

Low-End PA7100LC Adds Dual Integer ALUs

By Brian Case



At the Microprocessor Forum, Mark Forsyth of Hewlett-Packard presented the first details on the upcoming 7100LC PA-RISC implementation aimed at entry-level systems. Internally, the chip was known as "Hummingbird," and the 7100LC designation used at the Forum may be changed by the time the chip is formally introduced.

This single-chip, two-issue superscalar microprocessor integrates dual integer units, a floating-point unit, MMU, cache control, DRAM control, and a bus interface. Some "multimedia" hardware for manipulating pixel and audio sample data is also included but was not described in detail. This chip is probably intended to be the basis for a line of low-end, under-\$5000 PA-RISC workstations.

Since the introduction of HP's PA-RISC family of workstations, the goal of each new processor design has been to enable a performance increase over previous family members through improvements in clock speed, organization, and cache size. This emphasis on raw speed has kept PA-RISC at the forefront in workstation performance, at least in terms of SPEC ratings.

One weakness in the PA-RISC line of workstations—and arguably in most workstation families—has been the lack of an inexpensive low-end offering. To address the low-end deficiency, HP is well along in the development of the 7100LC. The 7100LC integrates system functions and reduces cache complexity to lower system cost, but it will still offer the potential of very high performance. To keep performance competitive with current HP workstations, the 7100LC uses the most aggressive superscalar organization yet revealed for a PA-RISC microprocessor.

Chip Overview

HP designed the 7100LC and will fabricate the chip in a 0.8-micron, 3-level-metal CMOS process. The die size of this 800,000-transistor chip is expected to be 14 mm on a side, which is quite large for what is supposed to be an inexpensive chip. This process will yield chips that operate at 50 to 75 MHz. Eventually, HP will move production to a 0.6-micron process, which will produce smaller and faster chips. HP plans to make the 7100LC available to its system partners, but there are no plans yet to offer it on the merchant market.

The 7100LC's 432-pin PGA package

has the potential to be expensive, but HP claims that it has worked directly with a PGA vendor to arrive at a cost reduction that makes the PGA package appropriate for low-cost systems.

As shown in Figure 1, the high pin count allows the 7100LC to have three separate interfaces. Like TI's microSPARC, the 7100LC has a high degree of system integration with DRAM and cache control on-chip.

The programmable DRAM interface accommodates various processor clock speeds, DRAM speeds, and memory configurations. With small DRAM arrays, no external address buffers are required.

The synchronous system bus is intended to be used solely for I/O and graphics and, consequently, is a multiplexed data/address interface. The system bus can be programmed to run at either one-half or one-third of the processor clock speed. The chip also supports JTAG boundary scanning and a mode that allows performance tuning and system diagnostic information to be gathered with external equipment.

The cache interface uses separate address/control and data buses for tag and data SRAMs. The cache is composed of standard asynchronous, TTL-I/O SRAMs.

Floating-point latencies are good, especially when compared to some other low-cost RISCs, but not quite as good as the current 7100 design. For add/subtract, latency is two cycles regardless of precision, and for multiply and multiply-add, single-precision operations have a two-cycle latency while double-precision latency is three cycles. Divide and squareroot have the same latencies: eight cycles for single precision and 15 cycles for double precision. These latencies make the 7100LC significantly faster than the PowerPC 601 and dramatically faster than the TI microSPARC.

