MICROPROCESSOR REPORT
THE INSIDERS' GUIDE TO MICROPROCESSOR HARDWARE
VOLUME 8 NUMBER 12 SEPTEMBER 12, 1994

Digital Leads the Pack with 21164

First of Next-Generation RISCs Extends Alpha's Performance Lead

by Linley Gwennap

Three years ago, unbeknownst to its participants, a race began among five major CPU vendors to bring to market the next generation of RISC technology. Each realized that pushing the performance envelope beyond 200 SPECint92 would require aggressive superscalar dispatch and high clock rates. While Hewlett-Packard, IBM, MIPS Technologies, and Sun struggled with more complicated designs, Digital has emerged from the pack to take the checkered flag with its new 21164 design, known internally as EV-5.

The company didn't skimp on performance to hasten the chip's debut; not only is the 21164 the first microprocessor to exceed 200 SPECint92, but it should reach a rating of 330 when running at 300 MHz, according to Digital's estimates, with an astounding SPECfp92 score of 500. These scores more than double those of any non-Alpha microprocessor shipping today, and they should keep Alpha in the performance lead even when other vendors deploy their own next-generation chips.

The design can issue four instructions per cycle into two integer units and two floating-point units, for a peak execution rate of 1.2 BIPS (billion instructions per second). It is the first microprocessor to include a secondary cache as well as primary caches on chip: the unified level-two (L2) cache is 96K in size, pushing the transistor count to 9.3 million, another first. (Previously, the most transistors on a general-purpose microprocessor were the 3.6 million on the PowerPC 604.) Digital already has working samples, which achieve the performance noted above, and plans to ship the 21164 in 1Q95.

Although the Alpha chip can really fly, it has a few drawbacks. The processor's vast die size (298 mm^2), transistor count, and advanced 0.5-micron IC process push its estimated manufacturing cost to a towering $430, and Digital is quoting a shocking initial price of $2,937 for the 300-MHz part. The 21164 breaks another less desirable record by dissipating nearly 50 W at its peak operating frequency.

Short Cycle Time Requires Two-Level Cache

With its 1,200-MIPS peak throughput, the 21164 requires tremendous instruction and data bandwidth to feed its ravenous engine, far more than could be supplied from external cache RAMs. The design requires a large on-chip cache to buffer the high-bandwidth CPU from the lower-bandwidth external world. With the 0.5-micron process, Digital knew it could push the on-chip cache well beyond the 32K used by most current RISC processors to 64K or even 128K.

But even with Digital's CMOS-5 process *(see **080504.PDF**)*, the design team could not create a large cache array that could return data in a single 3.3-ns clock cycle. In a large array, it simply takes too long for the address to propagate through the array and for the data to propagate back. The best the designers could do was a 16K cache similar to the ones used in the 275-MHz 21064A, which is also built in CMOS-5. This size didn't work for the 21164, because the data cache had to be dual ported, doubling the die area of the array. Thus, the new processor includes two primary caches—one for instructions, one for data—of 8K each that can be accessed in a single 3.3-ns cycle.

But the design needed more fast memory on chip than just 16K, leading to the two-level cache scheme. The second-level cache array is 96K in size and requires two cycles (6.7 ns) to access due to its larger physical size. Including cycles for tag access and level-one refill, the total cache-miss penalty for the primary caches is six cycles (20 ns) on an L2 hit. An external L2 cache, in contrast, requires at least 25 ns to service an L1 miss in the 21064. Thus, moving the L2 cache on chip reduces the cache-miss penalty, improving performance.

Putting the second-level cache on the chip has additional benefits. The 21164 uses a three-way set-associative L2 cache, which increases the hit rate compared with the direct-mapped L2 caches used by most processors. It is difficult to implement set-associative caches externally due to the high pin-count required, but this

difficulty is not an issue for on-chip caches.

The two-level organization allows a more efficient use of resources. The large unified cache offers a higher hit rate than split caches of the same total size. Because the L1 data cache must be dual ported to service two accesses per cycle, the two-level design also avoids the need for a large dual-ported memory, which would have required much more die area; instead, only the small primary data cache must be dual-ported.

Finally, incorporating a large cache on chip reduces the need for external cache. Once the price of the 21164 comes down, it will be feasible to include it in a midrange system with no external cache, reducing system cost. Digital believes that the performance reduction in this configuration should be less than 10% for many applications.

