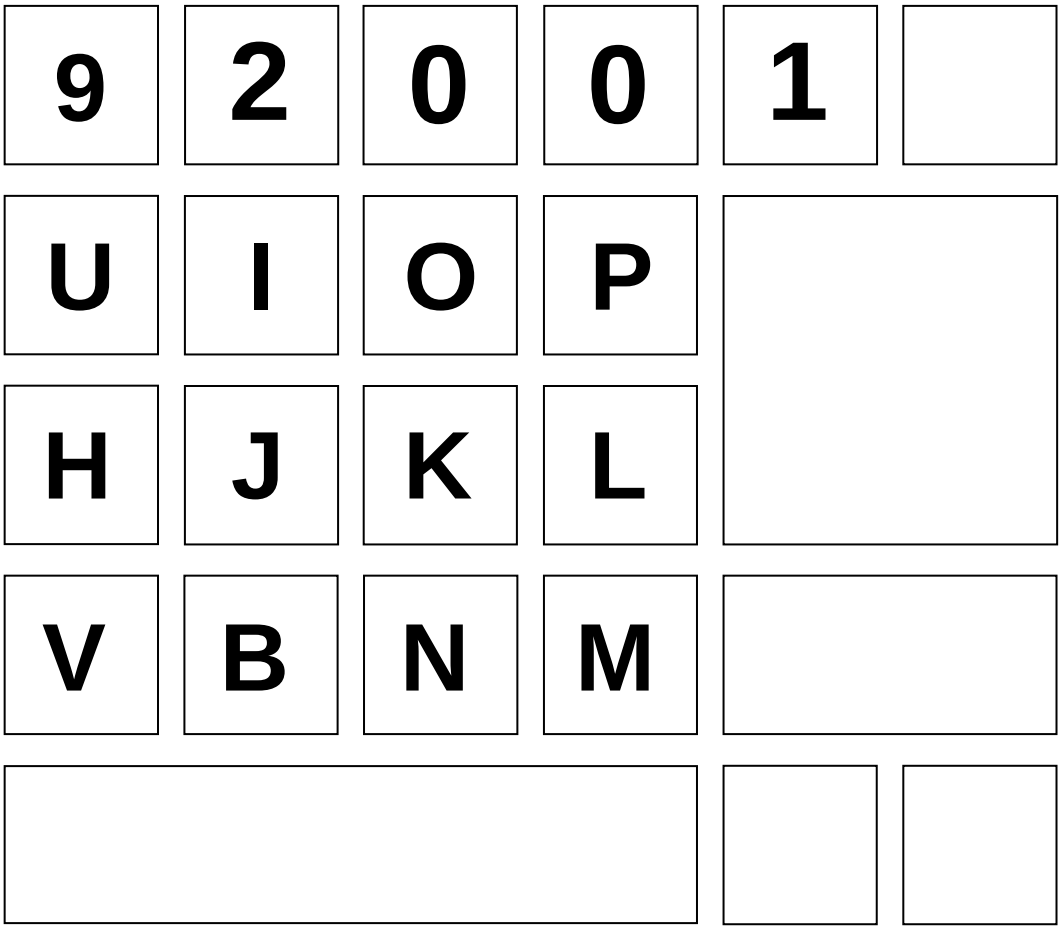
METI
Ministry of Economy,
Trade and Industry

Textbook for Fundamental Information Technology Engineers

NO. 1

INTRODUCTION TO COMPUTER SYSTEMS



Third Edition

REVISED AND UPDATED BY

Japan Information-Technology Engineers Examination Center
INFORMATION-TECHNOLOGY PROMOTION AGENCY,JAPAN

Contents

Part 1 COMPUTER SYSTEMS

1. Basic Theories of Information

Introduction **2**

1.1 Data representation **2**

1.1.1 Numeric conversion 2
1.1.2 Numeric representation 11
1.1.3 Operation and precision 22
1.1.4 Non-numeric value representation 23

1.2 Information and logic **26**

1.2.1 Proposition logic 26
1.2.2 Logical operation 26

Exercises **29**

2. Hardware

Introduction **33**

2.1 Information element **34**

2.1.1 Integrated circuit 34
2.1.2 Semiconductor memory 34

2.2 Processor architecture **36**

2.2.1 Processor structure and operation principles 36
2.2.2 Speed performance enhancement in processor 47
2.2.3 Operation mechanism 50
2.2.4 Multi-processor 54
2.2.5 Processor performance 55

2.3 Memory architecture **56**

2.3.1 Memory types 56
2.3.2 Memory capacity and performance 57
2.3.3 Memory configuration 58

2.4 Auxiliary storage devices 59

2.4.1 Types and characteristics of auxiliary storage devices 59
2.4.2 RAID types and characteristics 69

2.5 Input/output architecture and devices 71

2.5.1 Input/output control method 71
2.5.2 Input/output interfaces 73
2.5.3 Types and characteristics of input/output devices 76

2.6 Computer types 87

Exercises 91

3. Basic Software

Introduction 96

3.1 Operating system 96

3.1.1 OS configuration and functions 96
3.1.2 Job management 99
3.1.3 Process management 101
3.1.4 Main memory management 104
3.1.5 Virtual storage management 106
3.1.6 File management 108
3.1.7 Security management 112
3.1.8 Failure management 112
3.1.9 Supervisor 113

3.2 Types of OS 114

3.2.1 General-purpose OS 114
3.2.2 Network OS (NOS) 117

3.3 Middleware 118

3.3.1 DBMS 118
3.3.2 Communication management system 118
3.3.3 Software development support tool 119
3.3.4 Operation management tool 119
3.3.5 ORB 119

Exercises 120

4. Multimedia System

Introduction 125

4.1 What is multimedia? 125

4.1.1 Multimedia service 125
4.1.2 Platforms that implement the multimedia system 127
4.1.3 Multimedia technology 131

4.2 Multimedia applications 132

4.2.1 Voice and image pattern recognition 132
4.2.2 Synthesis of voice and image 133

4.3 Multimedia application system 134

Exercises 135

5. System Configurations

5.1 System classification and configurations 137

5.1.1 System classification 137
5.1.2 Client/server system 137
5.1.3 System configurations 140

5.2 System modes 144

5.2.1 System processing mode 144
5.2.2 System usage mode 146
5.2.3 System operating mode 150
5.2.4 Web computing 151

5.3 System Performance **152**

5.3.1 Performance calculation 152
5.3.2 Performance design 154
5.3.3 Performance evaluation 154

5.4 Reliability of the System **156**

5.4.1 Reliability calculation 156
5.4.2 Reliability design 159
5.4.3 Reliability objectives and evaluation 159
5.4.4 Financial costs 160

Exercises **162**

Answers to Exercises **166**

Answers for No.1 Chapter1 (Basic Theories of Information) **166**

Answers for No.1 Part1 Chapter2 (Hardware) **176**

Answers for No.1 Part1 Chapter3 (Basic Software) **184**

Answers for No.1 Part1 Chapter4 (Multimedia System) **193**

Answers for No.1 Part1 Chapter5 (System Configurations) **196**

Part 2 INFORMATION PROCESSING AND SECURITY

1. Accounting

1.1 Business Activities and Accounting Information 206

1.1.1 Fiscal Year and Accounting Information 206
1.1.2 The Accounting Structure 209

1.2 How to Read Financial Statements 214

1.2.1 How to Read the Balance Sheet 214
1.2.2 How to Read the Income Statement 221

1.3 Financial Accounting and Management Accounting 228

1.3.1 Financial Accounting 228
1.3.2 Management Accounting 229
1.3.3 Accounting Information System Configuration 236
1.3.4 International Standards 237

Exercises 246

2. Application Fields of Computer Systems

2.1 Engineering Applications 252

2.1.1 Automatic Control of Production 252
2.1.2 CAD/CAM/CAE 253
2.1.3 FA Systems and CIM 254

2.2 Business Applications 256

2.2.1 Head Quarters Business Support Systems 256
2.2.2 Retail Business Support Systems 257
2.2.3 Financial Systems 261

2.2.4 Inter-Enterprise Transaction Data Interchange 263

Exercises 266

3. Security

3.1 Information Security 269

3.1.1 What Is Information Security? 269

3.1.2 Physical Security 269

3.1.3 Logical Security 272

3.2 Risk Analysis 273

3.2.1 Risk Management 273

3.2.2 Types, Evaluation, and Analysis of Risks 273

3.2.3 Risk Processing Methods 277

3.2.4 Security Measures 277

3.2.5 Data Protection 277

3.2.6 Protection of Privacy 278

Exercises 280

4. Operations Research

4.1 Operations Research 283

4.1.1 Probabilities and Statistics 283

4.1.2 Linear Programming 296

4.1.3 Scheduling 300

4.1.4 Queuing Theory 310

4.1.5 Inventory Control 315

4.1.6 Demand Forecasting 325

Exercises 336

Answers to Exercises **344**

Answers for No.1 Part2 Chapter1 (Accounting) **344**

Answers for No.1 Part2 Chapter2 (Application Fields of Computer Systems) **355**

Answers for No.1 Part2 Chapter3 (Security) **361**

Answers for No.1 Part2 Chapter 4 (Operations Research) **368**

Index **382**

Photos provided by:

I-O Data Device, Inc.
Intel Corporation
Uchida Yoko Co., Ltd.
Canon Sales Co., Inc.
Sharp Corp.
Sumitomo Electric Industries, Ltd.
Seiko Instruments Inc.
Sekonic Co., Ltd.
Sony Corp.
IBM Japan, Ltd.
NEC Corp.
HAL Corporation
Fujitsu Ltd.
Microsoft Co., Ltd.
Ricoh Co., Ltd.

Textbook for Fundamental Information Technology Engineers

No. 1 INTRODUCTION TO COMPUTER SYSTEMS

First edition first printed September 1, 2001
Second edition first printed August 1, 2002
Third edition first printed August 1, 2003

Japan Information-Technology Engineers Examination Center
INFORMATION-TECHNOLOGY PROMOTION AGENCY, JAPAN
Center Office 15F, Bunkyo Green Court 2-28-8, Hon-Komagome, Bunkyo-Ku,
Tokyo 113-8663 , JAPAN

Part 1

COMPUTER SYSTEMS

Introduction

This series of textbooks has been developed based on the Information Technology Engineers Skill Standards made public in July 2000. The following four volumes cover the whole contents of fundamental knowledge and skills required for development, operation and maintenance of information systems:

No. 1: Introduction to Computer Systems
No. 2: System Development and Operations
No. 3: Internal Design and Programming--Practical and Core Bodies of Knowledge--
No. 4: Network and Database Technologies
No. 5: Current IT Topics

This part gives easy explanations systematically so that those who are learning computer systems for the first time can easily acquire knowledge in these fields. This part consists of the following chapters:

Part 1: Computer Systems
Chapter 1: Basic Theories of Information
Chapter 2: Hardware
Chapter 3: Basic Software
Chapter 4: Multimedia System
Chapter 5: System Configurations

1 Basic Theories of Information

Chapter Objectives

Understanding the computer mechanism of data representation and basic theories.
In particular, the binary system is an important subject to learn, indispensable for computer data representation. However, for people who are used to the decimal system, it is hard to become familiar with this representation, so it should be carefully studied.

① Understanding a computer's basic data units such as binary numbers, bits, bytes, words, etc. and their conversion from and to decimal and hexadecimal digits.
② Understanding basic concepts of computer internal data representation, focusing on numeric data, character codes, etc.
③ Understanding proposition calculus and logical operations.

Introduction

In order to make a computer work, it is necessary to convert the information we use in daily life into a format that can be understood by the computer. For that reason, the way information is actually represented inside a computer as well as the way it is processed will be learned here.

1.1 Data representation

1.1.1 Numeric conversion

For a computer to do processing it is first necessary to input into the memory the programs which are contents of a task or processing procedures. The binary system is what is used to represent this information. While the binary system represents information by means of the combination of "0" and "1," we ordinarily use the decimal system. Therefore, an important fundamental knowledge required by information processing engineers is to understand the relationship between binary and decimal numbers. This is the basic difference between computers and human beings as well as the point of contact between them.
Since the mechanism in which the computer operates is completely based on binary numbers, the relationship between binary and decimal numbers, as well as hexadecimal numbers combining binary numbers will be explained here.

(1) Data representation unit and processing unit

① Binary numbers

The internal structure of a computer is composed of an enormous number of electronic circuits. Binary numbers represent two levels of status in the electronic circuits, as in:

- Whether the electric current passes through it or not
- Whether the voltage is high or low

For example, setting the status where the electric current passes through (the power is on) to "1" and the status where the electric current does not pass through (the power is off) to "0," then by replacing the computer status or data with numerical values their representation can be easily performed in an extremely convenient way.
The representation of decimal numbers from "0" to "10" using binary numbers is shown in Figure 1-1-1.

Figure 1-1-1
Decimal numbers and binary numbers

	Decimal numbers	Binary numbers	
	0	0	
	1	1	
	2	10	A carry occurs
	3	11	
	4	100	A carry occurs
	5	101	
	6	110	
	7	111	
	8	1000	A carry occurs
	9	1001	
A carry occurs	10	1010	

As can be seen in this Figure, compared to the decimal system, a carry occurs more frequently in the binary system, but since besides "0" and "1," no other figure is used, it is the most powerful tool for the computer.

② Bits

A bit (binary digit) is 1 digit of the binary system represented by "0" or "1." A bit is the smallest unit that represents data inside the computer. 1 bit can represent only 2 values of data, "0" or "1," but 2 bits can represent 4 different values:

- 00
- 01
- 10
- 11

However, in practice, the amount of information processed by a computer is so immense (there are 26 values in the English alphabet alone) that the two bits, 0 and 1, are insufficient for an information representation method.

③ Bytes

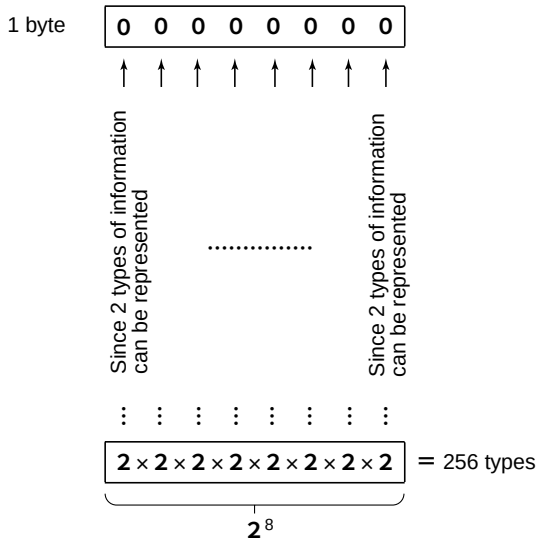
Compared to a bit, which is the smallest unit that represents data inside the computer, a byte is a unit that represents with 8 bits 1 character or number. Considering that a byte is equal to 8 bits, the following is the information which can be represented with one byte, by the combination of "0" and "1."

- 00000000
- 00000001
- 00000010

  →
- 11111101
- 11111110
- 11111111

The information represented by a string of "1's" and "0's" is called a bit pattern. Since 1 bit can be represented in two ways, the combination of 8 bit patterns into 1 byte enables the representation of 2^8=256 types of information. In other words, besides characters and figures, symbols such as "+" and "-" or special symbols such as "<" and ">" can also be represented with one byte.

Figure 1-1-2
Types of information that can be represented with one byte



However, since the number of kanji (Chinese characters) amounts to thousands, they cannot be represented with one byte. Therefore, two bytes are connected to get 16 bits, and one kanji is represented with two bytes. With 16 bits, 216 = 65,536 kanji can be represented.

④ Words

A bit is the smallest unit that represents data inside a computer and a byte is a unit that represents 1 character. However, if the computers' internal operations were performed on the bit basis, the operation speed would be too low. For that reason the idea of processing using a unit called word was born.

Over 10 years ago, personal computers operated on words each consisting of 16 bits. Currently mainstream PGs use words each consisting of 32 bits.

⑤ Binary system and hexadecimal system

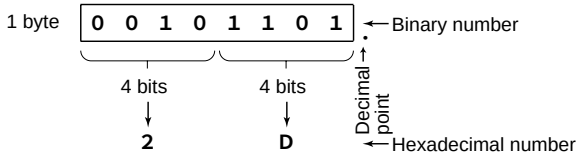
In information processing, the binary system is used to simplify the structure of the electronic circuits that make up a computer. However, for us, the meaning of string of "0's" and "1's" is difficult to understand. In the decimal system, the numeric value "255" has 3 digits, but in the binary system the number of digits becomes 8. Therefore, in order to solve the problem of difficulty in identification and of a large number of digits hexadecimal system is used.
A hexadecimal number is a numeric value represented by 16 numerals, from "0" to "15." When it becomes 16, a carry occurs. However, since it cannot distinguish between the "10" before a carry has been generated, and the "10" after a carry has been generated, for purposes of convenience, in the hexadecimal system "10" is represented by the letter "A," "11" by "B," "12" by "C," "13" by "D," "14" by "E" and "15" by "F."

Figure 1-1-3
Decimal numbers, binary numbers, and hexadecimal numbers

Decimal numbers	Binary numbers	Hexa-decimal numbers
0	0	**0**
1	1	**1**
2	10	**2**
3	11	**3**
4	100	**4**
5	101	**5**
6	110	**6**
7	111	**7**
8	1000	**8**
9	1001	**9**
10	1010	**A**
11	1011	**B**
12	1100	**C**
13	1101	**D**
14	1110	**E**
15	1111	**F**
16	10000	10
17	10001	11
18	10010	12
19	10011	13
20	10100	14

Figure 1-1-3 shows the notation of the numbers "0" to "20" of the decimal system in the binary system and the hexadecimal system.
Focusing on the relationship between the hexadecimal numbers and binary numbers in this table, it can be noted that 4 digits in the binary system correspond to 1 digit in the hexadecimal system. Therefore, binary numbers can be converted to hexadecimal numbers, by replacing each group of 4 bits with a hexadecimal digit, starting from the decimal point. (Figure 1-1-4)

Figure 1-1-4
Binary, and hexadecimal counting systems



(2) Representation of numeric data

By means of the combinations of "0's" and "1's," characters are represented as codes. However, a different representation method is used to process numeric data. Here, the radix and radix conversion, the addition and subtraction of binary numbers and hexadecimal numbers, the representation of negative numbers, among other points considered basic for the representation of numeric data, will be explained.

① Radix and "weight"

a. Decimal numbers' "weight" and its meaning

When quantities are represented using decimal numbers, 10 types of numerals from "0" to "9" are combined. Each of them, from the lower digit in the ascendant order has a "weight" as 10^0, 10^1, 10^2, 10^3... (Figure 1-1-5).
For example, using the weight, a decimal number 1,234 would be represented as follows:

$$1{,}234 = 1 \times 10^3 + 2 \times 10^2 + 3 \times 10^1 + 4 \times 10^0$$

Figure 1-1-5
Weight of each digit of the decimal number 21998

2	1	9	9	8	Decimal number
↑ Ten thousand	↑ Thousand	↑ Hundred	↑ Ten	↑ Unit	Name of each digit
10^4	10^3	10^2	10^1	10^0	Weight of each digit

In Figure 1-1-5 the weight of each digit is represented as 10^0, 10^1, 10^2, 10^3,... this "10" is called "Radix" and the value placed at the upper right of 10 is called the "Exponent." The notation and meaning of the weight in the decimal system is explained below.
In 10^0, the radix 10 is multiplied 0 times by 1, so it becomes 1, in 10^1, the radix 10 is multiplied 0 times by itself, so it becomes 10.
Likewise, in 10^2, 10 is multiplied 2 times by itself, so it becomes 100; in 10^3, 10 is multiplied 3 times by itself, so it becomes 1,000.
In this way, even when the number of digits increases, it can be easily represented by writing down in small numbers, to the upper right of 10, the numeric value that indicates the number of times the radix 10 is multiplied (exponent).

b. Binary digits "weight" and its meaning

The radix of the decimal system is 10, and the radix of the binary system is 2. As in the decimal system, the weight of each digit in the binary system is shown in Figure 1-1-6.

Figure 1-1-6
Weight of each digit of the binary number 11111001110

1	1	1	1	1	0	0	1	1	1	0	Binary number
↑ 2^{10}	↑ 2^9	↑ 2^8	↑ 2^7	↑ 2^6	↑ 2^5	↑ 2^4	↑ 2^3	↑ 2^2	↑ 2^1	↑ 2^0	Weight of each digit

The notation and meaning of the weight in the binary system is explained below.
In 2^0, the radix 2 is multiplied 0 times by itself, so it becomes 1, in 2^1, the radix 2 is multiplied only 1 time by itself, so it becomes 2. Likewise, in 2^2, 2 is multiplied 2 times by itself, so it becomes 4.
To verify that the decimal number 1,988 is represented as "11111001110" in the binary system, the weight of each of the digits represented by 1 in the binary representation should be added, as is shown below:

$$\begin{array}{ccccccccccccccccccccc}
1 & & 1 & & 1 & & 1 & & 1 & & 0 & & 0 & & 1 & & 1 & & 1 & & 0 \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & & & & & \downarrow & & \downarrow & & \downarrow & & \\
2^{10} & + & 2^9 & + & 2^8 & + & 2^7 & + & 2^6 & & & & & + & 2^3 & + & 2^2 & + & 2^1 & & \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & & & & & \downarrow & & \downarrow & & \downarrow & & \\
= 1{,}024 & + & 512 & + & 256 & + & 128 & + & 64 & & & & & + & 8 & + & 4 & + & 2 & & \\
= 1{,}998 & & & & & & & & & & & & & & & & & & & &
\end{array}$$

② Auxiliary units and power representation

Since the amount of information processed by computers is immense, auxiliary units that represent big amounts are also used.
Likewise, since computers operate at high speeds, auxiliary units that represent extremely small amounts are also needed to represent the performance.
Figure 1-1-7 shows the auxiliary units that represent large and small amounts as well as the exponent to which the radix is raised.

Figure 1-1-7
Auxiliary units

	Unit symbol	Exponent notation	Remarks
Units that represent large amounts	T (giga)	10^{12}	2^{40}
	G (tera)	10^{9}	2^{30}
	M (mega)	10^{6}	2^{20}
	k (kilo)	10^{3}	2^{10}
Units that represent small amounts	m (milli)	10^{-3}	$\frac{1}{1,000}$
	μ (micro)	10^{-6}	$\frac{1}{1,000,000}$
	n (nano)	10^{-9}	$\frac{1}{1,000,000,000}$
	p (pico)	10^{-12}	$\frac{1}{1,000,000,000,000}$

It is important to note that, as indicated in the Remarks column in Figure 1-1-7, kilo is equal to 10^3, but it is also almost equal to 2^{10}. In other words, the kilo we ordinarily use is equal to 1,000, however, since the binary system is used in computing, 2^{10} (1,024) is a kilo. Furthermore, if 2^{10} and 10^3 are almost equal, 10^6 that is a mega, is almost equal to 2^{20} and 10^9 a giga, is almost equal to 2^{30}.
Therefore, when it is said that a computer memory capacity is 1 kilobyte, strictly speaking, 1 kilobyte does not mean 1,000 bytes, but 1,024 bytes.

③ Addition and subtraction of binary numbers

a. Addition

The following are the 4 basic additions of the binary system:

- 0 + 0 = 0 (0 in the decimal system)
- 0 + 1 = 1 (1 in the decimal system)
- 1 + 0 = 1 (1 in the decimal system)
- 1 + 1 = 10 (2 in the decimal system) ← Main characteristic of the binary system that differs from the decimal system

Among these additions, a carry is generated in 1 + 1 = 10.

```
  [1]   ← Carry
    1
+   1
-----
  1 0
```

Example $(11010)_2 + (1100)_2$

```
  [1][1]   ← Carry
    1 1 0 1 0
+       1 1 0 0
---------------
  1 0 0 1 1 0
```

The result is $(100110)_2$.

b. Subtraction

The following are the 4 basic subtractions of the binary system:

- 0 – 0 = 0
- 0 – 1 = –1
- 1 – 0 = 1
- 1 – 1 = 0

Among these subtractions, if the upper digit of 0 is 1 in 0 – 1 = –1, a "borrow" is performed.

```
  ♥     ← Borrow
  1 0
–   1
-----
    1
```

Example $(10011)_2$ - $(1001)_2$

♥ ← Borrow

```
  1 0 0 1 1
– 1 0 0 1
  1 0 1 0
```

The result is $(1010)_2$.

④ Addition and subtraction of hexadecimal numbers

Basically, the addition and subtraction of hexadecimal numbers is similar to that of decimal and binary numbers.

a. Addition

Addition is performed starting at the lowest (first from the left) digit. When the addition result is higher than 16, a carry to the upper digit is performed.

Example $(A8D)_{16}$ + $(B17)_{16}$

```
  1  1   ← Carry
    A 8 D
+   B 1 7
  1 5 A 4
```

```
    10   8   13
+   11   1    7
    21   9   20
```

- First digit: D + 7 = (In the decimal system: 13 + 7 = 20) = 16 (carried 1) + 4
  The sum of the first column is 4 and 1 is carried to the second column.
- Second digit: 1 + 8 + 1 = (In the decimal system: 10) = A
  ↑ Carried from the first column
- Third digit: A + B = (In the decimal system: 10 + 11 = 21) = 16 (carried 1) + 5
  The sum of the third column is 5 and 1 is carried to the fourth column.

The result is $(15A4)_{16}$.

b. Subtraction

Subtraction is performed starting from the first column, and when the subtraction result is negative (minus), a borrow from the upper order column is performed.

Example $(6D3)_{16} - (174)_{16}$

♥ ← Borrow

```
    6 D 3
–   1 7 4
    5 5 F
```

♥ 16

```
    6   13   3
–   1    7   4
    5    5  15
```

- First digit: Since 3 – 4 = –1, a borrow is performed from D in the second digit (D becomes C). 16 (borrowed 1) + 3 – 4 = F (In the decimal system: 19 – 4 = 15)
- Second digit: C – 7 = 5 (In the decimal system: 12 – 7 = 5)
- Third digit: 6 – 1 = 5

The result is $(55F)_{16}$.

(3) Radix conversion

In order to process numeric values in a computer, decimal numbers are converted into binary or hexadecimal numbers. However, since we ordinarily use decimal numbers, it would be difficult to understand the meaning of the result of a process if it were represented by binary or hexadecimal numbers.

Therefore, the conversion amongst decimal, binary and hexadecimal numbers is necessary. This operation is called radix conversion.
A concrete explanation of the conversion amongst the radixes of decimal, binary and hexadecimal numbers, which are currently used the most, will be performed below. In order to avoid confusion, the respective radix will be written outside the parenthesis to distinguish them. For example:

- Notation of binary numbers: $(0101)_2$
- Notation of decimal numbers: $(123)_{10}$
- Notation of hexadecimal numbers: $(1A)_{16}$

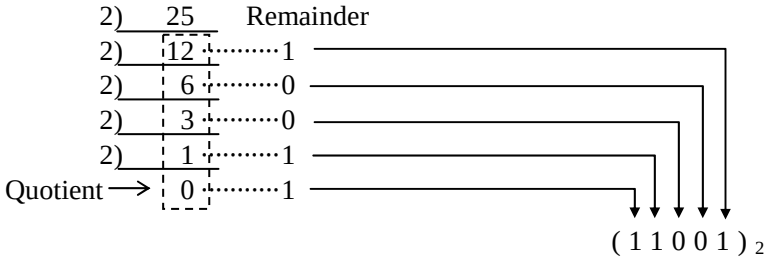
① Conversion of decimal numbers into binary numbers

The method of conversion for translating decimal numbers into binary numbers differs depending on whether the decimal number is an integer or a fraction.

a. Conversion of decimal numbers

The decimal integer is divided into 2, and the quotient and remainder are obtained. The resulting quotient is divided into 2 again, and the quotient and remainder are obtained. This operation is repeated until the quotient becomes 0.
Since a decimal integer is divided into 2, when the decimal integer is an even number the remainder will be "0," when it is an odd number the remainder will be "1." The binary digit is obtained by placing the remainder(s) in the reverse order.

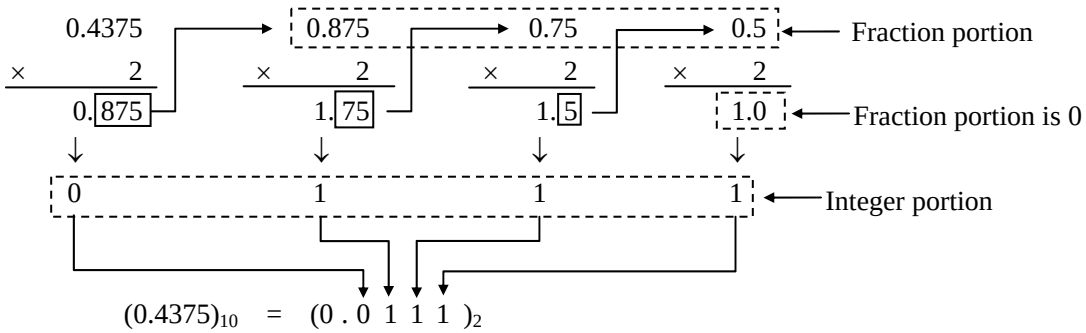
Example $(25)_{10}$



b. Conversion of decimal fractions

The decimal fraction is multiplied by 2, the integer and fraction portion of the product are separated, and the integer section is extracted. Since the integer portion is the product of the multiplication of the fraction portion by 2, it will always be "0" or "1." Next, setting aside the integer portion, only the fraction portion is multiplied by 2. This operation is repeated until the fraction portion becomes 0. The binary digit is obtained by placing the integer portions extracted in the order they were extracted.

Example $(0.4375)_{10}$



It should be noted that when decimal fractions are converted into binary fractions, most of the times, the conversion is not finished, since no matter how many times the fraction portion is multiplied by 2, it will not become 0. In other words, the above-mentioned example is that of a special decimal fraction, but most of the decimal fractions become infinite binary fractions.
The verification of the kind of numeric values which correspond to special decimal fractions is

performed below. For example, the result of the conversion of the binary fraction 0.11111 into a decimal fraction is as follows:

0.	1	1	1	1	1	← Binary fractions
	↓	↓	↓	↓	↓	
	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}	← Weight
	↓	↓	↓	↓	↓	
	0.5 +	0.25 +	0.125 +	0.0625 +	0.03125 = 0.96875	← Decimal fractions

From this example it can be understood that besides the decimal fractions that are equal to the weight of each digit (0.5, 0.25, 0.125, ...etc.) or the decimal fractions that result from their combination, all other decimal fractions become infinite binary fractions.

② Conversion of binary numbers into decimal numbers

The conversion into decimal numbers is performed by adding up the weights of each of the "1" digits of the binary bit string.

a. Conversion of binary integers

Example $(11011)_2$

$(1\ 1\ 0\ 1\ 1)_2$

$2^4 + 2^3 + 2^1 + 2^0$ ← Weight

$16 + 8 + 2 + 1 = (27)_{10}$

b. Conversion of binary fractions

Example $(1.101)_2$

$(1\ .\ 1\ 0\ 1)_2$

$2^0 + 2^{-1} + 2^{-3}$ ← Weight

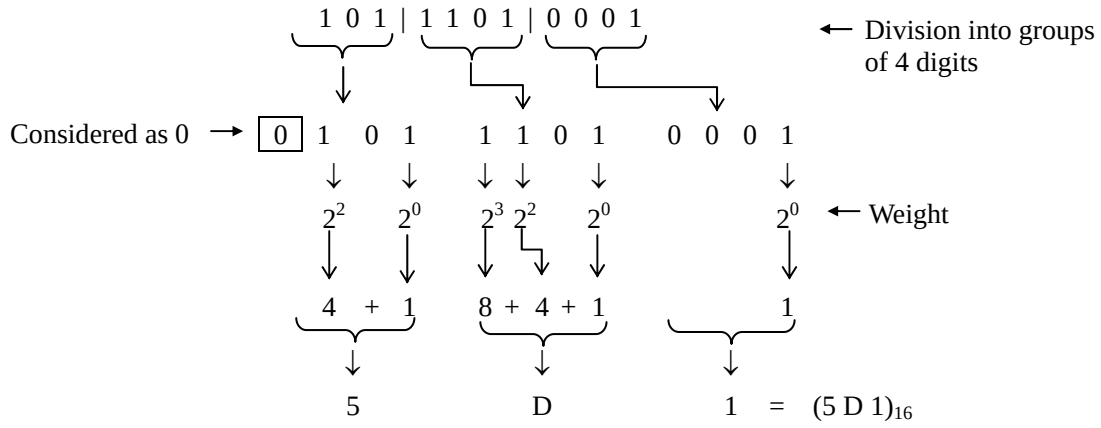
$1 + 0.5 + 0.125 = (1.625)_{10}$

③ Conversion of binary numbers into hexadecimal numbers

Since 4-bit binary strings are equivalent to 1 hexadecimal digit, in binary integers, the binary number is divided into groups of 4 digits starting from the least significant digit. In binary fractions, the binary number is divided into groups of 4 digits starting from the decimal point. Then, the conversion is performed by adding up the weights of each of the binary digits displayed as "1," in each group of 4 bits. In the event that there is a bit string with less than 4 digits, the necessary number of "0's" is added and the string is considered as a 4-bit string.

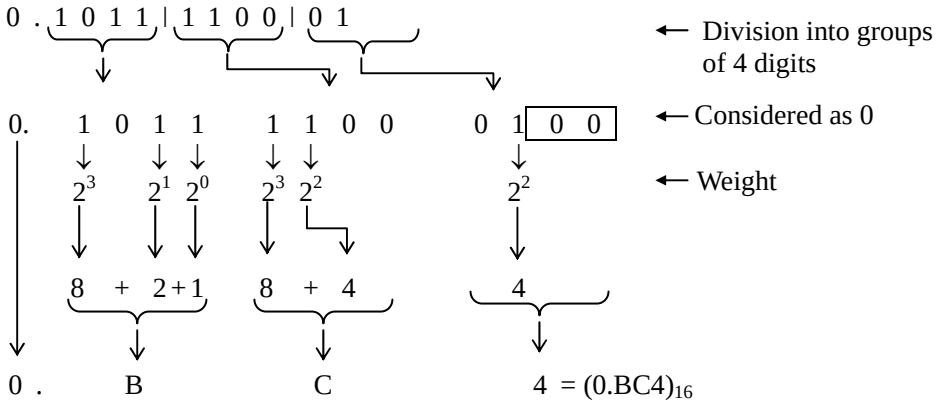
a. Conversion of binary integers

Example $(10111010001)_2$



b. Conversion of binary fractions

Example $(0.1011110001)_2$

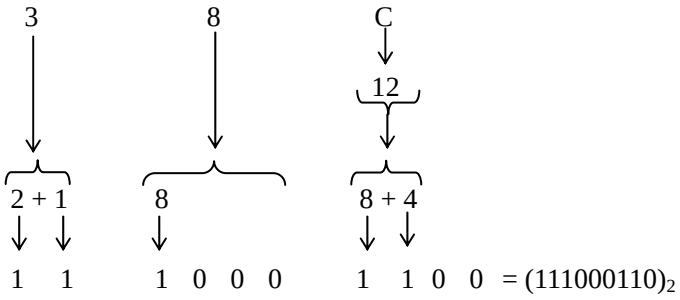


④ Conversion of hexadecimal numbers into binary numbers

Hexadecimal numbers are converted into binary numbers by performing the reverse procedure. In other words, 1 digit of the hexadecimal number is represented with a 4-digit binary number.

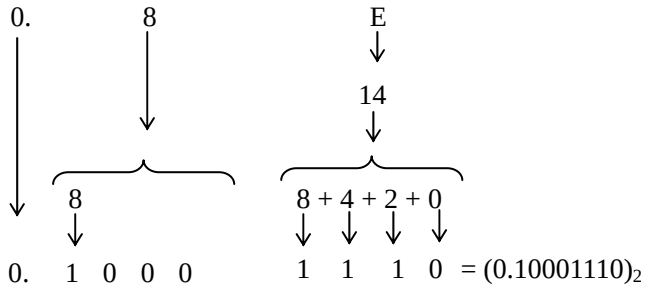
a. Conversion of hexadecimal integers

Example $(38C)_{16}$



b. Conversion of hexadecimal fractions

Example $(0.8E)_{16}$



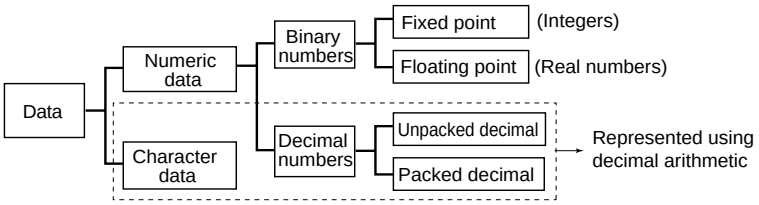
⑤ Conversion from decimal numbers into hexadecimal numbers and from hexadecimal numbers into decimal numbers

To convert them into binary numbers, decimal numbers are divided into 2, to convert them into hexadecimal numbers, and then they are divided into 16. Likewise, hexadecimal numbers are converted into decimal numbers by adding up the exponents whose radixes are 16.
It should be noted that due to the general unfamiliarity with the notation of hexadecimal numbers, ordinarily hexadecimal numbers are first converted into binary numbers to convert them into decimal numbers.

1.1.2 Numeric representation

In the computer, originally invented as a calculating machine, amongst other aspects involving the management of the data subject for processing, the precision and the easiness with which calculations can be performed have also been worked out. The representation format suitable for each type of data is explained here.

Figure 1-1-8
Data representation format



(1) Decimal digit representation

① Binary-coded decimal code

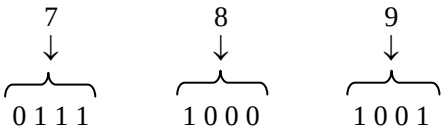
As a format of character data and decimal numbers, there is a representation method called binary-coded decimal code (BCD code) that, using 4-bit binary digits that correspond to the numbers 0 to 9 of the decimal system, represents the numeric value of each digit.

Figure 1-1-9 Binary-coded decimal code

Decimal number	Binary number	Binary-coded decimal code
0	0 0 0 0	0 0 0 0
1	0 0 0 1	0 0 0 1
2	0 0 1 0	0 0 1 0
3	0 0 1 1	0 0 1 1
4	0 1 0 0	0 1 0 0
5	0 1 0 1	0 1 0 1
6	0 1 1 0	0 1 1 0
7	0 1 1 1	0 1 1 1
8	1 0 0 0	1 0 0 0
9	1 0 0 1	1 0 0 1
10	1 0 1 0	0 0 0 1 0 0 0 0
11	1 0 1 1	0 0 0 1 0 0 0 1
⋮	⋮	⋮

Since the binary-coded decimal code is not a numeric value but a code, there are only 10 patterns, and the notation is performed by arranging the patterns for each digit.

For example, the representation of the decimal number "789" using the binary-coded decimal code would be as follows:



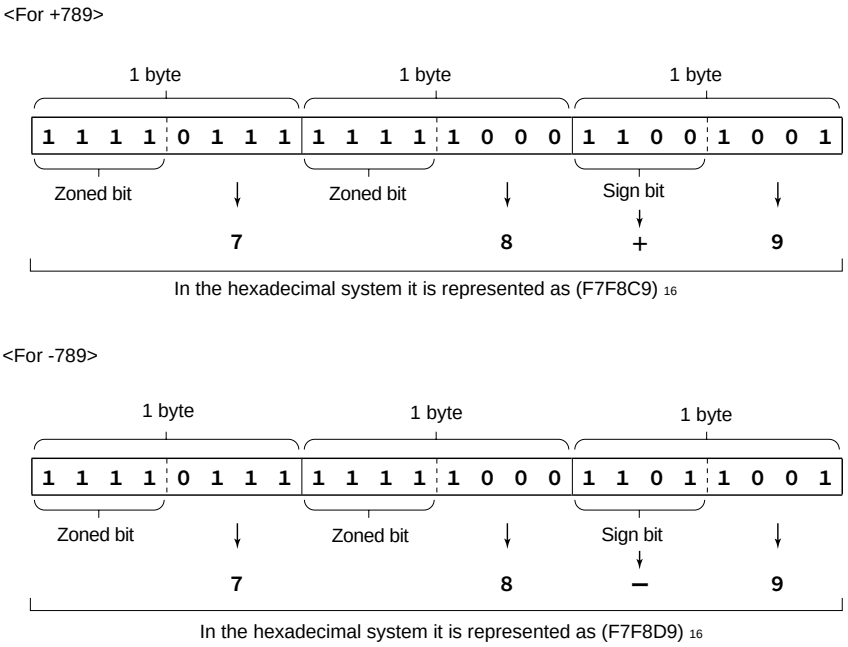
$(011110001001)_2$

In this way, as the number of digits of a decimal number increases, the length of the binary-coded decimal code increases as well (a group of 4 bits is added for each digit). This format is called variable-length format.
The same value as the binary-coded decimal code has also been set to the least significant 4 bits of the numeric characters of the EBCDIC, JISCII and other codes.
The binary-coded decimal code is mainly used for the numeric representation of office calculations, and according to the memory format of the computer, it is divided into unpacked decimal format and packed decimal format. And, since character codes as well as the unpacked decimal format or packed decimal format are represented by the use of the binary-coded decimal code, they are automatically processed using the decimal arithmetic system of the computer. It is not necessary for the user to be aware of this process.

② Unpacked decimal format

When representing signed decimals, the unpacked decimal format uses 1 byte for each digit of the decimal number.
The unpacked decimal format represents the values from 0 to 9 in the least significant 4 bits of 1 byte, and in the most significant 4 bits, which are called zoned bits, in the case of the EBCDIC code used in high-end mainframe machines, where ordinarily $(1111)_2$ is stored. However, in the zoned bits of the least significant digit, the 4 bits that represent the sign are stored, in both the case of 0 and positive numbers, $(1100)_2$, and in the case of negative numbers, $(1101)_2$. In the JIS code used for data transmission as well as in the low-end machines, $(0011)_2$ is stored in the zoned bits. The unpacked decimal format is also called zoned decimal format.
The bit pattern of the representation of the decimal numbers +789 and –789 in the unpacked decimal format is shown in Figure 1-1-10.

Figure 1-1-10 Unpacked decimal format

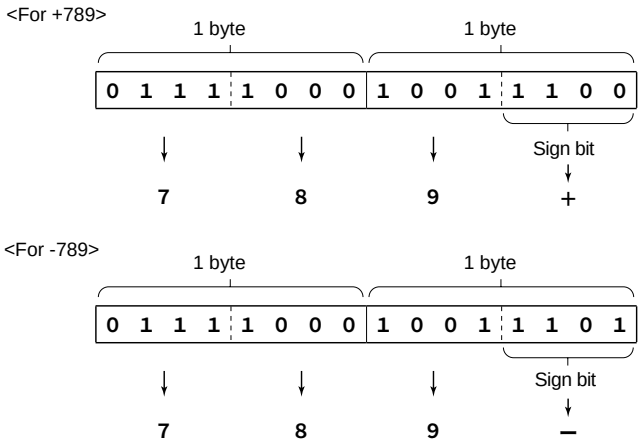


In the unpacked decimal format, excepting the least significant byte, only half of a byte is used. This was considered a waste of resources. This defect was eliminated by the packed decimal format.

③ Packed decimal format

In the packed decimal format, 1 byte represents a numeric value of 2 digits and the least significant 4 bits represent the sign. The bit pattern of the sign is the same as that of the unpacked decimal format, $(1100)_2$ for 0 and positive numbers, and $(1101)_2$ for negative numbers.
Figure 1-1-11 shows the bit pattern of the packed decimal format.

Figure 1-1-11 Packed decimal format



Compared to the unpacked decimal format, the packed decimal format has the following advantages:

- A numeric value can be represented by fewer bytes.
- The conversion into the binary system is easy.

(2) Binary representation

① Representation of negative integers

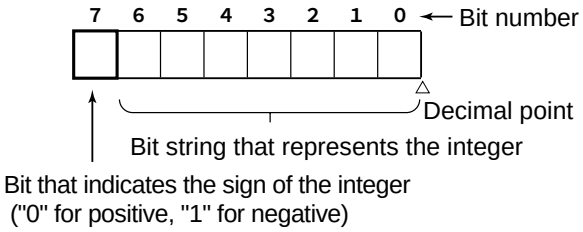
The typical example of methods to represent negative integers are mentioned below:

- Absolute value representation
- Complement representation

a. Absolute value representation of negative integers

As is shown in Figure 1-1-12, in the absolute value representation of negative integers, the first bit represents the sign and the 7 other bits represent the numeric value (absolute value).

Figure 1-1-12
Absolute value representation of negative integers



For example, in $(00001100)_2$, since the sign bit at the top is 0, it is a positive number. Likewise, since the 7 other bits, which are the absolute value of the numeric value, are $(0001100)_2 = 2^2 + 2^3 = (12)_{10}$, the decimal number 12 (positive number) is represented.
On the other hand, since in $(10001100)_2$ the sign bit at the top is 1, it is a negative number. The decimal number –12 = (negative number) is represented.
However, since in this representation method the numeric value 0 can be represented in two ways, as 00000000 (positive zero) or as 10000000 (negative zero), the operation becomes complicated, and for that reason it is not widely used.
It should be noted that when the absolute value representation of negative numbers is performed with 8 bits, the range of numeric values that can be represented is as follows (in decimal digits):

–127 to 127

b. Complement representation of negative integers

The complement is the number that indicates the quantity by which a numeric value falls short of a specific numeric value. There are 2 types of radix complements, the radix complement and the reduced radix complement.

● Decimal complement

There are 2 types of decimal complements, "10's complement" and "9's complement." For example, the 9's complement of a given numeric value will be the result of the subtraction of each of the digits of this numeric value from 9. Likewise, the 10's complement of a given numeric value will be the result of the subtraction of each of the digits of this numeric value from 10. As a result, the 10's complement is the result of the addition of 1 to the 9's complement.

Example "9's complement" of $(123)_{10}$

$$\begin{array}{r} 9\,9\,9 \\ -\ 1\,2\,3 \\ \hline 8\,7\,6 \end{array}$$

Example "10's complement" of $(123)_{10}$

$$\begin{array}{rl} 1\,0\,0\,0 & (= 999 + 1) \\ -\ 1\,2\,3 & \\ \hline 8\,7\,7 & \end{array}$$

● Binary complement

There are 2 types of binary complements, "1's complement" and "2's complement."

- 1's complement

  The "1's complement" of a given numeric value will be the result of the subtraction of each of the digits of this numeric value from 1, as a result, all the "0" and "1" bits of the original bit string are switched.

  For example, the "1's complement" of the bit string $(10110011)_2$ is shown below:

1 0 1 1 0 0 1 1
↓ ← All the "0" and "1" bits of the original bit string are switched
0 1 0 0 1 1 0 0 ←"1's complement"

- 2's complement
  "2's complement" is the "1's complement" bit string plus 1. Therefore, the "2's complement" of the bit string $(10110011)_2$ is obtained as follows:

1 0 1 1 0 0 1 1
↓ ← All the "0" and "1" bits of the original bit string are switched
0 1 0 0 1 1 0 0 ←"1's complement"
\+ 1 ←1 is added
0 1 0 0 1 1 0 1 ←"2's complement"

- "1's complement" and "2's complement" representation of negative integers
  - "1's complement" representation of negative integers
    - Sign bit: 0 for positive, 1 for negative, and both, +0 and -0, for 0
    - Numeric value: "1's complement"

For example, the "1's complement" representation of the decimal number -126 will be as follows:

0 1 1 1 1 1 1 0 ← + 126
Sign → ↓ ↓ ← All the "0" and "1" bits of the original bit string are switched
1 0 0 0 0 0 0 1 ← – 126

  - "2's complement" representation of negative integers
    - Sign bit: 0 for positive and 0, 1 for negative
    - Numeric value: "2's complement"

For example, the "2's complement" representation of the decimal number -126 will be as follows:

0 1 1 1 1 1 1 0 ← + 126
Sign → ↓ ↓ ← All the "0" and "1" bits of the original bit string are switched
1 0 0 0 0 0 0 1
\+ 1 ←1 is added
1 0 0 0 0 0 1 0 ← – 126

As can be observed, even for the same number the bit strings of the "1's complement" and the "2's complement" differ.
Figure 1-1-13 shows a comparison of the range of numeric values which can be represented with 3 bits in the "1's complement" and the "2's complement" representation. From this Figure it can be noted that the range of representable numeric values with the "2's complement" is wider. Likewise, as in the absolute value representation of negative integers, and in the representation of negative numbers using the "1's complement," 0 can be represented both, as +0 and as -0, so the operation becomes complicated. For that reason, a great number of today's computers have adopted the 2's complement method.
Figure 1-1-14 shows the range of representable numeric values when an n-bit binary number is represented with the "1's complement" and the "2's complement."

Figure 1-1-13
"1's complement"
and "2's complement"

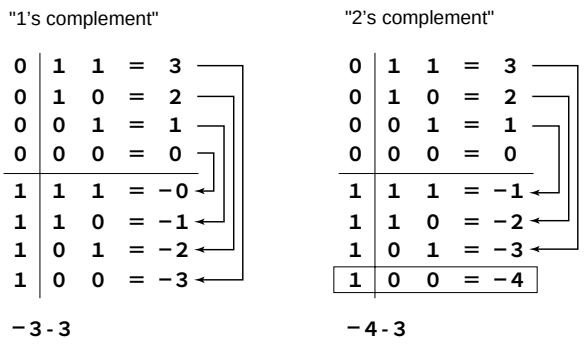


Figure 1-1-14
Range of representable numeric values with "1's complement" and "2's complement"

- Range of representable numeric values when an n-bit binary number is represented using the "1's complement" method

$$-(2^{n-1}-1) \text{ to } 2^{n-1}-1$$

- Range of representable numeric values when an n-bit binary number is represented using the "2's complement" method

$$-2^{n-1} \text{ to } 2^{n-1}-1$$

Another important reason for the adoption of the 2's complement method is illustrated by the following example:

Example When a decimal calculation of 100 - 90 is performed in a computer, the decimal numbers 100 and -90 are first converted into binary numbers. At this time, if –90 is represented using the "2's complement" representation, the minus sign will not be necessary, and the representation will be as follows:

$(100)_{10} = (01100100)_2$
$(-90)_{10} = (10100110)_2$

Therefore, the subtraction 100-90 can be replaced by the addition 100+(-90).

```
  0 1 1 0 0 1 0 0
+ 1 0 1 0 0 1 1 0
-----------------
1 0 0 0 0 1 0 1 0     (Decimal digit 10)
↑
```

Since the bit number is 8, the 9th digit resulting from the carry is ignored.

Therefore, the reason why negative numbers are represented using the "2's complement" method in computing is that subtractions can be performed as additions. In other words, since subtractions can be performed with the addition circuits, the subtraction circuits are unnecessary, simplifying the hardware structure.

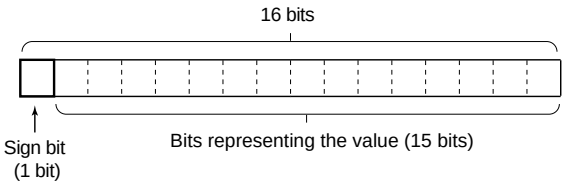
② Fixed point

a. Integer representation

The fixed point is a data representation format used mainly when integer type data is processed (Figure 1-1-15). Therefore, the fixed point format is also called an integer type.
In the unpacked decimal format or packed decimal format, depending on the digit number of the decimal number, the number of bytes changes, but in the fixed point format one word is represented in a fixed length such as 16 bits and 32 bits.
For that reason, if there is an attempt to represent a numeric value that exceeds the fixed length a problem called overflow will occur.

Figure 1-1-15
Fixed point

<When 1 word is represented by 16 bits>



Since in the fixed point format in Figure 1-1-15, where a value is represented with 15 bits, if a negative number is represented using the "2's complement" representation, the range of representable numeric values in the decimal system is as follows:

-2^{15} to $2^{15} - 1 = -32{,}768$ to $32{,}767$

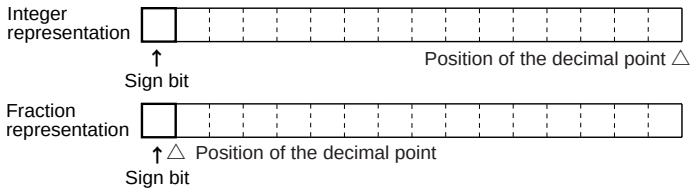
Likewise if one word is composed of n bits, and a negative number is represented using the "2's complement" representation, the range of representable numeric values in the decimal system is as follows:

-2^{n-1} to $2^{n-1} - 1$

b. Fraction representation

When integers are represented, the position of the decimal point is considered to be on the right-hand side of the least significant bit.
When fractions are represented, the position of the decimal point is considered to be immediately preceded by the sign bit.

Figure 1-1-16 Representation format of integers and fractions



③ Floating point

While the fixed point format represents integer-type data, the floating point format is used to represent real number type data. In ordinary mainframe computers, a maximum of 18-digit decimal numbers only can be represented. With 18 digits, there should be almost no problem in our daily life.
However, in a world where complicated calculations such as the ones mentioned below are required, correct results cannot be achieved with integer type data alone.

- Fluid mechanics calculations required for airplane design
- Calculations for weather forecasts
- Space flight planning and control
- Ballistic calculation
- CAD (Computer Aided Design)

For scientific and engineering fields requiring this kind of complicated calculation, the floating point format is used. Here, "complicated" means not only the calculation process itself is complicated, but also the either extremely large or small size of data is processed.
When we represent the number 1,500,000,000, instead of writing 8 zeros, we use the following exponent representation:

$$15 \times 10^{8}$$

In the floating point format it would be written as 0.15×10^{10}.

$$\underline{0.15} \times 10^{10}$$

↑
This part is represented as smaller than 1

The name of each of the numbers is shown below.

$$0.15 \times 10^{10} \leftarrow \text{Exponent}$$

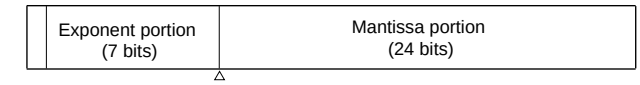
↑ ↑
Mantissa Radix

Here, to make it easier to understand, the decimal system is used, but in the computer, the binary system is used.
The floating point representation format varies depending on the computer. This is roughly classified into the format used in mainframe computers and that defined by the IEEE (Institute of Electrical and Electronics Engineering).

a. Floating point representation format in mainframe computers

The floating point representation format used in general-purpose computers is shown in Figure 1-1-17. This format was adopted in the first general-purpose computer in the world the "IBM System/360" and it was called Excess 64.

Figure 1-1-17 Floating point representation format in general-purpose computers



Mantissa sign (1 bit) The position of the decimal point is considered to be here

- Mantissa sign: Indicates whether the data represented in the mantissa portion is positive or negative (0: Positive, 1: Negative)
- Exponent portion: Here, 16 is considered to be the radix and a numeric value +64 is represented as the real exponent.
- Mantissa portion: Here, only a binary fraction lower than 1 can be represented.

● Exponent portion

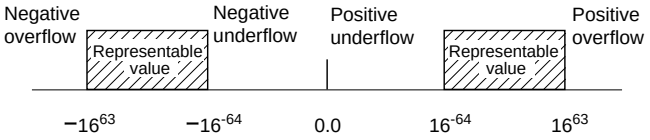
The exponent portion has 7 bits, and the range of numeric values representable in the binary system is $(0000000)_2$ to $(1111111)_2$, which in the decimal system is 0 to 127. However a numeric value 64 times larger than the real exponent is represented. For that reason, the real exponent is equivalent to –64 to +63.

Likewise, since the radix is considered to be 16, the numeric values that can be represented with the exponent portion range between

16^{-64} to 16^{63}

Then, including the sign bit, the range of numeric values that can be represented with the exponent portion is shown in Figure 1-1-18.

Figure 1-1-18 Range of numeric values that can be represented with the exponent portion



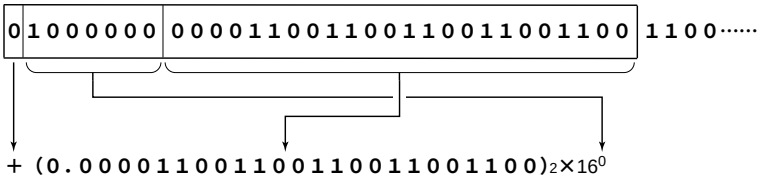
● Mantissa portion

When the decimal fraction 0.05 is converted into a binary fraction, it becomes a repeating binary fraction.

$(0.0000110011001100110011001100...)_2$

Figure 1-1-19 shows the representation of this fraction when the floating point format is used.

Figure 1-1-19 Representation of the binary fraction $(0.0000110011001100110011001100...)_2$

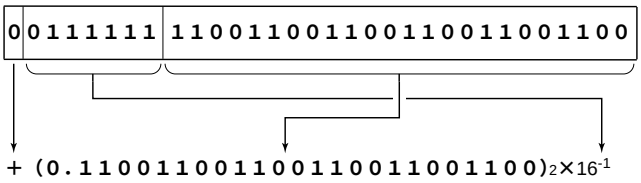


Since the mantissa portion has 24 bits, in this case, the decimal fraction 0.05 will not be represented correctly. (The error that occurs in this case is called a rounding error.)

However, if we look at the bit pattern of the mantissa portion, we can see that the 4 top bits are 0, if we then extract these 4 bits and shift the remaining bits to the left, 4 rounded bits can be represented. Here, as a result of shifting the mantissa portion 4 bits to the left, the original value of the mantissa portion was increased by $2^4 = 16$. In order to cancel this increase it is necessary to divide it into 16 ($\times 16^{-1}$). Since the radix is 16, the value of the exponent portion can be set to -1. The technique, used in this way, in order to reduce the rounding error to its minimum as well as to maximize precision is called normalization. Furthermore, as a result of this normalization technique, the bit strings that represent a value are standardized. This operation is performed automatically by the hardware.

The result of the normalization of the binary fraction representation in Figure 1-1-19 is shown in Figure 1-1-20.

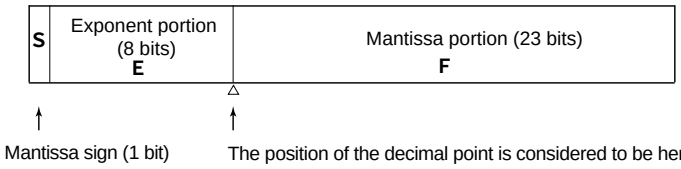
Figure 1-1-20
Normalization



b. IEEE Floating point representation format

The floating point representation format according to an IEEE standard is shown in Figure 1-1-21.

Figure 1-1-21 IEEE floating point representation format

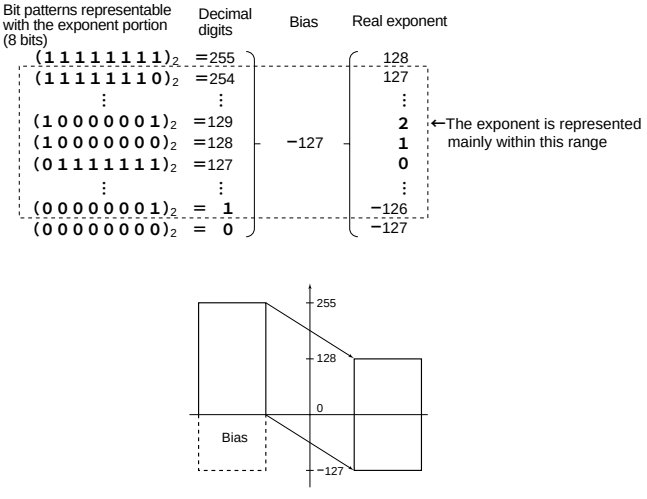


- Mantissa sign: Indicates whether the data represented in the mantissa portion is positive or negative
  (0: Positive, 1: Negative)
- Exponent portion: 2 is considered to be the radix, and a value resulting from the addition of 127 to the value of the original exponent portion is represented.
  However, When: E=255, F≠0 : Not a number
  When: E=255, F=0 : $(-1)^S\infty$
  When: E=0, F≠0 : $(-1)^S 2^{-126}$ (0.F)
  When: E=0, F=0 : $(-1)^S 0$ (Zero)
- Mantissa portion: Here, only a binary fraction lower than 1 can be represented.
- Value represented using the floating point format: $(-1)^S \times 2^{E-127} \times (1+F)$

The differences from the general-purpose computer floating point representation format are as follows:

- The exponent portion has 8 bits, and a value resulting from the addition of 127 to the value of the original exponent portion is represented. This addition to the original value is called bias (Figure 1-1-22).
- The mantissa portion has 23 bits and a binary fraction equivalent to the mantissa -1 is registered. In other words, 1 is considered to be omitted.
- The radix of the exponent portion is 2.

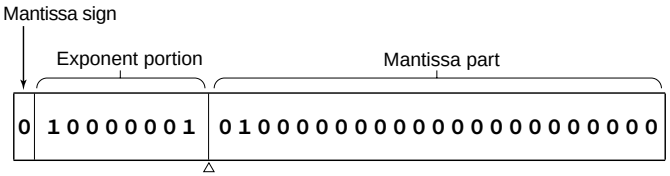
Figure 1-1-22 Representation of the exponent portion



For example, if the decimal number 5 is represented using this format, it will be represented as in Figure 1-1-23.

- Since it is positive, the mantissa sign will be 0.
- If $(5)_{10} = (101)_2 = (101)_2 \times 2^0 = (1.01)_2 \times 2^2$, then, the exponent portion will be $(1.01)_2$-$(1)_2 = (0.01)_2$
- If $(101)_2$ is shifted 2 bits to the right, it becomes $(1.01)_2$, 2^{-2} times the former value. In order to normalize it, 2 is added to the exponent 0, which becomes 2. As a result, since the exponent portion is 2 + 127 = 129, the representation will be $(10000001)_2$.

Figure 1-1-23
Representation of the decimal number 5



c. Shift operation

The mechanism of the shift operation performed at the normalization is explained below.
In the binary system, each digit has a weight which is a power of 2. This is called positional weight. Therefore, even though the same number 1 is represented, its meaning is different to that of the 1 positioned in the second digit and the 1 positioned in the third digit.

1 positioned in the second digit: $(10)_2 \rightarrow 2^1 = 2$

1 positioned in the third digit: $(100)_2 \rightarrow 2^2 = 4$

In the shift operation, by moving the position of 1 to the left (or to the right), the multiplication and division of numeric values can be easily performed.
The conversion into decimal numbers of $(100101)_2$ and $(1001010)_2$, resulting from shifting the first value 1 bit to the left, would be as follows:

Weight of each digit:		2^6	2^5	2^4	2^3	2^2	2^1	2^0	
$(100101)_2$	$\rightarrow$		1	0	0	1	0	1	$= 32 + 4 + 1 = (37)_{10}$
$(1001010)_2$	$\rightarrow$	1	0	0	1	0	1	0	$= 64 + 8 + 2 = (74)_{10}$

Through this conversion, it should be noted that the 1 that represented 2^5 before shifting, now represents 2^6, the 1 that represented 2^2 now represents 2^3, and the 1 that represented 2^0 now represents 2^1. In other words after the shift the value of each of the 1s was doubled, and the result of the conversion into the decimal system was also doubled from $(37)_{10}$, to $(74)_{10}$.
In short, the above-mentioned result shows that by shifting a binary digit 1 bit to the left, its value is doubled. Following this approach, the shift operation can be summarized by the following rules:
[Shift operation rules]

- When a binary number is shifted n bits to the left, its former value is increased 2^n times.
- When a binary number is shifted n bits to the right, its former value decreases 2^{-n} times. (The former value is divided by 2^n)

The shift operation can be used to calculate numeric values, as in the above-mentioned example, as well as to simply change the position of a bit.

● Arithmetic shift

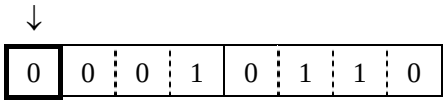
The arithmetic shift is the shift operation used to calculate numeric values. It is used in the fixed point format that represents negative numbers using the "2's complement" representation.
[Arithmetic shift rules]

- The sign bit is not shifted.
- The bit shifted out is lost.
- The bit to be filled into the bit position vacated as a result of the shift is:
  For left shifts: 0
  For right shifts: Same as the sign bit

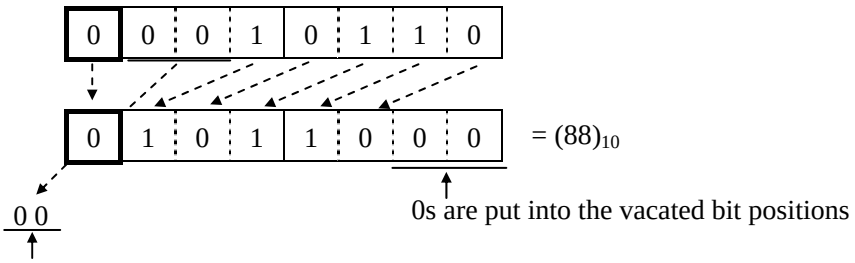
Example Calculation of $(22)_{10} \times 4$ using the arithmetic shift

① Represent $(22)_{10}$ using the fixed point format (8 bits)

$(22)_{10} = (10110)_2$



② Shift 2 bits to the left to increase it by 4 ($= 2^2$).



- Logical shift

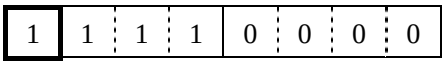
The logical shift is the shift operation used to change the bit position. The big difference from the arithmetic shift is that the sign bit is not treated differently.

[Logical shift rules]

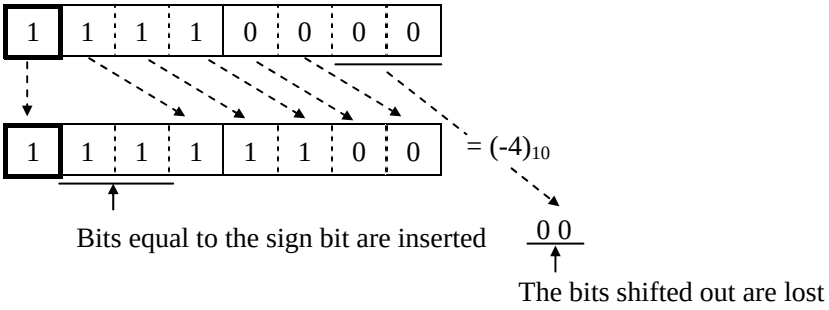
- The sign bit is also shifted (moved).
- The bit shifted out is lost.
- The bit to be filled into the bit position vacated as a result of the shift is 0.

Example After arithmetically and logically shifting $(-16)_{10}$ 2 bits to the right, convert each of the results into decimal digits.

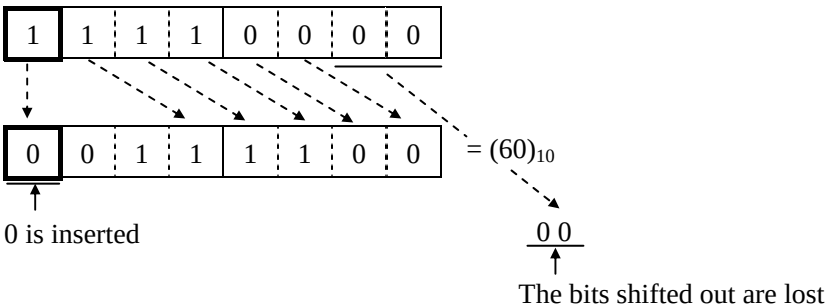
① Represent $(-16)_{10}$ using the fixed point format (8 bits).



② Arithmetically shift 2 bits to the right



③ Logically shift 2 bits to the right



1.1.3 Operation and precision

Since the storage capacity of a computer is limited, not all the numeric values we use can be represented correctly. In other words, a value represented in a computer is an approximate representation. As was mentioned above, in commercial data processing, operations are performed using decimal representation, while both the internal representation as well as the operation of scientific and engineering calculations, are performed using the binary representation. For that reason, the difference between the numeric value represented internally and the true value becomes a problem.

(1) Precision of the numeric value representation

The precision of a number is the range of its error, and so "high precision" means "small error". Focusing only on the integer part, if an enough number of digits to represent the conversion of decimal numbers into binary are available, no error occurs. However, the fraction part is not so simple; since many decimal fractions cannot be completely represented with the binary fractions containing a finite number of digits.

① Single precision

In scientific and engineering calculations, numerical values are represented with binary digits in word units. The word length depends on the hardware. For example, when 1 word = 16 bits, in general terms, the format in which 1 numeric value is represented with 1 word is called single precision. The range of numeric values representable with 16 bits, in case of an integer without a sign, is indicated below.

Minimum value = $(0000\ 0000\ 0000\ 0000)_2 = 0$

Maximum value = $(1111\ 1111\ 1111\ 1111)_2 = 65{,}535$

In other words, values higher than 65,535 cannot be represented.
Likewise, the range of numeric values representable with 16 bits, in the case of a fraction without a sign is indicated below.

Minimum value = $(0000\ 0000\ 0000\ 0000)_2 = 2^{-16} \leq 0.0000152587890625000$

Maximum value = $(1111\ 1111\ 1111\ 1111)_2 = 1\text{-}2^{-16} \leq 0.9999847412109370000$

In this case, values lower than 0.00001525878, and values higher than 0.999984741210937 can't be represented.

② Double precision

In order to widen the range of representable numeric values, the number of digits is increased. By representing 1 numeric value with 2 words, in comparison to the representation with 1 word, the range of representable numeric values becomes much wider. This format is called double precision. If 1 word = 16 bits, 1 numeric value is represented with twice as many bits, 32 bits.

Minimum value = $(0000\ 0000\ 0000\ 0000\ 0000\ 0000\ 0000\ 0000)_2 = 0$

Maximum value = $(1111\ 1111\ 1111\ 1111\ 1111\ 1111\ 1111\ 1111)_2 = 4{,}294{,}967{,}295$

In other words, values up to 4,294,967,295 can be represented.
Likewise, the range of numeric values representable with 32 bits, in case of a fraction without sign is indicated below.

Minimum value = $(0000\ 0000\ 0000\ 0000\ 0000\ 0000\ 0000\ 0000)_2 = 2^{-32}$
$\leq 0.00000000023283064365387$

Maximum value = $(1111\ 1111\ 1111\ 1111\ 1111\ 1111\ 1111\ 1111)_2 = 1 - 2^{-32}$
$\leq 0.99999999976716900000000$

(2) Operation precision

① Precision of fixed point representation

The range of representable numeric values with the fixed point representation depends on the computer hardware. Depending on the number of bits in one word, the range of representable numeric values differs. The step size of the integer part is always 1, regardless of the number of bits, and only the maximum value changes. However, in the step size of the fraction part the larger the number of bits assigned, the smaller the step size becomes and the error is also reduced.

② Precision of floating point representation

a. Overflow and underflow

When multiplication of extremely large values or extremely small values is performed, there are cases where the operation results exceed the range of numeric values that can be represented with the exponent portion. The condition that occurs when the product is higher than the maximum value that can be represented with the exponent portion, is called overflow (Maximum absolute value < Overflow). The condition that occurs when the product is lower than the minimum absolute value is called underflow (0 < Underflow < Minimum absolute value).

b. Cancellation

When subtraction of two floating point numbers of almost equal values is performed, since the result becomes extremely small, it is left out of the range of numeric values which can be represented. This condition is called cancellation.

c. Loss of information

When two values represented by using the floating point format are added, the exponents must coincide. Generally, exponents are adjusted to the largest value.
When an addition of an extremely small value and an extremely large value is performed, since exponents must be adjusted to the exponent of the largest value, the mantissa portion of the small value is shifted largely to the right. As a consequence of this shift, the information that should have been represented is lost. This condition is called loss of information.

In order to avoid this kind of error, it is necessary to think out strategies such as changing the order of the operations, etc. These strategies are worked out by the user.

1.1.4 Non-numeric value representation

When using a computer, in order to input numerals and characters (alphabetical characters, symbols, etc.) input devices such as keyboards, are used. Inside the computer, in order to represent characters using binary digits a concept called code is used.
Presently, different character codes are used depending on the computer. Here, the codes widely used around the world and in Japan will be explained.

(1) Character representation

The character codes widely used worldwide basically represent 1 character with 8 bits, that is, 1 byte.
The character codes used in the information processing field are sometimes called codes for information interchange.
By typing on a keyboard, these character codes are input in the computer as 1-byte codes.
The following keys are found in the keyboard of a personal computer, etc.

- Numeric keys: 10 types (0 to 9)
- Character keys: Alphabet: (Uppercase: A to Z and lowercase: a to z) 52 types
- Symbolic keys: 40 types
- Control character keys: 34 types (Space key, etc.)

To assign the unique bit pattern corresponding to these 136 types of characters and symbols, 256 types of bit patterns that can be represented with 8 bits are required.

(2) Character codes

The main character codes are listed below.

① ASCII (American Standard Code for Information Interchange) code

The ASCII code was established by the U.S. standards institution, ANSI (American National Standards Institute) in 1962. Character code of 8 bits composed by the code bit representing the alphabet, numeric characters, etc. (7 bits) and the parity bit used to detect errors. It is used in personal computers and in data transmission.

② ISO (International Organization for Standardization) code

The ISO code is a 7-bit character code that was established by the International Organization for Standardization (ISO) in 1967 based on the ASCII code. It is the base of the character codes used in all countries of the world.

③ JIS (Japanese Industrial Standards) code

The JIS code was established as JIS X 0201 by adding the Romaji, hiragana and other characters peculiar to the Japanese language to the ISO code. The "JIS 7-bit code" used to represent Romaji and the "JIS 8-bit code" used to represent katakana as well as the "JIS kanji code," that represents 1 character with 2 bytes (16 bits), and is used to represent hiragana and kanji, exists.

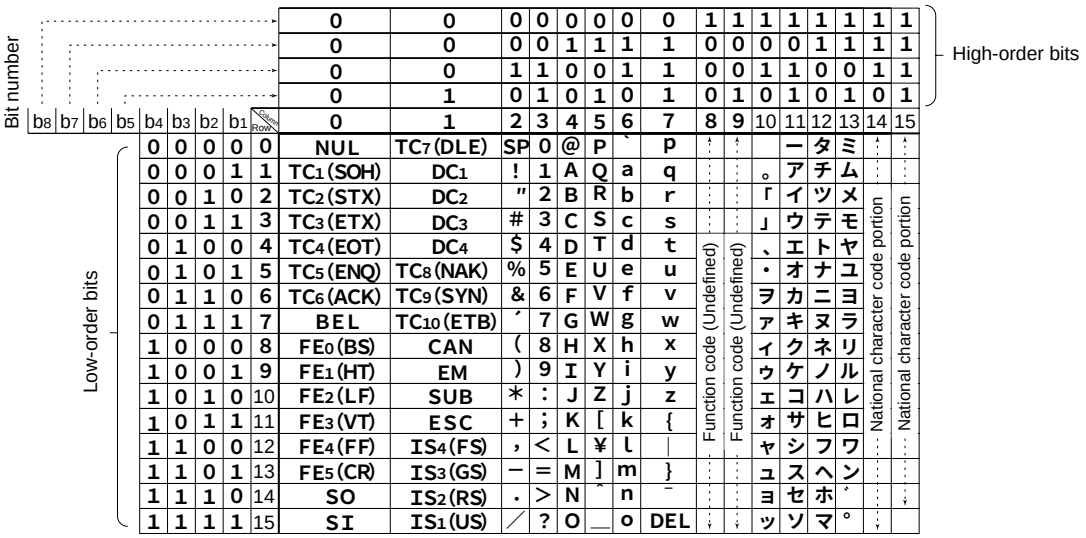
④ EBCDIC (Extended Binary Coded Decimal Interchange Code)

The EBCDIC is a character code developed by IBM. Compared to the above-mentioned character codes, which were established to be used as standards, the EBCDIC code was developed for IBM computers. Since IBM held the greatest share of the computer market when the third generation computers, in which this code was developed, were launched, other companies developed their computers according to this character code, and as a result it became a standard character code. Standards like this, resulting from the existence of a large number of users, are called *de facto* standards.

⑤ Shift JIS code

As was mentioned above, to represent kanji, the JIS kanji code represents 1 word with 2 bytes.

Figure 1-1-24 JIS X 0201 code table



There are times when the JIS code and the JIS kanji code are mixed. When 1-byte codes and 2-byte codes are mixed, their interpretation can't be performed. Therefore, a special escape code is added to the front and back of the kanji code string. For example, a bit string of 5 kanji becomes 10 bytes + 2 bytes. When data is missed during data transmission, etc. recovery is difficult.
In order to find a solution to this defect, the shift JIS code that converts the characters defined in the JIS Kanji code into another code system was created. The first byte of the shift JIS uses a code that is not used in the JIS (1 byte) code, and, at the same time, by avoiding control character codes in the second character, 1-byte codes and 2-byte codes can be mixed without using special escape codes.

⑥ Unicode

The unicode is a 2-byte code system unified to all the countries, which was proposed and designed by Apple Computer, IBM, Microsoft, and other U.S. companies in order to smooth the exchange of data amongst personal computers. This code was adopted by ISO as an international standard draft.

(3) Audio representation

As has been said, the current information technology provides multimedia support, and the data subject to processing is not limited to character data or numerical data. It also covers many kinds of information that are used in our daily life. One of the components which compose multimedia is audio.
The human voice is produced when the airflow generated in the lungs changes, vibrates and resonates due to a great number of organs such as the tongue, lips, teeth, jaw, nasal cavity and vocal cords. Since audio data has a complicated analog waveform, audio analysis is performed using a numeric formula and once it is converted into digital codes it is processed in the computer. Word processors that accept audio input and speaker recognition are examples of its recent applications.

(4) Image representation

In order to support current multimedia, not only audio but also image data must be processed.
Inside the computer, image data is processed as a set of dots. For that reason, the registration of the status of each of the dots that compose an image, is the registration of the image data itself. The easiest approach is to register two states, black and white, for each of the dots that compose an image. In this case, 1 bit is used to register the information of each dot. Today most of the image data is colored, so this method does not solve all the problems. Therefore, the representation method that combines the basic colors in each dot is used. Amongst computer screens, there are a great number of systems that combine the three primary colors (Red, green and blue) in 256 levels respectively and represent approximately 16,000,000 colors. In this case, since 8 bits are needed for 1 color, in order to register the information of 1 dot, 24 bits are used.

1.2 Information and logic

1.2.1 Proposition logic

Operations which can be processed in a computer are not limited to arithmetic formulas. By assigning a value to a sentence, sentence operations can be performed. For example, in logical mathematics, the sentences represented as "The wind is blowing," "It is raining," "x=5" and "y=2" are called propositions. Values of "truth" or "lie," in other words, "true" and "false" can be assigned to these propositions. However, one proposition will always be either "true" or "false." The same proposition can't be "true" at the same time it is "false." "The wind is blowing and it is raining" is possible, but "The wind is blowing, there is no wind" is impossible.
These propositions are represented by p, q, r, ... and other letters, and through the combination of their logical significance new synthetic propositions can be created. Each proposition relation is made clear, through logical operation by proposition logic. Whether a synthetic proposition is true or false is determined by the truth table. An example of this table is shown in Figure 1-2-1.
This truth table shows:

- The proposition "The wind is not blowing" is false when the proposition 1, "The wind is blowing," is true, it is true when the proposition 1 is false.
- The proposition "The wind is blowing or it is raining" is true when both, the proposition 1, "The wind is blowing," and the proposition 2, "It is raining," are true or when either of the two is true. When both of them are false, it is false.

Figure 1-2-1 Truth table

Proposition 1 The wind is blowing	Proposition 2 It is raining	The wind is not blowing	The wind is blowing and it is raining	The wind is blowing or it is raining	If the wind blows it rains
True	True	False	True	True	True
Wind	Rain	Lie	Truth	Truth	Truth
True	False	False	False	True	False
Wind	No rain	Lie	Lie	Truth	Lie
False	True	True	False	True	False
No wind	Rain	Truth	Lie	Truth	Lie
False	False	True	False	False	False
No wind	No rain	Truth	Lie	Lie	Lie

1.2.2 Logical operation

Since the expression of the logical significance with words becomes lengthy and it is not suitable for computer operations, logical relations are represented with symbols. The symbols that represent these propositional operations (or logical operations) are called logical symbols or logical connectors. The main logical symbols used in information processing are NOT, AND, OR, exclusive OR, etc. Their meanings are explained below.
Each logical operation will be explained using the examples shown in Figure 1-2-1, and proposition 1 "The wind is blowing," which will be denoted as p, and proposition 2 "It is raining" as q.

(1) Negation

By negating the proposition "The wind is blowing," a new proposition, "The wind is not blowing" can be created. In this case, the logical symbol "¬ (NOT)" is used and it is represented as "¬p."

Figure 1-2-2
Truth table for negation

p	¬ p
T	F
F	T

T(rue) : True
F(ales) : False

(2) Logical product

When two propositions are connected with the conjunction "AND" as in "The wind is blowing and it is raining," both "The wind is blowing" and "It is raining" are expressed simultaneously.
The connection of the two propositions p and q with the conjunction "AND" is called logical product. In this case, the logical symbol "∧ (AND)" is used and it is represented as "p∧q." The truth table is shown in Figure 1-2-3, and the result is true only when p and q are both true.

Figure 1-2-3
Truth table for
the logical product

p	q	p∧q
T	T	T
T	F	F
F	T	F
F	F	F

(3) Logical sum

When two propositions are connected with the conjunction "OR" as in "The wind is blowing or it is raining," either "The wind is blowing" or "It is raining" is expressed.
The connection of the two propositions p and q with the conjunction "OR" is called logical sum. In this case, the logical symbol "∨ (OR)" is used and it is represented as "p∨q." The result is true only when p and q are both true.

Figure 1-2-4
Truth table for
the logical sum

p	q	p∨q
T	T	T
T	F	T
F	T	T
F	F	F

(4) Exclusive OR

In the logical sum mentioned above, "The wind is blowing or it is raining," either "The wind is blowing" or "It is raining" is expressed. This logical sum is true when "The wind is blowing and it is also raining," or in other words, when both propositions are true. Ordinarily, the word "or" is used in many cases to express exclusive meanings as "either of the two." The exclusive OR is used to support these cases.
In the case of the exclusive OR, the logical symbol "$\veebar$ (EOR)" is used and it is represented as "p$\veebar$q." The result is true only when p or q, either of the two, is true. Therefore, the result is false when p and q are both true or false. This logical operation is frequently used in programming.

Figure 1-2-5
Truth table for
the exclusive OR

p	q	p∀q
T	T	F
T	F	T
F	T	T
F	F	F

(5) Negative AND (NAND)

It is the negation of the above-mentioned logical product. It is represented as "¬ (p∧q)." This logical operation is frequently used in the design of digital circuits.

(6) Negative logical sum (NOR)

It is the negation of the above-mentioned logical sum. It is represented as "¬ (p∨q)."
Figure 1-2-6 puts the six logical operations mentioned above together.

Figure 1-2-6 Truth table for the logical operations NOT, AND, OR, EOR, NAND and NOR (Summary)

p	q	NOT p	p AND q	p OR q	p EOR q	p NAND q	p NOR q
T	T	F	T	T	F	F	F
T	F	F	F	T	T	T	F
F	T	T	F	T	T	T	F
F	F	T	F	F	F	T	T

(7) Logical expression laws

The representation using the above-mentioned logical symbols is called logical expression. Along with the logical symbols presented earlier, the symbols shown in Figure 1-2-7 are also used.

Figure 1-2-7
Logical symbols

Meaning		Symbols		Notation example
Negation	NOT	¬	‾	$\overline{X}$
Logical product	AND	∧	•	X•Y
Logical sum	OR	∨	+	X+Y
Exclusive OR	EOR	∀	⊕	X⊕Y

As the logic becomes complicated, the logical expression also becomes extremely complicated. For that reason, in order to simplify the logical expressions, the following laws are used:

- Logical product law: $X \cdot X = X, \quad X \cdot \overline{X} = 0, \quad X \cdot 0 = 0, \quad X \cdot 1 = X$
- Logical sum law: $X + X = X, \quad X + \overline{X} = 1, \quad X + 0 = X, \quad X + 1 = 1$
- Exclusive OR law: $X \oplus X = 0, \quad X \oplus \overline{X} = 1, \quad X \oplus 0 = X, \quad X \oplus 1 = \overline{X}$
- Commutative law: $X + Y = Y + X, \quad X \cdot Y = Y \cdot X$
- Associative law: $X + (Y + Z) = (X + Y) + Z, \quad X \cdot (Y \cdot Z) = (X \cdot Y) \cdot Z$
- Distributive law: $X + (Y \cdot Z) = (X + Y) \cdot (X + Z)$
  $X \cdot (Y + Z) = (X \cdot Y) + (X \cdot Z)$
- Absorptive law: $X + (X \cdot Y) = X, \quad X \cdot (X + Y) = X$
- Restoring law: $\overline{\overline{X}} = X$
- De Morgan's law: $\overline{X + Y} = \overline{X} \cdot \overline{Y}, \quad \overline{X \cdot Y} = \overline{X} + \overline{Y}$

For example, the logical expression of the exclusive OR is represented as $X \oplus Y = (\overline{X} \cdot Y) + (X \cdot \overline{Y})$. By using the above-mentioned laws, this logical expression can be changed as follows:

$$
\begin{aligned}
X \oplus Y &= (\overline{X} \cdot Y) + (X \cdot \overline{Y}) \\
&= ((\overline{X} \cdot Y) + X) \cdot ((\overline{X} \cdot Y) + \overline{Y}) && \cdots\cdots \text{Distribution law} \\
&= ((X + \overline{X}) \cdot (X + Y)) \cdot ((\overline{Y} + \overline{X}) \cdot (\overline{Y} + Y)) && \cdots\cdots \text{Switching law and distribution law} \\
&= (\quad 1 \quad \cdot (X + Y)) \cdot ((\overline{Y} + \overline{X}) \cdot \quad 1 \quad) && \cdots\cdots \text{Logical sum law} \\
&= (X + Y) \cdot (\overline{X} + \overline{Y}) && \cdots\cdots \text{Logical product law}
\end{aligned}
$$

Exercises

Q1 **Which of the following represents correctly the size relation amongst the following 4 prefix symbols that represent integer exponents of 10: G (giga), k (kilo), M (mega) and T (tera)?**

a. G < k < M < T
b. k < G < T < M
c. k < M < G < T
d. M < G < k < T
e. M < T < G < k

Q2 **Which of these values corresponds to 1 picosecond?**

a. 1 nanosecond × 1,000
b. 1 microsecond / 1,000,000
c. 2^{-12} seconds
d. 10^{-10} seconds
e. 10^{-11} seconds

Q3 **Given the binary digits A and B, where**

A = 01010101 B = 01100110

which of these values corresponds to the result of the operation A + B?

a. 01101010
b. 01111010
c. 10011010
d. 10111011
e. 11010101

Q4 **Which of these values is the correct result of the subtraction of the hexadecimal numbers DD and 1F "00-1F"?**

a. AF b. BE c. CE d. EC e. FC

Q5 **Which of these values represents in decimal numbers the result of the addition of the binary numbers 1.1011 and 1.1101?**

a. 3.1 b. 3.375 c. 3.5 d. 3.8 e. 3.9375

Q6 **Which is the decimal number that can be represented without error in the binary floating point representation?**

a. 0.2 b. 0.3 c. 0.4 d. 0.5

Q7 **Which of these values represent the hexadecimal fraction 0.248 in decimal fractions?**

a. $\frac{31}{32}$ b. $\frac{31}{125}$ c. $\frac{31}{512}$ d. $\frac{73}{512}$

Q8 **Which is the correct bit pattern of the decimal number +432 when it is represented in the packed decimal format? Note that the sign is represented by the last 4 bits, and that "1100" represents positive numbers while "1101" represents negative numbers.**

a. 0000 0001 1011 0000
b. 0000 0001 1011 1100
c. 0001 1011 0000 1100
d. 0100 0011 0010 1100
e. 0100 0011 0010 1101

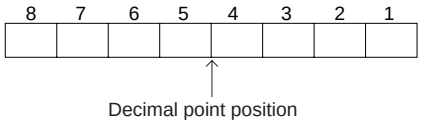
Q9 **What is the range of the integer representable with n bits using the fixed point format that represents a negative number with 2's complement representation? Here, the decimal point position is on the right side of the least significant bit (LSB).**

a. -2^{n} to 2^{n-1}
b. -2^{n-1} to 2^{n-1}
c. -2^{n-1} to $2^{n-1} - 1$
d. $-2^{n-1} - 1$ to 2^{n-1}

Q10 **Which of the following is the reason why in many computers the complement representation is used to simplify the operation circuit?**

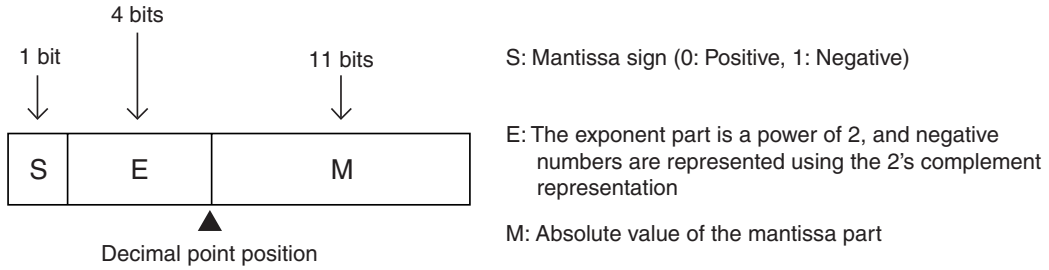
a. Additions can be processed as subtractions.
b. Subtractions can be processed as additions.
c. Multiplications can be processed by the combination of additions.
d. Divisions can be processed by the combination of subtraction.

Q11 **Which of these values corresponds to the representation of the decimal number -5.625 in binary number using the 8-bit fixed point format? Here, the position of the decimal point is between the fourth and fifth bits, and negative numbers are represented using the 2's complement representation.**



a. 01001100 b. 10100101 c. 10100110 d. 11010011

Q12 **Which is the normalized representation of the decimal number 0.375? Here, the numeric value is represented using the 16-bit floating point format and the format is indicated in the figure. The normalization performed is an operation that adjusts the exponent portion in order to eliminate the 0s of higher digit rather than the significant values of the mantissa portion.**



a | 0 | 0 0 0 0 | 0 1 1 0 0 0 0 0 0 0 0 0
b | 0 | 1 0 0 0 | 0 0 0 0 0 0 0 0 0 1 1
c | 0 | 1 1 1 1 | 1 1 0 0 0 0 0 0 0 0 0 0
d | 1 | 0 0 0 1 | 1 1 0 0 0 0 0 0 0 0 0 0

Q13 **Which is the value that corresponds to the result of logically shifting the hexadecimal number ABCD two bits to the right?**

a. 2AF3 b. 6AF3 c. AF34 d. EAF3

Q14 **The multiplication of binary number can be performed through the shift operation (digit shift), and addition. To increase the binary digit m by 2^n it is necessary to shift m n bits to the left.**

For example, m × 19 can be obtained through the following operation:

(Value of the result of shifting m one bit to the left) + (Value of the result of shifting m one bit to the left) + m

Which is the value of a?

a. 2 b. 3 c. 4 d. 5

Q15 **The decimal number -100 is registered using the 2's complement representation in a 8-bit register. Which of these values represent, in decimal numbers, the result of arithmetically shifting this registration three bits to the right?**

a. –33 b. –13 c. –12 d. 19

Q16 **Which of the following descriptions of the rounding error is correct?**

a. It is an error generated when an operation result exceeds the maximum value that can be processed by the computer.
b. Due to the limited number of digits in the number representation, it is an error generated as a result of rounding off, rounding up, or omitting portions smaller than the least significant digit.
c. It is an error generated due to the loss of the most significant values in the subtraction operation of numeric values, whose absolute values are almost equal.
d. It is an error generated due to the loss of the least significant values of the mantissa portion of the numeric value, with the lower exponent value in the subtraction operation of floating point numbers.

Q17 **What is the minimum number of digits required for representing uniquely with the same number of bits of the uppercase alphabetic characters (A to Z) and the numeric characters (0 to 9)?**

a. 5 b. 6 c. 7 d. 8

Q18 **The following truth table shows the operation results of the logical operation "x ★ y." Which of these expressions is equivalent to this operation?**

Truth table

x	y	x ★ y
True	True	False
True	False	False
False	True	True
False	False	False

a. x AND (NOT y)
b. x OR (NOT y)
c. (NOT x) AND y
d. (NOT x) AND (NOT y)
e. (NOT x) OR (NOT y)

Q19 **Which of the following expressions is equivalent to the logical expression $\overline{(A + B)\cdot C}$? Here, "·" represents the logical product (AND), "+," the logical sum (OR), and $\overline{A}$ is the negation of A (NOT).·**

a. $(A\cdot B) + \overline{C}$
b. $A \cdot B \cdot \overline{C}$
c. $\overline{A} + \overline{B} + \overline{C}$
d. $(\overline{A} \cdot \overline{B}) + \overline{C}$

2 Hardware

Chapter Objectives

Among computer related technology, the progress of hardware technology is particularly remarkable.
In this chapter, the learning objective is to understand the mechanism and functions of each of the five main units that are basic to computer hardware. To:

① Understand the roles and functions of the computer's five main units,
② Understand the basic operations and the registers used to read, decode and execute the instructions and data stored in the main storage unit by the processor (processing unit), which is composed of the control unit and the arithmetic unit,
③ Understand the basic approach and configuration circuits of the arithmetic unit that performs arithmetic operations and logical operations,
④ Understand the mechanism and functions of the main storage unit and the auxiliary storage devices used to store data, and know their types and characteristics,
⑤ Understand the types and mechanism of the input/output units as well as the input/output control system and the input/output interface,
⑥ Understand computer types and characteristics.

