Term Paper – Graphics Processing Unit

Abdullah Biary

Computer System Architecture

George Washington University

23/03/2015

Table of Contents

1. Introduction 3

2. History of Graphics Card 4

2.1 During 1980’s 4

2.2 During 1990’s 4

2.3 During 2000’s 5

2.4 From 2010 & Beyond 6

3. Features of Graphics Cards 7

3.1 Adaptive Vertical Sync 7

3.2 Anti-aliasing Capability 10

3.3 Multi Sampled Anti Aliasing 12

3.4 SLI Technology 13

3.5 3D Vision Ready 15

3.6 PhysX Technology 16

4. Comparison of Graphics Card 18

5. 0 Development of Gaming Console based on Graphics Processor Units 20

6. Conclusion 23

7. References 24

**Abstract**

Modern day graphics cards are more advanced and give real feel of graphics. At present we can see 3-D view of everything using such graphics cards. Graphics card development leaders like AMD and NVIDIA are giving too much emphasis on development of this technology. They are the leaders around the world but the NIVIDIA graphics cards are more popular than any other cards. The graphics card is able to power any computer graphics at extremely good frame rates and has really high clock speed. However not everyone would have a lot to benefits for this graphics card unless they are involved in computer gaming or using applications that have a high demand on the graphics processor. Most people would be sufficient with a graphics card that has way less processing power.

# 1. Introduction

Graphic processor unit helps in developing the 3 – D graphics on computers. Most of the graphic development software and video coders utilize the functions of GPU’s. Higher the performance of the GPU higher will be the experience while viewing the graphics. GPU’s have evolved from 1990’s to present times. Modern day GPU’s are more complex and advanced than the previous years.

# 2. History of Graphics Card

Since the development of graphics card has been started. I am representing the decade by decade development of Graphics cards.

## 2.1 During 1980’s

During 1980’s they were like tike TTL boards and relied on CPU. The word GPU is not introduced during 80’s for graphics processor as they have to rely on CPU’s for performance. IBM Professional Graphics Controller was the first video card developed for 2D/3D viewing.



GPU’s Architecture in 1980’s

## 2.2 During 1990’s

During this stage Nvidia comes in to the market and launched some new graphics card which is independent from CPU and uses some complex structure. Total two generation graphics cards are introduced during this phase and the technology development is on its high during this phase.



GeForce 3 Architecture during 1990’s developed by Nvidia

## 2.3 During 2000’s

In 2001, NVIDIA released the GeForce 4 which gave programmers the ability to program parts of the previously non-programmable pipeline [1]. Instead of sending all the graphics description data to the GPU and have it simply flow through the fixed pipeline



GeForce 6 (VI Generation GPU) by Nvidia

## 2.4 From 2010 & Beyond

The GPU hardware evolution thus far has gone from an extremely specific, single core, fixed function hardware pipeline implementation just for graphics rendering, to a set of highly parallel and programmable cores for more general purpose computation. Now, the architecture of many-core GPUs are starting to look more and more like multi-core, general purpose CPUs.



AMD Fusion Architecture

# 3. Features of Graphics Cards

## 3.1 Adaptive Vertical Sync

“Nothing is more distracting than frame rate stuttering and screen tearing. The first tends to occur when frame rates are low, the second when frame rates are high. Adaptive V-Sync is enabled to eliminate tearing, at low frame rates; it’s disabled to minimize stuttering. It gets rid of distractions so you can get on with the gaming. [3]”

First of all we have to understand the importance of video synchronization. When a video is displayed on a monitor we have to ensure that the scanning is synchronized with the incoming signal, so that the top of the screen when the picture runs up and down the screen or tears away from the left of the screen. Synchronization is achieved by including synchronization pulses in the horizontal and vertical retrace periods [4]. The basis for Adaptive Vertical Sync is Vertical Synchronization. Vertical Synchronization eliminates tearing by timing frame buffer fills to coincide with the vertical blanket interval, thus ensuring that only whole frames are seen on-screen [5]. The vertical sync pulses are made by prolonging the length of HSYNC pulses through almost the entire length of the scan line [5] basically The V-sync feature prevents drawing from getting more than one frame ahead.