Doubling Instruction Bandwidth

Figure 1 shows a block diagram of the 21164. Instructions are read in groups of four from the instruction cache and are placed into one of two four-word buffers. The dispatcher then issues as many instructions as possible from the current buffer; it must, however, completely empty one buffer before moving on to the next. For example, if three instructions are issued on one cycle, the fourth must be issued by itself on the next cycle.

To avoid this situation, the architects defined a "universal NOP" instruction that the dispatcher will recognize and discard. The compiler uses this NOP to pad odd groups of instructions, avoiding single-issued instructions. Single issue will be more frequent, however, on code that has not been optimized for the 21164.

The new chip doubles the issue rate of the 21064, which could issue two instructions per cycle among three function units: an integer unit, a load/store unit, and a floating-point unit. The 21164 has two fairly complete integer units, increasing the probability that the CPU can issue two instructions per cycle on integer code. The 21164 achieves about 1.1 SPECint92 per MHz, 50% better than the previous design (but significantly less than other very aggressive superscalar designs).

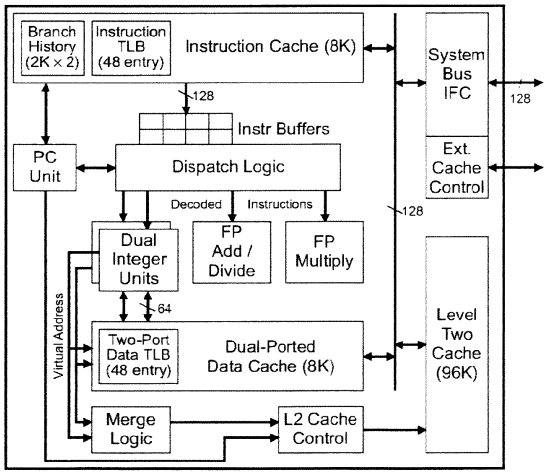

Figure 1. The 21164 can issue four instructions per cycle to two integer units and two floating-point units. The new processor is unusual in that it has a large secondary cache on chip.

The two integer units are not quite identical. Both can handle basic integer arithmetic and memory operations, but only one unit has a shifter and an integer multiplier; the other integer unit is the only one that can handle branches. Thus, the 21164 can pair most integer operations but cannot issue two shifts, two multiplies, or two branches in the same cycle.

Like most superscalar processors, the Alpha chip cannot issue instructions with data dependencies on the same cycle. The only exception is that a compare instruction can be paired with a conditional branch or conditional move instruction that uses the result of the compare. This case is handled by delaying the execution of the conditional instruction, as described later.

The new chip splits the floating-point unit into separate add and multiply units and can issue instructions to both of these units on the same cycle. This capability helps make up for the lack of a multiply-and-add instruction, as in PA-RISC and PowerPC. The FP adder also handles miscellaneous FP operations, such as FP branch and divide. Note that the 21164 can achieve its peak issue rate of four instructions per cycle only on floating-point applications; for integer code, the chip is limited to two instructions per cycle.

The dual-ported data cache supports up to two memory references per cycle. The data TLB is also dual ported and contains 64 entries. The instruction TLB has a single port and 48 entries. Both TLBs are fully associative. References that miss either the primary data or instruction cache are directed to the on-chip L2 cache. This cache is pipelined and can sustain a data rate of 16 bytes (128 bits) per cycle, or 4.8 Gbytes/s.

Cache Misses Don't Stall Processor

The 21164 always issues instructions in order and does not include the extensive out-of-order execution logic that will be used in other next-generation CPUs. The chip includes limited out-of-order execution, however, allowing instructions to continue executing while cache misses are being serviced.

The "merge logic" in Figure 1 includes a six-entry miss address file (MAF) that captures information on loads that miss the data cache. Once the integer unit generates the address for a load, it is finished with that instruction and can continue to accept new operations. If the load hits in the data cache, the desired data is loaded into the register file.

If a load misses the data cache, the load address

and target register are recorded in the MAF, which acts as a FIFO buffer. If the second-level cache is available, the load is processed immediately; if not, it is queued.

When a load request is entered in the MAF, the CPU checks to see if another load is pending for the same cache line. This is a fairly common occurrence, since many programs generate sequential load addresses. If indeed there are two pending loads to the same cache line, the two can share an entry in the MAF and are processed at the same time, reducing the number of accesses to the second-level cache. By merging loads, the six-entry MAF can keep as many as 21 loads pending at once. (Each entry can handle four loads except the last entry, which can hold only one.)

