Alternative Packages Emerge for Processors

TAB Used for MicroSPARC; Power2 Opts for Multichip Module

by Linley Gwennap

*This is the fourth in a series of articles on integrated-circuit manufacturing. The first covered basic manufacturing issues (see **070705.PDF**) , while the second discussed cost (see **071004.PDF**). Last issue, we described PQFP, PGA, and BGA packages (see **071203.PDF**). A future article will compare different vendors' manufacturing capabilities.*

While most microprocessors today are sold in either plastic quad flat packs (PQFP) or ceramic pin-grid arrays (CPGA), vendors are beginning to examine other, more radical alternatives. Last issue, we discussed the ball-grid array (BGA), which is used in IBM's RIOS chip set. IBM is also expanding the use of multichip modules (MCM) with its new Power2 design *(see **071301.PDF**)*. Texas Instruments is shipping its MicroSPARC chip exclusively in a tape automated bonding (TAB) package. In some applications, processors have eliminated packaging entirely by using chip-on-board (COB) attachment.

Tape Shrinks Package Size

Tape automated bonding has been around almost as long as integrated circuits. It is commonly used in digital watches and other products where package size and height are critical. TAB replaces the traditional package with a single piece of tape, or film, that holds the die to the board and provides the interconnect, as shown in Figure 1. Thus, the height of the package is not much more than the thickness of the die.

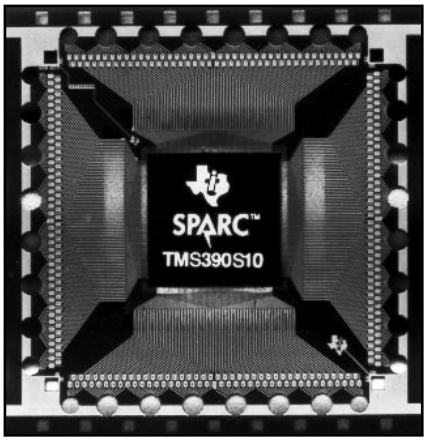
The figure shows the traces on the tape that carry signals from pads on the die to larger pads at the edge of the tape. Unlike Scotch tape, bonding tape is not adhesive; the tape pads are usually soldered to the board. The size of the tape is generally limited by the size and number of bonding pads; manufacturing processes that can handle very fine pitches can cut the tape close to the edge of the die itself.

Beside reducing package size, TAB can be less expensive than plastic packaging, particularly for high lead counts. The trace-on-film tape is easily mass-produced. The die is soldered to the tape, eliminating the lengthy wire-bond process. No plastic molding is required, although some designs use a simple plastic cap for protection. Like the BGA design, TAB also eliminates the problem of bent pins.

Electrically, TAB is a mixed bag. On the plus side, eliminating the bond wires creates a smoother signal path with fewer signal reflections. The single signal plane, however, has frequency limitations similar to PQFP; standard TAB devices max out around 40 MHz. Some vendors have experimented with adding a ground plane on the other side of the tape to reduce the inductance of high-frequency signals; this technique can significantly improve the operating frequency but adds to the cost.

Eliminating the bond wire has the additional merit of reducing pad pitch on the die itself. PQFP dice generally use a 130-micron pitch, while PGA designs, with their lower volumes, can push as low as 100 microns. With TAB, a pitch of 75 microns can be achieved in production. The smaller pad pitch offers some die size reduction, particularly for dice that are pad-limited *(see **071004.PDF**)*.

One downside of standard TAB devices is that, unlike PGAs, they cannot be socketed and later removed from the board. Socketing allows an expensive processor to be salvaged if a board fails. To combat this issue, Hewlett-Packard has developed a detachable TAB package (DTAB). As shown in Figure 2, the DTAB package places a metal cover over the tape. The cover screws into the board, aligning the contacts between the tape and



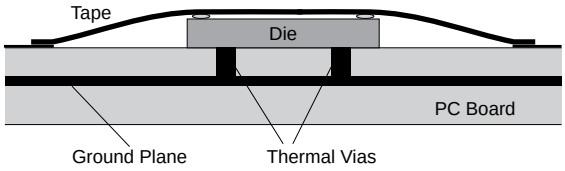


Figure 1. The photograph of a MicroSPARC chip in a TAB package shows the traces that carry signals from the die. The cutaway view (below) shows how the tape attaches to the die and board, and the heat dissipation path for MicroSPARC.

the board, and pressing them firmly together.

