Mockups – caching visualizations

CSC494 Summer 2025

Interactive Visualizations for Computer Architecture

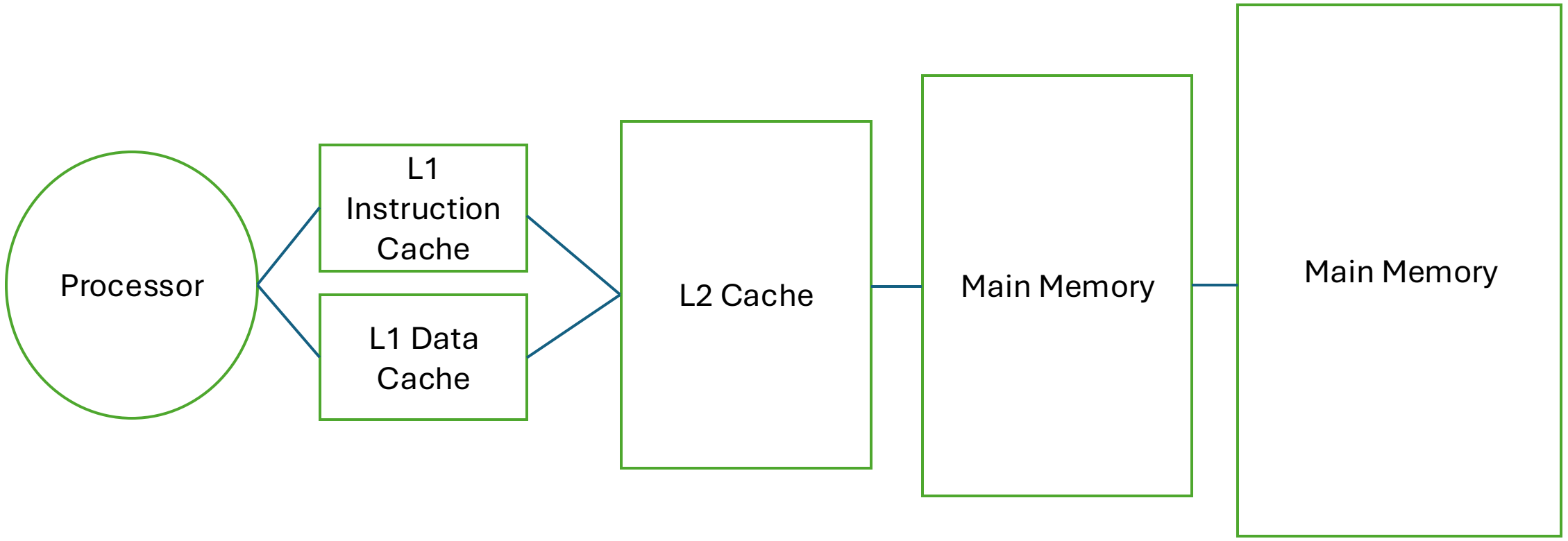
Updated for June 13

Cache Hierarchy Visualization

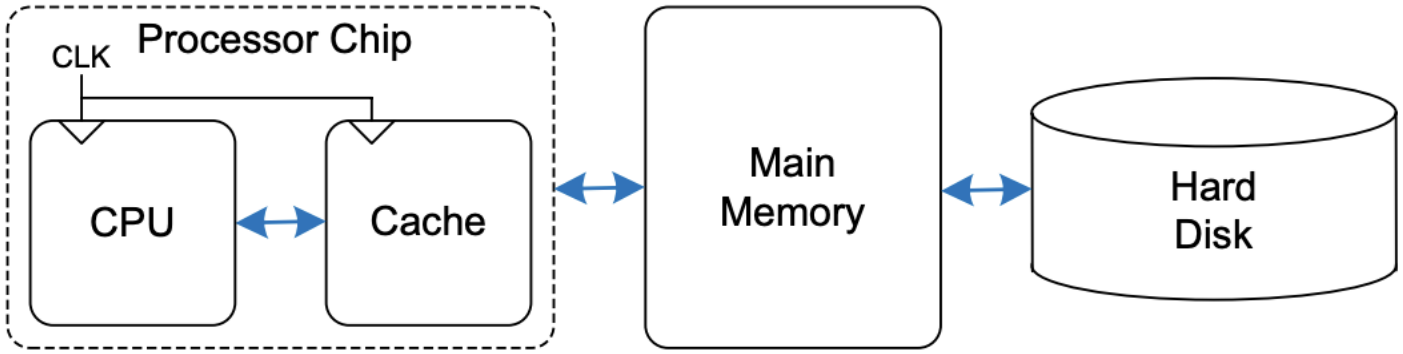
Changes from last week

- Simplicity
- Less configurations
- More focus on the cache hierarchy
- Smaller, condense viz
- Main outcome: going to mm takes a along time
- AMAT = t_cache + MRcache(t_MM + MR_MM * t_VM), Miss Rate, Hit Rate

