

Design Space Exploration Using Cycle Accurate Simulator

Nimrah Saeed, Misbahlrum, Shazadi Arfa Arshad,

Abstract—To improve system's speed modern 16-bit and 32bit processors contain cache memory for advance performance. Design space exploration is used for performance analysis of system and helps to find the best alternative. In this paper energy models for multilevel cache are evaluated. The benchmarks used for this purpose are BARNES, FMM, WATER-N and WATER-S. Furthermore, these benchmarks are explored to get a point where least energy is used to execute an instruction.

Index Terms—Cache, Benchmark, Design space exploration, BARNES, FMM, WATER-NSQARED, WATER- SPATIAL.

1 INTRODUCTION

Cache memory is located between CPU and main memory. It serves to reduce the average time taken by each memory access as it bridge the speed mismatch between processor and main memory and hence improves the overall system performance. In other words, cache acts as a buffer between CPU and main memory. Multilevel cache efficiently bridge processor-memory gap and it is beneficial to have multi-level cache system. This Memory hierarchy can consume up to 50% of the total energy spent by the microprocessor [9]. This fact has urged researchers to explore cache hierarchy design in terms of energy optimization. Design space exploration is used for system optimization and integration and to explore several design parameters.

In this paper, we explored two such parameters, the optimum sizes of cache at different levels of memory hierarchy and the number of cores that should be present in the system. We focused on the design space of two level memory hierarchies. We evaluated and improved the energy models of different benchmarks i-e BARNES, WATER-N, FMM and WATER-S. The models were presented by M.Y.Qadri.[7] To estimate these parameters we used cycle accurate simulator MARSS because it nearly give the exact number of cycles required to execute an instruction. Energy per access of the tag array for the L1 and L2 cache were obtained from CACTI tool.

We also explored design space to estimate the best point where a benchmark takes minimum number of cycles to execute an instruction. For this purpose we took different values of L1 and L2 cache and observed the number of cycles BARNES took to run completely for core 2 and 4.

The rest of paper is divided into four sections. Section 2 presents related work. Design Space exploration is discussed in section 3 and conclusion is described in section 4.

2 RELATED WORK

In the recent past years, cache energy consumption and throughput models have been the focus of researchers. There are several previous works related to cache power estimation some of them are presented in this section.

A. C. borty et al. [1] represents a new cache building that is multi-copy cache (MC^2) which gives significant reduction in

energy consumption of forceful voltage scaling in caches by producing the multiple copies of each data cache. From the experimental results, they obtain that using MC^2 60% reduction in energy can be achieved. Johnson Kin et al. [2] proposed an energy efficient memory structure. According to them L2 cache is placed behind the filter cache (a small memory), which is similar in structure and size to L1 cache in order to improve performance of processors. M. B.Kamble et al. [3] present analytical models for the energy dissipation in low power cache. The power obtained by these models was compared with that obtained by CAPE (Cache Power Estimator). The models for conventional caches are found to be accurate to within 2% error. C.Long. Su et al. [4] have proposed the power trade-offs in designing of caches and energy reduction using Gray code addressing and cache sub-banking. And experimental results show that direct-mapped caches consume less energy than set associative caches. Sheng Li et al.[5] presented McPAT, which is a frame work that supports comprehensive design space exploration for multicore and many core processor configurations. At micro architectural level, McPAT includes models for the fundamental components of a chip multiprocessor and at circuit and technology level it supports critical-path timing , area , dynamic ,short-circuit and leakage power modeling. McPAT help architects to use new standards combining performance with both area and power. Because die cost increases with area so area is a critical design constraint. Therefore good trade-off between performances and cost needs careful design of on-chip resources.

