GRAPHICS PROCESSING UNIT

SEMINAR REPORT

IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF DEGREE OF

BACHELOR OF COMPUTER APPLICATIONS

MAHATMAGANDHI UNIVERSITY

KOTTAYAM

SUBMITTED BY:

AKSHAY V SANTHOSH 200021093597

UNDER THE GUIDANCE OF

Mrs. Shajeni Justin

2020-2023

DEPARTMENT OF COMPUTER SCIENCE

SIENA COLLEGE OF PROFESSIONAL STUDIES

(Affiliated to Mahatma Gandhi University)

Edakochi– 682010

SIENA COLLEGE OF PROFESSIONAL STUDIES

Edakochi – 682010

(Affiliated to Mahatma Gandhi University, Kottayam)

DEPARTMENT OF COMPUTER SCIENCE

CERTIFICATE

*This is to certify that the seminar titled **"Graphics Processing Unit"** submitted in partial fulfillment of the requirements for the award of the **Degree in Bachelor of Computer Applications** is a bonafide report of work done by **AKSHAY V SANTHOSH 200021093597** during the year **2023**.*

Mrs.Shajeni Justin
Internal Guide

Mrs. Nancy Jismon
Head of the Department

Examiners:

Department Seal

1. External Examiner:

2. Internal Examiner:

DECLARATION

I hereby declare that this seminar entitled ***"Graphics Processing Unit"*** submitted to Department of Computer Science, Siena College of Professional Studies, Edakochi affiliated to Mahatma Gandhi University under the guidance of Mrs. Shajeni Justin, Assistant Professor, Department of Computer Science, and the empirical findings of the report are based on the studies and are not formed from any other Report. We declare that this report has not been submitted to any other Institution for the award of any Fellowship, Degree or Diploma.

Place: Edakochi
Date:

AKSHAY V SANTHOSH
200021093597

ACKNOWLEDGEMENT

If words are considered as symbols of approval and tokens of acknowledgements, then the words play the heralding of expressing our gratitude.

First of all, we thank the almighty for giving me the courage and enthusiasm to complete this seminar.

We wish to convey our deep sense of gratitude to Dr. M A Solomon, Principal, Siena College of Professional Studies, Edakochi, for providing us the facilities to conduct the Seminar.

We are deeply grateful to Mrs. Nancy Jismon, Head, Department of Computer Science, Siena College of Professional Studies, for granting us this opportunity to perform the seminar. Our heartily thanks to our internal guide Mrs. Shajeni Justin, Department of Computer Science, for the timely and intelligent guidance throughout for collecting materials for the seminar.

Finally, we thank our parents for their boundless support and for making our life so easy, and for helping us to tackle all those difficulties in our life.

AKSHAY V SANTHOSH

200021093597

CONTENTS

INTRODUCTION

A GPU, or Graphics Processing Unit, is a specialized processor designed for handling graphical and video rendering tasks. The seminar on GPU is an opportunity to learn about the inner workings of GPUs, their applications, and how they impact various industries. This seminar will provide a comprehensive overview of GPUs and their uses. Participants will gain a deeper understanding of the technology behind GPUs and how they are shaping the future of computing.

There are various applications that require a 3D world to be simulated as realistically as possible on a computer screen. These include 3D animations in games, movies and other real world simulations. It takes a lot of computing power to represent a 3D world due to the great amount of information that must be used to generate a realistic 3D world and the complex mathematical operations that must be used to project this 3D world onto a computer screen. In this situation, the processing time and bandwidth are at a premium due to large amounts of both computation and data.

The functional purpose of a GPU then, is to provide a separate dedicated graphics resources, including a graphics processor and memory, to relieve some of the burden off of the main system resources, namely the Central Processing Unit, Main Memory, and the System Bus, which would otherwise get saturated with graphical operations and I/O requests. The abstract goal of a GPU, however, is to enable a representation of a 3D world as realistically as possible. So these GPUs are designed to provide additional computational power that is customized specifically to perform these 3D tasks.

WHAT IS A GPU?

A Graphics Processing Unit (GPU) is a microprocessor that has been designed specifically for the processing of 3D graphics. The processor is built with integrated transform, lighting, triangle setup/clipping, and rendering engines, capable of handling millions of math-intensive processes per second. GPUs form the heart of modern graphics cards, relieving the CPU (central processing units) of much of the graphics processing load. GPUs allow products such as desktop PCs, portable computers, and game consoles to process real-time 3D graphics that only a few years ago were only available on high-end workstation.

Used primarily for 3-D applications, a graphics processing unit is a singlechip processor that creates lighting effects and transforms objects every time a 3D scene is redrawn. These are mathematically-intensive tasks, which otherwise, would put quite a strain on the CPU. Lifting this burden from the CPU frees up cycles that can be used for other jobs.

However, the GPU is not just for playing 3D-intense videogames or for those who create graphics (sometimes referred to as graphics rendering or content-creation) but is a crucial component that is critical to the PC's overall system speed. In order to fully appreciate the graphics card's role it must first be understood.

Many synonyms exist for Graphics Processing Unit in which the popular one being the graphics card .It’s also known as a video card, video accelerator, video adapter, video board, graphics accelerator, or graphics adapter.

Figure 1: GPU

The above figure represents NVIDIA GeForce3, which was a graphics processing unit (GPU) released by NVIDIA in 2000. It was the third generation of the GeForce series and was built on a 0.15-micron process technology. The GeForce 3 was one of the first GPUs to support hardware transform and lighting (T&L), which allowed for more realistic lighting and shadows in 3D graphics. It also introduced other features such as anisotropic filtering, which improved the quality of textures in 3D games and applications. The GeForce 3 was a popular choice for gamers and graphics professionals in its time, and it was widely used in high-end graphics cards.

