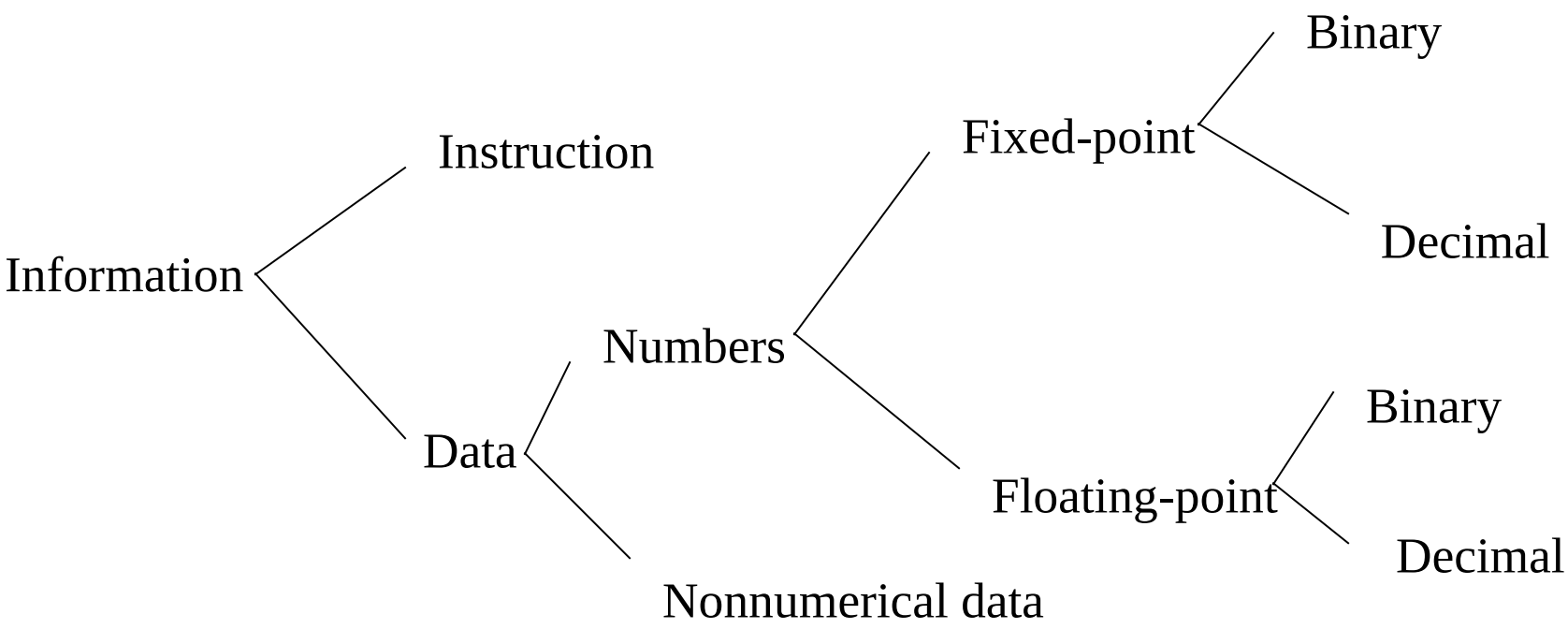
Lecture 2
Data Representation

CSE-2204: Computer Architecture and Organization

The Basic Information Types



Word Length

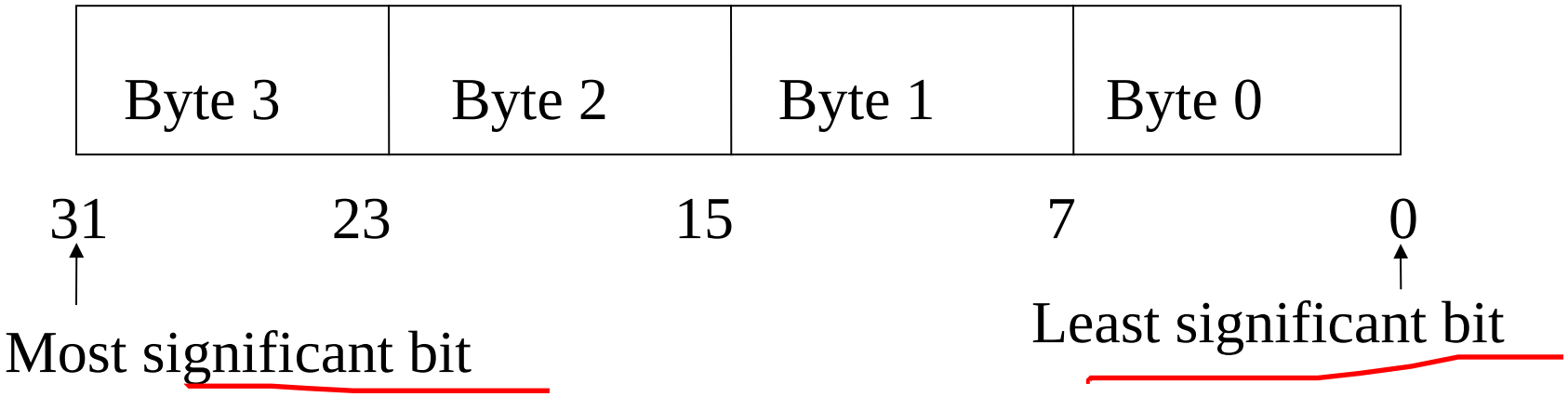
- ✓ A word is unit of information of some fixed length n.
- ✓ Word size is typically a multiple of 8, common CPU sizes being 8, 16, 32 and 64 bits.
- ✓ No single word length is suitable for representing every kind of information.

