PERFORMANCE EVALUATION USING GEM 5-GPU SIMULATOR

Sagar Namdev Sawal
Department of Electronics and Telecommunication
Goa College of Engineering
Farmagudi, Goa, India
sgsawal009@gmail.com

Dr. Nitesh Guinde
Department of Electronics and Telecommunication
Goa College of Engineering
Farmagudi, Goa, India
nitesh.guinde@gec.ac.in

Abstract— **with introduction of heterogeneous computing for developing and manufacturing next generation of computer it is necessary to have a simulator which would evaluate and simulate designed heterogeneous system showing how it would behave in actual application. The heterogeneous system refers to a system which uses more than one type of processor or cores. For the purpose of analyzing heterogeneous systems (CPU-GPU) behavior Gem5-gpu simulator was developed. The Gem5-gpu simulator was build combining two different simulators Gem5 simulator and GPGPU-sim simulator. The GPGPU-sim is used for simulating GPUs while Gem5 simulator is used for modelling CPUs. The Gem5-gpu combines best characteristics of both simulators.**

Keywords—GEM5;GPGPU-sim; GEM5-GPU

I. INTRODUCTION

Early computers were created in order to automate process of data handling and calculation which used to consume lot of time. The central processing unit (CPU) was created to perform this complex calculation using sequential logic at high speed and with perfect accuracy. As time passed by people noticed that computers can have much wider application than just performing complex computation and data handling, additionally they also began to be utilized for the purpose of entertainment such as multimedia, gaming etc. this resulted in increasing demand for high quality graphics. The traditional CPUs which used sequential processing logic reached bottleneck in field of speed at which it can deliver high quality graphics, this resulted in large amount of research work in area of improving computer graphics. The GPU that is graphical processing unit was developed in order to handle high quality graphic requirement. Unlike CPU which operates sequentially the GPU performs operation in parallel. Due to performing operation in parallel its processing speed increased dramatically as compared to CPUs. This resulted in computer being able to handle higher quality graphics. The recent technological development for handling graphics is heterogeneous computing in this type of system previously separate CPU and GPU were combined into single unit. This was done in order to reduce size by having CPU and GPU on same chip and by sharing same address space this resulted in increasing speed of operation. Having common memory space brought about accelerating correspondence amongst CPU and GPU subsequently speeding up overall performance. The heterogeneous computing has major application in field of smart-phones, tabs where size is constrained. Due to heterogeneous system [6] having both CPU and GPU on same chip its size is small enough to fit in devices like smart-phones thus enabling smaller devices to handle higher quality graphics.

II. RELATED WORK

In [1] authors introduced new type of heterogeneous simulator superior in performance compared to other simulator currently present. In this they introduced gem5-gpu simulator which is combination of gem5 and GPGPU simulator. It has flexible memory hierarchy compare to Multi2Sim [7] or FusionSim [8]. Gem5-gpu is the only simulator with all of the following advantages such as detailed cache coherence model, Full-system simulation, Check pointing, tightly integrated with the latest gem5 simulator, and Increased extensibility of GPGPU programming model and entire system architecture.

In [3] authors describes Rodinia benchmark its uses, and for which application this benchmark might be used in brief. The Rodinia benchmark is usually used for studying performance of GPU and multicore CPUs [9]. This benchmark can test various parameters of our GPU accurately. It can also be used to examine performance of cache in given system.

III. THE GEM5 SIMULATOR

The Gem5 simulator [2] was created by combining best features of M5 [10] and GEMS [11] simulator. It was created in order to handle latest development in computer architecture. The previous simulators became unable to simulate computer model accurately due to changes in latest technology as it included new cache coherence protocol, first time utilization of multi-core instead of just single core. So in order to simulate this new design of computer system new simulator was needed to be developed which would provide us correct result and capable of simulating various new features

introduced in system design. The Gem5 simulator is product of hard work of almost 10 years. It was created by collaboration of various famed institute such as MIT, AMD, ARM, HP, MIT, Princeton and various universities like Wisconsin, university of Texas, Michigan [2]. It was created with aim of providing all researchers across the world for free. Most of the software in field computer model simulation they have many license related restriction but unlike them Gem5 don't have such restriction this resulted in sharing of ideas and outlook of researchers on large scale resulting in rapid development of the simulator enhancing it even further by taking care of its previous shortcoming. Compared to other simulators in same field Gem5 simulator is exceedingly versatile it support broad combination of cache hierarchy [4], it can be used for simulating diverse CPU models be it ARM core or Intel's x86 CPU core. The Gem5 simulator also supports multiple device and memory models. It is available to both academic as well as corporate researchers. The other high level features of Gem5 are it can boot real operating systems such as Linux and android. There are various benchmarks available for Gem5 such as SPLASH-2 [5] and PARSEC [13], [14]. Gem5 also has multiple CPU models such atomic, in order and out of order CPU models.

