IC Makers Confront RC Limitations

IBM Leads Move to Copper Interconnect; TI Deploys Low-k Dielectric

by Linley Gwennap

With most microprocessor makers bringing 0.25-micron CMOS processes on line over the next year, many of these vendors are now looking at forging ahead to the 0.18-micron level or, as an intermediate step, a 0.22-micron process. At this level, however, chip makers must confront a threat that has been lurking in the darkness ahead: the RC monster. Unless these vendors come up with new solutions, the resistance and capacitance of the interconnect will sap the performance gains from faster transistors, derailing Moore's Law.

Fortunately, new materials are at hand to improve the interconnect layers, ideally without even adding cost to the process. To reduce the resistivity of the metal layers, vendors are exploring the use of copper or gold in place of the traditional aluminum. IBM is the first to reveal plans to use copper, starting with its 0.20-micron process due next year. Similarly, by replacing the normal oxide (SiO_2) layers with so-called low-*k* materials such as polyimide or even air, the dielectric constant, and thus capacitance, can be reduced.

These changes are likely to be phased into microprocessor fabrication processes over the next few years. As with any significant process change, implementation issues abound, and vendors that solve these issues quickly and effectively will have a technology edge. Conservative vendors that stay too long with the older materials, on the other hand, are likely to fare the worst.

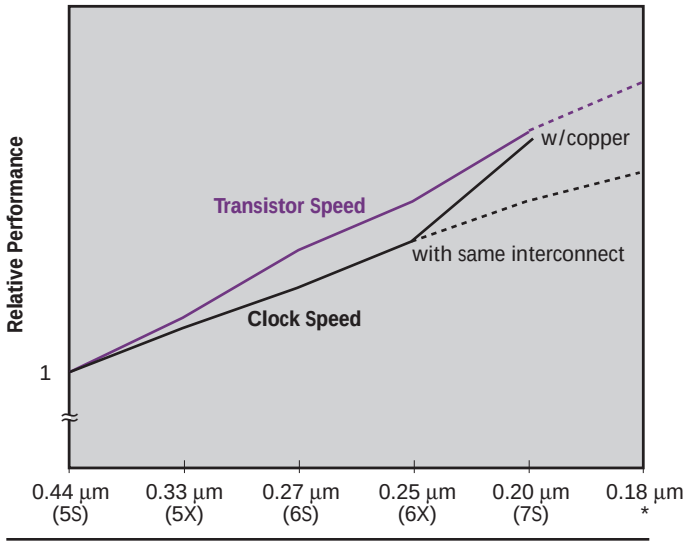


Figure 1. As geometries approach 0.18 microns, CPU clock speed falls well behind raw transistor speed without improvements in the interconnect layers. (Source: IBM except *MDR estimate)

Interconnect Saps Transistor Speed

Through the 0.5-micron generation, the method of improving clock speed was clear: reduce the size of the transistors to make them switch faster. Since the critical timing path was some number of gate delays, the increase in transistor speed caused a direct and equal increase in clock speed.

More recently, microprocessor designers have noticed that this is no longer the case. In many new processors, the critical path involves pushing a signal from one corner of the chip to another. Reducing the transistor delay does little to improve the speed of the chip. Shrinking the design improves clock speed mainly because the size of the die, and thus the length of the longest signal path, is reduced. This clock-speed gain is harder to achieve, however, when building a bigger chip on the new process.

The transition from transistor-dominated to interconnect-dominated performance is not instantaneous, but it is clearly occurring as we move to 0.35-micron, 0.25-micron, and ultimately 0.18-micron geometries. Figure 1 charts the clock speed of an IBM processor fabricated in various processes and compares it to the gain in transistor speed from each of these processes. At 0.33 microns, the clock speed tracks the transistor speed closely, but by 0.25 microns, the clock speed falls behind significantly. Extrapolating the curves, the gain in clock speed will lag severely at 0.18 microns despite the boost in pure transistor speed.