Word Length

Bits	Name	Description
1	Bit	Logic variable, flag
8	Byte	Smallest addressable memory item, Binary-coded decimal digit pair
16	Halfword	Short fixed-point number. Short address, Short instruction
32	Word	Fixed or floating point number, Memory address, Instruction.
64	Double word	Long instruction, double-precision floating point number.

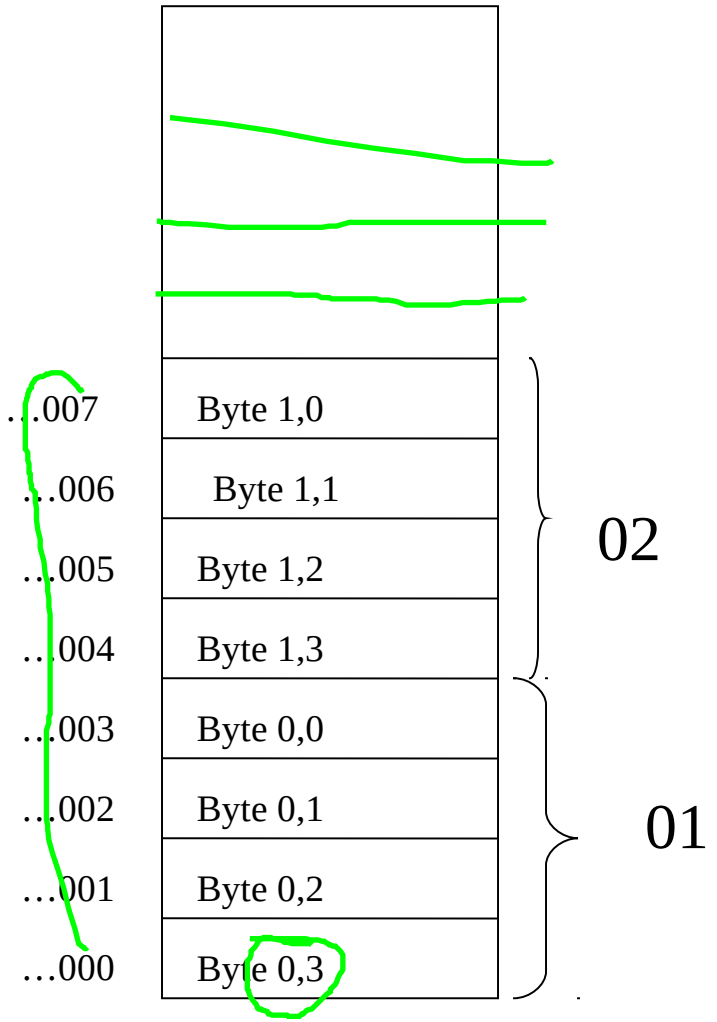
Information format of Motorola 680X0 microprocessor

