IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

MODERN APPLICATIONS OF GRAPHICS PROCESSING UNIT (GPU)

Mohish Khadse, Aditya Vaishale, Suraj Chaudhari, Aditya Ingale, Khushi Junnare,
Second Year B. Tech Students,
Department of Electronics and Telecommunication,
Vishwakarma Institute of Technology, Pune, India.

Abstract: The paper is mainly designed on the study of Graphics Processing Unit (GPU). We know that the GPU is a main component of the CPU and it has many functions to do. And nowadays it is used in most of the processing applications. This paper gives a glimpse about those applications which require the GPU for computing. The mentioned research papers give the validation that by integrating the GPU in the computing, system processes can be made fast.

I. INTRODUCTION

The GPU which stands for Graphics Processing Unit. For both private and business computing it has become a most vital form of computing technology. The GPUs are designed for the data processing and it is widely used in most important applications which includes video and graphics rendering. Although the GPUs are most popular due to its capabilities in gaming, but due to those capabilities, they are also used in creative production and in computer science most commonly in Artificial Intelligence (AI). By using the GPU graphics programmers can add more interesting realistic scenes and more interesting visual effects with advanced techniques like lighting and shadowing. To accelerate dramatically for the additional workloads in HPC (High Performance Computing), deep learning and other areas the developers are also using the abilities of GPU. Graphical Processing Unit (GPU) was first introduced by Nvidia and t Nvidia popularized the term GPU in 1999. and that we know that Nvidia is the most leading company in GPU and Graphic Card. [1]

II. LITERATURE REVIEW

The GPU may be a specific meat killer effective at maneuvering and affecting camcorder work. There are mathematically-exhaustive tasks, complex algorithms that set a pressure on the CPU. The GPU carries this load from the CPU, freeing the phase that follows additional work. The GPU is a chip that operates with the same functions as the CPU. It's nothing to try and do with any part of the system outside of the graphics package. The top results are a better managed graphics package and better latent period supported computer commands for gaming and other applications. The Graphics Processing Unit or GPU is assumed to lessen the burden of the CPU. The instrument is valuable generally to those that have exhaustive graphical functions on their calculating. If you play program tricks and presentation videos on a computer network, establishing a GPU will considerably help the conduct you avoid the complete act. Migrating large and whole-grain algorithms to a GPU coprocessor gives answers much faster and reduces fidelity times accordingly. Since it's very difficult to transfer these markups, individuals prevent utilizing complex dossier constructions and to foreshadow in their parallel algorithms. GPUs can never be a complete option for CPUs but they may be complete. [2]

In this article, they estimated a methodical structure for controlling different GPUs systems in a cloud environment by observing its performative and causative consumption. Preliminary results show that allocating some sufficient blocks and clothes to each block is causing success by 13.1% substance-dependent success in the Fermi GPU and 11.2% of capacity in the Kepler GPU. The virtualization of calculating resources to admit end-consumers to approach bureaucracy at an acceptable price is influences by cloud estimating. In the cloud foundation, it requires few things to do for staying fully responsive to the consumer needs. It involves methods that earn a tremendous amount of strength usage such as Virtual Memory (VM). The aim concerning this paper search out establish an orderly design to control the giving of GPU resources usually purpose Cloud calculating surroundings for comprehensive purpose use.[3]

Graphics Processing Unit (GPU) is an inseparable part of a computer and is used to render user interfaces, play games, and run complex algorithms for many non-graphical applications. As these applications become more over period, there is rise in demand for higher order GPU requirements in SoCs. There is a rising demand for high performance rendering sophisticated user interfaces, the new 3D games, and GPGPU computing applications. This article proposes the improvement of GPU performance. This is done by influencing a transparent DSP backward without affecting its software programming model. The expected answer includes new diverse approaches, namely the ability to offload the best shaders to DSP, 3-step sequence execution, and in-path reusability. [4]

We have a propensity to present an ascendible and flawless effectiveness prediction framework for gpu, ppt gpu, to overcome these difficulties. However, due of the very fast rendering, accessing the world memory might cause performance issues because the appropriate access latency is considerable when compared to other on-chip memories. The L1 cache is for SMs, while the L2

cache is for all SMs. Furthermore, each SM includes a shared memory connection that allows threads to communicate with one another. All threads that produce from the same kernel are categorized into a grid in CUDA terminology, and they all share the same world memory.

