CPU Vendors Deploy Half-Micron Processes

MIPS Processor Vendors Lead in Race to Shrink IC Feature Sizes

by Linley Gwennap

*This is the final article in our series on integrated-circuit manufacturing. The first covered basic manufacturing issues (see **070705.PDF**) while the second discussed cost (see **071004.PDF**). The third (see **071203.PDF**) and fourth (see **071304.PDF**) articles covered IC packaging. This article compares the leading-edge IC processes for microprocessors.*

Every couple of years, the wheel of IC process technology advances another notch. In 1992, a wave of new high-end processors using 0.8-micron CMOS technology appeared from HP, Digital, Intel, TI, and others. This year, the next wave is hitting the beach, bringing a new generation of 0.6-micron processors. With all chip vendors using similar manufacturing equipment, it is difficult for one to get too far ahead of the rest.

Upon closer examination, however, we see that the wheel of progress does not turn so smoothly or uniformly. Some vendors push technology aggressively to achieve better results while others, for a variety of reasons, lag behind. A plethora of "half-generation" and "shrink" processes further confuse the issue.

Although vendors typically use a single number to label their processes, a variety of parameters are important to speed, cost, and density. Transistor speed is most dependent on the effective gate length (L_{EFF}) and the gate oxide thickness (T_{OX}), while density primarily depends on the number and pitch of the metal layers *(see **070705.PDF**)*. Drawn gate length (L_{DRAWN}), the number of interconnect layers, and support for bipolar transistors affect the wafer processing cost *(see **071004.PDF**)*. Of course, a variety of other factors—such as circuit design, packaging, and microarchitecture—influence the chip-level cost and performance.

Differences in IC manufacturing are at least as important as design advantages in determining the performance and cost of a microprocessor. A chip vendor with a better process has an edge in delivering faster, cheaper CPUs to the market. As Table 1 shows, the MIPS chip vendors—IDT, NEC, and Toshiba—are in the lead with drawn gate lengths approaching 0.4 micron. Behemoths IBM and Intel have processes that exceed the MIPS process in certain critical dimensions and offer greater circuit density. HP, as usual, achieves good performance by marching to the beat of a different drum.

R4400 Drives Tight Geometries

A few years ago, many analysts criticized the MIPS technology model for its inability to take advantage of leading IC processes. Because MIPS had to ensure that five different vendors could build its processor designs, the argument went, its designs accepted the weakest features of the chip vendors. The R4000, for example, uses only two metal layers, since a few of the vendors could not handle a third layer. Other vendors, in contrast, tune their designs to their own in-house processes.

The company, now MIPS Technologies, Inc. (MTI), changed the model for the R4400. This time, MTI chose a set of aggressive design rules and told the vendors that they had to find a way to meet them—or they wouldn't be able to build the chip. As a result, the five vendors shrank to three, but those three vendors are using a leading-edge IC process.

There are actually two processes used to build the R4400, one for the 5-V version and the other for 3.3-V chips. The 3.3-V parts use gates drawn at 0.4 micron, with L_{EFF} of 0.3 micron, according to IDT. NEC says that the gate oxide thickness is about 90 angstroms, typical for a low-voltage process.

These fine-pitch devices, particularly with the thin oxides, cannot tolerate 5-V signals; the 5-V parts use a less aggressive process. NEC builds these chips with a 0.5-micron drawn gate and a T_{OX} of about 140 Å. The higher voltage mostly compensates for the slower speed of the larger transistors, but yield at 150 MHz is better for the 3.3-V parts.

The metal-1 pitch for the 3.3-V process is just 1.3 microns, better than most of the vendors in Table 1. The MIPS vendors have not placed as much emphasis on the other metal layers, however, reducing overall circuit density compared with IBM, Intel, and TI. In fact, while most vendors support four metal layers in their half-micron processes, the MIPS vendors have only three.

The MIPS processors make up for this in part by using a four-transistor (4T) SRAM cell instead of the 6T cell more common in microprocessors; the process supports a second polysilicon layer for the 4T cell. IDT, NEC, and Toshiba all adapted their logic processes from bulk SRAM processes that include two poly layers, so they have no problems building the MIPS chips.

These vendors plan to shrink this process later this year. The new 0.35-micron process will be used for the 200-MHz R4400, due to ship in 3Q94, and the forthcoming T5 processor. In addition to shrinking the gates, the new process adds a fourth metal layer, which the R4400 will not use but the T5 will. NEC will take advantage of this change to move to 200-mm wafers from the 150-mm wafers used by all three MIPS vendors today; it is not clear whether IDT and Toshiba will follow suit.

