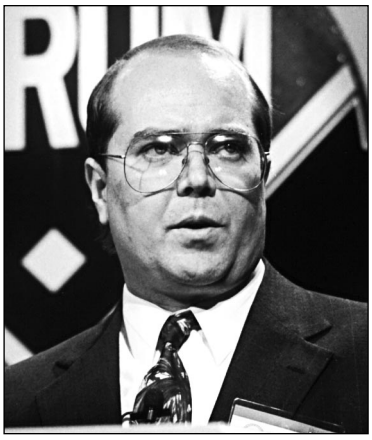
Integrated PA-7300LC Powers HP Midrange

In 3Q96, Today's High-End Performance with Low System Cost

by Linley Gwennap

Next year, Hewlett-Packard will rearm itself for the performance battle after losing ground over the past year or so. In 1Q96, HP expects to debut the PA-8000 in high-end systems, grabbing the overall performance lead. The company revealed at last month's Microprocessor Forum a second product, the PA-7300LC, that should revitalize HP's low-end and midrange systems as well, starting in 3Q96.

HP designer Tom Meyer expects the 7300LC to outperform every processor shipping today except Digital's 21164. Yet the chip's integrated design allows it to be used in relatively low cost systems. Like the PA-7100LC, its predecessor, the new chip includes cache and DRAM controllers. A new feature is 128K of on-chip primary cache, made possible by a move to HP's 0.5-micron CMOS process. Although high-end RISC chips from a number of vendors are likely to surpass the 7300LC's performance before it debuts, the new HP chip will greatly improve HP's competitiveness in its price class.



MICHAEL MUSTACCHI

At the Forum, HP chip designer Tom Meyer describes the highly integrated PA-7300LC CPU.

The new chip uses the same CPU core as the 0.8-micron 7100LC and borrows much of its system interface as well. Performance is increased through the new on-chip cache and a clock-speed increase from the new process. Meyer would not reveal the clock speed of the new chip, but the 0.5-micron process should allow the 7100LC core to reach speeds of 160 MHz or so. The chip taped out on October 11; with its proven CPU core and proven IC process, 3Q96 system shipments seem a realistic goal.

Simple Superscalar Core

Because it uses the 7100LC core *(see **061504.PDF**)*, the new chip is a two-way superscalar processor, relatively simple by current standards. Its dual integer units can pair two ALU operations or an ALU operation with a load or store. The cache is single ported, so loads or stores can be paired only if they access consecutive aligned words in the cache.

The design includes HP's hallmark, a low-latency floating-point unit. It also includes the PA-RISC multimedia extensions *(see **080103.PDF**)* pioneered by the 7100LC. These extensions perform some parallel arithmetic but are relatively modest compared with Sun's VIS extensions. Using its extensions, the 7100LC can decode MPEG-1 video in software, but this feat is now routinely achieved by processors, including high-end Pentiums, without multimedia instructions.

Large On-Chip Caches New for HP

The key to the 7300LC's performance, and the major distinction from its predecessor, is its massive on-chip cache memory. With 128K, the chip includes more on-chip memory than any other microprocessor, surpassing the 112K of the 21164. Instead of the two-level structure used by the Digital chip, in which all but 16K requires six cycles to access, the 7300LC divides its memory into two 64K caches that can each be accessed in a single cycle.

The mere existence of these caches is surprising in a PA-RISC chip; no previous HP design has included more than 2K of on-chip memory. In the past, HP has argued that small on-chip caches have too high a miss rate on many applications to make them worthwhile, and that they are not worth the overhead of managing multiple levels of cache. Meyer points out that the 0.5-micron technology used for the 7300LC allows, for the first time, large enough on-chip caches to provide a reasonable hit rate. In fact, low-end 7300LC systems may not implement any external cache, a significant cost savings.

A two-way set-associative design further improves the hit rate of the caches. Previous PA-RISC processors used direct-mapped caches, as associative external caches generally increase pin count, but this is not an issue with on-chip caches. The on-chip data cache is nonblocking; it can continue to respond to requests with up to two misses outstanding.

