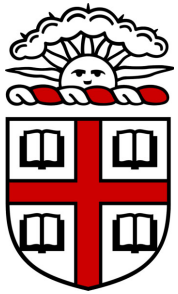
EN164: Design of Computing Systems
Lecture 25: Memory Systems 1

Professor Sherief Reda
http://scale.engin.brown.edu
Electrical Sciences and Computer Engineering
School of Engineering
Brown University
Spring 2011

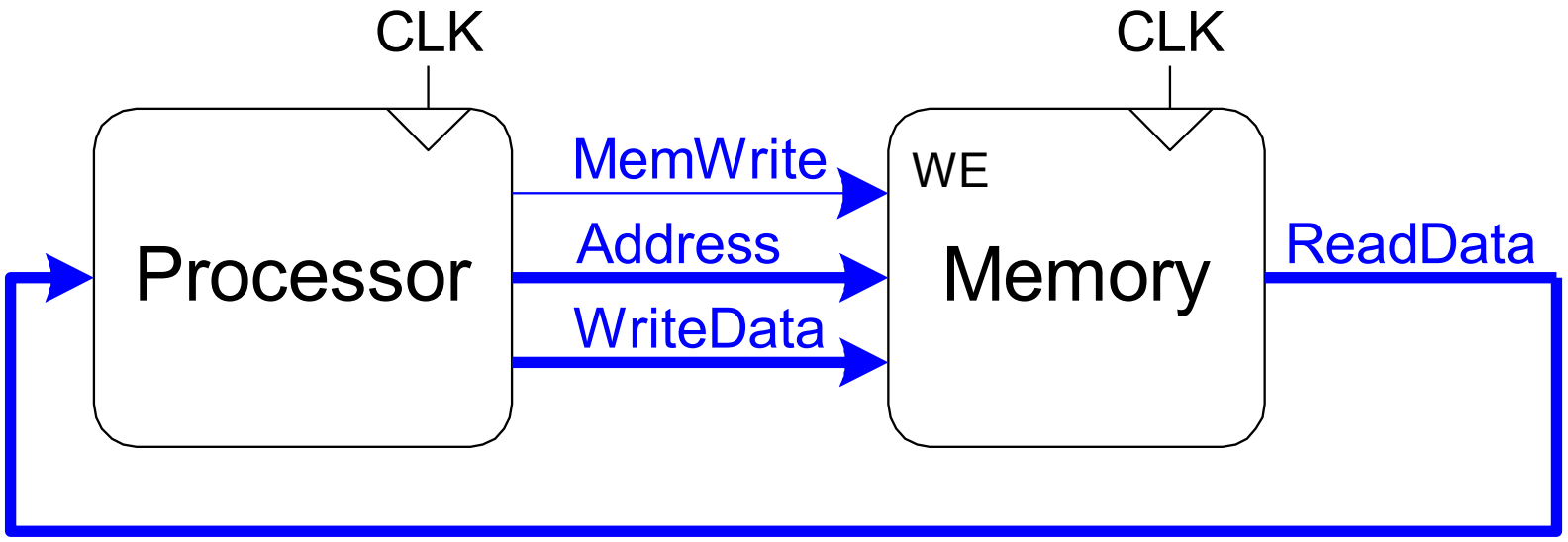


[material from Patterson & Hennessy, 4th ed and Harris 1st ed]

