HBRP PUBLICATION

Recent Trends in Analog Design and Digital Devices
Volume 6 Issue 2

Analysis of Cache Memory and its Importance in Processor Performance

Prajwal K P Bhat[1]*, K B Ramesh[2]
[1]UG Student, Department of Electronics and Instrumentation Engineering, R V College of Engineering, Bengaluru, India
[2]Associate Professor, Department of Electronics and Instrumentation Engineering, R V College of Engineering, Bengaluru, India

Corresponding Author
E-Mail Id: prajwalkpbhat.ei21@rvce.edu.in

ABSTRACT
Cache is a small, high-speed buffer memory between the CPU and the primary unit is a hardware component that stores data so that future requests for the data can be served faster. Cache memory must be small so that data can be used efficiently and it is cost-effective. This paper is going to analyze the hardware-based cache memory and its importance on processor performance, cache design and various cache levels. We also take a case study on how 3D V-Cache works and how it significantly enhances the performance of the processor.

Keywords: *Cores, threads, central processing unit (CPU), cache levels, thermal design power (TDP), hit ratio*

INTRODUCTION

CPUs, hard disk drives and solid-state drives use hardware-based cache, where cache is implemented as a block of temporary storage of data to be used again. Cache memory has significantly faster speeds than ROM (Read-only memory) and RAM (Random Access Memory), which makes it more efficient as it can transfer data at a much faster rate, instead of relying on the main memory to do the same. The reused data should be available at a faster speed than the stored data; this led to development of cache memory.

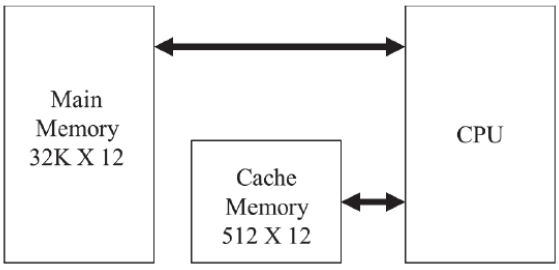

Fig. 1: *Location of cache memory. [1]*

MULTI-LEVEL CACHE

Cache memory initially was single cache. However, around 2000, multi-level cache became more common as there was a need for faster access to memory and more efficiency. Cache is usually split into 3 levels: L1, L2 and L3. The L1 cache is the fastest among the caches and is found in the processor. The L2 and L3 caches are present outside the processor and are much larger in capacity than L1.

If a word is to be searched, the CPU searches for the word's primary address in the cache memory. If it is found, it is a HIT, else it is a MISS. Misses usually can occur due to the inadequate size (happens if 2 blocks are simultaneously mapped on the same address) and the small size of cache.

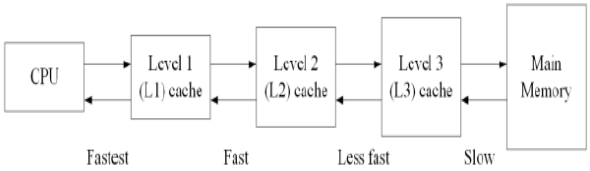

Fig. 2: *Hierarchy of cache memory[2].*

Hit ratio is defined as:

$$Hit\ Ratio = \frac{No.\ of\ hits}{No.\ of\ Hits + No.\ of\ Misses} \quad (1)$$

PERFORMANCE IMPROVEMENTS

In paper [1], Sonia, Monika and co. analyzed the various performance issues faced by cache memory and investigated various papers that offered solutions. The paper, after reviewing multiple papers, proposed architecture to improve the performance of the cache memory.

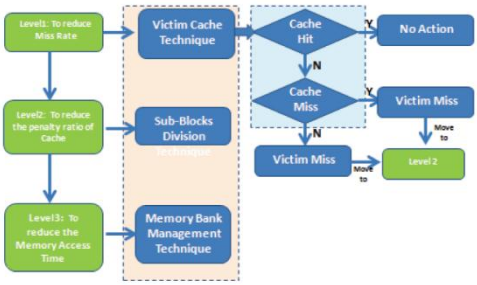

Fig. 3: *Proposed architecture to improve cache performance[1].*

The paper recommended the following steps in order to improve cache memory performance:

- Use victim cache to decrease the miss rate
- Reduction in penalty ratio of cache miss
- Decreasing memory access time

For direct-mapped cache, victim cache is used to enhance the hit latency rate and decrease the conflict miss rate.

Isha, Disha, and Vikram in paper [2], have investigated multi-level cache hierarchy. When dealing with bringing load balance between data and fetched instructions, unified cache has a better hit rate than split cache. However, split cache eliminates the disagreement between fetch/decode and execution instructions, which can reduce efficiency and affect performance.

