



Hammer Time!

cover feature



After much hype and anticipation, AMD is finally launching the first of its x86-64 CPUs. Previously code-named “Sledgehammer”, Opteron is now out in the wild. Paul Hudson shows us what it is, how it works, and how it aims to change the world...

AMD announced the x86-64 specifications all the way back in August 2000, and finally, in April 2003, the first x86-64 CPU was released. *Opteron*, previously known only as *Sledgehammer*, is the name for AMD's 8th-generation enterprise-class processor for servers and high-end workstations. Between the original announcement and the product release, x86-64 was renamed AMD64, and the two are interchangeable – although the official term is AMD64. The unique selling point of AMD64 is that it does what its name suggests: 64-bit operation combined with the same x86 technology that companies have relied on for so many years now. Not only does this allow adopters to preserve their investment in 32-bit technology, but it also allows them to ease the transition to the more advanced 64-bit model – one computer, powered by Opteron, can handle both 32-bit and 64-bit applications seamlessly.

If you've been watching the 64-bit market, you'll have seen both the Itanium and the Itanium 2 being

launched by Intel and HP since the AMD64 specification was first aired, and you might be forgiven for wondering when AMD will get its product out the door. Luckily for AMD, both the Itanium (*Merced*) and its successor the Itanium 2 (*McKinley*) – collectively known as the Itanium Processor Family, or IPF – have both received a cool reception from the industry, with the processor family even earning the title “Itanic” from several wry analysts. The problem with Itanium is that it wholly replaces x86-32 with a brand new architecture that bears little to no similarity to x86-32. While it can run in 32-bit compatibility mode, it requires special emulation to do so, and so performance suffers. Read the box *The Rise And Fall Of Itanium* on page 53 for more information.

The Opteron was officially launched on April 22nd with much fanfare at a special event in New York, which we attended (see box, *The Launch* on page 55). Now in the hands of consumers, it's time to prove its worth – will it have trouble like Itanium, or is Opteron the future?

Why should companies switch to Opteron? Before you can truly understand the benefits of AMD64, it's important to first understand how we got to the current state of play.

x86's heritage

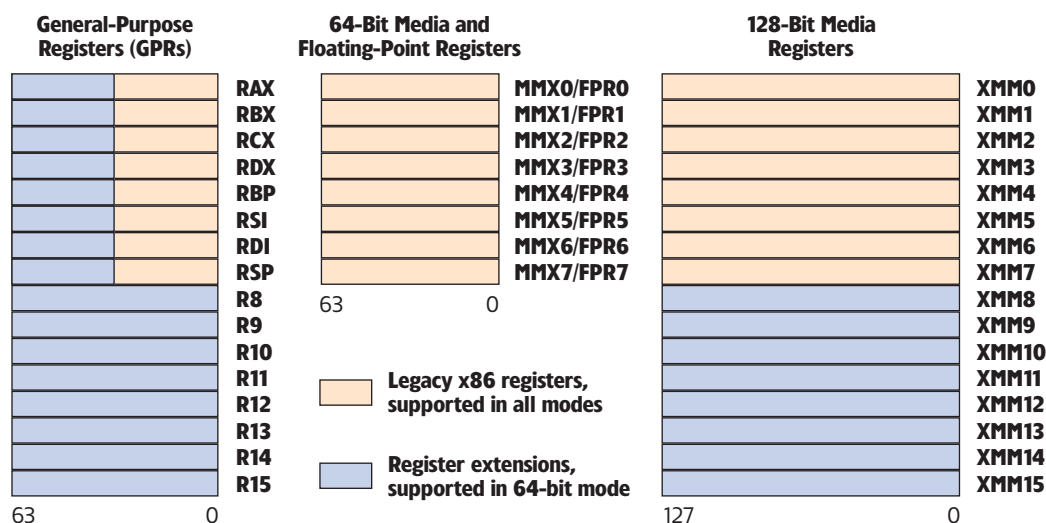
Unlike many other processor architectures, x86 has a very clearly defined history. Early on there was the 4004, then the 8008, then the 8080, and finally 8086. 8086 was Intel's first 16-bit CPU, capable of addressing up to 1MB of memory. The 286 followed (with a relatively uninteresting 80186 in between), introducing protected mode memory access, allowing lots more memory to be accessible to new programs without breaking existing ones. The 386 followed, introducing 32-bit flat addressing, which promised to do away with memory segmentation altogether. AMD and Cyrix started competing with Intel by manufacturing Intel-clone 386 CPUs.

The 486 was a key release, making several important changes to the architecture. Firstly, the 486 was the first to include level 1 cache on the chip, reducing the number of times main memory needed to be accessed. Furthermore, the later models had a ubiquitous maths co-processor (x87) built in, adding 80-bit floating point mathematics capabilities to the CPU which, at the time, made *Doom* run considerably faster!