M.Y.Qadri et al. [6] have presented the techniques for power efficiency at processor core level. They proposed that processor's speed can be improved by adding pipeline stages. Clock frequency, supply voltage and cache can be used efficiently for power reduction. They proposed various techniques such as DVFS, Power and Clock gating for power optimization. Qadri and maier [7] also presented mathematical models to calculate consumption of energy for multilevel caches using Ultra SPARC-2 and Power PC750 processors for two level cache. Then they extend their work [8] by keeping in mind the concept of battery powered embedded system i-e processor only turns on when required otherwise it remains in sleep mode and proposed improved energy and throughput models of data caches. These models are suitable for design of

optimized cache for processors.

In this paper, we evaluated the energy and throughput models presented by M.Y.Qadri [7] for multilevel cache using MARSSx86 simulator. This simulator is cycle accurate and is used for multicore implementation. So, it gives comparatively more accurate results. We also explored design space to estimate minimum number of cycles a benchmark takes to execute an instruction.

3 DESIGN SPACE EXPLORATION

In [7] M.Y.Qadri presented mathematical models to calculate consumption of energy for multilevel caches. The models presented by him analyze the energy consumption for multilevel data cache using PowerPC750, and UltraSPARC-II processors. In this paper we improve the throughput models by using MARSSx86 simulator. This is cycle accurate simulator and is used for multicore implementation. So, it gives comparatively more accurate results.

According to his proposed models

$$E_{total} = E_{ic} + E_{dc} + E_{l2c}[7]$$

E_{ic} = Energy consumed by instruction cache

E_{dc} = Energy consumed by data cache

E_{l2c} = Energy consumed by L2

Where;

$$E_{ic} = E_{ic-read} + E_{ic-mp}$$

$$E_{dc} = E_{dc-read} + E_{dc-write} + E_{dc-mp}$$

$$E_{l2c} = E_{l2c-read} + E_{l2c-write} + E_{l2c-mp}$$

E_{x-read} , $E_{x-write}$ are read and write energy for instruction, data or L2 cache. E_{x-mp} is miss penalty energy for corresponding cache.

To evaluate the above models we obtained number of read/write hits and miss from MARSS simulator. CACTI tool is used to calculate read and write energy and energy per cycle is obtained by simple formula of energy.

$$E = \frac{V * I}{f}$$

Energy of four benchmarks i.e BARNES, FMM, WATER-NSQUARED and WATER-SPATIAL is evaluated for this purpose.

For evaluation of energy we kept L1 instruction and data cache 32 Kbytes, L2 cache 6144Kbytes and two cores. Results of core one and two are presented in Table and 2 respectively.

Table 1: Energy evaluation for core1

Bench-marks	$E_{ic} (nJ)$	$E_{dc} (nJ)$	$E_{l2c} (nJ)$	$E_{total} (nJ)$
BARNES	930034.37	1877528.1277	52741139.298	55548701.298
FMM	1419665.393	23047857.59	49737310.15	74164833.1344
WATER-SPATIAL	1830823.931	24666379.48	63805772.92	90302976.33
WATER-NSQUARED	844077.9734	18740226.11	62852582.11	82436886.1883

Table 2: Energy evaluation for core

Bench-marks	$E_{ic} (nJ)$	$E_{dc} (nJ)$	$E_{l2c} (nJ)$	$E_{total} (nJ)$
BARNES	252.064	51951.646	52741139.298	52793343.008
FMM	218533.9278	31863.6381	49737310.15	2553712966.1514
WATER-SPATIAL	64791.4625	4337.5904	63805772.92	63874901.9719
WATER-NSQUARED	1021534.506	6356317.51	62852582.11	70230434.12138

We explore design space in order to find the best point where a benchmark takes minimum number of cycles to run completely. For this purpose different value of L1 and L2 caches has been taken and evaluated for core 2 and 4. We changed values of L2 cache, keeping L1 cache and cores constant. Specifications used were:

- L1 cache: 4,8,16,32,64 and 128Kbytes
- L2 cache: 32,64,128,256 and 512Kbytes
- Cores: 2 and 4.