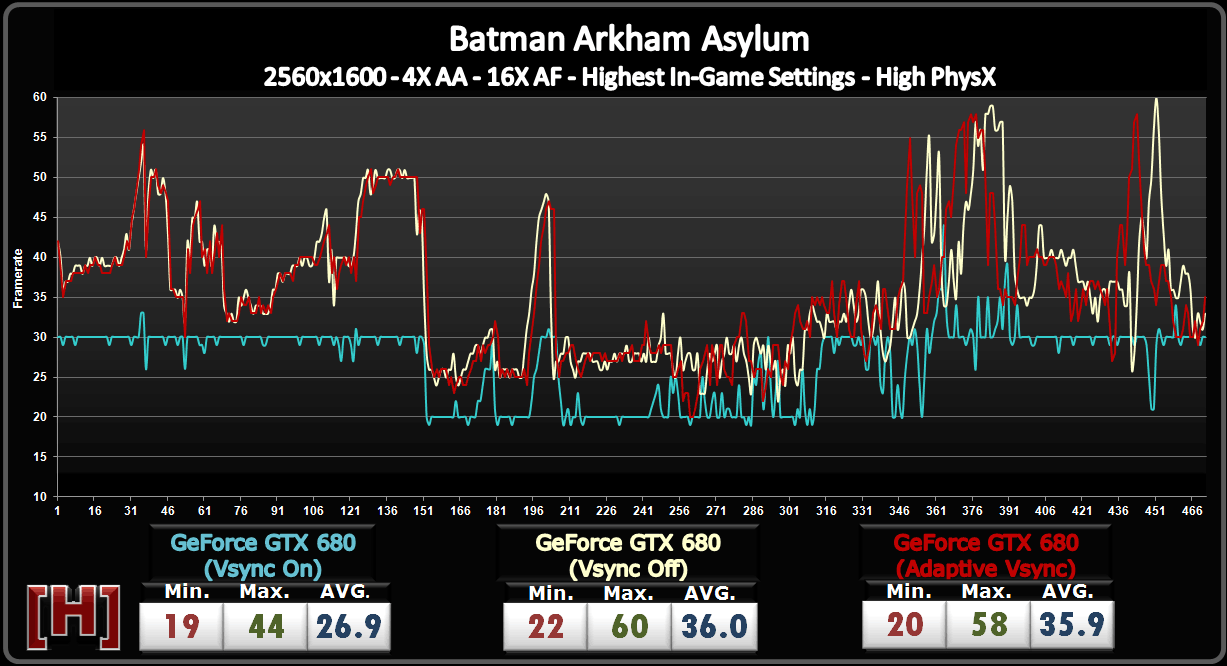
Adaptive VSync is an advancement of the VYSnc On and VYSnc Off technology. With the Vyscn On technology if the frame rate of the game is higher than the refresh rate of the display, the frame rate of the game is brought down to the refresh rate of the display to prevent tearing. The draw back with VSync turned on is that the frame rate can step down by a huge amount and fluctuations in the frame rate are perceivable and thus distracting to the gamer [6].To prevent this we can use VSync Off but has the side effects of screen tearing because when the frame rate exceeds the refresh rate of your display screen tearing becomes more visible [5].



VSync Difference between ATI and Nvidia

The NVIDIA Adaptive Vertical Sync enhances the game play experience [6]. With adaptive VSync if the frame rate of the game exceeds the refresh rate of the display the frame rate of the game is brought level to the refresh rate which prevents tearing in the display [6]. Also if the frame rate of the game is lower than the refresh rate of the display the frame rate of the game is kept at its original frame rate. Therefore the adaptive VSync behaves like the VSync turned on when the frame rate is higher than the refresh rate of the display and acts as the VSync Off when the frame rate is lower than the refresh rate of the display.