Introduction

The functions of the hardware composing a computer can be divided broadly into the following five categories:

· Input
· Storage
· Operation
· Control
· Output

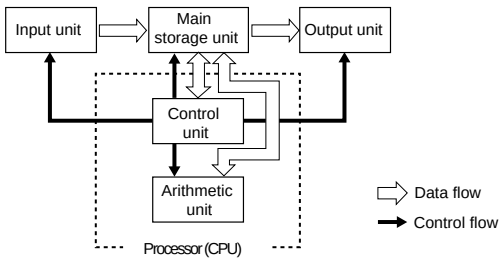
The following are the units that implement the above-mentioned functions:

· Input unit: This unit inputs the data and programs for computer processing. It is equivalent to the human eyes and ears.
· Storage unit: This unit stores the input data and programs. It is equivalent to the memory section of the human brain.
· Arithmetic unit: This unit conducts calculation and decision on the stored data according to the instructions of the program. It is equivalent to the thinking section of the human brain.
· Control unit: This unit controls the input unit, storage unit, arithmetic unit and the output unit. It is equivalent to the human central nervous system.
· Output unit: This unit outputs the results of computer processing in a format that can be understood by humans. It is equivalent to the human hands and mouth.

These five units are called the "computer five main units" (Figure 2-1-1).

Since the control unit and the arithmetic unit are handled as one unit, they are called the processor (processing unit) or central processing unit (CPU). The general term "peripheral devices" is used to refer to the input unit, the output unit and the auxiliary storage devices that are outside the processor and exchange data with the main storage unit. Likewise, the storage units are divided into main storage unit and auxiliary storage device, depending on their functions.

Figure 2-1-1
Computer five main units



2.1 Information element

2.1.1 Integrated circuit

In today's computers, an integrated circuit (IC) that integrates semiconductors to a high level is used. According to the integration level, ICs are classified as in Figure 2-1-2.

Figure 2-1-2
IC classification according to their integration level

IC	Integration level
SSI (Small Scale Integration)	10^1-10^2
MSI (Medium Scale Integration)	10^2-10^3
LSI (Large Scale Integration)	10^3-10^4
VLSI (Very Large Scale Integration)	10^5-

Note: The integration level indicates the range of the number of gates (number of transistors) contained in 1 IC.

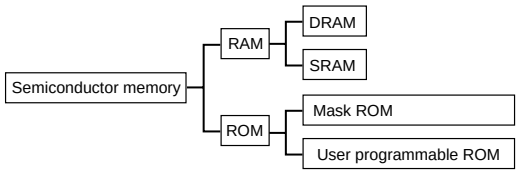
Likewise, according to their structure, ICs can be classified as follows:

- Bipolar IC: The speed and power requirements, as well as the costs, are high. It is used as a logic element.
- CMOS IC: The speed and power requirements, as well as the costs, are low. It is used as a storage element.

2.1.2 Semiconductor memory

Logic elements are used in logical operations while storage elements are used in data and instruction storage. Here the storage elements will be explained. (The logical circuit will be explained in detail in the next section, 2.2.) The storage element is called semiconductor memory (or IC memory) and is broadly divided into RAM and ROM.

Figure 2-1-3
RAM and ROM



(1) RAM (Random Access Memory)

A RAM is semiconductor memory in which data writing and reading is possible. When the computer is turned off, the stored data is lost. This property is called volatility. Since most main storage units are composed of RAMs, the processor can be made to read and write information from the main storage unit at random by specifying the address.
RAMs are classified into DRAMs and SRAMs.

① DRAM (Dynamic RAM)

A DRAM represents bits, and stores information depending on whether the part called capacitor is being charged (status "1") or is not being charged (status "0").
Since the circuits are simple and small, RAMs of large capacity can be created at low cost. However, since the charge stored in the capacitor is lost after a lapse of time, the memory needs to be rewritten (recharged) at regular intervals. This operation is called refreshing. Once, DRAMs were used in the main storage unit, but currently they are also used in storage units, etc., contained in the input/output units of

printers and other devices.

② SDRAM (Synchronous DRAM)

Due to the progress of IC technology, and the consequent substantial improvement of the performance of processors, the operating speed of the DRAMs that composed the storage unit could not keep up with the operating speed of the processors. For that reason, an external clock signal that indicates the processor operation timing is now set in the DRAM and through synchronization with this signal, complicated address specifications are reduced and simplified, enabling the development of DRAMs that operate at high speeds. These types of DRAMs are called synchronous DRAMs (SDRAM).

③ SRAM (Static RAM)

SRAMs are created with a circuit called the flip-flop. The flip-flop settles the output according to the previous input and the current input, and can preserve the status "1" and "0" inside the circuit. Since data is not lost unless the computer is turned off, memory refreshing is not necessary. However, since SRAM circuits are complicated, the memory capacity is smaller than that of DRAMs and the cost is higher. However, since its processing speed is high, it is used in devices such as the registers contained in main storage units and processors.

(2) ROM (Read Only Memory)

The ROM is semiconductor memory for read use only. Since programs and data are stored in the ROM from the beginning, the stored information is not lost even if the computer is turned off. This property is called nonvolatility.
ROMs are classified into mask ROMs and user programmable ROMs.

① Mask ROM

Since programs and data are already written in the Mask ROM before it is shipped by the manufacturer, the user cannot add any programs or data. Mask ROMs are used in the memories of game cassettes and IPL (Initial Program Loader), a program used to start the computer, etc.

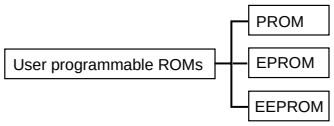
② User programmable ROM

The user programmable ROM is a type of ROM, but since at the time it is shipped by the manufacturer it has nothing stored in it, the user can write data into it once. The following types of user programmable ROM exist (Figure 2-1-4).

- PROM (Programmable ROM): Once data has been written, it cannot be erased.
- EPROM (Erasable PROM): It can be erased with ultraviolet light and rewritten.
- EEPROM (Electrically Erasable PROM): It can be erased through the application of electrical voltage and rewritten.

EEPROM is used in a storage medium called flash memory, which is used in the registration of image data of digital cameras, etc. Likewise, it is also used in the storage section of IC cards, etc.

Figure 2-1-4
User programmable ROMs



2.2 Processor architecture

2.2.1 Processor structure and operation principles

(1) Processor structure

Among the five main units that compose a computer, the control unit and the arithmetic unit are handled as one unit and are called processor (processing unit). The processor is also called central processing unit (CPU), and as the backbone of the computer, it plays the important roles indicated below.

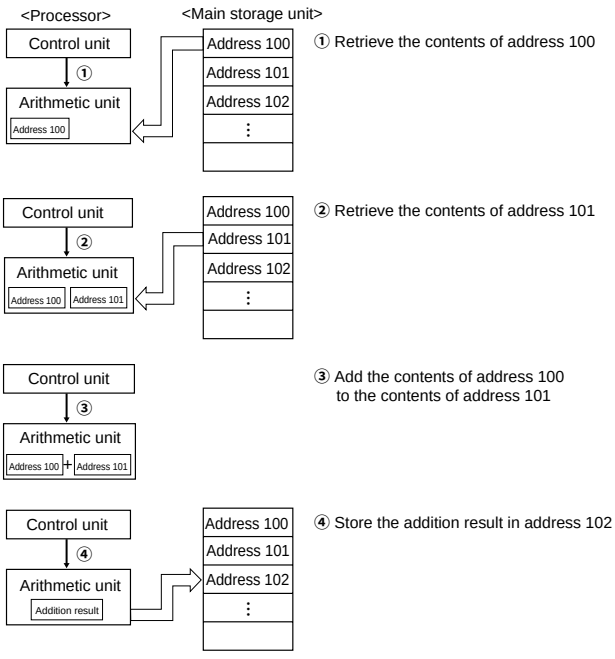
① Control unit

The control unit is the unit that controls all the operations of the computer. It retrieves, decodes and executes, one by one, in order, the instructions stored in the main storage unit.
The following are the main functions of the control unit:

- Retrieval of the instructions stored in the main storage unit
- Decoding of the retrieved instructions using the instruction decoder
- According to the decoding, transmission of the specifications required for the execution of the instructions to each unit

Through the repetition of these operations, the control unit controls each unit, and implements the functions of each of the units as a computer system. The system by which instructions are executed in this way, one by one, is called the sequential control system. The design of this sequential control system and the stored program system (Please refer to main storage unit in 2.3.1) was based on the concepts of John von Neumann. Computers of this kind are called Neumann computers.
In Figure 2-2-1, as an example to explain the mechanism of the sequential control system, the process of the "addition of the contents of address 100 and the contents of address 101 and the storage of the result in address 102" is represented.

Figure 2-2-1
Mechanism of the sequential control system



② Arithmetic unit

The official name of the arithmetic unit is arithmetic and logic unit (ALU). This unit performs arithmetic operations, logical operations, comparison, branch, and other processes on the data assigned to be subject to processing. The main instructions are shown in Figure 2-2-2.

Figure 2-2-2
Functions of the arithmetic unit

Basic operations	Basic instructions
Arithmetic operations	Addition, subtraction, multiplication and division
Logical operations	Logical sum (OR), logical product (AND), negation (NOT)
Comparison	Comparison instruction (size comparison)
Branch	Branch instruction (change of the sequence of instruction execution according to the conditions)

Depending on the representation method of data assigned to be subject to operations, arithmetic and logic unit has functions performing fixed point operation, floating point operation, and decimal operation.

(2) Processor operation principles

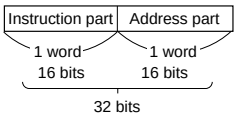
① Instruction readout and decoding

The data and programs retrieved from the main storage unit are transferred to the processor through the data bus. Then the data subject to processing is temporarily stored in the "general-purpose register," while a part of the program indicating the procedure of the process is transferred to the "instruction register."

a. Instruction and instruction format

- Instruction
  The program is a set of instructions that indicates "do …" to the computer. Since inside the computer, data is represented in the binary system, instructions are also represented in the binary system. The instructions represented in the binary system are called machine language instructions. Regardless of the program language in which a program is written, at the end it is converted into language that can be understood by the computer, machine language, in order to be decoded and executed. Machine language instructions and instruction formats differ depending on the computer, but, in general terms, they are composed of the following parts.
  - **Instruction part:** Indicates instructions and operations
  - **Address part:** Specifies the address and register of the main storage unit subject to processing

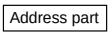
Figure 2-2-3
Example of the instruction structure



- Instruction format
  In practice, instruction formats differ depending on the computer. There are instructions that have several address parts, and according to the structure of the address part, there are four address formats, from zero-address format to three-address format.

- **Zero-address format**
  The zero-address format performs operations using a dedicated register called a stack pointer. Currently, zero-address format computers are not used. The stack pointer is the register that stores the address to be returned to (return address) after execution completion.

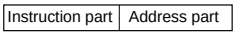
Figure 2-2-4 Zero-address format



- **Single-address format**
  The single-address format performs operations between the content of the main storage unit specified

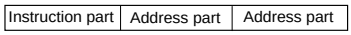
in the address and the accumulator data (Figure 2-2-5). The accumulator stores operation values and operation results. There are cases where general-purpose registers are also used as accumulators.

Figure 2-2-5 Single-address format



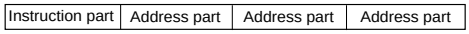
- **Two-address format**
  The two-address format specifies two addresses and uses the address data specified on the main storage unit.

Figure 2-2-6 Two-address format



- **Three-address format**
  The three-address format specifies two addresses to be used for the operation, and the address where the operation result is to be stored.

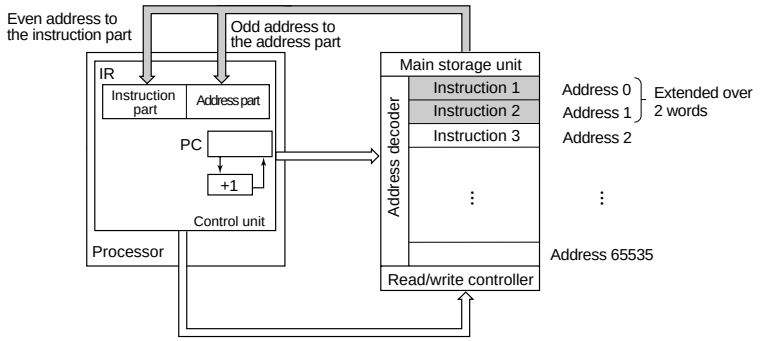
Figure 2-2-7
Three-address format



b. Instruction readout

The instruction received from the main storage unit is stored in the instruction register (IR).
The length of one word of the main storage unit is 16 bits. It is supposed that one instruction has 32 bits in the computers used. One instruction is stored in two words. Therefore, the content of the address of the main storage unit accessed is sent to the processor twice.
In practice, it is determined beforehand that in one instruction the instruction part is stored in an even address while the address part is stored in an odd address.

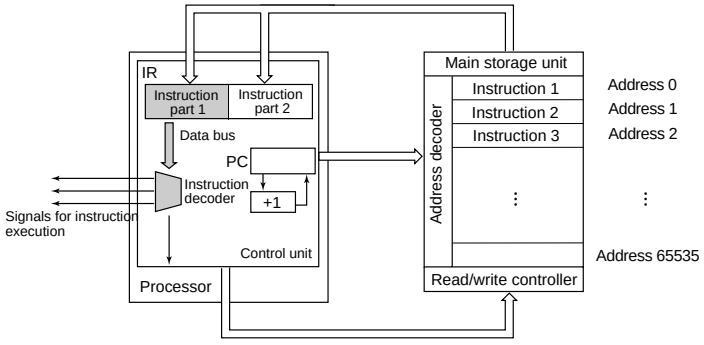
Figure 2-2-8 Instruction loading



c. Instruction decoding

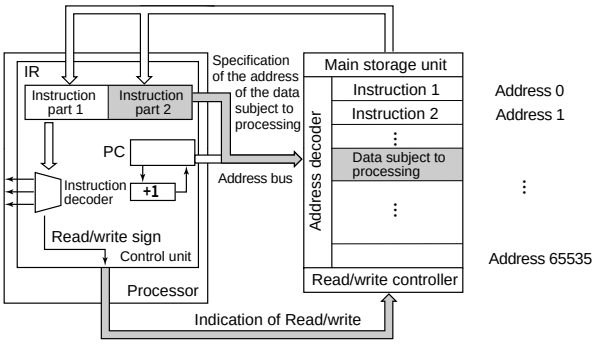
The content of the instruction part of the instruction register is transferred to a device called the decoder. The decoder decodes the type of work indicated by the instruction and sends the signals for the execution of the operation to each unit.

Figure 2-2-9
Instruction decoder



On the other hand, the content of the address part of the instruction register is transferred to the address bus. The content of the address of the main storage unit specified by the address bus corresponds to the data subject to processing. According to the instruction of the instruction part, a read/write signal is sent from the control unit to the data, subject to processing on the main storage unit.

Figure 2-2-10
Address part and data subject to processing



② Instruction execution

Once the instruction content and the address of the data subject to processing are obtained, the instruction is executed. Using the example of assembler language in which there is a one-to-one correspondence with the machine language, the instruction execution control and each type of register will be explained below.

a. Storing retrieved data

If, as a result of decoding the instruction part and the address part using the instruction decoder, the instruction is found to say "Retrieve and transfer to the processor the contents of address 100 of the main storage unit," a place to store the retrieved contents will be needed. Therefore, a general-purpose register is set in the arithmetic unit of the processor in order to store the retrieved data. In this example, it is assumed that there are five registers, and, for convenience, the numbers 0 to 4 will be assigned to them. Then, using the initials of each of the general-purpose registers, they will be represented as GR0, GR1, GR2, GR3 and GR4.

Figure 2-2-11
General-purpose register

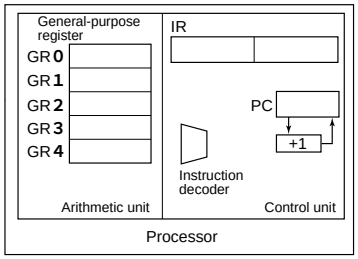
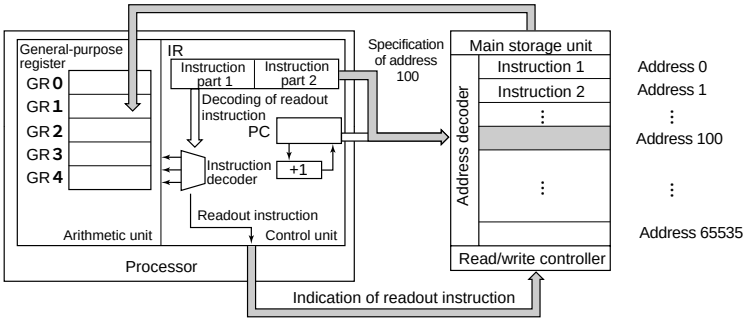


Figure 2-2-12 shows the mechanism by which the contents of address 100 of the main storage unit passes through the data bus to be stored in general-purpose register GR1.

Figure 2-2-12 Storage in the general-purpose registers



b. Instruction execution

If, as a result of decoding the instruction part and the address part of the instruction register, the instruction is found to say "Add the contents of address 100 of the main storage unit to the GR1 contents and store them in GR1," the retrieved contents of address 100 have to be added to the GR1 contents. The unit that performs this kind of addition and subtraction of numeric values is the ALU (Arithmetic and Logic Unit).

The ALU has the following arithmetic mechanism.

- Fixed point operation mechanism to perform operations of integer data
- Floating point operation mechanism to perform operations of floating point data

} For scientific and engineering calculations

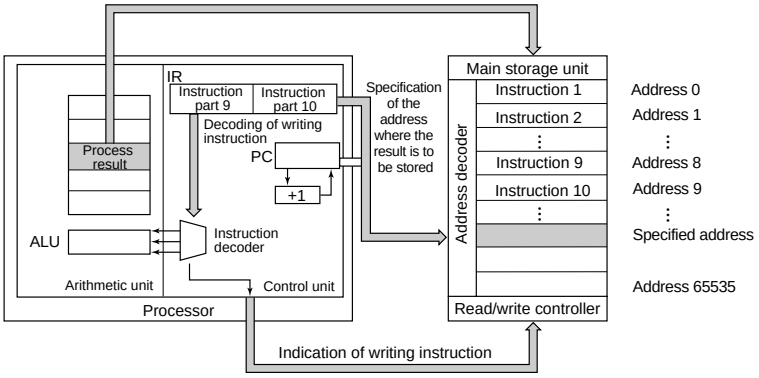
- Decimal operation mechanism to perform operations of binary-coded decimals in packet format ····· For commercial data processing

It should be noted that besides arithmetic operations such as addition and subtraction, the operation mechanism of the ALU performs logical operations such as logical products, logical sums and shifts. For a detailed explanation of the logical operations, please refer to Chapter 1 and Section 2.2.

c. Processing subsequent to the instruction execution

Based on the instructions and data retrieved from the main storage unit, the result of the process performed using the operation mechanism and the register contained in the processor is transferred to the main storage unit through the data bus. Then the address where the result is to be stored is specified by the program instruction.

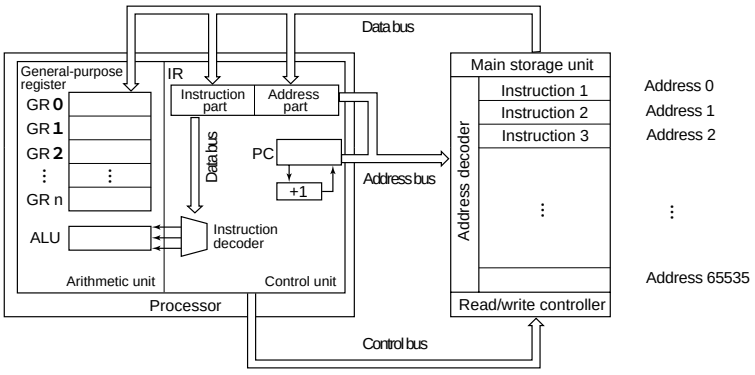
Figure 2-2-13 Storage of the process result



d. Flow of the instruction from its decoding to its execution and hardware structure

Figure 2-2-14 shows the hardware structure from the instruction readout to its decoding and execution.

Figure 2-2-14 Hardware structure



e. Various registers

Up to this point, the roles played by the instruction register, the general purpose register, etc., have been explained, but, besides these registers, the following registers exist:

- Program counter
- Accumulator
- Index register
- Base address register
- Program Status Word (PSW)
- Flag register
- Complement register

● Program Counter (PC)

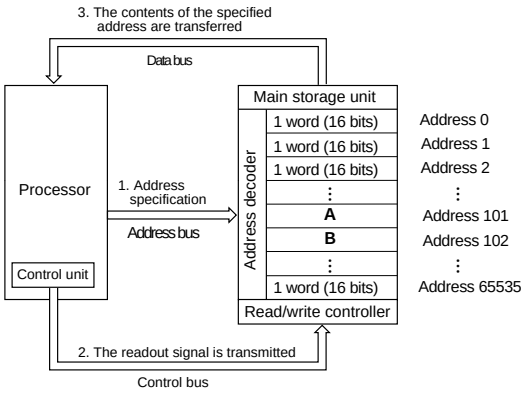
In Figure 2-2-15, considering the procedure in which instruction "A" stored in address 101 of the main storage unit is loaded to the processor, the following can be observed:

1. The processor specifies address 101 with the address bus
2. The control unit sends the readout sign to the main storage unit
3. The main storage unit transfers the contents of address 101 to the processor using the data bus.

At the beginning, the processor specified address 101, but under whose instructions? There is actually a storage unit that exclusively performs this kind of instruction inside the processor. It is called the program counter and is composed of 16 bits. The program counter is also called the instruction address register, instruction counter or sequential control counter.

The devices that, as is the program counter, are set inside the processor and temporarily store data are known generally as registers. Among the registers, there are specialized registers whose application is set in advance, as is the program counter's, and general purpose registers whose application is freely decided by the program.

Figure 2-2-15
Address reading

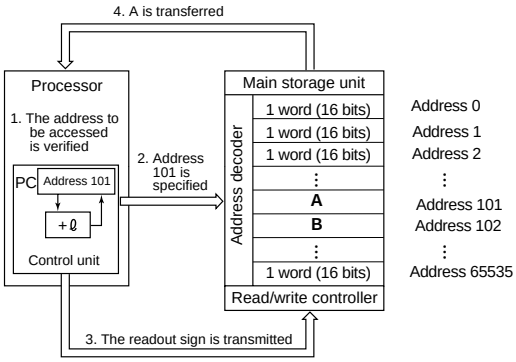


Computer hardware is set in such a way that when the computer is turned on, the content of the program counter is immediately read and the address of the main storage unit to be accessed is verified. Likewise, every time this program counter is referred to, the content stored is automatically "+∝ (instruction word length)."

Taking into account the role of the program counter, and if we consider the procedure in which the processor loads instruction "A" stored in address 101 of the main storage unit, the following can be observed:

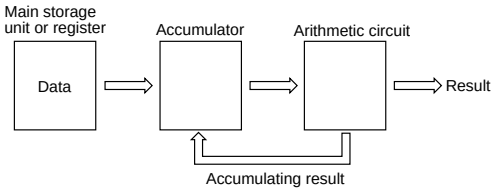
1. The content stored in the program counter is referred to and the address of the main storage unit to be accessed is verified. After it is referred to, "∝" is automatically added to the content of the program counter.
2. The processor specifies address 101 with the address bus
3. The control unit sends the readout signal to the main storage unit
4. The main storage unit transfers the contents of address 101 A to the processor using the data bus

Figure 2-2-16
Role of the program counter



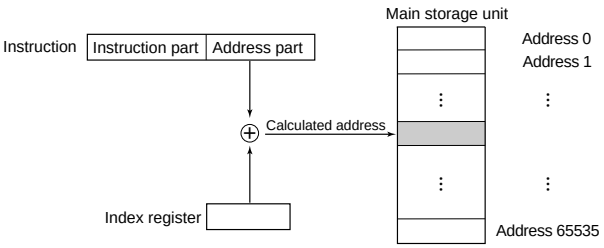
- Accumulator
  The accumulator is the register used exclusively to store operation results and operation values. Since it stores accumulating results, it is also called the accumulating device. There are cases where the general-purpose register is used as a substitute for the accumulator.
  When the accumulator is used, it is called accumulator mode, when the general-purpose register is used, it is called general-purpose register mode.

Figure 2-2-17
Accumulator



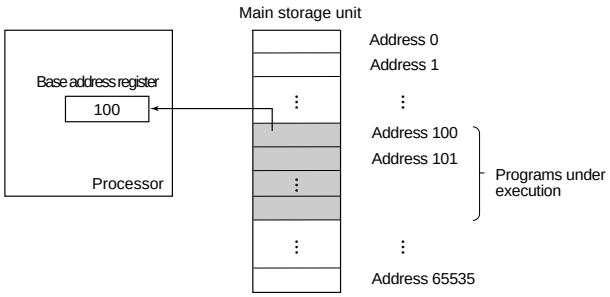
- Index register
  When an address in the main storage unit is specified, the act of changing the address of the address part of the instruction is called address modification. The register used to perform this change is the index register.

Figure 2-2-18
Index register



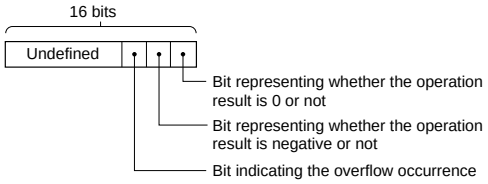
- Base address register
  The base address register is the register that stores the program top address.

Figure 2-2-19
Base address register



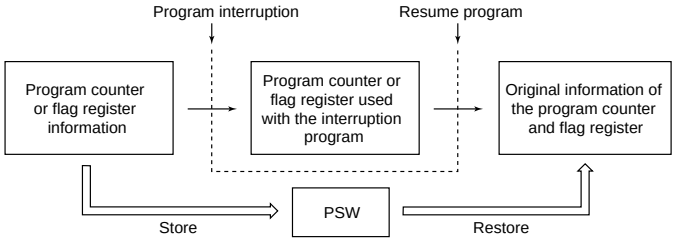
- Flag register
  The flag register stores information related to the operation result (if it is positive or negative or 0), to the existence of carry, overflow, etc.

Figure 2-2-20
Example of a flag register



- PSW (Program Status Word)
  The program counter, flag register and other information are registered in the PSW. In cases where an event that interrupts the program (interruption) in the processor occurs, the program execution can be resumed using the PSW information. Interruption is explained in Chapter 3 Section 3.1.

Figure 2-2-21
PSW



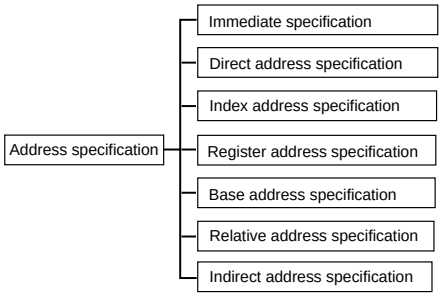
- Complement register
  The complement register generates integer complements in order to perform operations in the addition circuit.

Figure 2-2-22
Complement register

(3) Address specification mode

The address part of the instruction specifies the main storage unit address and the register subject to processing. This specification method is called the address specification method.
At the instruction execution, the value of the address part of the instruction is not used as the address subject to processing as it appears; the actual address is specified after performing calculations between the specified register and addresses. The act of obtaining the address through calculation is called "address modification" and the actual address obtained is called the "effective address." The types of address specification modes are listed in Figure 2-2-23.

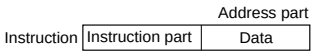
Figure 2-2-23
Types of address specification modes



① Immediate specification

In the immediate specification, the data itself is contained in the address part. Since it is not necessary to access the main storage unit address, it can be immediately executed.

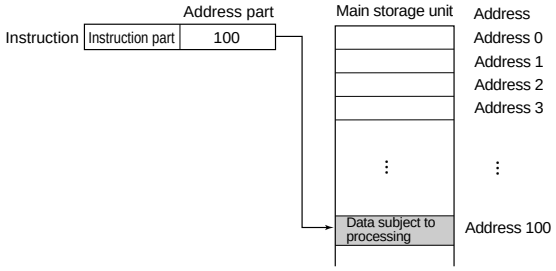
Figure 2-2-24
Immediate specification



② Direct address specification

In the direct address specification, the address of the data subject to processing is contained in the address part.

Figure 2-2-25
Direct address specification

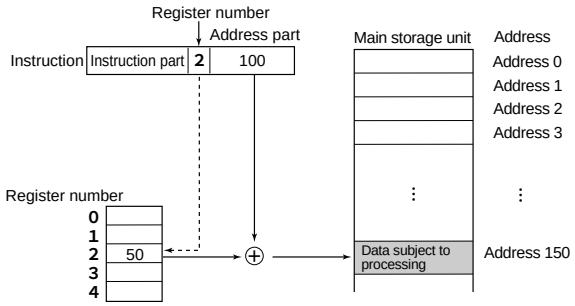


③ Index address specification

In the index address specification, the address part is divided into the section that specifies the number of the index register and the constant section, and the effective address is the result of the following addition:

(Content of the register content specified with the register number) + (Address constant)

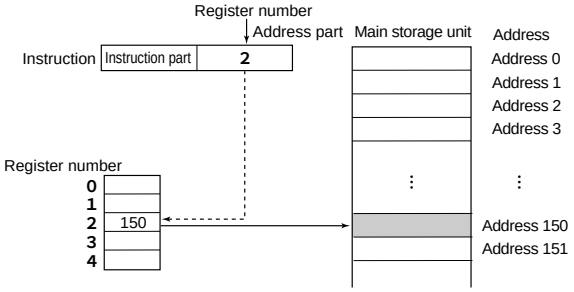
Figure 2-2-26
Index address specification



④ Register address specification

In the register address specification, the register number is stored in the address part and the address is stored in the register of that number.

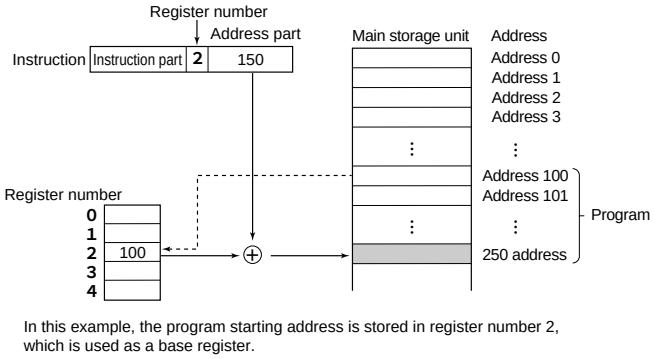
Figure 2-2-27
Register address specification



⑤ Base address specification

In the base address specification, the program starting address is stored in the base register. The result of the addition of the address contained in this base register and the address constant becomes the effective address.

Figure 2-2-28 Base address specification

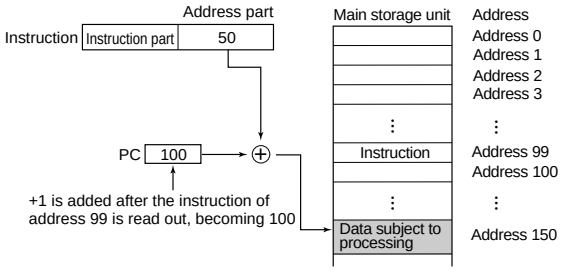


In this example, the program starting address is stored in register number 2, which is used as a base register.

⑥ Relative address specification

In the relative address specification, the result of the addition of the address of the instruction being executed at the present time (value of the program counter) and the address of the address part become the effective address.

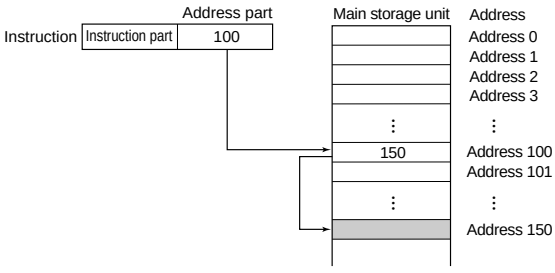
Figure 2-2-29
Relative address specification



⑦ Indirect address specification

In the indirect address specification, the address of the data subject to processing is contained in the address specified in the address part (Figure 2-2-30). There are cases where indirect address specification is performed on two or three levels.

Figure 2-2-30
Indirect address specification



(4) Instruction set

When the computer executes a job requested by the user, the hardware performs the process using several instructions needed from a group of instructions built into that computer. The set of instructions is defined as the interface of the software and the hardware of a specific computer, and, depending on the computer, the types and number of instructions differ. This group of instructions (the total of the instructions) is called the instruction set. As a result, computer software packages with identical instruction sets are basically compatible.

(5) Execution control of the instruction

The execution of a program consists of the repetition of the readout of the instruction from the main storage unit, and the decoding and execution of that instruction by the control unit. If we arrange the operations performed by the processor during program execution, we can divide them as follows:

- Instruction readout
- Instruction execution

① Instruction cycle and execution cycle

The series of operations consisting of the readout of the instruction stored in the main storage unit, its storage in the instruction register contained in the processor, and the instruction decoding using the instruction decoder, is called the instruction cycle. Likewise, the series of operations consisting of the reading out of the data subject to processing from the main storage unit, and writing and executing the instruction, is called the execution cycle.

a. Instruction cycle

The instruction cycle consists of the following two consecutive operations:

- According to the value of the program counter, the instruction to be executed is read out from the address of the main storage unit, where it was stored, to be stored in the instruction register.
- The instruction part of the instruction register is decoded using the instruction decoder, and the address of the data subject to processing is calculated based on the address part of the instruction register.

The instruction cycle is also called I cycle, and it is also called F (fetch) cycle.

b. Execution cycle

The execution cycle consists of the following two consecutive operations:

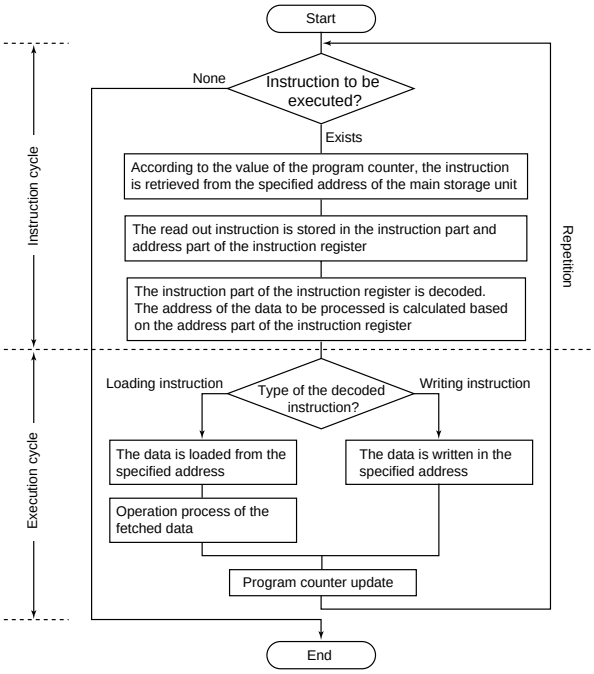
- If the decoded instruction is a readout instruction, the data subject to processing is read out from the specified address of the main storage unit; if it is a writing instruction, data is written in the specified address of the main storage unit.
- In the case of readout instructions, the read out data is used to perform the operation process.

Taking the initial of execution, the execution cycle is also called E cycle.

c. Instruction cycle and execution cycle operations

Figure 2-2-31 shows the flow figure of the operations of the instruction cycle and execution cycle in a clearly understandable way.

Figure 2-2-31 Flow figure of the operations of the instruction cycle and execution cycle



Note: There are cases where the program counter update is performed immediately after the instruction readout, and there are cases where it is performed after the instruction execution is completed.

The processor reads out, in order, the program instructions stored in the main storage unit and executes the program by repeating the instruction cycle and the execution cycle.

(6) Hardwired-logic control and microprogramming control

The instructions set in the computer are executed using logical circuits composed of several logic elements. In other words, logical circuits are the result of hardwiring among the logic elements; for that reason they are called hardwired-logic control system or wiring logic control system. Since the different instructions are implemented by the hardware, they have the advantage that the operation speed is fast. Opposite to this system, there is one that executes the instructions with the firmware.
As computer performance improves, instructions with more complicated functions become feasible. The higher the function level of the instructions, the more complicated the control procedure becomes. Therefore, instead of executing the instructions with hardware of complicated wiring, an execution method using easily modifiable firmware was designed. This method is called the microprogramming control system. Compared to the hardwired-logic control system, the operation speed of the microprogramming control system is slow, but the hardware is simpler and corrections (debugging) are easily performed. The microprogram, which is a string of patterns that determine whether the logic gate is to be on or off, is stored in a special memory called control storage inside the control unit. Instructions are followed by reading out the microprogram sequentially.

2.2.2 Speed performance enhancement in Processor

There are two types of architectures regarding instruction execution. There are

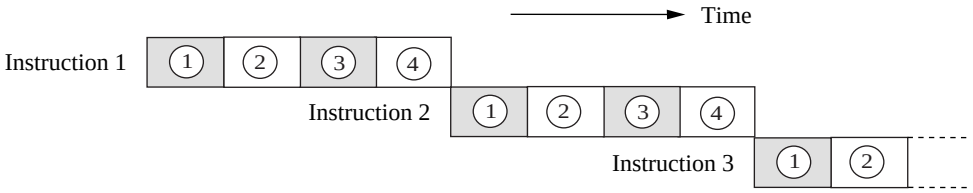
- sequential architecture
- speculative architecture

Another name for pre-fetched control is pipeline. RISC is an example of a processor that uses pipeline processing.

(1) Pipeline processing

Sequential architecture is the process of reading, decoding and execution of the instruction till it is finished before another instruction is fetched.
Program execution steps are
Instruction cycle ①: The instruction is read from the main storage unit
Instruction cycle ②: The instruction is decoded and the address calculated
Execution cycle ③: The reading and writing of the main storage unit is done
Execution cycle ④: The actions specified by the instruction is executed :

Figure 2-2-32 Serial control processing



Instruction cycle ① and execution cycle ③ accesses the main storage unit. Instruction cycle ②and execution ④ is executed within the processor.
The processor is idle when the instruction cycle ①and execution cycle ③ are executing. Conversely, when instruction cycle ②and execution ④ are executing, no access to the main storage unit is done.
The next instruction is read in and queued during this idle time. This increases processing efficiency and this method is known as speculative control method. Pipeline processing utilizes this speculative control method and the next set of instructions is read in and processed in queued to be processing.

Figure 2-2-33 Pipeline processing

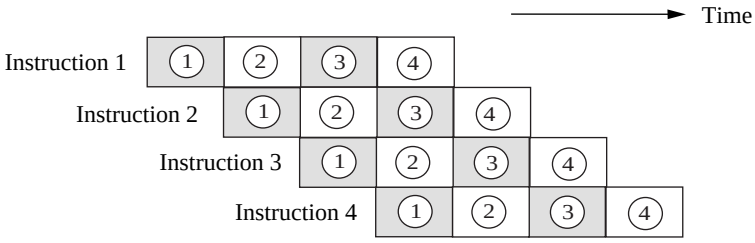


Figure 2-2-33 shows the execution of instructions 1 to 4.
1. Instruction one's instruction cycle ① is executed
2. When instruction one's instruction cycle ② is executing, instruction two's instruction cycle ① is executed
3. When instruction one's execution cycle ③ is executing, instruction two's instruction cycle ②is executed and instruction three's instruction cycle ① is executed
4. When instruction one's execution cycle ④ is executing, instruction two's execution cycle ③ is executed and instruction three's instruction cycle ② is executed and instruction four's instruction cycle ① is executed

In the pipeline processing architecture, when one instruction is being processed, the next instruction is simultaneously read and processing is continued.
This leads to an increased processing speed.
However, the size of the instruction and the execution time should be fixed as the cycles are concurrently executing. The following conditions will lead to degradation in the execution efficiency.

- Fork instructions are issued from within executing instruction. The order of execution of the instructions is changed.
- Program is interrupted and a different program is executed. This phenomenon is known as switching over.

(2) RISC and CISC

The length of one instruction and the execution time for each instruction should be fixed in pipeline processing. A computer designed with a set of simple instructions with each instruction having a fixed length and execution time is known as the RISC (Reduced Instruction Set Computer) computer.
The opposite of a RISC machine is a CISC (Complex Instruction Set Computer) computer. The computer is

designed with sets of complex instructions

Figure 2-2-34 Characteristics of RISC and CISC

RISC	CISC
• Make up of few simple instructions • Instructions are executed by the hardware • The size of the instruction and the execution time of each instruction is about the same	• Complex, high level type instructions • Instructions are executed by the micro-program • There is variation in the instruction size and length of execution

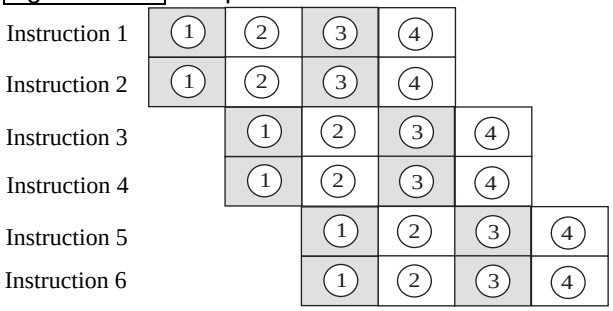
RISC is not optimized or the number of instruction steps is many, Efficient programs are dependent on the optimization functions found in the compiler.
Recently, many workstations are created with the RISC architecture

(3) Parallel method

① Super scalar architecture

Under the best conditions, the normal pipeline architecture is limited to one instruction per cycle. The super scalar architecture allows for multiple instructions to be executed in 1 cycle. Multiple parallel processors in the CPU allow the execution of these multiple instructions in 1 cycle.
Instructions that can executed in parallel have to be separated from the general instructions. The architecture is known as super scalar if on average, if multiple instructions are running in 1 cycle.

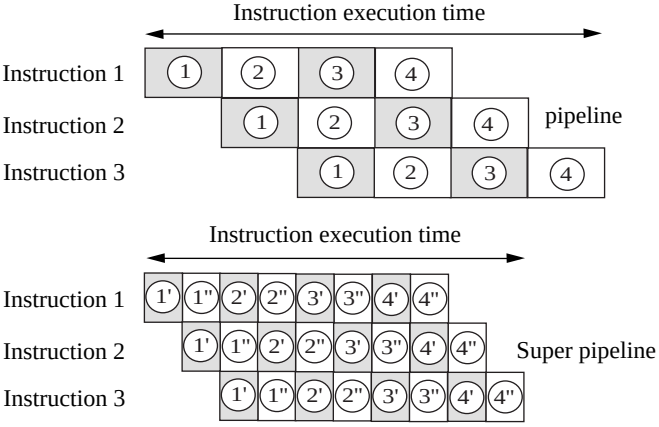
Figure 2-2-35 Super scalar architecture



② Super pipeline architecture

A super pipeline results when more stages are added to the normal pipeline where each stage does less work and leads to a higher speed.

Figure 2-2-36 Super pipeline architecture

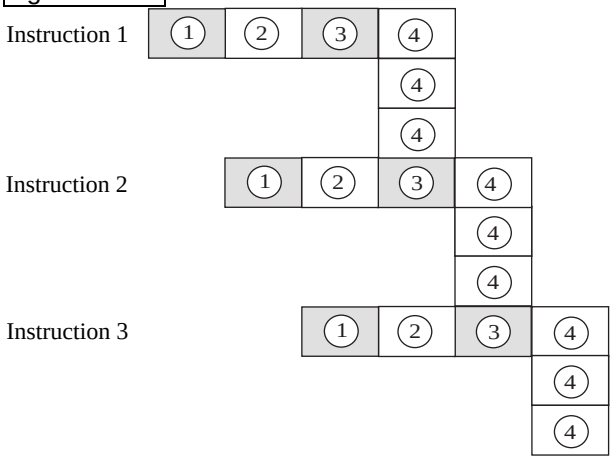


In this architecture, the next instruction starts to be executed before the end of the current instruction. The processing volume is higher than the normal pipeline which means the instructions have a high dependency.
The compiler's sequencing of instructions and the design of a easily processed instruction set is important to ensure a smooth flow of the pipeline

③VLIW (Very Long Instruction Word) method

The action of coding multiple tasks in one instruction instead of having short instruction cycles in pipeline processing is known as the VLIW method. The length of one instruction becomes long due to the multiple tasks that are defined within it.

Figure 2-2-37 VLIW method



2.2.3 Operation mechanism

The procedure of the execution of logic operations studied in Chapter 1 inside the computer will be explained using real logical circuits.
In the arithmetic unit there are numeric operation circuits that handle numeric values and logical circuits that perform logical operations. The following three operations are basic in logical operations:

- Logical product operation (AND operation)
- Logical sum operation (OR operation)
- Negation operation (NOT operation)

Through the combination of circuits that perform these three operations, a wide range of logical circuits is implemented.

Figure 2-2-38
Logical operations

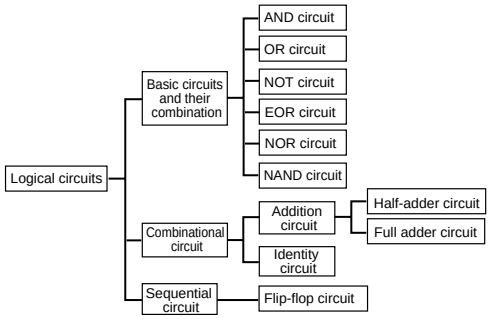
Logical operations	Operation symbols
Logical product (AND)	∧ or ・
Logical sum (OR)	∨ or +
Negation (NOT)	¯ or ¬
Exclusive logical sum (EOR)	∀ or ⊕
Negative logical sum (NOR)	
Negative logical product (NAND)	

Likewise, according to the combination of logical circuits, circuits can be classified as follows:

- Combinational circuits: Addition circuits, identity circuits, etc., that establish the output according to current input.
- Sequential circuits: Flip-flop circuits, etc., that establish the output according to current and past inputs.

The organization of these logical circuits is shown in Figure 2-2-39.

Figure 2-2-39
Logical circuits



(1) Basic logical circuits

The three operations "AND," "OR" and "NOT" are basic logical operations. The circuits that perform these three operations, AND circuits, OR circuits and NOT circuits will be explained below.

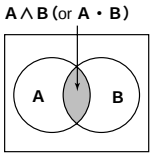
① AND circuit

AND circuits are the circuits that perform AND operations (logical product operations), as the one shown in Figure 2-2-40. In these circuits, if both A and B input values are not "1," "1" is not input. The table shown in Figure 2-2-40 is called the "truth table," and in this case, for the input "1" (true) or "0" (false) of A and B, the operation results "1" or "0" are indicated. Likewise, the Venn diagram, which clearly represents the operation result, is also displayed.
Besides "A AND B," AND operations are represented as "A · B" or as "A ∧ B"

Figure 2-2-40
Truth table and Venn diagram of AND operations

A	B	A AND B
0	0	0
0	1	0
1	0	0
1	1	1

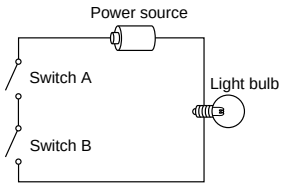
Truth table



Venn diagram

Figure 2-2-41 represents a comparison between the AND circuit that performs AND operations and a switch and a light bulb. Here, by establishing a correspondence between the switch "opens" and "1," and between "closes" and "0," as well as between the light bulb "lights" and "1" and "does not light" and "0," a truth table can be created.

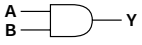
Figure 2-2-41 AND circuit of a switch and a light bulb



Switch A	Switch B	Light bulb
Opens (0)	Opens (0)	Does not light (0)
Opens (0)	Closes (1)	Does not light (0)
Closes (1)	Opens (0)	Does not light (0)
Closes (1)	Closes (1)	Lights (1)

The electric circuit is often represented using the US army MIL symbol (US MILitary standard, MIL-STD). Figure 2-2-42 shows the AND circuit represented with the MIL symbol.

Figure 2-2-42 AND symbol



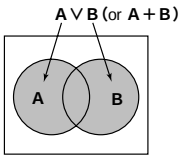
② OR circuit

OR circuits are the circuits that perform OR operations (logical sum operations) as the one shown in Figure 2-2-43. If either the A or B input value is "1," "1" is input.
Besides "A OR B," OR operations are represented as "A + B" or as "A ∨ B"

Figure 2-2-43
Truth table and Venn diagram of OR operations

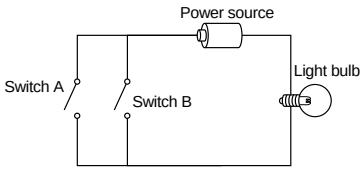
A	B	A OR B
0	0	0
0	1	1
1	0	1
1	1	1

Truth table



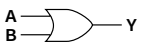
Venn diagram

Figure 2-2-44 OR circuit of a switch and a light bulb



Switch A	Switch B	Light bulb
Opens (0)	Opens (0)	Does not light (0)
Opens (0)	Closes (1)	Lights (1)
Closes (1)	Opens (0)	Lights (1)
Closes (1)	Closes (1)	Lights (1)

Figure 2-2-45 OR symbol



③ NOT circuit

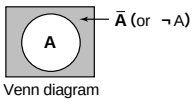
NOT circuits are the circuits that perform NOT operations (negation operations) as the one shown in Figure 2-2-46. The opposite of input value is output. The negation of "1" is "0" and the negation of "0" is "1."

Besides "NOT A," NOT operations are represented as " $\overline{A}$ " or as "¬ A."

Figure 2-2-46
Truth table and Venn diagram of NOT operations

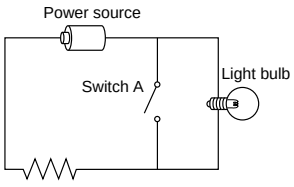
A	NOT A
0	1
1	0

Truth table



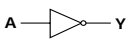
Venn diagram

Figure 2-2-47 NOT circuit of a switch and a light bulb



Switch A	Light bulb
Opens (0)	Lights (1)
Closes (1)	Does not light (0)

Figure 2-2-48 NOT symbol



(2) Combination of the basic circuits

Through the combination of the basic circuits, EOR circuits, NOR circuits and NAND circuits can be composed.

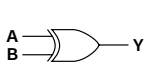
① EOR circuit (Exclusive logical sum operation circuit)

Figure 2-2-49 shows the truth table, MIL symbol and Venn diagram of EOR operations.
Besides "A EOR B," EOR operations are represented as "A ⊕ B" or as "A $\veebar$ B."
It should be noted that "A ⊕ B" is an operation that means the same as " $\overline{A} \cdot B + A \cdot \overline{B}$."

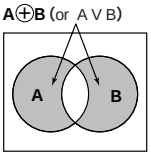
Figure 2-2-49 Truth table, EOR symbol and Venn diagram of EOR operations

A	B	A EOR B
0	0	0
0	1	1
1	0	1
1	1	0

Truth table



EOR symbol



Venn diagram

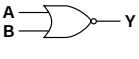
② NOR circuit (Negative logical sum operation circuit)

NOR operations are the negation of OR operations. NOR circuits result of the combination of a NOT circuit in the output side of an OR circuit.

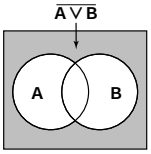
Figure 2-2-50 Truth table, NOR symbol and Venn diagram of NOR operations

A	B	A NOR B
0	0	1
0	1	0
1	0	0
1	1	0

Truth table



NOR symbol



Venn diagram

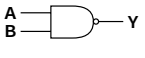
③ NAND circuit (Negative logical product operation circuit)

NAND operations are the negation of AND operations. NAND circuits result in the combination of a NOT circuit in the output side of an AND circuit.

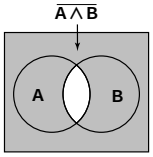
Figure 2-2-51 Truth table, NAND symbol and Venn diagram of NAND operations

A	B	A NAND B
0	0	1
0	1	1
1	0	1
1	1	0

Truth table



NAND symbol



Venn diagram

④ Addition circuit

Through the combination of several basic circuits and combinational circuits, circuits that perform 1-digit binary additions called addition circuits can be created. There are two types of addition circuits, half-adder circuits and full adder circuits.

a. Half-adder circuit (HA)

The addition of the 1-digit binaries A and B, A+B can be performed in the following four ways:

$$\begin{array}{r} 0 \\ +\,0 \\ \hline 0 \end{array} \qquad \begin{array}{r} 0 \\ +\,1 \\ \hline 1 \end{array} \qquad \begin{array}{r} 1 \\ +\,0 \\ \hline 1 \end{array} \qquad \begin{array}{r} 1 \\ +\,1 \\ \hline 10 \end{array}$$

↑
Carry

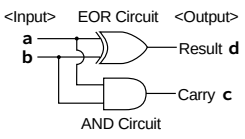
The circuit that performs these binary operations is created by the combination of one AND circuit and one EOR circuit. This circuit is called the half-adder circuit. (Figure 2-2-52)

Figure 2-2-52 Truth table of the half-adder circuit and half-adder circuit

<Addition of two 1-digit values>

$$\begin{array}{r} a \\ +\,b \\ \hline c\ d \end{array}$$

Input		Output	
a	b	c	d
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	0



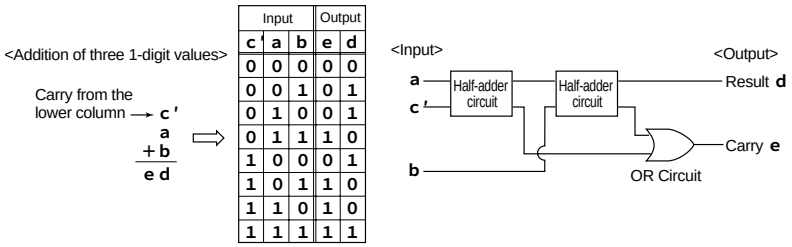
b. Full adder circuit (FA)

The addition of the 1-digit binaries A, B and C, A+B+C can be performed in the following eight ways:

0	0	0	0	1	1	1	1
0	0	1	1	0	0	1	1
+ 0	+ 1	+ 0	+ 1	+ 0	+ 1	+ 0	+ 1
0	1	1	10	1	10	10	11

The circuit that performs these binary operations is created by the combination of two half-adder circuits and one OR circuit. This circuit is called the full adder circuit.

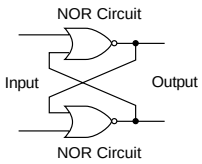
Figure 2-2-53 Truth table of the full adder circuit and full adder circuit



(3) Sequential circuit

The sequential circuit is a circuit in which the output is established according to the current input and the status preserved (past input). The sequential circuit, whose status changes with time, is composed of a flip-flop circuit and is used in registers, etc.

Figure 2-2-54
Flip-flop circuit



2.2.4 Multi-processor

Multi-processor systems are introduced to improve the performance and reliability of the system. Multiple processors in parallel with each processor having a dedicated function. When failure occurs, the processor will do a switch and the remaining processors will distribute the load among themselves.

(1) Symmetric Multi-processor

Symmetric multi-processors systems are systems where the memory is shared among all the processors executing the same OS. Competition for the use of memory among the processors since the memory is common to all. The means a large number of processors cannot be connected.
Message passing distributed memory multi processor systems are systems where each processor has its own private block of memory. A high speed input output port is used to transfer the data between the different blocks.

(2) Array processor

High speed scientific computing is done with array processors which using pipeline processing.
Large scale or dedicated mathematical processors

The sub units' acts are in a queue passing the result to the next unit after it has finished its part. This is known as vector processing. Most supercomputers utilized this method of high speed computing.

(3) Parallel

Multiple processors cooperate with multiple tasks being performed to execute one job.

SISD (Single Instruction Single Data Stream)

One instruction stream operating on a single data element and is not parallel

SIMD (Single Instruction Multiple Data Stream)

Each instruction may operate on more than one data element and is synchronous.

Parallel SIMD

The same instruction is executed by all processors operating on different sets of data.

MIMD (Multiple Instruction Multiple Data Stream)

Each processor has its own instruction stream acts on its own data stream independent of the other processors

2.2.5 Processor performance

The performance of the processor, which can be considered as the central nervous system of the units that compose the computer system, is measured using the number of instructions that can be executed in a unit of time as an index. These indexes are indicated below.

(1) MIPS

MIPS is an acronym of Million Instructions Per Second, and indicates in million units the number of instructions that can be executed in one second. In other words, a 1 MIPS processor is a processor that can execute one million instructions per second. Basically, the larger the number of instructions that can be executed, the higher the value. The term MIPS is mainly used to indicate the performance of processors of high-end mainframe computers. However, it is meaningless to use this index to compare processors of different types of machines that execute different instruction contents.

(2) Clock

In order to set the pace in which the micro-instructions, which are basic operations, are executed, the processor has a clock inside. A quartz crystal oscillator that pulses in regular intervals when electrical current passes through is used in this clock. The time taken for this oscillator to pulse once (one cycle) is called clock. The basic operations of the processor are performed according to this clock. The number of clocks varies according to the instruction.

The clock reciprocal number is called clock frequency. Clock frequency is used as an index to measure the performance of a personal computer.

Example Performance of a processor with a clock frequency of 500 MHz

500 MHz = 500×10^6 Hz = 500,000,000 Hz (times/second); 500 hundred million pulses per second

$\frac{1}{0.5 \times 10^9} = 2 \times 10^{-9} = 2$ nano (seconds/times); 1 pulse for every 2 nanoseconds

(3) CPI (Cycles Per Instruction)

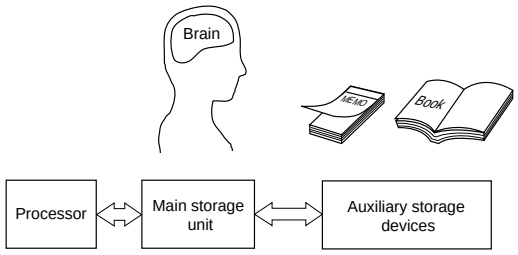
A CPI is the number of clocks required to execute one instruction. This index indirectly indicates the execution time of one instruction.

2.3 Memory architecture

2.3.1 Memory types

The storage function, which is the most important characteristic of the computer, is made possible by the storage units. According to the application, the storage units are divided into main storage unit and auxiliary storage devices.

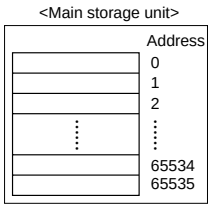
Figure 2-3-1
Main storage unit and auxiliary storage devices



(1) Main storage unit

The main storage unit, which is directly connected to the processor by a signal line called bus, is a device that stores the programs and data to be executed by the processor.
The place where data is stored is generally called memory and it has a series of consecutive addresses. By specifying one of these addresses, information is retrieved from or stored in the main storage unit. The act of exchanging data with the main storage unit performed in the above-mentioned cases is called access. In this way, the system that stores in advance the program to be executed in the main storage unit is called the stored-program system (or internally programmed system) and it is a basic approach in current computers.

Figure 2-3-2
Main storage unit address concept



The main storage unit is composed of semiconductor elements called RAMs (Random Access Memory). A RAM's property is such that when the computer is turned off, the contents stored are lost. Therefore, if there is data to be saved after the process is finished, it is stored in the auxiliary storage devices.

(2) Auxiliary storage devices

The auxiliary storage devices are devices that play the supporting role of making up for the shortage of storage capacity of the main storage unit. They are also called external storage units. The auxiliary storage devices are not included among the five main units of the computer, but they are indispensable for current computers. Large volumes of data and programs are stored/saved in the auxiliary storage devices and when the data or program required to perform a process is not found in the main storage unit, it is transferred (copied) from the auxiliary storage devices to the main storage unit in order to perform the process. Likewise, the data to be saved after the process is completed is transferred to the auxiliary storage devices and saved there. Since the auxiliary storage devices have the property that, even when the computer is

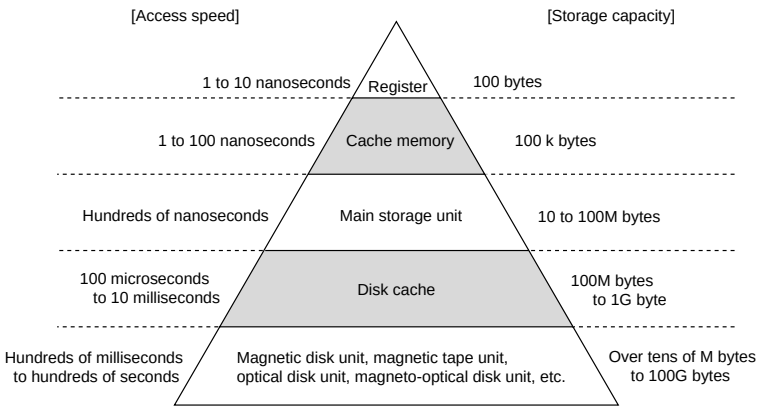
turned off, the contents stored are not lost, they can be used as input/output units of large volumes of data.
- The following are the main auxiliary storage devices.
- Magnetic tape unit
- Magnetic disk unit
- Floppy disk unit (Flexible disk unit)
- Optical disk unit
- Magneto-optical disk unit

2.3.2 Memory capacity and performance

(1) Memory hierarchical structure

The computer memory is composed of the register inside the processor, the main storage unit, the auxiliary storage devices, etc. The storage capacity and the processing speed of each of these devices differ. As is shown in Figure 2-3-3, the access speed is as follows:
(High speed) Register inside the processor > Main storage unit > Auxiliary storage devices (Low speed)
This access speed difference is absorbed by a device called the buffer.

Figure 2-3-3 Memory hierarchical structure



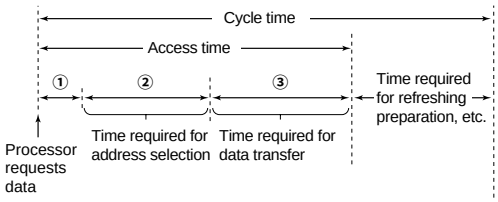
(2) Access time

The access time and cycle time indicate the operation speed of the storage units.
The access time is the time elapsed from when the processor sends the read/write instruction to the storage unit until the data delivery/acceptance is completed.
For the processor to access the main storage unit data, the following three stages are necessary:
① The time during which the processor requests the data readout
② The time during which the processor selects the main storage unit address with the address bus
③ The time during which the data of the selected address is transferred through the data bus.
In other words, ①+②+③ represent the time elapsed from when the data access request is sent until the data transfer is completed. This lapse of time is called the access time.

(3) Cycle time

Among the storage elements of the storage unit, when data is to be stored in the capacitor, there are some whose memory fades with time, as the DRAM. In this case, the refreshing operation that rewrites data at regular intervals becomes necessary. For that reason, after the data transfer is completed, a preparation time in order to receive the next request becomes necessary. The lapse of time that includes the point up to this preparation is called cycle time.

Figure 2-3-4
Access time and cycle time



2.3.3 Memory configuration

As was mentioned above, the memory used in the computer can be classified into hierarchies. To provide for the occurrence of malfunctions or failures, these devices are equipped with data error detection and error correction functions. These functions are implemented by several Error Correcting Codes (ECC).

(1) Magnetic disk

The series of errors caused by a small scratch, etc., on a magnetic disk are called burst errors. The Cyclic Redundancy Check code (CRC code) is adopted in the disk unit to detect these burst errors. Error detection is possible with the CRC code system.

(2) Magnetic tape

The magnetic tape indicates 1-byte data in the transverse direction of the tape. In order to detect bit errors in this transverse direction, a parity check system, which can detect odd numbers of bit errors by appending vertical parity bits, is adopted. In addition, CRC code is adopted to detect burst errors in a transverse direction.

(3) Main memory

In the main memory, due to the high probability of the occurrence of non-consecutive random errors, the Hamming code, which can detect single-bit errors and double-bit errors, is adopted.
It should be noted that, generally speaking, the main memory error detection is performed in general-purpose computers but it is not performed in personal computers.

(4) Memory protection system

Since various information is handled in a computer, depending on the characteristics of that information, a function that limits the users is necessary. This function is called memory protection and it protects the instructions and data stored in the main memory, auxiliary storage devices and other memories under specific conditions. When the memory is accessed, operations like the ones mentioned below are performed:

- Read
- Write
- (Instruction) execute

The right to perform these operations is called access right. Data has read/write rights, but instructions do not have a right to write. On the other hand, instructions have a right to execute, but data do not. When an illegal access that violates these access rights occurs, the control is transferred to the OS as a result of the interruption handling routine.
Likewise, as protection mechanisms of the main memory implemented by the hardware, the protective boundary register system, in which a dedicated register specifies the accessible domains, the TLB (Translation Look-aside Buffer) system, which applies the memory protection function in a virtual address space, etc., exist.

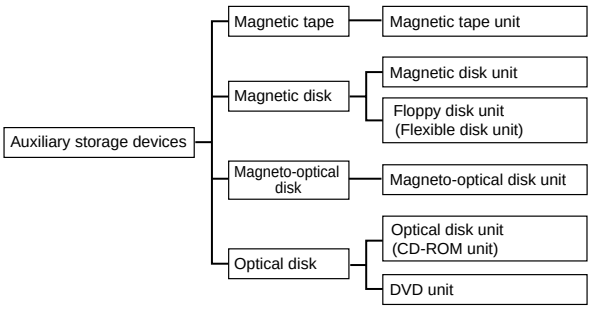
2.4 Auxiliary storage devices

2.4.1 Types and characteristics of auxiliary storage devices

As was mentioned above, computer storage units are divided as follows:
- Main storage unit
- Auxiliary storage devices

The main storage unit is equivalent to the human brain, while auxiliary storage devices are equivalent to notebooks and texts. Auxiliary storage devices are devices that store and save programs and data while they are not being executed. Likewise, as when one reads a text and writes down the necessary information, or when one writes in a letter the things to be transmitted to another person, these auxiliary storage devices also play the role of input devices and output devices.
There are two types of auxiliary storage device: devices that store data magnetically, as the magnetic tape unit, the magnetic disk unit, the floppy disk unit (flexible disk unit), and the magneto-optical disk unit, and devices that store data optically as the optical disk unit (Figure 2-4-1).
The main storage unit stores programs and data to be used by the processor for instruction execution, but it has a big problem: stored content is lost when the computer is turned off. On the other hand, compared to the main storage unit, the operating speed of auxiliary storage devices is low, but they can store a large volume of data and, even if the computer is turned off, the stored data is retained semi-permanently.
It should be noted that besides these devices, there is also a semiconductor disk unit that is, for example, the flash memory used as an auxiliary storage unit in digital cameras and notebook personal computers. This unit is composed of semiconductor(s) (EEPROM(s)), it does not operate mechanically, and it electrically performs data reading/writing processes at high speed. However, since it cannot store large volumes of data, it is used as the storage unit of small devices with low power requirements.
Here, the operation principles and characteristics of the typical auxiliary storage devices will be explained.

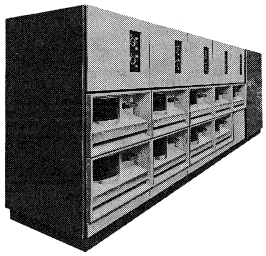
Figure 2-4-1
Diverse auxiliary storage devices



(1) Magnetic disk unit

Magnetic disk unit is devices that store data using magnetic disks. It is the auxiliary storage device most widely used in today's computer systems. Magnetic disks for personal computers or workstations are also called fixed disks or hard disks but the mechanism is the same.

Figure 2-4-2
Magnetic disk unit

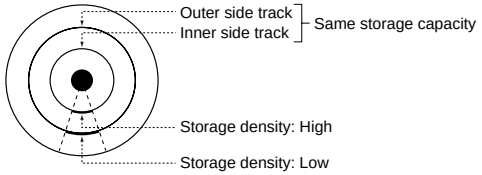


① Magnetic disk

a. Track

The magnetic disk is a circle-shaped magnetic body in which data is recorded along rings called tracks. There are several tracks concentrically set on the magnetic disk. The length of the outer tracks and that of the inner tracks differ, because of the difference of the storage capacity, the volume of data stored is the same in every track. (Figure 2-4-3).

Figure 2-4-3
Data recording side of the magnetic disk

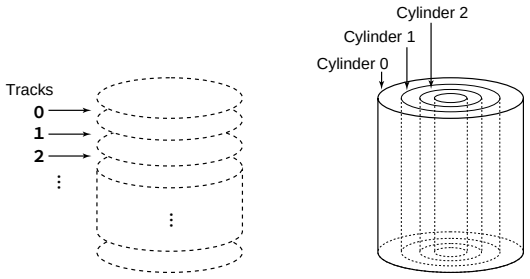


The storage density of the disk is based on the average track length and the storage capacity of the magnetic disk is determined by the number of tracks and the storage density of one disk.

b. Cylinder

In a magnetic disk unit, which is composed of multiple magnetic disks, the group of tracks with the same radius on each of the disks is set as one data storage area. This storage area is called a cylinder. When data is stored in a cylinder, if, for example, the data cannot be completely stored on track 0 of cylinder 1, it can be stored on track 1, track 2, etc. of the same cylinder. Therefore, since data access can be performed without moving the access arm (that is, the magnetic head), it is extremely efficient. To put it in another way, the cylinder is a group of tracks that can be read and written by multiple magnetic heads if the access arm of the magnetic disk unit is fixed.

Figure 2-4-4
Tracks and cylinders



c. Storage capacity

The storage capacity of the magnetic disk can be determined as follows:
Storage capacity of 1 track × Track number of 1 cylinder × Cylinder number of the magnetic disk

Example

Given a magnetic disk with the following specifications, the storage capacity of this magnetic disk is calculated:

[Magnetic disk specifications]

- Cylinder number: 800 cylinders
- Track number/cylinder number: 19 tracks
- Storage capacity/track: 20,000 bytes

The storage capacity per cylinder is as follows:

20,000 bytes/track × 19 tracks/cylinder = 380,000 bytes/cylinder = 380 kB (kilo bytes)

Since the number of cylinders on this disk is 800, the storage capacity of the magnetic disk is as follows:

380 kB/cylinder × 800 cylinders = 304,000 kB = 304 MB (Mega bytes)

An example of the calculation of storage capacity when blocking is performed is shown below.

Example

Given a magnetic disk with the following specifications, the number of cylinders required when 80 thousand records of 200 bytes each are stored in a sequential access file of 10 records/block per magnetic disk is calculated. It should be noted that block recording cannot be extended over multiple tracks.

[Magnetic disk specifications]

- Cylinder number: 400 cylinders
- Track number/cylinder number: 19 tracks
- Storage capacity/tracks: 20,000 bytes
- Inter-block gap (IBG): 120 bytes

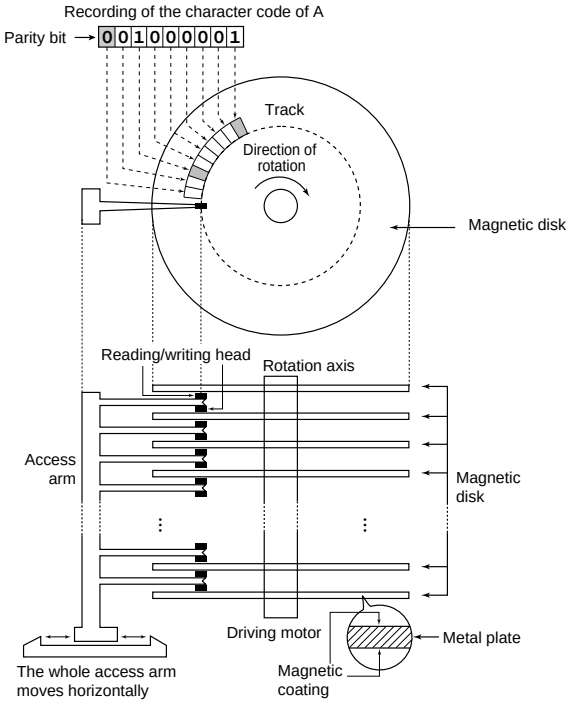
1. First, the number of blocks of the whole file is calculated.
   Since the number of records is 80,000 and the blocking factor is 10, the number of blocks is determined as follows:
   80,000 records ÷ 10 records/block = 8,000 blocks
2. The length of 1 block, including the inter-block gap is calculated.
   200 bytes/record × 10 records/block + 120 bytes/block = 2,120 bytes/block
3. The number of blocks that can be recorded in 1 track is calculated.
   20,000 bytes/track ÷ 2,120 bytes/block = 9.43… blocks/track
   Since a block cannot be recorded across multiple tracks, the decimals are omitted, and the number of blocks that can be recorded in 1 track becomes 9 blocks/track.
4. The number of tracks required for the whole file is calculated.
   8,000 blocks ÷ 9 blocks/track =888.88… tracks
   Rounding it up to the next whole number, it becomes 889 tracks.
5. The number of cylinders required to record the whole file is
   889 tracks ÷ 19 tracks/cylinder = 46.78…cylinders
   Rounding it up to the next whole number, it becomes 47 cylinders.

② Magnetic disk unit structure and operation principles

The magnetic disk unit has multiple magnetic disks, which it rotates at high speeds in order to record data along concentric tracks. On each recording side, an access arm with a magnetic head moves forward and backward to reach the track position where data is to be read or recorded.

Compared to the sequential access of the magnetic tape unit, in which access can only be performed in order from the beginning, in the magnetic disk unit, besides sequential access, direct access to the desired recording position can also be performed. Auxiliary storage devices in which this direct access can be performed are called direct access storage devices (DASD).

Figure 2-4-5 Structure of the magnetic disk unit

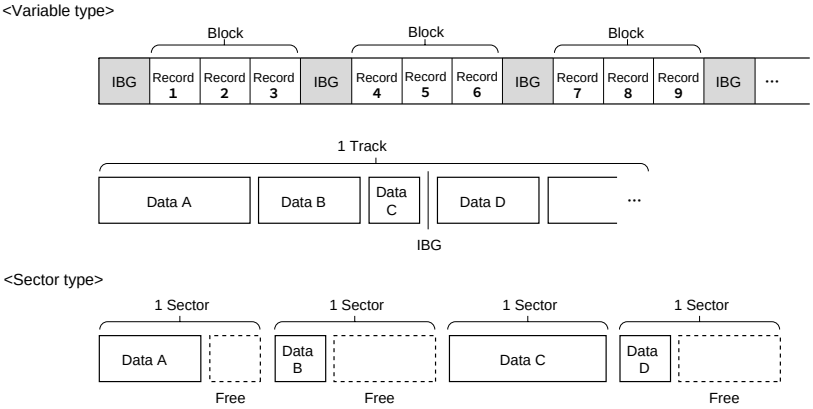


a. Variable type and sector type

By recording method, magnetic disk unit is classified into "Variable type" and "Sector type."

- Variable type
  In the variable type unit, data reading and writing is performed on a block basis, as in the magnetic tape. A block is a group of data called a record and there is an IBG between blocks. In the unit, there is no gap between data (or IBG). Reading and writing of any number of bytes can be started from any track position.
- Sector type
  In the sector type unit, one track is divided into approximately 20 small units called sectors. Data reading and writing is performed on a sector basis. The reading/writing position is specified with the sector number of the selected track.

Figure 2-4-6 Variable type and sector type



Generally, the variable type is used in magnetic disks, and the sector type is used in floppy disks and hard disks.

b. Parity check

When data is recorded on a magnetic disk, data is written on the track bit by bit using the magnetic head. The same method is used to read data. When this process is performed, as in the magnetic tape, in order to detect reading or writing errors, a parity bit (1 bit) is appended to perform the parity check.

c. Defragmentation

In personal computer hard disks, data is stored and deleted repeatedly. Since it is improbable that all the data to be stored will have the same size, a small volume of data can be stored after a big volume of data is deleted, or vice versa. As a consequence, there would be free sectors scattered about and a drop in access efficiency. This status is called fragmentation; in order to solve it, a function called defragmentation is implemented in the OS.

③ Magnetic disk unit performance

The performance of the magnetic disk unit is measured according to access time and storage capacity. Since the storage capacity was already explained in ①, here, the access time significance and the calculation method will be explained.

a. Access time

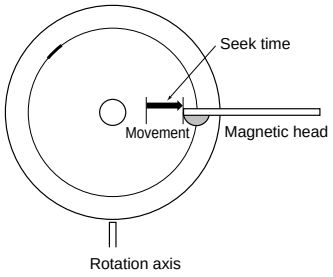
Access is the generic term for the act of reading specific data from the magnetic disk and writing it on a specific cylinder or track. Access time is calculated through the addition of the following:

- Seek time
- Search time
- Data transfer time

● Seek time

In order to access the target data, the magnetic head has to be moved to the position of the track where the target data is stored. The time it takes to move the magnetic head is called seek time.

Figure 2-4-7
Seek time

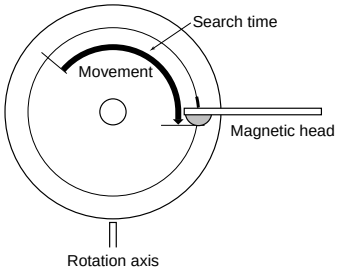


Since seek time differs depending on the distance between the position of the target track and the current position of the magnetic head, an average value is used as the actual seek time. This value is called average seek time.

● Search time

The search time is the lapse of time until the target data reaches the magnetic head position. (Figure 2-4-8).

Figure 2-4-8
Search time



As with average seek time, depending on the data position, there are cases where the search time is 0, as well as cases where there is a wait of 1 revolution. Therefore, 1/2 revolution of the magnetic disk is used as the search time. This value is called average search time.

- Data transfer time
  The time elapsed between when the magnetic head data access starts and when the transfer is completed is called data transfer time.

Therefore, the time elapsed between when the magnetic disk unit starts the data access and when the data transfer is completed, that is, the access time, is calculated as follows:
Seek time + Search time + Data transfer time
Strictly speaking, as in the above-mentioned formula, the time elapsed between when the access request occurs and the magnetic disk unit starts operating is the access time.

Access time of the magnetic disk unit = Average seek time + Average search time + Data transfer time

Example

Given a magnetic disk unit with the following specifications, the access time of this magnetic disk when a record of 9,000 bytes is processed is calculated.

[Magnetic disk unit specifications]

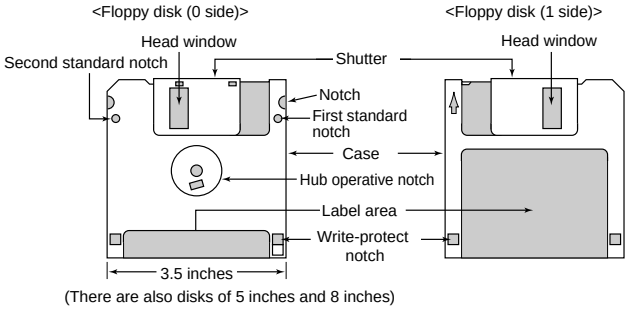
- Capacity per track: 15,000 bytes
- Magnetic disk rotation speed: 3,000 revolutions/minute
- Average seek time: 20 milliseconds

1. First, the average search time is calculated.
   Since the rotation speed of the magnetic disk is 3,000 revolutions/minute, through the following operation,
   3,000 revolutions/minute ÷ 60 seconds/minute = 50 revolutions/second,
   it is determined that the magnetic disk makes 50 revolutions per second. Therefore, the time required to make 1 revolution is as follows:
   1 revolution ÷ 50 revolutions/second = 0.02 seconds/revolution = 20 milliseconds
   Since the average search time is the time required to make 1/2 revolution, it is as follows:
   20 milliseconds ÷ 2 = 10 milliseconds
2. Since in 1 revolution, the information contained in 1 track passes through the magnetic head, considering that the disk makes 50 revolutions per second, the data transfer speed is as follows:
   Data transfer speed = 50 tracks/second × 15,000 bytes/track = 750×10^3 bytes/second
   Based on this data transfer speed, the time to transfer 9,000 bytes of data can be calculated as follows.
   $(9 \times 10^3$ bytes) ÷ (750×10^3 bytes/second) = 0.012 seconds = 12 milliseconds
3. Therefore, the access time is as follows:
   Average seek time + Average search time + Data transfer time
   = 20 milliseconds + 10 milliseconds + 12 milliseconds = 42 milliseconds

(2) Floppy disk unit

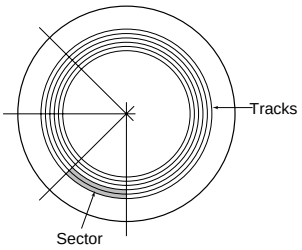
The floppy disk unit is also called a flexible disk unit. In floppy disk units data random access is possible, and, since the floppy disk itself, which is a storage medium, is low-priced and easy to carry about, its use has widely spread. As an auxiliary storage device of personal computers, it is the most ordinarily used device.

Figure 2-4-9
Floppy disk
(Flexible disk)



The recording method of the floppy disk is the sector method, and as it is shown in Figure 2-4-10; the track is divided into sectors, and the data is recorded on a sector basis.

Figure 2-4-10
Data recording side of the floppy disk



① Floppy disk

a. Types

Among floppy disks, there are magnetic disks that measure 8 inches, 5 inches and 3.5 inches, but, the most common disks today are 3.5 inch-disks, while 8- and 5-inch disks are almost never used. There are also the following 2 types of 3.5-inch floppy disks, depending on the storage density.

- 3.5 inch 2 HD (double side High Density)
  Storage capacity: 1.2 to 1.4 megabytes (MB)
- 3.5 inch 2 DD (double side Double Density)
  Storage capacity: 640 to 730 kilobytes (kB)

Figure 2-4-11
Example of the specifications of a floppy disk (2HD)

	1.4MB	1.2MB
Sides available for use	2	2
Track number/side	80	77
Sector number/track	18	8
Storage capacity (B)/sector	512	1,024

Likewise, there is a floppy disk whose storage capacity is 120MB (UHD) and a disk called Zip whose storage capacity is 100MB. Both of them are compatible with the 3.5-inch disk (2DD/2HD), but they have not come into wide use.

b. Storage capacity

The calculation of the access time of floppy disk units is the same as that for magnetic disk units. Therefore, here, the storage capacity of the sector method will be explained.
As was shown in Figure 2-4-11, among floppy disks, the sides available for use, the number of tracks per side, the number of sectors per track, etc. differ.
The storage capacity of a floppy disk is calculated using the following values:
Storage capacity per sector × Number of sectors per track × Number of tracks per side × Number of sides (One side or both sides)

Example

Given a floppy disk with the following specifications, the storage capacity is calculated.
[Specification of a floppy disk unit]

- Sides available for use: 2 sides
- Track number/side: 80 tracks
- Sector number/track: 9 sectors
- Storage capacity/sector: 1,024 bytes

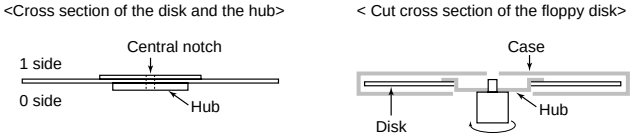
The storage capacity of 1 track is as follows:
1,024 bytes/sector × 9 sectors/track = 9,216 bytes/track
Therefore, the storage capacity of 1 side is as follows:
9,216 bytes/track × 80 tracks = 737,280 bytes ≤ 737kB
And, since the sides available for use of the floppy disk are 2 (sides), the following is the storage capacity:

737kB × 2 = 1,474kB
Approximately 1.474MB

② Floppy disk unit structure and operation mechanism

Basically, the floppy disk unit has the same structure as the magnetic disk unit. However, only 1 floppy disk is used and the sector method, in which the track is divided into sectors, is the recording method.
The number of tracks and the division of the sectors depends on the operating system used. Therefore, when the user uses a floppy disk, it has to be initialized in the format specified by the operating system. This process is called formatting.
When the floppy disk cartridge is installed in the floppy disk unit, the disk contained in the cartridge rotates. The magnetic head directly traces the magnetic disk surface of the disk and, in order to read/write information, the data access time is longer than that of the magnetic disk unit and the magnetic tape unit.

Figure 2-4-12 Floppy disk structure



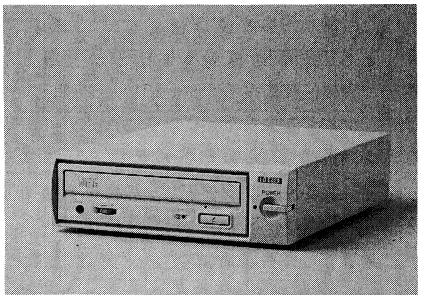
(3) Optical disk (CD, DVD) unit

Besides the magnetic disk unit, the magnetic tape unit and the floppy disk unit, there are various other kinds of auxiliary storage devices. Optical disk units, magneto-optical disk units, DVD units, etc. are used to store/save image processing data of extremely large volume or as storage devices of large volume packaged software. These devices can store large volumes of data through a mechanism that reads out information using light reflection.
In addition to floppy disk units and hard disk units, as auxiliary storage devices, today's standard personal computer systems are also equipped with CD-ROM units. The role played by the CD-ROM as a medium to supply software packages to the general marketplace, and as a multimedia storage medium, is extremely important.

① Optical disk

The surface of the optical disk is covered with a hard plastic that makes it resistant to scratches and dust. Furthermore, since a laser beam is used to read out data, the head does not touch the recording surface directly, so no friction is caused. Among optical disks, CD-ROM use in particular is expanding rapidly.

Figure 2-4-13
CD-ROM unit



Among optical disks, there is the music CD (Compact Disc), the CD-G (CD-Graphic) for image data, the CD-I (CD-Interactive) for interactive applications, the CD-R (CD-Recordable), etc. And, as a computer storage medium, the CD-ROM (CD-Read Only Memory) is widely used.
Furthermore, as an optical disk that supports the multimedia era, the DVD, which has great capacity and high image quality and is capable of storing animated images and audio, exists.
Here, the CD-ROM and the DVD specifically, will be explained.

② CD-ROM

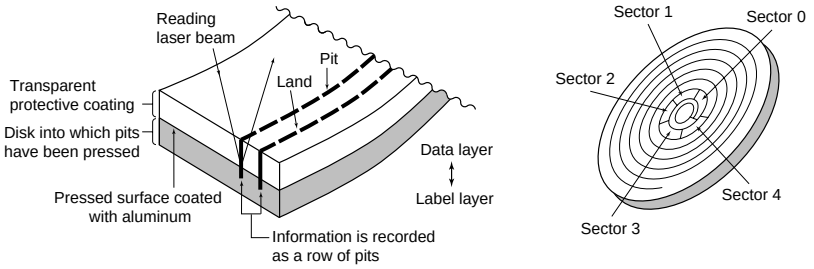
The CD used as a computer storage medium is the CD-ROM. The external appearance, diameter, thickness, etc. of the CD-ROM is the same as that of a CD (diameter: 12 cm, thickness: 1.2 mm, single

disk diameter: 8 cm), but the error correction function, file system, logical format, etc. differ. Since the CD-ROM logical format uses the international industrial standard ISO 9660, it has high compatibility. However, since the CD-ROM is a read-only disk, data can be read out but cannot be written.

a. Structure

The CD-ROM, which is a disc-shaped storage medium, does not have a concentric-track structure as does the magnetic disk or floppy disk. Tracks of continuous sectors are connected in a spiral as in vinyl records and the data is stored from the inner side to the outer side. Figure 2-4-14 shows a magnification of the data recording surface of the CD-ROM.

Figure 2-4-14 CD-ROM data recording surface



The CD-ROM stores "0" and "1" information using the pits and lands of the data recording layer. In order to read out data, a laser beam is applied and the optical head reads out the changes in intensity of the reflected light.

b. Storage capacity

By creating a master disk with a negative replica of these pits and lands and pressing it against plastic disks, a large quantity of CD-ROMs can be replicated at high speed and low cost.
1 CD-ROM (12 cm) has a storage capacity of approximately 600MB, which makes it an indispensable storage medium to process the enormous volume of information of multimedia data.

③ CD-ROM unit structure and performance

a. Structure

Basically the structure of the CD-ROM unit is the same as that of the magnetic disk unit. The difference is that data is not read out using a magnetic head, but an optical head that detects the laser beams.

b. Performance

The CD-ROM unit performance is measured according to the head seek time and the data transfer rate.

- Seek time
  The CD-ROM seek time is extremely slow compared to that of the magnetic disk unit. While the seek time of the magnetic disk unit is measured in tens of milliseconds, it is measured in hundreds of milliseconds in the CD-ROM unit. This is due to the use of a heavy lens in the read head and to the CD-ROM structure (the data storage format uses a spiral track as in a vinyl record).

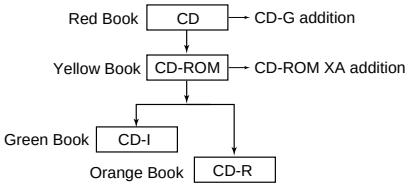
- Transfer rate
  The data transfer rate is expressed in numeric values that represent how much data can be transferred in comparison with audio CDs. The audio CD player can read out approximately 150k bytes of data in 1 second, which is an extremely low rate compared to computer processing speed. Therefore, units with transfer speeds 2 times or 3 times as fast as the transfer rate for audio CDs began to be developed and today the transfer rate has reached levels of 10 times and 20 times as fast.

④ Optical disk specifications

The optical disk, which was born from the audio CD and is widely used as a computer storage medium, has multiple variations that make the best use of its high storage capacity, portability, mass production through press replication processing, and other advantages. The standard specifications of these optical

disks have been established and the basic standard of each of them is called red book, yellow book, etc. These names were given after the color of the cover of the binder in which the standard specifications were kept.

Figure 2-4-15
Optical disks



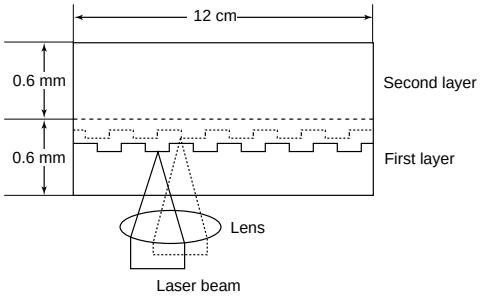
The outline of each standard is indicated below.

- Red Book
  The Red Book is the basic CD standard and describes the physical specifications of the CD. The standard of the CD-G, which made possible the storage of CG (Computer Graphics) in the audio CD, etc., has also been added.
- Yellow Book
  The Yellow Book is the basic CD-ROM standard and describes the physical format of the CD-ROM. As an extension to the Yellow Book, the CD-ROM XA, which made possible extended audio playback and multiple graphics recording, is also specified and its specification becomes the bridge between the CD-ROM and the CD-I.
- Green Book
  The Green Book describes the specifications of the CD-I (CD-Interactive), which is capable of storing audio, images, CGs, characters, programs, data, etc.
- Orange Book
  The Orange Book defines the physical structure of the CD-R (CD-Recordable), a writable CD. Among the CD-Rs, the CD-WO (Write Once), a recordable type in which once information is written, the content cannot be rewritten, and the CD-MO (Magneto Optical) a re-writable type which can be rewritten exist. This CD-R is used in photo CDs (camera film images recorded in a CD).

⑤ DVD

The DVD (Digital Versatile Disk or Digital Video Disk) is an optical disk capable of storing approximately 2 hours of animated images and audio data. The external appearance of DVD disks is the same as that of CD-ROMs, 12 cm of diameter and 1.2 mm of thickness. However, while the recording side of the CD-ROM consists of one side (1 layer), the DVD has a maximum of 2 layers, and data can be stored on both sides.

Figure 2-4-16
DVD structure



At present, DVD-ROMs are commercialized and can be played on DVD players. Likewise, the use of DVD-ROM units in personal computers is expanding.
DVD-ROM storage capacity is as follows:

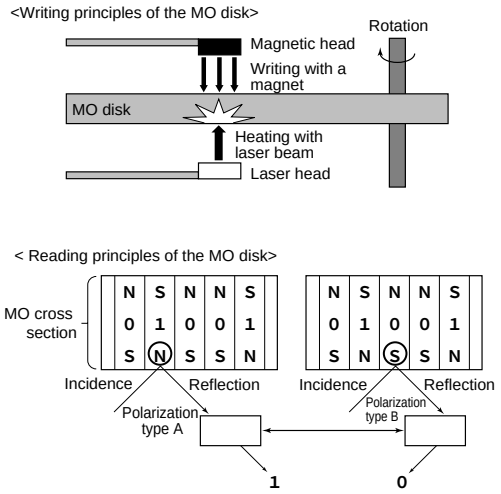
- Single layer single sided recording: 4.7 Gbytes
- Dual layer single sided recording: 8.5 Gbytes
- Single layer dual sided recording: 9.4 Gbytes
- Dual layer dual sided recording: 17 Gbytes

The DVD uses a compression method of animated images called MPEG2, which enables playback of extremely clear images. Due to its large capacity and high image quality, the DVD is attracting attention as a computer auxiliary storage medium/device. The standards of the read-only DVD-ROM, the recordable DVD-R and the re-writable DVD-RAM have been established.