DTAB can also handle more heat than a standard TAB package, since a heat sink can be mounted on the metal cover. Without the cover, the die/tape combination does not have the strength to support a heavy heat sink. The problem with DTAB is that the metal cover and other hardware greatly increase cost, making it nearly as expensive as a CPGA. Macrotek (Dortmund, Germany) will use DTAB for its PowerPC system-logic chip set *(see **071002.PDF**)*.

TAB Cuts the Cost of MicroSPARC

Despite its advantages, TAB has not been widely used in computer systems. Many vendors have tried it in the past and had problems working out the bugs. TAB has frequency and power limitations, common in plastic packages, which make it unsuitable for most high-performance processors.

Only one popular processor, MicroSPARC, is available in a TAB package. The TI chip is well-suited for standard TAB due to its moderate pin count (288 leads) and frequency. Although the chip runs at 50 MHz internally, the major external interfaces connect to relatively slow DRAM and to the SBus, which runs at 25 MHz. MicroSPARC does not support an external cache or other interfaces that would have to match the core CPU speed. It does require a single 100-MHz clock signal; this one input is easily shielded by surrounding ground signals to reduce switching noise, eliminating the expense of an added ground plane.

MicroSPARC's power dissipation, a moderate 4.5 W maximum, is still too high for a simple TAB design. Sun engineers solved the cooling problem by mounting the die with its back against the board, as shown in Figure 1. Thermal vias in the board (which are simply normal vias used to conduct heat instead of electricity) draw heat from the die and dissipate it through the board's internal ground planes. This design allows the board itself to act as a large copper heat sink at no extra cost to the system.

Although TAB allows for smaller pad pitches, the MicroSPARC designers chose not to take advantage of this option, since their chip was not pad-limited. MicroSPARC uses a pad pitch of 180 microns to simplify attaching the die to the tape. The lead pitch (at the outer edge of the tape) is a tight 0.25 mm (250 microns) to keep the package small.

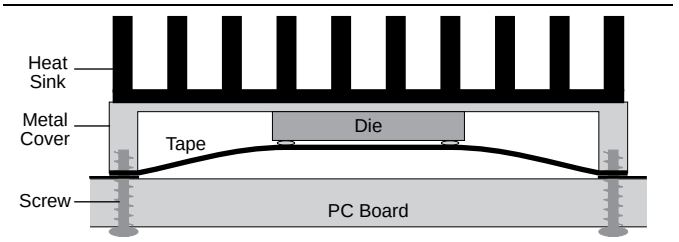


Figure 2. Hewlett-Packard's DTAB technology caps a standard TAB package with a metal cover that screws into the board. An optional heat sink can attach to the cover.

Most major board manufacturers now have the capability to handle TAB, and TI says that a surprisingly large number of customers are using MicroSPARC in its TAB package. Despite TI's success, other processor vendors say they have no immediate plans to market their chips in a TAB package. Large vendors such as Intel and Motorola can supply TAB packages as a special case.

One of the reasons that TAB has not been rapidly accepted is the competition provided by various direct-attach strategies, also called chip on board. As implied by the name, these techniques involve attaching the die directly to the board, eliminating the package entirely. The die is typically attached using bond wires, but some vendors have experimented with flip-chip attachment.

Flip-Chip Eliminates Bond Wires

Flip-chip, also known as solder bump, was developed by IBM as a high-density, high-performance packaging technology for its mainframes. IBM calls this technique controlled-collapse chip connection (C4). Like TAB, flip-chip eliminates wire bonds; pads on the chip are soldered directly to the board. This technique can be used with most types of chips.

