IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-10, Oct 2019 E-ISSN: 2347-2693

Evaluation of performance of Microprocessors of CPU-GPU CUDA architecture and EDA challenges facing Future Microprocessor Design

Shaikh Numan[1*], Sayed Sajjad Haider[2], Shaikh Rwitobaan[3], Shaikh Shakila[4], Shiburaj Pappu[5]

[1,2,3]Dept. of Computer Science, Rizvi College of Engineering, Mumbai, India
[4,5]Dept. of Computational Sciences, Rizvi College of Engineering, Mumbai, India

[*]*Corresponding Author: numanshaikh9@gmail.com*

DOI: https://doi.org/10.26438/ijcse/v7i10.177181 | Available online at: www.ijcseonline.org

Accepted: 08/Oct/2019, Published: 31/Oct/2019

Abstract: To increase the processor speed and its respective performance a more streamlined method of approach was required. Electronic Design Automation (EDA) is implemented. This article describes the changes brought about by EDA in the development of microprocessor systems and the challenges faced by EDA currently and in the future. This article will also describe the various tests and experiments conducted on the CUDA architecture based microprocessor and also highlight the overall performance and thermals of the system under such extreme test conditions.

Keywords— GPU, Cuda cores, Thermal;

I. INTRODUCTION

In the 1980's era the entire microprocessor development process was focused on verification of performance challenges. At that time of development, the processors were capable of 25MHz to 33MHz clocking speed. The process of this verification accelerated during the mid-90's.

In order to meet the challenges, the industry transitioned from a schematic based design to Hardware Description Language (HDL). With the help of HDL many peripheral systems could be integrated with the microprocessors at that time .This paved a way for development in integrated systems in the years to come.

In the current age of development, the main challenge in microprocessor design is to decrease the size of the chip and maximize its performance. This may seem trivial but there a number of factors and constraints that come into play while designing a microprocessor with decreased size. The factors can be thermal cooling, future proofing so that they are compatible with the upcoming latest peripherals.

II. LITERATURE REVIEW

The current microprocessors have tens of millions of transistors on a 0.18µm process chip. To improve the performance and increase speed further we need to increase number of transistors to approximately one billion on a 0.10µm process chip. This can only be achieved through EDA. However Even EDA faces a lot of issues and unprecedented challenges to further streamline and develop such fine processor technologies.