The Pentium CPU, so named as to allow Intel to trademark the brand, made few changes to the x86 architecture. However, it did introduce MMX in January 1997, which was the first major change to the x86 instruction set in a decade and added 57 new instructions designed to improve multimedia operations. However, at this point, both AMD and Cyrix were actually creating better CPUs than Intel. For example, AMD's K5 CPU was capable of executing six out-of-order instructions at once, allowed register renaming, and also speculative execution – three things not seen in Intel's chip until the Pentium Pro was released.

The Pentium Pro was the first CPU to be specifically designed for 32-bit code. The 386, 486, and Pentium CPUs could run 32-bit code, but they were designed to run both 16-bit and 32-bit code in equal amounts. The Pentium Pro also included an integrated level 2 cache, which greatly increased both the speed and the cost of the CPU.

Owing to a mixture of its price and the fact that it performed 16-bit operations very poorly at a time when Windows 95 still relied heavily on 16-bit code, the Pentium Pro didn't sell too well for Intel, and it took the company 'til May 1997 to produce the Klamath CPU, known on release as the Pentium II. The PII corrected most



As seen above, the AMD64 architecture extends all GPRs to 64 bits and also increases the amount of GPRs to 16.

of the problems with the Pentium Pro, and was generally a faster and better CPU than the Pentium MMX. AMD's clone of the PII was first the K6, which again outperformed the PII clock-for-clock, and also later evolved into the K6-2. This new CPU included AMD's new 3DNow! extensions to the x86 architecture, which offered another boost for multimedia applications, and was much more popular than the MMX extensions.

With the Katmai CPU, released as Pentium III, Intel added 70 more multimedia instructions to x86, although left much of the CPU entirely unchanged from the Pentium II. This turned out to be a big mistake when AMD released the Athlon CPU, a much-enhanced version of its K6-3 CPU, and managed to trounce Intel's CPUs quite drastically. In order to turn around its falling fortunes, Intel worked hard to entirely redesign their x86 platform, and in the end released 1.4 and 1.5GHz versions of the Pentium 4. The Pentium 4 is, architecturally, light-years ahead of the Pentium III, however in order to take advantage of its new architecture, significant optimisation is required – optimisation that, at the time of launch, was just not available. As a result, the P4 took quite a bit of hammering from tech sites for the first six to nine months of its existence. However, the P4 does have two advantages in its favour: it can ramp up clock speeds very quickly and to very high levels, and it also came with SSE2: 144 new instructions for x86 to further enhance multimedia applications.

This is where the tale of 32-bit processing now ends. Intel's 64-bit Itanium CPU is not part of the x86 family, which left a gap open for another company to squeeze into if they were capable of producing an x86-compatible 64-bit CPU. In August 2000, this materialised in the form of AMD64, designed and created by Intel's leading x86-32 rival: AMD.

Hammer Time!

The Pentium 4 CPU, despite being much faster than the 386, still bears a great resemblance to its ancestor. For example, the main set of x86 instructions (that is, not the MMX or SSE/SSE2 [collectively called XMM] additions) remain the same. They both have the same number of registers, and they both use the same style of

maths co-processor. While other architectures have been more willing to break backwards compatibility, x86-32 has mostly seen a history of tweaks and speed hikes.

When Intel came to design its own 64-bit CPU, it threw the x86 architecture out and started afresh. Unfortunately that has stifled adoption of the Itanium architecture because most companies have very large investments in the x86 CPU architecture in both the hardware and also software that runs on it.

AMD took a different option, and decided to create a CPU *evolution* as opposed to Intel's *revolution*. The Hammer family of CPUs, code-named *Sledgehammer* for servers and *Clawhammer* for workstations, were designed around a new AMD64



“Linux Enterprise Server for AMD64 enables customers to combine the stability and security of Linux with the performance enhancements available only through the 64-bit architecture”

RICHARD SEIBT, CEO
SUSE LINUX AG



Operating Mode		Operating System Required	Application Recompile Required	Defaults		Register Extensions	Typical
				Address Size (bits)	Operand Size (bits)		GPR Width (bits)
Long Mode	64-Bit Mode	New 64-bit OS	yes	64	32	yes	64
	Compatibility Mode		no	32		no	32
				16	16		16
Legacy Mode	Protected Mode	Legacy 32-bit OS	no	32	32	no	32
				16	16		
	Virtual-8086 Mode			16	16		16

The Opteron's variety of modes support full backwards compatibility, but 64-bit Long mode is the preferable one.

Opteron

« architecture, which retains complete compatibility with 32-bit code whilst adding the option of 64-bit computing. Moreover, it performs 64-bit and 32-bit operations natively – no emulation is required.