The below graph shows the, performance of the Adaptive VSync with the Vysnc On and VSync Off. The refresh rate for the display used is 60FPS [6].



“Adaptive VSync has the same performance profile as VSync Off in this scenario. What’s the great is that there is no tearing , and the frame rate is much better, not capped at 30 FPS or 20 FPS as it was with VSync turned on. [6]”

## 3.2 Anti-aliasing Capability

“Anti-aliasing smoothes out jagged edges but can be demanding on frame rates. FXAA is a new ant aliasing technology that produces beautiful smooth lines with minimal performance impact. And with Kepler based GPUs, you’ll be able to enable FXAA in hundreds of game titles through the NVIDIA Control Panel. Ant aliasing is used when the resolutions of the display screen is not sharp enough to display an image so lines in the display appear jagged. With ant aliasing the jagged lines are smoothed out.

FXAA and TXAA are the two ant aliasing techniques that NVidia use to perform ant aliasing. These techniques promise to be one of the best if not the best ant aliasing techniques used in top end game titles of today. FXAA is an acronym meaning Fast *Approximate Anti-Aliasing.* Below is an image of FXAA ant aliasing used in a game and when no ant aliasing is used in a game?

**

Anti aliasing utilization in GPU’s

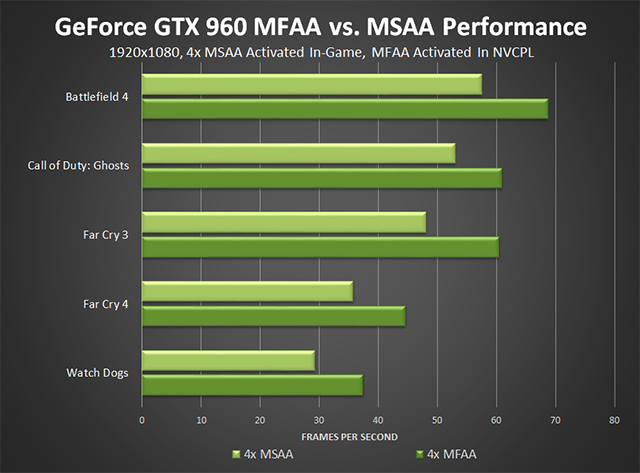
The effects of ant aliasing are clear that ant aliasing significantly enhances the overall look of the display and makes the game more realistic in its graphics. *“* A line has infinite resolution, but our digital displays do not. So when we “snap” a line to the pixel grid on our display, we can compensate by imagineering partial pixels along the line, pretending we have a much higher resolution display than we actually do.[7]*”*

*“*Anti-aliasing produces a superior image by using grey pixels to simulate partial pixels along the edges of a line. It is a hack, but as hacks go, it’s pretty darn effective. [7]*”*

TXAA combines MSAA with temporal filtering and post processing to make an efficient ant aliasing tool. “NVIDIA has designed an even higher-quality anti-aliasing mode called TXAA, which is designed for direct integration into game engines. Combining the raw power of MSAA with sophisticated resolve filters, similar to those employed in CG films, TXAA produces a smoother image, far in advance comparable technique. In addition, TXAA can also jitter sample the entire scene between frames to reduce between frames to reduce shimmering, known technically as temporal aliasing [8].” With TXAA ant aliasing we can combat shimmering to enhance the overall display output from a computer game where it is used.

## 3.3 Multi Sampled Anti Aliasing

Multi-Sampled Anti-Aliasing (MSAA) is found in most modern day graphics cards. The graphics card renders to a surface that is larger than the final image, but in shading each cluster of samples the pixel shadier is run only once. We save a lot of fill rate, but we still use a lot of memory bandwidth. MSAA does not work for a deferred renderer because lighting decisions are made after the MSAA is down-sized to its final image size[7].



