Review

1.Introduction

1.Digital Computer Classification

According to the scale, the function and the speed,etc, general purpose computer can be divided into**

> Supercomputer
>
> Mainframe
>
> Mediumcomputer
>
> Minicomputer
>
> Microcomputer
>
> Single-chip microcontroller

2.Computer system

The computer system includes two parts: hardware and software.

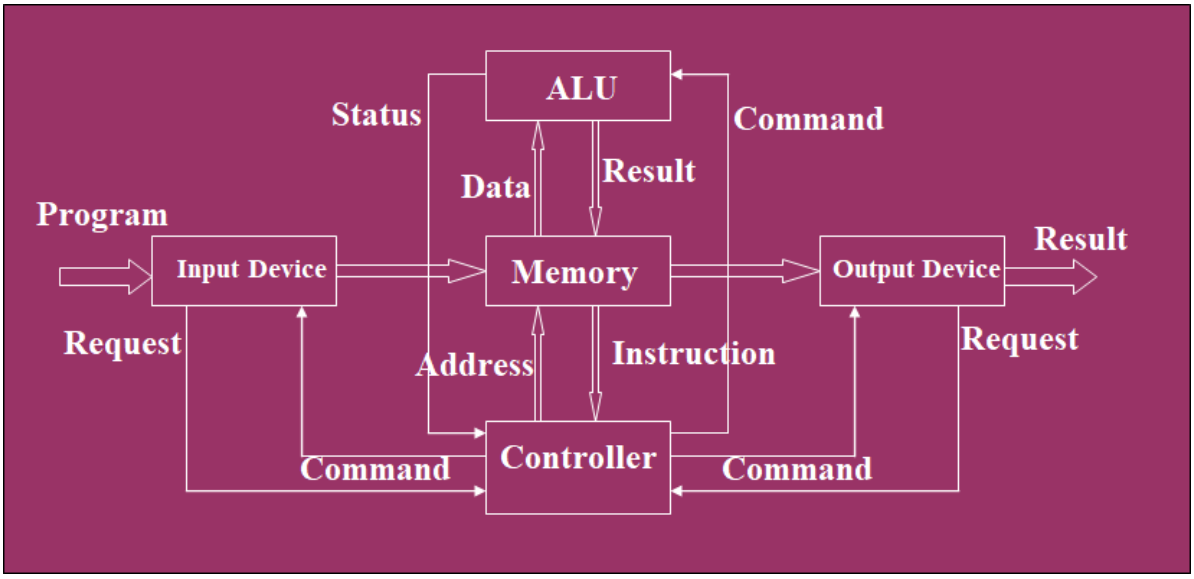
3.Machine basic word length

The basic amount of data digits which are participating in computer operation.

4.Von Neumann computer

Basic features of a Von Neumann computer: Instructions and data are represented in binary bits; Computers work in the form of stored program; Computer hardware consists of memory, arithmetic logical unit, controller, input device and output device

5.Computer Organization Diagram



6.Main Components of Computer

> 1) Input Device
>
> Input raw information into computer and convert these information into machine language.
>
> 2) Output Device
>
> Output the processing results in a familiar form which can be easily identified by people.
>
> 3) Memory
>
> Store program and data.
>
> Usually level 3 memory hierarchy.
>
> 4) ALU
>
> Arithmetic and logical calculation.
>
> 5) Controller
>
> Produce a series of control signal to command the entire computer system in an orderly way to run automatically.
>
> Combinational Logic Control Unit and Microprogram Control Unit

7.Computer System Software

1.System software

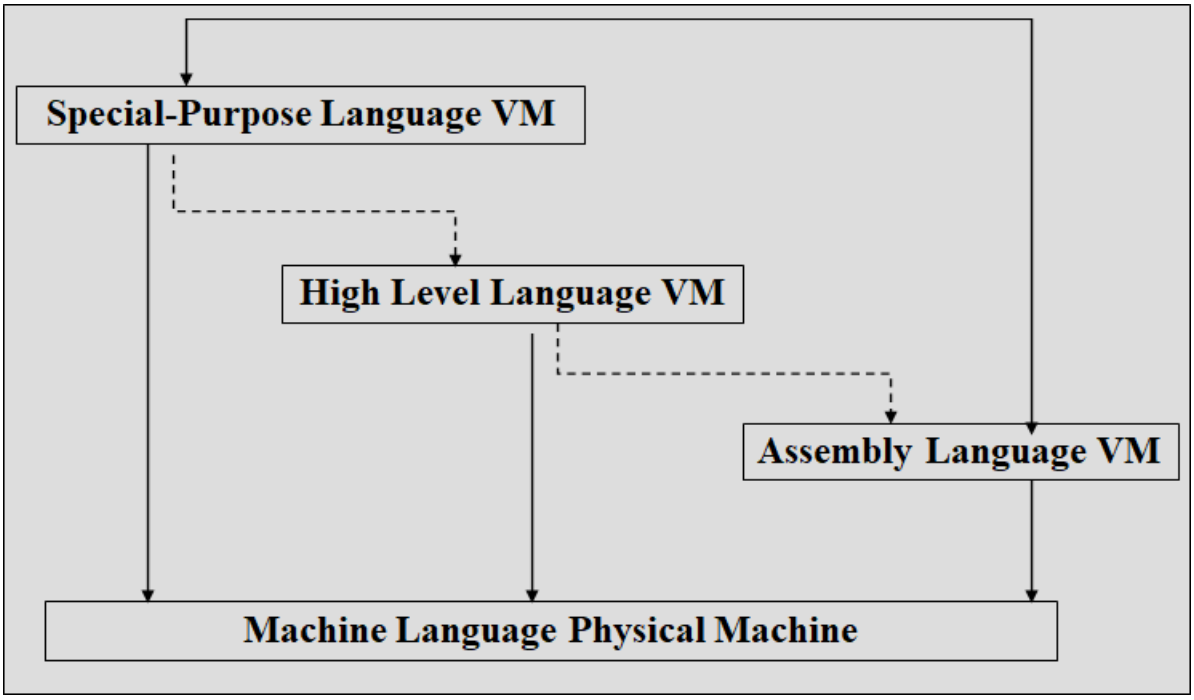
2.application software

8.Instruction System

The instruction system is the set of all the instructions that a computer can execute.

9.Machine language& assembly language

The function of computer hardware is to execute machine language, which can be directly recognized by the computer hardware. Assembly language is a kind of low- level programming language for machine structure.



10.Main Memory Capacity

Computer with word as unit:8192×16

Computer with byte as unit:1024B(1KB)

Main memory capacity is usually restricted by the length of addressing code. For example, if the length of addressing code is 16 bits, the amount of memory units which can be directly accessed is 2^16=65536

2.Data Information Representation

1.Counting System and Interconversion

1.Counting System

if N = N_n-1 N_n-2 **... N_0 • N-*1 N*-2 ... N_-m

$$(N)_R = \sum_{i=0}^{n-1} N_i R^i + \sum_{i=-1}^{-m} N_i R^i = \sum_{i=n-1}^{-m} N_i R^i$$

2.Interconversion

1.Convert R counting system into Decimal System

$$\sum_{i=n-1}^{-m} N_i R^i$$

2.Convert Decimal System into R counting system

(a)Conversion of integer part

Divide the decimal number by R to take the remainder until the quotient is equal to 0 so far. Each remainder belongs to the R counting system (the remainder obtained at the first time is the least significant bit).

(b)Conversion of decimal part

Multiply the decimal number by R to take the integer part until the decimal part is equal to 0 or the claimed precision. (the integer part obtained at the first time is the most significant bit).

2. Decimal Point Notation

1.Fixed Point Notation

If the amount of binary bits is n(not including sign bit)

$-(1-2^{-n})\leq N\leq 1-2^{-n}$

The Number Range of Fixed Point Integer

$-(2^{n-1})\leq N\leq 2^{n-1}$

2.Floating Point Notation

$N=R^E \cdot M=\pm R^{\pm e}\cdot M$

(E----Exponent, M----Mantissa, R----Radix)

> 1.Number Range
>
> if l indicates the amount of exponent digits and n indicates the amount of mantissa digits,
>
> $-(2^{(2^l-1)})(1-2^{-n})\leq N\leq (2^{(2^l-1)})(1-2^{-n})$
>
> 2.Normalization Floating Point Number
>
> Positive number: 0.1xx...x
>
> Negative complement number :1.0xx...x

3.Sign Notation

True form, one's complement, two's complement and biased exponent

dddd

1.The number range of two's complement

(1)The integer N with n bits(not including sign bit)

$-2^n\leq N\leq 2^n-1$

(2) The decimal N with n bits(not including sign bit)

$-1\leq N\leq 1-2^{-n}$

2.Floating Point Number with two's complement

(1) The number range of floating point number
Exponent is l bits，Mantissa is n bits (not including sign bit, two's complement representation)
(-1）2^(2^l-1)≤N≤2^(2^l-1)(1-2^-n)

4.Normalization Representation of Floating Point Number

5.Biased Exponent

1. Definition
If exponent is n+1 bits(including sign bit),the number range of exponent is -2n~+(2n-1)，then
[x]_B = 2^n + x ， -2^n≤x ≤2^n-1

dddd

6.Practical Floating Point Format

IEEE754 Standard Floating Point Format
Three parts: Sign bit(S), Exponent(E) and Mantissa(M).
Four basic formats
(1) Single precision(32bits)：E=8,M=23
(2) Extension single precision：E≥11,M≥31
(3)Double precision(64bits)：E=11,M=52
(4)Extension double precision：E≥15,M≥63

3.Numerical Calculation and ALU

1.Arithmetic and Logical Operation Basis

1. Fixed point add and subtraction

1) True form addition and subtraction
2) two's complement addition and subtraction
Basic formula：
[x + y]TC = [x] TC + [y] TC(mod M)
[x - y] TC = [x] TC + [-y] TC (mod M)

[y] TC → [-y] TC: All digits reverse and end bit plus one.

2. Overflow Detection

1) Testing according to the sign bit

If A_n, B_n are the sign bit of two operands respectively, S_n is the sign bit of the result,

$$\text{Overflow}=A_nB_n\overline{S_n} + \overline{A_n}\overline{B_n}S_n$$

63+66=129	(-63)+(-66)=-129
0,0111111	1,1000001
+0,1000010	+1,0111110
1,0000001	10,1111111

2) Testing according to the carry bit

If C_n is the carry bit of the sign bit， C_{n-1} is the carry bit of most significant bit,

$$\text{Overflow}=C_nC_{n-1} + \overline{C_n}\overline{C_{n-1}}=C_n \oplus C_{n-1}$$

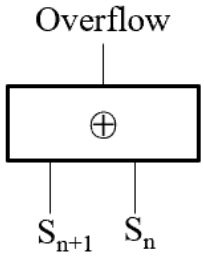
63+66=129	(-63)+(-66)=-129
0,0111111	1,1000001
0,1000010	1,0111110
1,0000001	10,1111111

3) Testing according to the double sign bit

- **Positive number: 00. $x_1x_2\ldots x_n$**
- **Negative number: 11. $x_1x_2\ldots x_n$**

If S_{n+1}、 S_n are the first sign bit and the second sign bit respectively,

Overflow = $S_{n+1} \oplus S_n$



63+66=129
00,0111111
+00,1000010
01,0000001

(-63)+(-66)=-129
11,1000001
+11,0111110
110,1111111

3.Shift

1)Logical Shift
Left shift: low bit fill 0 Right shift: high bit fill 0
Example: A=10110101
SHR one bit: 01011010
SHL one bit: 01101010

2)Cyclic Shift
Example: A=10011001
ROR one bit: 11001100
ROL one bit: 00110011

3)Arithmetic Shift
The sign bit is fixed, the value changes.
SAL one bit (*2)
SAR one bit (/2)
Arithmetic Shift Rules:

(1) Positive number: if shift left or right, vacancy fill 0
Example: A=0.0110
SAL one bit:0.1100 SAR one bit: 0.0011

(2) Negative number:
True form: maintaining the sign bit ,vacancy fill 0
Example: A=1.0110
SAL one bit: 1.1100
SAR one bit: 1.0011
Two's complement: if shift left, vacancy fill 0; if shift right, vacancy fill 1.
Example: A=1.1011
SAL one bit: 1.0110
SAR one bit: 1.1101

4. Logical Operation

1.Logical or (∨) Logical and(∧) Logical xor(⊕)

or:有1得1　and:有0得0　xor:不同得1，相同得0

2.Set the certain bit

```
    10010010 A
V   00000001 B
  ————————————
    10010011 A
```

3.Clear the certain bit

```
    10010010  A
∧   01111111  B
  ————————————
    00010010  A
```

4.Comparison

If A=10010010,B=10010011,compare A and B.

```
  10010010
⊕ 10010011
————————————
  00000001
```

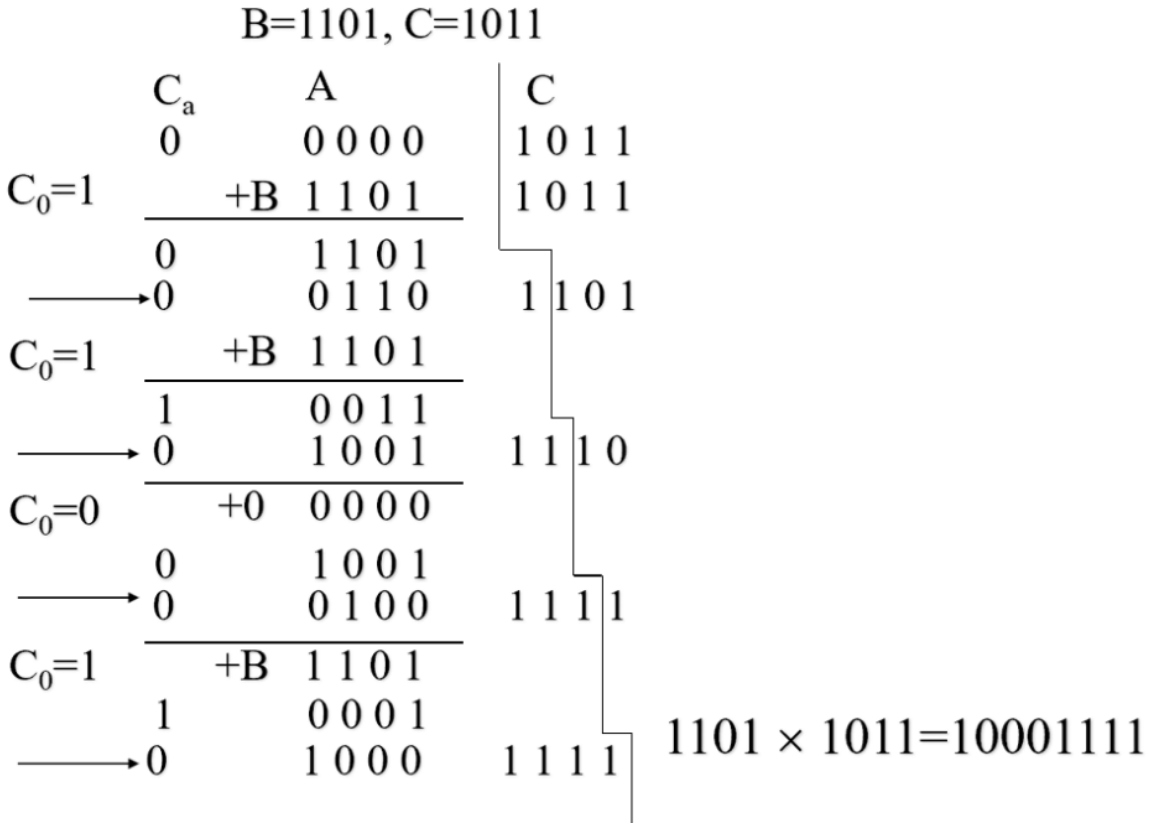
If the result is 0,then A equals B.

2.Fixed Point Multiplication

1. Unsigned Number Multiplication

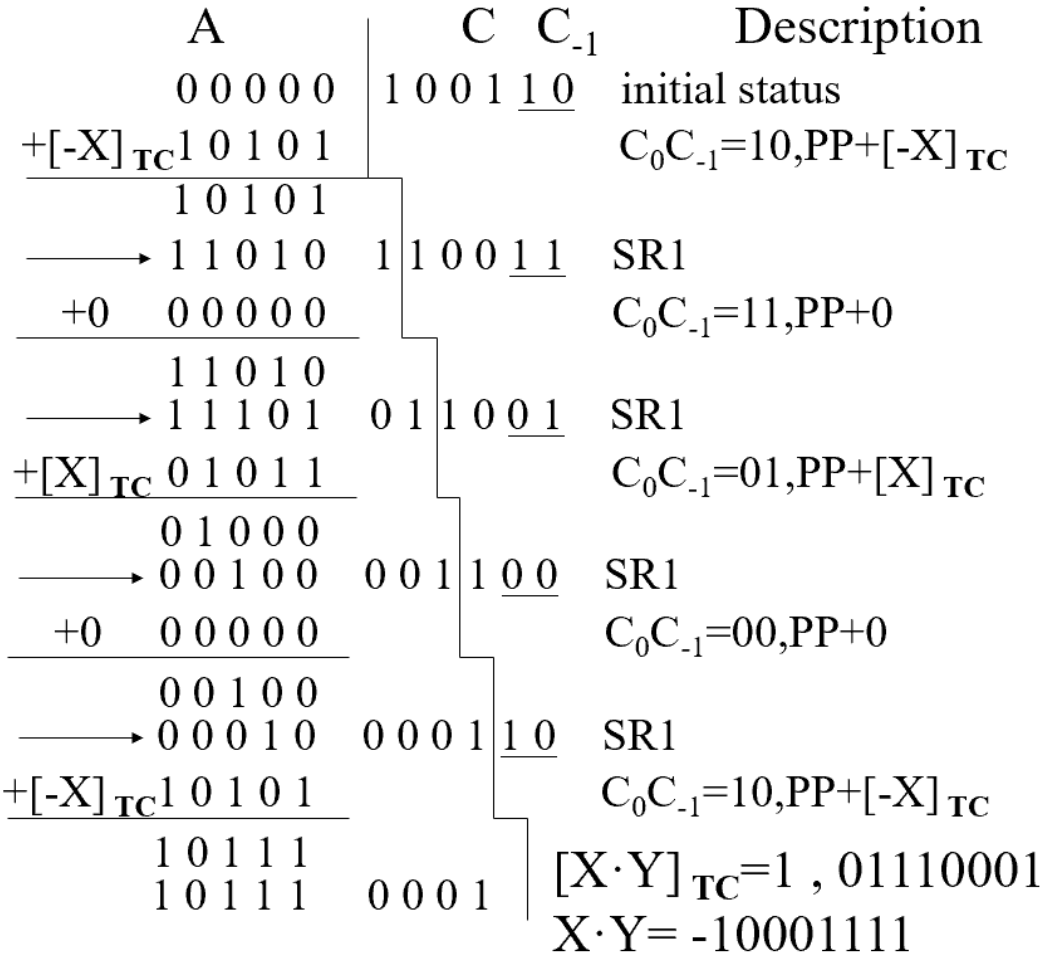
Example：　x=0.1101　　y=0.1011

$$
\begin{array}{r}
0.1101 \\
\times\ 0.1011 \\
\hline
1101 \\
1101 \\
0000 \\
1101 \\
\hline
0.10001111
\end{array}
$$

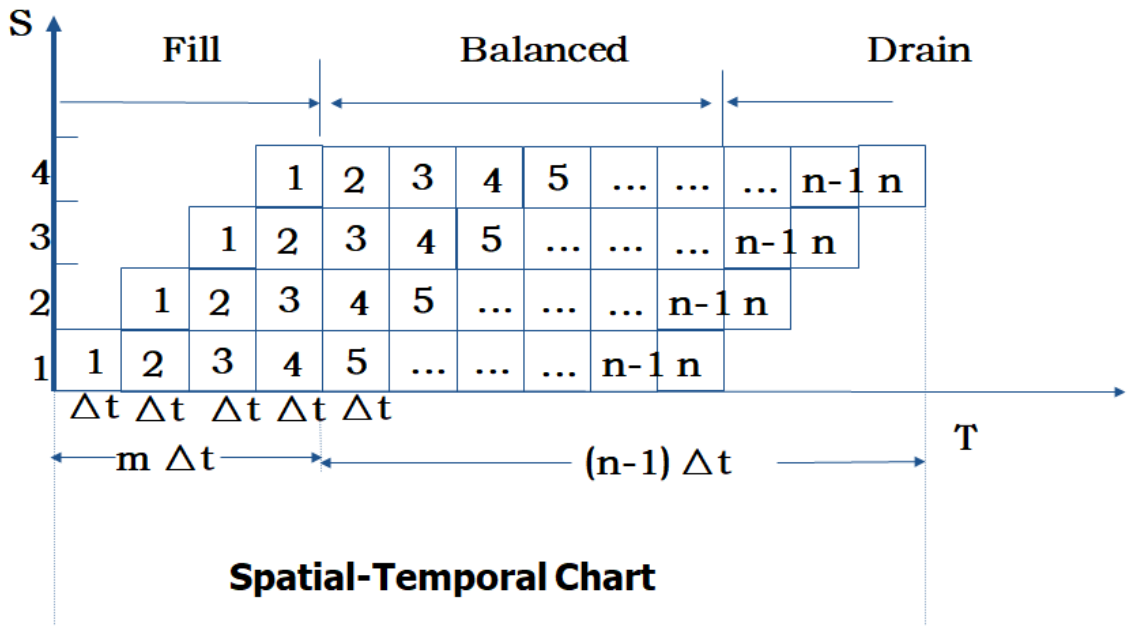


2.Signed Number one-bit Multiplication

Booth Algorithm



4.Pipelining Technology



Spatial-Temporal Chart

1. Performance Indicators

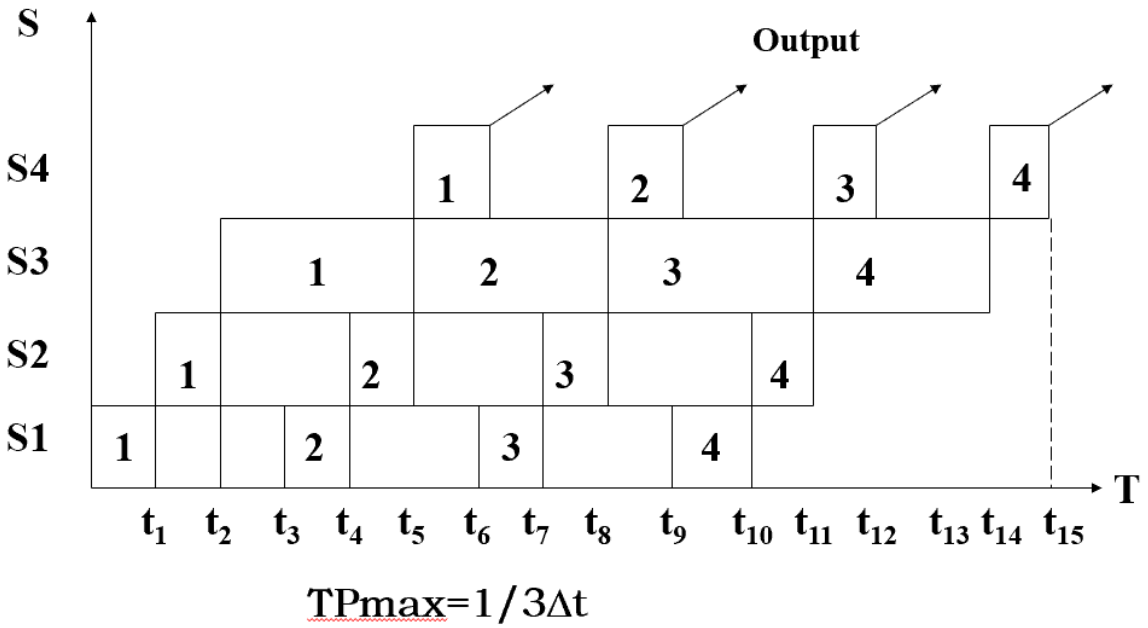
1)TP(Throughput Rate):

The count of tasks can output from Pipelining per unit time.

