

第 5 章 存储器技术

- 存储器概述
- 半导体随机存储器
- 高速缓冲存储技术
- 虚拟存储技术

● 存储器概述

存放程序和数据。

根据存储器件的性能及使用方法不同，存储器有不同分类方法。

■ 功能分类

1) 内存储器（亦称主存储器）

存取速度和CPU处理速度相当；存储器容量受到地址总线的限制；CPU可通过三总线（地址、数据、控制）直接对它进行访问；主要采用半导体存储器。

2) 外存储器（亦称辅助存储器）

存取速度比CPU处理速度慢；存储器容量不受地址总线的限制；CPU必须通过专用的I/O接口进行访问；主要采用光、磁介质存储器。

3) 高速缓冲存储器

存储内容为主存副本，速度快，不计入计算机存储系统的总容量。

● 存储器概述

■ 存取方式分类

- 1) 随机存取存储器
- 2) 只读存储器
- 3) 闪速存储器
- 4) 顺序存储器
- 5) 直接存取存储器

■ 存储介质分类

- 1) 磁芯存储器
- 2) 半导体存储器
- 3) 磁表面存储器
- 4) 光存储器

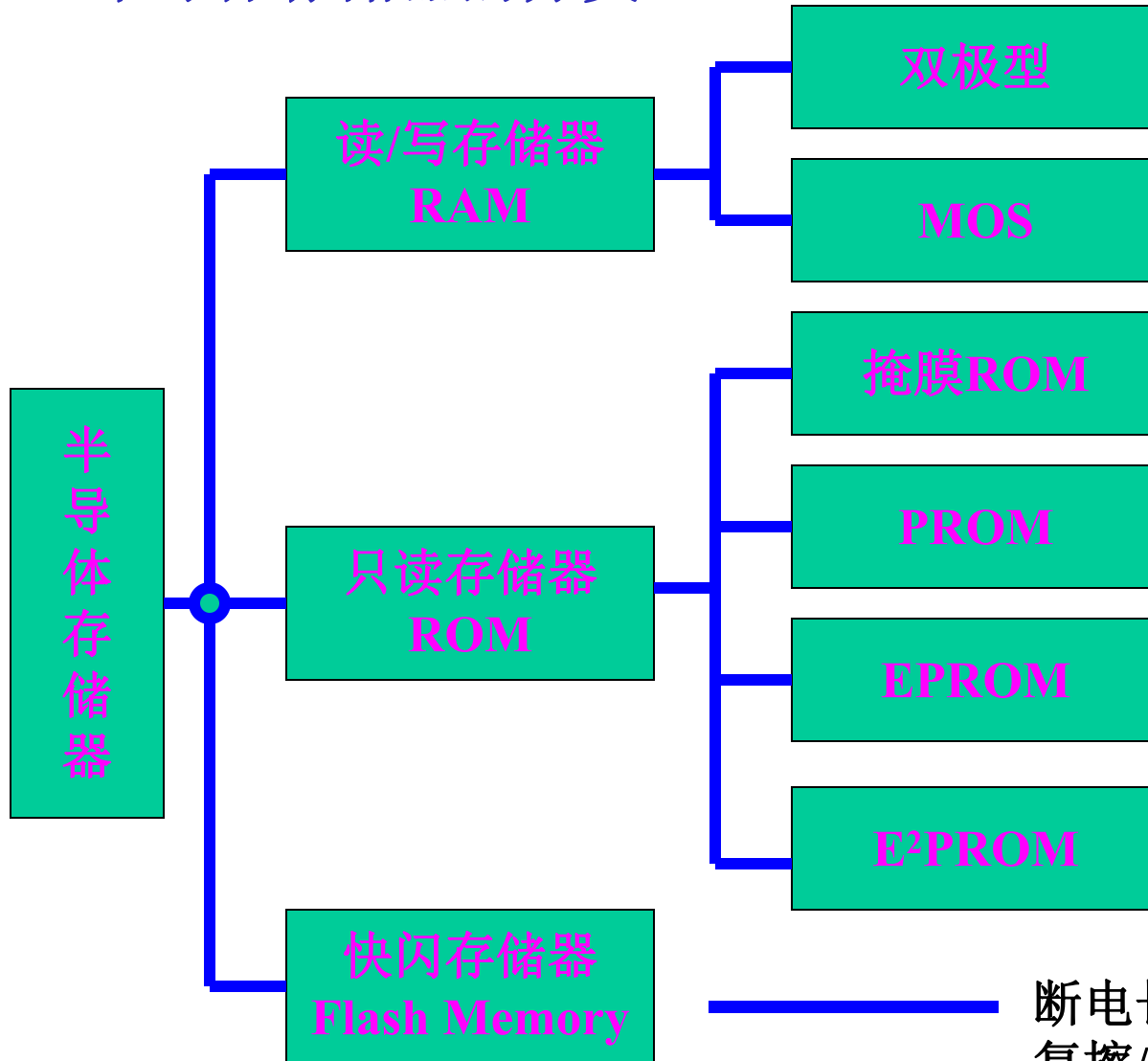
■ 信息可保存性

非永久性、永久性

■ 串并行存取方式

并行、串行

■ 半导体存储器的分类



—— 断电长期保存，快擦，多次重复擦/写

■ 存储器的主要技术指标

1. 速度 通常用存取时间或存取周期表示

存取时间：访问时间，启动一次存储器存取操作到完成该操作所经历的时间。

存取周期：存储周期或读写周期，对存储器进行连续两次存取操作所需要的最小时间间隔。

2. 存储容量 常用存储有多少个存储单元，每个存储单元又有多少个二进制位来表示，也就是说存储容量由二部分组成：

存储容量 = 存储单元数 * 每单元位数

如： $256 * 4$ $1024 * 4$ (1K * 4)

