Sun Tweaks SuperSparc to Boost Performance

New SuperSparc-2 Pushes SPARC Performance Beyond Pentium

by Linley Gwennap

Although the company's long-term performance hopes lie with UltraSparc, Sun will use an improved version of its original SuperSparc design to bridge the gap until its next-generation processor is available in 3Q95. The new SuperSparc-2 repairs the biggest problems with the constipated SuperSparc design, unclogging the pipeline and letting it flow much faster. The new processor pushes the clock frequency to 90 MHz, 50% faster than the top speed of SuperSparc.

The faster clock rate of the new chip moves SPARC performance out of the embarrassing sub-Pentium range, where it has loitered since the P54C Pentium began shipping last March. Sun has not yet announced systems using SuperSparc-2 but believes that the 90-MHz version will deliver 135 SPECint92 and 145 SPECfp92. This performance should put the chip ahead of even the 120-MHz Pentium expected from Intel early next year, about the same time as SuperSparc-2.

With time-to-market and performance as the two key design goals, processor cost was forced to take a hit. The new die is even larger than the original, increasing estimated manufacturing cost by 38%. Sun's SPARC Technology Business (STB) is quoting $999 for 75-MHz parts; the 90-MHz price has not been revealed but could be even more expensive. This price/performance lags that of other major RISC vendors.

Breaking the Register Bottleneck

Figure 1 compares the pipeline of the new chip with that of the original SuperSparc (see MPR 12/4/91, p.1). Both require four clock cycles, but the original design breaks down actions on half-cycle boundaries. For example, the register file is accessed twice per clock cycle: once in D1 to provide operands for an address calculation, and again in D2 to fetch operands for an integer ALU operation. If two integer ALU operations are paired, operands for the first are read in D1 and for the second in D2.

SuperSparc's designers chose to double-pump the registers to save the die area needed to add more read ports to the register file, already bloated to support the SPARC register windows. This double access created a critical timing path that made it impossible to push the clock speed beyond 60 MHz, even in a fast process like TI's 0.6-micron BiCMOS.

SuperSparc-2 eases this timing path by implementing a full set of read ports; the new register file is accessed just once per clock cycle. Instead of increasing the

die area, the designers borrowed a trick from UltraSparc and implemented the extra register windows underneath the metal routing needed to multiport the basic registers *(see **081301.PDF**)*. This technique provides adequate read ports with roughly the same amount of die area as the original register file.

Eliminating the register-file bottleneck also allows instructions to access the register file even as they are being fully decoded. The original design had to decode the instructions before allocating the limited register-file ports, creating an extra half-cycle delay (D0 stage).

SuperSparc-2 reads from the register file in D0, pulling the address calculation into D1 and eliminating the D2 stage entirely. The new arrangement also gets the address to the external cache sooner, shaving a cycle off of all data accesses to the L2 cache. Finally, the new design allows synchronization to occur only on clock-cycle boundaries; in SuperSparc, data had to be latched after every half cycle, adding the setup times of these latches to critical timing paths.

TLB Split in Two

Another timing problem in SuperSparc is the 64-entry unified TLB that, like the register file, is accessed twice per cycle. One of the reasons that the execute stage is spread across a clock boundary is to offset the instruction TLB access, which occurs in clock stage 1 ($\phi1$), from the data access to the TLB in $\phi2$. Each of these accesses must complete in one-half clock cycle; this requirement helps prevent SuperSparc from exceeding 60 MHz.

The new processor includes split instruction and data TLBs that each are accessed just once per cycle, easing this timing path. The two TLBs begin each access in $\phi1$ and complete by the end of $\phi2$; the caches are accessed in parallel. To maintain a similar TLB hit rate as in the original design, SuperSparc-2 keeps the data TLB size at 64 entries while adding a 16-entry instruction TLB.
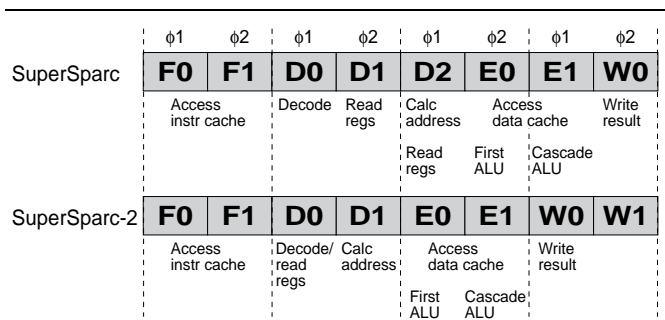
	$\phi1$	$\phi2$	$\phi1$	$\phi2$	$\phi1$	$\phi2$	$\phi1$	$\phi2$
SuperSparc	**F0**	**F1**	**D0**	**D1**	**D2**	**E0**	**E1**	**W0**
	Access instr cache		Decode	Read regs	Calc address	Access data cache		Write result
					Read regs	First ALU	Cascade ALU	
SuperSparc-2	**F0**	**F1**	**D0**	**D1**	**E0**	**E1**	**W0**	**W1**
	Access instr cache		Decode/ read regs	Calc address	Access data cache		Write result	
					First ALU	Cascade ALU		

Figure 1. SuperSparc-2 uses the same pipeline length as the original SuperSparc but shortens the decode stage.