Maximum Throughput Rate:

TPmax=1/□t (ideal condition)

TPmax=1/ max{□t1，□t2，□t3，□t4}



TPmax=1/3Δt

$$TP = \frac{n}{m\Delta t_0 + (n-1)\Delta t_0} = \frac{1}{\Delta t_0(1 + \frac{m-1}{n})} = \frac{TP\max}{1 + \frac{m-1}{n}}$$

(2) Efficiency

Each Segment Time is Equal

$$\eta = \frac{m \cdot n \cdot \Delta t_0}{m \cdot T} = \frac{n \cdot \Delta t_0}{m\Delta t_0 + (n-1)\,\Delta t_0}$$

Each Segment Time is Not Equal

$$\eta = \frac{n * \sum_{i=1}^{m} \Delta t_i}{m * \left[\sum_{i=1}^{m} \Delta t_i + (n-1)\Delta t_j\right]}$$

Δt_j is the processing time of bottleneck segment

(3) Speedup Ratio

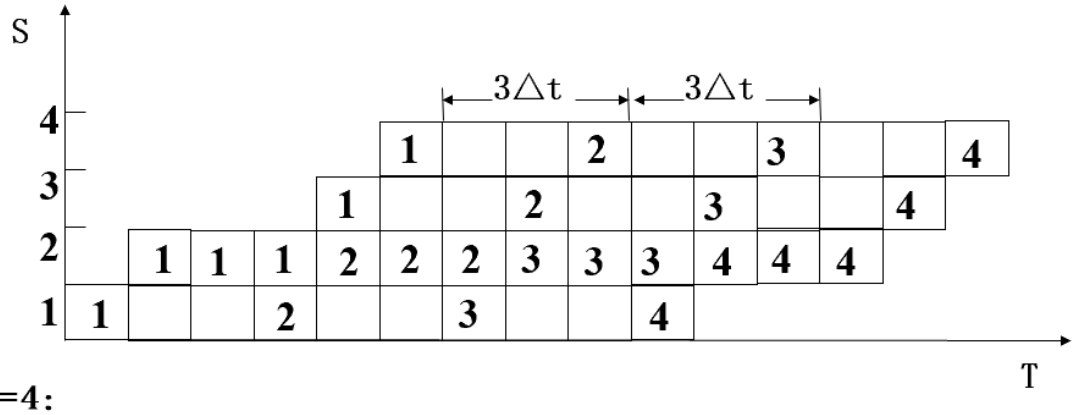
Each Segment Time is Equal

$$Sp = \frac{T_{\text{non-Pipelining}}}{T_{\text{Pipelining}}} = \frac{n*m*\Delta t}{m*\Delta t + (n-1)*\Delta t} = \frac{n*m}{m+n-1} = \frac{m}{1+\frac{m-1}{n}}$$

Each Segment Time is Not Equal

$$Sp = \frac{n * \sum_{i=1}^{m} \Delta ti}{\sum_{i=1}^{m} \Delta ti + (n-1)*\Delta tj}$$

(2) Spatial-Temporal Chart Method



n=4:

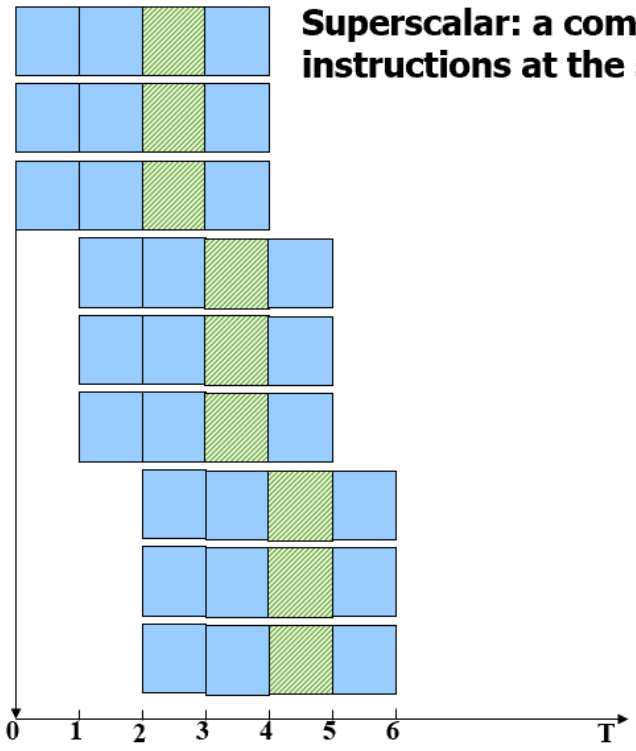
TP=4/((6+3*3) △t)=4/(15△t)=0.267/△t

η=6*4 △t /(4*15△t)=2/5=40%

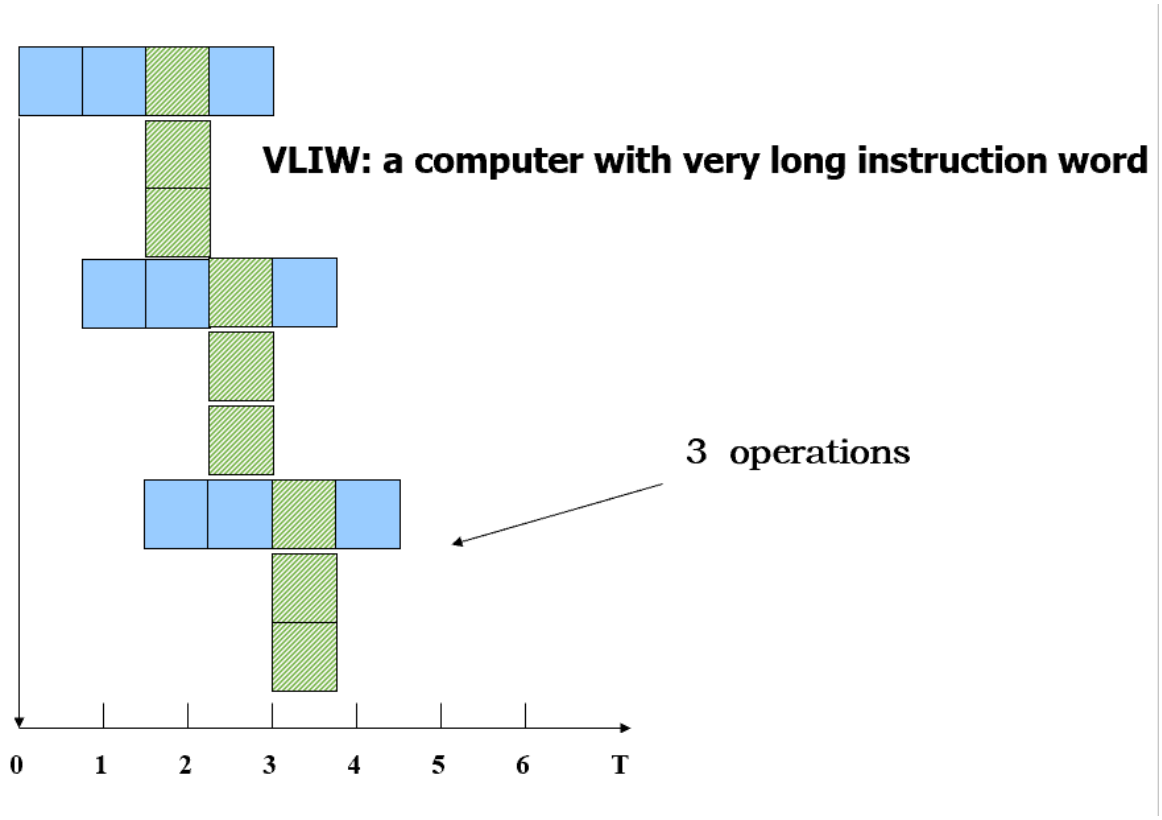
Sp=4*6△t /15△t=8/5=1.6

2.To Further Develop Instruction Parallelism

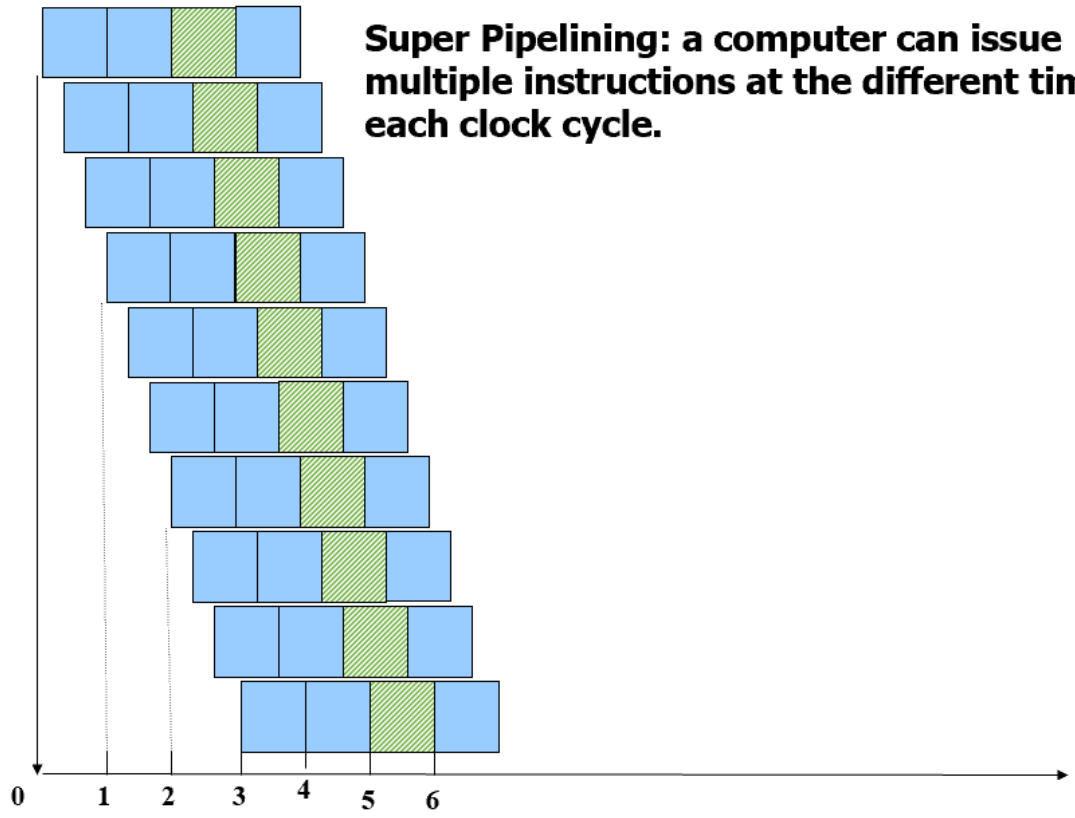
1)Superscalar



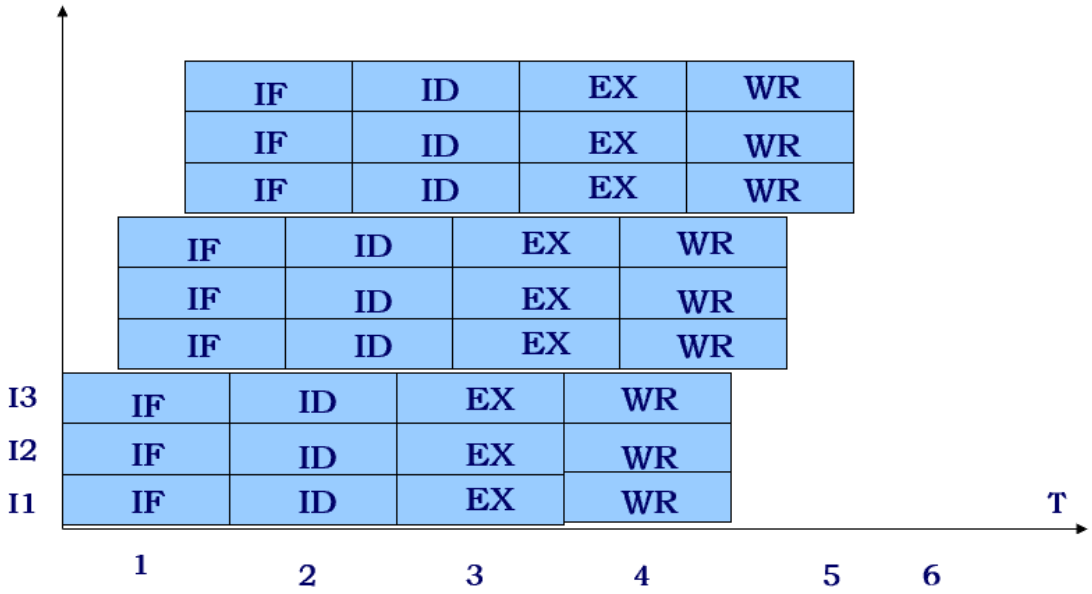
2)VLIW



3)Super Pipelining



4)Superscalar & Super Pipelining



5.Storage System

1.Storage System

1.Memory classification

1) Memory medium: Semiconductor, Magnetic, Optical, etc.

2) Easy to save: Volatile type, Non-volatile type.

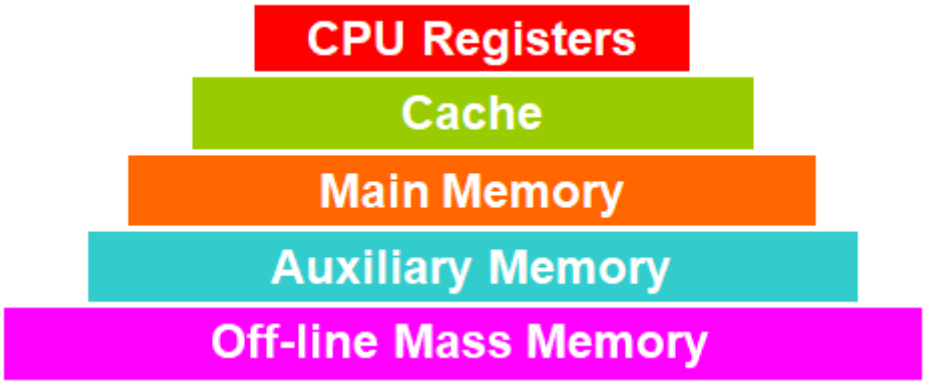
3) Access mode: RAM, ROM, SAM, DAM

4) Function: Main memory, Auxiliary memory, Cache and Control memory

2.Memory Key Performance Indicators:

Storage capacity, Access speed and Cost per bit.

3.Storage system hierarchy



2.RAM Chips

Device types: Bipolar and MOS
Storage Technology: SRAM and DRAM
SRAM: Fast access speed, Low integration, Large power consumption
DRAM: Slow access speed, High integration, Small power consumption

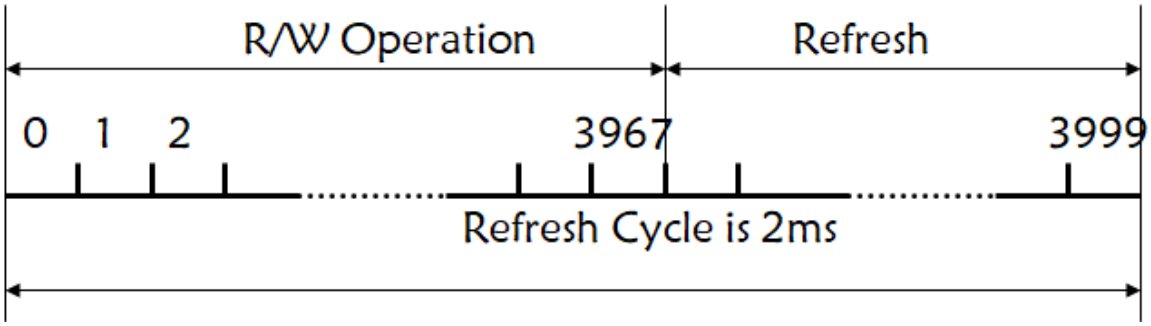
(1)SRAM Chip

(2)DRAM Chip

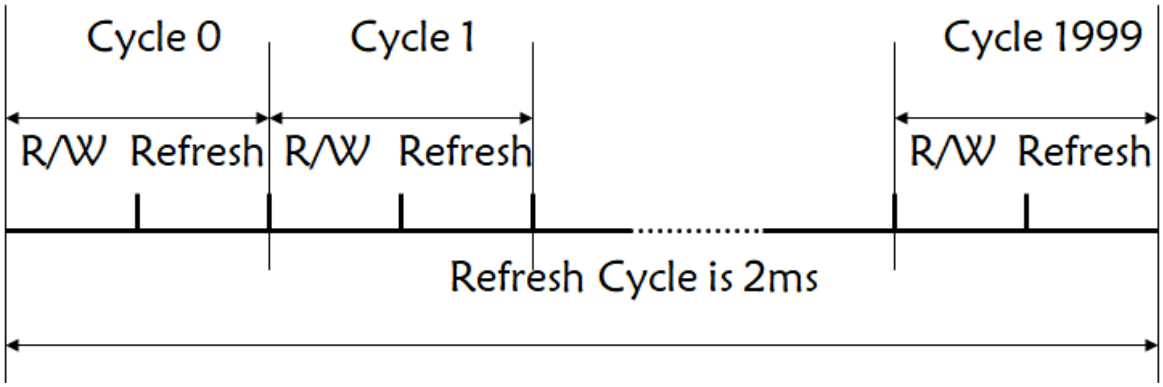
1) DRAM Refresh

Refresh: All storage capacitors are recharged every 2ms.
Maximum Refresh Cycle: The maximum interval of refreshing all storage capacitors.

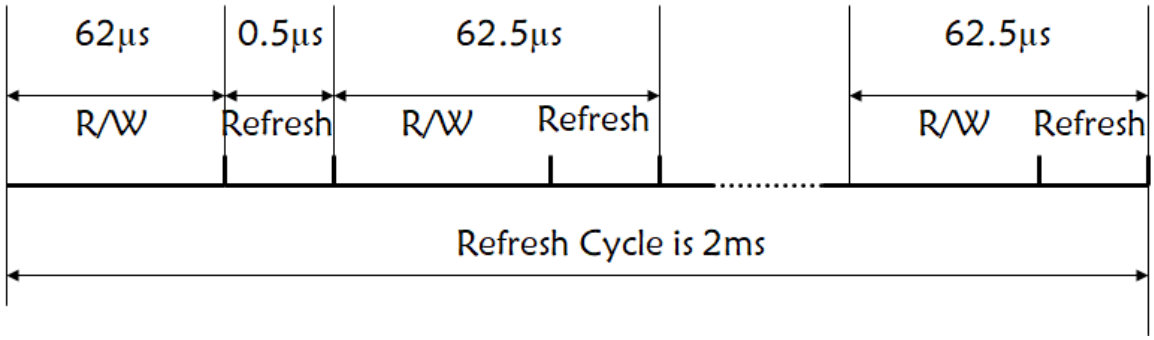
1Centralized Refresh



2Scattered Refresh



3Asynchronous Refresh



(3)Storage Capacity Expansion

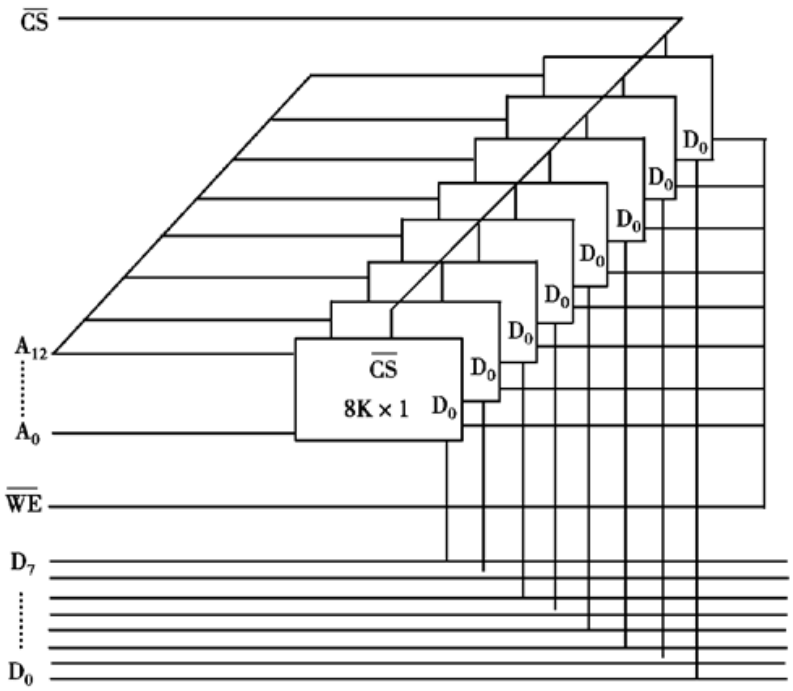
1) Bit Expansion

For example: $1M \times 1 \longrightarrow 1M \times 8$

- The data is joining together and sharing $\overline{CS}$

2) Word Expansion

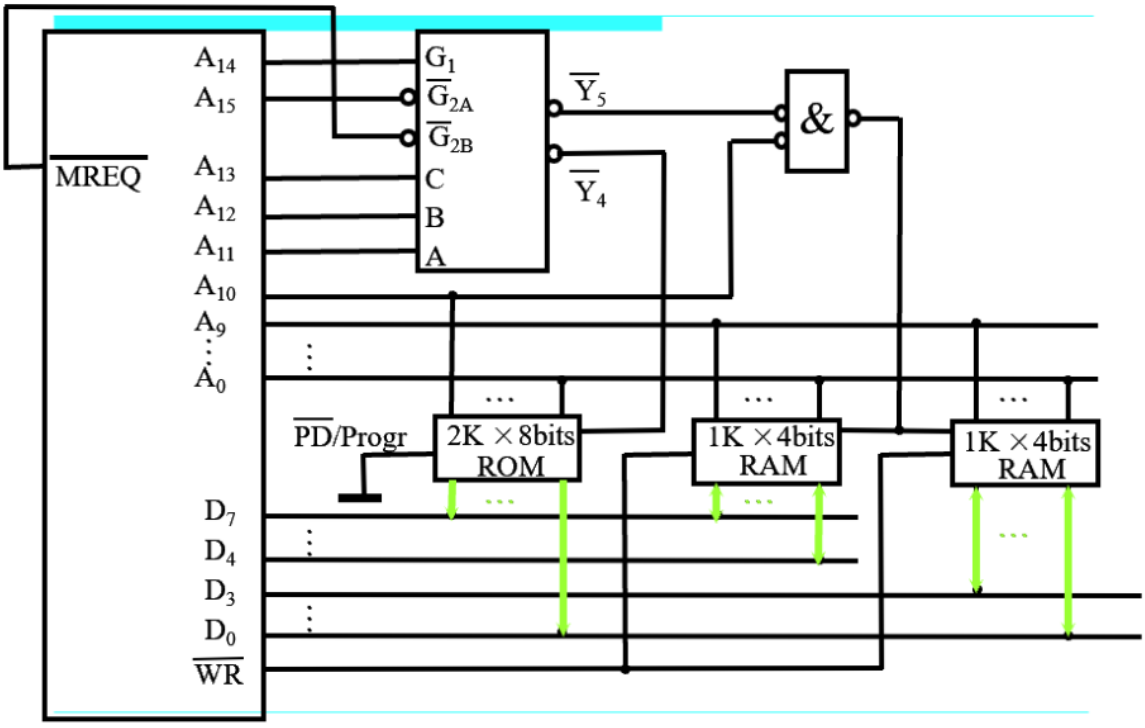
High addresses are decoded to produce several different chip select signals, according to each chip in the storage space distribution. Low addresses are decoded to select directly a certain storage cell in the chip .