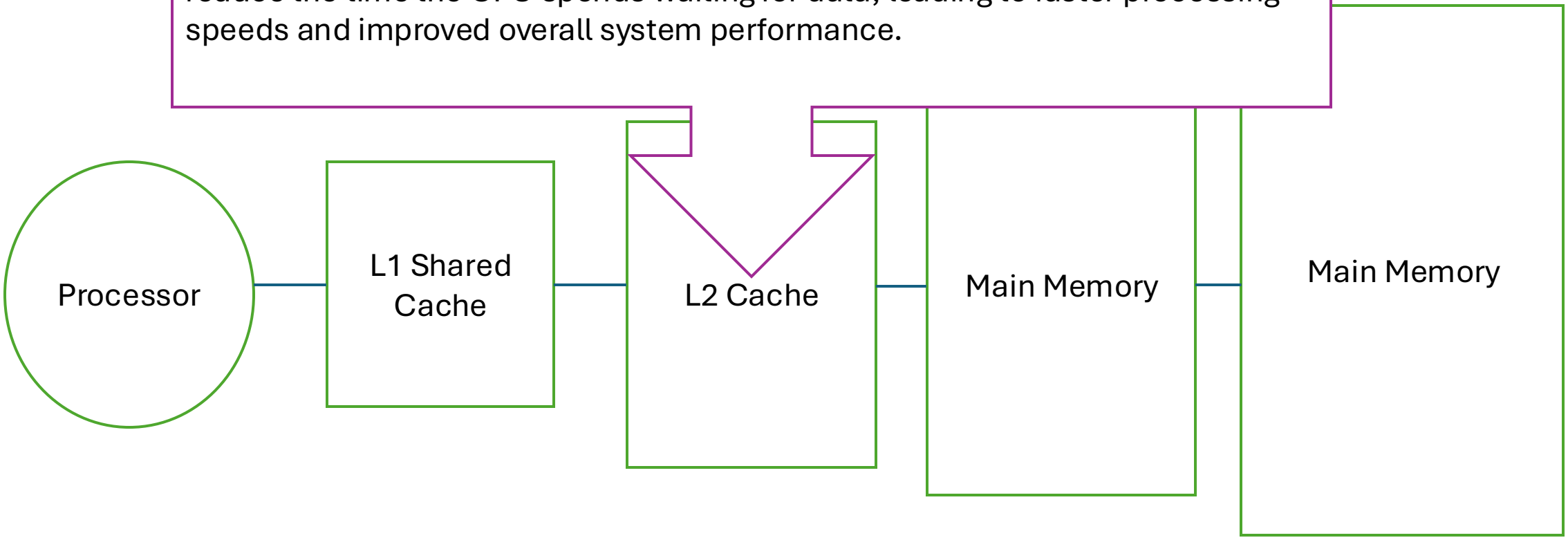
Harvard Architecture Visualization

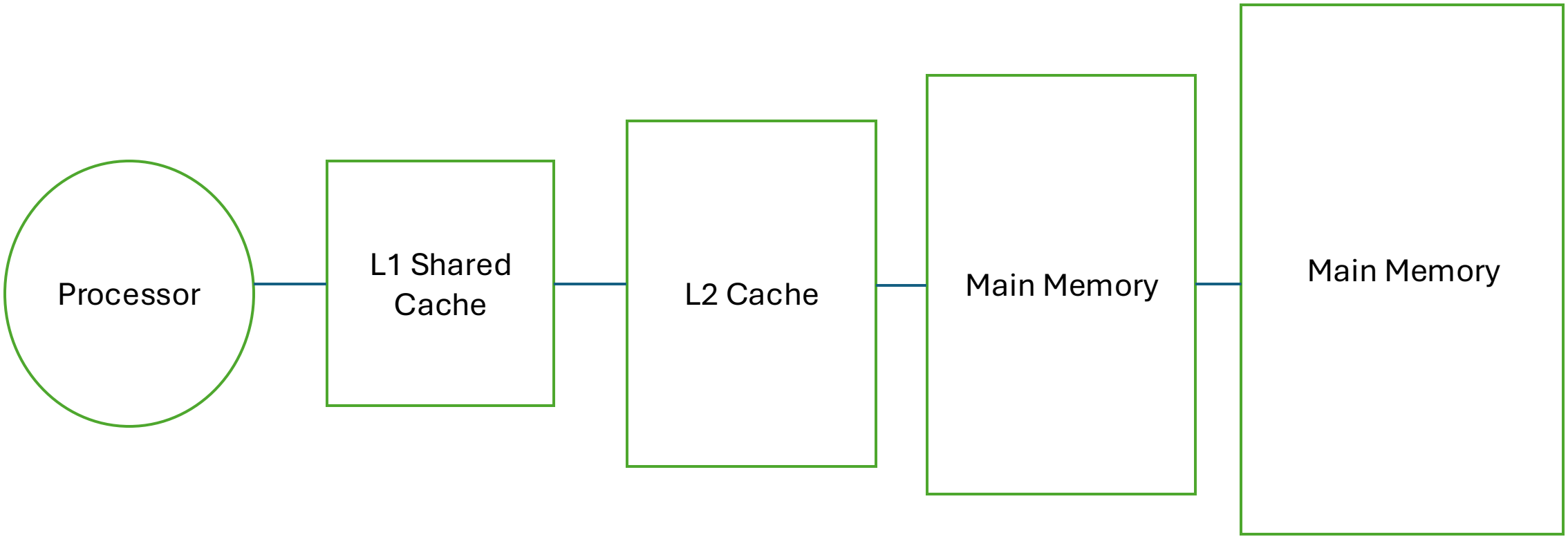


Actual Plan



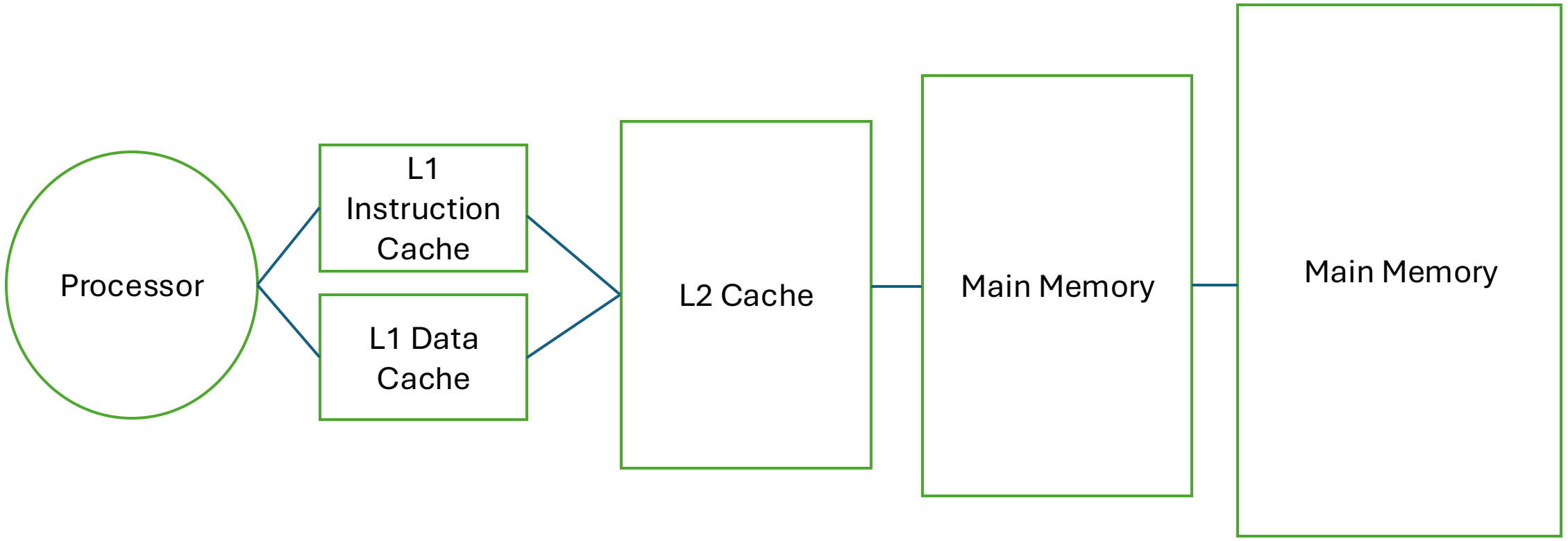
The primary purpose of L2 cache is to improve CPU performance by storing frequently accessed data and instructions, making them readily available for faster retrieval. It acts as a secondary cache between the faster L1 cache and slower main memory (RAM), bridging the gap between the two. This helps to reduce the time the CPU spends waiting for data, leading to faster processing speeds and improved overall system performance.