Once the cache miss is processed, the L2 cache returns the requested data to the L1 cache. The MAF entry also allows the data to be written directly to the register file and to an ALU if any pending instruction is waiting for the load data. If loads have been merged, the 128 bits of cache data may contain two load results, both of which can be written to the register file in one cycle.

The merge logic contains a six-entry write buffer for stores that miss the data cache. This buffer merges accesses to the same cache line, improving performance for programs that generate writes to sequential addresses.

The dispatch unit contains scoreboarding logic that prevents instructions from being dispatched until their operands are available. Because the 21164 always issues instructions in order, any instruction that is waiting for data (from a cache miss, for example) will prevent all subsequent instructions from being issued. The relatively simple approach of the merge queues allows the CPU to continue execution during the processing of the most common long-latency operation: data cache misses.

The compiler attempts to insert as many instructions as possible between an instruction that loads data and instructions that use that data. Even on integer code, however, as many as 12 instructions would be required to fill both issue slots during the six cycles that it takes to handle a data cache miss (assuming that the miss is handled in the L2 cache). Thus, the merge queues help the CPU perform useful work during a cache miss, but stalls will still occur.

Deep Pipeline Similar to 21064

The 21164 uses a deep, seven-stage pipeline similar to that in the 21064, as Figure 2 shows. Both pipelines have four stages to fetch and issue instructions, followed by two execution stages and a writeback stage. The new design includes several improvements that reduce pipeline stalls.

The instruction cache is accessed in stage 0. The branch history bits are read from the cache at the same time as the instructions. Instructions are partially decoded in stage 1, and the target address of the first

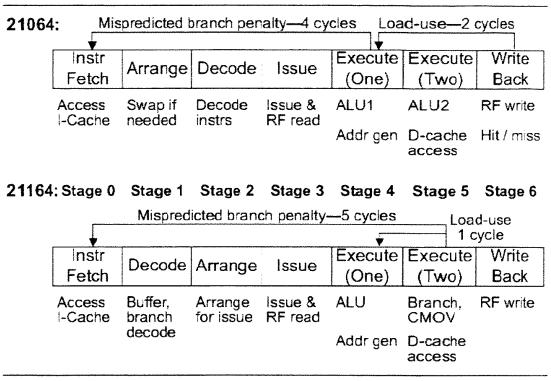

Figure 2. The 21164 pipeline has the same length as the pipeline of its predecessor, the 21064, but the execute stages are improved.

branch instruction (if any) is calculated; if the history bits indicate that the branch is predicted taken, the fetch stream is redirected to the new address.

Because the target address is not sent to the instruction cache until the end of stage 1, a predicted-taken branch creates a "bubble" of four instructions, fetched while the branch was in stage 1, that must be discarded. If the CPU is operating at its peak rate of four instructions per cycle, this bubble will propagate through the pipeline, creating a one-cycle penalty for taken branches. But if the dispatch logic takes two cycles to issue four instructions, as it does on all integer applications and some floating-point code, this bubble is quashed before it reaches the function units. The designers considered a branch-address cache, as used in the PowerPC 604 *(see **080501.PDF**)*, but it would have increased complexity without significantly improving performance, since the one-cycle penalty is rarely seen.

The 21164 uses an aggressive prefetch strategy that avoids instruction-cache misses for sequential code. If a miss does occur, the dispatch logic will consume the contents of both instruction buffers while the miss is being serviced, then wait for instructions from the L2 cache, which returns 128 bits at a time. These instructions are loaded directly into the instruction buffers at the same time that the instruction cache is updated. On integer code, the second-level cache can deliver instructions at the full issue rate while still retaining half of its bandwidth for data accesses.

During stage 2, the dispatch logic arranges the instructions so they can be sent to the appropriate function units and checks for resource conflicts (e.g., too many FP adds) and register dependencies. If any conflicts or dependencies are found, the dispatcher will issue fewer than four instructions.