how to solve:

(1) Write the corresponding binary address code
(2) Determine the number and type of chips
(3) Assign address line
(4) Determine the chip select signal

Example CPU and memory connection diagram



3.Hamming code composition

2 . Hamming code composition

The three elements that make up the Hamming code:

① The newly added detection bit *k* should satisfy:

$$2^k \geq n + k + 1$$

② Where should the detection bit located?

$$2^i \ (i = 0,\ 1,\ 2,\ 3,\ \cdots k\text{-}1)$$

③ What the value should be ?

The value of the detection bit along with its 'group' is related to the parity task.

Each detection bit C_i is responsible for group G_i

C_1 responsible for g_1 including: 1，3，5，7，9，11，⋯ bits

C_2 responsible for g_2 including: 2，3，6，7，10，11，⋯ bits

C_4 responsible for g_3 including: 4，5，6，7，12，13，⋯ bits

C_8 responsible for g_4 including: 8，9，10，11，12，13，14，15，⋯

g_i contains 2^{i-1} bit alone

g_i and g_j common contain $2^{i-1}+2^{j-1}$ bit

g_i、g_j and g_l common contain $2^{i-1}+2^{j-1}+2^{l-1}$ bit

Exercise

Find Hamming code of 0011 configured by "even parity"

Solve: ∵ $n=4$ $k=3$ According to $2^k \geq n+k+1$

index	1	2	3	4	5	6	7
name	C_1	C_2	0	C_4	0	1	1
	1	0		0			

$C_1 = 3 \oplus 5 \oplus 7 = 1$

$C_2 = 3 \oplus 6 \oplus 7 = 0$

$C_4 = 5 \oplus 6 \oplus 7 = 0$

∴ **The Hamming code of 0011 is 1000011**

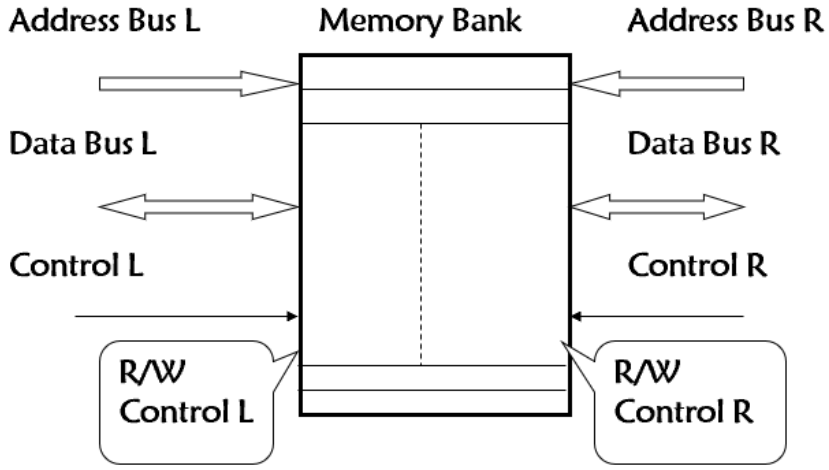
4.ROM Chips

(1) MROM
(2) PROM
(3) EPROM
2716 EPROM(2K*8)
(4) E^2PROM
1) Word Erase Mode 2) Data Block Erase Mode
Flash EPROM

5.High Speed Memory

Section3 High Speed Memory

1. Dual Ported Memory



2. Multiple Bank Parallel Interleaved Memory

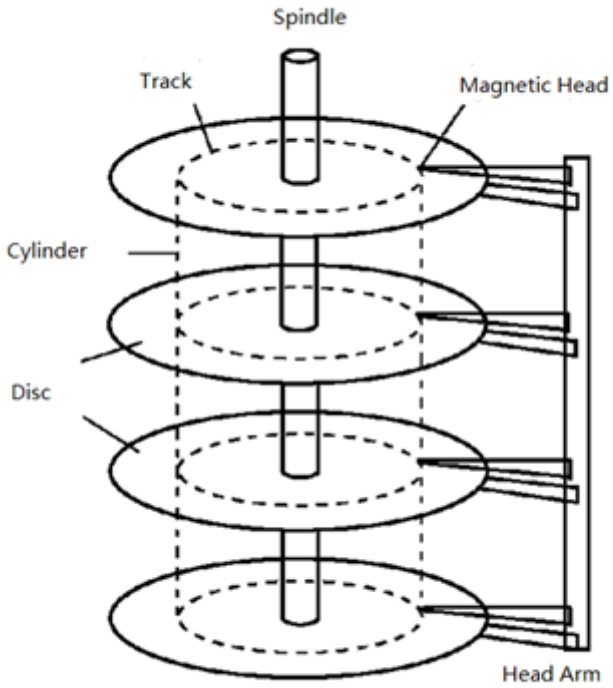
3. Associative Memory

6.Cache and Virtual Memory

	Cache	Virtual Memory
Function	Improve memory speed	Enlarge memory capacity
Implementation	Hardware	Software
Transparency	Transparent	Non-transparent
Address Translation	Simple	Complex , Low Speed

7.Auxiliary Memory

1. Magnetic Surface Memory
   1) Magnetic Tape Memory
   2) Magnetic Disk Memory
2. Optical Disc Memory
3. Removable Storage Devices



1.Magnetic Surface Memory

Magnetic Surface Memory

1. Characteristics Do not exchange information directly with the CPU

2. Magnetic surface memory specifications

(1) Recording density track density D_t bit density D_b

(2) Memory capacity $C = n \times k \times s$

(3) Average addressing time

$$\begin{cases} \text{Average seek time} \\ \text{Average wait time} \end{cases}$$

(4) Data transfer rate $D_r = D_b \times V$

(5) Error rate Error messages to the total number of bits read

Example

Assume that there are 6 pieces of disk in total, The two outer sides of the disk cannot be recorded, 204 tracks on each side, Each track has 12 sectors, and 512B per sector, the speed is 7,200 rpm, The average seek track time is 8ms.

（1）Calculate the capacity of the disk memory.

（2）Calculate the average addressing time of the disk memory.

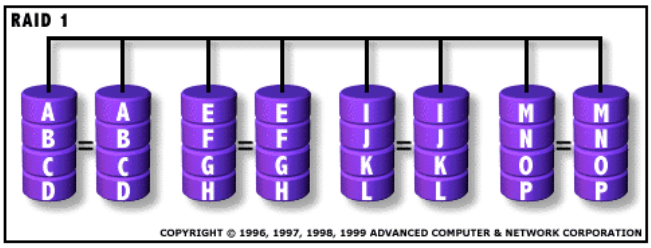
8.Disk Array(RAID)

Section6 Disk Array(RAID)

一、RAID0

二、RAID1

Disk Mirroring

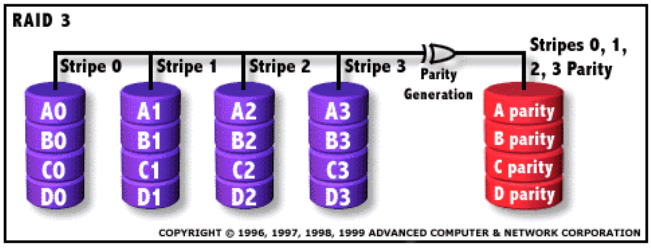


三、RAID2

Hamming Code

四、RAID3

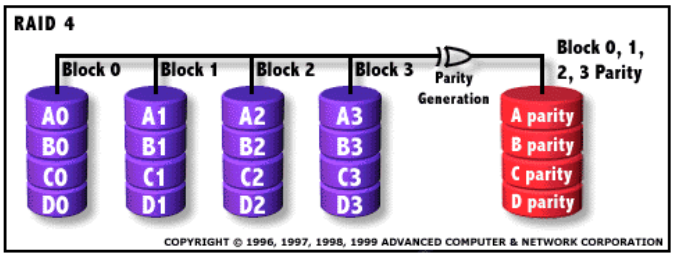
Parity Check



Section6 Disk Array(RAID)

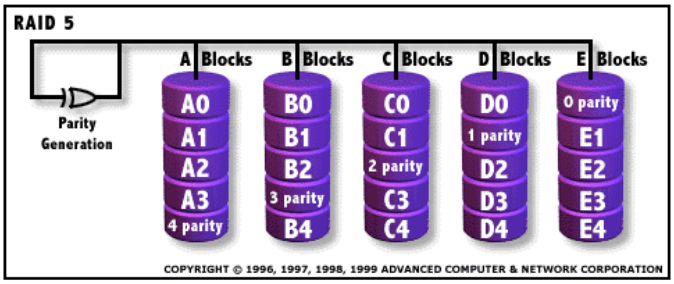
五、RAID4

Independent Access



六、RAID5

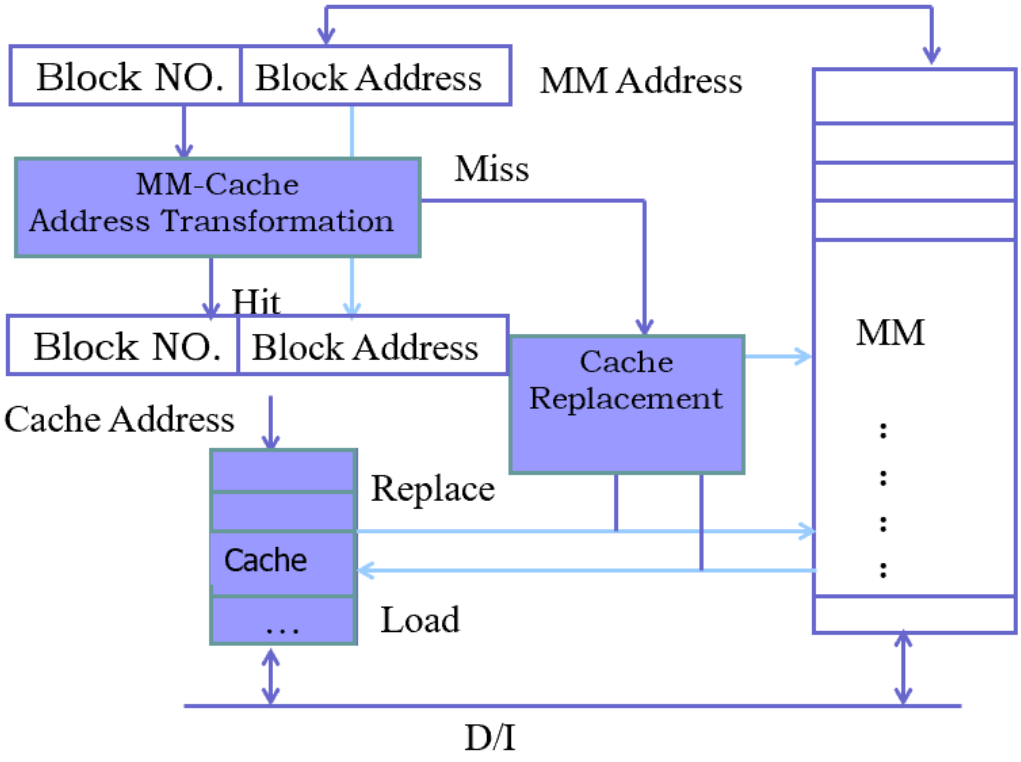
Check Block Circulation Distribution



6.CACHE

1.Structure

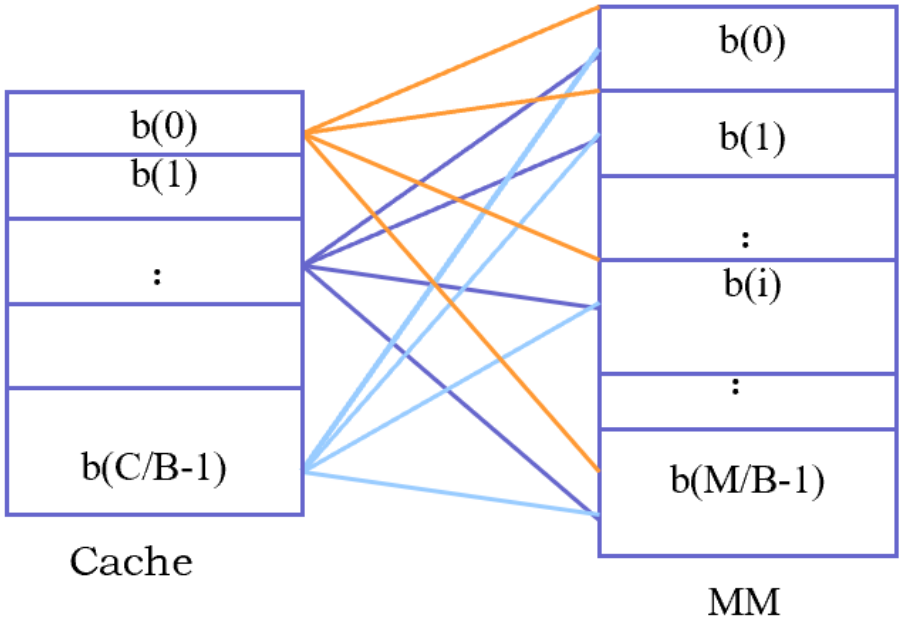
Cache Structure



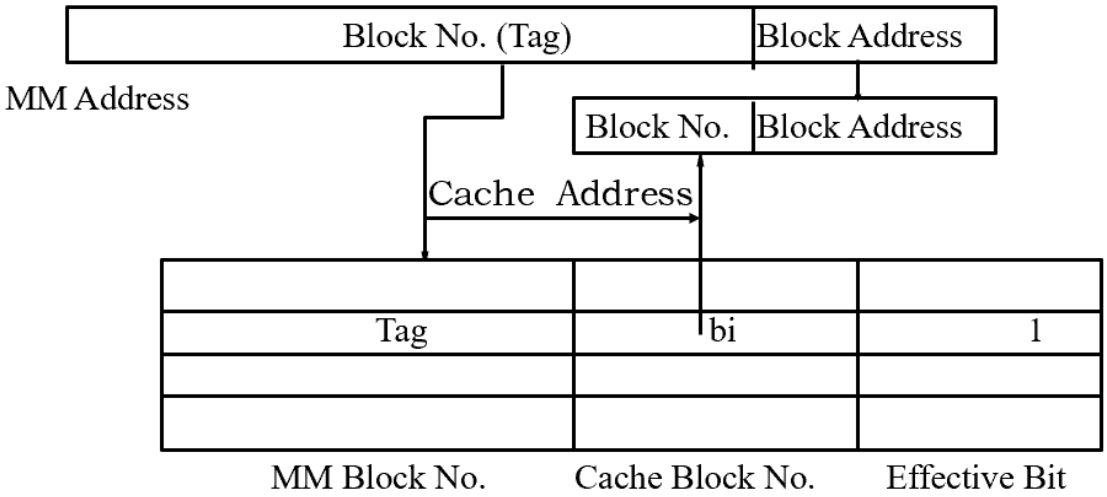
2.Address Mapping and Transformation

1.Fully Associative Mapping

•Fully Associative Mapping



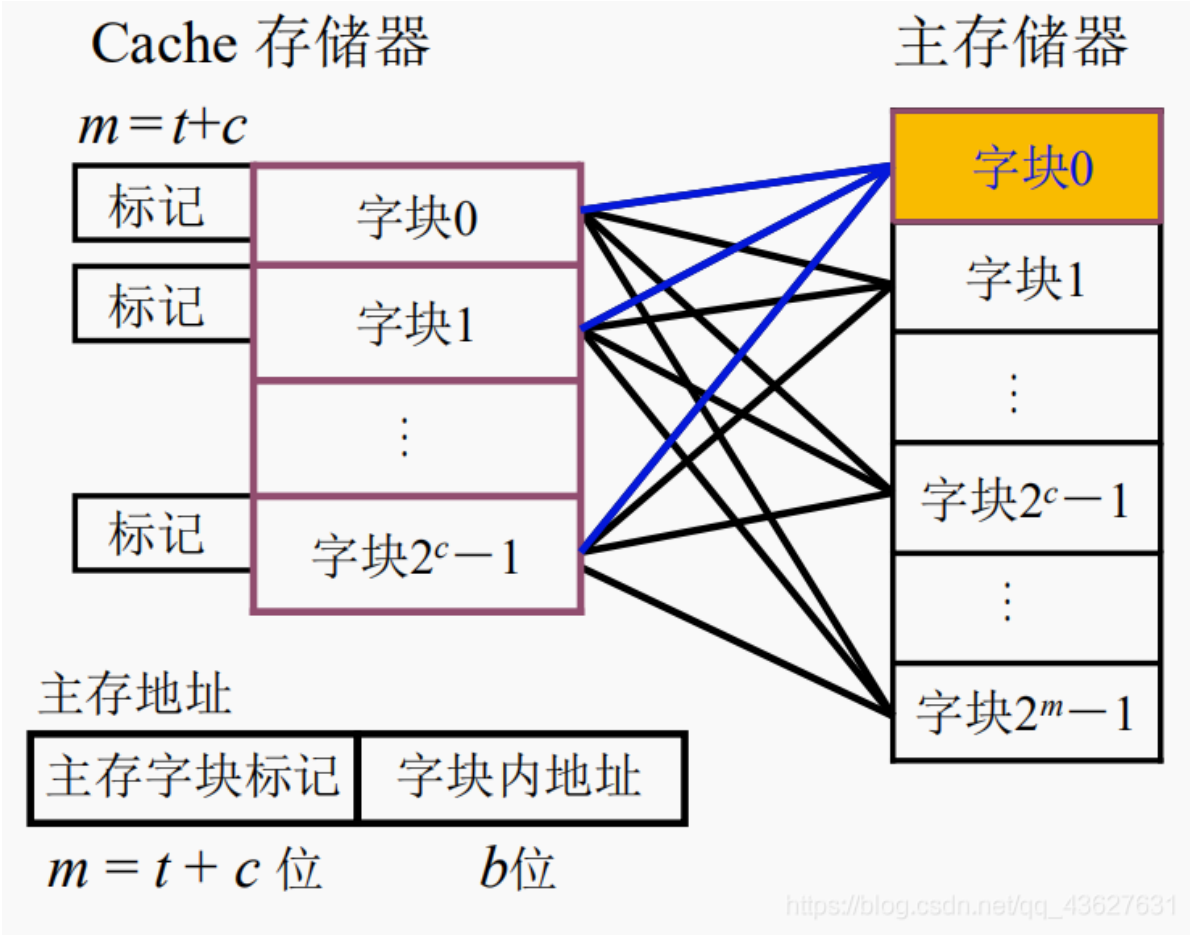
•Fully Associative Address Transformation



主存中的任一块映象到缓存中的任一块，将主存中一个块的地址（块号）与块的内容（字）一起存与cache的行中，其中块地址存与cache行的标记部分。cache的数据块大小称为行，主存的数据块大小称为块。cache与主存之间的数据交换以块为单位，CPU与cache之间的数据交换以字为单位。

优点：灵活，命中率高

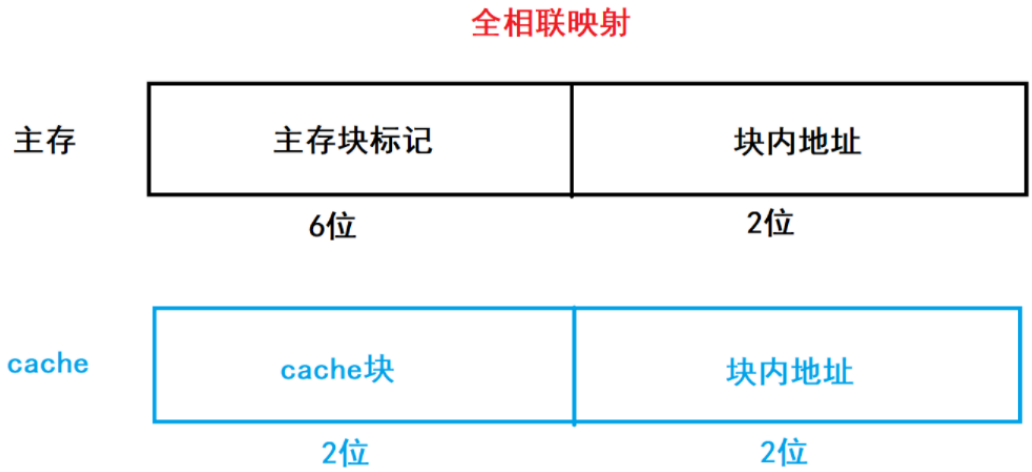
缺点：主存字块标记为全部块标记，访问cache时，主存的字块标记要和cache的全部标记位进行比较，所需的逻辑电路很多，成本较高，实际的cache还要采用各种措施来减少地址的比较次数。



