第三章：内部存储器

▼ 存储器概述

存储位元（最小单位）→ 存储单元 →存储器

分类

按存储介质：磁表面/半导体

按存取方式：随机/顺序存取（磁带）

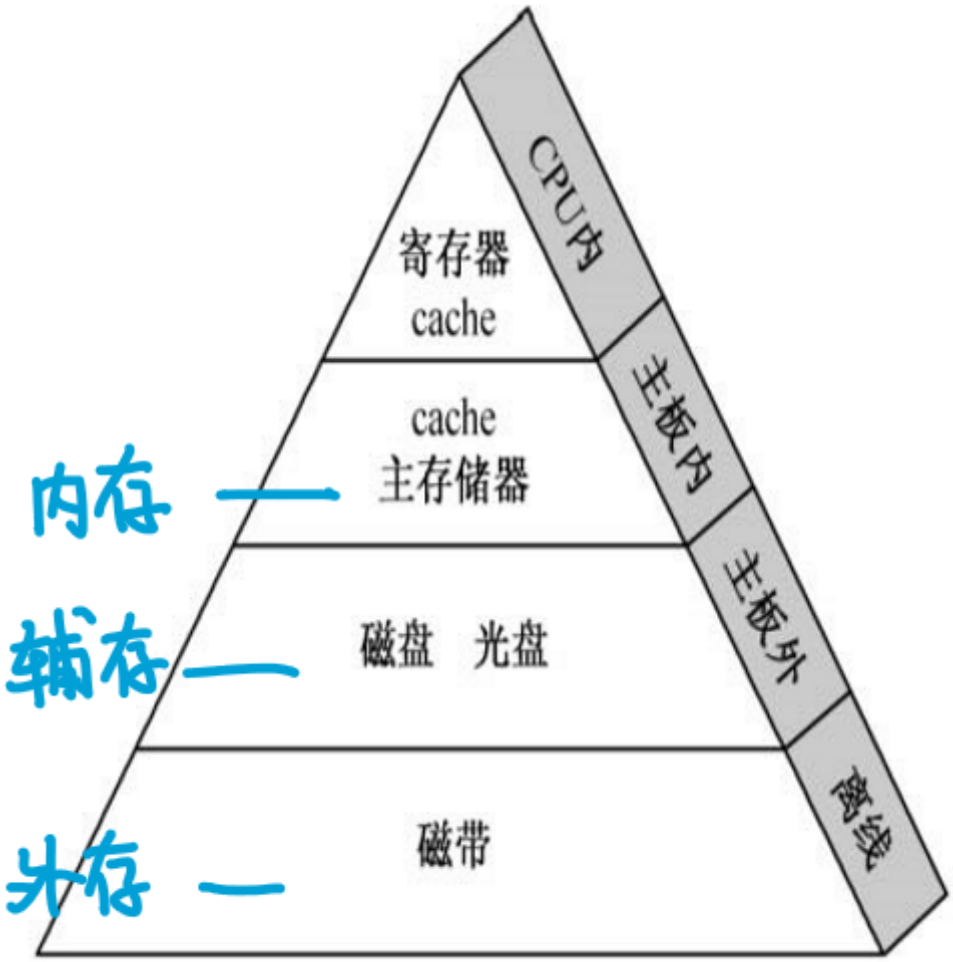
按读写功能：ROM（只读、非易失）/RAM（随机，易失）

按信息可保存性：永久性/非永久性（易失性/非易失性）

按作用：主存/辅存/缓存

- 速度快的容量小，价格高

分级结构



主存储器技术指标

- **存储容量=存储字数×字长（MDR位数）**
- **存取时间**：又称存储器访问时间，指一次读操作命令发出到该操作完成，将数据读出到数据总线上所经历的时间。
- **存储周期**：指连续启动两次读操作所需间隔的最小时间。
- **存储器带宽**：单位时间里存储器所存取的信息量。
  - 带宽=存储字长/存储周期