IV. The GPGPU-SIM Simulator

The GPGPU-sim (General Purpose Graphics Processing Unit Simulator) is simulator for GPU [12] and it is utilized for simulating parallel string execution. It can be used for modeling timing model of cache, memory controller, shader core, interconnection network and graphics DRAM [15]. During timing modeling GPGPU-sim it assume only GPU is active and CPU is idle, it continuously report number of cycles spent in running kernel. GPGPU-sim support various types of memory available for GPU, it also support various types of cache and DRAMs.

V. Using GEM5-GPU Simulator

For building Gem5-gpu simulator various software are required. It can be installed by following steps provide in "https://gem5-gpu.cs.wisc.edu/wiki/start" website.

The programming for Gem5-gpu simulator is done in python language. The Gem5-gpu simulator can be run in syscall emulation (SE) mode or in full system (FS) mode. The syscall emulation mode is easier to use compare to full system mode. For the full system mode its configuration script is much more complex compare to that of syscall emulation mode, since we need to specify all the information in detail which is related to hardware you are planning to simulate which is not required in syscall emulation mode. The advantage of full system mode is that the accuracy of your simulation will be as close as that of real hardware. To use full system mode we need to download kernel and disk image and depending upon way you build your Gem5-gpu simulator i.e. either in ARM or x86 mode type of kernel and disk image file used would change. When we are running Gem5-gpu simulator in SE mode, first we need to download benchmark and install it for purpose of simulating our designed system. There are various benchmarks available for simulating heterogeneous system. The installed benchmark can then be used for purpose of testing our designed system.

VI. Observation

Various CPU models were designed for purpose of study. In this paper we changed different attributes and impact of changing this attributes on the performance of our designed system was studied.

Initially a single attribute was selected at a time for purpose of studying its impact on overall performance system later we used combination of two or more parameters were chosen.

Most of the designed heterogeneous CPU-GPU model looks like Figure 1. The divergence between different models was number of CPU cores used, type of DRAM selected, what type of processor used. Figure 2 shows impact of cache size on the performance of designed system. It was observed that increasing size of cache increases performance of the designed model.

