Computer Organization & Architecture

Chapter 8 – Memory Basics

Zhang Yang　张杨

cszyang@scut.edu.cn

Autumn 2025

Contents of this lecture

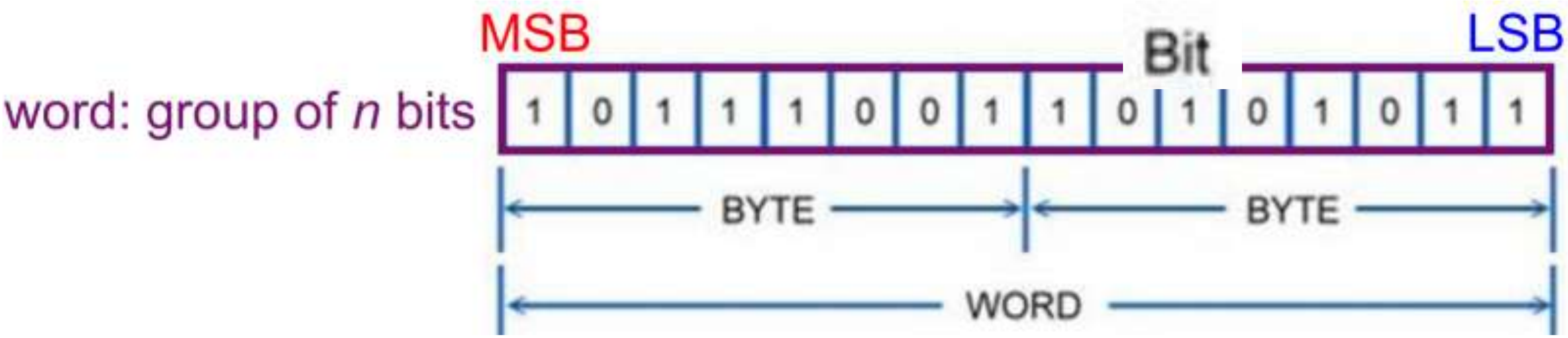
- 2.1 Memory Locations and Addresses & 2.2Memory Operations & 8.1Basic Concepts
  - □ Memory Locations and Addresses
  - □ Byte Addressability
  - □ Byte Ordering
  - □ Bit Ordering
  - □ Word Alignment
  - □ Main Memory Operations
  - □ Characteristic of Memory Systems
  - □ Summary

Internal Memory and External Memory

- The typical computer system is equipped with a hierarchy of memory subsystems.
- Internal Memory (Primary Storage)
  - Internal to the system, directly accessible by the processor.
  - Example: Main Memory, Cache, Processor Registers
- External Memory (Secondary Storage)
  - External to the system, accessible by the processor via an I/O module.

Memory Locations and Addresses (1)

- Main Memory Organization
  - The memory consists of many millions of storage cells.
    - Each cell can store a bit of information (0 or 1).
  - Cells (bits) are organized in groups of n bits.
    - Reason: A single bit represents very little information.
    - A group of n bits: a word (where n is the word length).
      - Example: A word of 16 bits

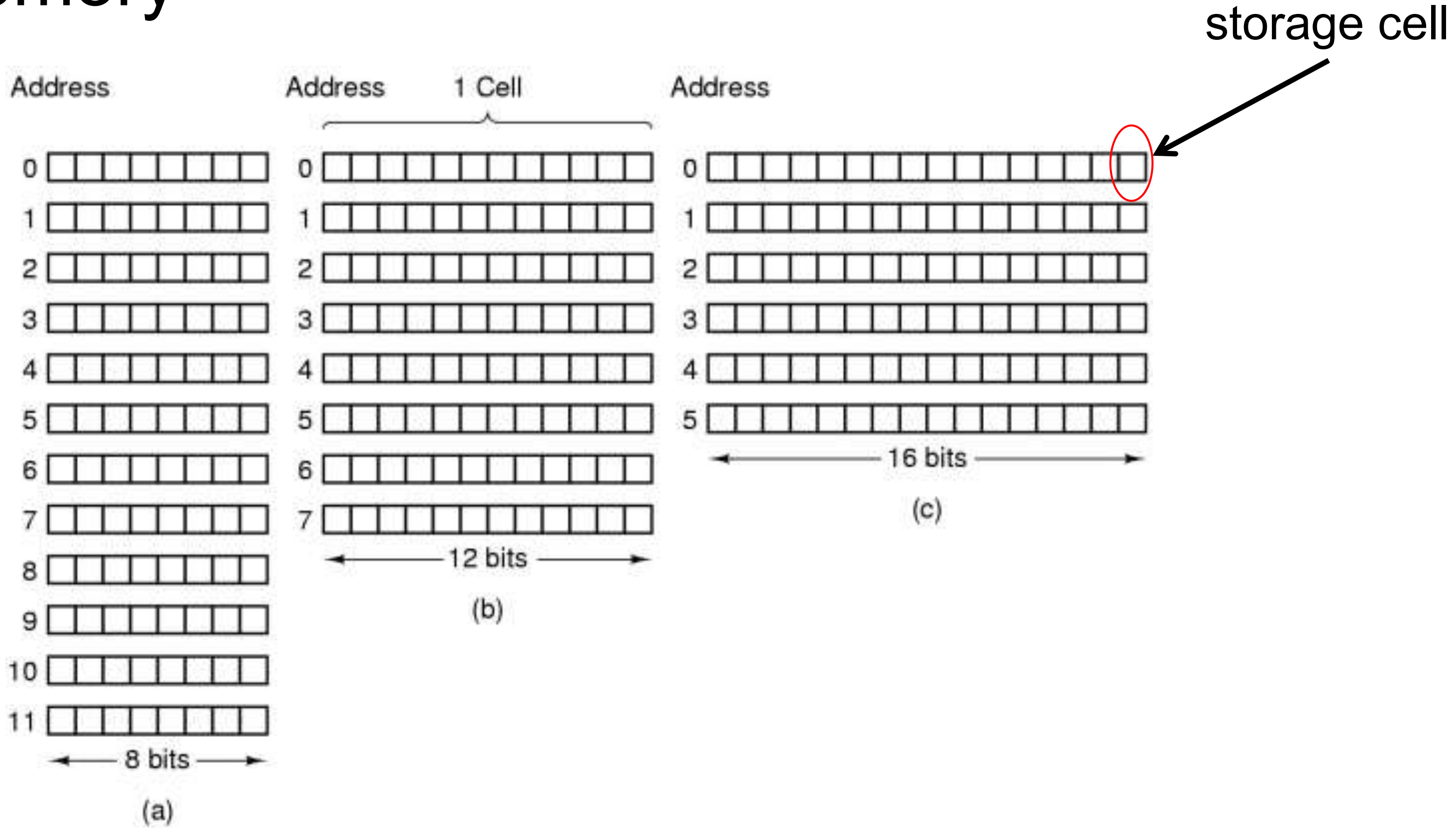


Memory Locations and Addresses (2)

- Main Memory Organization (ctd.)
  - Cells (bits) are organized in groups of n bits. (ctd.)
    - A word can be stored or retrieved in a single, basic operation.
    - Common word lengths in modern computers: 16 to 64 bits.
      - The number of bytes in a word is usually a power of 2.
      - A unit of 8 bits is called a byte (B).

