

The AMD Athlon™ XP Processor with 512KB L2 Cache

**Technology and Performance Leadership for x86
Microprocessors**

Jack Huynh

**ADVANCED MICRO DEVICES, INC.
One AMD Place
Sunnyvale, CA 94088**

Introduction: Continuing Performance Leadership of x86 Microprocessors

Founded in 1969, AMD has shipped more than 190 million PC processors worldwide. AMD processors are the power behind desktop and mobile PCs, and a new generation of servers and workstations. Since its introduction in 1999, the award-winning AMD Athlon™ processor has been known as an industry leader, enabling one of the highest system performance levels in the PC market. Since its launch in October 2001, the AMD Athlon XP processor and computer systems based on the AMD Athlon XP processor have won more than 30 awards worldwide, including the 2002 PC World's World Class Award for Computer Product of the Year. In all, the AMD Athlon processor family and systems based on such processors have won more than 100 awards worldwide, including PC World's World Class Award for overall Product of the Year in 2000. The AMD Athlon processor family has provided industry-leading processing power to pave the road to new levels of end-user capability with PC application areas from gaming to productivity applications, including digital photo editing, digital video, image compression, video encoding, audio compression, 3D modeling and animation, commercial desktop publishing, and speech recognition.

Engineering and technology leadership are key to performance leadership. AMD's engineering and technology leadership specific to the seventh-generation AMD Athlon processor family includes driving innovations such as instruction set extensions aimed at 3D applications (3DNow!™ Professional technology) at the processor instruction level, DDR memory and HyperTransport™ technology at the platform level, and 0.13-micron process with copper interconnect at the process technology level. With the introduction of the new AMD Athlon XP processor with 512KB L2 cache and 0.13-micron process technology,

AMD continues its tradition of technology innovation in enabling one of the highest levels of delivered desktop PC performance. The discussion that follows provides an in-depth look at how the new AMD Athlon XP processor with 512KB L2 on 0.13-micron process technology increases the performance scalability of QuantiSpeed™ architecture. The differentiating features as well as the real-world application performance benefits of QuantiSpeed architecture will also be discussed.

Manufacturing Technology Leadership with Leading Edge 0.13-Micron Process Technology

The new AMD Athlon XP processor with 512KB L2 cache, code-named "Barton," is manufactured on state-of-the-art, 130 nanometer copper technology and is the newest member to the family of seventh-generation AMD Athlon processors designed to meet the computationally-intensive requirements of software applications running on high-performance desktop systems. The AMD Athlon XP processor based on 0.13-micron process technology increases the performance scalability provided by QuantiSpeed architecture by delivering higher clock speeds at low thermal power. The 0.13-micron process technology provides the thermal headroom necessary to scale frequency within the thermal limits of desktop platforms, thus maximizing its overall performance. The new AMD Athlon XP processor based on 0.13-micron process technology, like all AMD Athlon XP processors, is also pin compatible with AMD's established Socket A infrastructure.

With the increased frequency scalability brought about by 0.13-micron technology combined with QuantiSpeed architecture, AMD continues to deliver compelling solutions for high-performance computing, and delivers superb integer, floating point, and 3D

multimedia performance for applications running on x86-based platforms.

QuantiSpeed™ Architecture: A More Optimally Balanced x86 Microarchitecture for Real-World Application Performance

The microprocessor is a key component in determining the effectiveness of a computer system to execute specific tasks in the shortest amount of time. The amount of time required to complete specific software tasks is referred to as real-world application performance. Application performance is the function of two elements.