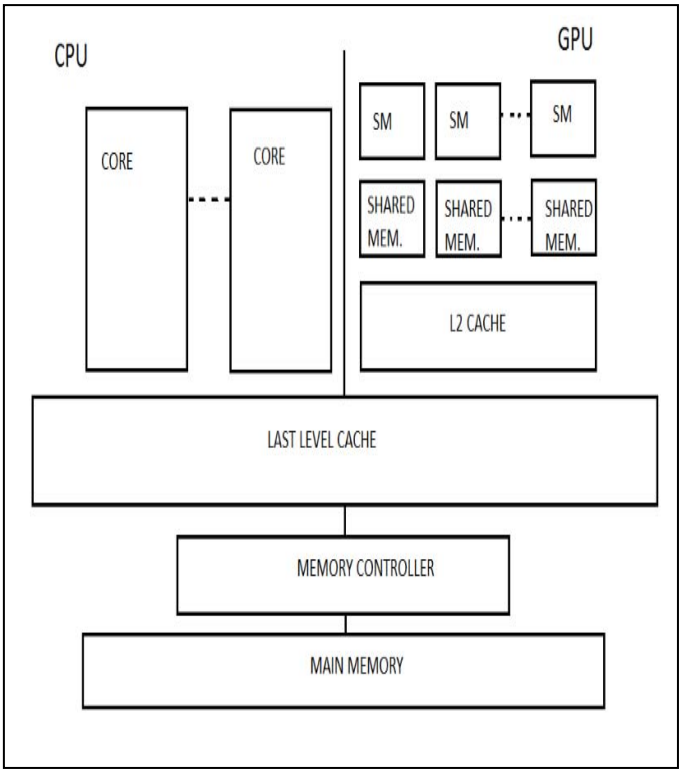

Figure 1: Heterogeneous CPU-GPU model

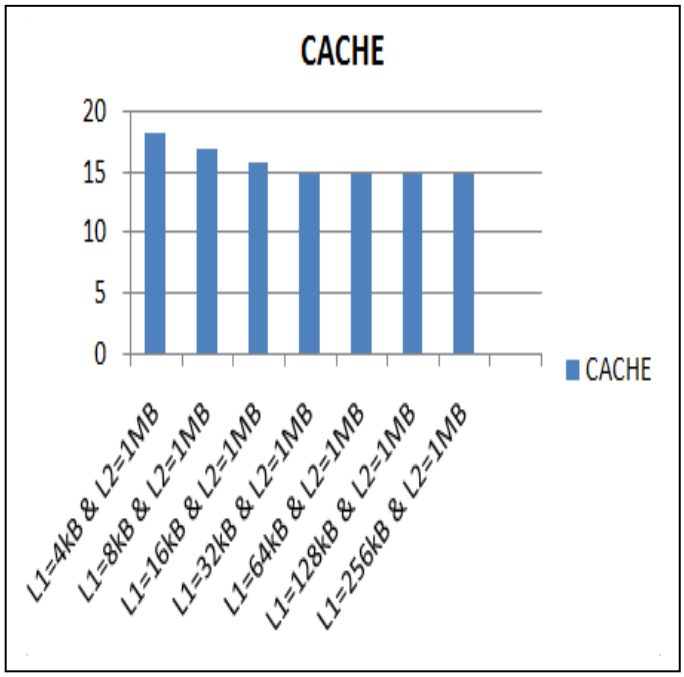

Figure 2: Designed systems output for various sizes of cache

The Figure 3 depicts time taken to execute benchmark for various clock frequencies. In this system designed clock frequency is changed continuously while keeping other parameters fixed. In figure 4 it shows how performance of computer system changes when DDR4 RAM for changing clock speed while Figure 5 shows comparison between DDR4 and DDR3RAM. In this experiment clock speed of computer model was continuously increased and for the every increase in clock rate time taken to execute benchmark when using DDR3 and DDR4 RAM was noted.

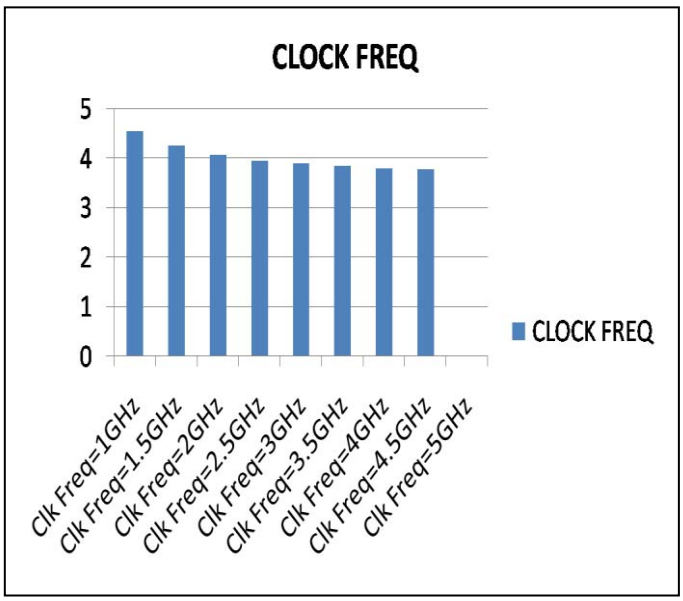

Figure 3: Designed system output for various clock speeds.

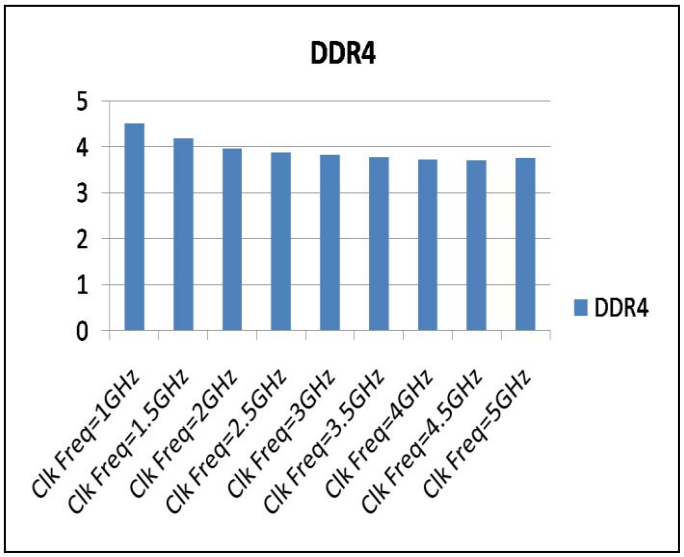

Figure 4: Designed system output for DDR4 for various clock Speeds.

