Excerpt from MindShare’s Upcoming Book:

**AMD K8 Processor Architecture**

Joe Winkles
MindShare, Inc.
joe@mindshare.com

November 2005

For training on this topic, visit [www.mindshare.com](http://www.mindshare.com) or call 1-800-633-1440
1

K8 Processors: Breaking Tradition

Notice

This material is copyrighted and is not to be reproduced without permission from MindShare, Inc. It is offered as a courtesy to MindShare subscribers.

Copyright © 2005 by MindShare, Inc. All rights reserved.

AMD, AMD Opteron, and combinations thereof are trademarks of Advanced Micro Devices, Inc.

Introduction

The following is an excerpt from the upcoming MindShare textbook on AMD K8 Processor Architecture. MindShare currently offers a course on AMD based processors which can be found at www.mindshare.com.

The K8 Microarchitecture

The terms “K8” and “Hammer” are AMD’s internal names for the processor microarchitecture that will be described in detail throughout this book. AMD uses the K8 microarchitecture for several lines of processors such as:

- AMD Opteron™
- AMD Athlon™ 64
- AMD Athlon™ 64 FX
- AMD Turion™
- AMD Sempron™ (a subset of this processor line uses the K8 microarchitecture, the early Semprons were based on the K7 microarchitecture)
The K8 Architecture

All of these processors use the same basic internal microarchitecture however they are targeting different markets and thus have different feature sets. A brief description of each processor line’s characteristics can be found later in the chapter. In this book, whenever the term K8 or Hammer is used, it applies to all processors based on this microarchitecture. If a specific processor is named, like AMD Opteron™, that reference only applies to that processor line. Figure 1-1 on page 3 shows a K8 based processor.

Another term that is often thrown around when discussing AMD processors is AMD64. AMD64 is a 64-bit instruction set architecture designed by AMD to add 64-bit extensions to the traditional 32-bit x86 architecture. It used to be called x86-64 during the development phase and was later changed to its current name, AMD64. This 64-bit architecture was widely adopted by the industry because it is backwards compatible with existing x86 software. Intel also came up with a version of AMD64 which they call EM64T (Extended Memory 64-bit Technology). EM64T is almost identical to AMD64, however there are a few minor differences which will be discussed later in this book.

Unfortunately the terms AMD64 and K8 are often used interchangeably which is not accurate. If a processor is compatible with the AMD64 architecture then it will support the legacy x86 instruction set as well as the 64-bit extensions defined in the AMD64 Programmer’s Manual (a five volume set). However this does not mean that the microarchitecture of the processor is a K8. On the flip side, a processor based on the K8 microarchitecture has the ability to support AMD64, however that ability may not be enabled. For example, some of the AMD Sempron™ processors (e.g. 3100+) are built with the K8 microarchitecture but are not AMD64 processors because the 64-bit extensions are disabled.
Chapter 1: K8 Processors: Breaking Tradition

Breaking the Mold

K8 based processors have several features which are new to x86 processors, such as 64-bit extensions, an integrated Northbridge, “glueless” multiprocessing capabilities, and a multi-core design. These characteristics differentiate K8 based processors from the traditional x86 processor design.

64-bit Extensions

The original x86 processors (8086, 80186, and 80286) were 16-bit processors, meaning they could operate on 16-bits of data at a time. These early processors, despite being 16-bit machines, could generate a 20-bit address, allowing them to target up to 1MB of memory. The 386 was then released in 1985 which extended the 16-bit architecture identified above to 32-bits. Extending the existing 16-bit architecture to 32-bits allowed all the software written for the 16-bit environment to function on the new 32-bit machine along with new 32-bit software.
The K8 Architecture

This extension maintained backwards compatibility which often plays a key role in the adoption of new technologies. All x86 processors since then have been 32-bit machines until the K8. The K8 microarchitecture was designed to support the AMD64 technology, which is 64-bit extensions on top of x86's existing 32-bit architecture.