1. Clock speed of the processor, measured in megahertz, or gigahertz
2. The amount of work the processor can accomplish in a given clock cycle, measured in instructions per clock cycle (IPC)

$$\begin{aligned} \text{Application Performance} &= [\text{work completed per clock cycle}] \times [\text{clock speed}] \\ &= \text{IPC} \times \text{Frequency} \end{aligned}$$

Different approaches can be taken to optimize the processor for application performance. AMD has worked to maintain a more balanced microarchitecture with a shorter pipeline designed for much higher IPC than competitive PC processors available in the market. Although other competitive processors enable deeper pipelines with fewer gates per clock to drive frequency improvements, deeper pipelines alone may translate into less work per clock cycle. This reduced work per clock cycle or reduced IPC can only be offset by improvements in other areas, such as branch prediction and cache hit rates. Taken to the extreme, processor performance can actually be reduced by forcing frequency improvements at the expense of IPC improvements.

This key point can be illustrated in PC office applications which tend to be branch-code intensive resulting in lower performance for deeper pipelines that must be flushed with a much greater performance penalty. As reaffirmed in the *Desktop Performance and Optimization for Intel Pentium® 4 Processor* paper (<http://developer.intel.com/design/pentium4/papers/249438.htm>), “Integer and basic office productivity applications, such as Word and spreadsheet processing, tend to have many branches in the code, thus reducing overall IPC capabilities. As a result, the associated branch penalties and performance on these applications does not generally scale as well with frequency and are more resistant to improvements in micro-architectural means, such as deeper pipelines.”

The AMD Athlon XP processor with QuantiSpeed architecture implemented on 0.13-micron technology continues to exhibit the AMD Athlon processor family’s balanced combination of improving clock frequency without compromising the amount of work done per clock cycle and therefore the IPC. The end result is a processor design that produces a high IPC as well as high operating frequencies, the optimum combination to deliver one of the highest levels of PC performance in real-world application environments.

QuantiSpeed architecture consists of four key differentiating features that enhance the application performance of the AMD Athlon XP processor:

1. Nine-issue, superscalar, fully pipelined microarchitecture
2. Superscalar, fully pipelined floating point unit (FPU)
3. Hardware data prefetch
4. Enhanced Translation Look-aside Buffers (TLBs)