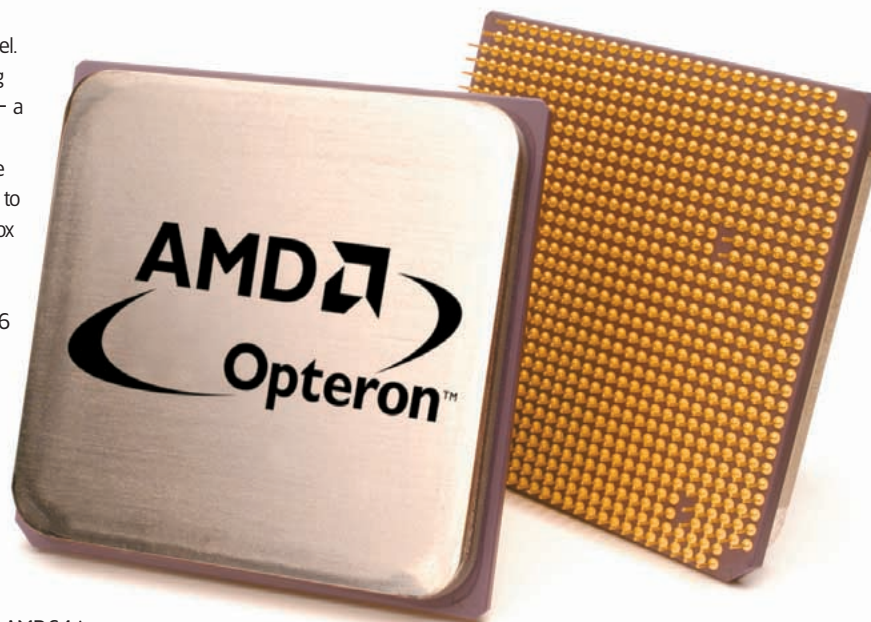
So, how does AMD64 handle both 32-bit code and 64-bit code? The architecture specification includes two modes of operation: “Long Mode”, and “Legacy Mode”. Legacy mode AMD64 CPUs operate as existing x86-32 chips do – they support 16-bit and 32-bit applications and also support 16-bit and 32-bit operating systems. Processors that implement AMD64, of which the Opteron is the first, boot into Legacy Mode by default, and therefore are entirely compatible with existing software – by default, the Opteron looks like a 32-bit CPU to software.

Long Mode, though, is where the magic happens, and can be split into two submodes: 64-bit Mode and Compatibility Mode. Both of these submodes require a 64-bit operating system to run; the differences lie in the applications: 64-bit Mode only supports 64-bit applications running on a 64-bit operating system, whereas Compatibility Mode supports 32-bit applications running on a 64-bit operating system.

Compatibility mode means that users can switch to a 64-bit operating system immediately whilst continuing to use their existing software – the operating system will perform better, which will yield some performance

boosts at the application level. However, it is of course Long Mode that is the ideal goal – a pure 64-bit processing environment. Apart from the inherent advantages offered to 64-bit programs (see the box *Why 64-bit Computing Is Needed*), AMD took the opportunity to revise the x86 platform to help bring it up-to-date a little. However, because of the need to retain backwards compatibility when running 32-bit programs, these new features are only available in 64-bit mode.

One of the most highly anticipated revision made in AMD64 is the addition of new registers. AMD has doubled the number of general purpose registers (GPRs) to sixteen, and also doubled the number of XMM registers to sixteen also. These 32 registers are, of course, lengthened to 64 bits each. Having more registers is key to improving performance, particularly when there were so few to begin with. Registers store small amounts of information directly in the CPU and provides instant access to that data – it doesn't need to be moved from cache or main memory. More registers mean more data can be held on the CPU, which means less shuffling around is required to get results of calculations. However, some are already saying that 16 GPRs aren't enough –



RISC architectures often have registers numbering in the hundreds, thereby allowing complete programming loop unrolling to take place inside the registers. For the time being, though, having just 16 registers is likely to make quite a drastic speed improvement by itself once programs are recompiled to use them – almost certainly a minimum of 10%.

Another key feature available exclusively in 64-bit Mode is instruction pointer-relative data addressing, which allows programs to address locations anywhere in the 64-bit address space relative to the instruction pointer.

These new features are only available in 64-bit Mode, where applications have specifically been compiled to make use of the functionality. However, fortunately it generally does just take a recompile to get the extra functionality offered by 64-bit Mode.

With AMD64, AMD took the opportunity to incorporate both the XMM extensions, SSE and SSE2. As mentioned already, there are twice as many XMM registers in the AMD64 specification as in Intel's own chips, which will mean a further increase in performance. Of course, the biggest XMM performance increase will be the XMM instructions themselves: thanks to Intel pushing so hard to have them adopted, these are now used extensively in multimedia programs such as *3ds max* and *Quake 3*. As a result, just by incorporating the XMM

instructions, the Opteron is going to be able to leap ahead in performance when compared to the Athlon chip.

