

# Announcement

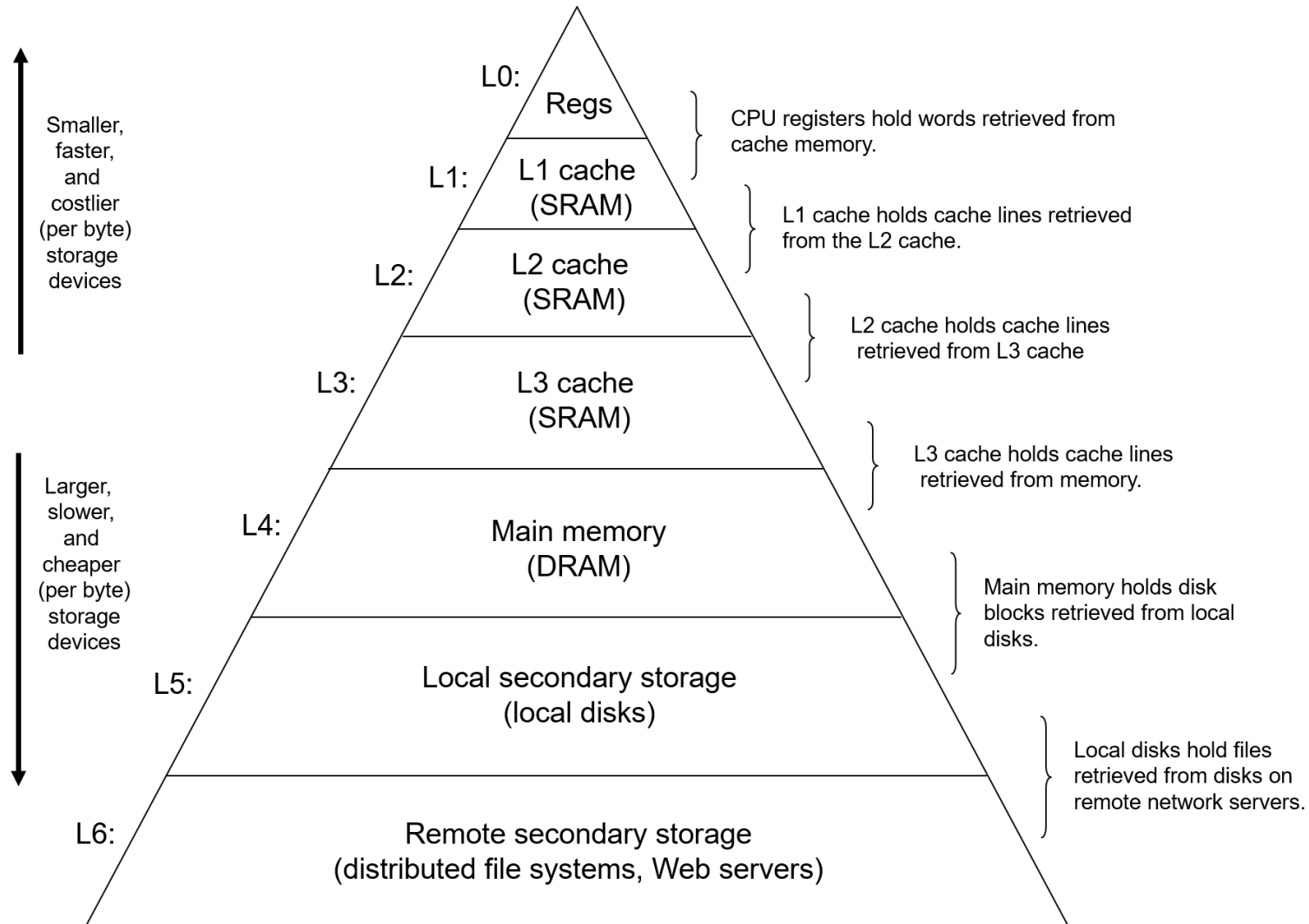
- Assignment 6
  - Due November 11
  - Sorting: Combining C and IA-32 assembly