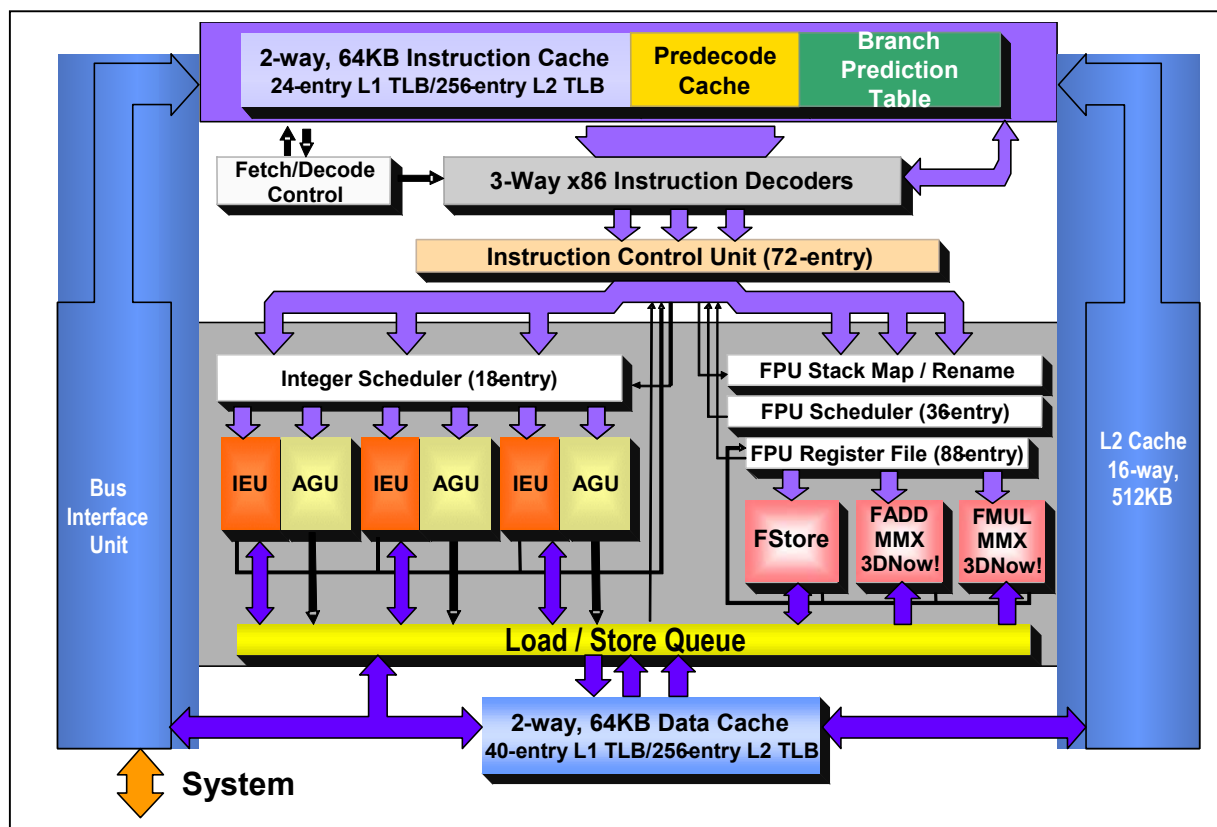


Figure 1: AMD Athlon™ XP Microarchitecture Block Diagram

QuantiSpeed™ Architecture: Nine-issue, Superscalar, Fully Pipelined Microarchitecture with High-Performance Cache Memory Architecture, and Three Full x86 Instruction Decoders

At the heart of QuantiSpeed architecture is a fully pipelined, nine-issue, superscalar processor core. The AMD Athlon XP processor provides a wider execution bandwidth of nine execution pipes when compared with competitive x86 processors with up to six execution pipes. The nine execution engines are comprised of three address calculation units, three integer units, and three floating-point units.

In order to supply such a highly superscalar microarchitecture, the AMD Athlon XP processor implements a very large on-chip cache architecture, particularly in the L1 caches closest to the core. The AMD Athlon XP processor's high-performance, on-chip cache architecture includes a dual-ported 128KB split-L1 cache with separate snoop ports, and an integrated full-speed, 16-way set-associative, 512KB L2 cache using a 72-bit (64-bit data + 8-bit ECC) interface. The AMD Athlon XP processor's large integrated full-speed L1 cache is comprised of two separate 64KB, two-way set-associative data and instruction caches which are much larger than the Pentium® 4 processor's L1 cache (128KB vs. 8KB + 12KB μ op). By featuring a larger L1 cache, applications running on the AMD Athlon XP processor perform exceptionally fast since more instruction and data information is local to the processor. Applications exploit the larger caches through benefiting from the increased support of instruction and data set locality. The data cache also has eight banks to provide maximum parallelism for running multiple applications. It supports concurrent accesses by two 64-bit loads or stores. The instruction cache contains predecode data to assist multiple, high-performance instruction decoders. Both instruction and data caches are dual-ported and

contain dedicated snoop ports designed to eliminate all system coherency traffic, common in systems with many devices, from interfering with application performance.

The AMD Athlon XP processor also includes an integrated, full-speed, 16-way set-associative, exclusive 512KB L2 cache. When the processor requests data, it first searches the data in its L1 cache. If the processor finds the data in its L1 cache, the result is what is known as a cache hit and the processor retrieves the data from the low latency L1 cache. If the processor can not retrieve the data from its L1 cache, the processor attempts to retrieve the data in its L2 cache and once again attempts to obtain a cache hit. In the event of a cache miss, the processor must then request this data from the slower system memory. With the additional 256KB L2 cache over previous AMD Athlon XP processors, the AMD Athlon XP processor with 512KB L2 cache increases the performance of applications such as high-end gaming and digital media by keeping more frequently accessed instructions and data close to the CPU. Higher set-associativity increases the hit rate by reducing data conflicts. This translates into more possible locations in which important data can reside in the L2 cache memory, instead of system memory. With an exclusive cache architecture, the contents of the L1 caches are not duplicated in the L2 cache. This enables 512KB of L2 cache and 128KB of L1 cache for a total usable storage space of 640KB.