The speed of a signal across a long metal trace is determined by the time required for the voltage at the destination to change from high to low or from low to high. This switching time depends on many factors, including the current drive of the source device and the voltage between the high and low points. Two key factors, however, are the resistance *(R)* and capacitance *(C)* of the metal trace. For clock speeds to increase, the product *RC* must decrease with each new process generation.

Thicker Traces Only a Partial Solution

Figure 2 illustrates some of the problems faced by IC process designers as process dimensions continue to shrink. The key dimension for interconnect is the metal pitch, which determines how closely the metal lines can be drawn and thus how small the design can be made. The metal pitch is typically about twice the width, *W*, of a single trace. As the figure shows, when moving from a 0.5-micron process to a 0.25-micron process, *W* is reduced by a factor of two, as one would expect. Because the chip area is proportional to the square of *W*, this change reduces die size by a factor of four.

The resistance of a metal trace, however, is inversely proportional to its cross-sectional area, *TW* in the figure. In

previous process transitions, the metal thickness, *T*, has been reduced or kept constant. In many 0.25-micron processes, however, *T* is actually increased, particularly for the metal layers used for global routing. Even so, the figure shows the resistance of this 0.25-micron metal trace is 33% higher than for the 0.5-micron metal trace due to the reduction in *W*. This increased resistance makes it more difficult to quickly drive a signal across a long metal trace.

Resistance can be managed by further increases in *T*, but this causes problems with capacitance. In general, capacitance rises as the parallel area between two metal traces increases. As Figure 2 shows, increasing *T* adds to the parallel area between adjacent metal-3 traces. Thus, the capacitance between traces will be worse. Clearly, making a tall skinny trace will hurt capacitance even as it improves resistance. In addition, current etching techniques are not capable of making traces that are much taller than they are wide.

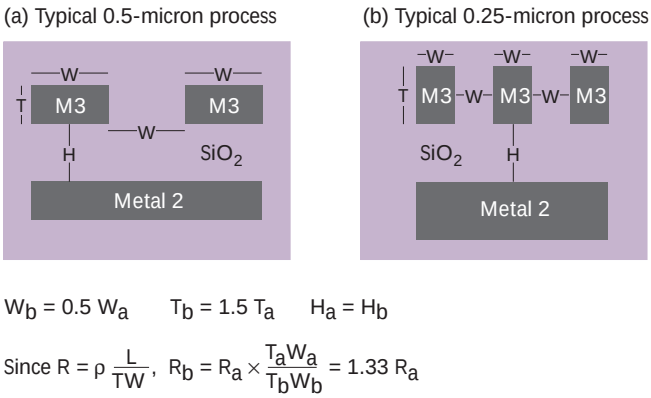
Capacitance Is Difficult to Improve

Calculating capacitance in an interconnect network is mathematically complex because it involves solving Laplace's equation, a second-order partial differential. In the simple case, where *W>>H*, the capacitance between a metal-3 trace and metal-2 trace below it can be stated as $C_{plate} = kWL/H$, where *k* is the dielectric constant of the insulator and *L* is the length of the trace (not shown in Figure 2). Thus, capacitance decreases at the same rate as *W*, assuming that *H*, *L*, and *k* are held constant. (In a simple design shrink, *L* will be reduced along with *W*, but we assume that vendors will eventually want to build an equally large chip in the new process.)

When *W*≤*H*, however, this simple equation for C_{plate} no longer applies, and the capacitance decreases with the log of *W*. Thus, for a given insulator height *H*, narrowing the metal traces will linearly decrease the capacitance only until the width of the traces reaches *H;* past this point, reductions occur at a slower logarithmic rate. As Figure 2 implies, this transition occurred around the 0.35-micron generation, so future improvements in capacitance will be difficult to come by through simple reductions in *W*.

We must also consider C_{edge}, the capacitance between adjacent metal traces. In Figure 2(a), where the traces are thinner and further apart, C_{edge} is roughly half of C_{plate}. In the 0.25-micron process of Figure 2(b), however, C_{edge} will dominate, because the traces are thicker than they are wide *(T>W)* and closer to each other than to the underlying metal layer *(W<H)*.