Since totally different ALU directions need unique cycles to end execution, instruction's tendency to be late, square measure additional to things within the task list whereas checking the PTX files looks at the sort of the instruction and therefore the GPU design is evolving. Then within the pipeline, for each memory instruction they have calculated the access time which is dependent on configuration in file dependencies and factors, as well as memory port numbers, information measures of memory and also the size of memory which allows the transfer from the memory queue in synchronic manner. We built each benchmark's kernel to run just once, and if a benchmark only has one kernel, we have a tendency to obsess over each kernel's findings separately.

Since we are able to change or disable caches on the $64000 device in time of compile time, we have a tendency to report the performance of the $64000 device(s) once whereas turning off L1 cache and victimization L2 cache solely and whereas victimization each cache.[5]

Machine learning development has become a highly vital developer resource with the rising use of human data such as face detection, biometric data, and significant statistics. As a result, facial footage and video inspection tools are becoming more prominent, resulting in a rise in permeable individuals. As we can see now, traditional systems like cryptographic network elements are very expensive, slow to act, and have very little security. As a result, they later worked on quantum computer assaults. The intended face security system was enhanced utilizing the high-performance parallel computing technology, which was based on Nvidia GTX 2080 TI (GPUs). Separate computations for each face frame (190* 190y px) essentially just 2.2, 2.7 MS entire total for cryptography and secret writing, respectively, with security settings n = 1024 & 2048. An evaluation of security criteria found that the proposed system was equivalent to previous methods in terms of secrecy.

All implementations were performed on a mainframe Intel i9-999KF with Windows 10 64-bit system, Nvidia GeForce GTX 2080Ti GPU with 8 GB DRAM and 32 GB DRAM to test the predicted cryptosystem effectiveness. CUDA Toolkit v10.1.243 and OpenCV v4.1.2 technologies were chosen.

They evaluated the topic's suitability for hidden writing and interpreting victimization detected facial parts from a clip. A two-minute-long input video in MPEG-4 format with 720 × 432 element quality at 30 frames per second was used to test the face detection approach. In addition to the chart study that was performed to facial color pictures, they have a tendency to compare the efficacy and security of the proposed cryptosystem with previous investigations, it was tested on a Lena River grey scales image with a size of 512 x 512 elements. Information entropy, CC, NPCR, and UACI these are some number of security methodologies, to verify the anticipated system lustiness.[6]

In this study, they began by doing an in-depth survey of the current BU techniques utilized in GPU representations of RSM. They experimented with CUDA to enforce all of the current GPU-based performance. One of the most modern solutions, that modified the matrix B column by column, was said to outperform all others. Its ability to deal with large obstacles, however, severely hampered it. As a result, the US developed a BU technique that is not cost-effective. They tend to improve a recent approach that used an element-by-element technique to change the matrix B in two steps for this aim.

Furthermore, it was fully capable of addressing all six examination issues' memory needs. On the other hand, the column-wise method proved unable to meet the shared memory needs of the two biggest examination questions. These findings, we believe, support the employment of our intended BU technique in dense platter RSM downside solvers. This claim was related to the fact that changing any portion of the matrix B with our method just requires the current price of an element and two alternative vectors.

