R5000 Improves FP for MIPS Midrange

New MIPS Processor Boosts Clock Speed, Floating-Point Scores

by Linley Gwennap

Silicon Graphics' popular Indy workstation line will get a big boost from the R5000 processor, due this spring. The new chip, a follow-on to the Orion design, combines improved integer speed with strong floating-point performance for a midrange processor. At 200 MHz, the R5000 is expected to deliver better than 5.2 SPECint95 and 5.2 SPECfp95 in initial systems. The compact design fits into just 84 mm^2 of silicon, allowing the 200-MHz version to sell for $365, with lower prices for slower speeds.

The original Orion chip, the R4600, helped the Indy line become a top seller by providing solid integer performance at a low price. The R4600 *(see* **061507.PDF***)* was designed for NT-based PCs, however, and so included a relatively slow floating-point unit, hampering its performance on some workstation applications. The R5000 features a much faster FPU and adds the ability to dispatch an integer and an FP instruction in the same cycle; these features help increase SPECfp/MHz by 30% compared with the R4600. Figure 1 shows a block diagram of the new part.

Like Orion, the R5000 was developed by Quantum Effect Design (QED), led by ex-MIPS designers Earl Killian (now back at MIPS) and Tom Riordan. Unlike Orion, the R5000 was funded by MIPS Technologies and has been licensed by IDT, NEC, and NKK. Both NEC and IDT are now sampling the device and expect production in March; NKK is a few months behind.

In addition to SGI, system maker Siemens-Nixdorf plans to use the R5000. The chip, particularly future derivatives, should also make its way into high-end printers and other embedded applications. Other MIPS partners, however, are ignoring the R5000 for now, focusing on the higher performance of the forthcoming R10000.
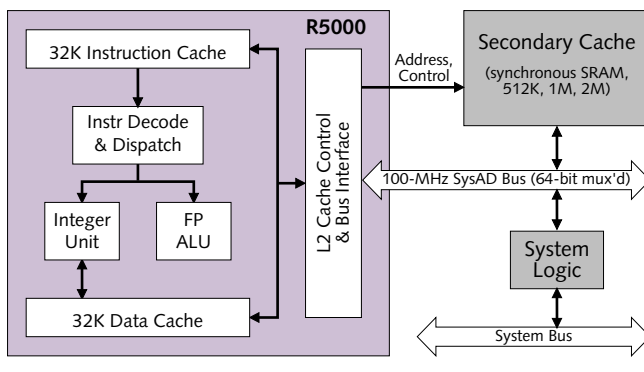
Limited Superscalar Dispatch

Although the R5000 is the intellectual heir to Orion, much of the logic design and most of the circuitry are changed in the new version. The main pipeline is much the same, but the addition of superscalar instruction execution caused significant changes in the control logic. To keep things relatively simple, the R5000 can pair instructions only if one is an FP calculation (add, multiply, etc.) and the other one isn't. Thus, for integer code, the R5000 is still a scalar device.

For floating-point code, however, this pairing algorithm works well. A series of FP calculations can be launched at a rate of one per cycle, with the second instruction slot used for either FP loads, to fetch data for a vector operation, or for integer instructions that update indexes and handle branching, for example. The R5000 includes a multiply-add instruction (as well as the other MIPS IV extensions found in the R10000) and can launch a single-precision multiply-add on each cycle, achieving 400 MFLOPS (peak) at 200 MHz. The second instruction slot allows the chip to sustain this rate for at least some meaningful calculations.

In contrast, a 133-MHz R4600 can generate just 44 MFLOPS. To save die area, that design uses a nonpipelined FPU that allows only partial overlapping of multiply and add operations, launching a new multiply-add combination every 6 cycles, with a 12-cycle latency. As Table 1 shows, the R5000 FPU is fully pipelined, with a latency of just 4 cycles, for most single-precision operations. Divide performance is also nearly doubled.

Although the R5000 FPU is a major improvement over its predecessor's, its latencies are on par with or slightly behind those of other midrange processors. The PowerPC 603, for example, is a cycle faster on most FP operations, and the PA-7100LC is two cycles faster. High-end processors are even better, of course, and most include a full double-precision FPU that eliminates the R5000's penalty cycle for DP calculations. While this penalty may not seem like much, it

Figure 1. The R5000 contains a simple superscalar core and large on-chip caches. It also directly controls an external cache.

	R5000		Orion	
	Single	Double	Single	Double
Integer multiply	4/5	8/9	6/8	6/8
Integer divide	36/36	68/68	32/32	61/61
FP add	1/4	1/4	4/4	4/4
FP multiply	1/4	2/5	6/8	6/8
FP multiply-add	1/4	2/5	6/12*	6/12*
FP divide	19/21	34/36	32/32	61/61
FP square root	36/38	66/68	31/31	60/60

Table 1. A comparison of the throughput and latency for various integer and floating-point math operations shows significant improvements for the R5000, particularly in FP throughput. *FP multiply-add requires two instructions in Orion. (Source: MIPS)

halves the R5000's peak MFLOPS rate, to 200, for double-precision calculations.

