Intel Raises the Ante With P858

Details of Next-Generation 0.18-Micron Process Disclosed

by Keith Diefendorff

The cost to play in the x86 processor game just went up. At last month's International Electron Devices Meeting (IEDM) in San Francisco, Intel disclosed details of its next-generation 0.18-micron P858 process; this month, tiny beads of perspiration have begun forming on the brow of Intel's x86 competitors.

The new process is fast, dense, power-efficient, and easy to manufacture. With 140-nm gates and an ultrathin gate oxide, the process delivers extraordinarily fast transistors. As Figure 1 shows, six layers of aluminum wires with low-*k* dielectrics provide a high-speed interconnect system with 25% tighter pitches than the company's current 0.25-micron P856.5 process (P856.5 is a 5% shrink of P856). Operating at between 1.3 V and 1.5 V, the new process will, at long last, bring desktop-speed processors to PC notebooks. With remarkably few mask layers (21) and conventional aluminum metallization, the process should be relatively inexpensive to manufacture.

Competitors should be worried. Process technology has always been Intel's most potent weapon against competition, and P858 will be no less formidable. The process should enable Intel to maintain its preeminent position, even with only modest enhancements to its existing processor designs.

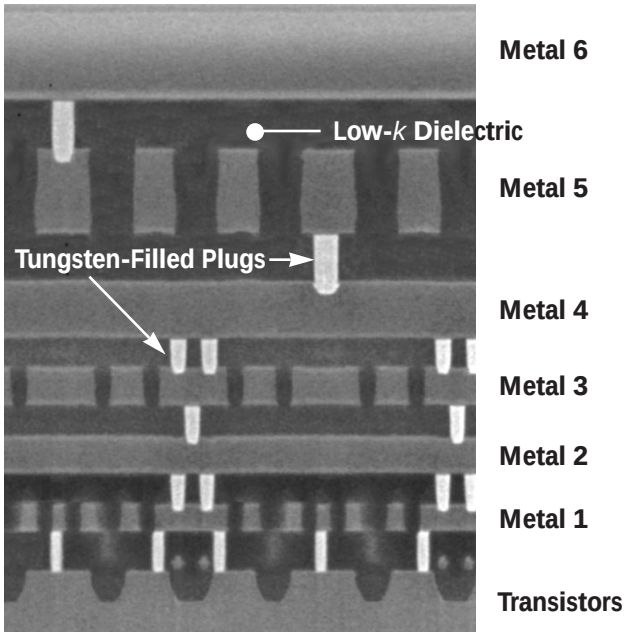

Figure 1. Cross section of P858's six-layer aluminum interconnect system. Chemical mechanical polishing (CMP) is used to planarize layers. Tungsten-filled plugs form the vias. (Source: Intel)

Fast Transistors Key to Speed

Intel believes transistor performance still dominates microprocessor speed. For this reason, the company placed its emphasis on P858's transistors and elected to stick with a conventional aluminum interconnect system.

P858 transistors are formed on an epitaxial wafer, separated by shallow-trench isolation (STI). Polysilicon lines and source-drain regions are coated with titanium salicide ($TiSi_2$) to reduce resistance. Precise control over the source/drain, well, and halo implants provides good short-channel behavior down to physical gate lengths (L_{gate}) as short as 130 nm (nFETs) and 150 nm (pFETs).

The gate is patterned using standard 248-nm deep-ultraviolet (DUV) lithography. The electrical thickness (T_{ox}^{EFF}) of the gate oxide is a mere 30 Å, providing excellent electrical control over the channel and high drive currents. Intel did not disclose the physical oxide thickness, but we estimate it to be less than 25 Å. This thickness is frighteningly close to the limit imposed by the breakdown of silicon dioxide from gate-to-channel tunneling current, as predicted in another IEDM paper by IBM. Intel reliability data, however, suggest that P858's gate oxide is thick enough to prevent breakdown for at least 10 years.

P858's nFET and pFET drive currents (I_{dsat}) are an astounding 940 and 420 µA/µm at 3 nA/µm of leakage (I_{off}). According to the MDR FET Performance Metric, this makes P858 transistors about 50% faster than those in P856.5. Such transistors will support Intel's roadmap for several years. The new process should boost the Katmai core (in the Coppermine processor) from its expected 533-MHz limit in P856.5 to 733 MHz or higher. The transistors are easily fast enough to support suitably designed processors with clock rates up to 1 GHz, as we expect from Intel's next-generation Willamette.

P858's transistors compare favorably with those in the fastest previously reported 0.18-micron process, IBM's CMOS-8S. On the basis of published data, Intel's transistors are faster; but our analysis of more recent data indicates that the transistors in both processes have roughly equivalent intrinsic speed. The two processes, however, are optimized differently: while Intel's thin gate oxide gives P858 transistors higher drive currents than 8S's, it also creates a higher gate capacitance. Thus, Intel's process has more power for driving heavy interconnect loads, while IBM's lighter gate loads offer better logic fan-out. The different approaches, as we will see, may have to do with their respective choices of interconnects.

The IBM process also appears to have a slight edge in transistor density, due to the use of cobalt-salicided polysilicon and diffusions. On very thin poly lines, Cobalt salicide offers lower sheet resistance than does titanium salicide. This

fact allows 8S's vertical process dimension to be reduced further, which, in turn, allows its horizontal dimensions to be made somewhat smaller. This is evidenced by 8S's poly pitch, which is about 12% tighter than that of P858.

The downside of cobalt salicide is that it's trickier to build; considerable cleverness is required to remove all the oxygen from the salicide surface. But once the technique is mastered, IBM claims the manufacturing costs are no higher than with titanium salicide.

Aluminum vs. Copper Debate Rages

Even though microprocessor speed may be dominated by the transistors, as Intel contends, interconnects cannot be ignored. Unfortunately, improving interconnect speed is more difficult than improving transistor speed: as transistors shrink they naturally speed up; not so for wires. Narrower wires have higher resistance, which slows signal propagation. Making wires thicker combats the resistance but increases parasitic capacitance, with equally undesirable results.

How best to improve interconnect speed is a subject of considerable debate. IBM and Motorola have adopted low-resistance copper in their initial 0.18-micron processes. Several other vendors plan to upgrade to copper over time. But Intel, along with a few others such as Mitsubishi and Toshiba, steadfastly insists that copper is neither required nor desirable at 0.18 micron. It is not clear whether these companies actually believe aluminum is superior to copper, or whether they simply wish to defer for another generation the burden of upgrading their factories with copper-deposition equipment.

The argument favoring copper over aluminum is that its lower resistance and higher current-carrying capacity (without electromigration) allow thinner wires with less wire-to-wire capacitance. The lower capacitance speeds signals and reduces crosstalk, allowing denser layouts.

One problem with copper is its propensity to diffuse into the surrounding material, contaminating adjacent structures. IBM, in its copper PowerPC 750, solves the problem with a 30-nm-thick tantalum-nitride antidiffusion seal around the copper traces, according to analyst firm Chipworks. This seal, however, reduces the cross-sectional area of the copper in the trace, increasing resistance. But data that Intel has published for P858 indicate that the titanium and titanium-nitride refractory layers in its metal stack increase the resistance of aluminum traces by a similar amount.

Thus, we see no reason that copper should not reduce RC delay and crosstalk, as promised. Still, copper opponents contend that the RC effects in real circuits are not sufficient to justify copper at 0.18 micron. But evidence to the contrary is mounting. In an IEDM paper on its 0.2-micron BiCMOS process, Hitachi reported that copper improved wire delay by 30% with signal lines half as thick as those of aluminum.

Electrical superiority aside, until copper processes mature, aluminum is clearly easier and less expensive to put into volume production. It is likely that these arguments are what's really leading opponents to deny the value of copper.

Intel Sticks With Aluminum

Despite the arguable advantages of copper, Intel has decided to stay with aluminum for P858. To achieve acceptable trace resistances with P858's tight pitches, Intel uses advanced aluminum metallurgy and aggressive aspect ratios (ratio of trace thickness to width).

As Figure 2 shows, P858's aspect ratios range from 10% to 25% higher than those used in P856.5. Even so, notice that P858's traces have significantly higher resistance. P858's sixth layer of metal compensates somewhat, while also improving routability, but not enough to offset the loss completely. Thus, routing delays are likely to figure more prominently in P858 designs than in P856.5 designs.

Figure 2 also shows that while the resistance of P858's interconnect layers is similar to those of 8S, 8S's metal layers are dramatically thinner. In addition, the dual-damascene copper process forms interlayer vias out of copper—rather than tungsten-filled plugs as in aluminum systems—further improving resistance or density. According to our calculations, had Intel used copper, P858's traces could have been about 40% thinner for the same resistance, noticeably reducing intrametal capacitance. The net effect on overall interconnect delays would depend on circuit design and layout details but could easily amount to 10% or more.

In addition, IBM's local-interconnect layer (M0) and seventh layer of wiring (M7) endow 8S with even more routing flexibility. Thus, IBM's copper metal system appears capable of delivering both faster and denser interconnects than Intel's aluminum system.

A factor that cannot be discounted, however, is Intel's nearly infinite design resources. To a large extent, design tools, elbow grease, and die area can overcome many of aluminum's deficiencies. Trading long wires for transistors, by replicating logic or judicious use of repeaters, for example,

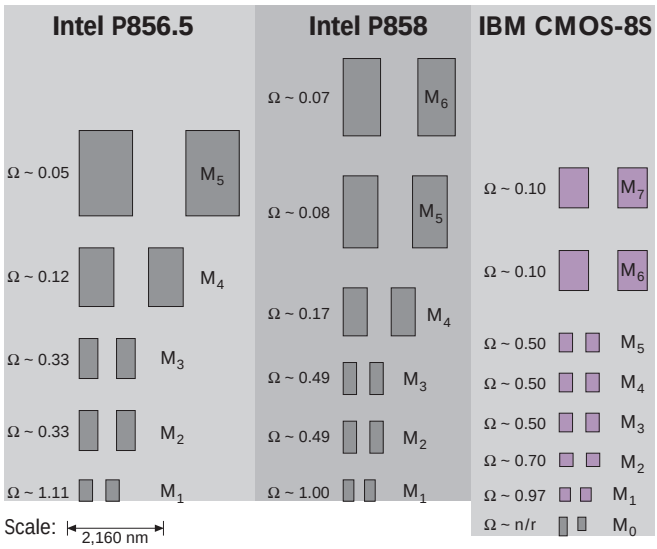

Figure 2. A comparison of the interconnect systems of Intel's P856.5 and P858 and IBM's CMOS-8S. Ω~ = relative resistance (MDR estimates); n/r = not relevant. (Source: vendors, except Ω)

can reduce RC effects at the cost of die area. To ensure that P858's aluminum interconnects don't hinder performance, Intel may be depending either on the truth of its claim that microprocessor speed is dominated by the transistors or on its circuit-design prowess to make it so.

Low-k Dielectric Counters High Aspect Ratios

To ameliorate the high capacitance of P858's thick wires, Intel employs a silicon-dioxide insulating material doped with 5.5% fluorine to lower the dielectric constant. This material, which Intel calls SiOF, is essentially the same as the fluorine-doped silicon glass (FSG) IBM employs in CMOS-8S. Pure SiO_2 has a dielectric constant of 4.1, about 15% higher than the 3.55 of SiOF or FSG. From experiments on interconnect-intensive ring oscillators, Intel found that SiOF improved frequency by 16% over pure SiO_2. (Intel did not explain how performance could have increased more than the improvement in dielectric constant.)

Since less energy is required to charge and discharge wires, the lower capacitance resulting from the lower dielectric constant also reduces active power consumption. Power savings from the use of SiOF are probably under 10%.

Smaller Features, Smaller Die

For most size-related parameters, P858's dimensions are 20–25% smaller than P856.5's. Thus, most logic circuits should be about 35–45% smaller than those in P856.5. The extra interconnect layer could save an additional 5–10% area.

SRAM cells will shrink by a similar amount. The process-development vehicle Intel used to demonstrate P858's performance and yield characteristics was a 900-MHz 16-Mbit SRAM. This SRAM utilized a 5.6-μm^2 bit cell, which is 60% smaller than the 9.3-μm^2 cell Intel uses in P856.5. The smaller cell will figure prominently in Intel's plans, as nearly all of its future processors will employ large on-chip caches.

As Table 1 shows, the IBM CMOS-8S bit cell is 4.2 μm^2, more than 30% smaller than the P858 cell. The difference is due mainly to IBM's tungsten local interconnect and partially to the tight pitch of its cobalt-salicided poly lines. Sources indicate that IBM has developed an even smaller cell that it will deploy in production 8S devices.

Intel doesn't use local interconnect, arguing that it has much less benefit in logic circuits and that the yield loss from the extra process complexity of the local interconnect is greater than the loss from a slightly larger die. Instead, Intel prefers to keep its process as simple as possible to facilitate rapid process shrinks. Intel's argument was certainly sound while microprocessors used small amounts of on-chip cache, but the argument is less convincing when processors have large portions of the die dedicated to SRAM.