Vendor	Digital	Digital	Fujtisu	HP	IBM*	IBM*	IBM	IDT**	Intel	TI	TI
Process Name	CMOS-4S	CMOS-5	CS-50	CMOS-14	CMOS-5L	CMOS-5S	CMOS-5X	CMOS7+	"0.6µ"	EPIC-2BE	EPIC-3
Example Product	21064	21064A	µSparc-2	PA-7200	PPC 603	PPC 620	PPC 601+	R4400	P54C	S'Sparc	MVP
First Production	3Q93	3Q94	1Q94	4Q94	3Q93	4Q94	4Q94	3Q93	1Q94	2Q94	3Q94
Supply Voltage	3.3 V	3.3 V	3.3 V	4.4 V	3.3 V	3.3 V	2.5 V	3.3 V	3.3 V	4.8 V	3.3 V
BiCMOS?	no	no	no	no	no	no	no	no	yes	yes	optional
Gate Length (drawn)	0.68 µm	0.50 µm	0.50 µm	0.55 µm	0.65 µm	0.50 µm	0.50 µm	0.40 µm	0.50 µm	0.60 µm	0.55 µm
Gate Length (effective)	0.46 µm	0.37 µm	0.45 µm	0.37 µm	0.46 µm	0.36 µm	0.25 µm	0.30 µm	0.37 µm	0.50 µm	0.47 µm
Gate Oxide Thickness	100 Å	90 Å	110 Å	120 Å	135 Å	90 Å	70 Å	90 Å†	80 Å	120 Å	90 Å
No. of Metal Layers	3 metal	4 metal	3–4 metal	3 metal	5 metal	5 metal	5 metal	3 metal	4 metal	3 metal	3–4 metal
Local Interconnect?	yes	yes	no	no	no	yes	yes	no	no	yes	no
Stacked Vias?	no	no	no	no	yes	yes	yes	no	no	yes	yes
M1 contacted pitch	2.3 µm	1.5 µm	2.1 µm	1.8 µm	1.4 µm	1.4 µm	1.2 µm	1.3 µm	1.4 µm	2.0 µm	1.8 µm
M2 contacted pitch	2.6 µm	1.8 µm	2.1 µm	1.8 µm	1.8 µm	1.8 µm	1.8 µm	1.7 µm	1.7 µm	2.0 µm	1.8 µm
M3 contacted pitch	7.5 µm	5.0 µm	2.1 µm	2.4 µm	1.8 µm	1.8 µm	1.8 µm	2.0 µm	1.7 µm	2.6 µm	2.4 µm
M4 contacted pitch	—	5.0 µm	210 µm	—	1.8 µm	1.8 µm	1.8 µm	—	3.5 µm	—	4.0 µm
Routing Index	10.8 µm^2	4.9 µm^2	4.4 µm^2	4.3 µm^2	3.1 µm^2	2.7 µm^2	2.5 µm^2	3.4 µm^2	2.9 µm^2	4.3 µm^2	4.1 µm^2
Wafer Cost Index†	1.4	1.8	1.1–1.2	1.2	1.4	1.8	1.8	1.6	1.5	1.4	1.2–1.6

Table 1. Many processor vendors delivered chips at 0.7-micron and below in 1993, and all the major CPU vendors are moving to 0.6-micron or below in 1994. Routing index is the product of the second and third metal-layer pitches, with adjustments for local interconnect and stacked vias. Cost index indicates relative cost to process a wafer with this process. Smaller is better for both indices. *Motorola uses a similar process for its PowerPC chips. **NEC and Toshiba use similar processes for their MIPS chips. (Source: vendors except †MPR estimates)

Digital Designs for Speed

While the R4400 gears up for 200-MHz operation, Digital has been shipping 200-MHz Alpha processors since last summer, despite using an L_{EFF} 50% larger than that of IDT's process. Both the MIPS and Alpha processors use a similar superpipelined design, so how does Digital achieve its speed advantage?

Digital's circuit design and process design work hand in hand to deliver fast clock rates. The Alpha chip is well-known for its giant clock driver, a pair of 250,000-micron transistors *(see **060301.PDF**)*. This clock signal is distributed throughout the chip using wide (7.5-micron) metal-3 traces. These traces, which are also thicker than usual, have a resistance of about 1 Ω/mm, a tenth that of typical metal; thus, clock traces from the central driver to the edge of the chip have a resistance of only a few ohms, minimizing clock skew.