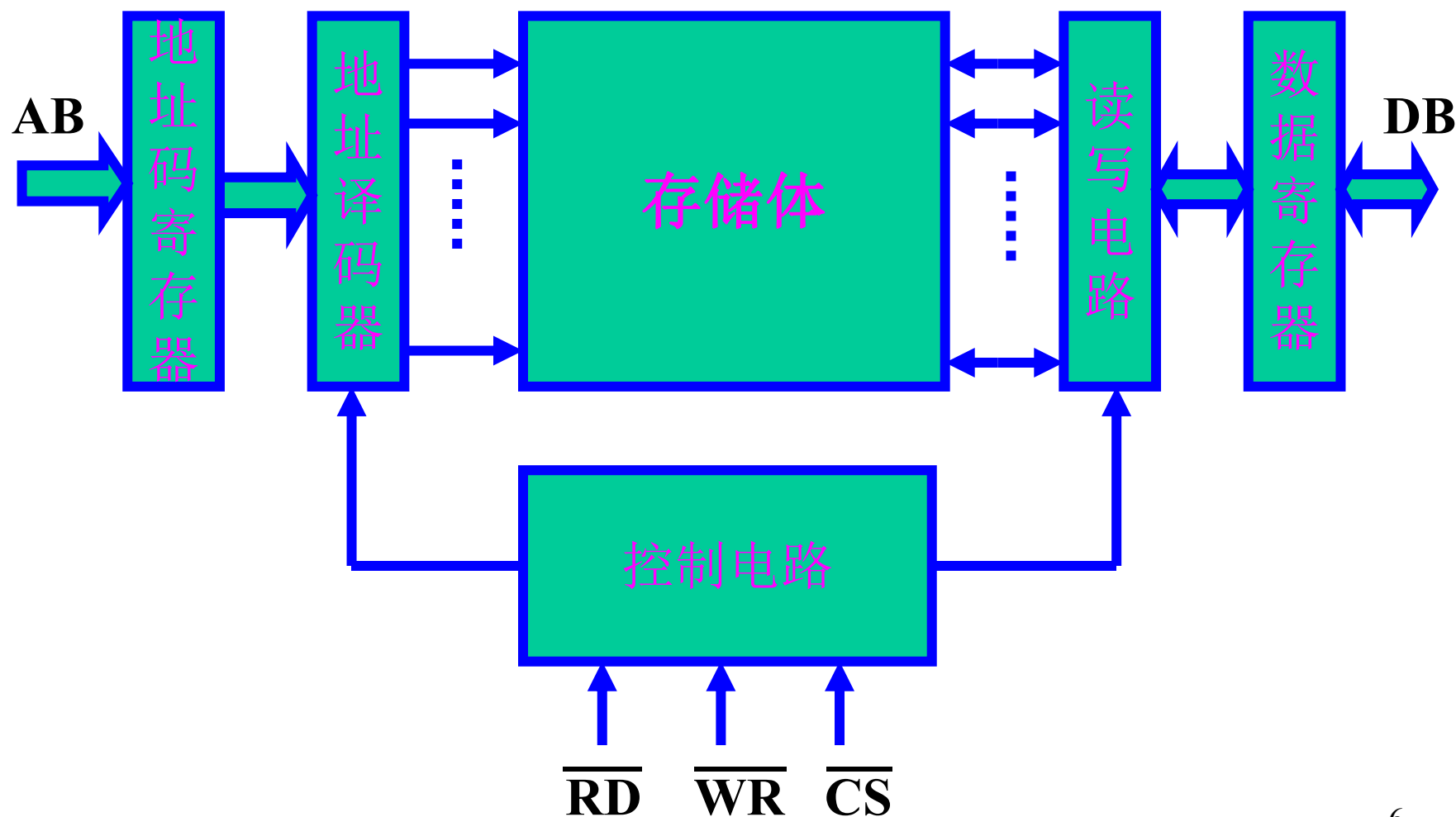
3. 存储带宽 单位时间内存储器所存取的信息量，位/秒，或字节/秒为单位，如B：带宽，F：存储器频率，D：存储器数据线位数

