CS 223: Computer Architecture & Organization

Cache Memory

J. K. Deka

Professor

Department of Computer Science & Engineering

Indian Institute of Technology Guwahati, Assam.

So you want fast?

- It is possible to build a computer which uses only static RAM
- This would be very fast
- This would cost high

- Alternatives??

Locality of Reference

- During the course of the execution of a program, memory references tend to cluster

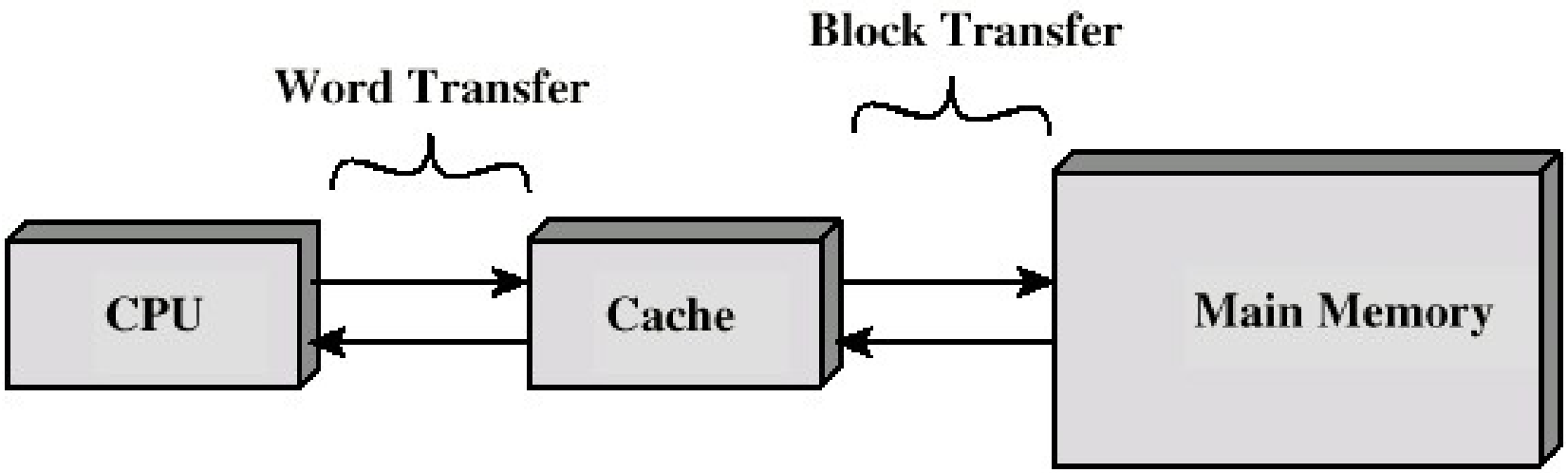
Locality of Reference

- During the course of the execution of a program, memory references tend to cluster

- e.g. Program executes in sequence, loops

Cache

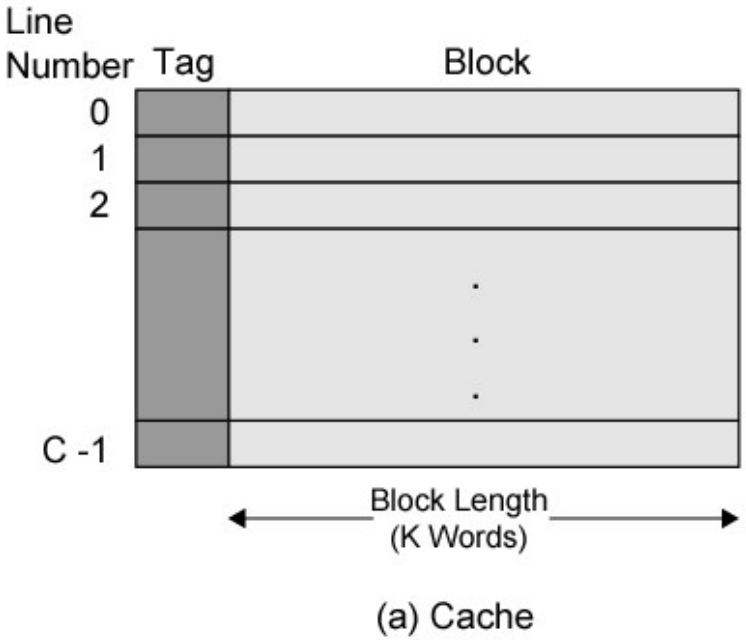
- Small amount of fast memory
- Sits between normal main memory and CPU
- May be located on CPU chip or module



