IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 6, June 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006083 |

A Study of GPU Computing Technologies: Architecture and Its Components

Darshan.N[1], Dr.Samitha.Khaiyum[2]

Post Graduate Student, Department of M.C.A, Dayananda Sagar College of Engineering, Bangalore, India

Head of Department, Department of M.C.A, Dayananda Sagar College of Engineering, Bangalore, India

ABSTRACT: In Nowadays, time is most important in computational field. Today every field in computing has a large amount of data, we need to process data to get valuable information out of it. Central Processing Units are latency-oriented processors, while Graphics Processing Units are data-parallel, throughput oriented processors. Besides their traditional use as graphics processors, the GPU computing has been used in recent years for general purpose computations**.** To reduce processing time and using of maximum capacity of processor, we divide a large computation problem in to small groups that is processed by individual processor. The development of graphics hardware led to an extensive use in both commercial and scientific applications. Many papers reported high speedups in various domains**.** In this paper we present the GPU based of Parallel Processing architecture, components, working and its applications for performing fast execution of a task.

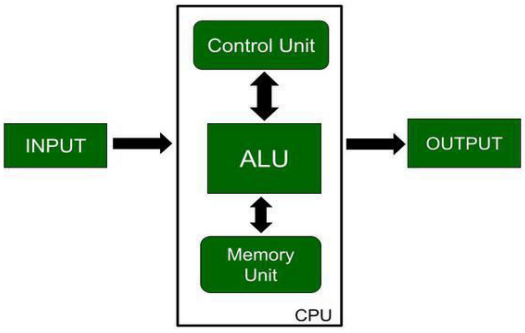
KEYWORDS: Computation, Parallel computing technique, GPU Architecture, Components of GPU.

I.INTRODUCTION

A GPU(graphic processing unit)is a microprocessor that discharges and accelerates 2D or 3D graphics. Its used in embedded systems, personal computers, smartphones , gaming consoles and workstations. GPU is mostly used for high performance and cost effective computing. In Current days GPU and CPU have also competency to solve complex problem with high performance. Every processor has its own physical architecture and maximum processing speed. To overcome this problem multiple processors are introduced where each processor co-ordinate to other that leads to Parallel Computing. In parallel computing architecture huge computation problem can divided into multiple problems that can be handled by different cores independently and result comes after recombination of all solutions of problem. The GPUs are found in many applications. At the beginning, they used mostly in academia, since they brought significant speedup in many domains they were adopted from wider research community. GPUs are nowadays used in computation chemistry, computational physics, life science, signal processing, finances, oil and gas exploration and many more.

II.CPU AND GPU : THE ARCHITECTURE

i. Architecture of CPU



Control unit gets the instruction or data from memory then decode the instruction and then it sequentially executes the programming task, whereas parallel computing also have same design but some changes in its processing. It helps input and

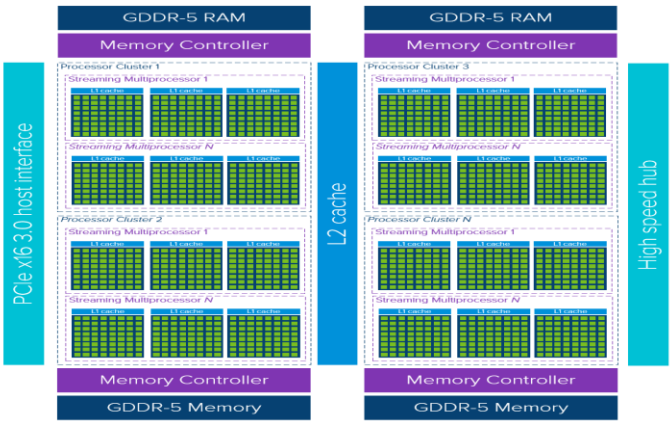
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006083 |

output devices to communicate with each other and perform respective operations. It stores data which is input, intermediate results in between processing, and instructions.

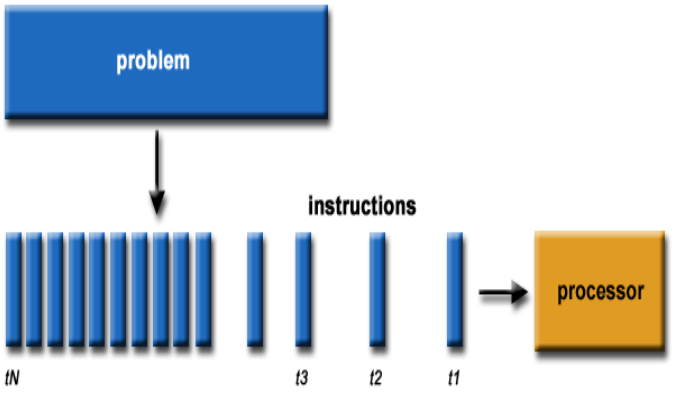
ii. Architecture of GPU



A single GPU device consists of multiple Processor Clusters that contain multiple Streaming Multiprocessors . Each SM accommodates a layer-1 instruction cache layer with associated cores. Typically, Streaming Multiprocessors uses a dedicated layer-1 cache and a shared layer-2 cache before pulling data from global GDDR-5 memory. Its architecture is tolerant of memory latency. Compared to a CPU, GPU works with few and relatively small, memory cache layers. Reason is that a GPU has more transistors dedicated to computational meaning it cares less how long it will takes the retrieve data from memory. The potential memory latency access is masked up as long as the GPU has enough computations at hand, keeping it busy.