The motivation for extending the x86 architecture to 64-bits was predominantly driven by large applications that needed to address significant amounts (more than 4GB) of virtual and physical memory. The traditional solution to this problem was to transition to an entirely different architecture which did support a 64-bit environment. However these alternative architectures were often extremely expensive (from both a hardware and software point of view) and were not as widely understood as the x86 architecture. In addition to that, these new architectures would either run x86 applications in an “emulation mode” (an instruction translator) which would have very poor performance for obvious reasons, or would not be able to run them at all. This was a major downside due to the fact that x86 applications comprise the largest installed software base in the world.

The solution to this problem was to extend the x86 architecture to 64-bits, which is what AMD64 has done. This provides an environment that can run both 32-bit and 64-bit software natively. In fact, AMD64 compatible processors can run 64-bit and 32-bit applications side-by-side under a 64-bit OS which allows customers to migrate to 64-bit applications at their own pace.

The AMD64 architecture actually incorporates a lot more than just increasing the data and address paths to 64-bits. A detailed discussion of all aspects of the AMD64 architecture can be found in subsequent chapters.

Integrated Northbridge

Another unique feature about the K8 microarchitecture is its integrated Northbridge functionality. In x86 systems, the Northbridge is the logic which serves as the processor’s interface to system memory and the I/O world. This logic has traditionally resided in a chip physically separated from the processor. A classic x86 based system is shown in Figure 1-2 on page 5, and a system built around a K8 processor can be seen in Figure 1-3 on page 6.

AMD decided to pull the Northbridge logic into the K8 processor design for several reasons. One of the integrated logic blocks was the memory controller. By having the memory controller directly on the processor itself, it allowed for lower latency memory accesses from the processor than in a solution where the
Chapter 1: K8 Processors: Breaking Tradition

memory controller is on an entirely separate chip. The on-chip memory controller is designed to run at the same speeds as the processor core, but is on a separate clock grid allowing the processor to go into a low power state while not affecting the latency or bandwidth of memory accesses from other devices.

Figure 1-2: Traditional x86 Single Processor System
The K8 Architecture

AMD had another thing on their mind when the decision to pull the Northbridge on-chip was made, and that was multiprocessing capabilities. AMD recognized that the existing multiprocessor solutions had limitations that needed to be overcome. As is shown in Figure 1-4 on page 7, both of the existing multiprocessor solutions required the Northbridge’s support in order to function.

The first solution shown, Figure 1-4a, was used for the AMD Athlon™ MP which was a processor based on the K7 microarchitecture (the predecessor to the K8). In this solution each processor is connected to the Northbridge with its own dedicated Front Side Bus (FSB). The FSB in this case was Alpha’s EV6 bus, a 64-bit wide point-to-point parallel bus that transmits data on both edges of the
Chapter 1: K8 Processors: Breaking Tradition

bus clock. This solution provides a significant amount of FSB bandwidth for each processor, but going beyond a 2-way system would be expensive due to the very high pin count Northbridge that would be required.

The second solution shown, Figure 1-4b, is the current solution for all of Intel’s x86 multiprocessor systems (e.g. Pentium® 4 Xeon). In this solution, all processors in the system share the same FSB which has one connection to the Northbridge. The FSB here is Intel’s proprietary version of the industry standard GTL (Gunning Transceiver Logic) specification, which Intel calls the AGTL+ (Assisted GTL+). Intel’s FSB allows up to 8 devices to reside on the bus. Since multiple processors are sharing the same FSB, the bandwidth of the bus is obviously divided among all the processors. This can be a bottleneck in terms of performance for the system because each processor must retrieve their instructions and the majority of their data from system memory. One way to alleviate this divided bandwidth is to increase the speed of the bus, however because of the loading limitations present in a multi-drop bus, the speed of the bus cannot be ramped very high (in comparison to a point-to-point bus).