Introduction

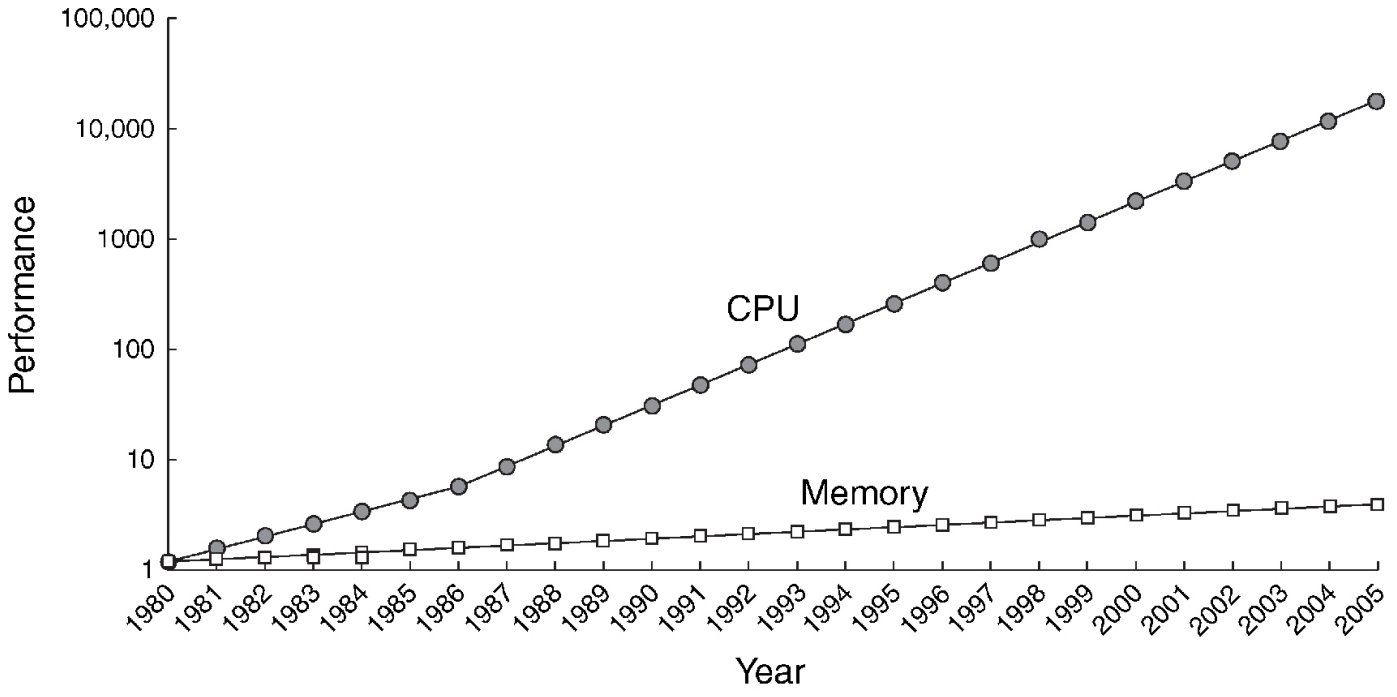
- **Computer performance depends on:**
  - **Processor performance**
  - **Memory system performance**

Memory Interface

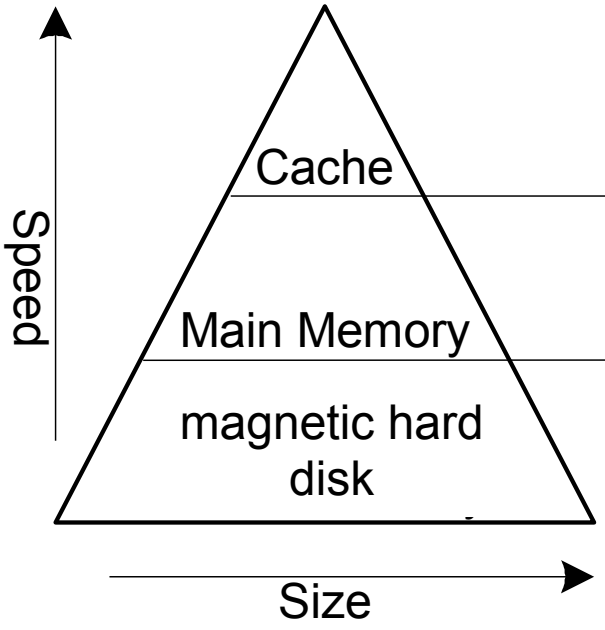


Introduction

- **Up until now, assumed memory could be accessed in 1 clock cycle**



Memory hierarchy



Technology	cost / GB	Access time
SRAM	~ $10,000	~ 1 ns
DRAM	~ $100	~ 100 ns
magnetic hard disk	~ $1	~ 10,000,000 ns

