Chapter 5
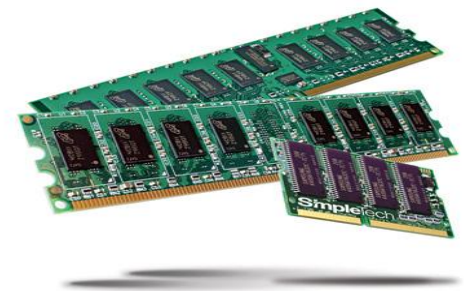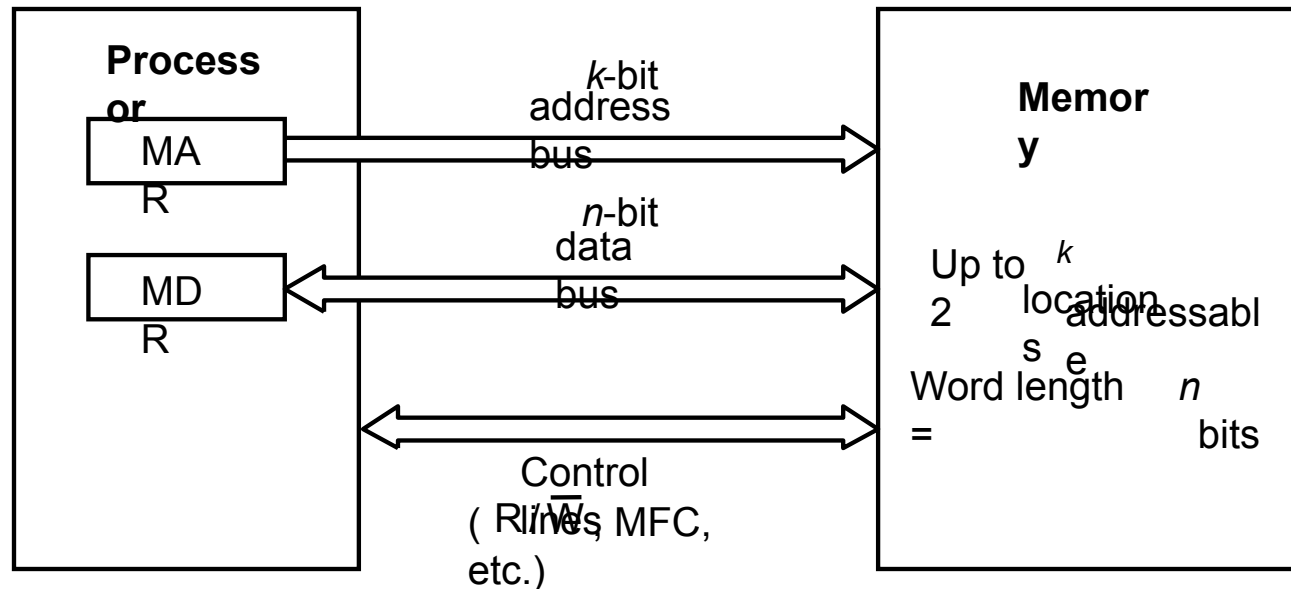
Memory Organization

Fundamental Concepts

The Memory System

Some basic concepts

- Maximum size of the Main Memory
- byte-addressable
- CPU-Main Memory Connection

Processor

MAR

MDR

k-bit address bus

n-bit data bus

Control lines ($R/\overline{W}$, MFC, etc.)

Memory

Up to 2^k addressable locations

Word length = n bits

Some basic concepts(Contd.,)

- ## Measures for the speed of a memory:

  - ### memory access time.
    It is a useful measure of the speed of the memory unit. It is the time that elapses between the initiation of an operation and the completion of that operation (for example, the time between READ and MFC)

  - ### memory cycle time.
    It is an important measure of the memory system. It is the minimum time delay required between the initiations of two successive memory operations

- ## An important design issue is to provide a computer system with as large and fast a memory as possible, within a given cost target.

- ## Several techniques to increase the effective size and speed of the memory:

  - Cache memory (to increase the effective speed).
 - Virtual memory (to increase the effective size).
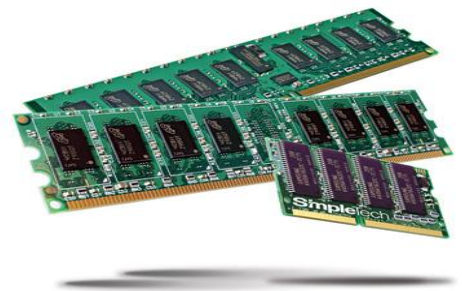
Semiconductor RAM memories