When operating in 64-bit Mode, the Opteron switches to using a new flat memory segmentation model. Segmented memory was previously used as a method to allow operating systems to isolate programs from each other, thereby increasing reliability. However, because most modern operating systems do this all in software, there is a lot of wasted space in the x86-32 architecture. This has been stripped out for 64-bit Mode Operons, which allows new 64-bit operating systems to have much simpler code to handle memory management.

So, not only does AMD64 allow much larger data to be worked on, but it also adds XMM support, cleans up a lot of very old legacy operations that no longer have a place in today's environment, and does so in such a way as to provide complete compatibility with existing 32-bit software.

AMD64 and Linux

Thanks to the hard efforts of the community, Linux has supported AMD64 for some time now. The new 64-bit kernel is based on the existing i386 port, and a lot of effort has been made to ensure that the new features of the Opteron are used to best possible advantage. SuSE and AMD have been working very closely to develop the new software, and

Learn More

Links and treeware

If you're interested in learning more about AMD64, AMD provide an excellent set of books under the banner, **AMD AMD64 Architecture**. If you're in the US or Canada, you can order a complete set of these books for delivery to your home. Readers outside of North America have to settle for the online version, available from the AMD site.

Though it hasn't been updated for some time, www.x86-64.org is still worth reading – there's quite a bit of info available there on how to write code for AMD64, and also how the effort to make AMD64-compatible versions of various free OSes is going.

AMD's website contains a great deal of info on Opteron – how it works, from where it's available, and such. It is



First-class reference for the IPF and kernel development in general.

of course a fairly one-sided opinion, and you'll find a better representation of Opteron by running a Google search.

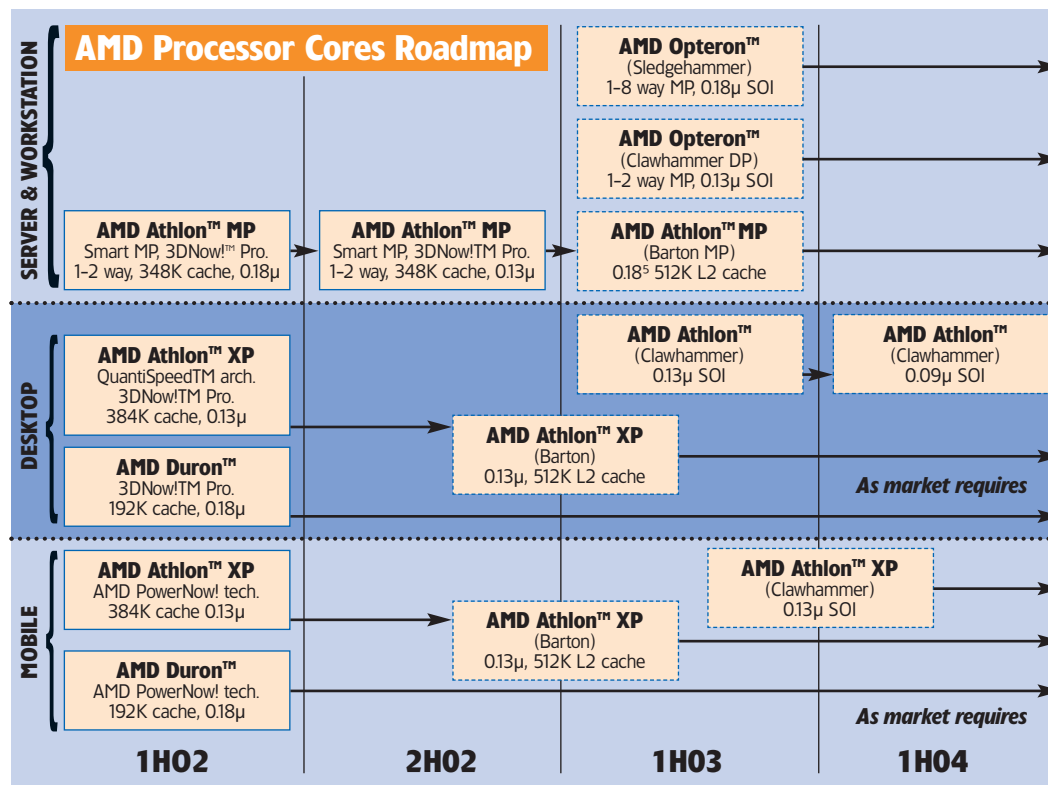
If you're interested in learning more about IA-64 and Itanium, the book *ia-64 Linux kernel* (Mosberger and Eranian, Prentice Hall ISBN 0-1306-1014-3), documents how the team at HP ported Linux to the IA-64 platform, and includes excellent information on the IPF and lots of insight into the Linux kernel itself. It's a bit pricey though.

naturally, it's released as GPL as part of the 2.5 kernel. Customers of UnitedLinux, such as companies who bought SuSE's excellent *Linux Enterprise Server 8* product, will be able to benefit from this code already, as the UnitedLinux kernel has many 2.5 backports, including AMD64 support. As such, SLES was the first server operating system available to fully support Opteron's architecture.

