

Best Solution for the Graphic Problem: A Single GPU or A Multi-GPU or An Upgradable GPU

R.M.C.L. Rathnayake^{#1}, D. Dhammearatchi^{#2}

^{#1}Department of IT, CINEC Campus

Millennium Drive, IT Park, Malabe, Sri Lanka

¹Lumiasky1502@gmail.com

²dhishan.dhammearatchi@cinec.edu

Abstract— Graphical Processing Unit (GPU) or in general graphic cards or VGA (Video Graphic Array) cards become major hardware of the computer world. After the development of the Graphical User Interfaces (GUIs) of Operating Systems (OSs), the most developed hardware after the processor is a graphic card. But a graphic card is consisting of a fixed amount of memory. It is the major limitation to use a graphic card for a certain big task because the memory limits the amount of data that can store in the graphic card. This paper briefly examines the possibility of installing more memory on a graphic card by the user according to the specific task.

Keywords— Graphical Processing Unit, Graphics Cards, Computer Hardware, Memory, Graphical User Interfaces

I. INTRODUCTION

When computer operating systems are introduced to the world the major part of the operating system was its interface. After adding more colorful graphical user interfaces for the operating systems the companies needed such enhanced methods rather than the Command Line Interface (CMD). The result was a GUI. By using GUIs users can handle the system very easier than the CMD. The computer needed a highly accelerated unit to display rich graphic content. Now we called it the computer graphic card.

Up to now, there are two major GPU (Graphical Processing Unit) vendors in the market. Both companies upgraded their graphic cards with the latest technology and more competitive technology. NVIDIA is a name in the graphic chipset industry with a tremendous brand name and a very big market value [1]. For every eighteen months' time gap NVIDIA releases their latest high technological graphic chipset. TURING architecture is the latest architecture that NVIDIA uses to create its new tremendous graphic cards [2]. Up to now, the best graphic chipset is called as NVIDIA RTX 30 series. Those chipsets are created with the latest technologies and the same TURING architecture NVIDIA used to create their previous graphic cards.

The next major graphic chipset vendor is AMD (Advanced Micro Devices) [3]. AMD graphic chipset series is called AMD Radeon and AMD Vega. AMD is using as their architecture called AMD VULCAN architecture. The VULCAN architecture is created by AMD to give a fast performance to the user [4]. Up to now, both companies have a very good competition throughout the graphic chipset

industry. The latest graphic chipset released by AMD is AMD Radeon RX 5000 series [5].

Most of the time those GPUs are used to create more powerful, more customized graphic cards to give more and more performance to the users. There are many more subsidiaries that created graphic cards. ASUS, MSI, Gigabyte, EVGA, and Zotac are the topper most famous graphic card manufacturers of the time.[6] Every graphic card brand has its technology and its unique performance when comparing with other graphic card manufacturers [4]. however, all those graphic cards are manufactured by the GPUs from major companies, NVIDIA and AMD.

By using a single graphics card, it gives some limitations to the user. There is a maximum number of tasks that can be done by using a graphic card. As a result, both companies created different technologies to use more than one card at a time for the workload. By using those technologies users can use more than one card at a time and the user can parallel use the cards at the same time. The technology that came from NVIDIA is called SLI (Scalable Link Interface) [5]. By using NVIDIA SLI technology, the user can add up to four graphic cards at the same time. To use the SLI technology user needs a module called NV Link bridge to connect those cards to each other.

From AMD the technology that came to the competition for multi-GPU is AMD Crossfire technology. As same in the SLI technology, using Crossfire technology users can use up to 4 graphic cards at the same time [6]. Crossfire technology was not much famous as SLI technology from NVIDIA because most of the time users create a scale, the best graphic cards are the cards that have NVIDIA chipset for multi-GPU scale. The reason was most of the latest software is recommended to use with NVIDIA GPUs to take the most UX (User Experience) from the software.

II. OBJECTIVES AND RESEARCH QUESTIONS

The section indicates three general and specific objectives that the author intends to study as indicated below:

General Objectives

1. Compare single GPU and parallel GPU technology.
2. To Implement a new graphic chipset which can upgrade its memory.

Specific Objectives

2. Is it improve the performance of a graphic card, by upgrading the VRAM?

The section covers three research questions that the author intends to research which are indicated as shown below:

1. Can multi-GPUs be replaced by a single GPU to gain a higher performance?
2. Is there any possibility to create a graphic card that has several memory card slots?