It is impractical to route buses and other signals using the heavy metal-3 traces. To compensate for this problem and improve circuit density, Digital implements a thin layer of titanium nitride as a local interconnect. This layer can be used only for short traces due to its high resistivity, but it is particularly useful in reducing SRAM cell size. The local interconnect adds an extra process step but is less expensive than a full metal layer.

Digital was concerned about large transient currents caused by rapidly switching signals on its 128-bit data bus. These currents can temporarily raise V_{SS} from its normal 0 V, possibly triggering some transistors and generating phantom signals. The low resistance of metal-3 helps route V_{SS} throughout the chip with minimal voltage deviation, but Digital needed to reduce the resistance at the bonding pads as well. Its process engineers added a special mask layer over the pad ring that minimizes the resistance of power and ground signals.

The company needed an optional ECL I/O mode for compatibility with its VAX 9000 systems; this mode is also used in the Cray T3D system. These ECL outputs must provide 50-Ω matched impedance. For this purpose, the 21064 includes on-chip precision resistors, which require an additional mask layer plus an ion-implant step. These resistors maintain 50 Ω within ±10% across the chip's full temperature and voltage specification.

Digital takes the unusual approach of using a cobalt salicide instead of a titanium salicide. Salicide covers (or "straps") the diffusion areas, reducing series resistance through the transistors. Cobalt has a lower electrical resistance than titanium, so the salicide layer can be thinner; this allows shallower junctions and thus faster transistors. Although many vendors find cobalt difficult to work with, Digital has been using it for several years and claims to have mastered its use.

The company continues to evolve its manufacturing process and plans to deploy a new, 0.5-micron process this summer. Digital says that this process, which it dubs CMOS-5, will boost the clock rate of the 21064 to 275 MHz. This process will also be used for the next-generation 21164 (EV-5) CPU. Both metal-3 and the new metal-4 are wide, low-resistance layers, giving this process the worst routing index in Table 1.

As a result, the 275-MHz 21064A requires more than twice the die area of the PPC 601+, despite having the same transistor count and the same cache size. Digital's added mask layers and extra process steps result in a high wafer cost that, combined with a large die area, further increases chip cost. These issues show the cost of high performance.

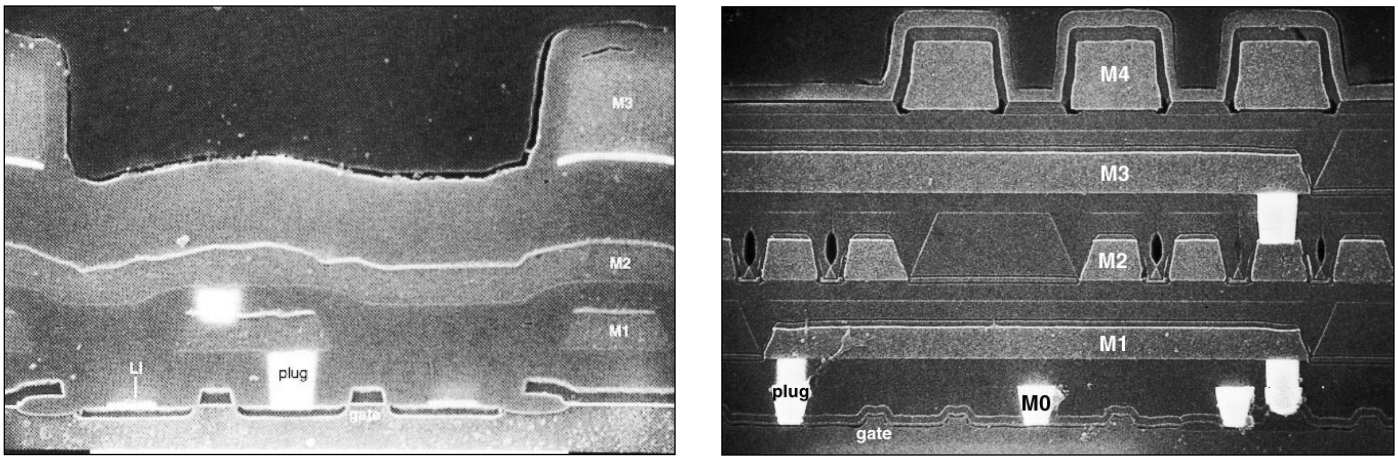

Figure 1. The photomicrograph at left shows a cross section of Digital's 21064 die, which uses limited planarization to support three metal layers plus local interconnect (LI). Note the thickness of the top metal layer, which reduces its resistance. At right, IBM's PowerPC 601 uses extensive planarization to support four thin metal layers plus local interconnect (M0), increasing circuit density but adding cost. (Source: ICE)

PowerPC Uses Family of Processes

IBM uses a variety of processes for its PowerPC chips. The original 601 uses a process called CMOS-4S, a "shrink" version of a 0.8-micron process originally used for 4M DRAMs. IBM's newer processes are derived from a 16M DRAM process it calls CMOS-5. The company labels all of these processes 0.5-micron, but the gate lengths vary significantly.

