## Comparison of Different CPU Cache Technology Used in Testing for Research Purposes.

**Seth Levine**

Department of Electrical and Computer Engineering

University of Central Florida

Orlando, FL 32816-2362

*Abstract*—

The objective for this paper is to understand different cache configurations. The higher the level of cache, there is an increase in cache capacity. There is a trend where level one cache will always be SRAM and the other two levels can have different RAM device technology. In the finding level two configuration is made up of mostly SRAM and STT-RAM RAM device technologies. Level three configuration has all three types of RAM (SRAM, STT-RAM, and DRAM). SRAM stands for static random access memories, STT-RAM stands for spin-transfer torque magnetic random access memories, and eDRAM stands for embedded dynamic random access memories [8]. Each of these technologies has advantages and disadvantages. SRAM has relative low latency (see figure two). STT-RAM has a relatively smaller size per capacity (see figure three). eDRAM allows for high capacity level three cache, while taking up little space [2]. STT-RAM is a form of Non-Volatile Memory which allows for low leakage and high density but has high latency [6].

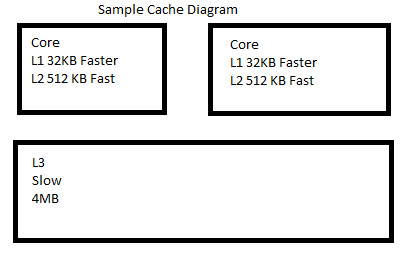
# Introduction

In this paper the researcher looked at different cache configurations to compare the advantages of different cache configurations. In paper’s, the CPU used multi-level configurations. The reason for this is because there are different types of cache for different purposes. Level one cache is usually a fast cache, with low latency, but with a small amount of capacity (table 1 and fig 2). This may be why SRAM is used in level one cache in all the reported configurations (table 1).

In the configurations, the cache used N-way association (table 1). This is for having a high hit ratio, which gets better with higher n. This also, has a low search time, but has a higher search time with the higher the value of n.

SRAM and eDRAM are volatile and STT-RAM is non-volatile. SRAM and eDRAM will lose memory after power is off, while STT-RAM will not lose memory.

Direct-mapped, the line contains either contains your slot in memory or it does not. This method has bad performance. The set strategies do not have 100% hit ratio, but increase the search speed. The hardware used in most cache’s is SRAM, STT-RAM, and eDRAM.



In the next section we will be looking at sources that research CPU cache configurations from 1995. In this paper five baseline systems and five other systems will be used to explain findings. The configurations found in these articles will explained in the following literature review section, as well as how different technologies affect system cache.

# Literature Review

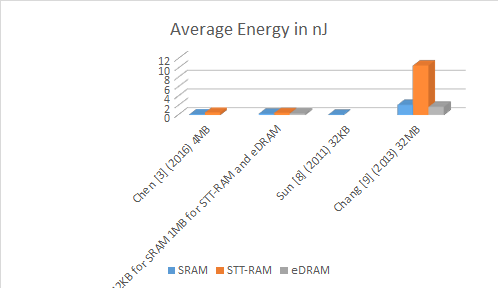
The first group of CPU’s looked at where manufactured from 2000-2006. In this group there were two processors [4][7]. These older processors seem to use SRAM technology for their cache. These are multi-core processors that increase the capacity of the cache as the level is increased.

The next group is from 2010-2016. This group also, utilized multi-core processors. The level one cache seems to be made up of SRAM. Probably because it has low latency and L1 cache is made to be fast [8]. The level two cache utilized SRAM or STT-RAM. This may be because STT-RAM makes it easier to have a higher capacity with a smaller area (See figure 3). SRAM also, gives the advantage of speed. Level three cache has all three types of cache RAMs depending on what the specs require. If speed is need then SRAM is used, if capacity is needed STT-RAM or eDRAM if more capacity is needed. If less energy is required, eDRAM would be preferable (figure 1). If scalability is required, then STT-RAM should be used [8].

# Data Analysis

This is table contain the average energy consumed in nJ, for different cache sizes in various CPU’s. The goal of this chart is to compare the cache energy consumed with different types of RAM.

<Fig. 1.> (Average Energy in nJ)



This is table contain the Average Latency in nS for different cache sizes in various CPU’s. The goal of this chart is to compare the cache latency with different types of RAM.

<Fig. 2.> (Average Latency in nS)

<Fig. 3.> (Average size in mm^2)

This table has the comparison between different sizes of cache configurations between device technologies. This table compares the size of cache for the same amount of capacity.

# Conclusion

From the papers referenced it can be concluded that SRAM, STT-RAM, eDRAM are the most used technologies in cache. SRAM has low energy (figure 1), STT-RAM has small area (figure 3), and eDRAM has a low latency. The cache configurations used n set arrays. This balances flexibility with tag matching complexity.

