



Original Link: <https://www.anandtech.com/show/1719>

## Microsoft's Xbox 360, Sony's PS3 - A Hardware Discussion

93  
Comments

by [Anand Lal Shimpi & Derek Wilson](#) on June 24, 2005 4:05 AM EST

Posted in [GPUs](#)

The point of a gaming console is to play games. The PC user in all of us wants to benchmark, overclock and upgrade even the unreleased game consoles that were announced at E3, but we can't. And these sorts of limits are healthy, because it lets us have a system that we don't tinker with, that simply performs its function and that is to play games.

The game developers are the ones that have to worry about which system is faster, whose hardware is better and what that means for the games they develop, but to us, the end users, whether the Xbox 360 has a faster GPU or the PlayStation 3's CPU is the best thing since sliced bread doesn't really matter. At the end of the day, it is the games and the overall experience that will sell both of these consoles. You can have the best hardware in the world, but if the games and the experience aren't there, it doesn't really matter.

Despite what we've just said, there is a desire to pick these new next-generation consoles apart. Of course if the games are all that matter, why even bother comparing specs, claims or anything about these next-generation consoles other than games? Unfortunately, the majority of that analysis seems to be done by the manufacturers of the consoles, and fed to the users in an attempt to win early support, and quite a bit of it is obviously tainted.

While we would've liked this to be an article on all three next-generation consoles, the Xbox 360, PlayStation 3 and Revolution, the fact of the matter is that Nintendo has not released any hardware details about their next-gen console, meaning that there's nothing to talk about at this point in time. Leaving us with two contenders: Microsoft's Xbox 360, due out by the end of this year, and Sony's PlayStation 3 due out in Spring 2006.

This article isn't here to crown a winner or to even begin to claim which platform will have better games, it is simply here to answer questions we all have had as well as discuss these new platforms in greater detail than we have before.

Before proceeding with this article, there's a bit of required reading to really get the most out of it. We strongly suggest reading through our [Cell processor article](#), as well as [our launch coverage of the PlayStation 3](#). We would also suggest reading through our [Xbox 360 articles](#) for background on Microsoft's console, as well as [an earlier piece published on multi-threaded game development](#). Finally, be sure that you're fully up to date on the [latest GPUs](#), especially the recently announced [NVIDIA GeForce 7800 GTX](#) as it is very closely related to the graphics processor in the PS3.

This article isn't a successor to any of the aforementioned pieces, it just really helps to have an understanding of everything we've covered before - and since we don't want this article to be longer than it already is, we'll just point you back there to fill in the blanks if you find that there are any.

Now, on to the show...

### A Prelude on Balance

The most important goal of any platform is balance on all levels. We've seen numerous examples of what architectural imbalances can do to performance, having too little cache or too narrow of a FSB can starve high speed CPUs of data they need to perform. GPUs without enough memory bandwidth can't perform anywhere near their peak fillrates, regardless of what they may be. Achieving a balanced overall platform is a very difficult thing on the PC, unless you have an unlimited budget and are able to purchase the fastest components. Skimping on your CPU while buying the most expensive graphics card may leave you with performance that's marginally better, or worse, than someone else with a more balanced system with a faster CPU and a somewhat slower GPU.

With consoles however, the entire platform is designed to be balanced out of the box, as best as the manufacturer can get it to be, while still remaining within the realm of affordability. The manufacturer is responsible for choosing bus widths, CPU architectures, memory bandwidths, GPUs, even down to the type of media that will be used by the system - and most importantly, they make sure that all elements of the system are as balanced as can be.

The reason this article starts with a prelude on balance is because you should not expect either console maker to have put together a horribly imbalanced machine. A company who is already losing money on every

ATI Ex. 2136  
IPR2023-00922  
Page 1 of 12

10/4/2022, 10:25 PM  
AMD1318\_0154194

console sold, will never put faster hardware in that console if it isn't going to be utilized thanks to an imbalance in the platform. So you won't see an overly powerful CPU paired with a fill-rate limited GPU, and you definitely won't see a lack of bandwidth to inhibit performance. What you will see is a collection of tools that Microsoft and Sony have each, independently, put together for the game developer. Each console has its strengths and its weaknesses, but as a whole, each console is individually very well balanced. So it would be wrong to say that the PlayStation 3's GPU is more powerful than the Xbox 360's GPU, because you can't isolate the two and compare them in a vacuum, how they interact with the CPU, with memory, etc... all influences the overall performance of the platform.

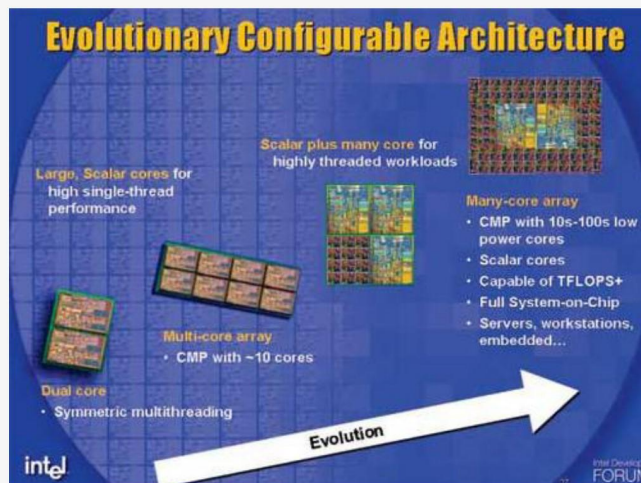
### The Consoles and their CPUs

The CPUs at the heart of these two consoles are very different in architecture approach, despite sharing some common parts. The Xbox 360's CPU, codenamed Xenon, takes a general purpose approach to microprocessor design and implements three general purpose PowerPC cores, meaning they can execute any type of code and will do it relatively well.

The PlayStation 3's CPU, the Cell processor, pairs a general purpose PowerPC Processing Element (PPE, very similar to one core from Xenon) with 7 working Synergistic Processing Elements (SPEs) that are more specialized hardware designed to execute certain types of code.

So the comparison between Xenon and Cell really boils down to a comparison between a general purpose microprocessor, and a hybrid of general purpose and specialized hardware.