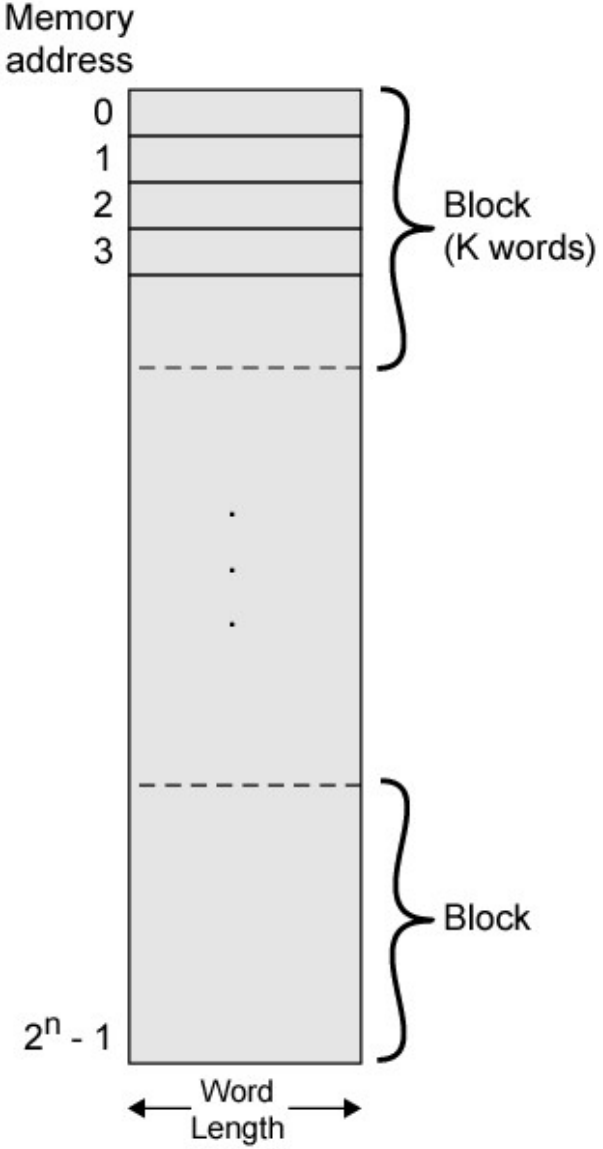
Memory Hierarchy

- Registers
- L1 Cache
- L2 Cache
- Main memory
- Disk cache
- Disk
- Optical
- Tape