The AMD Athlon XP processor cache architecture also supports error correction code (ECC) protection. With these cache architecture features, the AMD Athlon XP processor provides reliable high-performance computing.

When executing software, a processor begins by decoding the program's instructions and translating them into operations (or Ops) that the processor can execute. In order to continually feed the

execution engine with data, the AMD Athlon XP processor includes three x86 instruction decoders. Each decoder is capable of decoding three instructions per clock cycle. In comparison, the Pentium 4 is designed to decode only one instruction per clock cycle with the resource of only one x86 instruction decoder. Thus, the Pentium 4 has only one-third the maximum theoretical decode bandwidth of the AMD Athlon XP processor. The decode bandwidth of the AMD Athlon XP processor enables the processor to advantageously utilize the execution bandwidth capabilities of QuantiSpeed architecture, thereby improving IPC.

QuantiSpeed™ Architecture: Superscalar Fully Pipelined x86 Floating Point Unit (FPU)

The AMD Athlon XP processor offers one of the most powerful, architecturally advanced floating point units (FPU) delivered in an x86 microprocessor. The AMD Athlon XP processor's three issue, superscalar floating point capability is based on three pipelined, out-of-order floating point execution units, each with a one-cycle throughput. Using a data format and single-instruction multiple-data (SIMD) operations based on the MMX™ instruction model, the AMD Athlon XP processor can deliver as many as four 32-bit, single-precision floating point results per clock cycle.

FPU Microarchitecture

Three separate execution units in the AMD Athlon XP processor's floating point pipeline support x87 floating point instructions, MMX instructions, and 3DNow! Professional technology instructions. The three execution units are:

1. **Fstore** – This is the floating point load/store pipeline that handles FP loads, stores, and miscellaneous operations.
2. **Fadd** – This is the adder pipeline that contains 3DNow! Professional technology, add, MMX ALU/shifter, and FP add execution units.
3. **Fmul** – This is the multiplier pipeline that contains an MMX ALU, MMX multiplier, reciprocal unit, FP/3DNow! Professional technology instruction multiplier, and support for FDIV instructions.

In addition to its superscalar design, the AMD Athlon XP processor's FPU is super pipelined. This technique supports higher clock frequencies and enables the FPU to process complex floating point instructions more quickly and deliver high overall floating point instruction throughput. In comparison, the FPU of the Pentium 4 processor only offers two execution units, one for Fadd and Fmul and one for Fstore. Thus, as an example, AMD Athlon XP processor can do one floating point addition AND one multiplication per clock cycle, while the Pentium 4 processor can only do one multiplication OR one addition per clock cycle. The seventh-generation FPU of the AMD Athlon XP processor incorporates other features such as a 36-entry instruction scheduler and an 88-entry register file for independent, superscalar, out-of-order, speculative execution of floating point instructions. With three separate execution units,

the AMD Athlon XP processor's superscalar FPU can boost the performance of floating point-intensive applications varying from commercial applications such as 3D modeling to consumer applications such as high-end gaming for desktop systems.

3DNow!™ Professional Technology: FPU Innovation of the AMD Athlon™ XP Processor Core

The AMD Athlon XP processor with 3DNow! Professional technology adds 51 new instructions to the enhanced 3DNow! technology supported by the AMD Athlon processor. These 51 new instructions, along with the SIMD integer additions already included in enhanced 3DNow! technology, are compatible with Intel's SSE technology. Table 1 provides a breakout of the 3DNow! technology instruction set evolution.

Table 1: AMD processor support of SIMD instructions extensions to the x86 instruction set architecture.