(4) Magneto optical disk (MO) unit

The magneto optical disk unit has almost the same structure as the magnetic disk unit. While data can be only read from the CD-ROM unit, data can be read from and written onto the magneto optical disk unit. The main characteristic of the magneto optical disk is that the data reading and recording methods differ. Data recording is performed by applying the laser beam to the magnetized recording side to heat it then record high density information using the magnetic head. Data reading is performed focusing the laser beam on the magnetic disk and reading out the polarization direction of the reflected light. In other words, data writing is performed using a magnet while data reading is performed using a beam.
Figure 2-4-17 shows the writing and reading principles of the magneto optical disk unit. In this diagram, when the magnetization of the magneto optical disk from top to bottom is from N to S the value is "0" and when it is from S to N, "1" is the read out value.

Figure 2-4-17
Data reading and writing operation of the magneto optical disk unit



The magneto optical disk shipped today as a standard is the 3.5-inch disk, with a storage capacity of 128 MB or 230 MB. As a substitute for the floppy disk, the magneto optical disk has come into wide use as a computer auxiliary storage medium.

(5) Semiconductor disk unit

The semiconductor disk unit is a storage unit of high speed and large capacity, which uses flash memories and other devices. In most cases, it is used in high-end mainframe computers as a storage unit positioned between the main storage unit and the auxiliary storage devices. It has the advantage that, despite having several G bytes of storage capacity, its access time is 1/100 that of the magnetic disk unit.

2.4.2 RAID types and characteristics

Since computers have largely penetrated into our daily life, playing a critical role, high reliability is required together with high performance. Therefore, in order to achieve high reliability, a great variety of technologies are used. The method to duplicate system components to allow continued operation in case of a failure of a certain component is called fault tolerance technology. RAID (Redundant Arrays of Inexpensive Disk) is one such technology.
RAID is a method that consists of the parallel use of multiple hard disks (SCSI drives, etc.) in networks, etc. and has a high fault tolerance in the event that a failure occurs in more than one drive. There are 5 levels of RAID, which are used according to the objective.

(1) RAID 0

In RAID 0, data is distributed to more than 1 drive, but there is no spare drive.

(2) RAID1

In RAID 1, the same content is recorded on 2 hard disks with the same capacity. One of them automatically continues operating and the other is used as backup. This RAID 1 is called disk mirroring or disk duplexing.

(3) RAID 2

In RAID 2, bit interleave data is distributed to and recorded onto multiple drives, and the parity and error correction information is recorded in an extra drive.

(4) RAID 3

In RAID 3, bit interleave data is distributed to and recorded onto multiple drives, but only 1 drive is used as the parity drive.

(5) RAID 4

In RAID 4, data is not recorded by bits, but by sectors, and a separate drive is used as the parity drive for error detection.

(6) RAID 5

In RAID 5 the data is distributed and recorded by sectors and the parity information is added as separate sectors in the same way as ordinary data.

Among the above-mentioned RAIDs, the most used is RAID 5, which does not need a parity drive. A dedicated disk for the error correcting code (parity information) is required in RAID 2 and RAID 4. However, even if RAID technology is adopted, since it is only a countermeasure in the event that failure occurs in the disk itself, data backup at regular intervals is indispensable.

2.5 Input/output architecture and devices

Since there are many mechanical operations in the input devices and the output devices, a wide gap between the operation speed of these devices and that of the processors which perform electronic operations only, is generated. If, ignoring this operation speed gap, the processor and the input or output device are connected, the operation speed of the whole computer system will become slow. And as a consequence, the computer characteristic of high-speed processing becomes ineffective.
In order to solve this problem, input/output control and interruption are performed.

2.5.1 Input/output control method

When data is exchanged between the processor or the main storage unit and the auxiliary storage devices, input devices, output devices, etc., the following control methods are provided:
- Direct control method
- DMA method
- Input/output channel control method

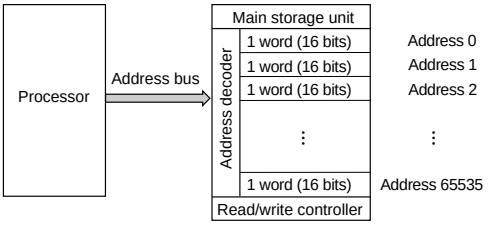
(1) Bus

The bus is a bunch of signal lines that connects units. In computers with a 16-bit word length, a bunch of 16 signal lines constitute a bus.
Access and information exchange are performed by a processor and a main storage unit using the following buses:
- Address bus
- Control bus
- Data bus

① Address bus

The address bus connects the main storage unit and the processor. This bus is used for the specification of the main storage unit address by the processor.

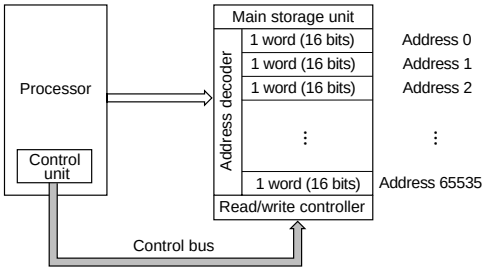
Figure 2-5-1
Address bus



② Control bus

The control bus connects the control unit and the main storage unit. This bus is used for the transmission of the instruction signal to the main storage unit from the control unit. (Figure 2-5-2).

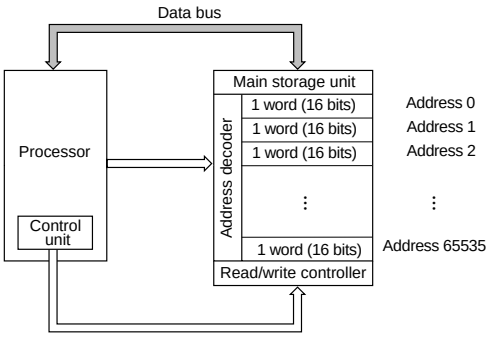
Figure 2-5-2
Control bus



③ Data bus

The data bus connects the main storage unit and the processor and is used to exchange data. Only this bus is used to exchange data between the main storage unit and the processor in both directions.

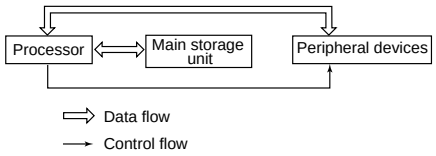
Figure 2-5-3



(2) Direct control method

The direct control method is the method by which the processor directly controls the input/output operations of the peripheral devices, and the data exchange is performed through the processor.
The structure of this method is simple, but it has a big drawback, which is that the processor cannot proceed to perform the next operation until the input/output operation is completed. For that reason, since processor efficiency is low, this method is not so widely used.

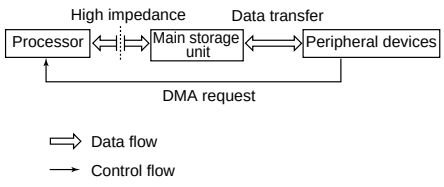
Figure 2-5-4



(3) DMA (Direct Memory Access) method

When a request signal is issued from an input device, output device or any a peripheral device such as an input device or output device, the connections between the processor and the main memory unit and between the processor and the peripheral device are set to the high impedance status and data transfer is performed between the main memory unit and the peripheral device. During this input/output process, the processor performs other processes, and it is not involved in the input/output process at all. This control method is called DMA method and is widely used as the input/output control system of personal computers.

Figure 2-5-5
DMA method



2.5.2 Input/output interfaces

An interface is an agreement for the connection of multiple devices and for the operation of these devices by humans. Among these, the interface related to the data input/output is called input/output interface.
The input/output control method explained above in Section 2.5.1 would not function correctly if the input/output interface were not established.
According to the transfer method, the input/output interface is divided as follows:

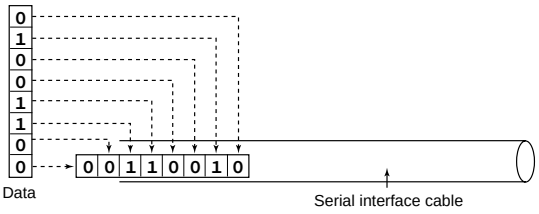
- Serial interface
- Parallel interface

(1) Serial interface

The interface that supports serial data exchange between the computer and the input device/output device is called the serial interface. The serial transfer is the transfer method conducted by lining up data of 8-bit or 16-bit processing units in one row and transferring one bit at a time (Figure 2-5-6). The data transfer rate is slower than that of the parallel interface, but it has the advantage that only one transmission channel is required. Since during a serial transfer there is no signal delay, long-distance transfers can also be performed.
The following serial interfaces are widely used:

- RS-232C (Recommended Standard-232C)
- USB (Universal Serial Bus)

Figure 2-5-6
Serial transfer



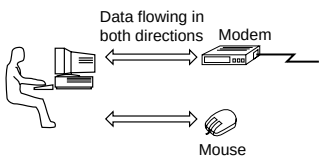
① RS-232C

The RS-232C is an interface that connects the computer, and the modem that converts digital signals into analog signals, or vice versa (Figure 2-5-7). This interface was standardized by EIA (Electronic Industries Association) and the physical connector format, pin number, pin role, etc. are strictly established.

<Characteristics>

- Transfer units: the start bit that indicates the start of 1-byte data: 1 bit, the top bit that indicates the end: 1 bit, the parity bit used for error detection: 1 bit. A total of 11 bits.
- Data exchange can be performed bi-directionally.
- The mainstreams in transfer rates are 28.8 kbps and 33.6 kbps.
- Besides the modem, it is widely used to connect image scanners, mice and other peripheral devices and personal computers.
- Data flow is bi-directional.

Figure 2-5-7
RS-232C interface connection example

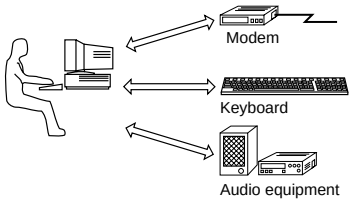


② USB

The USB is a new interface standard by which, besides the keyboard, modem and other peripheral devices, audio signals, image data and other input/output device data can be processed indistinctly with one connector (Figure 2-5-8). According to the USB specification, when a peripheral device supported by the USB is connected, the personal computer automatically is configured. The connection of cables has

also been simplified and its use as an input/output interface for multimedia systems is attracting attention.

Figure 2-5-8
USB interface
connection example



③ IEEE1394

IEEE1394 is a serial interface used to send animated image data in real time. Since real time transfer is supported, animated images can be smoothly represented. Therefore, IEEE1394 is used as a multimedia interface supporting connections such as those between digital video cameras and personal computers. The maximum data transfer rate is 400 Mbps and a connection of a maximum of 63 nodes can be performed.

④ IrDA (Infrared Data Association)

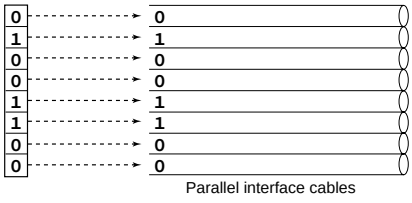
The IrDA is an interface for wireless (infrared) data transmission. It has the advantage that, since connection cables are not used, layout modifications inside the office can be easily performed. There are several IrDa versions, and the transmission speed ranges between 2.4 kbps and 4.0 Mbps. These versions are equipped in PDA (Personal Digital Assistants) and notebook-type personal computers.

(2) Parallel interface

In parallel interfaces, instead of transferring data in sequence, 1 bit at a time, as in serial interfaces, data is transferred in parallel using 8 or 16 cables (Figure 2-5-9). Compared to serial interfaces, the data transfer rate in parallel interfaces is high. However, since multiple transmission channels are required, the transmission channel maintenance cost becomes high.
The following parallel interfaces are widely used:

- Centronics interface
- SCSI (Small Computer Systems Interface)
- GPIB (General Purpose Interface Bus)

Figure 2-5-9
Parallel transfer



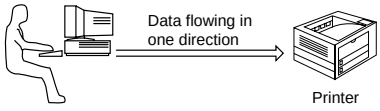
① Centronics interface

The Centronics interface is a printer interface developed by the U.S. company, Centronics Data Computer (Figure 2-5-10).
It is not an interface formally standardized by an international industrial standard organization, but, since it has been adopted by a very large number of manufacturers as the interface to connect printers and personal computers, in practice, the Centronics interface has become the (*de facto* standard printer interface).

<Characteristics>

- 8-bit parallel transfer is possible.
- Limited to one-direction data transfer.
- Limited to peer-to-peer connections.
- The transfer rate is 150 kbps

Figure 2-5-10



Centronics interface
connection example

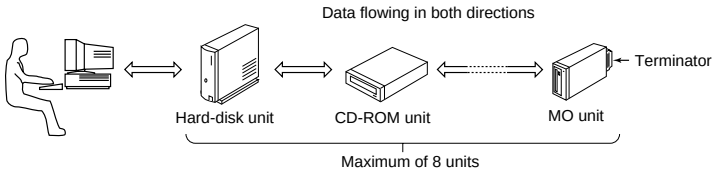
② SCSI

The SCSI was approved as a personal computer standard interface by ANSI (American National Standards Institute).

<Characteristics>

- 8-bit parallel transfers can be performed bi-directionally.
- The transfer rate ranges between 1.5 and 4 Mbps but the transfer rate of SCSI-2, an extensive enhancement of SCSI, is 20 Mbps.
- Up to 8 auxiliary storage devices, such as the hard disk unit and the CD-ROM unit, can be connected one after another. This is called daisy chain. An ID (a number to distinguish the devices) is assigned to the connected devices and a resistor, called a terminator, that indicates the termination, is attached (Figure 2-5-11).
- The pin number is 50 (or 25).
- The data flow is bi-directional.

Figure 2-5-11 SCSI connection example (daisy chain)



③ GPIB

The GPIB was originally approved as a standard interface to connect microcomputers and measurement instruments, but it is currently an interface with a wide range of uses that connects the microcomputer and its peripheral devices.

This interface was standardized as IEEE-488 by the U.S. Institute of Electrical and Electronics Engineers (IEEE).

<Characteristics>

- 8-bit parallel transfers are possible.
- It is composed of 24 signal lines.
- The transmission distance is within 20 m.
- The data transfer rate ranges between 1 kbps and 1 Mbps
- Connection of up to 15 devices is possible.

Figure 2-5-12 synthesizes the above-mentioned input/output interfaces.

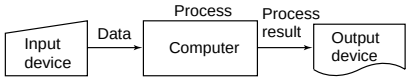
Figure 2-5-12 Types of input/output interfaces

	Name	Transfer rate	Connected devices	Content	Industrial standard
Serial transfer	RS-232C	• 28.8kbps • 33.6kbps • 56kbps	• Modem • Printer • Mouse • Plotter, etc.	• Connector format and role • Pin number and role	EIA standard
	USB	• 12Mbps	• Keyboard • Modem • Speaker, etc.	• Multimedia support • Audio, images, etc. can be processed with one connector • Cable connection simplification	—
	IEEE1394	• 400Mbps	• Digital video camera	• Multimedia support • Real-time function • Possibility to connect up to 63 nodes	IEEE standard
	IrDA	• 2.4kbps ～4.0Mbps	• Hard disk • Printer • Modem • Mouse	• Infrared data transmission	IrDA standard
Parallel transfer (8 bits)	Centronics interface	• 150kbps	• Printer • Plotter • Digitizer, etc.	• Wide use as printer standard interface	De facto standard developed by Centronics Data Computer Corp.
	SCSI	• 1.5～4Mbps	• Auxiliary storage devices	• Possibility to connect up to 8 devices in a daisy chain	ANSI standard
	GPIB	• 1kbps～ 1Mbps	• Measurement instruments • Peripheral devices	• Developed by the U.S. company, Hewlett-Packard • Possibility to connect up to 15 devices	IEEE-488

2.5.3 Types and characteristics of input/output devices

While the data we handle is made up of characters, numeric values (decimals), symbols, etc., only binary digits can be handled by the computer. Therefore, the data and information subject to process must be converted into a format that can be processed by the computer before being transferred to the processor. The devices equipped with this function are generally known as input devices.
Likewise, if we see the binary results processed by the computer, it will not be easy to understand their meaning. Therefore, the content processed using binary digits has to be converted into a format that can be understood by humans. The devices that perform this kind of function are generically known as output devices.

Figure 2-5-13
Roles of input devices and output devices



The input device is a device exclusively used to transfer information to the computer, and the output device is a device that represents the result of the computer process in a format that can be understood by us. But there is also a device equipped with both functions. It is called the input/output device.

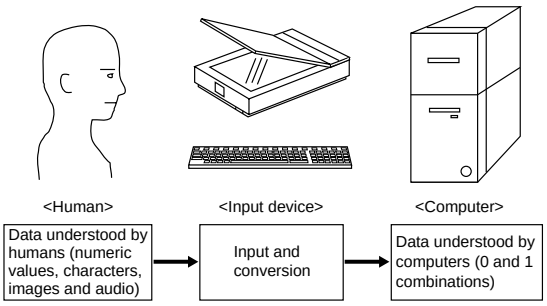
(1) Input devices and output devices

The device that enters data and programs into the computer is called the input device, and the device that represents/outputs computer data and programs is called the output device.

① Input device

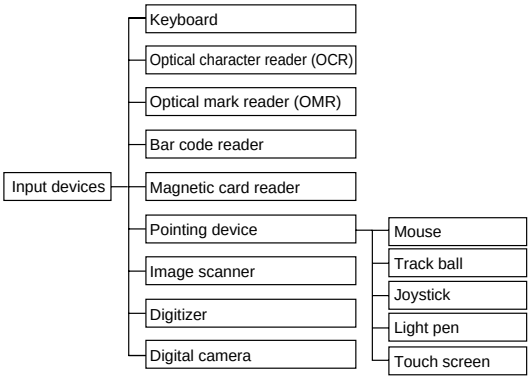
An input device is a device that converts data that can be understood by humans such as numeric values, characters, images and audio, into a data format (0 and 1 combinations) that can be understood by computers, and loads it into the computer main storage unit.

Figure 2-5-14
Input device's role



The early computers were limited to processing characters and numeric values, but with the progress of information technology today, computers can also process image and audio data. The different types of input devices that can process all these kinds of information are shown in Figure 2-5-15.

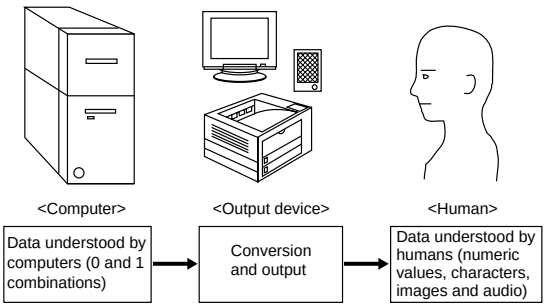
Figure 2-5-15
Various input devices



② Output device

Output device is the general term for the devices that convert the data processed in the computer (which processes all the data using 0s and 1s to produce results which are combinations of 0s and 1s) into data that can be understood by humans such as numeric values, characters, images, still images, animated images and audio, and output it.

Figure 2-5-16
Output device's role



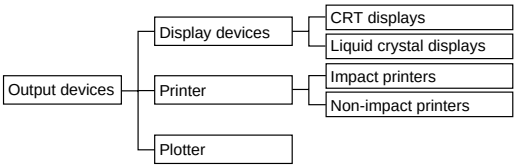
As in the input devices, there are different kinds of output modes for the output devices (Figure 2-5-17). For example, if the output is divided into "Display" and "Printing," it can be classified into the following two types:

- Display devices: The output is displayed on a television screen.
- Printer: The output is printed on the surface of a piece of paper.

Furthermore,

- Output into a display device is called "Soft copy"
- Output into a printer is called "Hard copy"

Figure 2-5-17
Various output devices



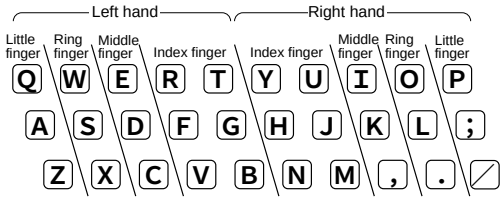
(2) Keyboard

The keyboard, which is the input device we find most familiar, inputs the code corresponding to the key of the character or symbol we press in the processing unit. A keyboard layout is specified by a JIS (Japanese Industrial Standard). However, in order to improve the efficiency of Japanese input, some word processor manufacturers have developed unique keyboard layouts of their own.

Figure 2-5-18
Keyboard

Figure 2-5-19 shows the correct finger position to be assumed to press the keys. Once one gets used to it, it becomes easy to press the correct keys without looking at the keyboard. This typing method is called touch typing.

Figure 2-5-19
Touch typing



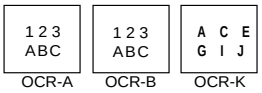
(3) Optical character reader (OCR)

The optical character reader is a device that, based on the intensity of the reflected light, reads out characters and symbols and inputs them (Figure 2-5-20).

Figure 2-5-20
Optical character reader (OCR)

At the beginning, the optical character reader could only read easily identifiable special printed characters and JIS OCR fonts (Figure 2-5-21). However, at present, the character pattern recognition has improved and even handwritten characters can be recognized.

Figure 2-5-21
OCR fonts



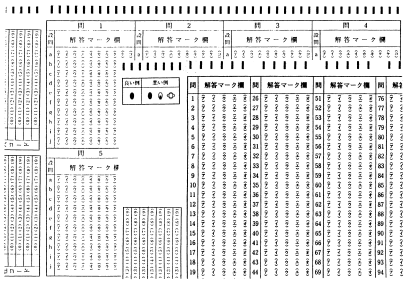
(4) Optical mark reader (OMR)

The optical mark reader is the device that reads and inputs data according to the marks made on marksheets. Alphanumeric characters are printed on marksheets; the part to be input is marked with a pencil, etc. The difference with the optical character reader is that this device does not directly recognize the patterns of characters or numeric values, but reads them out based on the marked position instead.
The information reading principle is the same as that of the optical character reader. Based on the reflected light, this device judges whether or not marks exist and inputs the character or numeric value corresponding to the marked position. Therefore, there are cases where reading errors occur or reading cannot be performed – when marksheets are dirty or folded.

Figure 2-5-22
Optical mark reader (OMR)

Since the mark reading position is set through the program, the pattern of the marksheet can be freely designed. In practice, marksheets are used in different fields, and are also used as answer sheets in the Information Processing Engineer Examination.

Figure 2-5-23
Marksheet example



(5) Bar code reader

The bar code reader is the device that reads and inputs the bar code attached to diverse products. The following types of bar code readers exist:

- Pen type
- Touch type
- Laser type

When the pen-type device is used, the bar code has to be traced with LED (light emitting diode). With the touch-type device, the LED only has to be focused on the bar code (Figure 2-5-24). Likewise, since it is not necessarily to directly touch the bar code with the laser-type device, it is widely used in convenience stores and supermarkets.

Figure 2-5-24
Bar code reader

Generally, in account processing using bar codes, not only product identification and accounting are performed, but also, based on the information input, stock control and order control are performed.

(6) Magnetic card reader

The magnetic card reader is a device that reads and inputs the information needed from a magnetic card. There are different types of readers depending on the magnetic card to be read. Automatic train ticket gate machines are also magnetic card readers.
The magnetic card is a paper or plastic card which has a magnetic stripe on the surface to store information such as numeric values and characters.
At present, as it is said, we live in a "card society" and magnetic card use has widely expanded as a paying method that replaces cash in our daily life. Among the most familiar cards, we have the following:

- Phone cards
- Cash cards
- Credit cards
- Tickets for automatic ticket gates

Magnetic cards have become indispensable in our daily life. There is also the IC card, which has increased the storage capacity of magnetic cards and incorporated information processing functions. It is more expensive than regular magnetic cards, but it is superior in security and functional aspects.

(7) Pointing device

"Pointing device" is the generic term for the devices that input positional information on the screen of a display device.
With the expansion of computer use, different kinds of pointing devices have appeared. Among the main pointing devices, the following can be mentioned.

- Mouse
- Track ball
- Joystick
- Light pen
- Touch screen

Without the troublesome operations on keyboards, etc., using any of these devices, anybody can easily input data while watching the screen of the display unit.

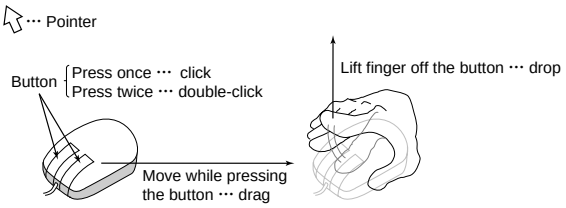
① Mouse

Along with the keyboard, the mouse is the most used input device. It was named mouse due to the similarity of its external appearance to a mouse.

Figure 2-5-25
Mouse

The mouse has a mechanism by which, when it is moved, the ball on the underside rolls, and the screen pointer moves according to the rolled distance and direction. When the pointer has been moved to the aimed position, by pressing the button of the mouse, the positional information is entered. This operation is called clicking. Likewise, as it is shown in Figure 2-5-26, besides clicking, there are other button operations such as double clicking, which means pressing the button two times, and dragging, which is the specification of an area by moving the mouse while holding the button down, etc.

Figure 2-5-26
Mouse operations and pointer



The biggest characteristic of the mouse is that, unlike the keyboard, it does not input data directly to the computer; instead, by pointing at the icons and windows indicating the operation locations which appear on the screen, it indicates and inputs operations to the computer. In other words, the mouse supports the GUI (graphical user interface) environment.

② Track ball

The principle is the same as that of the mouse, but, since the track ball is moved directly with the fingers, it does not require the moving space needed by the mouse. For that reason, it is often equipped in lap-top and notebook personal computers.

Figure 2-5-27
Track ball

③ Joystick

With the joystick, the stick is moved back and forth as well as to the right and left, and the pointer moves according to the direction and the angle in which the stick is moved. It can perform the same operations as the mouse, but, since indications have to be performed with buttons, etc. besides the stick, it is not so easy to handle as the mouse, which can be manipulated with one hand. For that reason, joysticks are widely used for game software manipulation.

Figure 2-5-28
Joystick

④ Light pen

The light pen is a device that inputs the coordinate information by pointing and tracing on the screen of the display device directly.
Since the optical sensor at the point of the light pen detects the position of the information on the screen and inputs it, the responsiveness is high.
As the uses of the light pen, data entry on palm top personal computers entry, and the entry of handwritten characters into a word process can be mentioned.

Figure 2-5-29
Light pen

⑤ Touch screen

The touch screen, which is also called a touch panel, takes advantage of the static electricity that passes through the human body. By directly touching with the finger the screen of the display device, the positional information is entered. In this mechanism, a transparent panel is attached to the screen surface, and the sensor on the panel senses changes in the voltage and detects the touched position.
Due to the dimensions of the area touched with the finger, detailed manipulations and instructions can not be performed, but since it can be easily manipulated by anybody, it is widely used in the automatic teller machines (ATMs) of banks, automatic ticket vending machines at train stations, reception/information at hospitals, etc.

Figure 2-5-30
Touch screen

(8) Image scanner

The image scanner is a device that reads and inputs figure and picture data from a sheet of paper with the same principle as a fax. The mechanism consists of decomposing a figure or image into a dot image, composed of small dots, and focussing light on it of order to load the intensity of the light reflected as electronic codes into the computer.
The devise which moves the reading mechanism over a fixed piece of paper is usually called the image sensor.

Figure 2-5-31
Image scanner

(9) Digitizer

The digitizer is a device that, by tracing a plane panel figure with a pen or cursor, detects the coordinates position and, based on this consecutive coordinate information, inputs the figure. This device is used by the

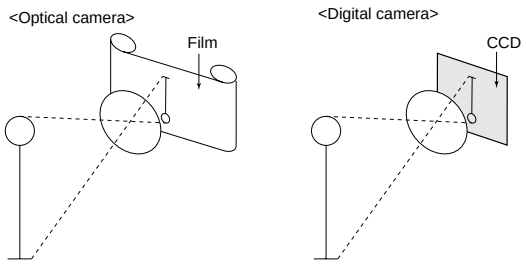
CAD (Computer Aided Design) application, in which the input of figures of high precision is required. Small-sized devices are sometimes called tablets for making a distinction.

Figure 2-5-32
Digitizer

(10) Digital camera

The digital camera is a camera that can input the picture taken to the computer as data. While optical cameras record images through the chemical change of the sensitized material of the film surface, digital cameras, using the image sensor of a semiconductor element called CCD (Charge Coupled Device), convert optical pictures into digital data and record this data as image files.
A semiconductor memory called flash memory is widely used as the storage medium.

Figure 2-5-33
Optical camera and digital camera mechanism



(11) Display device

In ordinary computer systems, operations are conducted verifying on the spot the input data and process results displayed on the screen. The display device is one of the devices which are indispensable for human use of computers. Displays are roughly divided into the following two types:

- Character displays: Capable of displaying characters only.
- Graphics displays: Capable of displaying characters and graphics.

Likewise, according to the colors that can be displayed, displays are divided as follows:

- Black and white displays
- Color displays, etc.

At present, color graphics displays are the standard. Furthermore, according to the structure of the display screen, display devices can be classified as follows:

- CRT displays
- Liquid crystal displays

Here, CRT displays and liquid crystal displays will be explained.

① CRT display

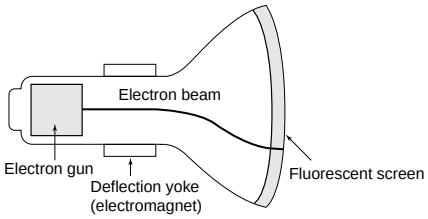
The display device that has the same structure as the television, and uses the cathode-ray tube, is called CRT (cathode-ray tube) display.

Figure 2-5-34
CRT display

a. Mechanism

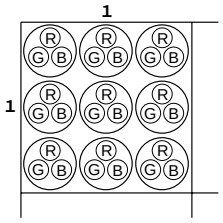
As it is shown in Figure 2-5-35, in the CRT display, when struck by the electron beam, the fluorescent screen emits light, which is displayed on the screen.

Figure 2-5-35
CRT structure



Color display is generally used in CRT displays. Color images are represented by striking the dots that contain the "three primary colors of light," R, G and B (Red, green and blue) with the electron beam.

Figure 2-5-36
Color screen diagram magnification



Besides the above mentioned, the screen display is equipped with an important function called the screen saver. In the CRT display, when the same screen is displayed for a long period of time, the image of that screen is burnt into the fluorescent screen of the cathode-ray tube. In order to prevent this, an animated image is displayed on the screen. This software is called screen saver.

b. Resolution

The screen size is represented by the diagonal length of the screen. According to this, there are screens of 15 inches, 17 inches, 21 inches, etc.
The screen resolution is represented by the value of the number of dots which can be represented in 1 screen (width × height), and resolutions of 640 × 480, 800 × 600, 1024 × 768, 1280 × 1024, etc. Today, as a result of the expansion of multimedia, 1280 × 1024 resolution has become a *de facto* standard due to its capacity to process high image quality. Likewise, the CRT display called multi scan monitor, in which the resolution can be switched according to necessity, is expanding.

② Liquid crystal display

The liquid crystal display is a display device widely used in the display screen of calculators, etc. (Figure 2-5-37). The liquid crystal material used has the property of aligning in one direction when voltage is applied, changing from non-transparent to transparent. The liquid crystal display takes advantage of this property, and by controlling whether or not light passes through it with the voltage applied, the appropriate display is produced.
Currently, the majority of liquid crystal displays are also color displays that use R, G, B color filters.
Unlike CRT displays that require a specific depth, liquid crystal displays are thin, and moreover, have low power requirements. Due to these reasons, they are widely used in lap top and notebook personal computers.

Figure 2-5-37
Liquid crystal display

The following two types of liquid crystal display exist:

a. Passive matrix type

The system in which multiple liquid crystal pixels are controlled by one semiconductor is called the passive matrix type. This system is adopted by the STN (super twisted nematic) liquid crystal display. Currently, the DSTN (Dual scan STN) liquid crystal display, which divides the liquid crystal panel into upper and lower sides enabling double scanning, is expanding.

b. Active matrix type

The system in which one liquid crystal pixel is controlled by one semiconductor is called the active matrix type. This system is adopted by the TFT (Thin Film Transistor) liquid crystal display. The TFT liquid crystal display uses a transistor as a switch to apply voltage. The screen contrast, response speed, viewing angle, etc. are dramatically superior to the STN liquid crystal display. However, due to the complex structure, the production cost is high.

(12) Printer

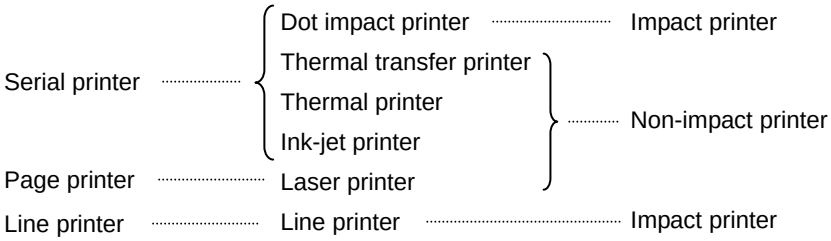
The printer is the oldest computer output device and, today, it is much more widely used.
There are many types of printers. According to the printing methods, printers are classified as follows:

- Impact printers: Print by mechanically hitting pre-formed characters against an ink ribbon.
- Non-impact printers: Print using heat, ink, laser, etc.

Likewise, according to the printing unit, printers are classified as follows:

- Serial printers: Print 1 character at a time, as typewriters.
- Line printers: Print 1 line at a time.
- Page printers: Print 1 page at a time.

Here, among the different types of printers, the printers indicated below will be explained:



① Dot impact printer

The dot impact printer is the printer that prints by striking a head against an ink ribbon, which, in turn, hits the paper. Alphanumeric characters are sets of dots, and since printing is performed by striking a head containing multiple small pins, dot impact printers are noisy. This is a drawback: however, they are convenient for performing multiple printing at once using carbon copies.
Compared to thermal transfer printers, etc., the quality of the printed characters is not so good. The density of the dots composing a character (width × height) determines the print resolution. The number of dots forming one character varies from 9 × 7, 16 × 16, 24 × 24. The higher these values, the better the resolution of the printed characters. Likewise, the characters that can be printed vary. Only alphanumeric characters can be printed with a 9 × 7 resolution, while kanji can also be printed with 16 × 16 and 24 × 24 resolution.

Figure 2-5-38
Dot impact printer

② Thermal transfer printer

The thermal transfer printer is a printer in which the print head heats and melts the ink of the ink ribbon to print the dot-composed character on normal paper. Since this printer uses a non-impact method, the noise level is lower than that of dot impact printers.
On the other hand, the thermal printer uses thermal paper as printer paper. Since thermal paper fades with time, it is not suitable for long-term conservation. Furthermore, the fact that the running cost becomes high due to the high price of thermal paper is a drawback.

③ Ink-jet printer

The ink-jet printer is a printer in which, according to the form of the dot-composed character, tiny ink nozzles in the print head squirt ink onto the paper. Color ink of 3 colors "Cyan: C, Magenta: M and Yellow: Y) or 4 colors" "3 colors + Black: K" are used in color printing. Today, the use of color ink-jet printers as personal computer printers has expanded.

Figure 2-5-39
Ink-jet printer

④ Laser printers

The laser printer is a page printer that, using toner (powder ink), creates the printing image of one page on the photoreceptive drum and transfers it to the paper through the application of laser beams.
The printing principle is the same as that of copy machines, and the character size, space between the lines, etc. can be selected freely. Figures, images, etc. can also be printed and, print quality as well as printing speed are high. For that reason, it is the mainstream printer for business use.

Figure 2-5-40
Laser printer

2.6 Computer types

Computers used in a great variety of fields will be explained below.

(1) Personal computer

As the name implies, personal computers are computers that were developed for personal use, commonly called PCs for short. Based on their external appearance, different types of personal computers have multiplied. These personal computers can be classified as follows:

- Desk-top type, which can be placed on a desk (Figure 2-6-1)
- Lap-top type, which can be placed on one's lap
- Notebook type, the size of A4 or B5 paper, thin and light (Figure 2-6-2)

Likewise, expansion of the palm-top type (Figure 2-6-3) ultra small-sized personal computers which can be held in one's palm, is starting.

On the other hand, according to the place where they are set up, and the main use purpose, computers can be classified as follows:

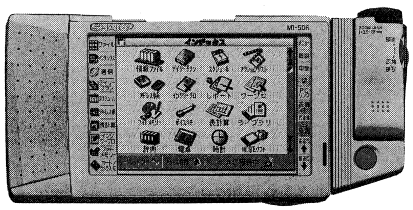
- Home: Used as word processors and to play games as home and hobby computers.
- Enterprise: Word processor, spreadsheet software and database software are used for business.
  Used in software development.
- School and enterprise: Used in CAI (Computer Aided Instruction) application for education.

Likewise, considering the use mode, up to now, the stand-alone system was used in most computers, but recently, the network system in which personal computers are connected by communication lines is becoming the mainstream.

Figure 2-6-1
Desk-top type

Figure 2-6-2
Notebook type

Figure 2-6-3
Palm-top type



(2) Workstation

Due to the expansion of use of personal computers, enterprises are adopting the system of one computer per person. However, personal computers lack the capacity to perform technological calculation processing, software development, etc. In order to solve this problem, computers called engineering workstations (EWS) were created. Compared to personal computers, workstations are capable of performing high quality image processing, etc. with high speed (Figure 2-6-4).
The main applications of these workstations are listed below:

- Research and development fields:
  High-speed processing of complex scientific and engineering calculations.
- Product design/manufacturing fields:
  Used in CAD (Computer Aided Design), CAM (Computer Aided Manufacturing), etc. application.
- Software development field:
  Use of CASE tools (Computer Aided Software Engineering tool), etc.
- Communication network field:
  Used as client machines or server machines in distributed processing systems.

Figure 2-6-4
Workstation

(3) General-purpose computer

The general-purpose computer is a computer which, literally, can be used for multiple purposes, capable of performing both office work as well as scientific and engineering calculations. Since it is the mainframe of a great number of computers used in an enterprise, it is also called mainframe (Figure 2-6-5).
The computers that conduct enterprise core business system processing, an automatic ticketing processing, bank services, etc. are all general-purpose computers.
Since most general-purpose computers are large sized, due to the high heat generation, it is necessary to install them in an air-conditioned room called a computer room.

Figure 2-6-5
General-purpose computers

(4) Supercomputer

There is no precise definition of supercomputers. Commonly, computers capable of performing enormous and complex calculations at extremely high speeds are called supercomputers. In other words, it can be said that supercomputers are computers whose design attaches importance to high-speed calculation.
Supercomputers are computers that compile computer high-speed technology using forefront semiconductor element technology as well as vector processors that perform floating point operations and vector operations, etc. However, their use purposes are limited. Among the main purposes, the following can be mentioned:

- Weather forecast
- Simulation of nuclear power generation
- Orbit calculation of artificial satellites

As a manufactured product, the U.S. Cray Inc.'s Cray supercomputer is famous. In Japan NEC's SX and Fujitsu's FACOM VP are produced.

Figure 2-6-6
Supercomputer

(5) Microcomputer

Microcomputers are small-sized computers into which a microprocessor is built. The computers that are imbedded into machines, especially household appliances such as washing machines, air conditioners and AV appliances, in order to control the machine operation are called microcomputers. These microcomputers are electronic parts with bare integrated circuits. According to the purpose, information on the temperature and number of revolutions can be entered using a sensor. Since their function is to repeat the same operation, control data is recorded in ROMs (Read Only Memory). Likewise, since the output devices are motors or electric switches, they are also called actuators.

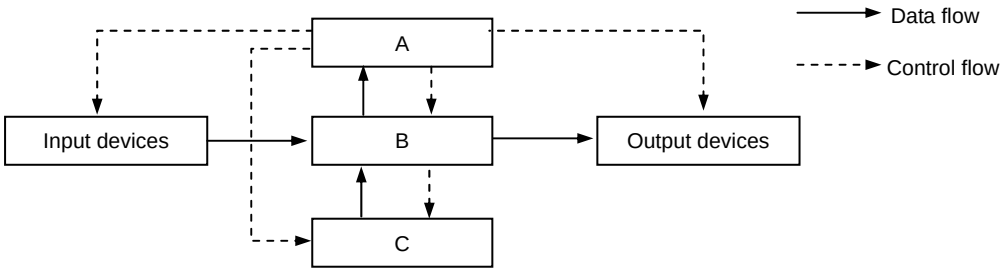
Figure 2-6-7
Microcomputer

(6) Process control computer

Process control computers are computers that control the different types of machines in steel mills, automobile plants, petroleum refineries, etc. Chemical plants, etc. are entirely automated with process control computers. When the supervising computer detects an abnormality, it immediately controls each machine and adjusts the production process.
In addition to this, centralized control and automation have been achieved through the use of process control computers in power system control, general building security systems, highway traffic control, etc.

Exercises

Q1 What is the combination of words that should fill in the blanks of the diagram representing the computer basic configuration?



	A	B	C
a	Arithmetic unit	Main storage unit	Main storage unit
b	Main storage unit	Control unit	Arithmetic unit
c	Control unit	Arithmetic unit	Main storage unit
d	Control unit	Main storage unit	Arithmetic unit

Q2 What is the appropriate explanation of a DRAM?

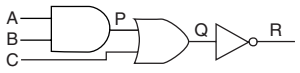
a. DRAM represents 1 bit depending on whether the capacitor is charged or not. It is commonly used as a main storage unit.
b. Data is written at the time it is manufactured. It is used as a microprogramming storage memory.
c. Data can be written using a special device and erased with ultraviolet light.
d. It is composed of flip flops. The speed is high but the manufacturing cost is high as well. It is used in cache memories, etc.

Q3 Regarding the index modification of the machine language instruction, which of the following would be the effective address?

Address in which the instruction is stored: 1000
Value of the instruction language: 100
Value of the index register: 10

a. 10 b. 100 c. 110 d. 1100 e. 1110

Q4 Given the following circuit, when the input values are A=1, B=0, C=1, what is the appropriate output value for P, Q and R? Here represents an AND gate, represents an OR gate and represents a NOT gate.



	P	Q	R
a	0	1	0
b	0	1	1
c	1	0	1
d	1	1	0

Q5 **When the sum of 1-bit values A and B is represented in 2-bit values, which of the following corresponds to the combination of the logical expression of the higher bit D and the lower bit S? Here, " · " represents the logical product (AND), "+," the logical sum (OR) and $\overline{A}$, the negation (NOT) of A.**

A	B	Sum of A and B	
		C	S
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	0

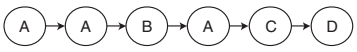
	C	S
a	$A \cdot B$	$(A \cdot \bar{B}) + (\bar{A} \cdot B)$
b	$A \cdot B$	$(A + \bar{B}) + (\bar{A} + B)$
c	$A + B$	$(A \cdot \bar{B}) + (\bar{A} \cdot B)$
d	$A + B$	$(A + \bar{B}) + (\bar{A} + B)$

Q6 **In a computer with 3 types of instruction sets, which of the following corresponds to the MIPS value when their respective execution speed and frequency rate are as follows?**

Instruction set	Execution speed (microseconds)	Frequency rate
A	0.1	40%
B	0.2	30%
C	0.5	30%

a. 0.25 b. 0.8 c. 1.25 d. 4

Q7 **A given program executes instructions A, B, C and D in the following order:**



The CPI required to execute each instruction is indicated in the following table. If 1 clock cycle of the CPU is 10 nanoseconds, how many nanoseconds will the CPU execution of this instruction string require?

Instruction	CPI
A	6
B	2
C	4
D	8

a. 20 b. 32 c. 200 d. 320

Q8 **Which of the following is the method of the ordinary computer basic architecture that loads programs and data together in a computer storage device and sequentially reads and executes them?**

a. Address method
b. Virtual storage method
c. Direct program control method
d. Program storage method

Q9 **Given the following magnetic disk unit specifications and conditions of the data subject to storage, how many tracks does the necessary area have when the blocking factor is 20? Here, the area is assigned by track and the file organization is sequential.**

Magnetic disk unit specifications

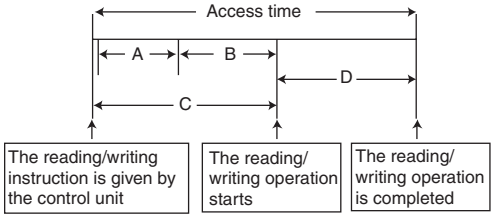
Storage capacity per track	25,200 Bytes
Inter-block gap	500 Bytes

Conditions of the data subject to storage

Record length	200 Bytes
Number of records	10,000 Records

a. 80 b. 83 c. 89 d. 100

Q10 **The following diagram represents the access time of the magnetic disk unit. Which is the A, B, C and D correct combination?**



	A	B	C	D
a	Seek time	Search time	Latency	Data transfer time
b	Seek time	Latency	Search time	Data transfer time
c	Latency	Seek time	Data transfer time	Search time
d	Latency	Data transfer time	Seek time	Search time

Q11 **Given the magnetic disk unit with the following performance, what is the average access time in milliseconds required to read the 2,000-byte-length-block data recorded in this magnetic disk?**

Magnetic disk unit performance

Storage capacity per track (bytes)	20,000
Revolution speed (revolution/minute)	3,000
Average seek time (milliseconds)	20

a. 30 b. 31 c. 32 d. 42

Q12 **Regarding the optical disk characteristics, which of the following descriptions is correct?**

a. CD-ROMs have large storage capacity, but since high-level technology is required for their manufacture, compared to magnetic disks, the cost is higher for the same amount of information.
b. In the magneto optical disk, which is one of the rewritable storage media, data is recorded by changing the medium magnetization direction.
c. In the recordable optical disk, in which data is recorded by making microscopic holes in the medium, data can be rewritten as many times as required.
d. Since the access mechanism of magneto optical disks is very similar to that of magnetic disks, the average access time is also of the same level.
e. Since magneto optical disks are susceptible to heat, light and dust, compared to magnetic disks, the magneto optical disk's durability is lower.

Q13 **Which of the following is the most appropriate explanation of mirroring, which is one of the methods used to improve the magnetic disk unit reliability?**

a. By giving a mirror-like finish to the disk surface the resistance at the time the disk rotates is reduced.
b. The data block and the parity block are stripped and stored across multiple disks.
c. Besides the disks that record the data, another disk for parity recording is used.
d. Identical data is recorded simultaneously in separate disks.

Q14 **Which is a feasible combination of interfaces for connecting the peripheral devices indicated below? Here ATA/ATAPI-4 represents the interface that is normally called IDE.**

	Hard disk, CD-ROM	Modem	Keyboard
a	ATA / ATAPI-4	GPIB	SCSI
b	GPIB	SCSI	RS-232C
c	SCSI	RS-232C	USB
d	USB	IrDA	ATA / ATAPI-4

Q15 **If an image, whose height and width in pixels is 480 dots and 640 dots, respectively, is represented in 256 types of colors: approximately how many kilo bytes would be required in order to save this data in a storage device? It should be noted that no compression process is performed.**

a. 170 b. 310 c. 480 d. 9,840 e. 78,650

Q16 **Which of the following printers uses a heating element to melt the ink of the ink ribbon and is capable of printing on normal paper?**

a. Ink-jet printer
b. Thermal printer
c. Dot impact printer
d. Thermal transfer printer
e. Laser printer

3 Basic Software

Chapter Objectives

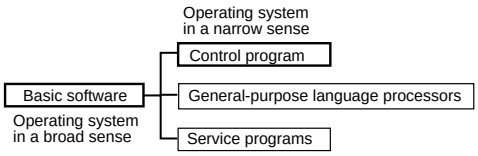
In order to use a computer, we need basic software.
In order to efficiently operate the hardware composing the computer system as well as the application software, it is necessary to understand the mechanism and functions of the operating system (control program), which performs different kinds of control/management. Therefore, we need to:

① Understand well the software names and classifications, functions and roles, including the relation with the hardware and the user.
② Understand the reasons why the operating system is necessary, its roles, structure, functions, etc.
③ Understand the types and characteristics of the major operating systems.

Introduction

Software that helps users to make effective use of the hardware functions is generically called system software. Systems software is roughly classified into basic software and middleware.
A basic software is a set of programs aimed at efficient use and control of the different types of resources provided by the hardware. It can be thought as the operating system in a broad sense.
As it is shown in Figure 3-1-1, the basic software itself is classified into control program, language processors and service programs. Software has a great number of complex functions, which will be explained in detail.

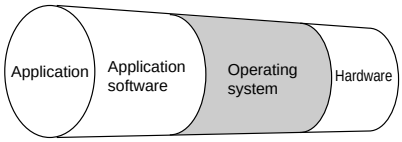
Figure 3-1-1
Basic software



3.1 Operating system

Software corresponding to its purpose is incorporated into computers utilized on various field. Such software is called "Application Software."
On the other hand, the software acting as a bridge between the application software and the hardware is the operating system (OS), which will be studied hereafter.
Here, the objectives and functions of the operating system, which is the closest system software to the hardware, will be studied.

Figure 3-1-2
Operating system
position



3.1.1 OS configuration and functions

The fully-fledged operating system was born in the 1960s. Since computers in that era were extremely expensive, the main consideration of the users was how to operate them efficiently. For example, if the computer was laying idle while the data to be processed by the computer was being prepared or while the results processed by the computer were being processed manually by humans, it can not be said that these expensive computers were being efficiently used.
Therefore, the operating system was born for the purpose of having the computer prepare the data to be processed and control the execution process by itself.

(1) OS role

The following are the purposes of the operating system that is incorporated in today's general-purpose computers, and is used in different fields:

- Efficient use of resources
- Consecutive job processing
- Multiple programming
- Reduction of the response time
- Improvement of reliability.

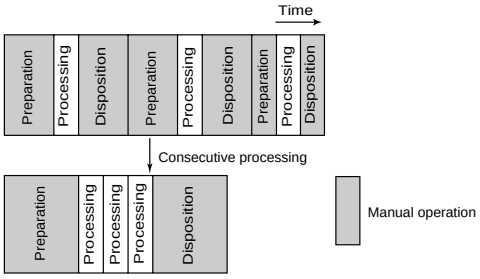
① Efficient use of the resources

From the operating system point of view, the processor, main memory unit, auxiliary storage devices, input devices, output devices, application software, and other components of the computer system are all resources for computer use. The purpose of the operating system is to efficiently use these resources without relying on humans and without waste.

② Consecutive job processing

The work done by a computer is called job. If human manual operation is required between jobs that process data at electronic speeds, the processor use efficiency would drop dramatically.
For that reason, by eliminating as much human intervention as possible, the operating system implements automatic consecutive processing and enhances the processing efficiency of the whole computer system.

Figure 3-1-3
Consecutive job processing



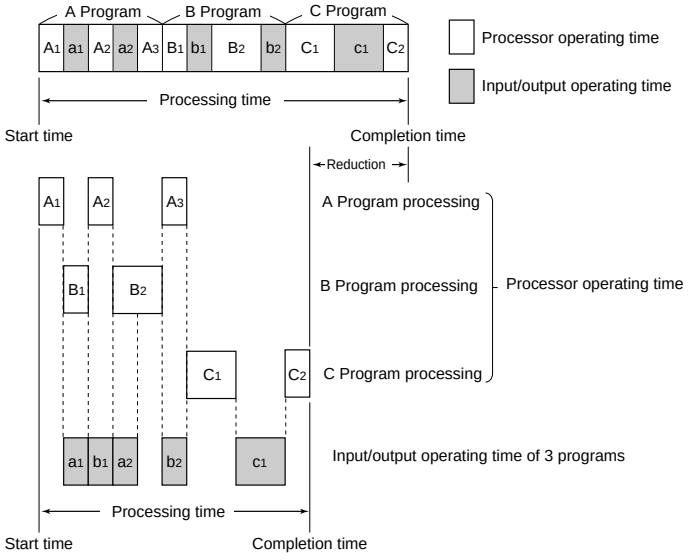
③ Multi-programming

Multi-programming consists of an approach to simultaneously processing multiple jobs with the same processor. If multiple jobs can be processed simultaneously, the computer processing efficiency will, of course, improve. The operating time of a job processed in a computer can be divided into the following:

- Time during which the data to be processed is entered
- Time during which calculations and other processes are performed using the processor
- Time during which process results are outputted

Most of the data input and process result's output are mechanical operations. Compared to them, the calculations and other processes performed using the processor are electronic operations. Therefore most of the time, the processor has to wait for the input/output operations.
For that reason, multi-programming was born from the need to separate the input/output operations and the processing operations so that the processor idle time can be used to process other job computing, etc.

Figure 3-1-4 Multi-programming approach



④ Reduction of the response time

The response time is the time elapsed since input from the terminal, etc. is completed until the system output resulting from that input is started. For example, when a reserved-seat ticket is bought via an automatic dispenser if the time elapsed since the necessary information is entered in the terminal device until the ticket is issued is too long, a long queue would be formed. For online transaction processing systems of this kind, the reduction of the response time is a factor of great importance.

⑤ Improvement of reliability

The improvement of the reliability of the different computer system components is also an important role played by the OS of general-purpose computers.

⑥ Other roles

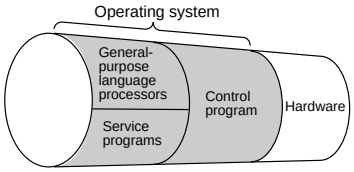
Information processing engineers do not have a thorough knowledge of every single detail of the computers. Therefore, one important function of the operating system is enabling "user friendliness" so that software development can be performed without having to keep in mind the hardware functions.
Likewise, "extensibility" of the operating system itself as well as the resources in order to support the increase of the information to be processed.

(2) OS configuration

The operating system with its complex and wide range of functions is composed of diverse programs. Figure 3-1-5 shows the relation among the following components of the operating system:

- Control program
- General-purpose language processors
- Service programs

Figure 3-1-5
Operating system configuration



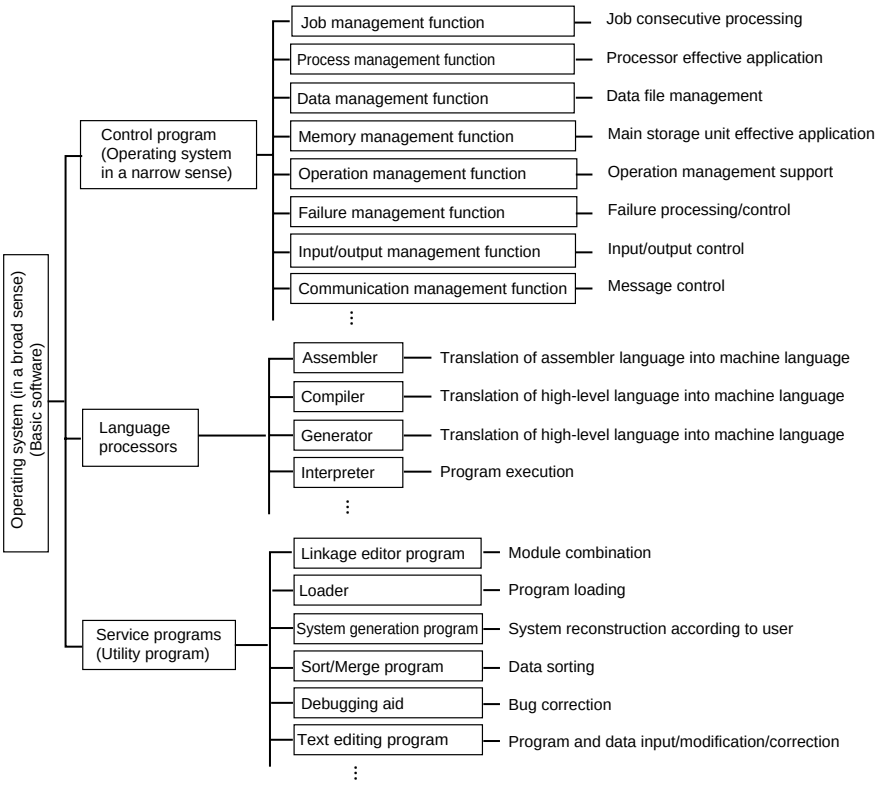
(3) OS functions

The control program, which is the nucleus of the operating system, is equipped with diverse functions such as the ones mentioned below:

- Job management function
- Process management function
- Data management function
- Memory management function
- Operation management function
- Failure management function
- Input/output management function
- Communication management function

The outline of the operating system functions is shown in Figure 3-1-6.
Here, among the different functions, mainly the control program functions, which are functions aimed at enabling efficient use of the hardware, will be explained. The language processors and service programs will be studied in Section 3.4.

Figure 3-1-6
Operating system functions



3.1.2 Job management

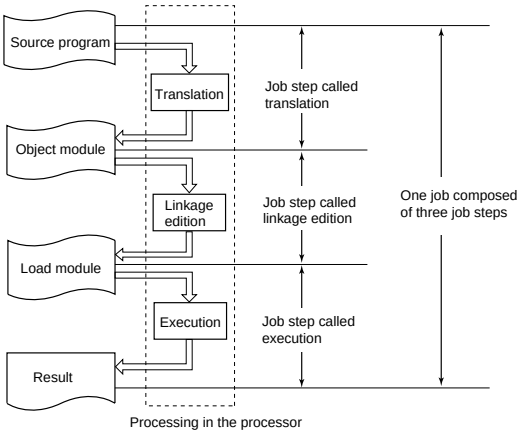
The main purpose of job management is to improve the computer system processing capacity by performing consecutive processing of the job.
In order to implement the consecutive job processing, the following are indispensable:

- Job control language
- SPOOL

(1) Job control language (JCL)

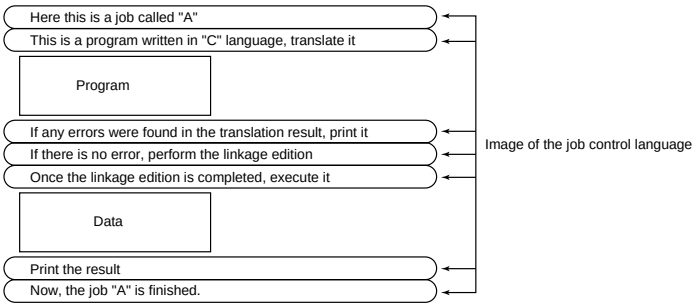
The unit of the works given to the computer by humans is called the job. Commonly, one job is composed of multiple job steps.
Figure 3-1-7, shows one job, that is, the process in which, after a given application software is submitted into the computer, data is entered and processed. The source programs written by humans become processable only after they are translated into the machine language (object module) by the language processor and edited by the linkage editor.

Figure 3-1-7
Job and job steps



In order to move from one job step to another, detailed instructions have to be provided to the computer system. The job control language is used to provide these instructions.
Since the job control language provides the instructions "Translation," "Linkage editing," "Execution," etc. to the job submitted to the computer, processing is conducted without having to rely on humans (Figure 3-1-8). The function of the job management is to decode and execute these detailed instructions written in the job control language.

Figure 3-1-8
Job control language functions



The syntax of the job control language differs depending on the operating system, but the main statements are as follows:

① JOB statement

The job to be submitted to the computer system is given a name and the job start is declared using the JOB statement.

② EXEC statement

Control information such as the order of execution of the programs performing the processing is indicated using the EXEC statement.

③ DD statement

The location where the files required for the process are located, etc. is indicated using the DD statement.

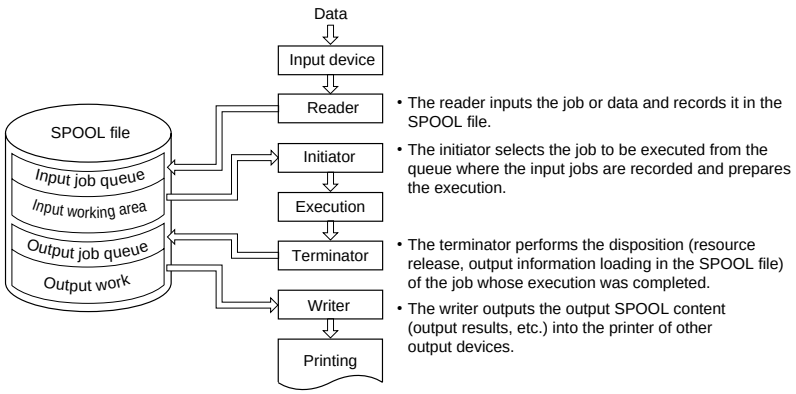
(2) SPOOL (Simultaneous Peripheral Operations Online)

SPOOL is an indispensable function in multi-programming environments.
If a given program occupies the printer to print the process result, even if the processor is free, it can not process other programs scheduled to use the printer. In order to solve this problem, for the execution of all the programs, the process result is once written out onto an auxiliary device before proceeding to print. In other words, the processor and the printer are physically separated. This is the SPOOL approach.

(3) Job scheduling

The series of controls performed after a job described in JCL, etc. is entered into the computer until the result outputted is called job scheduling (Figure 3-1-9). In practice, this processing is executed by the job scheduler using a dedicated program incorporated in the OS.

Figure 3-1-9
Job scheduling



3.1.3 Process management

For the operating system, a process (task) is the control unit of a job. The main purpose of the process management is to efficiently use the processor. In order to achieve this purpose, the operating system performs this process management.

(1) Execution control

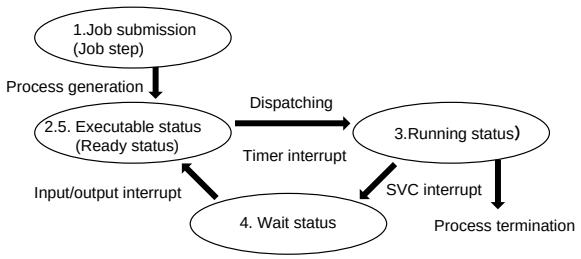
① State transition

After being subdivided into various job steps by the process management program, the jobs submitted to the computer are processed following a procedure such as the one mentioned below:

1. A job step is generated as a process that can be performed by the computer.
2. Immediately after being generated, the process turns into the executable status.
3. When the processor becomes vacant, the process in the executable status is immediately executed.
4. If input/output operations are generated while the process is being executed, the process turns into wait status.
5. When these input/output operations are completed, the process turns into the executable status again.

This procedure is called the processor state transition. The job step is converted into a processing unit called process and is processed while repeating the state transition through the process management function of the operating system.

Figure 3-1-10
Process state transition



Since the processor becomes free when the process being executed turns into the wait status, the approach of the process management consists of executing another process in the executable status during this time. The act of enhancing the efficiency of the processor use by controlling the status of multiple processes is called multiprocessing (multitasking).

② Dispatcher

The act of selecting the process to be executed from among the processes in the executable status for the processor allotment is called dispatching. The program that performs this operation is called the dispatcher. The following are the two main methods by which the dispatcher grants processor use rights:

a. Preemption

The preemption is the method by which an order of priority is given to each process and the processor is always assigned to the processes with high priority. In this method, when a process with a higher priority than the process being executed is generated, the execution of the process with lower priority is halted and the processor use is switched to the process with higher priority.

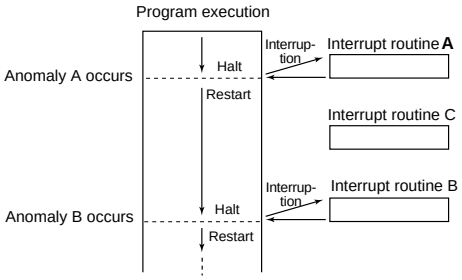
b. Round robin

The round robin is the method by which the processor use time is minutely divided (time slicing) and equally assigned to each process. In this method, once a process has used the processor for a specific time (time slice), its execution is halted and the process is sent to the last position in the process queue.

③ Kernel and interruption control

When the process state transition is performed, an interruption is performed in order to control the process execution. Interruption is the act of halting a program being executed to switch to the prevention program when a process transition or anomaly occurs (Figure 3-1-11).
This prevention program is called the interrupt routine and is stored in the computer in advance. Once the interrupt routine has been executed and the processing, etc. to fix the anomaly has been completed, the execution of the former program is restarted. The central part of the OS performing the interruption control is called the kernel.

Figure 3-1-11
Interruption



According to the location where the anomaly occurs, the interruption is divided into the following:

- Internal interrupt
- External interrupt

a. Internal interrupt

Internal interrupt is the general term for the interruptions that occur due to errors of the program itself. The internal interrupts that occur are as follows:

- Program interrupt
- Supervisor call interrupt

● Program interrupt
Program interrupt is the interruption that occurs due to an error generated during the execution of a program. For example, when the denominator of a division is zero, or when the number of digits of the result of an operation exceeds the acceptable limits, etc.

● Supervisor call (SVC) interrupt
This interruption occurs in cases where unless the operating system functions are used, a correct result can not be obtained, for example, when data input is requested during the execution of a program, etc.

b. External interrupt

External interrupt is the interruption that occurs due to external factors and not due to the program. The following external interrupts exist:

- Input/output interrupt
- Machine check interrupt
- Timer interrupt
- Console interrupt

● Input/output interrupt
Input/output interrupt occurs when an anomaly occurs in the input/output process completion report or in an input device or output device during processing.

● Machine check interrupt
Machine check interrupt occurs when a malfunction of the processor or the main storage unit or an anomaly in the power supply, etc. happen. The failure occurrence is reported to the operating system by the processor.

● Timer interrupt
Timer interrupt is an interruption generated by the timer contained inside the processor. Programs exceeding the execution time specified with the time sharing process, etc. are subject to forced termination by this interruption. Likewise, timer interrupt occurs when an abortion of programs of

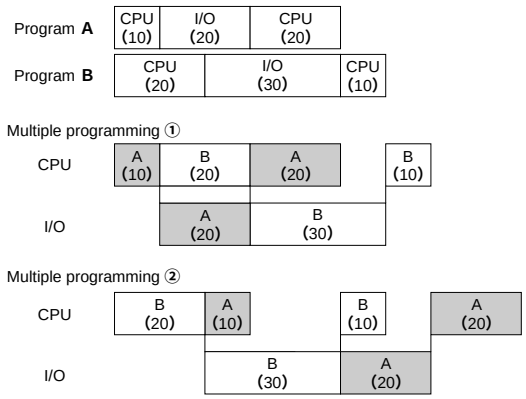
routines that never end, called infinite loops, is to be performed.

- Console interrupt
  Console interrupt is an interruption that occurs when a special process request was indicated from the operator console during the execution of a program.

(2) Multi-programming

Multi-programming is implemented by the process management to efficiently use the processor. This function enables the simultaneous execution of multiple programs through the execution of other processes while the process being executed turns into the wait status due to an input/output request.
This function is explained in Figure 3-1-12 using programs A and B as examples whereby an execution time of 50 seconds and 60 seconds for each program, respectively, is given. If program B is executed after the execution of program A, it will take 110 seconds to finish the execution of both programs (this is called the simple execution time).
However, using the multi-programming approach, which consists of the execution of one program while the input/output processing of another program is performed, the time needed to finish both programs will be 70 seconds (Multi-programming ①). It should be noted that in the event that program B is executed first, even if the same multi-programming approach is applied, the time needed to finish both programs will be 90 seconds (Multi-programming ②). From this, we can see that the execution order of the programs is very important for the processing efficiency.

Figure 3-1-12 Multiple programming



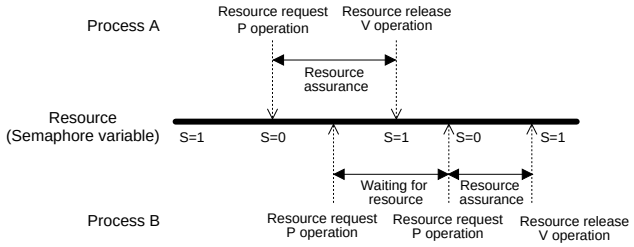
(3) TSS (Time sharing system)

By assigning equal CPU time to all the processes according to the round robin method, multiple users can simultaneously use one computer. This user format in which "one feels like the only user of the computer" is called TSS. TSS is one of the main interactive-type processing and it is used by a great number of centralized processing systems of the conventional host computer use.

(4) Exclusive control

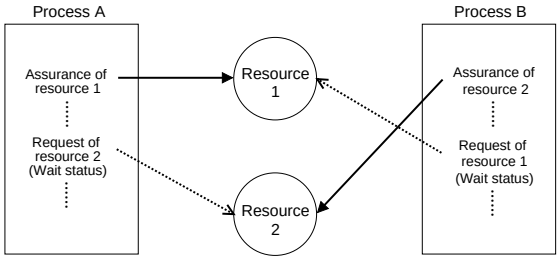
In the process generation stage, besides the processor, resources are assigned to each process. Here, the same resource can be shared by multiple processes; however, the same resource can not be used by all the processes at the same time. Therefore a semaphore is used to limit the resource use (exclusive control).
The semaphore, which in the broad sense is a word that means signal, is composed of semaphore variables and 2 operation instructions (P operation and V operation). The semaphore variables hold integer values according to the condition of each resource, and according to the integer value, the synchronization among the processes is conducted. In the binary semaphore, which is a typical semaphore, the semaphore values are 0 and 1. P operation is the operation for resource use that reduces the semaphore variable value. On the other hand, V operation is the operation for resource release that increases the semaphore variable value. Figure 3-1-13 shows an example of exclusive control using a binary semaphore.

Figure 3-1-13
Semaphore



Through the exclusive control performed using the semaphore, synchronization among processes is conducted and resource sharing is implemented. However, due to this control an event called deadlock can occur. Deadlock is the status in which two or more processes wait for the resource release of each other. Since the processes in this status are unable to assure a resource, processing is halted.

Figure 3-1-14
Deadlock



3.1.4 Main memory management

The main memory management controls the storage area of the main storage unit.
The following main memory management techniques exist:

- Partition method
- Swapping
- Overlay
- Memory protection

(1) Partition method

In the program storage method (or program built-in method) it is necessary to store programs and data in the main storage unit in advance. When programs are to be stored in the main storage unit, the method that divides the main storage unit into several parts, and stores programs in each of these parts is called the partition method, because these parts are called partitions.
The partition method can be roughly divided into the following three methods:

① Single-partition method

In the single-partition method, the main storage unit is controlled by dividing it into the area to store the control program and the area to store only one program. This method was applied in the early computers, but since it is not suitable for the efficient use of the main storage unit and other resources, it is practically obsolete today.

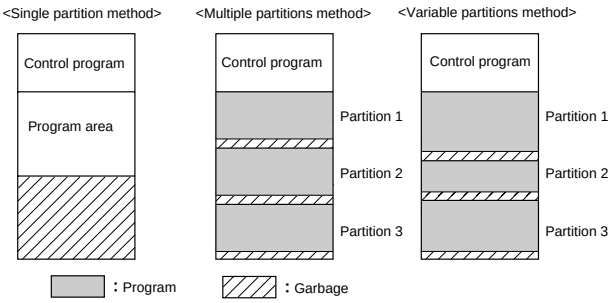
② Multiple partitions method

In the multiple partitions method, the program storage area is divided and multiple programs are stored in each of the partitions divided. This method was conceived for the implementation of multi-programming, however, since the main storage unit is subdivided, its results are inadequate for performing processing on programs that exceed the partition capacity.

③ Variable partitions method

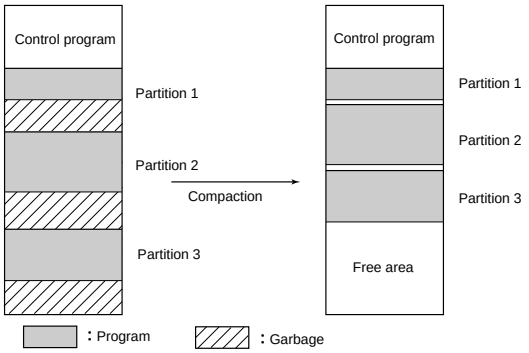
The variable partitions method is the method that sequentially assigns the area required by application programs in the program storage area. The main memory management program of the control program performs the area allotment.

Figure 3-1-15
Partition method



However, in these methods, areas that are not used (garbage) are generated in each partition of the main storage unit. This phenomenon is called fragmentation.
In order to solve this fragmentation, it is necessary to reset each partition at specific times or at specific intervals. This operation is called compaction (Figure 3-1-16).

Figure 3-1-16
Compaction

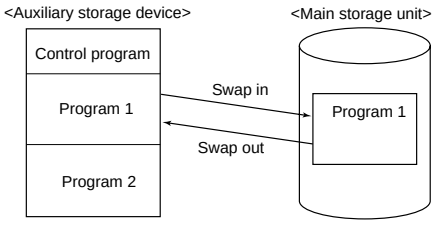


However, when compaction is performed, the address of each program instruction stored in the main memory unit changes. In order to solve this problem, it is necessary to reset and update the address of each instruction. This operation is called program relocation.

(2) Swapping

When multi-programming is performed in multiple partitions or other methods, if a job with high priority is generated, the job with low priority being executed has to be interrupted. In that case, in order to avoid letting the interrupted processing come to nothing, it is saved in an auxiliary storage device as it was when interrupted. This operation is called swap out (or roll out). On the other hand, the operation by which a job with high priority is transferred from an auxiliary storage device to the main storage unit is called swap in (or roll in). This kind of exchange of jobs between the main storage unit and the auxiliary storage devices is generically called swapping.

Figure 3-1-17
Swapping



(3) Overlay

Overlay is used for the execution of programs that are larger than the storage capacity of the partitions of the main storage unit.
Through the overlay technique, the application program is divided beforehand into units called segments, and after each of the segments is stored in the main storage unit, the program is executed.

(4) Memory protection

In order to avoid the misappropriation or destruction of the contents of the main storage unit, memory protection becomes necessary. Among the main memory protection methods, the following three methods can be mentioned.

① Boundary address method

The boundary address method is the method by which the address range that can be accessed is specified for each of the programs to be executed. The memory is protected by checking whether or not the access to the main storage unit to be executed is within the address range.

② Ring protection method

The ring protection method is the method by which a ring number is assigned to each program and the access is controlled according to the number size. A small number is assigned to important programs (OS, etc.) and a large number is assigned to user programs, etc. In this method, access from small numbers to large numbers can be performed, but in the opposite case, access can only be performed through service provision.

③ Keylock method

The keylock method is the method by which the main storage unit is divided into multiple partitions and each partition is locked for memory protection. Each program to be executed has its respective memory protection key(s) and access is authorized when the memory can be unlocked (the key and the lock match).

(5) Other main memory management

① Dynamic allocation

The dynamic allocation is the technique by which the main storage unit is dynamically assigned during the program execution.

② Memory leak

A memory leak occurs due to the failure to release the area that should have been released by a program that used the main storage unit, reducing as a consequence, the area of the main memory that is available for use. However, due to the volatility of the main storage unit, if the power is turned off, all the storage area is released. These kinds of events especially occur in servers, etc. that remain operational 24 hours a day.
Since the memory leak is not an event that occurs in all the OS, it is necessary to check the OS product information.

3.1.5 Virtual storage management

In the main storage unit, operations such as swapping and overlay become necessary in order to execute programs that are larger than the partition size of the storage area or to change the processing order. The development of programs under this kind of restriction can not be considered productive. For that reason, the approach of the virtual storage, which enables the execution of programs without worrying about the storage capacity of the main storage unit, was born.
The basic approach to implement virtual storage is as follows.

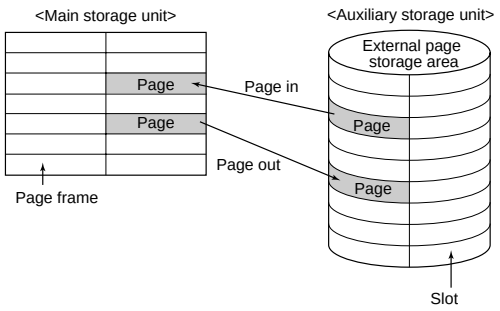
- The main storage unit is divided into partitions of a specific size. These partitions are called page frames.
- The program is temporarily stored in an area called the external page storage area of an external storage device.
- The external page storage area is divided into partitions called slots, which have the same size as the page frame. Therefore, the programs stored in the external page storage area are automatically divided into parts the size of a slot.
- The programs stored in the page frames or the slots are called pages. Generally the size of one page is 2 kilo bytes. Of course, the size of a page frame and the size of a slot are also 2 kilo bytes.
- The external page storage area of the main storage unit and the auxiliary storage device is called the logical address space.
- Among the programs stored in the external page storage area, the pages of the slots needed for execution are transferred to empty page frames of the main storage unit to be executed.

In this way, in the virtual storage method, execution is repeatedly performed by transferring the programs that are stored by page unit in the external page storage area to the page frames of the main storage unit. The act of transferring a program to the main storage unit is called load.

(1) Paging

The exchange of programs between the main storage unit and an auxiliary storage device is generically known as paging.
The transference of a slot from the external page storage area of the auxiliary storage device to the page frame of the main storage unit is called page in The transference in the opposite direction, when a page whose execution has been completed is transferred to the slot, is called page out.

Figure 3-1-18
Paging

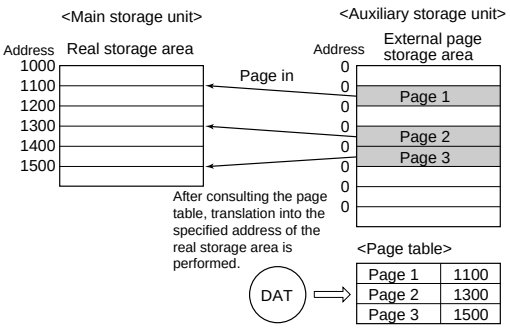


In the multi-programming method, there are cases where paging occurs frequently. This condition is called slashing.

(2) Address translation

An issue that often arises when paging is performed is that the page-in address of the main storage unit is unknown. Since in the virtual storage method, when a page frame becomes vacant, the next page to be executed is "paged in" to this page frame, it is necessary to translate the instruction address according to the address of the page frame. This conversion is called address translation.
The address assigned to each instruction of the programs stored in the external page storage area is called the static address and the address stored in the page frame of the main storage unit after the address translation is performed is called the dynamic address. The main address translation method is called the dynamic address translation (DAT), which is a method performed using the hardware.
The DAT performs address translation at the time the instruction paged in is executed. The addresses of the external page storage area start from address 0 and increase by page unit, while the addresses of the pages to be "paged in" are converted into dynamic addresses after consulting the page table.

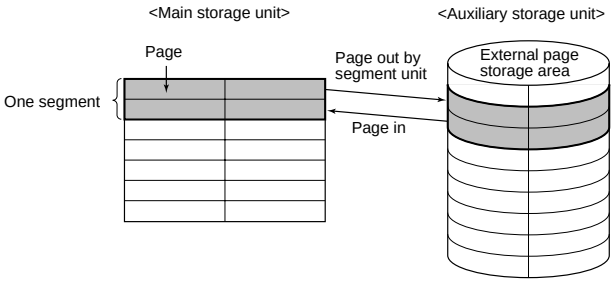
Figure 3-1-19
Address translation



(3) Segmentation paging

A group of pages logically related is called the segment. In segmentation paging, page in and page out are performed by these segments. Compared to the method in which paging is performed by page units, in this method paging occurs less frequently.

Figure 3-1-20
Segmentation paging



(4) Page replacement

In the page replacement, in order to achieve system processing efficiency, pages with a high application frequency are permanently stored in the main storage unit while pages with a low application frequency are stored in the external page storage area and are transferred to the main storage unit only when they are needed.
In this case, the following two methods are used to send out the pages from the main storage unit (paging algorithm).

① LRU (Least recently used) method

In the LRU method, among the pages of the page frame of the main storage unit, the page for which the time elapsed since it was used the last time is the longest is sent out.

② FIFO (First-in first-out) method

In the FIFO method, the page that was the first to be stored among the pages of the main storage unit, is sent out to the external page storage area.

3.1.6 File management

The data processed by the computer is controlled by the data management function of the operating system. Since most of the data is stored in auxiliary storage devices, file handling plays a central role in this data management. Therefore, it is also called file management.

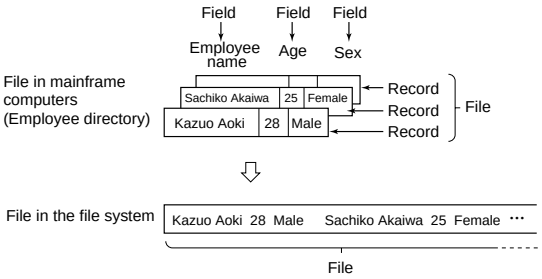
(1) File system

The concept of a file in personal computers and workstations differs from that of mainframe computers. The concept of low-end computer files, and the file system controlling those files, will be explained below.

① File concept and configuration in low-end computers

In personal computers or workstations, there are no concepts of records or fields as in the files handled in mainframe computers. Files simply record character strings, and there is no difference between data and programs.

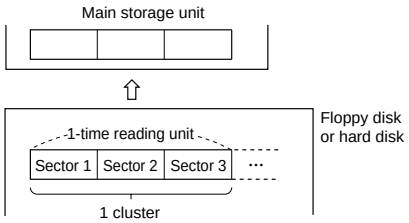
Figure 3-1-21 File concept in low-end computers



a. Cluster

The data sets composing a file are written on floppy disks (flexible disks) or hard disks in units called clusters. A cluster, which is a set of several sectors, is the input/output unit between these auxiliary storage devices and the main storage unit.

Figure 3-1-22
Cluster

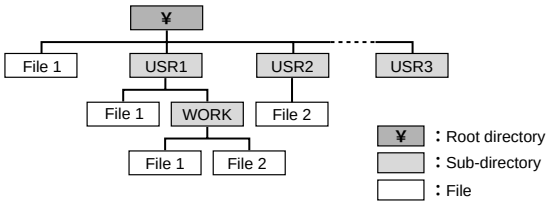


When data cannot be filed in one cluster, an unused cluster is coupled and the remaining data written in this cluster. When this operation is performed, it is not necessary for the coupled cluster to be a sequential cluster in the storage area.

b. Directory

The file system is composed of the directory and the file. The directory is the register where the file management information is recorded and stored. It is possible to have files and directories beneath a directory. The file system organizes these directories and files in a hierarchical structure to control them (Figure 3-1-23).
The highest directory of the hierarchical structure is called a root directory and it is an important directory in the aspect of volume. The directories positioned beneath the root directory are called sub-directories.

Figure 3-1-23 Hierarchical structure of the file system



② File operation

When personal computers and workstations start up, the directory is automatically set by the operating system. Normally, it is a root directory, but the user can freely set it.
It should be noted that the user has to move to the target directory in order to access a directory or file.

a. Directory

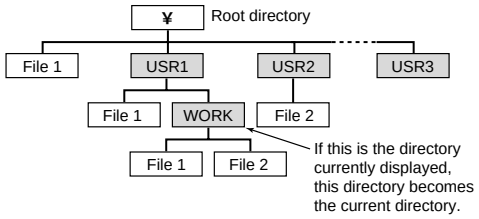
1) Home directory

The directory that can be freely used by the user is called the home directory. The user can freely create and access the sub-directories and files registered in the home directory.

2) Current directory

The current directory is the directory in current use. If the home directory is being used, the home directory will be the current directory.

Figure 3-1-24 Current directory



b. Path

When a file is sought inside the file system, the route along which to search for that file is specified. This route is called a path.

Depending on the specification method, paths are classified as follows:

- Absolute path
- Relative path

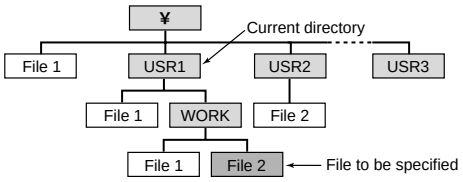
1) Absolute path

The absolute path is the path to the target directory or file from the root directory, which is at the highest position of the file system. In this specification method, all of the directories and files, from the root directory to the target directory or file, are written, using the ¥ sign or the / sign to separate them. In Figure 3-1-25, the absolute path to specify file 2 is as follows:

¥USR1¥WORK¥FILE2

The "¥" at the top represents the root directory.

Figure 3-1-25
Path specification



2) Relative path

The relative path is the path to the target directory or file from the current directory. In Figure 3-1-25, if the current directory is USR1, the relative path to specify file 2 would be as follows:

WORK¥FILE2

If the current directory is WORK, the relative path would simply be the following:

FILE2

Likewise, if the current directory is USR3, it is necessary to go up to the root directory once. Since ".. " is used to specify a directory that is one level higher, the relative path would be as follows:

..¥USR1¥WORK¥FILE2

c. Command

In the operating systems of personal computers and workstations, programs are executed as a result of the input of commands. The commands related to file operations in the case of MS-DOS are shown in Figure 3-1-26.

Figure 3-1-26
Commands used to perform file operations

Commands	Functions
DIR	Display the file name(s) contained in the specified directory
DEL	Delete the specified file(s)
REN	Change the specified file name
TYPE	Display the file(s) content on the screen
COPY	Copy the file(s) in other directory or volume
PRINT	Print the specified file(s) content
MKDIR	Create a sub-directory (or subdirectories) beneath the current directory
RMDIR	Delete the specified sub-directory
CHDIR	Transfer the current directory to the specified directory

For example, when the file name "File 1" is changed to "File 2," the change command REN is entered after the input prompt (A>) as follows:

A>REN FILE 1 FILE 2

The "A" written before the prompt (>) indicates the location of the device subject to the operation. It is called the current drive.

d. Extension and wild cards

The file names used in the file system are expressed using a file name of 8 or less alphanumeric characters (○○○) and an extension of 3 or less alphanumeric characters (ΔΔΔ) separated by a period. "○○○. ΔΔΔ"
Among the extensions, there are some that are given a special meaning by the operating system, as well as some that are freely set by the users, or uniquely set by application software.
When a file is specified, the following wild cards (? or *) can take the place of file names and extensions:

- ?: Any character can be placed in the ? position. That means any single character.
- *: Any character(s) can be placed in the * (and subsequent) position(s). The character types or character string lengths are not specified.

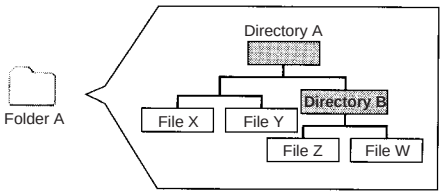
For example, the command, "Show all the files whose extension is BAK," would be specified as follows:

DIR*.BAK

e. File operations in GUI environments

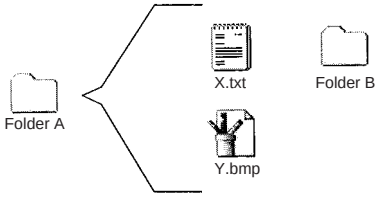
In the case of MS-DOS, files are operated using command operations consisting of character input. However, in recent years, GUI (Graphical User Interface) environments, in which icons on window screens are manipulated using a mouse, have become the mainstream.
In GUI environments such as MacOS and Windows, directories are represented by folder icons, and files are represented using icons set according to the extension. It should be noted that a user is free to design his/her own icons (Figure 3-1-27).

Figure 3-1-27
Folder and icon



Transfers to the current directory and access to files can be performed by clicking or dragging folders and icons with the mouse. To open a file means to display the contents of a folder on the screen. The screen displayed when Folder A of Figure 3-1-27 is opened is shown in Figure 3-1-28.

Figure 3-1-28
Opening Folder A



3.1.7 Security management

The protection of a computer and its resources from diverse menaces (natural disasters, failures, human errors and intentional malice) is called information security or computer security.
Security management aims at the achievement of the following three specific characteristics:

- Confidentiality
  Prevents the leakage of information contained in a computer due to illegal access, etc.
- Integrity
  Prevents the modification of information contained in a computer due to illegal access, etc.
- Availability
  Prevents the obstruction of the use (information reference or modification) by a legitimate user.

In general, the OS performs security control through access control and flow control.

① Access control

Access control is that which limits direct access to computer resources to legitimate users only.

② Flow control

Flow control is that which prevents the leakage of information to users that are not authorized as legitimate users.

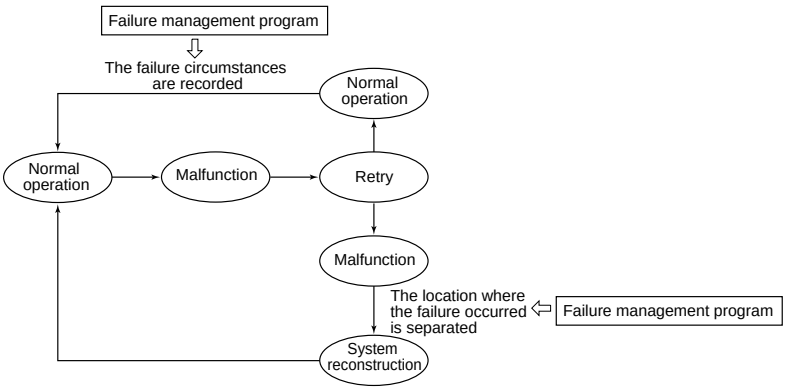
3.1.8 Failure management

As the system becomes larger, the impact of a failure becomes larger too.
Since a computer system is an extremely complex device, it is not easy to find what is wrong. In order to cope with these problems, the operating system is equipped with the following functions:

- Instruction retry
- Failure management program

The instruction retry is a function that automatically retries the execution of an instruction when a malfunction occurs, as is shown in Figure 3-1-29. If the computer operates normally after an instruction retry is performed, the failure management program records the failure circumstances. If a malfunction occurs again after an instruction retry is performed, the program isolates the location where the failure occurred and reconstructs the system. The failure circumstances record helps in selecting the check points to be emphasized at routine inspections, contributing to the MTTR (Mean time to repair) reduction as well.

Figure 3-1-29 Instruction retry and failure management program



3.1.9 Supervisor

The supervisor is a monitoring program functioning as the central part of the OS. It performs resource distribution and program control in order to implement the TSS, multi-programming, etc. The processing program sends an interrupt instruction called a supervisor call (SVC) or system call in order to request a special service from the supervisor. As a result of this instruction, the SVC interrupt is generated, the program (process) execution is temporarily interrupted and control is transferred to the supervisor.

3.2 Types of OS

Until the previous section, explanations were given with the general operating systems of mainframe computers in mind. Here, based on that knowledge, the operating systems that are actually widely used will be explained.

3.2.1 General-purpose OS

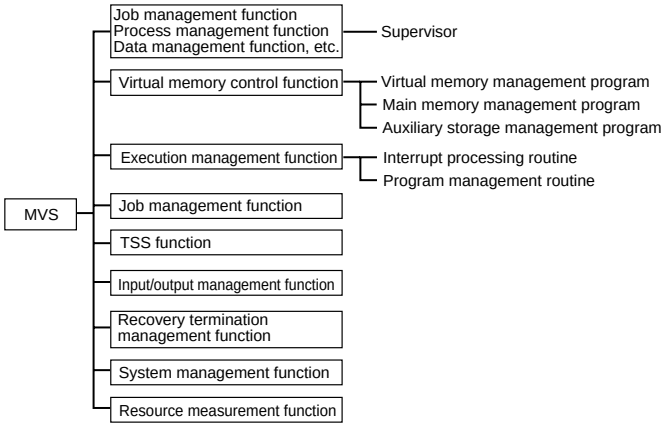
Diverse OSs are used in today's computers. As characteristics of OSs of recent years, the improvement of human interfaces using GUI, etc., the fulfillment of multimedia processing that enables easy use of audio and image data, etc., can be mentioned. Likewise, regarding the OS used in Japan, Japanese language processing functions have seen extreme improvement.
The main computer OS will be explained below.

(1) MVS (Multiple Virtual Storage)

The MVS is the most representative operating system adopted in high-end mainframe computers. Since this OS was developed by IBM for its own computers, it was introduced into the market as "MVS/370" in the 1970s, but was repeatedly upgraded thereafter. In MVS, 32-bit words are the foundation for everything.

① MVS configuration

Figure 3-2-1
MVS configuration



In MVS, besides the conventional logical address space, a data space and a hyper space, where only data is stored, are also provided. This was set up in order to reduce input/output frequency. In this system, data used frequently is stored in advance in the data space and hyper space, and the programs of the logical address space directly check the data space.

② MVS characteristics

The following points can be mentioned as MVS characteristics:

- It is the operating system for high-end general-purpose computers.
- It provides a multi-user function, which enables simultaneous MVS use by multiple users.
- It provides a multi-task function, which enables simultaneous processing of multiple tasks.
- It adopts the approach of multiple address space.
- One logical address space reaches up to 2 gigabytes.
- It has all the file organization functions.

(2) UNIX

UNIX is an operating system developed by AT&T Bell Laboratories, widely used in computer network systems that have personal computers, workstations, etc. connected by telecommunication lines. Since Version 1.0 was launched in 1969, it has since been upgraded in diverse ways.
The most distinctive characteristic of UNIX is that, unlike other operating systems, detailed contents of the operating system written in C language have been released to the public. For that reason, a great number of computer manufacturers, besides AT&T Bell Laboratories, can easily port it to the hardware of their own products. As a result, users are able to operate UNIX in all computer manufacturer products.
The following can be mentioned as the most representative examples of UNIX upgraded editions:

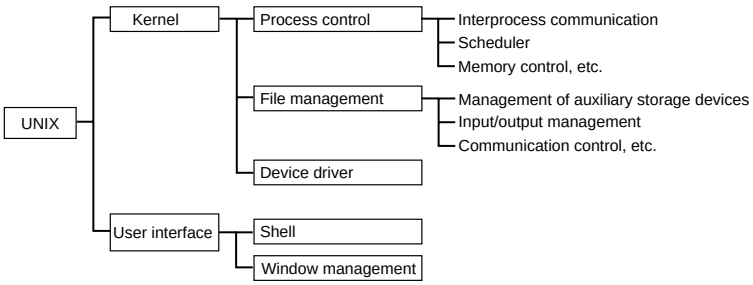
- XENIX (Microsoft)
- AIX (IBM)
- Ultlix (DEC)
- SunOS, Solaris (Sun Microsystems)

Even though their names differ, they are basically UNIX operating systems.

① UNIX configuration

UNIX is an operating system which can be simultaneously used by multiple users, and in which each user can simultaneously perform multiple job processing. Its configuration is shown in Figure 3-2-2.

Figure 3-2-2
UNIX configuration



UNIX has a control program called Kernel, which has the following functions:

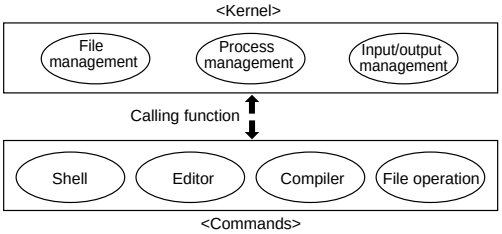
- It is the central part of the operating system, and controls the system resources.
- It performs the process management. (Since UNIX is distributed-processing oriented, jobs are called processes.)

Likewise, there are command sets that directly instruct jobs to UNIX. These command sets are composed of shells and commands. A shell has the following functions:

- It interprets the commands input by the users and calls the kernel function.
- It executes a program called a shell script, which combines commands.

Likewise, the command uses the devices connected to the system, and can call the kernel function to use the command sets.

Figure 3-2-3
Kernel and command



② UNIX characteristics

UNIX is an operating system which can perform distributed processing in computer network environments. Its characteristics are mentioned below.

- Distributed processing is presupposed.
- It was developed assuming that it would be used in workstations.
- It provides the multi-user function, which enables simultaneous use by multiple users. In operating systems that provide the multi-user function, a procedure called log-in, to receive the service, and a procedure called log-out, to report completion, are required.

- Through multi-programming, it can simultaneously process multiple jobs. In UNIX, this is called the multiprocessing function.
- As a technology to connect computers of different manufacturers, the communication protocol called TCP/IP has been established.
- It has instruction rights called commands that enable the user to use UNIX in an interactive mode. This function is called human interface, and has been implemented by X-Window.
- The program development tools are abundant.

(3) Windows

The operating system of more than half of the personal computers around the world is Windows.
There are the following Windows versions:

- Windows 3.1
- Windows 9x (Windows95/98/ME)
- Windows NT
- Windows 2000
- Windows XP

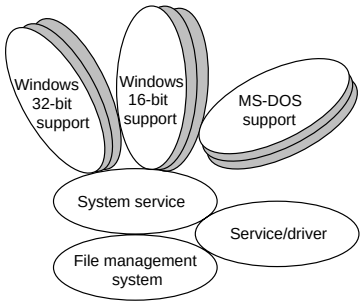
① Windows history

The first personal computers had 16-bit words, and IBM personal computers, called PCs, were the mainstream. The operating system adopted for these PCs was MS-DOS with single-task functions, developed by Microsoft.
Afterwards, with the appearance of 32-bit-word personal computers, Windows was born.
Inheriting the MS-DOS functions without changing them, Windows fulfilled the GUI environment and had outstanding operability, therefore becoming a worldwide best seller. (It is said that in the U.S., alone, 2 million sets were adopted.) However, since it inherited the basic concept of MS-DOS, it was not able to master the hardware functions of 32-bit words.
As a result, in 1995 Microsoft introduced into the market a new operating system that fulfilled multimedia functions, communication functions and network functions, while inheriting the unchanged concept of Windows. This operating system is Windows 95.
On the other hand, Windows NT was developed completely independently, without inheriting the restrictions of past operating systems. It is used as the server operation system in client/server systems, etc. (NT are the initial letters of New Technology). It should be noted that Windows NT upgrades, up to Version 4.0, have been put on the market, and Version 5.0, launched in 1999, was named Windows 2000.

② Windows 9x configuration

Windows 9x (Windows95/98/ME) is an operating system that capsules MS-DOS. In the time since the file system was newly created, constraints have been substantially reduced.
Figure 3-2-4 shows an image of the configuration of Windows 9x. However, since the functions of Windows are included, it still maintains the 16-bit-word MS-DOS environment.

Figure 3-2-4
Windows 9x configuration



③ Windows 9x characteristics

Since an outstanding GUI environment is provided, and it is widely used around the world, in practice Windows 9x has become a *de facto* standard.
GUI characteristics are mentioned below.

- Desktop approach

Considering the display screen as one desk, screens can be used as though several documents were spread over the desk. These screens are called windows.

- Multi-task function
  Not only can multiple windows be simultaneously displayed on the display screen, a multi-task function is also provided.
- Mouse pull down menu/dialog box manipulation
  With one mouse, a great variety of menus can be selected/operated.

④ Windows XP

Windows XP is the successor of Windows 2000 and Windows Millennium, internally, built on the enhanced Windows 2000 code base.
It is the Windows operating system developed by integrating the strengths of Windows 2000-standards-based security, manageability and reliability with the best features of Windows 98 and Windows Me such as Plug and Play, easy-to-use and familiar user interface and so on.
There are different versions of Windows XP aimed at home users and business users, Windows XP Home Edition and Windows XP Professional respectively.

(4) MacOS

MacOS is the OS developed by Apple Computer for its own product (Macintosh)
and:

- Almost all the operations can be performed with the mouse
- The operation method of application software is almost the same, etc.

The user interfaces are abundant. For that reason, it is said that MacOS is an OS that is easy for beginners to use.
The OS "MacOS X," for the client, and the OS "MacOS X Server," for the server, launched in 1999, integrate the former "MacOS 8" and "Rhapsody," and was announced as a new OS.

(5) Linux

Linux is the UNIX-based OS launched in 1991 by Linus Torvalds. The main characteristic is that the software is free. Since the source code has been released to the public, and redistribution and changes can be freely performed, a great number of people around the world have participated to make Linux a better OS. As a background factor, it should not be forgotten that the Internet expansion that enabled people around the world to communicate with each other allowed this participation.
It should be noted that the copyright is protected by GPL (GNU Public License).

3.2.2 Network OS (NOS)

The network OS is the OS used to construct LANs, in which computers are connected and used through a network. Besides providing the same services as a computer OS, based on the SNMP (Simple Network Management Protocol), it provides network management functions.
As the main network OS, NetWare and LAN manager will be explained.

(1) NetWare

NetWare is a network OS developed by Novell. It is the most common NOS, with file sharing and printer sharing functions.

(2) LAN manager

The LAN manager is a network OS developed jointly by Microsoft and 3 Com. The functions of this network OS were inherited from Microsoft's OS "Windows NT."

3.3 Middleware

Middleware is positioned between basic software and application software. This software provides the basic processing functions that are used in common by users.
Among main middleware, the following, whose applications are diverse, can be mentioned:

- DBMS (Database management system)
- Communication management system
- Software development support tool
- Operation management tool
- ORB
- Japanese word processor
- Spreadsheet software
- Graphic processing system

3.3.1 DBMS

DBMS (Database Management System) is dedicated software aimed at efficient database creation/maintenance/operation. The following are the three main characteristics:

① Integrity

Even when the database is simultaneously used by multiple users, it prevents the generation of data inconsistency.

② Security

It protects data secrecy by setting database access rights, etc.

③ Failure recovery

In the event that a failure occurs in a database, it promptly recovers that database.

3.3.2 Communication management system

The communication management system is software aimed at supporting computer network construction/operation. A recent tendency in software of this kind is to emphasize LAN control. The following are the three main characteristics of the communication management system.

① Network independence

In order to facilitate network construction, lines, communication equipment, and other network environments are separated from user programs.

② Network flexibility

Through the provision of flexibility to the devices and network mechanism that make up the network, the construction of network systems with high expandability is enabled.

③ Network transparency

This provides an environment in which network users can use the system without being aware of the network.

3.3.3 Software development support tool

A software development support tool is software that supports computer-aided software development. As software development support tools aimed at achieving development labor saving as well as quality improvement, CASE (Computer Aided Software Engineering) tools exist. Depending on the content supported, CASE tools are classified as follows:

① Upstream CASE tools

Upstream CASE tools support the high-end process (analysis, design, etc.) of software development.

② Downstream CASE tools

Downstream CASE tools support the lower-end process (programming, testing, etc.) of software development.

③ Maintenance CASE tools

Maintenance CASE tools support the operation and maintenance of the developed software.

④ Integrated CASE tools

Integrated CASE tools support overall functions from upstream CASE tools to maintenance CASE tools.

3.3.4 Operation management tool

An operation management tool is software aimed at supporting the operation duties of system operation managers. Among the operation management characteristics, the following can be mentioned:

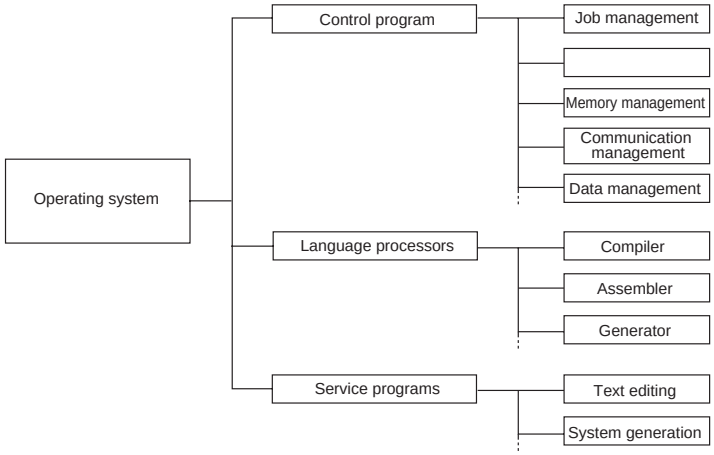
- Multiplicity optimization of multi-programming
- Allocation of system resources in order to reduce response time at peak periods
- Grasp of the operating conditions of system resources
- Recording of accounting information and creation of summaries
- Logging of operation records

3.3.5 ORB

ORB (Object Request Broker) is software used for the creation of object requests and responses, as well as for communication between objects in object environments. CORBA (Common Object Request Broker Architecture), which was completed as a standard specification by the Object Oriented Management Group (OMG), is among the most representative software of this kind.

Exercises

Q1 **The following diagram shows the relation of some of the functions of the operating system. Which is the appropriate function to fill in the blank?**

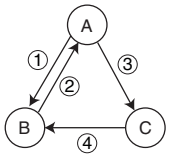


a. Overlay management b. Catalog management c. Process management
d. Project management e. Message management

Q2 **Which of the following is the most appropriate description of spooling?**

a. Provides a standard communication procedure regardless of the other devices and the communication network.
b. Using an external storage device, provides a virtual memory larger than the main storage unit.
c. Records the information related to the computer system operation process.
d. The operation of peripheral devices is separated and performed in parallel to the processor operation.
e. Enables processing on a logical record basis without having to worry about the physical record.

Q3 **The following diagram shows the process state and transition. Which of the following is the correct combination of A, B, and C states?**



The causes of status transition:
① The CPU use right was transferred to a process with a high execution priority.
② The CPU use right was provided.
③ Wait for the completion of the input/output operation.
④ The input/output operation has been completed.

	A	B	C
a	Executable status	Execution status	Wait status
b	Executable status	Wait status	Execution status
c	Execution status	Executable status	Wait status
d	Execution status	Wait status	Executable status

Q4 **Which of the following corresponds to the cause of internal interrupt?**

a. Occurrence of anomalies in the computer power-supply unit
b. The counter that measures clock time inside the processor has exceeded the preset value
c. Input/output device operation completion or failure occurrence
d. Occurrence of overflow in floating point operations

Q5 **Given the two programs, A and B, the occupancy time of the processor (CPU) and the input/output devices (I/O) when each program is executed separately is shown in the diagram. Considering that programs A and B are started simultaneously in the same CPU, how many milliseconds after the startup will program B be completed? The program execution conditions, etc., are as follows:**

① **A has a higher program execution priority than B.**
② **Programs A and B use the same input/output devices.**
③ **The execution of programs in the CPU is not interrupted until the input/output processing is started.**
④ **The execution of input/output processing in input/output devices is not interrupted until it is completed.**
⑤ **The time needed for the CPU task switching can be ignored.**

Program A — Milliseconds

CPU	→	I / O	→	CPU	→	I / O	→	CPU
20		30		20		40		10

Program B — Milliseconds

CPU	→	I / O	→	CPU	→	I / O	→	CPU
10		30		20		20		20

a. 120 b. 140 c. 160 d. 180

Q6 **Which of the following is used in process mutual exclusion (exclusive control)?**

a. Contention b. Semaphore c. Check point d. Hash

Q7 **In the operating system, a large number of small unused portions in the memory result from the repetition of the allocation and release of the memory space. What is the name of this phenomenon?**

a. Compaction b. Swapping c. Fragmentation d. Paging

Q8 **Which is the processing that transfers a program being executed to an auxiliary storage device in order to load and execute a program with a higher priority level?**

a. Overlay b. Swapping c. Paging d. Relocation

Q9 **Which is the method that divides the storage space into specific sizes, manages it, and implements virtual storage?**

a. Thrashing b. Swapping c. Blocking d. Paging

Q10 **Which of the following is the most suitable explanation of dynamic address translation?**

a. It is the translation of virtual addresses into real addresses in the virtual storage system.
b. It is the act of changing the base address of a program being executed in order to transfer and execute it in a new location.
c. It is the vicarious execution of the main memory reading and writing by the cache memory.
d. It is the act of resolving address references between modules in order to add a module during the execution of a program.

Q11 **Which of the following is the explanation of the LRU, which is one of the page replacement algorithms of the virtual memory?**

a. The page with the lower priority according to a priority level established in advance is expelled.
b. The page whose period of existence in the main storage unit is the longest is expelled.
c. The page whose period of existence in the main storage unit is the shortest is expelled.
d. The page that has not been referenced for the longest period is expelled.

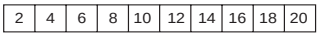
Q12 **Which is the most suitable explanation of indexed sequential organization, which is one the file organization methods?**

a. Direct access to the records can be performed using the address of each record. Sometimes, the efficiency of the medium use is low.
b. The records are recorded in the order in which physical writing is performed. Only sequential access can be performed.
c. It is composed of a data area called member and a directory area that controls the member information. It is suitable for storing programs.
d. It is composed of an area to store the records and an area to store the record key information.

Q13 **Considering that 10 records, whose keys are the numbers shown in the figure, are to be stored in direct organization files, if a division method in which 7 is the divisor is used as the hashing (address translation) method, how many records would be synonym records? It should be noted that in hashing using a division method,**

Key value ÷ Divisor = X with the remainder Y

Y is the record address.



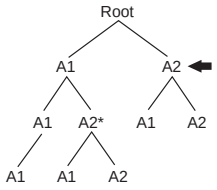
a. 1 b. 2 c. 3 d. 4 e. 5

Q14 **Given an operating system performing file management using a directory with a hierarchical structure, which of the following is specified to indicate the directory where the file is located?**

a. Extension b. Sub-directory c. Path
d. Root directory e. Wild card

Q15 **Directories A1 and A2 are managed with the structure shown in the diagram. In each directory a file, f, exists. Which is the method to specify the file, f, located beneath the directory pointed with an arrow, from the directory with the asterisk (current directory)? Here, the file specification method is based on the following:**

① **The directories on the route are sequentially specified, separating them with "¥," specifying the file in the following way.**
"**DIRECTORY NAME¥…¥ DIRECTORY NAME ¥FILE NAME**"
② **The current directory is represented with ".".**
③ **The directory that is one level higher is represented with ".. ".**



a. .¥A2¥f b. ..¥..¥A2¥f c. ..¥A1¥..¥A2¥f d. ..¥A2¥f

Q16 **Considering a character string composed of multiple alphabetic characters and 1 delimiter ". ", if "*" represents any character string larger than 0, and "?" represents 1 character, which of the character strings corresponds to the representation shown below?**
X*.Y??

a. XY.XYY b. XXX.YY c. XYX.YXY d. YXYX.YXY

Q17 **Which of the following is <u>not</u> a correct explanation of UNIX which is one of the operation systems (OS)?**

a. Provides an interactive human interface that uses character-based commands.
b. Since its specifications have been released to the public and it has a high portability, it has been adopted in a wide range of devices.
c. It is a single-user and multi-task OS.
d. Provides network functions that easily implement distributed processing.
e. It is the most representative workstation OS.

4 Multimedia System

Chapter Objectives

The objective of this chapter is to understand the multimedia system, which occupies the most important position in the current computer system. Likewise, the basic technology that implements the multimedia system will be studied.

① Understand the meaning of multimedia and the multimedia service outline.
② Understand the technology that supports the multimedia system, in particular audio and image related technologies.
③ Consider future multimedia-related application systems.

Introduction

The computer, which was created as a calculating machine, has seen its application range extended without limits, and audio and image processing, which formerly were considered as its weak points, have been made possible.
In this way, the system in which simultaneous processing of almost all human information transmission means is possible, is generically known as the multimedia system.
Here, a brief description of the technology supporting the multimedia service and multimedia processing, which have been attracting attention recently, is made.

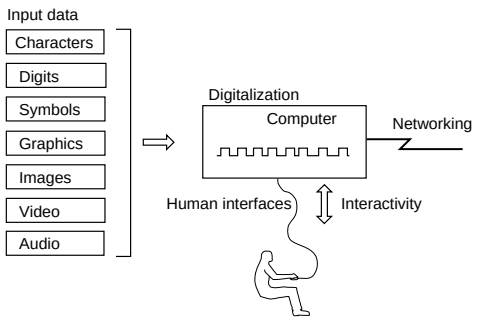
4.1 What is multimedia?

Use of the word "multimedia" began in 1993 as a result of the announcement of the "Information superhighway plan" by the U.S. government. This plan aimed at covering the entire U.S. territory with optical fiber networks in order to implement a bi-directional and high-level communication infrastructure to enable mutual understanding.
Multimedia is a medium, a method which has the following four factors:

- Digitalization
  Through the digitalization of audio, images and other information besides characters and numeric values, high-quality and easy-to-process information can be integrated and used.
- Networking
  Through the interconnection of computers using communication lines such as optical fiber, large amounts of information can be exchanged accurately at high speed.
- Interactivity
  As with the telephone, bi-directional interactive processing with a high response level can be performed.
- Inclusion of human interfaces
  Diverse types of information can be naturally and easily handled.

Overall, the multimedia system is a processing system that is based on multimedia technology, performs the digitalization of characters, digits, symbols, graphics, images, video and audio, exchanges information in real time using communication lines, etc. and can be easily operated by anybody (Figure 4-1-1).

Figure 4-1-1
Multimedia system outline



4.1.1 Multimedia service

The service provided to the users, based on the multimedia system, whose use in diverse fields is expanding along with the progress of computer and network technology, is generically called multimedia service. Here, application examples of multimedia service in the most representative fields, which are listed below, will be explained.

- Business field
- Medical care field
- Publication field
- Education field
- Game field

(1) Business field

Today, since PDAs (Personal Digital Assistants) and notebook personal computers are equipped with communication functions, mobile computing, which enables information exchange with the computer network of one's company or with the Internet through public telephones or cellular telephones, has become popular.
Systems that handle characters and digits are simply information processing systems, but, in multimedia systems, conferences can be performed while watching the face of the person(s) one is speaking to, and animated images and other information can be handled.

(2) Medical care field

Medical systems in which diagnoses are efficiently made through collective management of patients' personal information and medical records, radiographs, etc. as well as in-home medical systems in which patients for whom it is difficult to go to the hospital or patients who live in remote places can be diagnosed while watching the computer display screen, have been put to practical use.
In the system that offers support for remote medical diagnosis, general hospitals of large scale and clinics which do not have the necessary medical facilities are connected through communication lines, enabling patients of small scale clinics located in remote places to receive medical treatment of the same level as that at general hospitals.

(3) Publication field

Nowadays, large amounts of information contained in dictionaries, encyclopedias, illustrated reference books, etc., have been recorded on commercialized CD-ROM. Conventional encyclopedias and illustrated reference books contained only information based on printed characters and pictures. However, in encyclopedias and illustrated reference books for multimedia use, besides the conventional character-based information, images of flower petals unfolding can be displayed and the calling of birds can be heard over the speakers.

(4) Education field

In the education field, multimedia has begun to be used to present research results, exchange opinions, etc., providing image information on display devices as well as audio information conveyed through microphones and speakers. Through this trial, mutual understanding between students of schools located in depopulated areas and students of inner city schools can be promoted without regard to distance.
Education using computers in this way is generically known as CAI (Computer Aided Instruction).

(5) Game field

In the game field, virtual reality is widely used. Virtual reality is a world that imitates the real world on the computer display, created through the comprehensive use of three-dimensional graphics and three-dimensional sounds.
Virtual reality is not limited to the game field; it is also used in flight simulators at airline companies, etc., for pilot training.

(6) Interface technology

GUI (Graphical User Interface) is used as a multimedia system interface. In GUI, the use of graphics called icons, which can be understood at first sight, is basic.

Figure 4-1-2
Icon example

(7) Software production technology

The application software (application programs) that handles multimedia is called a multimedia title. Here, besides character and numeral text data, multiple audiovisual data with different properties such as still images, animated images, audio, etc. is handled. In order to create multimedia titles, tools that enable easy manipulation of multimedia data become necessary. The tools (software) are called an authoring tool. Nowadays, authoring tools are widely used for the production of multimedia titles.

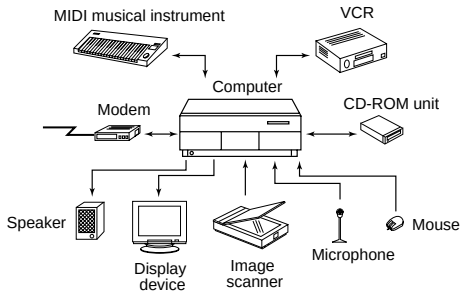
4.1.2 Platforms that implement the multimedia system

In many cases, we use the word "platform" to refer to the platforms of the stations. However, in information technology, it is used to refer to "the environment for the implementation of a given processing function." Here, the hardware configuration and software used as platforms for the implementation of multimedia systems will be explained.

(1) Hardware needed in multimedia systems

The hardware configuration required in order to implement a multimedia system in a personal computer is shown in Figure 4-1-3.

Figure 4-1-3
Example of a multimedia system hardware configuration



① Display device

Unlike conventional information processing systems that only handled characters and digits, in order to represent faithfully and beautifully multimedia data, which is complex and handles large amounts of information, high resolution displays are necessary. In order to support images of high picture quality, a resolution of 1,280 × 1,024 dots is required.

② Image scanner

Handy scanners and high-resolution image scanners are used. These devices input color pictures and other still images and process them as multimedia data.

③ Video equipment

Video cameras and VCRs are used. This viewdata can be recorded and played using QuickTime and other software.

④ Audio equipment

In video conferences, etc., that require interactive conversation, microphones and speakers are indispensable.

⑤ Digital sound equipment

In order to edit, create and play music using the computer, synthesizers and other MIDI musical instruments are necessary to input data.

⑥ Pointing device

As an input device, besides the keyboard, the mouse is widely used as a pointing device. The mouse is an indispensable input device in GUI environments.

⑦ Storage medium

In order to store enormous amounts of data, a storage medium of large capacity is necessary. Mainly hard disks, CD-ROMs, magneto-optical disks, etc., are used.
Currently, the mainstream is to use CD-ROMs as the medium to supply multimedia software. CD-ROMs have a large storage capacity (640 MB), are low-priced and convenient to carry about.

⑧ Modem

The modem is a device that connects telephone lines, dedicated lines and other analog lines with the computer in multimedia processing systems of communications network systems. It modulates digital computer signals into analog signals and performs the reverse, i.e., demodulation too. In order to connect a digital line, a DSU (Digital Service Unit) is necessary.

(2) Operating systems of multimedia systems

Among the operating systems of multimedia systems, the following can be mentioned:

- Apple's Macintosh OS + QuickTime
- Microsoft's Windows XP
- Microsoft's Windows NT as well as UNIX

These operating systems are generically known as multimedia operating systems.
The following characteristics can be mentioned for Windows XP:

- GUI adoption
- Multi-task implementation
- Provision of network functions
- Provision of multimedia functions

① GUI adoption

In former operating systems, in order to operate the computer, commands were input through the keyboard. In GUI, the screen is composed of windows and icons, and operations are instructed to the computer through the manipulation of a pointing device such as a mouse (Figure 4-1-4).

② Multi-task implementation

In multimedia operating systems, multiple application programs can be switched in short intervals to simultaneously perform multiple processing. Switching of application programs is compulsorily performed by the operating system. This operation is called preemptive multi-tasking.

Figure 4-1-4
Example of GUI
(Window screen)

③ Provision of network functions

The rules for communication or data exchange between computers are called protocols. Through the use of communications software, multimedia operating systems are enabled to connect to networks supporting protocols, mentioned below:

- TCP/IP (Internet support)
- IPX/SPX (NetWare support)
- NetBEUI (Windows network support)

④ Provision of multimedia functions

The following multimedia playback software is equipped as standards in multimedia operating systems:

- Video for Windows (animated images playback)
- CD player (music data playback)
- Media player (Diverse media playback)

(3) Creation of multimedia titles

The application software for multimedia systems is called a multimedia title. In order to create multimedia titles, the following are necessary:

- Editing software
- Authoring tools

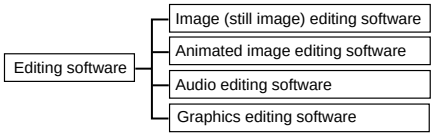
① Editing software

Editing software is software that creates still images, animated images, music and other media (material). Figure 4-1-5 shows the classification of this software.

a. Image (still images) editing software

Image (still images) editing software creates still image data, as well as edit and process data captured using a scanner. A large number of software packages handle this data as bitmap files, which are composed of sets of dots, and files in JPEG format.

Figure 4-1-5
Types of editing software



b. Animated image editing software

Animated image editing software creates and edits videos, animated images, etc. Besides multimedia titles, there is a large number of software packages that enable the creation of videos, animated images,

etc.

c. Audio editing software

Audio editing software is software that manages and controls sequencers, which automatically play synthesizers and other MIDI musical instruments, as well as create, edit and play MIDI data.

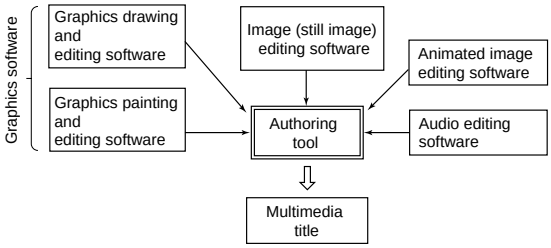
d. Graphics editing software

Graphics editing software is software that creates and edits graphical designs and illustrations. There is painting software that creates images using bitmaps, and drawing software that creates images through the combination of straight and curved lines. Furthermore, there is three-dimensional software that adds depth to input or created two-dimensional still images.

② Authoring tools

Authoring tools are software packages used to compile the media that compose multimedia titles. Music, still images, animated images, etc., which are the fundaments of multimedia titles are all created by the respective editing software.
In order to create multimedia titles using authoring tools, all one needs to do is to look at the screen, think about the design and story and paste the respective multimedia items. For example, animated images are created by combining multiple still images, specifying the movements, and adding audio.

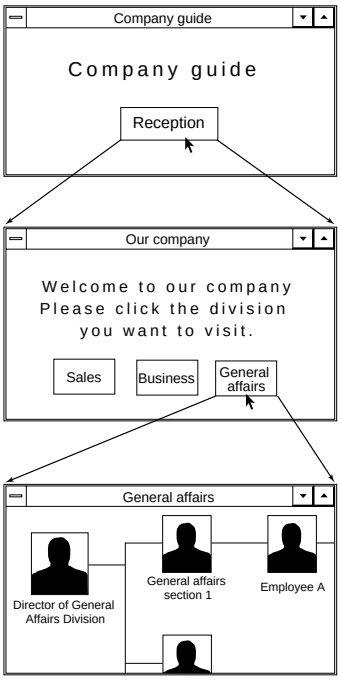
Figure 4-1-6
Role of the authoring tools



③ HyperText

CAI software and application software for presentations created with authoring tools have a structure that enables access to other specified information by clicking the image on the screen with the mouse. The function (concept) that enables free access to information by designating, one after another, the words, signs, images, etc. on the screen is called HyperText (Figure 4-1-7).

Figure 4-1-7
HyperText functions



4.1.3 Multimedia technology

The current expansion of the multimedia system is based on the diversity of technologies accumulated up to now. The technologies that result, indispensable for the implementation of the multimedia system, will be summarized here.

(1) AI

AI (Artificial Intelligence) is the research aimed at giving to computers functions found in humans, such as recognition, judgment, reasoning, problem solving and learning. AI is one of the technologies needed to implement the pattern recognition, etc., used in input operations of the multimedia system.

(2) Pattern recognition

In information processing, pattern recognition is the recognition of characters, images, audio, etc., using the computer. Pattern recognition is performed by extracting special characteristics of the input information (image, etc.) and comparing these special characteristics with a matching pattern. OCRs are an example of input devices that perform character pattern recognition. Pattern recognition will be explained in detail in the following section.

(3) AR/VR/CG

AR is an acronym for Artificial Reality and VR is an acronym for Virtual Reality. While AR creates an "artificial reality," VR creates a "virtual world." They tend to be deemed to have the same meaning, but in the U.S. they are clearly distinguished. Likewise, the technology needed to implement them is CG (Computer Graphics) technology.

(4) Agent

In information processing, the agent is the software that operates inside the computer on behalf of the user. The agent is software that supports user activities, and is capable of judging by itself when executing schedule management, seat reservations, etc. In order to play these roles, in addition to data and the procedures to process data, the agent is composed of a knowledge base to judge the situations.

4.2 Multimedia applications

Following the description of the multimedia system outline of Section 4.1, the multimedia actual implementation will be explained here.

4.2.1 Voice and image pattern recognition

In multimedia systems, besides characters, voice and images are also handled as digital data. The technology used to search for this voice and image data in an existing matching pattern is called pattern recognition. Here, pattern recognition methods for voice and image will be explained.

(1) Voice pattern recognition

The research and development of voice pattern recognition had gone forward before the word "multimedia" was born.
In the current voice recognition system, audio recognition is performed as follows:

1. Phoneme recognition processing
   Special characteristics of the voice input are detected and matched with a phoneme model and the phoneme candidates are obtained from those that match the best.
2. Word recognition processing
   Word candidates are obtained by combining the recognized phonemes and checking, in the dictionary, whether or not they have a meaning as a word.
3. Language processing
   The word candidates are subject to syntactic analysis and semantic analysis and the input voice is settled as data having a meaning.

When these processes are performed, dog words (er-r-r-, uh-uh, huh, er-hum-er-), etc., are excluded and conjecture of the next word based on the context is performed, in order to avoid misconceptions, using AI technology.

(2) Image pattern recognition

In a broad sense, OCRs, etc., that read handwriting are also included in image pattern recognition. The image pattern recognition is performed according to the following procedures:

1. Image input processing
   The image is scanned and entered as data.
2. Image recognition preparation processing
   Elimination of noise, highlighting the part to be recognized, color adjustment, etc., is performed on the input image.
3. Characteristics extraction processing
   The characteristics of the image that is subject to recognition are extracted.
4. Partial recognition processing
   Based on the characteristics extracted from each partial component of the image, the image patterns that match the existing patterns are obtained as candidates.
5. Total recognition processing
   The image patterns obtained in the partial recognition are combined and, in order to decide the image (meaning) they have as a whole, matching with existing models is performed.

4.2.2 Synthesis of voice and image

In the multimedia system, besides the technology performing pattern recognition of the input voices and images, technology to create (synthesize) voices and images is also necessary.
Here, the synthesis techniques of voices and images will be explained.

(1) Voice synthesis technology

Voice synthesis technology research has also been in place for a long time. Today, composite tones can be heard in train guidance information and household electric appliances.
The following are the three main audio synthesis technologies.

① Editing method

The editing method is the method that edits recorded voices and creates specified documents (conversations). The implementation of this method is easy and is widely used, but a sense of disharmony in the edited (connected) parts can be felt and there are times when the intonation becomes strange. However, today, research aiming at the elimination of this lack of naturalness has progressed, and it is possible to synthesize audio that sounds as real as conversations of human beings.

② Analysis method

The analysis method is the method that analyzes and encodes voices, and after storing them as information, synthesizes them while adjusting them to the specified documents (conversations). By encoding the voices in advance, it is possible to synthesize only the required voices necessary at the time. For that reason, since it can be implemented with small capacity storage devices, this method is frequently used in household electric appliances, etc.

③ Rule-based method

The rule-based method is the method that analyzes human voices and establish rules and in order to generate voices. That is, based on the characteristics of the analyzed voices, voices are generated by changing the base tones. However, in practice, there are many parts that sound unnatural, and therefore it is a method requiring further improvement.

(2) Image synthesis method

Image synthesis is a technology included in CG in the broad sense. It can be performed using existing image patterns or by creating new images. In particular, since the method that creates new images enables the creation of a diversity of things without having to stick to reality, this method is used as VR technology.
Generally, image synthesis is performed following the three stages mentioned below:

① Creation of the original picture

There are various methods to create an original picture, such as the preparation of existing image patterns, the creation of new images, etc. As the methods of geometric representation to create new images, the wire frame model, surface model, etc., are typical.

② Shape change

Shape change is to change the original picture to synthesize a new image. At this stage, in order to avoid any disharmony in the image, correction is performed from a three-dimensional viewpoint.

③ Image display

At the image display, which is the last stage, the display processing of the synthesized image is performed. In order to display the synthesized image in three dimensions, the position of the light source, etc. should be considered, and shading, filtering and other adjustments have to be performed.

4.3 Multimedia application system

The multimedia system has permeated diverse fields of the real world. Among recent multimedia application systems, the following can be mentioned:

(1) Internet broadcasting

Among broadcasting that have the Internet as medium, there are large-scale broadcasting provided by television stations, as well as small-scale broadcasting at a personal level, and, as a result, a great variety of programs are presented. Regarding the using modes, there are programs that can be enjoyed for free, while there are others for which the user has to sign an agreement with the broadcasting station of his choice and pay for the service.

(2) Nonlinear image edit system

Previous image editing was linear editing, which was performed based on video tapes. In this method, since the tape was used sequentially, in order to edit a one-hour tape, one hour was needed. Conversely, nonlinear editing is a method in which images are edited as digital data on the computer. Since, in this system, the position to be edited can be accessed directly, editing time can be shortened. Likewise, through the digitalization of video data, there is the advantage that video data can be easily processed. However, attention should be paid, since, as a result of repeated data compression, image quality might deteriorate.

(3) Video-on-demand

Video-on-demand is a service consisting of the distribution of video images at the user’s request. In this method, since service is provided to single users, the load of communications lines becomes too high. For that reason, there is a pseudo video-on-demand system, in which the program is distributed multiple times at specific intervals. In this case, the user is unable to see the video at the time he requested it, but since he only has to wait a specific period of time, it makes little difference. Currently, the system used in CATV, etc., is the latter.
Likewise, since the word "on-demand" means the provision of a service immediately after its request, besides videos, other on-demand services (such as "karaoke-on-demand," etc.) are expected to be available in the future.

(4) Other application systems

Multimedia application systems have spread from disaster monitoring systems, road traffic control systems, and other social systems to TV games, video shopping and other daily life uses. It is probable that in the near future, a use method with results unimaginable today will be born as a result of new technology.

Exercises

Q1 **Which of the following is a correct description of the concept of multimedia?**

a. It is the conversion of analog data into digital data
b. It is the use of the Internet to exchange electronic mails.
c. It is handling diverse data such as audio, animated images, etc., in a unified way.
d. It is watching television programs using personal computers.

Q2 **What is the name of the environment required to actually use multimedia software?**

a. Application b. Agent c. Authoring d. Platform

Q3 **Which of the following corresponds to the computer interface technology that uses icons, etc.?**

a. CAI b. CUI c. GDI d. GUI

Q4 **Which of the following is a correct description of a HyperText?**

a. By designating words and symbols displayed on the screen, information can be accessed one after another.
b. Detailed animated images can be displayed using high definition displays.
c. Not only texts, but also music, videos and all types of information can be represented.
d. A text created with word processing software can be directly converted into an HTML document.

Q5 **Which of the following is the general term for the technology capable of creating a virtual world with intense reality using the computer?**

a. AR b. IR c. OR d. VR

5 System Configurations

Chapter Objectives

The objective of this chapter is to learn about the classification of the information processing systems, which have largely permeated the daily life of the current society. Likewise, the system evaluation method will also be studied.

① Understand system characteristics based on the configuration, processing modes, using modes and other viewpoints.
② Understand meanings and calculation methods regarding the evaluation method of system performance.
③ Understand the meaning of the terminology related to system reliability as well as the reliability calculation methods.

5.1 System classification and configurations

5.1.1 System classification

Basically, "information processing system" refers to the diverse systems composed of hardware, communication equipment and software.

① Classification based on the difference in using modes

- Batch processing system
  - Center batch processing system
  - Remote batch processing system
- Online transaction processing system
- Real-time control processing system

② Classification based on the difference in processing modes

- Centralized data processing system
  - Batch processing system
  - Centralized online transaction processing system
- Distributed processing system
  - Computer network system (LAN, WAN)
  - Client/server system
  - Peer-to-peer system

③ Classification based on the difference in operating modes

- Non-interactive processing system (Batch processing system, etc.)
- Interactive processing system

④ Classification based on the difference in system configurations

- Reliability
  - Simplex system
  - Dual system
  - Duplex system
    - Cold standby
    - Hot standby
- Process efficiency
  - Multiprocessor system
    - Loosely coupled multiprocessor system
    - Tightly coupled multiprocessor system
  - Tandem multiprocessor system

5.1.2 Client/server system

The client/server system is the most representative system among the distributed processing systems. Computers scattered across the network are divided into clients and servers, and their respective roles are distributed as follows:

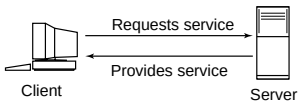
① Client

The client is the computer that receives a service from other computer (server). This term indicates terminal devices, workstations, and other devices.

② Server

The server is the computer that provides a service to other computer (client). This term indicates host computers, workstations and other devices.

Figure 5-1-1
Client/server model



(1) Characteristics of the client/server system

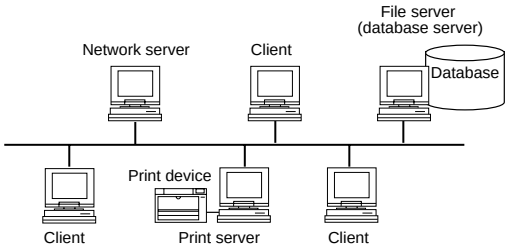
Among the strong points of the client/server system, the following can be mentioned:

- Since, in a great number of cases, interfaces have been standardized, systems can be made open. Therefore, the most suitable hardware or software can be selected without having to stick to a single manufacturer or vendor.
- The expansion of system functions and performance, as well as capacity, etc., is easy.
- System reliability regarding failures can be improved.
- By adopting a client machine with high processing efficiency, application programs using GUI (Graphical User Interface), which has good operability, can be developed.

On the other hand, among the weak points of the client/server system, the following can be mentioned:

- Due to distributed processing, several functions become redundant and system performance declines.
- The system manager is required to have a wide range of knowledge regarding hardware, software and networks of different manufacturers and vendors.
- System development methodologies have not been established.

Figure 5-1-2
Client/server system



(2) Platform functions

A platform is the hardware and software that serve as the base for the operation of a system. Regarding the functions of the platform of the client/server system, the following two can be mentioned:

- Data access function: Function that enables data exchange between computers.
- Program processing function: Function that enables processing request between computers.

These functions are implemented by the protocols and systems mentioned below.

① FTP (File Transfer Protocol)

FTP is the protocol used for the transmission and reference of files between computers connected to a network. Even if the operating systems of the computers differ, file transmission or reference can be performed oblivious to it.

② NFS (Network File System)

NFS is a network file system that enables the free use of files contained in other computers in the network.

③ RPC (Remote Procedure Call)

RPC sets an environment that enables the client's free use of various functions held by the server.

(3) Server types

The following are the main servers that provide service at the client's request:

① Print server

The print server temporarily stores the print data requested by the client and directs the printer to print.

② File server

The file server specializes in file input and output in order to share files and control them in a standard way.

③ Database server

The database server is a server that specializes in database management and is equipped with functions to search large capacity databases at high speed.

④ User interface server

The user interface server sends the commands that allow the user to direct processing to other servers using GUI.

⑤ Communication server

The communication server is the server that uses the network to communicate with another computer. It is equipped with interfaces supporting diverse networks such as LAN, WAN, and ISDN lines.

(4) Client server system application

① Three-tier architecture

Three-tier architecture is the architecture that divides the client/server system into the following three functional modules:

a. Data tier

In the data tier, the database is accessed and the needed data is referenced.

b. Function tier

In the function tier, message or data processing is performed.

c. Presentation tier

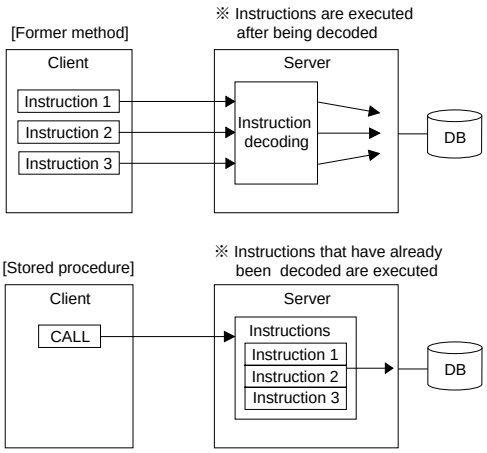
In the presentation tier, the user interface to exchange data with the users is implemented.

In the previous two-tier architecture, only database access was performed on the server side, and all of the remaining functions were performed on the client side. However, since in the three-tier architecture the server side is in charge of the functions of the data tier as well as the function tier, the volume of data transmitted between the client and the server can be reduced, lightening the transmission load. Likewise, since each function can be developed independently, it is also advantageous in the aspect of development efficiency. Currently, n-tier architecture, in which the functions are further subdivided, has also been implemented.

② Stored procedure

The stored procedure is one of the techniques to speed up the client/server system, and is a method that consists of storing in the server the instructions that are frequently used by the client (SQL statements, etc.). Since the client can execute the instructions stored in the server by just calling them, the volume of transmission data and transmission frequency can be reduced. Likewise, by translating beforehand the instructions stored on the server side into an executable format, execution efficiency can be further improved.

Figure 5-1-3
Stored procedure



5.1.3 System configurations

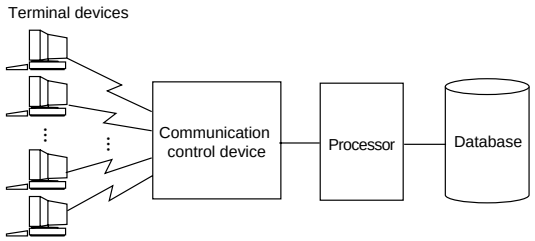
Information processing systems can also be classified according to the configurations of the main devices composing a system. Here, the explanation will be focused on the system configuration of online transaction processing systems.

(1) Systems that emphasize reliability

① Simplex system

The simplex system, which has a configuration that forms the nucleus of online transaction processing systems, operates without spare hardware. The cost to construct this system is low, but if one device breaks down, the whole system goes down. The weak points of the simplex system are overcome by the dual system and the duplex system.

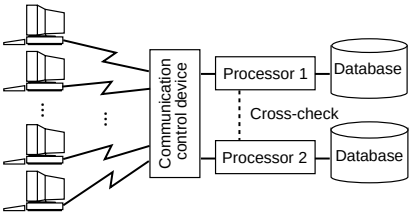
Figure 5-1-4
Simplex system



② Dual system

The dual system is a system in which each device is duplicated to compose a system that performs perfect parallel running of two courses (Figure 5-1-5). In this way, execution is performed while comparing the processing results of the devices of both courses at specific intervals in order to verify whether or not the processing is correct. This operation is called cross-check and is performed in ratios of 1 time every 10 milliseconds or 1 time every 100 milliseconds, etc. In the event of failure of any of the devices, the failed system is separated and processing is continued with the other processing system.

Figure 5-1-5
Dual system

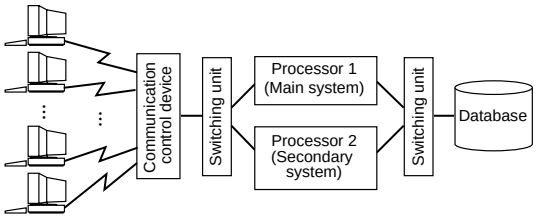


Since the dual system is constantly performing cross-checks of the processing results, it has high reliability and is used in areas in which any failure might endanger human life, such as medical care systems and aircraft control systems. However, due to the performance of cross-checks, the operating cost becomes high.
For example, the flight control system with which the space shuttle is equipped is a system that is expanding even more the idea of the dual system, with a multiple dual configuration of five processors.

③ Duplex system

The duplex system is also called the standby redundancy system. It is a system in which the processor, the main storage unit, the auxiliary storage devices, etc., each have their respective spare machines.
Under normal conditions, while one system (the main system) performs online processing and other high-priority functions, the other system (secondary system) performs batch processing, system development, etc. In the event that the main system fails, the secondary system interrupts processing in order to perform online processing.
Compared to the simplex system, duplex system reliability is much higher, but the equipment investment expenses are almost double. For that reason, few systems have this configuration, and it is used in banking systems or seat reservation systems in which the whole society would be affected if a system failure occurred.

Figure 5-1-6
Duplex system



According to the operating mode of the secondary system under normal circumstances, the duplex system is classified as follows:

- Cold standby mode
- Hot standby mode

a. Cold standby mode

In the cold standby mode, the main system performs online processing while the secondary system remains on standby without being turned on. For that reason, when a failure occurs, it takes time to switch to the secondary system.

b. Hot standby mode

In the hot standby mode, the main system performs online processing while the secondary system remains turned on and on standby, so that it can continue the main system processing at any time. For that reason, compared to the cold standby mode, switching time after a failure occurs is shorter.

The three system configurations mentioned above are compared in Figure 5-1-7.

Figure 5-1-7 Comparison of the three systems which were composed putting emphasis on reliability

	Simplex system	Dual system	Duplex system
Configuration (Spare machines existence)	No (Composed of the minimum equipment required)	Yes (in duplicate)	Yes (in duplicate)
Features	In the event one device breaks down, the whole system is stopped.	Perfectly identical processing is performed in two courses that are cross-checked at specific intervals. When a failure occurs, the system that broke down is separated and processing is continued.	Of the two systems, the main system performs the main functions and the secondary system performs batch processing, etc. When a failure occurs, processing is switched to the system working normally.
Reliability	Low	High	High
Real timeness	Low	Highest	High
Construction cost	Low	High	High
Operating cost	Low	Most expensive	High
Application field	General work	Medical care systems, flight control systems, etc.	Banking systems, seat reservation systems, etc.

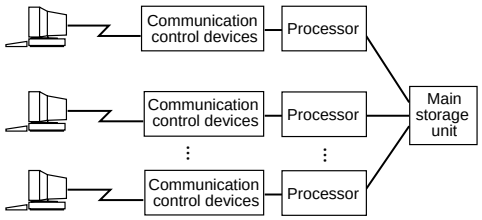
It should be noted that the systems such as the dual and duplex systems in which, due to the duplication of their configuration, processing can be continued as a whole even if one part breaks down, are called fault tolerant systems.

(2) Systems that emphasize processing efficiency

① Multiprocessor system

In the multiprocessor system, multiple processors share one main storage unit and auxiliary storage device, and each of the processors performs parallel processing under one operating system. For that reason, processing efficiency is high.

Figure 5-1-8
Multiprocessor system



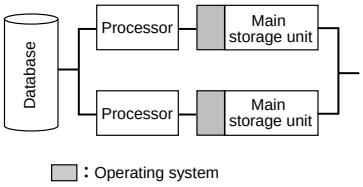
According to the type of processor coupling, multiprocessor systems can be divided into the following two:

- Loosely coupled multiprocessor system
- Tightly coupled multiprocessor system

a. Loosely coupled multiprocessor system (LCMP)

The loosely coupled multiprocessor system is a system in which the processors are loosely coupled so that, in the event of failure, the processor in which the failure occurred can be separated and the operation can be continued. For that reason, the system reliability is high.

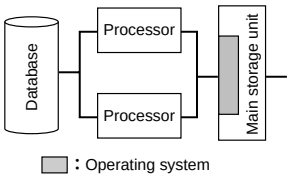
Figure 5-1-9
Loosely coupled multiprocessor system



b. Tightly coupled multiprocessor system (TCMP)

In the tightly coupled multiprocessor system, multiple processors share the main storage unit. For that reason, synchronization and information transmission between processors can be performed at high speed, and complex communication control programs are not required.

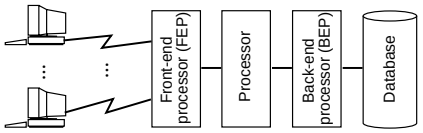
Figure 5-1-10
Tightly coupled multiprocessor system



② Tandem multiprocessor system

The tandem multiprocessor system is a multiprocessor system that connects multiple processors in series (tandem) and distributes the load by assigning functions to each of the processors.
Figure 5-1-11 shows an example of a tandem multiprocessor system in which a front-end processor and a back-end processor are placed in front and behind the main processor.

Figure 5-1-11
Tandem
multiprocessor system



a. Front-end processor (FEP)

The front-end processor mainly performs the communication control of a large number of terminal devices. When a front-end processor is in place, the main processor does not have to perform communication control processing, and, as a result, improvement of processing efficiency is enabled.

b. Back-end processor (BEP)

The back-end processor mainly controls a database in which an immense amount of data is stored. Through use of this database-dedicated machine, database access can be performed at higher speed.

(3) Back-up system

In order to be prepared for earthquakes, fires or other disasters, it is necessary to prepare in advance backup systems for computer systems. There are three methods to implement backup systems:

① Mirror site

A mirror site is a method by which files are simultaneously updated. As a result, a backup system in which switching can be performed in extremely short time is prepared.

② Hot site

In the hot site method, files are not simultaneously updated as in the mirror site. However, identical system environments are prepared, and a backup system in which switching can be performed in a comparatively short period of time is prepared.

③ Cold site

In the cold site method, hardware for backup is provided, but it is necessary to start construction of the system environment and other operations once the backup becomes necessary. Therefore, it takes time to perform switching.

(4) Cluster computing

Cluster computing is a method that, using communication media, connects multiple computers for use as a single computer. This configuration is called the cluster system, or simply cluster.
Clusters are classified into the following two types:

① Dedicated cluster

In the dedicated cluster, multiple computers of the same type (same OS and same architecture) are connected and used as a single computer.

② Distributed cluster

In the distributed cluster, multiple computers of different types are connected and, basically, each of the users uses one of these computers. However, in this configuration method, computer resources that are not being used by the formal users can be used by other users.

5.2 System modes

Here, the characteristics, differences, operating systems, etc., of the most representative systems in terms of the system processing mode, using mode, operating mode and system configuration will be explained.

5.2.1 System processing mode

According to the processing method, the information processing system can be divided as follows:

- Centralized processing system: By connecting multiple terminals to one host computer, processing is centralized in one location.
- Distributed processing system: By connecting multiple computers with communication lines, a network is constructed, and processing is distributed and performed by each of them.

Figure 5-2-1 Comparison between the centralized processing system and the distributed processing system

	Centralized processing system	Distributed processing system
Host computer load	Extremely heavy	Relatively light
Development and maintenance cost	High	Relatively low
Resource use	Limited	All of the resources can be effectively used
Data update	Real-time update is possible	Real-time update is not possible
Reliability	Low. If the host computer breaks down, the whole system is stopped.	High. If the host computer breaks down, other computers can cover it.
Flexibility	Low. When the amount of data increases, the hardware is switched.	High. Even if the amount of data increases, substitution by other computers can be performed.
Security	High	Low

(1) Centralized processing system

The operational aspect of the centralized processing system, which performs processing by concentrating data and information in one location, is extremely efficient. However, in batch processing systems and centralized online transaction processing systems the following problems exist:

- When the data subject to processing increases, switching to a computer with a processing capacity that fits that increase is required.
- When a failure occurs in the host computer, which is placed at the center of the system, all of the connected terminal devices are affected.
- Since most of the functions are concentrated in the host computer, when the software scale grows, the development cost, not to mention the maintenance cost, becomes enormous.

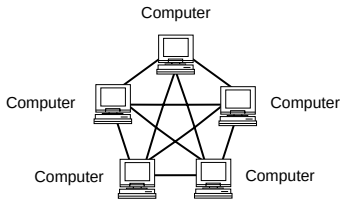
(2) Distributed processing system

The achievement of high performance in personal computers, etc., as well as the progress of network technology enabled the construction of computer networks in which multiple computers are connected using communication lines. As a result, the distributed processing system was conceived. In this system, data is distributed into each of the computers and the user can perform processing using all of the system resources through the network.
Here, the most representative computer network systems will be explained.

① Computer network system

The distributed processing system is a system that is implemented by the computer network system.
The computer network system, as is shown in Figure 5-2-2, is a system in which multiple independent computers are connected using communication lines.

Figure 5-2-2
Fully coupled computer network system



② Characteristics of the computer network system

The characteristics of the computer network system are also the characteristics of the distributed processing system.

- The processing functions and resources of the whole network can be efficiently used.
- Since even if a computer breaks down, processing can be continued in another computer, and the reliability of the system as a whole is high.
- When the work load of one computer is heavy, processing can be transferred to another computer with a light work load, providing flexibility to the system.

Currently, as a convention to implement computer networks, the OSI (Open Systems Interconnection) basic reference model established by ISO has been standardized.

③ Computer network system configuration

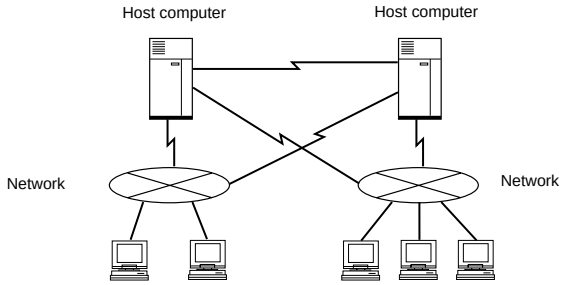
According to the connection method of each computer, computer network systems are roughly divided as follows:

- Vertically distributed configuration
- Horizontally distributed configuration

a. Vertically distributed configuration

Vertically distributed configuration is a configuration method that was widely used in the computer network systems constructed through the 1980s. In this configuration, data transfer among multiple host computers can be performed at high speed from the terminal devices through the switching network and LAN. Likewise, by providing intelligent functions to the terminal devices themselves, one processing can be divided between the host computer and the terminal devices, however the processing core is always in the host computer (Figure 5-2-3).

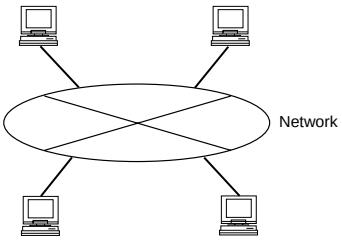
Figure 5-2-3
Vertically distributed configuration



b. Horizontally distributed configuration

In the 1990s, as a result of the appearance of high-performance workstations and personal computers, dependency between host computers and terminal devices connected to the network became extinct. In other words, the core of the computer network in the horizontally distributed configuration is the network, and the mainframes, workstations and personal computers are all connected to the network as host computers. Therefore, in the horizontally distributed configuration system, the user can select the host computer that suits him/her best.

Figure 5-2-4
Horizontally distributed configuration



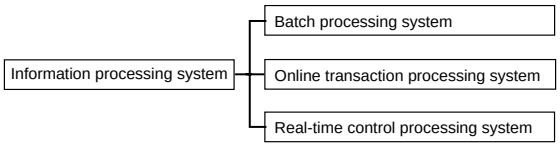
5.2.2 System usage mode

When diverse data and information is processed in a computer, the processing mode depends on the timing in which the data is to be processed, or the processor in which the data is to be processed.
According to the processing mode, the information processing system can be roughly classified as follows (Figure 5-2-5):

- Batch processing system
- Online transaction processing system
- Real-time control processing system

Here, the characteristics, use examples, OS, etc., of these information processing systems will be explained.

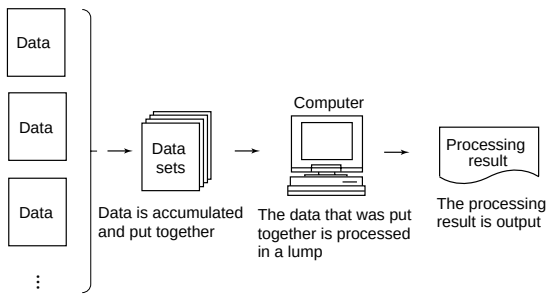
Figure 5-2-5
Using mode of the information processing system



(1) Batch processing system

In order to execute a job, it is necessary to process multiple data and information. If a central computer exists when one job is processed, the system is called a centralized processing system. The most representative of the centralized processing systems is the batch processing system.
For example, when the aggregation processing of a national census is performed, if the data of the all prefectures is not gathered in one location to be processed, correct aggregation results regarding population trends, etc., of the whole country could not be obtained. Likewise, when calculating the sales of one month of a convenience store, if all the sales data of the month subject to calculation is not collected, the sales of that month cannot be calculated.
In this way, the processing mode in which processing is performed after all of the data needed is gathered and lumped together is called batch processing. Batch processing is the oldest processing method, which has been in placed since computers were created.

Figure 5-2-6
Batch processing system



① Batch processing application example

Batch processing is most suitable to perform the following jobs:

- Payroll calculation, sales account and other processing that must be performed by daily or monthly
- Marking and aggregation of examinations such as for the University Testing Center Examination
- All sorts of statistical analysis processing

② Batch processing characteristics

When processing a job that substantially exceeds the processing capacity of the computer system, among the diverse processing modes, batch processing is the most efficient. This is because there is no human intervention during processing. However, as once processing starts, no human intervention is allowed up to processing completion, the establishment of processing order must be performed beforehand.
Likewise, since programs and data sharing can be performed, and the standardization of processing procedures is easy, the adoption of this processing expanded, especially in mainframe computers. At the time that "Grosch's law," which said that computers are expensive and computing power increases as the

square of the cost, held good, it was the best method. However, now that hardware prices have fallen and performance has increased, it cannot be said that centralized processing is the best processing method.

③ Batch processing modes

In batch processing, data is processed in a lump. Depending on whether processing is performed offline or online, it can be divided into the following two types:

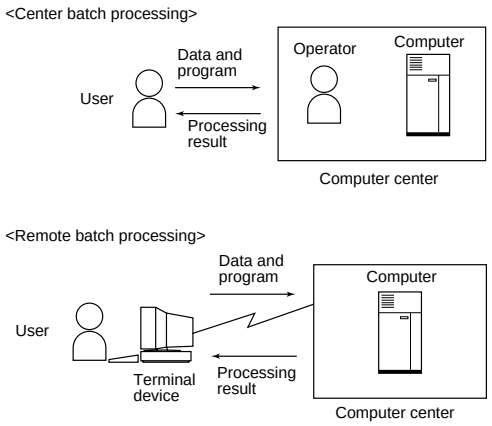
a. Center batch processing

In center batch processing, the data stored beforehand by the user, and the program of the processing procedure, are executed offline. In other words, data is transported by a human, input by an operator, etc., and processing is performed in a computer center.

b. Remote batch processing

In remote batch processing, data is sent from remote locations through communication lines (online), and once all of the data is gathered, processing is performed in a computer center. This processing is also called remote job entry (RJE).

Figure 5-2-7
Center batch processing and remote batch processing



Center batch processing can be further classified as follows:

- Open batch processing
  In open batch processing the user does everything, from data storage to computer manipulation.
- Closed batch processing
  In closed batch processing, the user hands over the processing procedure and the data to the operator and asks him/her to perform the computer processing.
- Cafeteria system
  In the cafeteria system, the user registers the processing procedure and the data in the computer and leaves the remaining operation to the operator.

④ Batch processing operating system

Considering that batch processing is the oldest processing system, it can be said that operating systems have been expanded in order to improve batch processing efficiency. Here, the functions of the operating system aiming at efficient batch processing performance will be briefly described.

a. Job control language (JCL)

Job control language was designed in order to implement automatic job processing. Processing is executed through the definition of the following:

- Job name
- Storage location of the program to be used
- Storage location of the data subject to processing
- Area of the work file and the output file

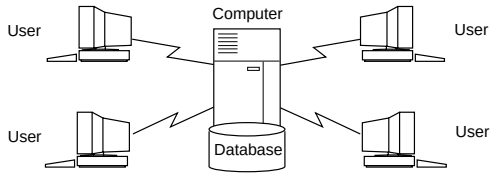
b. SPOOL (Simultaneous Peripheral Operations On Line)

Operating speed differs depending on the configuration of each of the devices composing the computer system. For example, the processor performs processing at an electronic speed (using nanosecond as a unit), while the printer only operates at a mechanical speed (using "second" as a unit). Considering speed differences of this kind, SPOOL, which is one of the techniques to effectively operate the system, is applied.
SPOOL is a function that enhances processing speed by separating low-speed input and output devices from high-speed processors. The data subject to processing, as well as the processing results, are stored at high speed in an auxiliary storage device, which is the only device with which the processor exchanges data.

(2) Online transaction processing system

The opposite of batch processing is real-time processing. Online processing system is the generic name for systems in which terminal devices at remote locations and computers are connected through communication lines. A great number of these systems are online transaction processing (OLTP) systems, in which the data generated as a result of a transaction is processed in real time.

Figure 5-2-8 Online transaction processing system



For example, in a bank's computer system, we perform a transaction to withdraw money from a cash dispenser and, based on that transaction data, the computer installed in the bank computer center performs the money withdrawal processing. Likewise, in a train seat reservation system, when the train to be boarded and the number of seats required are specified, it is instantaneously determined whether or not there are vacant seats in the requested train, and if there are vacant seats, booking can be performed.
In this way, there are many online transaction processing systems that support corporate activities as well as the foundations of daily life. It can be said that this type of system has a great impact on society.

① Characteristics of online transaction processing

In the online transaction processing system, the data and information subject to processing is normally managed as a centrally controlled database. For example, this is the case of the depositors' database of bank online transaction processing systems and the train seats database of train seat reservation systems. As conditions to control these databases, ACID attributes, Atomicity, Consistency, Isolation, and Durability, are required.

② Job contents of the online transaction processing system

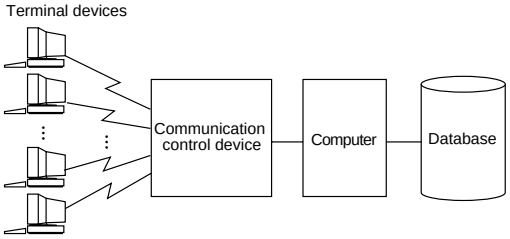
The main job contents of the online transaction processing system, adopted as a social and corporate infrastructure, are listed below.

- Production line control system and production management system in the manufacturing industry
- Seat reservation and ticketing system in the transportation business
- Deposit and money order systems as well as investment and loan systems in the finance sector
- Insurance system in the insurance business
- Stock exchange system in the securities sector
- Sales inventory management system and customer information control system in wholesale and retail businesses
- The public taxation system, social insurance system, car inspection registration system, postal savings and money order system, meteorological information system.

③ Configuration of the online transaction processing system

The system shown in Figure 5-2-9 is a centralized online transaction processing system. However, in order to emphasize system reliability, hardware duplication, etc., is necessary.

Figure 5-2-9 Centralized online transaction processing system



④ Conditions for online transaction processing

In order to implement the online transaction processing system, simultaneous execution control (exclusive control), which enables simultaneous response to the requests of multiple users, is an indispensable condition.
For that reason, the programs performing the online transaction processing must be reentrant programs. Reentrant programs are programs that can be executed again before their former execution is completed. The condition to run these programs is that the program area and the data area be separated. Based on this characteristic, simultaneous processing of multiple requests can be performed, and, moreover, correct results can be returned for each of these processing requests.
Likewise, since the resources are simultaneously shared by multiple users, it is necessary to perform simultaneous execution control (exclusive control) of the resources.

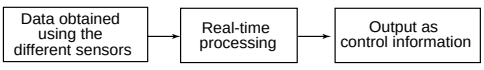
⑤ Failure recovery

In online transaction processing, hardware (processor, disk, printer, etc.) breakdowns, as well as application failures, must be detected and coped with promptly. For that reason, in the online transaction processing system, failure detection and failure recovery functions are robust, and when failures are detected, in many cases processing is abended (abnormally ended). Likewise, as methods to recover database physical and logical failures, the rollback and the roll forward methods are adopted.

(3) Real-time control processing system

In new jet airliners, in order to support the pilot, computers have been introduced and computer-controlled automatic piloting from takeoff to landing has been enabled.
In order to perform automatic piloting, it is necessary to provide operation instructions to perform the most suitable flight control based on the data related to air speed, outside temperature, wind direction, wind power, engine thrust, etc., obtained using the different sensors. Therefore, if it takes time to provide the operation instructions, correct flight control in ever-changing conditions will not be possible. The performance of immediate (real-time) calculation processing of the information obtained and the output of the results as control information is called real-time control. The system adopting the real-time approach is generically known as the real-time processing system.

Figure 5-2-10 Flow of the real-time control processing system



① Characteristics of the real-time control processing system

Since the real-time control processing system is mostly used as part of one device, it seldom uses input devices such as a keyboard or output devices such as a printer. In most cases, the input devices consist of diverse sensors, and the output devices of actuators and other control devices.
Likewise, the processors, mainly miniaturized microprocessors, are often composed of main storage units that store programs and data.

② Application examples of the real-time control processing system

Besides the flight control system adopted by jet airliners, real-time control processing systems are applied in such diverse fields as the ones mentioned below.

- Air-traffic control system
- Power supply monitoring system
- Industrial robot control
- Motor fuel control system and braking system
- Household electric appliances such as rice cookers, washing machines and air conditioners

③ Operating system implementing real-time control processing

In real-time control processing systems whose nucleus is a microprocessor, there are application programs and operating systems to control the programs. These operating systems are called real-time operating systems or real-time monitors. Real-time monitors have the following functions:

- Multi-task processing function
- Task switching function
- Function to minimize the load of the monitor itself

④ Interfaces needed for real-time processing

In the real-time control processing system, the data obtained using the different sensors is processed, and the electrical signals of the processing results are converted into mechanical operations by the actuators. The main interfaces used in the real-time control devices are listed below:

- RS-232C (Recommended Standard-232C)
- USB (Universal Serial Bus)
- Centronics interface
- SCSI (Small Computer Systems Interface)
- GPIB (General Purpose Interface Bus)

5.2.3 System operating mode

According to the operations (relations) generated when humans use computers to perform processing, the information processing systems can be further classified as follows:

- Non-interactive processing systems (batch processing systems, etc.):
  In non-interactive processing systems-- since processing is performed after the procedure is indicated, once it has started, humans cannot intervene in the processing.
- Interactive processing systems:
  In interactive processing systems, humans can provide indications or perform changes while interacting with the computer.

Here, only interactive processing systems will be explained.

(1) Interactive processing systems

Since at the time work processing is instructed to the computer, it looks as though computer and human are "talking" while performing the operation jointly, this system is called the interactive processing system.
TSS, in which one uses the computer as though one were the only user, and online transaction processing cases in which the next processing content is determined according to the processing results of the work requested to the computer by the terminal devices, among others, can be considered interactive processing. Needless to say, the operations performed in game software, word processing and spread sheet software, etc., can also be considered interactive processing.

① Characteristics of interactive processing

In batch processing the processing procedure is determined beforehand, but in interactive processing, since the processing content can be changed during processing, the indication of the procedure before processing starts can be vague.

② Functions of the software implementing the interactive processing

In interactive processing it is necessary to have a robust user interface. Therefore, the following are used:

- Window system
- GUI

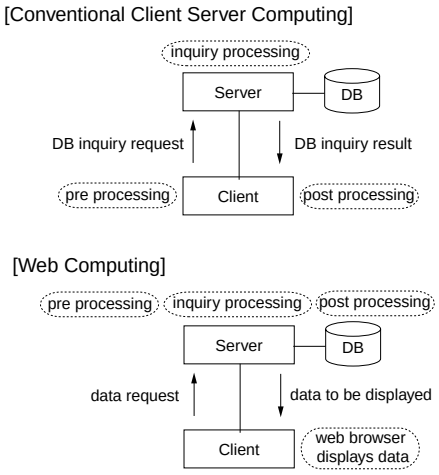
In practice, without using the keyboard, the mouse is used to select the processing from the following.

- Title bar
- Menu bar
- Pull-down menu
- Pop-up menu

5.2.4 Web computing

In former client/server systems, data processing and other processing was performed on the client side. Compared to this, Web computing does not provide processing functions to the client side, and all processing is performed on the server side (Figure 5-2-11). As a result, only the browser function that displays the information sent by the server side becomes necessary for the client side.
This mode has the drawback that the server load is increased, but the performance requirements of the client side can be low. Therefore, this mode is suitable for cases where multiple client terminals are required.

Figure 5-2-11 Web Computing



5.3 System performance

5.3.1 Performance calculation

The following measures are used for computer system performance evaluation

- Processing (TAT)
- Processing efficiency

(1) Processing time

Processing time means the time taken to execute data processing by the computer system. The following standards are used for the processing time.

- Turn Around Time (TAT)
- Response time

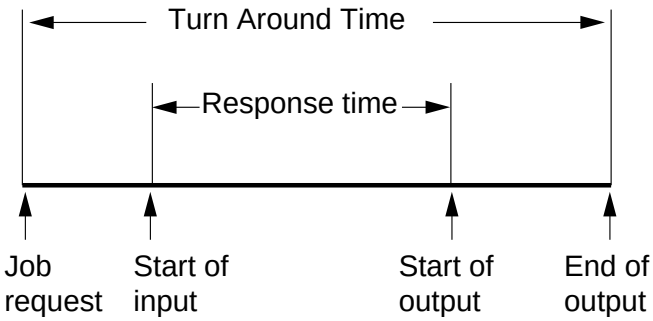
①Turn Around Time (TAT)

This is the time taken to return the results when a batch processing job is submitted to the computer in batch processing systems.
In systems with high processing capacity, the turn around time is reduced by allocating a high priority to special jobs and the computer will schedule appropriately the resources to be used.

②Response time

The time to get a produce a response from the computer from the time the transaction is input is known as the response time in online transaction processing systems.
The main aim during the development of online systems is to shorten the response time, thereby, improving the efficiency of the computer processing.

Figure 5-3-1 Turn Around Time and Response Time



(2) Processing efficiency

The processing efficiency measures the ability of the computer system to process the data. The following measures for used processing efficiency.

- Throughput
- Command mix
- MIPS
- FLOPS

①Throughput

This is the volume of work that can be processed by the computer system in a given time.

- Batch processing : The number of jobs that can be processed within a given time
- Online transaction processing: The number of transactions that can be processed within a given time

②Instruction mix

The combination of the execution time for representative instructions and frequency of such instruction occurrences found in programs represent the performance of the computer's processor.

Figure 5-3-2 Instruction mix

Type of instruction	Instruction Execution time	Frequency of occurrence
Transmission instructions	200 nanoseconds	65%
Calculation instructions	400 nanoseconds	20%
Decision instructions	300 nanoseconds	10%
Jump instruction	100 nanoseconds	5%
Total		100%

Representative programs with the individual instruction can be divided into two kinds.

- Commercial mix: These are frequently used in business processing and uses mainly transmission instructions
- Gibson mix: These are frequently used in scientific calculations and uses mainly the calculation instructions

Normally, the instruction set and the instruction mix differs for each manufacturer. It is not possible to assess the efficiency of the whole system based only on the processor's performance.

③MIPS (Million Instructions Per Second)

MIPS represents on average the number of instructions that a processor can execute in units of millions per second.

1 MIPS of processor performance

1,000,000 instructions / second

↓

10^6 instructions / 1,000 millisecond

↓

10^6 instructions / 1,000,000 microsecond

↓

10^6 instructions / 10^6 microsecond

↓

1 instructions / microsecond

It takes 1 microsecond to execute one instruction.

If we calculate the MIPS of the processor in figure 5-3-2

Instruction mix = 200 nanos x 0.65 + 400 nanos x 0.2 + 300 nanos x 0.1 +
100 nanos x 0.5
= 245 nanos

1 MIPS = 1 instruction / microsecond = 1 instruction / 1,000 nanos

MIPS rate = 1,000 nanos ÷ average instruction execution time
= 1,000 nanos ÷ 245 nanos
= 4.08

Currently, high performance processors can execute ten billion instructions in one second and they are measured in GIPS (Giga Instructions Per Second)

④FLOPS (Floating Point Operations Per Second)

MIPS are used as a representative measure of business processing performance evaluation. The number of floating point calculations possible in one
second is used as the measure called FLOPS for scientific calculations.
Example
If we represent in FLOPS the performance of a processor that performs 1,000,000,000 floating point calculations in 1 second.
1,000,000, 000 times / second

↓

1×10^9 times / second

↓

1 Giga FLOPS

5.3.2 Performance design

One of the main objectives of computer system design is to ensure that the processed data volume and contents meet the requirements in terms of performance. We need to understand the relationship between the requirements, form of processing by the system and any conflicts with the performance. For example, the response time is important in an online transaction system and represents the time taken when a user submits a job to the output of the result. In the case of batch processing, the need to decide which job to execute means the spool function is important.

5.3.3 Performance evaluation

The representative methods for measuring the computer system performance is

- Test program : Benchmark
  Kernel program
- System monitor

(1) Test program

A test program is a program that is used to simulate the actual business programs under heavy usage. The details of the business are modeled as processing within the program and the processing efficiency measured.

①Benchmark

In benchmark tests, actual working programs are executed to measure system's processing efficiency.

a. TPC

The TPC benchmark has become a de facto standard for OLTP performance evaluation. The American Transaction Processing Performance Council (TPC) is responsible for drawing up the contents of this benchmark. The following four types of TPC benchmarks

A to D was fixed by this body,

- TPC-A : This is used for the banking operations. It is based on the input/output model of the ATM
- TPC-B : This is modeled for a database system in a batch processing environment.
- TPC-C: This is based upon the order entry model
- TPC-D : This is based upon the decision support applications

b. SPEC

This is targeted towards OS that supported distributed processing especially computers that run UNIX

systems. These sets of benchmarks are fixed by the Standard Performance Evaluation Cooperation: SPEC. There are 2 kinds of such benchmarks

- SPEC-int : This is focused on integer type calculations
- SPC-fp: This is focused on floating point type calculations

②Kernel program

A simple program that does integer calculations and is repeatedly executed is known as a kernel program. This is sued to evaluate the processor's performance.

(2) System monitor

It is a combination of a diagnostic program with hardware and is used on the computer system to monitor the operations conditions. System monitors can take the form of programs known as software monitors or may also come with diagnostic hardware (hardware monitors).

5.4 Reliability of the system

5.4.1 Reliability calculation

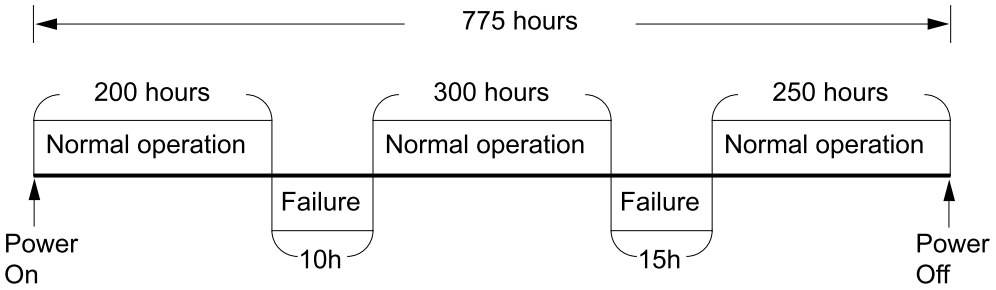
One of the measures used to indicate the safety and efficiency is known as RASIS.
In real terms, it includes MTBF (Average time between failures), MTTR (Average repair time) that covers the reliability, serviceability and the availability
RASIS stands for

- Reliability
- Availability
- Serviceability
- Integrity
- Security

(1) MTBF (Mean Time Between Failures)

MTBF represents the degree of reliability, i.e. the R in the RASIS. It represents the average time between the occurrence of one failure to the next.
MTBF = Total normal operation time ÷ total number of times normally executed

Figure 5-4-1 Example of the Operation condition of the equipment



For example, the MTBF for the equipment shown in figure 5-4-1 Taking the average normal operation time
MTBF = (200 h + 300 h + 250 h) ÷ 3 = 250 hours
The reciprocal gives the failure rate of the equipment.
Failure rate = 1 / MTBF

(2) MTTR (Mean Time To Repair)

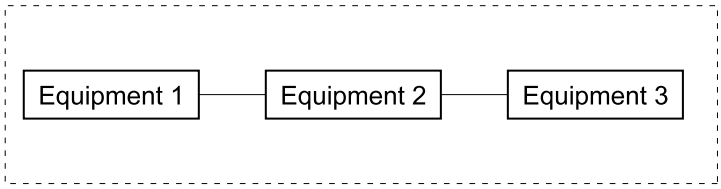
The degree of S (Serviceability) in RSAS is indicated by the average time in which the equipment is not acting normally.
The repair time and the and the downtime before repair. MTTR gives the average repair time.
MMTR is given by
MTTR = repair time ÷ no of failures
For example, the MTTR for the equipment shown in figure 5-4-1 is
MTTR = (10 h + 15 h) ÷12 = 12.5 hours
The (A Availability) in the word RASIS is given by
Availability = MTBF÷ (MTBF + MTTR)
For example, the availability for the equipment shown in figure 5-4-1 is

Availability = 250 h ÷ (250 h + 12.5 h)
The nearer to 1, the better is the equipment's availability.

①Availability of systems connected serially

The whole system will stop if anyone of the equipment in a serially connected system fails. This is like the simplex system.
The availability of serial systems are
Serial system's availability = availability of equipment 1 x availability of equipment 2 , ... availability of A machine.

Figure 5-4-2 Serial system

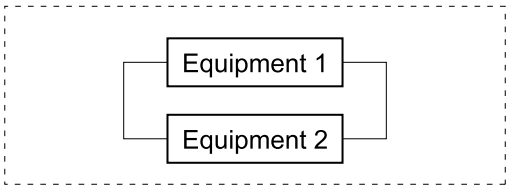


If 0.9 is the availability of each set of equipment for the 3 sets, shown in Figure 5-4-2. The availability of the entire system is given by
Availability of the entire system = 0.9 x 0.9 x 0.9 = 0.729

②Availability of serially systems connected in parallel

The duplex system and dual system is composed of two processors If one processor were to fail in a multi-computer parallel system, the remaining set can still operate normally and the system still operates normally. The only situation where the entire system is stopped is when both the processors were to breakdown simultaneously. The availability of a parallel system is given by
Availability of a parallel system
= 1 - ((1 - availability of equipment 1) x (1 - availability of equipment 1))
This first value of 1 in the formula represents a situation where they is no failure

Figure 5-4-3 Parallel system

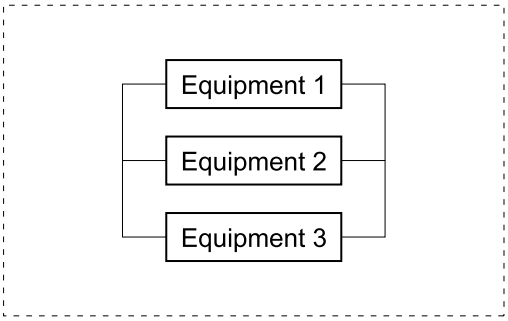


If 0.9 is the availability of each set of equipment for the 2 sets of equipment, shown in Figure 5-4-3. The availability of the entire system is given by
Availability of the entire system = 1 - ((1 - 0.9) x (1 - 0.9)) = 0.99

③Availability of 2 out of 3 systems

In a 2 out of 3 system, the system can function normally using only two of the three processors. This means the third processor is added for redundancy.

Figure 5-4-4 2 out of 3 system



However, the system will fail if either all the processors failed or two of the three processors were to fail.
The availability in such a parallel system is given by
Availability of parallel system = 1 - (1 + 2 x availability of equipment) x (1 - availability of equipment)^2
Figure 5-4-5 shows the availability of the total system and the individual equipment is shown in a table

Figure 5-4-5 2 out of 3 system availability

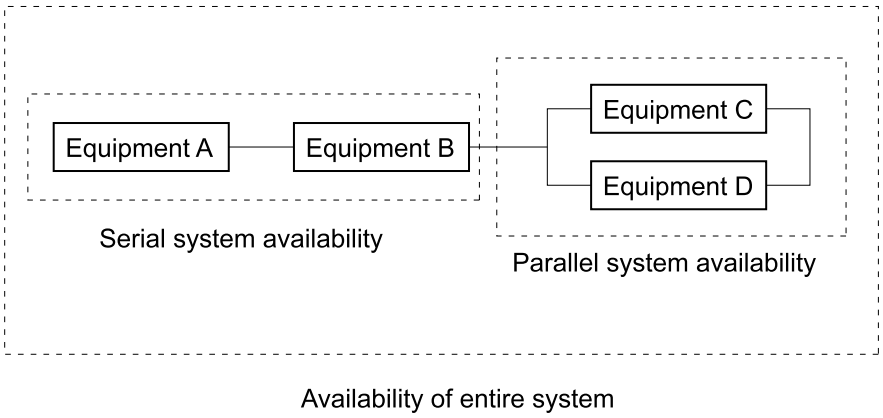
	Equipment 1	Equipment 2	Equipment 3	Total system
Case 1	Normal	Normal	Normal	Normal
Case 2	Normal	Normal	Failure	Normal
Case 3	Normal	Failure	Normal	Normal
Case 4	Failure	Normal	Normal	Normal
Case 5	Normal	Failure	Failure	Failure
Case 6	Failure	Normal	Failure	Failure
Case 7	Failure	Failure	Normal	Failure
Case 8	Failure	Failure	Failure	Failure

The system will operated for case 1 to 4. All the equipment has an availability of 0.9 and a failure of 0.1.
If these figures are entered, the resulting calculation is shown.
Availability in case 1 = 0.9 x 0.9 x 0.9 = 0.729
Availability in case 2 = 0.9 x 0.9 x 0.1 = 0.081
Availability in case 3 = 0.9 x 0.9 x 0.1 = 0.081
Availability in case 4 = 0.9 x 0.9 x 0.1 = 0.081
If you add these values from case 1 to case 4 to give the availability of the total system:
Availability of a 2 out of 3 system = 0.729 + 0.081 + 0.081 + 0.081 = 0.972
In addition, the availability of a 1 out of 3 system can be calculated easily.
In this condition, a minimum of 1 processor running is sufficient to keep the entire system running.
Availability in case 1 = 0.9 x 0.9 x 0.9 = 0.729
Availability in case 2 = 0.9 x 0.9 x 0.1 = 0.081
Availability in case 3 = 0.9 x 0.9 x 0.1 = 0.081
Availability in case 1 = 0.9 x 0.9 x 0.1 = 0.081
Availability in case 1 = 0.9 x 0.9 x 0.9 = 0.729
Availability in case 2 = 0.9 x 0.9 x 0.1 = 0.081
Availability in case 3 = 0.9 x 0.9 x 0.1 = 0.081
Availability in case 4 = 0.9 x 0.9 x 0.1 = 0.081
Availability in case 5 = 0.9 x 0.1 x 0.1 = 0.009
Availability in case 6 = 0.9 x 0.1 x 0.1 = 0.009
Availability in case 7 = 0.9 x 0.1 x 0.1 = 0.009
Availability of the total system is found by adding the results of case 1 to case 7.
Availability of a 1 out of 3 system = 0.729 + 0.081 + 0.081 + 0.081 + 0.09 + 0.09 + 0.09
= 0.999
Conversely, if we consider the above calculation in a reverse manner:

Case 8 failure = 0.1 x 0.1 x 0.1 = 0.001
We can extract the availability by taking 1 to represent no failure and subtracting the above value from it.
1 out of 3 system availability = 1 – 0.001 = 0.999

④Availability of combined systems

Figure 5-4-6 Complex system



The following are taken into considerations when designing systems with high reliability.

5.4.2 Reliability design

The following are considered in reliability design.

- Fail safe
- Fail soft
- Fool proof

①Fail safe

Failsafe is the taking into consideration of the safety aspects in order to minimize the effect on the other parts when failure occurs.
For example, the traffic would automatically switch to red if the traffic control light system were to go down. This would help to prevent accidents that may result from the system's failure.

②Fail soft

For example, if a power failure were to occur in a hospital, the minimum amount of lights world automatically be available and priority given to life support or life saving equipment when the generators are run.

③Foolproof

Foolproof means to prevent the effect of mis-operation by the human element. For example, input checking is done and re-entry is made to the mis-entered data.

5.4.3 Reliability objectives and evaluation

(1) RASIS

RASIS is the acronym containing the five words representing reliability.
Reliability
This is measured as the MTBF (Mean Time Between Failures). This can be considered as the normal

operation time of the system.

Availability

This represents the possible usage ratio of the computer system. This is computed as

A = MTBF ÷ (MTBF + MTTR)

Serviceability

This represents the ease of maintenance of the computer system. This is computed as MTTR (Mean Time To Repair). This can be considered as the down time for the system.

Integrity

This represents the ability to prevent the data from being corrupted

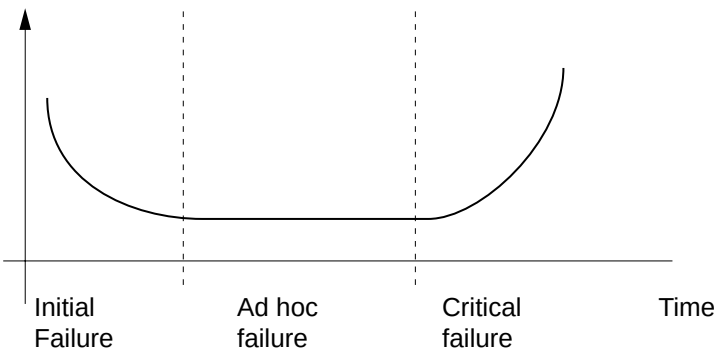
Security

This represents the ability to ensure the security of the data

(2) Bath tub curve

The bath tub curve is obtained when the failure rate is plotted against time

Figure 5-4-7 Bath tub curve



①Initial failure period (Burn-in)

This represents the initial failure rate when the system is initially installed. It shows a decreasing failure rate for stable operations.

②Ad hoc failure period (useful life)

This is caused by chance events. The failure is fairly constant. This represents the system's steady period.

③Critical failure (Wear out)

This is caused by fatigue or aging of the system. When the system enters the period, it will experience an increasing failure rate This is the time to change the system.

(3) Uninterrupted operation

There are usage situations which that do not allow the system to stop operation. This means continuous operation is required. Examples of such systems include life support systems in the hospitals or banking systems. Uninterrupted operation also becomes one of the important

To be concrete, uninterrupted operations are implemented by using UPS (Uninterruptible Power Supply) or by multiplexing systems.

5.4.4 Financial costs

It is necessary to consider the performance of the system from an economic point of view. It is necessary to consider the financial performance

- Development cost: The cost incurred by the development of the system
- Running costs: The cost incurred to run the system

It is not possible compare the direct costs as the costs for large scale systems will be high. Instead, the cost performance is used. The use of the above 2 rules to evaluate early systems is known as Grosch's law. However, this rule does not apply with the recent advances of technology.

An additional criteria of profitability is added. The cost incurred by the system can lead the generation of profit. A better evaluation of the system efficiency can be obtained by including this factor.

Exercises

Q1 **Which of the following corresponds to the function that is most suitable for processing on the server side in the client/server system?**

a. Output data display processing
b. Database update processing
c. Format checking of input data
d. Pull-down menu display processing

Q2 **Consider that computers A and B are connected using LAN, and the printer is connected only to computer A. When computer B is to print data, it sends the data to the computer A. Regarding this system, which of the following is the most suitable description?**

a. The same operating system must run in computers A and B.
b. The MIPS value of computer A must be higher than that of computer B
c. Printing can be performed even if computer A is not turned on.
d. The role of computer A is to be the print server for the client server model.
e. Until printing is completed, computer B cannot perform any other processing.

Q3 **Which of the following descriptions related to the computer system corresponds to the duplex system?**

a. Multiple processors share the main storage unit and are controlled by a single operating system. Even if one processor breaks down, processing can be continued with the rest of processors.
b. In order to improve processing capacity by distributing the processing load, multiple processors are connected in series.
c. Under normal conditions, one of the processors is on standby, and when the processor in operation breaks down, after switching to the processor on standby, processing is continued.
d. Multiple processors connected in parallel simultaneously perform the same processing and compare mutual results. In the event that a failure occurs, the processor that broke down is separated and processing is continued.

Q4 **Among the three jobs listed below, which is the most suitable combination of processing modes?**

[Jobs]
1. One-month salary calculation
2. Industrial robot automatic operation
3. Airplane seat reservation

[Processing mode]
A. Online transaction processing
B. Batch processing
C. Real-time processing

	1	2	3
a	A	B	C
b	A	C	B
c	B	C	A
d	C	A	B

Q5 **Which of the following is the most suitable description of the centralized processing system, when compared to the distributed processing system?**

a. In the event of disaster or failure, since the center can perform centralized recovery operations, the danger of having the system stopped for a long time can be avoided.

b. Since batch management is conducted in the system, it is easy to comply with requests for the addition, modification, etc., of system functions, and probabilities of the occurrence of backlog stacking are low.
c. By taking centralized measures in the center, the security and data consistency can easily be maintained and controlled.
d. The operation and management of the hardware and software resources become complicated, but expansion supporting new technology is easy.

Q6 **Which of the following is the appropriate term to represent the processing mode in which computer users exchange information with the computer by selecting the icons displayed on the screen, and entering commands using the keyboard, adding human judgement to the information processing?**

a. Online transaction processing
b. Time sharing processing
c. Interactive processing
d. Batch processing

Q7 **Which of the following is the term that represents the time elapsed between when a series of works is requested to the computer and the processing results are received, in the batch processing mode?**

a. Overhead
b. Throughput
c. Turnaround time
d. Response time

Q8 **Which of the following is the description of system performance evaluation?**

a. In OLTP (Online Transaction Processing), the MIPS value is used in system performance evaluation.
b. Response time and turnaround time are performance indexes from the point of view of the system operations manager.
c. Generally speaking, as the activity ratio of the system resources increases, the response time also improves.
d. The number of transactions and jobs that can be processed within the time unit is important for system performance evaluation.

Q9 **Which of the following is a correct explanation of the Gibson mix used for system performance evaluation?**

a. It is the average operating ratio based on the values of the failure occurrence record of a specific time period obtained through the online diagnostic program.
b. It is the estimated average processing capacity per time unit at the online transaction processing.
c. It is the average weighting value of the instruction execution time for scientific computation.
d. It is the record average execution time of multiple standard programs for business calculation.
e. It is the record processing capacity obtained by monitoring with measuring devices the internal signals generated when a monitoring program is executed.

Q10 **In a processor whose basic operating time (clock time) is 0.05 microseconds, when the values of the clock number required to execute an instruction and the instruction frequency rate are the ones shown in the table, approximately what is the MIPS average value of the processor performance?**

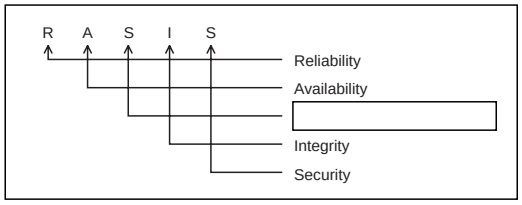
Instruction type	Clock number required for instruction execution	Use frequency
Operation between registers	4	40 %
Operation between memory and register	8	50 %
Unconditional branch	10	10 %

a. 3
b. 10
c. 33
d. 60
e. 132

Q11 Among the following descriptions related to computer performance evaluation criteria, which is the description related to SPEC-int?

a. It is the number of times floating point operations can be executed in one second. It is mainly used to measure scientific computation performance, but it is also used as the performance index of massive parallel processing computers.
b. It is the average number of times an instruction is executed in one second. It is not suitable for performance comparison between computers with different architectures.
c. It is an integer arithmetic benchmark whose main targets are computers in which UNIX can run. It was developed by the System Performance Evaluation Association and has expanded as a standard benchmark.
d. It is an online transaction processing system benchmark. According to the target models, four types of benchmark specifications, A, B, C, and D have been developed.

Q12 In the term "RASIS," which is related to system reliability, integrity and security, what does the third character, "S," stand for?

