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

AMD Takes Hammer to Itanium

Reveals First Details of 64-bit Hammer at MPF2001

By Kevin Krewell {11/26/01-02}

The first significant microarchitecture details of the Hammer architecture, AMD's entry into the enterprise multiprocessing server market, were released at **Microprocessor Forum 2001** by Fred Weber, vice president and chief technical official of AMD's computation products group.

Hammer will include a dual-channel DDR SDRAM controller on die and will use up to three coherent HyperTransport links to connect to neighboring Hammer processors without requiring extra chips. In contrast, in the Intel870 chip-set architecture, up to four McKinley (IA-64) or Foster processors share a bus to the memory hub that has an external memory-translation hub to the memory, and an additional switch chip is required to connect more than four processors. While one of the key architectural features of the Hammer processor family is the extension to 64 bits, the integration of the memory controller and the glueless multiprocessing capability are initially more relevant to the mainstream server market. The 64-bit capability may provide an "insurance sell" for future large-application support.

AMD is promising all this goodness in a very svelte die size. It revealed at a recent analysts meeting that the small-cache version of the Hammer family, Clawhammer, will be 104mm^2 in the 0.13-micron silicon on insulator (SOI) semiconductor process, and that it is currently in tapeout. AMD plans include, in 2H03, 90nm semiconductor technology that will shrink a Clawhammer-derivative to 64mm^2. The large-cache member of the Hammer family is called Sledgehammer and will be targeted at four-way and greater systems. The first Sledgehammer processor is expected to sample in 2H02 and to go into production in 1H03. We expect Sledgehammer will have a 1MB L2 cache on chip, while Clawhammer will stick to a 256KB L2 cache for the mainstream desktop, mobile, workstation, and two-processor server markets.

At the same analysts conference, AMD also revealed that the Hammer family will continue to use the model numbering system introduced with the Athlon XP (see *11/12/01-02*, "Athlon XP Eschews GHz"). AMD projects that a Hammer processor will ship as model number 3400 in 4Q02, reaching model number 4000 in 1H03 and 4400 in 2H03. It was not clear whether AMD will compare Hammer with Intel's Xeon or Itanium processor—or whether the model number references the improved performance over the Athlon XP core.

Haven't We Seen This Core Before?

From a microarchitecture viewpoint, the Hammer core shown in figure 1 looks remarkably like its predecessor, the K7/Athlon (see *MDR 10/26/98-01*, "K7 Challenges Intel"). AMD still adds in the instruction cache predecode bits that define x86 instruction boundaries. It still turns variable-length x86 instructions into fixed-length macroOps (MOP), which are later decoded into RISC ops (ROP). The Hammer architecture can still decode up to three x86 instructions per cycle; issue up to nine ROPs per clock cycle; and reorder up to 72 ROP instructions, with the number of functional units remaining unchanged. The L1 caches are each 64Kb two-way set-associative. Complex x86 instructions also follow a parallel decode path to a micro-ROM (MROM), as Athlon did.