Unlike C_{plate}, C_{edge} gets worse as each process generation shrinks *W*. C_{edge} also gets worse if the thickness of the trace is increased to improve resistance. Like C_{plate}, C_{edge} increases at a logarithmic rate if *T*≤*W*. But as *T* becomes significantly greater than *W*, C_{edge} will increase linearly with *T*, an even faster rate of degradation. Thus, increasing trace thickness improves the *RC* product only until *T* gets into the linear region of the Laplace equation, limiting *T/W* to approximately two.



$W_b = 0.5\ W_a \qquad T_b = 1.5\ T_a \qquad H_a = H_b$

Since $R = \rho \frac{L}{TW}$, $R_b = R_a \times \frac{T_a W_a}{T_b W_b} = 1.33\ R_a$

Figure 2. Despite increased metal thickness, the 0.25-micron process has a 33% higher resistance than the 0.5-micron process. The thicker traces also increase the capacitance between adjacent metal lines, further degrading the critical *RC* product.

In summary, if *T* and *H* are reduced at the same rate as *W*, capacitance will be held constant, but resistance will increase rapidly. Holding *H* steady improves C_{plate} but not C_{edge}, and resistance continues to increase. Holding *T* steady improves resistance but not capacitance, and the overall *RC* value improves only as long as *T* is close to *W*. Since the typical 0.25-micron process uses a fairly optimal ratio of *T/W*, there is little room for future improvement based on simple geometric changes.

IBM Uses Copper to Reduce Resistance

Given the fundamental resistance equation in Figure 2, another option for addressing this problem is by changing ρ, the resistivity of the metal trace. Most IC makers today use aluminum for metal traces. As Table 1 shows, the resistivity of aluminum is not as good as that of copper, which is the most common material for electrical wiring. Aluminum continues to be used in most integrated circuits, however, because it is easier to work with in a silicon environment.

Although copper's advantage in conductivity has long been known, its disadvantages for integrated circuits have been substantial. Copper has a higher mobility in silicon than aluminum; left on their own, copper atoms would spread throughout the chip, ultimately contaminating the transistors and preventing them from operating properly. To prevent contamination during manufacturing, equipment used to process copper must not come in contact with the

	Resistivity	Comments
Polysilicon	300 μΩ·cm	Used for short connections in gates
Tungsten	5.6 μΩ·cm	Used for vias, refractory layers
Aluminum	2.8 μΩ·cm	3.6 μΩ·cm with refractory layers
Gold	2.4 μΩ·cm	No cladding; limited IC experience
Copper	1.7 μΩ·cm	2.2 μΩ·cm with cladding
Silver	1.6 μΩ·cm	Similar to copper; limited IC experience

Table 1. Copper is a better conductor than the aluminum used in most integrated circuits today, but gold offers an alternative. (All resistivity figures are for pure material at room temperature.)

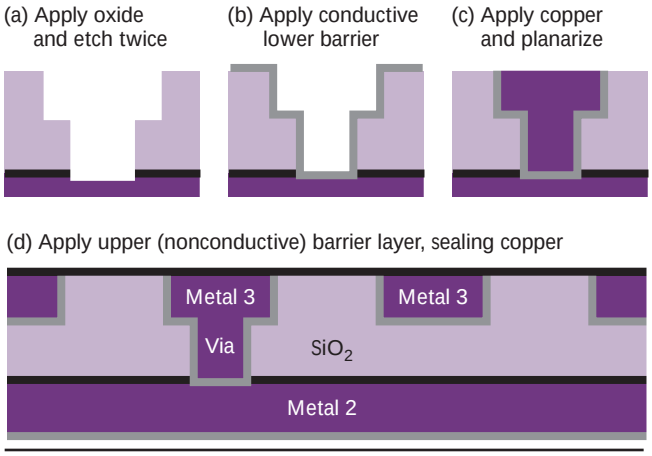


Figure 3. In IBM's dual-damascene process, vias and metal traces are etched out of the oxide and then filled, first with a conductive barrier material, then with copper. After planarizing, a nonconductive barrier is applied, completely sealing the copper.

silicon wafers. Copper must also be hermetically sealed to prevent oxidation; although aluminum also oxidizes, the initial oxidation forms a protective layer (self-passivation) that prevents further oxidation.