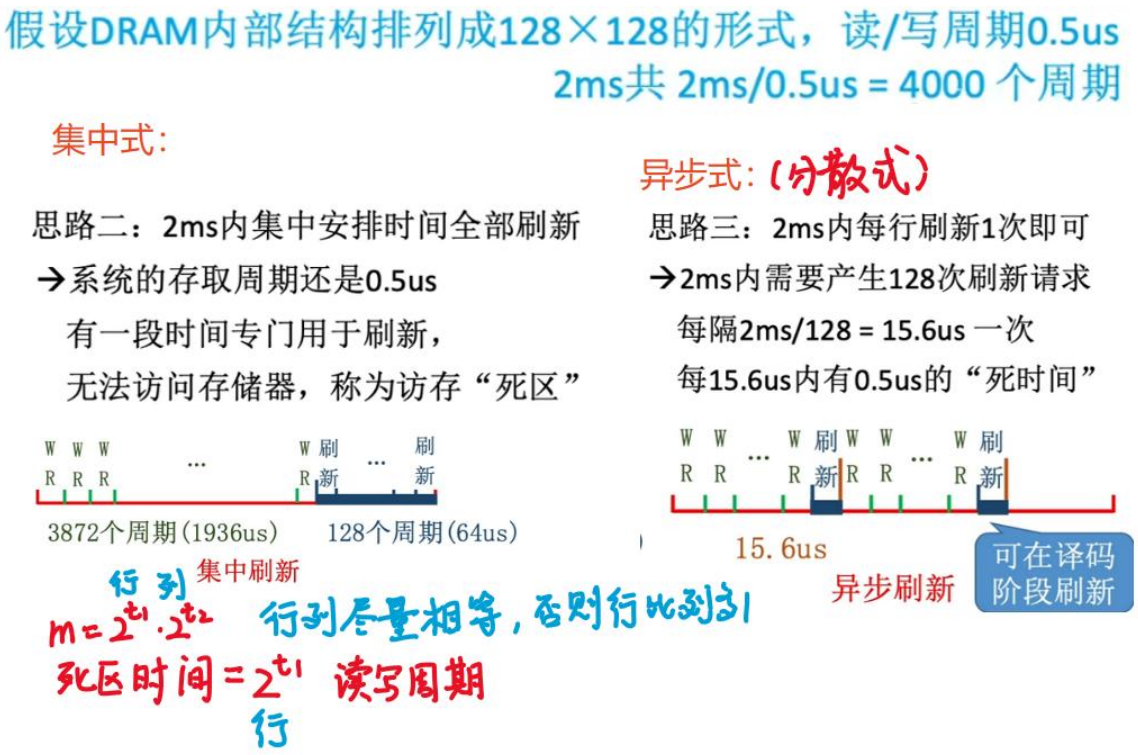
▼ RAM

根据信息存储的机理分为**静态读写存储器（SRAM）**和**动态读写存储器（DRAM）**

- SRAM常用作Cache，DRAM常用作主存

类 型 特 点	SRAM（静态RAM）	DRAM（动态RAM）
存储信息 存储位元	触发器	电容
破坏性读出	非	是
读出后需要重写？（再生）	不用	需要
运行速度	快	慢
集成度	低	高
发热量	大	小
存储成本	高	低
易失/非易失性存储器？	易失（断电后信息消失）	易失（断电后信息消失）
需要“刷新”？	不需要	需要（分散、集中、异步）
送行列地址	同时送	分两次送（地址线复用技术）

- **刷新周期**：上一次刷新结束到下一次全部刷新一遍为止，一般为2ms，占1个读写周期。
- **刷新信号周期**=刷新周期/行数，向下取读写周期的整数倍



▼ ROM

MROM：内容固定，不可重写

可编程ROM：

一次性编程：PROM

多次编程：EPROM（光擦除）、E2PROM（电擦除）

▼ 存储器扩容

字长扩展（数据线扩展）

[例2] 利用1M×4位的SRAM芯片，设计一个存储容量为1M×8位的SRAM存储器。

- 位数不足，8位需求 v.s. 4位芯片
- 所需芯片数量=(1M×8)/(1M×4)=2片

存储容量扩展（地址线扩展）

[例3]利用1M×8位的DRAM芯片设计2M×8位的DRAM存储器

- 容量不足：2M需求 v.s. 1M芯片
- 所需芯片数d=（2M×8）/（1M×8）=2(片)

混合扩展

- 系统程序区 → ROM
- 用户程序区、程序工作区 → RAM

- 由128K*8位的DRAM芯片构成1024K*32位存储器

1）总共需要多少DRAM芯片

2）画出存储器逻辑框图

3）存储器为读写周期0.5us，CPU在1us内至少访存一次，采用何种刷新方式?

4）刷新周期为8ms，刷新信号周期为?

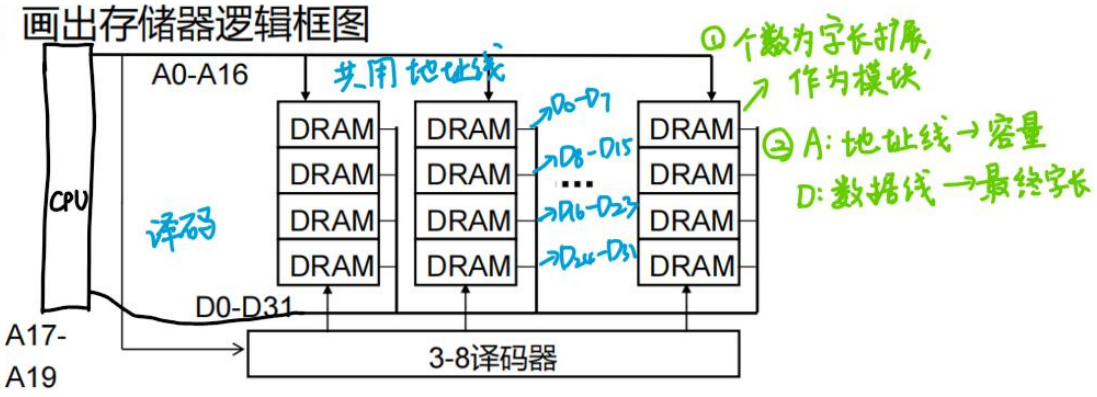
1）总共需要多少DRAM芯片

字长扩展 32/8=4 （4个DRAM组成32位）

容量扩展 1M/128K=8 （17位 20位，3-8译码器）

需要4*8=32片

2）画出存储器逻辑框图



3）存储器为读写周期0.5us，CPU在1us内至少访存一次，采用何种刷新方式？

128k=2^{17}→ 行列均分 or 行比列多1

行 列

假设存储器芯片为512*256*8bit（17位=9位+8位）

集中式刷新：0.5us*512=256us >> 1us，不可行

死区时间：刷新行的时间

采用分散式刷新

4）刷新周期为8ms，刷新信号周期为？

刷新最大时间间隔

刷新信号周期=刷新周期/行数

=8ms/512

=15.625us

=15.5 us （以读写周期向下取整）

向下取0.5us的整数倍

给定地址空间

容量 $2^{16}=64K$

习题）某16位计算机，地址总线16根（A15-A0，A0为低位），双向地址总线16根（D15-D0）。控制总线与主存相关包括MERQ（访存允许），R/W（读写控制）

