Alpha Hits Low End with Digital's 21066

First Microprocessor to Integrate PCI Bus Interface

By Linley Gwennap

Digital's 21066 microprocessor will significantly reduce the cost of Alpha systems while keeping much of Alpha's vaunted performance. The new chip cuts system cost by integrating cache, memory, and PCI interfaces—most of a system-logic chip set—onto the processor. The 21066 even includes a VRAM controller that can act as a low-cost graphics frame buffer.

The new chip, previously known as Low-Cost Alpha (LCA), has a clock rate of 166 MHz. While this is slightly faster than the current low end of the 21064 line, the performance of the 21066 is hampered by the 64-bit interface to cache and main memory, compared to the 21064's 128-bit memory interface. Digital expects the new chip to reach 70 SPECint92, about 20% less than its 150-MHz 21064-based workstation.

The company has priced the new Alpha chip quite aggressively at $424 (in quantities of 1,000). Intel's 50-MHz 486DX2 is listed at $406 but is actually more expensive than the 21066 when the cost of interfaces to cache, memory, and PCI is included. Digital's processor delivers nearly three times the performance of the DX2, based on SPECint92. The company is already sampling first silicon of the 21066; since the chip is leveraged from the 21064, Digital hopes that testing will go smoothly and volume shipments will begin in 1Q94.

The new processor could enable vendors to cut the price of Alpha PCs to around $3,000. The 21066 supports standard PC memory modules and, through its PCI interface and an external ISA bridge, standard PC peripherals; thus, the cost of a 21066-based PC should be close to that of a 486DX2 system, assuming similar volumes and sales channels. Many vendors sell DX2 systems today for $2,000–$3,000.

Although the announced 21066 pricing matches current 486DX2 prices, Intel's prices will undoubtably fall by the time the 21066 reaches volume production. A more accurate comparison would match the new Alpha chip against forthcoming 100-MHz 486DX3 processors.

When the 21066 debuts, 486DX3 chips will probably be priced about where the DX2 is today. Digital's strategy is thus to offer better-than-Pentium performance at the price of a 486DX3.

Highly Leveraged from 21064

The 21066 uses the entire processor core from the high-end 21064 *(see 060301.PDF)*. This includes the 64-bit, two-issue superscalar CPU and FPU, 8K of instruction cache, and 8K of data cache, as shown in Figure 1 (see below). The 21066 adds three major functional blocks: a memory interface for SRAM, DRAM, and VRAM; a PCI bus interface; and a phase-locked loop. The memory and PCI interfaces offer functions similar to the system-logic chip sets that Digital is developing for the 21064 *(see 070904.PDF)*.

A critical difference between the 21066 and the 21064/chip-set combination is that the new chip is restricted to a 64-bit data bus to the external cache. The 21064 uses a 128-bit cache bus, allowing it to refill an entire on-chip cache line in just two accesses. This change, made to reduce the cost of the chip and its memory system, lowers the performance by about 25%.

Other changes were made from the 21064 interface to reduce pin count, as shown in Table 1. The main address bus was completely eliminated by multiplexing the memory row and column addresses onto the cache index bus. The 21066 calculates a single ECC value across 64 bits, while the 21064 maintains ECC separately for each 32-bit section of its 128-bit bus. Control signals for cache and memory are simplified by eliminating multiprocessor support and other complex transactions. Even with the PCI interface, the 21066 fits into a 287-pin package, saving cost from its predecessor's 431-pin package.

SRAM, DRAM, VRAM on One Bus

To keep the pin-count low, the 21066 uses a single 64-bit data bus to connect to cache memory (SRAM), main memory (DRAM), and an optional graphics frame buffer (VRAM). An output-enable signal keeps the

SRAM off the bus when it is not being accessed, but external buffers are needed for the DRAM and VRAM, as shown in Figure 1.