HISTORY AND STANDARDS

The first graphics cards, introduced in August of 1981 by IBM, were monochrome cards designated as **Monochrome Display Adapters** (MDAs). The displays that used these cards were typically text-only, with green or white text on a black background. Color for IBM-compatible computers appeared on the scene with the 4-color **Hercules Graphics Card** (HGC), followed by the 8-color **Color Graphics Adapter** (CGA) and 16-color **Enhanced Graphics Adapter** (EGA). During the same time, other computer manufacturers, such as Commodore, were introducing computers with built-in graphics adapters that could handle a varying number of colors.

When IBM introduced the **Video Graphics Array** (VGA) in 1987, a new graphics standard came into being. A VGA display could support up to 256 colors (out of a possible 262,144-color palette) at resolutions up to 720x400. Perhaps the most interesting difference between VGA and the preceding formats is that VGA was analog, whereas displays had been digital up to that point. Going from digital to analog may seem like a step backward, but it actually provided the ability to vary the signal for more possible combinations than the strict on/off nature of digital.

Over the years, VGA gave way to **Super Video Graphics Array** (SVGA). SVGA cards were based on VGA, but each card manufacturer added resolutions and increased color depth in different ways. Eventually, the **Video Electronics Standards Association** (VESA) agreed on a standard implementation of SVGA that provided up to 16.8 million colors and 1280x1024 resolution. Most graphics cards available today support **Ultra Extended Graphics Array** (UXGA). UXGA can support a palette of up to 16.8 million colors and resolutions up to 1600x1200 pixels. Even though any card you can buy today will offer higher colors and resolution than the basic VGA specification, VGA mode is the de facto standard for graphics and is the minimum on all cards. In addition to including VGA, a graphics card must be able to connect to your computer. While there are still a number of graphics cards that plug into an Industry Standard Architecture (ISA)

or **Peripheral Component Interconnect** (PCI) slot, most current graphics cards use the **Accelerated Graphics Port** (AGP).

In the early 1990s, GPUs were developed as dedicated graphics acceleration hardware for use in computer games. The first consumer GPU, the Nvidia GeForce 256, was released in 1999. GPU technology continued to evolve and expand, eventually becoming a standard component in personal computers and powering other demanding applications such as scientific simulations, machine learning, and cryptocurrency mining.

In the mid-2010s, GPUs started to incorporate specialized hardware for machine learning, further expanding their capabilities and making them an essential tool for AI research and deployment. As of 2023, GPUs are critical components in many industries, ranging from gaming to high-performance computing, and AI. They continue to evolve, incorporating more transistors and advanced features, enabling more complex and demanding applications.

Following are some example of GPU with it's generations :

- 1st Generation: GeForce 256 (1999)
- 2nd Generation: GeForce 3 (2001)
- 3rd Generation: GeForce 4 (2002)
- 4th Generation: GeForce 8 (2006)
- 5th Generation: GeForce 500 (2010)
- 6th Generation: GeForce 600 (2012)
- 7th Generation: GeForce 700 (2013)
- 8th Generation: GeForce 900 (2014)
- 9th Generation: GeForce 1000 (2016)
- 10th Generation: GeForce 16 (2018)

PERIPHERAL COMPONENT INTERCONNECT(PCI)

There are a lot of incredibly complex components in a computer. And all of these parts need to communicate with each other in a fast and efficient manner. Essentially, a bus is the channel or path between the components in a computer. During the early 1990s, Intel introduced a new bus standard for consideration, the Peripheral Component Interconnect (PCI).It provides direct access to system memory for connected devices, but uses a bridge to connect to the front side bus and therefore to the CPU.

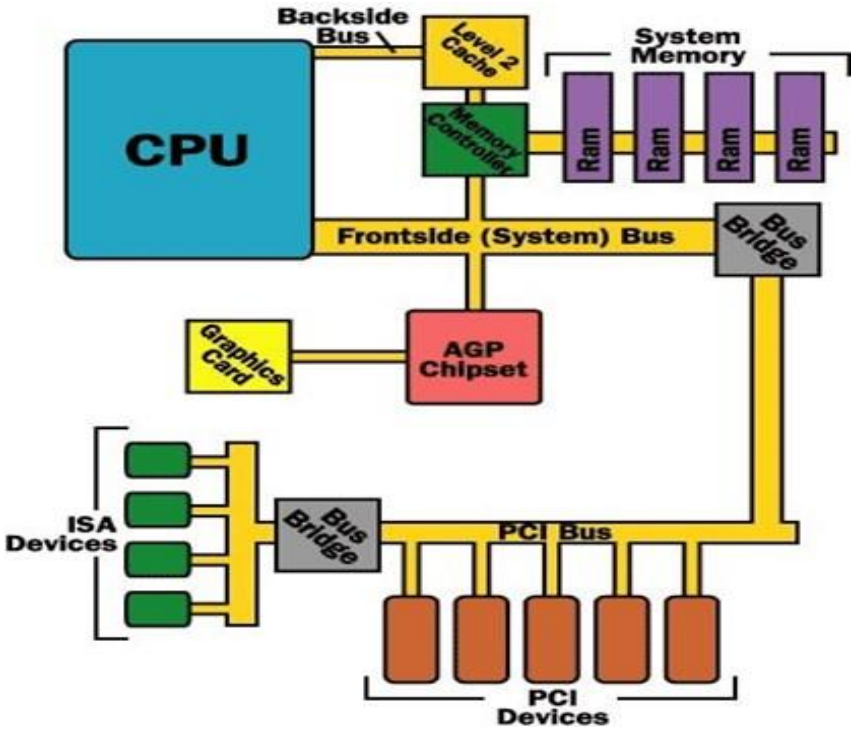


Figure 2

The illustration above shows how the various buses connect to the CPU. PCI can connect up to five external components. Each of the five connectors for an external component can be replaced with two fixed devices on the motherboard. The PCI bridge chip regulates the speed of the PCI bus independently of the CPU's speed. This provides a higher degree of reliability and ensures that PCIhardware manufacturers know exactly what to design for.