Cache/Main Memory Structure



(a) Cache

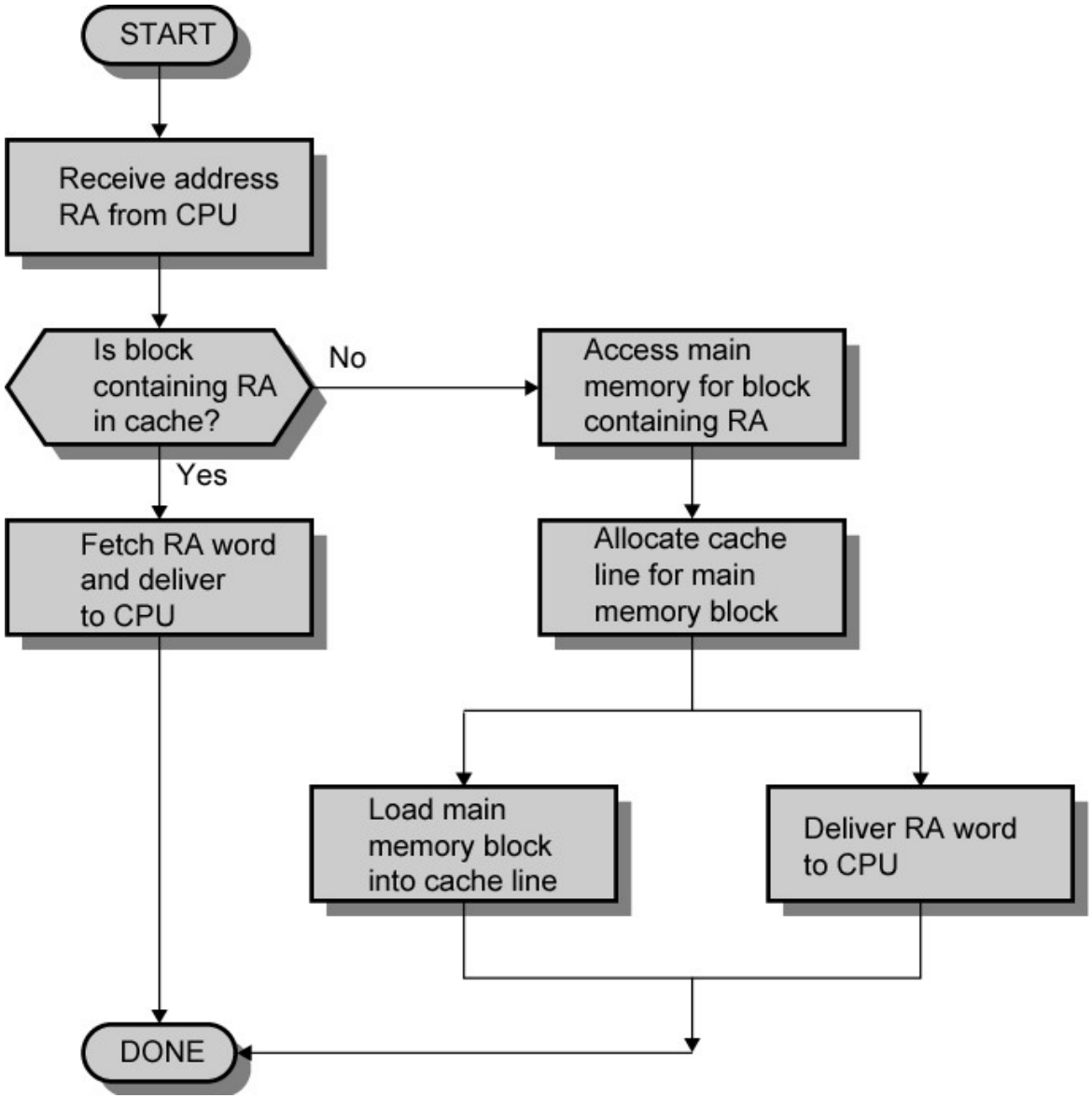


(b) Main memory

Cache operation – overview

- CPU requests contents of memory location
- Check cache for this data
- If present, get from cache (fast)
- If not present, read required block from main memory to cache
- Then deliver from cache to CPU
- Cache includes tags to identify which block of main memory is in each cache slot