A single address bus is used for all three types of memory. During a cache access, this bus carries the cache index. When accessing DRAM or VRAM, it carries the row and column address. To increase the drive current, these signals require external buffers in all but the smallest memory configurations.

The external cache size can range from 64K to 2M. Using 128K × 8 SRAMs, the cache requires eight chips totaling 1M. If ECC protection is desired, ×9 parts can be used instead. Like the 21064, the new chip supports a unified, direct-mapped, write-back cache. The cache timing is programmable.

The external cache uses a line size of eight bytes. This allows for one dirty bit for each 64-bit value, reducing the amount of write-back traffic. It also allows the 21066 to implement a simple allocate-on-write strategy, since the entire cache line is filled by the data from a single write.

For a 1M cache, the tags are nine bits wide to support the maximum memory size of 512M. An additional bit per line is needed for the dirty bits, plus an optional bit for tag parity. The 21066 eliminates the need for a separate valid bit by assuming that, after initialization, all lines are always valid. The tag SRAMs must have the same access time as the cache data SRAMs.

Using 12-ns SRAMs, the external cache can be read every four cycles at 166 MHz. It takes four 64-bit accesses to fill a line in the on-chip cache (4-4-4-4 access pattern), making a total of 16 cycles. While this miss penalty appears daunting, the CPU reduces the impact on throughput by loading the critical word first and restarting execution as soon as it is received. The processor can also continue instruction execution during a
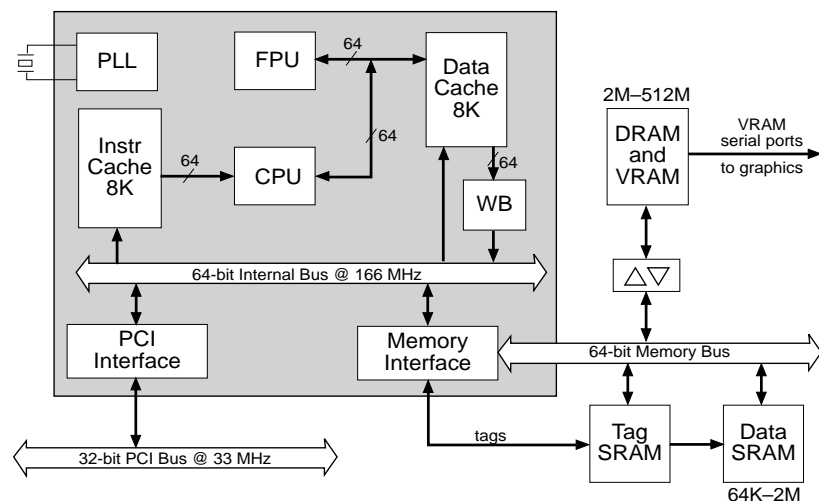
		21064	21066	Change
Processor Bus	Data	128	64	-64
(21064) or	ECC	28	8	-20
Main Memory	Address	29	n/a	-29
(21066)	Control	29	16	-13
Cache Control		40	31	-9
PCI Bus		n/a	50	+50
Interrupts		6	3	-3
Clocks, Test, etc.		31	17	-14
Total Signal Pins		291	189	-102
Power & Ground		140	98	-42
Total Package Pins		431	287	-144

Table 1. The 21066 uses 144 fewer pins than the 21064 despite adding a memory interface and a PCI interface.

cache miss if the requested data is not used in a subsequent instruction.

Cache write bandwidth is half of the read bandwidth, because a write must first read the tags, then turn the bus around to write to the data SRAM. This tradeoff was made because write timing is not as critical as read timing, since the processor does not wait for a write to complete before proceeding.

Memory Controller Does Graphics

If a read request misses the external cache, the 21066 accesses main memory directly. Using 70-ns page-mode DRAMs, it takes about 16 cycles to receive the first 64-bit value. Subsequent values are returned in nine cycles (16-9-9-9 pattern). These cycle counts seem huge by x86 PC standards, but remember that they are measured against a 6-ns processor clock, not the leisurely 15–30 ns clocks used by most x86 CPUs.