During stage 3, the chip sends instructions to the function units and reads their operands from the register file. The MAF is checked to make sure the current in-

	21164	21064
Most Integer Ops	1 cycle	1 cycle
Shift and Byte Ops	1 cycle	2 cycles
CMP with Branch	0 cycles	1 cycle
L1 Cache Hit	2 cycles	3 cycles
Mispredicted Branch	5 cycles	4 cycles
Integer Multiply	8–16 cycles	19–23 cycles
Most FP Ops	4 cycles	6 cycles
FP Divide (SP/DP)	10/23 cycles	31/61 cycles

Table 1. The new Alpha processor significantly improves instruction latencies compared with the 21064.

structions do not require data from a pending load. If an instruction cannot be issued because its operands are not available, dispatching halts until the operands are ready.

Execution Times Improved

Stages 4 and 5 are for instruction execution. Nearly all integer operations complete during stage 4; only conditional moves (CMOV) and branches execute during stage 5. In the 21064, in contrast, all shifts and all byte operations complete during the second execute stage.

In fact, branch and CMOV instructions execute in a single cycle and could be completed in stage 4. By delaying these instructions until the second execute stage, the CPU allows them to be paired with a compare (CMP) instruction even if they depend on the result of the compare. For example, in the sequence:

```
CMP R1, R2, R3;   BEQ R3, <target>
```

the compare executes in stage 4 and forwards its result to the conditional branch, which executes in stage 5. Without this accommodation, the 21164 would be forced to execute the instructions in this sequence sequentially, reducing the issue rate.

Pushing the branch evaluation into stage 5 extends the branch penalty to five cycles rather than four. This penalty is taken only when the branch is mispredicted, which is typically less than 15% of the time due to the large (2K-entry) branch history table. Even if branch execution were pushed into stage 4, dependent branches (like the one shown above) would still execute one cycle after the compare instruction because of the dependency; thus, delaying the branch has little downside and improves the issue rate.

The 21164 reduces the latencies of other operations. Table 1 compares the new chip to the 21064, showing improvements in integer multiply and most floating-point operations. The new FPU uses a nine-stage pipeline to perform most FP operations in four cycles, with an issue rate of one multiply and one add (or similar) operation per cycle. Only FP divide is not pipelined; the divider can calculate approximately 2.3 bits per cycle.

	Instruction Cache	Data Cache	Secondary (L2) Cache	External (L3) Cache
Cache Size	8K	8K	96K	1M-64M
Associativity	direct	direct	three way	direct
Line Size (bytes)	32 bytes	32 bytes	32 or 64	32 or 64
Number of Ports	one	two	one	one
Write Policy	n/a	write-thru	write-back	write-back
Allocation Policy	on read	on read	on write	on write

Table 2. The 21164 supports three levels of caches, two on the chip and a level-three external cache.

Second-Level Cache is Pipelined

For a load or store instruction, the virtual address is calculated during stage 4 and sent to the cache tags, cache array, and TLB in parallel. By the end of stage 5, these units return the requested data, which can immediately be sent to the ALUs. Thus, there is a single-cycle load-use penalty, one cycle better than the 21064.

If the access misses the data cache, it is sent to the merge queues. Accesses to the second-level cache are fully pipelined, extending the pipeline to 12 stages for an L2 cache hit. Assuming that the queue is empty, the chip performs a cache tag lookup during stage 6. The results are then checked in the next cycle to determine if there is a hit and, if so, which bank contains the data. Accessing the cache array takes two more cycles; this access is wave-pipelined, so the cache can deliver one result per cycle. In stage 10, the data is driven across the chip back to the CPU and is ready to be used by stage 11, when it is also written into the data cache.

The designers considered placing the tag lookup in parallel with the access to the cache array, as is done in the primary caches. By looking up the tag first, however, the chip can power sense amps only for the bank that contains the requested data, reducing cache power by roughly two-thirds. Because each 32K bank contains about 1.5 million transistors and dissipates up to 10 W, this strategy produces a major power reduction at the cost of extending the L2 latency by two cycles.

High-Performance System Interface

Although the 21164 has a 128-bit system bus similar to the 21064's bus, the protocols are different enough to require a new set of interface chips (see sidebar). One major improvement is that the new processor includes complete control logic for an external cache; the 21064 requires external PALs to control its external cache.

The 21164 supports an external level-three cache ranging in size from 1M to 64M. Smaller caches would be nearly useless when combined with the 96K secondary cache on the chip. This cache can be built using standard asynchronous SRAMs or synchronous (pipelined) parts. With 8-ns asynchronous SRAMs, the external cache can return data every three CPU cycles at 300 MHz (1.2 Gbytes/s). Table 2 shows the design parameters for this and the other caches in the 21164.