A third timing problem involved the floating-point multiplier, which serves double duty as an iterative divide/square-root unit. To enable higher clock rates, the new chip simplifies the multiplier by adding a separate unit for divides and square roots, retaining the three-cycle latency for multiplies even at higher clock speeds. Divide and square-root latencies, however, have been extended by about 50%. As a result, FP performance does not quite scale with clock speed from the original design: the 90-MHz SuperSparc-2 is rated at 147 SPECfp92, about 40% faster than a 60-MHz SuperSparc.

The remaining portions of the SuperSparc design are carried forward in SuperSparc-2 with little or no change. The sequential and target instruction queues are expanded to 12 entries each, twice the depth of the original design. The direct MBus mode has been removed, since the new chip operates at much higher frequencies than the 50-MHz MBus.

SuperSparc-2 connects to the same VBus support chips as SuperSparc, specifically the MXCC chip that provides secondary cache control and an asynchronous MBus (or XBus) interface. The current MXCC chip can be used with the 75-MHz SuperSparc-2; a new, faster version of the MXCC is needed for the 90-MHz part.

New Features Increase Cost

SuperSparc-2 uses the same 0.6-micron three-layer-metal BiCMOS process (dubbed EPIC-2BE) as the 60-MHz SuperSparc; this decision allowed parts of the circuit design to be copied from one to the other, reducing time to market. The team used automated layout tools for the new sections of the chip rather than compacting

Figure 2. SuperSparc-2 measures 17.4 x 17.2 mm and requires 3.1 million transistors in 0.6-micron three-layer-metal BiCMOS.

them by hand, as the original SuperSparc designers did; this choice increased the die area by about 6%, according to Sun, but helped get the chip to market sooner.

As Figure 2 shows, the new divide/square-root unit and ITLB in SuperSparc-2 increase its die area by 17% over SuperSparc, to 299 mm^2. Because die cost increases roughly with the square of the area, the estimated manufacturing cost of SuperSparc-2 is $255, 38% greater than that of its predecessor, according to the MDR Cost Model *(see 081203.PDF)*.

The faster clock speeds push power consumption as high as 16 W at 90 MHz. To support this increase in power, SuperSparc-2 adds 20 power and ground pins to the package, pushing the size to a 313-pin CPGA. This larger package contributes to the manufacturing cost increase; it also means that SuperSparc-2 is not socket-compatible with SuperSparc. The VBus interface and other signals are compatible, so the layout changes are fairly simple. System vendors using MBus modules will need to make no design changes at all.

SuperSparc-2 would surely be much smaller (and faster) if it were implemented in a more advanced process such as TI's 0.55-micron EPIC-3 *(see 080504.PDF)*; although EPIC-2BE uses 0.6-micron transistors, the metal layers use 0.8-micron design rules. But the designers could not afford the time to port the existing circuit design to a new process; if SuperSparc-2 had been delayed by six months, it would have appeared at the same time as UltraSparc.

As it is, the chip achieved first silicon in July and began sampling just three months later at 75 MHz. A second pass is expected to push the clock speed to the target of 90 MHz. To achieve this level of quality, the entire design was extensively verified and was emulated on a Quickturn system. Of course, the designers were careful to make as few changes as possible to the original design, minimizing the possibility of introducing a bug.

What Went Wrong with SuperSparc

Most of our readers are familiar with the sad story of SuperSparc, but it bears repeating one more time. The chip was originally announced (at the Microprocessor Forum in 1991) at 50 MHz, but initial shipments were made at 33 MHz, due to internal timing problems. This design was eventually pushed as far as 40 MHz in early 1993, but it took a 10% process shrink to achieve the 50-MHz goal. This version (also known as SuperSparc+) began shipping in 3Q93, one year later than promised. Even at 60 MHz, SuperSparc continues to lag other RISC chips in performance.

The announcement of SuperSparc-2 allows us to perform a definitive post-mortem on the original design. The design team knowingly sacrificed clock speed at the altar of complexity but did not realize how demanding its god would be. The inexperienced team never built a
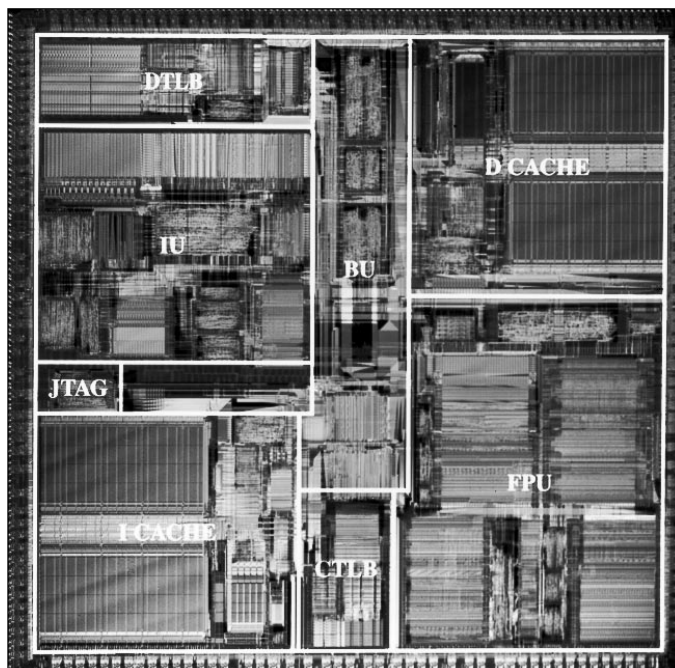
complete timing model of the chip, as such a task was beyond the capability of Sun's tools at that time. Critical paths were not apparent until the company received first silicon from TI and realized the design was in bad shape. The two companies went through several iterations to attempt to fix the problem; each time a critical path was repaired, a new version was fabricated, only to reveal another critical path slightly behind the first.