References

1. S. E. Crawford and R. F. DeMara, "Cache coherence in a multiport memory environment," *in Proceedings of the Second International Conference on Massively Parallel Computing Systems (MPCS-95)*, pp. 632-642, Ischia, Italy, May 2-6, 1995.
2. N. Khoshavi, X. Chen, J. Wang and R. F. DeMara, “Bit-Upset Vulnerability Factor for eDRAM Last Level Cache Immunity Analysis,” *Proceedings of 17th International Symposium on Quality Electronic Design (ISQED 2016),* Santa Clara, CA, USA, March 15 - 16, 2016.
3. X. Chen, N. Khoshavi, J. Zhou, D. Huang, R. F. DeMara, J. Wang, W. Wen and Y. Chen, “AOS: Adaptive Overwrite Scheme for Energy-Efficient MLC STT-RAM Cache,” *53rd Design Automation Conference,* Austing, TX, USA, 2016.
4. N. Khoshavi, X. Chen, J. Wang and R. F. DeMara, "Read-Tuned STT-RAM and eDRAM Cache Hierarchies for Throughput and Energy Enhancement,” arXiv preprint, 2016.
5. M. Lin, et al. "ASTRO: Synthesizing application-specific reconfigurable hardware traces to exploit memory-level parallelism" *Microprocessors and Microsystems 39.7* (2015): 553-564.
6. A. Jog, A. K. Mishra, C. Xu, Y. Xie, V. Narayanan, R. Iyer, and C. R. Das, “Cache Revive: Architecting Volatile STT-RAM Caches for Enhanced Performance in CMPs,” *in Proceedings of 49th Annual Design Automation Conference (DAC)*. 2012, pp. 243–252.
7. A. Jaleel, M. Mattina, and B. Jacob, “Last Level Cache (LLC) Performance of Data Mining Workloads on a CMP-a Case Study of Parallel Bioinformatics Workloads,” *in Proceedings of 12th International Symposium on High Performance Computer Architecture (HPCA)*, 2006, pp. 88–98.
8. Z. Sun, X. Bi, H. H. Li, W.-F. Wong, Z.-L. Ong, X. Zhu, and W. Wu, “Multi Retention Level STT-RAM Cache Designs with a Dynamic Refresh Scheme,” *in Proceedings of 44th annual IEEE/ACM International Symposium on Microarchitecture*. 2011, pp. 329–338.
9. M.-T. Chang, P. Rosenfeld, S.-L. Lu, and B. Jacob, “Technology Comparison for Large Last-level Caches (L 3 Cs): Low-leakage SRAM, Low Write-energy STT-RAM, and Refresh-optimized eDRAM*,” in Proceedings of 19th International Symposium on High Performance Computer Architecture (HPCA)*, 2013, pp. 143–154.
10. Z. Sun, X. Bi, and H. Li, “Process variation aware data management for stt-ram cache design,” *in Proceedings of the 2012 ACM/IEEE International Symposium on Low Power Electronics and Design, ISLPED,* 2012, pp. 179–184.
11. <CPU Cache comparistion table>

| ***Parameters for the below techniques (Year)*** | Processor | |  | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Level 1 (L1) for Instruction (I) or Data (D) | | | | | Level 2 (L2) | | | | | Level 3 (L3) or Last Level Cache (LLC) | | | | |
| # of cores | Freq. | Capacity | Set Assoc. | Device  Tech. | # of CL | Protocol | Capacity | Set Assoc. | Device  Tech. | # of CL | Protocol | Capacity | Set Assoc. | Device  Tech. | # of CL | Protocol |
| Khoshavi [2] (2016) | 8 | 3GHz | 32KB | 8-way | SRAM | 512 | MESI | 512KB | 8-way | SRAM | 8192 | MESI | 96MB | 16-way | eDRAM | ~1.5M | WB |
| Sun [8] (2011) | 4 | 2GHz | 32KB | 4-way | SRAM | 512 | N/A | 256KB | 8-way | SRAM | 4096 | N/A | 4MB | 16-way | STT-RAM | 65536 | N/A |
| C. Bienia [6] (2016) | 4 | 2GHz | 32KB | 4-way | SRAM | 512 | N/A | 1MB | 16-way | STT-RAM | 16384 | N/A | N/A | N/A | N/A | N/A | N/A |
| Chen [3] (2016) | 4 | 3.3 GHz | 32KB | 8-way | SRAM | 512 | WB | 4MB | 8-way | STT-RAM | 65536 | WB | N/A | N/A | N/A | N/A | N/A |
| Motlagh [4] (2000) | 8 | N/A | N/A | N/A | SRAM | N/A | N/A | 512KB-1MB | N/A | SRAM/STT-RAM | 8192-16384 | N/A | N/A | N/A | eDRAM | N/A | N/A |
| Lin [5] (2015) | 2 | 800 MHz | 32KB | N/A | N/A | 512 | MOESI | 512KB | N/A | N/A | 8192 | MOESI | N/A | N/A | N/A | N/A | N/A |
| Jaleel [7] (2006) | 8 | N/A | 32KB | 4-way | N/A | 512 | N/A | 256KB | 8-way | N/A | 4096 | N/A | 4MB-64MB | 16-way | N/A | 65536-  ~1M | N/A |
| Chang [9] (2013) | 8 | 2GHz | 32KB | 8-way | SRAM | 512 | MESI | 256KB | 8-way | SRAM | 4096 | MESI | 32MB | 16-way | SRAM/DRAM/STT-RAM | 524288 | N/A |
| Sun [10] (2010) | 8 | 2Ghz | 16KB | 2-way | SRAM | 256 | N/A | 8MB | 32-way | STT-RAM | 131072 | N/A | N/A | N/A | N/A | N/A | N/A |
| Carwford [1] (1995) | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |

Cache line

Calculation for “# of CL” columns:

Manually compute the number of cache lines given the capacity value as listed in capacity column, assuming the cache line size is always 64 Bytes

Protocol column = {Write Back (WB), Write Through (WT), MESI, MOESI, Not Available (N/A)}