layers of the MCM. MCC tests have confirmed that rise times as small as 200 ps can be supported over distances of several centimeters.

The MCC construction uses pillars, as shown in Fig. 10 [20], that achieve better interlayer interconnection than are possible with the standard vias shown above. The pillars may be stacked in several layers to form space-efficient vertical interconnections for electrical as well as thermal paths from the die to the substrate. Each copper pillar is constructed as a cylinder on top of another larger one for structural support; it is coated with nickel to protect against corrosion. The dielectric insulator—usually polyimide—is then spin-coated on top. The surface is then mechanically polished to very fine flatness and to expose the pillars. The pillar diameter is typically 15 $\mu$m. In this MCC construction, two copper planes are dedicated for power and ground.

After the interconnections are fabricated and tested, bare chips are mounted on the finished substrate by a variety of methods. If mounted face-up, wire bond or TAB may be used to make the electrical interconnections. Alternatively, the chip may be placed face-down in the so-called flip-chip method. In this method solder bumps are placed on the I/O pads of the chip and may be melted at a controlled degree to bond to the corresponding pads on the MCM. The solder bump can be used to connect to a flip-chip or to TAB fingers.

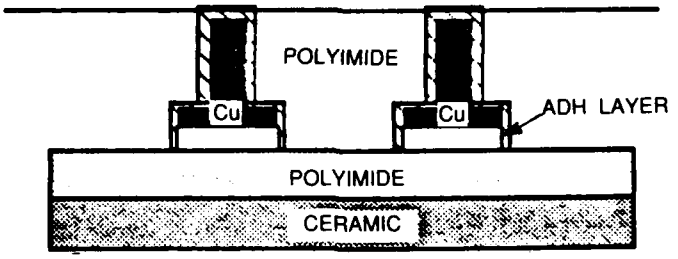


Fig. 10 — Profile of MCC nickel-plated copper pillar on a ceramic/polyimide substrate. Courtesy of Herrell and Hashemi [20].

*Programmable Silicon Circuit Board (PSCB) and Quick Turn Around Interconnect (QTAI)*

Multichip module substrates with prefabricated interconnections accelerate the integration of circuit dies into high-speed thin film modules. The Mosaic, Inc., PSCB is a thin-film MCM of the conventional construction type with a silicon substrate and several layers of metal lines that exist in segments in a variety of lengths and patterns. Figure 11 shows a cut-away illustration of a PSCB. Line segments can be connected by electrically shorting the prefabricated antifuses. The multitude of pads connected to each line segment bond the die I/Os and test the substrate before and after die attach. The main advantage of PSCB is the quick turn around time from design to test, which is between 2 and 4 weeks at the present time.

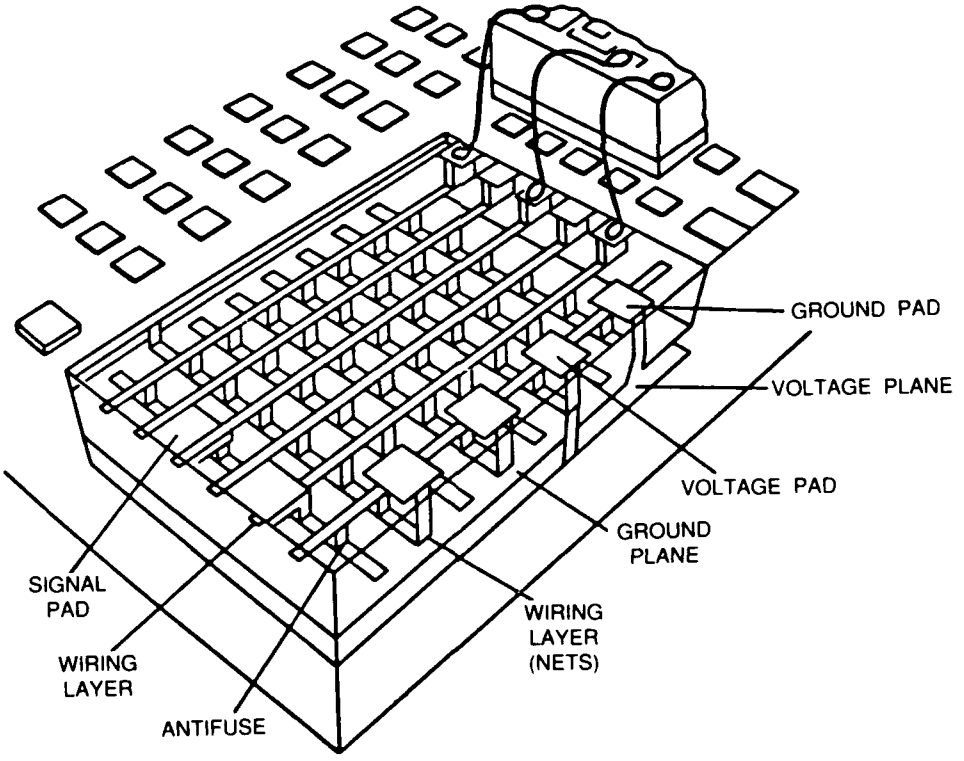


Fig. 11 — Mosaic system's PSCB

The antifuse uses amorphous silicon sandwiched between electrodes as the initial insulator. This is then broken down when a programming signal is applied and crosses certain threshold voltage. The Mosaic antifuse occupies only 3 $\mu$m and has an on-resistance of approximately 4 $\Omega$.

To make the so-called QTAI MCM, MCC fabricates copper/polyimide MCM substrates as "blanks" up to 70% completion. These blanks are then used for customizing particular circuit applications. The

personalization process involves metalization to connect various prefabricate signal segments and the metalization of the pads.

*Overlay MCM.*

In this construction, bare chips (dies) are mounted on a substrate before multiple layers of insulators and conductors are fabricated on top to make the interconnections. Die mounting is achieved by two methods. One method involves placing the dies face-up in cavities that are milled into a ceramic substrate at precise depths. The second method involves glueing them on a base plate then encapsulating the whole thing with an appropriate material that can be opened with vias to access the chip I/O pads. The I/O pads of the chips are flat on a planar surface covering the whole area. Advantages include: no separate steps are needed to make the bonding from the chip I/O pads to the MCM interconnections, and the areas over the dies are available for routing.

Figure 12 shows a typical profile of the overlay MCM approach used by General Electric (GE). Conductor and insulator thicknesses are similar to the MCC profile, enabling the MCM to perform at high speed on a 4 × 4-in. substrate.

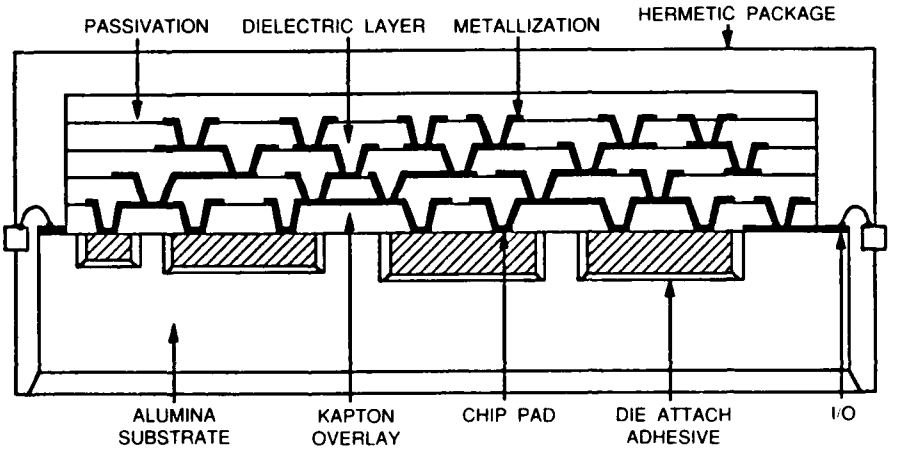


Fig. 12 — Profile of the GE's overlay MCM.
Courtesy of GE [36].

In this particular construction, the first level of overlay is a sheet of plastic that is glued on the substrate by an adhesive. This adhesive is a thermoplastic material—one that can be re-melted at some relatively low temperature—enabling the plastic sheet and everything above it to be easily removed. This capability is important because the chips lay underneath the interconnections. Another crucial difference between this and the conventional construction is the accurate placement of the dies on the substrate that is required. In the conventional design, the dies are attached one at a time to the MCM substrate at the I/O pads. This task is easy because only one die needs be positioned accurately relative to the substrate. In the overlay design, an interconnection line must be drawn to align with all the appropriate vias to the I/O of the dies in one step.