DDL: Temporal, Spatial, Random access

Harvard Architecture Visualization



Set Associative Visualization

Changes from last week

Visualize Set-Associative vs Direct Mapped caches

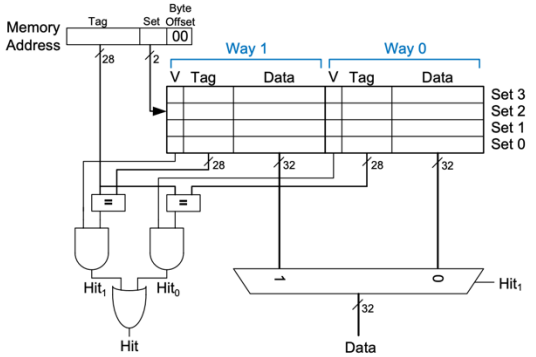


Figure 8.9 Two-way set associative cache

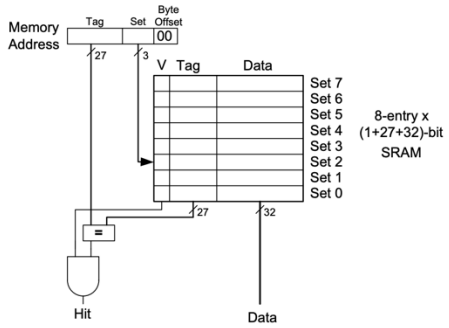