字长 16bit

主存空间分配如下（按字编址）：

- 0-8191为系统程序区，由ROM芯片组成
- 8192-32767为用户程序区
- 最后的2K地址空间为程序工作区

现有如下芯片：

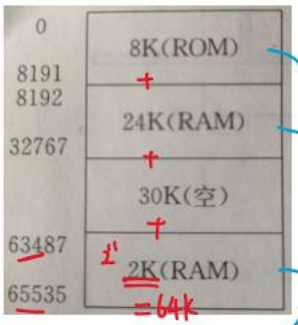
- ROM：8K*16bit、4K*16bit、8K*8bit（控制端CS）
- SRAM：16K*1bit、2K*16bit、4K*16bit、8K*16bit

1）画出地址分配示意图

2）从上述芯片中选择芯片设计该计算机系统存储器，说明选择哪些存储器，用多少片

3）画出主存逻辑图（可选译码器与门电路）

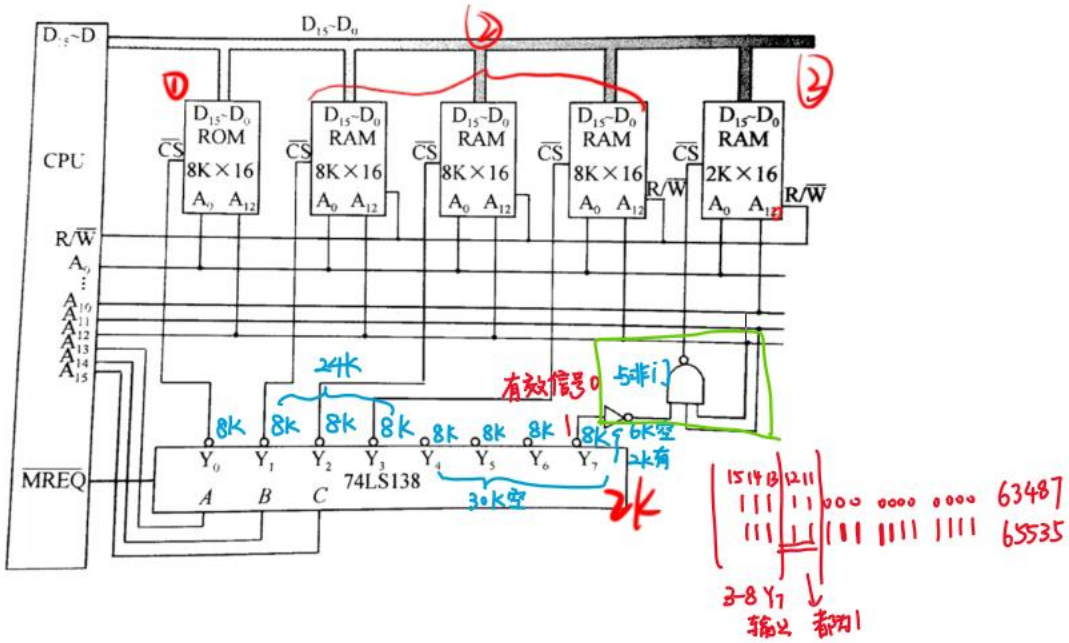
- 主存空间分配如下（按字编址）：
  - 0-8191为系统程序区，由ROM芯片组成
  - 8192-32767为用户程序区
  - 最后的2K地址空间为程序工作区



1）地址分配如图

2）芯片选择情况

- ROM选择8K*16bit，1片，片内地址13位，片选3位（0）
- SRAM选择8K*16bit，3片，片内地址13位，片选3位（1-3）
  2K*16bit，1片，片内地址11位，片选5位（7）

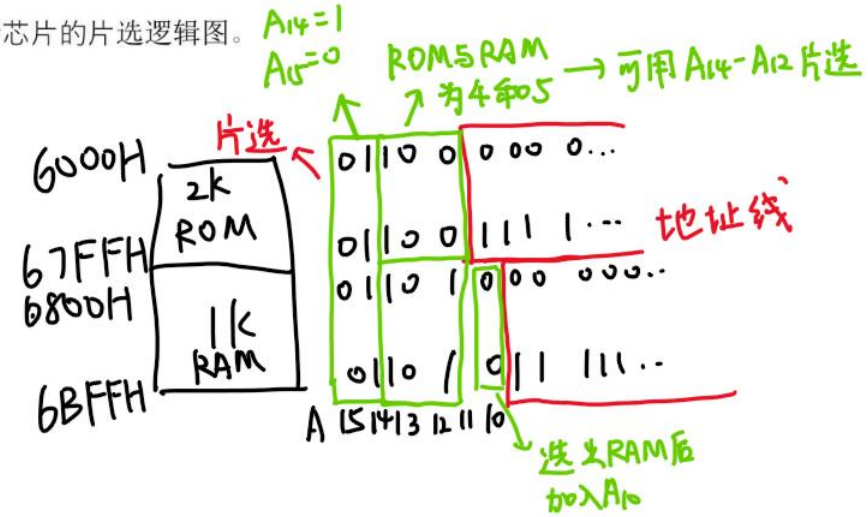


1. 设 CPU 有 16 根地址线， 8 根数据线，并用 $\overline{MREQ}$ 作为访存控制信号（低电平有效），用 $\overline{WR}$ 作为读 / 写控制信号（高电平为读，低电平为写）。现有下列存储芯片：$1K\times4$ 位 RAM，$4K\times8$ 位 RAM，$8K\times8$ 位 RAM，$2K\times8$ 位 ROM，$4K\times8$ 位 ROM ，$8K\times8$ 位 ROM 及 74LS138 译码器和各种门电路。画出 CPU 与存储器的连接图，要求：