The first of these new processes to reach production was CMOS-5L, a logic version of CMOS-5. The drawn gate length of this process is 0.65 microns. The gate oxide thickness (135 Å) is typical for DRAMs but lags the other processes in Table 1. The 5L process, used for the PowerPC 603 and 604 chips *(see **080501.PDF**)*, still achieves respectable clock rates. Motorola will use a similar process to build its own 603 and 604 processors.

The 5L process includes five full metal layers, all with fairly tight pitches, as Table 1 shows. When IBM builds this process, it includes a top layer, sometimes erroneously called an added metal layer, consisting of barrier-interface metal and solder balls for flip-chip attachment *(see **071304.PDF**)*. The 601, built in a similar process, has no pad ring at all *(see **070602.PDF**)*. Motorola, however, builds its chips using standard periphery wire bonding. Thus, the 603 and 604 designs include a pad ring, but IBM continues to use the added process step for its flip-chip technology. Because this step eliminates the need for wire bonding, IBM claims that flip-chip costs less for dice with lots of pads (such as most microprocessors).

The 5L process has been in production since last summer for IBM's 45-MHz RSC processor *(see **0709MSB.PDF**)*. IBM is developing two additional processes based on 5L. The first, CMOS-5S, shrinks the drawn gate to 0.5 microns and lowers T_{OX} to 90 Å, making that parameter competitive with other half-micron processes. The metal pitches for 5S remain the same as in 5L, but the new process adds a local interconnect layer similar to Digital's. (CMOS-4S also includes this layer.) IBM's local interconnect, or metal-zero, is a thick tungsten layer similar to its tungsten via plugs. Both IBM and Motorola will build the PowerPC 620 using the 5S process.

CMOS-5X takes a further step to accelerate the transistors. Although the drawn gate is still 0.5 micron, L_{EFF} is just 0.25 micron and the gate oxide is 70 Å. The metal-1 pitch is trimmed to 1.2 microns, although the remainder of the interconnect layers remain the same. The three changed parameters are better than any other vendor expects to achieve this year. The thinner gate oxide requires a supply voltage of 2.5 V, somewhat reducing the performance gain. IBM will build the 100-MHz PowerPC 601 *(see **080502.PDF**)* in this process.

Extensive planarization is required to support this tremendous number of metal layers. Without this leveling process, upper layers would break going over hill and dale. Figure 1 compares a photomicrograph of a Digital 21064 with a cross section of a PowerPC 601. The Digital chip is only partially planarized, with metal-3 showing the resulting bumpiness. In the IBM chip, each layer is carefully planarized using chemical-mechanical polishing (CMP), creating a smooth base for the next layer. This polishing allows the company to use as many layers of metal as needed and reduces the probability of reliability problems from thin, stretched metal traces.

CMP also increases the cost of the process, according to technology analysts at Integrated Circuit Engineering (ICE) in Scottsdale, Ariz., as do the added metal layers. As Table 1 shows, CMOS-5S and 5X are among the most expensive wafers to process. The extensive interconnect, however, results in the best circuit density of these processes. This high density reduces die size and thus increases the number of good chips per wafer, compensating for the higher wafer cost.

The PowerPC vendors, unlike most other RISC vendors, use 200-mm wafers for these chips. The larger wafers slightly decrease overall manufacturing cost and greatly increase capacity, since each one holds nearly twice as many chips as a 150-mm wafer. The bigger wafers, combined with the large throughput of IBM's Burlington (Vt.) fab and Motorola's Austin (Texas) facility, give the PowerPC vendors much greater manufacturing capacity than other RISC chip makers. Using a fraction of their capacity, these two fabs can produce a total of more than one million PowerPC chips in 1994.

Intel Aims for Density and Speed

Intel, the king of capacity, expects to fabricate 40 million x86 processors this year. The company's new flagship factory in Leixslip (Ireland) has the ability to run 5,000 wafers per week, using the same 200-mm size as IBM and Motorola. (This factory alone could produce nearly 40 million 486 chips per year.) To increase the number of chips per wafer and cut its production costs, Intel chose to focus on improving the circuit density of its processors.