PCI originally operated at 33 MHz using a 32-bit-wide path. Revisions to the standard include increasing the speed from 33 MHz to 66 MHz and doubling the bit count to 64. Currently, PCI-X provides for 64-bit transfers at a speed of 133 MHz for an amazing 1-GBps (gigabyte per second) transfer rate!

PCI cards use 47 pins to connect (49 pins for a mastering card, which can control the PCI bus without CPU intervention). The PCI bus is able to work with so few pins because of hardware multiplexing, which means that the device sends more than one signal over a single pin. Also, PCI supports devices that use either 5 volts or 3.3 volts. PCI slots are the best choice for network interface cards (NIC), 2-D video cards, and other high-bandwidth devices. On some PCs, PCI has completely superseded the old ISA expansion slots.

Although Intel proposed the PCI standard in 1991, it did not achieve popularity until the arrival of Windows 95 (in 1995). This sudden interest in PCI was due to the fact that Windows 95 supported a feature called Plug and Play (PnP). PnP means that you can connect a device or insert a card into your computer and it is automatically recognized and configured to work in your system. Intel created the PnP standard and incorporated it into the design for PCI. But it wasn't until several years later that a mainstream operating system, Windows 95, provided system-level support for PnP. The introduction of PnP accelerated the demand for computers with PCI.

ACCELERATED GRAPHICS PORT (AGP)

The need for streaming video and real-time-rendered 3-D games requires an even faster throughput than that provided by PCI. In 1996, Intel debuted the Accelerated Graphics Port (AGP), a modification of the PCI bus designed specifically to facilitate the use of streaming video and high-performance graphics.

AGP is a high-performance interconnect between the core-logic chipset and the graphics controller for enhanced graphics performance for 3D applications. AGP relieves the graphics bottleneck by adding a dedicated high-speed interface directly between the chipset and the graphics controller as shown below.

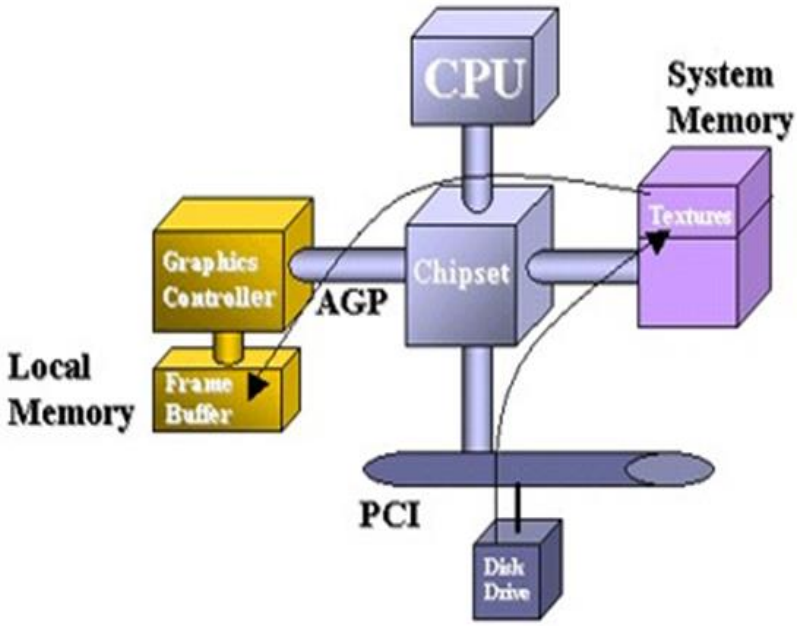


Figure 3

Segments of system memory can be dynamically reserved by the OS for use by the graphics controller. This memory is termed AGP memory or non-local video memory. The net result is that the graphics controller is required to keep fewer texture maps in local memory.

AGP has 32 lines for multiplexed address and data. There are an additional 8 lines for sideband addressing. Local video memory can be expensive and it cannot be used for other purposes by the OS when unneeded by the graphics of the running applications. The graphics controller needs fast access to local video memory for screen refreshes and various pixel elements including Z-buffers, double buffering, overlay planes, and textures.

For these reasons, programmers can always expect to have more texture memory available via AGP system memory. Keeping textures out of the frame buffer allows larger screen resolution, or permits Z-buffering for a given large screen size. As the need for more graphics intensive applications continues to scale upward, the amount of textures stored in system memory will increase. AGP delivers these textures from system memory to the graphics controller at speeds sufficient to make system memory usable as a secondary texture store.

- ## AGP Memory Allocation

During AGP memory initialization, the OS allocates 4K byte pages of AGP memory in main (physical) memory. These pages are usually discontiguous. However, the graphics controller needs contiguous memory. A translation mechanism called the GART (Graphics Address Remapping Table), makes discontiguous memory appear as contiguous memory by translating virtual addresses into physical addresses in main memory through a remapping table.

A block of contiguous memory space, called the Aperture is allocated above the top of memory. The graphics card accesses the Aperture as if it were main memory. The GART is then able to remap these virtual addresses to physical addresses in main memory. These virtual addresses are used to access main memory, the local frame buffer, and AGP memory.