III. LITERATURE REVIEW

Graphic cards are used in the computer industry not only used to display the output of the process from the system. They are used to do far more complex things. The most and the latest example is because of the high processing power graphic cards are used for cryptocurrency mining. For cryptocurrency mining, it needs very high processing power. The processor is not enough for all the tasks in the mining process because the same processor is used to control all the computer processes at the same time. Therefore, GPU processing power is a very rich power source for the mining process [8]. For the mining process, the miners used very high-end graphic cards because all high-end graphic cards are extremely powerful. Now the most famous cryptocurrency mining process is blockchain. Today the most single powerful graphic card for the mining process is NVIDIA GeForce RTX 2070 super [9]. By using more than a single GPU the mining process can be accelerated because always two or better than one. By accelerating the process, the mining process is enhanced. Therefore, in the cryptocurrency mining field multi-GPU is a good way to increase the performance of the whole system and it improves the efficiency of the process.

Deep Learning (DL) is another field that needs a very high GPU power [10]. Deep learning is a subset of the more extensive collection of Machine Learning (ML) techniques. According to "FRANKDENNEMAN.NL", it briefly explains that GPU is a requirement for ML because always a huge amount of data is used in ML. Without a GPU it took too much time for the processing. But the efficiency is increased with the use of a GPU [10]. Therefore, most of the time to have a higher performance there should be a GPU. With the use of multi-GPU, the performance can be more increased because several GPUs are doing the same task at the same time. Again, efficiency is more increased because the process is being in a parallel way. This paragraph proves that multi-GPUs are enhanced the performance of the DL systems. Therefore, this is another fact that multi-GPUs are high performance rather than a single GPU [11].

Another most famous sector that uses the most of the GPU power is 3D rendering and video editing [12]. For these kinds of graphical tasks, it needs very high-performance graphic cards. According to "bison-tech", the best graphic card for 3d rendering is NVIDIA RTX 2080 Ti when the paper is written [12]. The best way to rendering a graphical task is by using a multi-GPU platform. By using two or more graphic cards at the same time graphic cards enhance the performance and quickly render the task with a minimum amount of time. The time required for rendering is the biggest problem in the field

1. What is the most powerful technology, a single graphic card, or multiple graphic cards?
of 3d modeling because it needs more time for rendering. When the task has many more pixels then the time is increased. By using the latest technology, the time can be reduced. Because always using multi-platform the rendering task becomes very quick. Therefore, another use of parallel graphic cards is, that enhances the power of 3D modeling.

According to the above-mentioned situation always it proves that using two or more graphic cards is efficient rather than using a single graphic card. However, it did not depend on the number of cards. Most of the time using the latest technology users can reduce their time and gain a very rich user experience. With the latest GPU technology by using a very high-end graphic card can cost a lot of money, but without using several cards a single high-end graphic card can perform well with the system. But if the task needs more graphic power then the best solution is to use parallel graphic cards.

To use a computer very fast and accurately, it needs a good processor. Then the computer needs a RAM (Random Access Memory) to store the data. The motherboards have several RAM slots to add more RAM cards to the computer to speed up the computer system. The motherboards have a special kind of operating system to understand all the input and output devices to the motherboard. It is known as BIOS (Basic Input/Output System). BIOS can understand all the devices that are attached to the motherboard. Like in the motherboard graphic card needs a very big amount of VRAM (Video Random Access Memory) to give the best performance of the graphic card. However, all the released graphic cards have a fixed amount of VARM. The best and the latest graphic card from NVIDIA, NVIDIA GeForce RTX 2080 Ti. It has 11GB GDDR6 (Graphics Double Data Rate) memory. AMD Radeon RX 5700XT is the best graphic card from AMD and it has a memory of 8GB GDDR6 [13].