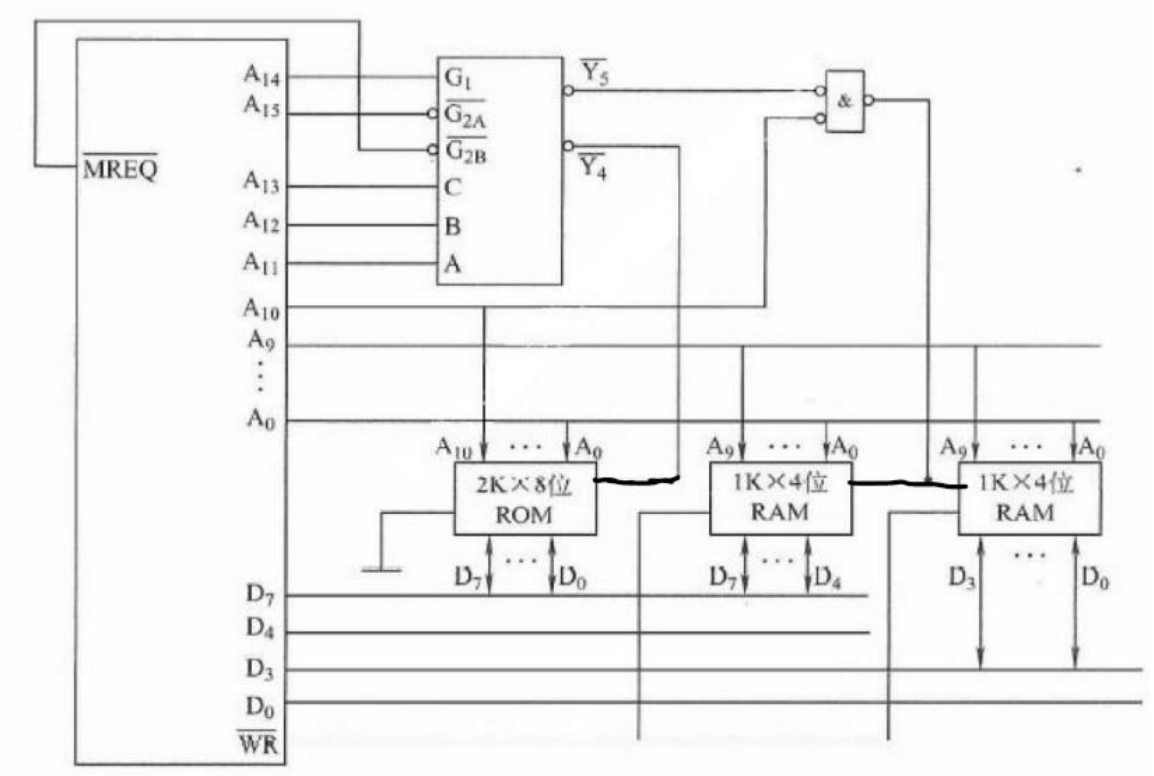
1) 主存地址空间分配： 6000H - 67FFH 为系统程序区； 6800H - 6BFFH 为用户程序区。

2) 合理选用上述存储芯片，说明各选几片。

3) 详细画出存储芯片的片选逻辑图。



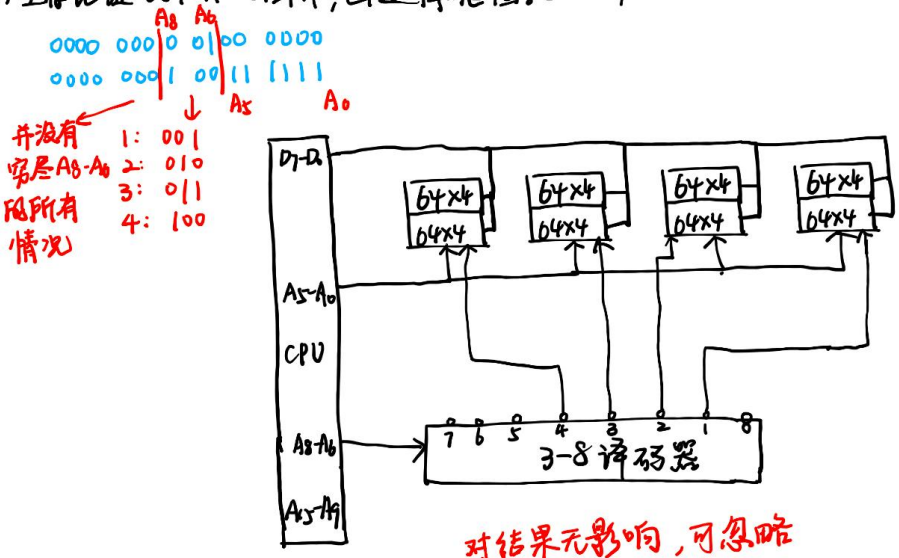
3）存储芯片的片选逻辑图如下图所示。



期中题

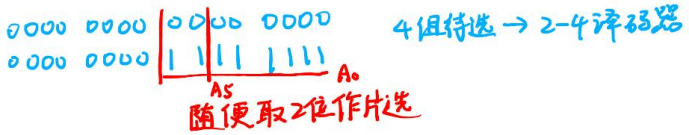
机字长8位，8条数据线（D7-D0），16条地址线（A15-A0）
用64×4位DRAM（8行8列）构成总容量256×8位主存。