Cache Read Operation - Flowchart



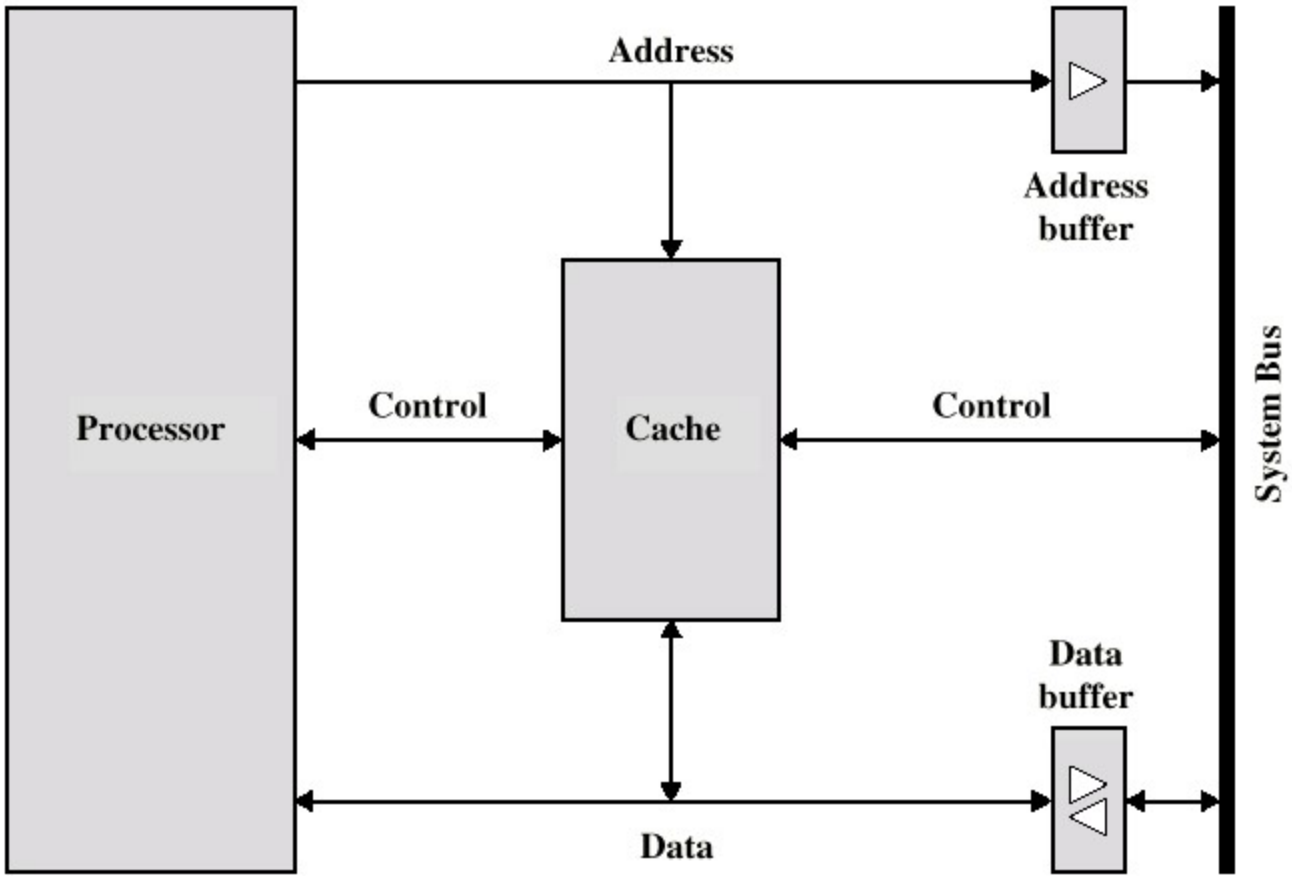
Cache Design

- Size
- Mapping Function
- Replacement Algorithm
- Write Policy
- Block Size
- Number of Caches

Size does matter

- Cost
  - More cache is expensive
- Speed
  - More cache is faster (up to a point)
  - Checking cache for data takes time

Typical Cache Organization



Write Policy

- Must not overwrite a cache block unless main memory is up to date
- Multiple CPUs may have individual caches
- I/O may address main memory directly

Write through

- All writes go to main memory as well as cache
- Multiple CPUs can monitor main memory traffic to keep local (to CPU) cache up to date
- Lots of traffic
- Slows down writes

- Remember bogus write through caches!

Write back

- Updates initially made in cache only
- Update bit for cache slot is set when update occurs
- If block is to be replaced, write to main memory only if update bit is set
- N.B. 15% of memory references are writes

Reference

Computer Organization and Architecture –
Designing for Performance
William Stallings, Seventh Edition

Chapter 04: Cache Memory

Computer Organization
Hamacher, Vranesic and Zaky, Fifth Edition

Chapter05: Page No.: 314 - 329