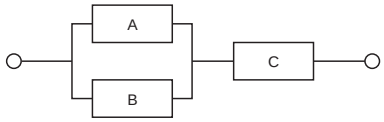


a. Safety
b. Selectivity
c. Sensitivity
d. Serviceability
e. Simplicity

Q13 Which of the following descriptions related to computer system reliability is correct?

a. System remote maintenance improves the operating ratio, by improving the MTBF.
b. The system operating ratio is improved by extending the MTTR and MTBF.
c. The more complicated the system configuration is, the longer the MTBF becomes.
d. System preventive maintenance is performed in order to extend the MTBF.

Q14 **When three computers, A, B and C are connected as displayed in the diagram, what is the operating ratio of the system as a whole? Here, the operating ratio is considered to be 0.8 for A, B, and C. Likewise, regarding the parallel connection part constructed by computers A and B, even when one computer, either A and B, is operating, the said parallel connection part is considered to be operating.**

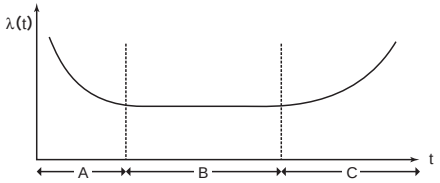


a. 0.512 b. 0.768 c. 0.928 d. 0.992

Q15 **Which of the following is the fail-safe measure taken when industrial robots are controlled with microcomputers?**

a. The circuits are designed to automatically stop when an abnormal operation signal is detected.
b. By making the circuits of each function easy to exchange, the failure recovery time is reduced to the utmost.
c. Using two hard disks, the same data is stored in each of the disks.
d. A manufacturer maintenance hot line is set up to give immediate assistance in case of emergency conditions.

Q16 **When the relation between the failure ratio, λ (t), of the equipment composing the system, and the time elapsed since the equipment started to be used, t, is represented in a graph, generally, the following bath-tub curve is drawn. Generally speaking, in which of the ranges do the failures generated by design/manufacture defects and inappropriate environments occur frequently?**



a. A b. A and C c. B d. C

Part 2

INFORMATION PROCESSING AND SECURITY

Introduction

This series of textbooks has been developed based on the Information Technology Engineers Skill Standards made public in July 2000. The following four volumes cover the whole contents of fundamental knowledge and skills required for development, operation and maintenance of information systems:

No. 1: Introduction to Computer Systems
No. 2: System Development and Operations
No. 3: Internal Design and Programming--Practical and Core Bodies of Knowledge--
No. 4: Network and Database Technologies
No. 5: Current IT Topics

This part gives easy explanations systematically so that those who are learning information processing and security for the first time can easily acquire knowledge in these fields. This part consists of the following chapters:

Part 2: Information Processing and Security
Chapter 1: Accounting
Chapter 2: Application Fields of Computer Systems
Chapter 3: Security
Chapter 4: Operations Research

1 Accounting

Chapter Objectives

Obtaining basic accounting knowledge essential for the understanding of business activities.

① Understanding the flow of accounting information in a business enterprise.
② Understanding the steps of preparing a balance sheet and an income statement and the method of settling accounts in a business enterprise.
③ Learning how to read the balance sheet and the income statement, and understanding the differences between financial accounting and management accounting.

1.1 Business Activities and Accounting Information

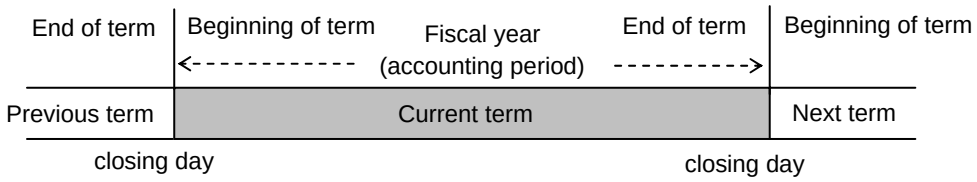
1.1.1 Fiscal Year and Accounting Information

A business processes its accounts on a fiscal year (accounting period) basis. A business summarizes the results of its activities during a fiscal year in statements of accounts (financial statements).
Statements of accounts show the business's operating performance during a fiscal year and its financial conditions. The main statements of accounts are a balance sheet and an income (profit and loss) statement, both prepared at the closing of the fiscal year.

(1) Fiscal Year

A business conducts activities on a continuous basis, and it is necessary to set a specific cycle of time for accounting purposes. This cycle is called a fiscal year (accounting period). A fiscal year is usually a 12-month period. The fiscal year of many Japanese companies runs from April 1 through March 31 of the next year.
The beginning of a fiscal year is called the beginning of a term, and the end of a fiscal year the end of a term. The end of a term is also called a closing day.

Figure 1-1-1 Fiscal Year and Closing Day

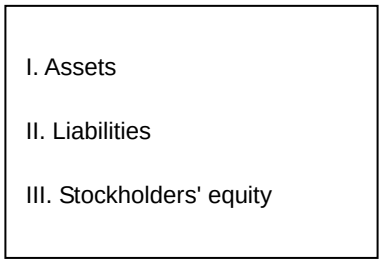


(2) Balance Sheet (B/S)

The balance sheet shows the business's financial position at a specific point in time, usually at the end of a term. It consists of assets, liabilities, and stockholders' equity. It shows assets on the left (debit) and liabilities and stockholders' equity on the right (credit).
In addition to this traditional format called the ledger account (T-account), there is a report form, which presents assets above and liabilities and stockholders' equity below.

Balance Sheet in Ledger Account

Balance Sheet in Report Form



① Assets

Assets include the cash, deposits with banks, buildings, furniture, machinery, and other goods of value held by an enterprise for business activities and the rights to receive cash from others in the future, such as receivable and loans.

② Liabilities

Liabilities are an enterprise's obligations to make payments in the future. Also called "borrowed capital," liabilities reduce assets.

③ Stockholders' equity

Stockholders' equity is the net assets remaining after subtracting an enterprise's total liabilities from its total assets. As opposed to borrowed capital, stockholders' equity is also called "owner's equity."
The following equation expressing stockholders' equity is called a "capital equation":

Assets - Liabilities = Stockholders' Equity

The balance sheet shows the conditions of the enterprise's assets, liabilities, and stockholders' equity at a specific point in time. Transposing liabilities to the right side of the capital equation gives the following equation:

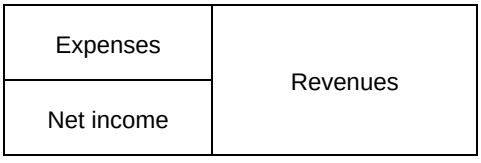
Assets = Liabilities + Stockholders' Equity

This equation is called a "balance sheet equation." This means that on the balance sheet, the total amount of assets on the left always equals (that is, balances with) the total amount of liabilities and stockholders' equity on the right. Hence the name "balance sheet."
As an enterprise conducts business, its assets, liabilities, and stockholders' equity change from their levels at the beginning of a term. When stockholders' equity at the end of a term exceeds stockholders' equity at the beginning of the term, the difference is called a "net income" (net profit for the term). In the opposite case, the difference is called a "loss" (net loss for the term).

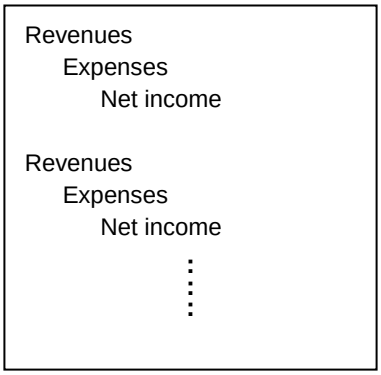
(3) Income Statement (Profit and Loss Statement; P/L)

The income statement consists of revenues and expenses. It presents the business's operating performance during a specific period (usually a fiscal year). It shows expenses on the left (debit) and revenues on the right (credit). If the total revenues exceed that of expenses, the difference is a profit. In the opposite case, the difference is a loss. A profit is recorded on the left (debit), and a loss on the right (credit).
Like the balance sheet, the income statement is also prepared in the account form or the report form. In the case of the income statement, the report form is more common.
For example, when companies publish their operating and financial results in newspapers, the balance sheets usually take the ledger account (T-account), and the income statements the report form.

Income Statement in Ledger Account



Income Statement in Report Form



① Revenues

Revenues are increases in stockholders' equity produced by an enterprise's business activities. Revenues include sales of products, commissions received, and rents received.

② Expenses

Expenses are decreases in stockholders' equity produced by an enterprise's business activities. Expenses are expenditures of the enterprise. Expenses include employee salaries, commissions paid, and advertising expenses.

③ Net income (net loss)

The difference between the total amount of revenues and that of expenses is a net income (net loss). This relationship is represented by the following equation:

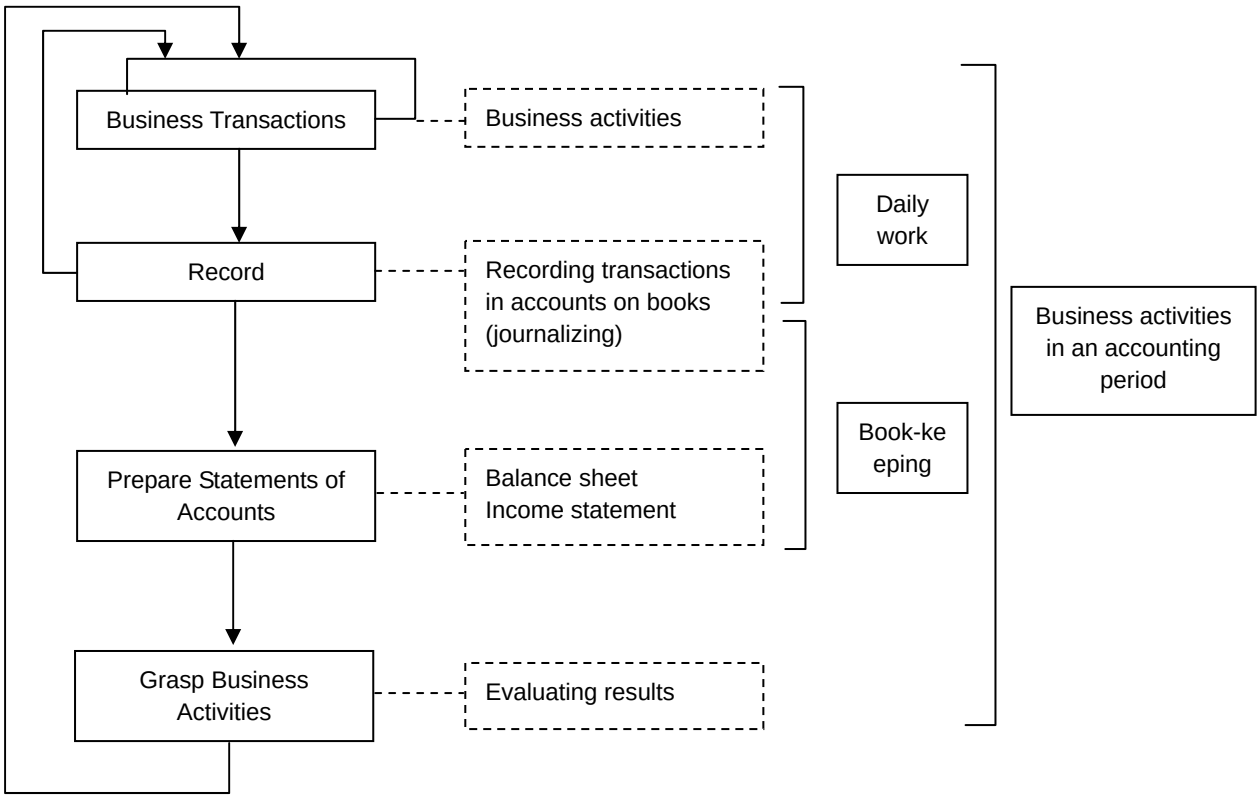
Expenses + Net income = Revenues

This equation is called an "income statement equation." The net income agrees with the balance remaining after subtracting the stockholders' equity at the end of a fiscal year from the stockholders' equity at the beginning of the term shown on the balance sheet.

(4) Flow of Transaction Information

In order to grasp the conduct of business activities, it is vital to understand the process of preparing statements of accounts based on transaction information (slips) and to correctly read the statements of accounts presenting the results of business activities.
The flow of transaction information is shown in Figure 1-1-2.

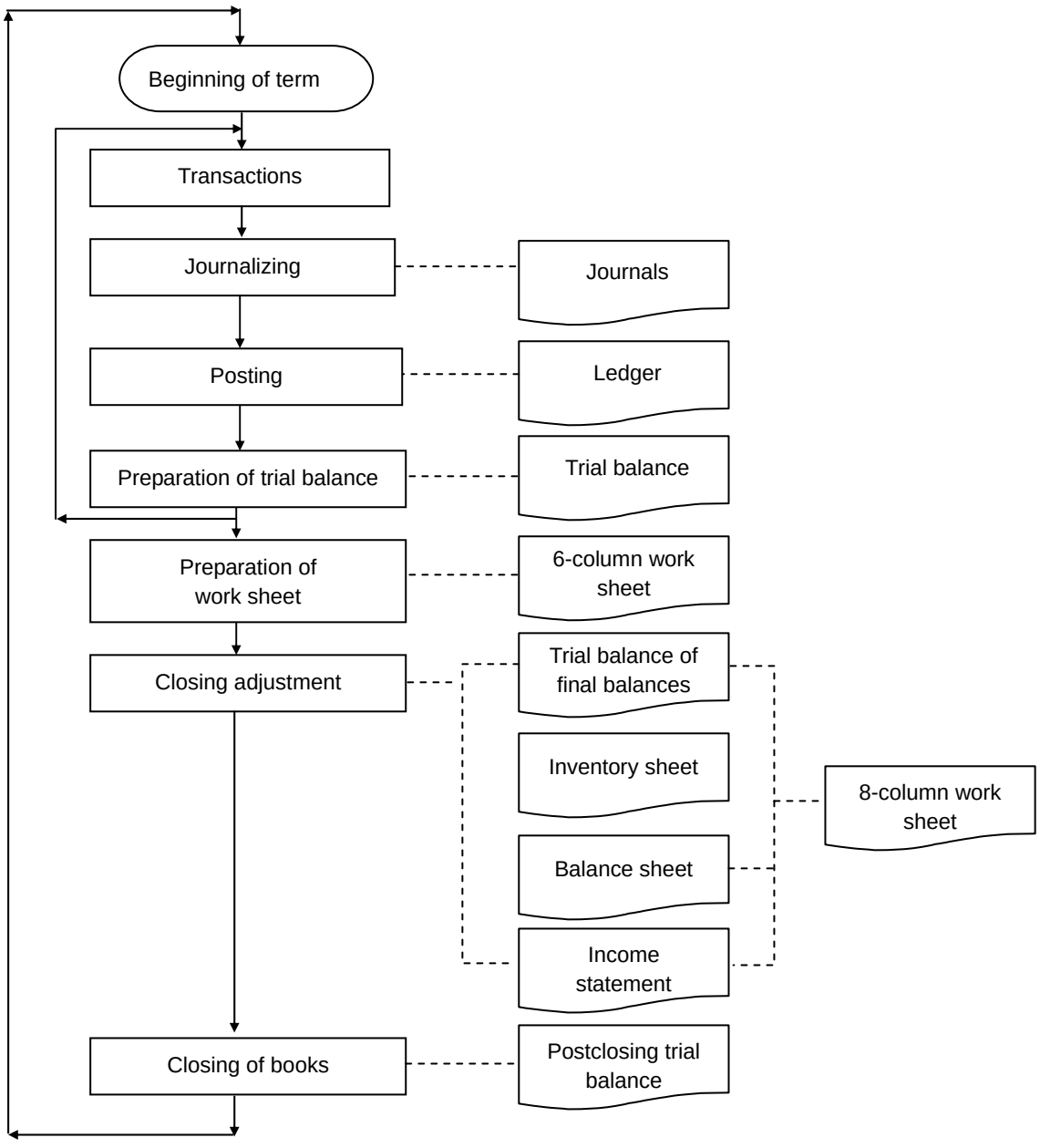
Figure 1-1-2 Flow of Transaction Information



1.1.2 The Accounting Structure

For the preparation of statements of accounts, all transactions arising from business activities are processed according to rules. These rules are called "(double-entry) bookkeeping." The accounting procedure under these rules is shown in Figure 1-1-3.

Figure 1-1-3 From Transactions to Closing of Books



(Source: "Class II Common Curriculums" edited by Central Academy of Information Technology, Japan Information Processing Development Corporation)

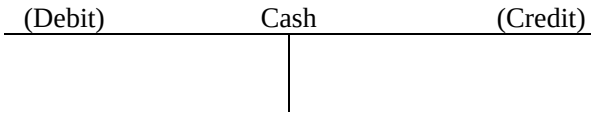
(1) Transactions

In bookkeeping, a transaction means an event that causes an increase or decrease in assets, liabilities, or stockholders' equity. The occurrence of revenue or an expense is also a transaction in bookkeeping, since it causes an increase or decrease in stockholders' equity.

(2) Journalizing

In bookkeeping, transactions as they are conducted are recorded, classified into detailed categories, and calculated to determine what increases or decreases they brought to assets, liabilities, and stockholders' equity and what revenues or expenses they brought about. The bookkeeping categories set for such recording and calculations are called "accounts," and their names are called "titles of account." Account columns are the columns set aside in a journal for the recording and calculation of increases and decreases in individual titles of account. Account columns are provided on the left-hand side (debit) and the right-hand side (credit) of ledger pages.

Example of an account in ledger account (T-account)



Each transaction is decomposed into debit and credit elements. Based on the results, it is determined:

- what amount should be entered on the debit side of which account and
- what amount should be entered on the credit side of which account.

This procedure is called "journalizing."
The results of journalizing are entered in the journal in chronological order of transactions. In recent years, it is also very common to use slips instead of a journal.

How to Make Entries

Here are the rules for recording in accounts the increases and decreases in assets, liabilities, and stockholders' equity or amounts of revenues and expenses arising from transactions:

① Enter an increase in assets on the debit side and a decrease in assets on the credit side.
② Enter an increase in liabilities or stockholders' equity on the credit side and a decrease on the debit side.
③ Enter revenue on the credit side as it accrues.
④ Enter an expense on the debit side as it accrues.

	Debit (left-hand side)	Credit (right-hand side)
① Assets	Increase (+)	Decrease (-)
② Liabilities	Decrease (-)	Increase (+)
Stockholders' equity	Decrease (-)	Increase (+)
③ Revenues	Decrease (-)	Accrual (+)
④ Expenses	Accrual (+)	Decrease (-)

(3) Posting

Posting means the transfer of records from the journal to account columns provided in a ledger (general ledger).

(4) Preparation of Trial Balances

There are three types of trial balances: the trial balance of totals, the trial balance of balances, and the trial balance of totals and balances. The main objective of preparing trial balances is to check whether posting from the journal to the ledger has been performed correctly.

① Trial balance of totals

A trial balance of totals is prepared by calculating the debit total and the credit total for each title of account on the ledger. An example is shown below.

(Thousands of yen)

Debit	Title of Account	Credit
2,430	Cash	1,400
170	Accounts receivable	50
865	Merchandise	650
75	Furniture	
300	Buildings	
	Accounts payable	120
100	Loans payable	500
	Stockholders' equity	1,000
	Sales revenue	350
20	Wages payable	
110	Advertising expense	
4,070		4,070

② Trial balance of balances

The trial balance of balances is prepared by calculating the balances (differences) of the accounts on the trial balance of totals. Here is the trial balance of balances prepared based on the trial balance of totals shown in ①:

(Thousands of yen)

Debit	Title of Account	Credit
1,030	Cash	
120	Accounts receivable	
215	Merchandise	
75	Furniture	
300	Buildings	
	Accounts payable	120
	Loans payable	400
	Stockholders' equity	1,000
	Sales revenue	350
20	Wages payable	
110	Advertising expense	
1,870		1,870

③ Trial balance of totals and balances

The trial balance of totals and balances is prepared by combining the trial balance of totals and the trial balance of balances into one. Here is the trial balance of totals and balances prepared by combining the trial balance of totals shown in ① and the trial balance of balances shown in ②:

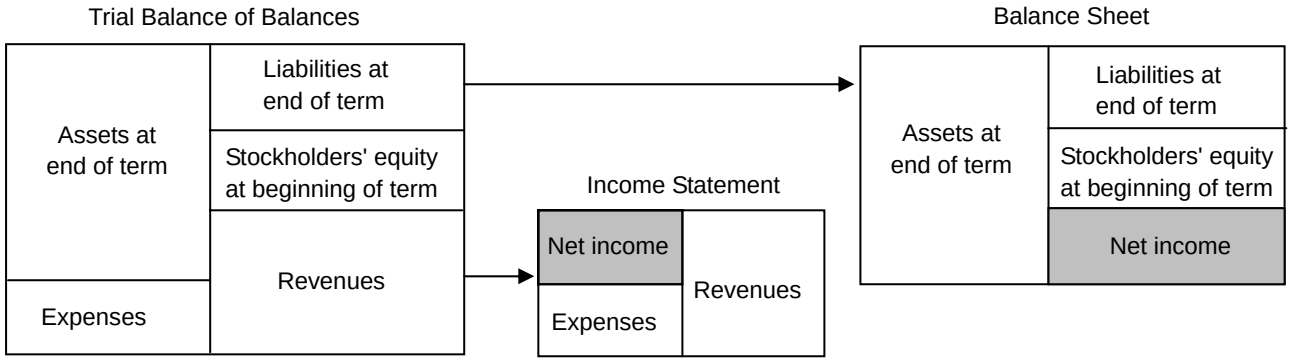
(Thousands of yen)

Debit		Title of Account	Credit	
Balance	Total		Total	Balance
1,030	2,430	Cash	1,400	
120	170	Accounts receivable	50	
215	865	Merchandise	650	
75	75	Furniture		
300	300	Buildings		
		Accounts payable	120	120
	100	Loans payable	500	400
		Stockholders' equity	1,000	1,000
		Sales revenue	350	350
20	20	Wages payable		
110	110	Advertising expense		
1,870	4,070		4,070	1,870

(5) Preparation of the Six-Column Work Sheet

A statement containing a trial balance of balances, an income statement, and a balance sheet is called a "six-column work sheet." This statement is helpful in understanding the general flow of the closing of the books. "Six columns" refers to the total number of the debit and credit columns on the trial balance of balances, the income statement, and the balance sheet. The six-column work sheet provides basic material for the preparation of an income statement and a balance sheet.

Title of Account	Trial Balance of Balances		Income Statement		Balance Sheet	
	Debit	Credit	Debit	Credit	Debit	Credit
Assets						
Liabilities						
Stockholders' equity						
Revenues						
Expenses						
Net income (Net loss)						



(6) Closing Adjustment

In bookkeeping, daily transactions are entered in the journal and these records are posted to the ledger in order to record and calculate increases and decreases in individual accounts. This work is performed on a daily basis. When a fiscal year ends, it is necessary to clarify the operating results during the period and the financial position at the end of the period. The closing of the books means the acts of closing the books at the end of a fiscal year, summarizing the records, and preparing a balance sheet and an income statement.

Closing adjustment means the acts of amending the records on the books at their closing so that the individual accounts can show correct actual balances or correct amounts of revenues and expenses. Adjustment required for this purpose is called "closing adjustment (closing adjustment entries)."

Closing adjustment includes such procedures as income account and capital account transfers, calculations of profits and losses from merchandise transactions, estimation of bad debts, and depreciation and amortization expense.

Items that require such closing adjustment are called "closing adjustment items." A sheet listing those items is called an "inventory sheet." Meanwhile, an extended trial balance prepared by adding columns for closing adjustment (closing adjustment columns) to a six-column work sheet is called an "eight-column work sheet."

Title of Account	Trial Balance of Balances		Adjustment Entries		Income Statement		Balance Sheet	
	Debit	Credit	Debit	Credit	Debit	Credit	Debit	Credit
Assets								
Liabilities								
Stockholders' equity								
Revenues								
Expenses								
Net income (Net loss)								

(7) Closing of Books

After the closing of accounts, it is necessary to check whether the amounts to be carried forward in individual accounts have been calculated and entered correctly. For this purpose, the amounts to be carried forward to the next term are collected to prepare an after-closing trial balance (postclosing trial balance).
The debit and credit totals on the after-closing trial balance are entered on the first line of the journal as "balance brought forward" with the first date of the next term.

1.2 How to Read Financial Statements

1.2.1 How to Read the Balance Sheet

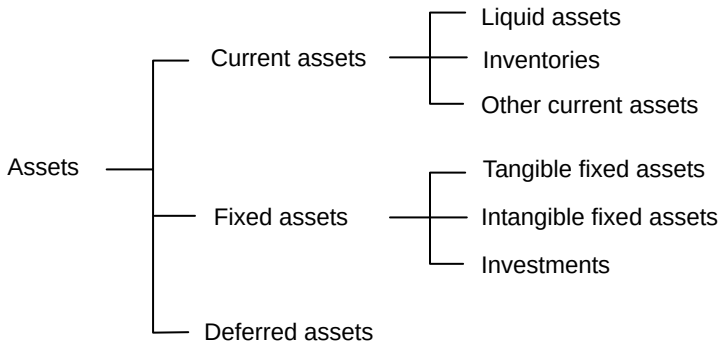
Figure 1-2-1 Example of the Balance Sheet (B/S)

Dated: (Thousands of yen)

Assets		Liabilities and Stockholders' Equity	
Title of Account	Amount	Title of Account	Amount
Current Assets	2,482	Current Liabilities	1,500
Cash and deposits	1,670	Accounts payable	1,200
Accounts receivable	600	Short-term loans	200
Securities	100	Advances received	100
Inventories	82	Fixed Liabilities	1,770
Others	30	Bonds	1,000
		Long-term debt	700
		Allowance for employee retirement and severance benefits	70
Fixed Assets	3,320		
Tangible fixed assets	1,840	Total Liabilities	3,270
Buildings	1,360		
Machinery and equipment	480	Capital stock	1,000
Intangible fixed assets	30	Legal reserves	650
Investments	1,450	Retained earnings	882
Shares in subsidiaries	890	(Of which, net profit or loss)	(340)
Investment securities	560	Total Stockholders' Equity	2,532
Total Assets	5,802	Total Liabilities and Stockholders' Equity	5,802

(1) Assets

According to the one-year rule and the normal operating cycle rule, assets are classified as follows:



- One-year rule

  Assets that are likely to be converted into cash within one year of a closing day are current assets. For example, a loan to be repaid within a year is a short-term loan and a current asset. A loan not to be repaid within one year is a long-term loan and a fixed asset.

- Normal operating cycle rule

When assets arise as part of a business's main activities (purchasing, production, and sales), they are classified as current assets even if they are not to be converted into cash within one year. The normal operating cycle rule applies to such titles of account as accounts receivable, notes receivable, and inventories.

For example, the price of an article sold under an installment contract is to be received in 36 monthly installments, these installments are not classified into current assets and fixed assets depending on whether they are receivable within one year or not; instead, the entire price is recorded as a current asset.

However, this rule does not apply to the uncollected amount for a fixed asset sold, since the sale is not a transaction made in the course of the main business activities. The amount should be recorded as a current asset or a fixed asset according to the one-year rule.

① Current assets

Current assets are assets likely to be converted into cash within one year according to the one-year rule or assets in the process of business activity according to the normal operating cycle rule. Depending on the characters of titles of account, current assets are divided into three categories: liquid assets, inventories, and other current assets.

a. Liquid assets

Liquid assets are cash and other assets that can be converted into cash in a short time. Liquid assets include checking and other deposits (excluding time deposits not maturing within one year), notes receivable, and securities being held temporarily.

Accounts receivable and notes receivable arising from business activity are called trade receivable.

Typical titles of account

Cash:	Legal tender, including banknotes and coins, checks received, stock dividends, and the like.
Accounts receivable:	Claims arising from the sale of products or services that have not been paid for yet.
Notes receivable:	Notes received in the course of ordinary transactions whose face amounts are to be received at promised dates.
Checking account:	A deposit account opened for conducting transactions using checks instead of cash.
Securities:	Shares, bonds, debentures, and the like purchased with the intention of holding them temporarily.

b. Inventories

Inventories include articles for sale, products manufactured for sale, and raw materials for manufacturing products. Since inventories require production and sales activities before they can be converted into cash, they are less readily convertible into cash.

Physical inventory means the act of counting the items of merchandise and product stock in the warehouse.

Typical titles of account

Merchandise:	Goods purchased from outside for resale.
Products:	Goods for sale manufactured or processed internally or externally from raw materials purchased.
Goods in process:	Semi-finished products still in the manufacturing stage.
Raw materials:	Materials to manufacture products.

c. Other current assets

Temporary claims arising from other than business transactions are collectively referred to as other current assets. Other current assets include non-trade accounts receivable, accrued revenue, prepayments, and prepaid expenses.

Typical titles of account

Non-trade accounts receivable:	Claims arising from the sale of goods other than merchandise that have not been paid for yet and are likely to be settled within one year.
Accrued revenue:	Revenue generated in the current fiscal year but not collected by the closing day. Accrued revenue is temporarily presented as an asset and transferred to the original revenue account at the beginning of the next fiscal year. However, it is more common to use more specific titles of account, such as house rent receivable, interest receivable, and land rent receivable.
Prepaid expenses:	That part of a payment corresponding to the next fiscal year onward. Prepaid expenses are temporarily presented as an asset and transferred to the original revenue account at the beginning of the next fiscal year. However, it is more common to use more specific titles of account, such as prepaid insurance premiums, prepaid interest, prepaid house rent, and prepaid land rent.
Prepayments:	Part of the price for an article paid in advance of the delivery of the article.

② Fixed assets

Fixed assets are assets requiring more than one year to be converted into cash according to the one-year rule or assets to be used for a long time for the enterprise's production or sales activities. Depending on the characters of titles of account, fixed assets are divided into three categories: tangible fixed assets, intangible fixed assets, and investments.

a. Tangible fixed assets

Tangible fixed assets are assets that have physical substance, such as land and buildings, and that are to be used for a long time for the enterprise's business activities such as production and sales.
Tangible fixed assets, except land, lose their value as time passes. In bookkeeping, the loss in value is recorded as depreciation and amortization. The value of each tangible fixed asset is reevaluated at the end of each fiscal year, and the loss in value is presented as an expense.

Typical titles of account

Buildings:	Such buildings as business offices, stores, factories, and warehouses.
Machineries:	Such equipment as working machines, machine tools, chemical machines, and conveyors.
Automotive equipment:	Cars, trucks, and other vehicles for business activities.
Land:	Land owned by the enterprise, such as store and office sites.
Furnitures:	Showcases, desks, chairs, and the like used for business.
Construction in process:	When the construction of a building, machinery, equipment, or the like extends over a long period, the payments already made are temporarily presented as assets. Construction in process is transferred to buildings or machinery, as the case may be, upon completion of the construction.

b. Intangible fixed assets

Intangible fixed assets are assets that do not have physical substance and are to be used for a long time for the operation of the enterprise, such as patent rights, trademark rights, and goodwill.
The cost of acquisition of these rights is the amount paid to acquire them. The value of these assets decreases as they are amortized over their life in years prescribed in the tax law or other laws.

Typical titles of account

Patent rights:	Legal rights held by inventors.
Trademark rights:	Legal rights to the registered trademarks of products.
Goodwill:	Also called a "going concern value." Goodwill is recognized when it is acquired for pay, as in a merger.

c. Investments

Investments are assets being held over a long time for profit making, such as long-term loans and shares in subsidiaries, or assets being held for the purpose of keeping subsidiaries or the like under control.

Typical titles of account

Long-term loans:	Claims arising from the extension of loans to others for a period of more than one year. However, the portion of such a loan to be repaid within one year is classed as a short-term loan and thus a liquid asset.
Investment securities:	Securities being held over a long time for profit making.
Shares in subsidiaries:	Shares in subsidiaries being held over a long time for purposes such as stabilizing their management.

③ Deferred assets

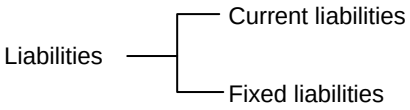
Deferred assets are expenses that are temporarily classed as assets, since their benefits extend to the next fiscal year onward.
There are eight types of deferred assets as shown below. All of them need to be amortized in each fiscal year as expenses.
Amortization periods are prescribed in the Commercial Code.

Titles of account

Bond issue costs:	Expenses incurred for issuing bonds. The amortization period is up to three years.
Stock issue costs:	Expenses incurred for issuing additional shares to increase capital. The amortization period is up to three years.
Start-up costs:	Expenses incurred for preparing to start business after the establishment of the company, such as advertising expenses, communication expenses, and salaries. The amortization period is up to five years.
Organization costs:	The general costs of launching a business concern, such as the expenses for preparing the articles of incorporation and for registering the establishment of the company. The amortization period is up to five years.
Development expenses:	Expenses incurred for the development of new products, new markets, and the like. The amortization period is up to five years.
Research expenses:	Expenses incurred for research on new products and new technologies. The amortization period is up to five years.
Bond issue discounts:	The difference between the face value of bonds and the amount of issue. The amount of issue is smaller than the face value of bonds. The amortization period is the bond redemption period.
Interest during construction:	The amount paid to shareholders for a certain period up to the start of business under the Commercial Code when the company remains idle for two or more years after its establishment. Each time interest exceeding 6 percent of the capital stock is paid per year, the amount equal to or larger than the excess needs to be amortized.

(2) Liabilities

According to the one-year rule and the normal operating cycle rule, liabilities are divided into current liabilities and fixed liabilities.



- One-year rule

  Liabilities that are to be settled within one year of a closing day are current liabilities. For example, a loan to be repaid within a year is a short-term loan and a current liability. A loan not to be repaid within one year is a long-term loan and a fixed liability.

- Normal operating cycle rule

  When liabilities arise as part of a business's main activities (purchasing, production, and sales) are classified as current liabilities even if they are not to be repaid within one year. The normal operating cycle rule applies to such titles of account as accounts payable and notes payable.

  For example, the price of an article sold under an installment contract is to be paid in 36 monthly installments, these installments are not classified into current liabilities and fixed liabilities depending on whether they are payable within one year or not; instead, the entire price is recorded as a current liability.

  However, this rule does not apply to the outstanding balance for a fixed asset purchased, since the purchase is not a transaction made in the course of the main business activities. The balance is recorded as a current liability or a fixed liability according to the one-year rule.

① Current liabilities

Current liabilities are liabilities that must be settled within year according to the one-year rule or liabilities arising in the process of business activity according to the normal operating cycle rule.

Typical titles of account

Accounts payable:	Liabilities arising for the purchase of merchandise, materials, and services yet to be paid for.
Notes payable:	Notes issued in the course of ordinary transactions whose face amounts are to be paid on promised dates.
Non-trade accounts payable:	Liabilities arising for the purchase of other than merchandise, materials, and services yet to be paid for.
Short-term loans:	Liabilities arising from receiving loans from banks and others repayable within one year.

② Fixed liabilities

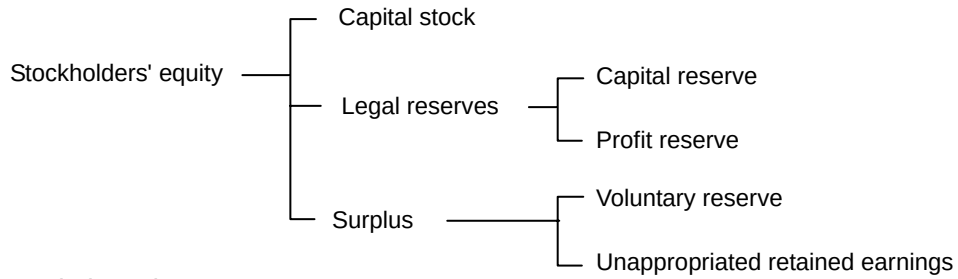
Fixed liabilities are liabilities not to be settled within one year according to the one-year rule. Fixed liabilities include long-term loans, bonds, and allowance for employee retirement and severance benefits.

Typical titles of account

Long-term loans:	Liabilities arising from receiving loans from banks and others not repayable within one year.
Bonds:	Debt instruments issued by an enterprise to borrow long-term funds from the general public.
Allowance for employee retirement and severance benefits:	The appropriation made by an enterprise by setting aside estimated amounts to prepare for the payment of benefits to retiring employees.

(3) Stockholders' Equity

Depending on the characters of titles of account, stockholders' equity is classified into three categories: capital stock, legal reserves, and surplus.



① Capital stock

Capital stock means the funds collected from shareholders for the operation of an enterprise.

Typical title of account

Capital stock:	The amount of funds paid in by shareholders. Strictly, capital stock includes the amounts corresponding to free share issues to shareholders and those corresponding to the conversion of convertible bonds.

② Legal reserves

The reserve specified by the Commercial Code. As legal reserves, there are profit reserve and capital reserve.
Legal reserves are used either to be converted into capital stock or to make up a deficit. Legal reserves may not be reversed and used for other purposes.

a. Profit reserve

Profit reserve is the reserve which an enterprise is obligated to have by the Commercial Code by setting aside at least 10 percent of its profit until the total amount reaches one-quarter of its capital stock.

Typical title of account

Profit reserve:	The amount of reserve, which is more than one-tenth, of the amount disbursed by a company as profit disposition out of the profits generated as a result of ordinary transactions conducted by the company.

b. Capital reserve

Capital reserve is the reserve which an enterprise is obligated to have by the Commercial Code by setting aside the amounts arising from capital transactions: stock issues, capital increase or decrease, and mergers. Set aside as capital reserve are additional paid-in capital in most cases.

Typical titles of account

Additional paid-in capital:	This is the part of the amount of a stock issue not converted into capital stock. That is, when the value on the stock market exceeds the face value of the issue, the board of directors may determine the part of the difference not to be converted into capital stock.
Surplus from reduction of capital stock:	This is the amount by which the reduced capital stock exceeds the stock canceled or redeemed or the deficit made up.
Gain from merger:	This is the amount by which the net worth of a company acquired in a merger exceeds the total amount of payments made to the shareholders of the acquired company or the total face value of shares delivered to those shareholders.

③ Surplus

Surplus is different from legal reserves, which are required by law. Surplus is the profit accumulated in accordance with a company policy adopted at a general meeting of shareholders.
There are two types of surplus: voluntary reserve and unappropriated retained earnings.

a. Voluntary reserve

Voluntary reserve is made by setting aside and retaining parts of the company's earnings. Unlike legal reserves, voluntary reserve may be used for the specific purpose for which it is made.

Typical titles of account

Reserve for construction:	Reserve for the construction of a new office building.
Reserve for dividends:	Reserve to pay dividends to shareholders.
Special reserve:	Reserve for no specified purpose.

b. Unappropriated retained earnings

Unappropriated retained earnings are profit that has not been appropriated yet at a general meeting of shareholders. At a general meeting of shareholders, unappropriated retained earnings are divided into the payout portion (i.e., dividends to shareholders and officers' bonuses) and the retained portion (i.e., profit reserve, voluntary reserve, and profit carried forward).

Typical title of account

Unappropriated retained earnings:	The amount obtained by adding the profit brought forward and so forth to net profit and subtracting interim dividends, provision for profit reserve, and so forth from the sum.

(4) Balance Sheet Principles

Balance sheet principles are part of the business accounting principles, which are the "constitution of accounting." Balance sheet principles contain detailed rules for the preparation of the balance sheet.
This section describes some of the balance sheet principles.

① Balance sheet integrity principle

First, the balance sheet principles stipulate that in order to clarify the financial conditions of the enterprise, the balance sheet must state all assets, liabilities, and stockholders' equity as of its date and present them fairly to shareholders, creditors, and other stake holders. What is important about this stipulation concerning the contents of the balance sheet is that "all" assets, liabilities, and stockholders' equity must be stated. This is called the "balance sheet integrity principle."

② Gross amount principle

The balance sheet principles also provide for the statement of amounts of assets, liabilities, and stockholders' equity. The principles stipulate that assets, liabilities, and stockholders' equity must be stated in their gross amounts in principle and that the amounts must not be totally or partly deleted from the balance sheet by offsetting capital items by liability or stockholders' equity items.
That is, assets, liabilities, and stockholders' equity must be presented in their gross amounts; it is prohibited to directly offset the amount of capital by that of liabilities and stockholders' equity. This principle is called the "gross amount principle."

③ Section and arrangement principles

The balance sheet principles also provide for balance sheet sections and the arrangement of balance sheet items. That is, the section and arrangement principles require that the balance sheet be divided into three sections, the assets section, the liabilities section, and the stockholders' equity section, and that the assets section be subdivided into current assets and fixed assets and the liabilities section into current liabilities and fixed liabilities (Figure 1-2-2). The section and arrangement principles further require that asset and liability items be arranged by current-first order.
Current-first order is the method of arranging asset or liability items by declining order of liquidity. The opposite of current-first order arrangement is fixed-first order arrangement.

Figure 1-2-2
Balance Sheet Sections

Balance Sheet

<Assets>	<Liabilities>
Current assets	Current liabilities
Liquid assets	Fixed liabilities
Inventories	
Other current assets	
Fixed assets	< Stockholders' equity >
Tangible fixed assets	Capital stock
Intangible fixed assets	Capital reserve
Investments	Profit reserve
	Surplus
Deferred assets	

1.2.2 How to Read the Income Statement

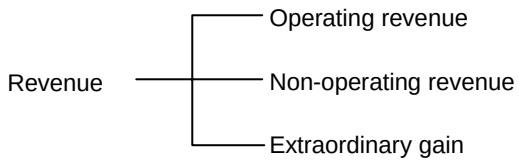
Figure 1-2-3 Example of the Income Statement (Profit and Loss statement; P/L)

From: the date when the term begins
Through: the date when the term ends (Thousands of yen)

Title of Account	Amount
Operating Revenue	35,200
Cost of goods sold	1,200
Gross Income	34,000
Selling, general and administrative expenses	32,000
Operating Income	2,000
Non-operating revenue	960
Non-operating expenses	750
Ordinary Income	2,210
Extraordinary gain	100
Extraordinary loss	210
Income before taxes	2,100
Provision for corporate income and inhabitant taxes	900
Net income	1,200
Balance brought forward	200
Interim dividends paid	50
Profit reserve	20
Unappropriated retained earnings	1,330

(1) Revenue

By character, revenue can be divided into three categories: operating revenue, non-operating revenue, and extraordinary gain.



① Operating revenue

Operating revenue is revenue arising from the main business activity of an enterprise. In the case of general companies, operating revenue is sales themselves. That is, operating revenue may be considered equal to sales revenue in these companies.

Typical title of account

Sales revenue:	Earnings obtained through the essential business activity of an enterprise.

② Non-operating revenue

Non-operating revenue is revenue arising recurringly from activities other than the main business activity of an enterprise. Typical examples are financial revenue such as stock dividends and interest received from financial engineering activities (financial activities).

Typical titles of account

Interest received:	Interest received on loans, deposits, and the like.
Interest on securities:	Interest received on public and corporate bonds and the like.
Gain on sale of securities:	The excess of the sales price of securities such as shares over their carrying value.

③ Extraordinary gain

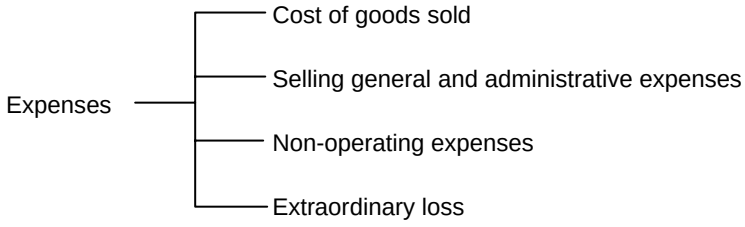
Extraordinary gain is revenue arising temporarily from activities other than the main business activity of an enterprise. The distinction between non-operating revenue and extraordinary gain is whether particular revenue is recurring or temporary. A typical example is a gain on sale of land or a building that was to be owned for a long time.
Other examples of extraordinary gain are the reversal of an allowance used for a purpose other than its originally intended purpose and an increase in revenue arising from the revision or correction of the gain or loss for the previous fiscal year.

Typical title of account

Gain on sale of real estate:	The excess of the sales price of real estate over its carrying value.

(2) Expenses

By character, expenses can be divided into four categories: cost of goods sold, selling, general and administrative expenses, non-operating expenses, and extraordinary loss.



① Cost of goods sold

Cost of goods sold means the expenses incurred for obtaining operating revenue, that is, the cost of merchandise or products themselves. Cost of goods sold is calculated by different methods in the retail business (commercial bookkeeping) and the manufacturing business (industrial bookkeeping). In both methods, however, it is equally important to calculate the cost corresponding to sales by accurately grasping the relations between purchases and inventory.

Typical title of account

Cost of goods sold: The cost of merchandise or products corresponding to sales revenue. It is the cost of goods purchased in the case of merchandise and the cost of goods manufactured in the case of products.

- In the case of the retail business:
  Cost of merchandise = beginning merchandise inventory + merchandise purchased - ending merchandise inventory
- In the case of the manufacturing business:
  Manufacturing expenses = materials expenses + labor expenses + other expenses
  Cost of goods manufactured = beginning inventory of goods in process + manufacturing expenses + ending inventory of goods in process
  Cost of goods sold = beginning product inventory + cost of goods manufactured + ending product inventory

Inventory Valuation Methods

Grasping the cost of goods sold requires accurate valuation of the existing merchandise or product inventory. In practice, the quantities, unit prices, and amounts of merchandise or products are recorded in the book "stock ledger." The goods are thus managed as inventories. At this time, unit prices and amounts are recorded on a cost basis. When goods of the same type were purchased at different unit prices, it is a question as to how to calculate the unit prices. In this case, the unit price is calculated by one of the following methods:

I. First-in first-out method (FIFO)
The unit price is calculated on the assumption that goods were delivered in the order of their purchase.

II. Last-in first-out method (LIFO)
The unit price is calculated on the assumption that goods were delivered in the reverse order of their purchase.

III. Moving average method
Each time goods are purchased, the unit price is calculated in accordance with the following formula:

$$\text{Unit price} = \frac{\text{inventory amount} + \text{purchase price}}{\text{inventory volume} + \text{purchased volume}}$$

② Selling, general and administrative expenses

Selling, general and administrative expenses are expenses incurred to obtain operating revenue. Selling, general and administrative expenses are divided into selling expenses incurred in carrying out selling activities and general and administrative expenses incurred for general business administration, such as accounting and general affairs.
In addition, cost of goods sold and selling, general and administrative expenses are collectively referred to as "operating expenses."

Typical titles of account

Advertising expense:	Fees for advertisements placed in newspapers, magazines, and the like for sales promotion.
Payroll (wages):	Personnel expenses, such as salaries to sales people and office workers.
Office-rent:	Rents for leased offices.
Communication expense:	Postage stamp and post card charges, telephone charges, and so forth.

③ Non-operating expenses

Non-operating expenses are recurring expenses arising from activities other than the main business activity of an enterprise. A typical example is financial expenses such as interest paid on loans.

Typical titles of account

Interest paid:	Interest paid on loans from financial institutions and others.
Loss on sale of securities:	The difference by which the sales price of securities, such as shares, is less than their carrying value.
Amortization of organization costs:	The amortization of organization costs, which are deferred assets.

④ Extraordinary loss

Extraordinary loss means temporary expenses arising from activities other than the main business activity of an enterprise. The distinction between non-operating expense and extraordinary loss is whether a particular expense is recurring or temporary. Examples are a loss on sale or retirement of real estate, such as land or a building, and damage suffered from a natural disaster, such as an earthquake or flood.

Typical titles of account

Loss on sale of real estate:	The difference by which the sales price of real estate is less than its carrying value.
Loss on retirement:	The carrying value of real estate retired (discarded).

(3) Income

Income is revenue less expenses. As described above, there are various types of revenue and expenses, and accordingly there are various types of incomes. The incomes in boxes below will be explained one by one.

Ordinary Income (Loss)

[Operating Revenue (Loss)]

Sales

Cost of goods sold

Gross income on sales (gross income)

Selling, general and administrative expenses

Operating income

[Non-operating Revenue (Loss)]

Non-operating revenue

Non-operating expenses

Ordinary income

Extraordinary Gain (Loss)

Extraordinary gain

Extraordinary loss

Net income before taxes

(Income and inhabitant taxes)

Net income

→

Unappropriated retained earnings

① Gross income on sales (gross income)

Gross income on sales, or simply gross income, is the income after the recovery of cost of goods sold. This is calculated by subtracting cost of goods sold from sales revenue.

Gross income on sales (gross income) = operating revenue (sales) - cost of goods sold

② Operating income

Operating income is the income derived from the main business activity of an enterprise. This is calculated by subtracting selling, general and administrative expenses from gross income on sales.

Operating income = gross income on sales - selling, general and administrative expenses

③ Ordinary income

Ordinary income is the income derived from the overall recurring activities of an enterprise. Since ordinary income is the result of the recurring activities of an enterprise, it describes the overall strength of the enterprise. It is thus the most important indicator of the five different incomes.
Ordinary income is calculated by adding non-operating revenue to operating income and subtracting non-operating expenses from the sum.

Ordinary income = operating income + non-operating revenue - non-operating expenses

④ Net income before taxes

Net income before taxes is the income derived as the result of all transactions during the fiscal year. This is the income on which corporation and other taxes are calculated. In reality, however, net income before taxes shown in the income statement does not necessarily agree with the taxable income in a report submitted to the tax bureau because of the different handling of expenses, losses, and so forth.
Net income before taxes is calculated by adding extraordinary income to ordinary income and subtracting extraordinary loss from the sum.

Net income before taxes = extraordinary income + ordinary income - extraordinary loss

⑤ Net income (net profits, net worth)

Net income is the final profit for the fiscal year. Therefore, the word "income" is used alone, it means net income.
Net income is calculated by subtracting corporate income and inhabitant taxes from net income before taxes.

Net income = net income before taxes - (income tax + inhabitant tax)

⑥ Unappropriated retained earnings

Unappropriated retained earnings represent the profit available to be appropriated for shareholders' dividends, bonuses for officers, profit reserve, voluntary reserve, and so forth.
Unappropriated retained earnings are calculated by adding earnings brought forward and so forth to net income and subtracting interim dividends and others from the sum. Unappropriated retained earnings shown in the income statement agree with unappropriated retained earnings shown under "surplus" in the stockholders' equity section of the balance sheet.

The appropriation of unappropriated retained earnings is stated in the appropriation statement, one of the financial statements.

Unappropriated retained earnings
= net income + (earnings brought forward + reversal of voluntary reserve + ...)
- (interim dividend + provision for profit reserve + ...)

(4) Income Statement Principles

Corporation accounting principles include income statement principles, which have detailed rules for the preparation of the income statement.
This section describes some of the income statement principles.

① Section principle

Income statement principles require that the income statement has sections for the calculation of operating income or loss, that of ordinary income or loss, and that of net profit or loss. This requirement is called the "section principle." In accordance with this principle, the income statement is divided into the ordinary income (loss) section and the extraordinary income (loss) section, the former section showing the income (loss) arising from recurring activities of the enterprise and the latter section showing the income (loss) arising from non-recurring activities. Furthermore, the ordinary income section is subdivided into the operating income (loss) section and the non-operating income section, the former section showing the income arising from the main business activities of the enterprise and the latter section showing the profit arising from other activities (Figure 1-2-4).

Figure 1-2-4 Sections of the Income Statement

Income Statement

<Expenses>	< Revenue >
Cost of goods sold	Operating revenue
Selling, general and administrative expenses	Non-operating revenue
Non-operating expenses	Extraordinary gain
Extraordinary loss	
(Net income)	

② Income statement integrity principle

Income statement principles first require that in order to clarify the operating performance of the enterprise, the income statement presents ordinary income, showing all revenue belonging to a fiscal year and all corresponding expenses, and presents net profit by adding and subtracting extraordinary revenue items to and from ordinary income. What is important about this stipulation concerning the contents of the income statement is that "all" revenue and expenses must be stated. This is called the "income statement integrity principle."

③ Gross amount principle

The income statement principles also provide for the statement of amounts of revenue and expenses.
The principles stipulate that revenue and expenses must be stated in their gross amounts in principle and that the amounts must not be totally or partly deleted from the income statement by offsetting revenue items by expense items.
That is, as in the balance sheet, revenue and expenses must be presented in their gross amounts; it is prohibited to directly offset the amount of revenue by that of expenses. This principle is called the "gross amount principle."

④ Accrual principle

The basis for income determination means a method for recognizing revenue and expenses in a particular year. There are a few different bases:

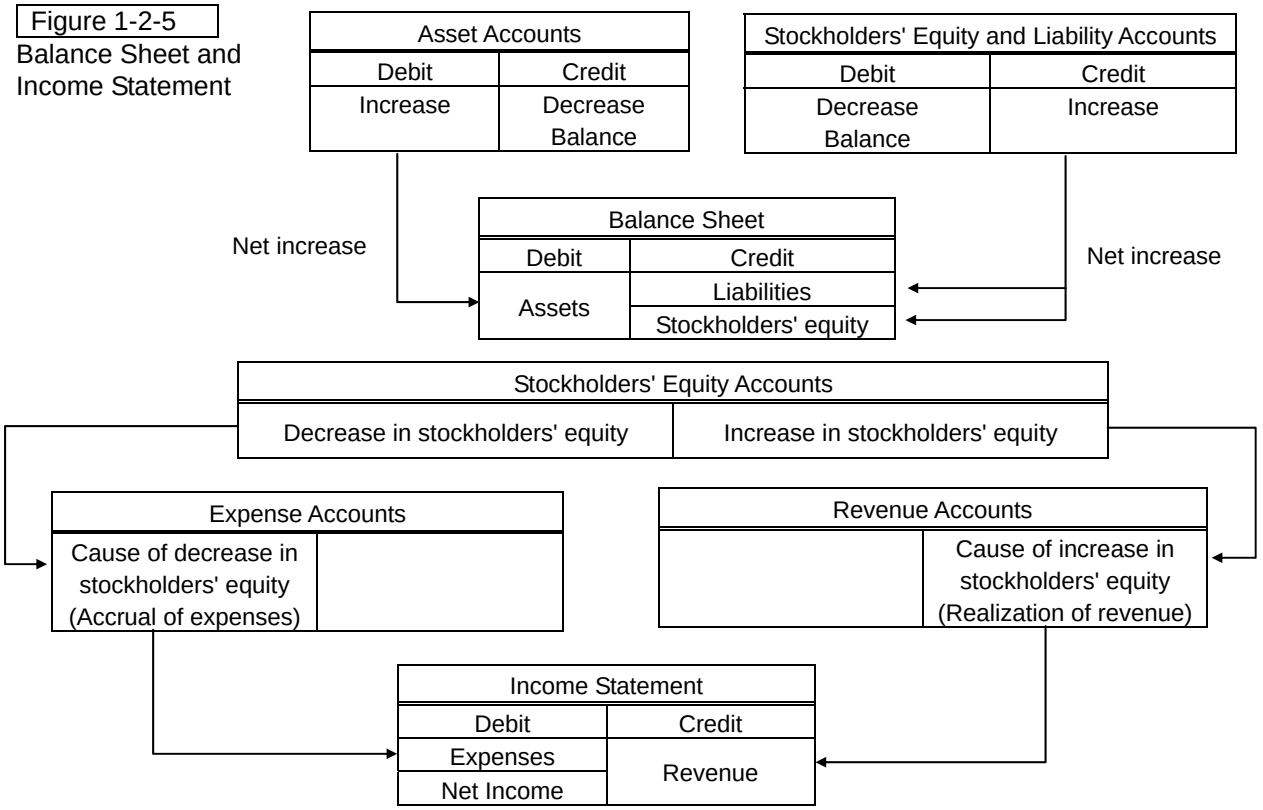
- Cash basis
  In cash basis accounting, revenue and expenses are recognized in the fiscal year in which cash is actually received and paid out. In this method, accounts receivable and accounts payable are not recorded, whereas advance receipts and advance payments are recorded, thus making it impossible to reasonably calculate the profit for the fiscal year.
- Accrual basis
  In accrual accounting, revenue and expenses are recorded completely irrespective of whether or not cash is received and paid out. That is, revenue and expenses are reflected in income determination as they accrue. In this method, advance receipts and advance payments are not recorded, whereas accounts receivable and accounts payable are.
- Realization basis
  In realization basis accounting, revenue and expenses are basically recorded on an accrual basis with some restrictions on the recording of revenue. That is, only realized revenue is recorded; revenue not yet realized is not. An exception, however, realization basis accounting permits the recording of gain from ongoing construction under a long-term contract.

In this respect, income statement principles stipulate as follows: "All expenses and revenue must be recorded as they are paid out and received, being allocated correctly to the year of their accrual. However, revenue not yet realized must not be reflected in income determination in principle. Prepaid expenses and unearned revenue must be excluded from income determination for the current fiscal year, whereas accrued expenses and accrued revenue must be reflected in income determination for the current fiscal year."
This means that expenses must be recorded on an accrual basis and revenue on a realization basis in principle.

⑤ Principle of matching costs with revenues

The principle of matching costs with revenues is one of the income statement principles. This principle is that expenses and revenues must be clearly classified according to the sources of their accrual and that revenue items and corresponding expense items must be presented in a corresponding manner in the income statement. This means that the expenses incurred in a fixed period and the revenues realized in the same period must be presented in a corresponding manner for the purpose of income determination.

Figure 1-2-5
Balance Sheet and Income Statement



(Source: "Class II Common Curriculums" edited by Central Academy of Information Technology, Japan Information Processing Development Corporation)

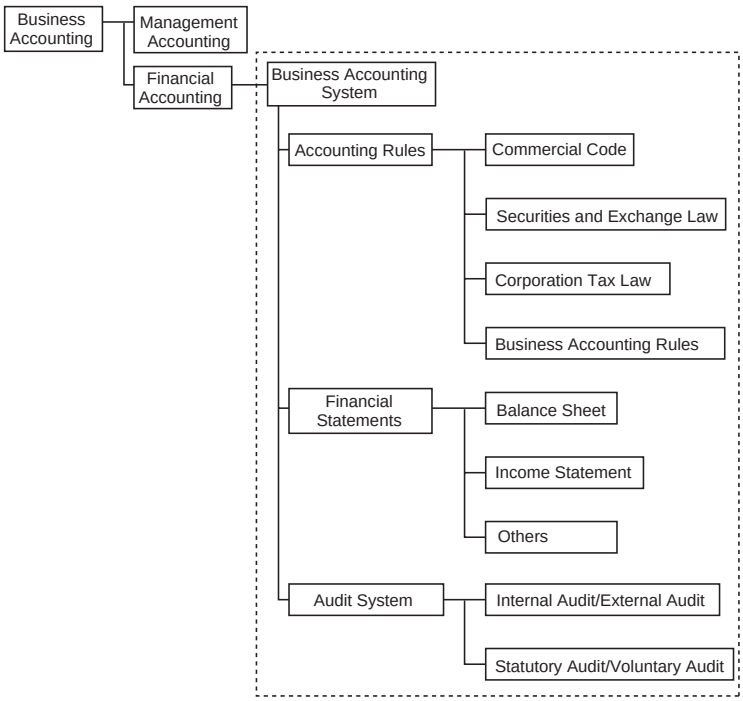
1.3 Financial Accounting and Management Accounting

Financial statements, including the balance sheet and the income statement and other accounting records provide useful accounting information for stakeholders inside and outside a business. Accounting information is divided, according to purpose, into financial accounting and management accounting.

1.3.1 Financial Accounting

The objective of financial accounting is to report the results of activities of a business to the stakeholders, namely, stockholders, employees, creditors, public institutions, and the community. The results of the activities of a business are made public through the balance sheets, the income statement, and other reports. Financial accounting is also called "corporation accounting." As shown in Figure 1-3-1, corporation accounting is governed by various laws and conventions concerning the obligation to prepare financial statements, the standards for preparation, and other regulations.
Note: The laws and explanations onwards are given as examples of accounting systems in Japan. International standards are described in 1.3.4 International standards.

Figure 1-3-1
Corporation Accounting System



Major laws and conventions concerning the corporation accounting system are outlined below.

(1) Commercial Code

The Commercial Code requires the preparation of financial statements from the viewpoint of creditor

protection. The financial statements as defined by the Commercial Code are the balance sheet, the income statement, the business report, and the proposal of appropriation of earnings (or disposition of deficit). All these statements must conform to the account statement rules ("Regulations Concerning the Balance Sheet, the Income Statement, the Business Report, and Supplementary Schedules of Joint Stock Companies"). Supplementary schedules to these statements must also be prepared.

(2) Securities and Exchange Law

The Securities and Exchange Law requires the preparation of financial statements from the viewpoint of investor protection. The financial statements as defined by the Securities and Exchange Law are the balance sheet, the income statement, supplementary schedules, and the earning appropriation statement. These must conform to the financial statement rules ("Regulations Concerning Terminology, Forms, and Method of Preparation of Financial Statements, etc.").
The appropriation of earnings is prepared as a "proposal" under the Commercial Code because it has to be submitted for approval to the annual meeting of stockholders and as a "statement" under the Securities and Exchange Law. It is prepared after the appropriation is approved at the annual meeting of stockholders.

(3) Corporation Tax Law

To ensure proper taxation, the Corporation Tax Law has various provisions requiring the preparation of financial statements. The law requires a corporation to submit a return on corporation tax accompanied by the balance sheet, the income statement, supplementary schedules, and the appropriation statement.

(4) Corporation Accounting Principles

Corporation accounting principles were established by the Corporation Accounting Council of the Finance Ministry in 1949 and have been amended a number of times since then. Although these principles are not law, they are fair accounting practices that must be always observed. In fact, related laws are based on the "constitution of accounting" based on related laws that are enacted.
The financial statements required by corporation accounting principles are the balance sheet, the income statement, supplementary schedules to financial statements, and the statement of appropriation of earnings (or disposition of deficit).
The corporation accounting principles consist of general principles, income statement principles, and balance sheet principles. They serve as theoretical and practical guiding principles for corporation accounting as well as guidelines for amending and abolishing laws and regulations and for audits.
The general principles are particularly important, spelling out seven fundamental concepts in corporation accounting.

1.3.2 Management Accounting

Management accounting is the type of accounting in which the internal management staff provides top executives with information to administer current affairs and make projections for the future. In management accounting, therefore, information on transactions is provided as well as management baselines or targets, such as numerical plans and budgets. These are established so that actual results can be compared for measurement and analysis. In addition, techniques such as multivariate analysis and econometric analysis are used to provide top management with information for decision making. This is why management accounting is also called "accounting for decision making."
At any rate, management accounting is performed based on the financial statements prepared under the corporation accounting system. The financial statements enable managers to read the enterprise's financial position, operating performance, prospects, and so forth. This process is called business analysis, financial analysis, or financial statement analysis.

(1) Financial Statement Analysis

Business analysis is the process of reading the balance sheet and the income statement and judging whether the enterprise is doing well or not.
Business analysis is classified into the following two types:

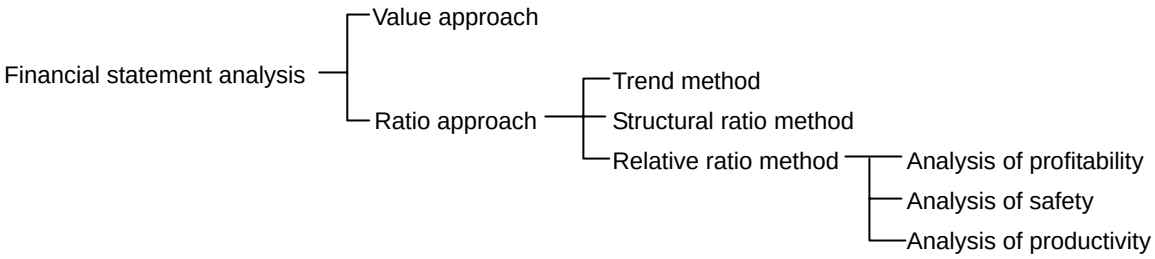
- External analysis:
  This is the analysis performed by outside people to objectively judge the enterprise's financial and operating conditions. External analysis corresponds to financial statements in corporation accounting. Examples of external analysis are investment analysis by investors and the examination of the credit standings of borrowers by financial institutions.
- Internal analysis:
  This is the analysis performed by people inside an enterprise to grasp the current conditions for determining policies for the future. Internal analysis corresponds to management accounting in corporation accounting. An example of internal analysis is the formulation of management plans by managers.

Business analysis techniques fall into the following two categories:

- Value approach:
  This is the approach in which analysis is performed using the values (amounts) stated in financial statements. In a typical method, year-by-year business conditions are compared using a comparative balance sheet, a comparative income statement, and other financial statements summarizing the financial and operating conditions for multiple fiscal years.
- Ratio approach:
  This is the approach in which analysis is performed using the ratios of various values (amounts) stated in financial statements.

The ratio approach can be subdivided into three methods:

- Trend method:
  This method analyzes changes of individual items from a base fiscal year (100).
- Structural ratio method:
  Also called the "percentage method," this method analyzes the ratio of each item to a total amount (100). The total amount is total stockholders' equity in the case of the balance sheet and sales revenue in the case of the income statement.
- Relative ratio method:
  This method analyzes the ratio of one item to another in financial statements.



The relative ratio method is most commonly used in business analysis. The ratios used in this analysis method can be divided into "static ratios," which are the ratios between items stated in the balance sheet, and "dynamic ratios," which are either the ratios between items stated in the income statement or the ratios between items in the balance sheet and those in the income statement.
By the ratios used or purpose, the relative ratio method is subdivided into three types:

- Analysis of profitability
- Analysis of safety (liquidity)
- Analysis of productivity

(2) Analysis of Profitability

The analysis of profitability is performed to check how efficiently an enterprise is making net income.
Five ratios are used for this analysis: the ratio of net income to stockholders' equity, the ratio of net income to sales revenue, the ratio of expenses to sales revenue, the equity turnover, and the asset turnover.

① Ratio of income to stockholders' equity

The ratio of income to stockholders' equity is the percentage of income to the stockholders' equity. This shows how much profit the stockholders' equity has produced. Naturally, the higher the ratio, the higher the profitability of the stockholders' equity.
The typical ratio of income to the stockholders' equity is the ratio of income to gross equity (borrowed equity + owner's equity):

$$\text{Ratio of income to gross equity} = \frac{\text{income before tax}}{\text{gross equity}} \times 100\ (\%)$$

Income in this case is usually net income before taxes.
In this equation, net income (after taxes) or ordinary income may be used as the numerator to determine how much net income (after taxes) or ordinary income the gross equity has produced.
Furthermore, owner's equity may be used as the denominator in this equation to determine how much income the owner's equity has produced:

$$\text{Ratio of net income to owner's equity} = \frac{\text{income}}{\text{owner's equity}} \times 100\ (\%)$$

② Ratio of income to sales

The ratio of income to sales is the percentage of income to sales revenue. This shows how much income was derived from sales, the base of revenue. Therefore, the higher the ratio, the higher the profitability.
This ratio has some variations, depending on different types of income used:

$$\text{Ratio of gross income to sales} = \frac{\text{gross income}}{\text{sales}} \times 100\ (\%)$$

$$\text{Ratio of operating income to sales} = \frac{\text{operating income}}{\text{sales}} \times 100\ (\%)$$

$$\text{Ratio of ordinary income to sales} = \frac{\text{ordinary income}}{\text{sales}} \times 100\ (\%)$$

$$\text{Ratio of net income to sales} = \frac{\text{net income before tax}}{\text{sales}} \times 100\ (\%)$$

The ratio of income to sales enables the comparison of the profitability levels of an enterprise over multiple years or with those of competitors.
In addition, the comparison of the variations of this ratio shown above enables the important work of determining at which level income is small or high.

③ Ratio of expenses to sales

The opposite of the ratio of income to sales, the ratio of expenses to sales is the percentage of expenses to sales revenue. Since smaller expenses means larger profit, the lower this ratio, the higher the profitability.
This ratio has some variations, depending on different types of expenses used:

$$\text{Ratio of cost to sales} = \frac{\text{cost of goods sold}}{\text{sales}} \times 100\ (\%)$$

$$\text{Ratio of selling, general and administrative expenses} = \frac{\text{selling, general and administrative expenses}}{\text{sales revenue}} \times 100\ (\%)$$

The sum total of the ratio of gross income to sales and the ratio of cost of goods sold to sales revenue is 1.
Meanwhile, it is important to use subdivided expenses as the numerator to learn which expenses are increasing or decreasing. For example, expenses affecting the performance of an enterprise, such as raw materials cost (a component of the cost of goods sold), advertising expenses (a component of selling, general and administrative expenses), and interest paid on loans are used as the numerator as shown below.

$$\text{Ratio of raw materials cost to sales} = \frac{\text{raw materials cost}}{\text{sales}} \times 100\ (\%)$$

$$\text{Ratio of advertising expenses to sales} = \frac{\text{advertising expenses}}{\text{sales}} \times 100\ (\%)$$

$$\text{Ratio of interest paid to sales} = \frac{\text{interest paid}}{\text{sales}} \times 100\ (\%)$$

④ Equity turnover

Equity turnover is the percentage of sales to stockholders' equity. This ratio shows how many times the stockholders' equity was used in an accounting period. The higher the ratio, the higher the profitability.
Equity turnover is either gross equity turnover or owner's equity turnover, depending on whether gross equity or owner's equity is used as the numerator:

$$\text{Gross equity turnover} = \frac{\text{sales}}{\text{shareholders' equity}}\ \text{(number of times)}$$

$$\text{Owner's equity turnover} = \frac{\text{sales}}{\text{owner's equity}}\ \text{(number of times)}$$

An enterprise invests capital, obtains revenue by using it, and records income. As a result, the enterprise can invest additional capital. This flow is a single turn of capital. Gross equity turnover shows how many times the invested capital was turned over within an accounting period and how much it contributed to sales.
High turnover means that relatively small capital has produced relatively large sales. That is, high turnover means that stockholders' equity has been used effectively.

⑤ Asset turnover

Asset turnover is the percentage of sales or cost of goods sold to assets. This ratio shows the number of times assets were used in an accounting period. The higher the ratio, the high the profitability.
Asset turnover has some variations, depending on what is used as the denominator as shown below:

$$\text{Merchandise turnover} = \frac{\text{cost of goods sold}}{\text{average merchandise inventory}}\ \text{(number of times)}$$

The higher the merchandise turnover, the fewer the number of days required for a single turn of inventory. That is, a high merchandise turnover means that merchandise is selling well. The average merchandise inventory is obtained as follows:
(beginning merchandise inventory + ending merchandise inventory) ÷ 2.
If cost of goods sold is unknown, sales may be used instead as the numerator.

$$\text{Fixed asset turnover} = \frac{\text{sales}}{\text{fixed assets}}\ \text{(number of times)}$$

The higher the fixed asset turnover, the more effectively the fixed assets are used. If the ratio is low, it means that equipment investment is excessive.

$$\text{Receivables turnover} = \frac{\text{sales}}{\text{notes receivable + accounts receivable}}\ \text{(number of times)}$$

The higher the receivables turnover, the more promptly receivables are collected. That is, a high receivables turnover means the enterprise is free from concern about cash flow.
Since the amount of total assets is equal to that of gross equity, a high turnover of each component of total assets leads to a high gross equity turnover.

(3) Analysis of Safety

A safety analysis is used to determine whether the assets necessary for business activities are operated in a sound manner and whether financial conditions, such as the ability to pay, are good enough. The analysis of safety is also called the "analysis of liquidity."
The safety of an enterprise is analyzed based on static ratios, or based on the relations between the asset, liability, and stockholders' equity items on the balance sheet, from the viewpoints of short-term and long-term safety.

① Short-term safety ratios

A short-term safety ratio is a ratio to examine the enterprise's current ability to pay. A high short-term safety ratio means that the enterprise is financially safe or has an adequate cash flow. There are two major short-term safety ratios, the current ratio and the acid test ratio, depending on whether current assets or liquid assets is used as the numerator.

$$\text{Current ratio} = \frac{\text{current assets}}{\text{current liabilities}} \times 100\ (\%)$$

The current ratio shows the enterprise's ability to pay liabilities. It is the percentage of current assets to current liabilities. More specifically, the ratio shows how much the enterprise has in assets that can be converted into cash in a short period to cover liabilities that need to be paid in the short period. It is generally considered desirable that the current ratio be 200 percent or more.

$$\text{Acid test ratio} = \frac{\text{liquid assets}}{\text{current liabilities}} \times 100\ (\%)$$

The acid test ratio is the percentage of liquid assets to current liabilities. Although this ratio also concerns the ability to pay, it shows how much the enterprise has in assets that can be converted into cash more readily. Naturally, the acid test ratio is lower than the current ratio. It is generally considered desirable that the acid test ratio be 100 percent or more.

② Long-term safety ratios

Long-term safety ratios measure the enterprise's potential ability to pay over a long term. High long-term safety ratios mean that the enterprise is financially safe.

$$\text{Owner's equity ratio} = \frac{\text{owner's equity}}{\text{total assets}} \times 100\ (\%)$$

The owner's equity ratio is the percentage of owner's equity to total assets. A high owner's equity ratio means a small amount of liabilities (borrowed equity and liabilities), that is, a sound financial position.

$$\text{Debt ratio} = \frac{\text{total liabilities}}{\text{owner's equity}} \times 100\ (\%)$$

$$\text{Debt/equity ratio} = \frac{\text{owner's equity}}{\text{total liabilities}} \times 100\ (\%)$$

The debt ratio is the percentage of total liabilities to owner's equity. The debt/equity ratio is the percentage of owner's equity to total liabilities. Both ratios are used to check whether the enterprise has too many liabilities as compared with its owner's equity. If the enterprise can cover all its liabilities with its owner's equity, its financial position is safe. Therefore, it is desirable that the debt/equity ratio be 100 percent or over. Conversely, the debt ratio should be low.

$$\text{Owner's equity to fixed asset ratio} = \frac{\text{owner's equity}}{\text{fixed assets}} \times 100\ (\%)$$

$$\text{Fixed ratio} = \frac{\text{fixed assets}}{\text{owner's equity}} \times 100\ (\%)$$

$$\text{Fixed assets to long-term equity ratio} = \frac{\text{fixed assets}}{\text{owner's equity + fixed liabilities}} \times 100\ (\%)$$

The owner's equity to fixed asset ratio is the percentage of owner's equity to fixed assets. The fixed ratio is the percentage of fixed assets to owner's equity. Both ratios show the how large a part of owner's equity is used as fixed assets. Financially, it is desirable that fixed assets be covered by part of owner's equity and that the remainder of owner's equity be applied as current assets. It is in turn desirable that the owner's equity to fixed asset ratio be 100 percent or over and that the fixed ratio be less than 100 percent. The fixed assets to long-term equity ratio is based on the idea that even if fixed assets cannot be covered by owner's equity, it should be covered by the total of owner's capital and fixed liabilities that need not be paid for the time being.

(4) Break-Even Analysis

The break-even point is a point at which an enterprise makes neither a profit nor a loss, that is, a point at which an enterprise makes no operating income. Sales at this point are called "break-even sales revenue." In commercial bookkeeping, values on financial statements are analyzed in terms of ratios for the purpose of profitability and safety analysis. In industrial bookkeeping (manufacturing industries), break-even analysis is widely used.

① Income planning

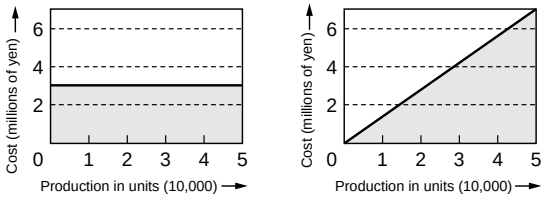
Income planning is the process of setting an income target for a certain future period and planning business activities to achieve the target. Break-even analysis is a particularly effective method for formulating a short-term profit plan.
That is, an income plan is made by grasping how costs will change when sales, production, and other business activities change. The method used at this time to control costs is called "direct costing."

② Fixed costs and variable costs

In direct costing, expenses are divided into fixed costs and variable costs for the purpose of control.

- Fixed costs:
  Expenses that remain constant in total, regardless of changes in sales or production. These expenses are required to maintain sales and production activities and are incurred even if sales or production is zero. Fixed costs include rents, insurance premiums, taxes, and depreciation and amortization costs.
- Variable costs:
  Expenses that increase or decrease in direct proportion to sales or production. Variable costs increase if sales or production increases and decreases if sales or production decreases. Variable costs includes direct materials expenses, packing and transportation expenses, commissions to consignees, wrapping expenses, and commissions to sales representatives.

Figure 1-3-2
Fixed Costs and Variable Costs



(a) Fixed costs (b) Variable costs

③ Fixed cost ratio and variable cost ratio

The percentages of fixed costs and variable costs to sales are called the "fixed cost ratio" and the "variable cost ratio," respectively.

$$\text{Fixed cost ratio} = \frac{\text{fixed costs}}{\text{sales}} \times 100\ (\%)$$

$$\text{Variable cost ratio} = \frac{\text{variable costs}}{\text{sales}} \times 100\ (\%)$$

④ Break-even sales revenue

If three figures—fixed costs, variable costs, and sales—are known, break-even sales revenue can be immediately calculated by the following equation:

$$\text{Break-even sales revenue} = \frac{\text{fixed costs}}{1 - \dfrac{\text{variable costs}}{\text{sales}}} = \frac{\text{fixed costs}}{1 - \text{variable cost ratio}}$$

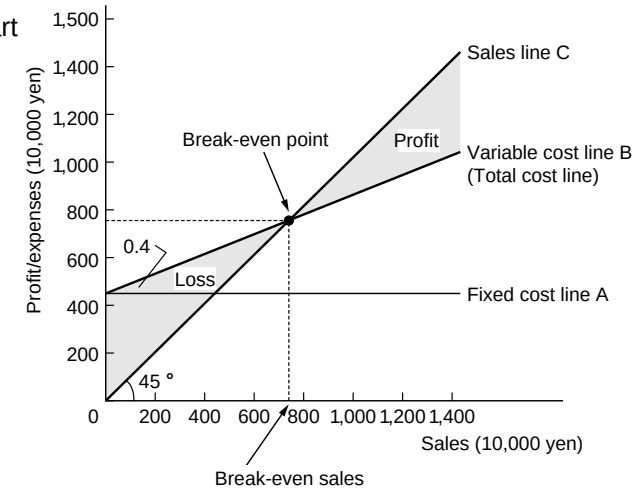
For example, when fixed costs are 4.5 million yen, variable costs 3.6 million yen, and sales 9.0 million yen, break-even sales revenue are calculated as follows:

$$\text{Break-even sales revenue} = \frac{4.5}{1 - \frac{3.6}{9.0}} = 7.5 \text{ (million yen)}$$

⑤ Profit and loss chart

Break-even sales can be calculated not only by the equation shown above but also by drawing a chart. A chart drawn for this purpose is called a "profit and loss chart" or a "break-even chart."
The profit and loss chart shows the relations between sales, expenses, and operating income. It shows how expenses and profit change when sales increase or decrease.
When fixed costs are 4.5 million yen, variable costs 3.6 million yen, and sales 9.0 million, the profit chart is drawn as follows:

Figure 1-3-3
Profit and Loss Chart



1. The horizontal axis represents sales, and the vertical axis represents profit and expenses.
2. Plot fixed cost of 4.5 million yen on the vertical axis, and from that point, draw line A in parallel to the horizontal axis. Line A is a fixed cost line.
3. From the position of 4.5 million yen on the vertical axis, draw gradient line B that represents the variable cost ratio. Line B is a variable cost line. The variable cost ratio is calculated as follows: variable costs (3.6 million yen) / sales (9.0 million yen) = 0.4. When fixed costs are also taken into account, line B represents total costs.
4. With respect to sales, draw gradient line C from the point of origin at an angle of 45 degrees. Line C is a sales line.
5. The break-even point is the point of intersection of the sales line C and the variable cost line B. The point of intersection of a line drawn from this point perpendicularly to the horizontal axis and the horizontal axis represents break-even sales (7.5 million yen).

⑥ Marginal profit

Marginal profit, also called "contribution profit," is calculated by subtracting variable costs from sales. Therefore, profit can be calculated by subtracting fixed costs from marginal profit.
In direct costing, expenses are considered in two stages. In the first stage, variable costs are recovered from sales, and in the second stage, fixed costs are recovered, figuring out operating income. Marginal profit means the gain calculated in the first stage.
The ratio of marginal profit to sales is called a "marginal profit ratio." The sum total of the marginal profit ratio and the variable cost ratio is 1.

$$\text{Marginal profit} = \text{sales} - \text{variable costs}$$

$$\text{Marginal profit ratio} = \frac{\text{marginal profit}}{\text{sales}} \times 100 \text{ (\%)}$$

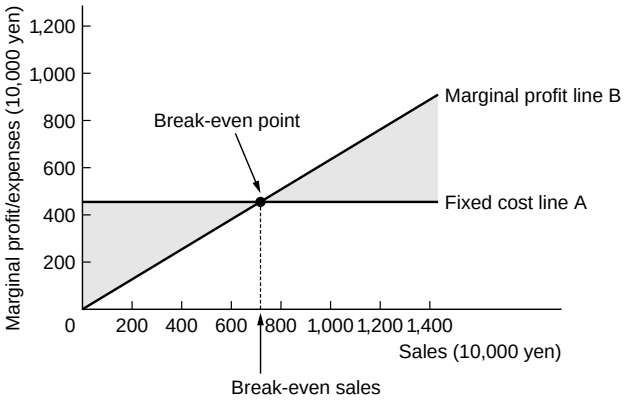
Meanwhile, the equation for calculating break-even sales mentioned above is written as follows using marginal profit:

$$\text{Break-even sales revenue} = \frac{\text{fixed costs}}{1 - \text{variable cost ratio}} = \frac{\text{fixed costs}}{\text{marginal profit ratio}}$$

⑦ Method of drawing a marginal profit chart

The marginal profit chart shows the relations between marginal profit, fixed costs, and profit (loss). Therefore, while the chart is unsuitable for the control of sales and variable costs, it makes it possible to quickly grasp the relation between fixed costs and profit (loss). This is a convenient chart for enterprises handling large varieties of products.
For example, let's draw a marginal profit chart when fixed costs are 4.5 million yen, variable costs 3.6 million yen, and sales 9.0 million yen.

Figure 1-3-4
Marginal Profit Chart

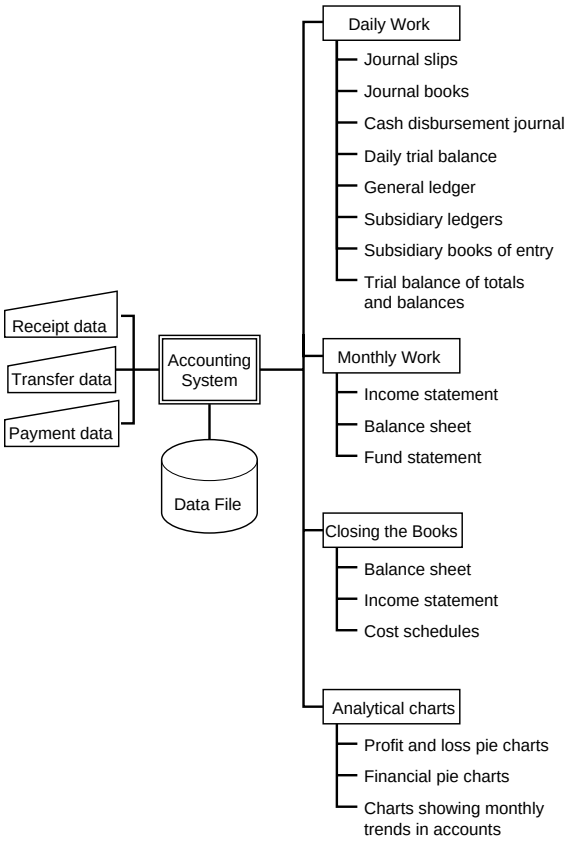


1. The horizontal axis represents sales, and the vertical axis represents marginal profit and expenses.
2. Plot fixed cost of 4.5 million yen on the vertical axis, and from that point, draw line A in parallel to the horizontal axis. Line A is a fixed cost line.
3. From the point of origin, draw gradient line B that represents the marginal profit ratio. Marginal profit can be calculated as follows: sales (9.0 million yen) - variable costs (3.6 million yen) = 5.4 million yen. Therefore, the marginal profit ratio is calculated as follows: 5.4 million yen ÷ 9.0 million yen (sales) = 0.6. Line B is a marginal profit line.
4. The break-even point is the point of intersection of fixed cost line A and marginal profit line B. The point of intersection of a line drawn from this point perpendicularly to the horizontal axis and the horizontal axis represents break-even sales (7.5 million yen).

1.3.3 Accounting Information System Configuration

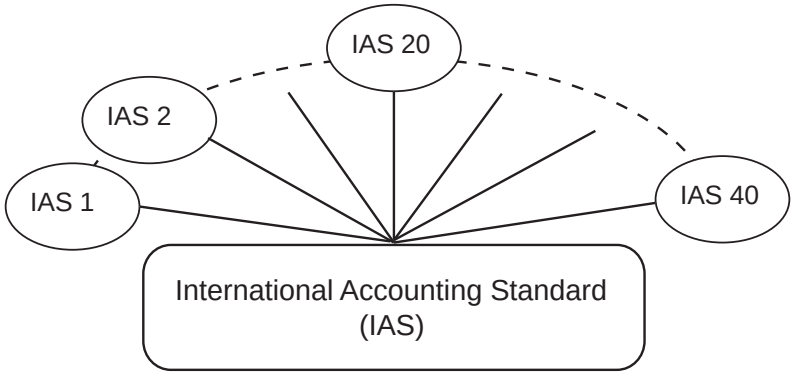
An accounting information system is a computer system for accurately and speedily performing business accounting as described above.
There have recently been demands for accounting information systems that will provide accurate information to external stakeholders, but that will also provide management accounting data.
A typical configuration of the accounting information system is shown in Figure 1-3-5.

Figure 1-3-5
Example of Accounting Information System Configuration



1.3.4 International standards

International Accounting Standard (IAS)
The major financial reporting standards used are the Generally Accepted Accounting Principles (GAAP) in the United States and IAS standards in Europe and other parts adopting the IAS standard.



Origins of IAS

In 1973, an agreement was established to set up the IASC (International Accounting Standards Committee) signed by representatives of the professional accountancy bodies in Australia, Canada, France, Germany, Japan, Mexico, Netherlands, United Kingdom/Ireland, and United States. IASB (IAS Board) opened an office in London.

Recently in 2002 the IASB Chairman and IASC Foundation Chairman testified at the US Senate hearing on Accounting and Investor Protection Issues Raised by Enron and other Public Companies.

Objectives of IAS

The objective of IAS is

(a) to develop, in the public interest, a single set of high quality, understandable and enforceable global accounting standards that require high quality, transparent and comparable information in financial statements and other financial reporting to help participants in the world's capital markets and other users make economic decisions

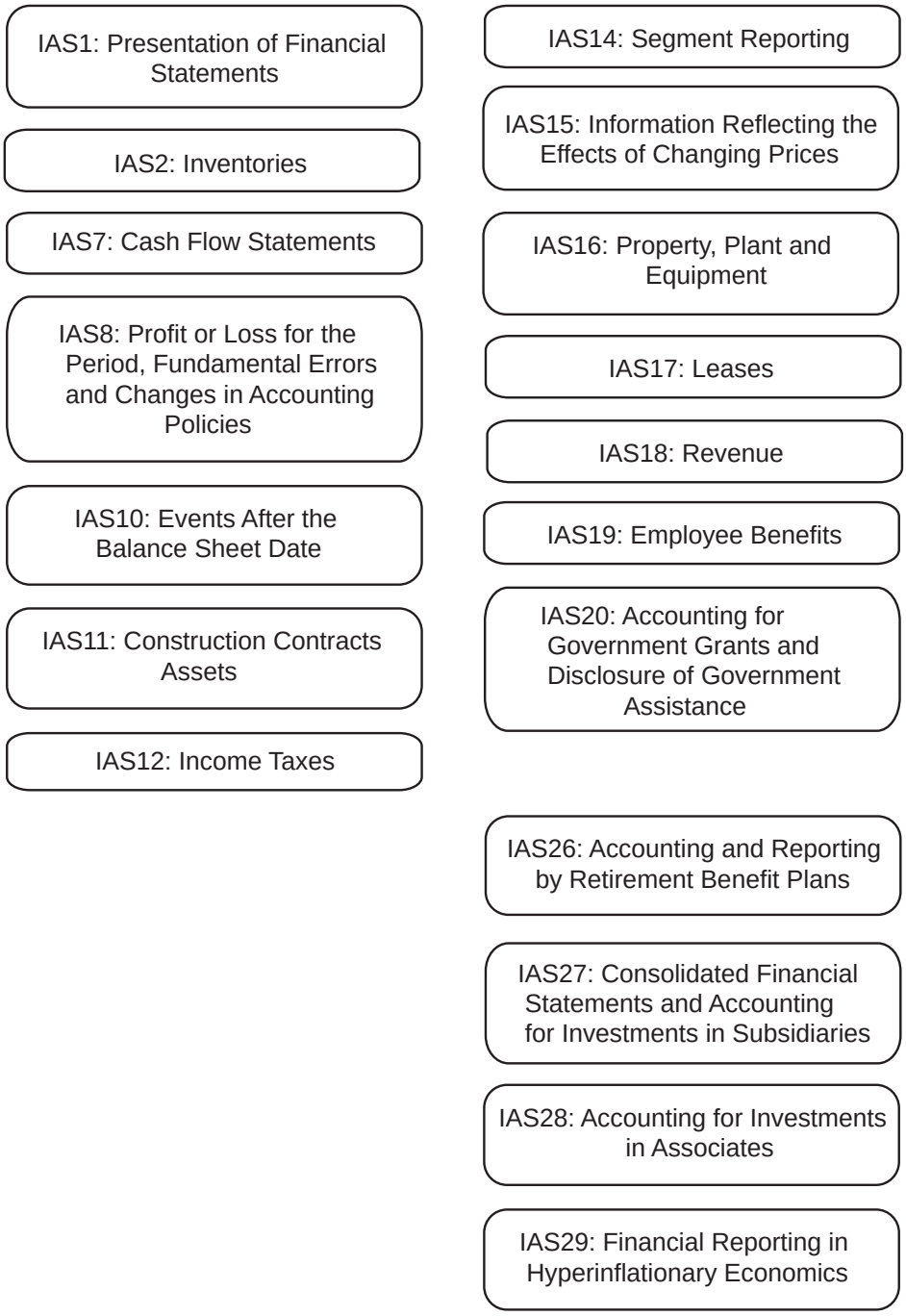
(b) to promote the use and rigorous application of those standards; and

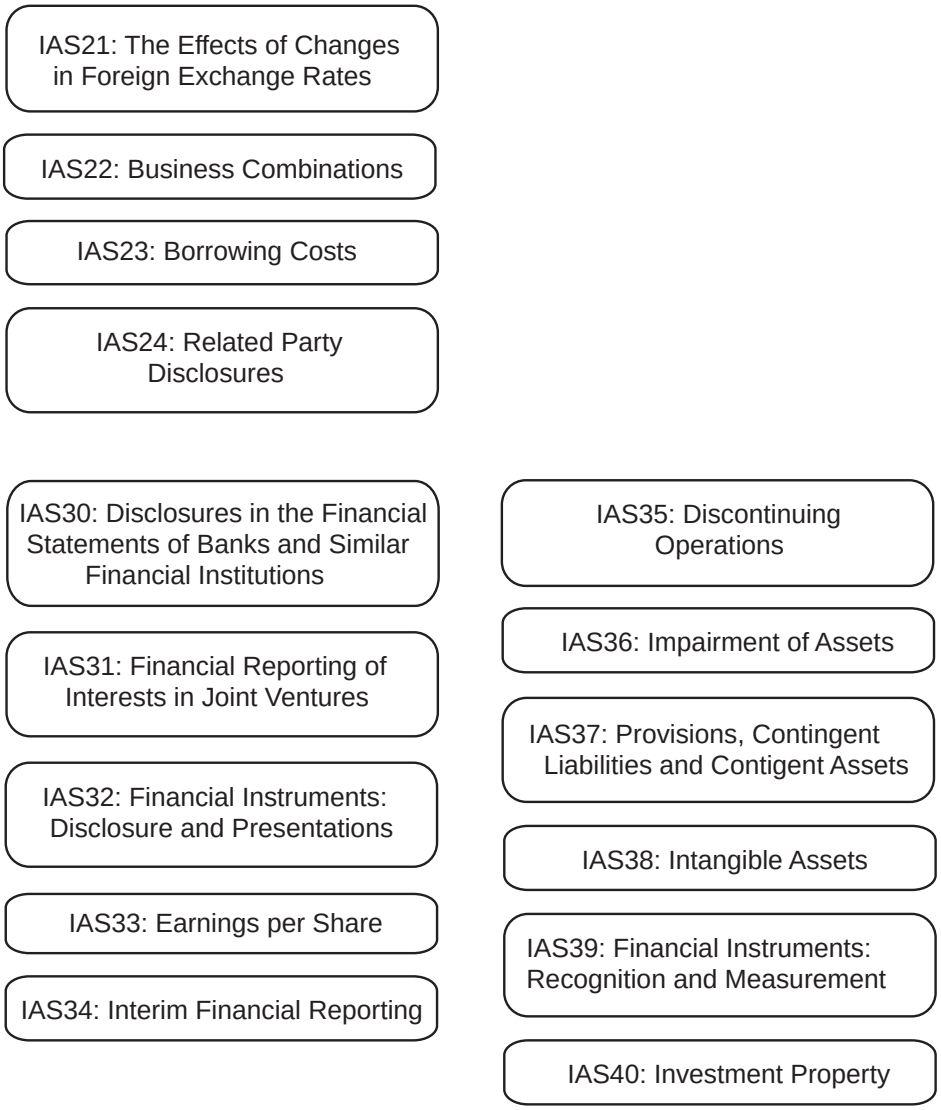
(c) to work actively with national standard-setters to bring about convergence of national accounting standards and IFRS to high quality solutions.

Countries utilizing the IAS standard

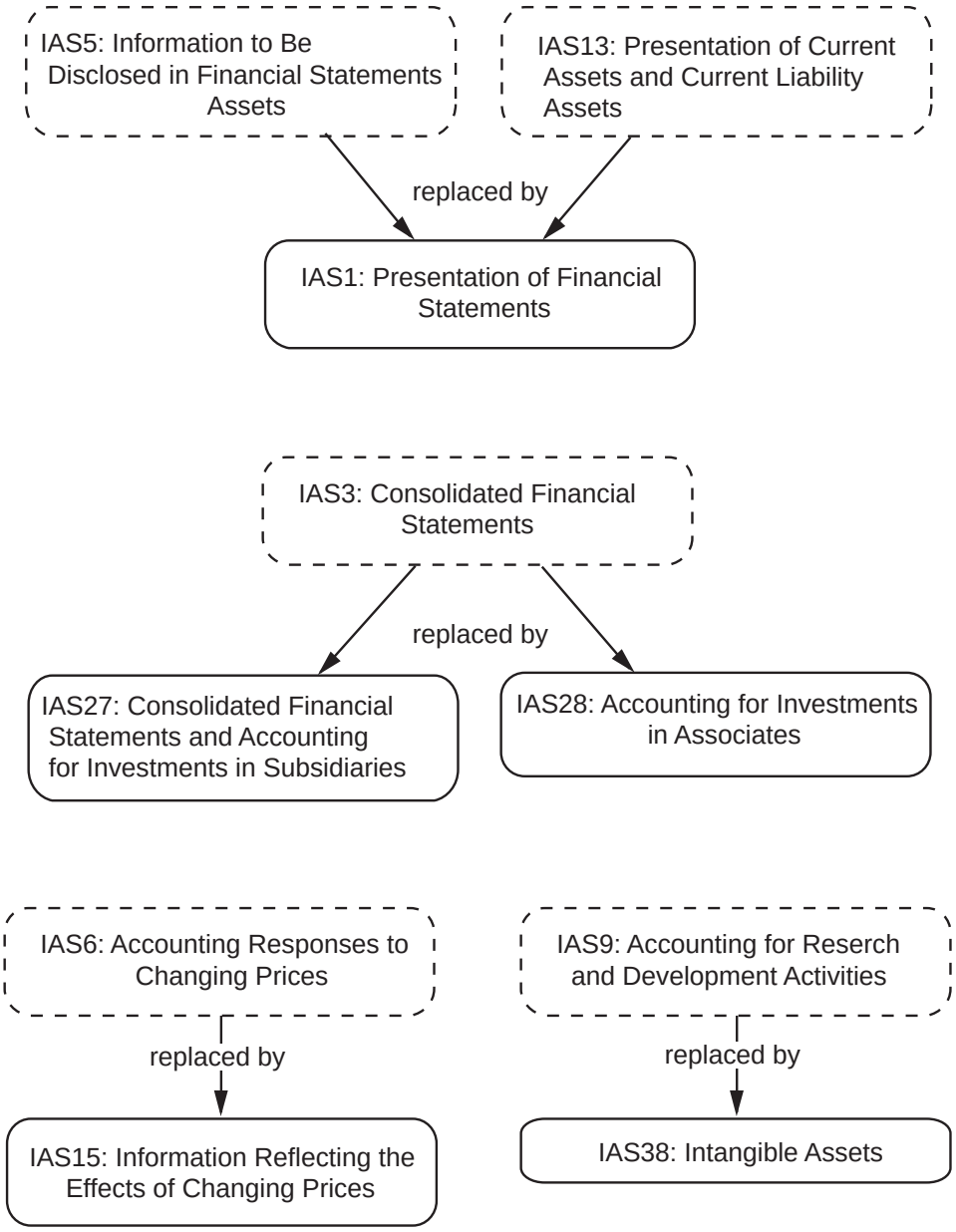
Australia	Malaysia
Austria	New Zealand
Brunei Darussalam	Pakistan
China	Philippines
Denmark	Portugal
France	Russia
Germany	Singapore
Greece	South Africa,
Hong Kong	Spain
India	Sweden
Indonesia	Sri Lanka
Israel	Taiwan
Italy	Thailand
Japan	United Kingdom
Korea	United States
Luxembourg	Vietnam

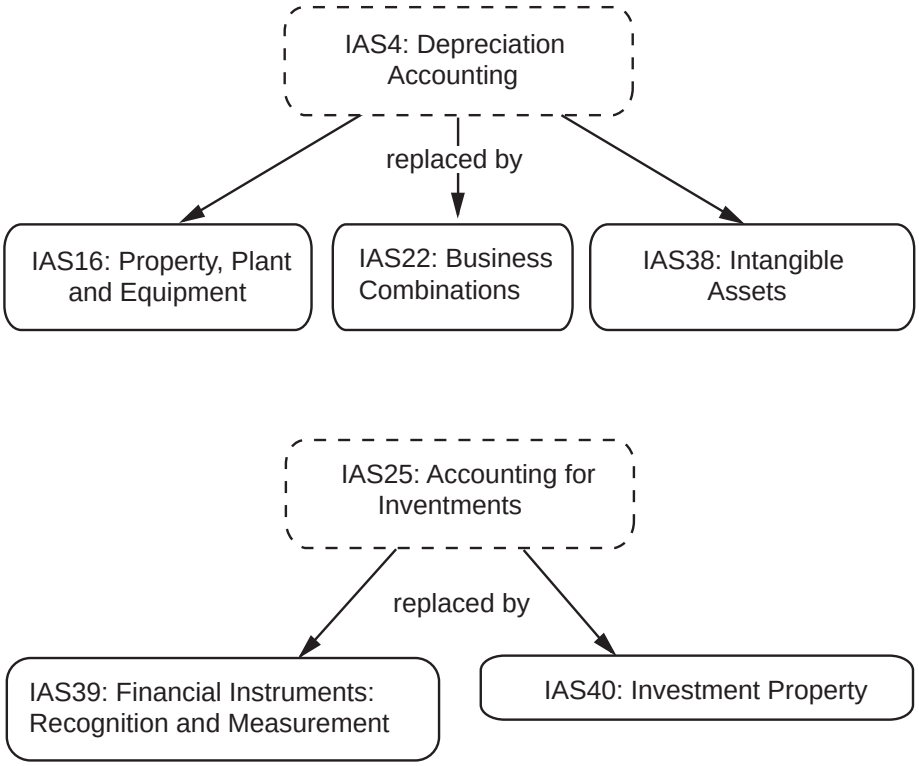
Sections in the IAS standard





The following standards in IAS has been superseded by the other IAS standards


IAS5: Information to Be Disclosed in Financial Statements Assets
IAS13: Presentation of Current Assets and Current Liability Assets
replaced by
IAS1: Presentation of Financial Statements
IAS3: Consolidated Financial Statements
replaced by
IAS27: Consolidated Financial Statements and Accounting for Investments in Subsidiaries
IAS28: Accounting for Investments in Associates
IAS6: Accounting Responses to Changing Prices
replaced by
IAS15: Information Reflecting the Effects of Changing Prices
IAS9: Accounting for Reserch and Development Activities
replaced by
IAS38: Intangible Assets



IAS 1 is the basic standard used for presenting the financial statement for the organization. The minimum requirements spelt out in this standard are shown below

IAS 1 Presentation of financial statement

Minimum items in balance sheet
property, plant and equipment
intangible assets
financial assets
equity method investments
inventories
receivables
cash and cash equivalents
payables
tax assets and liabilities
provisions
noncurrent interest-bearing liabilities
minority interest
issued capital and reserves

Minimum items on income statement
revenue
results of operating activities
financing costs
share of profits of equity method associates and joint ventures

income tax expense
profit or loss from ordinary activities
extraordinary items
minority interest
net profit or loss for the period
dividends per share

IAS 1 introduces a requirement to present a statement of changes in equity as a separate component of the financial statements, showing:

- the net profit or loss for the period;
- each item of income or expense, gain or loss which is recognized directly in equity and the total of those items; and
- the cumulative effect of prior period adjustments.

Either within this statement, or in a separate note, the enterprise is required to disclose:

- capital transactions with owners;
- the balance of accumulated profits at the beginning and at the end of the period, and the movements for the period; and
- a reconciliation between the carrying amount of each class of equity capital, share premium and each reserve at the beginning and at the end of the period, disclosing each movement.

The following other disclosures are required if not disclosed elsewhere in information published with the financial statements:

- domicile of the enterprise;
- country of incorporation;
- legal form;
- address of registered office or principal place of business;
- description of the enterprise's operations;
- name of its parent and the ultimate parent if it is part of a group; and
- number of employees - either end of period or average.

The Standard also specifies fundamental principles underlying the preparation of financial statements:

- the enterprise is a going concern (unless otherwise stated);
- financial statement presentation and classification are consistent with prior periods (unless otherwise stated);
- the accrual basis of accounting is used;
- materiality;
- timeliness: publish within six months of balance sheet date;
- disclosure if an International Accounting Standard has been applied before its effective date;
- the basis for selection of accounting policies and how they should be disclosed
- rules for the offsetting of assets and liabilities, and income and expenses; and
- a requirement for presenting comparative amounts.

Comparison between current accounting and IAS standard
Some of the major differences between the current accounting practice and IAS are outlined below.

	Current	**IAS**
General principles	Discounting is not currently considered	Consider the effect of discounting (present value).
Internally -generated intangible assets	All expenditure on research and development is recognized as an expense when it is incurred, capitalization is not allowed. Only registration fees and legal fees incurred for legal application of obtaining the asset can be capitalized.	• Expenditure on research is recognized as an expense when incurred. • Development costs are capitalized as intangible assets when certain criteria are met (such as technical feasibility, the availability of adequate resources to complete the development, and it is probable the intangible asset will generate future economic benefits).
Debt re-structuring	Gains arising from debt restructuring cannot be recognized as income, but recorded as capital surplus. Record any non-cash assets received at an amount equal to the carrying amount of the receivable to be restructured.	Gains arising from debt restructuring are recognized as income. Record the non-cash assets received at fair value, recognize the difference of the fair value and the carrying amount of the receivable to be restructured as the profits or losses.

	Current	IAS
Consolidated financial statements	• Consolidated financial statements should be prepared when an enterprise holds more than 50% of another enterprise's capital, or holds less than 50% of another enterprise's capital but controls the enterprise. • Unconsolidated subsidiaries include: - The parent company intends to dispose of its subsidiary in the near future (i.e. disposal intention does not require to be established at the time of listing). - The subsidiary operates under severe long-term restrictions that significantly impair its ability to transfer funds to the parent. - Total assets, sales revenue and profits of the subsidiaries do not exceed 10% of the corresponding amount of the group. (Not applicable for subsidiaries that incurred losses) - - Dissimilar activities, namely subsidiaries engaged in banking or insurance business.	• A parent (other than a parent that is a w holly owned subsidiary, or is virtually wholly owned and obtains the approval of the owners of the minority interest) should prepare consolidated financial statements. • A subsidiary is an enterprise that is controlled by another enterprise (known as the par ent). • Unconsolidated subsidiaries: - Control is intended to be temporary because the subsidiary is acquired and held exclusively with a view to its subsequent disposal in the near future. - Operate under severe long-term restrictions that significantly impair the ability to transfer funds to the parent.

Standardizing the electronic format of the financial statement

Many industries are now adopting a standard electronic representation format of the data by using XML. **XBRL** (e**X**tensible **B**usiness **R**eporting **L**anguage) is a royalty-free, open specification for software that uses XML data tags to describe financial information for public and private companies and other organizations. XBRL benefits all members of the financial information supply chain.

Origins

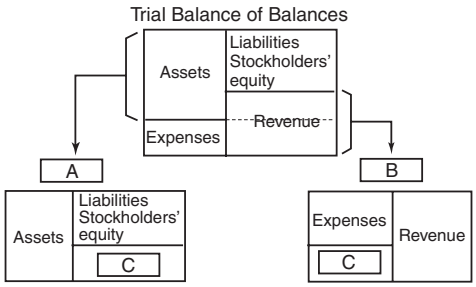
This was started in 1998 as an investigation into the use of XML for financial reporting.
The AICPA (American Institute of Certified Public Accountants) supported the initiative to set the standard. The steering committee was organized in August, 1999. 13 companies initially joined the effort (along with the AICPA) as members of the XFRML (XML based Financial Reporting) Steering Committee. The initial steering committee included: The AICPA, Arthur Andersen LLP, Deloitte & Touche LLP, e-content company, Ernst & Young LLP, FreeEDGAR.com, Inc. (now Edgar Online, Inc.), FRx Software Corporation, Great Plains, KPMG LLP, Microsoft Corporation,
PricewaterhouseCoopers LLP, the Woodbun Group and Cohen Computer Consulting.
XFRML was called XBRL.
Currently, there are 80 companies that are members of this organization.

Exercises

Q1 Which statement about financial statements is incorrect?

a.. P/L stands for the balance sheet, and B/S the income statement.
b. The balance sheet also shows the enterprise's net income.
c. Financial statements are prepared based on journal slips.
d. The income stated on the income statement includes operating income, ordinary income, and net income.
e. The balance sheet and the income statement are the most basic, most important financial statements.

Q2 Which is the appropriate combination of terms to put in the boxes in the following figure to show the relation between the balance sheet and the income statement?



	A	B	C
a	Income statement	Balance sheet	Costs
b	Income statement	Balance sheet	Net income
c	Balance sheet	Income statement	Costs
d	Balance sheet	Income statement	Net income

Q3 Which is the appropriate combination of terms to be put in the boxes in the following income statement?

Income Statement (From _____ through _____)

Sales	XXXX		
Cost of goods sold	XXXX	Gross income on sales	XXXX
Selling, general and administrative expenses	XXXX	A	XXXX
Non-operating revenue	XXXX		
Non-operating expenses	XXXX	B	XXXX
Extraordinary gains	XXXX		
Extraordinary loss	XXXX	Net income before taxes	XXXX
Corporation tax, etc.	XXXX	C	XXXX
Retained earnings from previous year	XXXX	D	XXXX

	A	B	C	D
a	Operating income	Ordinary income	Unappropriated retained earnings	Net income
b	Operating income	Ordinary income	Net income	Unappropriated retained earnings
c	Ordinary income	Operating income	Unappropriated retained earnings	Net income
d	Ordinary income	Operating income	Net income	Unappropriated retained earnings
e	Ordinary income	Net income	Operating income	Unappropriated retained earnings

Q4 **Shown below is an income statement for a fiscal year. How large is the operating income for the year?**

(Millions of yen)

Title of Account	Amount
Sales revenue	1,500
Cost of goods sold	1,000
Selling, general and administrative expenses	200
Non-operating income	40
Non-operating expenses	30

a. 270 b. 300 c. 310 d. 500

Q5 **Suppose that an income statement has been prepared from the manufacturing cost statement shown below. How large is the gross income on sales? (Amounts in thousands of yen)**

Manufacturing Cost Statement

Materials expenses	400
Labor expenses	300
Other expenses	200
Total manufacturing cost	
Beginning work in process inventory	150
Ending work in process inventory	250
Product manufacturing cost	

Income Statement

Sales revenue	1,000
Cost of goods sold	
Beginning product inventory	120
Product manufacturing cost	
Ending product inventory	70
(Cost of goods sold)	
Gross income on sales	

a. 150 b. 200 c. 310 d. 450

Q6 **Figure out gross income on sales from the following data. (Amounts in thousands of yen)**

[Manufacturing Cost Statement]		[Income Statement]	
1. Materials expects	500	1. Gross sales of products	2,000
2. Labor expenses	300	2. Sales returns	150
3. Other expenses	150	3. Sales discounts	150
4. Beginning work in process inventory	200	4. Beginning product inventory	600
5. Ending work in process inventory	300	5. Ending product inventory	400

a. 650 b. 750 c. 850 d. 950 e. 1,050

Q7 **When the analysis of financial statements at the end of a fiscal year produces figures ① to ③ shown below, what is the cost of goods sold? Note that the figures are all in thousands of yen and that the cost of goods sold ratio can be obtained by dividing cost of goods sold by sales revenue.**

① Ratio of cost of goods sold: 80%
② Ratio of operating profit to sales: 10%
③ Operating profit: 200

a. 1,200 b. 1,400 c. 1,600 d. 1,800 e. 2,000

Q8 **Of the following current assets, which is a liquid asset?**

a. Accounts receivable b. Work in process c. Short-term loan
d. Advance payment e. Non-trade accounts receivable

Q9 **Which is the equation for calculating the current ratio, which shows the degree of safety of short-term loans?**

a. $\frac{\text{current assets}}{\text{fixed assets}}$ b. $\frac{\text{current assets}}{\text{total assets}}$ c. $\frac{\text{current assets}}{\text{current liabilities}}$

d. $\frac{\text{current liabilities}}{\text{gross equity}}$ e. $\frac{\text{current liabilities}}{\text{total liabilities}}$

Q10 **Which is the correct statement about the break-even point?**

a. Where fixed costs remain unchanged, if the variable cost ratio rises, the break-even point lowers.
b. The break-even point means the level of sales at which the enterprise makes neither a profit nor a loss.
c. The break-even point indicates the degree to which assets are fixed.
d. Where the variable cost ratio remains unchanged, if fixed costs increase, the break-even point lowers.

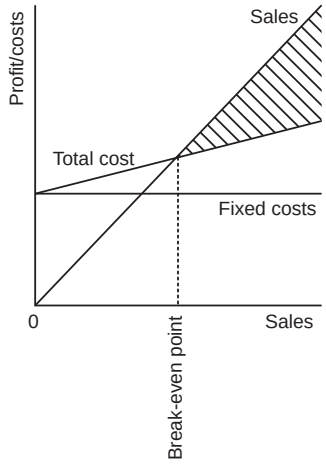
Q11 **Calculate break-even sales from the following income statement. (Amounts in thousands of yen)**

Table Income Statement

Title of Account	Amount
Sales	1,000
Variable costs	800
Fixed costs	100
Profit	100

a. 500 b. 600 c. 700 d. 800 e. 900

Q12 **In the following chart showing a break-even point, what is represented by the upper right area (the diagonally shared area above the break-even point) enclosed by the sales line and the total cost line?**



a. Operating loss b. Operating income c. Ordinary income d. Marginal profit

Q13 **There are goods whose unit purchase price is gradually rising. There was an inventory of these goods at the end of the last accounting period, and the goods were carried into and out of the warehouse several times during the current period. Which of the following valuation methods produces the highest valuation of the inventory at the end of the current period?**

a. Last-in first-out method
b. Moving average method
c. First-in first-out method
d. Average cost method

Q14 **When the first-in first-out method is applied to the receipt and delivery record shown below, what is the cost of goods sold for March?**

March	1.	Beginning inventory:	100 units	Unit price:	30 yen
	6.	Purchased:	50 units	Unit price:	50 yen
	10.	Sold:	50 units		
	17.	Purchased:	50 units	Unit price:	40 yen
	25.	Sold:	100 units		
	31.	Ending inventory:	50 units		

a. 4,000 b. 4,500 c. 5,000 d. 5,500 e. 6,500

Q15 **Shown below are the beginning inventory and purchases and sales during the current accounting period. When the inventory is evaluated by the last-in first-out method at the end of the current accounting period, how large is the inventory value?**

Purchases

Date	Volume (units)	Unit Price (yen)
Beginning inventory	10	100
May 1	15	90
Oct. 15	5	70

Sales

Date	Volume (units)
April 20	4
August 31	8
November 20	6

a. 840 b. 980 c. 1,080 d. 1,180

Q16 **Select from among the answers at the bottom the appropriate figures to be put in the boxes below based on the following statements regarding financial analysis:**

Store A's financial data for fiscal 2000 was as shown below.

<Data>

(1) The acid test ratio is $\frac{30{,}000{,}000\text{ yen}}{15{,}000{,}000\text{ yen}} \times 100\,(\%)$.

(2) The owner's equity ratio is $\frac{20{,}000{,}000\text{ yen}}{45{,}000{,}000\text{ yen}} \times 100\,(\%)$.

(3) The fixed ratio is 60%.
(4) The current ratio is 220%.
(5) The debt ratio is 125%.

Balance Sheet

(Thousands of yen)

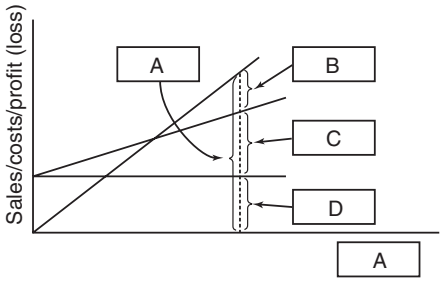
Assets	Amount	Liabilities & Stockholders' Equity	Amount
Liquid assets	A	Current liabilities	D
Inventories	B	Fixed liabilities	E
Fixed assets	C	Stockholders' equity	F
Total Assets	G	Total Liabilities & Stockholders' Equity	G

Answers

a. 3,000	b. 10,000	c. 12,000	d. 15,000
e. 18,000	f. 20,000	g. 25,000	h. 30,000
i. 45,000	j. 48,000		

Q17 Select from among the answers at the bottom the appropriate terms to be put in the boxes in the statements regarding break-even analysis.

(1) Select from among the answers the appropriate terms describing parts A through D of the break-even chart.



Break-Even Chart

(2) The amounts for A through D for an accounting period are as follows:

A. 10,000,000 yen
B. 2,000,000 yen
C. 6,000,000 yen
D. 2,000,000 yen

In this case, break-even sales are [E] yen.
If the amount for A becomes 20,000,000 yen, the amount for B is [F] yen.

Answers for A through D:

a. Selling, general and administrative expenses		b. Sales	c. Accounts receivable	
d. Fixed costs	e. Manufacturing cost	f. Loss	g. Variable costs	h. Profit

Answers for E and F:

a. 3,000,000	b. 4,000,000	c. 5,000,000	d. 6,000,000
e. 7,000,000	f. 8,000,000		

2 Application Fields of Computer Systems

Chapter Objectives

The reader is expected to learn some cases of computer applications in engineering fields such as CAD/CAM and in business system fields such as POS systems and EOS. The objectives of this chapter are as follows:

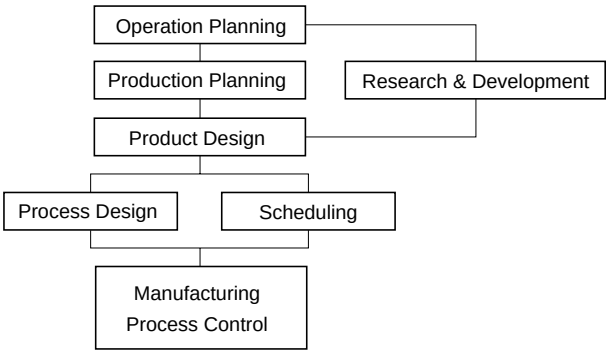
① Understanding the outlines of the mechanisms and functions of automatic production control, CAD/CAM/CAE, and factory automation (FA) systems.
② Acquiring knowledge about the computer applications in the business sector.
③ Acquiring knowledge about the computer applications in business-to-business commerce, such as EDI, CALS, and EC, and about business.

2.1 Engineering Applications

2.1.1 Automatic Control of Production

Human beings initially conducted production activities using their own power and simple tools. In the Industrial Revolution started in the late 18th century, tools dramatically evolved into machines, and human power into steam engines and electric power. Since the end of World War II, we have been seeing advances in computer and microelectronics technologies. Computers and microelectronic devices have automated production and complex control processes that used to be performed manually.
The production process flow is generally as shown below.

Figure 2-1-1
Production Process



(Source: "Class II Common Curriculums" edited by the Central Academy of Information Technology, Japan Information Processing Development Corporation)

The factors in the demand for production process automation include the following:

- Decreasing labor supply
- Mechanical substitution of dangerous and extreme human operations
- Cost reduction to deal with intensifying market competition
- Need to produce larger varieties of products in smaller quantities
- Advances in computer technology

Here are some examples of typical engineering systems:

- The direct numerical control (DNC) system in which a single computer controls multiple NC machine tools
- The automatic monitoring system that monitors machine tools and responds to any abnormality in real time
- The automatic warehousing system in which computers control a warehouse by operating robots, cranes, and so on
- The CAD (Computer Aided Design) system and the CAM (Computer Aided Manufacturing) system to design and manufacture products with computers
- The CAE system to help design and drafting on the computer display
- The office automation (OA) system to save labor in the clerical work
- The factory automation (FA) system to save labor in the manufacturing process in factories

2.1.2 CAD/CAM/CAE

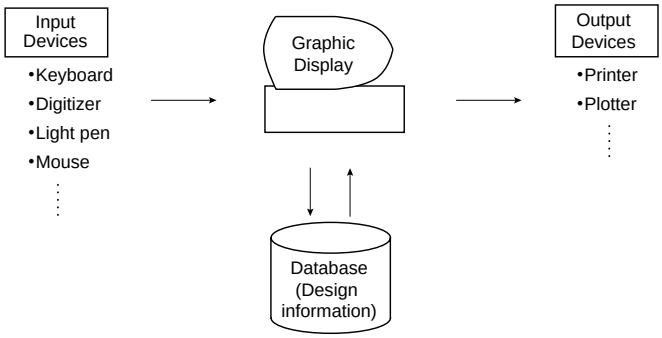
(1) CAD

CAD is an acronym for Computer Aided Design.
The objective of CAD is to automate product design as much as possible by using computers. To achieve this objective, CAD generally uses graphic display devices, digitizers, and other tools. CAD is the process of designing a product through dialog on a computer (EWS: engineering workstation) display using the stored design information. CAD requires substantially less time than manual design.
As peripherals, CAD uses various input and output devices:

- Input devices: keyboard, digitizer, light pen, and mouse
- Output devices: printer and plotter

In addition, CAD requires graphic display devices and a database to store and retrieve design information.
Graphic display devices are available in CRTs (Cathode Ray Tube) and in flat panel displays.

Figure 2-1-2
CAD Configuration



In CAD, actual assistance from the software is provided in the following steps:

- The computer provides accumulated design information and data and retrieves pertinent information and reference data.
- The computer provides geometric models to help to represent the object to be designed.
- The computer automatically creates part of a design proposal through a pre-arranged procedure.
- The computer examines and evaluates a design proposal by simulation and other procedures.

CAD is used for the design of electronic circuits, buildings, automobiles, and so forth. Combined with CAM, CAD is often referred to as CAD/CAM.

(2) CAM

CAM is an acronym for Computer Aided Manufacturing.
CAM means assisting the manufacturing products by using computers. It is the process of designing a manufacturing process based on the data in design drawings prepared in CAD and automatically controlling the assembly and processing of the product using numerically controlled (NC) machine tools and the like.
CAM makes it possible to automate machining and other operations with computers, thus saving labor.
CAM is based on the combination of several technologies, including the following:

- FMS (Flexible Manufacturing System)
- Computer aided process planning (process design)
- Computer aided scheduling
- Industrial robotics technology

(3) CAE

The objective of CAE (Computer Aided Engineering) is to reduce the time required for the development of a new industrial product with shorter prototyping and experimentation period. CAE helps to study characteristics of the product and its components through simulation and numerical analysis by such

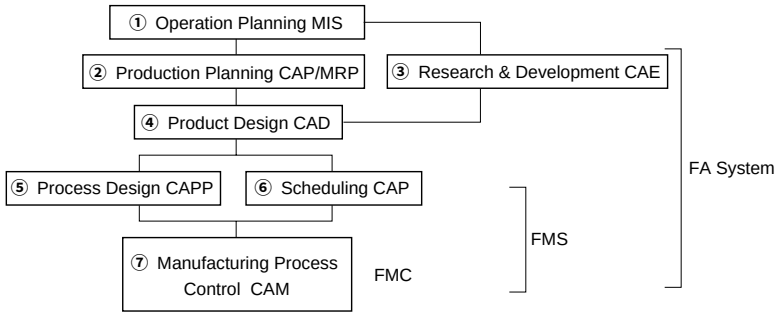
methods as the finite element method (FEM) using a computer.
The term CAE is used in the broad sense and the narrow sense. In its broad sense, CAE means the process of assisting in the stages from simulation to development, design, and drafting with the computer. In its narrow sense, CAE means only analytical work, leaving the subsequent work of design and drafting to CAD. In both cases, CAE does not include CAM in the manufacturing stage.

2.1.3 FA Systems and CIM

(1) FA Systems

The FA system is a system to perform work in a factory efficiently by automating the work as much as possible with the computer. It covers a very wide range of operations, from CAD/CAM to numerically controlled machine tools and robots.
The basic FA system configuration is shown in Figure 2-1-3.

Figure 2-1-3
FA System



(Source: "Class II Common Curriculums" (with some subsequent additions) edited by the Central Academy of Information Technology, Japan Information Processing Development Corporation)

① Operation planning

Management Information System (MIS): A management decision making system

② Production planning

Material Requirement Planning (MRP): A system for controlling the flow of materials, from raw materials to finished products, over time

③ Research and development

Computer Aided Engineering (CAE): A system for outline design based on the simulation of the strength and mechanism analysis of products

④ Product design

Computer Aided Design (CAD): A system for detailed design, including the creation of geometric models of products

⑤ Process design

Computer Aided Process Planning (CAPP): A system for determining work procedures, working machines, work time, and so forth

⑥ Scheduling

Computer Aided Planning (CAP): A system for determining work schedules, machine assignment schedules, and so forth

⑦ Manufacturing process control

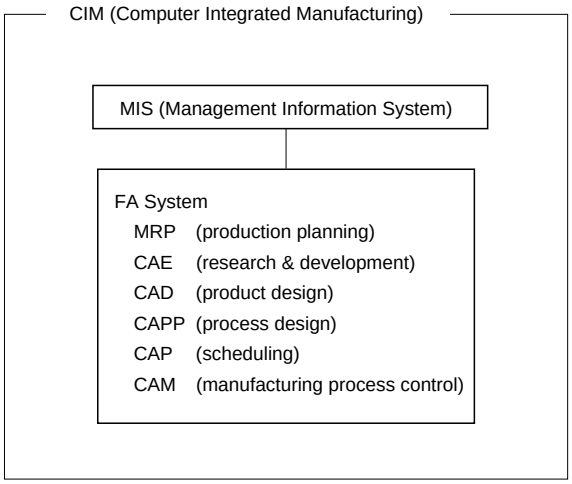
Computer Aided Manufacturing (CAM): A system for controlling working machines' manufacturing processes by computer

Of these systems, ② through ⑦ are collectively called an "FA system."
In addition, systems ⑤ through ⑦ are collectively referred to as "FMS" (Flexible Manufacturing System). The FMS links the automatic control of flexible manufacturing cells (FMCs), which are the units of processing and assembly in manufacturing, to an automated warehouse and automatic transport equipment for integrated control by computer. The FMS is drawing attention as an automatic production system to deal with the need to produce large varieties of products in small quantities.

(2) CIM

When an FA system is operated based on a management information system (MIS), the entire information system is called a CIM (Computer Integrated Manufacturing) system (Figure 2-1-4).
That is, CIM is built as an integrated corporate information system covering all stages from management strategies to production in order to improve the efficiency of all the operations of the enterprise.

Figure 2-1-4
CIM



In this case, the FA system is a subsystem of CIM. Other subsystems of CIM include an OA system, a POS system, and an EOS (electronic ordering system).

2.2 Business Applications

2.2.1 Head Quarters Business Support Systems

Some typical business support systems are described below.

(1) Accounting Information System

As mentioned in Chapter 2, an accounting information system processes corporation accounting with the computer. It digitizes slip data entries, automates accounting calculation, and prepares financial statements. This system can not only ensure speedy accounting but also reduce processing errors and personnel expenses.

(2) OA System

The OA system is aimed at improving the efficiency of clerical processing by introducing office automation (OA) equipment into the office space.
The functions required of an OA system vary depending on the line of the business, operation, division, and user.
To serve a variety of purposes, several systems are available as shown below:

- A document processing system based on word processors
- An electronic filing system aimed at minimizing the use of paper
- A videoconferencing system that reduces travelling time by taking advantage of advances in communications bandwidth
- An electronic secretary system that assists in secretarial work
- A decision support system that assists in the decision makings in business operation

It is now becoming increasingly common to build an office system by combining such systems as mentioned above or to construct intelligent buildings in which OA systems are accessible everywhere.

(3) Groupware

① Outline of groupware

Groupware is designed to enable members of a group to collaborate to each other by linking the members' personal computers via a LAN (local-area network) or WAN (wide-area network). For example, when a project is to be carried out, it is a common practice to form a team to work on it. Groupware is used to manage the roles and schedules of the individual group members and to ensure smooth execution of the project.

② How to use groupware

Groupware is still an emerging field, and there aren't many items available for practical use. The functions usually provided by groupware at present are electronic mail, bulletin board, and videoconference. Other groupware capabilities include scheduling and work flow management, and online approval/dismissal of business trips and expense proposals.

a. Electronic mail

Electronic mail is a digital version of traditional postal mail that is transferred by personal computers, in-house LANs, and the Internet. A computer center has a mail box for each user, so that you can write messages to other users and read messages addressed to you. Email can be sent between users anywhere in the world, in seconds.

For businesses, however, electronic mail is more than a substitute for letters and facsimiles. That is, since even field workers can send electronic mails directly to top management, it becomes possible for people at all levels to share information. Electronic mail thus ensures more efficient execution of collaborative work.

b. Bulletin board

While electronic mail is a means of communication between individuals, the electronic bulletin board is a broadcast system that enables multiple people to read and write freely. Some bulletin boards functions as the circulation of information and the reservation of conference rooms. The bulletin board system should be used effectively to ensure better communication among the members working on the same project.

c. Videoconferencing

Videoconferencing enables participants to have discussions and exchange opinions over the network. Videoconferencing is similar to the bulletin board but enables the registered members only. Unlike ordinary conferences, videoconference does not require participants to exchange opinions at the same time. This is one of the major advantages of videoconferencing.

d. Schedule management software

It is important but often difficult to manage schedules. When a large number of people are involved, it is quite a job to coordinate their schedules.
Schedule management software is used to manage schedules by shared calendars.
This software coordinates the work schedules of the members of a project team by recording their schedules and negotiating and reflecting changes over the network.

e. Bottom-up decision-making system

This is a system to perform "ringi," the unique Japanese method of decision making, by computer. In the conventional ringi system, a document stating a proposal and seeking approval is routed by the author to the section head, the department head, the director in charge, and the president. In the bottom-up decision-making system, similar document data is transmitted in this order and the managers signify their approval by using their electronic seals. Since the system makes it unnecessary for the author to physically take a proposal to the managers, it improves work efficiency and helps reduce the use of paper.

2.2.2 Retail Business Support Systems

It is vital for information processing engineers to properly understand what systems are required in particular situations of business activity. To this end, it is necessary for them to be able to build systems by accurately grasping the flow of information in various lines of business, detecting problems, and studying solutions. Such systems required in various situations of business activity are called business support systems in the broad sense.
The retail industries are also seeing the building of systems to ensure efficient operations. Such systems are called "retail information systems."
Typical retail information systems include the following:

- POS (point-of-sale) system
- EOS (electronic ordering system)
- Inventory management system

Since inventory management has already been learned in Chapter 3, this section describes the POS system and the EOS.

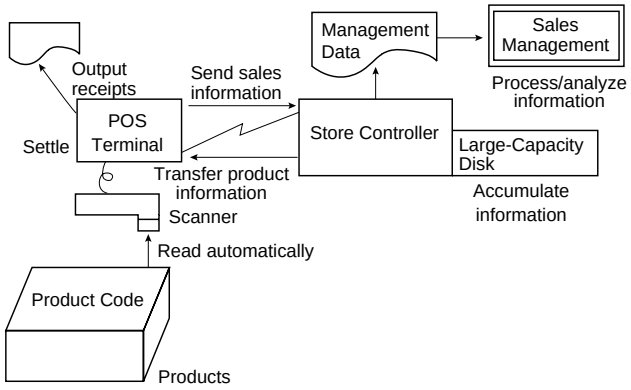
(1) POS System

① What is the POS system?

The POS system is a system to manage information at the point of sale. POS is an acronym for Point Of

Sale.
In the retail business, in which goods are purchased and sold, merchandise management is an important job. The POS system makes it always possible to grasp the number of goods sold in order level of not to run out of stock and to avoid storing excessive inventory. The system helps accurately grasp the level of inventory and the time to place orders.
In the POS system, a bar code reader linked to a cash register determines product names and prices by identifying the bar codes on the products customers are buying.
In addition to bar code readers, auxiliary computing and storage devices to indicate product prices and inventory levels are installed in convenience stores and supermarkets. These devices send data to computers located for the headquarters.

Figure 2-2-1
POS System



(Source: "Class II Common Curriculums" edited by the Central Academy of Information Technology, Japan Information Processing Development Corporation)

The POS system will provide the following benefits:
- Simpler and more accurate checkout by cashiers
- Automatic accumulation of sales data
- Proper merchandise offerings
- Less time required for sales staff training

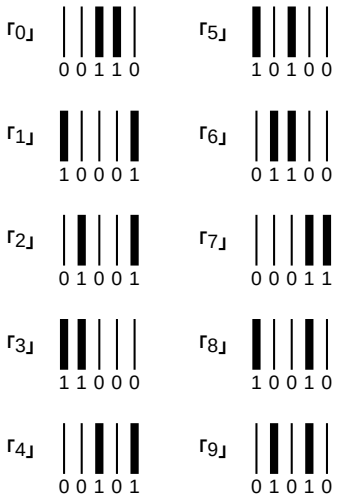
② Bar codes

A bar code represents characters with a combination of parallel vertical lines and spaces of varying thickness. A bar code can be optically scanned.
Today bar codes are attached to most foodstuffs and daily necessities and are playing an important role as point-of-sale input information. Bar codes are also used in libraries for the management of books.
A patent on bar codes was obtained in 1949 but bar codes were not widely used until around 1970 because devices to read these codes were astronomically expensive. Today bar code readers are inexpensive, and bar codes are in wide use.

a. Structure of bar codes

As noted, the bar code represents information by a combination of wide and narrow vertical bars.
In the case of the 2 of 5 system, each of five bars corresponds to one binary, a narrow bar representing a "0" and a wide bar a "1." One character is coded by five bars, on which two are always wide (Figure 4-2-2). Therefore, even if a wide bar is mistakenly printed or read as a narrow bar or vice versa, the resulting absence of two wide bars makes it possible to detect the data error. And yet, if in a set of five bars, one wide bar is mistakenly read as a narrow one and one narrow bar as a wide bar, then the data error cannot be detected.

Figure 2-2-2
Bar Codes



b. Bar code reader

Bar code readers are of:

- the pen type,
- the touch type, or
- the laser type.

The pen type bar code reader reads bar codes by scanning them with an LED (light emitting diode). The touch type reader can read bar codes by simply applying an LED to them. The laser type reader can read bar codes at some distance.

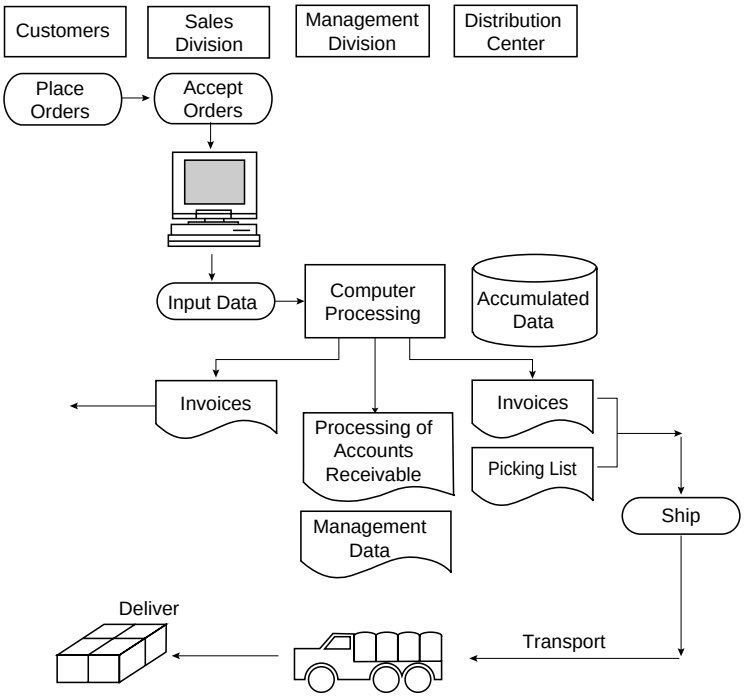
(2) EOS

EOS is an acronym for Electronic Ordering System.
In any line of business, accepting orders and placing orders are essential parts of business activity. These jobs require a great deal of time and manpower, and errors tend to occur. These jobs are important, since they are directly related to inventory management.
These jobs often involve the following problems:

- It takes time to receive deliveries after placing orders.
- It takes time to check inventory levels.
- It is possible to miss ordering necessary goods.
- It is necessary to have some knowledge about product inspection.

The EOS solves these problems by automatically managing these ordering tasks by computer.
The EOS receives such data as the product codes of the merchandise to be ordered, their quantities, suppliers or business partners through terminals and sends the data to the pertinent departments at headquarters, or suppliers or shippers in order to handle ordering tasks.

Figure 2-2-3
EOS
(Electronic
Ordering System)



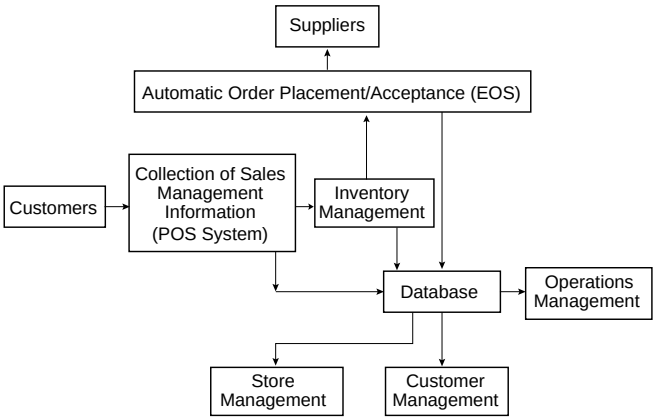
(Source: "Class II Common Curriculums" edited by the Central Academy of Information Technology, Japan Information Processing Development Corporation)

The EOS would bring the following benefits:

- Simpler inventory management
- Labor saving in product inspection
- Cost reduction in distribution
- Accurate merchandise management

When linked to the POS system, the EOS can be more effective from the viewpoint of merchandise management (Figure 2-2-4).

Figure 2-2-4
Sales and Distribution
Information System



(Source: "Class II Common Curriculums" edited by the Central Academy of Information Technology, Japan Information Processing Development Corporation)

To implement EOS, the following matters need to be agreed by business partners:

- Standardization of order acceptance and placement procedures
- Systematization of codes, including merchandise and supplier codes
- Protocol compatibility

2.2.3 Financial Systems

(1) What Are Financial Systems?

The first online system implemented in Japan is the score keeping system used in the Tokyo Olympics in 1964. The system was taken over by the banking industry the next year. The financial industry thus became the first user of a commercial online system in Japan.
Since then, the financial system has gone through the first-, second-, and third-generation online projects. Today it is providing a variety of services as a social system indispensable for people's lives.

(2) Banking Systems

A typical banking system consists of operational, clerical, and informational subsystems. Furthermore, the subsystems have the following subsystems:

- Operational subsystems: Accounting
  Fund and securities
  International exchange
  External connection, etc.
- Clerical subsystem: Retail banking
  Call and customer support center
- Informational subsystem: Management information

These functions have evolved through the processes described below.

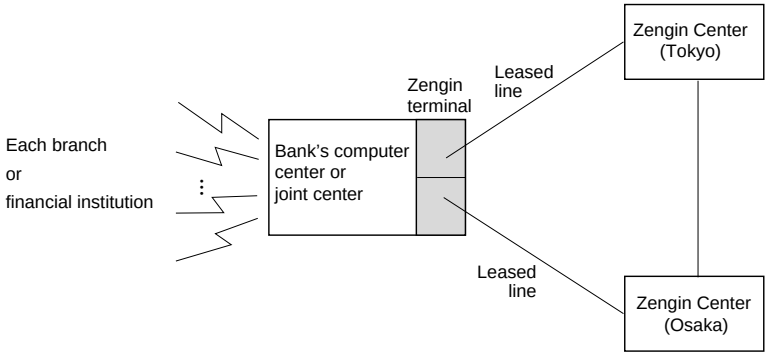
① First online project

Banks began to use computers for their operations in the 1960s. The systems they worked on were called the first online systems. To improve clerical processing efficiency and save labor, these systems made use of online storage of ledgers and the centrally controlled automatic fund transfers.

② Second online project

In the 1970s, banks began to work on the second online project in order to strengthen the system's functionalities. This project realized the interbank linkage of online cash dispensers, the linked processing of major accounts, and so forth. Furthermore, in 1973, banks put into operation the "Zengin Data Telecommunication System," a data telecommunication system of all banks in Japan. The Zengin System has Zengin centers in Tokyo and Osaka, which are linked to each financial institution's computer center through a Zengin terminal using leased lines. This system is functioning as the core of a domestic settlement system by performing such functions as sending and receiving messages on domestic exchange transactions and calculating exchange settlement amounts.

Figure 2-2-5
Zengin System



Internationally, SWIFT (Society for Worldwide Interbank Financial Telecommunication) was founded in 1973. SWIFT II is now in operation. This system is an international network of financial institutions. It handles communications on interbank transfers, customer remittances, and so forth concerning international financial transactions. Unlike the Zengin System, SWIFT II does not process interbank fund settlement.

③ Third online project

In the 1980s, in order to strengthen their information management functions, and customer networks to deal with financial deregulation, banks reconstructed their account systems and enhanced their information, international, securities, exchange and external connectivity.
In 1988 the Bank of Japan Financial Network System (BOJ-Net) went into operation. Handling foreign exchange settlements by yen, government bond operations, and other business, this system is contributing to efficient, speedy clerical processing in the entire banking industry.
In addition, the MICS nationwide cash dispensing service that started in 1990 made it possible for depositors to get cash through dispensers installed at any type of financial institutions. Previously, depositors had been able to get cash only through dispensers installed at financial institutions belonging to particular syndication of banks such as city banks, regional banks, and local credit unions.

(3) Electronic Banking

Electronic banking uses systems that electronically exchange data over the network connecting financial institutions' computers with corporate and individual customers' computers and terminals.
By the objects to be networked, electronic banking systems can be divided into:

- firm banking, which uses a system networking financial institutions and businesses, and
- home banking, which uses a system networking financial institutions and individuals.

Major types of processing performed in electronic banking include deposit balance inquiries, depositing and withdrawal operations through cash dispensers and automatic teller machines (CDs and ATMs), and account transfer transactions. Recently, electronic banking has been able to handle such operations as providing foreign exchange rates, payroll calculations, account transfer reservations, settlement of accounts payable, and so forth. Advising and inquiry services using touch-tone telephones began in 1981. After firm banking, banks kicked off home banking in which consumers can deal with their banks through personal computers, word processors, and game machines.

① Firm banking

Firm banking is a system that enables businesses to perform real-time transactions with their banks, such as inquiring about deposit balances, making deposits, and performing account transfers.
Firm banking also allows businesses to send such data as account transfers and salary payments directly to their financial institutions. Some city banks have realized firm banking utilizing personal computers.

② Home banking

Home banking is a social system that enables consumers to perform such transactions as checking the balances of their bank accounts and transferring funds from their accounts to others' through personal computers or game machines at home.

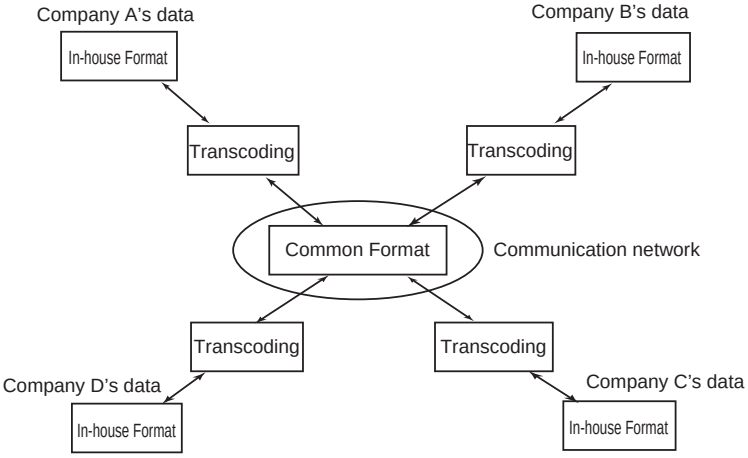
Banks are considering realizing general home shopping in the future by connecting homes, department stores, and banks so that consumers can do catalog shopping through home terminals and pay for the purchases through home banking.

2.2.4 Inter-Enterprise Transaction Data Interchange

(1) EDI (Electronic Data Interchange)

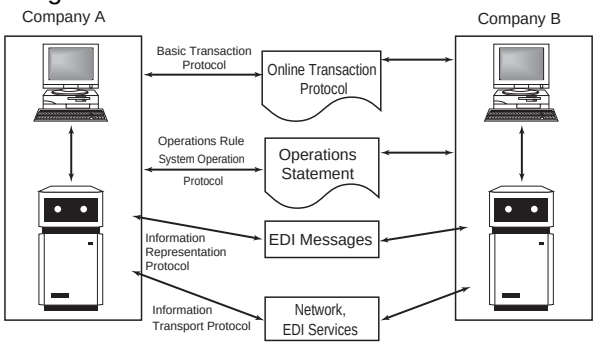
EDI (Electronic Data Interchange) means the process of digitizing order forms, quotations, and other information on business-to-business transactions and exchanging the digital data over a network. Electronic business-to-business transactions are realized by standardizing data exchange protocols and data formats. In Japan, EDI has progressed in the form of regional VAN and industry VAN.

Figure 2-2-6 Data Format Standardization



(Prepared from "Electronic Data Interchange Subcommittee Report," Computer Interoperation Environment Establishment Committee, the Ministry of Economy, Trade and Industry [former Ministry of International Trade and Industry])

Figure 2-2-7 Data Exchange Protocol Standardization



(Prepared from "Electronic Data Interchange Subcommittee Report," Computer Interoperation Environment Establishment Committee, the Ministry of Economy, Trade and Industry [former Ministry of International Trade and Industry])

The standard EDI protocol adopted by the United States and Europe in 1988 is called UN/EDIFACT (United Nations/Electronic Data Interchange For Administration, Commerce and Transport).
Web-EDI over the Internet has recently been widely in use.

① Benefits

EDI has the following benefits:
- Transaction cost reduction
- Manufacturing and sales cost reductions
- Relaxing time and physical restrictions on transactions

② Challenging issues

The wide acceptance of EDI may cause the following problems:
- Widening difference of information processing ability among businesses
- Decline in the competitive position of non-EDI companies in the industry
- Heavy burden on weak companies
- Data format confusion

(2) CALS (Commerce At Light Speed)

CALS is a system in which information on the product life cycle ranging from procurement through design, development, production, operation management to maintenance is managed digitally in an integrated manner to support the individual processes.
The origin of CALS is the concept "Computer aided Acquisition and Logistics Support" developed by the U.S. Department of Defense for materials procurement support systems.
While EDI primarily handles the interchange of transaction-related document data, CALS is a system to share information mostly concerning products.

① Benefits

CALS provides the following benefits:
- Sharing all information on the product life cycles in digital data
- Improving the quality of operations and products, and raising productivity
- Reducing costs in the entire life cycle

② Relations with EDI

The differences between CALS and EDI mentioned above include the following:
- CALS extends EDI to operations other than order acceptance and placement.
- CALS reduces costs throughout the life cycle of whole activity.

(3) EC (Electronic Commerce)

EC (Electronic Commerce) is defined by the Electronic Commerce Promotion Council of Japan (ECOM) as any commercial transaction part or all of which is performed over a network.
EC provides a mechanism in which individuals, businesses, governments, or organizations perform a series of activities such as selling, distribution, advertising, settlement, and various services.
EC is subdivided into:
- commerce between businesses (business to business: B-to-B)
- commerce between businesses and consumers (business to consumer: B-to-C)

① B to B (business to business)

B-to-B commerce can be further subdivided into commerce between specific businesses and commerce between any to any businesses.
EC between specific businesses is performed through an industry-specific system using CALS and EDI mentioned above. On the other hand, EC between any to any businesses is performed as EDI in an open network environment.
More specifically, B-to-B commerce is largely performed as inter-business transactions in such industries as manufacturing and wholesale. In these industries, B-to-B commerce is often aimed at cultivating new sales channels, increasing revenue, carrying out effective marketing, or improving customer relationship.

② B to C (business to consumer)

B-to-C commerce is the form of transactions in which businesses in the retail and service industries directly provide products and services for consumers. By eliminating an intermediate stage (middlemen)

in the complex process of product distribution, B-to-C commerce enables substantial reductions in costs and time.
In addition, B-to-C commerce allows transactions without having actual inventory, thus enabling businesses to reduce inventory burden. Furthermore, businesses can conduct operations 24 hours on 365 days a year, with an adequate customer support.

③ C to C (consumer to consumer)

C-to-C commerce is a special form of transactions in which products are bought and sold between consumers. In C-to-C commerce, consumers perform transactions and settle the prices between themselves, and the company providing the arena receives commissions from sellers. A typical example of C-to-C commerce is Internet auctions.

Meanwhile, B to B, B to C, and C to C may sometimes be written as B2B, B2C, and C2C, respectively.

Exercises

Q1. **Which is the correct statement about FA (Factory Automation) systems?**

a. An FA system performs geometric modeling of products using CAD.
b. CAM is an acronym for Computer Aided Modeling.
c. FA systems have nothing to do with CIM.
d. FMS is a subsystem of FMC.
e. The system to calculate the quantities of resources required for production is called MAP.

Q2 **For designing industrial products and building construction and for making industrial designs, which of the following technologies improves drawing and designing efficiency by using computers?**

a. CAD b. CAI c. CAM d. CIM e. GUI

Q3 **Which of the following systems displays objects using wire-frame models, surface models, and the like for efficient design work?**

a. CAD b. FA c. FMS d. MAP e. POP

Q4 **Which of the following systems constitutes part of an FA system and performs design and drafting interactively and automatically by using computers, graphic displays, computer aided drafting machines, and so on?**

a. CAD b. CAE c. CAM d. CAT

Q5 **Which of the following systems calculates necessary quantities of materials from basic production plans or manages manufacturing schedules by using parts configuration, inventory, and other files?**

a. CAD b. FA c. MRP d. Order entry e. Order picking

Q6 **Which provides overall support to a series of production activities by using computers?**

a. CIM b. EOS c. OA d. POS

Q7 **Which of the following systems is designed to enable the sharing of transaction, technology, and other information between purchasers and suppliers by creating an integrated data environment for the information to be used throughout the entire life cycle of products, ranging from planning, development, and design through purchasing, manufacturing, operation, to maintenance, in manufacturing industries?**

a. CAD b. CAE c. CALS d. CAM

Q8 **Which is the correct statement about groupware?**

a. Groupware is the technology to represent, store, and process graphics with computers.
b. Groupware is the technology to use microprograms instead of hardware to realize computer functionalities and instructions.
c. Groupware is a system to support collaborative work in an organization with computers.
d. Groupware is the use of software functionalities to provide an operating or running environment in which the user need not be conscious of hardware.

Q9 **Which of the systems collects and analyzes sales information on individual products separately in retail stores and is considered effective in tracking best-selling goods and preventing stock shortages?**

a. CAD b. CAM c. DSS e. OA d. POS

Q10 **Which is the correct statement about bank-POS systems?**

a. A bank-POS system analyzes day-by-day and temporal changes in a bank's over-the-counter business in order to improve operating efficiency.
b. A bank-POS system provides analysis of best-selling products and other services by connecting bank computers with POS terminals.
c. A bank-POS system performs online settlement of sales charges through POS terminals connected to bank computers.
d. A bank-POS system is a system that upon insertion of IC card issued by a bank into a POS terminal, subtracts a sales charge from the amount stored on the card and transfers it to the POS terminal.

Q11 **Which is the correct statement about the settlement methods under different card systems?**

a. Bank-POS cards, credit cards, prepaid cards, and loyalty cards all employ ID verification because they have a settlement function.
b. The settlement method with bank-POS cards is immediate payment.
c. The settlement method with credit cards is installment payments without interest.
d. The settlement method with prepaid cards is deferred payment.
e. The settlement method with loyalty cards is deferred payment.

Q12 **Following the evolution of computers, various cards have come to be used. Which of the following is the card whose main function is checking a limit amount and credit standing and allowing settlement at a later date?**

a. ID card b. Bank-POS card c. Credit card
d. Prepaid card e. Loyalty card

Q13 **Which is the <u>inappropriate</u> statement about EDI?**

a. For EDI, order placement and acceptance information formats are standardized in Japan.
b. EDI enables accurate, real-time transactions and settlements over a wide area.
c. It is expected that EDI-based placement and acceptance of orders will make large amounts of paperwork unnecessary.
d. EDI is the process of exchanging data on commercial transactions between different businesses via communication network.
e. The so-called FB (firm banking) is a kind of EDI.

Q14 **Which is the <u>inappropriate</u> statement about information systems for businesses?**

a. The DSS is an applications system that performs accounting, payroll work, and so on by computer. It is a support system designed to improve the efficiency of routine work.
b. The EOS is an automatic ordering system in which the codes and quantities of goods to be ordered are entered via data entry terminals and transmitted online.
c. The MRP system plans and manages the procurement of parts and materials based on bills of materials and the manufacturing of parts and products.
d. The POP is a system that integrates on a real-time basis the information (such as product names, quality, facilities conditions, and workers) required at the worksite to give appropriate instructions.
e. The SIS is a system that works out and executes strategies in order that the enterprise can expand its activities and strengthen its competitive position.

3 Security

Chapter Objectives

Advances in computer networks are being accompanied with increasing security risks such as the leakage of personal information, hacking of credit information, and computer virus infection. Accordingly, it is becoming increasingly important to take effective security measures.
In this chapter, the reader is expected to acquire knowledge about security and learn the necessity of security measures. The objectives are as follows:

① Learning the basic concepts and importance of information security.
② Understanding the kinds of risks involved in information processing systems and the management of those risks.

3.1 Information Security

3.1.1 What Is Information Security?

Information security means protecting information systems from various threats, including natural disasters, accidents, failures, errors, and crimes. In Japan, the Ministry of Economy, Trade and Industry (former Ministry of International Trade and Industry) has the Standards for Information System Safety Measures. Internationally, the OECD (Organization for Economic Cooperation and Development) has security guidelines.
The OECD guidelines, "Guidelines on the Security of Information Systems," defines "security" as protecting those who are dependent on information systems from hazards that may result from the absence of confidentiality, integrity, or availability.
In this context, the words "confidentiality," "integrity," and "availability" mean the following:

- Confidentiality means the state in which data, information, and the like can be disclosed only when an authorized person has gone through a prescribed procedure as authorized.
- Integrity, also called maintainability, means the state in which data and information have been maintained in an accurate, complete condition.
- Availability means the state in which data, information, and the like can be used at any time through a prescribed procedure.

3.1.2 Physical Security

Physical security means protecting information system facilities from intrusions, floods, lightning strikes, earthquakes, air pollution, explosions, fires, and other threats.

(1) RAS (Reliability, Availability, and Serviceability) Techniques

RAS is an acronym for Reliability, Availability, and Serviceability. These three elements are major yardsticks to measure the performance of information processing systems. RAS techniques are required to increase the time in which information processing systems can operate normally.
Major RAS techniques are described below.

① Redundancy system

A redundancy system means a system configuration in which a stand-by system is provided to prepare against equipment failures. Examples include parallel systems such as a duplex system and a dual system.

② Fail-safe system

"Fail-safe" refers to the idea of securing safety by preventing a failure of one part from affecting other parts. A fail-safe system is based on this idea.

③ Fail-soft system

"Fail-soft" refers to the idea of preventing a failure from halting major important functionalities at the sacrifice of some other functions. A fail-soft system is based on this idea.

(2) Standards for Information System Safety Measures

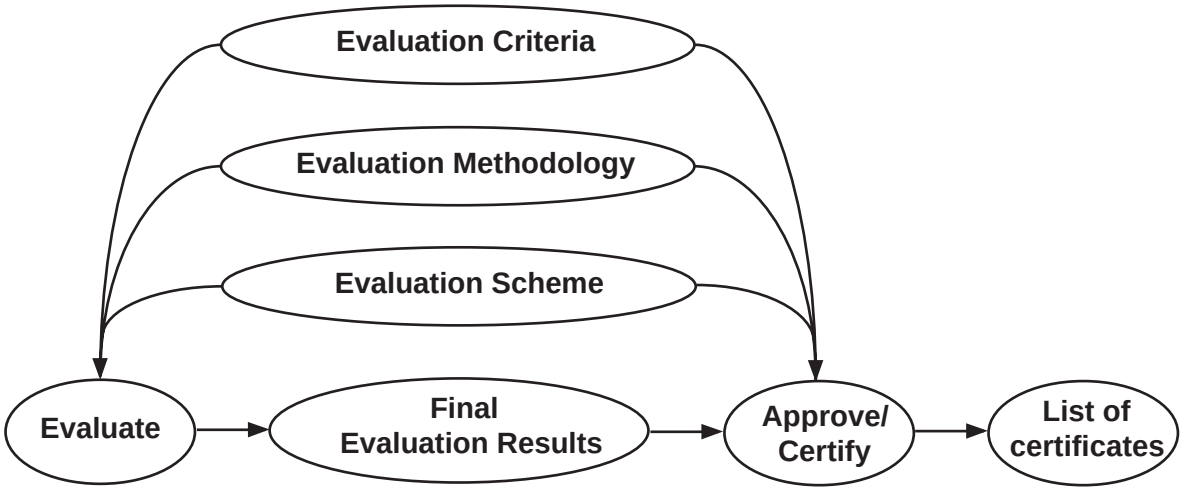
The Standards for Information System Safety Measures provide guidelines for securing the confidentiality, integrity, and availability of information systems. Last amended in 1995 by the Ministry of Economy, Trade and Industry (the former Ministry of International Trade and Industry), these standards enumerate the measures that must be taken by information system users.
The standards fall into three categories: installation standards (100 items), technological standards (26 items), and operation standards (66 items). By the magnitude of impact on society and industry, the contemplated threats are also divided into groups A, B, and C, and necessary measures are presented against them.
Other standards and guidelines regarding information systems include the following:

- Guidelines on the Security of Information Systems (1982, OECD)
- Standards for Preventing Illegal Access to Computers (1995, Ministry of Economy, Trade and Industry [former Ministry of International Trade and Industry])

CC (Common Criteria) or ISO 15408 standard

The (Common Criteria) CC represents the outcome of a series of efforts to develop criteria for evaluation of IT security that are broadly useful within the international community. In the early 1980's the Trusted Computer System Evaluation Criteria
(TCSEC) was developed in the United States. In the succeeding decade, various countries began initiatives to develop evaluation criteria that built upon the concepts of the TCSEC but were more flexible and adaptable to the evolving nature of IT in general. Work had begun in 1990 in the International Organization for Standardization (ISO) to develop international standard evaluation criteria for general use. ISO has recognized the CC and calls it the ISO 15408 standard.

Evaluation context is shown below



The CC is divided into 3 parts

a) **Part 1, Introduction and general model**

It defines general concepts and principles of IT security evaluation and presents a general model of evaluation. Part 1 also presents constructs for expressing IT security objectives, for selecting and defining IT security requirements, and for writing high-level specifications for products and systems. In addition, the usefulness of each part of the CC is described in terms of each of the target audiences.

b) **Part 2, Security functional requirements**

The functional requirements for the TOE (Target Of Evaluation) are expressed as a set of components. Part 2 catalogues the set of functional components, families, and classes.

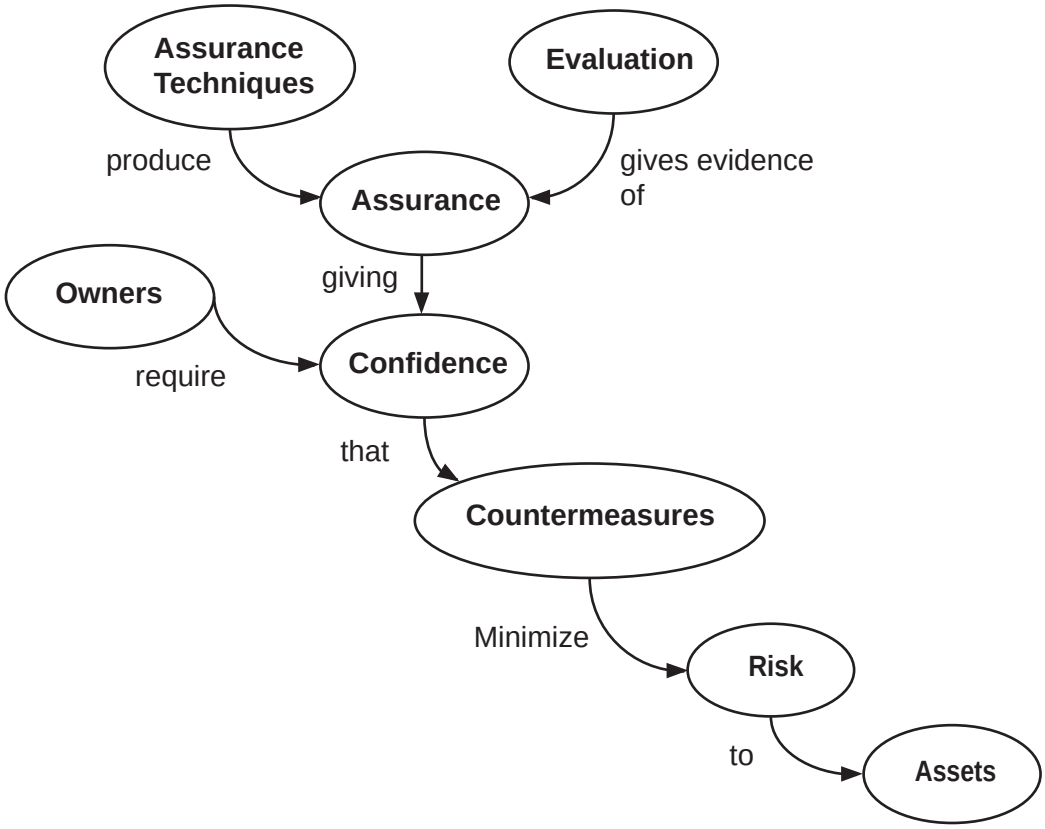
c) **Part 3, Security assurance requirements**

The assurance requirements for TOE are expressed as a set of components.

Part 3 catalogues the set of assurance components, families and classes. Part 3 also defines evaluation criteria for PPs(Protection Profiles) and ST(Security Target)s and presents evaluation assurance levels that define the predefined CC scale for rating assurance for TOEs, which is called the Evaluation Assurance Levels (EALs).

Security concepts and relationships are summarized below

Evaluation concepts and relationships



Vocabulary used in CC (Common Criteria)

Target of Evaluation (TOE) — An IT product or system and its associated administrator and user guidance documentation that is the subject of an evaluation.
Role — A predefined set of rules establishing the allowed interactions between a user and the TOE.
Protection Profile (PP) — An implementation-independent set of security requirements for a category of TOEs that meet specific consumer needs.
Security Target (ST) — A set of security requirements and specifications to be used as the basis for evaluation of an identified TOE.
Authorized user — A user who may, in accordance with the TSP, perform an operation.
TSP TOE Security Policy
Assets — Information or resources to be protected by the countermeasures of a TOE.
Evaluation Assurance Level (EAL) — A package consisting of assurance components from Part 3 that represents a point on the CC predefined assurance scale.
Evaluation — Assessment of a PP, an ST or a TOE, against defined criteria.
Human user — Any person who interacts with the TOE.
Role — A predefined set of rules establishing the allowed interactions between a user and the TOE.

3.1.3 Logical Security

Logical security means protecting information assets by encryption, user access control, and other systematic means of protection.

(1) Encryption

Encryption is a means of preventing tapping in communications. Encryption is the process of converting information into a ciphertext by using an encryption key so that it cannot be read by unauthorized people. The process of converting the ciphertext back into the plain text is called "decryption."
Decryption methods fall into two major categories:

- Common key cryptosystem: The same key is used for both encryption and decryption. The sender and the recipient need to have the same key. It is also called private key or symmetric key system.
- Public key cryptosystem: Different keys are used for encryption and decryption. The encryption key is made public, while the decryption key is kept confidential.

It should be noted that encryption is costly and requires management of the keys, which is a difficult task.

(2) Monitoring External Connection Points

It is becoming increasingly important to prevent intrusions from the outside by limiting or monitoring the points of connection with external networks, including the Internet. Routers and firewalls are monitored for this purpose.
A firewall has a filtering function to restrict the passage of data. It controls direct access to the internal network from the outside.

(3) User Authentication

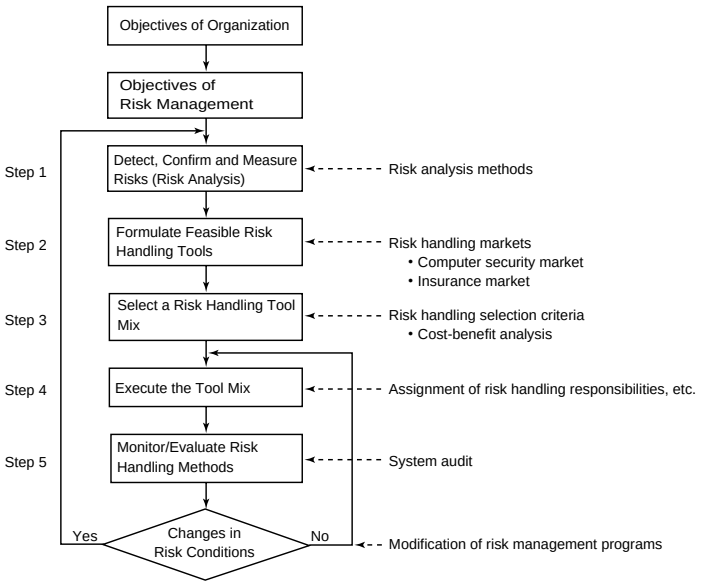
When an internal network accepts access from outside networks, it is necessary to authenticate users. A general method of user authentication requires users to enter their passwords, but this method loses its effectiveness once passwords are leaked. Hence the increasing popularity of the method using one-time passwords, which vary each time of use.

3.2 Risk Analysis

3.2.1 Risk Management

A logical process is required to cope with risks threatening an organization. It is necessary to identify possible accidents and other unfavorable events that could cause damage to an organization and take measures to deal with them in advance. This is called "risk management." It is defined as "planning, organizing, directing, and controlling the various activities of an organization in order to minimize the unfavorable operating and financial effects of contingent losses occurring in the organization."
Risk management is performed through such a procedure as shown in Figure 3-2-1.

Figure 3-2-1
Risk Management Procedure



3.2.2 Types, Evaluation, and Analysis of Risks

(1) Kinds of Risks

Risk analysis is the process of detecting risks present in an information system, determining their frequency and intensity, and analyzing how they will affect the achievement of the organization's targets. The causes of risks are referred to as "perils" or "threats." They include the following:

- Accidents and disasters
- Failures
- Errors
- Computer crimes and computer viruses
- Leaks of confidential or personal information

The factors promoting the occurrence or spread of perils are called "hazards." Examples of hazards are:

- Physical hazards: Losses resulting from physical factors such as the locations or structures of buildings and facilities
- Moral hazards: Losses caused intentionally or out of malice
- Morale hazards: Losses resulting from carelessness

(2) Risk Evaluation and Analysis

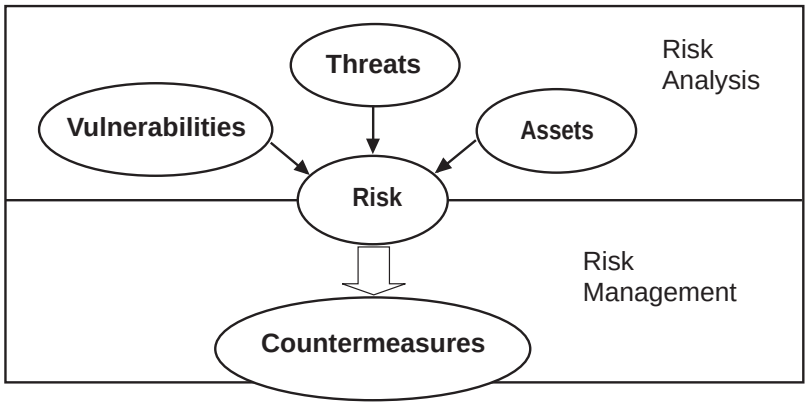
Risk analysis is performed by measuring deviations from standard values. The larger the deviations, the larger the risks.
There are two risk analysis methods: quantity method and quality method.

Risk Analysis standards

CRAMM (CCTA Risk Analysis & Management Method) methodology was devised in 1987. This methodology is used for the purpose of risk analysis. It provides a what-if ability of checking scenarios. It provides a catalog of threats and counter-measures

CRAMM asserts that risk is dependent on the asset values, the threats, and the vulnerabilities

The CRAMM method can be related to the ISO 17799 standard.



Origins of CRAMM

It was originally developed by CCTA (The United Kingdom Central Computer and Telecommunication Agency) in 1985 in response to a growing need for security in information systems.

Phases in CRAMM

It comprises 3 stages:

Stage 1: Asset identification and assessment.

There are three main types of asset involved in an IT system:

1. Physical i.e. equipment, buildings and staff
2. Software i.e. the system and application software
3. Data i.e. the information stored and processed

CRAMM reduces all items to a non-linear "value scale" of between 1 and 10. For example, anything valued at less than 1K UKP is valued as 1; for values between 1K UKP and 10K UKP the scale value is 2. Losses of over 30M UKP are scored as a 10.
CRAMM deals with all these circumstances by using a series of guidelines which map the scale of the impact onto the scale of 1 to 10 as used for simple asset values

Stage 2: Threats and vulnerabilities identification and assessment

The threats considered are:

1. Natural disasters e.g. fire, flood etc

2. Deliberate threats from outsiders
3. Deliberate threats from staff
4. IT equipment failures
5. Errors by staff

One popular model focusing on the threat impact uses two dimensions:

The first dimension divides threats into 3 categories (disclosure, modification and destruction). The second is made up of two categories (intentional and accidental)

Threat
A realized or potential event that would harm an information system
Vulnerability
This means it is susceptibility to injury or attack or the state of being vulnerable or Exposed. A weakness in the security system that might be exploited to cause loss or harm

Stage 3: Countermeasure selection.

After establishing what is to be protected, and assessing the risks these assets face, it is necessary to decide which are the controls to implement to protect these assets

The controls and protection mechanisms should be selected in a way so as to adequately counter the threats found during risk assessment, and to implement those controls in a cost effective manner

Based on the assessment , it identifies suitable and justifiable security and contingency solutions. It is used to identify security and/or contingency requirements for an information system or network.

Countermeasure Strategies

1) Security Checklist

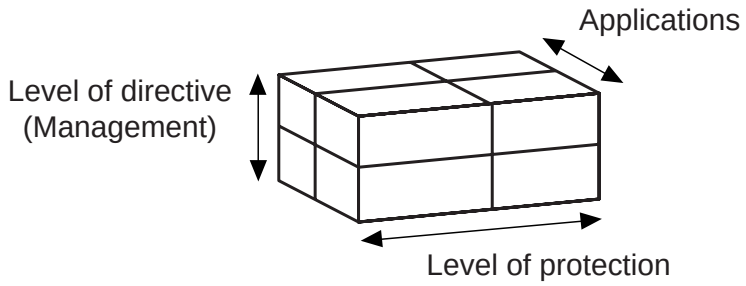
A model useful for reviewing security of current system as well as an aid when developing new security systems. Determine if controls exists and helps to identify the areas of concern where work needs to be done

There are several different general checklists available. Some checklists focus on only particular aspects of security. Descriptions of security actions / factors under a series of headings that forms a checklist

This is useful when security aspects or factors listed do not have a distinct sequential or layering Relationship

2) Matrix Model

Security (technical) objectives do not always align with management requirements of security
When deciding on controls we have to look from all possible viewpoints of the issue
The Matrix model provides us with a three-dimensional view of relationships between
Security Levels , Management Policy and Business Applications
Each intersection represents the conjunction of management directives, levels of protection and information applications
For each intersection, we decide on the security controls that should be implemented
Physical Level
Procedural Level
Logical Security (Access Management)



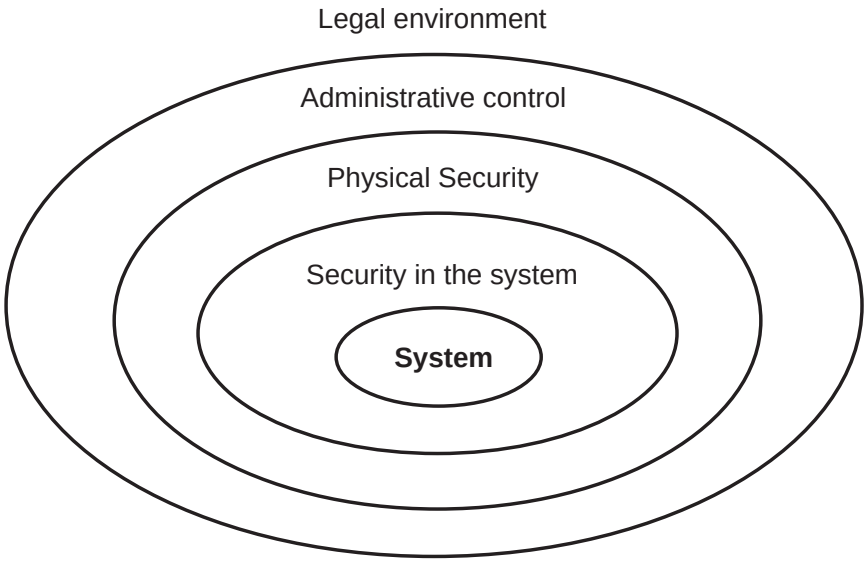
3) Ring (Onion Skin) Model

This is the most common model of information system security

Multi-layered security system enveloping and protecting the system and its data
A very early and influential statement of this model is that of Martin
Martin proposed this model to conceptualize the relationships between the aspects of security

It consists of 4 layers of control surrounding the computer system core

Each layer is considered to provide protection for the layers within, and to provide context for the operation of the inner layers



4) Filter Model

This was proposed by Smith to be used in education and training as a useful means of introducing the concepts of information security

Based on the premise that each action that may be taken to improve the security of information systems is limited in its effect Each action will only reduce the vulnerability to some security threats and not to others

Development of the checklist based models and the matrix models

Summaries, in the form of a matrix, the effects that various security actions have in providing protection against different categories of threat conditions
It comprises of Threat Categories and Action Categories
The categories have minimal overlap

ISO 17799

ISO 17799 has been developed to help organizations identify, manage and minimize, the range of threats to which Information is regularly subjected.

The following steps are required for Certification:
Developing an Information Security Policy
Defining a Scope Statement
Performing a Risk Assessment & Analysis
Defining a Statement of Applicability
Developing a Business Continuity Plan
Developing and implementing the Information Security Management System.
Completing a Certification Audit.
A cycle of periodic audits for re-certification takes place every three years

3.2.3 Risk Processing Methods

There are two risk processing methods:
- Risk control
- Risk finance

Information system security is based on risk control.

(1) Risk Control

Risk control is any of the methods of preventing the occurrence of risks or reducing their impact at their occurrence. Specific risk control methods include the following:
- Risk avoidance
- Loss prevention
- Loss reduction
- Risk separation
- Risk transfer by leasing contracts and the like

(2) Risk Finance

Risk finance refers to a financial means of ensuring a smooth recovery from the occurrence of a risk. Specific risk finance methods include the following:
- Risk holding
- Risk transfer by insurance

3.2.4 Security Measures

Procedures for risk analysis and security measures are described below.
First, risk analysis is carried out to clarify what risks are present and where in the information system. Annual losses are calculated based on the sizes and frequencies of losses. Next, security measures are worked out at a cost less than the amount of the losses.
That is, security measures are meaningless if they cost more than the losses that could result if they were not taken.

3.2.5 Data Protection

The information society is flooded with enormous volumes of data and information. Businesses hold huge

volumes of accumulated information and protect them as trade secrets. For the security of information systems, the Ministry of Economy, Trade and Industry formulated and released the System Audit Standards, the Standards for Information System Safety Measures, and the Standards for Preventing Computer Viruses.
Of the risks mentioned above, computer crimes and computer viruses are explained below from the viewpoint of data protection.

(1) Computer Crimes

Crimes in which computers are directly or indirectly involved are called "computer crimes." Data-related crimes such as those mentioned below could be committed:

① Illegal input

Illegal input is the entry of invalid data. It is difficult to prevent illegal input by online terminal operators.

② Destruction

Acts of destruction include data corruption by hackers via terminals as well as physical destruction by blasting.

③ Eavesdropping

Information could be stolen when recorded on paper or in storage media, when being processed by computer, or when being transmitted.

④ Falsification

Falsification means any unauthorized modification or deletion of data or programs.

(2) Computer Viruses

A computer virus is a program that destroys or falsifies the contents of memories and disks. It is often difficult to identify the route and time of virus infection. Some computer viruses remain dormant for some time after infection before becoming active. Typical symptoms of virus infection include the following:

- Program destruction
- Destruction of file data
- Sudden appearance of graphics or characters on the display
- Occurrence of trouble at a specific date or time (such as Friday, the 13th)

It is often too late to take some action after finding a symptom of infection. Therefore, floppy disks brought in from outside should be checked by anti-virus software before they are used. It is safe not to use media whose origins or owners are not known. On this issue, the Ministry of Economy, Trade and Industry formulated and released the Standards for Preventing Computer Viruses.
The type of virus that has been particularly prevalent in recent years is the macro virus. Macro viruses take advantage of the macro functions of applications programs sold on the market. A macro virus infects a data file of an applications program, and when the file is opened by the user, the macro function is executed without the user's knowledge. Macro viruses can spread more widely than the conventional types of viruses dependent on operating systems and hardware. One such example was the powerful "Melissa" virus, which emailed itself to all of a user’s address book entries.

3.2.6 Protection of Privacy

In their sales activities, businesses obtain personal information from order forms and applications prepared by consumers. The information obtained this way is usually stored in databases for use in subsequent sales activities. These databases hold enormous volumes of information, including address, gender, date of birth, family members earnings, and property held. Public organizations also hold huge volumes of personal information stored in the resident, taxpayer, driving license, social insurance, and other registries.
Personal information should naturally be kept confidential because of its character. Should it be disclosed by mistake or otherwise, privacy is inevitably violated. The protection of privacy is opposite to disclosure.

Any organization holding personal information must take every precaution to prevent the leakage of information.

For the protection of personal information, the OECD's privacy guidelines contain eight basic principles. In Japan, the Act for Protection of Computer Processed Personal Data held by Administrative Organs was established in 1988 to properly regulate the use of personal information (such as social insurance, tax payment, driving licenses, and resident registration) held by administrative agencies.

At present, however, Japan has only several guidelines in this field, including the Guidelines for Individuals' Information Protection established in 1989 by the Ministry of Economy, Trade and Industry and the Guidelines for the Protection of Personnel Information in Computer Processing in the Private Sector established in 1995 by the ministry. No legislation has been established yet to regulate the use of personal information in the private sector.

Exercises

Q1 **Which of the following measures is least effective for warding off, detecting, or eliminating computer viruses?**

a. Do not use software of an unknown origin.
b. When reusing floppy disks, initialize them in advance.
c. Do not share floppy disks with other users.
d. Clear the memory before executing a program.

Q2 **Which is the correct statement about the recent increase in macro viruses?**

a. The execution of an infected application loads the macro virus into the main memory, and in this process, the virus infects program files of other applications.
b. Activating the system from an infected floppy disk loads the macro virus into the main memory, and then the virus infects the boot sectors of other floppy disks.
c. A macro virus infects document files opened or newly created after an infected document file is opened.
d. Since it can be easily determined as to whether a macro function is infected by a virus, infection can be prevented at the time of opening a document file.

Q3 **Which is the appropriate term to describe the information given to users for the purpose of checking the authenticity to use a computer system and grasping the condition of use?**

a. IP address b. Access right c. Password d. User ID

Q4 **Which is the most appropriate practice for user ID management?**

a. All the users involved in the same project should use the same user ID.
b. A user having multiple user IDs should set the same password for all the IDs.
c. When privileges are set for a user ID, they should be minimized.
d. When a user ID is to be deleted, an adequate time interval should be taken after the termination of its use has been notified.

Q5 **Which is the inappropriate statement about the use or management of passwords?**

a. If a password is incorrectly entered a predetermined number of times, the user ID should be made invalid.
b. Passwords should be recorded in a file after being encrypted.
c. Users should try to use those passwords which are easy to remember, but those which are hard to be guessed by other people.
d. Users should be instructed to change their passwords at predetermined intervals.
e. Passwords should be displayed on terminals at the point of entry for the purpose of confirmation.

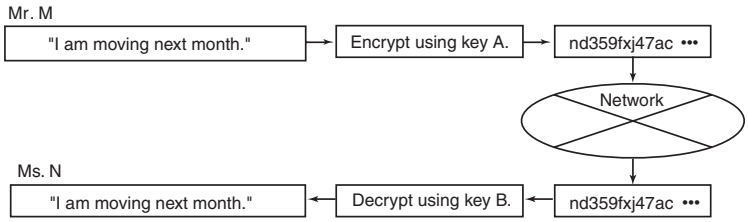
Q6 **Which is in an inappropriate way of handling passwords and a password file in the system management department?**

a. The security managers should regularly check whether or not passwords can be easily guessed, and recommend that problem passwords be changed.
b. The department should recommend that users record their passwords in their notebooks in order to minimize the frequency of inquiring about their passwords.
c. If it is possible to set the term of validity of passwords, the term should be used for checking password validation.
d. Even if a password file records encrypted passwords, the department should make it inaccessible to general users.

Q7 **From the viewpoint of security, which is the <u>inappropriate method of operating</u> a computer system using a public switched telephone network?**

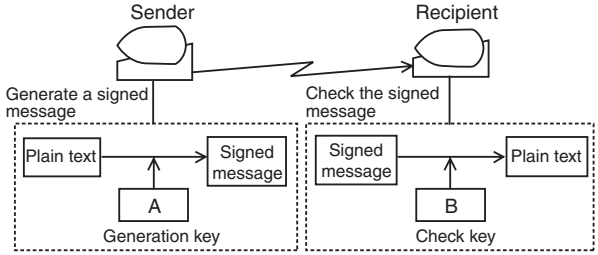
a. Make a password unusable for connection unless it is changed within predetermined intervals.
b. When a connection request is made, establish connection by calling back to a specific telephone number.
c. Display a password on a terminal at the point of entry so that the user will not forget the password.
d. Disconnect the line if a password is wrongly entered a predetermined number of times.

Q8 **When as shown in the figure below, Mr. M sends to Ms. N a message they want to keep confidential, which is the appropriate combination of the keys used for encryption and decryption?**



	Key A	Key B
a	M's private key	M's public key
b	N's public key	N's private key
c	Common public key	N's private key
d	Common private key	Common public key

Q9 **The figure shows the configuration of electronic signature used into the public key cryptosystem. Which is the appropriate combination of the terms to be put into a and b?**



	A	B
a	Recipient's public key	Recipient's private key
b	Sender's public key	Sender's private key
c	Sender's private key	Recipient's public key
d	Sender's private key	Sender's public key

Q10 **There is a transposition cryptosystem in which plain text is divided into four-character blocks and in each block, the first character is replaced by the third, the second by the first, the third by the fourth, and the fourth by the second. In this system, which is the correct cipher text for the plain text "DEERDIDDREAMDEEP"?**

a. DIDDDEEPDEERREAM
b. EDREDDDIARMEEDPE
c. ERDEIDDDEMRAEPDE
d. IDDDEPDEERDEEMRA
e. REEDDDIDMAERPEED

4 Operations Research

Chapter Objectives

In this chapter, the reader is expected to acquire knowledge about Operations Research that is useful and necessary in realizing an optimal information system for business operations. The objectives are as follows:

①Understanding the basic concepts of probabilities and statistics
②Understanding the basic concepts of linear programming
③Understanding the basic concepts of scheduling using PERT
④Understanding the basic concepts of queuing theory
⑤Understanding the basic concepts of inventory control
⑥Understanding the basic concepts of demand forecasting

4.1 Operations Research

4.1.1 Probabilities and Statistics

(1) Events and Sets

① Events

Consider the outcome of rolling a dice once. The number of dots on the side which ends face up is either 1 or 2 or ... or 6. The action of rolling a dice is called a "trial". Each of the expected outcomes of a trial is called an **"event"**.
The set of all possible outcomes of an experiment is called the "**certain event**" or "**sample space**" of the experiment, commonly denoted by Ω or U. Coming back to the rolling-a-die-once example, its certain event Ω is as follows.

$\Omega = \{1,2,3,4,5,6\}$

where each outcome i, for i =1, . . . , 6, corresponds to the face value.
Next, examine whether the result of a roll is an even number or an odd number. Let the event E be as follows.

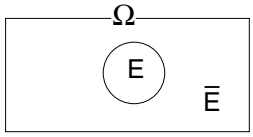
$E = \{2,4,6\}$

Then the case that the result of a roll is an even number is being referred to as "the event E has occurred". The opposite case, "the event E has not occurred" (or the result is an odd number) is defined as the "complement event" of E, denoted by $\overline{E}$ or E^c.

$\overline{E} = \{1,3,5\}$

The relationship between the certain event Ω (or U), the event E, and the complement event $\overline{E}$ described above is represented in the Venn diagram, in the Figure 4-1-1.

Figure 4-1-1



The complement event of the certain event Ω, denoted by $\overline{\Omega}$ means there are no possible outcomes. This is called the "**null event**", denoted by ϕ.
Suppose $\Omega = \{1,2,3,4,5,6\}$. Then $\phi = \{\}$. That is, in the rolling-a-die-once experiment, Ω occurs and ϕ never occurs.
Each of the possible outcomes, such as an outcome where the resulted face value is 1 is called a "**simple event (or elementary event)**".

② Relationships between events and sets

Additional types of events classified based on the relationships between multiple events are shown below.

a. Union of events

Suppose there are 2 sets of events, A and B. Under the phenomena in which either event A or B occurs; or both events A and B occur, we say this is the **union of events** of A and B, denoted by the union of sets "$A \cup B$".
For instance, the union of events of E, F and G is denoted by $E \cup F \cup G$. The union of an event H and its complement $\overline{H}$ is, $H \cup \overline{H} = \Omega$.

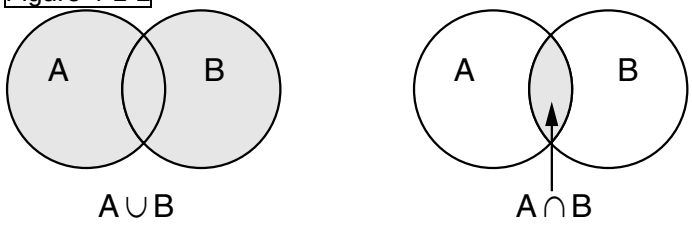
b. Intersection of events

When events A and B occur simultaneously, we say this is the **intersection of events** of A and B, denoted by the intersection of sets "$A \cap B$".

For instance, the intersection of events E, F and G is denoted by $E \cap F \cap G$. The intersection of an event H and its complement $\overline{H}$ is, $H \cap \overline{H} = \phi$.

The following Figure 4-1-2 shows the union of events $A \cup B$ and the intersection of events $A \cap B$.

Figure 4-1-2



c. Exclusive events

When events A and B never occur simultaneously, we say that A and B are mutually exclusive, or they are **exclusive events**. This means that if events A and B are mutually exclusive, $A \cap B = \phi$ is true.

The following Figure 4-1-3 shows the relationship between events and sets.

Figure 4-1-3 Relationship between events and sets

Event	Set	Notation
simple event	element	A,B, ...etc.
the certain event	universal set	Ω (or U)
null event	empty set	ϕ
complement of an event	complementary set	$\overline{A}$ $\overline{B}$,...or A^c,B^c,...etc.
union of events	union	$\cup$
intersection of events	intersection	$\cap$

(2) Probability

①Definition of Probability

Probability is the measurement of the likelihood of some occurrence. Assuming that an event E is associated with an experiment, the probability of the event E is defined as the number of times that the event E has occurred divided by the total number of experiments (or the ratio of E occurrences to all of the occurrences). This probability is denoted by P(E). P(E) is between 0 and 1. From the definition,

the certain event: $P(\Omega)=1$ null event: $P(\phi)=0$ $P(E)+P(\overline{E})=1$

②Mathematical Probability

Generally, the probability of an event E is calculated as:

P(E) = number of simple events in the event E / number of simple events in the certain event = $n(E) / n(\Omega)$

This is called the mathematical probability.