$$B = F \times D / 8$$

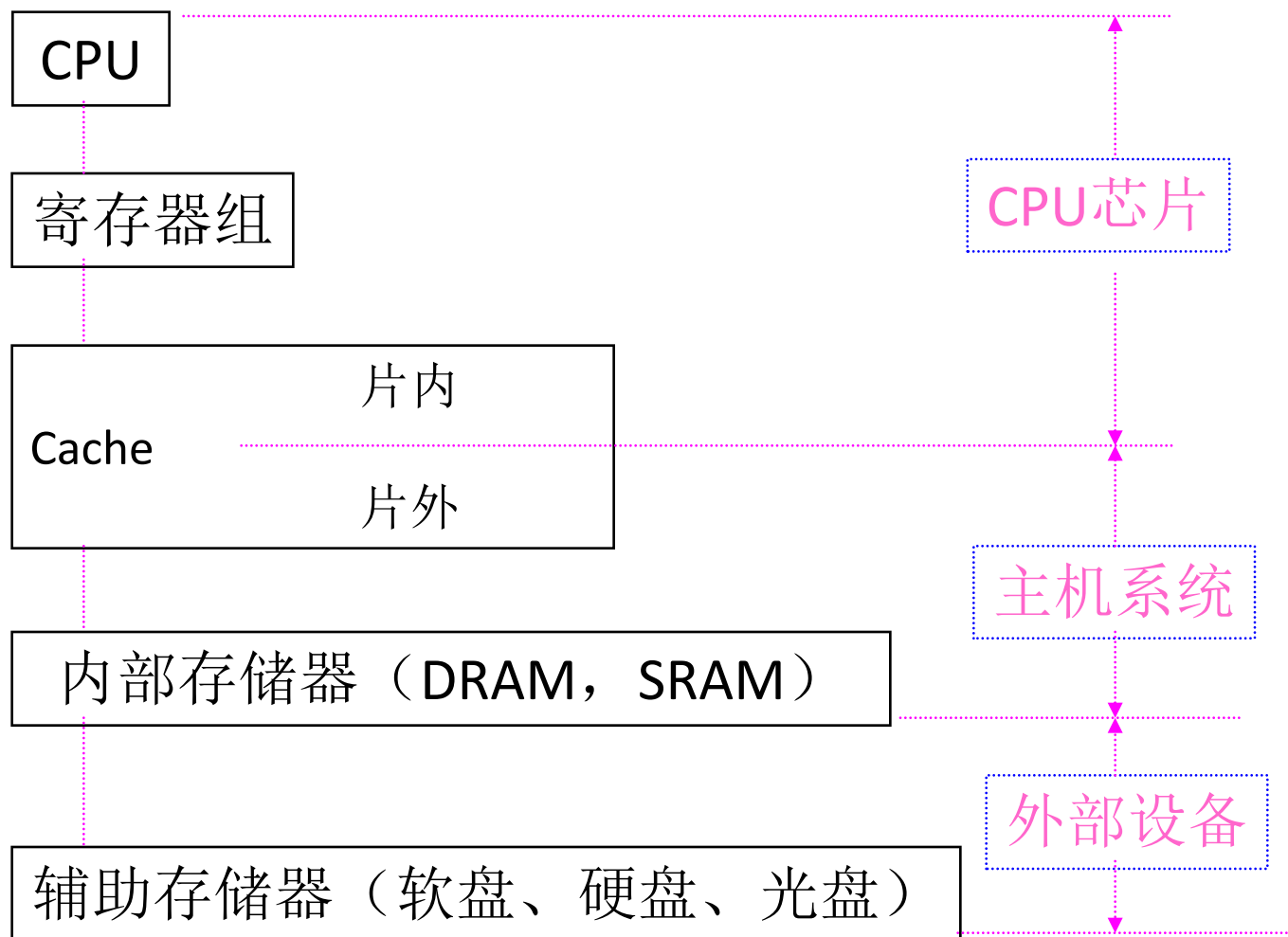
4. 存储器的可靠性 通常用平均故障时间衡量

■ 存储器的基本结构框架

内部组成一般可分为存储体、地址译码器、读写控制电路三大部分。



■ 存储系统的层次结构



■ 半导体随机存储器

1. 静态SRAM存储单元电路

