Cahners
MICROPROCESSOR REPORT
www.MPRonline.com
THE INSIDER'S GUIDE TO MICROPROCESSOR HARDWARE

Intel Delivers Cascades-Based Xeon

New Xeon Processor Ships at 700MHz With up to 2M L2 Cache

By Kevin Krewell {5/29/00-03}

Intel recently refreshed its Pentium III Xeon line with two products based on the new Cascades die. Cascades offers 1- and 2M on-chip L2 caches, using the same advanced transfer cache technology that Intel introduced on Coppermine (see *MPR 10/25/99-01*, "Coppermine Outruns Athlon"). The large-cache Xeons, coupled with the eight-way Profusion chip set, target server applications that support large corporate and Internet commerce applications, such as enterprise resource planning (ERP); Internet transaction servers and databases; back-end storage; and application-hosting servers. Intel expects the 700MHz Cascades to deliver 27–37% improvements over the 550MHz Pentium III Xeon on transaction-processing benchmarks.

Although Intel is rapidly converting standard Pentium IIIs to flip-chip PGA (FC PGA) for cost savings, it continues to sell Xeon processors only in the SC330 (Slot 2) cartridge. With the processor L2 located on chip, the brick-size cartridge is now largely unnecessary, but Intel offers additional capabilities, such as temperature and voltage sensors and system management (SM) bus, only in the SC330 cartridge. Intel is also reluctant to change form factors in the server market, where consistency and continuity are paramount. In addition, Intel continues to sell small-cache (256K L2) Coppermine-based Xeon processors for two-way (dual-processor) servers and workstations in the SC330 cartridge. OEMs designing two-way servers in the popular 1U and 2U (1.75 and 3.5 inches high, respectively) rack-mount form factors will use Pentium III processors, as the Xeon module is too tall to fit into those chassis. Intel recently began selling Pentium III processors in the lower-profile FC PGAs that are dual-processor capable, enabling two-way rack-mounted servers in a 1U height.

Intel will target the large-cache (2M) Xeon at maximum-performance four-way and above multiprocessing servers. The 2M L2 cache has the capacity necessary to contain large data sets, which reduces bus traffic and improves performance in these systems. The 1M L2 version is targeted at lower cost four-way servers running less demanding applications.

The eight-way set-associative advanced-transfer-cache design from Coppermine was adapted to the larger cache array of Cascades. The data path to the L2 is still 288 bits wide, allowing for 256 data bits and 32 ECC bits. Data is transferred every other cycle, delivering 2.5x the aggregate bandwidth of the external full-speed cache used on the earlier Katmai-based 550MHz Xeon. But the greatest performance benefit of moving the cache on chip is the reduction in the initial-access latency.

The 385mm^2 Cascades challenges the 477mm^2 PA-RISC 8600 for biggest processor die but falls short. The PA-8600, however, is built in 0.25-micron technology. HP will migrate to the 0.18-micron 8700 next year (see *MPR 5/22/00-02*, "HP Extends PA-RISC With 8700") with a 304mm^2 die size—21% smaller than Cascades—while offering 12% more on-chip cache. Manufacturing the 385mm^2 die will be demanding, but Intel has taken steps to increase yield. With large memory arrays on chip, redundant memory rows and columns are used to map out defective memory cells, increasing the number of recoverable die. Intel can also sell 1M parts—dies where, even with the redundant memory,

Price & Availability

The 700MHz Pentium III Xeon is available now. The product is offered with either a 2M L2 cache at $1,980 or a 1M L2 at $1,177. These list prices are for 1,000-piece quantities. Initial shipments have begun, but availability will be tight for the next two months.

half of the 2M cache is damaged beyond repair. Considering the problems Intel has had meeting volume production goals for Coppermine, the large Cascades die could eat up precious wafer capacity. We estimate that an 8-inch wafer will yield only 30 good Cascades parts. Despite the low numbers of die per wafer, however, the MDR Cost Model estimates that Cascades costs only $180 to manufacture.

To maximize yield, Intel chose a conservative 700MHz speed, such that the vast majority of the units will pass speed sorting. Intel is also under less pressure to push the speed boundaries in servers, where the latest RISC processors are running no faster than 700MHz. Intel has kept the 100MHz front-side bus for the large-cache Xeon processors, to maximize compatibility with the existing infrastructure—not because of yield concerns.

Xeon Gaining Market Share

The Xeon line has been gaining market share in mid-to-large servers over $10,000. Intel points to Dataquest numbers that, from 1998 to 1999, show Xeon processors going from number three to number one in four-way and greater servers. Intel may be on its way to dominating this market as it does the PC market. As Intel server platforms become commodities, server prices will drop. This will be good for e-businesses but is likely to reduce the margins of server manufacturers. Other beneficiaries of this trend will include Windows and Linux, the most popular operating systems for IA-32 processors.

While most RISC processor vendors, including IBM, HP, SGI, and Compaq, have adopted Intel server processors in addition to their RISC offerings, Sun remains the most conspicuous hold-out. Like Apple in the PC business, Sun appears to have a loyal following. But, unlike Apple, Sun's following is mostly in Fortune 1000 corporate IS departments. Those customers don't appear to be swayed away from buying Sun products, even though SPARC processors have had less than competitive performance. Service, support, scalability, and dependability seem to be the keys to these customers' hearts. There is also a strong market perception that Sun "got" the Internet earlier than the competition and understands it better.

Looking Ahead to 2001

The second half of 2001 will be interesting, when Willamette-based Foster processors will be available at 1.5GHz or better, and IA-64 McKinley processors will begin shipping at similar speeds. Both will share Intel's 870 chip set, which can scale from four-way and up (see *MPR 3/13/00-03*, "Intel Offers a Peek at 870 Chip Set"). In addition, AMD plans to ship the x86-64 SledgeHammer processor (see *MPR 10/25/99-06*, "AMD Shows Big Server Plans"), which, with the lightning data transfer (LDT) bus, can be configured as an *n*-way ccNUMA server. Foster will extend the conventional IA-32 performance and scalability; McKinley will offer an improved entry into Intel's 64-bit future; and SledgeHammer will offer the adventurous a transparent 64-bit extension of IA-32. The server market will then have multiple x86 architectural choices, which could be confusing to customers. Although the Cascades-based Pentium III Xeon, introduced May 22, is a very capable transition product, the real fireworks will happen next year.