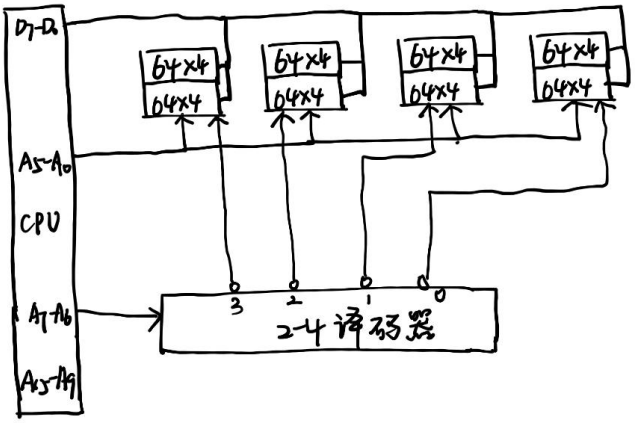
(1) 主存地址 0040H ~ 013FH，画逻辑框图。（3-8译码器）



对结果无影响，可忽略

(2) 主存地址 0000H ~ 00FFH，存储器采用多体交叉方式，画出逻辑框图（用2-4译码器）

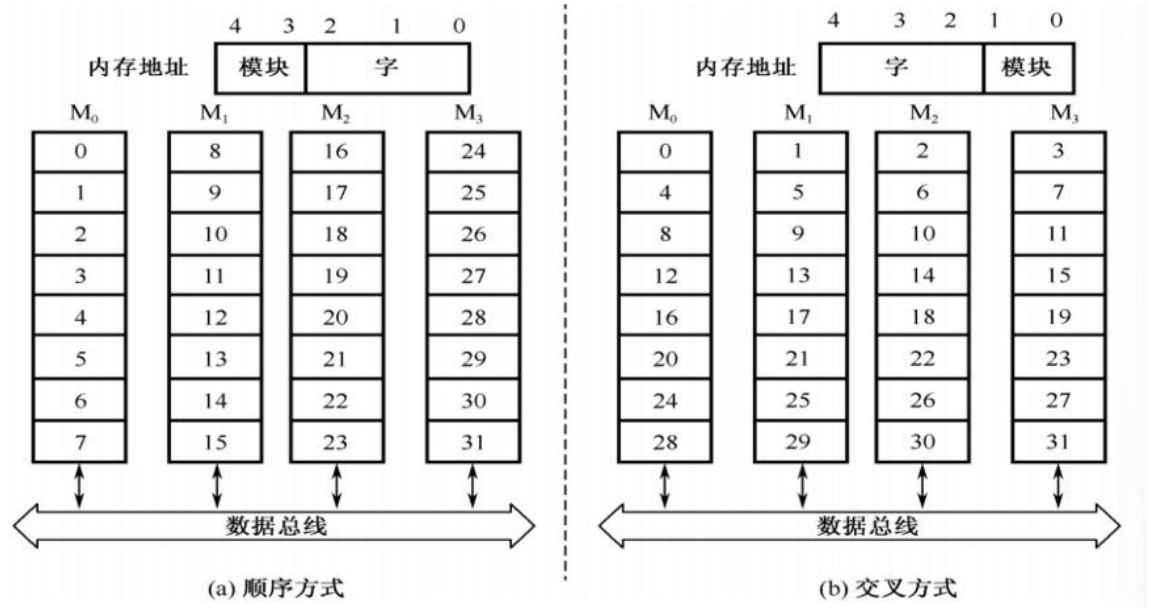




▼ 并行存储器

双端口存储器（空间并行）

多模块交叉存储器（时间并行）



顺序方式

优点：某模块故障时其他可以正常工作

缺点：串行工作，带宽受限

交叉方式

优点：对连续字的传送可以实现并行，提高带宽

对于给定地址x，判断属于第几个存储体：

1. 给出二进制地址：直接看体号
2. 对给出的十进制地址对m取余

存储周期与总线周期

存储周期：T

总线传送周期（存取时间）： τ

应保证 $n \geq T/\tau$

一般设计为 $\tau = T/n$，n为模块数

$t_{顺序} = xT$

$t_{交叉} = T + (n-1)\tau$

x 为模块数

信息量=字长×模块数

带宽=信息量/时间

例题：设存储器容量为32字，字长64位，模块数m=4，分别用顺序方式和交叉方式进行组织。存储周期T=200ns，数据总线宽度为64位，总线传送周期=50ns。→τ
若连续读出4个字，问顺序存储和交叉存储的带宽各是多少?