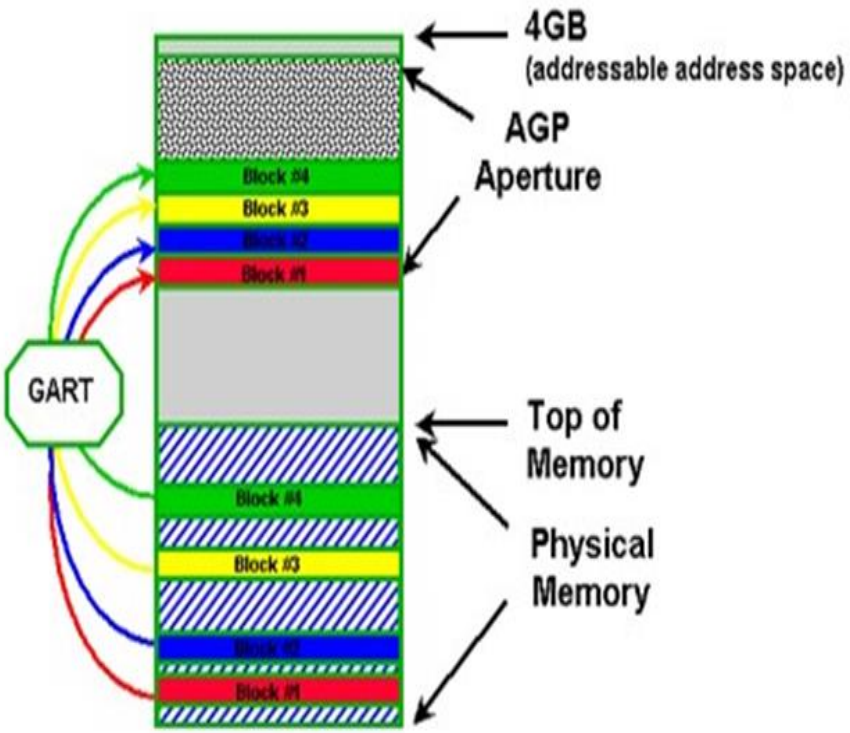


Figure 4

- ## AGP Transfers

AGP provides two modes for the graphics controller to directly access texture maps in system memory: pipelining and sideband addressing. Using Pipe mode, AGP overlaps the memory or bus access times for a request ("n") with the issuing of following requests ("n+1"..."n+2"... etc.). In the PCI bus, request "n+1" does not begin until the data transfer of request "n" finishes.

With sideband addressing (SBA), AGP uses 8 extra "sideband" address lines which allow the graphics controller to issue new addresses and requests simultaneously while data continues to move from previous requests on the main 32 data/address lines. Using SBA mode improves efficiency and reduces latencies.

- ## AGP Specifications

The current PCI bus supports a data transfer rate up to 132 MB/s, while AGP (at 66MHz) supports up to 533 MB/s! AGP attains this high transfer rate due to it's ability to transfer data on both the rising and falling edges of the 66MHz clock.

Mode	Approximate clock rate	Transfer rate (MBps)
1x	66 MHz	266
2x	133 MHz	533
4x	266 MHZ	1066
8x	533 MHZ	2133

The AGP slot typically provides performance which is 4 to 8 times faster than the PCI slots inside your computer.

COMPONENTS OF GPU

There are several components on a typical graphics card:

Graphics Processor: The graphics processor is the brains of the card, and is typically one of three configurations:

Graphics co-processor: A card with this type of processor can handle all of the graphics chores without any assistance from the computer's CPU. Graphics co-processors are typically found on high-end video cards.

Graphics accelerator: In this configuration, the chip on the graphics card renders graphics based on commands from the computer's CPU. This is the most common configuration used today.

Frame buffer: This chip simply controls the memory on the card and sends information to the digital-to-analog converter (DAC). It does no processing of the image data and is rarely used anymore.

Memory: The type of RAM used on graphics cards varies widely, but the most popular types use a dual-ported configuration. Dual-ported cards can write to one section of memory while it is reading from another section, decreasing the time it takes to refresh an image.

Graphics BIOS: Graphics cards have a small ROM chip containing basic information that tells the other components of the card how to function in relation to each other. The BIOS also performs diagnostic tests on the card's memory and input/ output (I/O) to ensure that everything is functioning correctly.

Digital-to-Analog Converter (DAC): The DAC on a graphics card is commonly known as a RAMDAC because it takes the data it converts directly from the card's memory. RAMDAC speed greatly affects the image you see on the monitor. This is because the refresh rate of the image depends on how quickly the analog information gets to the monitor.

Display Connector: Graphics cards use standard connectors. Most cards use the 15pin connector that was introduced with Video Graphics Array (VGA).

Computer (Bus) Connector: This is usually Accelerated Graphics Port (AGP). This port enables the video card to directly access system memory. Direct memory access helps to make the peak bandwidth four times higher than the Peripheral Component Interconnect (PCI) bus adapter card slots. This allows the central processor to do other tasks while the graphics chip on the video card accesses system memory.

HOW IS 3D ACCELERATION DONE?

There are different steps involved in creating a complete 3D scene. It is done by different parts of the GPU, each of which are assigned a particular job. During 3D rendering, there are different types of data the travel across the bus. The two most common types are texture and geometry data. The geometry data is the "infrastructure" that the rendered scene is built on. This is made up of polygons (usually triangles) that are represented by vertices, the end-points that define each polygon. Texture data provides much of the detail in a scene, and textures can be used to simulate more complex geometry, add lighting, and give an object a simulated surface.

Many new graphics chips now have accelerated Transform and Lighting (T&L) unit, which takes a 3D scene's geometry and transforms it into different coordinate spaces. It also performs lighting calculations, again relieving the CPU from these math-intensive tasks.

