INTRODUCTION

Information processing is used directly in computer products and indirectly as the brains in “smart products” such as electronic date books, central office telephone switches, medical monitoring instruments, and the latest luxury cars.

The incredible thirst we have for information processing power is fueling the migration of yesterday’s super computer abilities into tomorrow’s shirt pocket. Ever more versatile electronics-based products are creating a transformation in our lives and culture.

For example:

- The pocket calculator has single-handedly eliminated the slide rule and the desktop calculator.
- The introduction of electronic word processing has virtually eliminated the electric typewriter, which eliminated the manual typewriter.
- The CD player, based on laser diodes, high-resolution analog-to-digital converters, digital signal processors, and microprocessors, has virtually eliminated the production and sale of vinyl records (current revival here).
- Microprocessors have filtered so far into cars that electronics training is essential for car mechanics.
- Personal electronic information managers, such as the “Boss” and the “Wizard,” are quickly causing the daytimer pocket calendar to disappear.
- Cellular phone architecture, support electronics networks, and small, portable phones are changing our travel habits and ability to stay in touch.
- The introduction of the FAX machine has nearly eliminated courier service companies.
- The microcomputer and laser printer created the desktop publishing industry and put the power of the press in the hands of the individual.
- The office copier eliminated the carbon copy and transformed the reproduction and dissemination of information.
- A second-generation revolution will reduce the volume of printed paper in favor of electronically transmitted, stored, and displayed information, possibly replacing books, magazines, and newspapers.
• Personal computers revolutionized word processing, database management, and video and graphical presentation.
• Computer-aided design and simulation is replacing hand-drawn blueprints and three-dimensional models.
• SMART cards are allowing 24-hour access to personal information (accounts, medical records, etc.)
• Digital cameras and imaging are allowing instant video imaging, photo-enhancement, and previewing.
• The proliferation of the handheld video camera has made possible the ability to record spontaneous crimes and unique events, later showing these films to an audience of millions in some cases.
• The prevalence of faxes and e-mail has toppled governments because proliferation of information by individuals cannot be suppressed.
• The Information Superhighway has placed the resources of every great library in the world, in the hands of any individual, anywhere.
• The Web is also an enabler for personal investment as access previously available only to large companies and investment firms is now available to any individual with a PC and modem.

It is these products and others like them that ultimately are fueling the advance of IC technology by setting our expectations higher and higher, which in turn drives the advancement of packaging technology.

It is a given that IC devices will continue their steady advances of:

• Smaller features
• Shorter intrinsic gate delays
• Larger die size
• More functions per chip
• More power generated
• Higher bandwidth signals
• Lower cost per function

There will be a few IC advances that will relieve some of the increased packaging burden, such as:

• Lower rail voltages for CMOS devices yielding lower power dissipation.
• Massively parallel architectures that, at some integration level, will use an interconnect network closer to a bussed architecture than a random architecture, requiring less interconnect density.

The significant advances that will be part of the future roadmaps of packaging technology in the next five years will address the major limitations packaging imposes, illustrated in Figure 13-1.
All of the packaging features on the future roadmap and the associated benefits have been reviewed in this book. In addition, there are intriguing possibilities in both the packaging technologies and the information processing technologies visible on the horizon, that may significantly impact our lives.

The packaging technologies with potential impact are related to:

- Getting around the high resistance of fine lines
- 3D packaging
- New materials with lower dielectric constant and photo-sensitive properties
- Frost MOS
- Enhanced thermal management

The information technologies with potential impact are:

- IC integration enabling system on a chip
- Quantum computing
- Molecular computing

<table>
<thead>
<tr>
<th>Packaging Challenge</th>
<th>Packaging Roadmap Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>I/O off the chip</td>
<td>area-array bonding</td>
</tr>
<tr>
<td>Switching noise</td>
<td>low bonding parasitics</td>
</tr>
<tr>
<td></td>
<td>integrated decoupling caps</td>
</tr>
<tr>
<td>Higher bandwidth</td>
<td>controlled impedance interconnect</td>
</tr>
<tr>
<td>Wiring delay</td>
<td>higher packaging efficiency</td>
</tr>
<tr>
<td></td>
<td>higher interconnect density</td>
</tr>
<tr>
<td></td>
<td>higher via density</td>
</tr>
<tr>
<td>Smaller from factors</td>
<td>lower dielectric constant materials</td>
</tr>
<tr>
<td></td>
<td>3-D packaging</td>
</tr>
<tr>
<td>Power dissipation</td>
<td>higher packaging efficiency with BGA, CSP, DCA and MCMs</td>
</tr>
<tr>
<td></td>
<td>higher interconnect density substrate</td>
</tr>
<tr>
<td></td>
<td>lower voltage rails</td>
</tr>
<tr>
<td></td>
<td>heat spreaders</td>
</tr>
<tr>
<td></td>
<td>water cooling</td>
</tr>
</tbody>
</table>

Figure 13-1. Packaging Technology Advances
HIGH-DENSITY INTERCONNECTS

Ultimate Limit to Fine Lines

As the number of I/O increases and chips are brought closer together using either CSP or as bare dice, the interconnect density requirement will increase. One way of satisfying this increasing requirement is by using finer and finer lines with tighter via pitches. IC fabrication technology is capable of one micron and below linewidths.