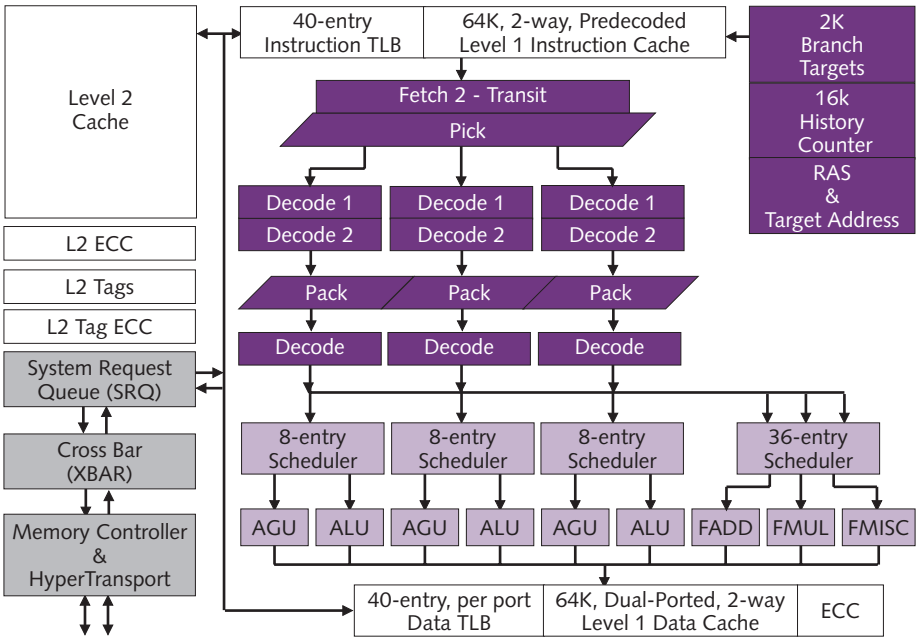

Figure 1. The Hammer microarchitecture looks much like an enhanced Athlon (K7). The fetch and decode sections of the pipeline exhibit have been redesigned to allow higher clock frequencies. The L1 instruction cache includes predecode bits, as did the Athlon core, but AMD's designers redefined the bits to include branch- prediction information.

Hammer clearly builds on the strong Athlon design and continues to improve it in areas such as the hardware decoders, where fewer instructions require the slower MROM. However, the predecode bits in the L1 instruction cache now include more information on branch prediction. The L1 data cache is now fully dual ported and can support two 64-bit stores in one cycle.

The microarchitectural improvements focused on streamlining the decode stage with a redefined pipeline (see Figure 2). The *fetch 2* stage turns out to be similar to the *drive* stage in Pentium 4, with most of the cycle time spent moving data across the die. The new *pick* stage is equivalent to the *scan* stage's deciding whether the instruction is for the MROM (not shown in Figure 1) or hardware decoders. The *decode 1* and *decode 2* stages are similar to the *align* stages in Athlon, but with additional information for the decoders extracted from the instructions. In the *pack* and pack/*decode* stages, it appears that information used to form the MacroOp bundle, and in *dispatch* stage the MacroOps, is targeted at the appropriate functional units and schedulers. The rest of the pipeline is similar to Athlon's, with the addition of a second clock cycle for the data cache. The final, unseen, architecture stage is the *commit* stage, where the architectural registers are updated to reflect the latest state information.

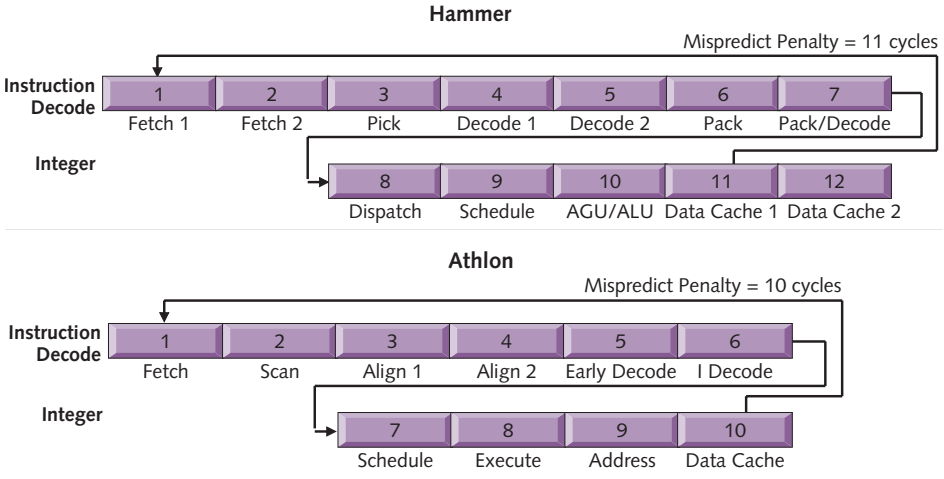

Figure 2. A comparison of the Athlon and Hammer pipelines shows similarities and some nomenclature changes.

Industrial-Strength Branch Prediction

Hammer beefs up the architecture in the branch-prediction and translation-lookaside buffers for large computing workloads. As Figure 3 shows, Hammer adds the two-clock-cycle branch target address calculator (BTAC) to accelerate address calculations. The branch selector selects between static predictor and the history table. A mispredicted branch penalty is 11 clock cycles.