Although P858 would reduce the size of Intel's existing processors by up to 40%, the company will primarily use the extra die area to increase functionality: mostly on-chip L2 cache. Although Coppermine will shrink the 0.25-micron Katmai core dramatically, from about 140 mm^2 to well below 100 mm^2, its expected inclusion of a 256K on-chip L2 will bring the die back into the 130-mm^2 range. Thus, we do not project any net decrease in Intel's average die size due to P858.

Vendor	Intel		IBM
Process	P856.5	P858	CMOS-8S
Process Generation	0.25 µm	0.18 µm	0.18 µm
Example Product	Mendocino	Coppermine	n/a
First Production	3Q98	3Q99	2H99
Supply Voltage	2.0 V	1.3–1.5 V	1.5 V
I/O Voltage (max)	2.5 V	2.5 V	2.5 V
Poly Half-Pitch	0.32 µm	0.24 µm	0.21 µm
Gate Length (L_{gate})	0.20 µm	0.14 µm	<0.13 µm
Gate Oxide (T_{ox}^{Eff})	41 Å	30 Å	36 Å
Substrate	Bulk Si	Bulk Si	Bulk Si
Metal Layers	5 Al	6 Al	7 Cu
M1 Contacted Pitch	0.61 µm	0.50 µm	0.49 µm
M2 Contacted Pitch	0.88 µm	0.64 µm	0.63 µm
M3 Contacted Pitch	0.88 µm	0.64 µm	0.63 µm
M4 Contacted Pitch	1.73 µm	1.08 µm	0.63 µm
M5 Contacted Pitch	2.43 µm	1.60 µm	0.63 µm
M6 Contacted Pitch	–	1.72 µm	1.26 µm
M7 Contacted Pitch	–	–	1.26 µm
Local Interconnect	–	–	0.42 µm (W)
Intrametal Dielectric (*k*)	SiO_2 (3.9)	SiOF (3.6)	FSG (3.6)
SRAM Cell Size	9.3 μm^2	5.6 μm^2	4.2 μm^2
Ring Oscillator Stage	22 ps	11 ps	11 ps
Routing Index* (μm^2)	0.60	0.30	0.25
Wafer Cost Index* ($)	4.2	5.4	6.0
FET Performance* (GHz)	29.9	46.0	48.0

Table 1. Intel's new P858 has transistors as fast as those in any previously reported 0.18-micron process, including IBM's CMOS-8S, but IBM's copper-interconnect system is about 20% more dense than Intel's aluminum system. The FET performance metric for CMOS-8S has been adjusted for the same 3-nA/µm I_{off} current Intel cited for P858. (Source: vendors, except *MDR estimates)

Full Speed Ahead for Notebooks

For the first time, P858 will allow Intel to field notebook processors with frequencies approaching those of its high-end desktop processors. Previously, Intel's mobile processors, including the recently announced Dixon (see MPR 1/25/99, p. 20), had to run at reduced frequency and voltage to fit within the 10-W thermal envelope for CPU and cache in notebooks. Dixon, for example, is limited to 366 MHz by its 9.5-W power dissipation at 1.6 V. Boosting the frequency would linearly increase power consumption but, worse, would require an increase in voltage, quadratically increasing power.

P858 will fix the problem. Coppermine, assuming a top speed of 750 MHz at 1.5 V, should run at 650 MHz at P858's 1.3-V operating point. At that voltage and frequency, it should comfortably fit within the 10-W notebook limit. Intel's Geyserville technology could be used to boost the part to full speed while the notebook is plugged into a wall socket.

Leading the Way With Organics

One reason Intel is able to justify sticking with aluminum interconnects is its aggressive push to flip-chip mounting on organic substrates. Packages based on organic substrates will

be the primary delivery vehicle for P858 processors, offering better performance and lower cost than ceramic substrates that are now commonly used for high-end microprocessors.

Although IBM and Motorola have for years used flip-chip mounting (which IBM calls C4) on PowerPCs, they have for the most part used ceramic substrates. Ceramic substrates have a dielectric constant of about 9, compared with about 3.5 for an organic substrate. Furthermore, ceramics must be fired at high temperature, restricting traces to refractory tungsten or molybdenum. In contrast, organic substrates, similar to the FR-4 in printed-circuit boards, use etched-copper traces with much lower resistance. The combination of lower capacitance and lower resistance gives organic substrates a performance advantage over ceramics.

Not only do organic packages themselves have better performance, they can also increase the performance of the mounted processor. Because P858 processors are flip-chip mounted, the substrate's copper traces provide the equivalent of an additional layer of coarse pitch, but very low resistance, chip interconnect, which can be extremely useful for distributing power and ground to the the chip. Of course, IBM could also, and probably eventually will, adopt organic substrates, gaining an eighth interconnect layer.

Steep Production Ramp Planned

In 1994, the Semiconductor Industry Association's *National Technology Roadmap for Semiconductors* placed technology generations on a three-year cycle, while Intel was operating on a 2.5-year cycle. In 1997, the SIA modified its roadmap to a 3/2/3-year cycle. Now, Intel says it is driving toward a two-year cycle, which it will achieve with P858 if it delivers Coppermine this fall. This pace will be difficult for Intel to sustain, due to the huge volumes through its fabs, but, if achieved, would place severe hardship on competitors, which have far fewer R&D dollars to work with.

In terms of transistors per year, P858 will produce an enormous increase in Intel's fab capacity. Coppermine, while similar in size to Katmai, will have nearly three times as many transistors. But in terms of units per year, P858 will have little impact. It may, in fact, cause a short-term hiccup in capacity as the new process comes on line. Ongoing fab upgrades, however, will increase Intel's total wafer throughput enough to compensate, allowing Intel to increase its total unit capacity in 1999 and 2000, despite the transition.

An important attribute of P858 is that it introduces no radically new manufacturing equipment. The process can be built with the same 248-nm steppers that Intel currently uses for P856.5; ion implants and salicide coatings are similar to those used in P856.5, and the interconnect system is a traditional aluminum deposition and etch with CMP. The new low-*k* dielectric has little impact on process flow. These factors will smooth the transition from 0.25- to 0.18-micron manufacturing, with great economic benefit to Intel.

Although initial yield may be lower than the current 0.25-micron yield, Intel has vowed to begin the production of each new process generation at a lower defect rate than that at which the previous generation began, and also to reduce defect rates more quickly each generation. Intel's 0.35-micron P854 generation achieved a learning curve of about 60%/year, improving to about 63% for 0.25-micron P856. So far, P858 appears to be on a 65% learning curve, which, if continued, will indeed meet both of Intel's stated objectives for the 0.18-micron generation.

Intel also says it intends to ramp P858 to volume production more quickly than any previous generation. The company estimates that the new process will achieve each level of wafer throughput in about 50% less time than P854 and about 20% less time than P856. By 4Q00, we expect nearly all of Intel's microprocessor capacity to be converted to 0.18-micron wafers. While other semiconductor vendors may introduce their 0.18-micron processes in the same time frame as P858, we doubt that any can match Intel's ramp rate.

Great Transistors, Good Interconnects

P858 is an impressive process with transistors as fast as those of any process reported to date, including IBM's aggressive CMOS-8S. The P858 interconnect system, while not as dense or as fast as that in 8S, is about as good as an aluminum system can get. Although 8S on SOI wafers may eventually claim the 0.18-micron speed title, it will be later than P858 and initially more expensive to manufacture. Intel's virtually unlimited design resources could nullify any technical disadvantages of P858 and magnify its advantage over lesser or equivalent processes from x86 competitors.

With P858, Intel has clearly demonstrated that a good 0.18-micron interconnect system can be built with aluminum. The high aspect ratios used, however, may extract a yield penalty, as the metal will be harder to etch without shorts and the spaces more difficult to fill without voids. Switching to copper would have solved this problem and given Intel valuable manufacturing experience for 0.13 micron, where copper will be mandatory. On the other hand, waiting simplifies the transition to 0.18 micron and gives Intel a chance to develop a more highly optimized copper technology for 0.13 micron.

Since 1995, when Pentium Pro was introduced, Intel has relied on the same basic P6 microarchitecture to power all its new processors; this will not change until Willamette appears in late 2000. Competitors have taken advantage of Intel's microarchitectural hiatus to encroach on the company's turf, and they will soon challenge Intel for the performance lead. So far, Intel's superior process technology and manufacturing prowess have kept it ahead of the pack. Fortunately for Intel, a half-generation lead in process technology is worth about as much as anyone is likely to gain from microarchitecture. If P858 can put such a distance between Intel and its x86 competitors—which it appears capable of doing (save for IBM)—process technology will again rescue the day for Intel. So, thanks to P858, Intel's leadership position seems secure—for at least another couple of years. M