The external cache is nonblocking, another improvement over the 21064. The 21164 system bus uses a split-transaction protocol, and the processor can have up

to two L2 cache misses pending at any given time. The L3 cache can also return data while a miss is being processed (hit under miss).

The CPU generates a clock for the system bus that can have a period of 3 to 15 times that of the CPU clock (i.e., 20 MHz to 100 MHz). Like its predecessor, the 21164 requires an input clock signal at twice the internal CPU frequency, or up to 600 MHz. This signal can be created by an oscillator adjacent to the CPU chip. For multiprocessor systems, a low-frequency reference clock synchronizes the processors.

The processor draws a maximum of 50 W at its top speed of 300 MHz, despite using a 3.3-V supply. This rating is far higher than even the 21064, one of the hottest chips around. Digital offers the 21164 with two heat sink options: a 1.75" tall version that requires 500 linear feet per minute (lfm) of airflow and a 1.25" heat sink for systems that can deliver 800 lfm. A typical PC case fan, for comparison, produces an airflow of 50–100 lfm, although some Pentium systems deliver up to 300 lfm. Digital says that its new chip can easily be cooled using the taller heat sink and a "common" package-mounted fan.

The 21164 is sold in a 499-pin ceramic PGA package that contains a heat slug to facilitate cooling the chip. Figure 3 shows a photograph of the die, which is 298 mm^2 in size. It is built in a 0.5-micron CMOS process with four metal layers. Although this die is larger than that of any other general-purpose microprocessor, its size is partially mitigated by the large cache, which is protected by redundant circuitry, that leaves only 60% of the die susceptible to defects. Even so, the MPR Cost Model *(see **071004.PDF**)* estimates the 21164's manufacturing cost to be about $430, again greater than any announced single-chip microprocessor, although it is 30% less than the two-chip R8000. The 21064A, by comparison, costs less than half as much to build as the 21164.

Running Rings Around the Competition

The performance of the new processor is staggering; at 330 SPECint92, it is 75% faster than the 21064A, currently the fastest microprocessor shipping. The new Alpha chip is at least 2.5 times faster than the best shipping processors in the SPARC, MIPS, PowerPC, and Pentium families. A performance gap of this size is nearly unprecedented in the microprocessor industry.

This enormous advantage will quickly shrink. Digital expects to begin volume shipments of the 300-MHz CPU in 1Q95. The PowerPC 604 and HP's PA-7200 will trim the Alpha advantage to a still-daunting 2× by the time the 21164 begins to ship. By midyear, other next-generation RISC processors—Sun's UltraSparc, the MIPS T5, and the PowerPC 620—could be shipping with performance in the 250–300 SPECint92 range. Against this strong competition, the 21164 should still keep Alpha in the overall performance lead for at least another year, even without possible clock-speed upgrades to the Digital processor.

Chip Set Available for 21164

Digital has upgraded the 21071 chip set *(see **070904.PDF**)* to work with the 21164. The new chip set, dubbed the 21171, consists of five chips but only two chip types. The control chip generates the control signals for DRAM and PCI and directs the flow of data to the system bus. Four small data-path chips buffer and route data between the 128-bit system bus, 256-bit interleaved DRAM subsystem, and 32-bit PCI bus. Because the 21164 generates all control signals for the external cache, the 21171 does not handle this function.

Like the 21164, the chip set operates at 3.3 V but can connect to DRAM and PCI devices at either 3.3 V or 5 V. It can support either a 32- or 64-bit PCI bus, although no peripherals currently support 64-bit PCI. The chip set allows either standard or synchronous DRAM for main memory; with the latter chips, the need for an external cache is reduced.

The company expects to release additional chip sets in late 1995, one that has a lower cost than the 21171 and another for multiprocessor configurations.

Digital also offers an evaluation board, the EB164, that contains a 21164 CPU, the 21171 chip set, 1M–8M of cache, eight SIMM sockets for DRAM, and three PCI slots. The company will also license the design of a similar board suitable for very high end PCs.

The new chip achieves its performance by continuing to emphasize clock speed over complexity. Digital clocks its processors two to three times faster than competitive high-performance products while accepting a lower number of instructions per cycle (IPC). Although other processors deliver much better IPC, none comes close to reaching a high enough IPC to counteract Digital's advantage in clock speed. The Speed Demons continue to outpace the Brainiacs *(see **0703ED.PDF**)*.