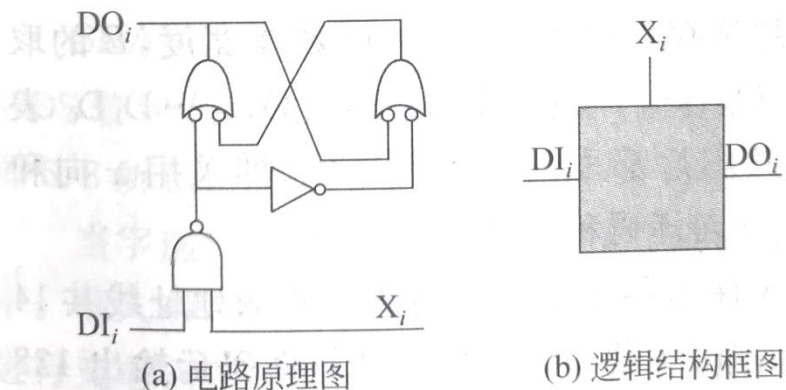


图 5.4 SRAM 存储元的电路原理图和逻辑结构框图

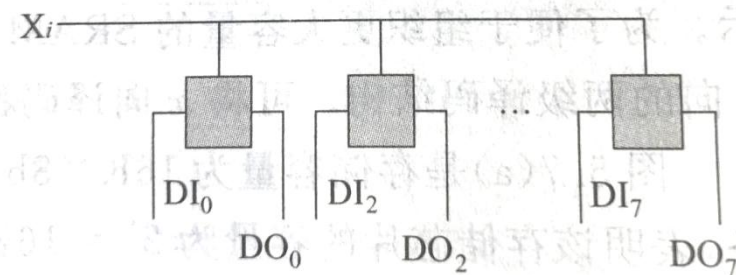


图 5.5 SRAM 存储单元连接结构框图

静态RAM的特点：

- 1) 只要电源接通所存储的信息就不会丢失；