Lecture 26

# Cache

CPSC 275  
Introduction to Computer Systems

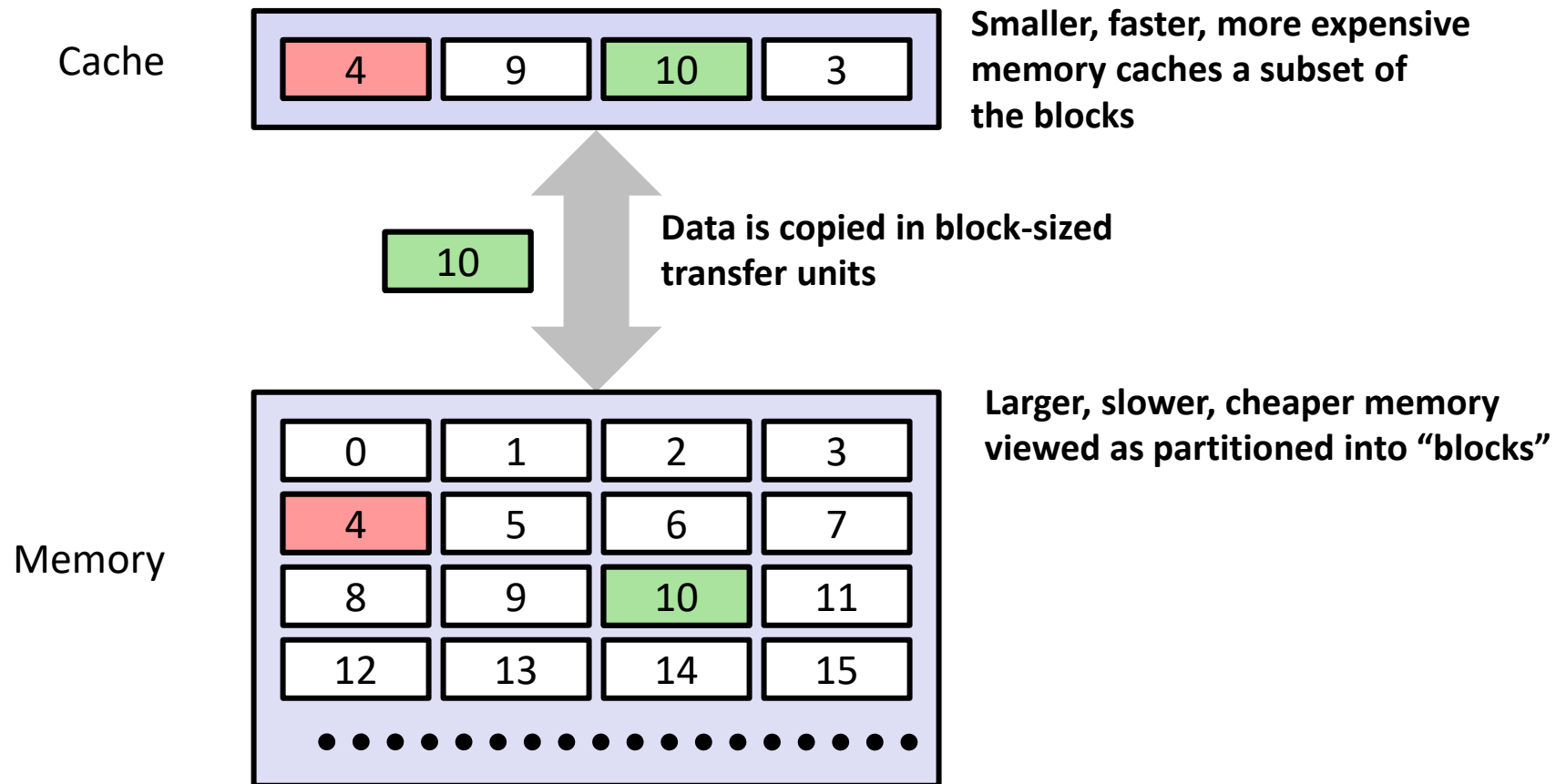
# Memory Hierarchy



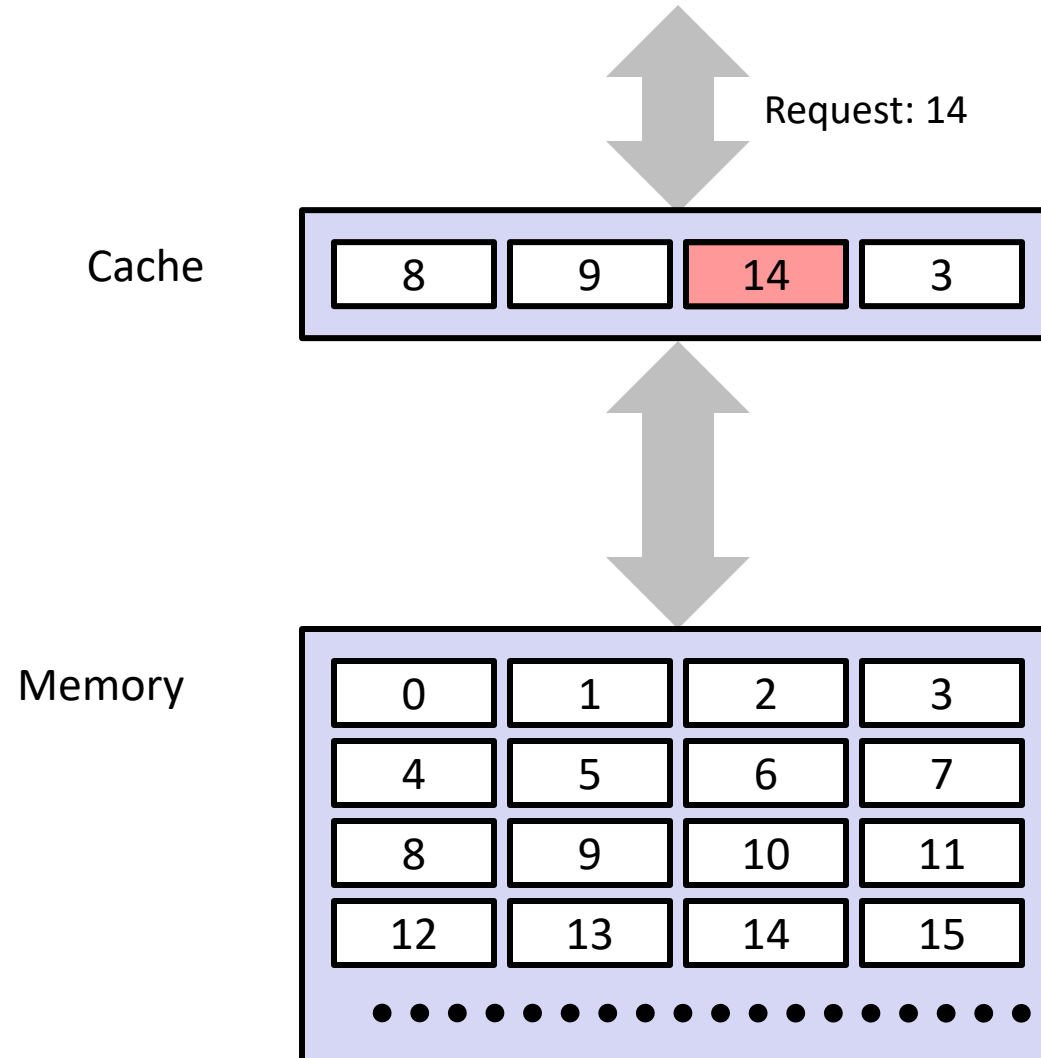
# Cache

- **Cache:** A smaller, faster storage device that acts as a staging area for a subset of the data in a larger, slower device.
- Fundamental idea of a memory hierarchy:
  - For each  $k$ , the faster, smaller device at level  $k$  serves as a cache for the larger, slower device at level  $k+1$ .
- Why do memory hierarchies work?
  - Because of locality, programs tend to access the data at level  $k$  more often than they access the data at level  $k+1$ .
  - Thus, the storage at level  $k+1$  can be slower, and thus larger and cheaper per bit.