III. PARALLEL COMPUTING

Parallel Computing is the study, design, and implementation of algorithms in such a way as to make use of multiple processors to solve a problem. The main purpose is to solve a problem faster and a bigger problem in the same amount of time by using more processors to share the work. Parallel computing is nothing but computation in which many calculations are carried out simultaneously, operating on the principle that larger problems can often be divided into smaller ones, which are then solved at the same time.



Normally parallel computing was considered as an advanced area of Computer Science. Parallel systems were huge, rare and expensive. Only a little percentage of software developers were creating parallel programs. With the movement into multi-core computers, all general purpose computers manufactured today are multi-core. Parallel systems are easily created by building a cluster of computers.

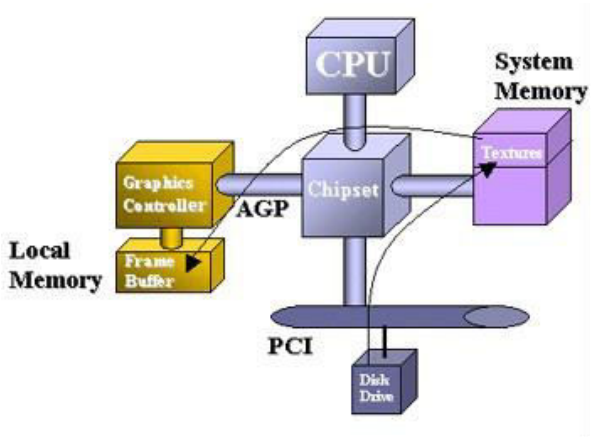
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006083 |

3.1. ACCELERATED GRAPHICS PORT (AGP)

In 1996, Intel introduced the Accelerated Graphics Port (AGP), a modified version of the PCI bus designed to facilitate the use of high-performance graphics and streaming video. AGP is a high-performance interconnector between the graphics controller for enhanced graphics performance for 3D applications and the core-logic chipset. AGP provides the graphics by adding a dedicated high-speed interface directly between the chipset and the graphics controller as shown below.



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

3.2. COMPONENETS OF GPU

- **Graphics co-processor**: this type of processor can handle all of the graphics chores without any assistance from the CPU. The Graphic co- processors are normally found on high-end video cards.
- **Graphics accelerator**: In this type, the chip is on the graphics card renders graphics based on commands from the computer's CPU.
- **Frame buffer:** it is a chip that controls the memory on the card and sends information to the digital-to-analog converter .
- **Memory –** The type of RAM used on graphics cards varies widely and the popular types uses a dual-ported configuration. The Dual-port cards are written to one section of memory while it is reading from another section.
- **Graphics BIOS –** Graphics cards have a small ROM chip containing normal basic information that tells the other components of the card how to function in relation to each other. The BIOS also performs tests on the card's memory and input or output to ensure that everything is functioning properly.
- **Display Connector –** Graphics cards uses standard connectors. Most cards use the 15- pin connector which was introduced with Video Graphics Array (VGA).

IV. TYPES OF GPU

There are substantially two kinds of GPUs, they are
. Those can handle all of the plates processes without any backing from the computer's
2. CPU. generally set up on high- end workstations. These are substantially used for Digital Content Creation like 3D vitality because it supports a lots of 3D functions.
• Some of them are … …
Quadro series from NVIDIA.
Wildcat series from 3D Labs.
FireGL series from ATI.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006083 |

2. The chip present on the plates card renders plates grounded on commands from CPU. This is the most common configuration used moment. Its used for 3D gaming and similar lower tasks. They 're set up on normal desktop PCs and are appertained as 3D accelerators. These support lower functions and they're cheaper.

• Some of them are … ….

Geforce series from NVIDIA.

Radeon series from ATI Technology ltd.

Kyro series from STM Microelectronics

Nowerdays GPU can do what was hoped for and beyond. Within In the last time a big vault have been made within in the GPU technology. The maximum quantum of RAM can be set up on a plates card has jumped from 16 MB to 128 MB. The premier company in GPU manufacturing ATI, who has held the position once number of times has given way to nVidia, whose new ground breaking technology is leaving ATI to follow.

V. APPLICATION DOMAINS

The GPU performance and calculation power offers particular classes of operations. A good check of characteristics that operation needs to satisfy and successfully counterplotted for GPU prosecution could be set up.