Memory Locations and Addresses (3)

- Main Memory Organization (ctd.)
  - Example: Three ways of organizing a 96-bit memory



Memory Locations and Addresses (4)

- Main Memory Organization (ctd.)
  - The memory of a computer can be schematically represented as a collection of words.

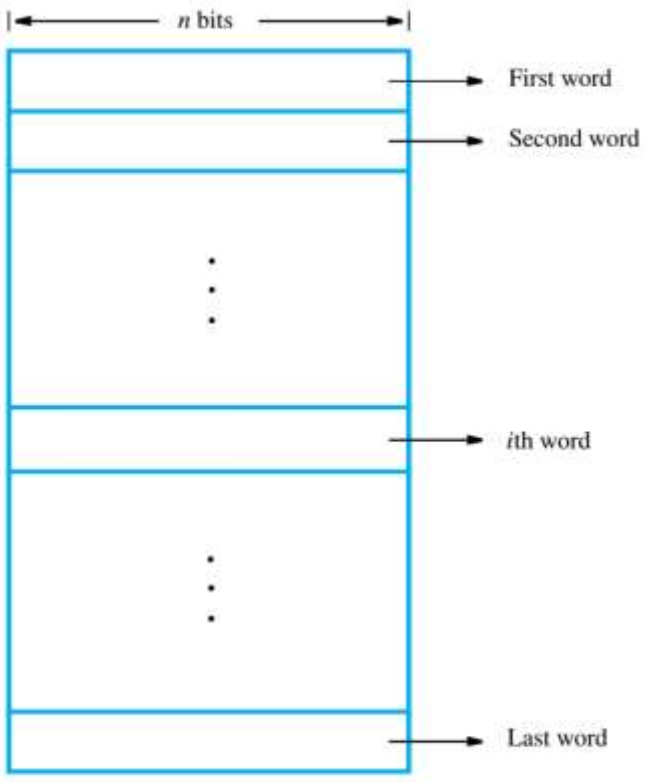


Figure 2.1 Memory words.

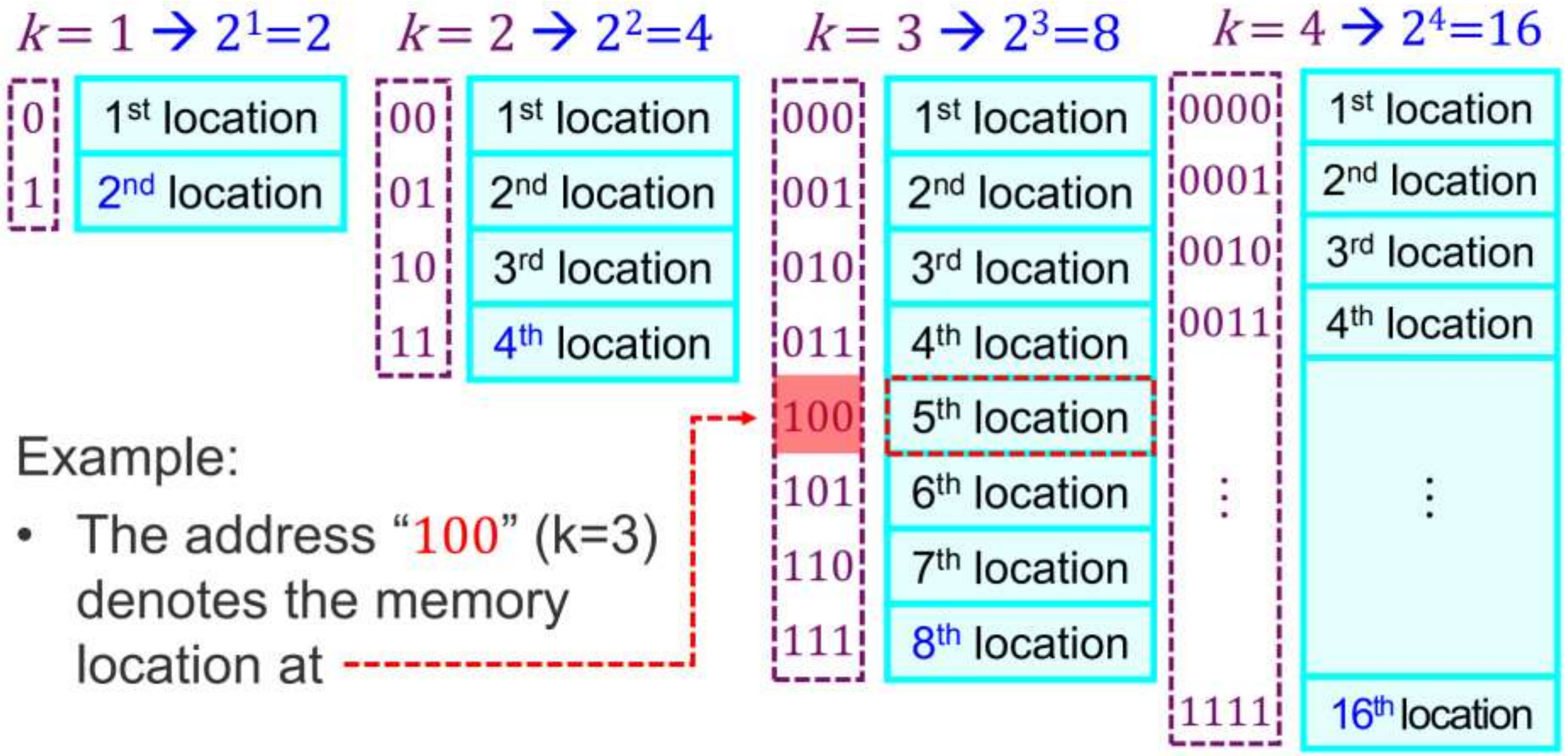
Memory Locations and Addresses (5)

- Question: How to access the contents of memory?

- Address and Address Space
  - □ Addresses are numbers that identify memory locations.
  - □ Accessing the contents of memory requires distinct addresses for each memory location.
    - Format: k-bit addresses can represent 2^k distinct locations.

Memory Locations and Addresses (6)

- Address and Address Space (ctd.)
  - □ Address Example



$k = 1 \rightarrow 2^1 = 2$

0	1st location
1	2nd location

$k = 2 \rightarrow 2^2 = 4$

00	1st location
01	2nd location
10	3rd location
11	4th location

$k = 3 \rightarrow 2^3 = 8$

000	1st location
001	2nd location
010	3rd location
011	4th location
100	5th location
101	6th location
110	7th location
111	8th location

$k = 4 \rightarrow 2^4 = 16$

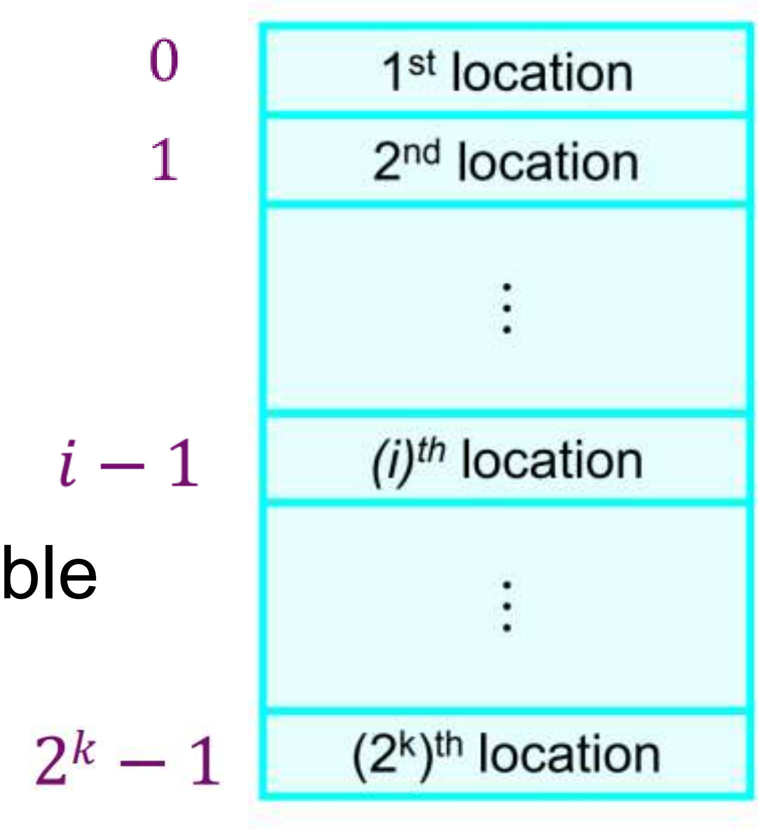
0000	1st location
0001	2nd location
0010	3rd location
0011	4th location
:	:
1111	16th location

Example:

- The address "100" (k=3) denotes the memory location at

Memory Locations and Addresses (7)

- Address and Address Space (ctd.)
  - □ General Rule: It is customary to use numbers from 0 through $2^k – 1$ as the addresses of successive locations in the memory.
  - □ The 2^k addresses constitute the address space of the computer, and the memory can have up to 2^k addressable locations.



Memory Locations and Addresses (8)

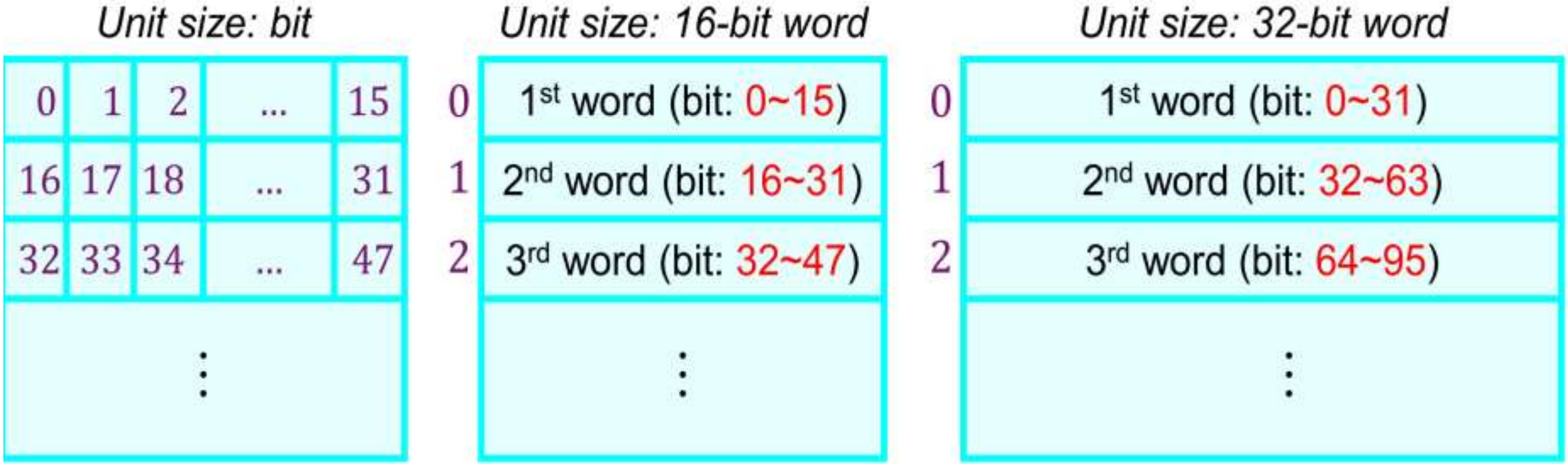
- Address and Address Space (ctd.)
  - □ Address Space Example
    - A 24-bit address can represent 2^{24} = 16,777,216 = 16M distinct locations.
    - A 32-bit address creates an address space of 2^{32} or 4G locations.
  - □ Notational conventions
    - 1K is the number 2^{10} = 1,024
    - 1M is the number 2^{20} = 1,048,576
    - 1G is the number 2^{30} = 1,073,741,824
    - 1T is the number 2^{40} = …

Contents of this lecture

- 2.1 Memory Locations and Addresses & 2.2Memory Operations & 8.1Basic Concepts
  - □ Memory Locations and Addresses
  - □ Byte Addressability
  - □ Byte Ordering
  - □ Bit Ordering
  - □ Word Alignment
  - □ Main Memory Operations
  - □ Characteristic of Memory Systems
  - □ Summary

Byte Addressability (1)

- Basic information quantities: bit, byte, and word.
  - □ A byte (B) is always 8 bits.
  - □ The word length typically ranges from 16 to 64 bits.
- What should be the unit size of an address?
  - □ It is costly to assign distinct addresses to individual bit.
  - □ The word lengths may be different in different computers.



Unit size: bit

0	1	2	...	15
16	17	18	...	31
32	33	34	...	47
⋮				

Unit size: 16-bit word

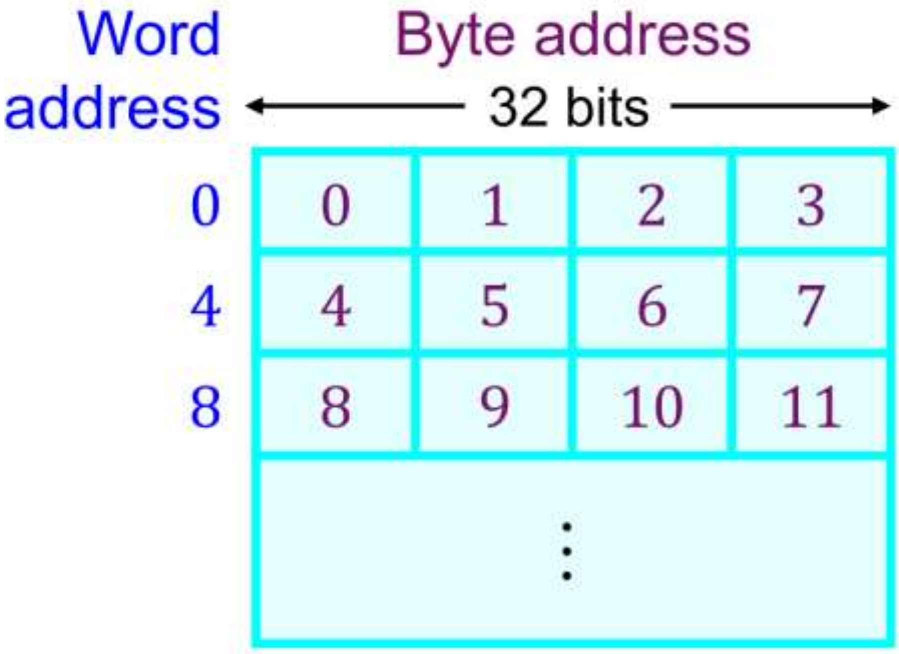
0	1st word (bit: 0~15)
1	2nd word (bit: 16~31)
2	3rd word (bit: 32~47)
	⋮

Unit size: 32-bit word

0	1st word (bit: 0~31)
1	2nd word (bit: 32~63)
2	3rd word (bit: 64~95)
	⋮

Byte Addressability (2)

- The most practical assignment: byte addresses
  - □ Successive addresses represents successive byte locations in the memory.
  - □ E.g. if the word length is 32 bits:
    - Byte addresses: 0, 1, 2, …
    - Word addresses: 0, 4, 8



Word address				
0	0	1	2	3
4	4	5	6	7
8	8	9	10	11
	⋮			

Byte Addressability (3)

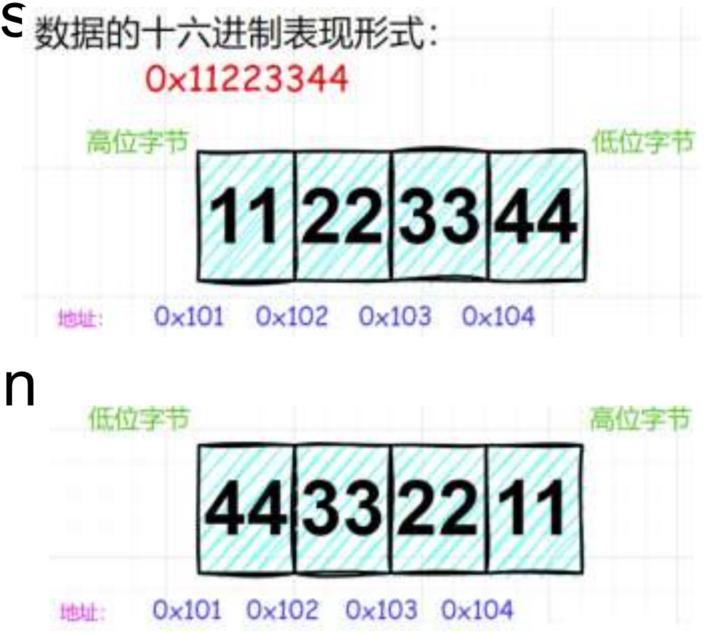
- Byte addressability
  - Each byte location in the memory has its own address and is addressable.
  - We need k-bit addresses to locate 2^k bytes.

Contents of this lecture

- 2.1 Memory Locations and Addresses & 2.2Memory Operations & 8.1Basic Concepts
  - Memory Locations and Addresses
  - Byte Addressability
  - Byte Ordering
  - Word Alignment
  - Main Memory Operations
  - Characteristic of Memory Systems
  - Summary

Byte Ordering (1)

- Endian
  - □ In what order do we load/store the bytes of a multi-byte value?
    - Depends on whether processor is
      - □ “big endian” or “little endian”
  - □ Big Endian:
    - load/store the MSB first
    - i.e., in the lowest address location
  - □ Little Endian:
    - load/store the LSB first
    - i.e., in the lowest address location
  - □ There is no superior endian!



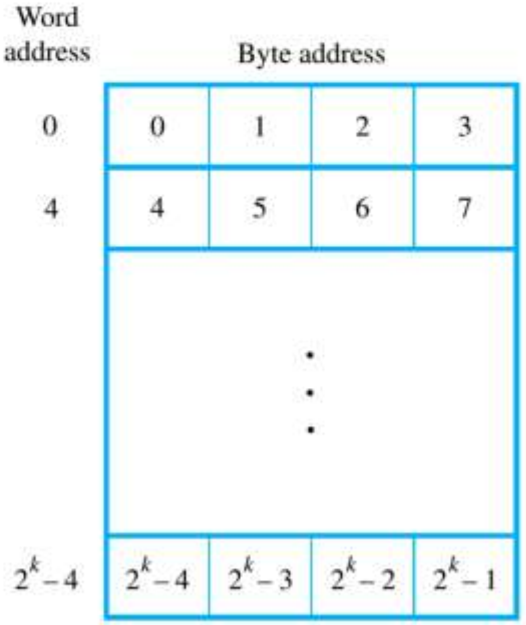
Byte Ordering (2)

- Endian matters when
  - 1) You store a multi-byte value to memory
  - 2) Then you load a subset of those bytes
  - Different endian will give you different results!
  - Different processors support different endian
    - Big endian: motorola 68k, PowerPC (by default)
    - Little endian: intel x86/IA-32, NIOS
    - Supports both: PowerPC (via a mode)
  - E.g.
    - Must account for this if you send data from big-Endian machine to a little-Endian machine.

Byte Ordering (3)

- Big-endian Byte Order
  - □ The most significant byte (the "big end") of the data is placed at the byte with the lowest address. The rest of the data is placed in order in the next three bytes in memory.
  - □ Example: : Figure 2.3 (a)

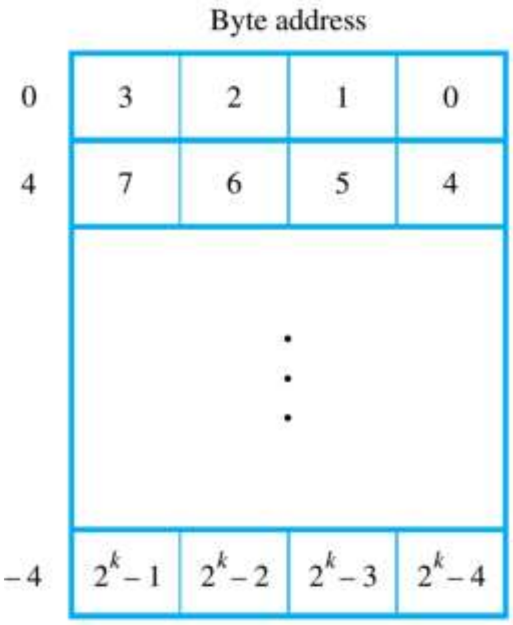
    Big-endian assignment for 32-bit word-length byte-addressable memory.



(a) Big-endian assignment

Byte Ordering (4)

- Little-endian Byte Order
  - □ The least significant byte (the "little end") of the data is placed at the byte with the lowest address. The rest of the data is placed in order in the next three bytes in memory.
  - □ Example: Figure 2.3 (b) Little-endian assignment for 32-bit word-length byte-addressable memory.



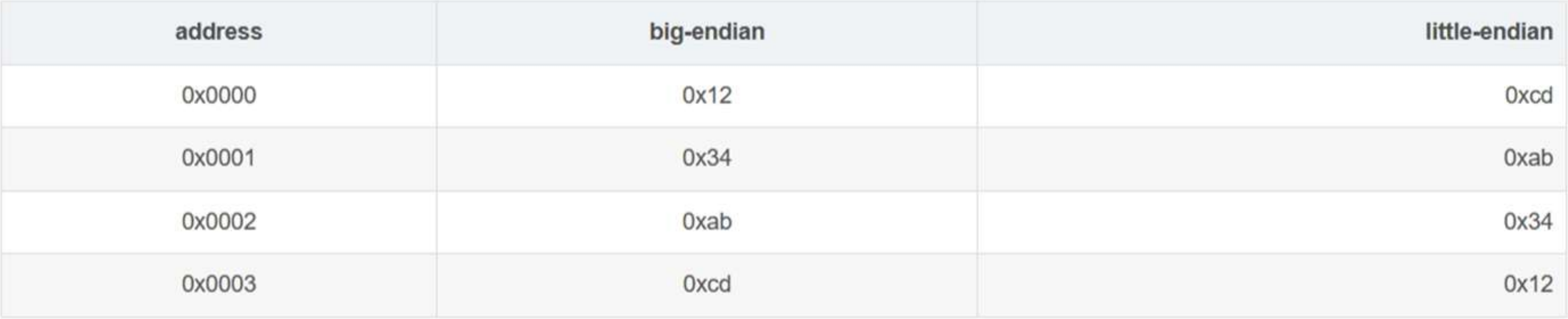
Byte address

0	3	2	1	0
4	7	6	5	4
	⋮			
−4	$2^k - 1$	$2^k - 2$	$2^k - 3$	$2^k - 4$

(b) Little-endian assignment

Byte Ordering (5)

- Note
  - □ In both cases, byte addresses 0, 4, 8, …, are taken as the addresses of successive words in the memory and are the addresses used when specifying memory operations for words.
- Example
  - □ Write 0x1234abcd to memory address 0x0000



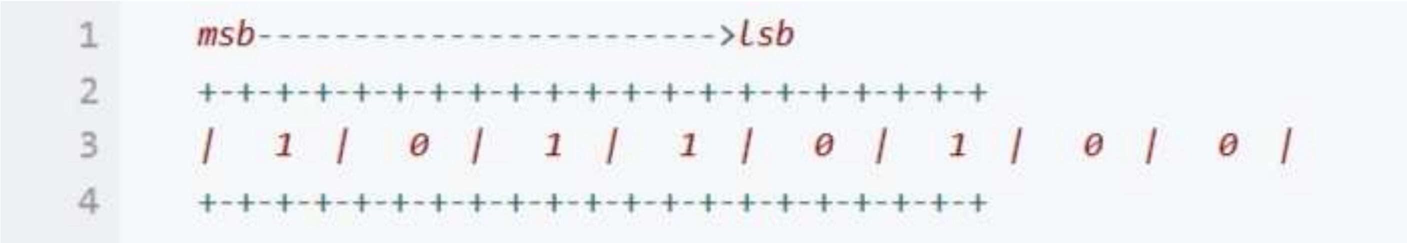
address	big-endian	little-endian
0x0000	0x12	0xcd
0x0001	0x34	0xab
0x0002	0xab	0x34
0x0003	0xcd	0x12

Contents of this lecture

- 2.1 Memory Locations and Addresses & 2.2Memory Operations & 8.1Basic Concepts
  - □ Memory Locations and Addresses
  - □ Byte Addressability
  - □ Byte Ordering
  - □ Bit Ordering
  - □ Word Alignment
  - □ Main Memory Operations
  - □ Characteristic of Memory Systems
  - □ Summary

Bit Ordering (1)

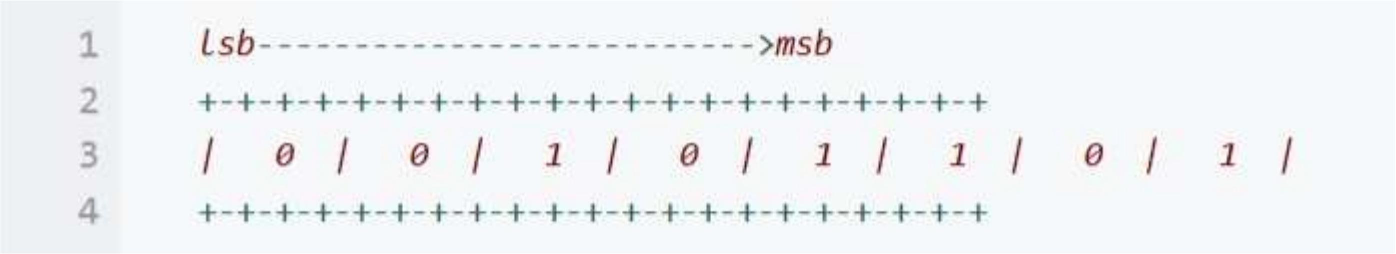
- Bit order refers to the direction in which bits are represented in a byte of memory.
- MSB
  - □ Most significant bit
- LSB
  - □ Least significant bit
- Big Endian
  - □ E.g. 0xB4(10110100)



```
msb------------------------>lsb
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   1  |   0  |   1  |   1  |   0  |   1  |   0  |   0  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Bit Ordering (2)

- Little Endian
  - □ E.g. 0xB4(10110100)



```
lsb-------------------------->msb
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   0   |   0   |   1   |   0   |   1   |   1   |   0   |   1   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Contents of this lecture

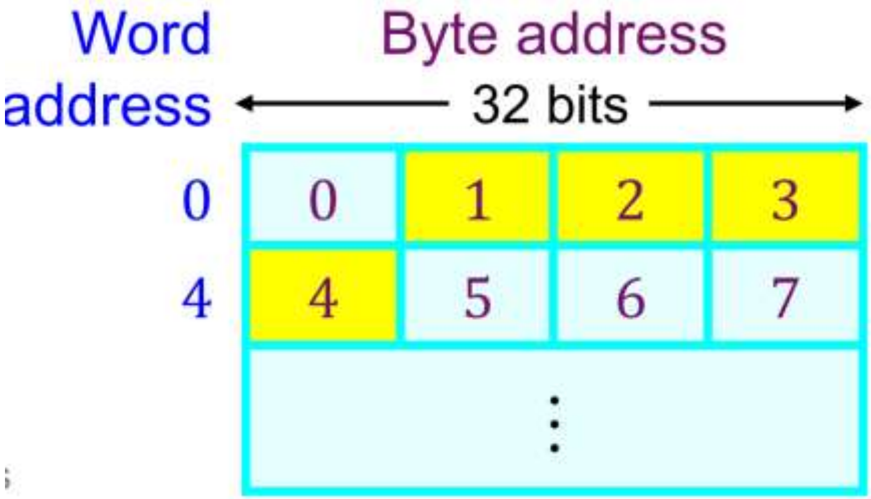
- 2.1 Memory Locations and Addresses & 2.2Memory Operations & 8.1Basic Concepts
  - □ Memory Locations and Addresses
  - □ Byte Addressability
  - □ Byte Ordering
  - □ Word Alignment
  - □ Main Memory Operations
  - □ Characteristic of Memory Systems
  - □ Summary

Word Alignment (1)

- 32-bit words align naturally at addresses 0, 4, 8, …
  - Aligned addresses: Word begins at a byte address that is a multiple of the number of bytes in a word.
  - The aligned addresses for 16-bit and 64-bit words:
    - 16-bit word: 0, 2, 4, 6, 8, 10, …
    - 64-bit word: 0, 8, 16, …

Word Alignment (2)

- 32-bit words align naturally at addresses 0, 4, 8, … (ctd.)
  - Unaligned accesses are either not allowed or slower.
    - E.g. read a 32-bit word from the byte address 0x01
      - Note: 0x represents the hexadecimal number system.



Contents of this lecture

- 2.1 Memory Locations and Addresses & 2.2Memory Operations & 8.1Basic Concepts
  - Memory Locations and Addresses
  - Byte Addressability
  - Byte Ordering
  - Word Alignment
  - Main Memory Operations
  - Characteristic of Memory Systems
  - Summary

Main Memory Operations (1)

- Connection of the Memory to the Processor

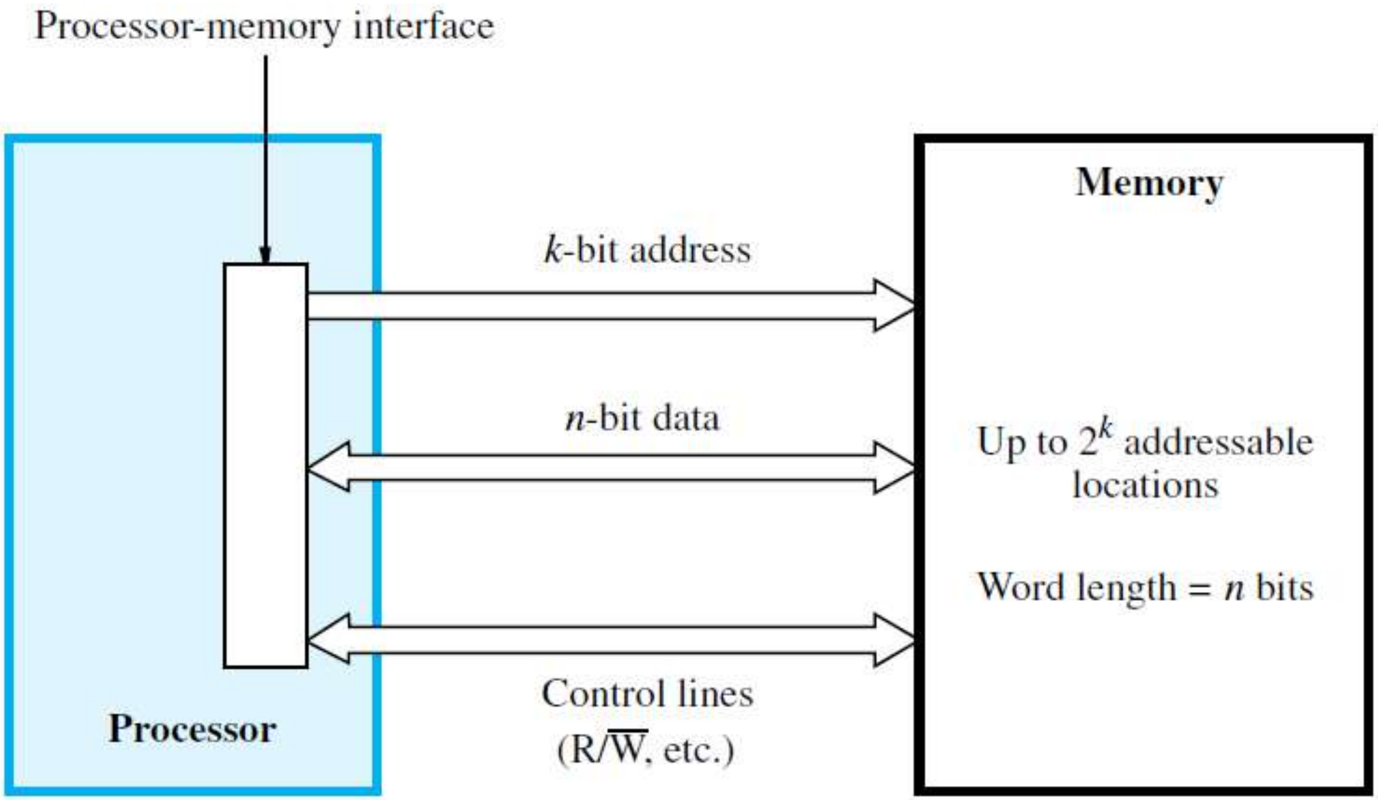


Figure 8.1 Connection of the memory to the processor.

Main Memory Operations (2)

- Processor-Memory Interface

Main Memory Operations (3)

- Read (Load) Operation
  - Transfers a copy of the contents of a specific memory location to the processor.
  - Processor: Load the address of the required memory location into the MAR register and set the R/W line to 1.
  - Memory: Place the data from the addressed location onto the data lines and confirm this action by asserting the MFC signal.
  - Processor: Upon receipt of the MFC signal, the processor loads the data on the data lines into the MDR register.

Main Memory Operations (4)

- Write (Store) Operation
  - Transfer an item of information from the processor to a specific location, destroying the former contents of that location.
  - Processor: Load the address of the specific location into MAR and load the data into MDR register. It also set the $R/\overline{W}$ line to 0.
  - Memory: When the data have been written, it responses processor with MFC signal.

Contents of this lecture

- 2.1 Memory Locations and Addresses & 2.2Memory Operations & 8.1Basic Concepts
  - □ Memory Locations and Addresses
  - □ Byte Addressability
  - □ Byte Ordering
  - □ Word Alignment
  - □ Main Memory Operations
  - □ Characteristic of Memory Systems
  - □ Summary

Characteristics of Memory Systems

Reference from 4.1 Computer Memory System Overview

《Computer Organization and Architecture, Designing for Performance, 9th Edition》

- Physical Types
- Capacity
- Unit of Transfer
- Access Methods
- Performance
- Physical Characteristics

Physical Types

- Semiconductor
  - Example: Semiconductor RAM and Semiconductor ROM
- Magnetic Surface
  - Example: Magnetic disk and Magnetic tape
- Optical
  - Example: CD, CD-R, CD-RW, DVD, Blu-Ray