Red Hat's *Enterprise Server* product does not support AMD64 at the time of writing, and it seems they have plans only to integrate it into the next version of the software. MandrakeSoft, for its part, launched Mandrake Corporate Server 2.1 for Opteron on the day of the Opteron launch, although we've yet to have time to review it fully. Red Hat has promised us that a beta will be available by the time you read this.

AMD64 & other OSes

In late 2002, Microsoft delivered a development release of AMD64 Windows to several of its industry partners, and has recently announced that it is developing native 64-bit versions of its Windows XP and Windows Server 2003 operating



systems for the Opteron and Athlon 64 platforms. Specific launch dates are a little vague, with MS saying that beta releases are expected in the middle of 2003. It's important to remember that the support of

Microsoft and Windows will provide an important boost to the Hammer architecture, and could spur large-scale adoption of the CPU.

Both FreeBSD and NetBSD have work underway to support the AMD64

architecture, with NetBSD already having the port ready for public use.

So, with Linux, Windows, and two of the three major BSDs developing support for Opteron, it is quite likely that adoption will be swift. With SLES



The Rise and Fall of Itanium

Lack of backwards-compatibility

The original Itanium CPU was launched on May 29th 2001, although it lacked fanfare. Indeed, some contend that it was never really launched at all, because it wasn't used much, had little media coverage, and was replaced by its successor a little over a year later.

Both the original Itanium and the Itanium 2 form the Itanium Processor Family (IPF), and constitute an entirely new 64-bit CPU architecture, named IA-64. With IA-64, Intel partnered with HP to produce a new 64-bit system that would be the best CPU they could design, and this involved coming up with a new processing paradigm, known as Explicitly Parallel Instruction Computing (EPIC). Borne from the Very-Long Instruction Word (VLIW) paradigm, EPIC relies on compilers to handle instruction scheduling in order to minimise CPU complexity.

This "feature" is Itanium's biggest advantage and also its biggest disadvantage. Traditional x86-32 CPUs use peep-hole optimisation at run-time to make code execute faster – that is, as instructions are loaded into the CPU, it checks to see whether it can reorder

things and/or speculatively execute code to improve speed. However, it does so using a "peep-hole" – it has a very narrow view of the code in order to not slow things down too much and also to keep the amount of instruction scheduling-related silicon down. EPIC relies heavily on the fact that during the compilation stage, the compiler is able to have knowledge of all parts of a program at the same time – what variables it has, when they will be accessed, how they will be accessed, etc. As such, it is able to generate far more optimised code than CPU-level peep-hole optimisation can. EPIC allows compilers to specifically mark chunks of code that can be executed in parallel, which can provide a major speed boost.

However, the problem relies in the fact that, as you can imagine, Itanium relies heavily on a *highly-optimised* compiler – without one, Itanium CPUs will certainly not perform at their best. As such, some people have rechristened EPIC to stand for "Expect Perfectly Intuitive Compilers", because compilers must be very advanced to properly take advantage of Itanium's architecture.

Another drawback is that because Itanium is so different from x86-32, it requires an emulation layer. As discussed earlier, this emulation layer isn't fast by any means despite being implemented in hardware (although a new, software based emulation system has just been announced which is claimed to increase performance of emulated code). This means that although companies opting to switch to Itanium can retain backwards compatibility with their existing applications, they must accept very low performance when running emulated software – satisfactory performance can only be achieved by changing software along with the hardware. Furthermore, the compatibility that exists for x86-32 is only for applications: Itanium must run on a 64-bit operating system.

Programming for Itanium is quite a jump from programming for x86-32 – it has 128 integer registers and 128 floating-point registers, and also has a much more complicated assembly language. While there's no doubt that Itanium is certainly a great chip in its own right, it currently has no place in the x86-32 world. It's possible

that Madison, the expected third member of the IPF, may change this by increasing the performance of x86-32 code, however, there is one other option for Intel: Yamhill.

Yamhill, the supposed codename for a rumoured Intel project to produce an AMD64-compatible CPU, is right now unconfirmed and merely circulating by means of Chinese whispers. However, given that AMD allows licensing of AMD64, Intel would be crazy to not be developing an Opteron competitor just in case the technology is a big hit. Of course, that would leave Intel following in AMD's footsteps for the indefinite future – probably not a situation Intel want to be in. However, if companies move over to 64-bit for new computers – which is seeming increasingly likely as the business benefits become more apparent – Intel might lose out either way.

A lot rides on Madison: if Intel can bring x86-32 performance up to speed, then release a fast and optimised compiler for little cost to take advantage of the technologically clever EPIC system, Itanium might have life in it yet.

Opteron

« 8, SuSE has worked very hard to make sure companies wanting to switch to Opteron have a reliable, well-supported Linux platform available.

Can't touch this!