AMD Processor:	AMD-K6®-2 Processor	AMD Athlon™ Processor	AMD Athlon XP Processor
3DNow!™ Technology Version Supported:	3DNow! Technology	Enhanced 3DNow! Technology	3DNow! Professional Technology
Description of Instructions Supported:	Original 3DNow! Technology Extensions	3DNow! Technology plus 19 MMX™ extensions (part of SSE) plus five DSP / communications extensions	Enhanced 3DNow! Technology plus 51 SSE extensions (completing SSE support)

3DNow! technology and SSE are largely complementary architectural enhancements. By implementing them in a variety of ways, software developers are able to determine how they can utilize the advanced architectural capabilities enabled by SIMD instruction set extensions. Examples of applications most able to benefit from the use

of these instruction set extensions include speed recognition, video encoding/decoding, and 3D graphics generations.

Many current software applications that are SIMD-optimized use different code paths to benefit from 3DNow! technology or SSE, depending on the processor architecture on which these applications are executed. AMD processor architectures preceding the AMD Athlon XP processor only supported 3DNow! or enhanced 3DNow! technology, which yielded the following three code path scenarios for developers:

1. Software optimized exclusively for AMD processor architectures with 3DNow! technology use their 3DNow! technology-optimized code path on AMD processors supporting 3DNow! technology.
2. Software optimized for both AMD processor architectures with 3DNow! technology, and other x86 industry processor architectures supporting SSE, use their 3DNow! technology-optimized code path on AMD processors supporting 3DNow! technology.
3. Software optimized exclusively for other x86 industry processor architectures supporting SSE use the non-optimized code path on AMD processor architectures.

With the advent of 3DNow! Professional technology, the AMD Athlon XP processor can seamlessly allow SIMD-optimized software in the third scenario above to recognize SSE support and run the optimized code path for increased performance. The recognition of SSE support in 3DNow! Professional technology is preformed automatically by PC software applications that use industry standard feature flags, provided in the CPUID instruction is designed to

automatically recognize SSE support and run the optimized code path. This means that with 3DNow! Professional technology's support for both 3DNow! and SSE technologies, the AMD Athlon XP processor is able to take advantage of the performance gains offered by SIMD-optimized software applications.

Not only is the AMD Athlon XP processor designed to benefit from existing software applications supporting 3DNow! and SSE technologies, but in the future, software developers will have the ability to utilize the strength of both 3DNow! and SSE technology when optimizing code paths for AMD processor architectures which support 3DNow! Professional technology. The AMD Athlon XP processor enables this advanced level of SIMD optimization by allowing 3DNow! and SSE instructions to be executed in the same code path.

QuantiSpeed™ Architecture: Hardware Data Prefetch

To further enhance processor IPC and therefore processor performance, the AMD Athlon XP processor also uses hardware data prefetch technology. This hardware data prefetch technology observes memory accesses, looks for regular access patterns, and speculatively fetches the cache line with the data into the processor's L2 data cache in advance of the actual data access, therefore reducing the average latency seen by the processor in accessing memory. In the past, data prefetch was supported through the instructions introduced in 3DNow! and SSE technologies. However, for the processor to take advantage of this capability, software applications had to be specifically optimized with the 3DNow! and SSE instructions. The AMD Athlon XP processor is designed to automatically optimize performance on existing software that has not previously been optimized using the hardware data prefetch instructions supported by 3DNow! Professional technology.

Benefits of the AMD Athlon XP processor's hardware data prefetching are observed more in high-end, data-intensive applications that access larger arrays of data. Performance also benefits by not occupying processor instruction execution bandwidth required by software prefetching instructions. The optimization is most effective when coupled with high-bandwidth system memory transfer capability, now available to the processor by platforms such as those optimized to support DDR memory.