An alternative to placing the dies in cavities is to encapsulate the dies with an appropriate material, then form a planar surface on top of the covered dies. This approach is being pursued by Integrated Systems Assembly (ISA). Figure 13 shows the typical construction. The basic idea was also briefly described by a Sharp Electronics Corp. engineer in Ref. 21.

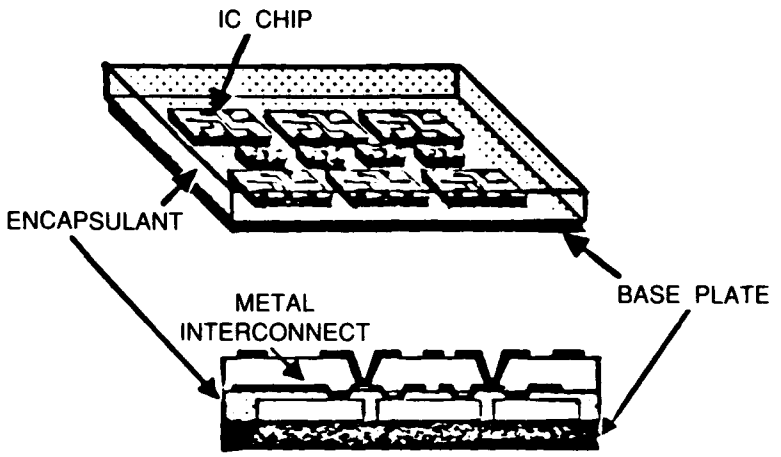


Fig. 13 — Overlay MCM with encapsulated dies.
Courtesy of ISA, Inc. [39].

**Status of Thick and Thin Film MCM Technologies**

The price for thick film MCM is projected to be about $25 per square-inch in 1992; for thin film MCM it is projected to be $75 per square-inch. It should be noted that the higher density of thin film means that a smaller area is required, so that the module price is equivalent.

At the time of this report, for a small quantity the price for each 4-in. MCM is about $20K to $40K excluding the price of the dies, with a nonrecurring cost of about $300K to $900K, depending on the quality of testing required.

Because it has been established for more than a decade, the thick film ceramic hybrid industry is rather mature. It has served primarily the military applications that require a better construction than the PWB and that can afford the very high cost of the low yield associated with the high-temperature, sequential firing process.

Other activities include:

• The Department of Defense has established a process to qualify manufacturers of thick film MCMs. The list includes almost 30 manufacturers among which are Boeing Electronics, Hewlett Packard, Honeywell, Hughes Aircraft, ITT, National Semiconductors, Raytheon, Teledyne, and many other smaller companies. Most of these firms are capable of producing 2 × 2-in.-thick film ceramic multichip modules.

• General Dynamics Hybrid Microelectronic Assembly facilities in California are certified to produce thick film MCMs at the maximum rate of 5,000 modules per week.

• Alcoa Electronic Packaging (AEP) in San Diego is gearing up a cofired ceramic thick-film MCM facility in San Diego and thin-film silicon on silicon MCM in New Jersey with design-through-acceptance-test turnaround time of 6 months in 1990, 3 months in 1991, and 10 months in 1992. AEP is the sole licensee of the advanced VLSI packaging technology (AVP) developed by AT&T.

• IBM's VHSIC chip on silicon (VCOS) and controlled collapse chip connect (C4) technology has qualified a 1.25 × 1.35-in. VCOS MCM line with 196 pins. The Manassas Virginia, plant produces more than 200 VCOS MCMs a month. Another line at this plant is currently allotted to process 50 5-in.

MCM wafers per day. Currently, a 308 I/O pin generic VCOS 2 × 2-in. MCM package is being qualified.

• At Hughes Aircraft, the two-year old HDMI-1 process can now produce 2 × 4-in. silicon or ceramic 5-conductor layer thin-film MCMs at the rate of 500 modules per year.

• The copper-pillar MCM technology was developed by MCC. In the last few years this and other related packaging technologies have been transferred to the share holders and partners, which number about 20 companies. Currently, the MCM program remains in the prototype status at MCC, while some shareholders have been trying to move into production either for in-house systems or as a commercial line.

• The latest Digital Equipment Corporation (DEC) mainframe computer, the VAX 9000 was produced by using 4 × 4 in. copper/polyimide MCMs on which MCA III ECL macrocell array chips are TAB mounted [22].

• Polycon of Ventura, California, is now opening a new facility in Arizona dedicated to thin-film MCM.

• Mosaic, Inc., of Belmont, California, is a commercial supplier of the programmable silicon circuit board (PSCB). The company has licensed this technology to ERIM of Michigan that is using the PSCB for various Air Force and Army contracts. The largest PSCB measures 1.2 × 1.2 in..

• The board-on-chip style is known in the industry as the overlay or GE-approach because it was invented by Chuck Echelburger at GE in 1985. Because of its high cost, so far only small prototypes have been produced for DoD-sponsored research. In September 1990, DARPA gave a multimillion-dollar contract to GE and its partner Texas Instruments (TI) to develop this MCM technique into a low-volume production merchant foundry for all DoD digital MCMs.

• Integrated Systems Assembly, a GE-spinoff company, was formed in April 1990 by Echelburger and a GE manager to set up a specialized commercial packaging company. The facility's start-up and operation are expected to begin in 1991. ISA has proposed and partially demonstrated a board-on-chip MCM technology generally similar to that of GE, but it addresses its main difficulty, namely accurate die placement and attachment to the substrate. In addition, very interesting 3-D MCM structures have been proposed and are being considered for funding by DARPA and NRL.

The computer and electronics industry, especially in the United States, has made a strong and definite commitment to the technology of advanced packaging as one of the principal means of staying competitive in the next decade and beyond. This commitment was expressed in recent articles [23, 24] by the management of the Semiconductor Research Corporation (SRC) and MCC.

• Formed in 1985, MCC is owned by 19 U.S. electronics, aerospace, telecommunications, and semiconductor corporations: Advanced Micro Devices, Bellcore, Boeing, Control Data Corporation, Digital Equipment Corporation, Eastman Kodak, General Electric, Harris, Hewlett-Packard, Honeywell, Hughes Aircraft, Lockheed, Martin Marietta, 3M, Motorola, National Semiconductors, NCR, Rockwell International, and Westinghouse. Packaging and Interconnect is one of the five research areas. The tape-automated-bonding program approach is to integrate into one process the steps of wafer passivation, wafer bumping, inner lead bonding, outer lead bonding, and chip-on-tape encapsulation. The goal of bonding is 400 leads to a 400-mil square die. The Interconnect Technology program approach and goal is to achieve higher wiring densities in each of the wiring layers by using substrates 4 × 4 in. and larger at a cost of $5 to $10 per square-inch. The Multichip Systems Technology program aims at providing U.S.

semiconductor, materials, component, and systems companies with technology that will allow a 10-million-gate computer/signal processor to be packaged, assembled, and tested at a cost of less than $1000 (not including the chips); this program has produced a software program and models that help analyze system cost, reliability, yield, assembly, and repair of liquid-cooled methods that achieve below 1°C/W/cm$^2$ and an air-cooled method that achieves 2°C/W/cm$^2$.

• SRC is formed by seven U.S. semiconductor device competitors and four U.S. computer competitors. Recognition of the need for advances in packaging technology has led the SRC to be involved in packaging research since its inception, with four areas defined since 1983: computer-assisted package design and optimization, thermal management, chip-to-package interconnection, and innovative materials for low-cost packaging. The 1994 packaging goal is 1000 input/outputs, flat chip profile, 2 × 2-cm chip-size, nonhermetic surface-mount, 256 buses, 100 ps rise time, multilayer capabilities, 10 or 100W power dissipation. With the 1984 technology as reference, the 10-year goal calls for a 250-fold increase in density, 10,000 times increase in functional throughput rate (gate Hz/cm$^2$), 500-fold decrease in cost per functional element and reliability of 10 FITS (1 failure in 1 million hours). The program aims to have real-time correlation of process, device, and circuit models in production. Advanced packaging research is done at the University of Alabama, Lehigh University, and Cornell University.