Cache memory is also vulnerable to cyber-attacks. Whenever the processor sends out an information request, it passes through cache. This information that was last accessed and requested by the processor is stored in the L2 cache. This increases the likelihood of cybercriminals trying to access information.

So, the proposed solution to this was the Interleaved Scrambling Technique (IST). This technique scrambles the data in write phase and then descrambles it in read

phase. All of this happens in the L2 cache; however, this can be extended to other levels as well.

3D V-CACHE™: A STUDY

In 2022, AMD released a new processor, the Ryzen 7 5800X3D that featured their latest 3D V-Cache™ technology. Unlike the traditional method, where cache memory is placed near the CPU, AMD placed the cache on top of the CPU. This resulted in a massive performance improvement in both their consumer-grade desktop and server processors. The non-3D V-Cache™ variant (Ryzen 7 5800X) for reference, has only 32MB of L3 cache, compared to the X3D's 96MB.

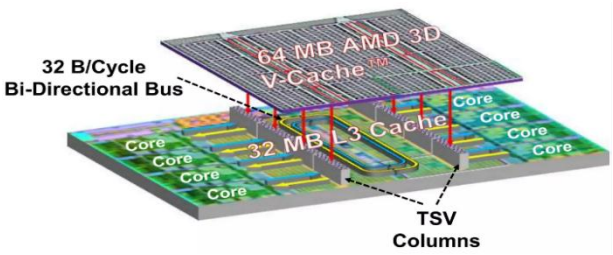

Fig. 4: *Stacking of cache atop the CPU[4].*

Gaming is one area where cache memory is very important, and this is where the 3D V-Cache™ excels.

The 5800X3D beat AMD's fastest processor, the Ryzen 9 5900X in a lot of high-end games, and edged Intel's fastest and most powerful processor (i9-12900K). The only area it lacked in was CPU intensive tasks as it has slower clock speeds than both the processors mentioned above.

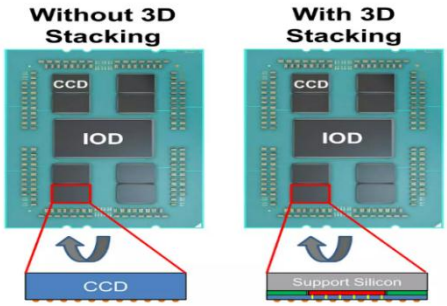

Fig. 5: *Server configurations[4].*

Table 1: *Specifications of 3D V-Cache™ processors.*

	R9 7950X3D	R9 7900X3D	R7 7800X3D	R7 5800X3D
CORES	16	12	8	8
THREADS	32	24	16	16
L2 CACHE	16MB	12MB	8MB	8MB
L3 CACHE	128MB	128MB	96MB	96MB
TDP	120W	120W	120W	105W

The server CPUs have also massively benefitted from 3D V-Cache. The number of jobs done per hour jumped from 24.4 to 40.6 with the 3D V-Cache implemented [4].

As of 2023, AMD released 3 new processors with the 3D V-Cache technology, two of which (Ryzen 9 7950X3D and 7900X3D) have more cores, more L2 and L3 cache and are significantly more powerful than the Ryzen 7 5800X3D.

CONCLUSION

In this paper, we analyzed cache memory and reviewed a solution offered by Sonia and Monika. In the future, one of the solutions can be implemented physically using hardware. This paper also took a case study on AMD's 3D V-Cache technology and observed the performance differences of the CPU with and without 3D V-Cache™. In a few years, other processor manufacturers like Intel could be using a similar technology to AMD as it allows the addition of more cache memory without increasing the actual size of the die. Given how expensive it is to manufacture cache memory, this technology could be a new breakthrough.

REFERENCES

1. Alsharef, A., Jain, P., Arora, M., Zahra, S. R., & Gupta, G. (2021, March). Cache memory: an analysis on performance issues. In *2021 8th international conference on computing for sustainable global development (INDIACom)* (pp. 184-188). IEEE.
2. Lokegaonkar, I., Nair, D., & Kulkarni, V. (2021, December). Enhancement of cache memory performance. In *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)* (pp. 1490-1492). IEEE.
3. Vamsi, B. S. (2022, November). High Performance Read and Write Policy of Multilevel Cache Memory. In *2022 IEEE 6th Conference on Information and Communication Technology (CICT)* (pp. 1-5). IEEE.
4. Wuu, J., Agarwal, R., Ciraula, M., Dietz, C., Johnson, B., Johnson, D., ... & Naffziger, S. (2022, February). 3D V-Cache: the Implementation of a Hybrid-Bonded 64MB Stacked Cache for a 7nm x86-64 CPU. In *2022 IEEE International Solid-State Circuits Conference (ISSCC)* (Vol. 65, pp. 428-429). IEEE.