In addition, the results of partial column computation by strands in Step 2 of the column-wise method deserve more investigation. Future research could look into the performance of our proposed BY method in a batch-processing system.[8]

The paper [9] talks about What and how actually the Graphics Processing is Done. So basically, GPUs are used to render the graphics on Display. The graphical information of the image is provided in the series of attributes and vertices to render the image. Basically, the vertices mean the desired geometry format of the image and the attributes means the associated data of the image like colors, lights etc. So, the Graphics Processing Unit processes this information in such a way that the new set of images is created and it is displayed. In this paper the architecture of NVIDIA's GeForce GTX285 GPU is given. In this there are many blocks that are used, like the host interface which is used to interface with a PCI express bus. The Geometric parameters of images like points line and rest of the associated data of that image is given as input to the input assembler. The architecture also uses the vertex work distribution which performs necessary computations. So, this paper shows a glimpse on massive parallel architecture and GPU operations. In the case of silicon technology, the performance of the GPU is controlled by the density of the transistor with low output stress and process robustness with minimum variability.[9]

They used a graphics processing unit (GPU) to software programmer basic debris along with high-strength physics in this work. Compared with the CPU packages, they received around 50 instances of common higher overall performance on GPU. GPU (Graphics Processing Unit) in the beginning evolved to carry out intense calculations important to render complicated transferring pictures on computer displays. Recently, they were used in simple debris and excessive strength physics. For physics techniques at the LHC, they have developed GPU variants of cross-phase computation tools. Those packages display pretty fifty instances of better overall performance in execution time at the common. The GPU's large number of multi-processors makes it possible to reduce the time it takes to compute these packages. [10]

Basically, in this paper though it is from 2008 but they talked about how the future GPU will be. Instead of increasing a single thread we will increase the core for better performance. They also tried to give an example of the Sony ps3, how it was a powerful machine of that time. and 2008 was that era where the work of graphics was growing. Then they talked about they have a full-fledged parallel programmable processor with extra special features GPU Architecture. in the part of architecture, we can't forget about graphics pipeline which is part of GPU itself, consists some steps or operation such as vertex, primitive assembly, rasterization, Fragment operation and the next point is the evolution of GPU Architecture. In 2008, their Shader model 4.0 was the most recent, which meant they were expecting 32-bit integers and 32-bit floating-point numbers, direct and indirect reads from global memory, and finally dynamic flow control and branches. They also discussed the GPU programming model, where they were attempting to discuss the GPU's programmable unit for efficiency, they imply that the procedure is simple to follow, but that it is harder to express [that general purpose computing terms]. They also talked about Software environments where mainly in the past they used graphics APIs. but for faster operation they also used other methods like pseudo assembly language and DirectX 9 higher level shader programming. In the computational and application part they talked about computational primitives, Algorithms

and applications and some theory about recurring themes. and they also consider the case study about game physics where how intense graphics of some games are processed. Finally at last they talked about future scope and prediction of application of these advanced and powerful GPUs.[11]

In the world of GPU we have two leading companies AMD and all-time best Nvidia, so in this research paper they are trying to explain us that how AMD GPUs are becoming an alternative of expensive Nvidia GPUs in real-time workloads, in this case study they discussed about AMD GPUs and how they are updating their algorithm for better performance in market against Nvidia, but we all know that Nvidia hold 77% of GPU market and it is currently (I mean in 2022) it is still at next level for delivery of performance in the field of gaming or in field of content/ graphic creation, even the great association like NASA consider Nvidia's GPU for their powerful missions, operation, extra-terrestrial/astronomical high intensive graphical work. and in this AMD is also trying to prove their GPUs worthy. AMD GPUs for real time Workloads in this AMD is working on how multiple applications can safely and efficiently share, these guys also concern about the weight, size and power consumption. but still there are some drawbacks of AMDs GPU like instability from less-mature software, less documentation support. In the part of Background and Related work, the subtopic of this is overview of GPU Programming in this there are some steps and these steps are common in Nvidia and AMD GPUs. GPU programming API. API Queues also called streams. and these process in FIFO order, for an Nvidia based system the API of choice is CUDA, and in AMDs their choice is HIP. In this they explain the organization of AMD RX 570 GPU with an understandable diagram. Now the most interesting part is the Comparison between Nvidia and AMD, 1. software for GPGPU Programming as they mentioned already Nvidia used CUDA software stack and AMD uses ROCm [Radeon Open Compute] and they also discussed some benefits of this ROCm that is open source and existing CUDA can often automatically be converted to HIP Programming language. and further they deeply studied AMD GPUs real time workload with some drawbacks, ROCm documentation and all also about the modification in ROCm. They were evaluating this AMD GPU using Pytorch deep learning tool along with some visualization before conclusion they compare some GPU of Nvidia, AMD and intel scores, in the conclusion they talked about that this potential GPU company is to reshape real time GPU research and how they are moving near to Nvidia with baby steps.[12]