解：
顺序存储器和交叉存储器连续读出m=4个字的信息总量都是：

$$q=64\text{（字长）}\times 4=256\ \text{bit}$$

顺序存储器和交叉存储器连续读出4个字所需的时间分别是：

$$t_{顺序}=mT=4\times 200ns=800ns=8\times 10^{-7}s$$

$$t_{交叉}=T+(m-1)\tau=200+3*50ns=350ns=3.5\times 10^{-7}s$$

顺序存储器和交叉存储器的带宽分别是：

带宽=信息量/时间

$$W_{顺序}=q/t_{顺序}=256b\div(8\times 10^{-7})s=320Mb/s$$

$$W_{交叉}=q/t_{交叉}=256b\div(3.5\times 10^{-7})s=730Mb/s$$

▼ Cache访存

N_c表示cache完成存取的总次数，N_m表示主存完成存取的总次数，
定义h为命中率 $h=\frac{N_c}{N_c+N_m}$
t_c表示命中时的Cache访问时间，t_m表示未命中时的主存访问时间，
则Cache/主存系统的平均访问时间 t_a 为：$t_a=h*t_c+(1-h)t_m$

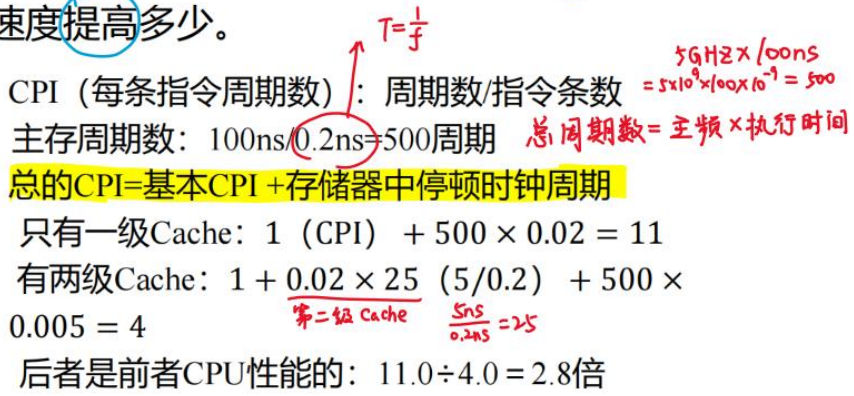
访问效率 $e=\frac{t_c}{t_a}$ **（Cache时间/平均时间）**

Cache与内存速度比（主存时间/Cache时间） $r=\frac{t_m}{t_c}$

性能提高多少？性能为原来的 $\frac{t_m}{t_a}$ 倍 **（主存时间/平均时间）**

多级Cache计算

[例10] 现有一处理器，假设其基本CPI为1.0，所有访问在第一级cache中命中，时钟频率5GHz。假定访问一次主存储器的时间为100ns，其中包括所有缺失处理。设平均每条指令在第一级cache中产生的缺失率为2%。若增加一个二级cache，命中或缺失的访问时间都为5ns，且容量大到可使必须访问主存的缺失率降为0.5%，问处理器速度提高多少。