For most microprocessors, density is controlled by the routing, so Intel includes four metal layers in its new process. The metal-1 pitch is just 1.4 microns, and the metal-2 and metal-3 pitches best all others in Table 1. Intel's routing index is better than any other process without a local interconnect layer. As a result, the 0.6-micron Pentium is 44% smaller than the original 0.8-micron, three-layer-metal version.

While focusing on circuit density, Intel did not lose sight of performance enhancements. Although the company conservatively quotes its process at 0.6 micron, the gate length and other parameters are comparable to a 0.5-micron process. In fact, Intel's effective gate length and gate oxide thickness, the two parameters that most affect transistor speed, are among the leaders. These parameters help boost the clock rate of the new Pentium by about 60% compared with the 0.8-micron version, despite cutting the voltage from 5 V to 3.3 V.

Intel is also the only vendor to exclusively use BiCMOS for its half-micron processors. Although bipolar transistors increase the speed of certain types of circuits, they also add steps, and thus cost, to the process. The fourth metal layer and tight metal pitches also increase costs compared with other half-micron processes. The table shows the Intel process to be moderately expensive, but the high circuit density helps reduce die area and increase the output of chips.

Although Intel's use of BiCMOS makes a strong argument in its favor, sources at both Intel and Texas Instruments (another BiCMOS vendor) indicate that the advantages of bipolar logic may not last to future process generations. At 0.35-micron or so, bipolar has very little performance advantage over CMOS and may not justify the cost of the extra bipolar steps. Most vendors are looking at pure CMOS processes at this level and beyond.

SPARC Vendors Keep Pace

Texas Instruments has been building the SuperSparc processor using a 10% shrink of its 0.8-micron EPIC-2B process. The company recently began to produce a second shrink, called EPIC-2BE, that reduces L_{DRAWN} to 0.6 microns. The new shrink greatly increases SuperSparc's yield at 60 MHz, a 50% speed increase over the original 0.8-micron version. The 2BE process will also be used for SuperSparc-2 later this year.

At the same time, TI is preparing EPIC-3, its next-generation process, for production this summer. This process will be used for a variety of chips, including the MVP DSP *(see **080405.PDF**)*, UltraSparc, and future x86 products. The first to reach production will be the MVP, due in the fourth quarter.

The EPIC-3 process can be configured slightly differently as needed. For example, the MVP uses three metal layers and no bipolar transistors, keeping costs low for the price-sensitive DSP market. UltraSparc, on the other hand, eschews cost tradeoffs in favor of performance. This processor will take advantage of the fourth layer of metal and bipolar circuits, driving wafer cost to the high end of the range shown in Table 1.

Unlike most vendors, TI implements stacked vias in its processes. This allows, for example, a connection between metal-1 and metal-2 to sit right on top of a contact between metal-1 and poly. This tactic requires careful alignment and extensive planarization to establish a flat base for the contacts, but it adds little cost. Both TI and IBM stack vias to improve circuit density slightly.

The other major SPARC processor maker is Fujitsu, which builds MicroSparc-2 (MS-2) in its CS-50 process. Fujitsu quotes the drawn gate length of this process as 0.50 micron, but L_{EFF} and T_{OX} are not competitive with most other half-micron processes. Thus, the transistor speed will not be as fast as some processes.

Fujitsu's metal pitches are also wider than other half-micron processes. The process includes three 2.1-micron layers and an optional fourth layer, with a vast 210-micron pitch, for routing power or for solder-bump pads (used in TAB packaging). These values give CS-50 a poor routing index. Fujitsu notes that the wider traces are less prone to defects, and the relative simplicity of the process makes it the least expensive in Table 1.

To support 5-V I/O, the company implements dual gate oxide thicknesses on a single die: thinner oxides speed 3.3-V gates while a thicker oxide supports the power of 5-V transistors. This duality simplifies the design of 5-V I/O signals in MS-2, which connects directly to DRAM and a 5-V SBus, and serves Fujitsu's gate-array customers. Unfortunately, each type of gate oxide requires a separate process step.

Fujitsu is also building the SPARC processor from HaL Computer; the company has not announced details of this chip, but it will probably use the CS-50 process. Fujitsu recently purchased Ross Technologies, which had been using a Cypress foundry to build its HyperSparc chip in a 0.65-micron process *(see **071502.PDF**)*. Ross is also porting its processor design to CS-50 and hopes to increase the clock rate from 66 to 100 MHz.