The Memory System

Internal organization of memory chips

- Each memory cell can hold one bit of information.
- Memory cells are organized in the form of an array.
- One row is one memory word.
- All cells of a row are connected to a common line, known as the "word line".
- Word line is connected to the address decoder.
- Sense/write circuits are connected to the data input/output lines of the memory chip.
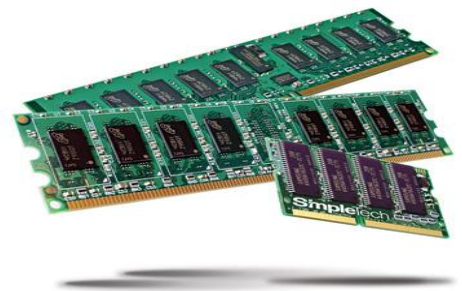
Static RAMs (SRAMs):

- Consist of circuits that are capable of retaining their state as long as the power is applied.
- Volatile memories, because their contents are lost when power is interrupted.
- Access times of static RAMs are in the range of few nanoseconds.
- However, the cost is usually high.

Dynamic RAMs (DRAMs):

- Do not retain their state indefinitely.
- Contents must be periodically refreshed.
- 1 referesh operation can referesh one row at a time.
- 1 chip refresh time=no. of rows * 1 refresh time
- Contents may be refreshed while accessing them for reading.
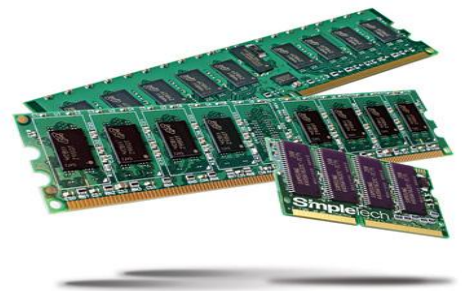
Memory controller

- Address is divided into two parts:
    - High-order address bits select a row in the array.
    - They are provided first, and latched using RAS signal.
    - Low-order address bits select a column in the row.
    - They are provided later, and latched using CAS signal.
- However, a processor issues all address bits at the same time.
- In order to achieve the multiplexing, memory controller circuit is inserted between the processor and memory.

Chip select (CS) signal is required to enable either ROM/RAM.
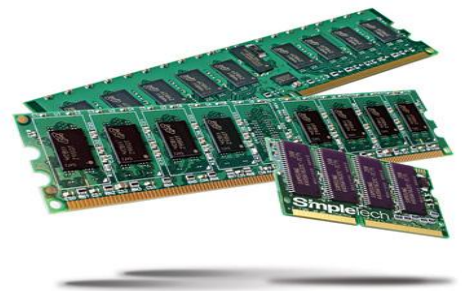
Memory controller (contd..)
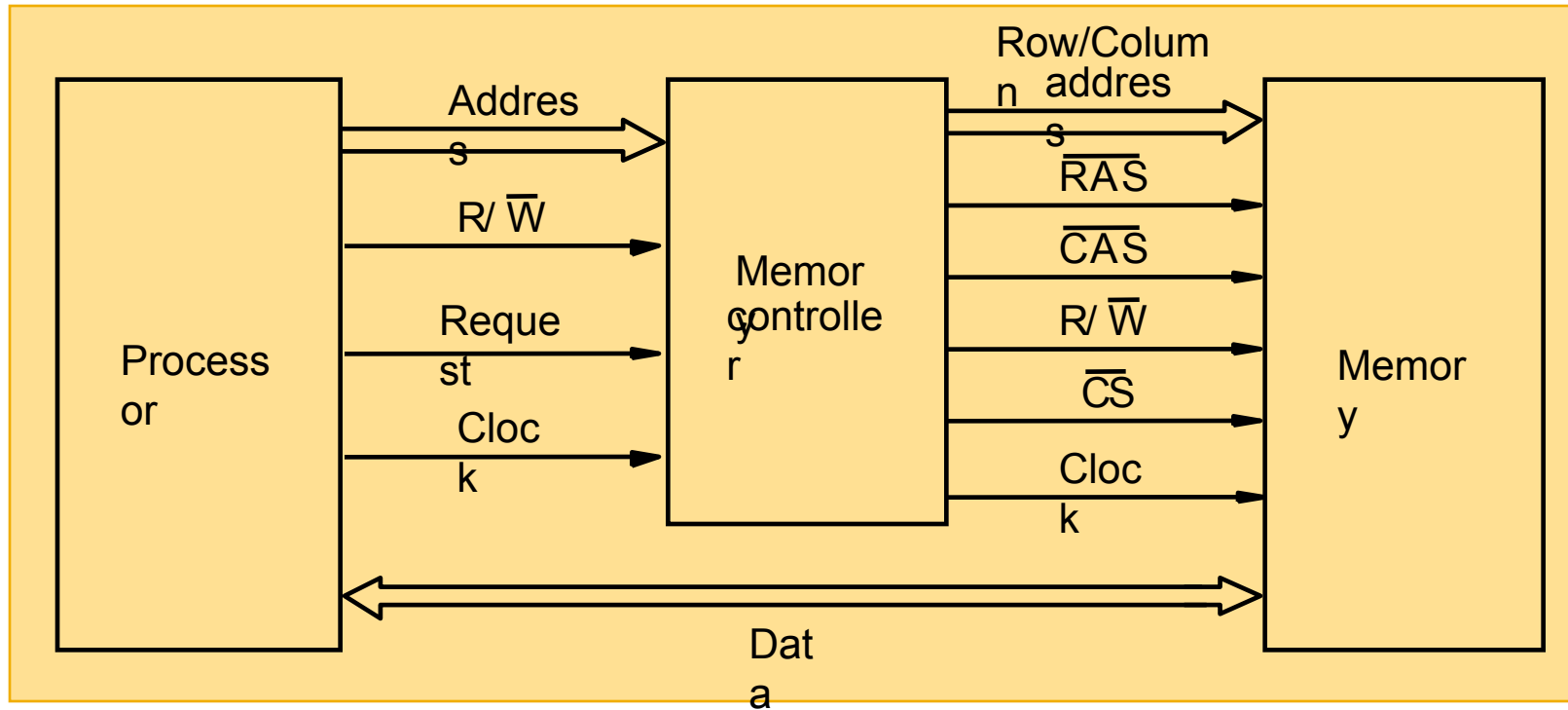
Read-Only Memories (ROMs)