Figure 1-4: Traditional x86 Multi-Processor Systems
The K8 Architecture

In addition to the limitations previously discussed, both of these solutions suffer from another potential bottleneck and that is the limited bandwidth of the memory bus. Each memory request targeting system memory must compete for the bandwidth of the memory bus. While processors are typically the largest consumer of memory bandwidth they are not the only devices that use a significant amount of bandwidth. Other devices that may generate a lot of memory requests and tax the memory bus are graphics devices and some I/O devices such as a gigabit ethernet card. In both solutions described in Figure 1-4 on page 7, each processor that is added significantly reduces the memory bandwidth allocated to each device. The K8 multiprocessor solution, as shown in Figure 1-5 on page 8, alleviates this problem because each new processor added comes with its own memory controller. So instead of decreasing the amount of memory bandwidth available for each device, adding processors actually increases the total available memory bandwidth! However, by physically distributing physical memory, another set of issues arise dealing with memory mapping, maintaining cache coherency, and optimizing a non-uniform memory access (NUMA) based system.

Figure 1-5: 2-Way K8 Based System
Chapter 1: K8 Processors: Breaking Tradition

AMD addressed these issues by taking a different approach than the two solutions previously discussed for multiprocessor systems. AMD developed a high-bandwidth, serial point-to-point bus technology called HyperTransport (HT). Each K8 based processor has a potential for having up to three viable HyperTransport links. Each HT link can connect either to an I/O chip or directly to another processor. The transfer protocols are slightly different based on what type of device it is connected to. For example, when connecting processors together, the HT link must support a protocol to maintain cache coherency (discussed in detail later in the book). This version of HT is proprietary to AMD and is known as coherent HyperTransport (cHT). When a link connects to an I/O device, the protocol is simpler because there is no need for the cache coherency semantics. This version of HT is a public standard and is managed by the HyperTransport Consortium (www.hypertransport.org). HT is very scalable in terms of link width and speed providing a flexible and configurable environment for x86 systems.

Glueless Multiprocessing

A subtle point that may have been missed in the last paragraph is that in K8 based systems, processors can be connected directly to each other with cHT in order to create a multiprocessor environment. This eliminates the need for having a separate chip, like an external Northbridge, to enable multiprocessing capabilities. This is known as “glueless” multiprocessing. The K8 microarchitecture can currently scale up to an eight processor system using the glueless multiprocessing capability. Knowing that each K8 can have up to three HT links, some viable multiprocessor system topologies are shown in Figure 1-6 on page 10.

The author would like to point out that K8 based systems are not limited to a maximum of eight processors. The eight processor limit only applies to the “glueless” capability. Due to the flexibility of HyperTransport, scaling a system beyond eight processors can be achieved simply by having an external chip that acts as a coherent HyperTransport switch between two (or more) clusters of K8 processors. There are solutions currently available which provide this capability. This topic will be discussed in more detail later in the book.
The K8 Architecture

The K8 microarchitecture was designed from the beginning to be a multi-core processor. The K8 dual-core microarchitecture is shown in Figure 1-7 on page 11. Multi-core processors can improve the performance of a system running multiple processor-intensive applications and/or multi-threaded applications. Multi-core processors can also benefit markets where board space or rack space is precious but performance is critical. For example, in the rack mounted server market, having a 1U or 2U server which holds 4 Dual-Core AMD Opteron™ processor.
Chapter 1: K8 Processors: Breaking Tradition

processors (an effective 8-way system) can provide a significant increase in processing power in comparison to 4 single-core processors (4-way system) without increasing the amount of rack space used.

In this book the term processor will be used to refer to the entire chip regardless of the number of cores which reside in that chip. The term core or processor core will be used to indicate the actual processing unit inside the processor.

*Figure 1-7: Dual-Core K8 Microarchitecture*

**Processors based on the K8 Microarchitecture**

The reader should keep in mind that MindShare’s book series often deals with rapidly evolving technologies. This being the case, it should be recognized that this book is a “snapshot” of the state of the K8 microarchitecture at the time this book was completed.