CPI（每条指令周期数）：周期数/指令条数

主存周期数：100ns/0.2ns=500周期

总的CPI=基本CPI +存储器中停顿时钟周期

只有一级Cache：1（CPI）+ 500 × 0.02 = 11

有两级Cache：1 + 0.02 × 25（5/0.2）+ 500 × 0.005 = 4

第二级 Cache　5ns/0.2ns=25

后者是前者CPU性能的：11.0÷4.0 = 2.8倍

用平均CPI表示性能。

▼ Cache地址映射

全相联映射

优点：Cache利用率高，命中率高

缺点：比较器难实现

适用于小容量Cache

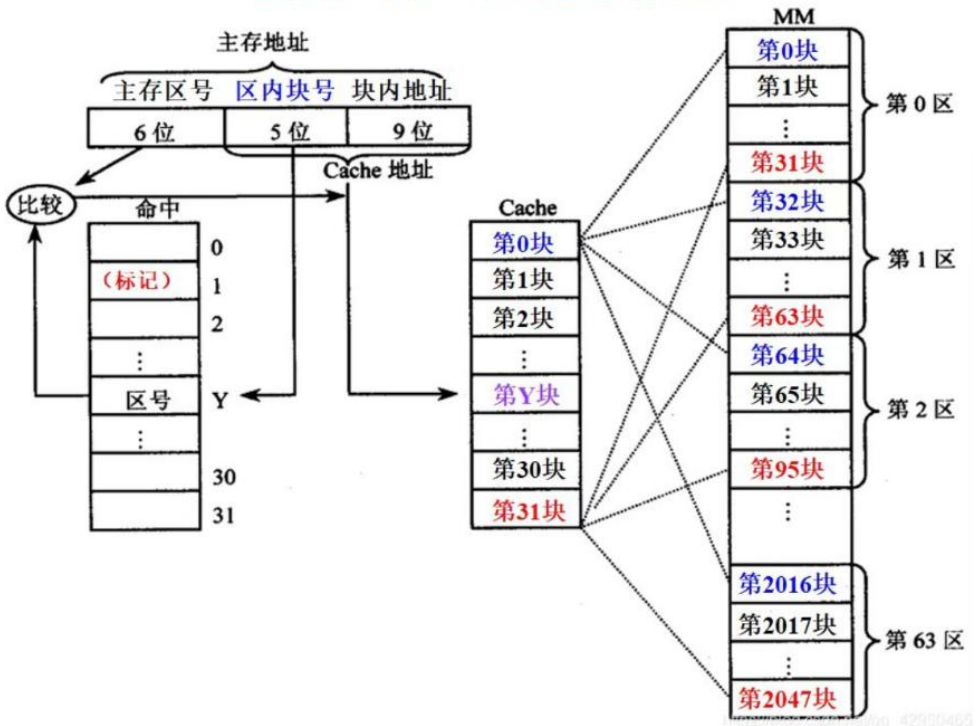
直接映射

优点：比较电路少m倍线路

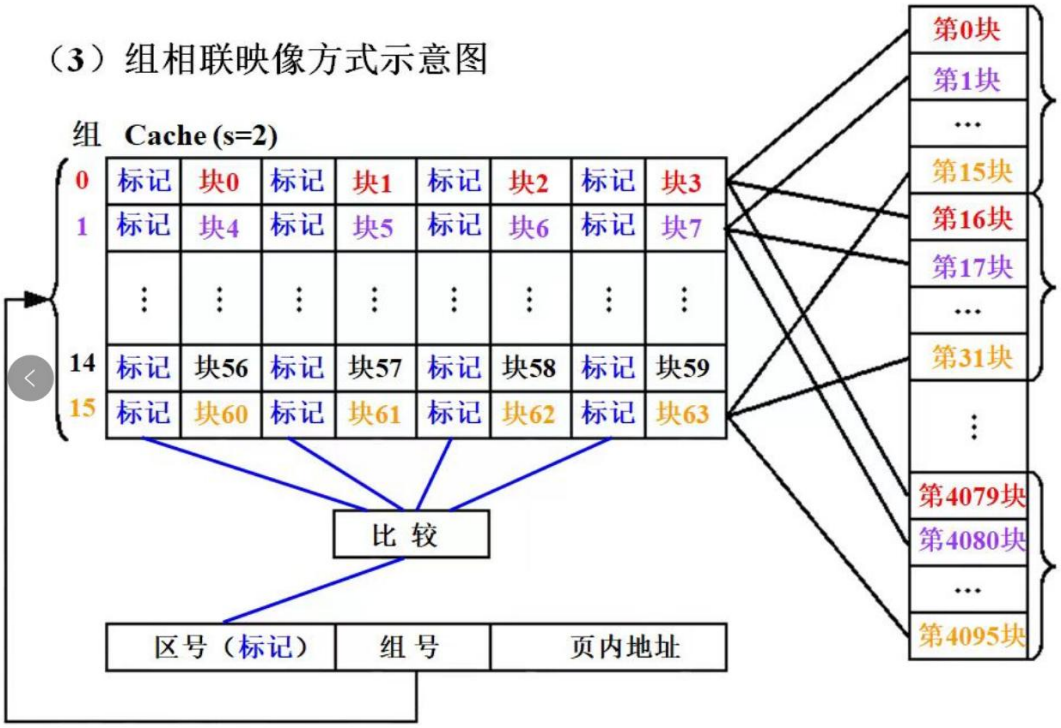
缺点：冲突概率高

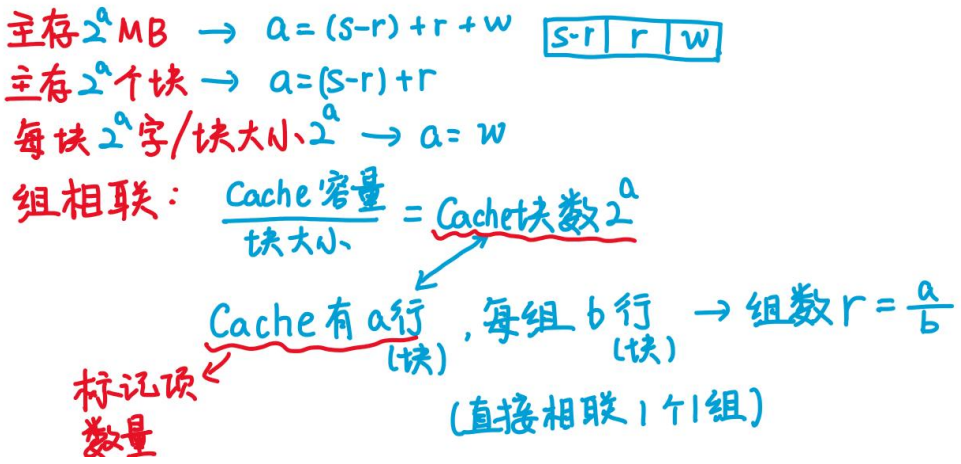
适合大容量Cache

直接方式地址映像及变换示意图



组相联映射方式





直接映射方式的内存地址格式如下所示：

标记s-r	行r	字w
8位$(BB)_{16}$	14位	2位

若主存地址用十六进制表示为BBBBBB，请用十六进制格式表示直接映射方法Cache的标记、行、字地址的值。

解：$(BBBBBB)_{16}=(1011\ 1011\ 1011\ 1011\ 1011\ 1011)_2$

标记s-r=$(1011\ 1011)_2=(BB)_{16}$

行r=$(1011\ 1011\ 1011)_2=(2EEE)_{16}$

字地址w=$(11)_2=(3)_{16}$

一个组相联cache由64个行组成，每组4行。主存储器包含4K个块，每个块128字。请表示内存地址的格式。

解：块大小=行大小=2^w个字，$2^w=128=2^7$，所以w=7

每组的行数k=4

cache的行数 = kv = $2^d \times k = 2^d \times 4 = 64$，所以d=4

组数v=$2^d=2^4=16$

主存的块数=$2^s=4K=2^2\times2^{10}=2^{12}$，所以s=12

标记大小=s-d=12-4=8（位）

主存地址长度=s+w=12+7=19（位）

主存寻址单元数=$2^{s+w}=2^{19}$

∴v=4路组相联的内存地址格式如下所示

8位	4位	7位
标记s-d	cache组号d	块内地址w

映射方式

- 主存容量1MB，字长8位，块大小16B，Cache容量64KB
- 1）采用直接映射，给出[F0010H]对应标记为 [填空1] 、行号为[填空2]、字号为 [填空3]。
  - F0010H= 1111 | 0000 0000 0001 | 0000
    （标记 | 行号 | 字号）
- 2）采用二路组相联映射，给出[F0010H]对应标记为 [填空1] 、行号为[填空2]、字号为 [填空3] 。
  - F0010H= 1111 0 | 000 0000 0001 | 0000
    （标记 | 组号 | 字号）
- 3）采用全相联映射，给出[F0010H]对应标记为 [填空1] 、字号为 [填空2] 。
  - F0010H= 1111 0000 0000 0001 | 0000
    （标记 | 字号）

设主存容量1MB，有16KB直接相联映像的Cache，假定该Cache的块为8个32位的字。解答下列问题：（1）写出Cache的地址格式。（2）写出主存的地址格式。（3）块表的容量有多大？（4）画出直接方式地址映像及变换示意图；（5）主存地址为DE8F8H的单元在Cache中的什么位置？

(1) $16K = 2^{14} \Rightarrow w+r = 14$

$\frac{8\times 32}{8} = 2^5 \Rightarrow w = 5$

块号 r	块内地址 w
9位	5位

(2) $1M = 2^{20} \Rightarrow (s-r)+r+w = 20$