Following the T&L unit on the chip is the triangle setup engine. It takes a scene's transformed geometry and prepares it for the next stages of rendering by converting the scene into a form that the pixel engine can then process. The pixel engine applies assigned texture values to each pixel. This gives each pixel the correct color value so that it appears to have surface texture and does not look like a flat, smooth object. After a pixel has been rendered it must be checked to see whether it is visible by checking the depth value, or Z value.

A Z check unit performs this process by reading from the Z-buffer to see if there are any other pixels rendered to the same location where the new pixel will be rendered. If another pixel is at that location, it compares the Z value of the existing pixel to that of the new pixel. If the new pixel is closer to the view camera, it gets written to the frame buffer. If it's not, it gets discarded. After the complete scene is drawn into the frame buffer the RAMDAC converts this digital data into analog that can be given to the monitor for display.

PERFORMANCE FACTORS OF GPU

There are many factors that affect the performance of a GPU. Some of the factors that are directly visible to a user are given below.

- **Fill Rate:**

It is defined as the number of pixels or texels (textured pixels) rendered per second by the GPU on to the memory . It shows the true power of the GPU. Modern GPUs have fill rates as high as 3.2 billion pixels. The fill rate of a GPU can be increased by increasing the clock given to it.

- **Memory Bandwidth:**

It is the data transfer speed between the graphics chip and its local frame buffer. More bandwidth usually gives better performance with the image to be rendered is of high quality and at very high resolution.

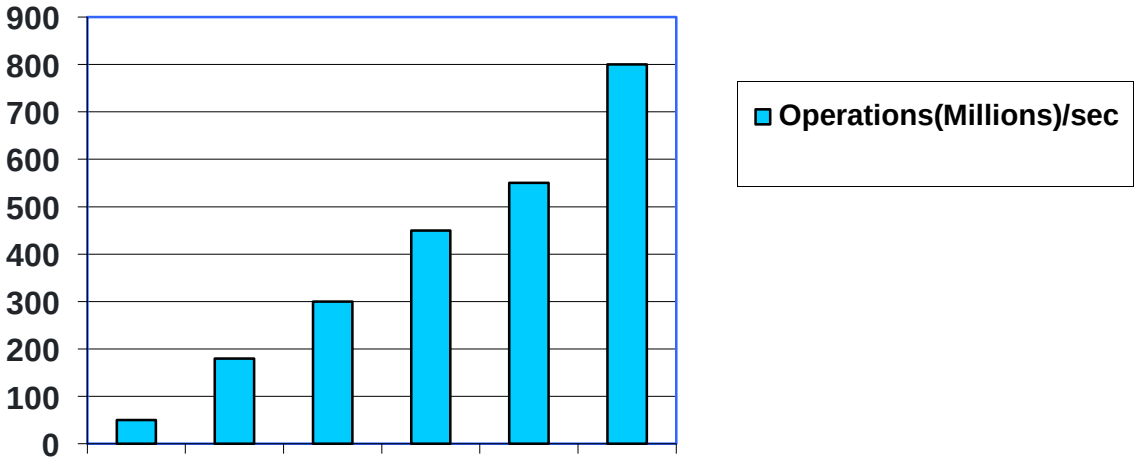
- **Memory Management:**

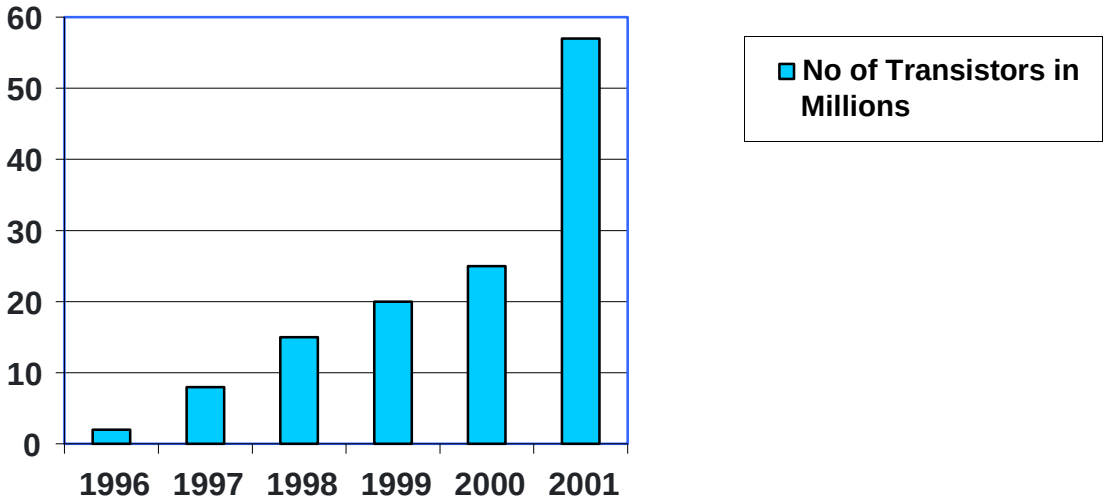
The performance of the GPU also depends on how efficiently the memory is managed, because memory bandwidth may become the only bottle neck if not managed properly.

- **Hidden Surface removal:**

A term to describe the reducing of overdraws when rendering a scene by not rendering surfaces that are not visible. This helps a lot in increasing the performance of GPU, by preventing overdraw so that the fill rate of the GPU can be utilized to the maximum.

Now lets see how far GPUs have come as far as performance is concerned.