The following sections briefly describe the differences between AMD’s current processor lines that are based on the K8 microarchitecture.
The K8 Architecture

**AMD Opteron™**

The AMD Opteron™ processor is targeted at the server and workstation market segments. The AMD Opteron™ processor line is the only line of K8 based processors that support coherent HyperTransport which provides the glueless multiprocessing capabilities.

The AMD Opteron™ line of processors is divided into three different series based on their multiprocessing capabilities. In every AMD Opteron™ processor, all three HT links are viable, however, the number of those links that can use the coherency protocol (cHT) may be limited. Table 1-1 on page 12 describes the differences between the three series.

*Table 1-1: AMD Opteron™ Processor Series Characteristics*

<table>
<thead>
<tr>
<th>Series</th>
<th>Processor Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>Can only be used for single-processor systems. The three HT links can only be connected to I/O devices and cannot be connected to other processors because they do not support the coherency protocol.</td>
</tr>
<tr>
<td>200</td>
<td>One and only one of the three HT links may be connected to another processor. The other two HT links are only allowed to connect to I/O devices. This processor can only be used in a 1 or 2-processor system.</td>
</tr>
<tr>
<td>800</td>
<td>All three of the HT links support the coherency protocol so all three HT links can be used to connect to other processors providing scalability for more than a 2-processor system. However it is not a requirement that all HT links be connected to other processors. Any number of HT links could be connected to I/O device(s) in which case the coherency protocol for that link or set of links would never be used and they would function as industry standard HT links.</td>
</tr>
</tbody>
</table>

The lower two digits of an Opteron’s model number reflect the relative performance of the processor. For example, an AMD Opteron™ 148 has better performance than an AMD Opteron™ 144. However, an AMD Opteron™ 248 and an AMD Opteron™ 148 have the same amount of computing power, the difference between the two processors resides in the capabilities of their HT links. One of the HT links on the AMD Opteron™ 248 can use the coherency protocol in order to connect to another processor, creating a multi-processor system. None of the HT links in the AMD Opteron™ 148 can use the coherency protocol, so it will be a single processor system.
Chapter 1: K8 Processors: Breaking Tradition

The dual-core AMD Opteron™ processors have model numbers that start at x6x and go up (x7x, x8x, etc.) based on performance.

Each AMD Opteron™ processor has a 128-bit wide interface to system memory providing a memory bandwidth of 6.4GB/s when using PC3200 DDR SDRAM.

AMD Athlon™ 64

The AMD Athlon™ 64 processor is targeted at the desktop market. Several variations of this processor line have emerged, two of them being the AMD Athlon™ 64 FX which is targeted at gaming enthusiasts, and the AMD Athlon™ 64 X2 which is a dual-core AMD Athlon™ 64.

There are basically three knobs which AMD can turn to tweak the overall performance of its AMD Athlon™ 64 processors:

- Processor Frequency
- Cache Size
- Width of Memory Controller (64 or 128-bit)

That overall performance is represented in its model number. The model numbers for these processors take the form of a number typically followed by a plus sign (e.g. 3600+). The higher the number the better the performance. The plus symbol is simply there for effect.

AMD Turion™ 64

The AMD Turion™ 64 processor is a low-power K8 processor targeted at the mobile market. This processor line differs from the AMD Athlon™ 64 processors because it was designed for the low-power market. For example, it implements a new lower power state, C3, and is built with transistors that consume less power than its desktop sibling.

AMD Sempron™

The AMD Sempron™ processor is targeted at the low-budget desktop market. This processor line is interesting because AMD switched microarchitectures from the K7 microarchitecture to the K8 in mid-stream. All AMD Sempron™ processors with a model number of 3100+ or higher are based on the K8
The K8 Architecture

microarchitecture and AMD Sempron™ processors with a model number smaller than 3100+ are K7 processors.

The K8 based AMD Sempron™ processors are simply AMD Athlon™ 64’s artificially limited to target the value computing desktop space. Currently, one of the features of the K8 based AMD Sempron™ processors that is artificially limited is its AMD64 support. In other words, these processors are only 32-bit machines because the 64-bit extensions are not enabled.