For example, the probability that the sum of the outcomes of two rolls of a die is 4 can be calculated as follows.

Let i = the outcome of the first roll, j = the outcome of the second roll, then the outcomes of two rolls can be represented as (i,j). The certain event Ω has 36 elements, that is

Ω ={(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), ...(6,5), (6,6)}

From the above, the elements in event E (the sum of i and j is 4) are

E={(1,3), (2,2), (3,1)}

Therefore, the probability that the sum of two rolls of a die equals 4 is

P(E) = number of simple events in E / number of simple events in Ω = 3/36 = 1/12

③Statistical Probabilities

There exists the **statistical probability** (or **experiment-based probability**) besides the mathematical probability.
Let m be the number of times the event E occurs out of n-time experiments, then if the value m/n is observed to get closer to a definite value p as n reaches infinite, this p is called the statistical probability of the event E.
Consider the statistical probability of the event E that the face value in the outcome of a roll of a die is 1. We can assume that if we roll a die repeatedly, as the number of rolls gets larger, the ratio of favorable outcomes (i.e. the face value is 1) will get as closer as possible to 1/6. We say that "the statistical possibility of E is 1/6".
This is equal to the mathematical probability 1/6. The statistical probability of an event corresponds to its mathematical probability. This is called the "**law of large numbers**".

(3) Axioms of Probability

①Addition Theorem

Let $A \cup B$ = the union of events A and B, $A \cap B$ = the intersection of events A and B, then the following relationship is always true. This is called the **addition theorem**.
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
If the event A and B are mutually exclusive, and therefore $A \cap B = \phi$ (or $P(A \cap B) = 0$), the following relationship is also always true.
$P(A \cup B) = P(A) + P(B)$
Consider a trial that draws one card out of a set of playing cards.
Event A : the event that the extracted card is club
Event B : the event that the extracted card is a picture card
Event C : the event that the extracted card is spade
Let us compute the probabilities, P(A), P(B), $P(A \cap B)$, $P(A \cup B)$, and $P(A \cap C)$.
First the number of simple events in A,B,C is to be counted respectively.
Number of simple events in the certain event
: 52 (= the total number of cards)
Number of simple events in A
: 13 (= the number of club cards)
Number of simple events in B
: 12 (= the number of picture cards)
Number of simple events in C
: 13 (= the number of spade cards)
Number of simple events in $A \cap B$
: 3 (= the number of picture club cards)
Number of simple events in $A \cap C$
: 0 (= the number of cards that are clubs as well as spades. Note that A and C are exclusive events)
The computation results are shown below.
P(A)=13/52=1/4 P(B)=12/52=3/13
P(C)=13/52=1/4 $P(A \cap B)$=3/52
$P(A \cup B)$=P(A) + P(B) - $P(A \cap B)$=22/52=11/26
$P(A \cap C)$=0

②Multiplication Theorem

The following is always true for two events A and B, if the possibility that one event occurs does not affect the possibility that the other event occurs.
$P(A \cap B) = P(A) \cdot P(B)$
Events A and B are called mutually independent, or are called **independent events**.
The probability that B will occur given that A is known to have occurred is called the "**conditional probability**", denoted by P(B|A). This means the probability that B will occur under the condition that A is

known to have occurred. Concerning this, the following is always true.
$P(A \cap B) = P(A) \cdot P(B|A)$
Suppose that there are 100 students in a classroom, 30 of which wear glasses. 60 of the students are male and 1/3 of the male students wear glasses. One student out of the 100 is randomly selected. What is the probability that the selected student is male and wearing glasses? Let events A and B be as follows.
Event A = the event that the selected student is a male
Event B = the event that the selected student wears glasses
Then, the event $\overline{A}$ is "the event that the selected student is female", the event $\overline{B}$ is "the event that the selected student does not wear glasses". The relationships among the above events are summarized below, in the Figure 4-1-4.

Figure 4-1-4 Relationships between events A, $\overline{A}$, B, $\overline{B}$

	A		$\overline{A}$		Total
Event	$A \cap B$	$A \cap \overline{B}$	$\overline{A} \cap B$	$\overline{A} \cap \overline{B}$	
Meaning of the event	the event that a randomly chosen student is "male and wearing glasses"	the event that a randomly chosen student is "male and not wearing glasses"	the event that a randomly chosen student is "female and wearing glasses"	the event that a randomly chosen student is "female and not wearing glasses"	
number of simple events	20	40	10	30	100

Let us compute the probabilities P(A), $P(A \cap B)$ and $P(\overline{A} \cap B)$ first.

P(A) = number of male students / number of all students = 60/100 = 0.6
$P(A \cap B)$ = number of male students with glasses / number of all students
= 20 / 100 = 0.2
$P(\overline{A} \cap B)$ = number of female students with glasses / number of all students
= 10 / 100 = 0.1
Then,
P(B|A) = number of male students with glasses / number of male students
= 20 / 60 = 1/3
In this fraction representing the probability, dividing the numerator and the denominator respectively by the number of all students does not change the probability itself. That is, the following two equations are true.

P(B|A) = (number of male students with glasses / number of all students) / (number of male students / number of all students) = $P(A \cap B) / P(A)$
$P(A \cap B) = P(A) \cdot P(B|A)$

Here, as shown below, the two possibilities P(B) and P(B|A) are different.
P(B) = 0.3
: the probability that event B (the selected student wears glasses) occurs
P(B|A) = 1/3
: the probability that event B occurs given that A (the selected student is male) is known to have occurred
That is $P(A \cap B) = P(A) \cdot P(B|A)$ is true because the event A and the event B are not independent. (i.e. conditional probability exists.)
Thus, when the probability that event B will occur varies depending on whether another event A has occurred or not, the event B is called a **dependent event** of the event A.
Compare the above with the following. Suppose that we draw one card out of a set of 52 cards. Let events A and B be as follows:
Event A : the event that the extracted card is a club
Event B : the event that the extracted card is a picture card
Then,
P(A) = 13/52 = 1/4

: the probability that event A (the event that the extracted card is a club) occurs
P(B) = 12/52 = 3/13
: the probability that event B (the event that the extracted card is a picture card) occurs
$P(A \cap B) = (1/4) \times (3/13) = 3/52$
Here,
P(B)=12/52=3/13
P(B|A)=3/13
$P(B|\overline{A})$=9/39=3/13

The probability that the outcome is a picture card is always the same, regardless of whether the outcome is a club card or not. That is, the events A and B are mutually independent and the following is true.
$P(B)=P(B|A)=P(B|\overline{A})$ and $P(A)=P(A|B)=P(A|\overline{B})$

Therefore from definition, $P(A \cap B) = P(A) \cdot P(B)$ is true.
As shown above, how to compute $P(A \cap B)$ depends on whether the probability that the event B will occur given that the event A is known to have occurred
is always the same, regardless of whether event A has occurred or not, or
varies depending on whether event A has occurred or not

③Independent Trials

If the probability that an event occurs is always the same, even if the trial is repeated several times, the trial is called an "**independent trial**".
Assume that the probability that an event E occurs is always the same and the value is p. Here, the probability P_r that the event E will occur r-times out of n-trials is computed as below.
$P_r = {}_nC_r p^r q^{n-r}$

$$(\text{Here, } {}_nC_r = \frac{n!}{r!(n-r)!} \qquad q=1-p)$$

Consider the probability that only one outcome is the side with 6 dots, out of the three trials of rolling a die. Since this is an independent trial, the possibility p that the resulting face value is 6 is 1/6. Therefore the answer is as follows.

$$P_1={}_3C_1 \left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2 = \frac{3!}{2!1!}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2 = \frac{25}{72}$$

(4) Statistics

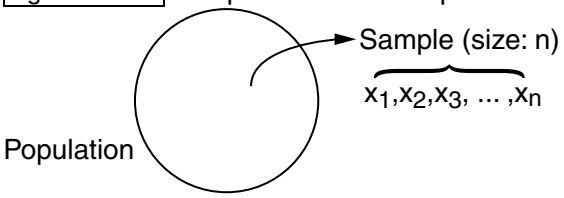
①What is statistics?

In natural phenomena, economic phenomena or social phenomena, even though the nature of data is already known, definite data values cannot be identified until individual data is observed. This is called irregularity of data. However, even so, in most of such cases some rules or regularities can be found as the result of examining the entire data. As above, the regularities that may be found in the entire data are called the statistical rules of data.
Statistics means techniques to find out these kinds of statistical rules of data. In general, the statistical methods analyze data by taking a series of steps as shown below.
How to collect data (mathematical statistics)
How to organize the collected data (descriptive statistics)
Using the organized data, how to identify the characteristics of the unknown data group and how to derive its regularities (mathematical statistics)
How to extract necessary information needed for decision making etc. from the collected statistical data (multivariate analysis and others)
To be simplified, statistics is something like a method used to find the regularities (e.g. a mean value) in large amounts of data, by first extracting a small amount of data out of the entire data, examining the extracted data to calculate its mean value, and finally extrapolating the mean value of the entire data.

②Population and Sample

All of the data within the data group to be observed by using a statistical method are called a "**population**". The method that some partial data out of the population (this is called a "**sample**") is to be extracted and examined is called a "**sampling survey**", the method that the entire data in a population is to be examined is called a "**complete survey**".

Figure 4-1-5 Population and sample



There has been much progress in the technology to derive statistical rules concerning a population by extracting just a small amount of data as an appropriate sample and examining it, instead of examining the entire data with a large amount.

(5) Frequency Distribution

①Variates

A **variate** is a thing that has some nature but whose value is not definite until actually observed. Values in statistical materials are variables. These variables are called variates in statistics.
For instance, the actual length of a product with a certain length specification, or students' scores on a test are variates.
Variates can be classified into the following two types.
Discrete variate : A variate that cannot take on all values. This takes integer values or non-integer discrete values.
Continuous variate : A variate that can take on any value. This can take arbitrary values within a given range.

②Frequency Distribution

Frequency distribution is convenient to understand the distribution of a variate. This shows the number of observations (frequency) in an individual interval assuming that the entire interval of the variate has been subdivided into a number of intervals.
<In the case of a discrete variate>
Possible values of x : $x_1, x_2, x_3, \ldots, x_k$
Frequency of x_i out of N observations: f_i (i=1,2,...,k)
<In the case of a continuous variate>
Possible ranges of x : The entire range is to be subdivided into k intervals (usually intervals of equal length). Each of the subdivided ranges is called a class. The median of a class is taken as its class frequency.
class frequency of i-th range : x_i
Frequency of the i-th class out of N observations: f_i (i=1,2,...,k)
In either case, the correspondence
$x_i \rightarrow f_i$ (i=1,2, ... ,k)
for the variate x is called the frequency distribution for x.

③Frequency Tables and Histograms

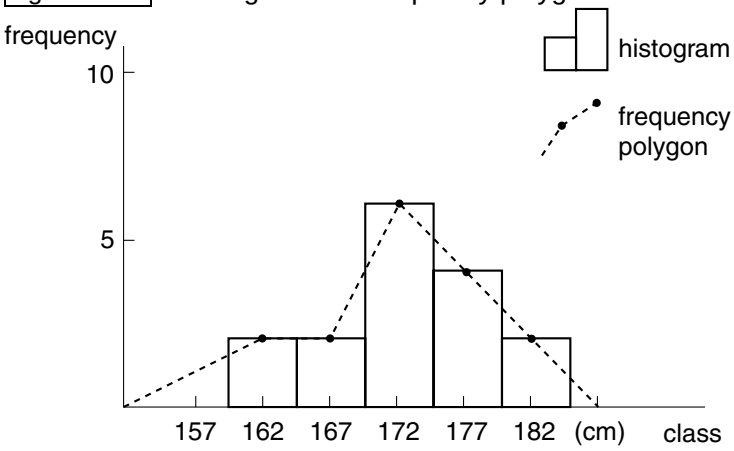
A frequency table shows the frequency distribution for a variate.
Suppose that heights of 16 players on a soccer team are as follows.
170cm, 181cm, 171cm, 184cm, 163cm, 178cm, 176cm, 174cm, 178cm, 164cm,
174cm, 168cm, 167cm, 177cm, 172cm, 174cm
These data can be summarized by the frequency table below.

Figure 4-1-6 Frequency table

Class	Class frequency x	frequency f	relative frequency	cumulative frequency	cumulative relative frequency
155≦x<160	157	0	0.000	0	0.000
160≦x<165	162	2	0.125	2	0.125
165≦x<170	167	2	0.125	4	0.250
170≦x<175	172	6	0.375	10	0.625
175≦x<180	177	4	0.250	14	0.875
180≦x<185	182	2	0.125	16	1.000
Total		16	1.000	16	1.000

Here, a relative frequency is the result of dividing a frequency by the number of observations (N=16 in the above table). A cumulative frequency is the cumulative sum of frequencies of each class. A cumulative relative frequency is the result of dividing a cumulative frequency by the number of observations.
What is shown in a frequency table can be graphically represented for ease of understanding. To visualize a frequency table, the following charts are commonly used.
Histogram : a chart in which classes and frequencies are shown either by a line or bars
Frequency polygon : a polygonal line that shows frequencies by connecting class frequencies (class medians).

Figure 4-1-7 Histogram and frequency polygon



(6) Characteristics of Distribution

①Central tendency

A **central tendency** is a single value that shows some characteristics of a variate.

a. Mean value

A **mean** is the value that is computed by dividing the sum of variate values by the number of observations. The mean of the previous data is
(170+181+171+184+163+178+176+174+178+164+174+168+167+177+172+174)÷16 = 173.2cm

b. Mode

A **mode** is the class that has the highest frequency of all the classes. In the previous example, it is the class "170≦x<175" whose frequency is 6.

c. Median

A **median** is the value that is the middle one in a set of variate values arranged in order of size. The median of the previous example is "174".

②Variations

Variations here refer to in what degree values of a variable are dispersed. This is represented by such factors as a variance and a standard deviation.

a. Variance

A **variance** is a measurement of data variations, calculated as the average of (the margin between each variate value and the mean)^2.

$$\text{Variance} = \frac{1}{number\ of\ observations} \sum_{1}^{number\ of\ observations} (\text{var}\,iate\ value - mean)^2$$

In the previous example, the variance is calculated as follows, assuming that the mean is 173.

$$(1/16)\times\{(170-173)^2 + (181-173)^2 + (171-173)^2 + (184-173)^2 + (163-173)^2 + \cdots\cdots + (174-173)^2\}$$
$$=519/16$$
$$\fallingdotseq 32.4$$

b. Standard Deviation

A **standard deviation** is also a measurement of data variations, calculated as $\sqrt{\text{var}\,iance}$. In the previous example, it is $\sqrt{32.4} \fallingdotseq 5.7$.

The fact that the variance (or the standard deviation) of a variate is large means its degree of variation is also large.

Conversely, if the variance (or the standard deviation) of a variate is small then its degree of variation is also small. That is, variate values are densely packed around the mean.

(7) Various Probability Distributions

①Random Variables and Probability Distributions

Let x be a variable associated with a trial. x is called a "**random variable**" if the probability that x takes a certain value can be defined. A "**probability distribution**" shows the relationship between a random variable x and the above probability.

For instance, in the die rolling, the probability that an outcome i will occur (i=1,2,...,6 which corresponds to the face value) is 1/6, respectively. Here the variable x represents the face value in an outcome, and this x is a random variable. And the relationship between the random variable x and its probability 1/6 is a probability distribution.

By regarding a variable as a random variable, the probability that the variable value exists in a given range can be computed. This result can be utilized as helpful information. There exist two different types of random variables.

a. Discrete random variables and Probability Distributions

A **discrete random variable** has a countable number of discrete values.

The precise mathematical definition is : In a finite set $\{x_1,x_2, \ldots ,x_n\}$, if the probabilities $P(x=x_1)=P_1$, $P(x=x_2)=P_2$, ... , $P(x=x_n)=P_n$ (probability that X equals each element of the set) can be defined, x is said to be a discrete random variable.

In this case, the following is always true.

$$0 \leqq P_i \leqq 1 \qquad \sum_{i=1}^{n} P_i = 1$$

The correspondence between individual x_i and $P(x=x_i)$ is called a probability distribution. In addition,

$$P(x \leqq x_i) = P(x = x_1) + P(x = x_2) + \cdots + P(x = x_i)$$

$\|$

the probability that a value of a variable x is equal to or smaller than x_i

is defined as the probability distribution function $F(x= x_i)$. i.e.,

$F(x=x_i) = P(x \leqq x_i)$

b. Continuous random variables and Probability Distributions

Let x be a random variable whose value changes continuously within a certain range, this x is called a **continuous random variable**.

The precise mathematical definition is : If the probability $P(a \leqq x \leqq b)$ that the value of x exists within a specific range (i.e. $a \leqq x \leqq b$) inside the given interval $[\alpha, \beta]$ can be defined as follows, x is said to be a continuous random variable.

$$P(a \leqq x \leqq b) = \int_a^b f(x)dx$$

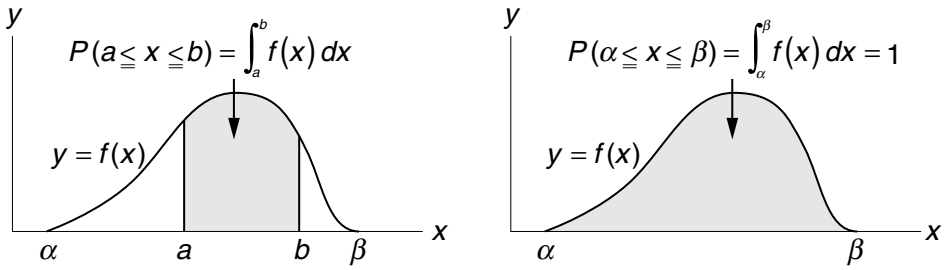
This is called a probability density function. It is a function unique to a continuous random variable

A density function : f(x) has the following characteristics

$$f(x) \geqq 0, \quad \int_\alpha^\beta f(x)dx = 1 \ (\textit{the same as } P(\alpha \leqq x \leqq \beta) = 1)$$

The interval where a random variable value exists $[\alpha, \beta]$ can be $[-\infty, \infty]$.

Figure 4-1-8 Probability density function f(x) of random variable x



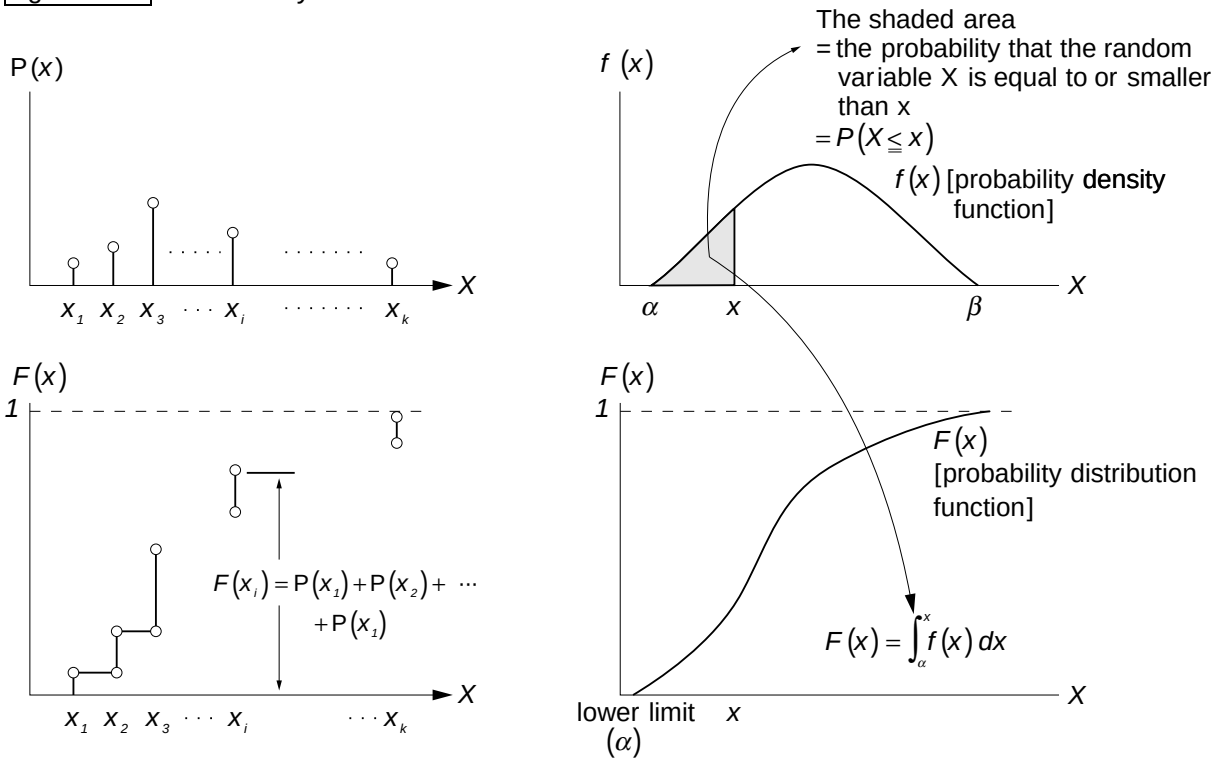
Here let X be a random variable, x be a certain value, the probability of $P(X \leqq x)$ is represented by the probability distribution function: F(x).

The probability density function: F(x) has the following characteristics.

$$F(X=x) = P(X \leqq x) = \int_a^b f(x)dx$$

$F(X=\alpha)=0$, $F(X=\beta)=1$

Figure 4-1-9 Probability Distribution Function



(a) In the case of discrete random variables

(b) In the case of continuous random variables

In general, a probability distribution of a random variable is determined by its "type of distribution and its parameters". From now on we will use the following as the notation of a distribution.
a random variable ∈ a name showing the distribution type (parameters for determining the distribution)

②Binomial Distributions

The **binomial distribution** is derived from the theorem of independent trials. It is represented by the equation $P_r={}_nC_r p^r q^{n-r}$ (here, q=1-p). For instance, suppose that the number of times the outcome is the face with 6 dots as a result of rolling a die 50 times is r, then the probability of r follows the binomial distribution. In the example above,

$$P_r = {}_{50}C_r \left(\frac{1}{6}\right)^r \left(\frac{5}{6}\right)^{50-r}$$

The representation of the binomial distribution is B, the parameters are n and p, as shown below. Assume that E(x) represents the expected value (or mean), V(x) represents the variance.
<Equations of binomial distribution>
the probability that the favorable event occurs at one trial : p
the number of times the favorable event occurs out of n-time trials : x
(x is not definite until the n-th trial finishes)

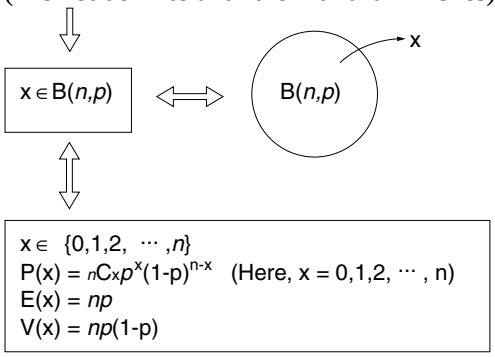
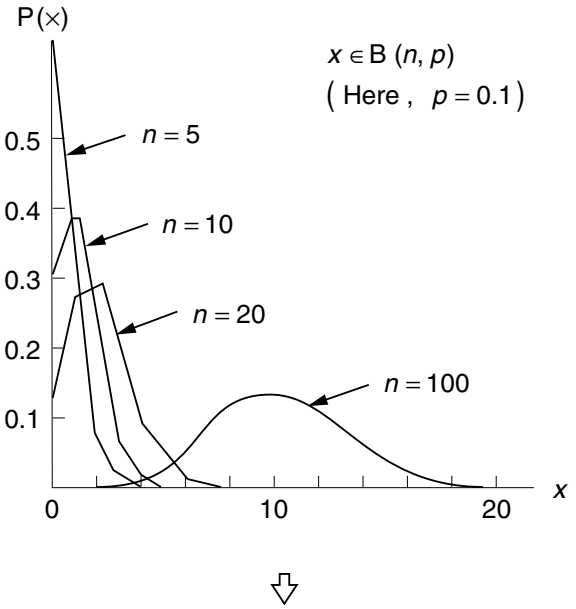


Figure 4-1-10 Relationship between the probability that follows the binomial distribution and n



When *n* gets bigger, the distribution becomes closer to the normal distribution

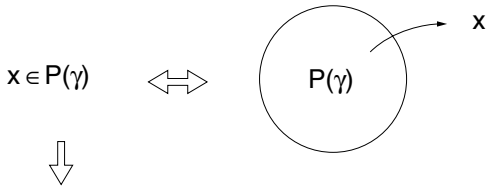
③Poisson Distribution

As for the random variable $x \in B(n,p)$ that follows the binomial distribution, the probability distribution with "E(x) = *np* = definite" and $n \to \infty$ is called the **Poisson distribution**. The type letter representing the Poisson distribution is P, with only one parameter γ. This is represented by the following equation.

$P(x) = (1/x!) \times \gamma^x e^{-\gamma}$ (γ represents the mean of *np*)

In the Poisson distribution, when a mean is given, everything concerning the distribution becomes definite. (Because the probability, mean, and variance are dependent on the parameter γ.)

<Equations of Poisson distribution>



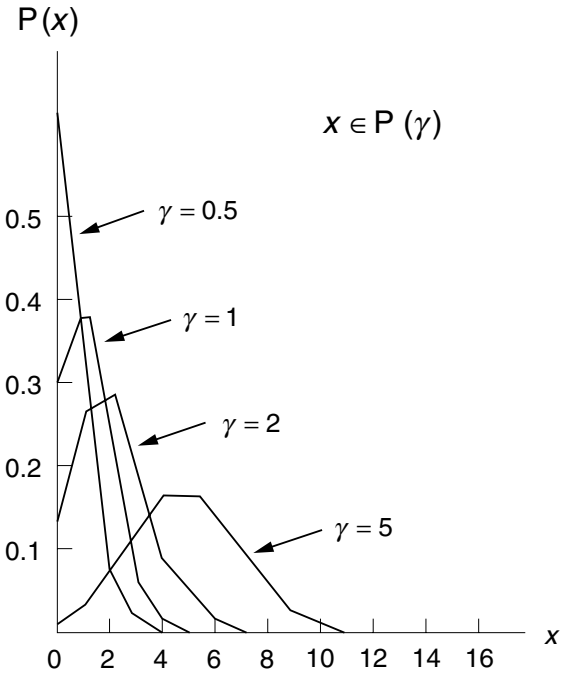
$x \in \{0,1,2, \cdots, n\}$

$P(x) = \frac{1}{x!} \gamma^x e^{-\gamma}$

$E(x) = \gamma$

$V(x) = \gamma$

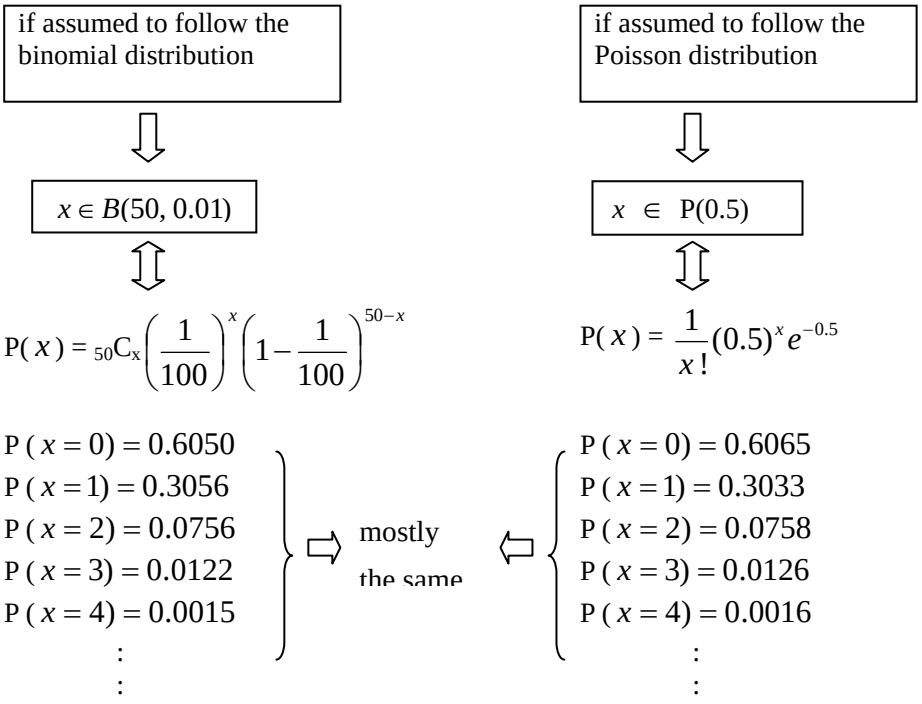
Figure 4-1-11 Probability with Poisson distribution for different means



For instance, suppose the defective rate of a product is 1/100 and its lot size is 50. Let x be the number of defective ones out of 1 lot. Precisely speaking this x follows the binomial distribution. In practice, when the probability is close to 0, it is easier to consider that it follows the Poisson distribution.
In this case, since $\gamma = np = 50 \times 0.01 = 0.5$, the probability of x is as follows.

$P(x) = (1/x!)(0.5)^x e^{-0.5}$

<A concrete example of the Poisson distribution>



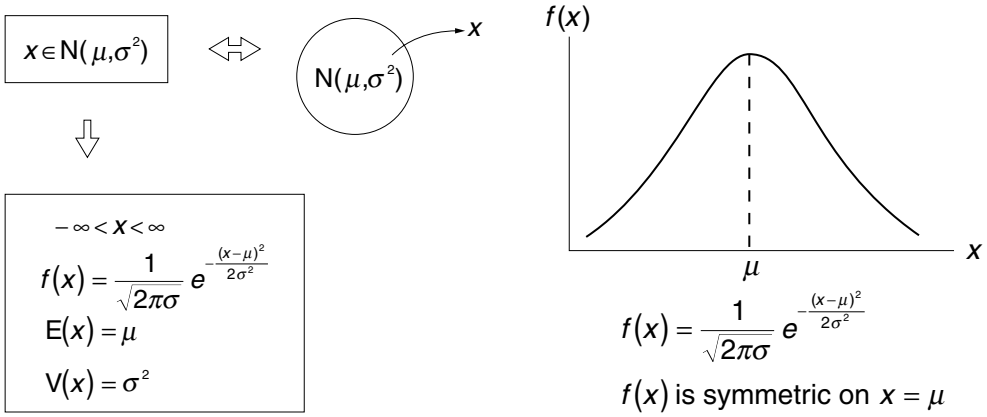
④Normal Distribution

The **normal distribution** is the most commonly used distribution for continuous variables. This has been studied sufficiently and its characteristics are well known. This distribution type is frequently observed in social phenomena and natural phenomena. The type letter for the normal distribution is N, with parameters μ and σ^2.

<Equations of normal distribution>

The equations in the Figure 4-1-12 show that "the random variable x follows the normal distribution with the mean μ and the variance σ^2".

Figure 4-1-12 Normal distribution



Some of the characteristics of a normal distribution are shown below.

In a normal distribution, every aspect concerning the probability is determined only by the mean μ and the variance σ^2

The probability density function is symmetric on x= μ

4.1.2 Linear Programming

Linear Programming is a management technique, used to find the optimal solution with certain given conditions.

(1) Formulation of Problems

Consider a company that produces and sells two types of products A and B. Production of these products requires materials P and Q. The company would like to find a production plan that maximizes its profit given the following constraints.
Production of Product A requires 6kg of material P and 2kg of material Q
Production of Product B requires 3kg of material P and 4kg of material Q
A maximum of 120kg of material P and a maximum of 100kg of material Q can be used for production of products A and B per month
Profit of product A per production unit is 30,000 dollars while that of product B is 40,000 dollars
Under the above assumptions, how many units of each product should the company produce and sell to maximize its profit? This problem can be solved by using the Linear Programming.

First, this problem can be shown in the following table.

	Product A	Product B	Constraint
Material P (kg)	6	3	120
Material Q (kg)	2	4	100
Profit per production unit (US100$)	30	40	

Next, based on the above table,
Let x = Number of production units of product A
y = Number of production units of product B
z = Maximum value of an objective function
The problem can be represented as shown below.

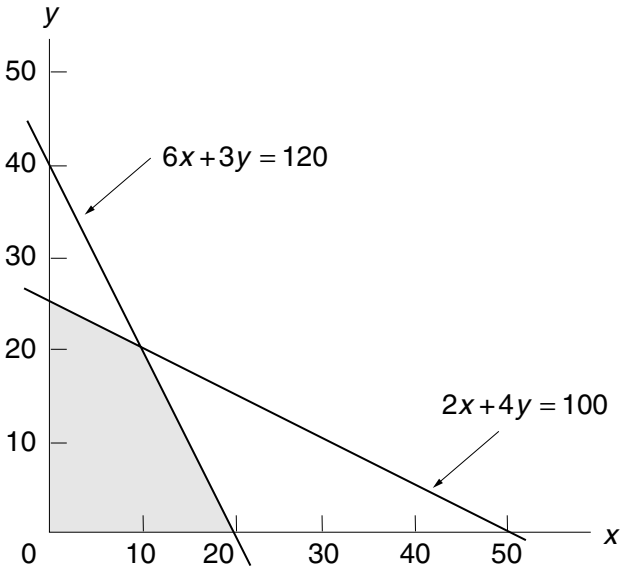
[Constraints]
Material P : $6x + 3y \leqq 120$
Material Q : $2x + 4y \leqq 100$
$x \geqq 0, y \geqq 0$ (x, y are non-zero integers: non-negative)

[Objective function]
$z=30x + 40y$ z is to be maximized
To represent a problem by mathematical formulas as shown above is called the "formulation of a problem". This is the first step in solving a linear programming problem.
In this question, the optimal solution (x,y), production units of product A,B respectively that maximizes the objective function, is to be obtained with the above constraints.

(2) Graphical Solution of Linear Programming Models

Graphs are frequently used to solve linear programming problems. Here, the procedure of solving a two-variable linear programming problem is explained below.
In the example (1) first, constraints are to be represented by a graph as follows (Figure 4-1-13).

Figure 4-1-13 Constraints graph



The shaded area in the Figure 4-1-13 is the region that satisfies the following two constraints.

Material P : $6x+3y \leqq 120$ ($y \leqq -2x+40$)

Material Q : $2x+4y \leqq 120$ ($y \leqq 0.5x+25$)

$x \geqq 0$, $y \geqq 0$

Therefore, the solution of this linear programming problem exists within the shaded area.

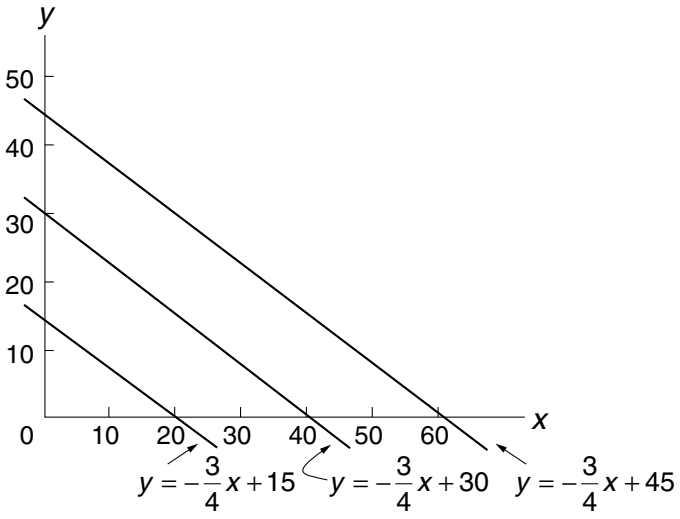
Next, the objective function is to be plotted.

The objective function of the above example is $z=30x+40y$. By transforming this equation so that it becomes an equation for y,

$$y = -\frac{3}{4}x + \frac{z}{40}$$

This equation can be represented as a straight line whose gradient is -0.75, with the horizontal axis x and the vertical axis y. Some concrete sample lines drawn by assigning some concrete values to z/40, are as follows.

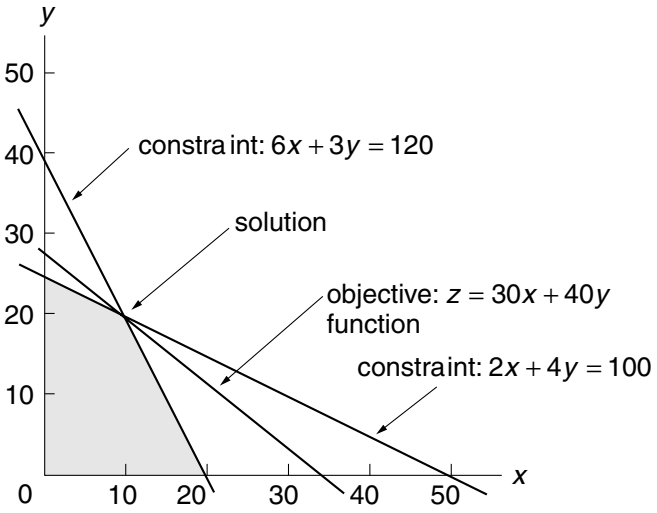
Figure 4-1-14 Graph of objective function



As you can see, the lines in this graph are the result of moving the line (gradient=0.75) in parallel. The value of the z function becomes larger for larger values of y when x=0.

Therefore, among the lines whose gradient is -0.75 and which intersects with the shaded area in the Figure 4-1-13 (Constraints chart), the line that maximizes the y value given that x=0, is the one to be obtained. (Figure 4-1-15)

Figure 4-1-15 Solution to linear programming problem(1)



In case of the graph in the Figure 4-1-15, the intersection point of the two constraints for material P and material Q gives the solution.

The solution of this linear programming problem can be obtained as the solution of the simultaneous equations that came from the constraints.

$$\begin{cases} 6x + 3y = 120 \\ 2x + 4y = 100 \end{cases}$$

By solving these simultaneous equations,

x=10, y=20

are obtained. That is, in this problem, the following maximizes the total profit.

Production of 10 units of product A

Production of 20 units of product B

The maximum total profit can be also obtained by substituting x, y values for the objective function z.

$z=30\times 10 + 40\times 20 = 1{,}100$ (unit: US 100$)

That is, the maximum total profit 1,100 (unit: US 100$) can be obtained by producing 10 units of product A and 20 units of product B.

(3) LP Theorem

In the previous example, the optimal solution that maximizes the objective function z was the intersection of two constraints. However, this is not always true.

For example, consider the situation where the profit per production unit for product A is now 10 (unit: US 100$) while that for product B is still 40 (unit: US 100$).

Let us formulate this problem first. Here, since there are no changes in the constraints, they are as follows.

[Constraints]

Material P : $6x + 3y \leqq 120$

Material Q : $2x + 4y \leqq 100$

$x \geqq 0$, $y \geqq 0$ (x, y are non-zero integers: non-negative)

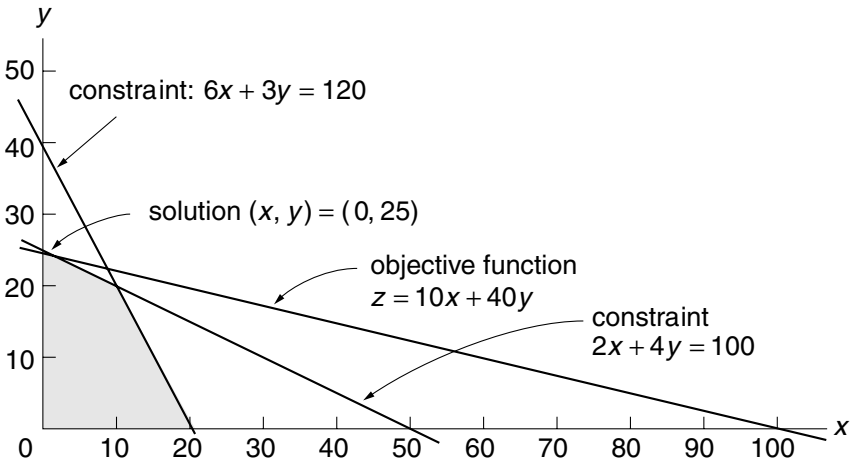
Then we need to modify the objective function as shown below, since the profit per production unit for product A has been changed.

[Objective function]

$z=10x + 40y \quad \left(y = -\frac{1}{4}x + \frac{z}{40} \right) \rightarrow$ z is to be maximized

The resulting graph is shown in the Figure 4-1-16.

Figure 4-1-16 Solution to linear programming problem(2)



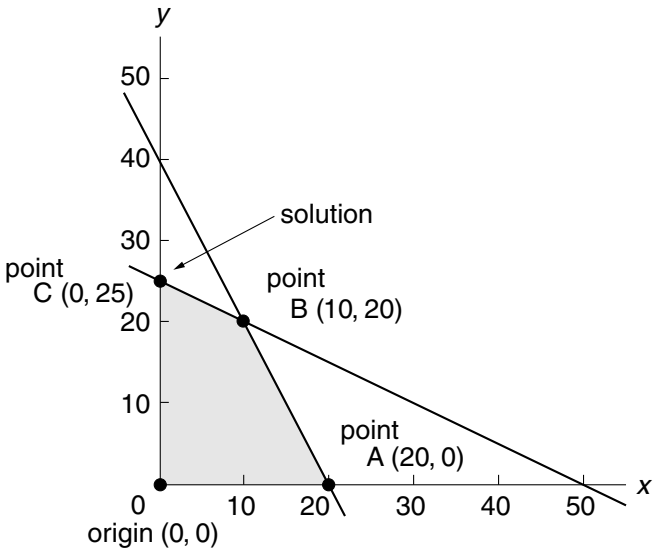
Here, the gradient -0.25 of the objective function is smaller and gentler than the gradient of the two constraints (-2 and -0.5).
Therefore, in this case, the solution is not the intersection point of the two constraints, but it is the point where the line representing the constraint for material Q and the y-axis intersect, and also which is on the line representing the objective function.
Thus the solution is x=0, y=25. Then z=1,000 is obtained by assigning these values to the objective function z.
That is, producing only product B without producing product A brings the maximum total profit 1,000 (unit: 100 US$)

The **LP theorem** used for solving linear programming problems is as follows.
The common region that satisfies the given constrains (this is called the "feasible region") is either a convex polygon or a convex polyhedron (or cone). If there is no common region satisfying the given constraints, the problem cannot be solved.
The optimal solution that maximizes (or minimizes) the objective function exists at one of the apexes of the convex shape.

The solution that exists at the apex of the region that satisfies a number of given constraints is called the "**basic solution**". In case of a standard question, the origin of the coordinate axes is always within the common region.

Therefore, another possible way to reach the optimal solution is to find the apexes of the common region first, then to compare the objective function z values at every apex to obtain the optimal one.
Using the above graph, let us find out the coordinates of the apexes of the common region by obtaining the intersection of the two constraints, the intersections of each constraint and the x-axis, and the intersections of each constraint and the y-axis. (Figure 4-1-17)

Figure 4-1-17 Corner points (apexes) of the common region (feasible region)



In this example, the z value at every corner point is as follows.

origin O (0,0) : $z = 10 \times 0 + 40 \times 0 = 0$

point A (20,0) : $z = 10 \times 20 + 40 \times 0 = 200$

point B (10,20) : $z = 10 \times 10 + 40 \times 20 = 900$

point C (0,25) : $z = 10 \times 0 + 40 \times 25 = 1,000$

From above, we conclude that the objective function z has its maximum value 1,000 at the point C (x=0,y=25).

The procedure of graphically solving a linear programming problem is summarized below.

Formulate the problem (i.e. clarify the constraints and the objective function)

⇩

Create a graph of constraints

⇩

Simulate the objective function lines

⇩

Find the objective function line which is within the common region satisfying the constraints (the feasible region) and which maximizes (or minimizes) the y value given x=0

⇩

The apex within the common region which also is on the line that maximizes (or minimizes) the objective function, is the solution

However, this graphical solution is not always practical, in the sense that if the problem has more than two variables we have to perform computations to decrease the number of variables to two. This is because the graphical solution is only possible when the number of variables is two.

In such cases the Simplex method (or the Simplex algorithm) is used. Although that is not described here.

4.1.3 Scheduling

(1) Scheduling

Scheduling means project planning and control. It includes such things as 1) creating schedules with given sequences of activities, 2) controlling schedules so that activities are proceeding as scheduled, 3) controlling delayed activities so that the difference between the original and the current schedules are

minimized as much as possible.
The scheduling creation and operational management of a large scale construction plan or a system development plan have been performed by bar graphs and Gantt charts. In addition to these, it is also possible to use the PERT (Program Evaluation and Review Technique) and CPM (Critical Path Method). PERT is a method used to make and manage schedules so that the total project duration becomes as short as possible, and CPM is a method which shortens the whole schedule while minimizing the cost increase.

(2) PERT Network

①What is a PERT network

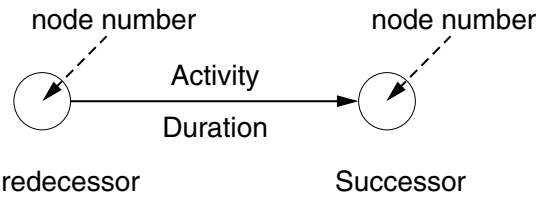
In order to manage a project systematically, the project should first be divided into activities, then a schedule should be created for each activity. Consider an example of a certain system development project. This project can be divided into the following activities.

List of Activities

Activity	Activity Description	Predecessor	Duration (days)
A	System Design		25
B	Program Design	A	40
C	Hardware Selection	A	20
D	Programming	B	50
E	Test Design	B,C	30
F	System Test	D,E	20

A job or an activity represents a task in a PERT network.
A predecessor is the activity that must be completed before a subsequent activity can begin and a successor is the activity that can not be started until previous activities are finished. In the above list, it is easy to understand the relationship between an activity and its required predecessors, but difficult to understand the relationships among activities of the whole project. In order to understand the whole project at a glance, the PERT method uses a **PERT network** as depicted graphically.
A PERT network is also called an **arrow diagram**. Each activity which constitutes a project is expressed by an arrow and is given a name. An activity name is placed above the arrow and a duration for the activity is placed under the arrow. A connection point (circle) is given to the both ends of the arrow and a number is put in the circle. In a PERT network, this connection point is called a **node** or an **event**. There must be relationship between the predecessor node number and the successor node number.
Predecessor node number < Successor node number
Thus a PERT network shows the whole project in the network with arrows representing the activities which constitute a project.

Figure 4-1-18 PERT network notation

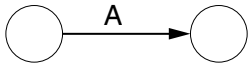


②Creation procedure of PERT network

The procedure of creating a PERT network is as follows.
1. First, divide a project into several activities and clarify all the precedence relationships (predecessor activity and successor activity) i.e. the order in which activities must be performed. The result of this step for the system development example was given before.
2. Identify the starting activity from the given activity list. Then draw two circles representing its starting point and completion point, and connect them with an arrow. A PERT network begins with a single node, and is finished with another single node.

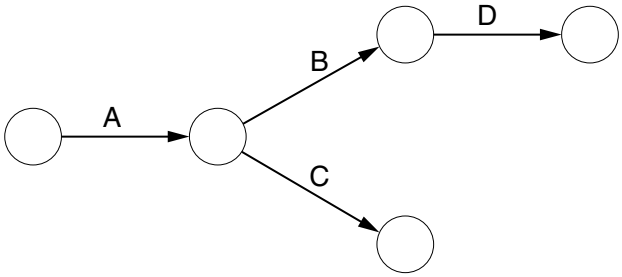
Here the activity A (system design) without predecessor activities turns into the initial activity.

Figure 4-1-19 First step



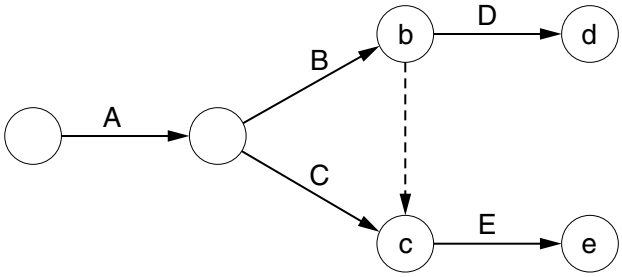
3. After that, connect activities based on their predecessor / successor relationships until the whole project network is completed. That is, if two activities have this relationship, the completion point of the predecessor activity should be the starting point of the successor activity. In the system development example, both activity B (Program design) and activity C (Hardware selection) are the successors of activity A. And activity D (Programming) is the successor of activity B. Therefore they appear in the PERT network as follows.

Figure 4-1-20 Activities A,B,C and D



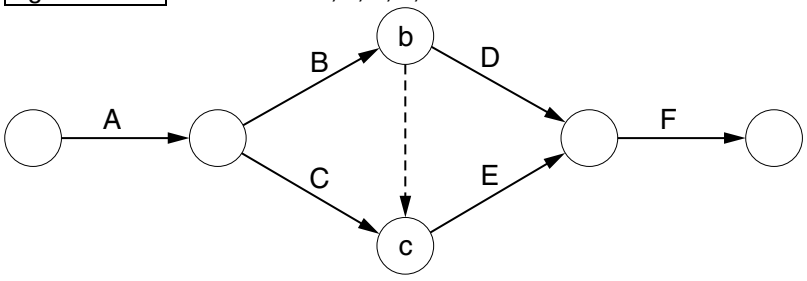
Activity E (Test design) is the successor of activities B and C. To represent this, define a non-existent activity with zero time and zero cost represented by a dotted-line from node b to node c. This activity is called a **dummy activity** which is used only to show the precedence relationship between a number of activities.

Figure 4-1-21 Activities A,B,C,D and E



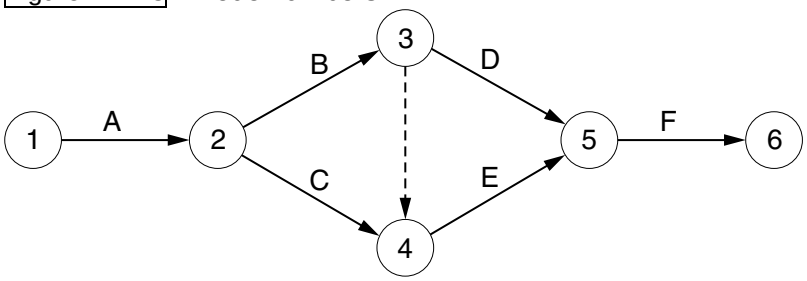
Then combine node d and node e to form a new node that is the starting point of activity F (System test) and the completion point of activities D and E. This is because activities D and E are the predecessors of activity F. Since activity F has no successor, activity F becomes the last activity.

Figure 4-1-22 Activities A,B,C,D,E and F



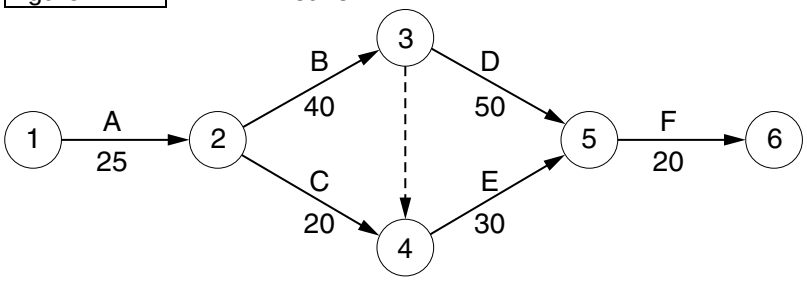
4. After you connect all nodes with arrows, fill in the nodes with numbers. Note that the numbers of predecessor nodes must be smaller than those of successors.
When this numbering procedure is done by using computers, it is called topological ordering.
An activity can be also represented as activity (i, j) with node numbers i and j in both ends. For example, activity D can be represented as activity (3,5)

Figure 4-1-23 Node numbers



5. Complete the PERT network by specifying the duration under arrows for each activity. (Figure 4-1-4)

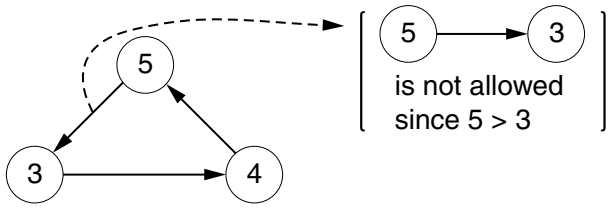
Figure 4-1-24 PERT network



③Creation rules of a PERT network

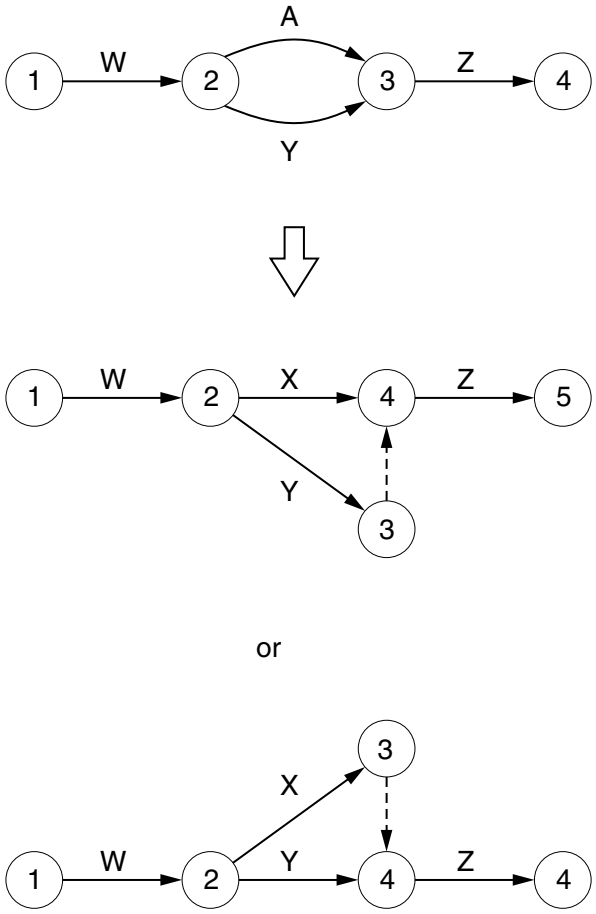
In addition to the rules shown above, there are some additional rules in the creation of a PERT network. One of them is the rule that a loop (circulating parts) in the PERT network is prohibited.

Figure 4-1-25 An example of a loop



When there are multiple activities between two nodes, this fact should be expressed by using a dummy activity. For example, when the successor activities of activity W are activities X and Y, and activities X and Y are the predecessor activities of activity Z, this should be represented by the following notation.

Figure 4-1-26 Dummy activity usage



(3) PERT Time Computation

After a network diagram has been developed, the next step is to calculate the total duration for the project and to distinguish between activities with delay allowance and without delay allowance. This is called **PERT/TIME computation**.

At first, the following terms will be defined for understanding PERT/TIME computation.

Duration of each activity (hours)

Time of each node

Time of each activity

Delay allowance of each activity

The method of PERT calculation will be explained below using the example of system development from section (2).

①Calculation of Node Times

First, consider the arrival time (or departure time) at each node. Here two different times at each node are considered.

Earliest node time (ET_i) : The earliest possible activity start time at node i. It is "the time at which an activity can be started at the node if the predecessor activities are started as early as possible".

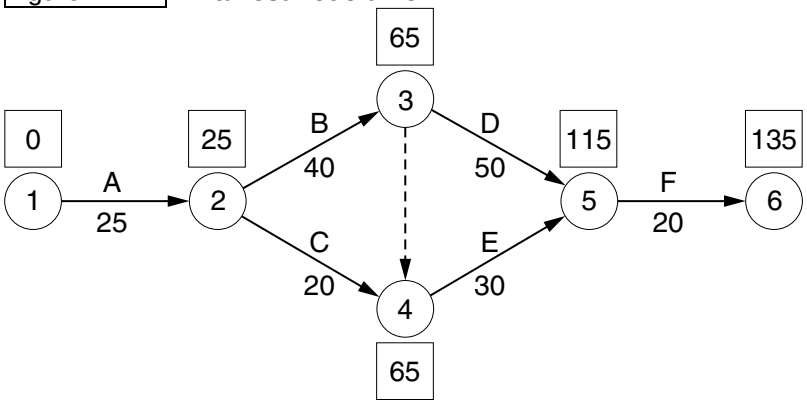
Latest node time (LT_i) : The latest possible time to start an activity at node i without causing delay to the completion time of the project. It is "the last time at which an activity can be started at the node without delaying the completion of the project beyond its earliest possible completion time".

a. Forward computation

Forward computation calculates the earliest node time at each node and duration of the whole project, starting with the initial node and working forward in minimum time toward the final node.

In this calculation method, the earliest node time for the initial node (node number 1) is 0 and then that of each node is calculated as the cumulative durations of all its predecessor activities. Note that this calculation is based on the relationships between activities.
For example, the earliest node times are calculated as shown below for the case of the system development in section (2).
Node1: 0
Node2: 0 + 25 = 25
Node3: 25 + 40 =65
Node4: Although the shorter one, 25+20=45, is preferable, the earliest node time here is 65 because activity E cannot start until both activities B and C are completed.
Note5: By comparing 65+50=115 and 65+30=95, the later one i.e.115 is the earliest node time here. This is because activity F cannot start until the predecessor activities D and E are completed,
Node6: 115+20=135
Filling in the earliest start time at each node in the Figure 4-1-24 (PERT network) leads to the Figure 4-1-27.

Figure 4-1-27 Earliest node time

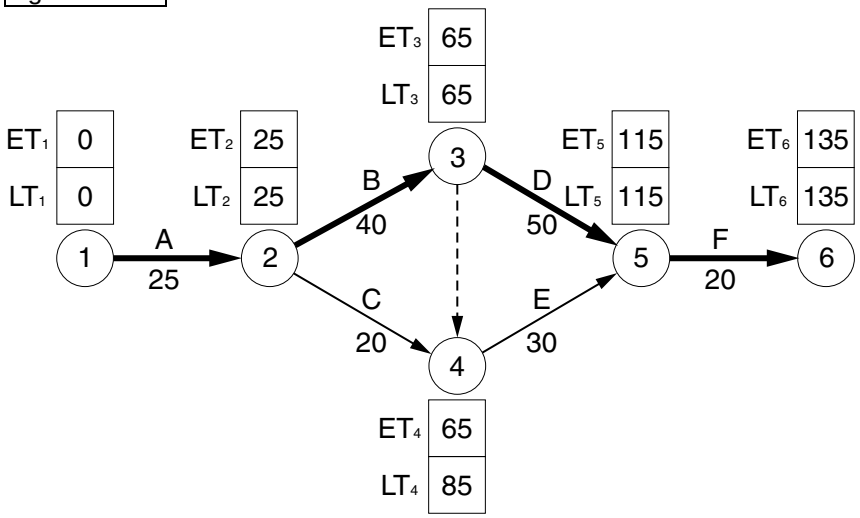


As shown above, the minimum time required for the project completion, that is the optimum project duration, is 135.

b. Backward computation

Backward computation calculates the latest start time of each activity without extending the optimum duration of the whole project.
In this calculation method, unlike the forward computation that starts computation with the initial node, the latest start times are calculated by a backward pass through the project. That is, the latest start times are calculated for each activity, starting with the final node and moving backward through the network toward the initial node. The final node of the project should be the optimum duration. And the latest node time at each node is computed by subtracting the predecessor activity durations from the earliest node time of the last node.
For example, the latest node times are calculated as shown below for the case of the system development in section (2).
Node6: 135
Node5: 135-20=115
Node4: 115-30=85
Node3: 115-50=65
Node2: By comparing 65-40=25 with 85-20=65, the smaller, i.e. 25 is taken. Because if 65 is selected, the start of activity B will be delayed. As a result, the time required for the project will be increased.
Node1: 25-25=0
The result of placing the latest node time LT_i below the earliest node time ET_i for each node in Figure 4-1-27 appears in Figure 4-1-28.

Figure 4-1-28 Earliest/Latest node times



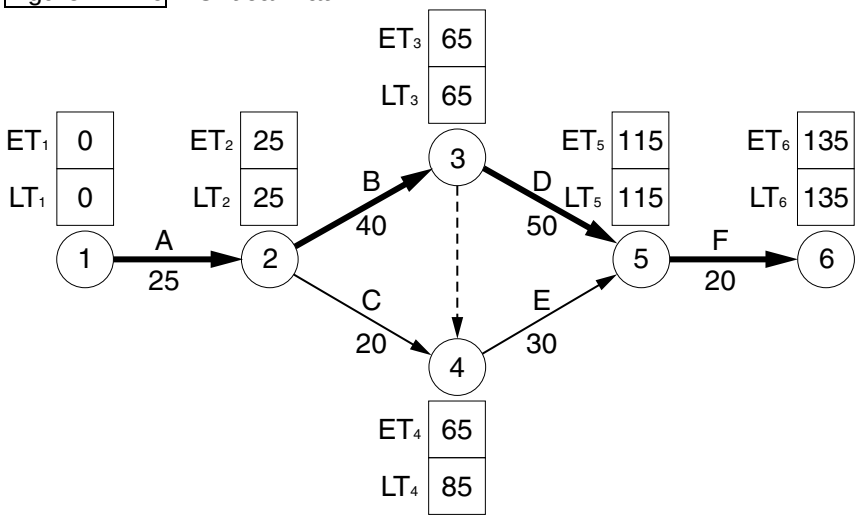
Here, [latest node time] - [earliest node time] at each node is called **slack** (allowance). It indicates delay allowance and how much delay in the latest start time can be tolerated without delaying the project's completion.

②Critical Path Method

A **critical path** is a path through the network where all of the nodes have no delay allowance, or [latest node time]= [earliest node time]. Since a critical path is a path with zero slack, a delay in the time of any critical path activity results in a delay in the project completion time. Because of their importance for completing the project, the importance of managing those activities should be emphasized. The method to identify important activities (activities on the critical path) as above, is called **CPM** (Critical Path Method).

In the example of system development in section (2), the path of A→B→D→F (dark arrow) is the critical path.

Figure 4-1-29 Critical Path



③Calculation of Start/Finish Times

The calculation of start/finish times for an activity will be explained here, in addition to such times for a node already explained in section (1). The starting time and the end time for an activity are calculated by using the node times calculated before.

There are 4 kinds of activity times for each activity.

Earliest start time

Earliest finish time

Latest start time

Latest finish time

a. Earliest start time (ES_{ij})

The **earliest start time** (ES_{ij}) is the earliest start time for activity (i,j), which is equal to the earliest node time at node i.

Earliest start time (ES_{ij})= Earliest node time (ET_i)

The earliest start time for each activity in the example of system development from section (2) appears in the following table

Activities (i,j)	Earliest start time (ES_{ij})
A(1,2)	0
B(2,3)	25
C(2,4)	25
D(3,5)	65
E(4,5)	65
F(5,6)	115

b. Earliest finish time (EF_{ij})

The **earliest finish time** (EF_{ij}) is the earliest finish time for activity (i,j) when the activity (i,j) is started at the earliest start time. The earliest finish time for an activity (i,j) is calculated by adding the activity duration (D_{ij}) to the earliest start time for the activity.

Earliest finish time (EF_{ij})= Earliest start time (ES_{ij}) + Activity duration (D_{ij})

The earliest finish time for each activity in the example of system development from section (2) is shown below.

Activities (i,j)	Earliest start time (ES_{ij})	Earliest finish time (EF_{ij})
A(1,2)	0	25
B(2,3)	25	65
C(2,4)	25	45
D(3,5)	65	115
E(4,5)	65	95
F(5,6)	115	135

c. Latest finish time (LF_{ij})

The **latest finish time** (LF_{ij}) means the latest time at which the activity (i,j) can be started without delaying the completion of the project beyond its earliest possible time, i.e. without delaying the times in a critical path. The latest finish time for an activity (i,j) corresponds to the latest node time of its second node j.

Latest finish time (LF_{ij})= Latest node time (LT_j)

The latest finish time for each activity in the example of system development from section (2) appears in the following table.

Activities (i,j)	Latest finish time (LF_{ij})
A(1,2)	25
B(2,3)	65
C(2,4)	85
D(3,5)	115
E(4,5)	115
F(5,6)	135

d. Latest start time (LS_{ij})

The **latest start time** (LS_{ij}) is the start time to finish the activity (i,j) by the latest finish time. The latest start time for an activity (i,j) is calculated by subtracting activity duration (D_{ij}) from its latest finish time.

Latest start time (LS_{ij})=Latest finish time (LF_{ij})-Activity duration (D_{ij})

Latest start time for each activity in the example of system development from section (2) is shown in the following table.

Activities (i,j)	Latest start time (LS_{ij})	Latest finish time (LF_{ij})
A(1,2)	0	25
B(2,3)	25	65
C(2,4)	65	85
D(3,5)	65	115
E(4,5)	85	115
F(5,6)	115	135

(4) Task floats

The task float in activity (i,j) is the allowance between the finish time of the activity (i,j) and the start time of the successor activity in the range which does not delay the activity schedule of a critical path. There are two types of task floats: total float and free float.

a. Total float (TF_{ij})

The **total float** (TF_{ij}) is the maximum delay allowance that can be consumed by the activity (i,j) itself in the range which does not delay the activity schedule of a critical path, under the condition that other activities do not consume their delay allowance at all.
Total float for an activity is calculated by subtracting its earliest start time from its latest start time (or earliest finish time from latest finish time).

Total Float (TF_{ij}) = Latest start time (LS_{ij}) - Earliest start time (ES_{ij}) = Latest finish time (LF_{ij}) - Earliest finish time (EF_{ij})

Total float for each activity in the example of system development appears in the following table.

Activities (i,j)	Latest start time (LS_{ij})	Earliest start time (ES_{ij})	Total Float (TF_{ij})
A(1,2)	0	0	0
B(2,3)	25	25	0
C(2,4)	65	25	40
D(3,5)	65	65	0
E(4,5)	85	65	20
F(5,6)	115	115	0

b. Free float (FF_{ij})

The **free float** (FF_{ij}) is the maximum delay allowance that can be consumed by the activity (i,j) itself in the range which does not delay the activity schedule of a critical path, regardless of whether other activities consume their delay allowance or not. In other words, free float is the delay allowance that activity (i,j) can consume and that does not affect the allowance of any successor activity.
Free float is calculated by subtracting earliest finish time (EF_{ij}) from earliest node time (ET_j)

Free float (FF_{ij})= Earliest node time (ET_j)－ Earliest finish time (EF_{ij})

Free float for each activity in the example of system development from section (2) is shown in the following table.

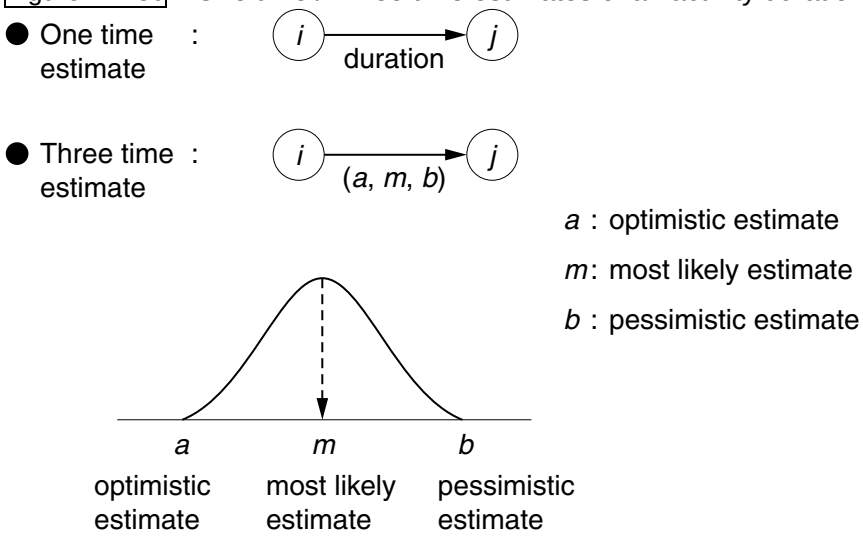
Activities (i,j)	Earliest node time (ET_j)	Earliest finish time (EF_{ij})	Free Float (FF_{ij})
A(1,2)	25	25	0
B(2,3)	65	65	0
C(2,4)	65	45	20
D(3,5)	115	115	0
E(4,5)	115	95	20
F(5,6)	135	135	0

As mentioned before, in the case of total float, consumption of total float in an activity (i,j) removes delay allowances from successor activities. Those successor activities have to start at latest time, having no more delay allowances.
On the other hand, in the case of free float ,even if the free float has been consumed in activity (i,j) it has no influence on the successor activities. Successor activities still can be started at the earliest scheduled time.
The difference between total float and free float, and the following relationship between them should be clearly understood.
Total Float >= Free Float

(4) Estimate of activity time

In the PERT/TIME calculation explained in (3), we used a constant as an estimate for the duration of activity (i,j). This estimate method is called “one time estimate.” However, in fact, it is difficult to give an exact estimate for activity duration by using this estimate method because of its uncertainties. Therefore, there is a more practical method that estimates the activity duration by doing statistical calculation with three values obtained from different points of view. This is called **three time estimates**. (Figure 4-1-30)

Figure 4-1-30 One time / Three time estimates of an activity duration



In the three time estimates, the expected value (or mean: μ) and variance (σ^2) of activity duration are calculated by the following formulas using a: Optimistic estimate, m: Most likely estimate and b: Pessimistic estimate.

expected value (or mean: μ)= (a+4m+b)/6

variance (σ^2) = ((b-a)/6)^2

Here, an expected value is a mean and a variance indicates the degree of variation in data. By calculating these values, the total duration of a project will be calculated as a random variable. Note that this method is not suitable to apply to a project that has many indefinite elements.

(5) Scheduling and Project Costing

The schedule of the whole project is closely related to the schedules of a critical path. Managing activities on a critical path should receive the top priority in order not to delay the whole schedule.

In order to shorten the whole project schedule, what is necessary is just to shorten the activities on a critical path. However, if activity duration is shortened the cost will increase. As a method of solving this trade-off, the concept of **cost slope** is used. Let the [normal time] = normal duration for activity, [normal cost] = normal cost for activity, [crash time] = crash duration for activity, [crash cost] = crash cost for activity, the cost slope will be calculated by the following formula.

cost slope = (crash cost - normal cost)/(normal time - crash time) = additional cost / reduced time

An optimum project schedule is to be planned using the cost slope, considering the reduced times and additional costs for the whole project.

For example, in the project example of system development from section (2), suppose the days that can be reduced and the additional cost for each activity are defined as follows.

Activity	CP	Normal Time for activity (days)	Days that can be reduced	Additional cost	Cost slope
A	*	25	-	-	-
B	*	40	10	40	4
C		20	5	5	1
D	*	50	25	125	5
E		30	5	15	3
F	*	20	-	-	-

Here, assume that the duration can not be shortened for activities A and F, even though additional costs are allocated.

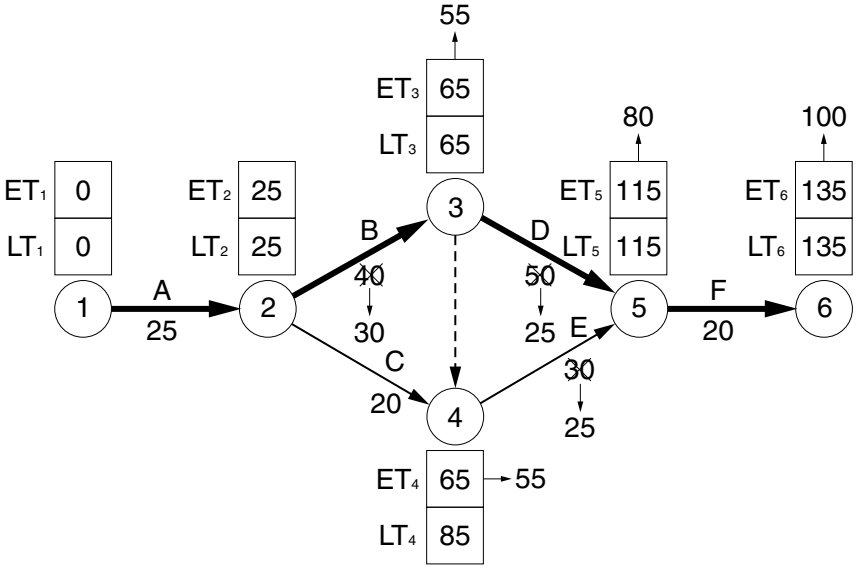
Case 1:

Consider how we can reduce the necessary duration for the whole project by 5 days. In this case, although the additional cost for activity C is the minimum of all (its cost slope being the smallest), the whole duration can not be reduced by reducing the duration of this activity, since this activity is not in the critical path. Instead, the expected reduction can be achieved by reducing the duration of activity B by 5 days. Because this activity's cost slope is the minimum among the activities in the critical path.

Case 2:

Consider how we can shorten the necessary duration for the whole project by 35 days. In this case, although the answer seems simple, to shorten activity B by 10days and activity D by 25days, it is not. Because this causes a change in the critical path from A→B→D→F to A→B→(Dummy) →E→F, this by itself cannot achieve the goal (to shorten the necessary duration of the whole project by 35 days). Therefore, in addition to this, to shorten activity E by 5 days is also necessary.

Figure 4-1-31 Reduction of total duration by 35 days



4.1.4 Queuing Theory

(1) Queuing Theory

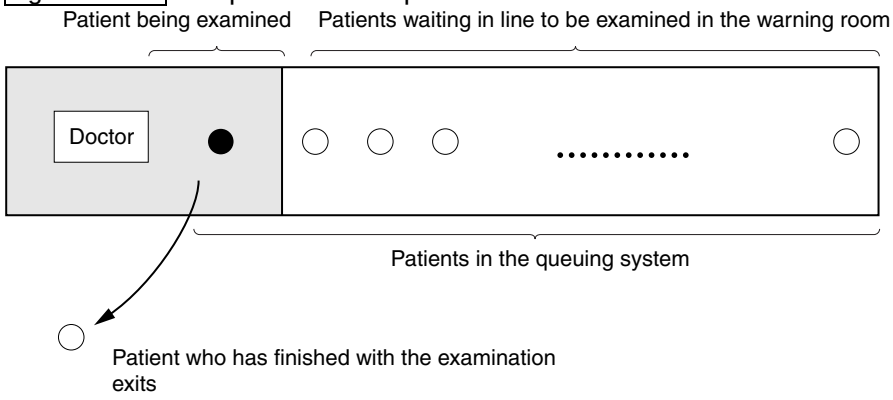
When a customer requiring service arrives at the service area, if there is any available server, the customer can receive service immediately. If not, the customer has to wait in a **queue** until it is his/her turn to receive service.

The formation of lines of people can be widely observed in the real world, such as cash dispensers in a bank, ticketing machines at a train station and checkout lines in a supermarket.

Another example is medical care in a hospital. A patient may have to wait in a waiting room until it is

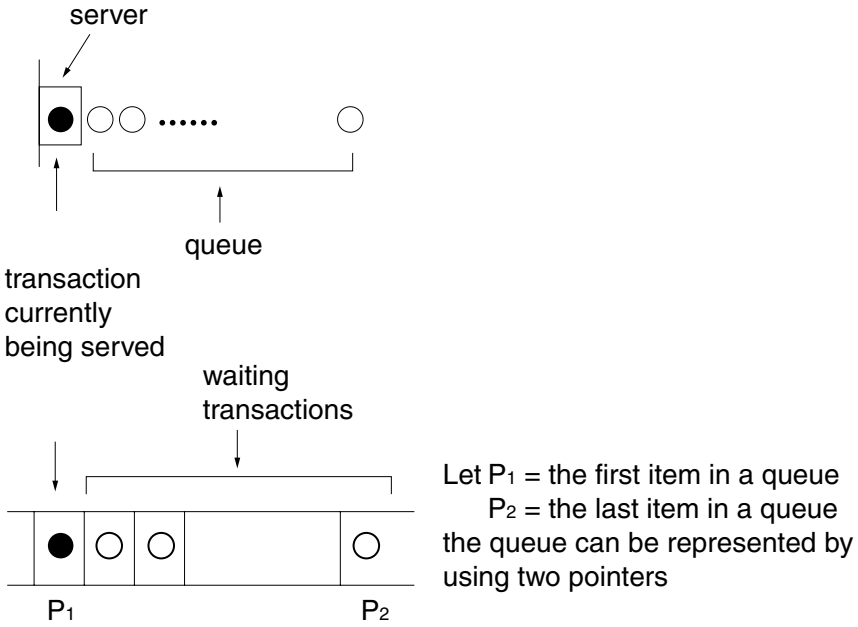
his/her turn to be treated by a doctor. If there aren't any other waiting patients, he or she can receive medical care immediately. If the waiting room is crowded with waiting patients, he or she has to wait for some time.
See the example shown in the Figure 4-1-32. How long will a patient have to wait to be examined by a doctor? This question is solved by computation giving considerations to such factors as how long it takes for a doctor to see one patient, how many waiting patients there are in the waiting room etc.
In case the waiting time is too long, the hospital should take some actions such as increasing the number of doctors. However, the hospital may not be sure whether increasing the number of doctors is the most effective solution to this problem. Because the time needed to examine one patient may vary depending on the patient's symptoms, and the congestion in the waiting room may change from time to time.
Queuing theory can be used to build models that describe a waiting-line situation. It provides vital information needed when solving problems like "how long a patient has to wait before getting a checkup", "How long the queue is?" etc.

Figure 4-1-32 A queue in a hospital



In the computer data structure, this waiting line is called a "queue". A key feature of a queue is that items in a queue are processed one after another such that the first element stored is the first one retrieved. This data processing is called FIFO (First-In First-Out).
This data structure can be diagrammatically represented as the Figure 4-1-33.

Figure 4-1-33 Data structure in a queue



The following shows how to manipulate items in this data structure.
1. Serve the "P_1"-th transaction
2. When the above is completed, $P_1+1 \rightarrow P_1$

3. When a new transaction arrives, $P_2+1 \rightarrow P_2$
4. The new transaction is "P_2"-th

A queuing model is composed of the following.

・Number of arrivals of transactions (customers) per unit time
・Service time per transaction
・Number of parallel servers in the queuing system
・Possible number of transactions in the system

Naturally, queue length (i.e. number of customers waiting for service) varies depending on the interarrival time of transactions (or number of arriving transactions per unit of time), service time and number of servers.

Note that the following assumptions are made in a queuing system.

・No transactions in a queue are cancelled.
・No arriving transactions are rejected, regardless of queue length

Under the above assumptions, processing in a queuing system is FIFO-based and "non real-time" (that is, a transaction has to wait until service is available)

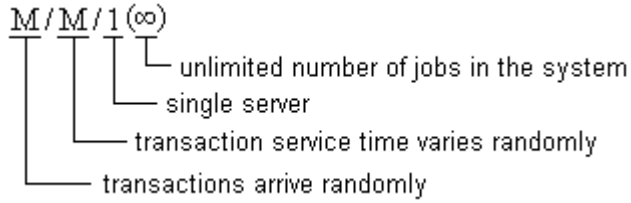
(2) Kendall Notation

Queuing models are conventionally represented using the **Kendall** notation as follows

A/B/C(D)

A represents the distribution of time between arrivals, B the distribution of service times, and C the number of parallel servers. D represents the possible number of jobs in the system.

The most fundamental queuing model "M/M/1(∞)" designates a single server queuing model with a Poisson arrival rate and exponential service time distribution, and with an unlimited number of jobs in the system.



Concerning random events, the following theorems are known.

Number of occurrences per unit of time follows the Poisson distribution.

In the Poisson distribution, mean=variance

Time between occurrences follows the exponential distribution.

In the exponential distribution, $(\text{variance})=(\text{mean})^2$

Therefore, when transactions arrive randomly,

Number of arriving transactions per unit of time follows the Poisson distribution

Time between transaction arrivals (interarrival time) follows the exponential distribution

(3) Mean Arrival Rate (λ)

The **mean arrival rate** indicates "the expected number of transaction arrivals per unit of time", represented by " λ "(lambda).

Mean interarrival time can be obtained by using the following expression.

Mean interarrival time = 1 / Mean arrival rate = 1 / λ

For example, if there are 4 expected arrivals per minute,

Mean arrival rate(λ) = 4 transaction arrivals per minute

Mean interarrival time = 1/4 minute per transaction arrival = 15 sec per transaction arrival

That is, on average a transaction arrives every 15 seconds.

(4) Mean Service Rate (μ)

The **mean service rate** means "the expected number of transactions completing service per unit of time",

represented by " μ "(mu).
Mean service time can be obtained by using the following expression.

Mean service time = 1 / Mean service rate = 1 / μ

For example, if service to 5 transactions can be completed every minute,

Mean service rate(λ) = 5 transactions per minute

Mean service time = 1/5 minute per transaction = 12 sec per transaction

That is, on average it takes 12 seconds to serve one transaction.
If $\mu > \lambda$ or $1/\mu < 1/\lambda$ is true in a queuing system, the system is said to be in a "steady-state condition".

(5) Traffic Intensity (ρ)

The **traffic intensity** represents "the expected fraction of time individual servers are busy", denoted by " ρ "(rho). This can be obtained by using the following expression.

Traffic intensity (ρ) = Mean arrival rate / Mean service rate = λ / μ = $1/\mu$ / $1/\lambda$
= Mean service time / Mean interarrival time < 1

For example, if there are 4 expected transaction arrivals per minute, and service to 5 transactions can be completed per minute,

Traffic intensity (ρ) = 4/5 = 0.8 → 80%
or 12(sec) / 15(sec) = 0.8 → 80%

This means each server is busy for 80% of the time.
Traffic intensity should be less than 100%. Because when it is equal to or more than 100%, there are always waiting transactions in the queue. Therefore, in such a case, some measures (e.g. to assign some additional servers) should be taken so as to make it less than 100%.

(6) Mean Number of Transactions in the System (L)

The **mean number of transactions in the system** is the "expected number of transactions in queuing system, both waiting for service and currently being processed", denoted by L. We can calculate this from traffic intensity ρ using the next equation.

Mean number of transactions in the system (L) = traffic intensity / (1- traffic intensity)
= ρ / (1- ρ)

For example, if 4 transactions occur per minute and 5 transactions can receive service per minute,

Mean number of transactions in the system (L) = 0.8 / (1-0.8) = 4

This indicates that on average 4 transactions are in queuing system, waiting for or receiving service.

(7) Mean Time Transaction Spends in the System (W)

The **mean time transaction spends in the system** is the "expected waiting time in system (includes service time)", denoted by W. This can be calculated from mean number of transactions in the system (L) and mean arrival rate (λ) (or mean interarrival time ($1/\lambda$)), using the following equation.

Mean time transaction spends in the system (W)
= mean number of transactions in the system × (1 / mean arrival rate) = L × 1 / λ
= mean number of transactions in the system × mean interarrival time

For example, if 4 transactions occur per minute and 5 transactions can receive service per minute,

Mean time transaction spends in the system (W) = 4 × 1/4 = 1 minute

This means the time between the transaction arrival and the completion of the service is one minute on average.
There exists another equation for calculating the mean time transaction spends in the system (W). The next equation calculates W as the sum of mean service time ($1/\mu$) and mean time transaction spends in the queue (W_q). (The mean time transaction spends in the queue (W_q) will be introduced later.)

Mean time transaction spends in the system (W)
= mean time transaction spends in the queue + mean service time = $W_q + 1/\mu$

(8) Mean Number of Transactions in Queue (L_q)

The **mean number of transactions in queue** is the "expected queue length (excluding transactions being served)", denoted by L_q.
This is calculated by the following equation, using the mean number of transactions in the system (L) and traffic intensity (ρ).

mean number of transactions in queue (L_q)
= mean number of transactions in the system (L) × traffic intensity (ρ)
= (traffic intensity)^2 / (1 - traffic intensity) = $\rho^2 / (1-\rho)$

For example, if 4 transactions occur per minute and 5 transactions can receive service per minute,
mean number of transactions in queue (L_q) = 4 × 0.8 = 3.2
or $0.8^2 / (1 - 0.8) = 3.2$
This indicates that on average 3.2 transactions are in the queue, waiting for service.

(9) Mean Time Transaction Spends in Queue (W_q)

The **mean time transaction spends in queue** is the "expected waiting time in queue (excluding service time)", denoted by Wq.
This can be calculated from mean number of transactions in queue (L_q) and mean arrival rate (λ) (or mean interarrival time ($1/\lambda$)), using the following equation.

Mean time transaction spends in queue (Wq)
= mean number of transactions in queue × (1 / mean arrival rate) = $L_q \times 1 / \lambda$
= mean number of transactions in queue × mean interarrival time

For example, if 4 transactions occur per minute and 5 transactions can receive service per minute,
Mean time transaction spends in the queue (Wq)
= 3.2[transactions] × (1[minute] / 4[transactions]) = 0.8[mimutes] = 48[seconds]
This indicates that it takes 48 seconds on average for an arrived transaction until it starts receiving service.

As shown below, formulas of queuing theory explained so far are related to each other.

Calculate ρ from λ and μ	$\rho = \lambda / \mu$
Calculate L from ρ	$L = \rho / (1-\rho)$
Calculate W from L	$W = L \times (1/\lambda)$
Calculate L_q from L	$L_q = L \times \rho$
Calculate W_q from L_q	$W_q = L_q \times (1/\lambda)$

Each of the above formulas can be applied either as it is, or after transformation(?). Select a formula that is the most suitable for the problem to be solved.

(10) Probability of Exactly n Transactions in the System (P_n)

The **probability of exactly *n* transactions in the system** is the probability that there are exactly *n* transactions in the system (including transactions being served). This is denoted by P_n.
This can be calculated from traffic intensity (ρ) by using the next equation.

Probability of exactly *n* transactions in the system (P_n)
= (mean service rate)^n × (1- mean service rate) = $\rho^n(1-\rho)$

For example, if 4 transactions occur per minute and 5 transactions can receive service per minute, the possibilities of exactly 0 / 1 / 2 transactions are

0 transaction i.e. n=0 $P_0 = 0.8^0(1-0.8) = 0.2$
1 transaction i.e. n=1 $P_1 = 0.8^1(1-0.8) = 0.16$
2 transactions i.e. n=2 $P_2 = 0.8^2(1-0.8) = 0.128$

The above result indicates that
probability of no transactions in the system = 20%
probability of exactly one transaction in the system = 16.0%
probability of exactly two transactions in the system = 12.8%

4.1.5 Inventory Control

(1) Inventory Control

①Necessity of Inventory Control

Once, it was said, "Stock is a business's treasure". But since the Great Depression following World War I, it has been commonly understood that "Stock is a business's tomb". Furthermore, during the early 1970s, at the time of an oil crisis, many businesses suffered from having excessive stocks and tried hard to optimize them.
Therefore, the management in a business wants to reduce stock as much as possible, and also wants to minimize the holding cost. However, if the amount of stock is not sufficient, the sales opportunities may be lost and marketing may be hindered. So, it is important for a business to give careful considerations on how they control their inventory in order not to cause any trouble for their everyday business.
For instance, it is a big problem for a PC shop to decide how many personal computers it should stock. When the stock is too big, there may be some dead stock, and storage space and insurance for them may become quite costly. On the other hand, when the stock is small, there may not be enough goods to fill the received orders, consequently the shop may miss their chance to earn profit.
In order to solve such problems, inventory control systems that can systematically manage the inventory have emerged. This has made it possible to identify the inventory level that minimizes loss and maximizes profit, by mathematically processing the indefinite factors (such as demand, time of delivery) which had complicated inventory control in the past.
Recently those systems have been developed into the so called “total inventory control systems” that cover procurements, transmission, production and sales processes utilizing information technology.

②Goal of Inventory Control

In terms of management science, the goal of inventory control is to "find out the management methods to minimize the total inventory cost under given conditions".

a. Total Inventory Cost

The **total inventory cost** includes all of the expenses that are related to the inventory control. It is roughly categorized into two types, the ordering cost (procurement cost) and the holding cost (storage cost).

●Ordering cost (procurement cost)

The **ordering cost** (also called the **procurement cost**) is the cost of placing an order. This includes the fixed cost needed for an order regardless of the order quantity, e.g. personnel expenses, communication expenses, insurance expenses and office management expenses.

Annual ordering cost	= ordering cost per order × annual number of orders
	= ordering cost per order × annual demand / amount ordered per order

●Holding cost (storage cost)

The **holding cost** (also called the **storage cost**) represents all costs associated with the storage of the inventory until it is sold or used. This varies depending on the stock amount. Included are warehousing expenses, personnel expenses, insurance expenses, utility charges etc.

Annual holding cost	= annual holding cost per stock item × average stock amount
	= annual holding cost per stock item × amount ordered per order / 2

Actually, the ordering cost includes the cost of the items being purchased, calculated by "unit price X amount". However, this is not taken into account in solving inventory control problems because the unit price is fixed in many cases, regardless of the amount ordered, order date etc.

b. Minimizing Total Inventory Cost

To minimize the total inventory cost, each of the following should be minimized since they affect the total inventory cost.

●Stock-out rate

The **stock-out rate** is the percentage of orders where stock-out occurs between the order and the delivery of goods (i.e. within the procurement period). If stock-out occurs only once out of 20 orders, the stock-out rate is 1/20 (=0.05). An example of a condition put on this figure is "the stock-out rate should be equal to or

less than 0.1". The antonym of this is service rate. The relationship between the two is as follows.

Stock-out rate + Service rate = 1

●Extra stock loss

The **extra stock loss** is the loss caused by extra stocks that result in additional costs for stocks not being purchased. An example of a condition put on this is "the extra stock loss should be minimized".

●Shortage cost

The **shortage cost** is the loss of not being able to sell an item when needed because of the stock-out. It means a loss resulting from losing sales opportunities because of no stock. An expected condition put on this is "the shortage cost should be minimized".

In other words, the goal of inventory control is "to obtain the ideal stock level that minimizes the total inventory cost under given constrains". The obtained value is called the optimal inventory.

(2) Economic Order Quantity

To have the optimal amount of stocks, we need to answer the question "When and how much do we order?". "When" is relevant to "order date" while "how much" is related to "order quantity". The ultimate goal of inventory control is to find out the order date and order quantity that minimize the total inventory cost.

Example

A PC shop operates 250 days in a year. It costs this shop 500 yen to hold a PC for one year and 1,250 yen to place an order. To simplify the problem, the following assumptions are made.

Once an order is placed, ordered items (PCs) are delivered to the shop immediately

The amount of PCs sold per day is always 2 units, regardless of what day it is

A new order for a fixed amount is placed when the shop is out of stock

Consider the optimal order timing and the optimal order quantity for this shop.

First, assume that the shop places an order once a year, with the order quantity 500 (= annual sales amount). In this case, annual average stock is 250, half of 500.

Annual ordering cost: 1,250 [yen per order]×1 [order] = 1,250 [yen]

Annual holding cost: 500 [yen per PC]×500 [PCs] × 1/2 = 125,000 [yen]

Annual total inventory cost = 126,250 [yen]

Next, assume that the shop places an order once a day, with the order quantity 2 (= daily sales amount). In this case, the shop places 250 orders per year and the annual average stock is 1.

Annual ordering cost: 1,250 [yen per order]×250 [order] = 312,500 [yen]

Annual holding cost: 500 [yen per PC]×2 [PCs] × 1/2 = 500 [yen]

Annual total inventory cost = 313,000 [yen]

①Relationship between ordering cost and holding cost

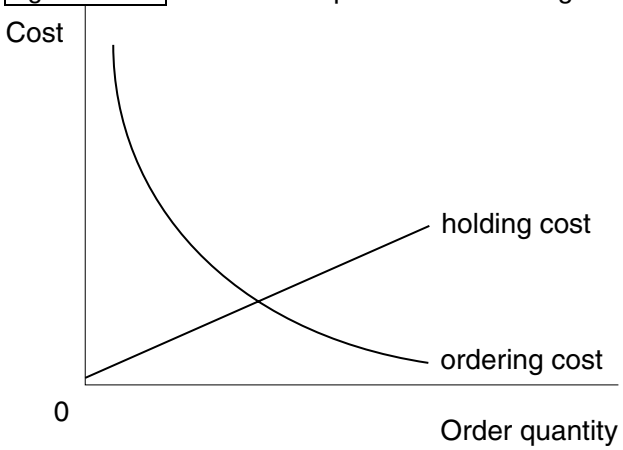
The above example is an extreme case. However, when the order quantity per order is big, ordering cost is small. On the other hand, when the order quantity per order is small, the number of orders as well as the ordering cost increase but the holding cost decreases.

Therefore, it is necessary to obtain the number of orders and the order quantity that minimizes the total inventory cost, i.e. the sum of ordering cost + holding cost.

Annual total inventory cost = annual ordering cost + annual holding cost = (annual sales amount / order quantity per order) ×ordering cost per order + (order quantity per order / 2) × annual holding cost per unit

The Figure 4-1-34 shows the ordering cost and holding cost for one year in a order quantity to cost chart.

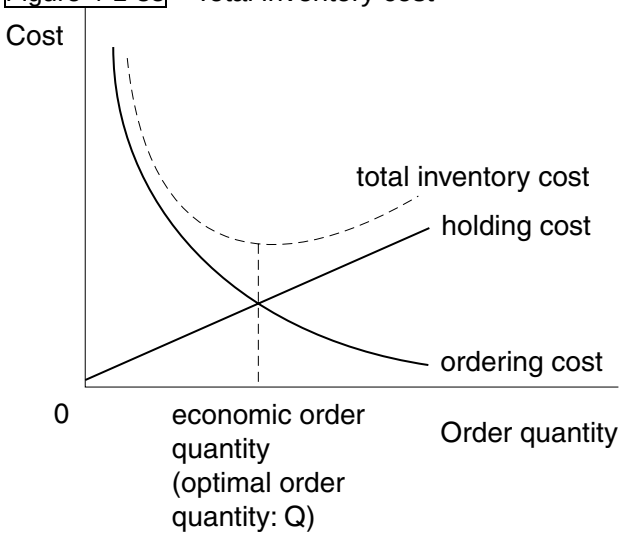
Figure 4-1-34 Relationship between ordering cost and holding cost



②Calculating Economic Order Quantity (EOQ)

Figure 4-1-35 includes a new curve representing the total inventory cost (= ordering cost + holding cost), added to the previous Figure 4-1-34.

Figure 4-1-35 Total inventory cost



Here, the lowest point in the total inventory cost curve is the **economic order quantity** (EOQ) or the **optimal inventory**.
This EOQ can be calculated by using the following formula.
Annual sales (demand) amount : R
Ordering quantity per order : Q
Ordering cost per order : A
Annual holding cost per unit : P
Then
Annual ordering cost = (R/Q) A
Annual holding cost = (Q/2) P
Therefore,
Total inventory cost (TC) = (R/Q) A + (Q/2) P
In the above formulas, the TC can be regarded as the function of the order quantity Q given that P, A and R are constants. Therefore, the Q that minimizes the value of TC is the economic order quantity (EOQ). There exist 2 methods to calculate this.

<method 1>

$$Let\ \frac{dTC}{dQ} = \frac{P}{2} = \frac{AR}{Q^2} = 0$$

$$Q = \sqrt{\frac{2AR}{P}} = \sqrt{\frac{2 \times \text{ordering cost per order} \times \text{annual demand}}{\text{annual holding cost per unit}}}$$

$$\left(\frac{\mathrm{dTC}}{\mathrm{dQ}} \text{ can be also represente d as } \frac{\partial \mathrm{TC}}{\partial \mathrm{Q}} \text{ using the partial differenti al notation} \right)$$

<method 2>

$$In\ TC = \frac{P}{2}Q + \frac{AR}{Q},\ let\ \frac{P}{2}Q = \frac{AR}{Q}$$

$$Q = \sqrt{\frac{2AR}{P}} = \sqrt{\frac{2 \times \text{ordering cost per order} \times \text{annual demand}}{\text{annual holding cost per unit}}}$$

This Q is the order quantity that minimizes the total inventory cost, called the economic order quantity (EOQ). The formula for obtaining this is called the EOQ formula.
The following is the result of applying this formula to the earlier PC shop sample,

The economic order quantity $(Q) = \sqrt{\frac{2 \times 1{,}250_{\text{yen}} \times 500_{\text{units}}}{500_{\text{yen}}}} = \sqrt{2{,}500_{\text{units}}} = 50_{\text{units}}$

Number of orders: 500 units per order /50 units = 10 orders
Ordering cost: 1,250 yen per order × 10 orders = 12,500 yen
Holding cost: 500 yen per unit × 50 units ×1/2 = 12,500 yen
Total inventory cost= 25,000 yen
We can conclude that at this PC shop, if an order is issued 10 times per year and the order quantity per order is 50 units, the total inventory cost is minimized to 25,000 yen.

(3) Ordering Method

① Types of ordering methods

Concerning inventory control methods focusing on ordering, there exist a number of methods such as the fixed quantity ordering method (reorder point method), periodic ordering method, the two bin method and others.
a. Fixed quantity ordering method (reorder point method)
The **fixed quantity ordering method** is also called the **reorder point method**. In this method, the order quantity is set to be fixed and the order cycle is not fixed or to be determined at every order. The optimal order quantity is calculated so that the total inventory cost will be minimized. When the reorder point is reached (i.e. an order date comes), a new order for the fixed amount is placed.
b. Periodic ordering method
In the **periodic ordering method**, the order cycle is fixed and the order quantity varies from time to time depending on the situation. On the pre-determined order interval (order cycle), the optimal order quantity for the order date is calculated based on demand forecasting. This method is generally applied to goods with few type variations and a high unit price.
c. Two bin method
In the **two bin method**, two bins A, B are to be prepared first. The stock is to be retrieved from bin A until the stock there reaches zero. After that, the stock is to be retrieved from bin B while a new order is to be placed to refill bin A.
Since this method is not much important in the computerized inventory control system, further description for this method is not given.
Characteristics of the first two methods are summarized in the table below.

Method	Order interval	Order quantity	Demand fluctuation	Changes of Item specification	Cost
Fixed quantity ordering	not fixed	fixed	not assumed	not assumed	Low
Periodic ordering	fixed	not fixed	allowed	allowed	High

② Selection of ordering method using ABC analysis

Inventory control is to be done by using an ordering method selected from above. The ABC analysis (grade analysis) is commonly used to decide which ordering method to use.
In general, the selection of ordering method is done according to the following steps.

Perform the ABC analysis for all the items that require inventory control so that they are categorized into A, B and C classes.

⇩

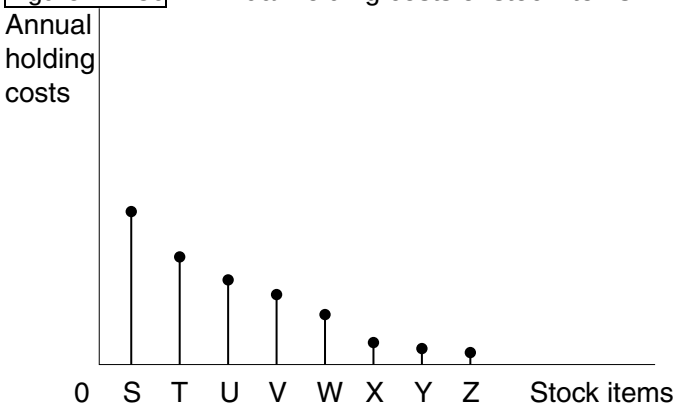
Clarify the characteristics of items in each group, such as the existence of "demand fluctuations", "frequent changes in its specification or standard" etc.

⇩

Decide which method to apply to the inventory control of each item. The “periodic ordering method" or "fixed quantity ordering method" or others.

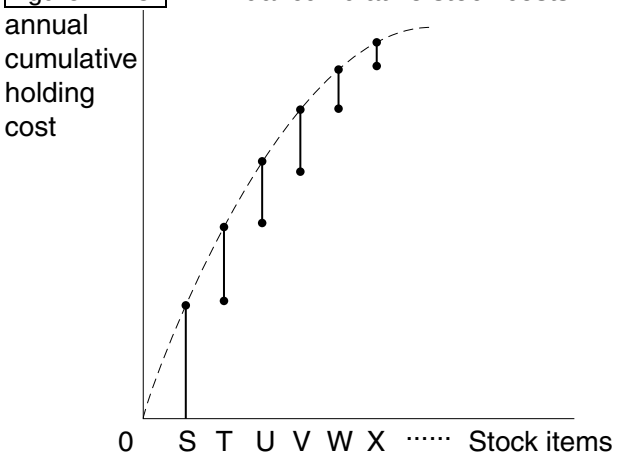
To be more concrete, the ABC analysis will be performed in the following procedure.
1. Arrange the stock items in descending order by their annual holding costs.
Here, the annual holding cost refers to the sum of annual total storing and retrieving cost (purchase cost + the total inventory cost), or the annual total retrieving amount (= sales amount).

Figure 4-1-36 Annual holding costs of stock items



2. Calculate the cumulative annual holding cost.
Obtain the cumulative sum by adding the annual holding cost for an individual stock item. (Figure 4-1-37)

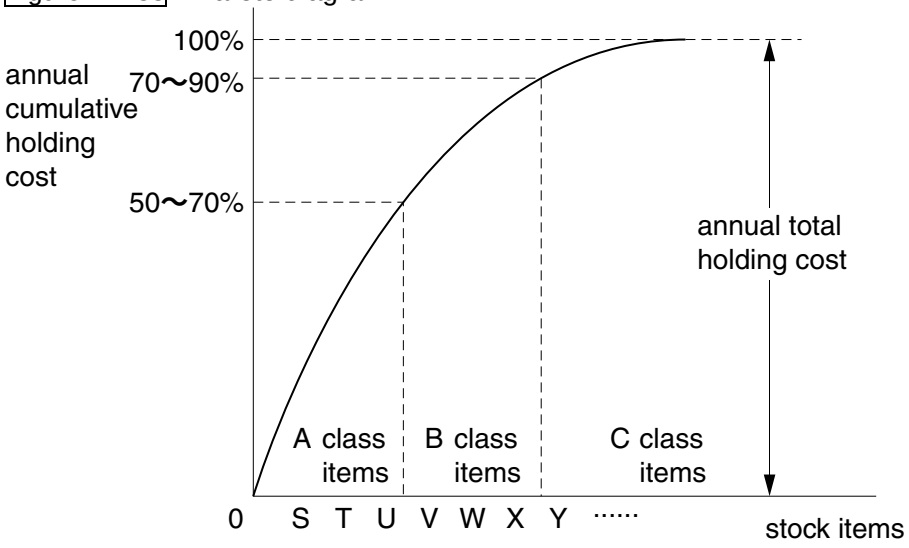
Figure 4-1-37 Annual cumulative stock costs



3. Categorize items into these three classes depending on their annual cumulative holding costs.
class A : stock items the sum of whose annual holding costs occupy 50% through 70% of the annual total holding cost
class B : stock items the sum of whose annual holding costs and that of the items in class A occupy 70% through 90% of the annual total holding cost

class C: the remaining stock items
The chart completed here is called the **Pareto diagram**.

Figure 4-1-38 Pareto diagram



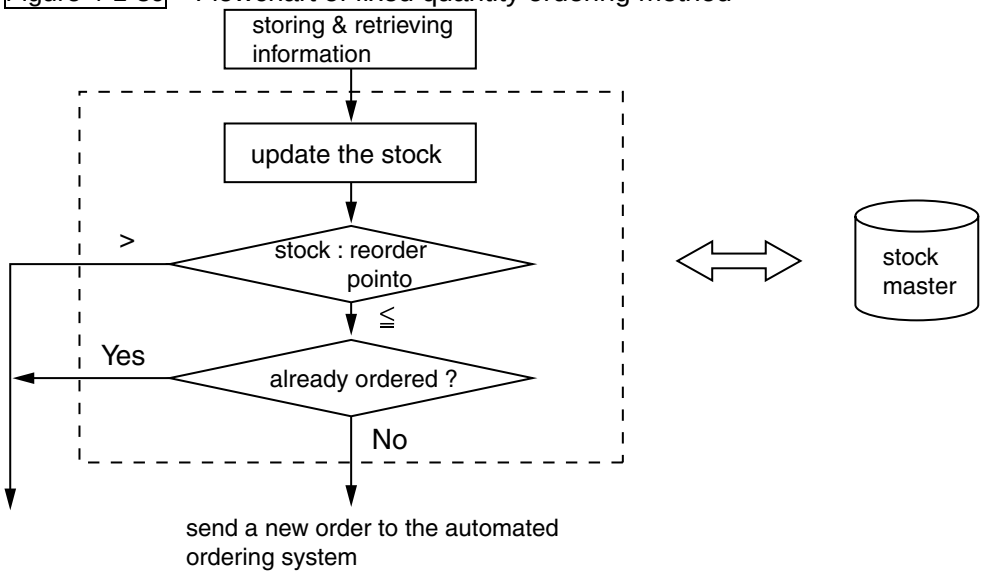
The selection of ordering methods based on only the ABC analysis is as follows.
The class A stock items' inventory should be very carefully monitored using the periodic ordering method because they are critical stock items.
The class B stock items should be mainly managed by the fixed quantity ordering method.
The class C stock items are to be managed by using such methods as the two bin method.

(4) Fixed Quantity Ordering Method (Reorder Point Method)

①What is the fixed quantity ordering method

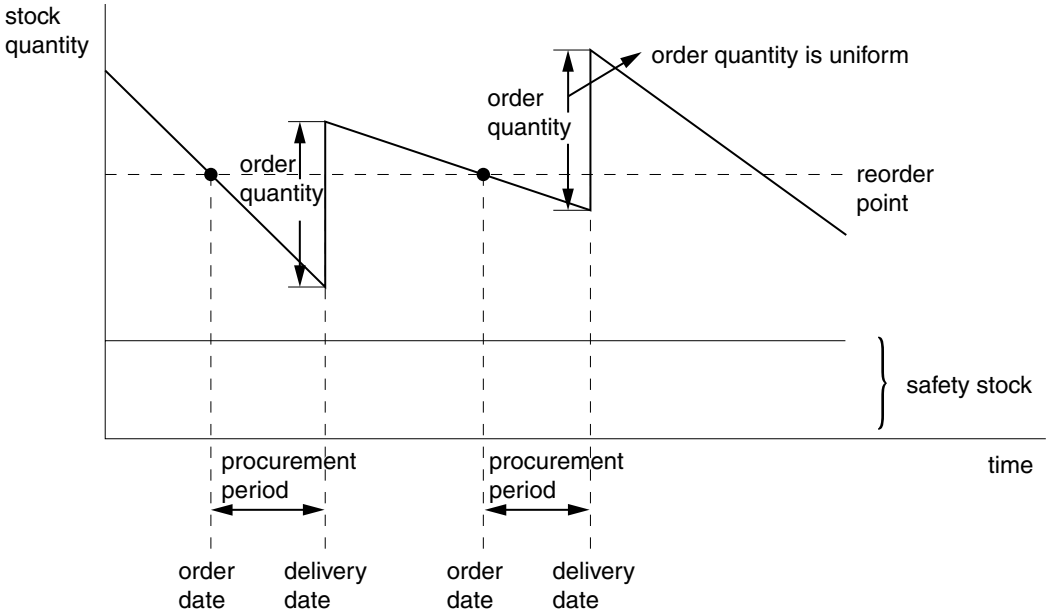
The **fixed quantity ordering method** is also called the **reorder point method**. In this method, a new order with the economic order quantity is placed when the stock level decreases and reaches the pre-determined fixed stock level (this is the reorder point). (See the Figure 4-1-39)

Figure 4-1-39 Flowchart of fixed quantity ordering method



The relationship between the time and stock quantity in the fixed quantity ordering method is shown in the Figure 4-1-40.

Figure 4-1-40 Inventory control using the fixed quantity ordering method



In this method, order timing (order date) is determined by demand for the stock item. If demand gets higher the order interval (order cycle) becomes shorter. If the demand gets lower the order interval (order cycle) becomes longer. Thus the order interval is not fixed.

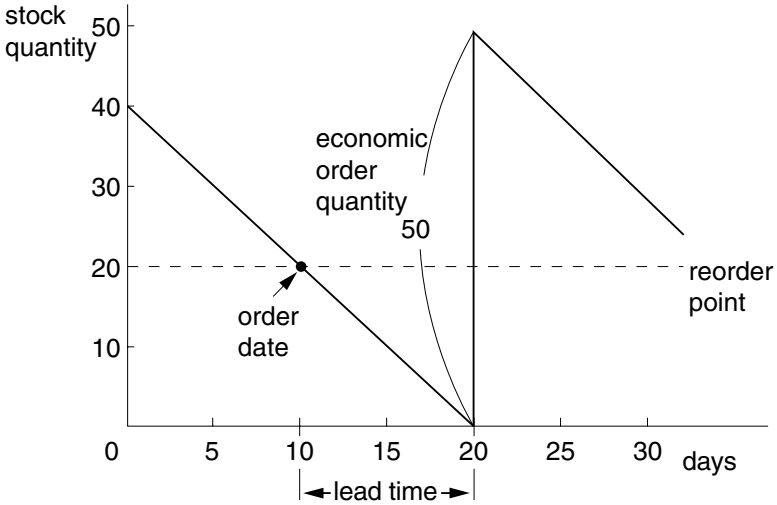
②Deciding the reorder point

The question in the fixed quantity ordering method is to what level the reorder point should be set. The reorder point here is to be decided so that "it suppresses the shortage rate up to a specific level" or "it keeps the service rate above a specific level".
If the lead time and the demand are both fixed, stock level will decrease at a fixed rate. Therefore in such cases, an order should be placed considering the point where the stock will reach 0.
For instance, if those two are fixed as follows,
the procurement period of PCs is 10 [days]
the sales of PCs is 2 [units per day]
Then the reorder point is as follows.

Reorder point =demand (sales) per day×the procurement period
=2 [units per day]×10 [days]
=20 [units]

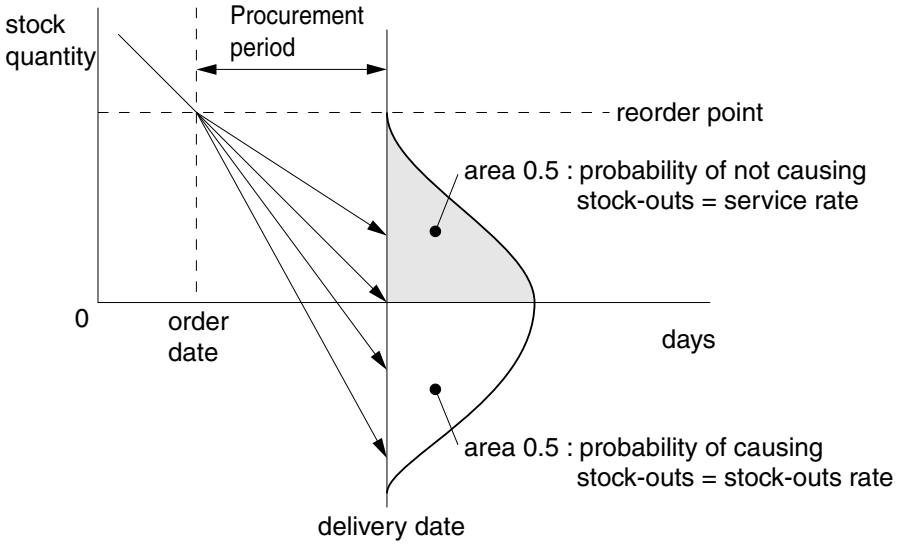
Thus we conclude that when the stock quantity reaches 20, a new order for 50 sets (=EOQ) should be placed.

Figure 4-1-41 Deciding the reorder point



However, in the real world, it is not practical to assume that the procurement period and demand are fixed. If the procurement period becomes equal or more than 10 [days], or sales amount (=demand) exceeds 2 [units per day], stock-outs will happen.

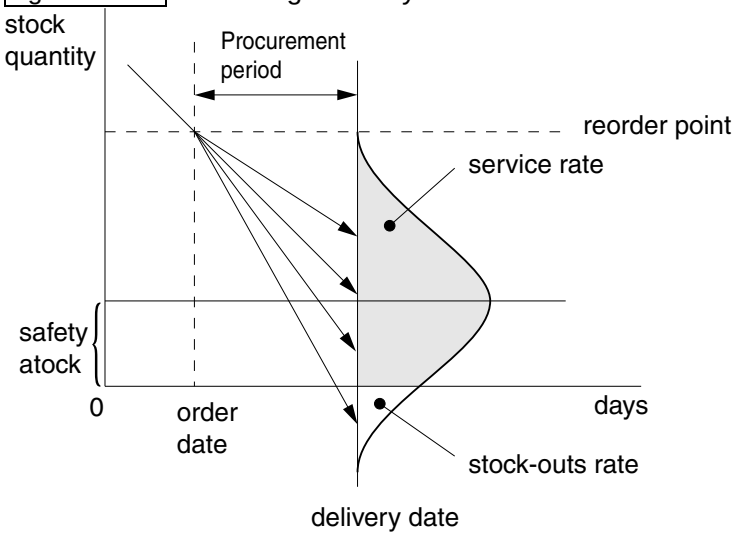
Figure 4-1-42 Meaning of reorder point



③Safety stock

Assume that the demand in the procurement period follows the normal distribution in the Figure 4-1-42, the expected value of the demand within that period is 0.5. That is, the probability that stock-out does not occur is 50%, and the probability that stock-out occurs is 50%.
To better avoid stock-outs, some extra stock quantity is needed. This is called the "**safety stock**".

Figure 4-1-43 Meaning of safety stock



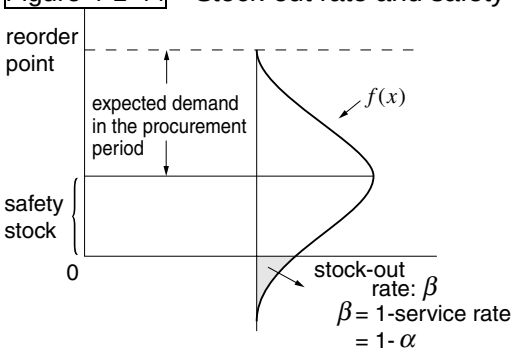
The safety stock is determined to keep the stock-out rate below a certain permitted level. Therefore if either a stock-out rate or the service rate is given, the safety stock can be calculated (Figure 4-1-44). The following is the formula for this.

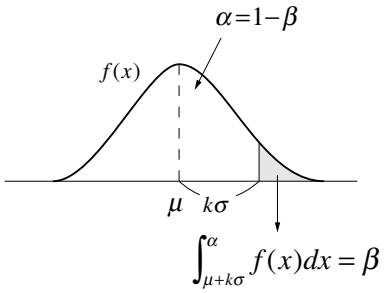
Safety stock = safety factor × standard deviation in lead time

Then the reorder point can be obtained as follows.

Reorder point = expected demand value in lead time + safety stock

Figure 4-1-44 Stock-out rate and safety factor





$f(x)$: distribution function of a randam variable representing the demand for the procurement period
μ: expected vale of demand for the procurement period
σ: standard deviation of demand for the procurement period
β: stock-out rate
α: service rate
k: safety factor

If β (or α) is given, the safety factor K is determined.
For example, to set the stock-out rate to 5% or less, K=1.65.

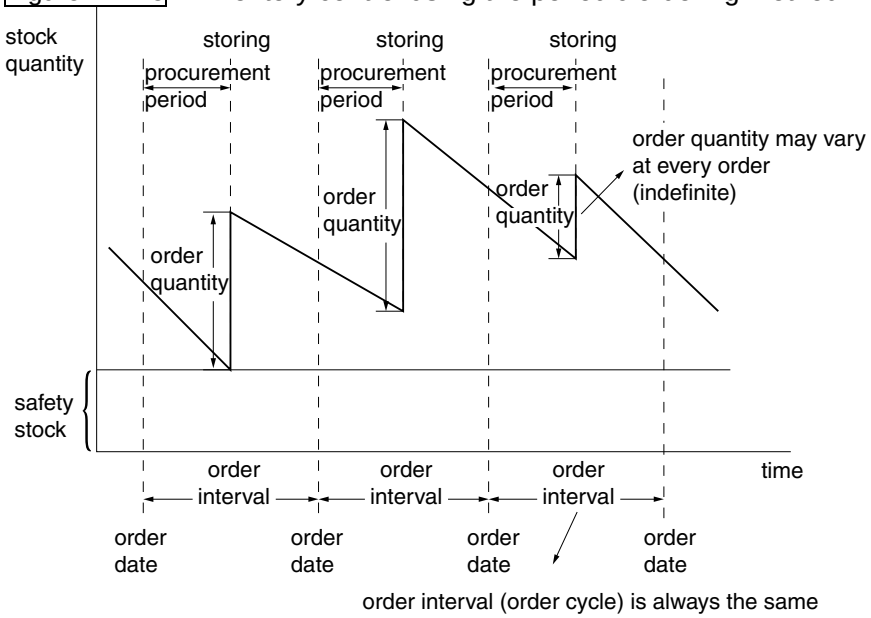
Stock-out rate(β)	safety factor(k)
1% (α :99%)	k=2.33
5% (α :95%)	k=1.65
10% (α :90%)	k=1.28

(5) Periodic Ordering Method

①What is the periodic ordering method

In the **periodic ordering method**, the order interval (order cycle) is pre-determined and fixed. That is, on every pre-determined order date, demand forecasting is done to obtain the optimal order quantity for that day. Different businesses may set different order dates, e.g. the first day of every month, the 10-th day of every month, or every weekend.
When applying the periodic ordering method, the inventory of the goods with frequent changes in their demand or with a high price should be carefully monitored in shorter order cycles compared to others.

Figure 4-1-45 Inventory control using the periodic ordering method



②Deciding the order interval

In the periodic ordering method, how much to order is decided based on the demand forecast performed at every order.
If economical considerations have to be given in this decision, the order quantity should be the EOQ (economic order quantity). In this case, the order interval and the number of orders per year are as follows.
annual number of orders = annual demand / EOQ
order interval = number of days in operation / number of orders
= (number of days in operation × EOQ) / annual demand

③Deciding the order quantity

The order quantity is computed with the following formula using the result of the demand forecast for the period between the order date and the next storing date.

order quantity = the forecasted demand in (procurement period + order cycle) - number of unsatisfied orders - remaining stock + safety stock

Here the number of unsatisfied orders, the forecasted demand in (procurement period + order cycle) are taken into account. Because generally the procurement period tends to be long in this method, consequently the case that previously ordered items haven't been delivered happens frequently.
The safety stock is calculated using the following formula.

safety stock = safety factor × the standard deviation of the demand in (procurement period + order cycle)

④Demand forecasting

If the demand is uniform, the sum of the demand in the procurement period and the demand in the order cycle determines the order quantity. If not, demand forecasting is not so simple.
Methods used for forecasting changing demand include, the moving average method, the method of least squares, and the exponential smoothing method among others. Those are discussed in more detail in the

next section, 4.1.6.

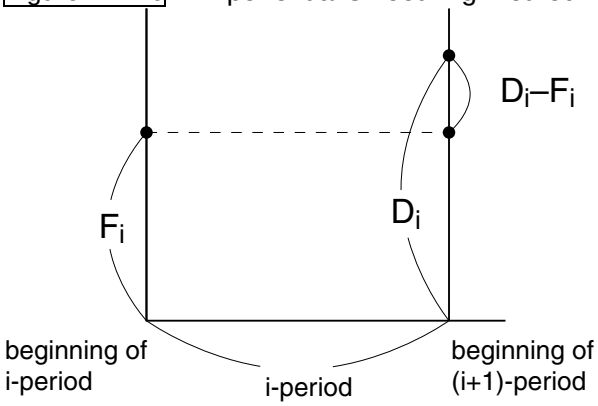
a. Demand forecasting by using the exponential smoothing method

Demand forecasting by using the exponential smoothing method
The exponential smoothing method is commonly used for demand forecasting in a periodic ordering system.
Let F_i = forecasted demand value for period i, calculated at the beginning of period i
D_i = actual observed demand value for period i,
the next forecast F_{i+1}, forecasted demand value for period i+1 is computed as follows.

$$F_{i+1} = F_i + \alpha (D_i - F_i)$$

$(D_i - F_i)$ is the forecast error (the difference between the observed value for period i and the forecast for period i). The exponential smoothing method uses F_{i+1}, the result of adding a specific ratio of this forecast error to the period i forecast, as the forecast for period i+1.
α is called the smoothing constant that usually is assigned a value between 0.1 and 0.5. As for the selection of an α value, it has been suggested that α should be chosen to be small if the demand fluctuation is small.

Figure 4-1-46 Exponential smoothing method



This forecast, although calculated by a relatively simple expression, is actually a weighed average of past observed values D_i, D_{i-1}, D_{i-2} ....
<Proof of the fact that F_{i+1} is a weighed average of past observed values D_i, D_{i-1}, D_{i-2} ...>

$$F_{i+1} = \alpha (D_i - F_i)$$
$$= \alpha D_i - (1-\alpha) F_i \qquad \text{replace } F_i \text{ by } \alpha D_{i-1} - (1-\alpha) F_{i-1}$$
$$= \alpha D_i + \alpha (1-\alpha) D_{i-1} + (1-\alpha)^2 F_{i-1} \qquad \text{replace } F_{i-1} \text{ by } \alpha D_{i-2} - (1-\alpha) F_{i-2}$$
$$= \alpha D_i + \alpha (1-\alpha) D_{i-1} + \alpha (1-\alpha)^2 D_{i-2} + (1-\alpha)^3 F_{i-2}$$
$$\ldots$$
$$= \alpha D_i + \alpha (1-\alpha) D_{i-1} + \alpha (1-\alpha)^2 D_{i-2} + \alpha (1-\alpha)^3 D_{i-3} + \ldots$$

The sum of weighs added to the past observed values D_i, D_{i-1}, D_{i-2} ...

$$\alpha + \alpha (1-\alpha) + \alpha (1-\alpha)^2 + \alpha (1-\alpha)^3 + \ldots$$

is the sum of the terms in an infinite geometrical progression whose first term is α and whose common ratio is $1-\alpha$. This progression converges because its |common ratio| < 1.

first term / (1 - common ratio) = α / (1- (1- α)) = 1

Therefore, F_{i+1} has been proved to be a weighed average of past-observed values.

4.1.6 Demand Forecasting

(1) What is forecasting

Forecasting means formulating information needed for planning, decision-making or business operation optimization, based on the past or current information.
However, because of the existence of uncertainty, there may exist some forecasting errors or differences

between observed values and forecasted values. Therefore we need to keep in mind that when we take actions based on forecast, we are at some risk. What is important in forecasting is how we minimize this risk as much as possible.
There are various forecasting methods, classified into the following three:
Forecasting using past trends
Forecasting using current indexes
Forecasting using model calculations

①Forecasting using past trends

Forecasting methods using past data trends assume the idea that "past trends are projected onto the future". Some of these methods are shown below.

a. Time-series analysis

In the **time-series analysis**, future data are forecasted as a result of analyzing past time-series data and identifying a certain trend in the given data.

b. Regression analysis

In **regression analysis**, a forecast is obtained by identifying the cause of changes on past time-series data. This analysis is also called "time-series regression analysis," since it generally performs the analysis of cause-and-effect relationships, starting from time-series analysis.
For example, the number of new students in a private school is greatly affected by such factors as birth rate. Here we can say birth rate is a cause, and the number of new students is a result.

c. Correlation analysis

The **correlation analysis** is a method that analyzes cause-and-effect relationships so as to find out what caused the changes on the past time-series data

②Forecasting using current indexes

Forecasting method using currently available indexes are also called "similarity methods". The following are some of these methods.

a. Cross section method

The **cross section method** is also called the "simultaneous comparison method". This method obtains a future forecast by comparing the data in question with similar other data, at the same point in the past.
An example application of this method is to forecast the lifecycle of goods whose sales are currently growing by comparing these goods with other goods whose current sales are decreasing, but whose sales were once growing.

b. Leading index method

Forecasting in the **leading index method** is based on the selected index that shows future trends of the given data. This index is selected out of the current or past statistical data.
An example application of this method is to forecast future fashion trends. This is done by identifying indications of fashion trends, e.g. a successful movie in the U.S. is also successful in Japan

c. Delphi method

In the **Delphi method**, also called the "converging with questionnaires" method, a panel of experts are asked to fill out a series of questionnaires until a consensus is reached, and the resulting forecast is obtained.
This method can be applied, for example, to forecast future politics based on the randomly selected responses to the questionnaire that contains such questions as one concerning the public approval ratings for the Cabinet.

③Forecasting using model calculations

The methods that build a model on the given data, and perform computation and analysis on the model to forecast the future are called "jurimetrics". Some of them are shown below.

a. Econometric analysis

The **econometric analysis** is also called the "prediction equation method". This method obtains a forecast

by solving associated equations modeled after given data. This method is applied when forecasting in the macro-level.

b. Inter-industry relations analysis

In the **inter-industry relations analysis**, linear programming models are first created using the inter-industry relations table, and then necessary computations are done. This method is applied when forecasting in the micro-level.

(2) Time-series Regression Analysis and variations

In the time-series regression analysis, forecasting is done by clarifying cause-and-effect relationships in the time-series data, or chronologically collected data. There exist a number of time-series regression analysis methods, some of which are described in more detail in 3.

①Types of variations

Variations found in time-series data are classified into the following four.

a. Trend variations

The variations in a trend, which are observed in the long term, are called the **trend variations.**

b. Cyclical variations

Like economic (business) fluctuations, variations that show some cyclical trend (from several years to approximately ten years) are called the **cyclical variations**. Some of well-known business cycles are the Kitchin's short-wave cycle of 3-5 year duration and the Jugler's cycle of approximately 10-years cycle.

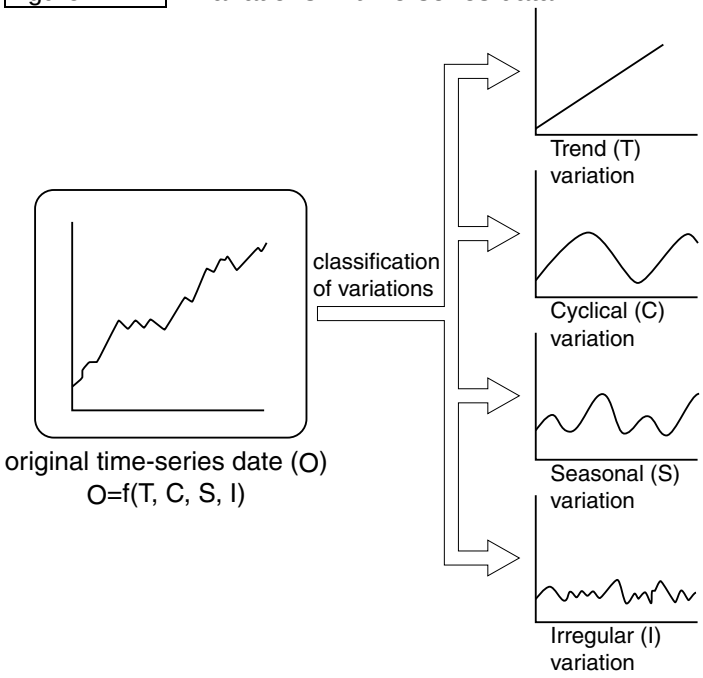
c. Seasonal variations

The **seasonal variations** are variations affected by natural conditions, social practices etc. Usually, this has a 1-year (or 12-month) cycle. The data values in the same month of different years show similar fluctuations, affected by such social practices as annual bonuses, and accounting terms.

d. Irregular variations

The **irregular variations** are also called "accidental variations". This refers to the variations that cannot be explained by the above-listed three variation elements. This may include weather and disaster factors. But in fact it is impossible to predict this.

Figure 4-1-47 Variations in time-series data



②Analysis of variations

To perform the time-series regression analysis, we first need to analyze variations in the original time-series data. That is, we should use the data that have only meaningful variations for forecasting. If forecasting is done based on past data that have been affected by irregular variations, inappropriate forecasts may be obtained.

Therefore, when forecasting, inappropriate variations of original data should be removed. This is called the "**seasonal adjustment method**" and the data whose inappropriate variations have been removed is called "seasonal adjustment data".

There exist a number of seasonal adjustment methods. Commonly the seasonal variations and irregular variations are removed from the original data. In some cases, cyclical variations are removed instead of seasonal variations. The decision of which variations to be removed is made by considering the characteristics of the goods to be forecasted.

(3) Method of selecting a best-fitted line

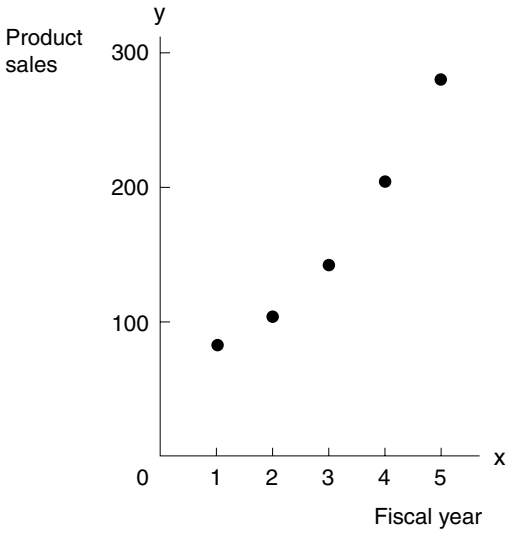
The **method of selecting a best-fitted line** is a method that selects a line that best fits the given data. In this method, using the given time-series data, a plot of observed data values versus time is created first. Then a best fitted line (a trend line with minimal forecast errors) for this plot, either a straight line or a curve, is to be identified. Lastly, future data values are forecasted by extrapolating the graph.

Suppose that the annual sales amounts of a product are given for 5 fiscal years. These data appear in the following table.

Fiscal Year	1st	2nd	3rd	4th	5th
Product sales	80	100	140	200	280

Let us create a chart containing a plot of sales amounts versus a fiscal year, using the above table data. The resulting plot is given in the Figure 4-1-48.

Figure 4-1-48 Method of selecting a best-fitted line



Sales in the next year are to be forecasted using this chart. There exist two different forecasting methods, the "selection by visual observation" method and the "sum of least squares" method.

①Selection by visual observation

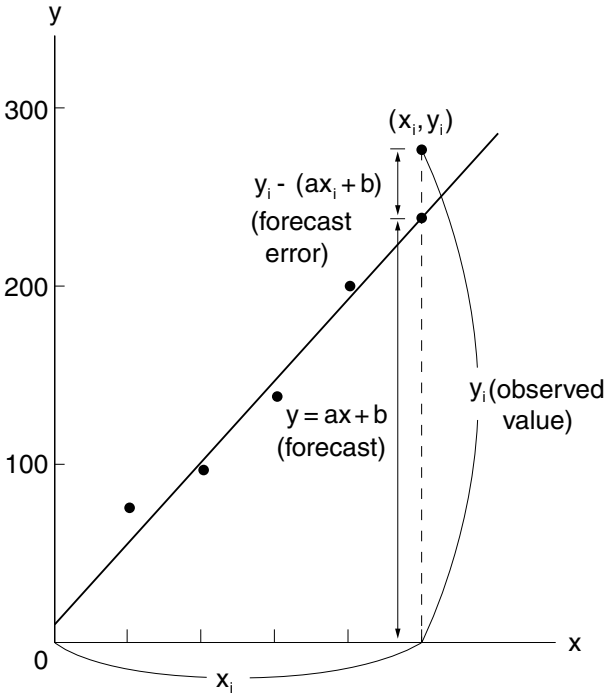
In the **method of selection by visual observation**, a trend line is drawn by guessing after observing the plotted data in a chart. A trend line may not be a straight line. It is drawn freely, without using any mathematical computation. Because of its simplicity, this is often used in practical situations.

②Method of least squares

The **method of least squares** is the mathematical method that identifies the best-fitted trend line.
A measure of how well a trend line fits the data can be obtained by computing the sum of squares of the vertical deviations of the actual points from the trend line ; the line that minimizes this value is the best-fitted one. The concrete steps of this are shown below.
1. Let $(x_1,y_1),(x_2,y_2)$, ... (x_i,y_i) represent the x,y coordinates plotted in a chart. A forecast error (difference between the observed value and the forecast) can be represented by $y_i - (ax_i + b)$. (Figure 4-1-49) A simple sum of forecast errors may not be meaningful because there may be positive values and negative values, possibly canceling out each other's values. Instead, forecast errors are first squared so that they all become positive, then the sum of squares is calculated. The trend line that minimizes this sum of squares to be obtained.

Sum of least squares $(S) = \sum_{i=1}^{n} \{ y_i - (ax_i + b) \}^2$ → this is to be minimized

Figure 4-1-49 Forecast error (difference between the observed value and the forecast)



2. The expression calculating the sum of least squares (S) is a function of a and b. S is to be partially differentiated by a and b respectively, and results are set to be 0.

$$\begin{cases} \dfrac{\partial S}{\partial a} = 0 & \dfrac{\partial}{\partial a}\sum_{i=1}^{n}(y_i - ax_i - b)^2 = 0 \\ \Rightarrow & \\ \dfrac{\partial S}{\partial b} = 0 & \dfrac{\partial}{\partial b}\sum_{i=1}^{n}(y_i - ax_i - b)^2 = 0 \end{cases}$$

The resulting equations are as follows.

$$\begin{cases} 2\sum_{i=1}^{n}(y_i - ax_i - b)\times(-1) = 0 \\ 2\sum_{i=1}^{n}(y_i - ax_i - b)\times(-x_i) = 0 \end{cases}$$

3. The following simultaneous equations are obtained by transposing and organizing terms accordingly. These equations are referred to as "normal equations".

$$\begin{cases} \sum_{i=1}^{n} y_i = a\sum_{i=1}^{n} x_i + nb \\ \sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i^{\,2} + b\sum_{i=1}^{n} x_i \end{cases}$$

4. The trend line is obtained by solving the above simultaneous equations for a and b.
For example, the earlier example can be solved as follows.
To apply the least sum of squares method to this example, first, summarize items needed for this method in a table. Necessary items for this sample are the figures related to the sales trend in the past 5 years.

item n	year x_i	product sales y_i	year x product sales $x_i y_i$	year x year x_i^2
1st year	1	80	80	1
2nd year	2	100	200	4
3rd year	3	140	420	9
4th year	4	200	800	16
5th year	5	280	1,400	25
Total Σ	15 $\sum_{i=1}^{n} x_i$	800 $\sum_{i=1}^{n} y_i$	2,900 $\sum_{i=1}^{n} x_i y_i$	55 $\sum_{i=1}^{n} x_i^2$

After assigning the above table values to the normal equations,

$$\begin{cases} 800 = 15a + 5b & (1) \\ 2{,}900 = 55a + 15b & (2) \end{cases}$$

a=50, b=10 are obtained as a result of calculating (2) - (1) × (3), . These represent the trend line

y=50x + 10

The forecasted value for next years product sales y=310 is obtained by assigning x=6 to the above equation. (Here, 6 represents the 6th year.)

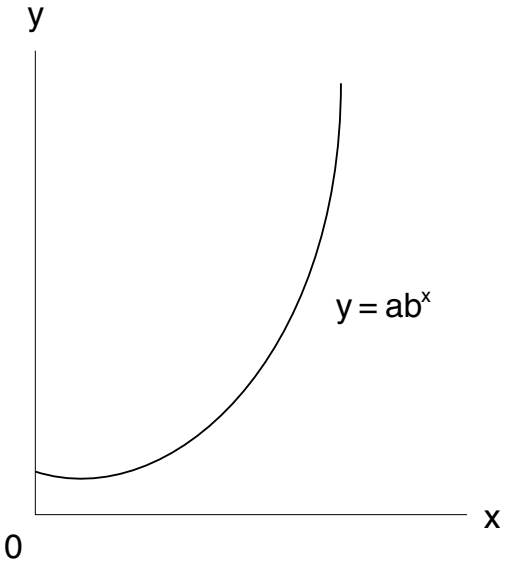
③Other trend lines

In the previous section (2), we have solved the problem by using the sum of least squares method, under the assumption that the trend line is linear, y = ax + b. However, there are various cases to which linear trend lines do not fit. In practice, we first need to observe the nature of the data to be forecasted to decide what kind of trend line is likely to fit. Then we obtain the actual trend line by using the method of least squares. When a trend line is not linear (or it is a curve), normal equations are quadratic.
Some non-linear trend lines are shown below.

a. exponential curve : $y=ab^x$

Numerous social phenomena such as increase in population, bacteria propagation show **exponential curve** trend lines. The growth of this curve is similar to that of a geometrical progression. (Figure 4-1-50)

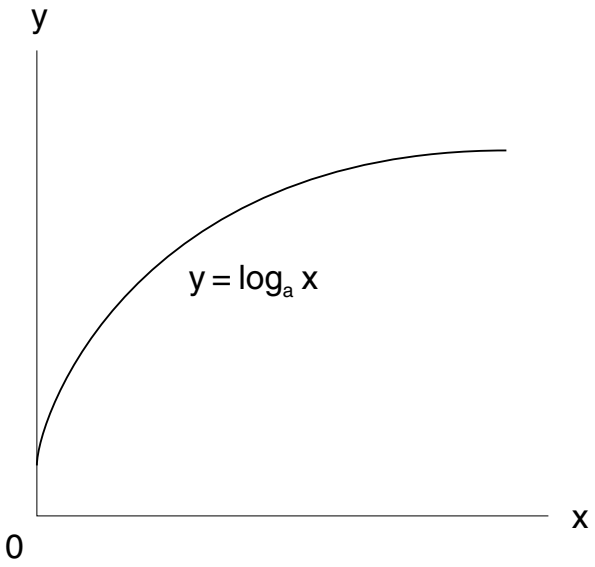
Figure 4-1-50 Exponential curve



b. logarithmic curve : $y = \log_a x$

Phenomena with an upper limit in its growth such as average height, average weight and the production amount of goods whose lifecycle is reaching an end, show **logarithmic curve** trend lines. The growth in this curve is sharp in the beginning, gradually slows down, and finally flattens out.

Figure 4-1-51 Logarithmic curve

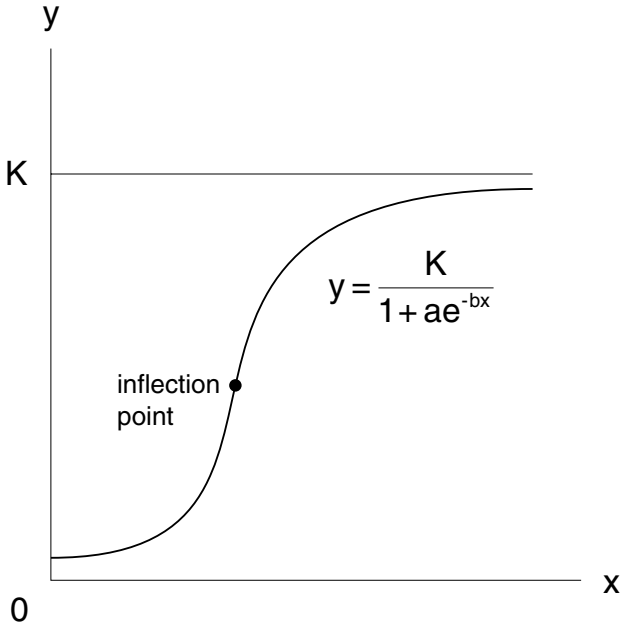


Growth curve $y = \dfrac{K}{1 + ae^{-bx}}$ (e: the base of natural logarithm K: constant)

The saturation level of air conditioners or VCRs, or a product's lifecycle fit **growth curve** trend lines. A growth curve originally simulates the growth of an organism in a given environment. This curve begins to show a rapid growth at the inflection point, or the turning point of the curve. At the end of this curve, the growth rate flattens. (Figure 4-1-52)

There are two well-known growth curves, the logistic curve and the Gompertz curve. Compared to the logistic curve, the Gompertz curve has its inflection point in an earlier stage. Therefore, if a sharp growth is forecasted, the Gompertz curve is commonly used.

Figure 4-1-52 growth curve



(4) Moving Average Method

The moving average is a method that does not try to fit a single trend line to the entire time-series data. Instead, it gives forecasts by calculating the average of the partial time series data, moving along the chronological axis.

The above-mentioned method of least squares basically uses original data for time-series analysis and forecasting. Since original data may contain variations, in the long run, they may not fit the same trend line (a straight line or a curve). Therefore, for each observed value in the time series data, the moving average method computes its partial average i.e. the average of "the value in question and some of its preceding and succeeding values for a certain period". Then it uses the computed values instead of observed values to forecast future values. Thus the moving average method smoothes out the irregular variations and seasonal variations to better track the averages. This method smoothes out seasonal variations.

3-period moving averages of original data (number of elements : n) $\bar{y}_i$ are calculated as follows.

Original data : $x_1, \underline{x_2}, \underline{x_3}, x_4 \ldots\ldots\ldots\ldots\ldots\ldots, \underline{x_{n-1}}, x_n$

$$\bar{y}_2 = \frac{x_1 + x_2 + x_3}{3} \qquad \bar{y}_3 = \frac{x_2 + x_3 + x_4}{3} \qquad \bar{y}_{n-1} = \frac{x_{n-2} + x_{n-1} + x_n}{3}$$

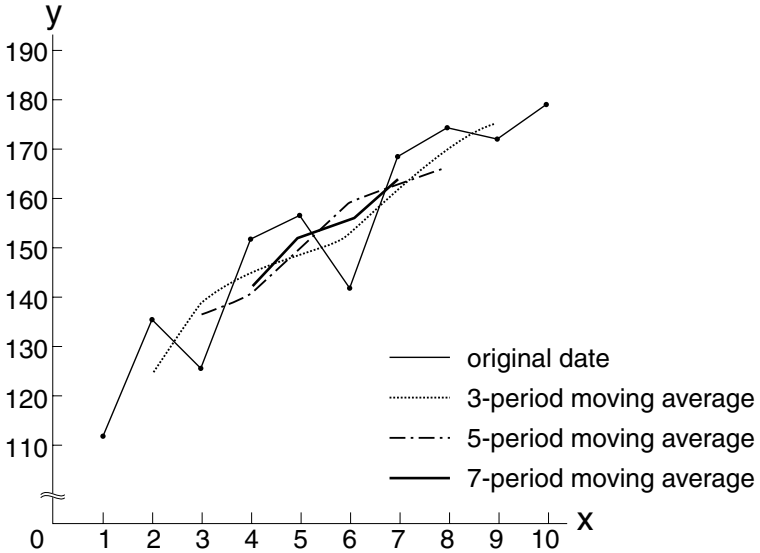
Note that the first element (x_1) and the last element (x_n) do not have moving averages. Those are called "missing terms".

Consider N-period moving averages (N=3,5 and 7) of 10 original data shown in the following table.

No.	1	2	3	4	5	6	7	8	9	10
data	112	136	126	152	156	142	168	174	172	178
3-period	-	125	138	145	150	153	161	171	175	-
5-period	-	-	137	142	149	159	162	167	-	-
7-period	-	-	-	142	151	156	163	-	-	-

The graphical representation of this is shown in the Figure 4-1-53.

Figure 4-1-53 Moving average method



It is obvious that the lines of moving averages are smoother than the line of original data. Also, the line of 7-period moving average is smoother than the line of 3-period moving average. This is because irregular variations in each period have been canceled by obtaining moving averages. Also, as the number of periods used for moving averages increases, seasonal adjustment proceeds, consequently only trend variations come to be shown.

In the moving average method, moving averages calculated as above are used for forecasting instead of original data. It uses the least sum of squares method to obtain forecasts.

(5) Exponential Smoothing Method

The **exponential smoothing method** was developed by R.G. Brown to forecast demands in inventory control.

A disadvantage of the moving average method is that an N-period moving average requires N elements of time-series data. Therefore, if a large number of items are to be forecasted by this method, a large amount of data needs to be processed.

On the other hand, in the exponential smoothing method, it does not matter whether the amount of collected data is small or large, or whether the number of items to be forecasted is big or small. The following is the formula used in this method.

$y_i = y_{i-1} + \alpha (Y - y_{i-1})$

y_i : smoothed value for period i (next period)

y_{i-1} : smoothed value for period i-1 (this period)

Y : actual observed value for period i-1 (this period)

α : a smoothing constant ($0 < \alpha < 1$)

In this method, the smoothed value for the "next + L" period is the same as that for the next period.

The forecast based on the above formula changes more smoothly compared to the original data. Consider the α value. The forecast with α set to be smaller than 0.1 shows a quite smoothened curve, while the forecast with α closer to 1 shows a curve that is very similar to that of the original data. Because of its affect on the forecast result, the selection of the α value is very important.

Transforming the above basic formula results in the following equation

$y_i = \alpha Y + (1-\alpha) y_{i-1}$

This is a weighted average, computed by weighing the actual observed value Y and the smoothed value (this period) y_{i-1}. Note that the moving average method does not weigh observed values.

The formula below calculates y_i on only observed values. This is a transformation of the previous one.

$y_n = \alpha Y_{n-1} + \alpha (1-\alpha) Y_{n-2} + \alpha (1-\alpha)^2 Y_{n-3} + ...$

This means that the smoothed value for the period n is computed as the sum of weighed observed values each of which is given the weight α, $\alpha(1-\alpha)$, $\alpha(1-\alpha)^2$ ... respectively. As for α, $1 > \alpha > \alpha(1-\alpha) > \alpha(1-\alpha)^2$ ... is always true.

That is, this method gives more weight to more recently observed values, therefore it gives us a forecast considering recent trend variations.

Let us apply this method to the example of the 10 original data used in the moving average method. Consider two cases, one with α=0.2, the other with α=0.4.

No.	1	2	3	4	5	6	7	8	9	10
data	112	136	126	152	156	142	168	174	172	178

Using the above table data, compute a forecast by the exponential smoothing method by assigning data values to the formula : $y_i = \alpha Y + (1-\alpha) y_{i-1}$. Since the smoothed value for this period forecasted previously is not given, we use the average of the first two values i.e. 124, as the initial value.

No.2 $(112+136) \div 2 = 124$

No.3 α=0.2 $0.2 \times 126 + (1-0.2) \times 124 = 124$

α=0.4 $0.4 \times 126 + (1-0.4) \times 124 = 125$

No.4 α=0.2 $0.2 \times 152 + (1-0.2) \times 124 = 130$

α=0.4 $0.4 \times 152 + (1-0.4) \times 125 = 136$

No.5 α=0.2 $0.2 \times 156 + (1-0.2) \times 130 = 135$

α=0.4 $0.4 \times 156 + (1-0.4) \times 136 = 144$

No.6 α=0.2 $0.2 \times 142 + (1-0.2) \times 135 = 136$

α=0.4 $0.4 \times 142 + (1-0.4) \times 144 = 143$

No.7 α=0.2 $0.2 \times 168 + (1-0.2) \times 136 = 142$

α=0.4 $0.4 \times 168 + (1-0.4) \times 143 = 153$

No.8 α=0.2 $0.2 \times 174 + (1-0.2) \times 142 = 148$

α=0.4 $0.4 \times 174 + (1-0.4) \times 153 = 161$

No.9 α=0.2 $0.2 \times 172 + (1-0.2) \times 148 = 153$

α=0.4 $0.4 \times 172 + (1-0.4) \times 161 = 165$

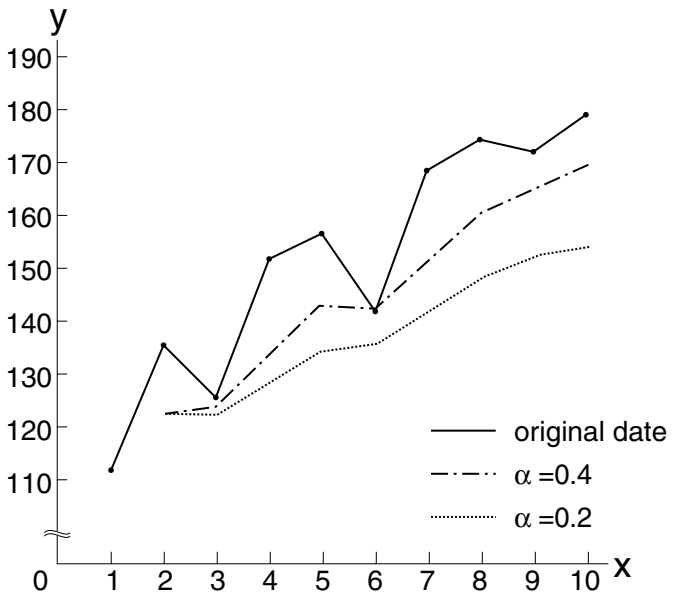
No.10 α=0.2 $0.2 \times 178 + (1-0.2) \times 153 = 158$

α =0.4 $0.4\times178+(1-0.4)\times165=170$

No.	1	2	3	4	5	6	7	8	9	10
data	112	136	126	152	156	142	168	174	172	178
α =0.2	-	124	124	130	135	136	142	148	153	158
α =0.4	-	124	125	136	144	143	153	161	165	170

The resulting plot is shown in Figure 4-1-54.

Figure 4-1-54 Exponential smoothing method



By comparing the two results obtained by α =0.2 and α =0.4 respectively, we can see that both of them show a more smooth curve than the original data does. Their trend changes (increase or decrease) are not as drastic as the original's.

Also in this comparison, we find that the result obtained by α =0.2 is smoother than that by α =0.4. It is more common to use α =0.4 in practice for the actual demand forecasting.

Exercises

Q1 Which of the following chart descriptions is incorrect?

a. A Pie chart is used to show the contribution of each value to a total. It is suitable for visualizing the ratio.
b. A 100% stacked bar chart compares the percentage each value contributes to a total over time. It is suitable for visualizing the contents' changes over time.
c. A Scatter chart is drawn by plotting two attributes on the horizontal and longitudinal axis of a graph. It is used to check whether those 2 attributes are related to each other.
d. A Pareto chart shows ranked items and their frequencies. Items are arranged in a descending order according to the frequency. The cumulative sums of their frequencies are also shown. We use this to find out important items.
e. A Radar chart is often used to check the balance of data characteristics. There are two methods. One is to compare characteristics with idealistic figures or standard figures. The other is to compare characteristics by overlaying several figures.

Q2 Which of the following combinations is inappropriate?

	Objectives of Expression	Graph
a.	To express the traffic change in a day	Line chart
b.	To express each company's share in the market	Z graph
c.	To express the position of new product in the market	Portfolio chart
d.	To express comparison of machines based on several assessment items.	Radar chart
e.	To express the work schedule and subsequent progress	Gantt chart

Q3 Which of the following is the suitable description of a Radar chart?

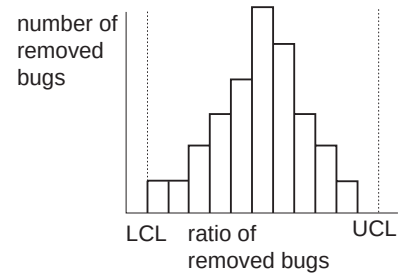
a. A chart that shows the relationship between fixed cost and proportional cost to a sales amount, and is used to analyze the commercial profit.
b. A chart with a web-like figure that is used to check the balance of several characteristics.
c. A chart which is used to judge the correlation between two characteristics from the variance of plotted data on the x-y coordinates.
d. A chart which is used to analyze sales performance of a certain period of time, showing the monthly sales amount, the cumulative sales amount and the moving average within one chart.

Q4 Which of the following is correct as the description of quality control?

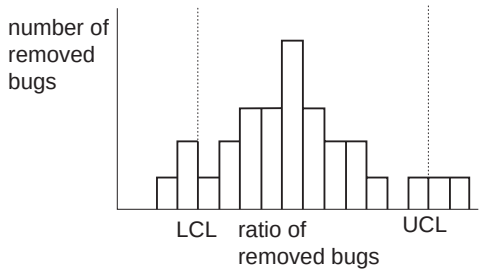
a. A Scatter chart is useful to check the data spread for a variable; the mean and standard deviation can easily be found.
b. An Affinity diagram (KJ diagram) is used to arrange and sort-out intricate problems, loose opinions and ideas.
c. A Cause and effect diagram is useful to express the interrelation of more than two variables.
d. A Frequency table contrasts cause and result; it is often used to seek the cause for defective products.
e. A Pareto chart is characterized by its shape: pie or sector form; it is used to judge the size of data all at once.

Q5 It is known that in a unit test operation, the bug removal rate per 1,000 steps is almost within normal distribution. Suppose there are a number of development teams and that each of the following histograms shows the bug removal rate for one development team. From these histograms, we are able to understand there is a team that made changes to their bug removal data so that their data is between UCL and LCL, because they were not satisfied with their high bug-removal-rate. Which of the following histograms corresponds to this? Here UCL stands for the "upper control limit", LCL stands for the "lower control limit".

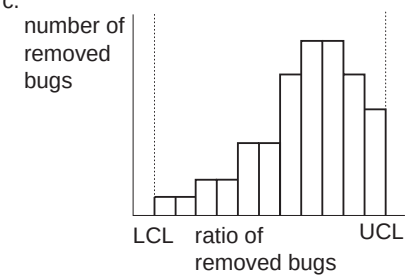
a.



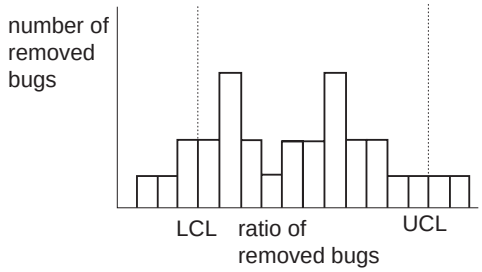
b.



c.



d.

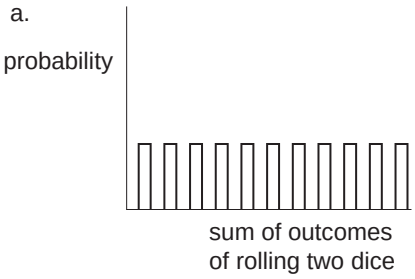


Q6 **When extracting a card out of a set of 52 cards, what is the probability that the extracted card is the queen(12) of spades or any card of hearts?**

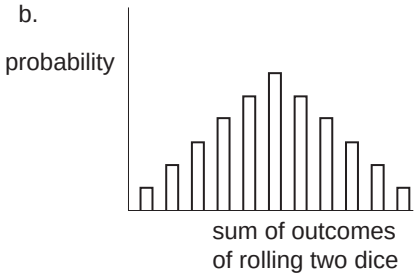
a. 1/208 b. 1/26 c. 1/4 d. 7/26

Q7 **Consider the outcome of rolling two dice. Which of the following charts shows the probability distribution of the sum of dots on the two dice?**

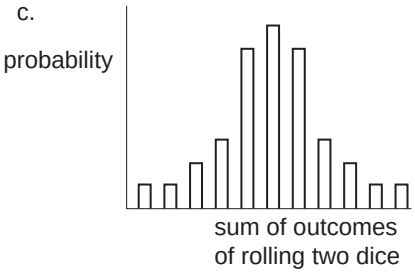
a.



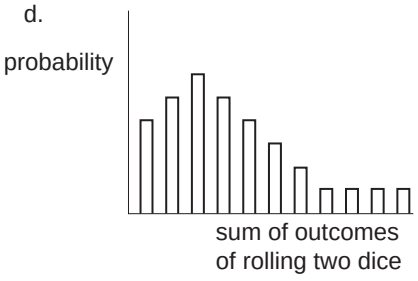
b.



c.



d.



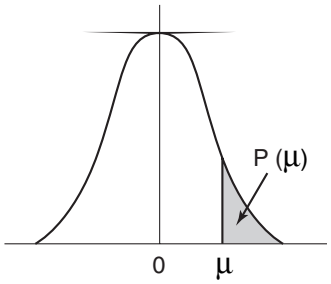
Q8 **What is the difference (margin) between the mathematical mean value and the medium value of the following 5 datum? 25 31 17 17 27**

a. 0.6 b. 1.6 c. 2.6 d. 6.4

Q9 **The size of a product produced in a process flow has normal distribution with an average of 200mm and a standard deviation of 2mm. When the size specification of this product is 200± 2mm, what is the probability of a product being defective?**

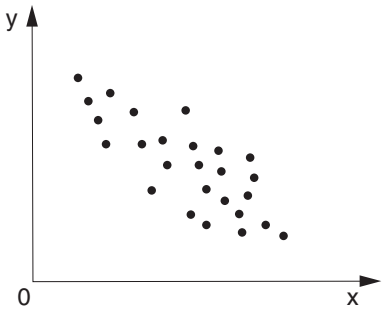
Normal distribution table

μ	P (μ)
0.5	0.3085
1.0	0.1587
1.5	0.0668
2.0	0.0228
2.5	0.0062
3.0	0.0013



a. 0.0228
b. 0.0456
c. 0.1587
d. 0.3174

Q10 **The following chart shows the relationship between a value x, a certain factor in manufacturing, and value y, a quality characteristic. Which of the following describes this chart correctly?**



a. There is positive correlation between x and y.
b. There is negative correlation between x and y.
c. There is little correlation between x and y.
d. The regression function for estimating y from x is as same as the one for estimating x from y.
e. In order to estimate y from x, it is necessary to calculate the secondary regression coefficient.

Q11 **What is the most suitable method to solve the following problem?**

Some plant is producing three products; A, B and C made from material M. Each of A, B, C 's approximate time required for manufacturing per 1 kg, necessary amount of material M, and its profit are shown in Table 1. Also, resources that can be allocated per month in this plant appear in Table 2. Under the above assumptions, how many units of A, B, and C should the plant produce every month to maximize the total profit?

Table 1 Production Constraints

Product	**A**	**B**	**C**
Required time for production (hours/kg)	2	3	1
Amount of material M needed for production (liters/kg)	2	1	2
Profit (1,000yen/kg)	8	5	5

Table 2 Allocatable Resources

Production time (hours/month)	240
Amount of material M (liters/month)	150

a. Moving average method
b. Method of least squares
c. Linear programming
d. Fixed quantity ordering method

Q12 Which of the following (x,y) minimizes x-y under the given constraints?

Constraints $x + y \leqq 2$
$x, y \geqq 0$

a. (0,0) b. (0,2) c. (1,1) d. (2,0) e. (2,2)

Q13 Which is the most relevant method to solve the following question?

Some plant produces products A and B made from petroleum. The following table shows the resources that can be allocated per month in this plant, A and B's respective required time for production, the necessary amount of petroleum, and its profit per kilogram. Under the above assumptions, how many units of A and B should the plant produce every month to maximize the total profit?

Table : Production constraints

Product	A	B	Allocatable resources in this plant
Time required for production per kg	3 hours	2 hours	240 hours per month
Amount of petroleum required for production	1 liter	2 liters	100 liters per month
Profit per kg	50,000 yen	80,000 yen	

a. PERT
b. Moving average method
c. Sum of least squares
d. Linear Programming
e. Fixed quantity ordering method

Q14 Some plant produces products A and B. Production of A (1 ton) requires 4 tons of material P and 9 tons of material Q. As for production of B, 8 tons of P and 6 tons of Q are required. Profit of product A is 2×10,000 yen per ton, and that of product B is 3×10,000 yen per ton. A maximum of 40 tons of P and a maximum of 54 tons of Q can be used for production.

Which of the following is the formulation result of a linear programming model for the question which obtains the production amount that maximizes the total profit. Assume that the production amount of A and that of B are represented as x and y, respectively.

a. Constraints $4x + 8y \geqq 40$
$9x + 6y \geqq 54$
$x \geqq 0, y \geqq 0$
objective function $2x + 3y$ --> to be maximized

b. Constraints $4x + 8y \leqq 40$
$9x + 6y \leqq 54$
$x \geqq 0, y \geqq 0$
objective function $2x + 3y$ --> to be maximized

c. Constraints $4x + 9y \leqq 40$
$8x + 6y \leqq 54$
$x \geqq 0, y \geqq 0$
objective function $2x + 3y$ --> to be maximized

d. Constraints $4x + 9y \leqq 2$
$8x + 6y \leqq 3$
$x \geqq 0, y \geqq 0$
objective function $40x + 54y$ --> to be minimized

Q15 When creating the scheduling of a project, which is the most suitable OR (Operations Research) technique?

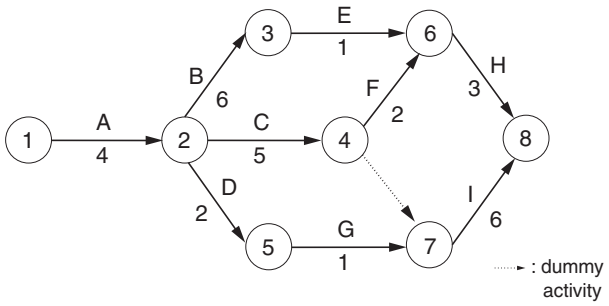
a. PERT
b. Regression Analysis
c. Time Series Analysis
d. Linear Programming

Q16 **Suppose that we created an execution plan of a system development project using PERT and calculated a critical path. Which of the followings is the best way of utilizing a critical path?**

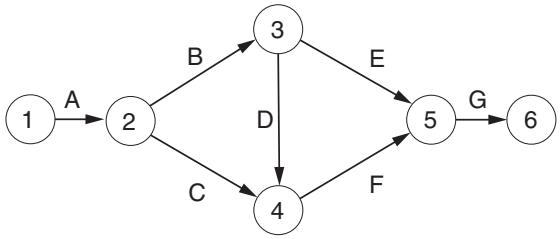
a. To identify the activity that requires the most careful attention from the viewpoint of system quality.
b. To identify the activity whose execution order is changeable.
c. To identify the activity which directly causes the delay of a whole project.
d. To identify the most costly activity.

Q17 **Concerning the following project, in order to shorten the necessary duration of a critical path by one day, which of the followings is the suitable action to be taken? In the figure, an alphabetic letters above an arrow represents the name of the activity; and the number represents the necessary duration for the activity.**

a. To shorten the activity B by one day.
b. To shorten the activity B and F by one day respectively.
c. To shorten the activity H by one day.
d. To shorten the activity I by one day.



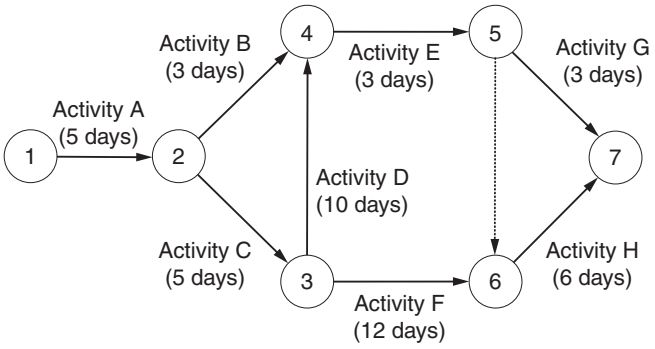
Q18 **Which of the following is the latest start time for the activity E in the scheduling diagram shown below?**



Activity	Normal time for activity (days)
A	3
B	6
C	5
D	3
E	4
F	5
G	3

a. 7 b. 9 c. 12 d. 13

Q19 **By examining each activity in the following arrow diagram, we found out that only activity D can be shortened by three days. As a result of this reduction in D, how many days can be eliminated from the necessary duration of the entire project? In the following diagram, a dotted-line arrow indicates a dummy activity.**



a. 0 b. 1 c. 2 d. 3

Q20 **Which of the following describes the M/M/1 queuing model correctly?**

a. Mean number of arriving customers has the exponential distribution
b. A customer may leave from the system before service is finished
c. Mean service time has exponential distribution
d. The length of the queue is finite
e. There may be a number of service windows

Q21 **Assume that there exists an online system with a single server. The mean number of arriving transactions in this system is 0.6 per second. The mean service rate is 750 milliseconds per transaction. How long is the mean response time (in seconds) for this system? Note that the mean time a transaction spends in the system (W) is represented as follows.**
$W = (\rho/(1-\rho)) \times Es$ ρ: Traffic intensity, Es : Mean service time

a. 0.45 b. 0.61 c. 1.25 d. 1.36

Q22 **Suppose there exists a transmission system that uses communication networks. By applying the queuing theory M/M/1 model to this system, the relationship among the mean waiting time in the queue, the mean transmission time, and the traffic intensity is represented as follows.**
mean waiting time in queue=mean transmission time × (traffic intensity / 1-traffic intensity)

If the traffic intensity(%) exceeds one of the following values, the mean waiting time in the queue becomes greater than the mean transmission time. Which is the value?

a. 40 b. 50 c. 60 d. 70 e.80

Q23 **Which of the following is the most appropriate as the description of inventory control based on ABC analysis?**

a. It is advisable to decide the reorder point for each group i.e. A, B, C in advance, from the statistical and mathematical point of view.
b. It is advisable to control the inventory of an individual item in group A. Because the holding cost of such items, although the number of items is small, is high.
c. It is advisable to control the inventory of items in group B as much as possible. Because this group has a relatively large number of items but their holding cost is small.

d. It is advisable to examine the requisition quantity and stock quantity of items in group C regularly to decide the order quantity.

Q24 Read the following descriptions concerning system capacity. Then fill in the blanks by choosing the appropriate words from the given list. A word in the list can appear in more than one space.

Consider a simple transaction processing model.
-Both the mean interarrival time and the mean service time have negative exponential distribution
-There exists a single service window. Transaction processing is done FIFO-based
Let λ = the mean number of arriving transactions per unit of time (i.e. the mean arrival rate)
μ = the mean number of transactions completing service per unit of time (i.e. the mean service rate)

(1) Mean queue length N is computed by the following formula
$N= \rho /(1- \rho)$ $\rho =$[A]

(2) The probability that at the time of a transaction arrival an existing transaction is receiving service is [B].

(3) If the response time is defined as the time between the transaction arrival and the service completion for the transaction, including the duration waiting in the queue, the mean response time T is calculated as follows.
T=[C]

(4) Let the mean arrival rate λ = 12 transactions per minute and
the mean service rate μ = 15 transactions per minute.
Then the mean queue length N is [D] transactions, and the mean response time T is [E] seconds.

(5) Assume that the mean arrival rate is 12 transactions per minute. In order to let the mean response time T be equal or less than 10 seconds, the mean service time μ should be equal or greater than [F] transactions per minute.

Answers for A through C
a. $\lambda \mu$ b. $\mu - \lambda$ c. $1/\lambda$ d. $1/\mu$
e. λ / μ f. μ / λ g. $\lambda /(\mu - \lambda)$ h. $\rho ^2/(1- \rho)$
i. $(1/\lambda)(\rho /(1- \rho))$ j. $(1/\mu)(\rho /(1- \rho))$

Answers for D though F
a. 3 b. 4 c. 5 d. 16 e. 18
f. 20 g. 36 h. 48 i. 60 j. 84

Q25 Read the following descriptions concerning inventory control. Then answer the sub questions 1 and 2.

Mr.Y works at a factory. The factory has decided that it should improve its inventory control. Their inventory control has been based on the experience and guessing of the persons in charge. Therefore the factory asked Mr.Y for his opinion concerning the new ordering method that applies to the raw materials used in the factory.

Sub question (1)
In the factory, there are 100 raw materials (items) used for production. Mr.Y came up with the idea to identify the most important items for production then he will propose the improvement of inventory control based on this identification system. He understood by referencing some books that inventory of items with high annual consumption costs (= unit price × annual consumption amount) have to be monitored in detail as critical items as compared with items with lower consumption costs. Therefore, as the first step, he

examined annual consumption cost for every item. He summarized the results in the following table.

material number	item code	unit price	annual consumption amount	annual consumption cost
1	AA01	30	56,380	1,691,400
2	AA07	200	1,500	300,000
3	AC01	1500	23,400	35,100,000
...	...	...	...	...
100	ZQ80	10	2,875	28,750
Total annual consumption cost				231,730,960

He then identified the items with high annual consumption costs by performing the following analysis.

i. Arrange all items in descending order according to the annual consumption cost
ii. Calculate the cumulative sums of annual consumption costs, starting from the highest one down to the lowest one.
iii. Create a bar chart with ordered items on the horizontal axis and annual consumption costs on the longitudinal axis.
iv. To the above bar chart, add a line representing the cumulative sums of annual consumption costs.

Which diagram/chart is suitable to express the result of this procedure?

Answer

a. Arrow Diagram b. Gantt Chart c. Control Chart
d. Pareto Chart e. Portfolio Chart

Sub question (2)

The commonly used inventory control methods include "periodic ordering" and "fixed quantity ordering" methods. In order to decide which method is more suitable for the important items identified in the sub question (1), Mr.Y summarized the characteristic of each method as follows. Fill in the blanks in the following descriptions by choosing correct answers from the given list.

To decide an ordering method for a raw material, various conditions should be taken into account, such as unit price of the raw material, amount consumed during a unit period, degree of variance in the consumption amount, ease of forecast, procurement period (i.e. time from order to storing), the ordering cost per order e.g. transportation cost, labor cost and the holding cost according to the stock quantity.

In case of the "periodic ordering" method, an order cycle of an item is pre-determined based on the item's procurement period. In every cycle, the consumption demand in the next cycle is forecasted and the [A] is decided based on this forecast. Thus inventory control of an item can be achieved up to a detail level. It does not have the risks of stock-out, even if the fluctuation of consumption amount is big. This method is useful to manage raw materials whose products' demand in the market changes drastically, products whose production plan changes frequently, and whose unit price is high.

In the case of the "fixed quantity ordering" method, a new order of an item with the pre-determined quantity is placed when the [B] is reached. The reorder quantity is decided by considering the item's procurement period. Thus, compared with the periodic ordering, this method is easier to manage but with some risks of stock-out if the variance in consumption is big. Therefore, this is a more suitable method to apply to control the inventories of the raw material whose product demand and procurement period are stable, and whose stock level can be precisely monitored.

Answer

a. Safety stock b. 50% of safety stock c. Order interval
d. Reorder point e. Order quantity f. Average stock quantity

Index

[Numerals]

1's complement 14
2's complement 14
9's complement 14
10's complement 14
3-period moving average 334

[A]

absolute path 110
access 56
access arm 60
access control 112
access right 58
access time 57, 63
account column 210
accounting information system 236
accounting period 206
accounts 210
accrual basis 227
accrual principle 226
accumulator 42
acid test ratio 233
active matrix type 85
actuator 89, 149
addition circuit 53
addition theorem 285
address 56
address bus 39, 71
address format 37
address modification 42, 43
address specification method 43
after-closing trial balance 213
agent 131
AI 131
ALU 37, 40
analysis of profitability 230
analysis of safety 232
AND circuit 51
AND operation 51
AR 131
arithmetic and logic unit 37
arithmetic operation 37
arithmetic shift 20
arithmetic unit 33, 37
arrow diagram 301
ASCII code 24
asset turnover 232
assets 207
audio representation 25
authoring tool 127, 130
auxiliary storage device 33, 56, 59
auxiliary units 5
availability 269
average search time 63
average seek time 63

[B]

back-end processor 142, 143
Backward computation 305
balance sheet 206, 207
balance sheet integrity principle 220
balance sheet principles 220
banking system 261
bar code 79, 258
bar code reader 79
base address register 42
base address specification 45
basic software 96
basic solution 299
batch processing 146
batch processing system 146
BCD code 11
BEP 143
bias 19
binary numbers 2
binary-coded decimal code 11
binomial distribution 292
Bipolar IC 34
bit 3
bookkeeping 209
borrow 6
borrowed capital 207
bottom-up decision-making system 257
boundary address method 106
branch 37
break-even analysis 234
break-even chart 235
break-even point 234
break-even sales revenue 234
B-to-B 264
B-to-C 264
buffer 57
bus 71
business accounting principles 220
byte 3

[C]

CAD 88, 253, 254
CAE 253, 254
CAI 87, 126
CALS 264
CAM 88, 253, 254
cancellation 23
CAP 254
capital equation 207
capital reserve 219
capital stock 219
CAPP 254
carry 3, 6
CASE tool 88, 119
cash basis 227
CCD 83
CD 66
CD-ROM 66
center batch processing 147
central processing unit 33
central tendency 289
Centronics interface 74
certain event 283
CG 131
character representation 23
CIM 255
clicking 80
client 137
client/server system 137
clock 55
closing adjustment 212
closing adjustment entries 212
closing day 206
cluster 109
CMOS IC 34
code 23
codes for information interchange 23
cold site 143
cold standby mode 141
command 110
Commercial Code 228
common key cryptosystem 272
communication server 139
Compact Disc 66
compaction 105
comparison 37
complement 14
complement register 43
complete survey 288
computer five main units 33
computer network 144
computer security 112
computer virus 278
confidentiality 269
Constraints 296
continuous random variable 291
control bus 71
control unit 33, 36
CORBA 119

corporation accounting principles 229
Corporation Tax Law 229
correlation analysis 327
cost of goods sold 222
cost slope 309
CPI 55
CPM (Critical Path Method) 301, 306
CPU 33
CRC code 58
critical path 306
Critical Path Method 306
cross section method 327
cross-check 140
CRT display 83
C-to-C 265
current assets 215
current directory 110
current liabilities 218
current ratio 233
cycle time 57
cyclical variations 328
cylinder 60

[D]

daisy chain 75
DASD 61
DAT 107
data bus 72
data transfer time 64
database server 139
DBMS 118
de facto standard 24, 74, 116
De Morgan's law 28
deadlock 104
debt ratio 233
debt/equity ratio 233
decimal arithmetic system 12
decimal operation 37
decimal operation mechanism 40
decoder 38
dedicated cluster 143
deferred assets 217
defragmentation 63
Delphi method 327
dependent event 286
depreciation and amortization 216
desk-top type 87
destruction 278
digital camera 83
digitalization 125
digitizer 82
direct access 61
direct access storage device 61
direct address specification 44
directory 109
disk mirroring 70
dispatcher 101
dispatching 101
display device 77, 83
distributed cluster 143
distributed processing system 144
DMA method 72
dot impact printer 85
double clicking 80
double precision 22
double-entry bookkeeping 209
dragging 80
DRAM 34
DSTN liquid crystal display 85
dual system 140
dummy activity 302
duplex system 141
DVD 68
dynamic address translation 107
dynamic allocation 106

[E]

earliest finish time 307
Earliest node time 304
earliest start time 307
eavesdropping 278
EBCDIC code 24
EC 264
ECC 58
econometric analysis 328
Economic Order Quantity 316, 317
EDI 263
EDIFACT 263
EEPROM 35
effective address 43
eight-column work sheet 212
electronic banking 262
electronic bulletin board 257
electronic mail 256
Electronic Ordering System 259
engineering workstation 88
EOR circuit 52
EOR operation 52
EOS 259
EPROM 35
equity turnover 232
error 23
Error Correcting Code 58
event 283, 301
Excess 64 17
exclusive control 103, 149
exclusive events 284
exclusive logical sum operation circuit 52
exclusive OR 27
executable status 101
execution control 101
execution control of the instruction 46
execution cycle 46
execution order 103
expenses 208, 222
experiment-based probability 285
exponent 5
exponential curve 332
exponential smoothing method 335
extension 111
external analysis 230
external interrupt 102
extra stock loss 316
extraordinary gain 222
extraordinary loss 224

[F]

F cycle 46
FA 254
fail-safe system 269
fail-soft system 269
failure recovery 149
falsification 278
fault tolerant system 142
FEP 143
fetch cycle 46
FIFO method 108
file server 139
financial accounting 228
financial statements 206
firewall 272
firm banking 262
first-in first-out method 223
fiscal year 206
fixed assets 216
fixed assets to long-term equity ratio 233
fixed cost ratio 234
fixed costs 234
fixed disk 59
fixed liabilities 218
fixed point 16
fixed point operation 37
fixed point operation mechanism 40
fixed quantity ordering method 318, 320
fixed ratio 233
flag register 43
flash memory 35, 83
flip-flop 35
flip-flop circuit 54
floating point 17
floating point operation 37
floating point operation mechanism 40
floppy disk 65
floppy disk unit 64

flow control 112
FMC 255
FMS 255
folder 111
Forward computation 304
fragmentation 63, 105
free float 308
free software OS 117
Frequency distribution 288
frequency table 288
front-end processor 142, 143
FTP 138
full adder circuit 54

[G]

general ledger 210
general-purpose computer 88
general-purpose register 37
GPIB 75
Green Book 68
gross amount principle 220, 226
gross income 225
gross income on sales 225
groupware 256
growth curve 333
GUI 81, 127, 128

[H]

half-adder circuit 53
hamming code 58
hard copy 77
hard disk 59
hardwired-logic control system 47
hexadecimal system 4
holding cost 315
home directory 110
horizontally distributed configuration 145
hot site 143
hot standby mode 141
human interface 114, 116, 125
HyperText 130

[I]

I cycle 46
IC 34
IC card 80
IC memory 34
icon 81, 127
IEEE1394 74
illegal input 278
image representation 25
image scanner 82
image sensor 82
immediate specification 44
impact printer 85
income (profit and loss) statement 206
income planning 234
income statement equation 208
income statement integrity principle 226
income statement principles 226
independent events 285
independent trial 287
index address specification 44
index register 42
indirect address specification 45
infinite binary fraction 8
information security 112
ink-jet printer 86
input device 76
input unit 33
input/output interface 73
instruction cycle 46
instruction execution control 39
instruction register 37, 38
instruction set 45
intangible fixed assets 216
integrated circuit 34
integrity 269
interactive processing system 150
interactivity 125
inter-industry relations analysis 328
internal analysis 230
internal interrupt 102
internally programmed system 56
International Organization for Standardization 24
interruption 102
intersection of events 284
inventories 215
Inventory Control 315
investments 217
IR 38
IrDA 74
irregular variations 328
ISO 24
ISO code 24

[J]

Japanese Industrial Standards 24
Japanese language processing 114
JCL 99, 147
JIS 24
JIS 7-bit code 24
JIS 8-bit code 24
JIS code 24
job control language 99, 100, 147
job scheduling 100
journal 210
journalizing 210
joystick 81

[K]

Kendall notation 312
kernel 102, 115
keyboard 78
keylock method 106

[L]

LAN control 118
lap-top type 87
laser printer 86
last-in first-out method 223
latest finish time 307
Latest node time 304
latest start time 307
law of large numbers 285
LCMP 142
leading index method 327
ledger 210
liabilities 207
light pen 81
line printer 85
Linear Programming 296
Linux 117
liquid assets 215
liquid crystal display 84
load 107
logarithmic curve 333
logical circuit 50
logical operation 26, 37, 50
logical product 27
logical product operation 51
logical shift 21
logical sum 27
logical sum operation 51
logical symbol 26
long-term safety ratios 233
loosely coupled multiprocessor system 142
loss of information 23
LP theorem 299
LRU method 108

[M]

MacOS 117
macro virus 278
magnetic card reader 80
magnetic disk 60
magnetic disk unit 59
magneto optical disk unit 69
main storage unit 33, 56
mainframe 88
management accounting 229
manufacturing process control 254

marginal profit 235
marginal profit chart 236
marginal profit ratio 235
marksheet 79
Mask ROM 35
Mathematical Probability 284
mean arrival rate 312
mean number of transactions in queue 314
mean number of transactions in the system 313
mean service rate 313
mean time transaction spends in queue 314
mean time transaction spends in the system 313
Mean value 289
Median 289
memory 56
memory hierarchical structure 57
memory leak 106
memory protection 58, 106
method of least squares 330
method of selecting a best-fitted line 329
method of selection by visual observation 330
microcomputer 89
microprogramming control system 47
middleware 118
MIL symbol 51
MIPS 55
mirror site 143
MIS 254
MO 69
mobile computing 126
Mode 289
mouse 80
moving average 334
moving average method 223
MRP 254
MS-DOS 116
MTTR 112
multi scan monitor 84
multimedia 125
multimedia operating system 128
multimedia processing 114
multimedia service 125
multimedia system 125
multimedia title 127, 129
multiprocessing 101
multiprocessing function 116
multiprocessor system 142
multi-programming 97, 103
multitasking 101
multi-user function 115
MVS 114

[N]

NAND circuit 53
NAND operation 53
negation 27
negation operation 52
negative AND 28
negative logical sum 28
net income 208, 225
net income before taxes 225
net loss 208
networking 125
NFS 138
node 301
non-impact printer 85
non-operating expenses 224
non-operating revenue 222
nonvolatility 35
NOR circuit 53
NOR operation 53
normal distribution 295
normal operating cycle rule215, 218
normalization 18
NOT circuit 52
NOT operation 52
notebook type 87
n-tier architecture 139
null event 283

[O]

objective function 297
Objective function 296
OCR 78
OCR font 78
OLTP 148
OMR 79
one-time password 272
one-year rule 214, 218
online transaction processing system 148
open system 138
operating income 225
operating revenue 222
operating system 96
operation planning 254
optical character reader 78
optical disk 66
optical disk unit 66
optical mark reader 79
optimal inventory 317
OR circuit 51
OR operation 51
Orange Book 68
ORB 119
ordering cost 315
ordinary income 225
OS 96
OSI basic reference model 145
other current assets 216
output device 76, 77
output unit 33
overflow 16, 23
overlay 106
owner’s equity 207
owner's equity ratio 233
owner's equity to fixed asset ratio 233

[P]

packed decimal format 13
page 107
page frame 107
page in 107
page out 107
page printer 85
page replacement 108
paging 107
paging algorithm 108
palm-top type 87
parallel interface 74
parallel transfer 74
Pareto diagram 320
parity check system 58
partition method 104
passive matrix type 85
path 110
pattern recognition 131
PDA 126
periodic ordering method 318, 325
peripheral device 33
personal computer 87
PERT (Program Evaluation and Review Technique) 301
PERT network 301
PERT Time Computation 304
platform 127, 138
pointing device 80
population 288
POS information management system (system to manage information at the point of sale) 257
POS system 257
posting 210
precision 22
precision of fixed point representation 23
precision of floating point representation 23
preemption 101
preemptive multi-tasking 128
principle of matching costs with revenues 227
print server 139

printer 77, 85
Probability 284
probability distribution 290
probability of exactly *n* transactions in the system 314
process 115
process control computer 90
process design 254
processing unit 33, 36
processor 33, 36
procurement cost 315
product design 254
production planning 254
profit and loss chart 235
profit reserve 219
program counter 41
program relocation 105
PROM 35
proposition 26
proposition logic 26
protective boundary register system 58
protocol 129
PSW 43
public key cryptosystem 272

[Q]

queue 310
Queuing Theory 310

[R]

radix 5
radix conversion 7, 8
RAID 69
RAM 34, 56
random variable 290
RAS 269
ratio approach 230
ratio of expenses to sales 231
ratio of income to sales 231
ratio of income to stockholders' equity 231
realization basis 227
real-time control 149
real-time monitor 150
real-time processing system 149
Red Book 68
redundancy system 269
reentrant program 149
refreshing 34
register 41
register address specification 44
regression analysis 327
relative address specification 45
relative path 110
relative ratio method 230
remote batch processing 147
remote job entry 147
reorder point method 320
research and development 254
response time 98
revenue 208, 221
ring protection method 106
risk analysis 274
risk control 277
risk evaluation 274
risk finance 277
risk management 273
RJE 147
roll in 105
roll out 105
ROM 35
root directory 109
round robin 101
rounding error 18
router 272
RS-232C 73

[S]

safety stock 323
sales revenue 222
sample 288
sample space 283
sampling survey 288
schedule management software 257
scheduling 254
Scheduling 300
screen saver 84
SCSI 75
SDRAM 35
search time 63
seasonal adjustment method 329
seasonal variations 328
section and arrangement principles 220
section principle 226
sector 62
sector type 62
Securities and Exchange Law 229
seek time 63
segment 106, 108
selling, general and administrative expenses 223
semaphore 103
semiconductor disk unit 59, 69
semiconductor memory 34
sequential access 61
sequential circuit 50, 54
sequential control system 36
serial interface 73
serial printer 85
serial transfer 73
server 137
shell 115
shell script 115
shift 18
shift JIS code 24
shift operation 20
shortage cost 316
short-term safety ratio 233
simple event 283
simple execution time 103
simplex system 140
simultaneous execution control 149
single precision 22
single-address format 37
six-column work sheet 212
slashing 107
slip 210
slot 107
SNMP 117
soft copy 77
SPOOL 148
SRAM 35
stand-alone system 87
Standard Deviation 290
state transition 101
statements of accounts 206
statistical probability 285
Statistics 287
STN liquid crystal display 85
stockholders' equity 207
stock-out rate 316
storage cost 315
storage unit 33
stored procedure 139
stored-program system 56
structural ratio method 230
sub-directory 109
supercomputer 89
supervisor 113
supervisor call 113
surplus 220
SVC 113
swap in 105
swap out 105
swapping 105

[T]

tablet 83
tandem multiprocessor system 142
tangible fixed assets 216
TCMP 142
terminator 75
TFT liquid crystal display 85
thermal printer 86
thermal transfer printer 86
three time estimates 309
three-address format 38
three-tier architecture 139

tightly coupled multiprocessor system 142
time-series analysis 327
titles of account 210
TLB system 58
total float 308
total inventory cost 315
touch panel 82
touch screen 82
touch typing 78
track ball 81
traffic intensity 313
transaction 209
Translation Look-aside Buffer system 58
trend method 230
trend variations 328
trial balance 210
trial balance of balances 211
trial balance of totals 210
trial balance of totals and balances 211
truth table 26, 51
TSS 103, 150
two bin method 318
two-address format 38
two-tier architecture 139

[U]

UHD 65
UN/EDIFACT 263
unappropriated retained earnings 220, 225
underflow 23
unicode 25
union of events 283
UNIX 115
unpacked decimal format 12
USB 73
user authentication 272
user interface server 139
user programmable ROM 35

[V]

value approach 230
variable cost ratio 234
variable costs 234
variable type 62
Variance 290
variate 288
vector processor 89
Venn diagram 51
vertically distributed configuration 145
videoconferencing 257
virtual reality 126
volatility 34
voluntary reserve 220
VR 131

[W]

wait status 101
Web computing 151
wild card 111
Windows 116
Windows NT 116
word 3
work sheet 212
workstation 88

[Y]

Yellow Book 68

[Z]

zero-address format 37
Zip 65
zoned decimal format 12