Applications of GPU

Graphics Processing Units (GPUs) have a wide range of applications in various fields, including:

1. Gaming:

   GPUs are essential for high-quality gaming experiences, providing fast and smooth graphics rendering, 3D effects, and other visual features.

2. Professional Graphics and Design:

   GPUs are used by artists, designers, and engineers to create complex visual content, including 3D animations, product visualizations, and architectural designs.

3. Video Processing and Streaming:

   GPUs are used for encoding, decoding, and processing video, enabling faster and more efficient video streaming and playback.

4. Artificial Intelligence and Machine Learning:

   GPUs are widely used for training and running deep learning algorithms, providing the massive parallel processing power required for these tasks.

5. Scientific and Engineering Applications:

   GPUs are used in a variety of scientific and engineering fields, including simulations, modeling, and data analysis, enabling faster and more accurate results.

6. Virtual and Augmented Reality:

   GPUs are critical for delivering high-quality and immersive virtual and augmented reality experiences, by providing fast and smooth graphics rendering and interaction.

7. Autonomous Systems:

GPUs are used in autonomous systems, such as self-driving cars, to provide the computational power required for real-time object detection and decision making.

Overall, GPUs are versatile and powerful computing devices, used in a wide range of applications and industries, to provide high-performance graphics and computational capabilities.

TYPES OF GPUS

There are mainly two types of GPUs, they are

1. Those that can handle all of the graphics processes without any assistance from the computer's CPU. They are typically found on high-end workstations. These are mainly used for Digital Content Creation like 3D animation as it supports a lot of 3D functions. Some of them are

   Quadro series from NVIDIA.

   Wildcat series from 3D Labs.

   FireGL series from ATI.

2. The chip on the graphics card renders graphics based on commands from the computer's CPU. This is the most common configuration used today. These are used for 3D gaming and such smaller tasks. They are found on normal desktop PCs and are better known as 3D accelerators. These support less functions and hence are cheaper. Some of them are

   Geforce series from NVIDIA.

   Radeon series from ATI Technology ltd.

   Kyro series from STM Microelectronics

Today's GPU can do what was hoped for and beyond. In the last year a giant leap have been made in the GPU technology. The maximum amount of RAM that can be found on a graphics card has jumped from 16MB to a whopping 128MB. The premier company in GPU manufacturing ATI, who has held the position past couple of years has given way to Nvidia , whose new ground breaking technology is leaving ATI to follow.

GEFORCE4

The NVIDIA GeForce4 is a series of graphics processing units (GPUs) released by NVIDIA in 2002. It was the fourth generation of the GeForce series.

NVIDIA introduced the groundbreaking, top-to-bottom GeForce4 family of GPUs—delivering new levels of graphics performance and display flexibility to desktop and mobile PCs. NVIDIA's latest creation, the GeForce4 GPU is the fourth edition in the famed GeForce lineup. It has wowed gamers and artists alike by having the capability to make graphics "better than life" with 128MB of DDR memory and a super fast processor. What this means is that the GeForce4 is capable of rendering graphics better than the eye can see. Although this is extraordinary, there are no available monitors that can handle displaying such graphics. Even so, the NVIDIA GeForce4 is truly an extraordinary and groundbreaking graphics processing unit.

GeForce4 is the most complete family of graphics solutions—from the ferocious graphics power of the GeForce4 Ti, the world's fastest GPU; to the multidisplay flexibility of the mainstream GeForce4 MX; to the most advanced mobile graphics available, GeForce4 Go.