- 2) 不要刷新电路；
- 3) 容量较动态RAM为小。

2. 静态SRAM存储基本存储阵列

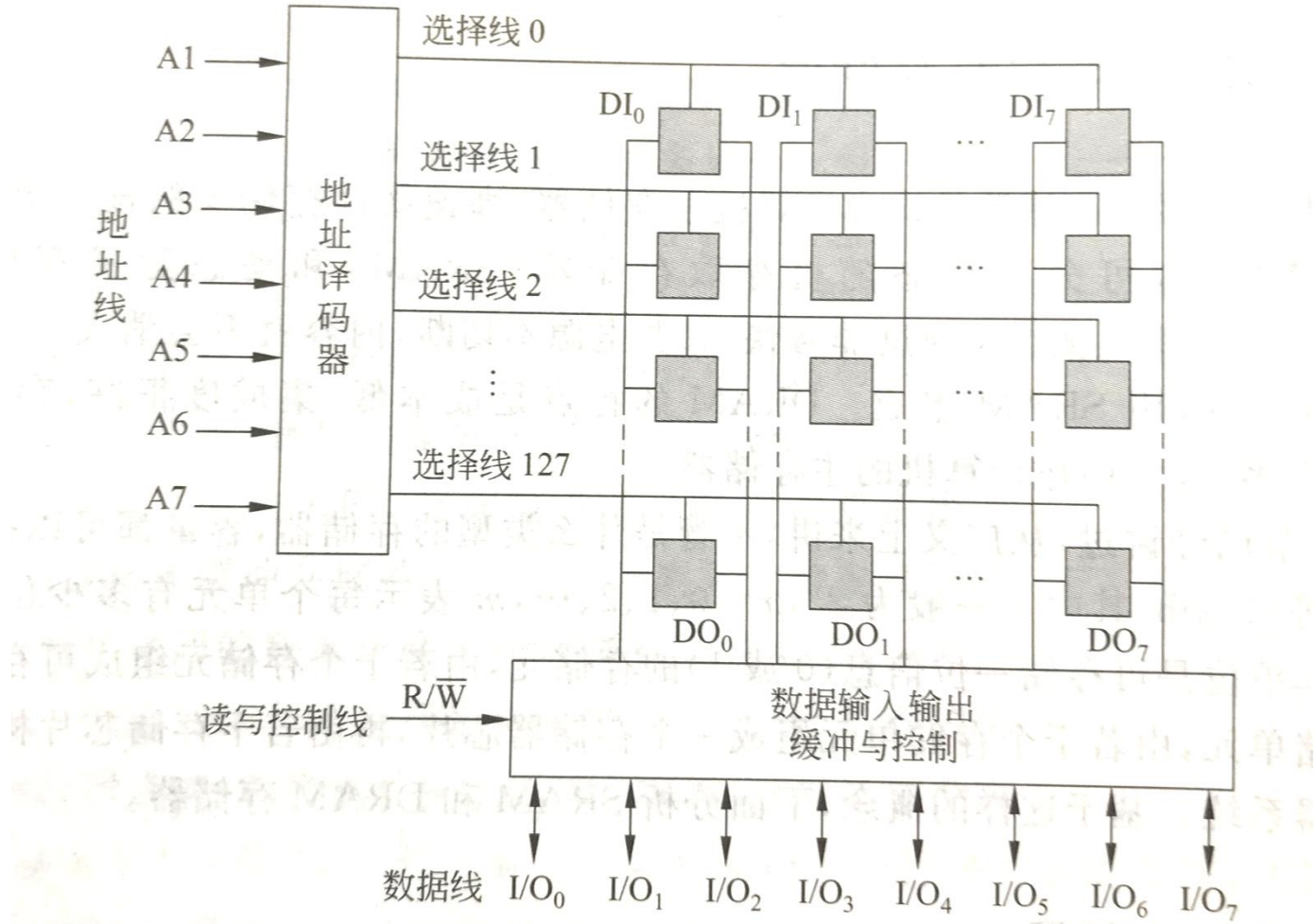
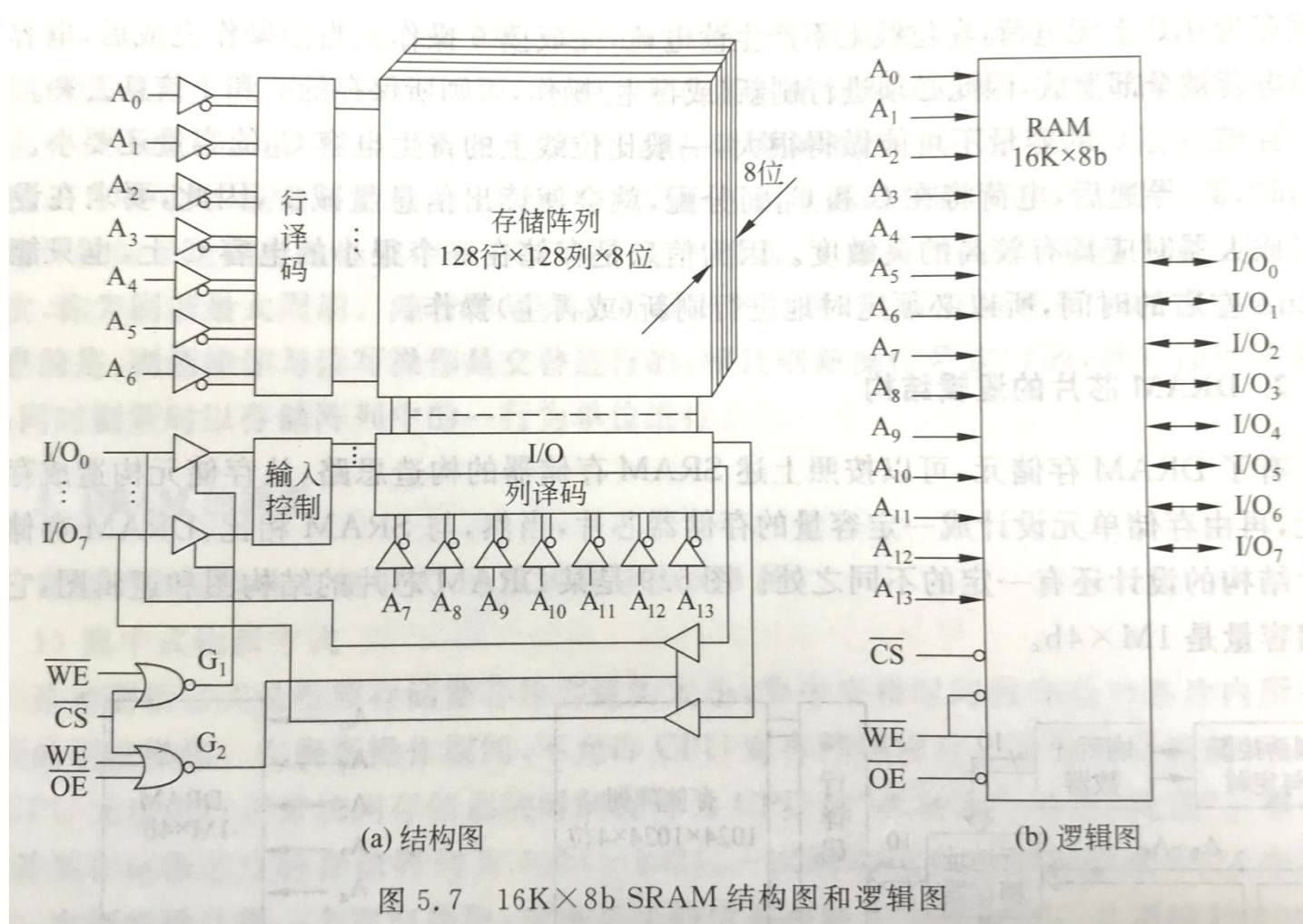
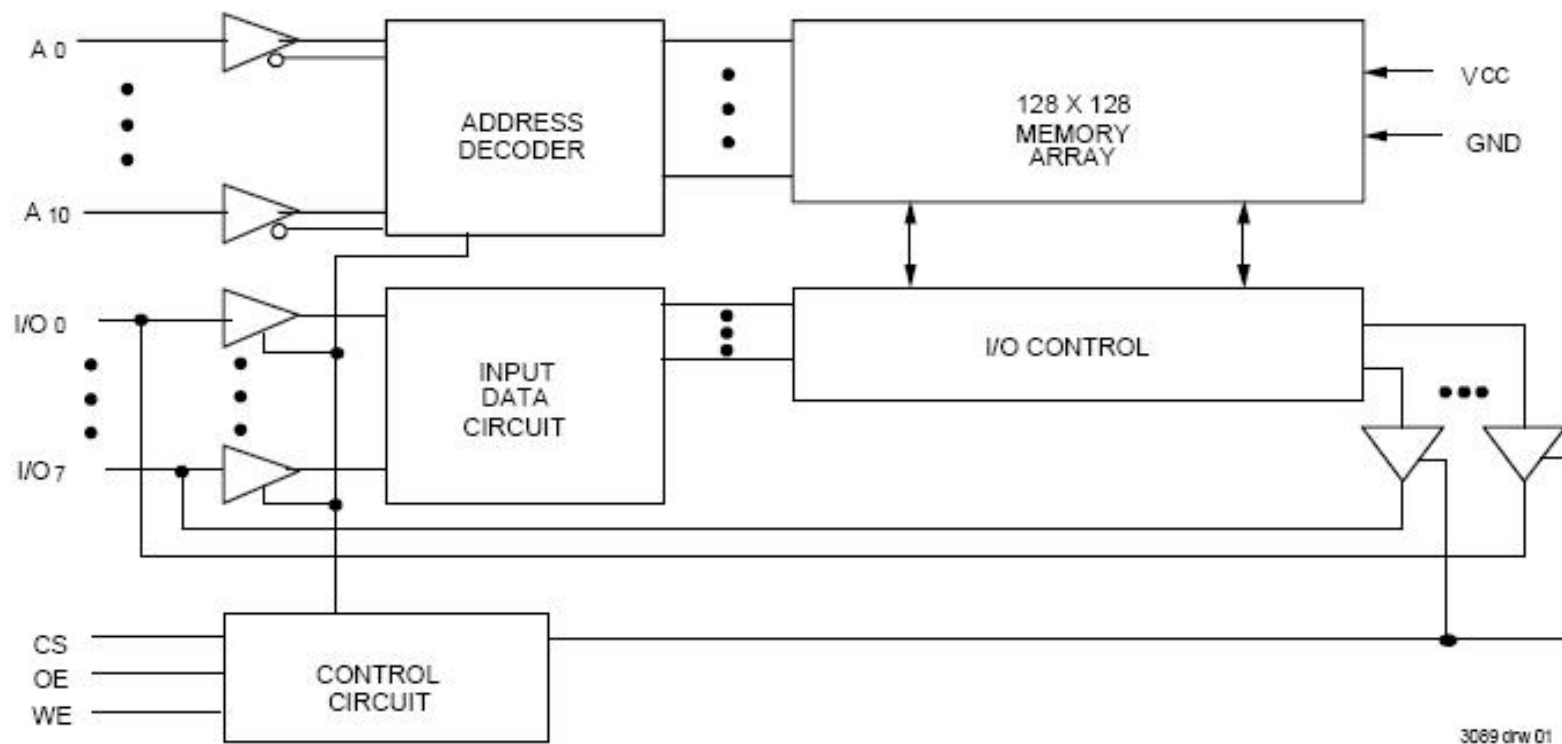