True C4 technology involves a unique style of chip design that uses an extra layer of metal to place the pads on top of the circuitry instead of in a pad ring around the edge of the die. Arranging the pads in an array on a separate layer can reduce the die size by up to 15%, significantly improving yield *(see **071004.PDF**)*. For a pad-limited design, the die size reduction can be even greater. The added cost of the extra metal layer, however, may negate any gains in die manufacturing cost.

C4 requires a different set of chip layout tools, a die manufacturing process that can handle the added metal layer, and a different manufacturing process. IBM is the only microprocessor vendor to put all of these items into place to produce C4 chips. In fact, because IBM has optimized its processes for C4, most of its chips use this technology. The PowerPC 601, for example, is mounted in its CQFP using C4 instead of bond wires.

Although IBM claims otherwise, flip-chip attachment appears to be more costly than wire-bonding. Both Intel (under license from IBM) and Motorola have experimented with flip-chip but rejected it due to cost.

Another problem with flip-chip designs occurs if there is a mismatch between the thermal expansion of the die and the substrate. As the die heats up, it will expand slightly; the underlying substrate, which is made of a different material, may expand faster. The attachment points are die bonding pads only 100 microns (0.1 mm) wide, so even a small amount of expansion can tear open a contact, particularly for large dice. One solution to this problem is to use a substrate with a similar coefficient of

expansion as the silicon die, but this prevents the use of a standard, low-cost PC board.

Chip On Board: The Packageless Package

Wire-bonded COB does not suffer from the thermal mismatch problem, as the flexible wires have enough slack to handle moderate expansion. It can be more expensive than flip-chip attachments, however, and requires that the board vendor perform an operation that is normally done by the chip vendor.

COB has been used for years to reduce both height and board size. A standard 120-pin PQFP, for example, consumes about 800 mm^2 of board area while a 120-pad die would take less than 40 mm^2. A critical factor in this example is the ability to route 120 PC-board traces into such a small area. Very small devices, such as digital timers, might consist of a single IC mounted directly on a board connected to the buttons and display.

Connecting the signal pads directly to the board provides an ideal electrical environment with minimal lead lengths, few reflection points, and extensive ground planes in the PC board. Thermal expansion issues generally prevent the use of high-power die, however.

Chip-on-board is potentially the lowest-cost approach to packaging. It gets rid of the package entirely, but some of the assembly costs remain: the cost of wire-bonding is simply shifted to the board manufacturer. A big problem today is obtaining unpackaged die.

Few processor vendors are selling bare die today, and even fewer have bare die on their price lists. Bare die are most common for embedded applications. Motorola, for example, reports that several customers are buying its low-end microcontrollers as bare die for COB applications, but these sales are handled as special deals. Intel has said that it will sell unpackaged 386 processors in a program aimed at embedded customers.

Known-Good Dice Improve the Odds

The major barrier to widespread adoption of COB is the issue of known-good dice, known in packaging circles as the KGD problem. The normal chip manufacturing process *(see **070705.PDF**)* includes an initial wafer test that detects 90–95% of the bad chips, and a final packaged-part test that finds the rest. If a part passes the wafer test but fails the final test, the manufacturer has wasted the cost of packaging the die; this cost is not huge, particularly for a plastic package.

If a chip vendor sells unpackaged die using its standard wafer test, the buyer will receive 5–10% bad parts. These parts will not be detected until the board is assembled and tested, requiring time-consuming rework or the complete board to be discarded. Multiple unpackaged die on a single board increase the chances of a failure. This situation is unacceptable to most manufacturers.

For COB applications, most buyers demand chips that have been thoroughly tested and are known to be good. The expense of testing a part without its package reduces the cost savings of the packageless solution; Motorola says that most of its KGD sales are at about the same price as packaged parts.