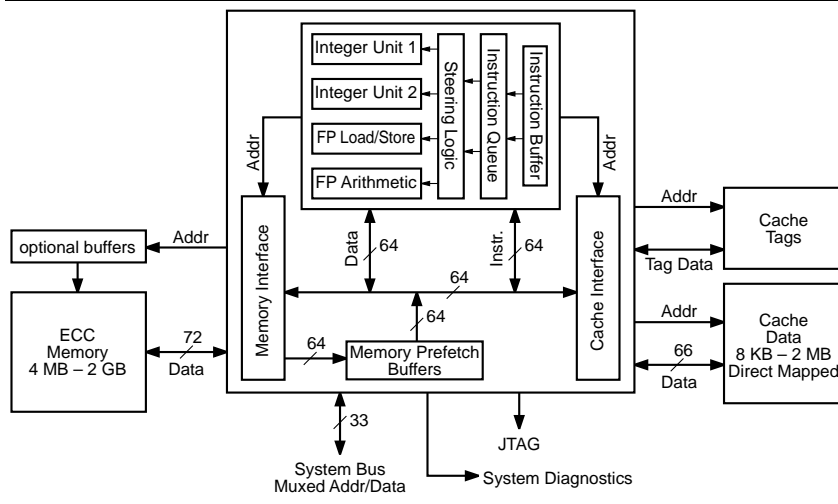


Figure 1. Block diagram of a 7100LC-based CPU.

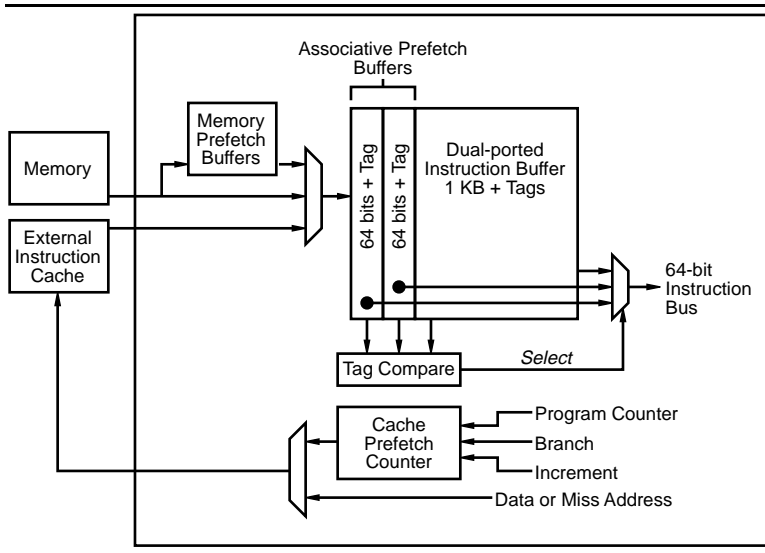


Figure 2. HP 7100LC cache structure.

The 7100LC also implements two PA-RISC architecture extensions. Support for non-cached memory pages has been added. More significantly, the 7100LC is the first PA-RISC implementation to support little-endian byte ordering in addition to big-endian. It is difficult to know for sure, but HP was probably compelled to add little-endian support to be able to run Windows NT more easily. (HP has not publicly committed to supporting Windows NT.)

Caches

The lack of on-chip cache has been the main distinguishing characteristic of PA-RISC implementations: while every other major microprocessor family has migrated to the use of on-chip cache for the first level, HP's designers have been staunch in their preference for off-chip SRAMs for both cache data and tags.

For the 7100LC, some compromises were made to keep costs down. First, while other PA-RISC implementations use separate external instruction and data caches, the 7100LC implements only a single, combined cache: a single set of SRAMs cache both data and instructions, but one half stores only data while the other half stores only instructions. This is not the same as a unified cache, which can store instructions and data anywhere in the cache.

The main problem with a combined cache is that a simple implementation will cause execution pipelines to stall on every load or store instruction. The 7100LC addresses the problem of contention for a single cache port by implementing a small, specialized instruction cache on chip. The on-chip I-cache together with a single, large off-chip cache is one way that the 7100LC uses integration to reduce total system cost.