At the present time, the MCM market is mostly driven by military applications that require the fastest electronics in the smallest box. Indeed, most of the pioneering MCM projects in this country were funded by DoD. Some of these projects are:

- DARPA 85-90: GE Digital Signal Processors MCMs based on AT&T DSP chips. The programs demonstrated the DSP1, 2, and 4.

- *AFWAL 85-90: GE MCM Contracts* that demonstrated the packaging of memory devices. One contract demonstrated the rework and repair feasibility of the overlay approach.

- Air Force, RADC and DARPA: Space Computer Corporation (SCC) and its partner Alcoa received a number of contracts to research 3-D WSI and Packaging, HWSI Processor and Space Based Surveillance Technology.

- DARPA/ISTO: Information Sciences Institute (ISI) affiliated with the University of Southern California has been working with DoD in microelectronics. Most well known is the MOSIS VLSI fabrication services set up originally by DARPA/ISTO. The current Advanced Packaging Technology (APT) contract is a precursor of a similar brokerage service in the area of advanced packaging.

- DARPA/ISTO: MCC Laser Direct Write MCM aims at producing a quick-turnaround MCM capability based on the MCC QTAI technology and a specialized software router. Another contract is the ES-Kit that provides a small modular software and hardware framework to experiment with new high-performance computer architectures.

- Air Force: IBM Advanced Spaceborne Computer Modules are fabricated with the VCOS technology.

- DARPA 90-92: In 1990, GE-TI was awarded a multimillion-dollar contract to advance the thin-film MCM prototype facility into a foundry that can serve the DoD.

- NOSC 89-91: Hughes VLSIC Packaging Technology (VPT) contract develops the HDMI MCM process.

## Some Design and Fabrication Issues

### *Chip Attachment and Maximizing Active Circuitry Area*

On a 2-D thin-film MCM, where interconnection requirements rarely approach the available wiring density on several layers, maximizing active circuitry to board area means placing the dies as close together as possible.

For the overlay construction, dies may be placed as close together as placement tolerance allows. Both GE and ISA have demonstrated die separation on the order of several mils, thus making use of more than 90% of the substrate footprint for active circuitry. For these methods the limiting factor is how accurately the dies may be placed relative to the interconnection pattern on the substrate. Pick-and-place robot arms at the MIT Lincoln laboratory can place dies with better than 8 $\mu$m accuracy, placing them no more than a few mils apart. Unistructures, Inc. possess a proprietary process that controls the positioning to roughly 5 $\mu$m in the x-, y-, and z-axes [25]. Without this accuracy, the final connection from the input/output pads on the die to the corresponding interconnection lines are determined by computer vision and adaptive connection, as is done in the GE process. The adaptive connection method actually complicates the routing. In addition, space must be reserved where bridges are to be built to facilitate the final bridging process.

For conventional construction, the die separation depends on the type of chip attachment used: wire-bonding and TAB require extra footprint on the MCM substrate, while flip-chip bonds are underneath the die area. These attachments are described as follows.

- *Wire bonding* — In the most conventional method, the die is glued to the substrate faceup. A wire is manually soldered to the IC's bonding pad and to the corresponding bonding pad on the substrate surface. The typical wirebond pitch is 200 $\mu$m (8 mils). Besides the inefficiency and high cost of precise manual work, the wire bond represents an uncompensated line that exhibits excessive inductance. A typical value of 1 nH has a severe effect on the power supply to the chip if a large number of I/O signals simultaneously draw current—a very likely situation in synchronous circuits.

- *Tape automated bonding (TAB)* — Thin film interconnection patterns are fabricated into a standard-size plastic tape that is rolled into a reel. In a typical automated process, a segment of the tape is positioned exactly over a faceup wafer and pressed down under precisely controlled and localized heat to make a thermocompression bond [26] (Fig. 14). In the Bumped-Tape TAB technology, both ends of each TAB finger are processed to contain a drop of solder. In the Bumped-Chip TAB technology, wafers are processed to put bumps at each bonding pad. The pads may be placed at the periphery of the chips or distributed over the entire area. One-, two- and three-layer tapes are available to make the simple peripheral lead pattern or a multilayer interconnection for pad arrays. TAB tapes are typically made from polyimide and copper foil, each several mils thick. The patterns are made by either photoetching or electroplating copper on a carrier tape with or without polyimide. The standard TAB tape sizes are 35, 48, and 70 mm. The outer lead bond standard pitches are 10, 15, and 20 mil. The inner lead bond pitch is determined by the I/O pad pitch on the die, which is typically 4 mil.

TAB is a 30-year old technology that has recently received wide interest as the number of leads per VLSI package increase into the hundreds. It will take a few years for the market to gear up to make TAB widely available and at competitive cost to small volume applications.

The advantages of TAB are numerous. It has the possibility of gang-bonding that will further increase automation, and it has the convenience of burn-in and testing of TAB-bonded unpackaged dies. High-frequency performance improves marginally—the typical inductance per lead is about 0.7 nH.

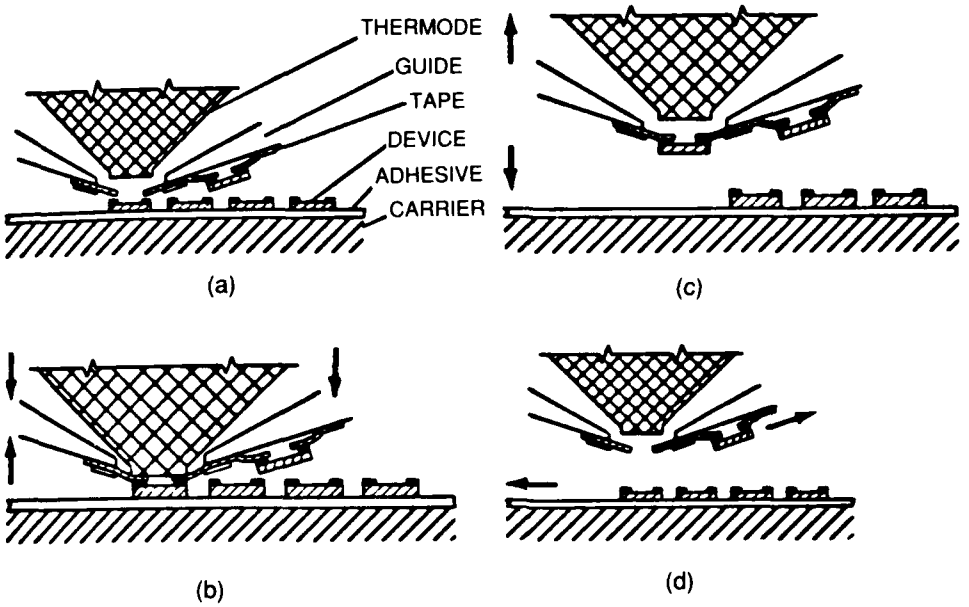


Fig. 14 — Inner lead bonding sequence. Courtesy of Lau et al. [26].

The disadvantages include a lower thermal conductance path to remove heat from the chip. This is true only when the dies are placed facedown. And as with wire-bonding, a peripheral area expansion is required around the footprint of the die on the substrate.

• *Flip Chip Bonding* — After the bonding pads of a die and the substrate are bumped, the die is mounted facedown on the substrate. The solder bumps are of a controlled pressure and temperature to reflow into a bond. The solder bump diameter is typically 4 mil. A flip-chip connection exhibits less than 0.05 nH and is excellent as a thermal conduit. The largest drawback is the unavailability of bumped wafers [27].

*Material Selection*

Since the VLSI fabrication process works only with silicon and silicon dioxide, the new wave of microelectronics and their vast selection of material for substrate, conductor, insulator, and encapsulant present new opportunity for improvement as well as failure. A thorough discussion of the subject is beyond the scope of this report. To introduce the key issues involved, the key criteria for HWSI material selection are reviewed. These include: high-speed performance, thermal conductivity, thermal expansion elasticity, mechanical durability, cost and availability.