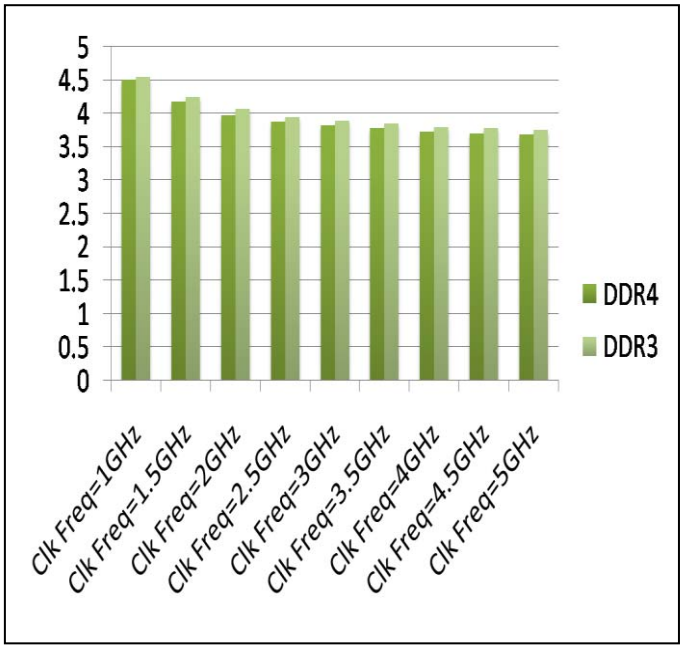

Figure 5: Comparison between DDR3 and DDR4.

VII. Conclusion

In this paper we studied impact of various attributes on performance of heterogeneous system. We saw that size of cache as well as clock rate plays important role in computer systems performance.

Acknowledgment

We would like to thank all the people who have dedicated their time in research and development of Gem5-gpu simulator.

References

[1] Jason Power, Joel Hestness, Marc S. Orr, Mark D. Hill, David A. Wood, "gem5-gpu: A Heterogeneous CPU-GPU Simulator" IEEE Computer Architecture Letters, vol. 14, Issue: 1, 2015.

[2] N. Binkert, B. Beckmann, G. Black, S. K. Reinhardt, A. Saidi, A. Basu, J. Hestness, D. R. Hower, T. Krishna, S. Sardashti, R. Sen, K. Sewell, M. Shoaib, N. Vaish, M. D. Hill, and D. A. Wood, "The gem5 simulator" in ACM SIGARCH Computer Architecture News, vol. 39, no. 2, 2011.

[3] S. Che, M. Boyer, J. Meng, D. Tarjan, J. W. Sheaffer, S.H. Lee, and K. Skadron, "Rodinia: A Benchmark Suite for Heterogeneous Computing" in IISWC, 2009, pp. 4454.

[4] Neethu Bal Mallya, Geeta Patil and Biju Raveendran, "Simulation based Performance Study of Cache Coherence Protocols" in 2015 IEEE International Symposium on Nanoelectronic and Information Systems.

[5] Vikas B."On the Cache Behavior of SPLASH-2 Benchmarks on ARM and ALPHA processors in Gem5 Full System Simulator" in 2014 3rd International Conference on Ecofriendly Computing and Communication Systems.

[6] P. Rogers, "Heterogeneous System Architecture Overview," in Hot Chips 25, 2013.

[7] R. Ubal, B. Jang, P. Mistry, D. Schaa, and D. Kaeli, "Multi2sim:A simulation framework for CPU-GPU computing," in PACT '12, 2012.

[8] V. Zakharenko, T. Aamodt, and A. Moshovos, "Characterizing the performance benefits of fused CPU/GPU systems using fusionsim," in DATE '13, 2013.

[9] Shuai Che; Jeremy W. Sheaffer; Michael Boyer; Lukasz G. Szafaryn; Liang Wang; Kevin Skadron "A characterization of the Rodinia benchmark suite with comparison to contemporary CMP workloads" Workload Characterization (IISWC), 2010 IEEE International Symposium on Year: 2010

[10] Binkert, Dreslinski, Hsu, Lim, Saidi, Reinhardt, "The M5 Simulator: Modeling Networked Systems" Micro, IEEE, vol.26, no.4, pp.52, 60, July-Aug. 2006

[11] M.M.K. Martin, "Multifacet's General Execution-Driven Multiprocessor Simulator (GEMS) Toolset" Sigarch Computer Architecture News, vol. 33, no. 4, Sept. 2005, pp.92-99

[12] A. Bakhoda, G. Yuan, W. Fung, H. Wong, and T. M. Aamodt, "Analyzing CUDA workloads using a detailed GPU simulator," in ISPASS, 2009.

[13] Christian Bienia, Sanjeev Kumar and Kai Li, "PARSEC vs. SPLASH-2: A Quantitative Comparison of Two Multithreaded Benchmark Suites on Chip-Multiprocessors" in Workload Characterization 2008, IISWC 2008, IEEE International Symposium.

[14] Nick Barrow-Williams, Christian Fensch and Simon Moore, "A Communication Characterisation of Splash-2 and Parsec" in Workload Characterization, IISWC 2009, IEEE International Symposium.

[15] Tor M Aamodt, Wilson WL Fung, A El-Shafiey, T Hetherington, A Gubran, T Rogers, A Bakhoda, H Jooybar, "GPGPU-Sim 3.x manual" 2012.[Online],Available:http://gpgpusim.org/manual/index.php/GPGPU-Sim_3. x_Manual. [Accessed: 9- June -2017]