IBM plans to use copper for all metal layers and vias in its forthcoming 0.20-micron CMOS-7S process. To solve the problems associated with copper, the company has developed a barrier solution that envelopes all copper traces inside a protective sheath. This cladding keeps the copper away from all silicon structures and prevents oxidation. The company would not reveal the materials used for the barrier layer. The thickness of the barrier is less than 1,000 Å, or 0.1 µm; IBM would not be more specific.

Because of the higher resistivity of the barrier material, the overall resistivity of IBM's clad copper traces is about 2.2 µΩ·cm, a bit higher than the 1.8-µΩ·cm resistivity of pure copper but 40% lower than the resistivity of the traces in IBM's previous processes, in which the resistivity of the aluminum is impaired by the titanium refractory layers typically used to encase the aluminum.

The switch to copper enables a more efficient manufacturing technique known as dual-damascene construction. In most fabs today, a metal layer is constructed by applying a layer of aluminum and the refractory material, etching away unwanted metal (a masking step), applying oxide to fill in between the metal traces, and planarizing the wafer to produce a flat surface. For vias, however, the sequence starts with oxide instead of metal, reversing the order but requiring the same number of steps. This latter technique, in which the metal is inlaid in the oxide, is called damascene.

In a dual-damascene process, illustrated in Figure 3, a layer of oxide is produced and etched twice before the metal is applied and planarized, eliminating one metal step and one planarization step. This technique can be used only when the via and metal trace use the same material—in this case, copper. Most IC processes today combine aluminum metal layers with tungsten vias, but IBM developed the dual-damascene process for its 4-Mbit DRAM, which has a tungsten interconnect layer. IBM's experience with producing this DRAM in high volume gives it a leg up in implementing dual-damascene construction for its CMOS-7S parts.

As Figure 3 shows, the dual-damascene process allows the barriers around the copper to be constructed efficiently, since a single set of barrier layers protects both the via and the metal trace. These layers eliminate the refractory layers used with aluminum traces and the barrier typically placed around a tungsten via. Thus, the dual-damascene technique provides a cost saving over the etching used in traditional aluminum/tungsten processes. In addition, IBM believes the switch to copper will close the gap between clock speed and transistor performance shown in Figure 1, at least for the 0.20-micron process.

Reducing the resistivity provides designers with two options. One is to leave the shape of the traces as shown in Figure 2(b) and simply substitute copper for aluminum; this reduces overall resistance in proportion to ρ. Alternatively, one could reduce the thickness of the traces, keeping R constant while improving C_{edge}. Some combination of these two approaches may lead to the optimal solution, which depends on the specific parameters of each process.

Gold Also Reduces Resistance

Gold is used in the metal layers of some high-speed bipolar chips today, and MicroUnity used gold interconnect in its ill-fated BiCMOS fab (see MPR 10/23/95, p. 11). Although gold has a higher inherent resistivity than copper, it is essentially inert in silicon and thus does not require cladding or refractory layers. As Table 1 shows, the resistivity of a pure gold trace is within 10% of that of a clad copper trace.

By eliminating the barrier layers, gold could reduce the number of process steps and thus the wafer cost compared with aluminum or copper, assuming a dual-damascene process is also used for the gold traces. Gold is also more scalable to finer geometries. As the metal width gets smaller, the barriers will occupy an increasing portion of the metal trace, ultimately degrading its resistance.

Gold sounds expensive, but for the nanowires on an integrated circuit, the material cost is not significant. Processing two metal layers with gold instead of aluminum requires less than one gram of unrecoverable gold per 200-mm wafer—about $10 at current prices. Like copper, gold will contaminate silicon transistors, and so must be handled carefully, but this does not add significant cost in a large fab.