Superscalar dispatch is accelerated by predecoding instructions, a technique used in several recent devices. In the R5000, four bits per instruction pair indicate whether the instructions can be paired and the target function unit for each. In addition to multiply-add, the chip also includes the MIPS IV conditional move (CMOV) instruction, which is particularly useful for graphics cliptest functions.

By focusing the R5000 on single-precision math, MIPS has tailored the part for the 3D visualization market, SGI's stronghold. Single-precision operations are adequate for visualization due to the limitations of the human eye and monitors; double-precision math is used mainly for scientific calculations. The R5000 has the power to manipulate complex objects in a three-dimensional space using only a low-cost graphics subsystem to perform simple tasks such as scaling and color conversion. Current Indy systems require a more expensive graphics engine to perform similar visualization applications as quickly.

Cache Subsystem Improved

Improvements in integer performance are due mainly to the improved cache subsystem. The on-chip caches are expanded to 32K each, twice the size of Orion's, while retaining the same two-way associativity. As Figure 2 shows, the die continues to be dominated by these caches, which (including tags and control logic) consume as much area as the rest of the circuitry combined. Doubling the cache increases SPEC-int/MHz by 15%; another good reason for tue increase is that the die would have been pad-limited with a smaller cache, so the extra performance comes with little cost.

Performance can be improved by adding an external level-two cache. While Orion allows for an L2 cache, this cache must sit on the main system bus and be controlled by an external agent. Since the system bus typically runs at 50 MHz, L2 cache bandwidth is limited.

As Figure 1 shows, the R5000 includes its own L2 cache controller, improving the timing of cache accesses. The external agent no longer needs to control the L2 cache; instead, in an R5000 system, the agent acts as a bus buffer. This buffering permits the L2 cache to run faster, typically 100 MHz, while allowing the main system bus to run at a more relaxed pace. With custom synchronous SRAMs, the latency to the critical doubleword is just two cycles. These improvements alone add 5–10% in overall performance, or up to 20% compared with a system with no L2 cache at all, a common Orion configuration.

In other respects, the bus remains compatible with the venerable 64-bit SysAD bus used by several other MIPS processors. The R5000 clocking is improved from that of the R4600, reducing the number of clock signals from three to one. Although it is not pin-compatible with its predecessors, the R5000 offers a straightforward modification to existing Orion-based and R4000-based systems.
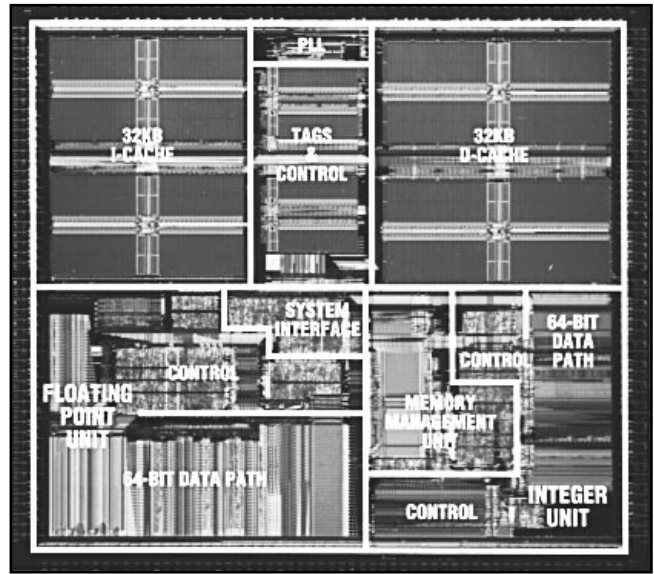

Figure 2. IDT's version of the R5000 contains 3.6 million transistors and is built in a 0.35-micron CMOS process with three metal layers and two poly layers; it measures 8.7×9.7 mm (84 mm^2). Other vendors' versions vary slightly in size and transistor count.

Like Orion, the initial R5000 operates only in uniprocessor systems. The new chip, however, reserves eight pins for multiprocessor support. MIPS promises that future versions of the R5000 will take advantage of these pins to operate in configurations of up to four processors.

New Process Speeds Clock

Additional performance gains come through an increase in clock speed. The 0.65-micron R4600 operates at 133 MHz, while the 0.5-micron R4700 jumps to 175 MHz. Moving to a 0.35-micron process pushes the R5000 to 200 MHz. Although this 14% increase appears small for such a process shrink, it's just the start: MIPS expects its partners to deliver 250-MHz parts off the same 0.35-micron process later this year. The R5000A, due in 1997, will further increase clock speed, probably through a shrink to 0.25-micron geometries, and add some features.

The process shrink keeps the R5000 die to about the same size as the R4600 despite the enhanced FPU and larger caches. IDT's version of the R5000, pictured in Figure 2, measures 84 mm^2 using a two-poly process, like most other MIPS chips. NEC and NKK both have switched to a single-poly process; this move reduces processing costs but forces a move to six-transistor (6T) SRAM cells rather than the 4T cells in IDT's design. For these vendors, the larger SRAM cells swell the die slightly, to 87 mm^2, but the elimination of the extra poly layer results in an overall cost reduction.

The vendors are offering the R5000 in a thermally enhanced plastic ball-grid array, Amkor's SuperBGA *(see* 091304.PDF*)*. The 272-lead package has a lower cost than conventional ceramic BGA or PGA products, yet it can handle the 10-W maximum dissipation of the R5000, which

runs at 3.3 V. The MDR Cost Model estimates that the chip, in this package, costs about $50 to build, $10 more than the R4600, due to the higher wafer cost and defect density of the advanced 0.35-micron process. The new chip is also available in a 233-pin ceramic PGA package.

The R5000 has about the same build cost as the PowerPC 603e and is 20% less costly than Pentium. All three chips have a similar die size in a 0.35-micron process; Pentium is a bit more expensive due to its four-layer-metal BiCMOS process and PGA package. The 166-MHz Pentium *(see* **1001MSB.PDF***)* offers integer performance similar to the 200-MHz R5000 but far less floating-point performance; the 166-MHz 603e trails the MIPS chip in both areas.

MIPS claims that the R5000 is capable of delivering 5.5 SPECint95 and 5.5 SPECfp95 (base) at 200 MHz. These ratings are not yet measured, however, and assume a 2M pipelined L2 cache running at 100 MHz and a 200-ns main-memory system. The first generation of R5000 systems, based on R4600 motherboards, is likely to come closer to 5.2 on both SPECint95 and SPECfp95, due to less powerful system configurations.

R4400 to Fade Out

In combination with the R10000, the R5000 will completely revamp the lineup of MIPS desktop processors. The R5000 carries the midrange cost structure of its Orion parentage but actually outperforms the fastest R4400 processors on most benchmarks. Thus, the R5000 will quickly obsolete the R4400 in desktop systems. The older device holds an edge in large servers due to its ability to support a larger cache (4M), higher cache bandwidth, and multiprocessor support. Most of these systems, however, are moving to the R10000; the R5000 will probably take care of the remaining designs once the chip gains multiprocessor support. One advantage for servers: the R5000 supports up to 16G of main memory.

From a CPU viewpoint, the R5000 delivers unprecedented price/performance of 1.7 SPECint95 and 1.7 SPEC-fp95 per $100. The PowerPC line, in contrast, is around 1.0 and Intel's line is below 1.0 on the integer side and closer to 0.5 for FP. The significant spread between the R5000's $275 price and $50 cost will allow its vendors to aggressively reduce the chip's price over time.

Silicon Graphics' price structure, however, dictates that the R5000 will appear in systems that sell for between $5,000 and $15,000. IBM is already selling 604-based systems in this price range *(see* **090803.PDF***)* with performance similar to the R5000's; the 604e, due shortly after the MIPS chip, should outrun the R5000 on SPECint95. HP's midrange 7300LC, expected in 3Q96, is slated to outperform the R5000 as well, particularly on SPECfp95.

More brutal competition could come from Pentium Pro systems, which will also appear in this price range despite a processor price nearly four times greater than the R5000's. At equivalent clock speeds, Pentium Pro delivers about the same SPECfp95 performance as the R5000 but 50% better SPECint95 scores.

R5000 Targets 3D Market

The R5000 has one advantage over these other parts: its single-precision FP performance, which is not measured by SPECfp95, is significantly better than its double-precision performance. Thus, it could outperform its rivals on applications, such as 3D visualization, that run in single-precision mode. This advantage plays well to SGI's core markets but may not help in the broader general-purpose market. As the clock speed of the R5000 increases, its performance will become more attractive over time.

IDT has had significant success pushing the low-cost Orion design into the embedded market, with design wins at Cisco and Canon, among others. The company is focusing the R5000 as well on embedded designs. The enhanced FPU will significantly improve Postscript rendering speed, making the chip ideal for high-end printers.

Most embedded designers, however, don't care about FP; for these users, IDT is likely to develop derivative versions that eliminate the FPU, simplify the bus interface, and possibly reduce the on-chip cache size. Taken together, these changes will cut the die size even further, allowing the chip to address high-volume designs. Future process shrinks will also help reduce the chip's cost and power dissipation.

Some embedded users may jump to the R5000 to take advantage of its multiple sources. The R4700, for example, is sold only by IDT. Since AMD's pullback of the 29000, system designers are more eager to avoid single-sourced parts.

The design changes in the R5000 reflect changes in the MIPS marketing strategy. Three years ago, Orion was designed specifically for Windows NT PCs, and there were high hopes that the NT/MIPS combination would break into the volume PC market. This strategy has failed, and today neither SGI nor IDT have any interest in NT. The R5000 is clearly focused on boosting SGI's position in the workstation market, and it is well designed to meet these needs. ♦