Hot topics right now are: "how fast can Opteron users expect their new computers to be", and "will it perform better in 32-bit mode than the IPF"? The second question is easier to answer, so we'll cover that first. Put simply, the answer is: almost certainly. Because Itanium shares no relation with the x86 architecture, it's difficult but not impossible to emulate a 32-bit CPU with software only. As a result, Itanium defines a special interface to translate x86-32 instructions into Itanium (IA-64) instructions that can be (and currently is) implemented in hardware. This translator, known as the Intel Value Engine, has the tricky job of converting x86-32 instructions into IA-64 instructions, which can then be executed. As you can imagine, this is quite a speed hit.

The Opteron, however, defaults to 32-bit mode by default, and becomes a particularly fast Athlon with SSE and SSE2 enabled for all intents and purposes. That is, it executes 32-bit instructions at the same speed or faster than a clock-matched Athlon.

With regards to the speed users will find their shiny new Opterons running at, it depends on how you look at the equation. Firstly, if you use applications that make use of SSE and



SSE2 (that's pretty much any 3D software/game), you'll see an immediate and considerable speed boost. Once software is specifically compiled for 64-bit mode, you'll see a further 10-15% speed improvement from the use of the new registers, although this could rise much higher for computationally-intensive code.

One factor that might muddy the issue is AMD's choice of model rating for the new Opteron. Each Opteron will be given a three-digit number to represent its speed and scalability. The first digit will either be a 1 for uniprocessor boxes, 2 for two-way systems, 4 for four-way, or 8 for eight-way. Digits two and three are a relative performance figure, and will start at 40. So, an Opteron 140 will be the slowest uniprocessor Opteron machine, whereas an Opteron 141 will be faster than the 40 (faster by an unknown amount), and an 899 will presumably be the fastest eight-way Opteron machine available. Confused yet?

The reason, so we've been informed, that the second two digits started at 40 is purportedly so that people don't confuse the numbers as a three digit clock speed. Precisely how numbers such as 866 are don't look like a clock speed, we're not sure. However, this is the naming scheme that is to be used, and hopefully it won't confuse potential customers too much.

What's next?

The Opteron is the first chip to be built on the AMD64 architecture, but it almost certainly won't be the last. As Opteron is designed to be used in servers and workstations, AMD are working on a new version of its Athlon XP desktop CPU, dubbed Athlon 64/Clawhammer, that will also make use of the AMD64 architecture. This will allow AMD to push 64-bit computing to everyone. However, there are no plans as yet to discontinue production of the x86-32 Athlon XP or Duron chips, despite the

"I like to say they [Intel] are big, strong, and wrong... they can't stop us!"

JERRY SANDERS

Founder and chairman of the board, AMD

fact that a Clawhammer CPU would be able to emulate (and outperform) each of them.

There are also plans afoot to launch a Mobile Athlon 64 at the same time as the desktop Athlon 64, which will presumably contain much of the same power-saving functionality seen in the existing Mobile Athlon XP.

In the first half of 2004, AMD are slated to release 0.09 micron version of the Opteron (code-named *Athens*), the Athlon 64 (code-named *San Diego*), and the Mobile Athlon 64 (*Odessa*), which should allow them to continue increasing clock speeds.

Whether or not there will be an Opteron 2 out with a revamped architecture is as yet unknown. Intel are still making quite drastic changes to the Itanium, trying to get it right. With the Opteron, due to the fixed x86 heritage behind it, it's more likely that the interface will stay the same, with some parts being shuffled and tweaked internally to increase performance.

Why 64-bit computing is needed

But what's this I've heard about a diminished cache-hit rate?

We've had 32-bits for such a long time now, that it might seem a bit unnecessary to double things. 32-bits means 32 binary digits, so 32 1s and 0s are used to represent numbers. If we have a 32-bit unsigned integer (that is, it can only be positive), it can hold numbers from 0 to 4,294,967,295 (2^{32}). Now, because memory is allocated using bytes, it is also referenced using bytes – when a program wants to read from a byte, it simply looks up its number. As seen, numbers max out at 4 billion or so, which means that the maximum byte of memory that a 32-bit program can reference is number 4,294,967,295. This means that the range of memory that can be accessed is 4,294,967,296 bytes, which is divisible by 1024. So 4,294,967,296 bytes is 4,194,304 kilobytes, which is 4096

megabytes, which is 4 gigabytes. What this means is that the maximum amount of memory a 32-bit program can reference is 4GBs – sounds like a lot, no? Well, consider the topic of databases. Large companies buy very powerful servers designed to hold their entire database in RAM for the fastest access – in situations like this, 4GBs soon runs out.

Switching to 64-bit numbers squares the maximum value of integers, so the largest amount of memory that can be addressed becomes 18,446,744,073,709,551,616 bytes – 16 Exabytes! Of course, this is all virtual – there are other limitations generally involved when things are implemented in reality. With the Opteron, 40 of the 64 bits can be used, which gives 1 Terabyte (a thousand gigabytes) of memory. Furthermore, the

Opteron has a 48-bit virtual memory address space, allowing 256 Terabytes to be accessed.