# General Cache Concepts



# General Cache Concepts

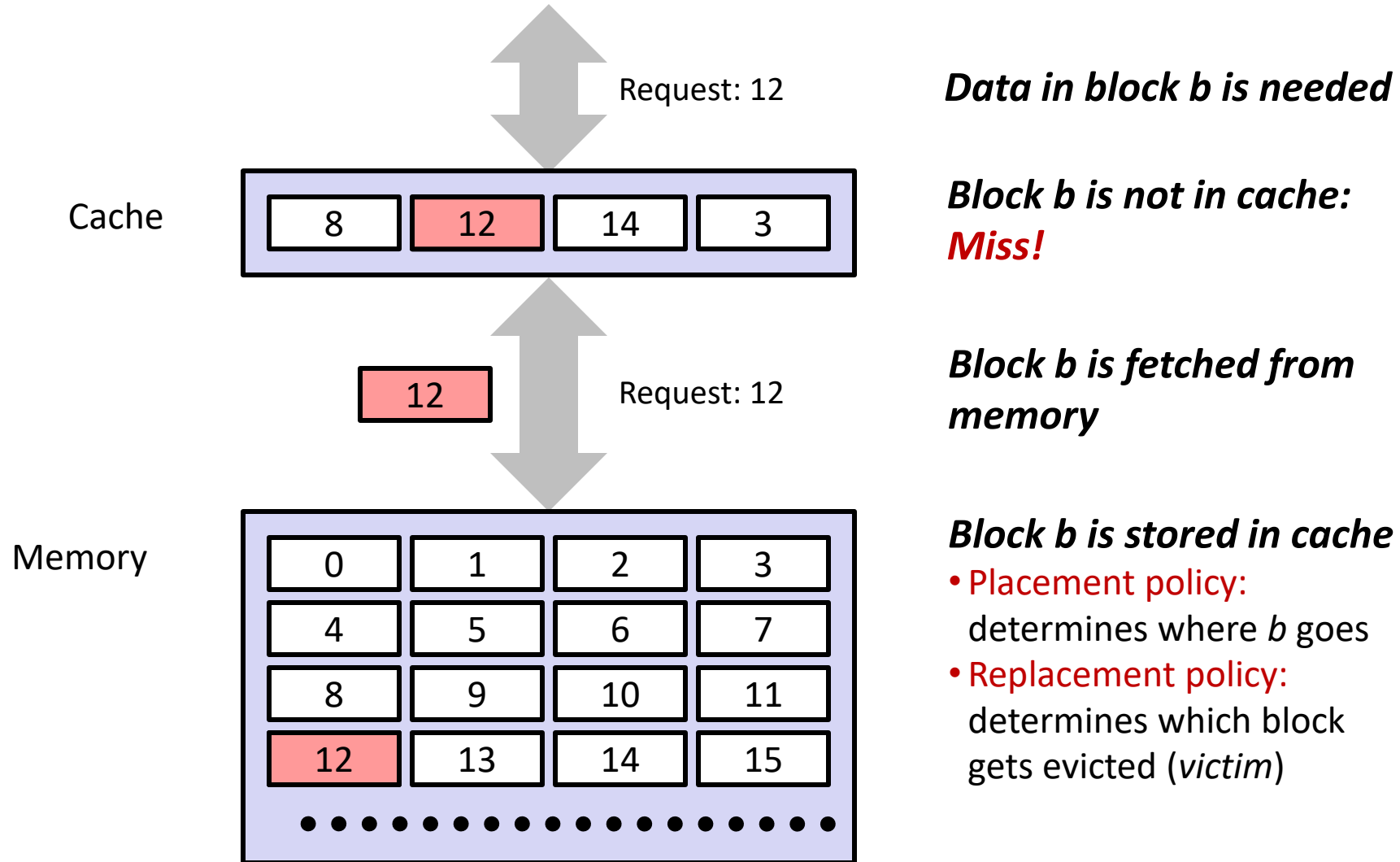


*Data in block b is needed*

*Block b is in cache:*

*Hit!*

# General Cache Concepts: Misses

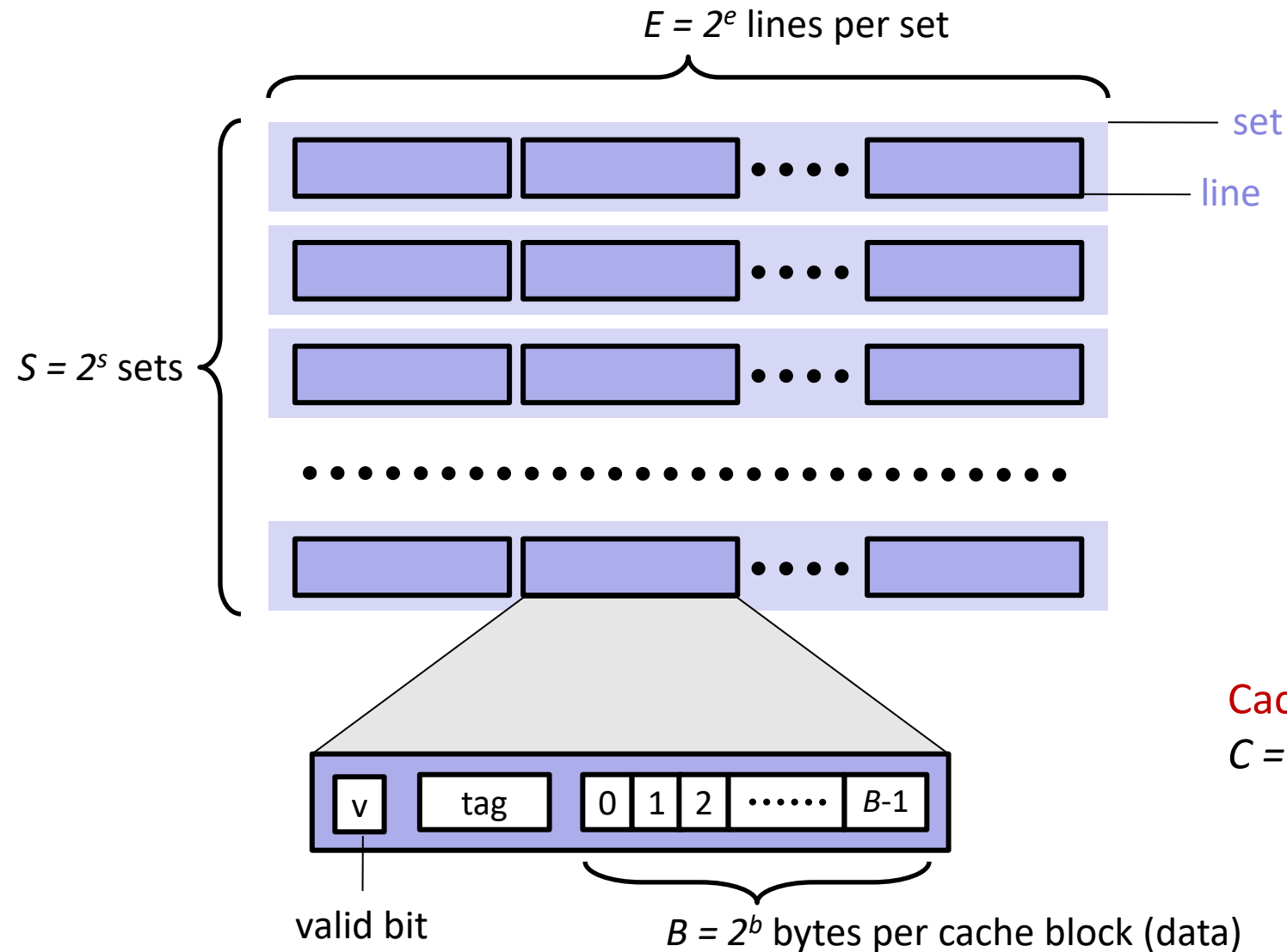


# Types of Cache Misses

- Cold (compulsory) miss
  - Cold misses occur because the cache is empty.
- Conflict miss
  - Most caches limit blocks at level  $k+1$  to a small subset of the block positions at level  $k$ .
  - Conflict misses occur when the level  $k$  cache is large enough, but multiple data objects all map to the same level  $k$  block.
    - e.g., Referencing blocks 0, 8, 0, 8, 0, 8, ... would miss every time.
- Capacity miss
  - Occurs when the set of active cache blocks (known as the *working set*) is larger than the cache.



# General Cache Organization ( $S, E, B$ )



Cache size?

$$C = S \times E \times B \text{ bytes}$$

# Cache Read