Three of its kind has been released…

- GeForce4 Ti series (NV25)
- GeForce4 MX series (NV17)
- GeForce4 Go series

Chip Architecture

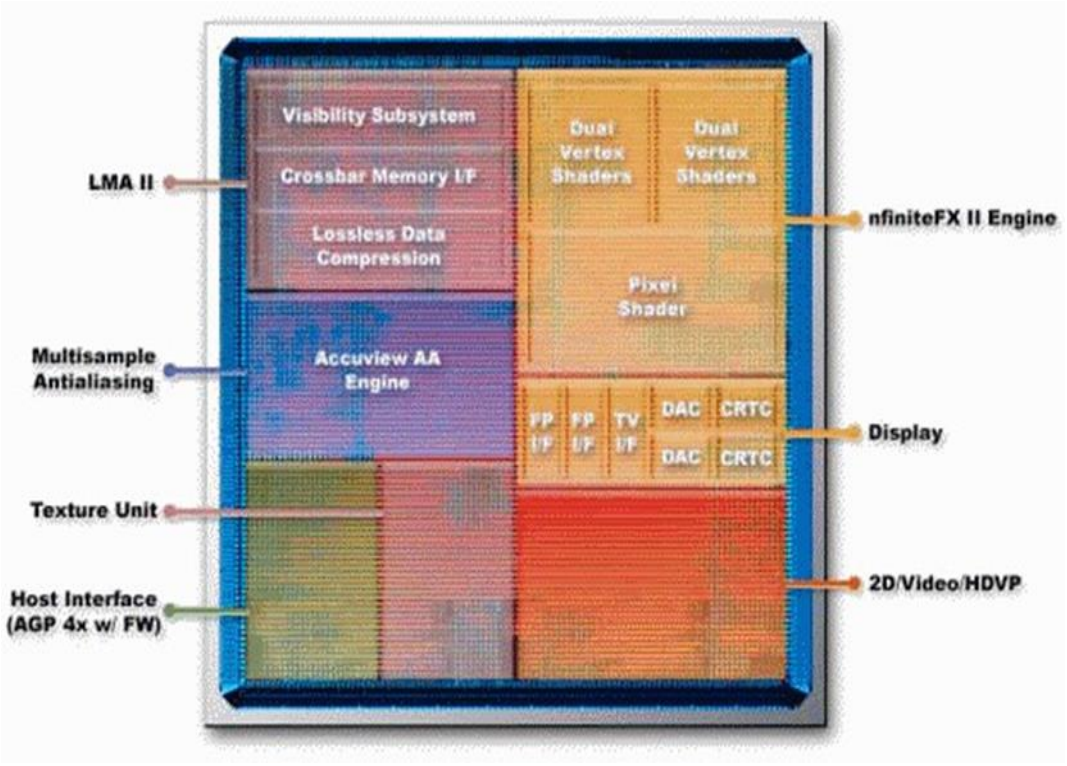


Figure 5

The LMA II

In the upper left hand corner lies the LMA II. The LMA II controls the flow of data from the chip to the GPU's memory (the DDR memory on the graphics card). It controls how much data is sent to the memory and how fast it is sent to the memory.

The Accuview AA Engine

The Accuview AA Engine, located to the left and in the middle of the chip, does the antialiasing for the GPU.

The Interface Unit

At the bottom of the GeForce4 chip and to the very left is the interface unit. This simply determines what interface (2X or 4X AGP) the computer has, and adjusts to comply with it**.**

The Texture Unit

At the bottom of the chip and to the left-center is the texture unit. This unit is dedicated to processing textures that must be rendered.

The InfiniteFX II Engine

This is the striking feature of GeForce4 series.

One among the classic computer graphics problem have been regarding the renderingthe realistic hair and fur. Animals with skin texture were easy to render than those with fur.

For the first time ever, and only through the power of nfiniteFX II engine, which includes support for dual vertex shaders,advanced pixel shader pipelines , 3D textures, shadow buffers and z-correct bump mapping ,is it now possible to render those.

It also improves rendering of shading.

Figure 6

A demo shot of a wolf man which has been rendered by GeForce4

The Display Unit

In the center to the right of the chip is the display unit. This unit is simple and its only job is to determine the optimal display resolution for the display being used. It also determines what type of display it is and optimizes its performance.

The 2D/Video/HDVP Unit

In the bottom right corner of the chip is the 2D/Video/HDVP unit. This part of the chip is dedicated to all those tasks that don't require 3D rendering. These tasks include movies, 2D pictures, 2D games and almost any other program that doesn't have 3D graphics

nView Technology:

Simply put, NVIDIAs Multi-Monitor and Dual Independent Display technology, now called "nView", has been polished up nicely. nView is available on both the GF4 Ti and MX and enables the following:

Multi-desktop tools

- Multi-desktop integration
- Full featured interface including explorer browser with birds-eye views of desktops
- Toolbar control available as well for those needing a streamlined, low real- estate interface

Window management

- Individual application control
- Window & dialog repositioning

Application management

- Transparency & colored transparency window options
- Extends functionality of all applications
- Pop- up menu control

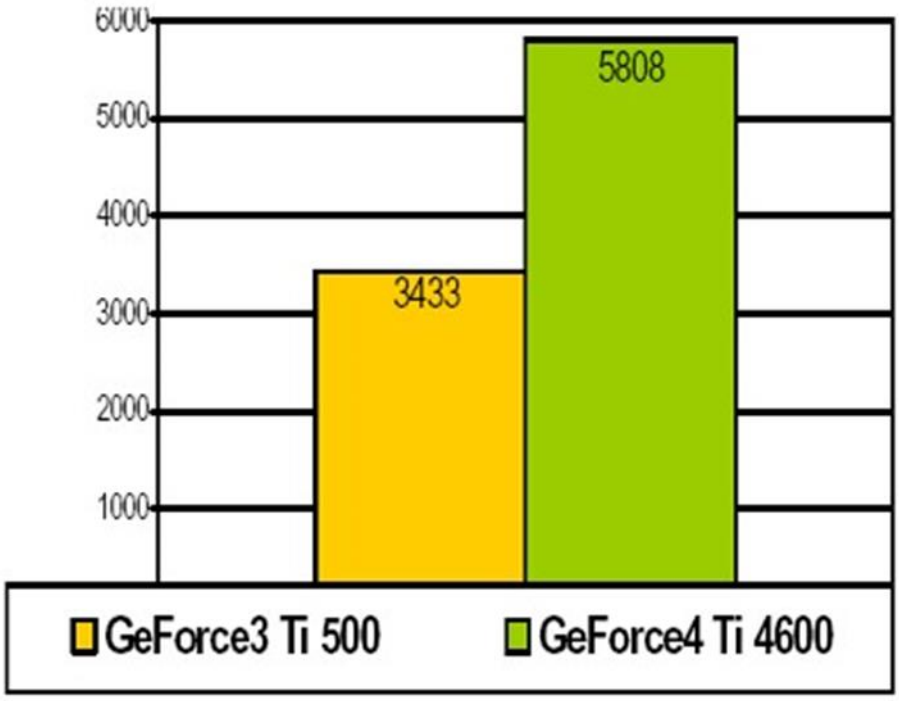
GEFORCE4 TI

NVIDIA's crown graphics card is the GeForce4 Ti, a creation of admirable brilliance. Its basic specifications include

- 63 million transistors (only 3 million more than GeForce3)
- Manufactured in TSMC's .15 µ process
- Chip clock 225 - 300 MHz
- Memory clock 500 - 650 MHz
- Memory bandwidth 8,000 - 10,400 MB/s
- T&L Performance of 75 - 100 million vertices/s
- 128 MB frame buffer by default
- InfiniteFX II engine
- Accuview Anti Aliasing
- Light Speed Memory Architecture II
- nView

It has a 128 bit bus , double the size of previous busses and the only graphics card to have one. This extra bus size improves performance dramatically. Its considered as the **world's fastest GPU** available today**.**