Capacity

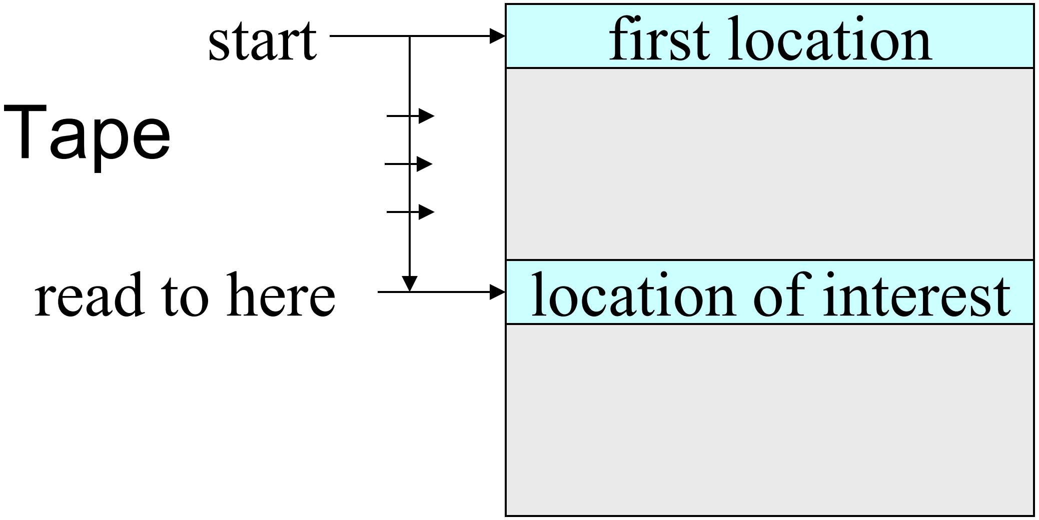
- For main memory, it is expressed using
  - Number of Word
    - Specifies the number of words available in the particular memory device.
  - Word size
    - Number of bits in natural unit of organization
    - The common word size is 8-bits, 16 bits, 32 bits and 64 bits.
- Example
  - If memory capacity is 4K×8, then its word size is 8, and the number of word is 4K.

Unit of Transfer

- Number of bits read from, or written into memory at a time.
- Internal Memory
  - Usually equal to the number of data lines into and out of the memory module.
  - It is often equal to the number of the word length, but it may not be.
- External Memory
  - Data are often transferred in much larger units than a word, and these are referred to as blocks.

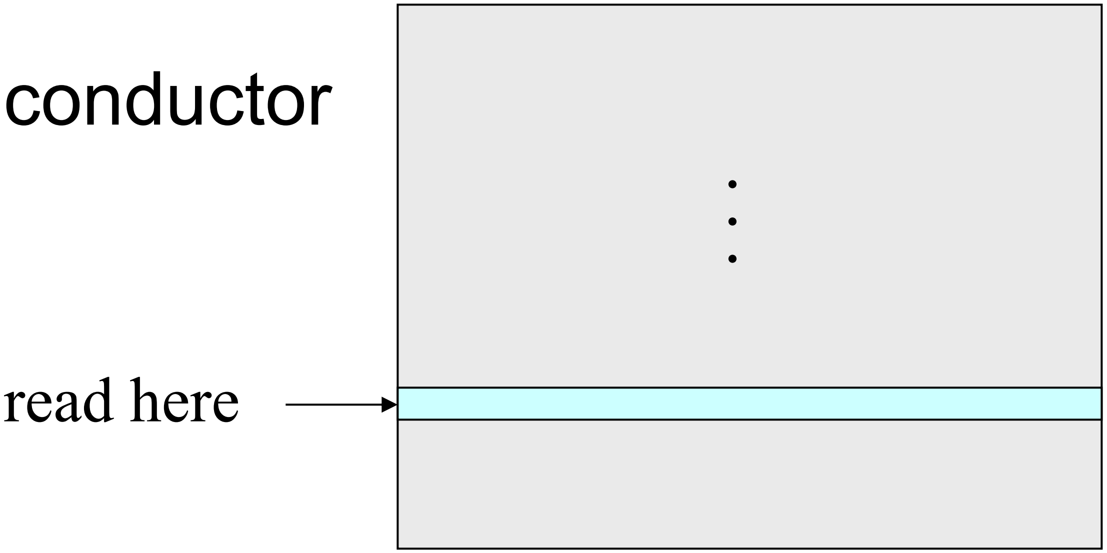
Methods of Accessing (1)

- Sequential Access
  - □ Memory is organized into units of data, called records.
  - □ If current record is 1, then in order to read record N, it is necessary to read records 1 through N-1.
  - □ The time to access an arbitrary record is highly variable.
  - □ Example: Magnetic Tape



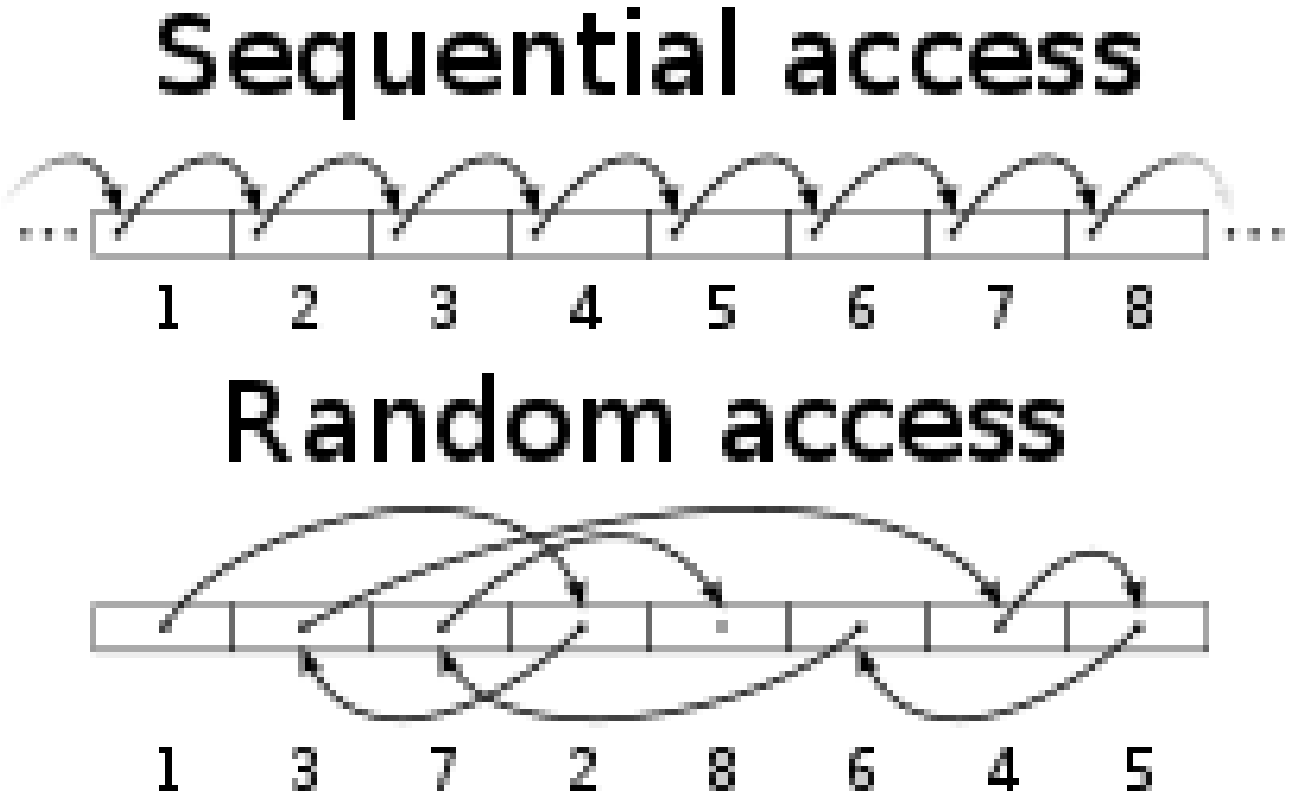
Methods of Accessing (2)