This paper presents the method for power consumption of graphics processing unit with respective the computing unit which are in operating state. It relies on the multiplication of the matrices and also the first goal of the paper is to research various modes which are functioning, of graphics processing unit and power consumption with respective to the number of operating gpu units. This paper also shows the glimpse how power consumption varies as the number of functioning units or streams varies. As per the energy measuring platform, the current sensor with the microcontroller is used. For testing purposes, the NVIDIA GeForce GTX 489 used a program for summing matrices with stable threads and varying number of blocks.

Results: The power consumption of the GPU relies on the number of locks and streams of video chip. Also, by using the optimal GPU resources and power consumption in the computing process can be reduce to 17%.[13]

The paper [14] covers how the development of communications in networks and technologies in multimedia are taking place rapidly. So, to meet such high-performance processes or we can say high performance computing, the Graphics processing unit (GPU's) like parallel architectures are used. But there are many ways to program this GPU and also there are many ways to get the most effective execution time of the GPU but that is commonly seen as an art. Study on the threads of GPU is performed in this paper. Also, inside the High Efficiency Video Coding, the number of Blocks that end up in a Prediction Unit of 64x64 (PU64) compute (HEVC) and The Compute Unified Device Architecture (CUDA) proposes it. They used many methods that are used to describe the optimal execution time for the GPU for better performance. The experiments that are performed in the papers result that the most effective method that is used to run the GPU is obtained at 128 blocks and 32 threads. The topology used in the paper gives the minimum time for the execution of the GPU as compared with the CPU where the speed of operation is around 50% more than CPU. [14]

In blockchain, angered transactions, together with non-legal ones, can't be changed just like bank transactions. To pause the harm due to non-legal transactions, speedy anomaly supervision of transactions is needed as a result of transactions that are often changed before their approval. However, existing abnormality detection strategies should method all transactions in blockchain, technology and the time interval are longer as compared to standard approval time. This paper put forward a model which is based of subgraph-irregularity detection technique to detect an employing as an integral part of the blockchain knowledge. The preplanned structure of the subgraph is appropriate for graphics process units (GPUs) to faster the detection by victimization multiprocessing. In associate analysis victimization real Bitcoin transaction knowledge, once the amount of aimed transactions was 100, the planned method was eleven.1x quicker than associate preexisting technique which is based on the gpu while not reducing the detection accuracy. Structure illustrates the general structure of the projected methodology divided into the CPU facet and GPU facet. At first, the User Graph is made from authentic transactions on the CPU facet. Then this Graph is sent to the GPU and it is inserted in the cached int the GPU. Once the detection requested for the GPU facet, the subgraph is created int the GPU. If many simultaneous transactions square measure targeted, the subgraphs square measure incorporates throughout subgraph creation. The anomaly detection, which incorporates a feature extraction and detection formula, is performed at the same time for a number of transactions victimizing the incorporated subgraph. If one dealing is aimed, the subgraph isn't incorporated and can be used for fault detection. The incorporated subgraph and also the aimed many dealings type is that the extension supported the case which is being targeted transaction. In a system, where block chained is used it's important to notice abnormal transactions at super-fast speed as a result of transactions that can't be changed after approval.[15]