So, lots more RAM is made available – but what else does switching to 64-bit help? One simple advantage is in the field of encryption, where very large numbers need to be processed quickly. 32-bit CPUs can perform 64-bit mathematics (such as long long integers in GCC), however the operation needs to be split into two for calculation, then recombined. 64-bit CPUs can natively perform 64-bit mathematics, and so avoid this speed hit.

As you can see, there's lots to be gained by switching to 64-bit – and that's before we consider maths-intensive code such as digital content creation! However, are there any downsides? Well, there is one: switching to 64-bit processing

does mean a slight decrease in the cache-hit rate. That is, because 64-bit data is twice the size of 32-bit data, less information can be stored in cache, which means it's less likely that requested data will be found in the cache (a *cache hit*). Digital Equipment Corporation found that this increase will cause a 5% performance penalty in programs that were similar apart from their bit size. To get around this, 64-bit CPUs support 32-bit data operands in order to allow programmers to keep their operands short when possible to minimise this penalty. AMD64 defaults to 32-bit data operands when working in 64-bit mode, and only uses the larger numbers when specifically instructed – this provides the best performance solution, and is easy for compilers to make use of.

Glossary

3DNOW! AMD-designed multimedia extensions to x86-32

ATHLON AMD's x86-32 CPU brand

ATHLON 64 AMD's proposed AMD64 desktop CPU brand

BIT Binary digit, a 0 or a 1

BYTE Eight binary digits

CISC Complex Instruction Set Computer, uses complex instructions that are often decoded to RISC instructions

CLAWHAMMER Codename for AMD's desktop 64-bit CPU, Athlon 64

CPU Central Processing Unit, forms the core of your PC's processing abilities

EPIC Explicitly Parallel Instruction Computing, allows compilers to specify which instructions can be executed in parallel. An extension of VLIW.

FPU Floating Point Unit, handles mathematics-specific capabilities of your PC. Usually built into the CPU

GIGABYTE 1024 Megabytes

GPR General Purpose Register, a register that can be used for general programming

IA-64 Intel's Itanium architecture

INSTRUCTION Simple binary digit that translates to one CPU operation when called

INSTRUCTION SCHEDULING the act of re-arranging instructions in a non-destructive manner to improve performance

IPF Itanium Processor Family, includes Itanium and Itanium 2

ITANIUM Intel's IA-64 implementation

KILOBYTE 1024 Bytes.

MADISON Codename for Itanium 3

MCKINLEY Codename for Itanium 2

MEGABYTE 1024 Kilobytes

MERCED Codename for Itanium

MMX Intel-designed multimedia extensions to x86-32

OPTERON AMD's server and workstation 64-CPU CPU

REGISTER A storage area directly on a CPU able to hold a small amount of data

RISC Reduced Instruction Set Computer, uses simpler instructions for faster execution at the expense of larger software

SLEDGEHAMMER Codename for AMD's Opteron

SSE Streaming SIMD Extensions. More Intel-designed multimedia extensions to x86-32

SSE2 Streaming SIMD Extensions 2. Even more Intel-designed multimedia extensions to x86-32

SIMD Single Instruction, Multiple Data. The technique of allowing one CPU instruction to act on multiple registers

TERABYTE 1024 Terabytes

VLIW Very-Long Instruction Word, method for allowing compilers better control over execution, while making CPUs simpler

X86 Common name for 386-compatible CPUs

X86-32 32-bit x86 CPU architecture

X86-64 Original name for AMD64 64-bit architecture

AMD64 64-bit x86 CPU architecture

X87 The mathematics architecture used in x86 PCs

XMM Internal name for SSE and SSE2 combined

Note: recent standardisation has changed the meaning of kilobytes, megabytes, etc. In order to preserve ease of reading, we have used the traditional usage of megabytes: 1024 kilobytes.


Conclusion

The Opteron is the biggest step forward AMD has ever made, and consequently also its biggest risk. However, having worked so hard these past years, AMD hasn't rushed to market with the Opteron – it has put a lot of thought into how to give customers the most value for money, and has also worked hard to recruit the best partners to help make AMD64 a success.

Will Opteron be a success? That's a hard question to answer, given companies' general dislike of changing what works. The Itanium fell into this trap by breaking native compatibility with existing popular hardware, but the Opteron doesn't do that – indeed, users can switch to Opteron immediately and the absolute worst-case scenario is that they'll continue to run their 32-bit systems and therefore only get a particularly powerful 32-bit CPU.