Despite what many have said, there is support for Sony's approach with Cell. We have discussed, in great detail, the architecture of the Cell processor already but there is industry support for a general purpose + specialized hardware CPU design. Take note of the following slide from Intel's Platform 2015 vision for their CPUs by the year 2015:



The use of one or two large general purpose cores combined with specialized hardware and multiple other smaller cores is in Intel's roadmap for the future, despite their harsh criticism of the Cell processor. The difference is that Cell appears to be far too early for its time. By 2015 CPUs may be manufactured on as small as a 32nm process, significantly smaller than today's 90nm process, meaning that a lot more hardware can be packed into the same amount of space. In going with a very forward-looking design, the Cell processor architects inevitably had to make sacrifices to deal with the fact that the chip they wanted to design is years ahead of its time for use in general computation.

### Introducing the Xbox 360's Xenon CPU

The Xenon processor was designed from the ground up to be a 3-core CPU, so unlike Cell, there are no disabled cores on the Xenon chip itself in order to improve yield. The reason for choosing 3 cores is because it provides a good balance between thread execution power and die size. According to Microsoft's partners, the sweet spot for this generation of consoles will be between 4 and 6 execution threads, which is where the 3-core CPU came from.

The chip is built on a 90nm process, much like Cell, and will run at 3.2GHz - also like Cell. All of the cores are identical to one another, and they are very similar to the PPE used in the Cell microprocessor, with a few modifications.

The focus of Microsoft's additions to the core has been in the expansion of the VMX instruction set. In particular, Microsoft now includes a single cycle dot-product instruction as a part of the VMX-128 ISA that is implemented on each core. Microsoft has stated that there is nothing stopping IBM from incorporating this

support into other chips, but as of yet we have not seen anyone from the Cell camp claim support for single cycle dot-products on the PPE.

The three cores share a meager 1MB L2 cache, which should be fine for single threaded games but as developers migrate more to multi-threaded engines, this small cache will definitely become a performance limiter. With each core being able to execute two threads simultaneously, you effectively have a worst case scenario of 6 threads splitting a 1MB L2 cache. As a comparison, the current dual core Pentium 4s have a 1MB L2 cache per core and that number is only expected to rise in the future.

The most important selling point of the Xbox 360's Xenon core is the fact that all three cores are identical, and they are all general purpose microprocessors. The developer does not have to worry about multi-threading beyond the point of getting their code to be thread safe; once it is multi-threaded, it can easily be run on any of the cores. The other important thing to keep in mind here is that porting between multi-core PC platforms and the Xbox 360 will be fairly trivial. Anywhere any inline assembly is used there will obviously have to be changes, but with relatively minor code changes and some time optimizing, code portability between the PC and the Xbox 360 shouldn't be very difficult at all. For what it is worth, porting game code between the PC and the Xbox 360 will be a lot like Mac developers porting code between Mac OS X for Intel platforms and PowerPC platforms: there's an architecture switch, but the programming model doesn't change much.

The same cannot however be said for Cell and the PlayStation 3. The easiest way to port code from the Xbox 360 to the PS3 would be to run the code exclusively on the Cell's single PPE, which obviously wouldn't offer very good performance for heavily multi-threaded titles. But with a some effort, the PlayStation 3 does have a lot of potential.

### Xenon vs. Cell

The first public game demo on the PlayStation 3 was Epic Games' Unreal Engine 3 at Sony's PS3 press conference. Tim Sweeney, the founder and UE3 father of Epic, performed the demo and helped shed some light on how multi-threading can work on the PlayStation 3.

According to Tim, a lot of things aren't appropriate for SPE acceleration in UE3, mainly high-level game logic, artificial intelligence and scripting. But he adds that "Fortunately these comprise a small percentage of total CPU time on a traditional single-threaded architecture, so dedicating the CPU to those tasks is appropriate, while the SPE's and GPU do their thing."

So what does Tim Sweeney see the SPEs being used for in UE3? "With UE3, our focus on SPE acceleration is on physics, animation updates, particle systems, sound; a few other areas are possible but require more experimentation."

Tim's view on the PPE/SPE split in Cell is far more balanced than most we've encountered. There are many who see the SPEs as utterly useless for executing anything (we'll get to why in a moment), while there are others who have been talking about doing far too much on SPEs where the general purpose PPE would do much better.

For the most part, the areas that UE3 uses the Cell's SPEs for are fairly believable. For example, sound processing makes a lot of sense for the SPEs given their rather specialized architecture aimed at streaming tasks. But the one curious item is the focus on using SPEs to accelerate physics calculations, especially given how branch heavy physics calculations generally are.

Collision detection is a big part of what is commonly referred to as "game physics." As the name implies, collision detection simply refers to the game engine determining when two objects collide. Without collision detection, bullets would never hit your opponents and your character would be able to walk through walls, cars, etc... among other things.

One method of implementing collision detection in a game is through the use of a Binary Search Partitioning (BSP) tree. BSP trees are created by organizing lists of polygons into a binary tree. The structure of the tree itself doesn't matter to this discussion, but the important thing to keep in mind is that to traverse a BSP tree in order to test for a collision between some object and a polygon in the tree you have to perform a lot of comparisons. You first traverse the tree finding to find the polygon you want to test for a collision against. Then you have to perform a number of checks to see whether a collision has occurred between the object you're comparing and the polygon itself. This process involves a lot of conditional branching, code which likes to be run on a high performance OcO core with a very good branch predictor.

Unfortunately, the SPEs have no branch prediction, so BSP tree traversal will tie up an SPE for quite a bit of time while not performing very well as each branch condition has to be evaluated before execution can continue. However it is possible to structure collision detection for execution on the SPEs, but it would require a different approach to the collision detection algorithms than what would be normally implemented on a PC or Xbox 360.

We're still working on providing examples of how it is actually done, but it's tough getting access to detailed information at this stage given that a number of NDAs are still in place involving Cell development for the PS3. Regardless of how it is done, obviously the Epic team found the SPEs to be a good match for their physics code, if structured properly, meaning that the Cell processor isn't just one general purpose core with 7

others that go unused.

In fact, if properly structured and coded for SPE acceleration, physics code could very well run faster on the PlayStation 3 than on the Xbox 360 thanks to the more specialized nature of the SPE hardware. Not to mention that physics acceleration is particularly parallelizable, making it a perfect match for an array of 7 SPEs.

Microsoft has referred to the Cell's array of SPEs as a bunch of DSPs useless to game developers. The fact that the next installment of the Unreal engine will be using the Cell's SPEs for physics, animation updates, particle systems as well as audio processing means that Microsoft's definition is a bit off. While not all developers will follow in Epic's footsteps, those that wish to remain competitive and get good performance out of the PS3 will have to.