The on-chip memory controller (see Figure 4) reduces memory latency and provides the processor with more bandwidth. It supports either an 8- or 16-byte interface, and either unbuffered or registered DIMMs. AMD plans to support PC1600, PC2100, and PC2700 DDR memory. The 16-byte interface also supports direct connection with up to eight registered DIMMs and supports chipkill ECC.

Glueless Multiprocessing Is Very Slick

Using up to three HyperTransport (HT) links, a Hammer processor can be arranged in a square configuration for a four-way server, a rectangle for a six-way, and a cross-connected rectangle for an eight-way (see Figure 5). Each

Hammer processor has up to three HyperTransport links, with one link that can be configured for multiprocessor communications or for I/O connectivity.

AMD will make HyperTransport transparent to software by building on the PCI software model. It will also send APIC (advanced priority interrupt controller) signaling in-band over HT.

Intel's Itanium processor continues to rely on a shared front-side bus to a chip set, much as the PC architecture does. Up to four processors share one bus and one channel to memory through the chip set. The Hammer architecture places a memory channel at each processor, and, as more processors are added, more memory bandwidth is available. The memory channel can be either 8 or 16 bytes wide. The 16-byte-wide configuration can support up to eight registered DIMMs. AMD's memory approach is very similar to that of Sun's UltraSPARC III and allows scaling as more processors are added.

At the time Hammer ships, Intel's McKinley will uses the Intel870 chip set. Beyond four processors, the 870 chip set uses a scalability port to connect to a switch chip, providing a coherent connection to other 870 chip sets. This hierarchical approach increases latency for memory accesses that cross the coherency switch and the 870 chip set. The Intel chip set also adds to the component count and memory latency by using an external chip (MRHD) to drive main memory.

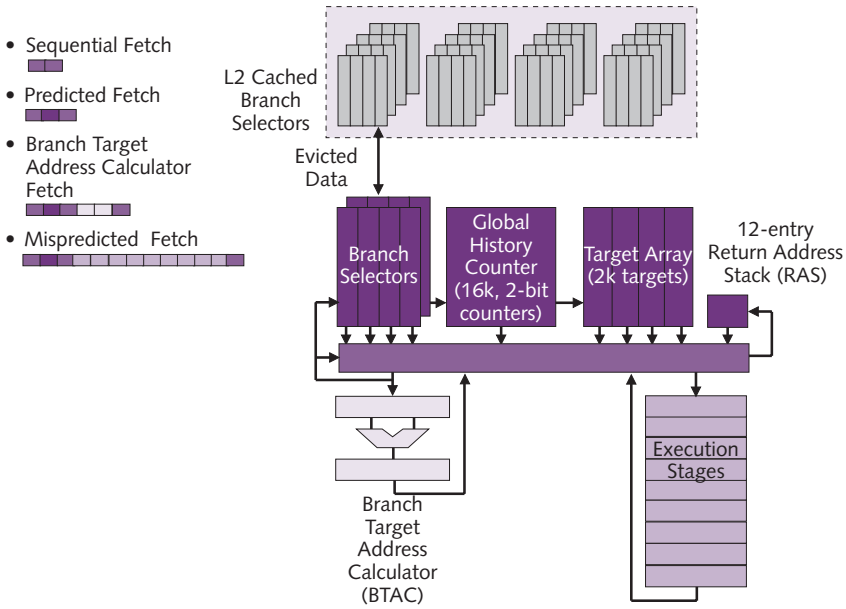

Figure 3. The Hammer branch-prediction logic has been significantly improved over Athlon's. The Hammer design goal is to perform better than Athlon on applications found in servers. One unique trick in Hammer is that instruction branch selectors are saved in the L2 cache when the cache line is evicted. AMD used the L2 cache bits assigned to ECC to store the instruction branch-prediction information. The instructions stored in the L2 cache are still protected by parity.