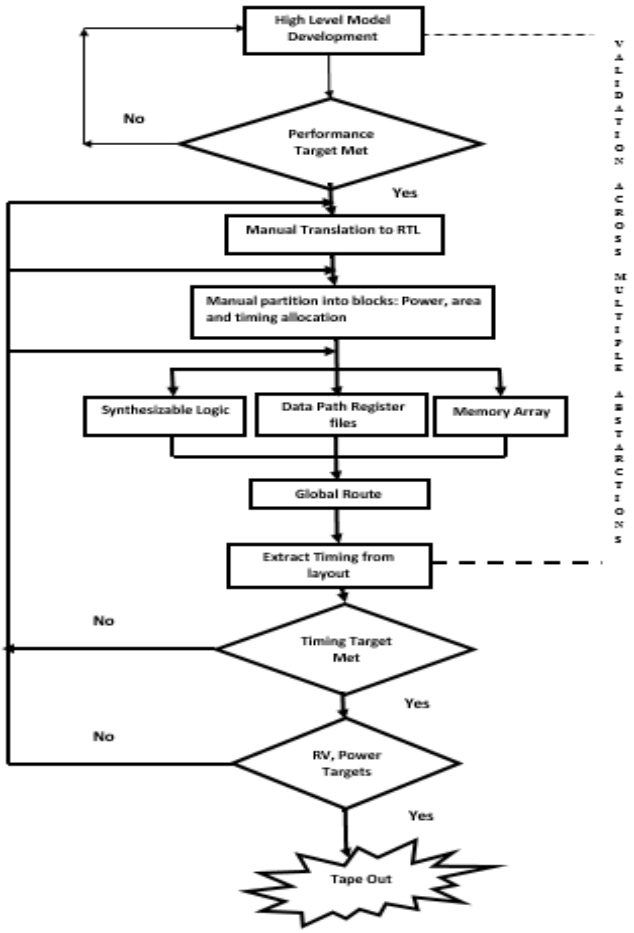


Fig 2.1 Flowchart of EDA design process

However Even EDA faces a lot of issues and unprecedented challenges to further streamline and develop such fine processor technologies. For microprocessor designs in 2006 there could be only 10 Million transistors in each block. Now the number of transistors that can exist in each block has increased to a lot more than that but to increase this count to 1 Billion and then further reducing the chip size is still proving to be quite difficult. A high performance microprocessor design flow is quite complicated. During the design process various sub flows occur based on design characteristics and performance targets. Figure 1 shows a view of microprocessor design flow.

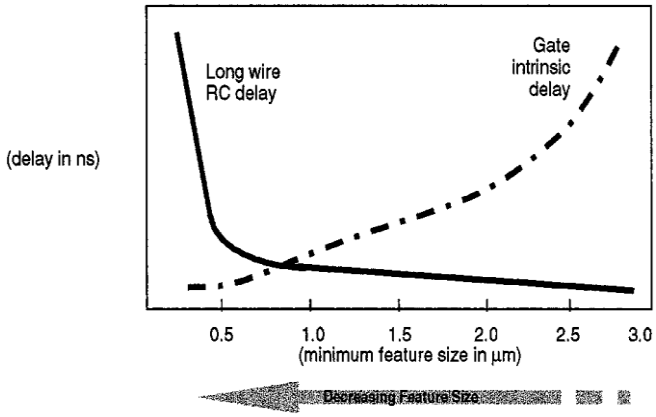


Fig.2.2 Variation in long wire RC delay and intrinsic delay with reduction in feature size.

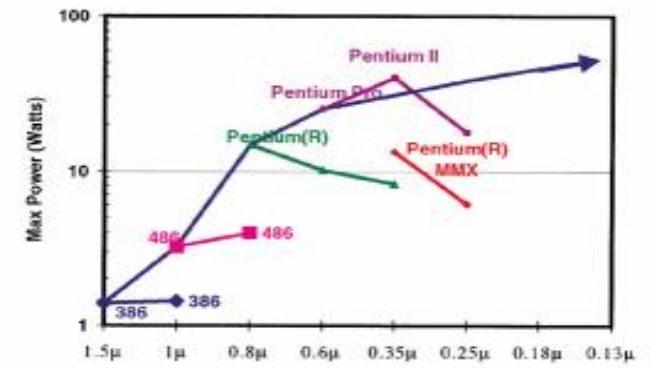


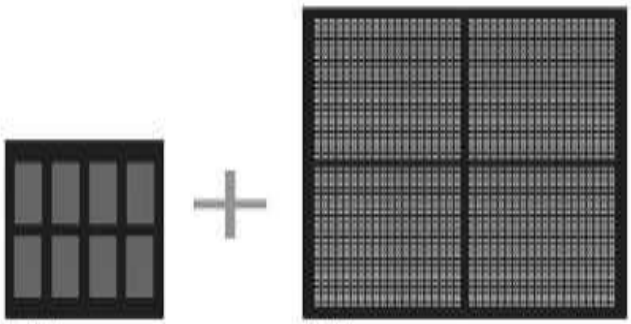
Fig.2.3 Historic power trends for lead and proliferation products.

We identify three problems to be of utmost importance during the design of next generation microprocessors namely Design Correctness, Power management and Performance verification. Achieving faster design convergence and avoiding design errors are some of the most critical issues faced during designing a next gen microprocessor.

III.PROPOSED METHODOLOGY

In order to increase the microprocessor performance, a new Approach was proposed, according to which performance booster system such as Graphic Processing Unit (GPU) can be made accessible to the microprocessor so that it can perform complex operations by using the GPU's resources.