The bottom line is that Sony would not foolishly spend over 75% of their CPU die budget on SPEs to use them for nothing more than fancy DSPs. Architecting a game engine around Cell and optimizing for SPE acceleration will take more effort than developing for the Xbox 360 or PC, but it can be done. The question then becomes, will developers do it?

In Johan's Quest for More Processing Power series he looked at the developmental limitations of multi-threading, especially as they applied to games. The end result is that multi-threaded game development takes between 2 and 3 times longer than conventional single-threaded game development, to add additional time in order to restructure elements of your engine to get better performance on the PS3 isn't going to make the transition any easier on developers.

### Why In-Order?

Ever since the Pentium Pro, desktop PC microprocessors have implemented Out of Order (OoO) execution architectures in order to improve performance. We've explained the idea in great detail before, but the idea is that an Out-of-Order microprocessor can reorganize its instruction stream in order to best utilize its execution resources. Despite the simplicity of its explanation, implementing support for OoO dramatically increases the complexity of a microprocessor, as well as drives up power consumption.

In a perfect world, you could group a bunch of OoO cores on a single die and offer both excellent single threaded performance, as well as great multi-threaded performance. However, the world isn't so perfect, and there are limitations to how big a processor's die can be. Intel and AMD can only fit two of their OoO cores on a 90nm die, yet the Xbox 360 and PlayStation 3 targeted 3 and 9 cores, respectively, on a 90nm die; clearly something has to give, and that something happened to be the complexity of each individual core.

Given a game console's 5 year expected lifespan, the decision was made (by both MS and Sony) to favor a multi-core platform over a faster single-core CPU in order to remain competitive towards the latter half of the consoles' lifetime.

So with the Xbox 360 Microsoft used three fairly simple IBM PowerPC cores, while Sony has the much publicized Cell processor in their PlayStation 3. Both will perform absolutely much slower than even mainstream desktop processors in single threaded game code, but the majority of games these days are far more GPU bound than CPU bound, so the performance decrease isn't a huge deal. In the long run, with a bit of optimization and running multi-threaded game engines, these collections of simple in-order cores should be able to put out some fairly good performance.

### Does In-Order Matter?

As we discussed in our Cell article, in-order execution makes a lot of sense for the SPEs. With in-order execution as well as a small amount of high speed local memory, memory access becomes quite predictable and code is very easily scheduled by the compiler for the SPEs. However, for the PPE in Cell, and the PowerPC cores in Xenon, the in-order approach doesn't necessarily make a whole lot of sense. You don't have the advantage of a cacheless architecture, even though you do have the ability to force certain items to remain untouched by the cache. More than anything having an in-order general purpose core just works to simplify the core, at the expense of depending quite a bit on the compiler, and the programmer, to optimize performance.

Very little of modern day games is written in assembly, most of it is written in a high level language like C or C++ and the compiler does the dirty work of optimizing the code and translating it into low level assembly. Compilers are horrendously difficult to write; getting a compiler to work is a pretty difficult job in itself, but getting one to work well, regardless of what the input code is, is nearly impossible.

However, with a properly designed ISA and a good compiler, having an in-order core to work on is not the end of the world. The performance you lose by not being able to extract the last bit of instruction level parallelism is made up by the fact that you can execute far more threads per clock thanks to the simplicity of the in-order cores allowing more to be packed on a die. Unfortunately, as we've already discussed, on day one that's not going to be much of an advantage.

The Cell processor's SPEs are even more of a challenge, as they are more specialized hardware only suitable to executing certain types of code. Keeping in mind that the SPEs are not well suited to running

branch heavy code, loop unrolling will do a lot to improve performance as it can significantly reduce the number of branches that must be executed. In order to squeeze the absolute maximum amount of performance out of the SPEs, developers may be forced to hand code some routines as initial performance numbers for optimized, compiled SPE code appear to be far less than their peak throughput.

While the move to in-order architectures won't cause game developers too much pain with good compilers at their disposal, the move to multi-threaded game development and optimizing for the Cell in general will be much more challenging.

### How Many Threads?

Earlier this year we saw the beginning of a transition from very fast, single core microprocessors to slower, multi-core designs on the PC desktop. The full transition won't be complete for another couple of years, but just as it has begun on the desktop PC side, it also has begun in the next-generation of consoles.

Remember that consoles must have a lifespan of around 5 years, so even if the multithreaded transition isn't going to happen with games for another 2 years, it is necessary for these consoles to be built around multi-core processors to support the ecosystem when that transition occurs.

The problem is that today, all games are single threaded, meaning that in the case of the Xbox 360, only one out of its three cores would be utilized when running present day game engines. The PlayStation 3 would fair no better, as the Cell CPU has a very similar general purpose execution core to one of the Xbox 360 cores. The reason this is a problem is because these general purpose cores that make up the Xbox 360's Xenon CPU or the single general purpose PPE in Cell are extremely weak cores, far slower than a Pentium 4 or Athlon 64, even running at much lower clock speeds.

Looking at the Xbox 360 and PlayStation 3, we wondered if game developers would begin their transition to multithreaded engines with consoles and eventually port them to PCs. While the majority of the PC installed base today still runs on single-core processors, the install base for both the Xbox 360 and PS3 will be guaranteed to be multi-core, so what better platform to introduce a multithreaded game engine than the new consoles where you can guarantee that all of your users will be able to take advantage of the multithreading.

On the other hand, looking at all of the early demos we've seen of Xbox 360 and PS3 games, not a single one appears to offer better physics or AI than the best single threaded games on the PC today. At best, we've seen examples of ragdoll physics similar to that of Half Life 2, but nothing that is particularly amazing, earth shattering or shocking. Definitely nothing that appears to be leveraging the power of a multicore processor.

In fact, all of the demos we've seen look like nothing more than examples of what you can do on the latest generation of GPUs - not showcases of multi-core CPU power. So we asked Microsoft, expecting to get a fluffy answer about how all developers would be exploiting the 6 hardware threads supported by Xenon, instead we got a much more down to earth answer.

The majority of developers are doing things no differently than they have been on the PC. A single thread is used for all game code, physics and AI and in some cases, developers have split out physics into a separate thread, but for the most part you can expect all first generation and even some second generation titles to debut as basically single threaded games. The move to two hardware execution threads may in fact only be an attempt to bring performance up to par with what can be done on mid-range or high-end PCs today, since a single thread running on Xenon isn't going to be very competitive performance wise, especially executing code that is particularly well suited to OoO desktop processors.

With Microsoft themselves telling us not to expect more than one or two threads of execution to be dedicated to game code, will the remaining two cores of the Xenon go unused for the first year or two of the Xbox 360's existence? While the remaining cores won't directly be used for game performance acceleration, they won't remain idle - enter the Xbox 360's helper threads.

The first time we discussed helper threads on AnandTech was in reference to additional threads, generated at runtime, that could use idle execution resources to go out and prefetch data that the CPU would eventually need.

The Xbox 360 will use a few different types of helper threads to not only make the most out of the CPU's performance, but to also help balance the overall platform. Keep in mind that with the 360, Microsoft has not increased the size of the media that games will be stored on. The dual layer DVD-9 spec is still in effect, meaning that game developers shipping titles for the Xbox 360 in 2006 will have the same amount of storage space as they did back in 2001. Given that current Xbox titles generally use around 4.5GB of space, it's not a big deal, but by 2010 9GB may feel a bit tight.

Thanks to idle execution power in the 3-core Xenon, developers can now perform real-time decompression of game data in order to maximize storage space. Given that a big hunk of disc space is used by audio and video, being able to use more sophisticated compression algorithms for both types of data will also help maximize that 9GB of storage. Or, if space isn't as much of a concern, developers are now able to use more sophisticated encoding algorithms to encode audio/video to use the same amount of space as they are today, but achieve much higher quality audio and video. Microsoft has already stated that in game video will essentially use the WMV HD codec. The real time decompression of audio/video will be another use for the

extra power of the system.

Another interesting use will be digital audio encoding; in the original Xbox Microsoft used a relatively expensive DSP featured in the nForce south bridge to perform real-time Dolby Digital Encoding. The feature allowed Microsoft to offer a single optical out on the Xbox's HD AV pack, definitely reducing cable clutter and bringing 5.1 channel surround sound to the game console. This time around, DD encoding can be done as a separate thread on the Xenon CPU - in real time. It reduces the need for Microsoft to purchase a specialized DSP from another company, and greatly simplifies the South Bridge in the Xbox 360.

But for the most part, on day 1, you shouldn't expect Xbox 360 games to be much more than the same type of single threaded titles we've had on the PC. In fact, the biggest draw to the new consoles will be the fact that for the first time, we will have the ability to run games rendered internally at 1280 x 720 on a game console. In other words, round one of the next generation of game consoles is going to be a GPU battle.

The importance of this fact is that Microsoft has been talking about the general purpose execution power of the Xbox 360 and how it is 3 times that of the PS3's Cell processor. With only 1 - 2 threads of execution being dedicated for game code, the advantage is pretty much lost at the start of the console battle.

Sony doesn't have the same constraints that Microsoft does, and thus there is less of a need to perform real time decompression of game content. Keep in mind that the PS3 will ship with a Blu-ray drive, with Sony's minimum disc spec being a hefty 23.3GB of storage for a single layer Blu-ray disc. The PS3 will also make use of H.264 encoding for all video content, the decoding of which is perfectly suited for the Cell's SPEs. Audio encoding will also be done on the SPEs, once again as there is little need to use any extra hardware to perform a task that is perfectly suited for the SPEs.

### The Xbox 360 GPU: ATI's Xenos

On a purely hardware level, ATI's Xbox 360 GPU (codenamed Xenos) is quite interesting. The part itself is made up of two physically distinct silicon ICs. One IC is the GPU itself, which houses all the shader hardware and most of the processing power. The second IC (which ATI refers to as the "daughter die") is a 10MB block of embedded DRAM (eDRAM) combined with the hardware necessary for z and stencil operations, color and alpha processing, and anti aliasing. This daughter die is connected to the GPU proper via a 32GB/sec interconnect. Data sent over this bus will be compressed, so usable bandwidth will be higher than 32GB/sec. Inside the daughter die, between the processing hardware and the eDRAM itself, bandwidth is 256GB/sec.

At this point in time, much of the bandwidth generated by graphics hardware is required to handle color and z data moving to the framebuffer. ATI hopes to eliminate this as a bottleneck by moving this processing and the back framebuffer off the main memory bus. The bus to main memory is 512MB of 128-bit 700MHz GDDR3 (which results in just over 22GB/sec of bandwidth). This is less bandwidth than current desktop graphics cards have available, but by offloading work and bandwidth for color and z to the daughter die, ATI saves themselves a good deal of bandwidth. The 22GB/sec is left for textures and the rest of the system (the Xbox implements a single pool of unified memory).

The GPU essentially acts as the Northbridge for the system, and sits in the middle of everything. From the graphics hardware, there is 10.8GB/sec of bandwidth up and down to the CPU itself. The rest of the system is hooked in with 500MB/sec of bandwidth up and down. The high bandwidth to the CPU is quite useful as the GPU is able to directly read from the L2 cache. In the console world, the CPU and GPU are quite tightly linked and the Xbox 360 stands to continue that tradition.

Weighing in at 332M transistors, the Xbox 360 GPU is quite a powerful part, but its architecture differs from that of current desktop graphics hardware. For years, vertex and pixel shader hardware have been implemented separately, but ATI has sought to combine their functionality in a unified shader architecture.

### What's A Unified Shader Architecture?

The GPU in the Xbox 360 uses a different architecture than we are used to seeing. To be sure, vertex and pixel shader programs will run on the part, but not on separate segments of the hardware. Vertex and pixel processing differ in purpose, but there is quite a bit of overlap in the type of hardware needed to do both. The unified shader architecture that ATI chose to use in their Xbox 360 GPU allows them to pack more functionality onto fewer transistors as less hardware needs to be duplicated for use in different parts of the chip and will run both vertex and shader programs on the same hardware.

There are 3 parallel groups of 16 shader units each. Each of the three groups can either operate on vertex or pixel data. Each shader unit is able to perform one 4 wide vector operation and 1 scalar operation per clock cycle. Current ATI hardware is able to perform two 3 wide vector and two scalar operations per cycle in the pixel pipe alone. The vertex pipeline of R420 is 6 wide and can do one vector 4 and one scalar op per cycle. If we look at straight up processing power, this gives R420 the ability to crunch 158 components (30 of which are 32bit and 128 are limited to 24bit precision). The Xbox GPU is able to crunch 240 32bit components in its shader units per clock cycle. Where this is a 51% increase in the number of ops that can be done per cycle (as well as a general increase in precision), we can't expect these 48 pipelines to act like 3 sets of R420 pipelines. All things being equal, this increase (when only looking at ops/cycle) would be only as powerful as a 24 piped R420.

What will make or break the difference between something like a 24 piped R420 and the unified shaders of the Xbox GPU is how well applications will lend themselves to the adaptive nature of the hardware. Current configurations don't have nearly the same vertex processing power as they do pixel processing power. This is quite logical when we consider the fact that games have many more pixels displayed than vertices. For each geometry primitive, there are likely a good number of pixels involved. Of course, not all titles will need the same ratio of geometry to pixel power. This means that all the ops per clock could either be dedicated to geometry processing in truly polygon intense scenes. On the flip side (and more likely), any given clock cycle could see all 240 ops being used for pixel processing. If game designers realize this and code their shaders accordingly, we could see much more focused processing power dedicated to a single type of problem than on current hardware.

ATI is predicting that developers will use lots of very small triangles in Xbox 360 games. As engines like Epic's Unreal Engine 3 have shown incredible results using pixel shaders and normal maps to augment low geometric detail, we can't tell if ATI is trying to provide the chicken or the egg. In other words, will we see many small triangles on Xbox 360 because console developers are moving in that direction or because that is what will run well on ATI's hardware?

Regardless of the paths that lead to this road, it is obvious that the Xbox 360 will be a geometry power house. Not only are all 3 blocks of 16 shaders able to become vertex shaders, but ATI's GPU will be able to handle twice as many z operations if a z only pass is performed. The same is true of current ATI and NVIDIA hardware, but the fact that a geometry only pass can now make use of shader hardware to perform 48 vector and 48 scalar operations in any given clock cycle while doing twice the z operations is quite intriguing. This could allow some very geometrically complicated scenes.

### Inside the Xenos GPU

As previously mentioned, the 48 shaders will be able to run either vertex or pixel shader programs in any given clock cycle. To clarify, each block of 16 shader units is able to run a shader program thread. These shader units will function on a slightly higher than DX9.0c, but in order to take advantage of the technology, ATI and Microsoft will have to customize the API.

In order to get data into the shader units, textures are read from main memory. The eDRAM of the system is unable to assist with texturing. There are 16 bilinear filtered texture samplers. These units are able to read up to 16 textures per clock cycle. The scheduler will need to take great care to organize threads so that optimal use of the texture units are made. Another consideration to take into account is anisotropic filtering. In order to perform filtering at beyond bilinear levels, the texture will need to be run through the texture unit more than once (until the filtering is finished). If no filtering is required (i.e. if a shader program is simple reading stored data), the vertex fetch units can be used (either with a vertex or a pixel shader program).

In the PC space, we are seeing shifts to more and more complex pixel shaders. Large and larger textures are being used in order to supply data, and some predict that texture processing will eclipse color and z bandwidth in the not so distant future. We will have to see if the console and desktop space continue to diverge in this area.

One of the key aspects of performance for the Xbox 360 will be in how well ATI manages threads on their GPU. With the shift to the unified shader architecture, it is even more imperative to make sure that everything is running at maximum efficiency. We don't have many details on ATI's ability to context switch between vertex and pixel shader programs on hardware, but suffice it to say that ATI cannot afford to have any difficulties in managing threads on any level. As making good use of current pixel shader technology requires swapping out threads on shaders, we expect that this will go fairly well in this department. Thread management is likely one of the most difficult things ATI had to work out to make this hardware feasible.

Those who paid close attention to the amount of eDRAM (10MB) will note that this is not enough memory to store the entire framebuffer for displays larger than standard television with 4xAA enabled. Apparently, ATI will store the front buffer in the UMA area, while the back buffer resides on the eDRAM. In order to manage large displays, the hardware will need to render the back buffer in parts. This indicates that they have implemented some sort of very large grained tiling system (with 2 to 4 tiles). Usually tile based renderers have many more tiles than this, but this is a special case.

Performance of this hardware is a very difficult aspect to assess without testing the system. The potential is there for some nice gains over the current high end desktop part, but it is very difficult to know how easily software engineers will be able to functionally use the hardware before they fully understand it and have programmed for it for a while. Certainly, the learning curve won't be as steep as something like the PlayStation 2 was (DirectX is still the API), but knowing what works and what doesn't will take some time.

### ATI's Modeling Engine

The adaptability of their hardware is something ATI is touting as well. Their Modeling Engine is really a name for a usage model ATI provides using their unified shaders. As each shader unit is more general purpose than current vertex and pixel shaders, ATI has built the hardware to easily allow the execution of general floating point math.

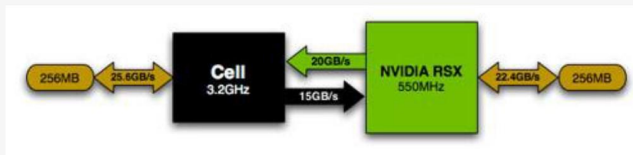
ATI's Modeling Engine concept is made practical through their vertex cache implementation. Data for general purpose floating point computations moves into the vertex cache in high volumes for processing. The implication here is that the vertex cache has enough storage space and bandwidth to accommodate all 48 shader units without starvation for an extended period of use. If the vertex cache were to be used solely for vertex data, it could be much less forgiving and still offer the same performance (considering common vertex processing loads in current and near term games). As we stated previously, pixel processing (for now) is going to be more resource intensive than vertex processing. Making it possible to fill up the shader units with data from the vertex cache (as opposed to the output of vertex shaders), and the capability of the hardware to dump shader output to main memory is what makes ATI's Modeling Engine possible.

But just pasting a name on general purpose floating point math execution doesn't make it useful. Programmers will have to take advantage of it, and ATI has offered a few ideas on different applications for which the Modeling Engine is suited. Global illumination is an intriguing suggestion, as is tone mapping. ATI also indicates that higher order surfaces could be operated on before tessellation, giving programmers the ability to more fluidly manipulate complex objects. It has even been suggested that physics processing could be done on this part. Of course, we can expect that Xbox 360 programmers will not implement physics engines on the Modeling Engine, but it could be interesting in future parts from ATI.

### PlayStation 3's GPU: The NVIDIA RSX

We've mentioned countless times that the PlayStation 3 has the more PC-like GPU out of the two consoles we're talking about here today, and after this week's announcement, you now understand why.

The PlayStation 3's RSX GPU shares the same "parent architecture" as the G70 (GeForce 7800 GTX), much in the same way that the GeForce 6600GT shares the same parent architecture as the GeForce 6800 Ultra. Sony isn't ready to unveil exactly what is different between the RSX and the G70, but based on what's been introduced already, as well as our conversations with NVIDIA, we can gather a few items.



Despite the fact that the RSX comes from the same lineage as the G70, there are a number of changes to the core. The biggest change is that RSX supports rendering to both local and system memory, similar to NVIDIA's Turbo Cache enabled GPUs. Obviously rendering to/from local memory is going to be a lot lower latency than sending a request to the Cell's memory controller, so much of the architecture of the GPU has to be changed in order to accommodate this higher latency access to memory. Buffers and caches have to be made larger to keep the rendering pipelines full despite the higher latency memory access. If the chip is properly designed to hide this latency, then there is generally no performance sacrifice, only an increase in chip size thanks to the use of larger buffers and caches.

The RSX only has 60% of the local memory bandwidth of the G70, so in many cases it will most definitely have to share bandwidth with the CPU's memory bus in order to achieve performance targets.

There is one peculiarity that hasn't exactly been resolved, and that is about transistor counts. Both the G70 and the RSX share the same estimated transistor count, of approximately 300.4 million transistors. The RSX is built on a 90nm process, so in theory NVIDIA would be able to pack more onto the die without increasing chip size at all - but if the transistor counts are identical, that points to more similarity between the two cores than NVIDIA has led us to believe. So is the RSX nothing more than the G70? It's highly unlikely that the GPUs are identical, especially considering that the sheer addition of Turbo Cache to the part would drive up transistor counts quite a bit. So how do we explain that the two GPUs are different, yet have the same transistor count and one is supposed to be more powerful than the other? There are a few possible options.

First and foremost, you have to keep in mind that these are not exact transistor counts - they are estimates. Transistor count is determined by looking at the number of gates in the design, and multiplying that number by the average number of transistors used per gate. So the final transistor count won't be exact, but it will be close enough to reality. Remember that these chips are computer designed and produced, so it's not like someone is counting each and every transistor by hand as they go into the chip.

So it is possible that NVIDIA's estimates are slightly off for the two GPUs, but at approximately 10 million transistors per pixel pipe, it doesn't seem very likely that the RSX will feature more than the 24 pixel rendering pipelines of the GeForce 7800 GTX, yet NVIDIA claims it is more powerful than the GeForce 7800 GTX. But how can that be? There are a couple of options:

The most likely explanation is attributed to nothing more than clock speed. Remember that the RSX, being built on a 90nm process, is supposed to be running at 550MHz - a 28% increase in core clock speed from the 110nm GeForce 7800 GTX. The clock speed increase alone will account for a good boost in GPU performance which would make the RSX "more powerful" than the G70.

There is one other possibility, one that is more far fetched but worth discussing nonetheless. NVIDIA could

offer a chip that featured the same transistor count as the desktop G70, but with significantly more power if the RSX features no vertex shader pipes and instead used that die space to add additional pixel shading hardware.

Remember that the Cell host processor has an array of 7 SPEs that are very well suited for a number of non-branching tasks, including geometry processing. Also keep in mind that current games favor creating realism through more pixel operations rather than creating more geometry, so GPUs aren't very vertex shader bound these days. Then, note that the RSX has a high bandwidth 35GB/s interface between the Cell processor and the GPU itself - definitely enough to place all vertex processing on the Cell processor itself, freeing up the RSX to exclusively handle pixel shader and ROP tasks. If this is indeed the case, then the RSX could very well have more than 24 pipelines and still have a similar transistor count to the G70, but if it isn't, then it is highly unlikely that we'd see a GPU that looked much different than the G70.

The downside to the RSX using the Cell for all vertex processing is pretty significant. Remember that the RSX only has a 22.4GB/s link to its local memory bandwidth, which is less than 60% of the memory bandwidth of the GeForce 7800 GTX. In other words, it needs that additional memory bandwidth from the Cell's memory controller to be able to handle more texture-bound games. If a good portion of the 15GB/s downstream link from the Cell processor is used for bandwidth between the Cell's SPEs and the RSX, the GPU will be texture bandwidth limited in some situations, especially at resolutions as high as 1080p.

This option is much more far fetched of an explanation, but it is possible, only time will tell what the shipping configuration of the RSX will be.

### Will Sony Deliver on 1080p?

Sony appears to have the most forward-looking set of outputs on the PlayStation 3, featuring two HDMI video outputs. There is no explicit support for DVI, but creating a HDMI-to-DVI adapter isn't too hard to do. Microsoft has unfortunately only committed to offering component or VGA outputs for HD resolutions.

Support for 1080p will most likely be over HDMI, which will be an issue down the road. If you're wondering whether or not there is a tangible image quality difference between 1080p and 720p, think about it this way - 1920 x 1080 looks better on a monitor than 1280 x 720, now imagine that blown up to a 36" 60" HDTV - the difference will be noticeable.

At 720p, the G70 is entirely CPU bound in just about every game we've tested, so the RSX should have no problems running at 720p with 4X AA enabled, just like the 360's Xenos GPU. At 1080p, the G70 is still CPU bound in a number of situations, so it is quite possible for RSX to actually run just fine at 1080p which should provide for some excellent image quality.

You must keep one thing in mind however; in order for the RSX to be CPU limited and not texture bandwidth limited at 1080p, the games it is running must be pixel shader bound.

For example, Doom 3 is able to run at 2048 x 1536 at almost 70fps on the 7800 GTX, however Battlefield 2 runs at less than 50 fps. Other games run at higher and lower frame rates; the fact of the matter is that the RSX won't be able to guarantee 1080p at 60 fps in all games, but there should be some where it is possible. The question then becomes, as a developer, do you make things look great at 720p or do you make some sacrifices in order to offer 1080p support.

One thing is for sure, support for two 1080p outputs in spanning mode (3840 x 1080) on the PS3 is highly unrealistic. At that resolution, the RSX would be required to render over 4 megapixels per frame, without a seriously computation bound game it's just not going to happen at 60 fps.

Microsoft's targets for the Xbox 360 are far more down to earth, with 720p and 4X AA being the requirements for all 360 titles. With a 720p target for all games, you can expect all Xbox 360 titles to render (internally) at 1280 x 720. We've already discussed that the 360's GPU architecture will effectively give free 4X AA at this resolution, so there's no reason not to have 4X AA enabled as well.

Most HDTVs will support either 1080i or 720p; those that natively support 720p will simply get a 720p output from the 360 with no additional signal processing. We'd be willing to bet that the game will still render internally at 720p and rely on either the Xbox 360's TV encoder to scale the output to 1080i, or you can rely on your TV to handle the scaling for you. But for all discussion here, you can expect the Xbox 360 GPU to render games at 1280 x 720 with 4X AA enabled.

The support for 4X AA across the board is important, because on a large TV, even 720p is going to exhibit quite a bit of aliasing. But the lack of 1080p support is disturbing, especially considering it is a feature that Sony has been touting quite a bit. The first 1080p displays just hit the market this year, and the vast majority of the installed HDTV user base will only support 720p or 1080i, not 1080p. In the latter half of the Xbox 360 and PS3 life cycle, 1080p displays will be far more common place but it may be one more console generation before we get hardware that is capable of running all games at 1080p at a constant 60 fps.

In the end, Sony's support for 1080p is realistic, but not for all games. For the first half of the console's life, whether or not game developers enable AA will matter more than whether 1080p is supported. By the second half, it's going to be tough to say.

Microsoft's free 4X AA is wonderful and desperately needed, especially on larger TVs, but the lack of 1080p support is disappointing. It is a nice feature to have, even if only a handful of games can take advantage of it, simply because 1080p HDTV owners will always appreciate anything that can take full advantage of their displays. It's not a make or break issue, simply because the majority of games for both platforms will still probably be rendered internally at 720p.

## Storage Devices

Both the PlayStation 3 and the Xbox 360 feature removable 2.5" HDDs as an option for storage; the difference being that the PS3 won't ship with a hard drive, while the Xbox 360 will.

In the original Xbox, developers used the hard drive to cache game data in order to decrease load times and improve responsiveness of games. Compared to the built-in 5X CAV DVD drive in the Xbox, the hard drive offered much faster performance. With the Xbox 360, the performance demands on the hard drive are lessened, the console now ships with a 12X CAV DVD-DL drive. You can expect read performance to more than double over the first DVD drive that shipped with the original Xbox, which obviously decreases the need for a hard drive in the system (but definitely doesn't eliminate it).

This time around, Microsoft has outfitted the 360 with a 20GB removable 2.5" HDD, but its role is slightly different. While developers will still be able to use the drive to cache data if necessary, its role in the system will be more of a storage device for downloaded content. Microsoft is very serious about their Xbox Live push with this next console generation, and they fully expect users to download demos, game content updates and much more to their removable hard drive. The fact that it's removable means that users can carry it around with them to friends' houses to play their content on other 360s.

It is important to note that disc capacity remains unchanged from the original Xbox, the 360 will still only have a maximum capacity of 9GB per disc. Given that the current Xbox titles generally use less than half of this capacity, there's still some room for growth.

Microsoft has also reduced the size of the data that is required to be on each disc by a few hundred megabytes, combine that with the fact that larger game data can be compressed further thanks to more powerful hardware and game developers shouldn't run into capacity limitations on Xbox 360 discs anytime soon.

The PS3 is a bit more forward looking in its storage devices, unfortunately as of now it will not ship with a hard drive. The optical drive of choice in the PS3 will be a Blu-ray player, which originally looked quite promising but now is not as big of a feature as it once was.

The two main competitors for the DVD video successor are the HD-DVD and Blu-ray standards. Around the announcement of the PS3 at E3, there was a lot of discussion going on surrounding an attempt to unify the HD-DVD and Blu-ray standards, which would obviously make the PS3 Blu-ray support a huge selling feature. It would mean that next year you would be able to buy a console, generally estimated to be priced around the \$400 mark, that could double as a HD-DVD/Blu-ray player. Given that standalone HD-DVD/Blu-ray players are expected to be priced no less than \$500, it would definitely increase the adoption rate of the PS3. However, talks between unifying the two standards appear to have broken down without any hope for resolution meaning that there will be two competing standards for the next-generation DVD format. As such, until unified HD-DVD/Blu-ray players are produced, the PS3 won't have as big of an advantage in this regard as once thought. It may, however, tilt the balance in favor of Blu-ray as the appropriate next-generation disc standard if enough units are sold.

When we first disassembled the original Xbox, we noticed that it basically featured a PC DVD drive. From what we can tell, the Xbox 360 will also use a fairly standard dual layer DVD drive. As such, it would not be totally unfeasible for Microsoft to, later on, outfit the Xbox 360 with a HD-DVD or Blu-ray drive, once a true standard is agreed upon.

The one advantage that Sony does have is that developers can use BD-ROM (Blu-ray) discs for their games, while if MS introduces Blu-ray or HD-DVD support later on it will be strictly as a video player (game developers won't offer content for only owners of Blu-ray/HD-DVD Xbox 360 versions). The advantage is quite tangible in that PS3 developers will be able to store a minimum of 23.3GB of data on a single disc, which could mean that they could use uncompressed video and game content, freeing up the CPU to handle other tasks instead of dealing with decompression on the fly. Of course, Blu-ray media will cost more than standard DVD discs, but over the life of the PS3 that cost will go down as production increases.

## System Costs

One thing that a surprising number of people seem to overlook is the idea that consoles are built to take a loss on the hardware itself. If the Xbox 360 retails for \$299, it may very well cost Microsoft \$399, or even more. This has been the way consoles have been manufactured for quite some time now, and it has not changed with the latest generation of consoles.

However, given the very high system costs of the original Xbox, it isn't surprising to see that Microsoft is quite

concerned with keeping costs down to a minimum this time around. There are a number of decisions that Microsoft has made in order to limit their loss on the 360 hardware.

First and foremost, Microsoft owns the iP in the Xbox 360 and thus they can handle manufacturing on their own without having to re-negotiate contracts with ATI or IBM. It remains to be seen how much of a money saver this will be for Microsoft, but it does present itself as a departure from the way things were done the first time around for the folks at Redmond.

Assuming Xenon is nothing more than 3 PPEs put on the same die coupled with twice the L2 cache, it looks like Xenon is a smaller chip than Cell.

The Xenos GPU features a higher transistor count than the RSX (332M vs. 300.4M), but a lower clock speed.

Microsoft didn't skimp much on the CPU or GPU hardware, which isn't surprising, but it is in the auxiliary hardware that the console ends up being cheaper in. The best way to understand the areas that Microsoft didn't spend money in, is to look at the areas that Sony did spend money in.

The Xbox 360 is using a tried and true 12X dual layer DVD drive, probably very similar to what you can buy for the PC today. A very popular drive format with mass produced internals is a sure fire way to keep costs down. Sony's solution? A very expensive, not yet in production, Blu-ray drive. As we mentioned earlier, the first Blu-ray players are expected to retail for more than \$500. The PlayStation 3 isn't going to be successful as a \$800 console, so we'd expect its MSRP to be less than \$500, meaning that Sony will have to absorb a lot of the cost (initially) of including a Blu-ray player, until production picks up.

Both the Xbox 360 and the PS3 feature wireless controller support, although Sony supports a maximum of 7 Bluetooth controllers compared to Microsoft's 4 2.4GHz RF controllers.

The PS3 also ships with built in 802.11b/g and three Gigabit Ethernet ports so the system can act as a Gigabit router right out of the box. Adding wireless support isn't a huge deal, but the physical layer as well as the antenna do drive costs up a bit. The same goes for getting controllers to drive the three Gige ports on the unit.

Sony also offers built in support for more USB 2.0 ports (6 vs 4), media card slots (Memory Stick, SD and Compact Flash) where the 360 has none and two HDMI outputs where the 360 only offers component. Again, not major features but they are nice to have, and do contribute to the overall price of the system.

The one difference that favors Microsoft however is the inclusion of a 2.5" HDD with the Xbox 360 console; Sony's hard drive will be optional and won't ship with the system.

In the end it seems that Microsoft was more focused on spending money where it counts (e.g. CPU, GPU, HDD) and skimmed on areas that would have otherwise completed the package (e.g. more USB ports, built in wireless, router functionality, flash card readers, HDMI support in the box, etc...). Whereas Sony appears to have just spent money everywhere, but balanced things out by shipping with no hard drive.

## Final Words

Game consoles have always been different, architecturally. The PlayStation 2 was very different from the original Xbox, and thus it is no surprise to see that the two platforms continue to be quite different this time around.

Given what we've discussed thus far, there are a number of conclusions we can draw:

The most important thing to keep in mind is that the revolution in physics engines and collision detection isn't going to happen over night. The first games for both consoles will, for all intents and purposes, be single threaded titles. More adventurous developers may even split up execution into two concurrent threads, but for the most part don't expect to see a dramatic change in the quality and reality of the physics simulation of the first titles, especially when compared to titles like Doom 3 and Half Life 2.

However, a change is coming and by the end of next year multi-threaded game engines should be commonplace on both consoles and PCs, which will hopefully lead to much more entertaining experiences. The approach to that change will be different according to the platform; without a doubt, developers will have their work cut out for them.

The transition to multi-threaded development alone will increase development time 2 or 3 fold. Not to mention that the approach to architecting game engines will differ whether you are porting to the Xbox 360 or the PlayStation 3. The Xbox 360 is clearly going to be the easier of the two to develop for once a game engine is multi-threaded, just because of the general purpose nature of its hardware. That being said, it won't be impossible to get the same level of performance out of the PS3, it will just take more work. In fact, specialized hardware can be significantly faster than general purpose hardware at certain tasks, giving the PS3 the potential to outperform the Xbox 360 in CPU tasks. It has yet to be seen how much work is required to truly exploit that potential however, and it will definitely be a while before we can truly answer that question.

Cell's on-die memory controller is a blessing for game performance; it most definitely will keep the PPE fed far better than the Xbox 360's external memory controller. Even the cache size advantage of the 360 won't be

able to offset the reduction in memory latency thanks to an on-die memory controller.

The on-die memory controller is not all an advantage however, a big part of its inclusion is out of necessity. Remembering back to our discussion about the SPEs as being in-order with no cache, threads run on these processors only have access to 256KB of local memory, which is reasonable for a cache, but not much in the way of memory. So these SPEs will depend on having low latency access to memory in order to keep their pipeline filled and actually contribute any useful performance to the system.

At the end of day 1, when running mostly single threaded code, the performance difference from a CPU standpoint between the Xbox 360's Xenon and the PS3's Cell processor is basically a wash. The 360 has more cache, while the Cell has a lower latency path to main memory. In the end, the first generation or two of games will mainly be a GPU battle between the two consoles, and both will offer significant improvements over what we have with current consoles.

Graphics-wise the 360's Xenos GPU and the PS3's RSX are fairly different in implementation, but may end up being very similar in performance. Treating Xenos as a 24-pipe R420, it could be quite competitive with a 24-pipe RSX despite a lower clock speed. The unified shader architecture of the Xenos GPU will offer an advantage in the majority of games today where we aren't very geometry limited. The free 4X AA support offered by Xenos is also extremely useful in a console, especially when hooked up to a large TV.

If the PS3's RSX isn't much more than a higher clocked G70 then at least we have a good idea of its performance. NVIDIA has mentioned that by the time the RSX launches we will have a faster GPU on the PC, which leads us to believe that the performance advantages of the RSX are mostly clock speed related. At 550MHz, the RSX GPU should have no problems handling both 720p and 1080p resolutions, although the latter won't be possible in all games, mainly those that are more texture bandwidth bound. We do think it was a mistake for Microsoft not to support 1080p, even if only supported by a handful of games/developers. At the same time, by not imposing strict AA implementation regulations like Microsoft, Sony does open themselves up to having some PS3 games plagued by jaggies despite the power of the console. Given the amount of power in both of these consoles, we truly hope that their introduction will mark the end of aliasing in console games, but some how we have a feeling it won't. Aliasing has plagued console games for too long for it to just disappear, that has to be too good to be true.

With at least 5 months before the official release of Microsoft's Xbox 360, and a number of still unanswered questions about the PlayStation 3, there is surely much more to discuss in the future. The true nature of NVIDIA's RSX GPU, the real world programming model for Cell, even final hardware details for both consoles has yet to be fully confirmed. As we come across more information we will analyze and dissect it, but until then we hope you've gained more of an understanding of these consoles through this article.