Robust Reliability Features for Servers

To be considered as a serious contender for enterprise server applications, AMD has installed significant reliability features in the Hammer processor. Hammer protects the L1 data cache and the L2 cache and cache tags with ECC. The on-chip DRAM controller also includes ECC protection with Chipkill ECC support. The on-chip and off-chip ECC-protected arrays include background hardware scrubbers to remove soft errors.

Remaining arrays are parity protected: L1 instruction cache, TLBs, tags. These arrays are generally read-only and can be recovered by performing a reload from memory or by recalculating the stored values.

Hammer also includes a machine-check architecture that reports failures and predictive failure results. Mechanisms for hardware/software error containment and recovery are essential parts of enterprise-class RAS (reliability, accessibility, and serviceability) feature sets. While all these capabilities will provide for a robust server system, they will also enhance the reliability desktop and mobile designs.

32 or 64 Bits: Hammer Gives You Both

AMD's 64-bit architecture, x86-64, represents an extension and evolution of the 32-bit architecture (see *MPR 9/4/00-01*, "AMD Drops 64-Bit Hammer on x86") rather than the clean-slate and divergent design approach Intel (along with partner HP) took in developing IA-64/EPIC (Itanium) architecture. AMD's Hammer should provide a significant

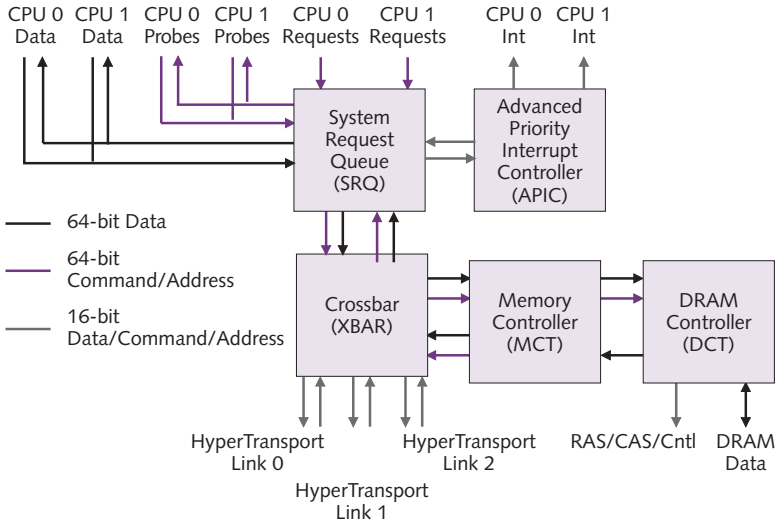

Figure 4. The on-chip memory controller runs at CPU core speeds, speeding data transfers and coherency checking within the chip.

advantage for those who plan to migrate to 64 bits slowly and who expect to run mostly 32-bit code.

At present, AMD lacks the wide OEM acceptance that Intel's Itanium enjoys. Intel already has commitments to ship Itanium systems from most major OEMs, including Compaq, HP, IBM, SGI, and Unisys. The only major server OEM that does not have an Itanium program is Sun Microsystems. Intel also has the financial resources to fund major software ports and thousands of "pilot systems." AMD has, by necessity, taken a grassroots approach, focusing on support from the more freewheeling Linux community.

