AMD Gets the IIIrd Degree

K6 III Positioned Against Pentium III, But Will PC Makers Buy It?

by Linley Gwennap

Seeking to escape the profitless pit of Celeron competition, AMD has begun shipping the long-awaited K6 III. With both performance and nomenclature, the company shows that its part is competitive not only with Intel's Pentium II but with the just-released Pentium III (see MPR 3/8/99, p. 1). The new K6 is a convincing argument that consumers, and even businesses, should consider an AMD processor instead of an Intel chip. But PC makers are a staid lot, and it will take time for AMD to convert its technical prowess into design wins in more expensive PCs.

With the new name, the company has adopted the Roman numerals that are *sine qua non* for Super Bowls and PC processors, what with Pentium II and III, Cyrix's MII (that's M2, not 1002), and Rise's forthcoming mP6 II. (Any day now, IDT will surely release a WinChip II.) The K6's III is both a numerical increment from the K6-2 and an unsubtle indication that the processor is ready to rumble with Pentium III, Intel's top-of-the-line PC processor.

The AMD chip—previously known as the K6+ 3D, the K6-3, and Sharptooth—has been in the works for some time; initial details of the part were disclosed at Microprocessor Forum 1997, and AMD has had working silicon since last summer. Its introduction one week before Pentium III is no coincidence; the company had been waiting until it needed to deploy the chip, as its larger die size will reduce AMD's processor output. But with the K6-2 under fire from Celeron, and with a juicy target like Pentium III available, the time is right for AMD to advance its technology.

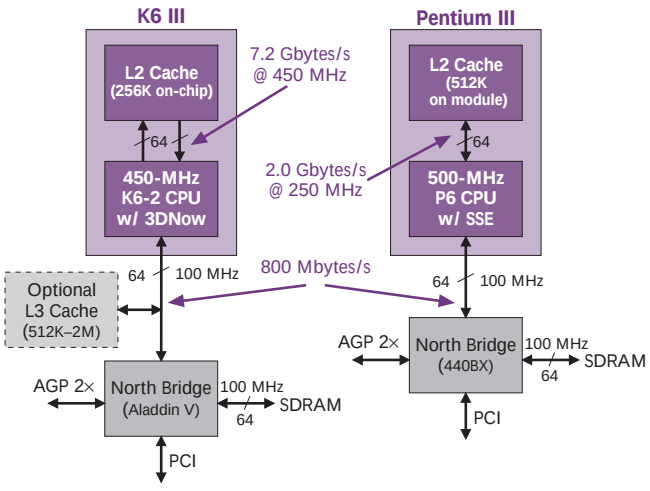

Figure 1. The K6 III matches the dual-bus structure of Pentium III; in fact, its dual-ported full-speed cache delivers nearly four times the bandwidth of Intel's single-port half-speed cache.

Integrated Cache Breaks Socket 7 Bottleneck

As previously disclosed (see MPR 10/27/97, p. 19), the K6 III combines the CPU core from the K6-2 with 256K of on-die level-two (L2) cache. This design breaks a major constraint of the Socket 7 architecture, giving the K6 III the same dual-bus architecture used in Intel's P6 processors. As Figure 1 shows, a K6 III system has a nearly the same bus structure as a typical Pentium III system. The only external difference is that the K6 III supports an optional level-three (L3) cache, which is impractical in a Pentium III system.

By placing the L2 cache on the CPU chip, AMD can run it at the full CPU speed, delivering at least twice the bandwidth of Intel's half-speed external cache. Taking further advantage of the integration, AMD made its L2 cache dual ported, allowing it to process one read and one write per cycle. This design further increases the peak cache bandwidth, delivering 7.2 Gbytes/s in a 450-MHz K6 III. This is four times the bandwidth achieved in a Pentium III at the same clock speed.

We expect Intel to integrate an L2 cache in the 0.18-micron Pentium III, code-named Coppermine, in 2H99. This cache, like AMD's, should run at the full CPU speed. If Intel uses the same cache design as in its current Mendocino (see MPR 8/24/98, p. 1) and Dixon chips, however, it will not be dual ported. This situation will prevent Coppermine from matching the K6 III's L2-cache bandwidth.