The Memory System

Read-Only Memories (ROMs)

- SRAM and SDRAM chips are volatile:
  - Lose the contents when the power is turned off.
- Many applications need memory devices to retain contents after the power is turned off.
  - For example, computer is turned on, the operating system must be loaded from the disk into the memory.
  - Store instructions which would load the OS from the disk.
  - Need to store these instructions so that they will not be lost after the power is turned off.
  - We need to store the instructions into a non-volatile memory.
- Non-volatile memory is read in the same manner as volatile memory.
  - Separate writing process is needed to place information in this memory.
  - Normal operation involves only reading of data, this type of memory is called Read-Only memory (ROM).
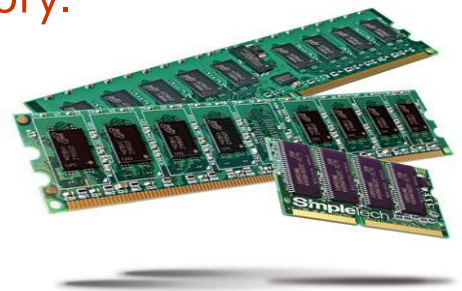
Read-Only Memories (Contd.,)

- **Read-Only Memory:**
  - Data are written into a ROM when it is manufactured.

- **Programmable Read-Only Memory (PROM):**
  - Allow the data to be loaded by a user.
  - Process of inserting the data is irreversible.
  - Storing information specific to a user in a ROM is expensive.