This paper describes the effective implementation of the streaming application on the Graphics Processing unit. Basically, it is an automatic code generation solution. This model shows the effective realization of a streaming application framework on a single and number of GPU simultaneously by dividing and automatically mapping the streaming application. This framework takes into consideration the Graphics Processing pipeline architecture and individual memory hierarchy. While doing so it provides the maximum performance and scale of application This model has built on the backside or you can say underline to the Stream It Programming language compilers. It ensures the effective working of on-chip memory hierarchy using the combinational mix of computation and memory access threads. First, the methodology used is not a standard approach for graphics processing unit (GPU) computing, and this is because it uses a huge number of similar threads. Another, it implements the stream graph partitioning algorithm to hold number of applications. And last it also implement the complexed applications on the number of Graphics processing unit (GPU), this is done by using the efficient pipeline execution scheme.[16]

The paper [17] contains some of the methods which Ares used to improve the mobile phones performance mainly the smart

phones. The methods mentioned in the paper are such that they not only improve the execution speed but also reduce the power consumption on the GPU in cell phones. In these for the improvement in performance of GPU there are some mechanisms are added like data forwarding and hazard control. And for the reduction in the power consumption the methods used in the paper can disable the unusable memory units or executions. So firstly, to get the feature of high performance mentioned above there is a change made in the architecture of the GPU. After getting improvement in performance there is one analysis takes place which shows that when the accessing of memory takes place the specific portion consumes more power. So, to get into rid of that they have used certain method. After all the experiment they found that the actual operating frequency for the original designed is 133MHz and the Version or we can say design they have made has reached the operating frequency of 247MHz after running the ultra-commands to generate the optimal results. One important point they have mentioned that the improved version has only 4 pipelines where as for the original design there are 5 pipelines.[17]

Nowadays a huge exponential growth is being observed in the use of mobile devices. This in terms also led to the usage of hunger number of mobile applications, this application nowadays has a huge requirement of computational space, as the tasks are quite real-world problems. So here mobile experiences some challenges regarding their resources, power battery-life, low computational power and storage. Mobile cloud computing technology is a thing which is going to solve all these problems, which provides the cloud enabled resources for processing the and storage of the mobile data outside the mobile. This cloud provides additional development tools such as on cloud CPU and Gumby using the gpu, the solution for the real time face detection is proposed in this paper. The solution which is given in the paper shows the mobile GPU cloud computing which affects the factors like speed and accuracy in positive manner for the facial detection system. In this paper they have evaluated the applications of mobile devices like face detection and cloud structure. and result shows that mobile cloud computing experienced the a less execution time and also saves the mobile device resources as well.[18]

Latest generation graphics processing units (GPUs) are very strong, they provide a very fast memory access and have excellent video decoding support. In a recent era, these GPUs are mainly used for a gaming purpose, though the modern applications are shifting on GPU for computing, but the major part is in the gaming. But these GPUs also provide the efficient computational performance for Image Processing, which makes them applicable for the use of CCTV rendering purposes. Old orthodox CCTV system was based on an array of video decoding cards with non-digital output, i.e., output was analogous in nature. This type of the system can be restored with the computers having suitable Graphics cards. Doing this will create tremendous advantages over the traditional system in terms of network installation, it can render from any source which Computer receives that can be HTML Captures VNC. Not only this given application but also, we can do the video analysis, with the help of GPU. In this paper authors provide the system to render and decode the multiple videos from various platforms with the help of modern GPU. They have used the Raspberry Pi for this application, which is capable of rendering 100 video sources or you can say streams and decoding as well. They have done this in 5 steps.