The one thing that might endanger the success of Opteron is that AMD has long been considered "almost as good

as Intel, but much cheaper". While this belief has been true at times in the past, it would likely damage Opteron sales if it were seen as a cheap version of Itanium, or "the poor man's 64-bit CPU". The server market carries with it a much higher margin level than the desktop market, which should in theory allow AMD to reach profitability quickly if they can keep its prices at a good level. However, if AMD somehow manage to sink back into its old persona of being billed as the "value option", this increase in revenue will certainly be jeopardised.

It's our opinion that AMD64 has a strong proposition for both the consumer and enterprise market. Long term popularity is likely to be dictated by the availability of software to take advantage of it, but that doesn't seem much of an obstacle in the Linux world at least. Time will tell, and the stakes may be high for AMD, but the Opteron and the AMD64 chips that will follow certainly seem to have a future. 

The AMD64 Launch

Paul Hudson reports from New York, Tuesday 22nd April.

I flew to NY to the Opteron launch event, and were mightily impressed – AMD has spared no expense in getting Opteron noticed by the media at large, and are eager to make sure that everyone who wants to know about Opteron can get lots of information.

Many VARs attended, including MSI, Nvidia, IBM, Oracle, Microsoft, and others. Crucially, SuSE, Red Hat, and Mandrake were all there, and more than willing to disclose lots of information about their plans for Opteron. As such, Opteron has a lot of support from launch – several of the speakers got up and declared that Opteron was pretty much the best thing that had happened to the x86 industry since, well, ever. A recurring theme was "It only took us X days to port to AMD64". For example, IBM ported DB/2 to Opteron/Linux in just two days, despite it having 10,000,000 lines of code – quite an achievement!

Perhaps the best quote from the trip was from Jerry Sanders himself, overheard at a private VAR dinner the night before Opteron launched: "With Opteron, AMD is going to change the world, and unlike Steve Jobs we're going to change the *real* world." With unsubstantiated rumours flying that AMD and Apple might have something in the works (OS X for Opteron?!), this seems like quite a risky comment to make!

Both Hector Ruiz (AMD CEO) and Jerry Sanders (Chairman of the Board), gave an excellent presentation making the situation quite clear: AMD64 is the way forward. Some laughter was heard during the media Q&A session when one reporter questioned whether there had been any "poisoning of the wells" from Intel or AMD. Humorous, yes, but less so considering that both myself and several other reporters there questioned company representatives at the launch, asking what Intel thought of them supporting AMD. The (scary!) response was more often than not that they had received a call from an Intel representative before they came to the Opteron launch asking the vaguely threatening question about whether they *really* wanted to go, and that some vendors had pulled out almost certainly as a result of the call. Co-incidence? Possibly – we'd like to think so. Eager to give Intel the chance to say it was all just a coincidence, we gave them a call "All we can say is that the reports are speculation and rumour" came the response. I'll leave you to make your own mind up.

Gene Kim, VP of Global Sales and Marketing, RackSaver said: "We're very

proud to announce that based on our Quadrix-64 platform, our 4-Way Opteron, we have reached the highest throughput score ever achieved for a 4-way system as measured by the TPC council ever. Not only that, but that score based on that same platform also has the lowest price and performance ratio, beating even Dell."

At the launch, Opteron models 240, 242, and 244 were announced. Clearly heralding the start of mass confusion regarding this naming convention, AMD representatives from the Strategic Marketing department were unable to tell us precisely how fast these CPUs were in Gigahertz, although continued questioning from other AMD sources led to the answer 1.4, 1.6, and 1.8GHz respectively – not terrifically high speeds, but it leaves the Opteron a lot of space to grow.

On the subject of performance, early reports are already coming in of Opteron blowing competition out of the water. AMD's official results show model 444 Opterons are 37% faster than Xeon 2.8GHz running SPECint, despite the massive clock speed difference. Microsoft's own MMB benchmark results gave model 424 Opterons a score of 15,520 in real world tests, compared to just 13,200 for the next best (4 2GHz Xeon MPs). The same test had two model 244 Opterons trouncing four 1.6GHz Xeon MPs – wow!

I had the chance to talk to people from Red Hat, Mandrake, and SuSE – both were strongly behind the Opteron movement. Red Hat intend to support Opteron in all three of its Enterprise Linux platforms in version 3 of the family. Importantly, Red Hat said they will allow users to upgrade to an Opteron-enabled v3 system as long as the company's subscription is paid up. As to whether Red Hat will ship the x86-32 and AMD64 versions in the same box a la SLES 8, "That decision has not yet been taken". Further uncertainty arises from the fact that the current release of Red Hat's Enterprise Linux isn't supported for use on Opteron systems. Nick Carr, Red Hat's Enterprise OS Marketing Manager, informed us that while customer support will dictate its Opteron policy on existing releases, v3 of its Enterprise Linux platform will definitely support the chip.

SuSE didn't really have much to say, which was odd considering nearly all of the Linux machines there were running SLES8. MandrakeSoft gave us an early copy of its new Opteron Corporate Server 2.1, and we're looking forward to reviewing that in a future issue.