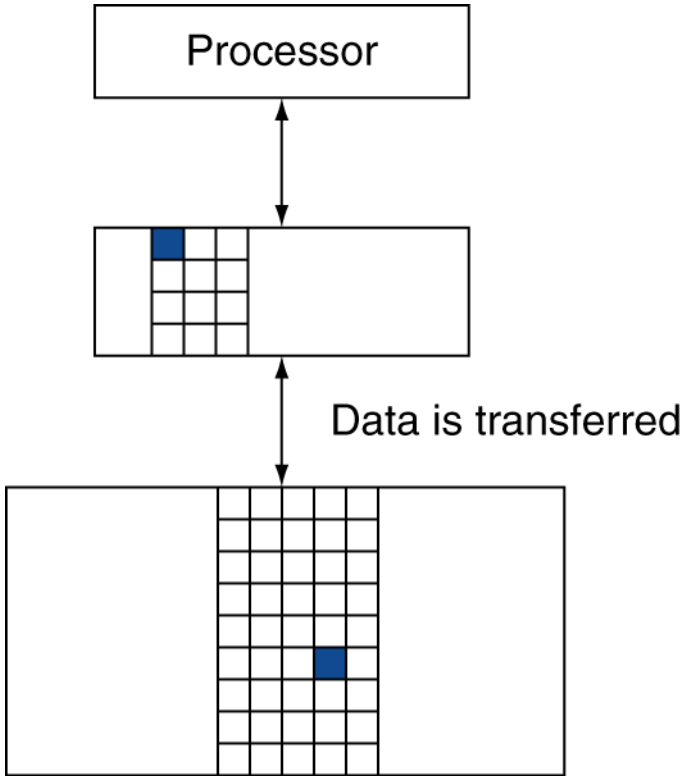
Ideal memory

- Access time of SRAM
- Capacity and cost/GB of disk

Principle of locality

- Exploit locality to make memory accesses fast
- **Temporal Locality:**
  - Locality in time (e.g., if looked at a Web page recently, likely to look at it again soon)
  - If data used recently, likely to use it again soon
  - **How to exploit:** keep recently accessed data in higher levels of memory hierarchy
- **Spatial Locality:**
  - Locality in space (e.g., if read one page of book recently, likely to read nearby pages soon)
  - If data used recently, likely to use nearby data soon
  - **How to exploit:** when access data, bring nearby data into higher levels of memory hierarchy too

Memory hierarchy levels



- **Block (aka line): unit of copying**
  - **May be multiple words**
- **If accessed data is present in upper level**
  - **Hit: access satisfied by upper level**
    - **Hit ratio: hits/accesses**
- **If accessed data is absent**
  - **Miss: block copied from lower level**
    - **Time taken: miss penalty**
    - **Miss ratio: misses/accesses = 1 – hit ratio**
  - **Then accessed data supplied from upper level**

Example

- **A program has 2,000 load and store instructions**
- **1,250 of these data values found in cache**
- **The rest are supplied by other levels of memory hierarchy**
- **What are the hit and miss rates for the cache?**

Hit Rate = 1250/2000 = 0.625

Miss Rate = 750/2000 = 0.375 = 1 – Hit Rate

Example

- **Suppose processor has 2 levels of hierarchy: cache and main memory**
- **t_{cache} = 1 cycle, t_{MM} = 100 cycles**
- **What is the average memory access time (AMAT) of the program from Example 1?**

$$\text{AMAT} = t_{cache} + MR_{cache}(t_{MM})$$
$$= [1 + 0.375(100)] \text{ cycles}$$
$$= 38.5 \text{ cycles}$$