As the linewidths of metal interconnects decrease, the penalty that is paid is higher resistance, which contributes to an interconnect-related RC delay time. In addition to the dielectric-related time of flight delay, there is an RC delay that scales with the square of the interconnect length, as discussed in Chapter 11.

There will not be an infinite trend toward ever finer linewidths. There is a fundamental limit to the finest line that has a reasonable wiring delay. This limit has been reached for submicron chip design. Though the propagation delay for the gates decreases as the feature size gets smaller, the interconnect RC delay increases. Lower than about 0.7 microns interconnect delay (using aluminum wire and SiO₂ dielectric) begins to dominate. This is illustrated in Figure 13-2. This means care in the layout and routing on the chip is essential to obtain the highest performance.

![Figure 13-2. Wiring (Interconnect) Delay Versus Gate Delay](image_url)
To get the most performance out of the interconnect, the intrinsic interconnect RC should be kept a small fraction of the time of flight from dielectric delay. Consider a best case example using a square cross section of copper with a width, \( w \), in microns. Using a dielectric such as polyimide or SiO\(_2\), the longest length for a trace, \( L_{\text{max}} \), in inches, for which the intrinsic interconnect RC delay is less than the dielectric-related time of flight delay, is roughly:

\[
L_{\text{max}} < \frac{w^2}{10}
\]

With copper for example, a line six inches long should be fabricated with a width of at least 7.7 microns to avoid an excessive wiring delay. For a linewidth of three microns, a trace one inch long will suffer a total wiring delay of 310psec, equal to twice its dielectric-related time of flight of 155psec. Three microns may be the practical limit for linewidths and thickness for thin-film multilayer interconnect substrates.

How can we break through this linewidth limitation, both for on chip and global system wiring? Today there are three possible paths:

- Operating at cryogenic temperatures
- A hierarchy of linewidths and technologies for different trace lengths
- The use of superconductors

Cryogenics

The resistivities of copper and aluminum decrease with lower temperature. This is basically because mobility and electron scattering decrease at lower temperatures. The bulk resistivities of aluminum, copper, and several other metals versus temperature are shown in Figure 13-3. From room temperature (298\(^\circ\)K) to 77\(^\circ\)K, the resistivities of both copper and aluminum decrease by about a factor of 10.

This means that a copper trace could be 10 times longer at cryogenic temperatures than at room temperature before it contributes to a doubling of the interconnect wiring delay. Likewise, traces can be a factor of three more narrow and thinner at cryogenic temperatures and have the same interconnect RC penalty as at room temperature.

At 77\(^\circ\)K, copper traces one micron in width and thickness could be practically used for interconnects of one inch or less in length.
Mixed Thin Film, Cofired, and PWB

Another approach to minimizing the detrimental effect of high-resistance interconnects is to use a mixture of narrow lines for the short traces, which are most numerous, and wider lines, with consequent thicker dielectric, for the longer traces. Though the linewidth and metal thickness determines the resistance of the line, the characteristic impedance of the line is set by the linewidth, dielectric thickness, and the dielectric constant.

If the linewidth increases to allow lower resistance, the dielectric thickness must increase to keep the characteristic impedance at roughly 50 Ohms. If anything, the target value of characteristic impedance is likely to increase, requiring a further increase in the dielectric thickness. As linewidths increase and dielectric thickness increases, at some point different via generation methods might have to be employed. This means mixing interconnect process technologies.
In a large multichip module most of the interconnect will be between nearby chips, requiring short traces. These can be of the finest linewidths, possibly down to three microns. The longer traces, such as clock lines, can be routed in layers of wider trace widths. This is effectively what is done by NEC in the SX-X chip.

As shown in Figure 13-4, there are three interconnect technologies used with different linewidths and length scales. Multilayer thin-film interconnect, with 25 micron linewidth on a 75 micron pitch contains most of the signal lines in four signal layers. These layers and their three ground planes are built on top of a multilayer, cofired base used mainly for power, ground, and clock distribution. Two of these substrates are mounted into an 18-layer printed circuit board.