Like Pentium III's, the K6 III's L2 cache is four-way set-associative. A read and a write must access different sets to execute in the same cycle. The cache is nonblocking; up to two cache misses can be in progress without stalling the CPU on a cache hit. The L2 cache has a latency of three cycles; at 450 MHz, this is 4.5× faster than an external L2 cache on the K6-2. AMD did not implement 256-bit on-chip buses, so it still takes four cycles to transfer an entire cache line.

The K6 III continues to use the "Super" Socket 7 bus to access main memory and I/O, maintaining compatibility with K6-2 systems. Like Pentium III's Slot 1 bus, this bus runs at 100 MHz and delivers a peak bandwidth of 800 Mbytes/s. One drawback of the Socket 7 bus is that only one transaction can be pipelined. This limitation can bog down performance when many lines are being flushed to DRAM.

The L3 cache helps reduce the impact of this problem. AMD's measurements indicate that a 512K L3 cache improves the K6 III's application performance over a system with no L3 cache by about 4%. This sizable boost is surprising, given an L3 cache that is only twice as large the L2 cache. But the L3 acts as a cast-out buffer, holding lines that are flushed from the L2 cache without stalling the bus during a long main-memory write.

Performance Comes at a Cost

As Figure 2 shows, the integrated L2 cache consumes a third of the K6 III's 118-mm^2 die. According to the MDR Cost Model, the new chip costs about $45 to build, $10 more than the K6-2. Worse yet, the increase in die size reduces the net die per wafer by 40%, according to our model. The reduction would be greater had AMD not used redundant rows to map out some defects in the cache array. The lower yield means that every wafer AMD devotes to K6 III produces just over half as many good chips as a K6-2 wafer.

This explains the vendor's hesitation in announcing the new chip. During 1998, AMD's shipments were limited by the number of chips it could produce. By focusing on the K6-2, AMD boosted its output from 1.5 million in 1Q98 to more than 5 million in 4Q98. It could not have achieved such rapid growth had it shifted production to the larger K6 III.

Even in 1999, AMD will probably phase in the new part slowly to prevent a decline in unit shipments. Ironically, Intel's newly aggressive Celeron tactics (see MPR 1/25/99, p. 18) will help AMD ramp the K6 III more quickly. If Intel succeeds in blocking AMD's unit-share gains, AMD can devote more of its growing wafer capacity to the larger die. If AMD continues to rapidly gain design wins, it will have to rely more heavily on the K6-2 to meet demand.

The increase in manufacturing cost is less significant. If all goes well, the K6 III's average selling price (ASP) will be much higher than that of the K6-2, far outweighing the $10 increase in cost. In fact, the company would be better off selling fewer processors if such action results in greater profits.

Even at $45, the K6 III costs about $10 less to build than Intel's Celeron (Mendocino) and $25 less than the initial Pentium III (Katmai) modules. These cost advantages are due to two factors. First, the K6 core is smaller than the P6 core by about a third. Intel argues that the P6 core is more powerful, but most benchmarks show little performance difference between the two.

Second, AMD's L2-cache design is also compact; the K6 III's 256K cache is a third smaller than the 256K cache on Intel's Dixon (see MPR 1/25/99, p. 20). The metal pitches of AMD's CS-44 process (see MPR 9/16/96, p. 11) are similar to those of Intel's P856.5 process, but CS-44 provides a local interconnect layer not found in P856.5. This extra feature is valuable for reducing SRAM cell size.

In fact, AMD's die-size advantage is so significant that the 0.25-micron K6 III is likely to be smaller than Intel's 0.18-micron Pentium III (Coppermine), which we expect will also include 256K of on-chip cache. In previous processor generations (such as the 386 and 486), AMD's parts were always larger than Intel's, because the smaller company couldn't devote as many engineers to circuit design and IC process development. Today, the shoe is on the other foot, as AMD is trying to kick Intel's booty with a more compact CPU and a better manufacturing process.

It seems odd that AMD is introducing on-die cache at the top of its product line while Intel is using on-die cache mainly in its low-end Celeron parts. Intel, however, had already made the leap of including L2 cache as part of the "processor." Because the original Pentium II featured a discrete L2 cache on a module, moving that cache onto the die provided a cost reduction. In AMD's case, the L2 cache is a performance enhancement that adds cost, so the company is using it as a premium feature. With its Coppermine Pentium III, Intel will also introduce on-chip cache at the high end.