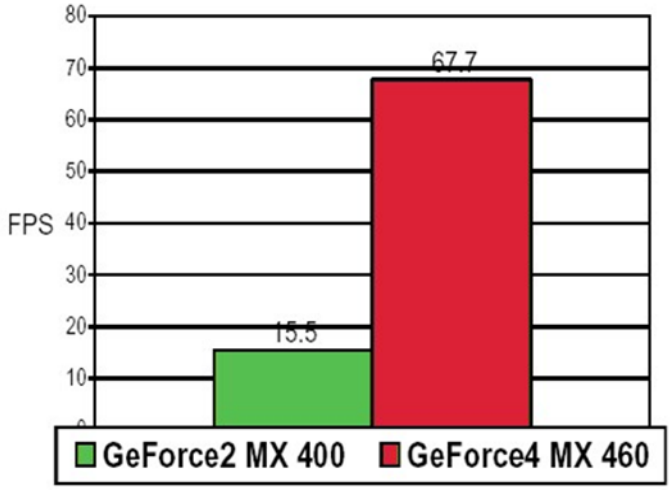
NV25 series include four- GeForce4 Ti 4600, GeForce4 Ti 4400, GeForce4 Ti 4200 & GeForce4 Ti 4200 w/AGP 8x



This chart shows comparison of performances of GeForce4 Ti 4600 and its predecessor in FPS

GEFORCE4 MX

With the GeForce4 MX graphics processing units (GPUs), NVIDIA provides a new level of cost-effective, high-performance graphics to the mainstream PC user. The GeForce4 MX is the cheapest and worst performing of the three lines. The standard graphics card comes with 64MB of RAM, a hefty amount that can handle almost any task. The GeForce4 MX has a 64bit bus, which is also pretty standard for today's graphics cards. For the most part the only real improvement from the GeForce3 MX to the GeForce4MX is the graphics processing unit, which beats its predecessors, the GeForce2 and GeForce3, hands down.



Performance comparisons

GEFORCE4 GO

NVIDIA introduces the fastest, most comprehensive and feature-rich computing experience ever realized on a mobile platform— the GeForce4 Go. With a revolutionary core of integrated technologies, the GeForce4 Go ensures unparalleled performance, battery life, and DVD and video playback. From notebooks powerful enough to replace your desktop PC, to those that are both thin and light, NVIDIA's GeForce4 Go mobile GPUs provide unprecedented mobile computing experiences.

The GeForce4 Go was based on the NVIDIA infiniteFX II engine and included features such as programmable shaders, advanced shading techniques, and Accuview Antialiasing technology to improve the quality of 3D graphics. It also included support for nView technology, allowing multiple monitors to be connected to a single GPU and used as a single large desktop

Figure 7

GeForce RTX 30 series

The NVIDIA GeForce RTX 30 series GPUs are some of the most widely used and highly regarded GPUs on the market in 2021. These GPUs are designed for gaming and offer excellent performance, ray tracing capabilities, and support for 4K and VR gaming. The RTX 3080, RTX 3070, and RTX 3060 Ti are some of the most popular models in this series.

In the professional and scientific market, the NVIDIA Quadro and Tesla series GPUs are widely used for applications such as 3D rendering, machine learning, and scientific simulations. The NVIDIA Quadro RTX 8000, Quadro RTX 6000, and Tesla V100 are some of the most popular models in this market.

It's important to note that the GPU market is constantly evolving, and new models and technologies are constantly being introduced. The best GPU model for you will depend on your specific use case, budget, and performance requirements.

CONCLUSION

In conclusion, Graphics Processing Units (GPUs) have come a long way since their inception and have become an integral part of modern computing. GPUs are specialized processors that are designed to perform the demanding mathematical calculations required for rendering high-quality 3D graphics and visual effects.

Over the years, GPUs have evolved to include a wide range of advanced features and technologies. GPUs also have important applications beyond gaming and graphics, such as in scientific computing, machine learning, and data visualization. With the increasing demand for computing power and the growing complexity of computational tasks, GPUs are poised to play an even more important role in the future of computing.

Overall, the GPU is a fascinating and important technology that has had a profound impact on the world of computing, and it is sure to continue to play a central role in shaping the future of technology. By observing the current pace at which work is going on in developing GPUs we can surely come to a conclusion that we will be able to see better and faster GPUs in the near future.

FUTURE ENHANCEMENT

In the future, Graphics Processing Units (GPUs) are likely to continue to evolve and become even more powerful and versatile. Some of the key areas in which GPUs are likely to see enhancements and improvements include:

- Performance: GPUs are likely to become even faster and more efficient, enabling them to handle even more complex and demanding computational tasks.

- Energy efficiency: As the demand for more computing power increases, it will be important for GPUs to consume less energy and run cooler, making them more suitable for use in a wider range of devices.

- AI and Machine Learning: GPUs are already widely used in machine learning and AI applications, and in the future, they are likely to become even more specialized and optimized for these tasks, providing faster and more accurate results.

- Cloud Computing: GPUs are likely to play an increasingly important role in cloud computing, enabling the delivery of high-performance graphics and computational services to users over the internet.

- Virtual and Augmented Reality: As virtual and augmented reality technologies continue to mature, GPUs are likely to play a central role in delivering high-quality, immersive experiences to users.

Overall, the future of GPUs looks bright, and as new technologies emerge and computational demands continue to increase, GPUs are sure to continue to play a critical role in driving innovation and shaping the future of technology.

BIBLIOGRAPHY

1) https://www.wikipedia.org/
2) https://chat.openai.com/chat
3) www.google.com