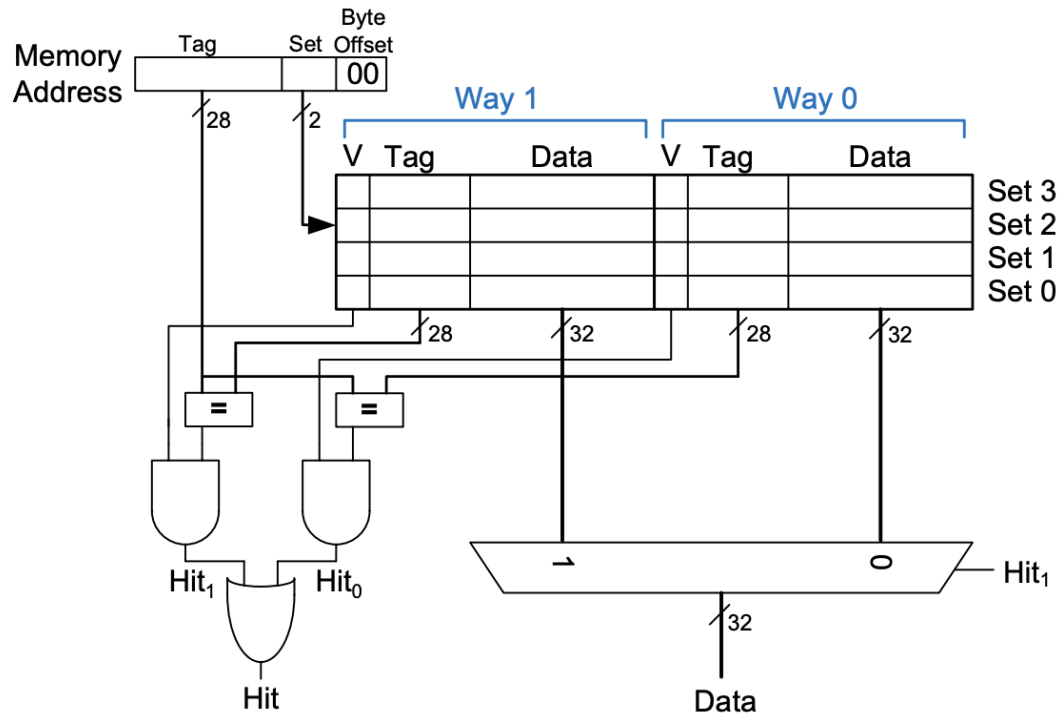
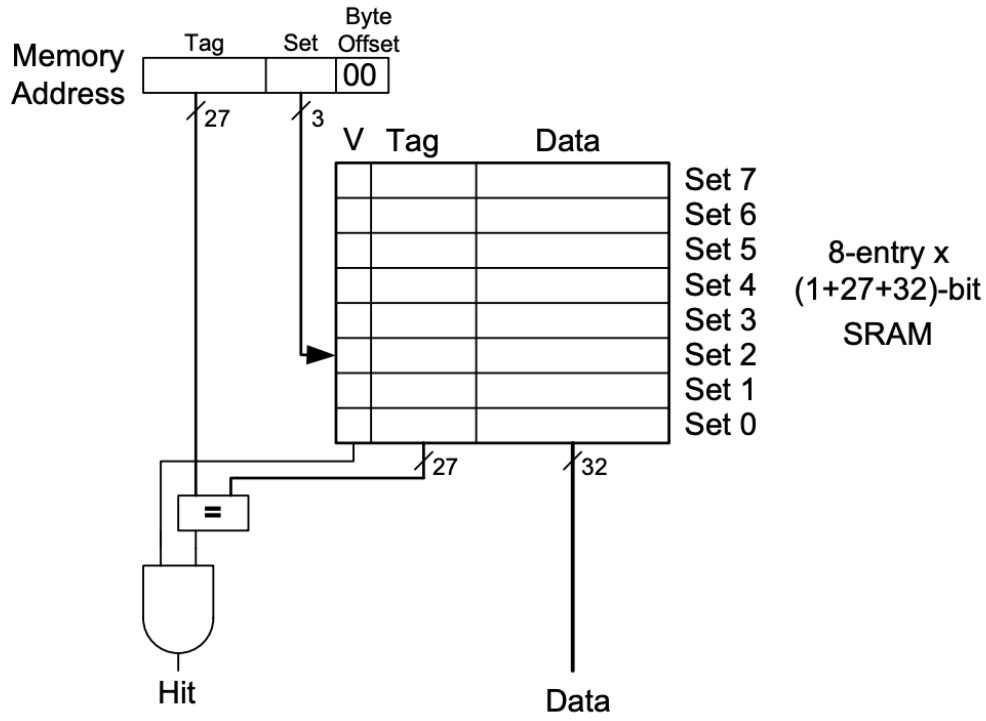


Figure 8.7 Direct mapped cache with 8 sets

$$Miss\ Rate = \frac{Number\ of\ misses}{Number\ of\ total\ memory\ accesses} = 1 - Hit\ Rate$$

$$Hit\ Rate = \frac{Number\ of\ hits}{Number\ of\ total\ memory\ accesses} = 1 - Miss\ Rate \qquad (8.1)$$