Price & Availability

The AMD K6 III is available now at 400 MHz; the 450-MHz version is sampling, with volume shipments set for March. In 1,000-unit quantities, the two parts list for $284 and $476, respectively. For more information, access *www.amd.com/K6*.

K6 III's Performance Matches Pentium III's

According to AMD's testing, the K6 III is more than a match for Pentium III at the same clock speed. On Winstone 99, which represents typical PC productivity applications, a 450-MHz K6 III rated 5% faster than a 450-MHz Pentium II with Windows 98. (Intel's tests show a Pentium II–450 and a Pentium III–450 are equivalent on Winstone 99.) This difference seems small, but AMD claims the K6 III is actually faster than a Pentium III-500 as well, as Figure 3 shows.

Previous versions of the K6 fared less well on Windows NT, which generates more memory traffic and thus encounters the Socket 7 bottleneck. With its dual-bus architecture, the K6 III does well on NT, rating 4% better than Pentium II at the same clock speed. In this case, however, the new 500-MHz Pentium III barely tops the K6 III-450.

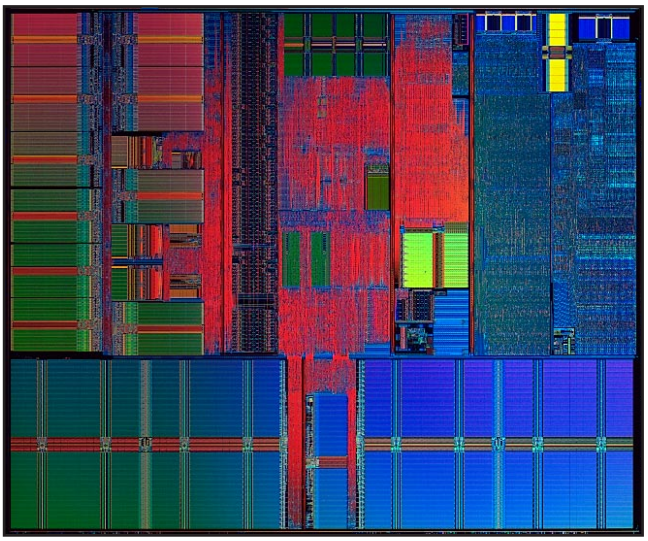

Figure 2. The K6 III contains 21.3 million transistors and measures 9.8 × 12.0 mm in AMD's 0.25-micron five-layer-metal process. The 256K L2 cache (at bottom) consumes 34% of the die area and 56% of the transistors.

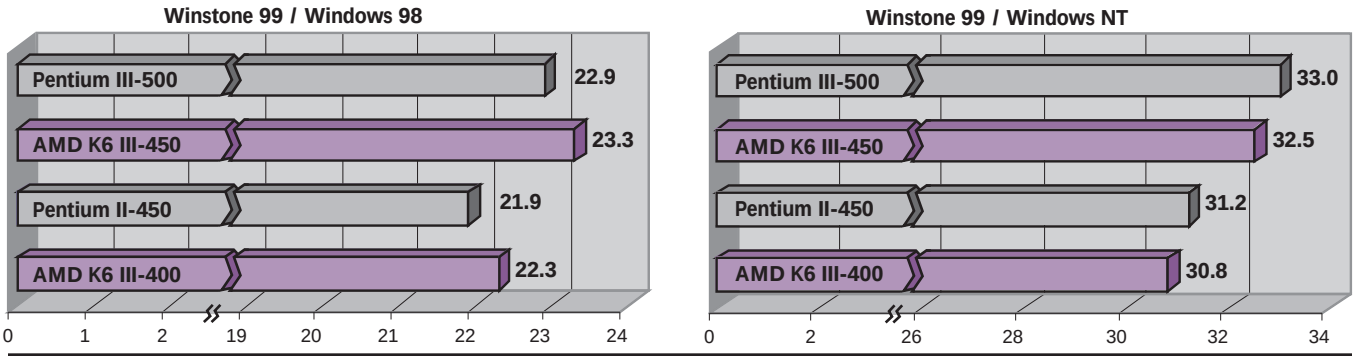

Figure 3. When running Ziff-Davis's Winstone 99 benchmark, the K6 III-450 is faster than a Pentium III-500 on Windows 98 and faster than a Pentium II-450 on Windows NT. Both systems use a high-end Maxtor DiamondMax 2500 Plus IDE disk, an STB Lightspeed 3300 16M AGP graphics card, and 64M of SDRAM (128M for NT). The K6 III uses the ALi Aladdin V chip set with 512K of L3 cache on an Asus P5A motherboard, while the Pentium II/III uses the 440BX chip set and no L3 on an Asus P2B motherboard. (Source: AMD)