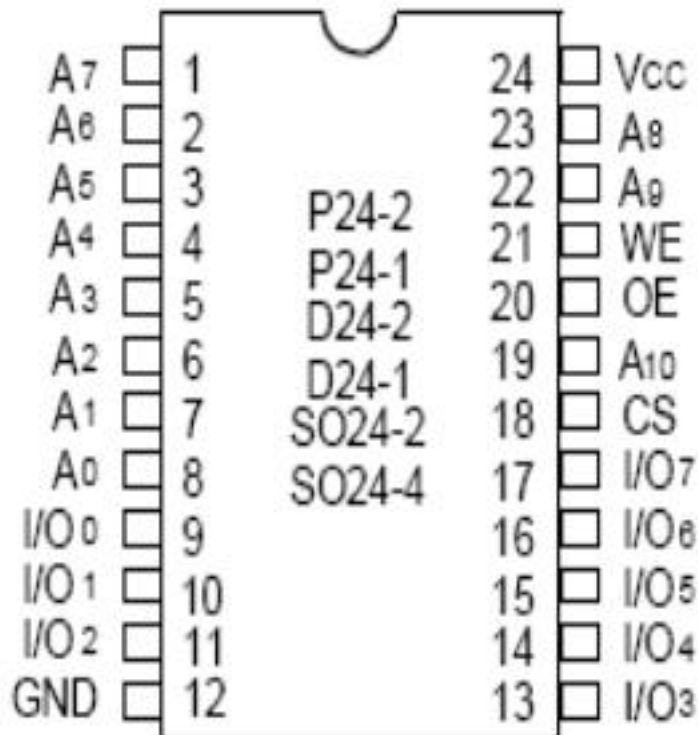
图 5.6 基本的静态存储元阵列

3. 存储器芯片的逻辑结构



4. 典型芯片 6116 (2K×8)





| Name | Description |
|-------------------------------------|-------------------|
| A ₀ - A ₁₀ | Address Inputs |
| I/O ₀ - I/O ₇ | Data Input/Output |
| \overline{CS} | Chip Select |
| \overline{WE} | Write Enable |
| \overline{OE} | Output Enable |
| V _{cc} | Power |
| GND | Ground |

6116的运行方式

| Mode | \overline{CS} | \overline{OE} | \overline{WE} | I/O |
|---------|-----------------|-----------------|-----------------|---------------------|
| Standby | H | X | X | High-Z |
| Read | L | L | H | DATA _{OUT} |
| Read | L | H | H | High-Z |
| Write | L | X | L | DATA _{IN} |