To avoid the limitation of the VRAM of a graphic card the graphic card manufactures can create an area to install VARM cards. To create such a customized graphic card the graphic card, should have the capability to read the input. Like BIOS to the motherboard graphic card should have a system called BIS because there should be a way to read the newly installed VARM. According to the authors' opinion by adding more VRAM to the graphic card, it can increase the amount of data that can store. By using this technology, it is a better way to create a new graphic card that can customize by the user to the user's method. But the question with the technology is the GPU must be created to understand the new hardware installations to the graphic card's motherboard.

The new technology mentioned in the above paragraph will help to create a mid-range high-end graphic chipset and there is another marketplace for VRAM cards. The most important thing is, for a certain specific task the users can create a specific unique graphic card according to their purpose. Therefore, it will become a low-cost graphic technology with different kinds of chipsets and VRAMs.

IV. METHODOLOGY

This research was conducted using primary and secondary quantitative data. Firstly, a study was conducted to analyze the data collected from all major graphic-card vendors and all major GPU vendors. In addition to that, another qualitative study was done to analyze all the gathered data from previously published research papers. Secondly, a study was conducted to calculate the quantitative data. A survey was done by focusing on most of the graphic designers, game developers, 3D modelers and researchers related to the subjects of data science and Artificial Intelligence. The above group was used because the mentioned category needs more graphical power in their industries.

A questionnaire was created by focusing on the above group because for the study, the author needs to use the people who are related to the industries which use more graphical power. The sample was not a large one, because there is a smaller number of professionals who related in the mentioned fields. A few interviews have been done with some experts who use the latest technology and the latest graphic cards with their systems. All the information in this paper was collected by the mentioned quantitative and qualitative data.

V. RESULTS

From the research conducted by the authors, there are so many solutions for the graphic problem. The best answer is to buy a high-end graphic card that costs most of the money of the user. From the research, the authors found an answer that, most of the users are going to buy or bought the high-end cards who can go for a highly costing one. Therefore, a very high-performance graphic card can easily use for the tasks that are mentioned in the paper. But there is a question with the high-performance single cards. It doesn't have any ability to run parallel tasks. The solution is user need to buy two or more cards for parallel processes. Those cards must have same specifications with the same technology. Again, it is costing more money of the user.

There are advantages and also disadvantages by using multiple graphic cards in a single system. The most unique advantage is the system becomes an extremely powerful computer system with those two graphic cards. The system is much more enough to perform all the tasks mentioned in the paper. Parallel GPUs give high-performance as well as work as an accelerator for parallel applications [14]. That is the major advantage of the parallel GPU computing. Then as mentioned above in the paper, mining cryptocurrency become very easy because of multi-GPU systems

In other high-performance tasks like DL, ML, and Neural Networking, everything became very easy to perform using the mentioned technology. As a result, Artificial Intelligence (AI) and related technologies become very familiar to our day to day lives. Therefore, most of the industrial and day to day life's computational activities took the advantage of parallel GPU computing. However, the main disadvantage of the multi GPU platform is it needs a very big power to power up the system. And a very powerful cooling system to control the heat generated inside the computer system.

According to the survey conducted by the author, the results are as follows. The survey was conducted by using a group of people who are related to the industries such as game development, 3D modeling and animation, machine learning, and deep learning with the addition of data science.

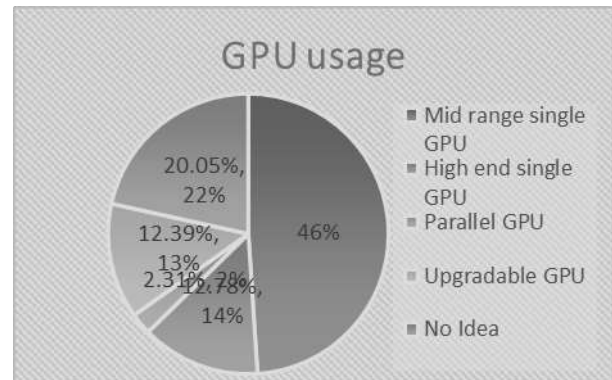


Fig. 1 GPU Usage

From the results, it generates more ideas about the knowledge and the usage of the graphics cards. Most of the users around 45% are agree with mid-range cards according to their budget range because mid-range cards can add more performance to the systems but less than high-performance cards. Then the maximum number represents the users that they don't have any idea about the graphic cards and also the performance of the graphic cards. It shows us the knowledge of the users about this kind of technology. 14% of all users are agree with the high-end graphic cards. Therefore that 14% have the chance to go with the high-cost graphic cards. And most of the industry level game developers and 3D animators are in this category because they used very high-end technology for their workload.

Then the maximum number of users are agreeing with the technology that is going to be implemented in the future. That is 12.39% of all the users. The statistical data from the survey, shows us that most of the users are need to use high-end technology and performance. But they are facing some limitations in their lives. Therefore the newly implementing technology is the best solution for them.

The least number of users are belonging to the category who are using or going to use parallel GPU technology. Again, some of the game developers and data science researchers belong to this category. They also agree with the implementing idea because they have simple solutions, rather than being buying another graphic card for their computer system.

The last technology described in the paper was to create a graphic card that has the opportunity to upgrade its VRAM. That means the user can customize the graphic card according to the task that the user needs. To create such a graphic card, new technology should be implemented with a new type of graphic chipsets and a new type of graphic motherboards. It will be a low-cost high-end graphic technology that everyone can afford. As a result of it, another new technology should be created. A RAM card that consists of GDDR technology because all the graphic cards are consisting of GDDR memory.

Or like other mentioned technologies a new technology should be created for the memory cards.

With the discussions had with the tech experts, game developers, Artificial Intelligence researchers who have much more experience in the graphic industry, they have much more different ideas. Most experts take the side of a single GPU which consists of the most tremendous and high-end technology. The reason for that was the cost that they have to pay to buy a graphic card is less than the cost that they need to buy two cards. By the way, their condition is fine with their budget range and the workload that they have to do. This is not a complete survey because the number of participants is very less and the only category that the authors choose was only the experts who are in the fields of graphic designing, 3D modeling, game development, data science, and machine learning.

VI. CONCLUSION

Due to the lack of technological knowledge and experience, most of the users are going to use technology that is not enough for their specific tasks. Therefore, most of those users face much more difficulties using their computer systems because of the previously mentioned situation. The selection of graphic cards has also become difficult because the user does not know which kind of graphic card that they need and which kind of specifications should in the graphic card. Especially the clock speed, and the VRAM. Because of that reason users are spending more money on the graphic cards even it is affordable or not. A very few can take this chance but most of the people remaining with a mid-range card. Therefore, those users can buy cards with less amount of memory. It is a very big question when doing a specific task that needed a big amount of VRAM. But the graphic cards VRAM is a fixed amount and it cannot be upgraded. There is a new technology for the question that is mentioned in the paragraph.

Throughout the paper, it describes a new graphic technology that can upgrade its VRAM of the graphic card. With the implementation of the new technology, there will be another great solution for the graphic problem. Therefore, all the users will have another option to go with when buying graphic cards. It will be a low-cost high-performance solution for all the users mentioned in this document.

In conclusion, modifications must be implemented in all sectors of graphic chipset and graphic card manufacturing industry, as well as motherboards. With the results that generated from the survey show us most the users are waiting for new high-tech solution for their graphic problem. Therefore, the new technology will be the best answer and the only solution for their budget and also to their requirement.

VII. ACKNOWLEDGEMENTS

I would like to express my full gratitude to Mr. Dhishan Dhammearatchi and department of Information Technology for their primary encouragement and supervision of this research.

REFERENCES

- [1] Nvidia popularity & fame | YouGov", *Today.yougov.com*, 2020. [Online]. Available: <https://today.yougov.com/topics/technology/explore/brand/Nvidia>. [Accessed: 04- Dec- 2020].
- [2] Nvidia.com. [Online]. Available: <https://www.nvidia.com/content/dam/en-zz/Solutions/design-visualization/technologies/turing-architecture/NVIDIA-Turing-Architecture-Whitepaper.pdf>. [Accessed: 31-Aug-2020].
- [3] 2020. [Online]. Available: <https://www.amd.com/en>. [Accessed: 04- Dec- 2020].
- [4] B. Thomas and B. Thomas, "Nvidia GeForce RTX 3080 release date, price and specs: all the RTX 3080 rumors," *TechRadar*, 27-Aug-2020. [Online]. Available: <https://www.techradar.com/news/rtx-3080>. [Accessed: 31-Aug-2020].
- [5] F. F. (Author), "GRIN - design and Implementation of a Vulkan Engine," *Grin.com*. [Online]. Available: <https://www.grin.com/document/456305>. [Accessed: 31-Aug-2020].
- [6] A. Verma, "Top Graphics Card Manufacturers & Brands for Nvidia & AMD GPUs", *Graphics Card Hub*, 2020. [Online]. Available: <https://graphicscardhub.com/graphics-card-manufacturers-brands/>. [Accessed: 04- Dec- 2020].
- [7] Amd.com. [Online]. Available: <https://www.amd.com/en/graphics/radeon-rx-graphics>. [Accessed: 01-Sep-2020].
- [8] A. Verma, "Top Graphics Card Manufacturers & Brands for Nvidia & AMD GPUs," *Graphicscardhub.com*, 06-Feb-2017. [Online]. Available: <https://graphicscardhub.com/graphics-card-manufacturers-brands/>. [Accessed: 01-Sep-2020].
- [9] Nvidia.com. [Online]. Available: http://developer.download.nvidia.com/whitepapers/2011/SLI_Best_Practices_2011_Feb.pdf. [Accessed: 01-Sep-2020].
- [10] Deep Learning with GPUs - Run:AI", *Run:AI*, 2020. [Online]. Available: <https://www.run.ai/guides/gpu-deep-learning/>. [Accessed: 04- Dec- 2020].
- [11] Jon Story AMD Graphics Products Group, "Harnessing the performance of CrossFireXTM," *Amd.com*. [Online]. Available: <https://developer.amd.com/wordpress/media/2012/10/Harnessing-the-Performance-of-CrossFireX.pdf>. [Accessed: 01-Sep-2020].
- [12] NVIDIA GPU Rendering Solutions for 3D Designers", *NVIDIA*, 2020. [Online]. Available: <https://www.nvidia.com/en-us/design-visualization/solutions/rendering/>. [Accessed: 04- Dec- 2020].
- [13] T. Warren, "Bitcoin mania is hurting PC gamers by pushing up GPU prices," *The Verge*, 30-Jan-2018. [Online]. Available: <https://www.theverge.com/2018/1/30/16949550/bitcoin-graphics-cards-pc-prices-surge>. [Accessed: 01-Sep-2020].
- [14] M. Hanson and M. Hanson, "Best mining GPU 2020: the best graphics cards for mining Bitcoin, Ethereum and more," *TechRadar*, 18-Aug-2020. [Online]. Available: <https://www.techradar.com/news/best-mining-gpu>. [Accessed: 01-Sep-2020].
- [15] About, "Multi-GPU and Distributed Deep Learning - frankdenneman.nl," *Frankdenneman.nl*, 19-Feb-2020. [Online]. Available: <https://frankdenneman.nl/2020/02/19/multi-gpu-and-distributed-deep-learning/>. [Accessed: 01-Sep-2020].

- [16] H. Cui, G. R. Ganger, and P. B. Gibbons, "Scalable deep learning on distributed GPUs with a GPU-specialized parameter server," Cmu.edu. [Online]. Available: <https://www.pdl.cmu.edu/PDL-FTP/BigLearning/CMU-PDL-15-107.pdf>. [Accessed: 01-Sep-2020].
- [17] About, "Multi-GPU and Distributed Deep Learning - frankdenneman.nl," Frankdenneman.nl, 19-Feb-2020. [Online]. Available: <https://frankdenneman.nl/2020/02/19/multi-gpu-and-distributed-deep-learning/>. [Accessed: 01-Sep-2020].
- [18] D. James and D. James, "The best graphics cards in 2020," PC Gamer, 26-Aug-2020. [Online]. Available: <https://www.pcgamer.com/the-best-graphics-cards/>. [Accessed: 01-Sep-2020].
- [19] "Ubiquity: GPUs: High-performance accelerators for parallel applications," Acm.org. [Online]. Available: <https://ubiquity.acm.org/article.cfm?id=2618401>. [Accessed: 01-Sep-2020].