First, reading the videos and images from multiple sources. Second, step involves decoding the images and video streams. This is done by CPU decoding and GPU decoding. Third, Composition after creating the content there is a need to generate the final frames using the combination of all the generated frames, this process includes the scaling, positioning and smoothing. Fourth, this process involves the rendering of the images and video streams. Fifth is Multiple GPU. It is very rare that GPU to have more than 6 output spaces, so this gives the need for integrating the multiple Graphics Processing Unit. [19]

Nowadays face acknowledgment use is widely used in miscellaneous activities to a degree in traffic, security, medical architecture, etc. In this article, they have provided an effective match against ability and resistance to promote the use of facial recognition. It includes Xilinx Zynq (ARM + FPGA) and Jetson TK1 from NVidia (ARM + GPU) with PCIe test. The JetsonTK1 board is based on the NVIDIA Terga SoC. This application includes a two directional ARM Cortex-A9 based CPU. On-chip mind, extrinsic thought interfaces, and a diverse collection of peripheral relatedness interfaces are also included. Face acknowledgment use is established by the LBPH algorithm and eigenvalues. LBP and Scan consume higher opportunities than other tasks in the facial recognition process. We performed this use in a pipelined mode to capture representation and inspect written description of past events by equating it accompanying the prepared images. They then scrutinized the results and established two main variations: speed of adoption when differentiated from additional tactics, and effectiveness of established resistance to legitimate capacity calculations. They have used half HD resolution faces to capture the following imaging equipment installed with the Zynq FPGA. They achieved fifty-one times the rate of sequential queries executed in single core indebtedness. Energy efficiency is the biggest hurdle faced by geeks and rooted system designers. In this paper, they intend to strengthen effective miscellaneous platforms and real-opportunity structures. This example leverages two sets of FPGAs for threading and the GPU for extreme compute frequency ranges with a full 0.4 second wait time. It surpasses sequential kill by 65 into quick and 70 into less powerful for queries. [20]

The paper [21] mainly talks about image recognition which is based on deep learning. It also contains the applications based upon deep learning of video analysis. The analysis done in the paper is used to accelerate the predefined algorithm models. In the paper an example of the intelligent business is taken for the video analysis. This example is used to find the different types of attributes of customers in a business environment like Face attribute, Age Attribute, then we have gender attribute. Also, this paper covers the applications of Convolution Neural Network as it has a good performance in the computer vision field. Criteria's that are needed for Implementation of embedded GPU are also covered in the paper. To meet that kind of intelligence for the deep learning there should be a proper hardware so it is recommended to use ASIC specific chips such as Google's TPU, Nvidia jetson TX1 or FPGA(Xilinx). For the optimization of algorithm, they have used Google Net to analyze the attributes in the videos and the method of accelerating in frame is used in it to realize the applications in deep learning.[21]

They present a groundbreaking GPU-BASED PARALLEL AUTOMATED ECG ANALYSIS PROGRAMME that greatly minimizes overall duration expense for detecting seven specific different heartbeats patterns. For 168 hr. ECG data, GPU-AECG takes just under two seconds to process. W They leverage the CUDA GPU architecture's concurrent capability to construct GPU-AECG, ECG analysis automated a concurrency-based technique that improves calculation performance of GPU-AECG at the same time by meningeal concurrent signals and improving GPU compute resource allocation.

When compared to the CPU version employing only one processor which is intel Xeon E2600, achieves 35.0 and accordingly is being handled 24-hour. Emg signal with a one Nvidia tesla K20xm card. In a few seconds they can process up to 1000 users in 24 hours of ECG signal and provide quickest results using only 2 gpu servers for a total cost which was less than $100,000, with 16 GPU cards each, significantly improving the user experience of decrease the cost of constructing ECG evaluation systems.[22]

A GPU-based parallel genetic technique for distribution feeder reconfiguration was reported in this paper (DFR). This is a major

contribution because this is the first time a graphic processor has been used to solve this problem. The GPU, which was built with thousands of cores, gave a 66.2x speedup over a sequential CPU implementation. This speed is significant because it helps for speedy restructuring of issuing system in the case of an interruption, a failure, or merely a desired variation, allowing for an extremely efficient architecture to be maintained. In order to retain a circular architecture when generating alternatives, this study also suggests a unique encoding option depending on tree structure and makes a second contribution.