Despite the lengthy pipeline, Digital eschews the term "superpipelined" for its newest processor. If this hoary term has any technical meaning, it indicates a pipeline that splits instruction execution and/or cache accesses across multiple stages. The 21064 did, in fact, take two cycles to execute certain instructions as well as two cycles for data cache accesses. The newer chip, however, performs these functions in a single cycle. The 21164 pipeline is much more akin to the classic five-stage RISC pipeline, except that in this case the complexities of superscalar dispatch require three decode stages rather than one.

One peril of high clock rates is the problem of implementing single-cycle caches. It seems scary that signals cannot propagate through an on-chip memory array in a single cycle, even if that clock cycle lasts only three nanoseconds. As noted previously, the two-level cache

Price and Availability

Digital expects to sample the 21164 in October, with volume shipments of the 266-MHz speed grade in January (1995), followed by 300-MHz shipments in March. In quantities of 1,000, the 266-MHz part will cost $2,052 and the 300-MHz version will sell for $2,937.

The 21171 chip set is expected to sample in October and ship in January; it will sell for $295 in 5,000-unit quantities. The EB164 evaluation board, with 1M of cache and 16M of DRAM, will be available in December for $7,500 each. Contact Digital at 800.332.2717 or 508.568.6868; fax 508.568.6447.

design used in the 21164 makes sense for a number of reasons other than the short clock cycle, and this technique will become more common in future high-performance microprocessors.

Performance at Any Cost?

Another problem with fast clocks is their impact on system design. Digital eases most of these problems by allowing the system bus to clock at a reasonable rate, and the on-chip secondary cache avoids the need for an ultrafast external cache. System designers must cope with the very high power dissipation of the 21164 chip.

Digital has announced pricing of $2,937 for the 300-MHz part in quantities of 1,000 units, with a 266-MHz speed grade available for $2,052. The company is taking advantage of its new performance point by charging an extreme price. Other RISC vendors price their processors at around $4 per SPECint92, but the 21164 comes in at $7 to $9 per SPECint92.

At this inflated price, the 21164 is attractive only for scientific systems (Cray is considering the chip for the successor to its T3D), high-end servers, and other niche products. Digital will probably deploy the chip in a workstation that sells for even more than its $42,000 21064A machine *(see **0811MSB.PDF**)*. Although the company is maximizing its per-chip profits, it is doubtful that Digital will ship enough parts at these prices to make a significant amount of revenue.

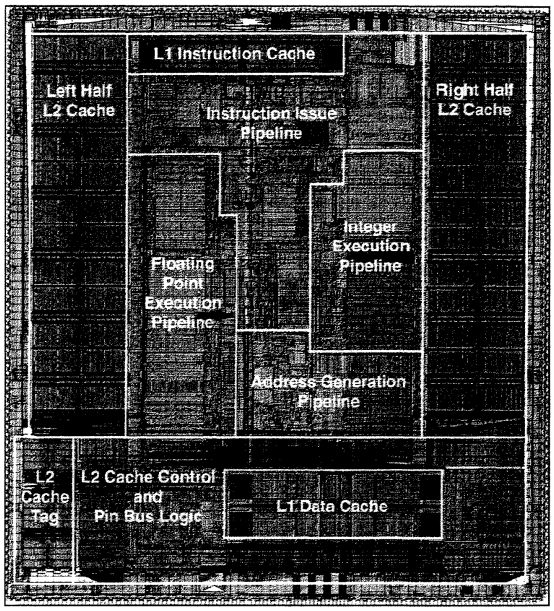

Figure 3. The 21164 measures 16.9 x 18.6 mm and includes 9.3 million transistors in a 0.5-micron, four-layer-metal CMOS process.

The chip price will have to come down significantly before it will make an impact in the Windows NT market, supposedly a key Alpha target. Digital cut the price of the 21064 by a factor of three over a two-year period and, after a process shrink, the 21164 could end up under $1,000. We suspect that competitive pressures will force Digital to cut its prices even faster for the 21164; after all, the bigger they come, the harder they fall.

The 21164 sets another performance record for Alpha and demonstrates Digital's commitment to delivering the fastest microprocessors in the world. With its stunning performance, however, comes a set of drawbacks: the chip is big, hot, and expensive. Of these, the most critical is the high price, and this is one problem that Digital can solve over time. Despite its problems, the 21164 is an impressive technical achievement. ♦