There are several approaches to achieving KGD. One is to increase the accuracy of the wafer test. Most wafer testers are not capable of testing dice at full speed and high temperature; many of the chips that fail the final test have defects that are only detectable under these conditions. Wafer test can be improved by using more expensive equipment that tests at speed and temperature. Using a more lengthy test can also increase accuracy. These changes, however, significantly increase the cost of the test.

Another method is to use a temporary package to test each die. For most chips, the cost of wire-bonding the die just for testing would negate any cost savings from COB. Some vendors have found that TAB can be used effectively for this purpose, as it is relatively easy to attach (and detach) the die. A TAB-packaged die can be connected to a system and quickly tested at a full range of frequencies and temperatures.

A third alternative is to use a pressure attachment to place the die in a test fixture without bonding. Micro-Module Systems (Cupertino, Calif.) supplies custom thin-film chip carriers to connect a die to a standard test fixture for this purpose. This method requires careful handling of the die and accurate, consistent alignment of the die and the test fixture.

So far, none of these alternatives has been proven to deliver KGD in volume at a cost less than that of a packaged die. Until this milestone is reached, COB will be limited to applications that are willing to accept higher manufacturing costs to reduce board height and area. Many vendors are working to solve the KGD problem, however, and new solutions may eventually emerge.

Multichip Modules

Multichip modules are a single package containing more than one die. With the expanding use of COB, however, the definition of an MCM can become blurred; two die mounted directly to a small PC board can be considered an MCM or simply a COB design.

Unlike COB, MCMs have historically been used to improve electrical performance (and thus clock speeds). The most extreme example is IBM's thermal conduction module (TCM), used for years in high-end mainframes. A typical TCM contains 121 dice connected by 63 routing layers to 2772 pins; it dissipates a turkey-roasting 2300 W through a water-cooled heatsink the size of a toaster. The cost of such a device must be staggering, but it probably increases the speed of IBM's mainframes by 30–40% over more conventional packaging.

Most MCMs are much less complex, combining only

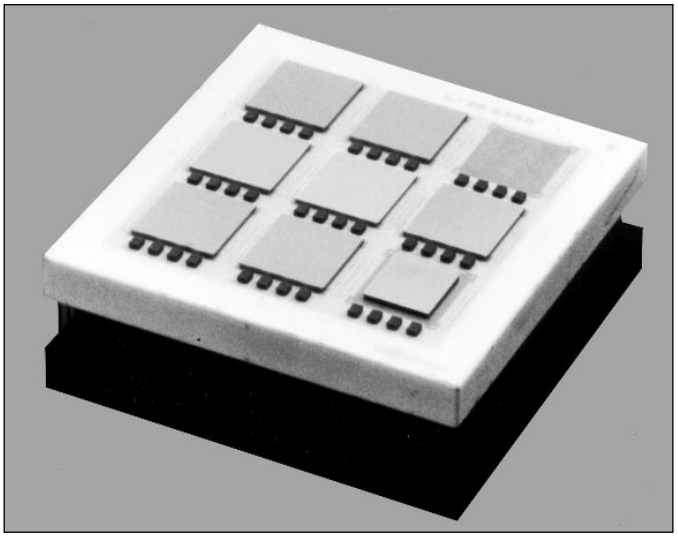


Figure 3. IBM's Power2 module includes eight chips mounted on a single package. The substrate under the dice contains interchip routing, and PGA pins are on the bottom of the package.

a few chips. The most common today are ceramic MCMs, known as MCM-C. These are simple extensions of other ceramic packages, modified to support a handful of dice. The ceramic offers the advantages of high heat dissipation and nearly unlimited signal routing, since additional signal layers can always be added. These advantages are helpful in high-performance applications.

MCM-C packages have been widely used in military, aerospace, and supercomputer systems that are willing to accept their high costs in return for better performance and increased density. These packages typically connect to the board using a PGA interface, but MicroModule has discussed using a ball-grid array or QFP instead.

A less expensive approach is to use a plastic laminate substrate similar to a standard PC board. This technique, designated MCM-L, has become more popular in the past few years; Integrated Circuit Engineering (ICE) estimates the current MCM-L market to be $25 million but expects it to grow to $100 million in 1995.