MSAA Performance in GTX 960

In other words, multi-sample anti aliasing is a brute-force technique for achieving order invariant single-pass anti aliasing. Single-pass anti aliasing is a multi-pass accumulation buffer anti aliasing using an accumulation buffer is order invariant, and produces high-quality images in 10 to 20 passes [9].

The advantage of MSAA is that it reduces the cost of computing the ﬁnal pixel color by decoupling visibility determination from shading and by limiting per-pixel shading to the rate of one sample per primitive while still evaluating visibility at full rate [9].

## 3.4 SLI Technology

“SLI features an intelligent communication protocol embedded in the GPU, a high-speed digital interface to facilitate data flow between the two graphics cards, and a complete software suite providing dynamic load balancing, advanced rendering, and compositing to ensure maximum compatibility and performance in today’s latest games.[10]” First of all, SLI stands for Scalable Link Interface. Basically the NVidia SLI Technology is the technology that allows you to hook up a second graphics card to your motherboard to enhance the graphics processing power.

It is possible for a computer to use multiple graphic processing units because computational tasks often contain exploitable concurrency that enables different parts of the problem to be solved simultaneously by multiple closely coupled processors which is known as parallel programming. Using multiple GPU’s enhances the graphics performance because usually high-speed communication channels exist between these processors and fast synchronization is also possible [10]. This also allows less time for processing and allows high data-intensive problems to be solved by optimizing the resources of the graphic processor units [11].

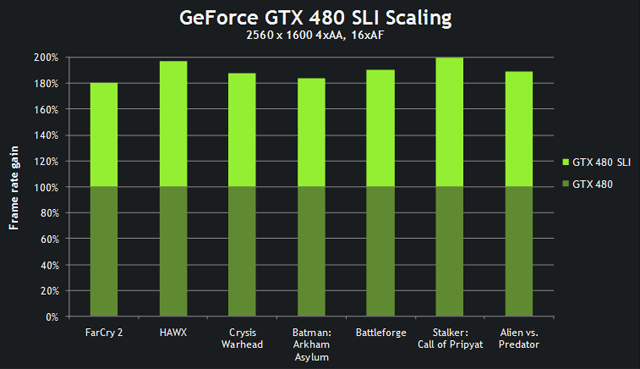


GPU with SLI Technology

The reason that multiple GPUs can be used to solve problems is because they are based on programming models such as DirectX and OpenGL that use single instruction stream multiple thread models of computation exploited by hardware multi-threading and single instruction multiple data (SIMD) cores[12].

The drawback of the SLI Technology is that SLI technology may not be supported by some software. One advantage of SLI is that with NVidia graphics cards using SLI Technology could prove to be cost effective. When the two GTX 670 graphics cards are used in one computer using the SLI technology equals the performance levels of the GTX 690 graphics card [13]. The GTX 670 costs 200 pounds whiles the GTX 690 costs over 800 pounds [13]. Therefore buying using two GTX 670 graphics cards is half the cost of the GTX 690 whiles having almost the same performance level as the GTX 690.

Performance of the SLI technology in the frame rates is shown below [9]:



SLI Scaling in GeForce GTX 480

## 3.5 3D Vision Ready

NVIDI 3D Vision Ready is NVIDIA’s technology used in displaying 3D displays of images and other graphics. It is the top of the line technology. It’s not only used to create the 3D experience but is used to make 2D display images look more realistic by rendering 3D images in a 2D display. A few decades 3D display was a thing only seen in sci-fi movies now the technology has been developed for us to watch movies and other video sources in 3-D. Now in order to view stereoscopic 3D images you would first have to have the NVIDIA graphics card that supports 3D stereoscopic display. You would also have to have the 3D Vision glasses from NVIDIA and also have a monitor that supports 3-D display.