The team eventually pushed the design to 40 MHz in the original process, 20% less than the target, but could go no further. The convoluted pipeline resulting from the double-pumped TLB and register file, as well as the bottleneck in the FP multiplier, kept the chip from going faster; later performance gains came primarily from the gate shrinks. The cascaded ALUs apparently were not a performance problem, as they remain in the new version. But if the design team had recognized the problems it was getting into, the original SuperSparc would have looked more like SuperSparc-2.

Another effect that bit the SuperSparc design: bipolar transistors have a much narrower frequency-yield curve than CMOS transistors. A CMOS chip, when manufactured in volume, will yield significant numbers of chips at 20% or even 30% better speeds than the center frequency of the curve. Due to the bipolar component, a BiCMOS chip may have a range of only 5–10%. Thus, Sun couldn't even skim fast chips off the top for its high-cost systems. (This effect is seen in the BiCMOS Pentium, for which speed grades vary by only 10%.)

Sun does not expect to achieve significant yield at both 75 and 90 MHz. The initial design yields mainly at 75 MHz, but by fixing a few timing paths, Sun expects to move the yield curve to 90 MHz for the production version. There is no obvious path to further speed increases, as TI currently has no 0.5-micron BiCMOS process.

Treating the Symptoms

While the new SuperSparc-2 processor should ease the woes of SPARC users begging for more performance, it does not significantly improve the position of SPARC processors in comparison with other RISC chips. Figure 3 compares the 90-MHz SuperSparc-2 with the fastest processors expected to be available from other vendors in 1Q95. Although this chip closes the gap with MIPS, it still leaves SPARC well behind the other RISCs.

Like its predecessor, SuperSparc-2 does not make up for this lack of performance by offering lower cost; indeed, it is just the opposite. According to our estimates, SuperSparc-2 is the most costly processor of its generation, 15% more than the PA-7200 and at least 60% more than any of the other chips shown in Figure 3.

The price is no bargain either: the 75-MHz SuperSparc-2 costs twice as much as a 90-MHz 604, for example, but delivers less performance. Even the stately 21064A, at its top speed, sells for only 20% more than the

SPARC chip while achieving 75% better integer performance. In addition, the Alpha chip blows SuperSparc-2 out of the water on FP code.

By offering minor changes to SuperSparc, Sun's latest effort is merely an ice pack soothing its high-end workstation users. To truly cure Sun's price/performance problems, the SuperSparc design must be completely exorcised from Sun's product line. Unfortunately, it appears that UltraSparc-1 will simply push SuperSparc-2 into the midrange rather than displace it entirely, leaving Sun with a part that is too expensive and underpowered to compete in that market.

For the past two years, Sun has retained its industry-leading share of the workstation market despite SuperSparc's shortcomings. Sun has achieved this feat by aggressively pricing its SuperSparc systems and essentially giving up on the very high end of the market, which simply isn't that large in unit volume. The midrange, however, is Sun's most important market, and with SuperSparc-2 filling that spot, the company could see some hiccups in either its market share or its gross margins. But help is on the way: MicroSparc-3 *(see 081301.PDF)* is due in early 1996 and should finally rid Sun of the curse of SuperSparc. ♦
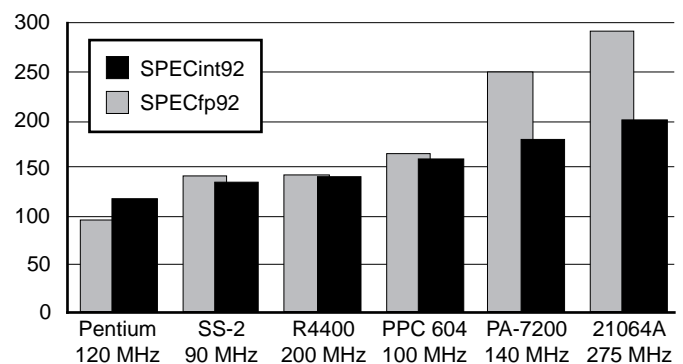

Figure 3. SuperSparc-2 exceeds the performance of Pentium but lags all other high-end RISC processors.