The 21066 generates all the control signals needed for one to four banks of memory. Like the address lines, many of these control signals require external buffers to drive a large number of DRAMs. The main memory can be as small as 2M or as large as 512M. Each bank of memory has individually programmable timing and supports optional ECC.

One of the memory banks can be configured for VRAM instead of DRAM and used as a graphics frame buffer, eliminating the need for a graphics accelerator chip. In this arrangement, the CPU performs all graphics operations and writes the results to the frame buffer; a simple ASIC (or set of PLDs) serially transfers the VRAM contents to a RAMDAC that updates the screen.

To improve graphics performance in this mode, the 21066 implements some simple raster operations and pattern fills. It can also write one bit at a time in the frame buffer. For better graphics performance, a separate PCI-based graphics accelerator can be used instead.

Figure 1. The 21066 includes the processor core from the 21064 plus a memory controller and a PCI bus interface. WB indicates the write buffer.

PCI Provides Direct Peripheral Interface

The 21066 is the first microprocessor to integrate a PCI interface. This is somewhat ironic, as PCI was developed by rival processor vendor Intel. The x86 vendor has talked about integrating PCI onto its processors at some future time, but Digital has beaten them to it.

The Alpha chip includes a 32-bit PCI bus interface. It supports burst transactions and bus speeds up to 33 MHz. The chip uses a separate PCI clock input so the PCI bus can run asynchronously to the CPU. Like most PCI interfaces, the 21066 uses a scatter/gather unit that provides virtual-to-physical address translation for PCI-initiated DMA transactions, removing this burden from the processor.

The PCI interface allows the emerging set of PCI-based peripherals, primarily intended for x86-based PCs, to be used with the 21066. Not leaving anything to chance, Digital is developing its own PCI chips for graphics and networking, which it plans to market. The 21066 can use Intel's SIO chip *(see 061602.PDF)* to connect to ISA peripherals. The 21066 supports the non-PCI "side signals" that are used by the Intel bridge chip, which is available for about $21.

To further simplify system design, the 21066 eliminates the need for a high-frequency clock input by implementing an on-chip phase-locked loop. With its PLL, the 21066 can multiply the input clock by any integer between two and nine, greatly reducing the oscillator frequency. For example, a 166-MHz 21066 could be clocked from the 33.3-MHz PCI clock simply by setting the PLL to use a multiplier of five. The 21064, in contrast, requires a 2× clock input—400 MHz for the 200-MHz version.

Advanced Process Reduces Die Area

The 21066 uses the same 0.68-micron process as the current 21064. (The original 21064 used a 0.75-micron process.) It is a three-layer-metal CMOS process that is among the most advanced in the world for microprocessors, keeping the die area down to 209 mm^2. Figure 2 shows the small size of the added memory and PCI interfaces. The MPR Cost Model *(see 071004.PDF)* estimates the manufacturing cost to be $230, about 33% more than an R4400PC but much less than Pentium.

Digital is already working on a 0.5-micron version of the 21066, which it expects to be ready for shipments next year. This could cut the die size to under 150 mm^2, reducing the estimated manufacturing cost to under $150. The shrink could give the 0.5-micron version a cost advantage over the R4400 and the future 0.6-micron Pentium—assuming Digital can build enough of its chips to make its fab overhead competitive.

The 21066 operates at 3.3V, but its inputs and outputs are 5V-tolerant, allowing the chip to connect directly to standard 5V memories and logic without volt-

age-translation buffers. The chip's PCI interface can drive up to 10 bus loads directly.

Since it uses most of the circuitry of the 21064, the new processor is also a power hog. Digital has not yet made measurements on the 21066 but expects it to dissipate about 20–23 watts maximum, about 30% more than the power-hungry Pentium. Digital is able to cool the 21064 with a 3" square heat sink that is only slightly larger than the chip's 431-pin package; the 21066 uses a package that is a bit smaller but adds a new heat slug to maintain adequate cooling.