We are able to view stereoscopic 3-D because of two phenomena. The first phenomena is where your eyes focus on the 3-D image, specifically they focus at what depth our individual eyes are adjusted to see, just like a camera lens focuses. The other phenomena are convergence, which is basically the phenomenon where your eyes cross or uncross in order to bring two similar features in an image together. This focus and convergence conflict causes visual fatigue. It can be minimized by linear mapping the depth budget to the comfort zone, the range that we can see stereoscopic images comfortably [14]. In order to have comfortable viewing of stereoscopic 3D images consistency between left and right. An image is consistent when objects appearing in both views match visually. If the color, texture, or rendering of an object differs between views, then the pair is inconsistent and the image is less comfortable to view [15]. In order to improve the display the following techniques are used: visual fatigue reduction and depth perception enhancement.

Cropping and warping techniques are used to employ depth perception enhancement. The algorithm for depth perception enhancement has three steps: preprocessing, cropping, and warping. In the preprocessing stage, we obtain the sparse correspondences, saliency maps, and the segmentation from the stereoscopic image pair. In the cropping stage, we implement the black frame and avoid the stereo window violation by the correspondences and segmentation. In the warping stage, we define three constraints for the energy function to adjust the disparities of the stereoscopic image pairs [16]

## 3.6 PhysX Technology

PhysX technology is developed by NVidia. “NVIDIA® PhysX® technology adds an element of realism never before seen in gaming. With an NVIDIA® GeForce® GPU in your PC, experience dynamic PhysX® effects like blazing explosions, reactive debris, realistic water, and lifelike characters[17].” “NVIDIA® PhysX® is a powerful physics engine enabling real-time physics in leading edge PC games. PhysX software is widely adopted by over 150 games and is used by more than 10,000 developers. PhysX is optimized for hardware acceleration by massively parallel processors. GeForce GPUs with PhysX provide an exponential increase in physics processing power taking gaming physics to the next level [17].” NVIDEA PhysX Technology is basically a physics engine that is able to perform many physical calculations in real-time in order to display images that look real; objects that react to the environment as it would in the real world. I myself have had first-hand experience with PhysX technology used by EA Games. In FIFA 13, a soccer computer game PhysX is used to make the players fall to the ground in numerous different ways because they fall depending on the physical environment and the forces that are impacted on the players in the game.

The image below shows how NVidiaPhysX is used to make the paper on the ground reacts realistically with the environment (when it comes in contact with the person):



PhysX makes it possible to make objects in a scene appear realistic when they come in contact with each other. The way PhysX works to do this is by using Collision detection. CD uses geometrical approaches to identify bodies that are in contact and appropriate contact points. A space in CD contains geometric objects that represent the outline of rigid bodies. Spaces are used to accelerate collision detection by allowing the removal of certain object pairs that would result in useless tests [18].

The PhysX engine can be used to dynamically simulate objects, collisions with the terrain and the static geometry of a scene. PhysX is also used to simulate several fuzzy phenomena such as fire, water, fog and rain that are not simply done using traditional rendering techniques. PhysX makes the particles from phenomena such as fire and rain possible to be able to look realistic enhancing the visual appearance by the user [19].

# 4. Comparison of Graphics Card

There are various developers of Graphics card but the leaders are AMD and NVIDIA. Gamers will benefit from this invention also Hyper transport allows networking communications to increase, this doesn't mean your internet connection will go faster but it would receive all the date much faster causing less clutter and less lag. The AMD Athlon 64 FX processor includes a 16-bit Hyper Transport technology interface capable of operating up to 1600 mega-transfers per second with a resulting bandwidth of up to 6.4 Gbytes. The AMD Athlon 64 FX processor supports HyperTransport technology fast clocking mode. In gaming there are several processes going on and on a processor like Intel something that is going 1 at a time would cause too much clutter and more latency.

This means this allows the Video Card to render all the graphics much faster. Even though the graphic cards usually have their own GPU which is almost like a Computer Processor they are dedicated in creating the images. The computer processor would distribute everything the video card is processing at a much faster rate, which allows smoother game play. Another factor is the motherboard AMD is also releasing a motherboard which has 300 millimeter fabrication. This also allows everything on the board to communicate much faster even though it gets smaller that means it will travel much faster, but smaller doesn't mean its better it all depends on how it is built. The Intel Pentium is referred as the dominant corporation in processing with its high speed but because of their early releases before AMD and advertisements they have through the media they have the ability to brainwash people. Another competitor is Motorola.

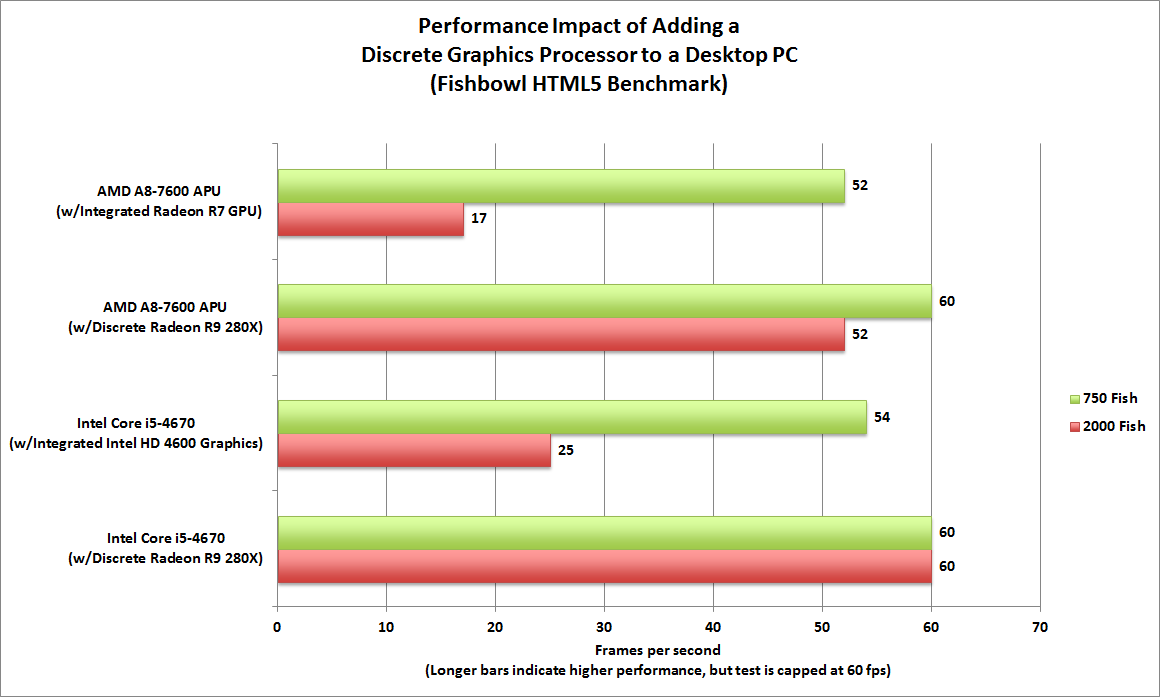
Motorola makes a processor for Apple computers who are famous for their G4 processor with L3 cache and .13 micron similar to the AMD then their G5 processor with L2 Cache but with a faster bus with .13 micron. Apple is known as a competitor to AMD because how they advertise to the media. They are actually quite fast, faster than the Intel processors. I would say AMD is almost similar to the G4 and G5 processors but with their inferior designs they lack the ability to run anything with direct X. Direct X is what allows PC users to play video games with beautiful graphics which AMD 64 contributes tremendously, this doesn't mean the G4 or G5 has horrible video it means they can’t run as much beautiful games as the pc but when it comes to video editing the apple beats the pc competitor by a lot. That's what they are primarily used for and secure servers but now since the Opteron has been released, AMD will probably take over in business servers.

This invention of .13 doesn't just contribute to the processors but Graphic processors to. ATI and NVIDIA competitors in the video card technology need faster GPU's and since they are primarily designed for the pc for AMD to invent this design it opens the door for ATI and NVIDIA to design .13 Gpus which run beautifully with the Opteron. Another major factor with the opteron is its low heat output because its small design and .13 micron it produces less head which allows businesses who have a lot or rack servers and the room don't have to be freezing temperatures for the servers to be running smoothly. Continuing the TITAN legacy of supercomputer-inspired performance, NVIDIA CEO Jen-Hsun Huang unveiled the GeForce GTX TITAN Z today at our annual GPU Technology Conference. Built around two Kepler GPUs and 12GB of dedicated frame buffer memory, with two GK110 chips, TITAN Z is powered by a total of 5,760 processing cores, or 2,880 cores per GPU.

Unlike traditional dual-GPU cards, Titan Z's twin GPUs is tuned to run at the same clock speed, and with dynamic power balancing. So neither GPU creates a performance bottleneck. And that performance is delivered in a card that is cool and quiet, rather than hot and loud. Low-profile components and ducted base plate channels minimize turbulence and improve acoustic quality. So if you want to build the ultimate ultra-high definition gaming rig that can harness the power of quad GPUs working in tandem

# 5. 0 Development of Gaming Console based on Graphics Processor Units

Sony launched their first console in 1995, and soon after “the PlayStation became the most popular, best-selling 32-bit system” . This lightweight, compact, gray console that stood out from the stereotypical manufactured black configurations was groundbreaking and extremely well designed. The top of the unit had two buttons that provided the power and open options, while the front of this machine provided a place for a pair of controllers with matching memory cards to be plugged in. The controllers matched the innovative design of the machine, they featured elongated handle grips that fit well into hands and four shoulders buttons, neither of which had ever been seen. The small, easy to use memory cards fit into slots right above each controller port. Sony also focused on the internal aspect of the PlayStation; optimizing this system to handle three-dimensional game play. Eventually, this best-selling console would establish Sony as the market leader in the business allowing success to carry over into other endeavors.



While the Atari 2600 was the first system to use cartridges, the Nintendo 64 was the last video game console to use them. Although the N64 release was a "hyped, long-anticipated moment", it was often delayed nevertheless, the system delivered state-of-the-art 3-D graphics and an unusual controller. The design allowed for a power button, reset button, cartridge slot, and expansion port to be located on the top of the machine, but the front edge contained enough space to house four controller ports. Memory cards were of no use with this console. The Nintendo 64 allowed for games to be saved directly to the console, there was no need to purchase memory cards. Analog thumb sticks are standard on all game controllers these days, but in 1996 it was revolutionary for the oversized controllers. This machine was ideal for multi-player tournaments. But, even with a faster load time and the ability to generate high resolution 3-D environments, the Nintendo 64 never flourished as well as the PlayStation did. While the system definitely had its moments, the high prices and lack of games proved for the console slowly lead the N64 into decline.

Microsoft’s initial entrance into the video game console market was marked in 2001 with an impressive machine, the Xbox. The PC technology utilized in the production of the Xbox caused many to compare it to a high-end PC that was only a fraction of the price. This ten pound rectangle contained a 10 GB hard drive, provided space for four controller ports, power and open buttons, and an electrically powered disc tray but also supported high definition videos. Xbox games revealed faster load times and there was no need for memory cards. At first the Xbox controllers that were manufactured were extra-large with rounded buttons. But soon after receiving complaints, Microsoft switched to the “Controller-S” which was much less bulky. As prices began to drop on each console, consumers quickly opted for Xbox versions titles partly because Microsoft offered an online platform that could not be rivaled. While the Xbox was meant to compete with the PlayStation 2, the sales never quite matched up. Yet Microsoft managed to position itself into second place making itself a competitor in the video game industry.

In the 1990s, the storage used for games shifted from cartridges to compact discs because of the increasing memory needs of progressing games. This evolution paved the way from the Atari 2600 to the PlayStation. While Nintendo used cartridges for its system, Sony changed everything by putting their games onto discs. It took Nintendo an additional five years to switch discs, but that year Microsoft brought more competition to the table with their Xbox. And now we have arrived at where we are today, with three major companies in the video game console industry. But with the continuous technological advancements society strives for, will we be seeing another company soon? Or will the top three companies continue to struggle to outdo each other?