• *High speed performance* — The resistivity of the conductor material and the dielectric of the insulating material affect the transmission line behavior of signal lines as outlined previously. The effect is quite indirect since improvements of these parameters only increase the critical distance $d_c$ for a given set of conductor and insulator thickness. As Table 1 shows, a desirable $d_c$ is obtainable for a rather large value of insulator thickness, which results in proportional increase of the thermal resistance through the layer. For polyimide, which has very poor thermal conductivity (0.25 W/°C-m), any reduction in insulator thickness is useful.

• *Thermal conductivity* — Heat generated from the movement of electrons and holes as the circuit operates must be removed and transferred to the ambience, which could be air or liquid. Each layer of material on the path, from the chip substrate to the colder air of liquid ambience, contributes some amount of heat resistance that is inversely proportional to the material axial thermal conductivity. The

lateral spreading of heat by the material is a secondary effect that helps disperse the heat from the heat source to a larger area. Lateral thermal conductivity is also very important where heat is not removed from the surface of module or board. The military dual-standard electronic-module (dual SEM) that mounts chips on both sides of the board is a case in point.

• *Thermal expansion and elasticity* — When layers of different material are stacked, the layers must expand and contract together to avoid stress that results in the warpage of the planes that promotes cracking. Expansion and contraction occur when the module undergoes a sudden temperature change, such as when power is first applied or when it is turned off. Thermal stress is also induced through the fabrication process that subjects some of the structures at temperatures from 150°C to over 1000°C. The traditional approach has always been to match the coefficient of thermal expansion (CTE). However, recent developments are considering the fact that a larger degree of thermal expansion may be tolerated if the interfacing layers are sufficiently flexible and the internal stress generated is less severe, [28-30].

• *Mechanical durability* — Large, flat surfaces have the lowest natural frequency and the expected vibration amplitudes and dynamic stresses are the greatest. To increase natural frequency, the material selected should have a high ratio of stiffness to mass.

• *Corrosion-free* — After mechanical stress, corrosion is the most important factor that affects microelectronic package reliability. The chemical or electrochemical attack normally occurs in metals that have electrons that are free to move. Moisture, dirt, and dust are the major sources of corrosion in semiconductor devices. A typical substrate undergoes high temperature curing to dissolve undesired substances. Also, through this process any moisture that has been absorbed evaporates. Organic material is water permeable, therefore, must be protected to some extent by a layer of coating or by a hermetic seal. Moisture uptake, measured as the percentage change in dimension as the material swells up when water is absorbed, is an important measure especially for polymers.

• *Manufacturability* — A typical lithographic process exposes the substrate to solvents during pattern definition; insulator material must, therefore, be insensitive to these solvents. The patterning process can be shortened by using photosensitive material. A UV light source, for example, may be used to expose a pattern on the insulator. Curing temperature must also be considered when choosing insulator material. A low curing temperature is advantageous because it results in less thermal by induced stress.

*Some Design Issues*

• *Heat Removal* — Temperature has a dramatic effect on the performance and reliability of an electronic device. An approximately 10°C increase in temperature reduces the mean time to failure by a factor of two [31]. As dies are placed closer together and the module clock rate increases, the heat generated per unit area increases and places more severe requirements for removing the heat. For example, a factor of 4 area compression and a factor of 2 frequency increase for CMOS circuitry means 8 times more wattage is produced per unit area.

The ability of a material in a packaging system to conduct heat away from the heat source is usually characterized by the thermal conductivity in units of W/m-K. To get a feel of the thermal design of a package, a first-order calculation is usually made based on the simplistic model of 1-D thermal resistances of parallel slabs. For a plane slab with faces maintained at temperatures $T_1$ and $T_2$, the heat transfer is

$$Q = (kA/L)\ (T_1 - T_2) \tag{6}$$

where $k$ is the thermal conductivity, $A$ is the area, and $L$ is the thickness.

The thermal resistance in units of °C/W is

$$R = (T_1 - T_2)/Q = (L/kA) \tag{7}$$

The 1-D model may be extended in an analogous manner to electrical resistance to calculate series and parallel thermal resistance from the heat source (I/O pads on the chip) to the heat removal interface. The temperature differential between chip I/Os and the heat removal interface may then be estimated as the product of the thermal resistance and the chip power.

For the thermal path from the die substrate to MCM substrate, the overlay and wire-bond/TAB MCMs have the best conduction because of the large contact area between the die substrate and the MCM substrate. Flip-chip MCMs rely on an array of bumps made of solder, which has a relatively low thermal resistance.

For high-power modules, coolant channels are fabricated inside the MCM substrate. An advanced design uses microchannels distributed in an H-tree topology.

The temperature differential in the profile of a typical overlay MCM is calculated. The MCM contains 25 1-cm$^2$ dies placed perfectly side-by-side to fill up the module. The MCM interfaces to a silicon carbide microchannel liquid-cooled heat sink through a button board. The button board is a ceramic board with 25-mil copper buttons placed in an array; the distance between buttons is 50 mil. Thermal resistance is calculated by using Eqs. (6) and (7). For the button board and the air gap between the button board and the interfacing layers, the effective thermal conductivity is calculated as follows:

$$k_{\text{eff}} = R\,K_{\text{button}} + (1-R)\,K_{\text{ceramic}} \tag{8}$$

where $R = \text{Area}_{\text{button}}/\text{Area}_{\text{cell}}$, and $K$ denotes thermal conductivity.

Table 2 shows the parameters and calculated thermal resistances.

Table 2 — Calculations of Thermal Resistance of Various and Encapsulated Overlay MCM Construction

| Layer | Material | Thickness ($\mu$m) | Conductivity (W/mK) | Therm. Res.K/W |
|---|---|---|---|---|
| MCM conductor | Aluminum | 25 | 216 | 4.63E-05 |
| MCM insulator | Polyimide | 35 | 0.25 | 5.60E-02 |
| Encapsulant | Polyimide | 25 | 0.25 | 4.00E-02 |
| Circuit die | Doped silicon | 250 | 98.4 | 1.02E-03 |
| Adhesive | Silicon gel | 12.5 | 0.4 | 1.25E-02 |
| Base plate | Al nitrite | 1250 | 140 | 3.57E-03 |
| Button board | Ceramic/cop-per | 625 | 56 | 4.46E-03 |
| Airgap | Air/copper | 50 | 33 | 6.06E-04 |
| Adhesive for heat sink | Silicon gel | 12.5 | 0.4 | 1.25E-02 |
| Heat sink | Silicon carbide | 2000 | 270 | 2.96E-03 |

Assuming that the heat is adequately drawn to the heat sink through convection, the temperature through different layers in the MCM is calculated for a set of heat flux, given in terms of W/chip.

28W/cm$^2$ is typical of the MCA-3 ECL gate array; 7 to 14 W/cm$^2$ typify 1$\mu$m-CMOS gate arrays clocking at 100 MHz with 20% and 10% devices clocking, respectively. 0.5-1W/cm$^2$ is typical of most CMOS VLSI circuits clocking at 5 to 15 MHz today. Figure 15 shows the result. It is clear that the polyimide layers are responsible for over 70% of the temperature differential from heat sink to heat source.

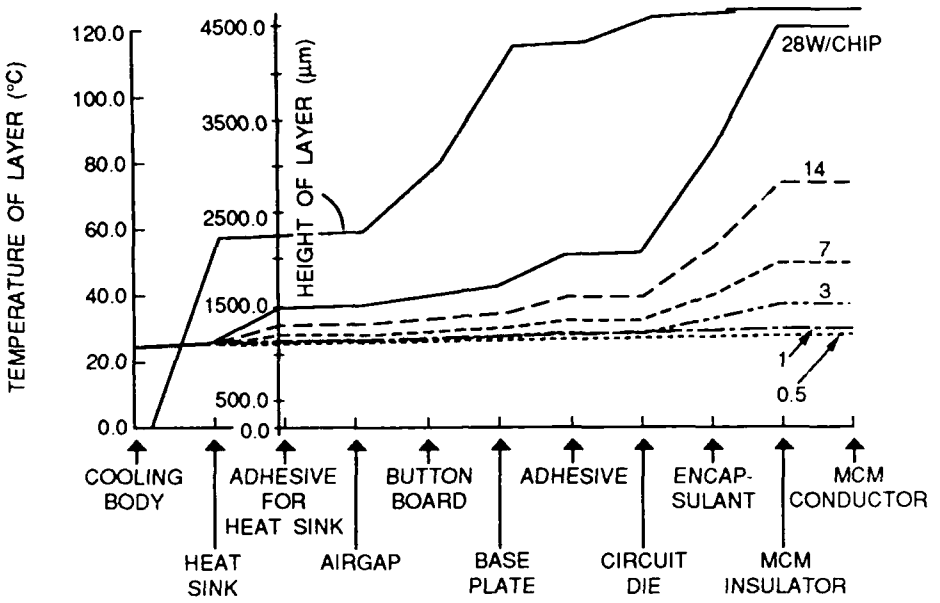


Fig. 15 — Temperatures at each MCM layers for a variety of heat flux conditions. The thick polyimide layers are responsible for over 70% of the temperature drop.