$\frac{8\times 32}{8} = 2^5 \Rightarrow w = 5$

∵直接相联 ⇒ 组数=块数 $= \frac{16K}{2^5} = 9$ ∴ $r=9$

6位	9位	5位
标记 s-r	Cache 组号 r	块内地址 w

(3) 块表：2^9 B

(5) DE8F8H = 1101 1110 1000 1111 1000

标记 110111 块号 101000111 字号 11000

Cache中地址 101000111 11000

$\Rightarrow (0010\ 1000\ 1111\ 1000)_2 = 28F8H$

一个组相联映像Cache由64个存储块组成，每组包含4个存储块。主存包含4096个存储块，每块由128字组成。访存地址为字地址。

（1）写出主存地址位数和地址格式

（2）写出Cache地址位数和地址格式；

主存地址的组成为8位区号（256区）+4位区内块号（16个存储块）+7位块内地址（一块128字，按字编址）

Cache地址为4位组号（16组）+2位组内块号（每组4快）+7位块内地址（一共128字，按字编址）

▼ Cache替换策略

1. 直接映射直接替换
2. 组相联和全相联Cache，需要替换策略

读不命中替换

LFU（最不经常使用，频率）

被访问的行+1，换掉最小的。计数器=最近访问次数

LRU（近期最少使用，时间）

被访问的行置0，其他+1，换掉最大的。计数器=上次访问时间间隔

随机替换

读未命中的替换

例：设cache有1、2、3、4共4个块，a、b、c、d等为主存中的块,访问顺序一次如下：a、b、c、d、b、b、c、c、d、d、a ,下次若要再访问e块。
问，采用LFU和LRU算法替换结果是不是相同?

	LFU（最不经常使用）					LRU（近期最少使用）				
	说明	1块	2块	3块	4块	说明	1块	2块	3块	4块
a	a进入	1	0	0	0	a进入	0	1	1	1
b	b进入	1	1	0	0	b进入	1	0	2	2
c	c进入	1	1	1	0	c进入	2	1	0	3
d	d进入	1	1	1	1	d进入	3	2	1	0
b	命中	1	2	1	1	命中	4	0	2	1
b	命中	1	3	1	1	命中	5	0	3	2
c	命中	1	3	2	1	命中	6	1	0	3
c	命中	1	3	3	1	命中	7	2	0	4
d	命中	1	3	3	2	命中	8	3	1	0
d	命中	1	3	3	3	命中	9	4	2	0
a	命中	2	3	3	3	命中	0	5	3	1
e	替换a	1	0	0	0	替换b	1	0	4	2

二路组相联Cache采用LRU替换算法
某次读Cache未命中，需进行替换操作
对应组内块计数器计数器值分别为：5、15
Cache此时选择替换掉块计数器值为？ 15

读命中无需替换

写策略

写回法

写命中时，只修改Cache内容，增加修改位，换出时，根据修改位进行写回或舍掉。

全写法

写命中时，Cache与内存一起写。

写一次法

第一次采用全写，其他采用写回。

写命中的替换

→ 1位16进制

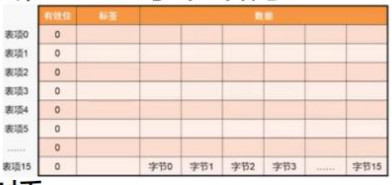
$2^4 \times 2^4$ 行4位，字4位（=进制）

有如下程序段，Cache采用写回法，16*16字节结构

#1 MOV [2011H], AL

#2 MOV [2011H], BL （AL ≠ BL）

#3 MOV DL, [4011H]



执行第3条语句，Cache的操作包括

写[2011H]单元，替换[4010H]数据块至表项1

[2011H]
↓ ↓ ↓
标签 行号 字号

① 找1行1字节，对比表项标签，#1 未命中

② 表项1标签更换为20，并把整个块 2011H～201FH写入后面

③ #2命中，只修改Cache中内容，AL改为BL

④ #3未命中，1行1字节处标签改为40，写入4011H～401FH，2011写入主存（修改后的）