- Random Access
  - □ Any memory location can be selected at random and directly addressed and accessed.
  - □ The time to access a given location is independent of the location’s address and is constant.
  - □ Example: Semiconductor RAM Memory



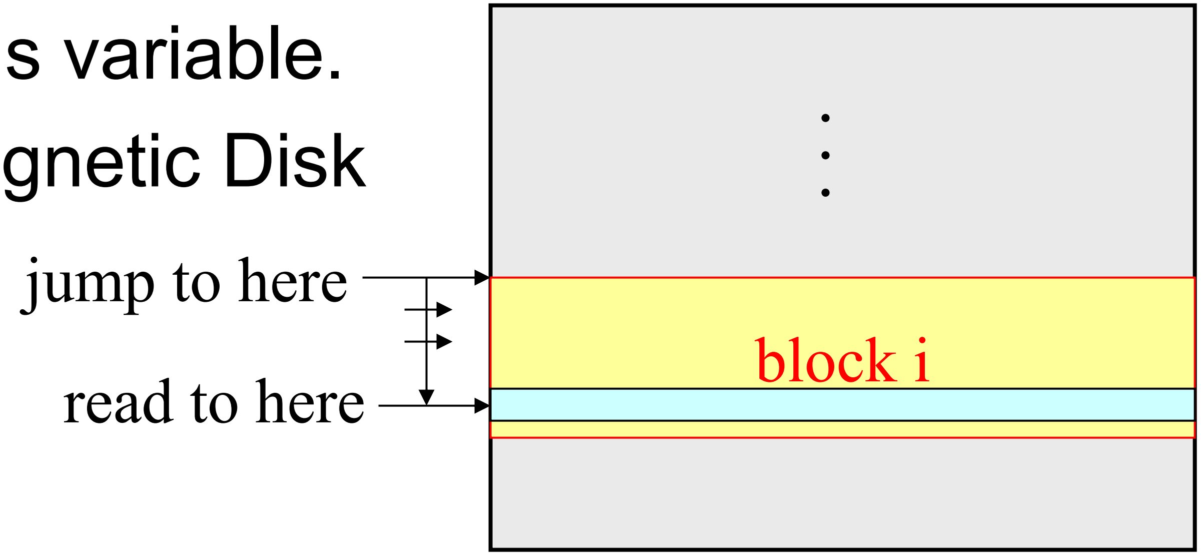
Methods of Accessing (3)

- Sequential vs. Random Access



Methods of Accessing (4)

- Direct Access
  - Individual blocks or records have a unique address based on physical location.
  - Access is accomplished by direct access to reach a general vicinity plus sequential searching, counting or waiting to reach the final location.
  - Access time is variable.
  - Example: Magnetic Disk



Methods of Accessing (5)

- Associative Access
  - This is a random access type of memory that enables one to make a comparison of desired bit locations within a word for a specified match, and to do this for all words simultaneously.
  - Access time is constant.
  - Example: Cache Memory

Performance (1)

- Access Time
  - Time between presenting the address and getting the valid data (memory or other storage)
  - Random-access memory
    - The time it takes to perform a read or write operation.
  - Non-random-access memory
    - The time it takes to position the read-write mechanism at the desired location.

Performance (2)

- Memory Cycle Time (for random access memory)
  - Time from a memory access to the next memory access
  - It consists of the access time plus any additional time required before a second access can commence.
  - Memory Cycle Time = Access Time + Recovery Time
  - Access time and memory cycle time are all measures of the speed of memory units.

Physical Characteristic (1)

- Volatile/Nonvolatile
  - Volatile: Information decays naturally or is lost when electrical power is switched off.
  - Nonvolatile: Information once recorded remains without deterioration until deliberately changed. No electrical power is needed to retain information.
  - Example: Magnetic surface memories are nonvolatile. Semiconductor memory may be either volatile or nonvolatile.

Physical Characteristic (2)

- Erasable/Non-erasable
  - Erasable
    - The contents of the memory can be altered.
    - E.g. Semiconductor RAM
  - Non-erasable
    - The contents of the memory can not be altered, except by destroying the storage unit.
    - E.g. Semiconductor ROM

Summary (1)

- 知识点:Basic concepts of memory
  - Word
  - Word length
  - Address
  - Address space
  - Byte addressable memory
  - Byte ordering: big-endian, little-endian
  - Read and write operation of main memory
  - Physical types of memory: semiconductor, magnetic surface, optical

Summary (2)

- 知识点:Basic concepts of memory (ctd.)
  - Main memory capacity
  - Unit of transfer: internal memory, external memory
  - Access methods: sequential, random, direct access
  - Performance: memory access time, memory cycle time
  - Physical characteristic: volatile/nonvolatile, erasable/nonerasable

Exercise (1)

- 1. The memory is organized so that a group of n bits can be stored or retrieved in a single, basic operation. n is called the ________.
  - □A. word
  - □B. word length
  - □C. address
  - □D. cell

Exercise (2)

- 2. The 32-bit value 0x30A79847 is stored to the location 0x1000. If the system is little endian, the value of the byte ______ is stored in address 0x1002.

  - □ A. 0x30
  - □ B. 0xA7
  - □ C. 0x98
  - □ D. 0x47

Exercise (3)

- 3. In a main memory, its word size is 16, the number of word is 8K, what is the capacity of this main memory?
  - A. 16K×16
  - B. 16K×8
  - [✓ C. 8K×16]
  - D. 8K×8

Exercise (4)

- 4. For random access memory, the time it takes to perform a read or write operation is called ______.
  - □ A. memory cycle time
  - □ B. memory hit time
  - □ C. memory recovery time
  - □ D. memory access time

Exercise (5)

- Memory in which any location can be reached in a short and fixed amount of time after specifying its address is called ______.

□ A. direct access memory

□ B. sequential access memory

□ C. associative access memory

□ [✓ D. random access memory]

Exercise (5)

- If you turn off the power to the computer, items stored on_______device will be lost.

  - □ A. disk
  - □ B. DVD
  - □ [✓ C. RAM]
  - □ D. CD-ROM

Homework

- 补充题：

1.If a byte-addressable machine with 32-bit words stores the hex value 98765432 at address 0x00, indicate how this value would be stored on a little- endian machine and on a big-endian machine.