- Test, Repair and Rework

*Bare die testing, rework and repair* — The importance of thorough testing of the component dies before assembling are illustrated by considering a module with 40 chips that have a failure rate of 1%. The probability of the module to be defective is (0.99)40= 0.669.

Commercial integrated circuits typically go through extensive design maturity tests (DMTs). The DMT is accomplished by operating a set of sample devices for two to four months to simulate actual field operation. Failures and their causes are identified and corrected repetitively until the mean time to failure reaches a certain statistical threshold [32].

Complex VLSI chips require extensive incoming testing by using automatic test equipments that connect to all the pins of the chips. It is necessary to mount a bare die on a temporary carrier for testing. In a typical procedure, the bare chips are sprayed with a protective coating then attached to the test carrier by using an epoxy that can be easily dissolved. The test carrier and the chips on it are then interconnected into a temporary MCM that can be tested at high speed. In another approach, a die may be vacuum held faceup on a temporary test carrier, then a short TAB-like ribbon cable is placed on the die and formed into a gull wing shape. The inner leads of the ribbon are bonded to the die pads, and the outer leads are bonded to the test probes. After testing, the outer leads are laser cut, leaving leads just long enough for bonding to the final circuit.

For overlay MCMs, testing, rework, and repair are made possible through the use of a solvent sensitive or thermoresin layer of material. This material is coated on the encapsulated chips before the interconnection is fabricated. Dies are mounted on a test substrate or carrier, and a special test interconnection pattern is constructed on this test substrate. After testing, the test circuit is destroyed by

removing everything above the solvent sensitive or thermoresin layer. An 11 times rework of a GE MCM module has been successfully demonstrated under an AFWAL sponsored work.

*Substrate testing* — The testing of the interconnection substrate may be separated from the MCM testing in some approach where the chips are attached after the interconnection layers are formed. In this case, the most common methods used in the printed circuit board industry may be used to check for open and short against an expected database: flying probes and network capacitance probing. Advanced testing, based on an electron beam being investigated under DARPA sponsorship, promises to overcome the limited speed of mechanical probes.

Flying probe systems usually use two electrical probes that are moved by mechanical arms over a circuit, each probe position at a test point or power and ground. Typically, these systems are intended for point-to-point spacing of 20 to 10 mils, with placement accuracy of 2 mils, which are inadequate for the typical 20 $\mu$m spacing in thin film MCMs.

Network capacitance probing use a single probe to measure the capacitance of a net to a ground reference plane. Errors are indicated when two points on the same net get a substantially different capacitance reading, or when the net capacitance value is sufficiently different from the computed value. The typical system performs at approximately half a second per net [33].

A voltage contrast electron beam (VCEB) test system uses an electron beam and thus avoids all problems associated with mechanical probes. The beam reads the voltage present at each node in a net, then injects a charge into the net before reading the other nodes in the net. Electrical discontinuity (open circuit) error is indicated at the nodes that show a different voltage. Short circuit error is indicated if a charge is detected at a node at the initial read because it is shorted to a net that has been previously charged.

**Comparing 2-D MCM Approaches**

*PWB vs MCM* — It is easy to make a decision between the conventional PWB and MCM technology based on the ratio of active circuitry area (e.g., *silicon*) over the footprint on the board. Though-hole multilayer PWB use less than 1% area for active circuitry, and the most aggressive surface mount PWB uses no more than 10% [34]. To achieve better than 10%, multichip modules of wafer sizes must be used.

*Thick- vs Thin-film MCM* — To choose between thick-film and thin-film MCM is also easy for high performance applications, where the signal frequency exceeds 50 MHz, and interconnect density is more than 2000 in./in.$^2$. The lack of fine thickness control of the screen-printing process used in thick-film MCM and skin effect limit the high-frequency performance of thick film. To achieve 2000 in./in.$^2$ interconnection density, a 3-mil-width thick-film MCM requires more than 20 signal layers. Too many layers result in high cost and technical problems such as increased thermal resistance.

Thick film is inherently more suitable for high power applications [35]. The thin-film conductors break down at 2.5 to 7.6 mA per unit width in mils, while thick-film conductors breakdown at 50 to 125 mA/mil. The starting breakdown voltage for the dielectrics used in thin and thick film respectively are 2.5 and 13 V/$\mu$in.

For performance, thin- and thick-film MCM technologies overlap a great deal in the low density and multi-MHz speed area, which may be defined to the first order by a line density of 2000 in./in.$^2$ and signal frequency of 50 MHz. Since thin film provides higher quality signalling and twice silicon-to-footprint area, the only tradeoff is cost. At the present time, the cost reflects that more companies are

capable and willing to compete in the thick-film technology while most companies with new thin-film lines are still in the learning curve. This is expected to change in a few years.

*Comparing thin-film MCMs* — Although thin-film MCM profiles described above differ in some details, these differences are not inherent. The final success of a thin-film MCM process depends on the best combination of thorough and robust design (test and repair methodology, chip attachment method, thermal removal mechanism, etc.) and selection of material and geometric dimensions for substrate, conductor, and insulator.

Some of the construction approaches described above are still in the development stage, and most are still undergoing iterative improvements. A definite comparison is therefore difficult or impossible. Nevertheless, some comparative comments can be ventured here:

- The overlay construction eliminates a separate bonding process. This reduces the number of manufacturing steps significantly. GE contends that the number of processing steps involved in its overlay is 50% less than in the flip-chip, and 42% less than in the flip-tab construction [36].

- The overlay approach construction can pack the most active circuitry per module area.

- The overlay approach construction potentially has the lowest thermal resistance path from die junctions to the heat sink. The encapsulation overlay construction is even better because the substrate can be metal.

- The cavity-in-ceramic overlay approach could be handicapped because of the need to mill cavities in ceramic.

- The conventional construction that uses hard masks is potentially better than the laser-based construction in handling very large volume.

- Silicon substrate is a poor choice unless active devices are designed on the substrate. There is no evidence of the significance of such a need, however.

- A substrate based on metal or its alloys is preferable to silicon or ceramic for thermal and mechanical reasons.

- The PSCB provides the least amount of active circuitry area, per module area, and shares most disadvantages of the conventional approach.

- The overlay constructions and PSCB are best for testing after module completion because of the large number of pads.

- Die test and burn-in can be done more easily in the TAB-based conventional approaches than in the overlay.

## 5. THREE-DIMENSIONAL HWSI MODULES (3-D HWSI)

In the quest for packing the most active circuitry in a given volume, stacking of dies, wafers, or MCMs seems to be the natural thing to do. By packing circuits closer than ever, the distance between pipelines in a circuit could be dramatically decreased, especially if interconnections can be made through the horizontal layers. The promise of such high payoffs has created much excitement and activities at the present time.

Naturally, one expects many complications. Indeed some key issues stand out: the severity of heat concentration, how to remove heat from such a concentrated heat source, and the vertical interconnection structure through the layers of the 3D structure.

These issues are addressed in this section after the representative design approaches are described, and their status reviewed.

**Construction**

In general, 3-D packaging approaches can be classified in four categories based on the way vertical interconnections are made through the layers: peripheral-connector-based, sideways-interconnection, pyramid, and through-layer-area-interconnections.

*Peripheral-connector-based* — MCMs are mounted on a base plate or on a PCB. The periphery of the plate or PCB is reserved for I/O pads that interface to a z-axis connector. Such a connector can be based on a variety of basic technologies, exemplified by the following:

• The TRW/Cinch button-based connector is a thin section of a stable material such as ceramic with prefabricated holes in which buttons are inserted. A typical 20-mil diameter button is a wadded wire cylinder made from 1-mil diameter gold-plated silver copper wire. In a typical application, the buttons protrude 3 to 5 mils above and below the button board. When compressed, the protruding wire is forced into the hole. The connection must be secured by clamping pressure through the boards. Figure 16 shows a TRW/Cinch button. This technology has been thoroughly tested by TRW for reliability.

Fig. 16 — TRW/Cinch button

• As Fig. 17 shows, a button/plunger connector that uses a typically 20-mil button is also available from Clinch. The height of the connector is about 1/4 in..

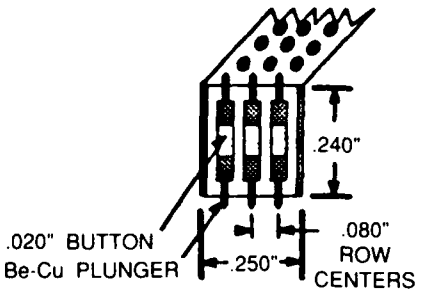


Fig. 17 — Button-based plunger connector.
Courtesy of Isi.

• Figure 18 shows the AMP part 54503 ZIF connector that provides both z-direction connection and an edge connector beam contact. The connectors may be directly stacked on top of one another.

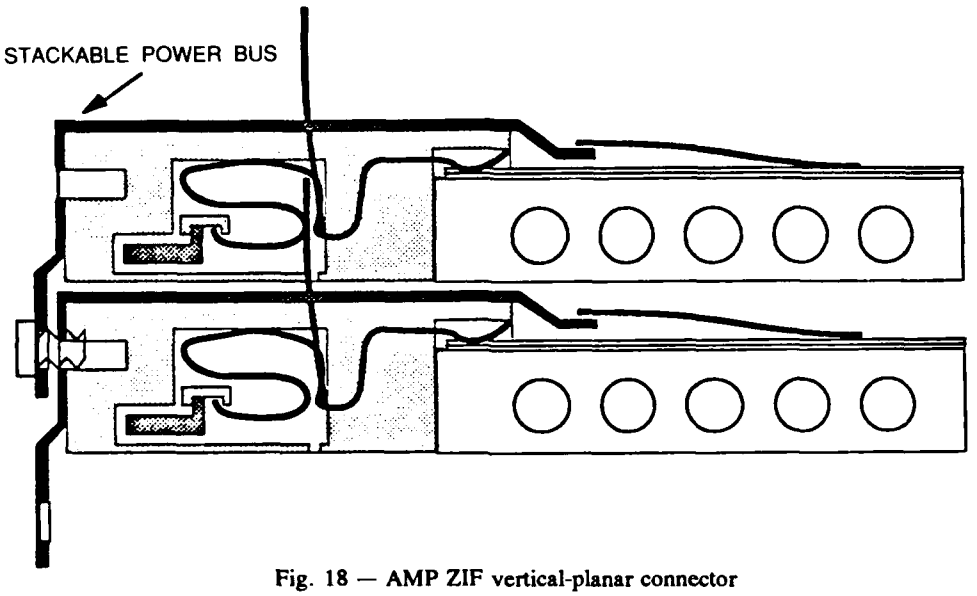


Fig. 18 — AMP ZIF vertical-planar connector

*Sideways-interconnection construction* — A metal layer of routing is deposited on the die to route the signals to one edge. The laminated dies are glued together by using a very thin layer of adhesive. The face of the resulting cube is lapped to expose leads, and the silicon edges are passivated. Metals are then vacuum-deposited to connect the exposed leads, and solder bumps are deposited. The module is flip-chip bonded to a carrier substrate. Figure 19 illustrates this concept. In an alternate design, each interconnection layer could be done entirely by a TAB film.

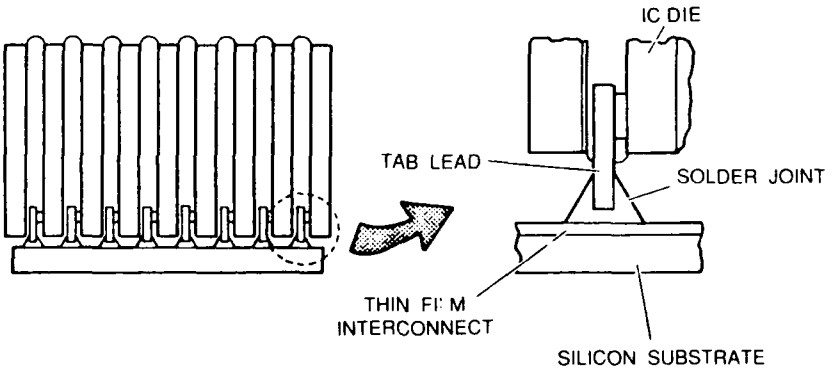


Fig. 19 — Layers of a TAB-based 3-D Memory Cube interconnect sideway by a TAB film. Courtesy of TI [36].

This structure is almost all doped silicon that has a good thermal conductivity of 98 W/°C-m, except for the adhesive layers that could add up in terms of high resistance. A thermal simulation was done for an 80-layer memory cube with all layers active. The result showed that each IC can dissipate more than 0.6 W without the function temperature rising above MIL standards [37].

*Pyramid* — A smaller die is glued on top of another. Interconnections between layers are done by wire bonding the I/O pads at the periphery of the top one to those on the lower ones.

*Through wafer area interconnections* — In this construction approach the interconnection points designated for the z-direction are distributed over the entire area of each wafer or MCM and are made present at both faces of the MCMs by feed-throughs. The MCMs are then pressed together to make the contacts. This requirement basically excludes constructions that are not flat. Only overlay MCMs are appropriate for this type of 3-D packing. A variety of techniques exist to make the feed-throughs:

- 20-mil diameter through-holes may be laser drilled through all the layers. The holes are electroplated to fill with metal. Since these feed-throughs are to be made during the MCM fabrication process, the actual sequence of construction is more complex because the drilled holes are filled with different material at different steps, and they must be removed before the final metal filling steps.

- For silicon wafers, hot aluminum drops may be dropped at the location and allowed to eat through the material in the so-called thermo migration used by Hughes Research in the 3-D computer [38]. In the 3-D computer, the feedthroughs interface at the wafer surfaces through a pair of microbridges to provide some alignment flexibility in the plane and to absorb the stress in the vertical direction. Figure 20 shows these mechanisms.

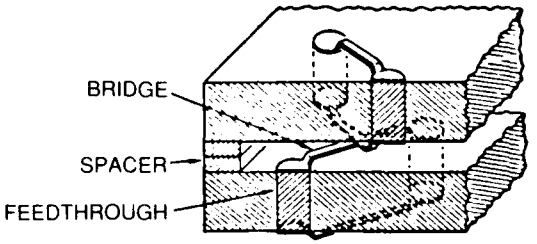


Fig. 20 — Microbridges and thermo-migrated feedthroughs. Courtesy of Little et al. [38].

- The metallic feedthrough on the substrate can be prefabricated. A via may be etched in the normal way through all of the interconnect layers (the vias will have to be larger because of the extra distance) to connect to the feedthrough through a layer filled with button contacts at prescribed or regularly spaced positions. The button contact absorbs stress in the vertical direction. Figure 21 illustrates this concept [39].

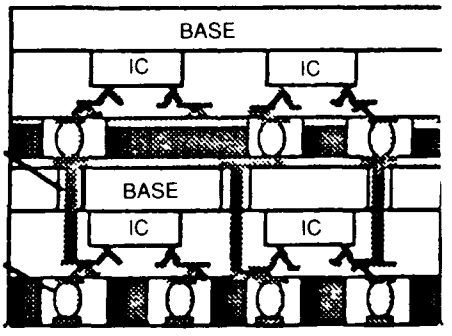


Fig. 21 — Vias made on a button-board with prefabricated feed throughs. Courtesy of Isa [39].

## 3-D MCM Design Issues

Since volume contraction is a major reason for 3-D packaging, the total height of the stack of MCMs must be sufficiently reduced from the typical pitch in traditional systems. As modules are stacked closer,

the thermal management becomes even more important. Another special consideration is the replaceability of each layer of a 3-D module for system repair, as boards are replaced and repaired in the traditional systems. The quality of the vertical interconnections will affect the high-speed capability of the module.

*Layer separation* — The key design parameter that differentiates the various constructions is the intimacy with which two layers can be made to contact.

• Apparently designs based on z-axis connectors have chosen not to let the layers come in contact; thermal and mechanical designs appear to be less daring. Several designs based on the AMP and button/plunger connectors specify the layer separation to be as much as one half of an inch.

• The most compactness is achieved in the 3-D memory modules produced by several places. The inherently simple I/O and low-power requirements make memory structures perfect applications of the sideways-interconnection method. The layers are simply glued together with micron-thick epoxy. Each routing layer is a simple single-conductor layer no thicker than several mils. In addition, each memory die is lapped to no more than a few mils.

• The ISA encapsulated overlay approach includes a base plate, an adhesive layer, the encapsulant thickness above the dies, and between 5 and 10 insulator/conductor layers depending on speed requirements. Table 2 shows a typical profile. The total thickness is 171 mils excluding the heat sink.

• For high-power requirements, MCM stacks may be interspersed with liquid-cooled micro-channel heat sinks. The thickness of one such design [40] is 160-mil plus a 32-mil thick adhesive layer. The total thickness per stack is 362-mil, if a heat sink exists for every stack. If one is used every 10 stacks, the average thickness in a 10-stack 3-D modules is a little less than 200 mils.

*Thermal Removal* — Traditionally, the heat transfer path is designed to be perpendicular to the plane of each MCM stack. The temperature at each stack in 3-D MCM, based on the encapsulated ISA design described above, can be calculated by using the same methodology described in Eqs. (6), (7), and (8). By making the worst-case assumption that heat is drawn by convection in one direction from the top stack to the bottom, the temperature at each stack can be calculated in a manner analogous to the elementary electrical circuit theory. In fact, in an $N$-layer stack the temperature $T_i$ at the surface of layer $i$, $i = 1$ for the bottom stack and $i = N$ for the top, is calculated as follows:

$$T_i = T_0 + R\,F[1 + \sum_{i=1}^{N-1} (i)] \tag{9}$$

where

$T_0$ is the temperature at the bottom of the bottom stack, presumably of the ambience or heat sink,
$R$ is the thermal resistance that is the same for all stacks, and
$F$ is the heat flux generated at the top of each stack, which is considered the same for simplicity.

Figure 22 shows the temperature profile through the layers of a set of heat flux density as discussed previously for the 2-D thermal analysis. For a maximum temperature differential of 100°C, the number of stacks that can be assembled for a given heat flux density is indicated in the figure as the place on the horizontal axis where the corresponding curve stops.

Thermal removal techniques must be considered to decide if a design is feasible. If air-cooling is used, the typical thermal resistance of the heat sink is 2 °C-cm$^2$/W. For heat-flux densities 7W/cm$^2$ and below, the temperature drop across the heat sink is 14°C or less. Liquid-cooled heat sinks must be used to keep temperature drop at 7°C and 14°C for the 14 W/cm$^2$ and 28 W/cm$^2$ cases, respectively.

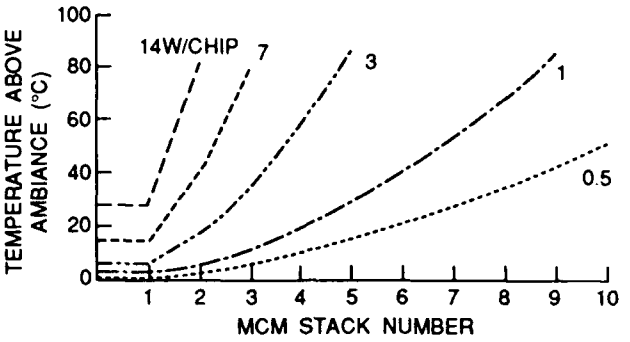


Fig. 22 — Temperature profiles and number of stacks that can be assembled for a set of heat flux densities. Each stack is an encapsulated overlay MCM with 25 1-cm$^2$ chips.

In the low-power applications where each chip generates less than 3 W, the calculation shows that 5 to 10 layers may be stacked without raising the top stack 100°C higher than the bottom one. For 100 MHz CMOS, which generates between 7 and 14 W/chip, two to three layers are possible. For the state-of-the-art ECL gate arrays that generate 28 W, stacking is not possible.

### Status of 3-D Packaging

So far, the only significant activity in this technology is in 3-D memory. There are many announcements and design proposals along the lines reported above, but only a few have been implemented. These efforts are outlined as follows:

• Irvine Sensors, Inc. 3-D Memory — After a period of research and development under DARPA sponsorship since 1987, the company's RAM pack that contains up to 80 ICs per module is now being offered commercially. However, at the present, ISI possesses only prototype capability.

• Harris Corp. 3-D Memory — In 1988, Harris produced experimental 8 64-K memory dies based on the side-interconnection approach similar to that of ISI. The difference is each interconnecting layer is a TAB film, and therefore a resin has to be used to encapsulate each layer to keep the structure in place. The company is actively working on improved versions, especially a nonorganic version for military application.

• Texas Instruments produces 3-D memory by using TAB-based sideways-interconnections approach.

• Hughes Research has successfully demonstrated a small version of the 3-D computer that features 32 × 32 array of processors in five stacks of 1-in. wafers, employing the microbridges and feedthroughs as discussed. The program continues with the design of the 128 × 128 version.

• Space Computer Corp, Polycon are working on a z-connector-based stacked hybrid WSI version of the WASP machine. No results have been reported yet.

## SUMMARY AND RECOMMENDATIONS

The main findings of this report are:

• An advanced packaging model is needed for the next generation of electronics systems. Such a model will use optical signals for communication and 3-D semiconductor modules.

• High density packaging technology will help achieve higher performance systems in two ways: make high clock rate systems faster by reducing signal lengths, and make scaling-up of massively parallel computers possible by providing the necessary bandwidth for interconnections and by holding down the growth in the physical dimensions.

• Application-specific integrated circuits (ASIC) are abundantly available in all logic technologies for different needs.

• Success in MWSI is seen only in memory devices and cellular architectures.

• Techniques for successfully extending VLSI technology into WSI have been identified but not applied.

• Many MCM technologies exist in a complementary way. The appropriate technology may be selected along the lines suggested in Fig. 23.

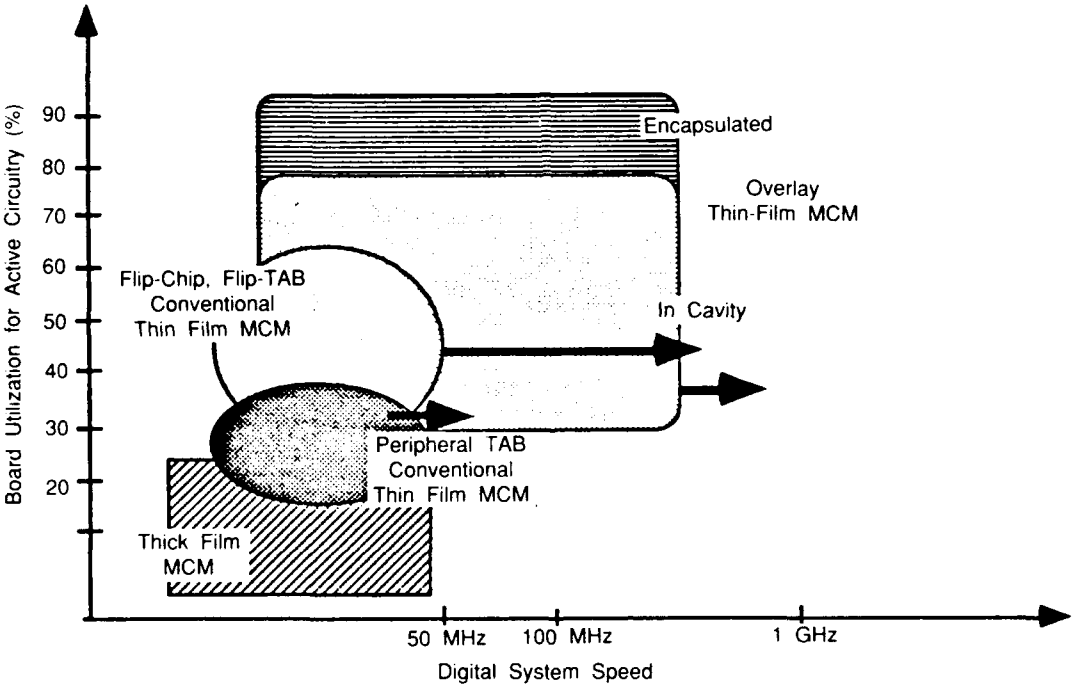


Fig. 23 — Summary of 2D MCM techniques

These 2D MCM technologies are still in the start-up cycle, and therefore they are not cost effective at present. However, the cost will go down to less than one thousand dollars per module in 2 years.

