

Intel Reveals Next-Generation 960 H-Series

Larger Caches, Higher Clock Rates Distinguish New Chips

by Brian Case

Bringing Pentium-class performance to the embedded world, Intel has announced the new high end of its 960 family of embedded control microprocessors. Like the original K-series processors, the three new H-series chips are built from a single silicon design, which is internally code-named the P110. This design retains the superscalar organization of the C-series parts but has larger caches and operates at up to 75 MHz. The H-series chips also have slightly improved on-chip peripheral functions and offer clock doubling and tripling *à la* Intel's 486DX2 and DX4 chips.

Intel expects to fabricate first silicon before January 1995 and to sample shortly thereafter. Volume production is planned for 2Q95. Prices are expected to range from about \$80 to \$160 in quantities of 10,000, depending on package and speed. This price range overlaps the price structure of the CA/CF parts, but the H-series does not obsolete them. Intel says it is more concerned about gaps in a product line than overlaps and currently plans to keep all five 960 product lines alive.

The new chips are too expensive to make a big impact on the mainstream embedded market, but they should generate significant interest among the makers of color and high-resolution laser printers and image setters. Some of the new features in the H-series parts are designed to address problems specific to internetworking applications, so the chips should generate interest from vendors of bridges and routers as well.

H-Series Builds on C-Series Core

The H-series chips combine the C-series core and J-series instruction-set extensions with larger caches. There are no major microarchitectural changes over the C-series design, but the data RAM (part of which can be used to cache the register stack) is doubled in size to 2K and the DMA controller has been eliminated, as Figure 1 shows.

Intel says that many peripheral chips, particularly networking devices, have evolved into bus masters that perform their own DMA. This has rendered the on-chip DMA controllers irrelevant.

Perhaps surprisingly, the superscalar core of the H-series is the same as that of the C-series (see MPR 9/89, p. 1). Because the superscalar capabilities of the C-series parts were underutilized in some applications because of a too-small instruction cache, it made sense for Intel to spend design time and silicon area on larger caches instead of more aggressive instruction-issue logic and execution units. The underutilization of the core is also probably what led Intel to remove the superscalar capabilities for the less expensive J-series.

The biggest improvements of the H-series chips over the C- and J-series chips are:

- Larger caches. The P110 die integrates a 16K instruction cache and an 8K data cache. The instruction cache is four times bigger than the largest cache previously offered in the 960 family, while the data cache is eight times larger.
- Higher clock rates. The HT will operate with an internal clock rate of up to 75 MHz. The highest clock rate offered before was 40 MHz for a superscalar core (CF) and 50 MHz for a scalar core (JD).
- One new instruction capability for cache invalidation (the H-series includes all the new instructions introduced in the J-series).
- Hardware for unaligned accesses (vs. microcode).

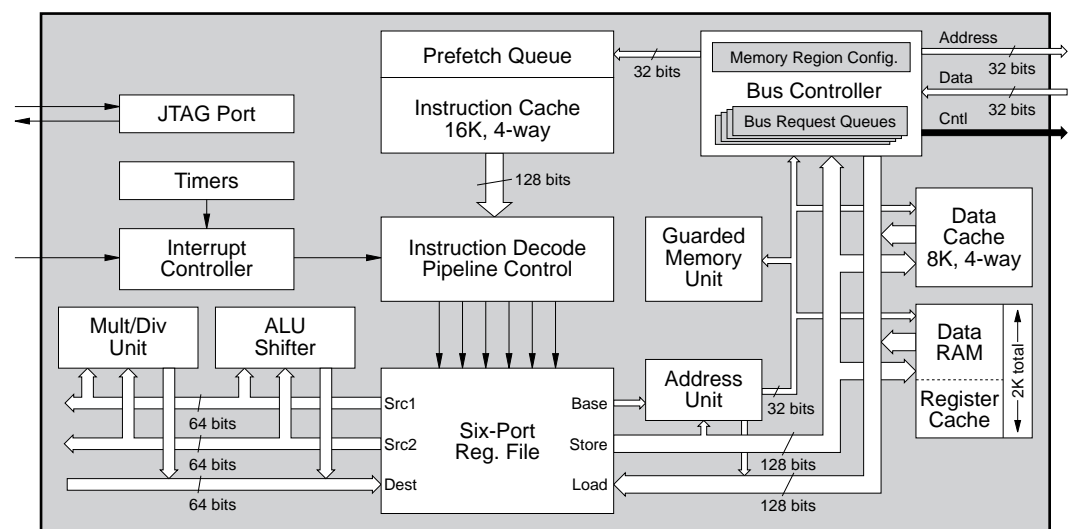


Figure 1. This block diagram shows that the organization of the H-series is identical to that of the C-series. The major improvements over the C-series are the larger caches and the enhanced bus controller.

	960KA/KB	960SA/SB	960CA/CF	960JA	960JF/L960JF	960JD	960HA/HD/HT			
Core Technology	First Generation	Cost-reduced K-series core	Superscalar	CA-derived scalar	CA-derived scalar	CA-derived scalar	Superscalar			
FPU	KB only	SB only	no	no	no	no	no			
Instruction Cache	512 byte data memory	512 byte data memory	1K, 2-way/CA 4K, 2-way/CF	2K, 2-way	4K, 2-way	4K, 2-way	16K, 4-way			
Data Cache	none	none	none/CA 1K, direct/CF	1K, direct	2K, direct	2K, direct	8K, 4-way			
Data RAM	none	none	1K	1K	1K	1K	2K			
Register Sets	4	4	5–15	8	8	8	5–15			
DMA	no	no	4 channels	no	no	no	no			
Buses	32/32 adr/data muxd	32/16 adr/data muxd	32 adr, 32 data	32/32 adr/data muxd	32/32 adr/data muxd	32/32 adr/data muxd	32 adr, 32 data			
Clock (MHz)	10/SA only 16, 20	10/SA only 16, 20	16, 25, 33 40/CF only	16, 25, 33	16, 25, 33, 40/JF only	33, 40, 50	25, 33, 50, 66, 75			
Clock Multiplier	no	no	no	no	no	2×	2×/HD, 3×/HT			
VAX MIPS*	12 (20 MHz)	7 (16 MHz)	40 (CA-33) 60 (CF-40)	28 (33 MHz)	33 (40 MHz) 28 (33 MHz)	41 (50 MHz)	65 (HA-40) 83 (HD-50) 125 (HT-75)			
Supply Voltage	5 V	5 V	5 V	3.3 V	5 V, 3.3 V/LJF	5 V	3.3 V			
Typ. Power	1.8 W (25 MHz)	1.1 W (20 MHz)	3.8 W (CA-25) 5.0 W (CF-33)	0.5 W (33 MHz)	1.2 W (JF-33) 0.5 W (LJF-33)	1.9 W (50 MHz)	4.0 W (66 MHz)			
Price (10K)	\$18.55 \$20.40 \$22.45 (KA)	\$20.40 \$22.45 \$24.70 (KB)	\$9.75 \$11.70 \$12.80 (SA)	\$14.45 \$15.90 (SB)	\$43.90 \$54.30 \$81.10 (CA)	\$52.70 \$87.90 \$113.10 \$135.70 (CF)	\$16.60 \$20.80 \$25.95	\$24.90 \$31.10 \$38.95 \$48.65	\$29.45 \$44.45 \$50.30 \$57.35	\$78.70/HA-25 \$90.80/HA-33 \$81.07/HD-33 \$152.00/HD-66 \$158.10/HT-75
Availability	Now	Now	Now	3/95 (6/95-33 MHz)	12/94 (JF-33) 3/95 (LJF-25)	6/95	1Q95/samples 2Q95/volume			

Table 1. The five series of 960 processors span a wide range of price and performance topped by the new H-series (rightmost column). *VAX MIPS based on Dhrystone 2.1 in a system with zero-wait-state SRAM.

- Improved on-chip memory control and debugging functions.

Like all current 960 chips, the H-series does not have an MMU. So far, the only 960 that implements an MMU is the 960MC, which uses the original K-series core. The MC is still available, but its performance—equivalent to other K-series chips—is meager compared to newer 960 designs.

Three Products from One Die

Three members of the H-series will be introduced:

- The HA with a 1× bus-clock-to-core-clock multiplier
- The HD with a 2× multiplier
- The HT with a 3× multiplier

Unlike the S-series and the original K-series—in which letter-distinguished family members had different capabilities—the H-series parts differ only in the bus-to-core clock multiplier and the corresponding ability to operate at a higher internal clock speed.

As shown in Table 1, Intel's 960 product catalog now consists of five families: the original K-series (see MPR 4/88, p. 1); the S-series (see MPR 10/17/90, p. 15) for the lowest cost designs; the new J-series (see *080803.PDF*),

which sports low cost and moderate performance; the previous high-end C-series (see MPR 9/89, p. 1), which now offers medium cost and superscalar execution for midrange performance; and the H-series, which will offer a high-performance upgrade path through higher clock rates and larger on-chip caches.

The SB and KB retain their niche as the only 960 chips that offer on-chip floating-point support. None of the H-series parts has floating-point hardware—contrary to earlier reports—but Intel may introduce such a chip if sufficient demand develops.

New Capability Aids Cache Consistency

The H-series uses the C-series core but also implements the new instructions added to the J-series. These instructions—conditional arithmetic and moves, and 8- and 16-bit compares—increase code density and eliminate branches. The byte-swap instruction simplifies data handling in networking applications that must deal with both little- and big-endian data packets.

The H-series adds one new instruction variation that accesses the “quick-invalidate” capability of the caches. The memory controller allows arbitrary regions

of memory to be assigned the quick-invalidate attribute (see "Bus Controller Improves Memory Handling" below). When the quick-invalidate cache-control instruction is executed, any valid cached locations that fall within the attributed memory regions are invalidated.

This capability is useful for dealing with off-chip DMA, which is often performed by peripheral chips that can become bus masters. When the processor knows that DMA is going to be performed into a memory region, it can issue a quick-invalidate for that region to guarantee that no stale data will remain in the on-chip caches. This provides the chief benefit of hardware cache consistency enforcement without the hardware expense.

Cache Sizes Quadrupled

Functionally, the caches on the H-series are similar to those of the C- and J-series. The instruction cache is 16K with a four-way set-associative organization. A pseudo-LRU replacement algorithm is used. The tag array stores a valid bit for every two instructions. Each 4K block associated with one way of associativity in the instruction cache can be selectively locked to keep important or time-critical code permanently in the cache.

Three predecode bits are stored along with each instruction word in the cache to aid parallel decoding and dispatch. As in the C-series, these bits are computed when the instruction words are written into the cache, which makes it easier for the decoding and dispatching logic to discover opportunities for parallel instruction dispatch. This simplifies the decoding and dispatching logic, which can save a pipeline stage and/or shorten a critical timing path to help keep clock rates high.

The data cache is 8K in size and also has a four-way set-associative organization. The data cache is write-through and uses a write-allocate policy. To conserve bus bandwidth, only the needed data word (or the word containing the data) is fetched on a load miss, and the cache has a valid bit per word to facilitate this policy.

Mitigating the Impact of the Slow Bus

One of the performance limitations of the 960CA is its small instruction cache (1K, two-way set-associative). For routines that fit in the cache, performance is good, but instruction cache misses—which happen frequently with such a small cache—reduce performance considerably because only one instruction can be delivered from memory per clock cycle. The CF quadruples the instruction cache size to reduce the miss rate but does not increase associativity.

To help reduce the system cost associated with a fast processor, the HD and HT run internally at double or triple the bus speed, so slow, cheap memory can still be used. As a result, unfortunately, the performance penalty of a cache miss is two or three times greater than for the C-series. The H-series' large, 16K four-way set-

Price & Availability

Samples of the 960HA and 960HD in all speed grades are promised for 1Q95. Samples of the 960HT-75 are promised for 2Q95. Production is promised for one quarter after sampling. See Table 1 for price information.

For more information, contact your local Intel sales office or call 800.628.8686. Information can also be obtained from Intel's FAXBack service at 800.628.2283; request document number 2068.

associative instruction cache helps reduce cache misses and decouple processor performance from slow memory for most applications.

The larger cache with increased associativity will benefit real-time applications as well. With four-way associativity, locking critical code on chip creates less of a penalty for code not locked in the cache. This is because locking one block gives a generous amount—4K—of space for time-critical code while leaving a still-effective 12K, three-way set-associative instruction cache for caching dynamic program activity. With the C-series, locking one block provides only a small area for time-critical code and can decrease general performance severely, because only a small, direct-mapped cache remains for nonlocked code.

Some minor microarchitectural changes have been made to the queue structure in the bus controller to further decouple internal performance from bus speed. In the C-series, all bus requests—loads, stores, and instruction fetches—are buffered in a single three-entry queue. For the H-series, separate load/store and instruction-request queues are implemented. The load/store queue can hold any combination of requests, up to a total of four 128-bit (quad-word) accesses. The separate instruction-fetch queue can buffer two requests.

The store queue acts as a write buffer, allowing the processor to dismiss stores and continue executing instructions even though the stores have not executed externally. The 128-bit queue entries also help exploit the concurrency possible with the internal 128-bit buses.

Bus Controller Improves Memory Handling

Unlike other embedded processors that adhere more strictly to RISC tenets, the 960 family has always had support for unaligned accesses. This support has probably been to the 960's advantage in markets like networking where packet alignment can be arbitrary.

In the C-series, an unaligned access invokes a microcode routine to handle the multiple accesses and data merging. This is faster than trapping the access as an exception, as some RISC processors do.

To achieve even better performance, the H-series processors implement unaligned-access support in hard-

ware. Hardware is used because the microcoded routine requires the full resources of the processor, which means that normal instruction processing stops during an unaligned access. Using microcode would waste a lot of performance in the clock-multiplied H-series chips because of the relatively slow bus. With hardware, the processor can continue to execute application code while hardware processes the unaligned access (always true for stores and for loads when independent instructions can be scheduled during the load delay).

The H-series memory controller retains the address-space partitioning implemented in the C-series chips: sixteen 256M memory partitions. The C-series memory controller gives each 256M partition individually programmable attributes, such as memory size and speed.

The H-series memory controller augments the partitioning and capabilities implemented in the C-series chips. The sixteen separate 256M partitions have the following programmable characteristics, called “physical attributes”:

- Memory width (16 or 32 bits)
- Memory speed (wait states, etc.)
- Burst capability
- Address pipelining capability
- Byte parity (odd, even, none)

All of these attributes are also implemented in the C-series, except for byte parity.

In addition, the H-series chips implement a separate facility for programmable “logical attributes,” which are byte order and caching policy (cachable/not-cachable and quick-invalidate). (In the C-series chips, byte order is a physical attribute.) Unlike physical attributes, the logical attributes are assigned to regions of memory defined by a start address and an address mask. Thus, a region of memory governed by a set of logical attributes can span part or all of a 256M physical region or even multiple regions. There are 15 sets of logical attributes assignable to programmable regions and one set of default logical attributes that governs all memory addresses not covered by a programmable logical-attribute region.

As explained above, the “quick-invalidate” logical

attribute is new. When a logical region is marked with this attribute, executing the data-cache-control instruction with quick-invalidate set invalidates any cached locations that fall within the attributed region(s).

Guarded Memory Unit Enhances Debugging

The new guarded memory unit eases the problems associated with developing and debugging code that executes largely from on-chip memory. This unit provides the capability to protect two memory regions from accesses and to detect accesses to six memory regions.

The extents of the two access-protected memory regions are defined by registers that contain the start address and an address mask, similar to the programming of the logical-attribute regions in the memory controller. Any writes to these two memory regions will be detected *and* prevented.

The extents of the six access-detection memory regions are defined by registers that contain a start address and a stop address. Accesses, including writes, are not guaranteed to be prevented to these memory regions, but any disallowed access will cause a trap.

The processor checks user/supervisor, read/write, and execute/data privileges for all eight guarded regions.

In addition to the guarded regions, the H-series chips also enhance the breakpoint facilities of past chips, increasing the number of instruction- and data-address breakpoints to six each.

One Board Design Can Accept Cx or Hx

The pin-out of the H-series chips allows a single socket to accept either a C-series part or an H-series part. H-series parts cannot, however, be plugged into an existing C-series board. The major differences between the chips are the power-supply voltage—5 V for the C-series, 3.3 V for the H-series—and the timing reference signal—CLKOUT for the C-series, CLKIN for the H-series. New board designs can accommodate these differences. Intel hopes to encourage customers now using C-series parts to begin designing and debugging boards that can easily be upgraded by dropping in an H-series chip.

In addition, all three H-series parts are completely pin compatible, making it a simple matter to upgrade the performance of a board that uses an H-series chip. A 25-MHz HA system could nearly double or triple CPU-bound performance with an HD-50 or HT-75.

Hx Chips Use Intel's Best IC Process

The P110 die will be fabricated in the same 0.6-micron four-layer-metal BiCMOS process that is used for the P54C Pentium and DX4 (*see 080504.PDF*). Although no real P110 chips are available to characterize, simulation results

	960Sx	960Kx	960CA	960CF	960Jx	960Hx
Transistors	346k	350k	600k	800k	750k	2,300k
Die size	51 mm ²	59 mm ²	137 mm ²	120 mm ²	64 mm ²	100 mm ²
IC Process	1.0μ, 2-metal CMOS	1.0μ, 2-metal CMOS	1.0μ, 2-metal CMOS	0.8μ, 3-metal CMOS	0.8μ, 3-metal CMOS	0.6μ, 4-metal BiCMOS
Package	80-pin PQFP, 84-pin PLCC	132-pin PGA or PQFP	168-pin PGA, 196-pin PQFP	168-pin PGA, 196-pin PQFP	132-pin PGA or PQFP	168-pin PGA, 208-pin SQFP
Est. Mfg. Cost	\$5	\$7	\$20	\$40	\$15	\$50

Table 2. With almost three times as many transistors as the CF, the H-series is by far the most aggressive implementation of the 960 yet.

and a knowledge of the process allow Intel to predict silicon behavior. Table 2 summarizes some of the chip characteristics.

Power dissipation is expected to be about 4 W with a 66-MHz internal clock. The chip requires 3.3-V power, but can use either 5-V or 3.3-V levels for I/O signals. A voltage reference pin determines which I/O levels are used.

One of the new features of the chip is a power-down mode invoked by the HALT instruction. This mode reduces power consumption by 90% and should help reduce cooling costs. Eliminating a fan from a product design is significant for some communications products.

Intel expects the P110 die to be about 400 mils on a side, or about 100 mm². Note that, due to the 0.6-micron process, the H-series die is actually smaller than either the CA or CF. Due to the extra 8K of cache and 2K of RAM, however, the die size is 30% larger—about 23 mm²—than the DX4, which is built in the same process. The MPR Cost Model (*see 081203.PDF*) estimates the cost of the P110 to be roughly \$50, about 25% more than the DX4 at \$40.

Performance More Than Doubles

Based on Dhrystone 2.1, the HT-75 is expected to reach 125 VAX MIPS, outrunning all other significant embedded processors and even a 66-MHz Pentium. Figure 2 shows Intel's estimates for relative performance within the 960 family. With DRAM memory, chips from the CA-33 through the HT-75 span a performance range of more than 4.5-to-1. There seems to be little incentive to design a system with SRAM memory.

As expected, the H-series chips show a much less severe performance decline with DRAM memory than do the C-series parts. This testifies to the effectiveness of the large on-chip caches of the H-series chips. Note especially that the CA—with its tiny cache—loses a whopping 43% of its performance by using DRAM instead of SRAM.

Note also that, as a percentage, the HT's performance loss is more than twice that of the HA. With 80-ns DRAM, the HT-75 is only about 43% faster than the HA-40 despite its 88% faster clock. With 60-ns DRAM, the HT-75 performance is better, at about 56% faster than the HA-40. This indicates that the extra expense of the HT may not be justified in a system with really slow memory.

Intel Aims for One-Stop Shopping

With the H-series, the 960 family spans a wider range of processor performance than any other high-end embedded control family. It is also the most successful, with higher unit volumes than any other high-end chip. The 960 has a very visible design win in the HP LaserJet 4, but even without it, the 960 would still be the vol-

Composite Networking Benchmark

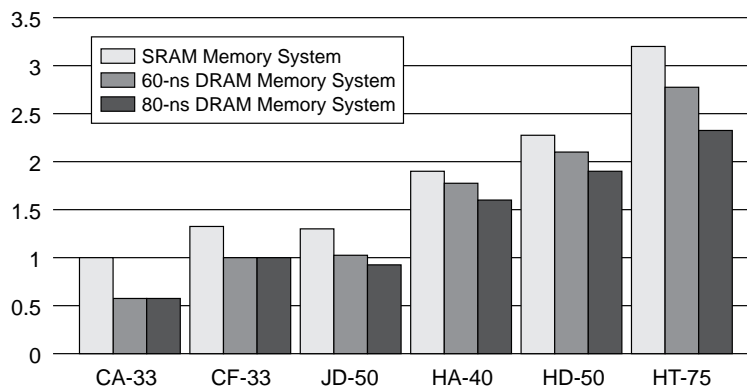


Figure 2. Relative performance of 960 microprocessors with various memory systems. As expected, memory speed matters more to clock-doubled and clock-tripled chips. (Source: Intel)

ume leader. Currently, over 60% of new design wins are in internetworking applications such as bridges and routers. RAID and ATM controllers may also be valuable markets for very fast embedded processors.

The H-series will help Intel strengthen the market-leading position of the 960 by showing customers it is committed to continuous upgrades of the family and that it is willing to use its most valuable process technology. It is possible to stick with the 960 and get almost any combination of price and performance.

On the surface, it seems senseless for Intel to build any H-series chips; the wafer starts would produce much more value for the company as DX4 or P54C die. For example, the 77-mm² DX4 sells for about \$400, while the larger P110 generates about one-fourth the revenue. Perhaps, though, by the time Intel is shipping the H-series in volume, it will have sufficient capacity. In any case, the volume of H-series chips—perhaps measured in the tens of thousands per quarter at first—will not displace much production of the more profitable x86 chips—measured in the millions per quarter.

Another issue is the tremendous number of products and the overlap among them in price and performance. Though Intel claims all 960 products are alive and well, some of the parts must either drop in price or be phased out of the product line.

As for the competition, AMD will reveal its super-scalar 29000 later this month at the Microprocessor Forum. The new 29K and recent MIPS R4200 and R4600 chips, in versions with integrated peripherals, might combine equal or better performance with lower system costs than H-series chips. Although the 960 family still lacks chips with the specialized, integrated peripherals of some of its competitors, Intel seems to be finding success by simply supplying a wide range of cost-effective, simple embedded processors. ♦