To date, however, no one has produced highly integrated CMOS circuits using gold interconnect, and with MicroUnity out of the fab business, no vendor has even announced plans to use gold. Gold's advantages over copper are not compelling for most applications, so most research by chip manufacturers and equipment makers has been focused on copper. We expect copper to replace aluminum in most advanced CMOS processes over the next few years.

New Materials Reduce Capacitance As Well

New materials can also be used to address the capacitance equation. Most current VLSI processes use SiO_2, which has a dielectric constant (*k)* of about 4.0, as the insulating material. Reducing *k* has a direct effect on capacitance. Texas Instruments today uses a material called HSQ (hydrogen silsesqui-oxane, $HSiO_{1.5}$) to fill the gaps between adjacent metal traces, although it still uses SiO_2 for the less-critical gaps between metal layers. HSQ, developed by Dow Corning, has a *k* of about 3.0 and is a polymer with properties similar to those of SiO_2, making it a straightforward substitution.

Other vendors are exploring polyimide, which has a *k* of about 2.0. While switching to polyimide reduces capacitance by a factor of two, there are some drawbacks. Its organic nature makes it more permeable than SiO_2 to aluminum electromigration. Thus, combining polyimide with traditional aluminum traces is not a good idea.

Once vendors switch to clad copper traces, however, the barriers, not the dielectric, protect the metal traces, so the permeability of polyimide is no longer an issue. We expect IC makers to adopt polyimide or a similar material once they have moved from aluminum to copper metal layers. HSQ provides an intermediate solution that is compatible with aluminum but offers a smaller benefit. These new dielectrics also reduce power dissipation, since $P=Cv^2f$.

Air Bridges Offer Potential for Improvement

The best possible dielectric is a vacuum, but for practical purposes, air (or any other low-pressure gas) is just as good, with a *k* of just 1.0. This technique, called air bridging, removes the dielectric material entirely, leaving a metal trace suspended in midair. MicroUnity used this technique in its fab but never got it into production; air bridges are in production today in a few high-speed bipolar fabs.

As Figure 4 shows, air bridges (like larger bridges) must be supported by piers if they span a long distance. These piers are not to support the weight of the trace; even under the high G-force caused by dropping a chip, the weight of a micrometer-size metal trace is negligible. Given the small distance between metal traces, however, electromagnetic forces are much greater and can bend unsupported traces.

Consider that an unsupported trace, like a piano wire, has a resonant frequency. If nearby metal traces carry an electrical signal that fluctuates at this resonant frequency, the air bridge can quickly build up enough momentum to destroy itself. The piers must be inserted often enough to raise the resonant frequency of the bridge above that of any signals in the chip. Of course, the piers add unwanted capacitance to the system, but the overall dielectric value of the bridge can be kept to within 10% or 20% of that of pure air, still far less than the *k*-value of polyimide.

Fabrication issues are straightforward: a "sacrificial" dielectric is used during the initial construction and etched out in a single step after the interconnect layers are complete. A second extra step is needed to construct the piers, which

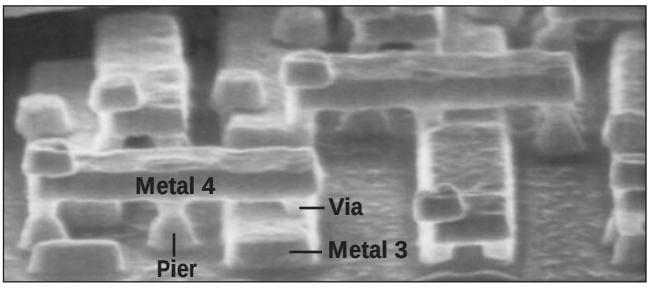


Figure 4. The photo shows an air-bridged metal trace with a support pier built in MicroUnity's 0.5-micron BiCMOS process. (Source: MicroUnity)

are fabricated from a dielectric substance that must remain after the sacrificial dielectric is removed.

One drawback is that the air bridges are not tolerant of the flux used in solder bonding. Air-bridged die must be wire bonded or pressure bonded. A second problem is that the air bridges must be hermetically sealed, ruling out the use of a standard plastic package.