Results for core 2 are as follows

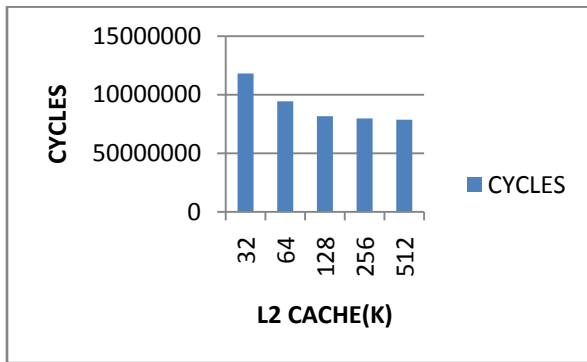


Figure 1: Number of cycles for $L1=4k$

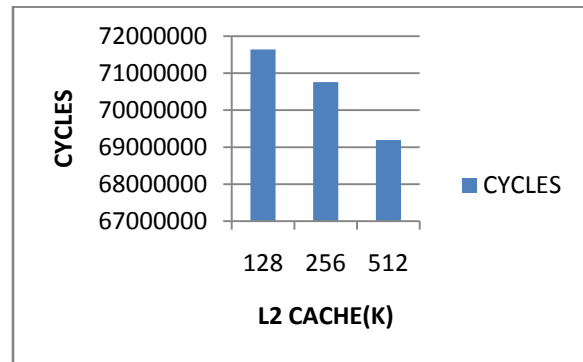


Figure 5: Number of cycles for $L1=64k$

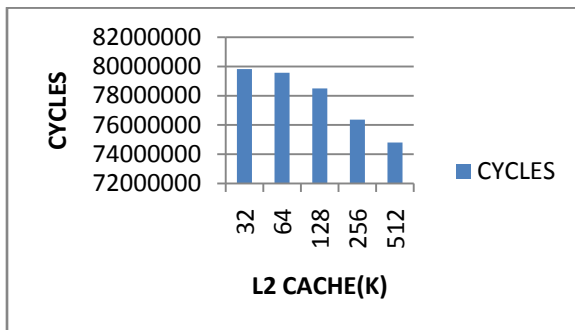


Figure 2: Number of cycles for $L1=8k$

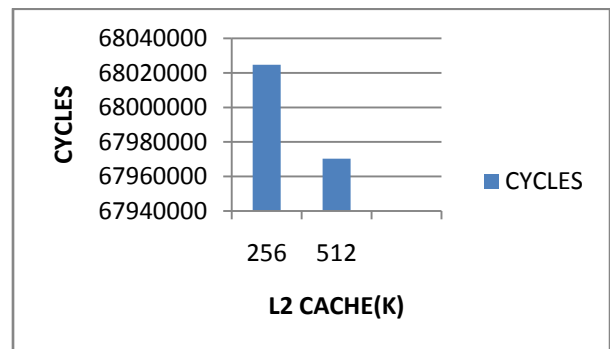


Figure 6: Number of cycles for $L1=128k$

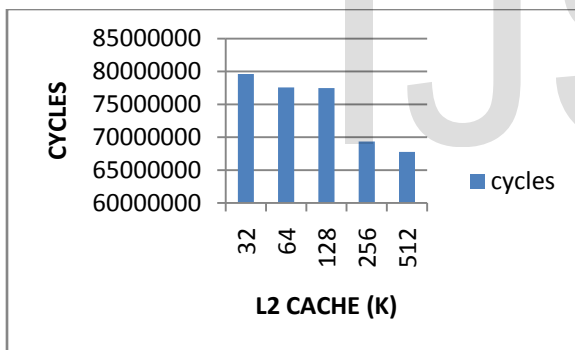


Figure 3: Number of cycles for $L1=16k$

Following are results of core 4:

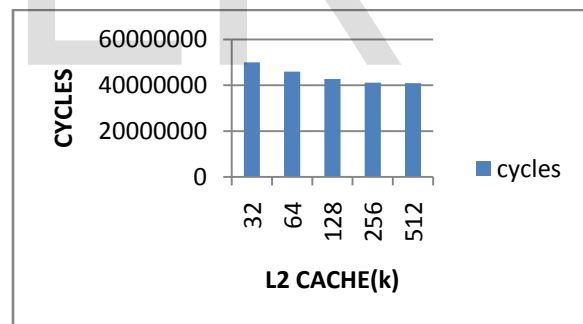


Figure 7: Number of cycles for $L1=4k$

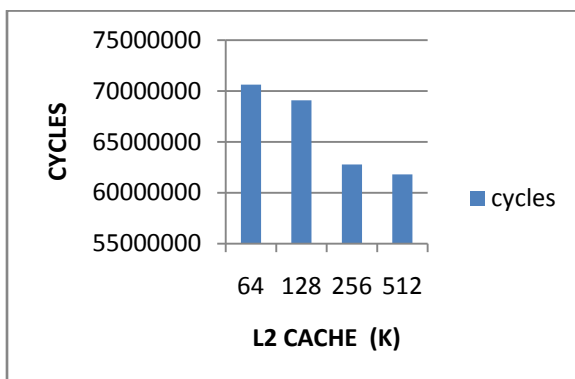


Figure 4: Number of cycles for $L1=32k$

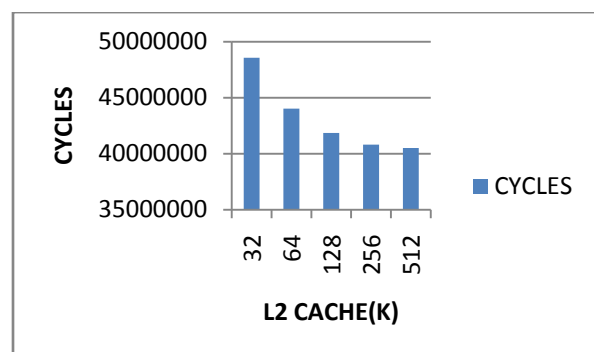


Figure 8: Number of cycles for $L1=8k$

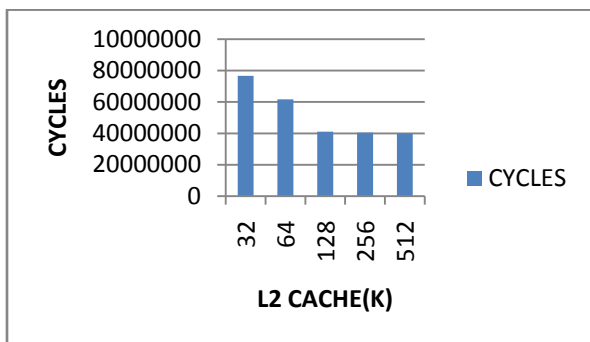


Figure 9: Number of cycles for **L1=16k**

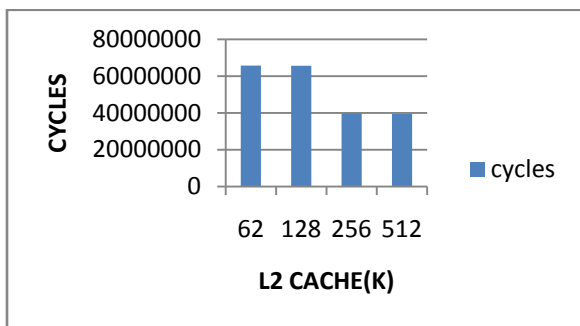


Figure 10: Number of cycles for **L1=32k**

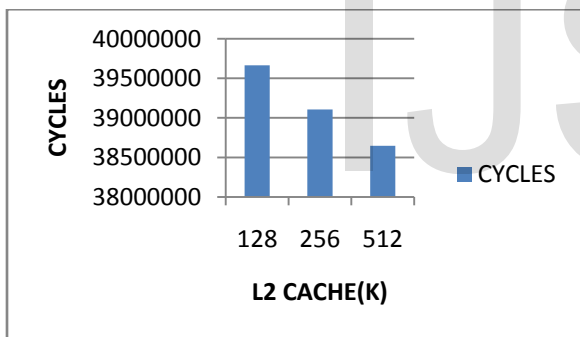


Figure 11: Number of cycles for **L1=64k**

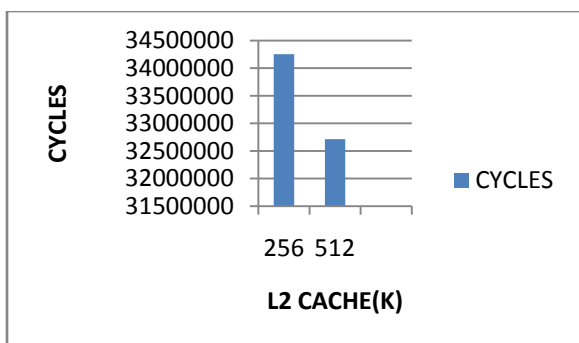


Figure12: Number of cycles for **L1=128k**

- Benchmark takes less number of cycles to execute an instruction, if value of L1 cache increases.
- Large value of L2 takes less number of cycles to run.
- Cores and cycles have inverse relationship. Greater number of cores will take less energy to complete task.

4. CONCLUSION:

We evaluate energy models for different benchmarks i-e BARNES, FMM, WATER-NSQAURED and WATER-SPATIAL for two cores. For this purpose we use MARSS simulator and CACTI tool. The results showed that for core 1, BARNES is 25.1% better than FMM and 32.61% and 38.4% better than WATER-NSQUARED and WATER-SPATIAL respectively. For core 2 we observed that BARNES provide 25% improved results from FMM and 24.8% and 17.34% improved for WATER-SPATIAL and WATER-NSQUARED respectively.

We also explore design space exploration, for this purpose we changed values of L2 and kept cores and L1 cache constant. We observed that by increasing cache values (L1 and L2), benchmark take less number of cycles to run completely. Performance becomes better for greater number of cores. Machine took less number of cycles to run for core 4 than core 2.