![Multilayer Substrate Cross Section](image)

**Figure 13-4. Multilayer Substrate Cross Section**

The combination of ultra fine lines from thin-film technology, fine lines in BUM technology and the low-resistance lines in the conventional PWB technology, described in Chapter 11, is also a powerful combination offering the entire spectrum of highest possible interconnect density, yet lowest possible interconnect delay.
Superconductors

The discovery in 1986 of ceramic materials that become superconductors at 30°K heralded a new era for electronics. Just one year later, the liquid nitrogen barrier at 77°K was beaten by a YBa2Cu3O compound showing a transition temperature at 90°K. These steady advances are diagrammed in Figure 13-5.

![Figure 13-5. Sharply Rising Critical Temperatures in Superconductors Stem From the Cuprate Materials](image)

If interconnect traces are less than six inches long in a module, and linewidths of eight microns and comparable thickness are used, then replacing the copper with a room temperature superconductor will not have a significant impact on the wiring delay of that module.

If this module were cooled to 77°K, to allow the use of a ceramic superconductor, copper traces could be as narrow as three microns before they would begin to appreciably affect the wiring delay. This sets the upper limit to the useful regime for liquid nitrogen superconducting interconnects.
When the interconnect density requires linewidths of less than three microns on a pitch of five microns, a superconductor may offer better performance than copper. Two pairs of signal lines at five microns pitch and 50 percent efficiency would offer an interconnect density of about 10,000in/in².

At these geometries, current density becomes a major concern. Assuming a 50 ohm line and a one volt signal level, the current through the trace is 20mA. With a three micron width and thickness, the cross sectional area is $9 \times 10^{-8} \text{cm}^2$. This corresponds to a current density of over 200,000A/cm².

In the lab, typical critical current densities before experimental superconducting materials are quenched to normal conductivity are about 10,000A/cm². However, highly crystalline thin films have been fabricated that allow sustained current densities of as high as 4,000,000A/cm². There is nothing fundamentally preventing the use of superconductors for interconnects.

It is likely that if superconductors play a role, it will only be at the very extreme end when interconnect densities of more than 10,000in/in² are needed for modules that are over three inches on a side, and are randomly interconnected.

Of course, in addition to the development of a reliable superconductor material and deposition process, a compatible dielectric material with low dielectric constant is essential.

**3D INTERCONNECTS**

Once the packaging efficiency has increased and the chips are very close together and the interconnect density is available to perform all the routing, it may still be possible to decrease the wiring delay by adjusting the topology to route in three dimensions rather than two. This will decrease the interconnect path lengths. There are a number of technologies that are aiming toward 3D packaging.

**Extension Of Conventional Approaches**

The most straight-forward methods of implementing 3D packaging do not require invention of new technologies. Some clever implementations have been available for years. An example of an approach of small building block modular daughter cards that snap together for embedded applications from Ampro is shown in Figure 13-6. The vertical connections are made with standard pins and connectors. The pins are on the bottom side of each modular board. On the top side of each modular board are sockets. There is a standard bus that each board uses. Rapid prototyping is possible by just snapping boards together.
Another approach from Advanced Electronics Packaging uses DIMM sized modules with memory and logic devices that plug vertically into a motherboard. An example is shown in Figure 13-7. These daughter boards in turn plug into a backplane.

![Figure 13-6. Snap-Together Mini-Modules From Ampro](source)

Conventional peripheral packages can be stacked one on top of the other, as long as there is a bus architecture between all the chips, and there is a means of affecting the vertical interconnect. The Chiprack approach uses a standard chip carrier format, with approximately 88 to 132 leads. Each
layer of the stack is a frame with one or more chips. These frames are stacked on top of each other in a special connector that makes contact with the one above it. An example of a frame and a 6 chip stack composed of a 68070 processor, ASIC, ROM and RAM, is shown in Figure 13-8.

Dense-Pac offers memory modules with stacks of custom ceramic single chip layers interconnected along their edges with solder. Examples of the company’s module is shown in Figure 13-9. Each stackable leadless chip carrier (SLCC) starts as a bare die, mounted into a custom ceramic leadless package with castillations around the edges. The chip is assembled into the package and wire-bonded in place. The lid is sealed and the unit is tested.

In the typical SRAM applications, the enable line on each chip in a stack must be unique. This is programmed by cutting out links on the outside of the package before the SLCCs are stacked. The SLCCs are then solder dipped to connect the edges. This process flow is shown in Figure 13-10. The bottom footprint can be a PGA, gull wing or J lead, by changing the bottom package configuration.
Figure 13-9. Dense-Pac 3D Memory Modules

Figure 13-10. Dense-Stack Assembly Flow Stackable Leadless Chip Carrier (SLCC)
Vertical module stacking allows dramatic board area reduction. Figure 13-11 compares the board area needed to assemble 128 chips in various package styles. The standard SOIC form factor takes up about 65 square inches of board space. Sixteen, 8 high stacks would take up less than 10 square inches.

![Chart showing PCB area required for different package types]

Source: Dense-Pac Microsystems, Inc./ICE, "Roadmaps of Packaging Technology"

**Figure 13-11. Comparisons of PCB Space Required for a 16MB SRAM System for Different Package Types**

Staktek uses a similar method, but with plastic packaged chips. Each die is mounted in a molded plastic unit with leads out two sides, as shown in Figure 13-12. These can use a standard TSOP style or a customized part with half the thickness. These layers are then stacked, 4-8 chips high and held in place with a metal frame. The frame is then soldered to the package leads. With the ultraslim module, 512 chips can be assembled to a card less than 4 inches by 8 inches. An example of a double sided memory card used by Cray Research and assembled by Celestica is shown in Figure 13-13.

**Stacking Bare Die**

Methods pioneered by TI and Irvine Sensors involve creating a “cube” of functional silicon. An example of Irvine Sensors’ design is shown in Figure 13-14. So far, this technology has been applied to memory chips, which require low interconnect densities. A metallization layer is added to a wafer of standard dice to bring all the leads to one side of each die. The wafer is thinned to six mils, and the chips are diced.
Good dice are laminated together into a cube using epoxy. The edge with the leads is lapped down to expose the pads, which are connected and finally solder dipped to provide the vertical connections. The cube is soldered to a base plate that converts the side connections to a pad array. Solder balls on the base turn the cube into a BGA. An example of a short stack, a tall stack and the base are shown in Figure 13-15.

TI has developed a similar process. It uses a TAB-like leadframe deposited on top of the die and overhanging the edge, transforming the standard peripheral leads into edge tabs. Chips are glued together into a cube with the tabs sticking out, which are also soldered to a substrate. An example of an MCM using their memory cubes is shown in Figure 13-16.
Issues related to thermal management, manufacturing, and reliability still need to be worked out. However, the memory density possible is very considerable. A RAM module with a footprint of less than one square inch could hold over 32 die. If each one were a 64M DRAM, this module would store 256Mbytes of memory.
Stacking MCMs and Boards

TI has extended this notion to include the stacking of multichip modules. This is shown in Figure 13-17. Connections are brought out to the edge of each substrate and interconnected by a motherboard, which can be a thin-film multilayer substrate.

Button Board

A limitation to the above approach to 3D interconnection is the need to provide connections along the edge of the module in a single row only. This is simply a scaled-down version of the traditional approach of printed circuit cards plugged into a common backplane. The reason it has not been adopted by IBM, NEC, or DEC, for example, instead of one or two large planar printed circuit boards, is the limit in the number of connections available between boards.

One approach to provide a greater number of contacts between boards, while distributing the contacts over the area of the device rather than just around the periphery, is to use the “Button Contact.” It has been extensively used by TRW and is available commercially from Tecknit. An example of TRW’s board is shown in Figure 13-18.

Figure 13-16. TI Cube MCM
The basic element of a button contact is a short length of wire that has been randomly twisted and bent into the shape of a cylinder, typically 20 to 50mils in diameter, and twice as long as it is wide. The wire, from 0.5mil to 5mils in diameter, is composed of copper, beryllium-copper, tungsten, or stainless steel, which has been gold plated. Examples of the wire buttons are shown in Figure 13-19.

These buttons are inserted into insulated holes of slightly smaller diameter and act as conductive, compliant interconnects between flat surfaces. In the boards developed by TRW, the interconnects have been mated and de-mated hundreds of times with no effect on the signal integrity.

The use of button contacts may allow an area array of contacts between modules for a truly high-density, 3D interconnected structure. These contacts are used in BGA and LGA sockets as well.

NEW ELECTRONIC MATERIALS

The driving force of reducing the wiring delay of the interconnects is also influencing the introduction of lower dielectric constant materials. As mentioned in Chapter 11, glass-ceramics that have dielectric constants from 5.8 to 7.0 are beginning to appear as LTCC. These materials will be replacing the 90 to 95 percent alumina with dielectric constants over 10 that have been used for cofired ceramics for over 30 years.
For polymer-based printed wiring boards, fluoropolymer materials are being introduced. Teflon has the lowest dielectric constant of any solid, homogeneous material, at about 2.0. The same non-polar and low polarizability of its molecular structure that keeps its adhesion so low, also keeps its dielectric constant low.

This value of dielectric constant represents the basic polarizability of the electronic structure of the polymer chain. The polymer chains are so symmetrical, and the C-F bonds hold its electrons so tightly, that they do not easily contribute to either van der Waals forces with foreign atoms, or polarization from external fields.

As mentioned in Chapter 11, fluoropolymer PWB laminates are available from Gore, DuPont, and Rogers with fillers added to enhance their mechanical properties and manufacturability in a PWB process. The dielectric constant is increased up to approximately 2.9 depending on the filler material.
Up to recently, fluoropolymer materials have been suitable only for use as laminates, and limited to thickness greater than three mils. DuPont has recently announced a Teflon™ material that can be spin coated. Termed Teflon AF for “Amorphous Fluoropolymer,” it can be spun onto wafer substrates in micron-thick films. Patterning methods based on laser ablation or plasma etching as well as metal adhesion still need to be fully developed. Teflon AF’s very low dielectric constant of about 1.9 and ability to form thin films makes it an attractive material.

Another approach to low dielectric constant has been tried by nCHIP, based on structure rather than materials’ engineering. It has often been said that a microstrip transmission line is “faster” than a stripline structure. This is because the electric field of a microstrip extends into the air above the trace, and the contribution from the air lowers the effective dielectric constant of the trace.

In a stripline geometry, the electric field is fully contained within the dielectric and the effective dielectric constant is the bulk dielectric constant of the material.

nCHIP has applied this principle of fringe fields in air decreasing the effective dielectric constant in its “Isostrip,” or Isolated Microstrip Transmission lines. A microstrip structure is used with all the dielectric material not directly under the trace removed. This increases the field that is in air, reducing the effective dielectric constant. Figure 13-20 shows a cross section of an Isostrip trace. The effective dielectric constant is reduced to 1.9 using this configuration.
In addition to performance, as related to decreased wiring delay, another driving force on new materials is reduced manufacturing time and cost. Photosensitive polymers offer the possibility of eliminating half of the polymer processing steps. For example, the Proimid photosensitive polyimides from Ciba-Geigy Corp., have been used to fabricate thin-film multilayer structures with the vias created in the polyimide by direct imaging from a mask. This process is outlined in Figure 13-21.

**FROST-MOS: CYRO-COOLING OF CMOS**

In CMOS-based systems, a large source of critical net propagation delay is the intrinsic gate delay. This is continually being reduced as feature sizes are made smaller. In addition, it can be lowered by cooling the devices.

The other dominant source of propagation delay is the intrinsic RC delay of the interconnect. As pointed out in a previous section of this chapter, this can also be decreased by an order of magnitude by cooling the chip.

In Figure 13-22, the clock frequency versus temperature for a single-chip CMOS microprocessor is shown. This indicates that if the clock is running at 15MHz with a junction temperature of 125°C (398°K), cooling it down to 80°K will result in a clock frequency of 35MHz, a 2.3x improvement.

ETA applied this principle in the ETA-10 by immersing an entire CPU board containing over 200 CMOS gate arrays in liquid nitrogen dewar. Speed increases of from 2 to 5 over room temperature operation were reported.
On a smaller scale, a micro-refrigerator technology is available from MMR Technologies. An early prototype, shown in Figure 13-23, can extract one watt while keeping a chip at 82 K. This first prototype was built in a 40-pin ceramic DIP.

These refrigerators use the Joule-Thomson effect, in which the expansion of a gas extracts heat. Pressurized gas enters the module, and through the use of precisely shaped channels, low-pressure, chilled gas contacts the cold plate where the chip is mounted.

Figure 13-21. Polyimide Pattern Generation Processes
Figure 13-22. Speed of CMOS Clock as a Function of Temperature

Figure 13-23. Single-Chip Cryogenic Cooler
Superconductor Technologies Inc. (STI), has introduced a cooler for chilled CMOS module applications. Called the Radically Accelerated Cold Electronic (RACE) refrigeration system, it is 5 in x 5 in x 12 inches and handles a heat load of 30 watts at a temperature of −25°C. At this temperature, it is estimated there is a 60% increase in frequency for core limited CMOS. An example of the RACE 30 is shown in Figure 13-24. The chilled CPU is actually in the external chamber, connected to the rest of the computer by a ribbon cable.

The RACE cooler operates based on the Sterling cycle of compressing and expanding a gas. There are three versions, the RACE 30 cools to −25°C, the RACE 60 cools to −55°C and the RACE 77k can pump 4 watts at 77K. An example of the RACE 77k is shown in Figure 13-25.

Through the use of such localized coolers, the critical chips in a CMOS-based CPU could be cooled to decrease the gate delay of a current generation technology to that of a few generations in the future. This could mean a few years’ jump in the time to market for a new CMOS-based workstation or server.
THERMAL MANAGEMENT

At the very high end, both the power densities generated by the individual devices and the resulting power densities generated by the CPU boards are increasing. For example, the MCA III currently has a power density at its surface of greater than 200W/in². The CPU of a large mainframe or super computer can contain about 1 million gates. If each of these gates dissipates about 3mW of power, the cooling for the CPU must allow the extraction of 3kW of power.

Two methods will see growing application: the use of high thermal conductivity heat spreaders to transport the heat from the die itself to a cooling medium. The second method uses a cooling fluid, possibly water or some other chemical, to remove heat from the system.

In the DEC VAX 9000, the heat spreader is a block of copper to which the dice are directly attached, using electrically insulating, but thermally conducting, epoxy, filled with diamond powder. In the Motorola/Siemens multichip modules, the backsides of the dice are pressed to a polished and anodized aluminum plate. In both cases the metal plates spread the heat to a larger area for extraction — using air for the VAX 9000, water for the Motorola/Siemens module.

Another attractive heat spreader material is aluminum nitride. When processed with a minimum oxygen content, it can have a thermal conductivity approaching aluminum metal, 2.3W/cm·°C. Figure 13-26 shows the thermal conductivity dependence on the oxygen content. Processing of aluminum nitride will require careful control.

Aluminum nitride offers the special advantage of allowing multilayer interconnect when cofired with tungsten, for example. Coors and WR Grace are developing an aluminum-nitride/tungsten cofired process. Toshiba has built a number of multilayer, multichip modules with cofired aluminum nitride, using the substrate as the integrated heat spreader.

The further advantage of aluminum nitride is the good TCE match with silicon and the possibility of allowing a true hermetic seal to enclose the chips.

As pointed out in Chapter 6, when air is used as the thermal extraction medium, it offers a best case thermal resistance of 100°C/watt for a 1in³/sec flow. To extract 3kW of power with a 25°C case-to-ambient temperature drop would require a flow rate of 12,000in³/sec, which is approximately 7ft³/sec. In moving through an aperture 6 inches wide by 4 inches high, the air velocity would have to be about 30 miles per hour!
As power densities increase, water cooling will tax the engineering heroics less than air cooling, and may result in a lower cost system. Extracting 3kW of heat under the same conditions as above with water cooling requires only 4in³/sec of water. The Motorola/Siemens module has integrated water cooling channels in the aluminum heat spreader. NTT has proposed integrating cooling channels in the cofired ceramic interconnect substrate, as shown in Figure 13-27.

As power densities increase, water cooling will tax the engineering heroics less than air cooling, and may result in a lower cost system. Extracting 3kW of heat under the same conditions as above with water cooling requires only 4in³/sec of water. The Motorola/Siemens module has integrated water cooling channels in the aluminum heat spreader. NTT has proposed integrating cooling channels in the cofired ceramic interconnect substrate, as shown in Figure 13-27.

Figure 13-26. Thermal Conductivity Dependence on Oxygen Impurity in AlN

Figure 13-27. NTT Water-Cooled Cofired Aluminum Nitride Substrate Packaging Concept
Yet another application for cryo-cooling of CMOS is for enhanced heat extraction. Liquid nitrogen can be used as a cooling medium by its heat of vaporization. If a chip or module is immersed in liquid nitrogen and if the nitrogen fluid is flowing over the surface, to minimize the boiling film, large heat fluxes can be removed and the surface can be kept at low temperature. Figure 13-28 is an example of the heat fluxes that can be removed in a simple arrangement of a stream of liquid nitrogen, cooled to 68°K, 9 degrees below its boiling point, impinging on a heated chip. 80 watts/cm² have been demonstrated, with a rise in temperature of only 20 degrees above the nitrogen temperature of 68°K. This approach may easily remove over 200 watts/cm² while keeping the junction temperature well below 0°C.

![Figure 13-28. Constant Flow Rate Liquid Nitrogen Jet Impingement With 8.8°K Subcooling](image)

**SILICON ADVANCEMENTS: SYSTEM-ON-A-CHIP**

A high-end CPU may need at most only 10 million transistors. By the year 2000, 100 million transistor chips will be fabricated. What will the extra gates be used for? One proposal is to integrate all the features of the system on a single chip. This includes RAM, ROM, EPROM, disk and peripheral controllers, multimedia, 3D graphics, communications, and DSP functions.

The first trend in this direction is the introduction of DRAM with on-chip microprocessors. In this low-end case, the chips are dominated by memory, with a relatively small area taken up by the logic functions. In this case, it is really a microprocessor embedded on a DRAM. An example of the architecture and implementation of the M32R/D from Mitsubishi is shown in Figure 13-29. It consists of 2Mbits of DRAM and a 32 bit microprocessor. It targets embedded applications.
IBM has converted an 8 chip MCM with over 1,000 pins into a single chip, with 15 million transistors, using 0.25 micron technology. It is packaged into a ceramic BGA with a 42.5mm body size and dissipates 30 watts. This part is used in the RS/6000 workstation as well as the SP-2 supercomputer.

The transition toward systems-on-a-chip will be driven by the availability of 100 million transistor ASICs. This will radically change the way systems are designed. There are specialized companies that only supply DRAMs or only supply MPEG decoders, or only supply peripheral chip sets.

In the conventional system design approach, each company could specialize and provide a clean interface to the system designer in a single chip package. The design of MCMs caused complications when more than one companies’ chips were to be included. Who took responsibility if a chip was bad, or didn’t meet specifications?

In designing a system-on-a-chip, who will supply the design information for all the system components, or cores, that will go on the single chip? Recognizing that there are new relationships that must be forged in the industry to support the technical features that will be soon available, a new industry alliance has emerged, called the Virtual Socket Interface, VSI. This group is composed of

Source: Electronic Products/ICE, “Roadmaps of Packaging Technology”
representatives from the EDA design community, the semiconductor vendors, the core design vendors, and the system end users. A list of the current membership is shown in Figure 13-30. The excitement over the birth of the system-on-a-chip and the role that the VSI alliance may play in the industry has been put to verse in Figure 13-31.

<table>
<thead>
<tr>
<th>EDA Vendors</th>
<th>Intellectual-Property Vendors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspec</td>
<td>Advancel Logic</td>
</tr>
<tr>
<td>Cadence</td>
<td>ARM</td>
</tr>
<tr>
<td>Compass</td>
<td>CompCore Multimedia</td>
</tr>
<tr>
<td>Mentor</td>
<td>DSP Group</td>
</tr>
<tr>
<td>Summit</td>
<td>Excellent Design</td>
</tr>
<tr>
<td>Synopsys</td>
<td>Integrated Silicon Systems</td>
</tr>
<tr>
<td>Viewlogic</td>
<td>iReady</td>
</tr>
<tr>
<td>VLSI Libraries</td>
<td>Object Oriented Hardware</td>
</tr>
<tr>
<td>Semiconductor Vendors</td>
<td>Phoenix</td>
</tr>
<tr>
<td>Actel</td>
<td>PrairieComm</td>
</tr>
<tr>
<td>Alcatel-Mietec</td>
<td>Rapid*</td>
</tr>
<tr>
<td>Altera</td>
<td>Sand</td>
</tr>
<tr>
<td>Cirrus Logic</td>
<td>Sun/Java</td>
</tr>
<tr>
<td>Fujitsu</td>
<td>3Soft</td>
</tr>
<tr>
<td>Hitachi</td>
<td>VLSI Libraries</td>
</tr>
<tr>
<td>National</td>
<td>Systems Vendors</td>
</tr>
<tr>
<td>NEC</td>
<td>Alcatel</td>
</tr>
<tr>
<td>Texas Instruments</td>
<td>Silicon Graphics</td>
</tr>
<tr>
<td>Toshiba</td>
<td>Sun</td>
</tr>
<tr>
<td>TSMC</td>
<td>Sony</td>
</tr>
<tr>
<td>VLSI Technology</td>
<td></td>
</tr>
<tr>
<td>Xilinx</td>
<td></td>
</tr>
</tbody>
</table>

Source: EE Times/ICE, “Roadmaps of Packaging Technology”

Figure 13-30. Virtual Socket Interface Member Companies

The introduction of the system-on-a-chip will have significant impact on packaging as well. As the functional integration increases, there is a point where the I/O needs of the chip start to decrease. When the entire system is integrated on one chip, the I/O demands should eventually be limited to power and ground distribution and a few narrow busses for input and output, such as video drivers, keyboards and speaker drivers. Packaging may be relegated to the role of physically embedding the systems-on-a-chip into the system, rather than focusing on interconnecting higher and higher pin count devices. The new technologies that are being driven by the current demands for smaller package form factor, higher via density and higher interconnect density may have a limited future.
LIMITS TO SILICON: QUANTUM COMPUTING

When the physical size of a transistor gets below 0.1 micron, it contains less than a million atoms. The behavior of junctions changes radically in this size regime. After all, the statistical fluctuations of carrier concentrations from a doping level of 1ppm will result in less than half of all the channels having no minority atoms!

In this size scale, with structures about 1,000 atoms wide, macroscopic quantum effects dominate. The vision slowly emerging of what a transistor will look like below 0.1 micron feature size is that of a quantum dot. A single, isolated island of conductor will act as a confining trap for a single electron. The conductor might be a GaAlAs heterostructure. One possible design of a quantum dot logic element is shown in Figure 13-32.
A quantum dot that contains one bit of information has been termed a quantum bit or “qubit”. In the qubit illustrated above, there are five stable locations for a free electron to be located: the four corners, and the middle.

In an array of adjacent quantum dots, each would be loaded with two electrons. Though no electron would propagate between quantum dots, the location of the two electrons in each dot is dependent on the proximity of the electron in adjacent dots. When the state is changed at one end of the chain of quantum dots, other dots down the line might change as a result, depending on the logical operation. For example, Figure 13-33 illustrates a stable pattern of a three input gate. The majority value of the inputs is transmitted to the output.

In Figure 13-34, an example of how this structure could be turned into a programmable gate is illustrated. The principles of quantum computing are just now being developed in anticipation for the period 10 years from now when the limits to conventional semiconductor operation will be reached and quantum computing might be required.
The limits of quantum computing may be when individual molecules are encoded with information. The Buckminsterfullerene molecule, C_{60}, offers an intriguing possible molecule to manipulate. IBM has demonstrated the positioning of Buckyballs on a copper surface. Rows of 10 Buckyballs were lined up in grooves on a copper plate, using a scanning tunneling microscope tip. By careful positioning, the balls mimic an abacus. However, this abacus, shown in Figure 13-35, is only 10 x 40nm on a side. No one has proposed “abacus computing” on the molecular level. However, this feat demonstrates what might be possible.

![Figure 13-34. A Possible Qubit Programmable Logic Element](image)

The limits of quantum computing may be when individual molecules are encoded with information. The Buckminsterfullerene molecule, C_{60}, offers an intriguing possible molecule to manipulate. IBM has demonstrated the positioning of Buckyballs on a copper surface. Rows of 10 Buckyballs were lined up in grooves on a copper plate, using a scanning tunneling microscope tip. By careful positioning, the balls mimic an abacus. However, this abacus, shown in Figure 13-35, is only 10 x 40nm on a side. No one has proposed “abacus computing” on the molecular level. However, this feat demonstrates what might be possible.

![Figure 13-35. A Molecular Abacus Using Buckyballs](image)
BEYOND QUANTUM COMPUTING: DNA COMPUTING

Molecular computing in a biological cell proceeds continuously and at a furious pace. One cell, approximately 5 microns on a side, has approximately 1 trillion molecules. Each of these molecules has combinatory rules of interaction. Harnessing the computing power of biologically active molecules has been demonstrated by computer scientist Leonard Adleman of the University of Southern California in Los Angeles. He used the DNA molecule as the computing engine to solve a specific problem.

DNA is composed of sequential nucleotide segments, each one strung across a rung in a double helix pattern. An example of a 10 segment DNA molecule is shown in Figure 13-36. There are four basic nucleotides, symbolized by the letters A, T, C and G. Information is encoded in a strand of DNA by the order of the nucleotides. The problem Adleman was able to solve is often called the traveling salesman problem:

There are seven cities, and a total of 14 one way flights between the cities. What is the path a salesman can take to travel starting at a specific city and ending at a specific city while passing though each of the other cities only once.

Aldeman encoded each of the cities as a 20 segment nucleotide. He then created a complimentary segment of nucleotides that would attach to each one of the 20 nucleotide cities. To form the 14 one way flights, he synthesized DNA sequences that had the last 10 nucleotides of the origin city and the first 10 segments of the destination city.

Figure 13-36. Short Segment of a DNA Molecule, With 10 Nucleotide Segments
Next, he added pinches of these 21 powders: 14 one way flight segments and 7 city compliment segments, to a test tube of water, where they dissolved. With all the molecules free to interact with each other, the city complements would bind to the ends of the one way link segments, forming all possible chains.

The information encoding was in the nucleotide segments for each city and one way link. The combination rules were encoded by the choice of the nucleotide sequences and the nature of the A, T, C, and G chemistry. The information processing was the chemical reaction that created all possible one-way-link-city-one-way-link polymer chains that the segments and encoding rules allowed. The interpretation of the information relied on finding the resulting chain that was exactly (7 cities) x (20 nucleotides per city) segments long, had the right start-and-end segments and did not have repetitious segments. These were found by electrophoresis.

The resulting DNA segment chain with the path through the 7 cities encoded in it was then read out. The answer is shown in Figure 13-37. This shows the 7 cities, the 14 one way links, and the complete path proceeding from city number 0 to city number 6.

It is not clear what the future of DNA computing is, even if there is a future, or whether this one example of DNA computing is a novelty. However, the interest has started molecular geneticists talking to computer scientists. Many outcomes may result from this association.

FINAL COMMENTS

For as long as machines require ever more powerful microprocessors, technology will steadily advance and packaging technology will continue to play catch-up.
The myriad of packaging choices available today will be a part of the portfolio of solutions for all electronic products for years to come. After all, DIPs, which are over 30-years old, are still today an integral part of many electronic products.

The next generation solutions will arise from natural engineering extensions and creative combinations of new designs, new materials, and new processes for the ultimate goal. As summarized in Figure 13-38, this goal involves packing higher performance in a smaller volume at lower cost, or in other words, developing systems that are “…faster, smaller, less expensive, NOW.”

“...We’ve finally paid up on this $300,000 baby, and now they’ve come up with one that can do the same job for $79.95.”

© 1982 by Sidney Harris - ‘What’s So Funny About Computers?’, William Kaufmann, Inc.

Figure 13-38. Next-Generation Computing Patterns