Air bridges can be made to work with copper, but the copper must still be clad where it contacts silicon, adding processing complexity and cost. Fabs that use air bridges today, as well as MicroUnity's fab, generally use gold for the metal layers that are bridged, but current IC processing techniques are not optimized for the use of gold.

These issues, as well as the added processing cost, are likely to keep air bridges out of mainstream microprocessors in the near future, but as the pressure increases to improve the capacitance of the interconnect layers, CPU makers are sure to take another look at this technology.

Interconnect Problems Won't Get Easier

Over time, RC issues will become even more difficult. Changing materials is a one-time improvement; even if the industry gets all the way to air bridges, there are only a few factors of two along the way, and these will be consumed by ongoing reductions in trace width. After a few process generations, improvements in RC could come to a halt.

At that stage, about 5–10 years from now, BiCMOS technology could reappear. Bipolar transistors drive more current than CMOS processors and could be used to speed critical signals once methods of improving RC peter out. Bipolar transistors were popular in 0.5-micron microprocessors, but silicon-based bipolars don't operate well at the low core voltages used in 0.25-micron processes.

IBM, however, has recently sampled products based on its new silicon-germanium (SiGe) technology, which provides low-power bipolars compatible with advanced CMOS processes. SiGe is best suited for small mixed-signal devices today, but it could boost the performance of high-end microprocessors in several years.

Motorola, IBM Lead in 0.25-Micron Deployment

In the emerging generation of 0.25-micron processes (see MPR 9/16/96, p. 11), Motorola and IBM have taken the lead

Vendor	IBM†	IBM†	TI	TI	TI	TI	Intel	UMC
Process name	CMOS-7S	CMOS-6X	C05	C07b	C07a	C10c	P856	0.25µ Logic
Example product	PPC G3	Mach 5	n/a	n/a	n/a	UltraSparc-2	Tillamook	n/a
First production	2H98	3Q97	1H00	4Q98	1Q98	3Q97	3Q97	2Q98
Supply voltage	1.8 V	1.8 V	1.5 V	1.8 V	1.8 V	2.5 V	1.8 V	2.5 V
I/O voltage (max)	3.3 V	3.3 V	n/a	3.3 V	3.3 V	3.3 V	2.5 V	3.3 V
Gate length (drawn)	0.20 µm	0.25 µm	0.13 µm	0.18 µm	0.21 µm	0.25 µm	<0.25 µm	0.24 µm
Channel length (effective)	<0.15 µm	0.15 µm	n/a	0.14 µm	0.17 µm	0.21 µm	n/a	n/a
Gate oxide thickness	<40 Å	40 Å	33 Å	40 Å	40 Å	54 Å	45 Å	50 Å
Number of metal layers	6 metal	6 metal	6 metal	6 metal	6 metal	5 metal	5 metal	5 metal
M1 contacted pitch	0.63 µm	0.7 µm	0.60 µm	0.85 µm	0.85 µm	1.1 µm	0.64 µm	0.64 µm
M2 contacted pitch	0.81 µm	0.9 µm	0.60 µm	0.85 µm	0.85 µm	1.1 µm	0.93 µm	0.76 µm
M3 contacted pitch	0.81 µm	0.9 µm	0.60 µm	0.85 µm	0.85 µm	1.1 µm	0.93 µm	0.76 µm
M4 contacted pitch	0.81 µm	0.9 µm	0.60 µm	0.85 µm	0.85 µm	1.1 µm	1.6 µm	0.76 µm
M5 contacted pitch	0.81 µm‡	0.9 µm‡	1.75 µm	2.5 µm	2.5 µm	2.5 µm	2.6 µm	1.2 µm
M6 contacted pitch	0.81 µm‡	0.9 µm‡	1.75 µm	2.5 µm	2.5 µm	—	—	—
Metal composition	Copper	Aluminum	2 Cu + 4 Al	Aluminum	Aluminum	Aluminum	Aluminum	Aluminum
Intrametal dielectric	SiO_2	SiO_2	HSQ	HSQ	HSQ	HSQ	SiO_2	SiO_2
Local interconnect?	Yes	Yes	No	No	No	No	No	No
SRAM cell size	6.8 µm²	8.6 µm²	n/a	10.5 µm²	10.5 µm²	22 µm²	10.3 µm²	6.0 µm²
Routing index*	0.43 µm²	0.53 µm²	0.26 µm²	0.53 µm²	0.53 µm²	0.85 µm²	0.67 µm²	0.44 µm²
Wafer cost index*	$4.9	$4.7	$5.5	$4.6	$4.5	$3.6	$4.0	$4.1