The external cache is implemented with standard,

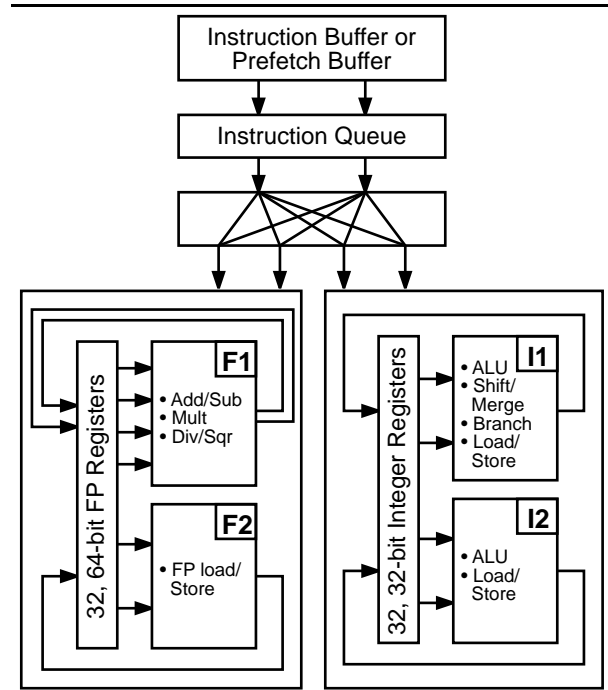


Figure 3. HP 7100LC instruction flow and execution units.

TTL I/O SRAMs and can range in size from 8K to 2M. The cache is direct mapped with a line size of 32 bytes. As shown in Figure 1, the cache access path is a separate, 64-bit bus with two extra bits for word parity. Whatever the cache size, exactly one half is used for instructions and the other half for data.

The number of SRAM chips needed for the cache depends on the desired size and the SRAM organization ($\times 8$, $\times 16$, or $\times 32$). For example, a 64K cache can be implemented with four $2K \times 8$ tag SRAMs and eight $8K \times 8$ data SRAMs.

The on-chip instruction cache consists of two parts: a 1K direct-mapped buffer and two associative prefetch buffers, as shown in Figure 2. The processor can fetch instructions from either the main buffer or the prefetch buffers. The prefetch buffers hold instructions that are likely to be executed based on the prefetching algorithm. Prefetched instructions progress from the first prefetch buffer, to the second, and finally to the main buffer. An instruction that is prefetched will eventually be stored in the main buffer even if it is not executed.

The miss rate of the relatively small buffer is kept low by an aggressive prefetch policy: whenever the external cache is not occupied with a load or store, the prefetch algorithm soaks up free cycles by bringing another instruction pair into one of the buffers. The prefetch algorithm uses static branch prediction to begin speculatively prefetching to minimize the impact of a taken branch.

When instructions are fetched from the on-chip instruction buffer, no penalty cycles are incurred. When

an instruction must be fetched directly (i.e., it has not been prefetched) from the off-chip cache, one or two cycles of penalty are incurred.

For data or instructions that miss in the off-chip cache, a reference to memory is required. Cache misses are satisfied using the critical-word-first strategy to make sure the processor continues with useful work as soon as possible. A miss in the cache that hits in the memory prefetch buffer typically incurs a four-cycle latency, and a miss that must go all the way to DRAM incurs a seven-cycle latency. Different DRAM and cache configurations can affect these latencies.

The cache is non-blocking, which means that if a miss occurs for data that is not needed immediately, the processor can continue executing instructions—including memory references—until a true dependency is encountered.

Superscalar Capabilities

So far, all of HP's PA-RISC implementations have relied on brute force—high clock frequencies and large first-level caches—to achieve their state-of-the-art performance. The current 7100 processor is the first and only superscalar PA-RISC implementation, and it has only modest superscalar capabilities: it can issue two instructions per cycle only if one is an integer operation and one is a floating-point operation.

As shown in Figure 3, the 7100LC goes farther by providing two integer units. The two integer units are not completely symmetrical since only I1 can issue branch and shift instructions, but they do allow two integer ALU instructions to be issued together. Like most superscalar processors with two integer units but unlike SuperSPARC, the two integer ALU instructions must be independent. The general rules for issuing two instructions in a cycle are shown in Figure 4.