The proposed encoding was shown to be exceptionally efficient, allowing meta - heuristics to handle networks 5 times bigger than all the reference articles investigated. On circuits with 16 - 4400 buses, the suggested solution revealed methods with lower power consumption and a significantly quicker deadline than previous benchmarks.[23]

For high-intensity proton rings, particle tracking computations with space charge effects are critical. They have just developed a GPU-based particle tracking simulation code. The simulation of single-particle mechanics can be carried out as soon as possible with complete parallelization thanks to GPUs' great parallel processing capabilities. It has to generate charge densities that can't be totally parallelized. Before creating the overall histogram for the charge density construction, we fill sub-histograms in shared memory then each process can not only avoid frequent clashes but also access sub histograms rapidly. To start making use of the cuFFT package, which is tuned for high speed on CUDA-enabled GPUs, we employ DFT for the Poisson solution. In addition, the new code can simulate 117 rpm on a single tesla-V100 gpu, compared to 9 for SCTR (13) on an intel Xeon(R) Gold 6126. (2.6GHz). [24]

At present, HD television performers become to a greater extent main in the broadcast area. By applying it to the processor, HDTV animators can't completely ignore adversarial intelligence skills. GPU is time and again secondhand in augment computing rapidly. Therefore, a brand-new HD broadcast performer technique hooked up GPU electronics is projected withinside the paper. It had used the open-source TV codec VAAPI along with VDPAU backed support along with NVIDIA GPU for the purpose of fast translation. It provides the results using the computer's graphics processor on higher entertaining settings. The order given in this article is believed to favor certain video structures. [25]

REFERENCES

[1] GPU-Graphics Processing Unit Ramani Shrusti K., Desai Vaishali J., Karia Kruti K. Department of Computer Science, Saurashtra University Rajkot, Gujarat.

[2] An Introduction to Graphical Processing Unit; Jayshree Ghorpade1, Jitendra Parande2, Rohan Kasat3, Amit Anand4. (Department of Computer Engineering, MITCOE, Pune University) International Journal of Modern Engineering Research (IJMER)

[3] A Holistic Resource Management for Graphics Processing Units in Cloud Computing; Abdulaziz Alnori1 and Karim Djemame2 School of Computing University of Leeds, Leeds, UK.

[4] M. Mody, H. Hariyani, A. Balagopalakrishnan, J. Jones, A. Jayaraj and Y. A. Prithvishankar, "GPU Assist using DSP Pre-processor," 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), 2020, pp. 1-4, doi: 10.1109/CONECCT50063.2020.9198650.

[5] PPT-GPU: Scalable GPU Performance Modeling By Yehia Arafa, Abdel-Hameed A. Badawy, Gopinath Chennupati Nandakishore Santhi, Stephan Eidenbenz 2018.

[6] Efficient NewHope Cryptography Based Facial Security System on a GPU. PHAP DUONG-NGOC , TUY NGUYEN TAN (Member, IEEE), AND HANHO LEE , (Senior Member, IEEE) 2020.

[7] Performance Evaluation of cuDNN Convolution Algorithms on NVIDIA Volta GPUs. MARC JORDÀ, PEDRO VALERO-LARA, AND ANTONIO J. PEÑA 2019.

[8] Accelerating Revised Simplex Method Using GPU-Based Basis Update. USMAN ALI SHAH , SUHAIL YOUSAF IFTIKHAR AHMAD 1, SAFI UR REHMAN AND MUHAMMAD OVAIS AHMAD 2020.

[9] GPU Technology Trends and Future Requirements, John Y. Chen, NVIDIA Corporation, Santa Clara, CA USA, 2009.

[10] Application of graphics processing unit (GPU) to software in elementary particle/high energy physics field. J. Kanzaki KEK, Oho 1-1, Tsukuba, Ibaraki 305-0801, Japan.

[11] GPU Computing by John D. Owens, Mike Houston, David Luebke, Simon Green, John E. Stone, and James C. Phillips, 2008.