  - Providing programming capability to a user may be better.
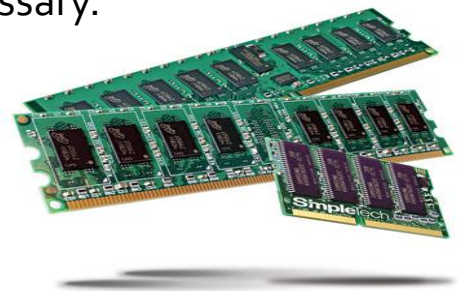
- **Erasable Programmable Read-Only Memory (EPROM):**
  - Stored data to be erased and new data to be loaded.
  - Flexibility, useful during the development phase of digital systems.
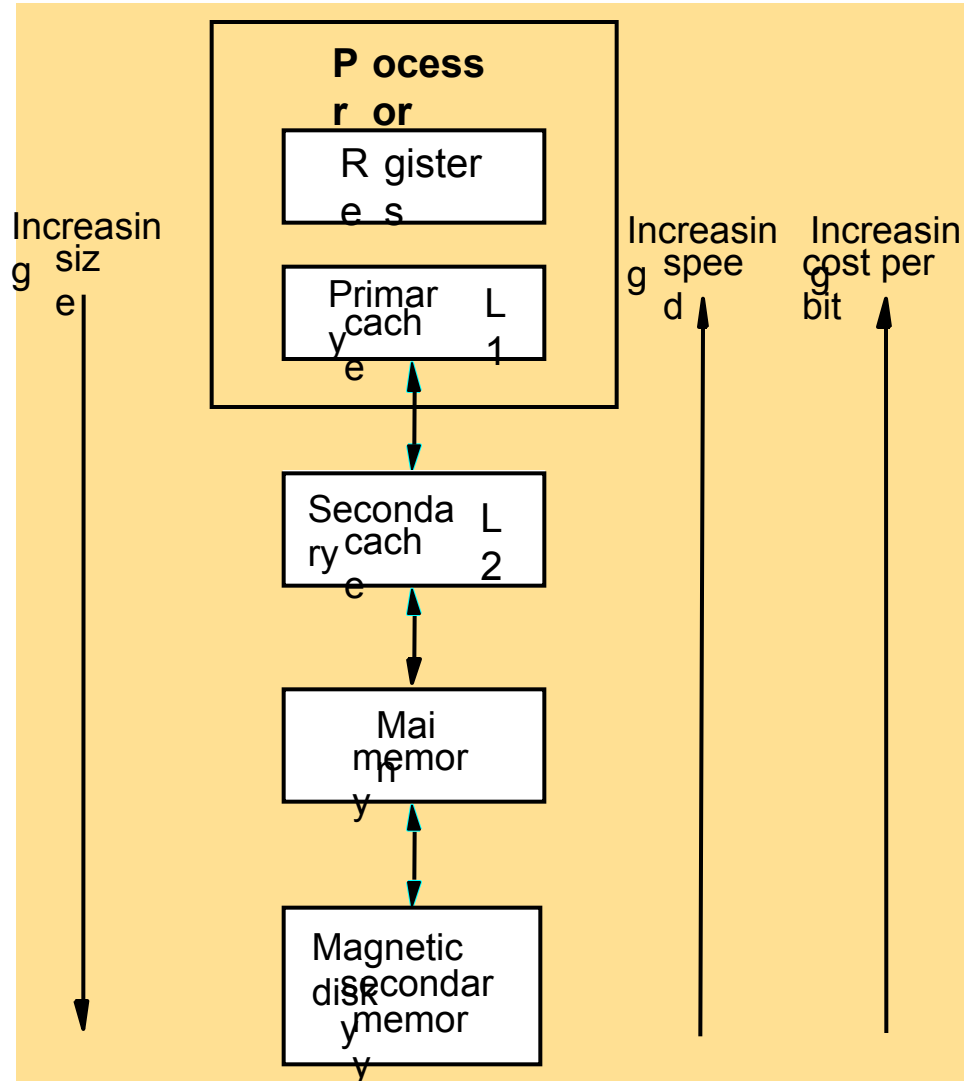 - Erasable, reprogrammable ROM.
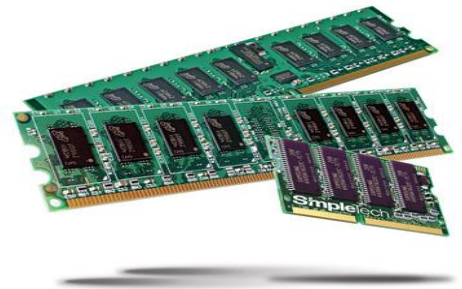
Speed, Size, and Cost

- A big challenge in the design of a computer system is to provide a sufficiently large memory, with a reasonable speed at an affordable cost.
- Static RAM:
  - Very fast, but expensive, because a basic SRAM cell has a complex circuit making it impossible to pack a large number of cells onto a single chip (flipflops).
- Dynamic RAM:
  - Simpler basic cell circuit, hence are much less expensive, but significantly slower than SRAMs (capacitors).
- Magnetic disks:
  - Storage provided by DRAMs is higher than SRAMs, but is still less than what is necessary.
  - Secondary storage such as magnetic disks provide a large amount of storage, but is much slower than DRAMs.

Memory Hierarchy

- **P ocess r or**
  - R gister e s
  - Primar y cach e   L1
- Seconda ry cach e   L2
- Mai n memor y
- Magnetic disk secondar y memor y

Increasin g siz e

Increasin g spee d

Increasin g cost per g bit

- *Fastest access is to the data held in processor registers. Registers are at the top of the memory hierarchy.*
- *Relatively small amount of memory that can be implemented on the processor chip. This is processor cache.*
- *Two levels of cache. Level 1 (L1) cache is on the processor chip. Level 2 (L2) cache is in between main memory and processor.*
- *Next level is main memory,*
- *Much larger, but much slower than cache memory.*
- *Next level is magnetic disks. Huge amount of inexepensive storage.*
- *Speed of memory access is critical, the idea is to bring instructions and data that will be used in the near future as close to the processor as possible.*