The world of high performance computing is a rapidly evolving field. To improve the performance of modern computing systems it is becoming more and more difficult to rely on the Microprocessor chip (CPU) alone. Hence to give a boost to the performance of a system it is necessary to incorporate and appropriately divide the workload between the CPU and GPU. GPUs can provide astonishing performance boosts by making use of the hundreds of cores available to them. The CUDA architecture developed by NVIDEA is specifically designed to harness the power of the GPUs and increase the processing performance by developing a complete synergy between the CPU and the GPU. It is a parallel computing architecture. The programmer can choose to express the program in higher or lower level languages using this architecture.



CPU: MULTIPLE CORES GPU: THOUSANDS OF CORES

Fig.3.1 Multi-core Arrangement of CPU-GPU

Execution time = (Process_start_time)-(process_end_time) (In Clocks per second)

Applying the above formula the performance of a CUDA architecture based system is calculated. The most important factor of performance in heterogeneous CPU/GPU computing is the workload assignment.

If we assign too little work to CPU, it is not enough to keep the CPU busy during GPU kernel launch and memory transfer, and thus the latency cannot be well hidden. On the other hand, if we assign too much work to the CPU, when the GPU kernel finishes, it has to wait for the CPU to finish the searching job before generating the resultant vector, which will also result in inferior performance. Thus what should be the optimal GPU workload? To answer this question, we vary the GPU workload from 60% to 100% for each core searching function under two configurations, GPU+1thraed CPU and GPU+2 threads CPU, to check the performance impact. The speedup is measured as:

[runtime of X% GPU workload]
[runtime of 100% GPU workload].

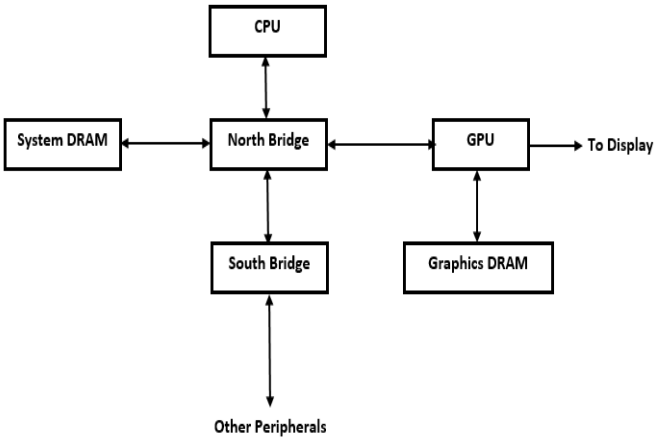


Fig 3.2 GPU Architecture

We run three types of tests to understand, record and determine the performance and efficiency of the GPU-CPU CUDA architecture based system.

Geeks3D GpuTest Benchmark is the GPU stress test and OpenGL benchmark. GpuTest is the first public version of a new cross-platform GPU stress test and benchmarking utility. Here are some experimental results from the given GPU test.

Table 3.1 Test case results for GpuTest Benchmark

Time (in ms)	CPU	GPU
10	6	25
20	9	29
30	17	35
40	26	47
50	29	53
60	34	61

The following graph Fig.3.3. Shows the comparison of CPU and GPU for the Gputest Benchmark:

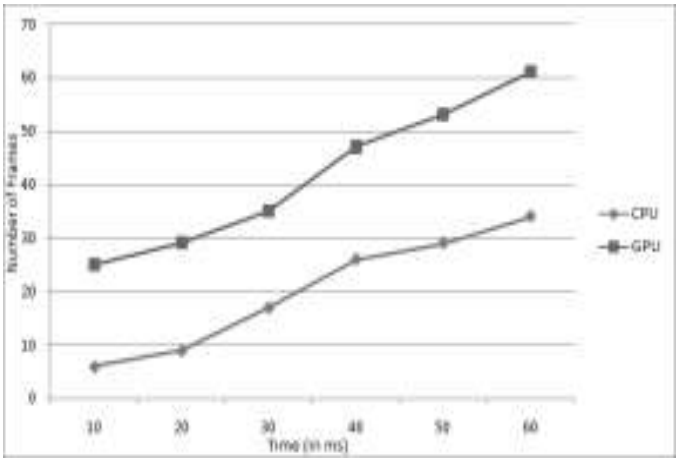


Fig 3.3**:** Comparison of CPU-GPU with GpuTest

FurMark is a very intensive OpenGL benchmark that uses fur rendering algorithms to measure the performance of the graphics card. Here are some results using the FurMark test.