[12] AMD GPUs as an Alternative to NVIDIA for Supporting Real-Time Workloads by Nathan Otterness, James H. Anderson, 2020.

[13] Energy Consumption Powered by Graphics Processing Units (GPU) in Response to the Number of Operating Computing Units by Huz Miev I.K., Chipirov Z.Ah. North Caucasian Institute of Mining and Metallurgy Vladikavkaz, Russian Federation.

[14] Execution-Time Optimization Based on Thread and Block Repartitions on a Graphic Processing Unit.
By Randa Khemiri, Fatma Ezahra Sayadi, Haythem Bahri, Marwa Chouchene, Mohamed Atri Laboratory of Electronics and Microelectronics, University of Monastir, Environment Street, 5019 Monastir, Tunisia.

[15] Scalable anomaly detection in blockchain using a graphics processing unit.
By Shin Morishima Faculty of Engineering, Toyama Prefectural University, 5180 Kurokawa, Imizu-shi, Toyama, 939-0398, Japan.

[16] H. P. Huynh, A. Hagiescu, W. Wong, R. S. M. Goh and A. Ray, "Abstract: Mapping Streaming Applications onto GPU Systems," 2012 SC Companion: High Performance Computing, Networking Storage and Analysis, 2012, pp. 1488-1489, doi: 10.1109/SC.Companion.2012.279.

[17] S. Hsiao, S. Li and K. Tsao, "Low-power and high-performance design of OpenGL ES 2.0 graphics processing unit for mobile applications," 2015 IEEE International Conference on Digital Signal Processing (DSP), 2015, pp. 110-114, doi: 10.1109/ICDSP.2015.7251840.

[18] M. Ayad, M. Taher and A. Salem, "Mobile GPU Cloud Computing with real time application," 5th International Conference on Energy Aware Computing Systems & Applications, 2015, pp. 1-4, doi: 10.1109/ICEAC.2015.7352209.

[19] V. B. Saiz and F. Gallego, "GPU: Application for CCTV systems," 2014 International Carnahan Conference on Security Technology (ICCST), 2014, pp. 1-4, doi: 10.1109/CCST.2014.6987028.

[20] S. K. Rethinagiri, O. Palomar, J. A. Moreno, O. Unsal and A. Cristal, "An energy efficient hybrid FPGA-GPU based embedded platform to accelerate face recognition application," 2015 IEEE Symposium in Low-Power and High-Speed Chips (COOL CHIPS XVIII), 2015, pp. 1-3, doi: 10.1109/CoolChips.2015.7158532.

[21] J. Xu, B. Wang, J. Li, C. Hu and J. Pan, "Deep learning application based on embedded GPU," 2017 First International Conference on Electronics Instrumentation & Information Systems (EIIS), 2017, pp. 1-4, doi: 10.1109/EIIS.2017.8298723.

[22] Toward Automated Analysis of ECG Big Data by GPU XIAOMAO FAN1,2, RUNGE CHEN1 , CHENGUANG HE3 , YUNPENG CAI1 , PU WANG1,2 , AND YE LI1 , (Member, IEEE)

[23] V. Roberge, M. Tarbouchi and F. A. Okou, "Distribution System Optimization on Graphics Processing Unit," in IEEE Transactions on Smart Grid, vol. 8, no. 4, pp. 1689-1699, July 2017, doi: 10.1109/TSG.2015.2502066.

[24] Y. Kurimoto, "Particle Tracking With Space Charge Effects Using Graphics Processing Unit," in IEEE Transactions on Nuclear Science, vol. 68, no. 8, pp. 1912-1919, Aug. 2021, doi: 10.1109/TNS.2021.3084214.

[25] D. Min, Q. Rongcai, W. Ruiping, B. Sheng, C. Wenyi and X. Jiayi, "A new high-definition video player method based on GPU technology," 2012 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), 2012, pp. 388-392, doi: 10.1109/CYBER.2012.6392585.