Storage Order



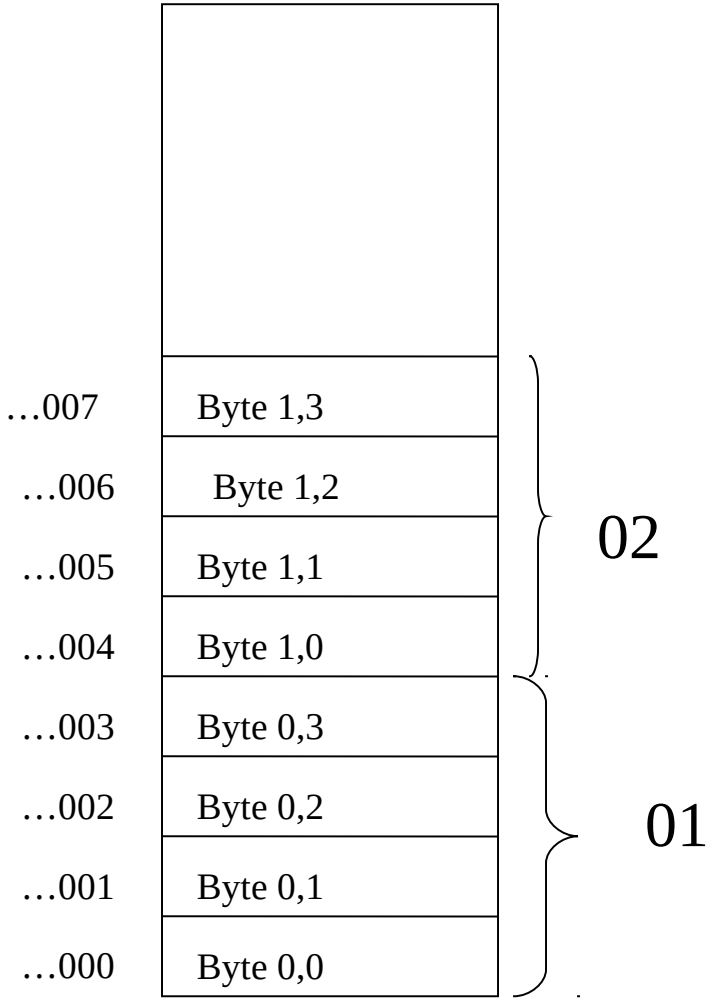
Indexing convention for the bits and bytes of a word

Big-Endian Storage Method



- ✓ Suppose a word W_i is represented as $B_{i,3}B_{i,2}B_{i,1}B_{i,0}$
- ✓ Most significant byte $B_{i,3}$ of W_i is assigned the lowest address and the least significant byte $B_{i,0}$ is assigned highest address.
- ✓ Assigns highest address to byte 0 of a word.

Little-Endian Storage Method



- ✓ Suppose a word W_i is represented as $B_{i,0}B_{i,1}B_{i,2}B_{i,3}$
- ✓ Most significant byte $B_{i,3}$ of W_i is assigned the highest address and the least significant byte $B_{i,0}$ is assigned lowest address.
- ✓ Assigns lowest address to byte 0 of a word.

Tags

- ✓ It is a technique of determining the type of a word.
- ✓ This is done by associating with each information word a group of bits, called a tag, that identifies the word's type.

Advantages:

- ✓ It simplifies instruction specifications.
- ✓ Tag inspection permits the hardware to check for software errors.

Disadvantages:

- ✓ They increase memory size.
- ✓ Add system hardware costs without increasing computing performance.