Table 3.2: Test case results for FurMark benchmark

Time (in ms)	CPU (Frames per second)	GPU (Frames per second)
10	2254	3082
20	4501	6156
30	6748	9233
40	8993	12314
50	11239	15391
60	16582	18466

The following graph shows the FurMark comparisons of CPU and GPU:

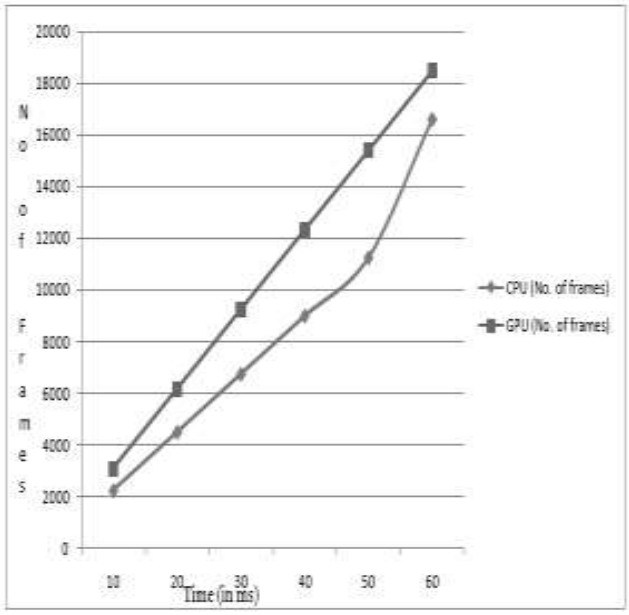


Fig 3.4 **:** Comparisons of CPU and GPU with FurMark

Table 3.3**:** Test case results for Matrix multiplication on CPU

Matrix Size / Processors	200 * 200	400 * 400	600 * 600	800 * 800	1000 * 1000
2	0.46	3.74	12.56	48.62	97.63
4	0.12	1.23	9.63	30.30	59.48
8	0.10	1.02	6.12	19.14	47.54
16	0.07	0.16	3.34	9.86	13.73

The following graph Fig.3.3 Shows matrix size versus the time taken on CPU with 4 processors**:**

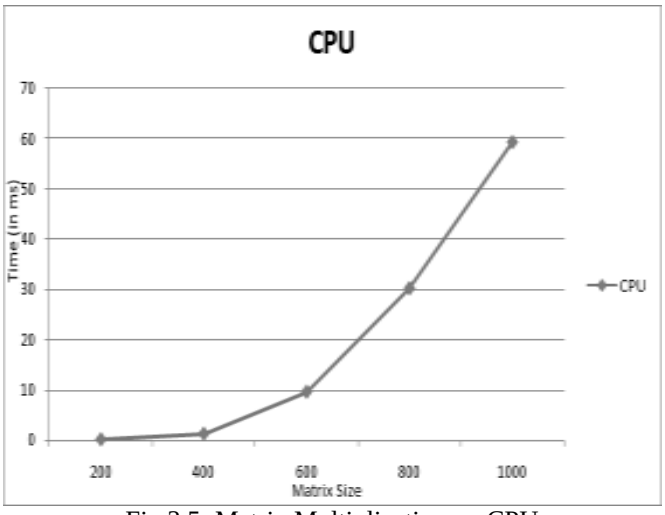


Fig 3.5: Matrix Multiplication on CPU

The following table shows the test case results of Matrix Multiplication on GPU with various matrix sizes by varying the number of cores in Graphic card:

Table 3.4**:** Testcase results for Matrix multiplication on GPU

Matrix Size / Cores	1000 * 1000	2000 * 2000	3000 * 3000	4000 * 4000
100	9.17	81.419	351.361	817.255
200	8.946	81.152	354.745	777.089
300	9.124	81.173	333.788	760.036
400	8.998	81.280	330.569	807.365
448	9.114	81.617	336.744	784.841

The following graph shows matrix size versus the time taken on GPU with 300 GPU cores:

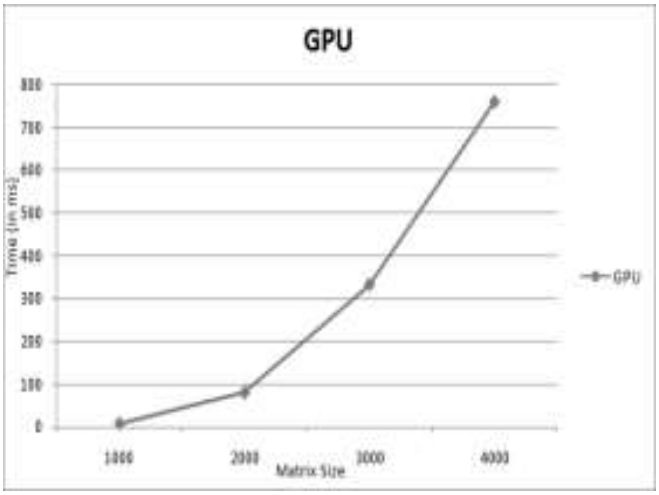


Fig 3.6**:** Matrix multiplication on GPU Graphical analysis

In this research paper we thoroughly studied thermal behaviour of different CPU and GPU based on different platforms and it showed that both CPU and GPU contribute towards rise in temperature of the chip and there should be various techniques to resolve the problem of heating, by implementing various measures we can achieve high performance with robust thermal management.

Different algorithms and models are designed to estimate the temperature of CPU and GPU and for cooperative thermal management. These approaches reduced the variance in on-chip temperature by more than 90%, it also helps to reduce the chip temperature. At the same time the average Frames per Second (FPS) increased by 17%, 19% and 23% when compared to the existing Linux**.**

	Average Temperature				Max Temperature			
Apps	Linux	Naive	Ref[15]	Co-op	Linux	Naive	Ref[15]	Co-op
Epic	77	75	76	75	83	79	79	76
Dday[	75	75	76	75	79	79	79	77
Anomally	78	75	76	75	83	79	79	77
Bikerally	76	75	76	75	80	79	79	77
Robocop	74	76	76	75	78	79	79	77
Farmville	74	75	75	75	77	77	79	76

Fig 3.7**:** Avg. Frame rates in Mobile games

In this experiment, six popular games from play store were tested with various techniques and algorithms. From the above games few of them only require GPU for their performance enhancement but some of the games depend on both CPU and GPU for achieving better performance.

Depending upon the requirements comparisons between Naïve and Linux average and maximum temperature of six android games were done.

IV.ANALYSIS

"Moore's law is the observation that the number of transistors in a dense integrated circuit doubles about every two years."

This Law proved to be accurate in the period from 1990-2010. It must be noted that this law is not a natural or fixed law but rather an observation or projection based on the trends.

Though this law was accurate in the olden days, this however has become redundant and will soon reach its saturation

point. Therefore, we cannot decrease the size of Micro-chip beyond a certain point. However due to the growing demand of microprocessor there has been an increase in competition to create such microprocessor with cutting edge performance benchmarks and minimum size as possible. To meet these demands integration with newer systems such as GPU's is the most viable solution.

Heat sink: a major factor while creating new microprocessors also comes into play. In order to solve this problem, newer materials should be experimented with in order to find a better component that solves the problem of heat sink. Also better cooling systems can be designed to troubleshoot this problem.