■ 动态DRAM存储器

1. 动态RAM存储单元电路

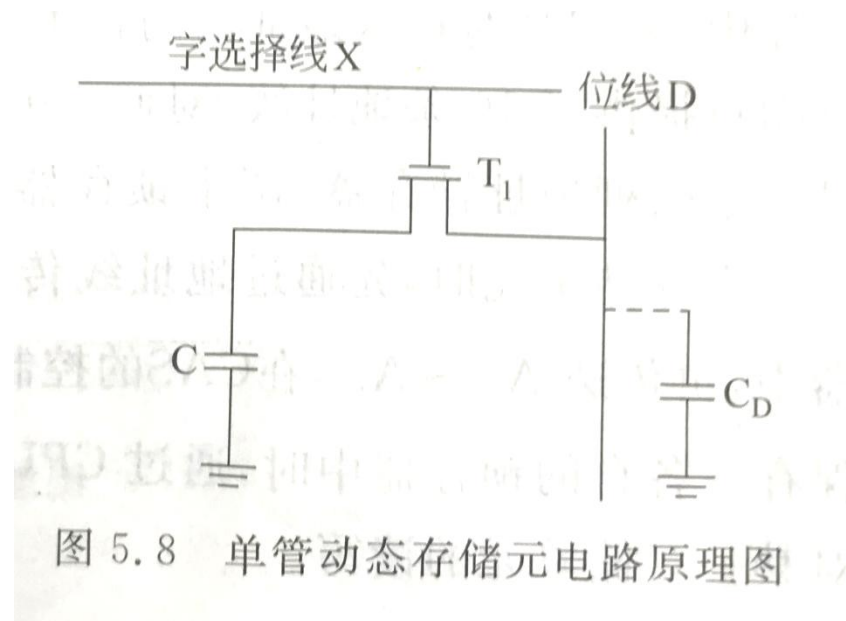


图 5.8 单管动态存储元电路原理图

动态RAM的特点：

- 1) 每隔大约2ms必须对数据进行刷新，否则所存储的信息就会丢失；
- 2) 要有刷新电路；
- 3) 容量较静态RAM为大。

2. 动态DRAM存储阵列

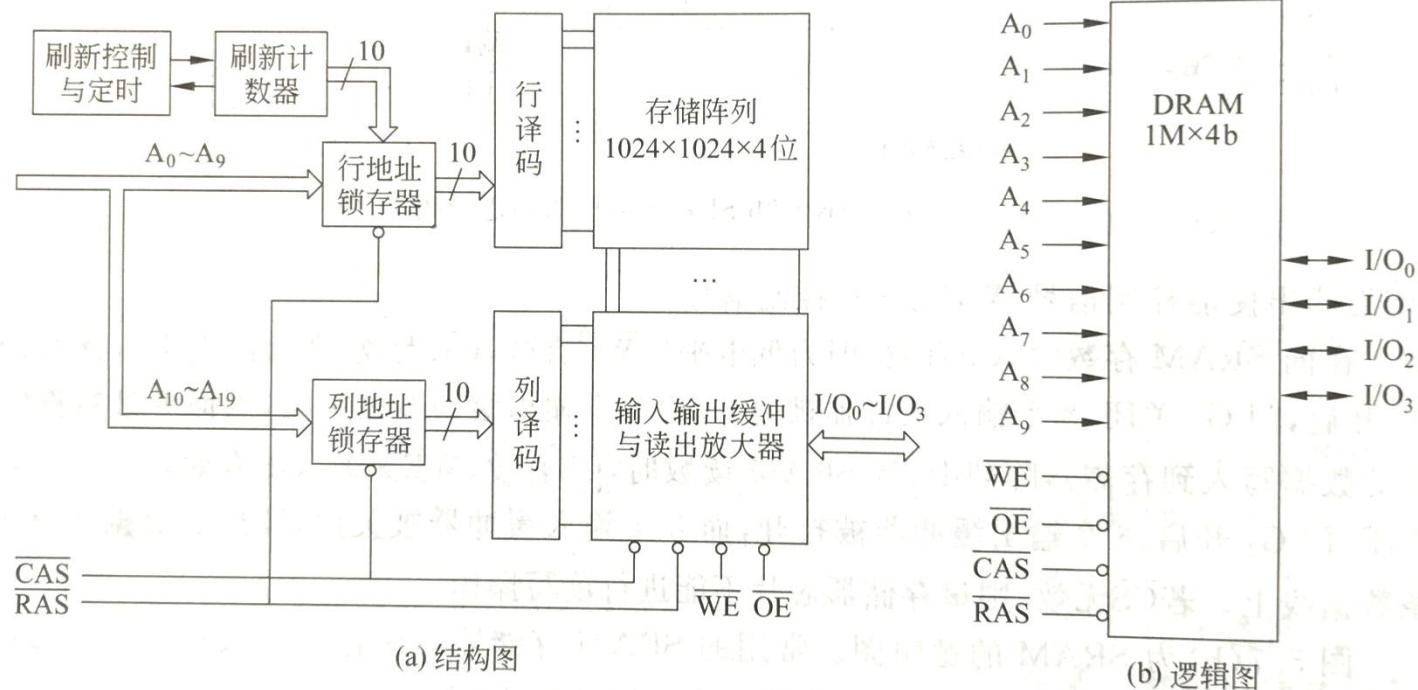


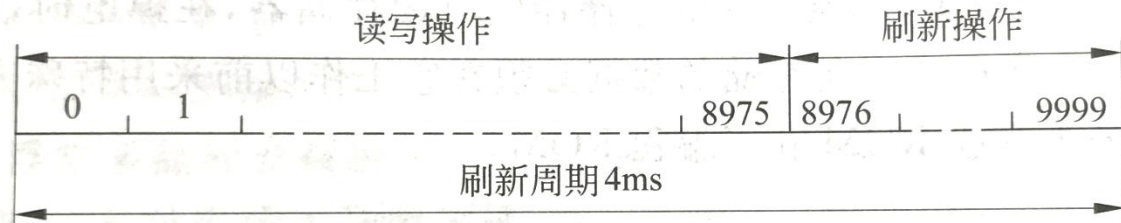
