TED ÜNİVERSİTESİ

Computer Organization – CMPE361

Department of Computer Engineering
TED University- Fall 2023

Memory Systems 4- Caches

These Slides are mainly based on slides of the text book (downloadable from the book’s website).

Capacity Misses

- Cache is too small to hold all data of interest at once
- If cache full: program accesses data X & evicts data Y
- ***Capacity miss*** when access Y again
- How to choose Y to minimize chance of needing it again?
- **Least recently used (LRU) replacement:**
  - the least recently used block is evicted

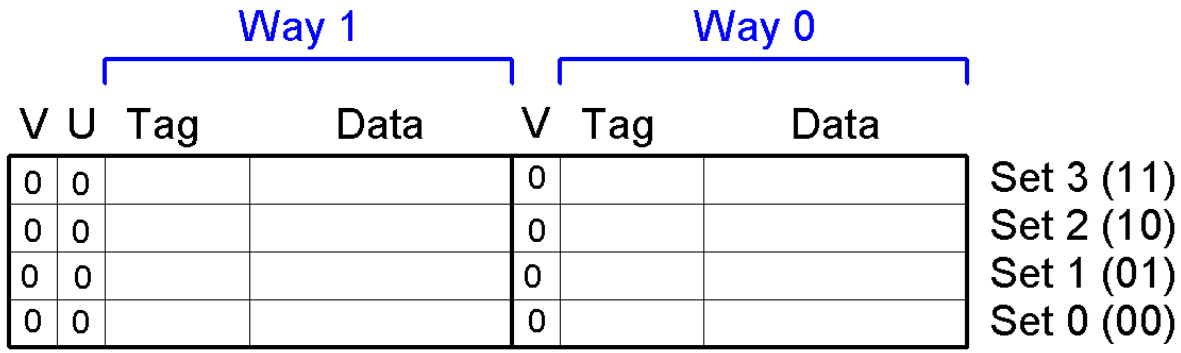
Types of Misses

- **Compulsory:** first time data accessed
- **Capacity:** cache too small to hold all data of interest
- **Conflict:** data of interest maps to same location in cache

Miss penalty: time it takes to retrieve a block from lower level of hierarchy

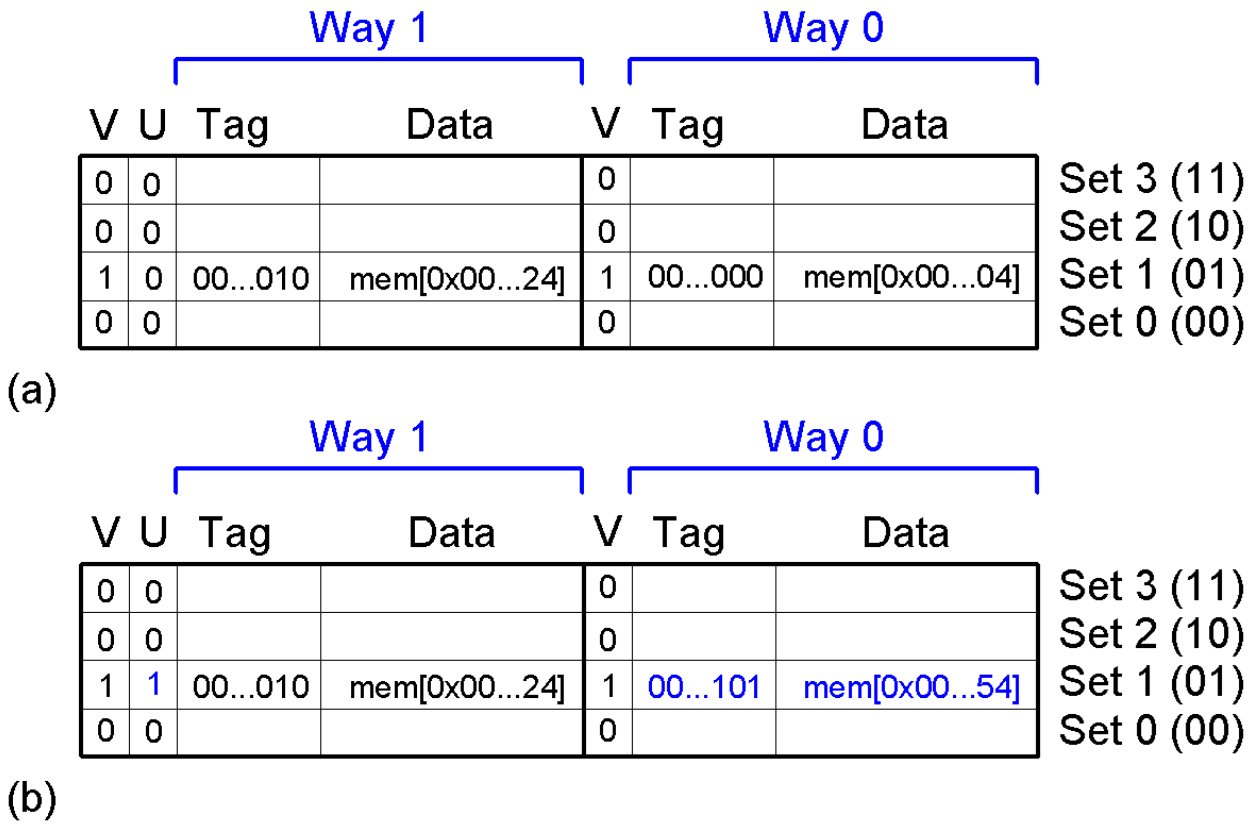
LRU Replacement

```
# MIPS assembly

lw $t0, 0x04($0)
lw $t1, 0x24($0)
lw $t2, 0x54($0)
```



LRU Replacement

```
# MIPS assembly

lw $t0, 0x04($0)
lw $t1, 0x24($0)
lw $t2, 0x54($0)
```



U=0 in (a) indicate way 0 is LRU
U=1 in (b) indicates way 1 is LRU

Cache Summary

- **What data is held in the cache?**
  - Recently used data (temporal locality)
  - Nearby data (spatial locality)
- **How is data found?**
  - Set is determined by address of data
  - Word within block also determined by the address
  - In associative caches, data could be in one of ways
- **What data is replaced?**
  - Least-recently used way in the set