The most impressive aspect of the 7100LC superscalar organization is its ability to issue two integer memory references simultaneously. To be able to dual-issue memory references, they must be of the same type—both loads or stores—and the addresses must be consecutive and aligned.

While these are significant restrictions, there are still many important graphics and sound algorithms that can take full advantage of this capability. For such algorithms, a system with a 7100LC and a fairly large

-
- (1) F1 + F2
 - (2) F1 + I1
 - (3) F2 + I1 (except no integer load/store)
 - (4) I1 + I2 (two loads/stores must be consecutive and aligned)
-

Figure 4. Rules for 7100LC dual-instruction issue.

cache should offer a considerable performance advantage over other high-performance processors. It should also be possible to take advantage of this capability in many traditional applications.

The P5 also allows two simple integer instructions with memory references to be issued simultaneously, but its issue rules are much more general. It will, however, execute them simultaneously only if they reference different cache blocks. The 7100LC can issue two memory references together only if they are guaranteed to be simultaneously executable. The P5 issue and execution rules allow it to achieve superscalar performance in a greater number of cases than does the 7100LC, but both take advantage of the most important cases.

Multimedia Features

HP has decided not to divulge many details about the multimedia hardware on the 7100LC, but Forsyth's presentation provided some information. The hardware supports pixel and digital audio data with addition, subtraction, averaging, saturation, and acceleration of multiplication by a constant. For pixel data, the hardware operates at a peak of four pixel operations (eight pixels in, two pixels out) per clock cycle.

The claimed applications are image processing, digital video (motion compensation and discrete cosine transforms), and digital audio (sample rate conversion, digital filtering, and mixing). This description makes it sound like the multimedia hardware is aimed at MPEG

full-motion video/audio applications. If so, the 7100LC could eliminate the need for a separate MPEG chip in a multimedia workstation.

Conclusions

With the introduction of 7100LC-based systems, HP will set a new performance standard for low-end workstations if it can meet its under-\$5000 goal. With a reasonably large cache and 75 MHz clock frequency, the integer performance of a 7100LC system will be lower than the current high-end PA-RISC systems, but not by



PHOTO BY ANNE HEMERSKY

The sizes of on-chip caches practical in today's technologies are still relatively small, and hence they can exhibit very high miss rates on some applications....On-chip caches also stress CPU die size and density limits. Even though multimillion-transistor CPUs are very impressive from a technical point of view, they can still be difficult to produce inexpensively and in high volumes.

Mark Forsyth, Hewlett-Packard

much. When HP introduces these systems, they will make the new, low-end, microSPARC-based workstations from Sun look truly anemic.

Instead of scaling back design objectives for its low-end processors, HP has instead improved processor organization to make up for the system-level compromises needed to reduce overall system cost. The result is its most aggressive superscalar design yet and performance that should—given a 75-MHz 7100LC with a large cache—challenge all but the fastest of its competitors.

7100LC-based workstations will be HP's low-end sometime in 1993, possibly in the second quarter. Even though the 7100LC is aimed at commercial rather than technical markets, it is still capable of equalling the current midrange PA-RISC machines. With this in mind, it is worth wondering what HP has planned for its midrange and high-end machines.

The multimedia hardware is a wild card. It is inevitable that full-motion video at CD-ROM data rates will become an important capability for desktop personal computers. When this capability becomes important depends on its cost. Many companies are working on dedicated MPEG encoder/decoder chips or chip sets with the hope of cashing in on the next wave of popular personal and business applications.

If HP has included sufficient hardware to enable MPEG or MPEG-like compression/decompression on the CPU chip using software, it will have a full-motion video product long before its competitors. By beating the competition to the punch, the importance of PA-RISC in the market could increase dramatically, and, together with the rumored port of Windows NT, HP could be right in the thick of the battle for dominance in the next wave of personal computers. ♦