V. CONCLUSION

Thus we have successfully discussed the challenges faced by the EDA in microprocessors design. From this study we have quantitatively proved that newer generation of microprocessors can be developed with the integration of newer systems to achieve ground breaking benchmark performances. Benchmarking will be a constant process to satisfy different systems as well as different devices. With the information in this report, some light is shed on the processing power for different applications between CPUs and GPUs. The requirement of a dedicated GPU in modern computing is becoming more of a necessity. Hence our efforts showed that GPU performance is always a step ahead of CPU performance and is a very important tool to ensure faster processing and results if used together with a CPU.

REFERENCES

[1]Cuda Architecture Link:http://developer.download.nvidia.com/compute/cuda/docs/CUDA_Architecture_Overview.pdf

[2]Electronic Design Automation Link:https://www.sciencedirect.com/topics/engineering/electronic-design-automation

[3]The End of Moore's Law? Link:https://www.technologyreview.com/s/400710/the-end-of-moores-law/

[4]Transient Thermal Analysis of a Microprocessor using a heat spreader with variable thermal storage characteristics Link:https://www.researchgate.net/publication/229050507_Transient_thermal_analysis_of_a_microprocessor_using_a_heat_spreader_with_variable_thermal_storage_characteristics

[5]Temperature sensitive microprocessor design Link:https://pdfs.semanticscholar.org/afea/435fc7fb23ba7b31dc4c4bc75840b0470e20.pdf

[6] Tiwari, R. Sam and S. Shaikh, "Analysis and prediction of churn customers for telecommunication industry," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, 2017, pp. 218-222. doi: 10.1109/I-SMAC.2017.8058343

[7] S Navadia, P. Yadav, J. Thomas and S. Shaikh, "Weather prediction: A novel approach for measuring and analyzing weather data," *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam, 2017, pp. 414-417. doi: 10.1109/I-SMAC.2017.8058382

[8] S. Shaikh, S. Rathi and P. Janrao, "IRuSL: Image Recommendation Using Semantic Link," *2016 8th International Conference on Computational Intelligence and Communication Networks (CICN)*, Tehri, 2016, pp. 305-308. doi: .1109/CICN.2016.6

[9] S. Shaikh, S. Rathi and P. Janrao, "Recommendation System in E-Commerce Websites: A Graph Based Approached," *2017 IEEE 7th International Advance Computing Conference (IACC)*, Hyderabad, 2017, pp. 931-934. doi: 10.1109/IACC.2017.0189

[10] A. Fasiku, Ayodeji Ireti, B. Olawale, Jimoh Babatunde, C. Abiola Oluwatoyin B., "*Comparison of Intel Single-Core and Intel Dual-Core Processor Performance*", International Journal of Scientific Research in Computer Science and Engineering, Vol.1, Issue.1, pp.1-9, 2013

[11] M. Sora, J. Talukdhar, S. Majumder, P.H Talukdhar, U.Sharmah, "*Word level detection of Galo and Adi language using acoustical cues*", International Journal of Scientific Research in Computer Science and Engineering, Vol.1, Issue.1, pp.10-13, 2013

[12] Manish Mishra, Piyush Shukla, Rajeev Pandey, "*Assessment on different tools used for Simulation of routing for Low power and lossy Networks(RPL)*", International Journal of Scientific Research in Network Security and Communication, Vol.7, Issue.4, pp.26-32, 2019

AUTHORS PROFILE

Mr. Numan Shaikh is pursuing Bachelor of Engineeiring in Computer Engineering from Rizvi college of Engineering which is affiliated with the Mumbai University.

Mr. Sayed Sajjad Haider is pursuing Bachelor of Engineeiring in Computer Engineering from Rizvi college of Engineering which is affiliated with the Mumbai University.

Mr. Rwitobaan Sheikh is pursuing Bachelor of Engineering in Computer Engineering from Rizvi college of Engineering which is affiliated with the Mumbai University