Fujitsu plans to improve performance by the end of this year with its CS-55 process, which uses the same metal layers as CS-50 but shrinks the drawn gate to 0.4 microns. A next-generation process, due in mid-1995, further reduces L_{DRAWN} to 0.35 microns and includes four full metal layers with 1.45-micron pitches. MicroSparc-3 will probably use this process.

HP Takes Big Step Forward

Hewlett-Packard has been lagging other CPU makers in process development. Although HP rolled out a 0.8-micron process in 1992 along with many other CPU vendors, other chip makers advanced a half-step in 1993, reaching 0.7-micron or so. HP has been stuck at 0.8-micron for its PA-7100, 7150, and 7100LC processors, although its fab has been able to shrink L_{EFF} somewhat.

The company plans a leap directly to a 0.55-micron process, CMOS-14, around the end of this year, making it the last major vendor to reach this level. This process, originally expected to be in production last summer *(see **0702MSB.PDF**)*, has slipped. The first CPU to take advantage of it is the PA-7200 *(see **080302.PDF**)*.

Because HP has the luxury of designing its chips primarily for its own systems, it can take a nonstandard approach. For example, the new CMOS-14 process operates at 4.4 V using a 120-Å gate oxide. This voltage allows the chip to operate at a higher clock rate than it would at 3.3 V, and the thicker gate oxide is easier to build than the thinner oxides used by other vendors.

HP's 0.8-micron process was recognized for its tight metal pitches, but the company has not made a comparable advance in the newer process. The 1.8-micron pitch for metal-1 trails most comparable processes. HP has chosen three metal layers for the PA-7200, reducing wafer cost but decreasing circuit density compared with the four layers used by other processor vendors.

By sticking to three metal layers and a relatively large T_{OX}, HP keeps its costs down but does not offer the density of other processes. The higher supply voltage keeps the thicker gate oxide from dragging down the transistor speed.

Vendors Emphasize Different Priorities

Despite the apparent similarity of the half-micron IC processes used by these vendors, the details of these processes are as different as the processors built with them. Although the companies are working with similar equipment and the same laws of physics, each uses its equipment to achieve different goals. Ideally, close cooperation between the process designer and the CPU designer results in the two working together to reach the same goals.

Intel ships an order of magnitude more processor chips than all the RISC vendors combined. Certainly, these prodigious shipments depend on several large factories, but Intel also increases its output by using larger wafers and smaller chips. The company is not afraid to invest heavily in equipment and up-front design to reduce the size of its processors. Even with its focus on output, Intel also keeps its transistor speed competitive with that of performance-focused CPU vendors.

IDT, NEC, and Toshiba have striven to increase transistor speed by pushing down both L_{DRAWN} and L_{EFF}. Digital, while not ignoring L_{EFF}, takes a more holistic approach to improving circuit speed via process changes. Both of these approaches increase wafer cost without a corresponding increase in circuit density.

IBM and Motorola have taken a less aggressive approach with the CMOS-5L process, particularly with metal pitches and oxide thickness. This process delivers good performance and high density with a moderate wafer cost. The more advanced PowerPC processes are designed for higher performance with associated increases in wafer cost, although these costs are mitigated by further improvements in circuit density.

HP also takes a less aggressive approach and gets greater cost savings by sticking with three metal layers. HP regains some transistor speed by boosting the operating voltage to 4.4 V.

TI's EPIC-2 process lagged the industry somewhat and produced poor results for SuperSparc. Two shrinks have improved the situation, but TI is just now beginning to produce enough 60-MHz chips for Sun's mainline workstation business. EPIC-3 promises to bring TI up to the state of the art; perhaps it will come on line faster than TI's previous efforts. Sun has carefully evaluated both EPIC-3 and Fujitsu's 0.5-micron process and believes that TI's process offers better performance but that Fujitsu delivers lower cost.

One thing is clear: it is difficult to optimize a process for both cost and performance. Sun has the luxury of working with different vendors for its high-performance and low-cost processors. Digital, on the other hand, may have problems converting its performance-at-any-cost foundry to deliver low-cost processors; perhaps Mitsubishi can help in this regard once it begins producing Alpha parts. IBM, Intel, and the MIPS vendors produce enough chips to support multiple fabs with different processes, but HP and Digital do not. Ultimately, these two smaller vendors may have to pick a particular niche and optimize both their processes and processors for that one area—or work with a second foundry. ♦