主存中的任一块映象到缓存中的任一块

例:主存有64块，cache有4块,一个块为4个字节

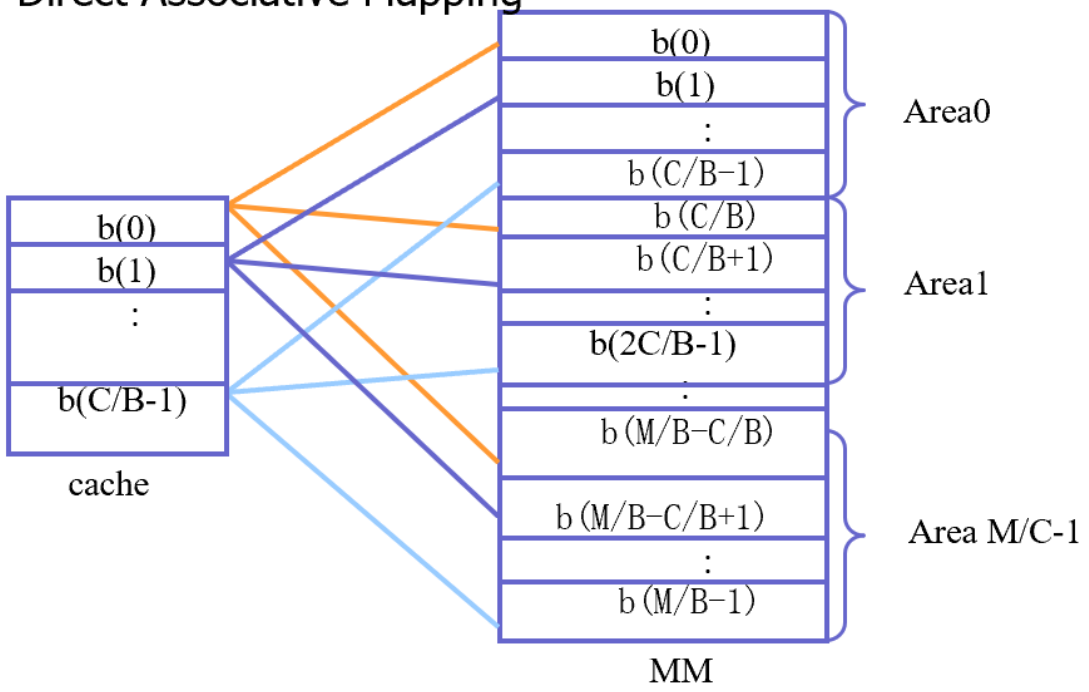
64->2^6,4->2^2



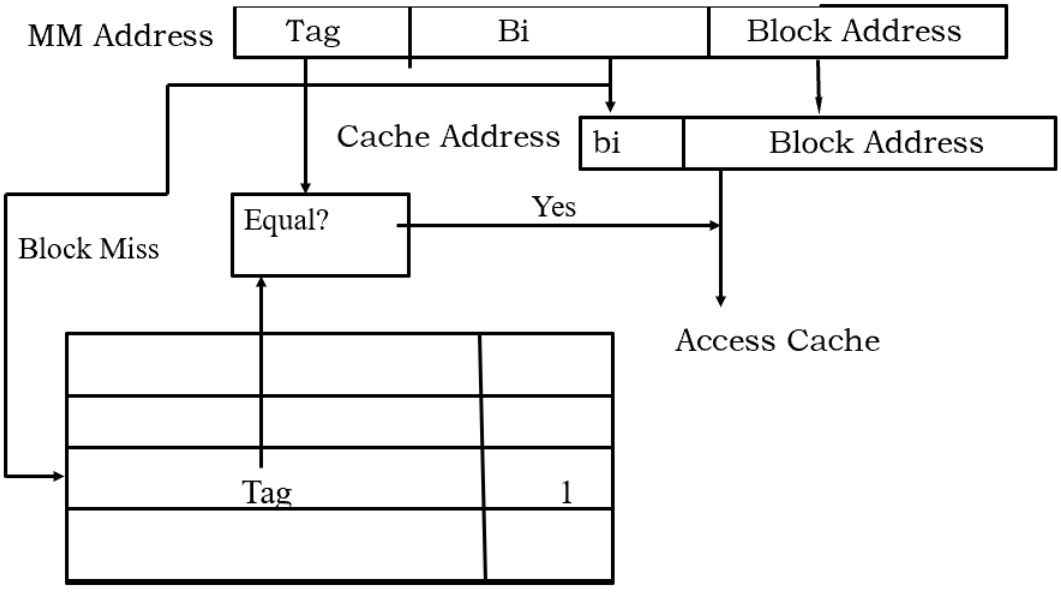
2.Direct Associative Mapping

b=A mod C/B
b： Cache Block No.
A： Main Memory No.
C/B：Cache Block Count

•Direct Associative Mapping



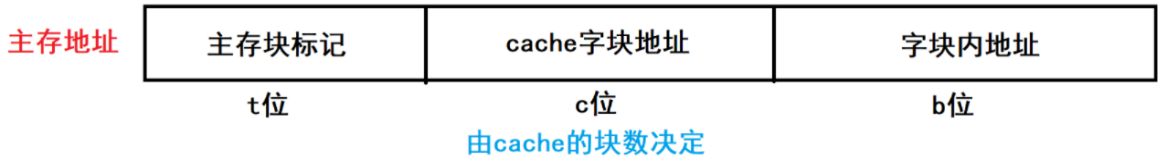
•Direct Associative Address Transformation

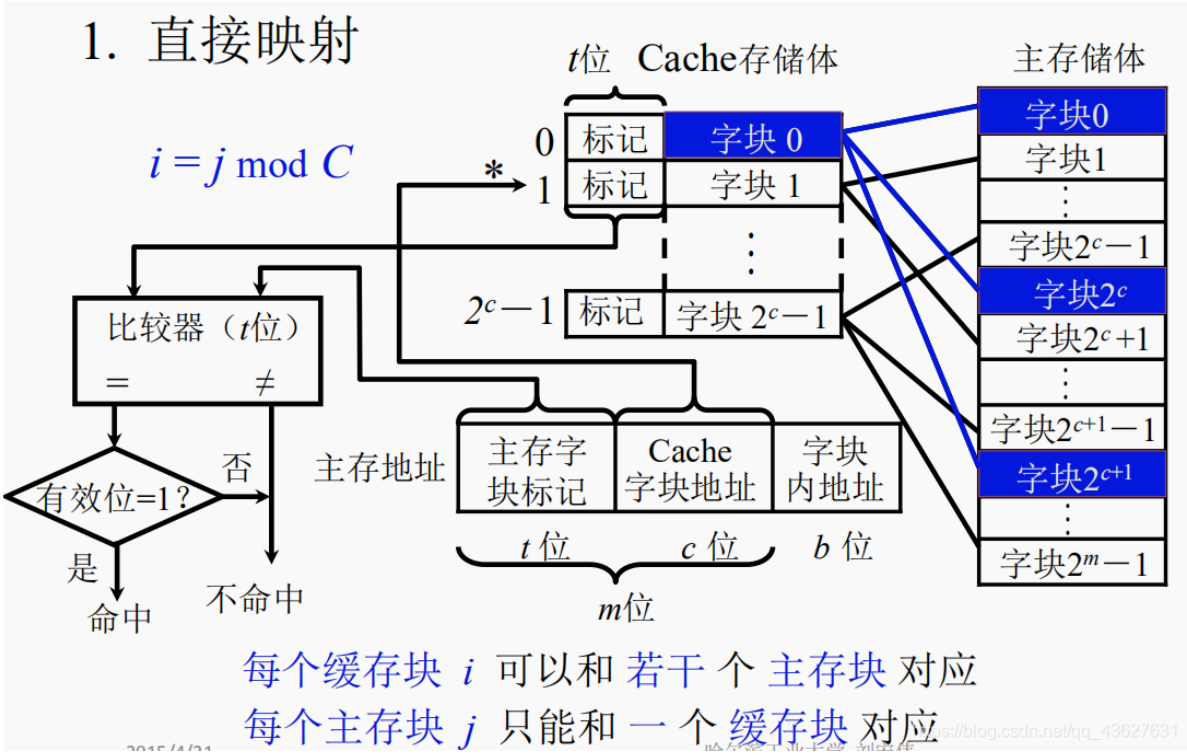


Directory Table

主存中的任意块映射到缓存中的唯一块，每个主存块只与一个缓存块相对应。

i为缓存块号，j为主存块号，C为缓存块数，映射关系式：i = j mod C
特点：不灵活，每个主存块只能与固定对应某个缓存块，即使还空着许多位置也不能使用。
主存地址的格式：



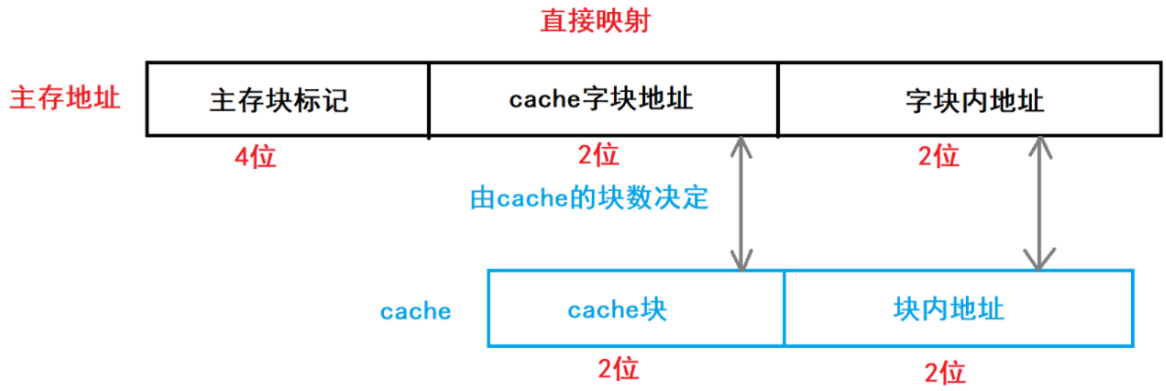


例题：某内存64块，cache有4块，一个块为4个字节，按字节编址，采用直接映射方式，写出cache和主存地址格式。

我们一个个的分析，先分析cache
cache：4块=16字节=2的4次方字节，cache的总位数是4。
主存：64块=64*4字节=2的8次方字节，主存的总位数是8

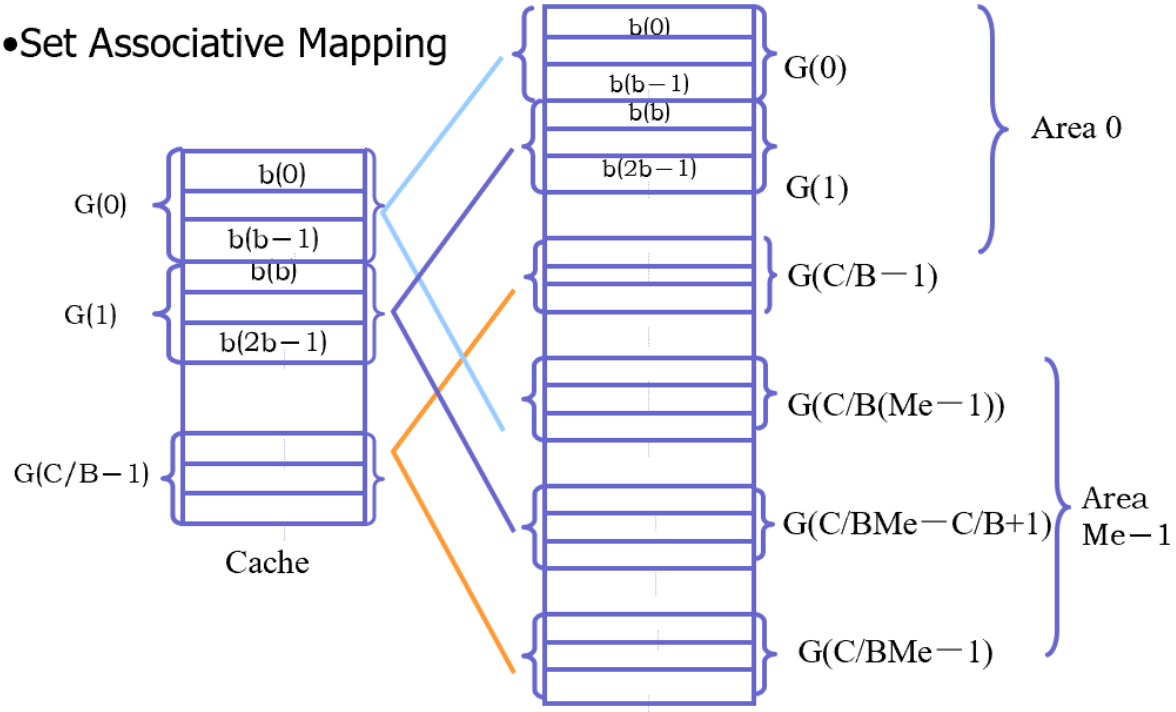
由题意得：一个块4个字节，就是2的2次方，可以得出cache的块内地址为2位，又因为cache的总位数是4，所以得出cache块有（4-2）位