While MCM-L is a less expensive packaging technique than, for example, a ceramic PGA, it cannot handle the high-power and high-pin-count chips that typically use a PGA. Thus, MCM-L is generally used to combine chips previously packaged in PQFPs to achieve a smaller physical size. This technology is quite similar to chip-on-board and suffers from many of the same costs and limitations. One advantage over COB is that critical chips can be combined into a single module and tested; if there is a bad die, only the module need be discarded rather than an entire board.

A third MCM technology, called MCM-D, uses a substrate built by depositing dielectric and conductive materials onto a base using thin-film technology. While this technique is even more expensive than MCM-C, it provides optimal performance due to vastly improved routing density; 10-micron line widths and 15-micron vias are roughly one-tenth the size of similar features on an MCM-C or MCM-L.

One example of this technology is a five-chip module developed by nChip containing a complete SPARC 601 processor (see MPR 5/30/90, p. 8). The module, used in Tadpole's SPARCbooks, uses a silicon substrate to avoid a thermal mismatch with the chips. Ross Technology, which marketed the module, claims that better electrical characteristics improved the yield of the parts, resulting in manufacturing costs comparable to discrete packages. Ross plans to package the next generation of its hyperSPARC chip set in a similar multichip module.

IBM Deploys MCM in PCs, Workstations

MCMs are beginning to appear in several computer memory systems. IBM has incorporated a four-chip memory module in some PS/2 systems. LSI Logic supplies a four-chip module to a major workstation vendor. MicroModule has announced two- and four-chip SRAM modules for Pentium and other high-speed processors. Memory chips are ideal for MCMs because many of the signals are common among multiple chips, reducing the number of pins needed for the package.

Both IBM and LSI are using MCM-C, while MicroModule is using MCM-D. Despite the added expense of the thin-film process, MicroModule expects that the cost of its SRAM modules will be competitive with monolithic (single-chip) devices, since it is combining four chips from a relatively mature CMOS process while single devices of the same capacity and speed are using a more expensive leading-edge process.

IBM's Power2 is the first workstation processor to take advantage of MCM packaging. As shown in Figure 3, eight dice are mounted in the package using C4 flip-chip bonding, and the package connects to the board using a 736-pin PGA. The ceramic substrate contains 20 signal routing layers (44 layers total) that include over 90 meters of signal traces. IBM claims that the MCM offers a 20% increase in clock speed over traditional discrete packaging. The company admits that the MCM is more expensive than packaging the parts individually but would not specify the cost increase.

Other processors may take advantage of MCM packaging in the future. MIPS Technologies is investigating placing its two-chip TFP processor into an MCM. It is widely rumored that Intel's next-generation P6 processor will be a multichip design packaged in an MCM.

As CPU clock speeds continue to increase, MCMs offer a method of placing a large memory in close proximity to the processor, increasing memory bandwidth. MCM-C and MCM-D are currently expensive but offer the performance needed for high-speed processors.

At the low end, MCM-L could become popular for

portable systems in which a small cost premium is willingly traded for minimum size. PCMCIA cards might take advantage of MCM-L; TQFPs, BGAs, and COB technologies will also be wrestling over this design space.

Cost of Packaging Varies Widely

Figure 4 depicts the estimated cost per pin of various packaging alternatives. At this time, it appears that the least expensive package for most chips is still a PQFP, primarily due to the high volume and available infrastructure for the plastic package. Single-layer TAB and bare KGD have similar costs, however, and can be less costly than a PQFP at high pin counts.

The graph shows that standard pricing for PQFPs ranges from about 2¢ per pin for small packages (and even lower for the smallest PQFPs) to about 6¢ per pin for the largest packages. Above 208 pins, the per-pin cost increases dramatically because the larger body sizes require more material for the package and lead frame. A 304-pin PQFP costs about $18 in volume. At these pin counts, TAB packages and KGD have lower costs, since they don't have to pay for the larger body size.