Factors in Selecting Number Representation

1. The number types to be represented: integers or real numbers.
2. The range of values.
3. The precision of the numbers (refers to maximum accuracy of the representation)
4. The cost of hardware required to store and process the numbers.

Fixed-point Binary Numbers

- ✓ The unsigned binary fixed-point format takes the form

  $b_N \ldots b_1 b_0 . b_{-1} b_{-2} \ldots b_M$ where each b_i is 0 or 1, representing the number $\sum b_i 2^i$ where $M <= i <= N$
- ✓ This is a positional notation in which each digit has a fixed-weight according to it's position relative to the binary point.
- ✓ The signed binary number is represented as

$$X_{n-1} \underbrace{X_{n-2} X_{n-3} \ldots X_2 X_1 X_0}$$

sign — magnitude

Signed Numbers

- ✓ 3 representations:
  1. Sign magnitude [+75=01001011, -75=11001011]
  2. 1's complement [+75= 01001011, -75 = 10110100]
  3. 2's complement [+75= 01001011, -75 = 10110101]

Signed Numbers

- ✓ Sign magnitude number employs the positional notation for magnitude and simply change the sign bit to represent + or -.
- ✓ In both complement code x_{n-1} retains the role as sign bit, but remaining bits no longer form a simple positional code when the number is negative.
- ✓ The advantage of the 2's complement code is that subtraction can be performed by logical complementation and addition only. 2's complement addition and subtraction can be implemented by a simple adder for unsigned numbers.
- ✓ Multiplication and division are more difficult to implement of 2's complement code.

Signed Number

Decimal Representation	Sign Magnitude	1's Complement	2's Complement
+7	0111	0111	0111
+6	0110	0110	0110
+5	0101	0101	0101
+4	0100	0100	0100
+3	0011	0011	0011
+2	0010	0010	0010
+1	0001	0001	0001
+0	0000	0000	0000
-0	1000	1111	0000
-1	1001	1110	1111
-2	1010	1101	1110
-3	1011	1100	1101
-4	1100	1011	1100
-5	1101	1010	1011
-6	1110	1001	1010
-7	1111	1000	1001

Decimal Codes

- ✓ Binary-decimal conversion is carried out by encoding each decimal digit separately by a sequence of bits.
- ✓ Some representations:

  1. BCD (Binary coded decimal) [971=1001 0111 0001]
  2. ASCII
  3. Excess-three code [formed by adding 0011 to the corresponding BCD number]
  4. Two-out-of-five

Decimal Codes

BCD Codes:

- ✓ In BCD format each digit d_i of a decimal number is denoted by a 4-bit equivalent $b_{i,3}b_{i,2}b_{i,1}b_{i,0}$.
- ✓ BCD is a weighted number code where each $b_{i,j}$ has the weight 10^i2^j.

ASCII:

- ✓ Represent 10 decimal digits by a 4-bit BCD field and the remaining 4-bits have no numerical significance.

Excess-three:

- ✓ 0011 is added to the BCD number.
- ✓ Addition can be performed same way as binary code.
- ✓ Non-weighted code.

Decimal Codes

Two-out-of five:

- ✓ Each decimal digit is represented by a 5-bit sequence containing two 1s and three 0s. There are exactly 10 distinct sequence of this type.
- ✓ It is error detecting code.
- ✓ It is non-weighted code and uses 5 rather than 4 bits per decimal digit.

Decimal Codes

Decimal Digit	BCD	ASCII	Excess-three	Two-out-of-five
0	0000	00110000	0011	11000
1	0001	00110001	0100	00011
2	0010	00110010	0101	00101
3	0011	00110011	0110	00110
4	0100	00110101	0111	01001
5	0101	00110101	1000	01010
6	0110	00110110	1001	01100
7	0111	00110111	1010	10001
8	1001	00111001	1100	10100

Disadvantages of Decimal Code

- ✓ Requires more memory space.
- ✓ The circuit required to perform arithmetic using decimal operands is more complex.

- **Chapter 3 section 3.2 of the Book of John P. Hayes**