Memory Address
Tag
Set
Byte Offset
00
28
2
Way 1
Way 0
V
Tag
Data
V
Tag
Data
Set 3
Set 2
Set 1
Set 0
28
32
28
32
=
=
Hit1
Hit0
Hit
1
0
Hit1
32
Data


Memory Address
Tag
Set
Byte Offset
00
27
3
V
Tag
Data
Set 7
Set 6
Set 5
Set 4
Set 3
Set 2
Set 1
Set 0
8-entry x (1+27+32)-bit SRAM
27
32
=
Hit
Data

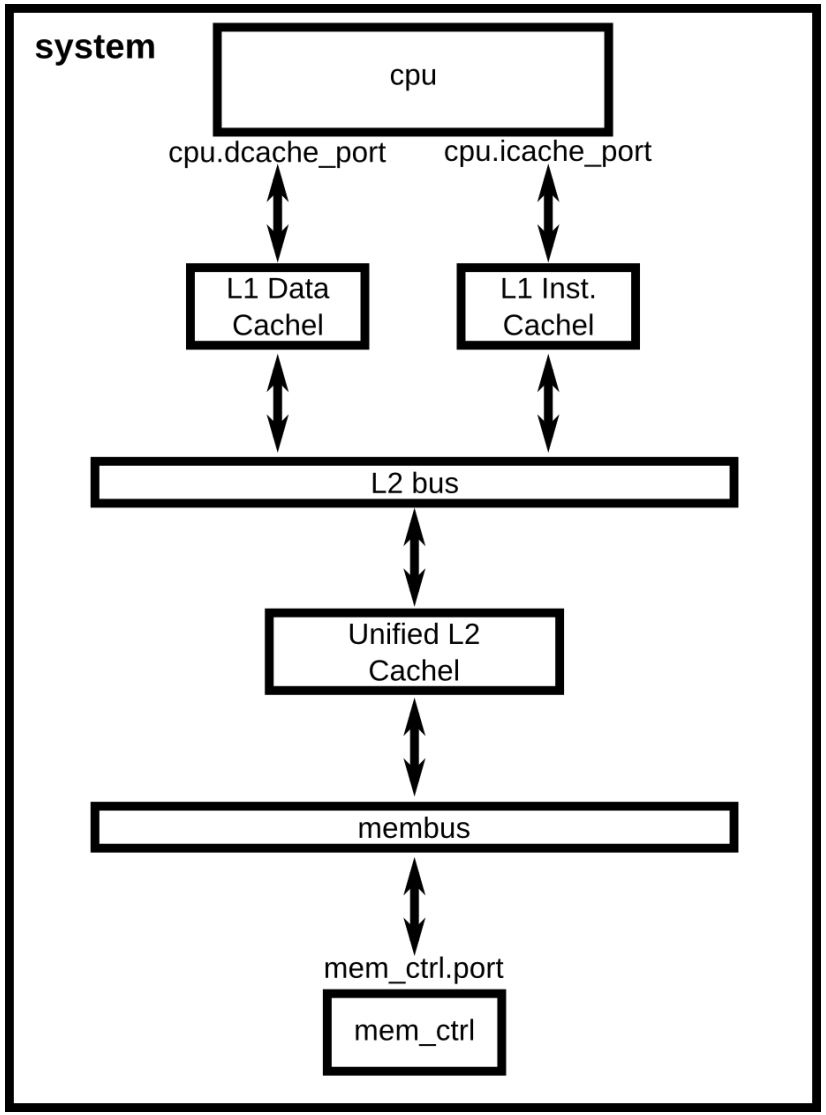
References

https://www.gem5.org/documentation/learning_gem5/part1/cache_config/

This explains the idea of having a L2 unified cache.

Some ideas for tag latency and data latency and response latency.

Miss handling status



References

Digital Design and Computer Architecture (4th Edition)
chapter 8 on Set Associative and caches