The high-integration strategy of the Alpha chip is reminiscent of the ill-fated 486SL. When Intel added DRAM control and an ISA bus interface to its 486DX, however, the die size ballooned from 82 to 167 mm^2, an increase of 104%. This increase made the 486SL much more expensive to manufacture, a poor tradeoff considering the low price of system-logic chip sets for the 486. Ultimately, Intel terminated development of the 486SL family in favor of relying on external system logic.

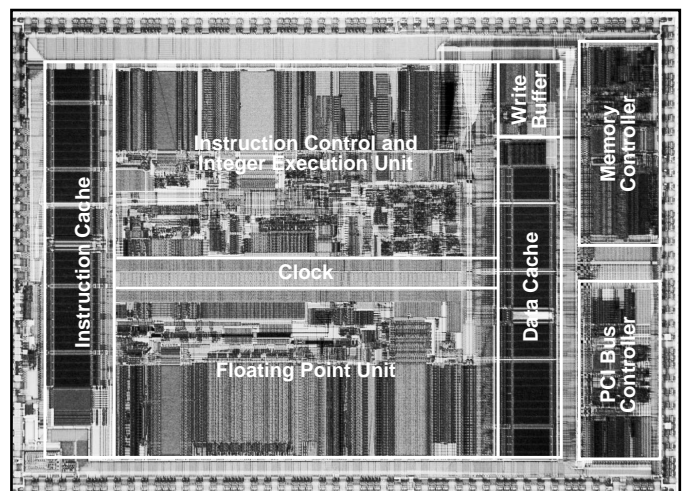The 21066 should fare better, as its die area is only 8% larger than the 21064 when it is built with the same

Figure 2. The 21066 uses 1,746,892 transistors on a 12.3 × 17.0 mm (484 × 670 mil) die. The features that have been added to the original 21064 core—including the memory and PCI interfaces—take only 10% of the die.

21068 For Embedded Apps

Digital also announced the 21068, the first member of a planned 21x68 family of embedded processors. The chip is functionally identical to the 21066 but is rated at just 66 MHz. Even at this speed, the chip generates about 70 Dhrystone MIPS, about 35% faster than the 33-MHz i960CF *(see 0712MSB.PDF)*, Intel's top of the line. The marketeers picked this frequency to reduce power consumption below 10 W, a figure that the company feels is an upper limit in the embedded market. At 66 MHz, the 21068 is rated at 8.5 W maximum.

Digital has priced the 21068 at $243. While this seems high (the aforementioned 960CF costs $159), the company believes it can penetrate the high end of the embedded market in applications such as single-board computers and network routers. The 21066 includes system-logic functions that the i960 does not, though it lacks the '960's interrupt and DMA controllers. Digital also sees an advantage in using its workstations as native development platforms for embedded designs; i960 applications are built using cross-development tools on x86 PCs. The i960, however, has a much more extensive set of tools, including in-circuit emulators.

The MPR Cost Model predicts that the 21068 will be a loss leader for Digital, but the company sees its chip as a foot in the door until new 21x68 chips can be deployed. The next products will be derivative designs. An easy change would be to eliminate the large clock driver *(see 060301.PDF)*, which is not needed at such low speeds; this would cut the power nearly in half, or allow for twice the performance at the same power level.

Eventually, the 21x68 line will encompass the low-power chips being developed at Digital's new Palo Alto design center *(see 0710MSB.PDF)*. These chips, based on a new CPU core, are meant for mobile applications but their low cost will appeal to embedded users .

IC process. The MPR Cost Model predicts that the manufacturing cost difference between the two Alpha chips is less than $20, significantly less than it would cost to build an external system-logic chip set. Digital has kept the area increase small even while it has added memory and bus interfaces that are more complex than those on the 486SL.