The cost of plastic BGAs is slightly higher than that of PQFPs for most package sizes. At the highest pin counts, however, the plastic BGA becomes more cost-effective than the PQFP because its body size does not increase as dramatically. By arranging the contacts in an array rather than around the perimeter of the package, the BGA is better able to handle pin counts above 208.

Ceramic BGAs take another step upward in price, ranging from about 3¢ to 7¢ per pin. These packages can accommodate far more pins than the plastic packages, however. Ceramic QFPs are yet more expensive, and ceramic PGAs are the most expensive of all the packages shown here. The high cost of the ceramic package is mainly due to the substrate area and number of layers.

The cost of the MQUAD package is kept artificially high for low-pin-count designs due to patent issues and a decreased market supply *(see **071203.PDF**)*. For higher pin counts, however, these packages become quite cost-effective.

MCMs are not shown on the graph because there are more variables involved in estimating their cost. For a single-chip package, the substrate carries signals only from the pads to the pins; the interconnect complexity is directly proportional to the number of pins. In an MCM, the substrate must also route signals between the chips. Thus, one must understand the complexity of the inter-chip routing to estimate the cost.

The cost model of an MCM is also slightly different. If wafer-tested dice are bonded in the MCM using a standard process, then the packaging cost will be similar to other packages but the yield will be much lower, since any single bad die causes the module to fail. Some modules can be reworked to remove bad die, but this cost must also be considered. Another way to build MCMs is to use known-good die, but then the cost of KGD testing must be added.

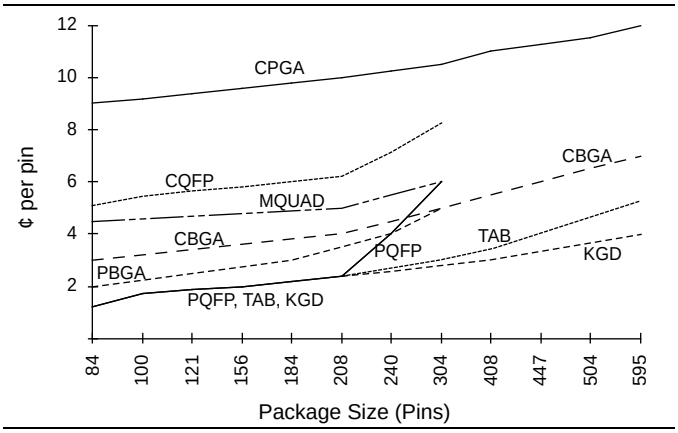


Figure 4. Per-pin costs increase with pin count for all package types, as greater numbers of pins require either a larger physical package or a tighter (and more expensive) pin pitch.

Many Packaging Options Available

Today's microprocessor designer faces a wide array of packaging options. For microcontrollers and low-end microprocessors, the plastic quad flat pack is the most popular choice, although a few vendors opt for minor variations such as the MQUAD and PLCC packages. These packages are suitable for low-cost, high-volume manufacturing and provide adequate performance for these products.

Where board area and overall physical size are limited, designers should consider options such as plastic BGAs, TAB, and chip-on-board attachment. Of these, COB is the most compact and may be the best solution if known-good die are available and affordable. A successful COB product requires the chip manufacturer to be willing and able to provide known-good die, and the board manufacturer to perform the required wire bonding or flip-chip attachment.

High-performance processors are typically limited by their need for high pin counts, high frequency signals, and high power dissipation. Ceramic PGAs are the package of choice, but ceramic BGAs can offer similar features at a lower cost. For maximum performance, MCM-C or MCM-D can give a 20–30% boost in speed but at a significant cost premium.

As demonstrated by nChip, MicroSPARC, RIOS, and Power2, the status quo of PQFP and CPGA is not always the best solution. Innovative package designs can deliver an edge in the on-going price/performance competition. ♦

The author thanks John Gallagher of LSI Logic, Jim Reinhart of Motorola, and Brian Matas of ICE for providing much of the background for these articles and helping verify the information presented here.