This feature allows limited out-of-order execution, as instructions can continue executing while previous loads wait for data. The chip avoids complicated out-of-order hardware by checking for most traps before a load or store is sent off chip; for example, a TLB protection fault is detected before the results of any subsequent instruction have been written to the register file. Traps that occur during an access to external memory, such as a double-bit ECC error, are not recoverable—the process

that received the trap must be aborted.

As Figure 1 shows, the caches consume about half (47%) of the 259-mm^2 die. The caches include numerous redundant columns, allowing most defects in these areas to be repaired. Thus, the yields of the 7300LC should be relatively high despite its large die size. The MDR Cost Model estimates that the 7300LC will cost about $175 to build, about the same as the PA-7200 and one-third the cost of the PA-8000.

The unified TLB is similar to the 7100LC's. It is fully associative and has 96 entries, 50% more than its predecessor.

Fast External Cache

As Figure 2 shows, the 7300LC has a single external cache. This level-two (L2) cache is built from synchronous SRAMs that typically cycle at one-half the CPU clock speed, although slower speeds are supported to reduce cost. With a 128-bit interface, a typical 1M cache uses eight 64K×16 SRAMs for the data plus two 32K×8 parts for the tags.

For systems with an L2 cache, the minimum recommended size is 1M, as smaller caches provide little performance benefit over the 128K of on-chip cache. The 7300LC supports up to 8M of external cache.

When the primary data cache is accessed, the address is sent to the TLB and the cache in parallel, as the primary cache is virtually indexed. One cycle later, the TLB produces a physical address, and the chip determines if the access hit in the primary cache. On a miss, the physical address is immediately sent to the L2 cache. Since the physically indexed L2 cache must wait for the TLB access to complete in any case, there is essentially no delay in starting the L2 access.

With a half-speed cache, the 7300LC has six cycles of latency on an L1 cache miss that hits in the L2 cache. Using its wide bus, the L2 cache returns a complete 32-byte L1 cache line in just two accesses (6-2 access time with a half-speed cache). HP claims this cache delivers a peak bandwidth of 1.3 Gbytes/s, supporting our estimate of a 160-MHz CPU clock speed for the 7300LC.

The requested data word is always read from the cache on the first access and bypassed directly to the CPU. The full cache line is written into a buffer, and the primary cache is then updated in a single cycle, reducing the number of cycles that cache is blocked from CPU accesses. The data cache has two of these buffers, so the chip can try to fit in cache updates when the CPU is not accessing the cache.

Memory Controller on Chip

Like the 7100LC, the new chip includes a DRAM controller on chip. This direct connection eliminates the overhead of sending memory requests as external bus transactions and improves performance. With a low memory latency and large on-chip caches, the performance without any external cache at all is just 10–20% less, depending on the application, than with a 1M cache.

The main memory shares the 128-bit data bus with the L2 cache. As Figure 2 shows, a set of FET switches disconnects the main memory from the bus when the L2 cache is accessed. This design essentially combines the separate 64-bit buses used for cache and DRAM in the 7100LC, doubling the bandwidth to memory without increasing the pin count. The single bus also allows

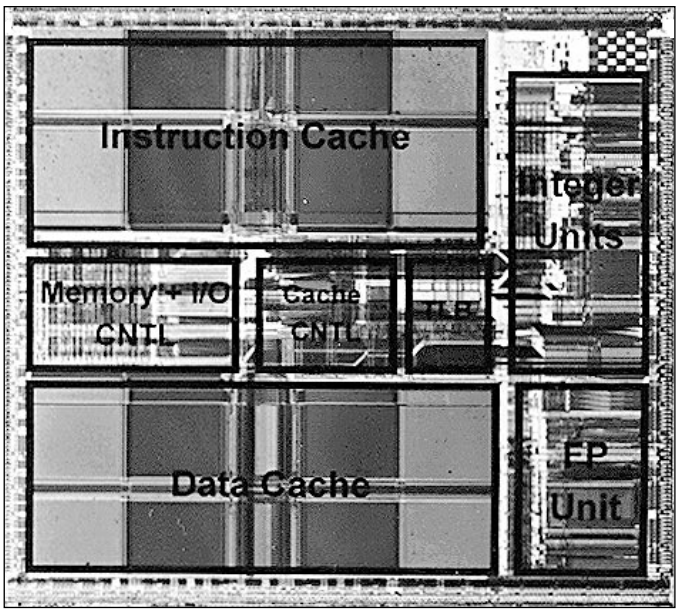


Figure 1. Die photo of the PA-7300LC shows the large on-chip caches. The die contains 9.2 million transistors and measures 15.3 × 16.9 mm (259 mm^2) in 0.5-micron four-layer-metal CMOS.

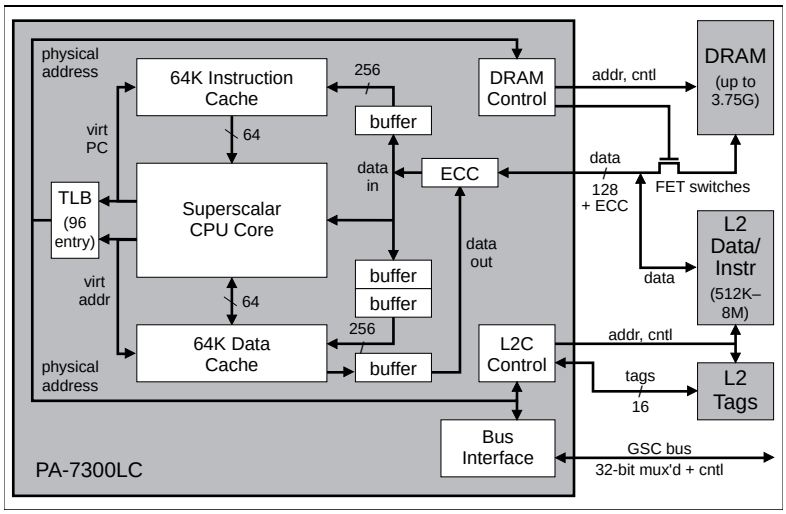
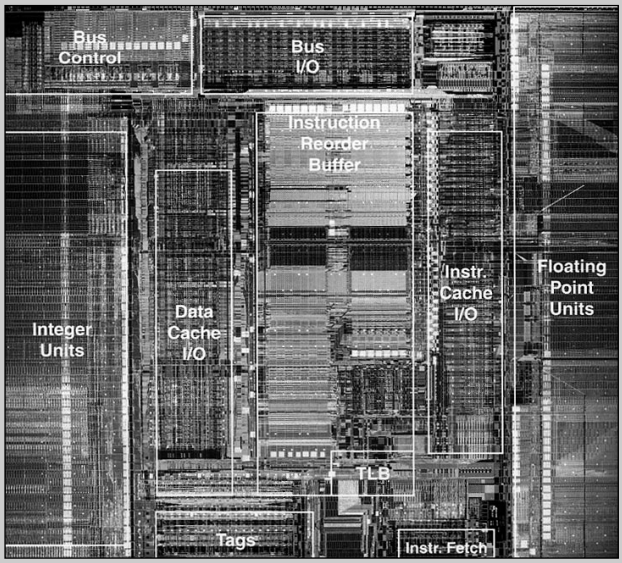


Figure 2. The 7300LC connects directly to an optional L2 cache and to main memory, while graphics and I/O devices can connect to the GSC bus, an HP standard design.

PA-8000 Stays on Track

After taping out last March, the PA-8000 is on track for initial system shipments in 1Q96, as the company indicated when the design was first disclosed *(see **081501.PDF**)*. HP is now sampling to its partners prototypes that run at speeds up to 150 MHz, which is more than 80% of the target frequency. The company still has not disclosed the official target clock speed, but simple math indicates that it must be around 180 MHz, slightly less than the 200-MHz speed we predicted last year. HP claims that it has found no major functional defects after six months of testing.

The performance of the device is also meeting expectations. When the PA-8000 was first disclosed, an HP spokesperson boasted that the processor would be the fastest in the world when it shipped. The company now believes that the PA-8000, at its target clock speed, will deliver more than 8.6 SPECint95 and more than 15 SPECfp95 (baseline). These figures exceed those of any processor expected to ship by 1Q96, giving HP a strong chance to make good on its boast.



The PA-8000 die, shown above, is the largest of the next-generation RISC chips, measuring 347 mm^2 in a 0.5-micron four-layer-metal process (the same as the 7300LC), despite containing no on-chip cache. The dual floating-point multiply-accumulate units are mirror images and occupy 19% of the die, while the integer units take up 17%. The heart of the chip is the instruction reorder buffer, which consumes 15% of the die. The two instruction queues can be seen clearly.

Because it uses flip-chip bonding, the die has no pad ring. Much of the space, however, is used for I/O drivers for the more than 700 signals that travel off chip. These drivers are spread across the die rather than being concentrated around the edge, but the large number of I/O signals helps make this an immense die.

main-memory data to be stored directly into the L2 cache as it is bypassed into the CPU. The downside is the small cost of the external switches and the need to implement a 128-bit-wide DRAM array.

If there is no other memory activity in progress, the CPU sends an address to the main memory at the same time that it is broadcast to the L2 cache (if present), minimizing DRAM latency. If the access misses the L2 cache, the DRAM's row-access (RAS) time is already past; the DRAM access is then completed by asserting CAS.

With 60-ns fast-page-mode DRAM, the CPU stalls for only 14 cycles on an L1 cache miss (16 if there is an L2 cache, because RAS must be delayed until the L2 miss is detected). Because of the nonblocking cache, some of this latency can be hidden if the requested data is not immediately required by another instruction. With the 128-bit bus, it takes only two accesses to fill an L1 cache line, for a total of 22 (or 24) cycles.

The memory controller is quite flexible in its timing, allowing it to support a variety of memory chips, including EDO DRAM. The two-access refill reduces the advantage of EDO for typical desktop applications, but it is advantageous for transaction processing and other applications that require frequent DRAM accesses. HP's quoted performance estimates for the 7300LC assume a more conservative fast-page-mode memory subsystem.

The main memory can be up to 3.75G, the maximum allowed in the PA-RISC architecture, and includes 16 bits of ECC. Because the external cache shares the same data path, it also supports ECC. With a wide data path, ECC requires the same ×9 or ×18 SRAMs as a per-byte parity scheme and allows error correction, making it an increasingly popular feature for high-end processors that support large external caches.

Flexible System Configurations

The 7300LC can be used in much less expensive systems than high-end PA-RISC chips because of its integrated design and flexible system configurations. The PA-7200 and PA-8000 both require two 128-bit-wide caches that cycle at 120–180 MHz. A complete system thus requires a large number of expensive SRAMs. The 7300LC uses a single cache built from 50–83-MHz synchronous SRAMs that are much less costly than the faster parts.

The new processor can thus achieve a variety of cost/performance points. Entry-level systems will use the 7300LC without any external cache. More expensive systems can support a 1M cache at perhaps a third of the CPU speed. Midrange desktops will couple the 7300LC with a half-speed cache of 1M; this is the configuration used to generate HP's performance estimates. Midrange servers will probably increase the external cache to 2M–4M, possibly at a slower speed, to improve performance when juggling a larger number of tasks.

The I/O bus is the same as the 7100LC's, which HP calls GSC. Thus, the new chip can use the same graphics and peripherals designed for the existing systems, and end users can retain their add-in cards when upgrading to a 7300LC box. GSC is a 32-bit multiplexed bus that operates at 40 MHz. Since memory traffic is handled by a separate bus, this bandwidth should be adequate for graphics and I/O devices.

Crashing the Performance Boards

Figure 3 compares the projected 3Q96 performance of the five major RISC vendors and Intel for both high-end and midrange chips. The PA-8000 (see sidebar "PA-8000 Stays on Track", above) should take the SPEC95 performance lead if it ships in 1Q96. We project that by 3Q96, however, both the 21164A and Pentium Pro will surpass it on integer code, and the 21164A and perhaps the R10000 may beat the HP chip on SPECfp95.

Among midrange processors, HP should be in the thick of the performance battle with the 7300LC. We expect that only the PowerPC 604e will beat the HP chip on SPECint95, and only the entry-level UltraSparc will come close on SPECfp95. Of course, these projections assume that the vendors deliver their products as planned, but the midrange products are all versions of existing products—or, in the case of the 7300LC, highly leveraged devices—so the risks are relatively small.

(The chart also shows a general increase in competitiveness among the various vendors. Although Digital has held nearly a 2:1 performance advantage over many of its competitors throughout much of 1995, by the middle of next year, the gap between the integer performance leader and the performance laggards is less than 35%. The gap is only 25% for midrange chips, even including Pentium. But on SPECfp95, there is still a large gulf separating the best from the worst.)

To even challenge for the performance lead is a step forward from HP's current position. With its PA-7200 reaching the market late and below performance expectations, the company's high-end position has eroded, leaving PA-RISC 40% behind Alpha on SPECint95. The 7100LC has been carrying HP's midrange for nearly two years and now trails chips like the PowerPC 604 by 35% in integer performance. The 7300LC is the key for HP to improve its competitiveness in low-end and midrange systems, the bulk of its desktop and server lines.

Performance Surges, But for How Long?

In the past, HP's competitive edge has ebbed after new processor introductions, as the company has failed to improve the performance of its base designs. For example, by 3Q96, the 7100LC will have carried HP's high-volume products for two and a half years without a performance upgrade, an untenable situation in the fast-moving microprocessor game. (Contrast this arrangement with the aggressive UltraSparc roadmap.) *(see **091505.PDF**)*

The failure to upgrade the 7100LC for so long is indicative of HP's lack of processor design resources. A simple port of the 7100LC design to the 0.5-micron process would have offered a kicker until the 7300LC could be completed, but such a product never appeared.

Another problem is HP's lagging IC process technology. It seems ludicrous for a new processor design to debut in 3Q96 using 0.5-micron CMOS; by that time, every other major microprocessor vendor will have 0.35-micron devices in production, and some will have already moved below 0.3 microns. The 7300LC would be a phenomenal product in a 0.35-micron process, but by the time HP delivers such a device (if ever), the market will have moved on.

HP's partnership with Intel should provide a long-term solution to these problems. Sharing the burden of processor development with another vendor should greatly increase the number of CPU designs and the rate at which they appear. HP also gains access to Intel's industry-leading manufacturing technology and is likely to cede all processor fabrication to its partner. The fruits of this alliance, however, probably won't appear before 1998. The 7300LC, along with the PA-8000, should make 1996 a good year for HP, but a slow decline may follow until the new HP/Intel products debut. ♦

Price & Availability

HP expects the PA-7300LC to ship in systems starting in 3Q96. The company does not sell its processors on the merchant market and does not reveal pricing. For more information on HP systems, contact the company at 800.637.7740; fax 800.231.9300 or access the Web at *www.hp.com/go/computing*.

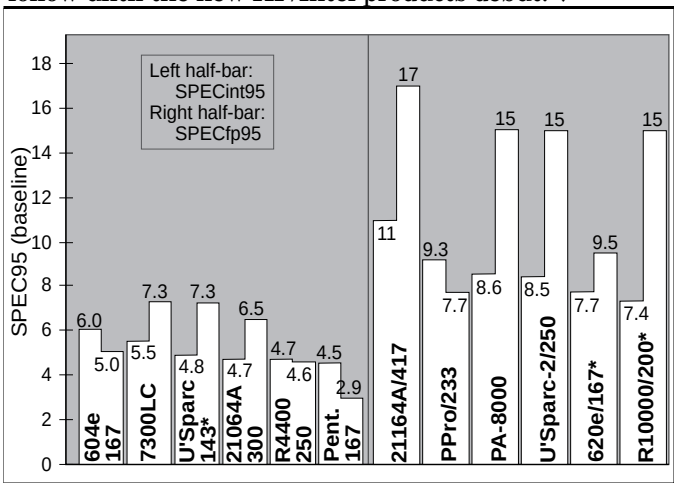


Figure 3. A comparison of midrange processors expected to ship in 3Q96 shows the PA-7300LC near the performance lead. Among high-end CPUs in the same timeframe, the PA-8000 also fares well. (Source: MDR based on vendor SPEC95 estimates except *based on vendor SPEC92 estimates)