Table 2. IBM and Texas Instruments are leading the way in using new materials for metal traces and dielectric layers. Foundries such as UMC lag the process leaders by about one year. See MPR 9/16/96, p. 11, for a description of routing and cost indices. †Motorola offers processes similar to IBM's, called HIP5 and HIP4, respectively. ‡optionally double width for lower resistivity. (Source: vendors except *MDR)

by getting a 0.25-micron microprocessor into production before any other vendor (see MPR 7/14/97, p. 4). Although Intel has not yet formally announced its 0.25-micron Tillamook processor, the company has already started volume shipments of the part, putting it in nearly a dead heat with the PowerPC vendors. AMD says it is on track to deliver a 0.25-micron version of its K6 processor by the end of this year, and Fujitsu also expects to reach this milestone by Q4.

Texas Instruments has quietly started shipping a 0.25-micron version of UltraSparc-2, claiming to be the first microprocessor vendor to the 0.25-micron level. The company had originally planned to put this process, called C10c, into production in 1Q97, but it ran into problems with the predecessor C10b process, used for the first UltraSparc-2 chips. TI began shipping some C10b chips to Sun in 3Q96 but had problems yielding the part, and it was 1Q97 before there were enough chips available for Sun to announce a product (see MPR 2/17/97, p. 3). Although most other vendors moved to DUV steppers at the 0.25-micron level, TI stayed with less precise i-line steppers for the C10 processes, and this choice may have contributed to the initial yield problems. TI says the C10c process is now yielding well with the i-line steppers.

Despite the delay, the company has rushed to near the head of the pack. TI still plans to deploy a 0.18-micron C07 process in 1998, although the metal pitches will be more characteristic of a 0.25-micron process. This process will use some DUV steppers. The company also previewed its 0.13-micron process, due in early 2000. As Table 2 shows, the process will combine the low-*k* dielectric of the C07 process with copper traces for the top two metal layers.

Other microprocessor vendors, including Digital and IDT, continue to target 1H98 for their 0.25-micron processes. While IC foundries have historically lagged as much as a full process generation behind the leading edge, demands from the emerging 3D market (see MPR 6/2/97, p. 16) and from makers of other high-speed chips are driving companies such as TSMC, Chartered, and UMC to advance their process technology more rapidly. Both TSMC and UMC plan to produce 0.25-micron chips in 1H98, and as Table 2 shows, UMC's process parameters are quite competitive with those of the microprocessor makers.

Changes Ahead in 0.18-Micron Generation

For the leading vendors, 1998 will be a transition year, with 0.25-micron processes ramping into high volume but 0.18-micron processes still over the horizon. Some, such as the PowerPC vendors, will deploy "half-generation" processes with reduced transistor and metal pitches. Others, such as TI, will use hybrid processes, combining the metal layers of a 0.25-micron process with smaller transistors. Intel will take the latter approach but probably won't describe its advances as a new process, continuing to use the P856 name as it tweaks and improves its transistors.

The leap to true 0.18-micron CMOS processes will begin in 1999 as the leading vendors—likely to be TI, Intel, IBM, and Motorola again—ship their first microprocessors at that level. We expect most vendors will move to copper in the 0.18-micron generation, adopting polyimide or another low-*k* dielectric at the same time or shortly thereafter. The PowerPC makers will get a jump on this transition by adding copper to their processes in 1998, which could give them an advantage in moving to the 0.18-micron generation. With more changes than usual ahead, some CPU makers are likely to stumble during this transition. M