At this point, it's too early to say how the two chips will compare on 3D applications. AMD claims that 3DNow delivers 20% better performance than Pentium II on applications that have been enhanced for 3DNow. Although AMD touts an impressive list of 3DNow applications, many are only partially optimized and thus do not currently see the full performance benefit.

Pentium III contains Intel's new SSE instructions, which also boost 3D performance for applications that take advantage of them. Intel has not published 3D application benchmarks for Pentium III, but we expect SSE to deliver a boost of between 15% and 30%, similar to 3DNow's. Both new processors have a peak execution rate of 4 FLOPS per cycle, so neither has an inherent advantage on optimized 3D code. Pentium III has an edge on unoptimized 3D applications, as it delivers about 25% better performance on standard x86 floating-point code.

The performance data in Figure 3 was measured with an expensive cached hard drive typical of a high-end configuration. AMD would not provide results for a more mainstream configuration. On the other hand, AMD is giving up a bit of performance by benchmarking its chip with a 512K L3 cache. The Aladdin V chip set supports up to 2M of L3 cache, which would boost performance by another 4%, according to AMD. Conversely, the K6 III can be used with no L3 cache, with only 4% performance loss.

The K6 III is likely to be used without an L3 cache in notebook systems, where the extra performance isn't worth the power consumed by the cache. AMD has not yet announced notebook versions of the K6 III, but the new chip should fit into the mobile power envelope. By definition, the K6 III CPU core dissipates the same power as the K6-2; the cache adds 2 W, about what is used by the external L2 cache in a K6-2 system.

AMD Assumes the Position

With a $476 list price, the K6 III-450 sets a new high-water mark for AMD. More important, this price puts the chip in direct competition with Pentium III. In fact, the K6 III-450 is only $20 less expensive than the Pentium III-450 and the same price as the Pentium II-450. Similarly, the K6 III-400, at $284, has the same price as the Pentium II-400. This seems untenable, as AMD has always had to offer a 20% or greater discount for the same performance as Intel's.

AMD's plan is to position the K6 III against Intel's premium products while the K6-2 continues to get down and dirty with Celeron. The K6-2 will provide the bulk of AMD's shipments for some time, as it will continue to thrive in the sub-$1,000 PC market, where only MHz matters. With Intel keeping even the fastest Celerons priced below $150, AMD needs to target Pentium II and Pentium III to boost its prices. The K6 III's role is to take on these chips, appealing to more sophisticated PC buyers who have bigger budgets.

As Figure 3 shows, the K6 III has the performance to compete well with Pentium II and even Pentium III. AMD hopes that the new chip's stronger performance on Windows NT will help it break into the business PC market, where AMD has had little success to date. Because businesses buy two-thirds of all PCs, this market is crucial to AMD's hopes of securing a 30% market share.

Although AMD assumes this position will hold, the flaw is that in the PC market performance is becoming less relevant than branding. On most applications, Celeron, a Pentium II, and a Pentium III at the same clock speed all have virtually the same performance, yet Intel positions these parts very differently. Because of this performance overlap among its own lines, Intel could easily show that a 400-MHz Celeron delivers nearly the same Winstone 99 performance as a 400-MHz K6 III.

Intel must be careful in making such an argument, however, as it would undercut its own fragile positioning of Celeron and Pentium II/III. AMD would like to duplicate this positioning with the K6-2 and K6 III, respectively. We expect the K6 III will succeed in pushing AMD into new PC segments, but volumes will ramp slowly, as building momentum in high-end markets takes time. On its technical merits alone, the K6 III clearly deserves to be more than the core of a sub-$1,000 PC. M