Pentium-Level Performance

Digital has not yet measured the performance of the 21066, but predicts that the chip will reach 70 SPECint92 and 105 SPECfp92, based on its performance simulations. This positions 166-MHz 21066 systems just below the 133-MHz 21064-based Model 300 in Digital's product line. Despite its clock-rate advantage, the new processor is slowed by its narrower interface to cache and main memory, which reduces performance by about 25%. (High-frequency processors are quite susceptible to

changes in the memory system.)

More importantly, the 21066 is about 10% faster than a 66-MHz Pentium processor with a fast memory system. With its smaller die, Digital could keep the cost of the 21066 below that of Pentium, although this will become increasingly difficult as the volume difference between the two processors grows.

At 100 MHz, the 486DX3 should achieve about 40–50 SPECint92. This would give the Alpha processor a performance advantage of around 60% over the fastest 486 chip expected in 1994. If Intel follows its historical pricing pattern, the 21066 and the 486DX3 should have about the same price, giving the Alpha chip a significant price/performance advantage. However, the DX3, in a 0.6-micron process, will have a tiny die, giving Intel plenty of room for price cuts if the 21066 proves to be more than a minor annoyance.

Although Digital probably chose to offer the 21066 at 166 MHz to maximize yield, there is no apparent reason why the chip shouldn't have significant yield at 200 MHz; it uses the same CPU core and 0.68-micron process as the 200-MHz 21064, and the memory and PCI timing are programmable. The company may be saving the higher-frequency parts for its own systems, as it did initially with the 21064, although repeatedly withholding the best chips could damage Digital's credibility as a chip vendor. In any case, the 0.5-micron 21066 should easily reach 200 MHz or perhaps higher; the 0.5-micron version of the 21064 is expected to reach 300 MHz.

Instant PC—Just Add Memory

The 21066 greatly simplifies the design of an Alpha-based system. It eliminates the need for a system-logic chip set and supplies popular PC interfaces, allowing a system designer to treat the Alpha processor much like an x86 chip set. The new chip's PLL gets rid of the 400-MHz input clock that frightens many potential 21064 customers. The system designer now has little need to know the details of Alpha technology and can build a system using standard memory, buffer, and I/O chips.

The new chip reduces system cost over current Alpha designs in several ways. The 21066's relatively low price (compared to the 21064) offers some cost savings. Moving the system logic on-chip eliminates the many PALs and ASICs used for these functions today. The biggest break, however, comes from savings in the memory and I/O systems. Switching to a 64-bit-wide cache and main memory reduces the number of parts needed for a minimum system. PCI, coupled with a low-cost ISA bridge, allows system vendors to use high-volume PC interfaces.

These moves simply put Alpha on parity with x86 systems, however. System logic for Pentium is available for $84, including PCI and ISA interfaces, and already

uses 64-bit paths to cache and main memory. Intel's Pentium chip set also features integrated cache tags.

This puts the onus on Digital to maintain a significant price gap with Pentium. It appears that a 21066 system will outperform a similarly configured Pentium system by about 20%. This in itself is not enough of an edge to convince people to switch to a new architecture with few applications and fewer system vendors. The 21066's price is about half of Pentium's current tag; Digital must maintain that differential if, as expected, Intel drops its prices in 1994.

The 21066 again demonstrates Digital's technology leadership. It is the first microprocessor with a PCI interface, and the first to offer such a complete level of system integration: processor, cache, external cache control, memory control, bus interface, and even some graphics features. Digital's initial pricing opens a 2× gap in price/performance over Pentium; this wide gap could attract the PC system and application vendors that the company needs to make its Windows NT strategy succeed. To close these deals, Digital needs to overcome the lack of an installed base for Alpha and its own reputation as a high-priced processor vendor. If the company fails, it won't be for lack of trying. ♦