The GPU application depends on the degree of community in the workload. The GPU also needs a workload to hide memory quiescence, so computation conditions should be big. Hence the high outturn is more important than quiescence in GPU. The Number of numeric calculations performed per memory sale should be an order of magnitude advanced in order to save computation intensity. On the other hand, prosecution divergence should be avoided as much as possible. Since vestments are run in batches, if vestments within a batch diverge, there's a penalty in prolonged prosecution. Eventually, bandwidth application has to be sustainable. The GPUs have high peak bandwidth application to and from their onboard memory. But, for operations, similar as sorting, that have a low calculation to bandwidth rate, it's of great significance to maximize coherent memory accesses.

VI .PERFORMANCE FACTORS OF GPU

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

- **Memory Bandwidth**:

Its the data transfer speed between the graphics chip and local frame buffer. The More bandwidth usually gives better performance with the image to be rendered and very high resolution.

- **Fill Rate:**

Fill rate is defined as the number of pixels or textured pixels rendered per second by the GPU to the memory . It shows the power of GPU. The Modern GPUs have fill rates upto 3.2 billion pixels. The fill rate of GPU is increased by increasing clock given to it.

- **Hidden Surface removal:**

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

- **Memory Management:**

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

6.1 Applications Of GPU

GPU is important for making super computer like Tianha - 2 (China), It has 3120000 cores, Titan (USA), It uses a total of 561,000 Opteron 6274 16-core 2.2GHz processors along with NVIDIA GPUs, Sequoia (USA), IBM machine with 1.6 million 16- core processors. It has a performing speed of petaflops. Parallel computing also provides concurrency execution of task, solve huge and complex problem, the real world is monolithic, save time and cost and also better performance of software and hardware. Parallel computing plays a vital role on different fields like science, engineering, industrial & commercial, space, aeronautics and medical area.

- **Medical:** Used For the diagnostic of cancer patient, By Using image processing it is easy to describe specific description of tumor.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006083 |

- **Science & Engineering:** The NVidia CUDA is an extension to C language offer programming efficiency to General purpose GPU (GPGPU). It offers the high performance computing platform in remote sensing areas .
- **GPU Industry:** There are many applications have been developed to use GPU for supercomputing in various fields Like Scientific Computing: Quantum Electrodynamics, Molecular Dynamics, Mechanical Simulation, Image Processing, Registration, feature detection, recognition, filtering the Data Analysis Databases, sorting and searching and also data mining.

VII. CONCLUSION

In this paper, we studied in detailed about the armature, resemblant computing, factors, types and operations of GPU grounded parallel computing. Using GPU calculating our system performance is increases. We've to use community for the increased performance needed in order to deliver further value to druggies. A GPU that's well optimized will deliver resemblant performance more forcefully than a CPU. Then We tried to explain the significance and difference between CPU and GPU with their infrastructures. In Current days it's important to work on GPU calculating so that we can explore the real power of Digital World and as well as Artificial.

REFERENCES

[1] Rafiqul Zaman Khan, Md Firoj Ali, "Current Trends in Parallel Computing, International Journal of Computer Applications" (0975 – 8887) Volume 59– No.2, December 2012.

[2] Bhavna Samel, Shubhrata and Prof. A.M. Ingole, "GPU Computing and Its Applications, International Research Journal of Engineering and Technology (IRJET)," Volume: 03 Issue: 04 | Apr-2016.

[3] Shuichi Asano, Tsutomu Maruyama and Yoshiki Yamaguchi, "Systems and Information Engineering," University of Tsukuba 1-1-1 Ten-ou-dai Tsukuba Ibaraki 305-8573 JAPAN, 978-1-4244-38921/09/$25.00 ©2009 IEEE

[4] Ben Cope,Department of Electrical & Electronic Engineering, Imperial College London benjamin.cope@imperial.ac.uk

[5] Gurindar Singh and Aman Kaura, "Recent Trends in Parallel Computing," GIAN JYOTI E-JOURNAL, Volume 6, Issue 1 (Jan-Apr 2016).

[6] Bhavna Samel, Shubhrata and Prof. A.M. Ingole, "GPU Computing and Its Applications, International Research Journal of Engineering and Technology (IRJET)," Volume: 03 Issue: 04 | Apr-2016.

[7] Rafiqul Zaman Khan, Md Firoj Ali, "Current Trends in Parallel Computing, International Journal of Computer Applications," (0975 – 8887) Volume 59– No.2, December 2012.

[8] George Teodoro, Rafael Sachetto, Olcay Sertel,Metin N. Gurcan, Wagner Meira Jr., Umit Catalyurek, Renato Ferreira1, "Coordinating the Use of GPU and CPU for Improving Performance of Compute Intensive Applications," 978-1-4244-5012-1/09/$25.00 ©2009 IEEE.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details