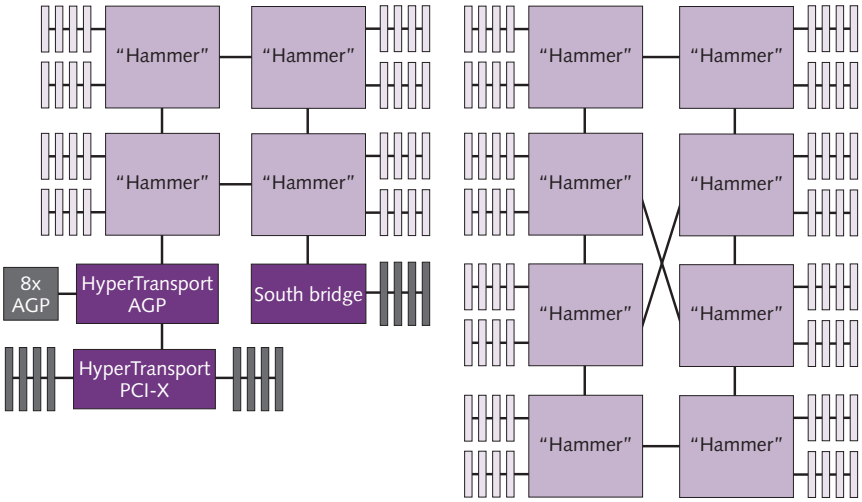

Figure 5. Using up to three HyperTransport links, Hammer can be configured using from one to eight processors without the need for an external crossbar switch. Each Hammer processor can support up to eight registered DIMMs with a 16-byte-wide memory bus. This figure shows four-way and eight-way configurations. In each configuration, a processor is no more than two processors away from any memory segment. The variance in memory access time is less than a DRAM page-conflict delay. Therefore, software does not have to view this as a NUMA (nonuniform memory architecture) design.

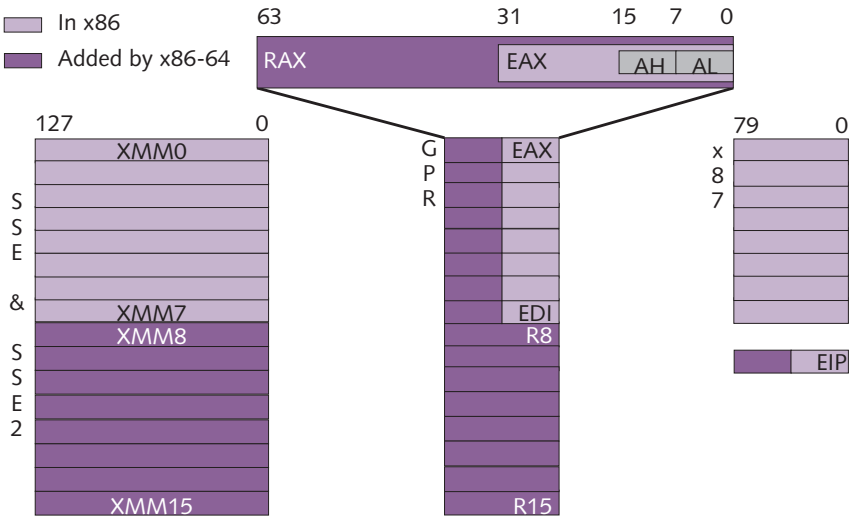

Figure 6. In the x86-64 programmer model, AMD has widened all the standard x86 registers from 32 to 64 bits. In addition, AMD added eight new general-purpose registers and doubled the number of SSE registers from eight to sixteen. The operating system must be modified to save and restore the wider general-xpurpose registers, program counter, and new registers during a context switch. AMD claims that implementing the compiler changes to support the new and wider registers is straightforward.

While x86-64 is an elegant improvement over the standard 32-bit instruction set, AMD may find itself in a chicken-and-egg dilemma. Major software companies, including Microsoft, will be reluctant to commit resources to x86-64 unless major system-OEM commitment and market-volume potential exist. Unfortunately, most OEMs won't commit resources to a new architecture unless there is wide software support. Sun Microsystems could make a very subversive move by supporting Solaris on Hammer after failing to support Itanium. But Sun's Solaris group is tightly focused on the SPARC microarchitecture support and would likely be reluctant to divert resources to an unproven architecture. Without large OEM support and lacking sufficient ported 64-bit application software, the x86-64 ISA could be relegated to being an unused extension to AMD's PC processors.

The irony of AMD's predicament is that the Hammer microarchitecture and system design is more consistent with recent developments in RISC-server microprocessor design, while Intel's Itanium represents a more radical approach, combining a PC-style shared system bus and a controversial VLIW-derived microarchitecture. The Hammer design offers an exciting alternative to Intel and needs only a major OEM design and Microsoft support to truly fulfill its ambitions. However, even if it falls short of its server aspirations, AMD will still probably be the first vendor to offer a 64-bit processor for mainstream desktop systems—and even for notebooks (in 2H03). ◇