主存块标记的位数就相当于主存的总位数减去cache的总位数，就是4位，最后就求出cache和主存的地址格式

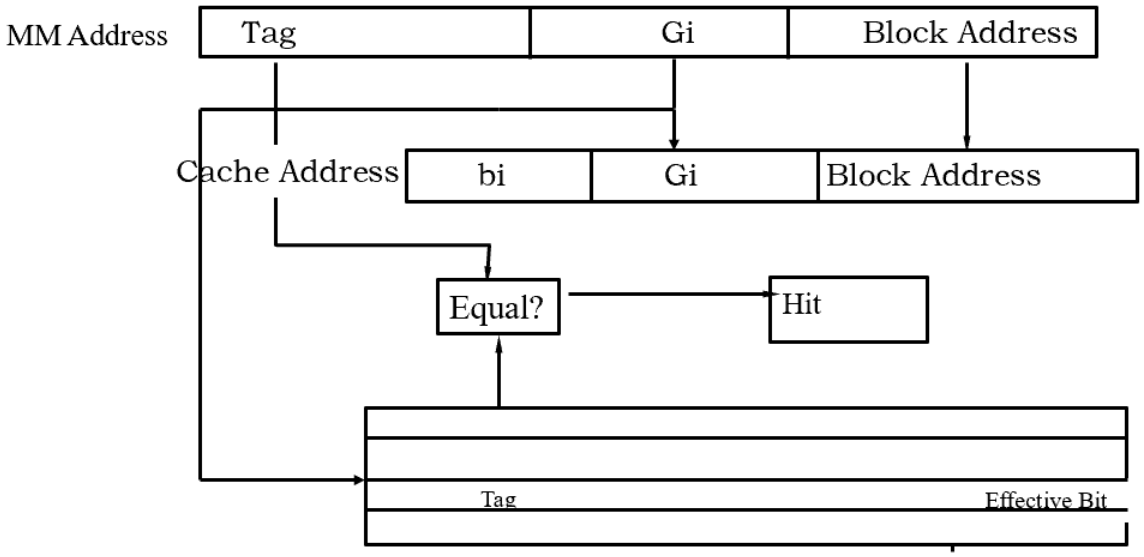


3.Set Associative Mapping

•Set Associative Mapping

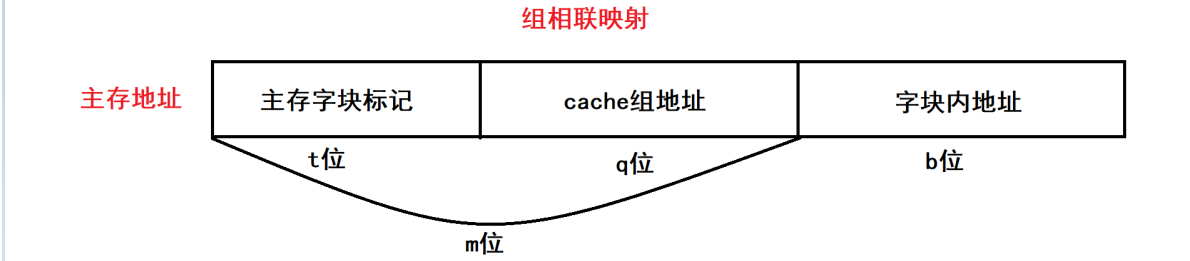


•Set Associative Address Transformation



组间直接映射，组内全相联映射
主存块j按模Q（组数）映射到缓存的第i组中任一块，cache分为Q组，每组R块
对应关系有：i = j mod Q
i为缓存的组号，j为主存的块号

组内n块，组相联映射定义为n路组相联



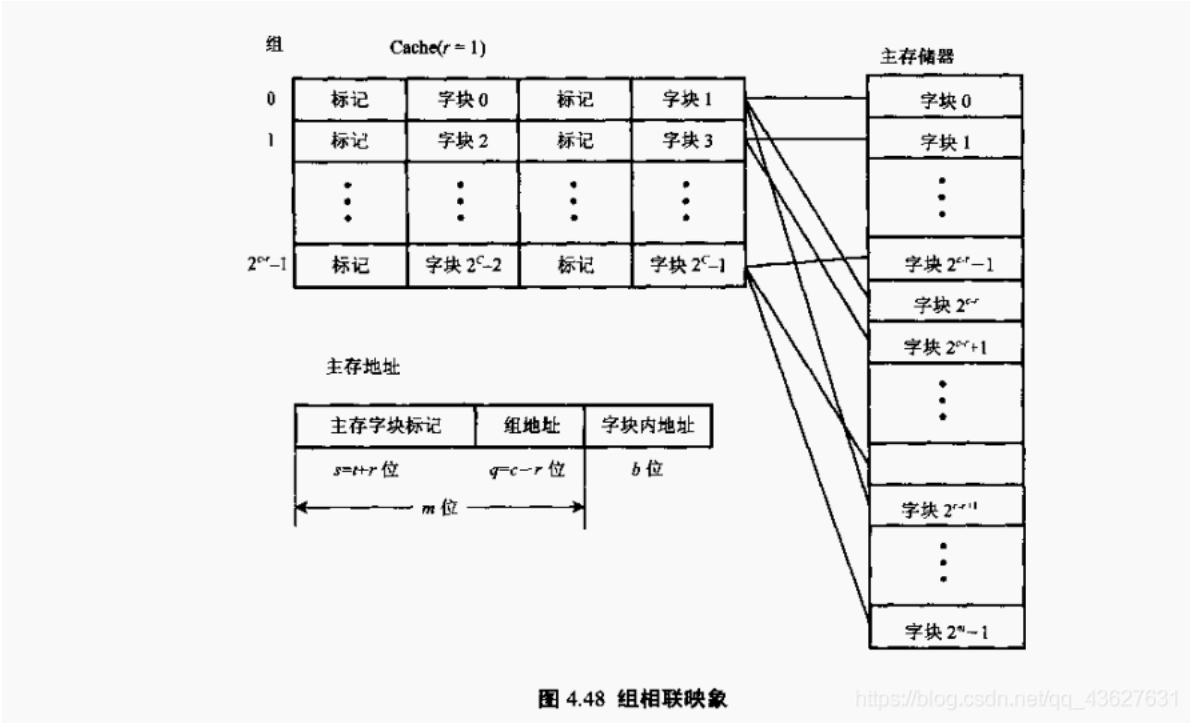
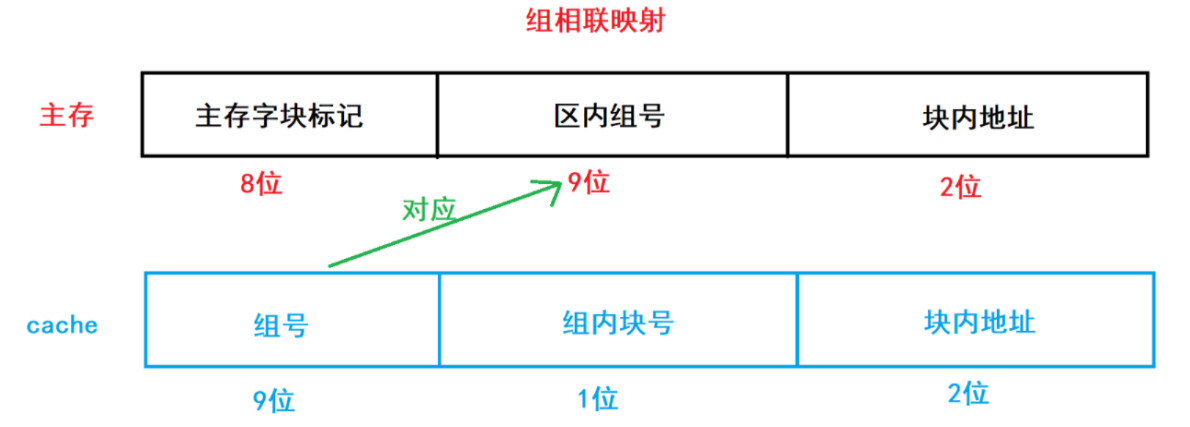


图 4.48 组相联映象

例题：主存容量512K×16位，cache容量4096×16位，块长为4个16位的字，按字地址访存写出二路组相联映射下主存的地址格式。

cache地址的位数：4096=2的12次方，12位
主存地址的位数：512K=2的19次方，19位
由块长可得块内地址位：2位
二路组相联，cache的组数：2的10次方/2=2的9次方。所以主存的区内组号是9位
主存字块标记的位数：19-2-9=8



3.Replacement Policy

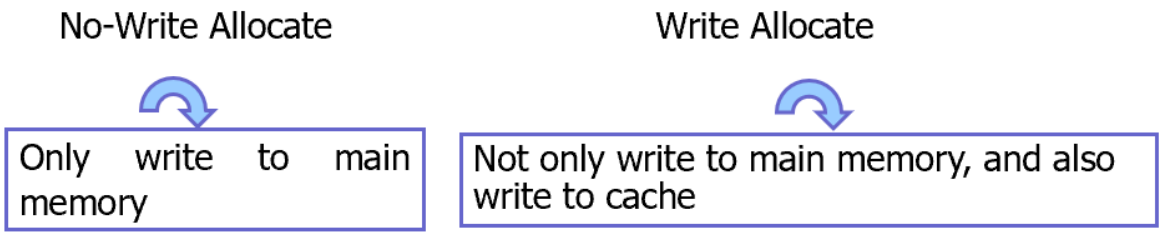
RAND (Random)
FIFO (First-In First-Out)
LRU (Least Recently Used)
LFU (Least Frequently Used)

4.Write Operation of Cache

(1)WT——Write through：Not only write to cache, and also write to main memory.
Pros: Simple and Good Consistency
Cons: Low Speed

(2)WB——Write back：Only write to cache.
Pros：High Speed
Cons: Poor Reliability

(3)Write Miss：



5.indecator

- Speedup Ratio of Cache

Average Memory Access Time: T

$$T = H_c T_c + (1 - H_c) T_m$$

Tc： Cache Access Time

Tm： Main Memory Access Time

Hc： Cache Hit Rate

$$S_p = \frac{T_m}{T} = \frac{T_m}{H_c T_c + (1-H_c)T_m} = \frac{1}{H_c \frac{T_c}{T_m} + (1-H_c)}$$

When Hc→1，

$$S_p \to \frac{T_m}{T_c}$$

Memory Access Efficiency

$$e = \frac{T_c}{T} = \frac{1}{H_c + (1-H_c)\frac{T_m}{T_c}}$$

7.Point

1.The most widely used character encoding scheme is the ASCII code, in which one character occupies one byte cell of main memory, if it is a sequence of characters, usually occupy several continuous byte cells of main memory.

2.Dynamic random access memory (DRAM) must be refreshed regularly due to the leakage of capacitor charge. There are mainly three kinds of refresh modes, which are centralized refresh mode, scattered refresh mode and asynchronous refresh mode.

3.Interrupt processing includes: interrupt request, Interrupt priority arrangement, interrupt response, interrupt processing and interrupt return.

4.The DMA transfer process includes: DMA initialization, DMA request, DMA response, DMA transfer, and DMA end.

5.Methods for Increasing memory bandwidth include parallel memory, buffer memory, cache.

6.The parallel adder is divided into serial carry and parallel carry according to the form of carry chain.

7.Pipelining hazard includes structural hazard, data hazard and control hazard. Data hazard includes RAW, WAR and WAW.

8.EPROM refers to erasable programable read only memory.

9.There are four ways of bus communication, namely synchronous, asynchronous, semisynchronous and separate.