We got best result for core 4 when L1=128k and L2=512k.

REFERENCES

- [1] Arup Chakraborty, Houman Homayoun, Amin Khajeh, NikilDutt, Ahmed Eltawil, FadiKurdahi, "E < MC2: Less Energy through Multi-Copy Cache".
- [2] Johnson Kin, Munish Gupta and William H. Mangione-Smith, "The Filter Cache: An Energy Efficient Memory Structure",
- [3] Milind B.Kamle and KanadGhose, "Analytical energy dissipation models for low power caches", NY 13902-6000.
- [4] Ching-Long Su and Alvin M. Despain, "Cache Design Trade-offs for Power and Performance Optimization".
- [5] Sheng Li, Jung Ho Ahn, Richard D. Strong, Jay B. Brockman, Dean M. Tullsen, Norman P. Jouppi, "McPAT: An Integrated Power, Area, and Timing Modeling Framework for Multicore and Manycore Architectures".
- [6] Muhammad YasirQadri, Hemal S Gujarathi and Klaus D. McDonald-Maier, "Low Power Processor Architectures and Contemporary Techniques for Power Optimization - A Review" CO4 3SQ, UK.
- [7] Muhammad YasirQadri, Klaus D. McDonald-Maier, "Analytical Evaluation of Energy and Throughput for Multilevel Caches", Colchester CO4 3SQ, UK
- [8] Muhammad YasirQadri and Klaus D. McDonald-Maier, "Data Cache-Energy and Throughput Models: Design Exploration for Embedded Processors", in Hindawi Publishing Corporation
- [9] S.Segars. "Low power design techniques for microprocessors," International Solid State Circuit Conference, February 2001

Observation: