Reed Electronics Group

MICROPROCESSOR REPORT

www.MPRonline.com

THE INSIDER'S GUIDE TO MICROPROCESSOR HARDWARE

INTEL ADOPTS STRAINED SILICON

New Fab Technology Will Boost Clock Speeds

By Tom R. Halfhill {9/3/02-01}

Intel's next-generation 90nm fabrication process will use strained silicon, a new technique that boosts circuit performance and is also under development at other companies (see *MPR 4/22/02-01*, "AmberWave Commercializes Strained Silicon, Part 1," and *MPR 4/29/02-02*, "Strained Silicon, Part 2"). The technique should significantly increase the clock frequency and only slightly increase the manufacturing cost of Intel's Prescott Pentium 4 when it debuts in 2H03.

For now, Intel isn't saying how much faster the NetBurst-microarchitecture Prescott will run by using strained silicon. Experiments with similar technology at other companies—including Hitachi, IBM, and startup AmberWave Systems—have shown improvements ranging from 30% to 120% in electron mobility and transistor current flow (drain current), which translate into higher clock frequencies because of faster transistor switching and wire speeds. Intel is claiming a 10–20% increase in drive current from its first-generation strained silicon, which it says it developed independently. Anything close to a double-digit boost of the core frequency would be significant, because Intel says strained silicon will add only about 2% to the manufacturing cost of a wafer.

As Figure 1 shows, strained silicon improves circuit performance by increasing the mobility of electrons and holes in the silicon layer. It adds several manufacturing steps to a bulk CMOS process and requires some design modifications to recharacterize circuit performance. Although the exact process may vary at different companies, it generally starts with deposition of one or more layers of silicon germanium (SiGe) on the wafer in a substrate below the silicon. SiGe has a larger atomic spacing

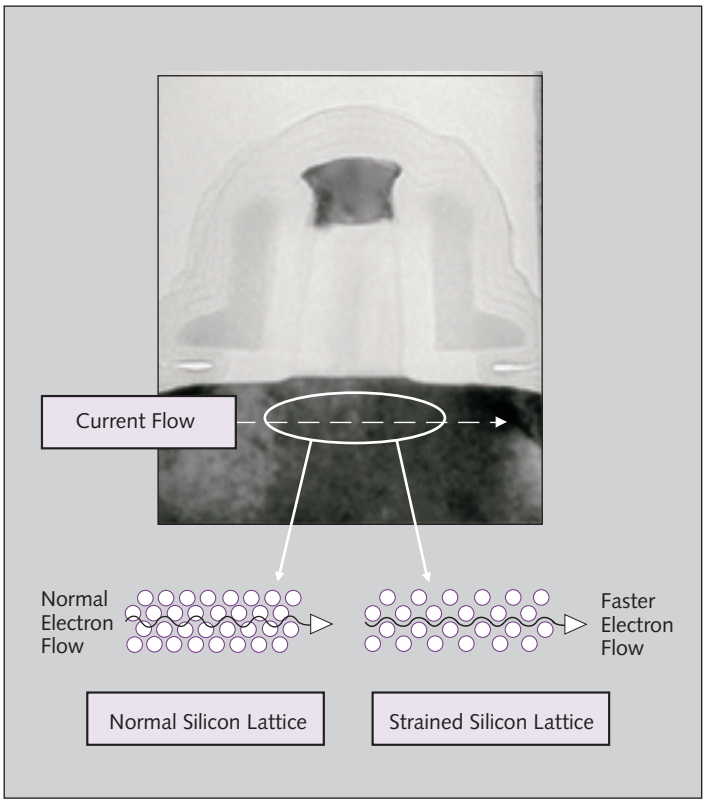

Figure 1. Strained silicon improves electron mobility by stretching the atoms in the silicon layer further apart.

(lattice constant) than silicon, and it exerts tensile stress or "strain" on the tightly bonded silicon layer, stretching the silicon atoms further apart. The resulting improvement in electron and hole mobility increases the drain current of the transistors, which translates into faster switching and greater performance.

Microprocessors built using strained silicon require some design modifications, partly because the technique accelerates n-channel transistors to a greater degree than p-channel transistors. In one experiment at Hitachi, the difference was as great as 3×. Engineers say the circuit-level modifications are relatively minor, especially in view of the performance payoff.

Intel will probably be the first company to use strained silicon in high-volume manufacturing, and it will almost certainly produce more chips with the technology than any competitor. Sticking to its aggressive Moore's Law roadmap, Intel plans to introduce strained silicon with its next-generation 90nm copper process on 300mm wafers in 2H03, only two years after introducing a 0.13-micron (130nm) copper process on 300mm wafers. Intel's 90nm process has an effective gate length (L_{eff}) of less than 50nm. It also has the industry's thinnest gate oxide (1.2nm, less than five atomic layers), carbon-doped-oxide low-*k* dielectrics (18% lower capacitance than the 0.13-micron dielectrics), and up to seven layers of copper (compared with six layers in the 0.13-micron process). Figure 2 is a photograph of Intel's 90nm transistor.

Rolling the Dice

Introducing a new technology like strained silicon at a new process geometry in combination with three other recent innovations (copper interconnects, low-*k* dielectrics, and 300mm wafers) is a risky gamble for Intel, especially considering the enormous stakes. A last-minute glitch anywhere along the line could idle one or more multibillion-dollar fabs and choke the company's inventory for months.

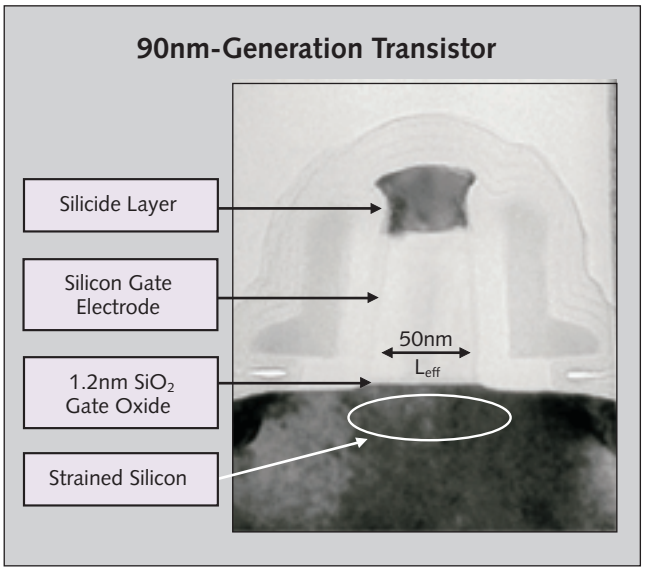

Figure 2. Intel's new transistors have some of the smallest features in the industry.

Intel is undeterred, however. The company has a 0.13-micron line with copper interconnects and 300mm wafers humming smoothly at its fab in Hillsboro, Oregon, and it claims the 90nm strained-silicon line is already demonstrating high yields with a 52Mb SRAM. That memory chip crams six transistors into an area of only 1.0 square micrometer—the industry's smallest SRAM cell. Such tiny cells will allow Intel to integrate larger on-chip caches on microprocessors without inflating die sizes.

The impressive integration of 90nm lithography, copper interconnects, low-*k* dielectrics, strained silicon, and the Pentium 4's 20-stage superpipeline will eventually boost clock frequencies beyond 4.0GHz. To make the actual application performance correlate more closely with the stratospheric clock speeds, the Prescott Pentium 4 will also use Intel's Hyper-Threading technology.

Hyper-Threading is Intel's brand of simultaneous multithreading (SMT), which allows instructions from two different software processes to share a pipeline without time-consuming context switching (see *MPR 09/17/01-01*, "Intel Embraces Multithreading"). The FosterMP version of Intel's Xeon processor introduced Hyper-Threading earlier this year, and the technology will soon migrate to Intel's desktop processors. Hyper-Threading is a boon for the Pentium 4's from-here-to-eternity pipeline, which efficiently crunches data-intensive code but frequently suffers from misprediction penalties when it's running heavily branched code, such as office-productivity software.

Strained silicon may also compensate for Intel's slow adoption of silicon-on-insulator (SOI) technology. SOI speeds up circuit switching by reducing transistor-junction capacitance by as much as 50% (see *MPR 08/24/98-02*, "SOI to Rescue Moore's Law"). PowerPC processors have been using SOI since 1998, and it helped push the pipeline-challenged PowerPC beyond 1.0GHz. An AMD-Motorola technology partnership will bring SOI to the x86 in 4Q02, when AMD rolls out the first 64-bit Hammer-family processors.

AMD hasn't publicly estimated the speedup it expects from SOI. IBM, which pioneered the technology, claims improvements as great as 35% over a bulk CMOS process at the same geometry. AMD's vice president for technology development, Craig Sander, says he's seen 20–35% speedups in test circuits. Sander is reluctant to estimate the real-world improvement with an Athlon processor, because the effects of SOI can vary widely, depending on the circuit design.

Intel doesn't plan to use SOI before moving to a 65nm process in 2005, so strained silicon could help counterbalance any advantage AMD gains from adopting SOI sooner. (SOI and strained silicon are compatible and complementary in the same process.)

Intel's Clock-Speed Lead

At this point, AMD needs every advantage it can get. Intel's superpipelined NetBurst microarchitecture is opening a wider clock-speed gap with the Athlon microarchitecture. The latest Pentium 4 (announced August 26) currently tops out at 2.8GHz, whereas the latest Athlon XP (announced August 21) is the 2600+, which runs at 2.13GHz. Their actual application performance is much closer than the clock-frequency gap implies, but the market is notoriously focused on frequency.

	Die Size	Gross Die Per Wafer 200mm	Gross Die Per Wafer 300mm
Pentium 4 (NorthwoodX)	131mm^2	201	482
Athlon XP (Thoroughbred)	84mm^2	325	Not In Production

Table 1. In-Stat/MDR's estimates of gross die per wafer for the latest desktop processors from Intel and AMD, excluding defects.

Now Intel is threatening to open a manufacturing-cost gap by migrating more rapidly to 90nm lithography on 300mm wafers. At AMD's Fab 30 in Dresden, Germany, the production lines are tooled for 0.13-micron lithography and 200mm wafers. According to AMD's roadmap, 90nm production is scheduled to commence in 2H03 with a 64-bit Athlon desktop processor. That's about the same time Intel's transition to 90nm occurs, except that AMD will still be using the smaller wafers.

When it comes to wafers, size matters. Intel's largest wafers are 50% larger in the linear dimension than AMD's, which means 2.25 times more area. Assuming equal die sizes and defect densities, Intel could obtain at least twice as many gross die per wafer. In practice, the advantage of using 300mm wafers is potentially even greater, because the larger circumference wastes less usable area around the margins, and there's an additional bonus in yields. According to an article in *Semiconductor Magazine* by Nun-Sian Tsai, senior director of TSMC's 300mm pilot project, the absolute number of dust particles that tends to fall on wafers of either size is comparable, owing to the way the wafers are processed. If that's true, actual defect density would also decrease about 2.25 times with 300mm wafers, providing correspondingly higher yields.

Fortunately for AMD, the die sizes in question are not equal. The Athlon XP's 0.13-micron Thoroughbred die is only 64% the size of the Pentium 4's 0.13-micron NorthwoodX die: 84mm^2 compared with 131mm^2. (NorthwoodX is In-Stat/MDR's code name for the Intel Northwood core that recently underwent a 5% linear optical shrink. We use that name to avoid confusion with the older Northwood core in the *Intel Microprocessors Service*, a periodic report In-Stat/MDR publishes.) The appreciable size difference between the Athlon XP and Pentium 4 not only allows AMD to squeeze more die on a wafer, it also decreases the chance that a random defect will spoil a particular die, assuming comparable defect densities—admittedly a major assumption in the absence of closely guarded company data.

Setting aside the issue of defects to isolate the benefit of using larger wafers, we estimate that AMD obtains about 62% more gross die per 200mm wafer than Intel does, thanks to the Athlon XP's small die. But 300mm wafers allow Intel to get about 48% more die per *wafer start*, despite the Pentium 4's greater girth. Table 1 shows our estimates.

Calculating the financial impact of migrating to 300mm wafers is a challenge because of numerous other factors. Two factors on the plus side for Intel: the greater output of 300mm wafers requires the company to build fewer fabs and also fattens profits, because Intel's chips fetch a higher average selling price (ASP). Two factors on the plus side for AMD: Intel's 300mm fabrication equipment is more expensive to buy and takes longer to amortize, and the raw materials cost for 300mm wafers is much higher than it is for 200mm wafers. The cost difference can be as great as eight times per wafer start when a company first migrates to 300mm, according to TSMC. But Intel has had more than a year of high-volume production to drive those costs down; in any case, they are relatively minor compared with other chip manufacturing costs like testing and packaging.

Realistically, of course, AMD cannot match Intel's capital expenditures on new 300mm fabs and process technologies. Intel will spend about $5.5 billion this year, actually a slowdown from $7.5 billion in 2001 and $6.7 billion in 2000; AMD will spend about $850 million.

But AMD doesn't necessarily have to match Intel's spending. One (ironic) advantage of having less than 20% market share is that AMD can roll out new process technologies more rapidly because it has fewer fabs to retool. Indeed, virtually all Athlon production is now concentrated at the Dresden fab (which luckily escaped the recent European floods).

The lighter burden of fewer fabs has helped keep AMD in the process-technology race. AMD was able to introduce copper interconnects at the 0.18-micron level, a whole generation ahead of Intel. AMD's existing 0.13-micron process has eight layers of metal, compared with six for Intel's. That same 0.13-micron process is scheduled to use carbon-doped low-*k* dielectrics sometime around 1Q03, several months ahead of Intel's 2H03 rollout of low-*k* dielectrics in the 90nm generation. Next year's Hammer-family processors will have up to nine layers of metal, compared with seven in Intel's 90nm process. And AMD says it will fully convert to 90nm production only three quarters after the new process debuts in 2H03; Intel's conversions typically take longer.

Alliances Fill the Gap

Still, AMD anticipates a need for more plant capacity and technology development than it can fund alone, so the company is increasingly turning to alliances with other companies for fabrication technology and foundry services. AMD's SOI partnership with Motorola was one example. (That partnership expires after this quarter.)

Price & Availability

Intel's Prescott-core Pentium 4 processor, built with strained silicon in a 90nm process, is scheduled to ship in 2H03. Clock frequencies and prices will be announced in 2003.

Another example is a foundry agreement with Taiwan's UMC to produce chips in a 0.13-micron fab by the end of this year. That agreement could lead to 300mm-wafer production for AMD. UMC is already making 0.13-micron chips on the larger wafers for Xilinx. Manufacturing the diminutive Athlon XP on 300mm wafers wouldn't narrow the clock-speed gap with the Pentium 4, but it would slash costs and significantly relieve the pressure on AMD from Intel's relentless price-cutting.

Fatter margins would allow AMD to plow more money into R&D. In a separate deal, announced earlier this year, AMD formed a joint venture with UMC called AU Pte Ltd. The new company plans to open a fab in Singapore using 65nm lithography on 300mm wafers in mid-2005. Intel's roadmap calls for 65nm production to begin at the same time, so AMD can't afford to slip the schedule. Table 2 summarizes the process roadmaps of the two companies.

	0.13µ	90nm	65nm	300mm Wafers	Strained Si	SOI	Low-*k*
AMD	Now	2H03	2005	2003-2005	Before 2005	4Q02	1Q03
Intel	Now	2H03	2005	1Q02	2H03	2005	2H03

Table 2. Intel and AMD both have aggressive roadmaps for process shrinks, but they have different priorities for new technologies, such as strained silicon and SOI.

UMC doesn't expect to adopt strained silicon until the 45nm generation—the next process shrink below 65nm—which probably won't begin volume production until 2007. Not that UMC is lazy, but almost all its other customers make chips for the embedded market, where demand for nose-bleed clock speeds is less obsessive than in the PC market. (Though judging from the latest sales figures, perhaps PC buyers are now less obsessed with high clock speeds than they once were.)

However, AMD doesn't necessarily have to wait for UMC to adopt strained silicon in order to take advantage of the technology. Sander says his roadmap is independent of UMC's, and that strained silicon could appear in AMD processors before the jointly owned 65nm fab comes online in 2005. Sander notes that AMD's partnership with UMC is more than a standard foundry arrangement, in which the customer merely buys finished chips. The partnership includes joint technology development, with Infineon as another partner. Among other things, AMD will contribute expertise in transistor design, and the companies are working with AmberWave to develop their own flavor of strained silicon. (AmberWave has filed more than 100 related patents and is working with other customers as well.)

Until then, SOI should help AMD maintain parity at the fabrication level, while smaller die sizes will help hold the line on costs. If UMC accelerates Athlon's transition to 300mm wafers, the move should improve AMD's bottom line. Meanwhile, AMD must avoid falling too far behind Intel in raw clock speeds—as the gap widens, it becomes harder to excuse—and try to gain ground on the architectural front with the new 64-bit Hammer processors. ◇