# 6. Conclusion

The graphics card has extremely outstanding graphics capabilities. It’s use of Vertical Synchronization, SLI capability to use two or more graphics cards, top end anti-aliasing techniques, being able to have 3D display and to have a physics engine that is able to create realistic virtual scenery make the graphic card a really cool graphics card. The graphics card is able to power any computer graphics at extremely good frame rates and has really high clock speed. However not everyone would have a lot to benefits for this graphics card unless they are involved in computer gaming or using applications that have a high demand on the graphics processor. Most people would be sufficient with a graphics card that has way less processing power.

# 7. References

[1] NVIDIA, 2012, GeForce. http://www.nvidia.com/page/geforce6.html

[2] NVIDIA, 2012, GeForce GTX 670. http://www.geforce.com/hardware/desktop-gpus/geforce gtx-670

[3] NVIDIA, 2012, GeForce GTX 670 Features. http://www.geforce.com/hardware/desktop-gpus/geforce-gtx-670/features

[4] Greg Lehey. 1995. Setting up X11: A no-tears guide to XFree86 configuration. Linux Journal Volume 1995 Issue Article No. 4

[5] WIKIPEDIA, 2012, Analog Television.http://en.wikipedia.org/wiki/Analog\_television

[6] HARDOCP, 2012, NVidiaAdaptive Sync Technology Review. http://www.hardocp.com/article/2012/04/16/nvidia\_adaptive\_vsync\_technology\_review

[7] Kotaku, 2012, What is FXAA?http://www.kotaku.com.au/2011/12/what-is-fxaa/

[7] GeForce, 2012, TXAA-Anti-Aliasing Overview. http://www.geforce.com/landing-page/txaa

GREINER, R. 1999. Explanation-based learning. In The Encyclopedia of Cognitive Science, R. WILSON AND F. KEIL, Eds. MIT Press, Cambridge, MA, 301-303.

[9] Marco Salvi.2012.Surface based anti-aliasing. Proceeding I3D '12 Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, New York, NY, USA. 159-164

[10] GeForce, 2012, SLI Technology. http://www.geforce.com/hardware/technology/sli/technology

[11] Kurt Akeley. 1993. Reality Engine graphics. SIGGRAPH '93 Proceedings of the 20th annual conference on Computer graphics and interactive techniques, New York, USA 1993, 109-116

[12] Gregory Diamos.2011.SIMD re-convergence at thread frontiers. Proceedings of the 44th Annual IEEE/ACM International Symposium on Microarchitecture, New York, NY, USA, 2011, 477-488

[13] TechRadat, 2012, Top 11 Graphics Cards. http://www.techradar.com/us/news/computing-components/graphics-cards/11-best-graphics-cards-in-the-world-today-654141/page:2#articleContent

[14] Eden Ashley Umble. 2006. Making it real: the future of stereoscopic 3D film technology. Volume 40 Issue 1, May 2006. Article No. 3

[15] KOOI, F. L., AND TOET, A. 2004. Visual comfort of binocular and 3D displays. Displays 25, 23, 99 – 108.

[16] Hong-Shang Lin. 2011.Stereoscopic 3D experience optimization using cropping and warping. ProceedingSA '11 SIGGRAPH Asia 2011 SketchesArticle No. 40[17] NVidia, 2012, FAQs. http://www.nvidia.in/object/physx\_faq\_in.html

[18] Thomas Y. Yeh. 2006. Enabling real-time physics simulation in future interactive entertainment. Proceedings from Sandbox '06 Proceedings of the 2006 ACM SIGGRAPH symposium on Videogames, New York, NY, USA 2006, 71-81

[19] Salvatore A. Catanese. 2011. Rendering of 3D dynamic virtual environments.SIMUTools '11 Proceedings of the 4th International ICST Conference on Simulation Tools and Techniques, 351-358