The encapsulated overlay MCM is a new invention that holds promise of a low cost, high performance, and convenient package. The company behind the technology is starting up this year and expects operation by next year.

Hybrid memory cube technology based on TAB is available from several sources. It could be a very competitive alternative to monolithic WSI memory.

Several MCM stacking approaches based on z-axis connectors are being explored. However, a 10:1 improvement in vertical space utilization is most likely achievable through the button-board technology, coupled with a flat MCM such as the encapsulated overlay MCM.

Given these observations, the Navy could take some proactive steps as part of a technology development and insertion strategy.

In the short term, the following are recommended:

- ASIC technology should have the first opportunity for technology insertion to increase circuit density and performance. Retrofit and new system development should not do without this technology.

- Hybrid memory cubes should be recognized as the easiest and most robust implementation of large scale memory. Procurement should start now to anticipate the low-cost, high-volume availability in 2 years.

- 2-D thin film MCM should also be a candidate for immediate use in packaging systems that require either high-speed or very high-density capabilities or both.

- Monolithic WSI implementations of memory should be evaluated critically vis-a-vis the hybrid alternative.

- While the RVLSI program holds the best promise of technology, a serious potential user should be satisfactorily enlightened about the implementation of transmission lines on the wafer and the low utilization of silicon.

- More research funds should be directed to the development of encapsulated overlay structures for 2-D and 3-D modules.

In the midterm, i.e., the next 3 to 6 years the following are recommended:

- Build and deploy miniaturized versions of some massively parallel signal processors in the field. New parallel computers are breaking into a new trajectory of throughput increase, currently in the tens of gigaFlops, within two years in the teraFlops. These signal processors also have a supply base of software and hardware in the commercial world and are unlike previous exotic military machines that got too expensive because they were one-of-a-kind. The new packaging technology will make available for the first time commercial supercomputers for field deployment.

- The use of high density packaging technology is also beneficial in accelerating and perhaps lowering the cost of the implementation of the next generation—the teraFlops machine.

Looking ahead into the next decade, one can see the replacement of electronics equipment wherever possible with programmable processors. The only way to hold down the cost of the defense weaponry is to anchor the logistics mechanisms into the commercial base.

## REFERENCES

1. S. K. Tewksbury and L. A. Hornak, "Communication Network Issues and High Density Interconnects in Large-Scale Distributed Systems," *IEEE J. Sel. Areas Comm.* **6**(3), 597-598 (1988).

2. L. A. Hornak and S. K. Tewksbury, "On The Feasibility of Through-Wafer Optical Interconnects for Hybrid WSI Architectures," *IEEE Trans. Electron Devices* **ED-34**(7), (1987).

3. A. Guha et al. "Optical Interconnections for Massively Parallel Architectures," *Applied Optics*, **29**(8) (1990).

4. R. Selvaraj et al., "Integrated Optical Waveguides in Polyimide for WSI," *J. Lightwave Tech.* **6**(6), (1988).

5. A. Hussain, Honeywell, private conversation.

6. T. A. Lane et al., "Gigabit Optical Interconnects for the Connection Machine," *SPIE OR/Fibers*, (1989).

7. A. Barna, *VHSIC, Very High Speed Integrated Circuits Technologies and Tradeoffs* (John Wiley and Sons, 1981).

8. J. Yuan and C. Svensson, "High Speed CMOS Circuit Technique," *IEEE J. Solid State Circuits* **24**(1), 62-70 (1989).

9. J. Gabay, "ASIC Technologies Boost Chip Density and Speed, High Performance Systems," 32-53 (1990).

10. D. H. Hartman and C. C Shen, "GaAs Photodetectors" p. 343, and "Optical Sources" p. 409, in *Gallium Arsenide Technology*, D. K. Ferry, ed. (H. W. Sams and Co. Inc., 1985).

11. G. H. Chapman, J. M. Canter, and S. S. Cohen, "The Technology of Laser Formed Interactions for Wafer Scale Integration," Proc. Int. Conf. on Wafer Scale Integration, Jan. 3-5 1989, pp 21-29.

12. The Programmable Gate Array Data Book, (Xilinx, Inc., 1988).

13. H. Stopper, "Wafer Scale Integration," in *Electronics Materials Handbook*, Vol. 1 Packaging, May 1990), p.362.

14. R. Bourassa, T. Coffman, J. E. Brewer, "Progress in WSI SRAM Development".

15. J. Trile, "A 4 Mbit Static RAM," Proc. Int. Conf. WSI 1989, pp.193.

16. I. P. Jalowiecki, K. D. Warren and R. M. Lea, "WASP: A WSI Associative String Processor," Int. Conf. WSI 1989, pp.83-93.

17. M. Sage et al., "Staff Report, BPA Technology and Management," 1989.

18. H. Stopper, "Wafer Scale Integration," in *Electronics Materials Handbook*, **1** Packaging, p.362 (1990).

19. W. Borland, "Thick-Film Hybrids," in *Electronics Materials Handbook*, **1** Packaging, p.333.

20. D. Herrell and H. Hashemi, "Hybrid Wafer Scale Integration," MCC Technical report P/I-329-89.

21. M. Matsunami et al., "A Study of High Density Multilayer LSI," Proc. Int. Conf. WSI, Jan. 1990, p.322.

22. Susan Godsell Baust, Richard J. Dishler, Scott Westbrook, "Implementing A Packaging Strategy for High Performance Computers, High Performance Systems," January 1990, pp. 28-31.

23. R. K. Calvin III, L. W. Sumney, and R. M. Burger, "The Semiconductor Research Coporation: Cooperative Research," Proc. IEEE, 77(9), 1327-1344 (1989).

24. G. A. Dove, "Cooperative Research at MCC: A Focus on Semiconductor-Related Efforts," Proc. IEEE, 77(9), pp.1364-1375 (1989).

25. M. Busby, "Interconnect Advances Bare-Chip Packaging, High Performance Systems," Jan. 1990, pp. 22-

26. J. H. Lau, S. J. Eramus and D. W. Rice, "Overview of Tape Automated Bonding Technology," in *Electronics Materials Handbook*, 1 Packaging, p.281.

27. C. Barlett, J. M. Segelken and N. A. Teneketges, "Multichip Packaging Design for VLSI-based Systems," IEEE Trans. CHMT, **CHMT-12**(4) Dec. 1987, p. 649.

28. M. Shinohara, "Encapsulants—Overview," in *Electronics Materials Handbook*, 1 Packaging, p. 808.

29. C. Echelburger, Private conversation.

30. J. W. Balde, "Overview of Multichip Technology," in *Electronics Materials Handbook*, 1 Packaging, p 305.

31. Mali Mahalingam, "Design Considerations," in *Electronics Materials Handbook*, 1 Packaging, p. 408.

32. D. P. Sieworek, R.S. Swarz, "The Theory and Practice of Reliable System Design," (Digital Press 1972) p. 7.

33. C. Hilbert and C. Rathmell, MCC Technical report P/I-431-89.

34. J. W. Balde, "Overview of Multichip Technology," in *Electronics Materials Handbook*, 1 Packaging, p. 298.

35. H. K. Charles, Jr., and G. V. Clatterbaugh, "Thin-Film Hybrids," in *Electronics Materials Handbook*, 1 Packaging, p. 322.

36. GE-TI proposal to BAA 90-10.

37. S. N. Shanken and J.C.Carson, "Very High Density 3-D Packaging of Integrated Circuits," ISHM '89 Proc.

38. M. J. Little et al., "The 3-D Computer", Proc. Int. Conf. Wafer Scale Integration, 1990, pp. 55-82.

39. ISA White Paper to BAA 12-90.

40. R. Darveaux et al., "Thermal/Stress Analysis of A Multichip Package Design," IEEE 1989, pp. 668-671.