QuantiSpeed™ Architecture: Exclusive and Speculative Translation Look-aside Buffers (TLBs)

The AMD Athlon XP processor features advanced, two-level Translation Look-aside Buffer (TLB) structures for both instruction and data address translation. The AMD Athlon XP processor's Level 1 (L1) Instruction TLB (I-TLB) holds 24 entries, the L1 Data TLB (D-TLB) holds 40 entries, and the L2 I-TLB and D-TLB each hold 256 entries.

To reduce the incidence of TLB entry conflicts, The L1 and L2 TLB structures adopt an exclusive architecture design. With an exclusive TLB architecture, the L1 TLBs can contain entries that are not duplicated in the L2 TLBs, enabling the combination of L1 TLB and L2 TLB sizes for a larger total available entry space on both the instruction and data TLBs. By reducing the number of conflicts caused by holding more TLB entries within the processor, performance increases on high-end, data-intensive applications that encounter instruction sequences that no longer have to wait for TLB entries to be reloaded during execution.

The TLB structures of the AMD Athlon XP processor also have the ability to enter data TLB misses in the TLBs speculatively. The AMD Athlon XP processor allows TLB entries to be written speculatively

before the first instruction is completed, while preserving proper instruction execution ordering which removes the serialization effect and results in improved system performance.

Conclusion: Technology and Performance Leadership of x86 Microprocessors

With these key differentiating features of the new AMD Athlon XP processor with QuantiSpeed architecture...

- **0.13-micron process technology** – Provides further thermal headroom necessary to scale frequency within the thermal limits of desktop platforms of the AMD Athlon XP processor, thus maximizing overall performance
- **512KB L2 cache** – Increases the performance of applications such as high-end gaming and digital media by keeping more frequently accessed instructions and data close to the processor.

QuantiSpeed Architecture:

- **Nine-issue, superscalar, fully pipelined microarchitecture** – Provides a wide executing bandwidth to improve overall productivity
- **Superscalar, fully pipelined FPU** – Increasing performance of floating point-intensive applications while offering 3DNow! Professional technology support
- **Hardware data prefetch** – Increasing performance on high-end software applications using high-bandwidth system capability, especially with DDR memory
- **TLB enhancements** – Increasing performance of high-end, data-intensive applications

AMD continues to accelerate technology innovations while meeting the computationally intensive requirements of software applications including:

- **3D applications** – 3D modeling, high-end gaming, etc.
- **Multimedia/digital content creation applications** – photo and video editing, video encoding and decoding, image compression, soft DVD, MP3 encoding and decoding, etc.
- **High-end applications** – digital publishing, speech recognition, CAM, digital prototyping, etc.

With industry leading performance across these and a number of other applications, the AMD Athlon XP processor implemented on 0.13-micron technology continues to increase the performance scalability provided by QuantiSpeed architecture by delivering higher clock speeds and industry-leading processor performance at lower thermal power. The AMD Athlon XP processor continues in the tradition of the AMD Athlon processor family in providing compelling levels of delivered system performance for today's and tomorrow's applications.

AMD Overview

AMD is a global supplier of integrated circuits for the personal and networked computer and communications markets with manufacturing facilities in the United States, Europe, Japan, and Asia. AMD, a Fortune 500 and Standard & Poor's 500 company, produces microprocessors, Flash memory devices, and support circuitry for communications and networking applications. Founded in 1969 and based in Sunnyvale, California, AMD had revenues of approximately \$2.7 billion in 2002. (NYSE:AMD).

© 2002 and 2003 Advanced Micro Devices, Inc. All rights reserved.

AMD, the AMD Arrow logo, AMD Athlon, AMD-K6 and combinations thereof, and 3DNow! and QuantiSpeed are trademarks and AMD-K6 is a registered trademark of Advanced Micro Devices, Inc. Pentium is a registered trademark and MMX is a trademark of Intel Corporation in the United States and/or other jurisdictions. HyperTransport is a trademark of the HyperTransport Technology Consortium in the U.S. and other jurisdictions. Other product and company names used in this publication are for identification purposes only and may be trademarks of their respective companies.