图 5.9 1M×4b DRAM 结构图和逻辑图

DRAM刷新控制：

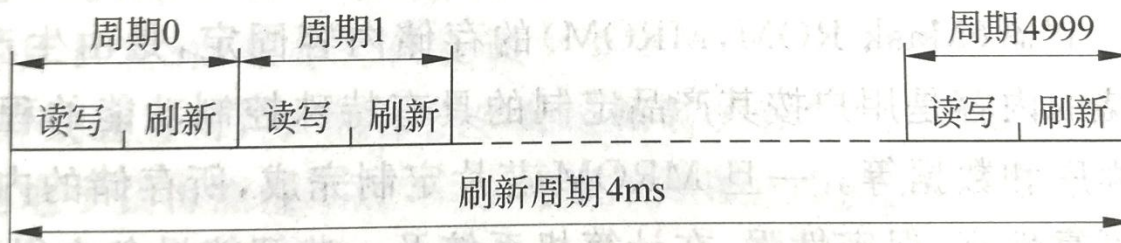
- 1) 时序功能；
- 2) 地址处理功能；
- 3) 仲裁功能。

3 DRAM的刷新

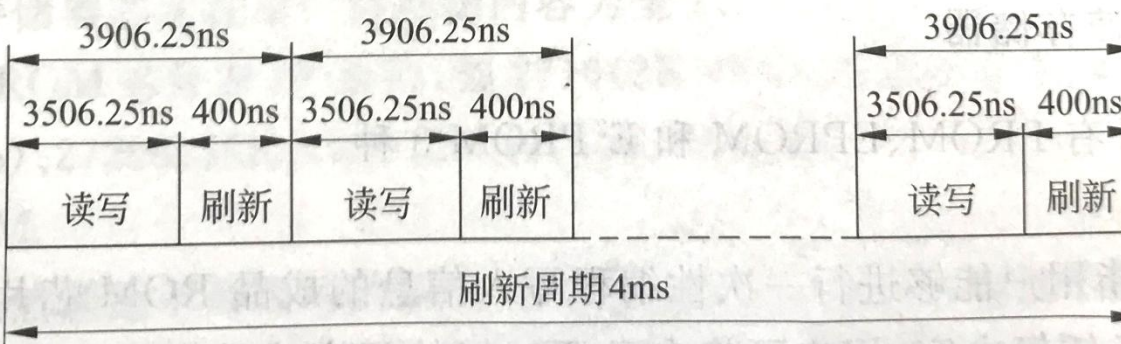
- 1) 集中式刷新
- 2) 分散式刷新
- 3) 异步刷新



(a) 集中式刷新方式

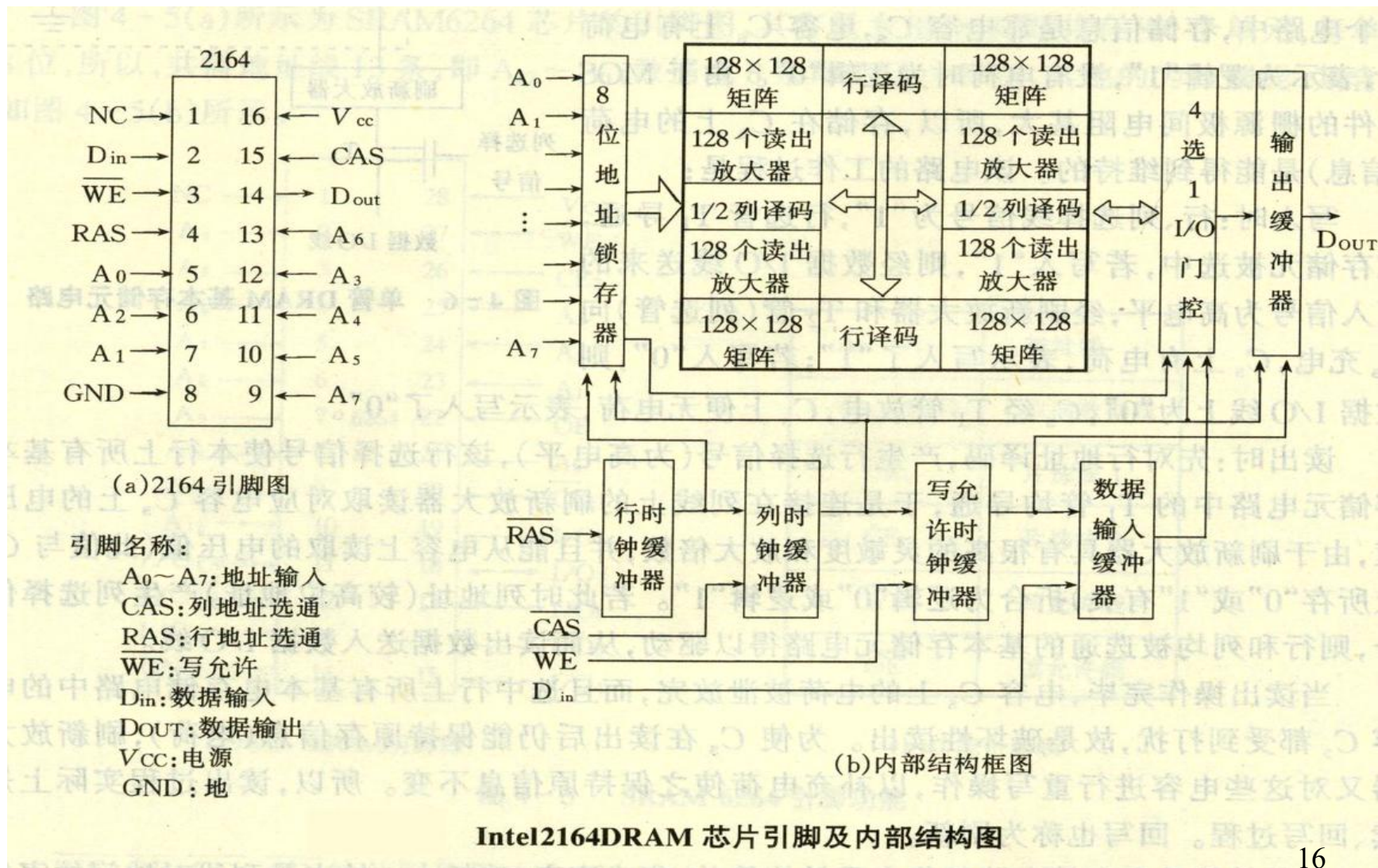


(b) 分散式刷新方式



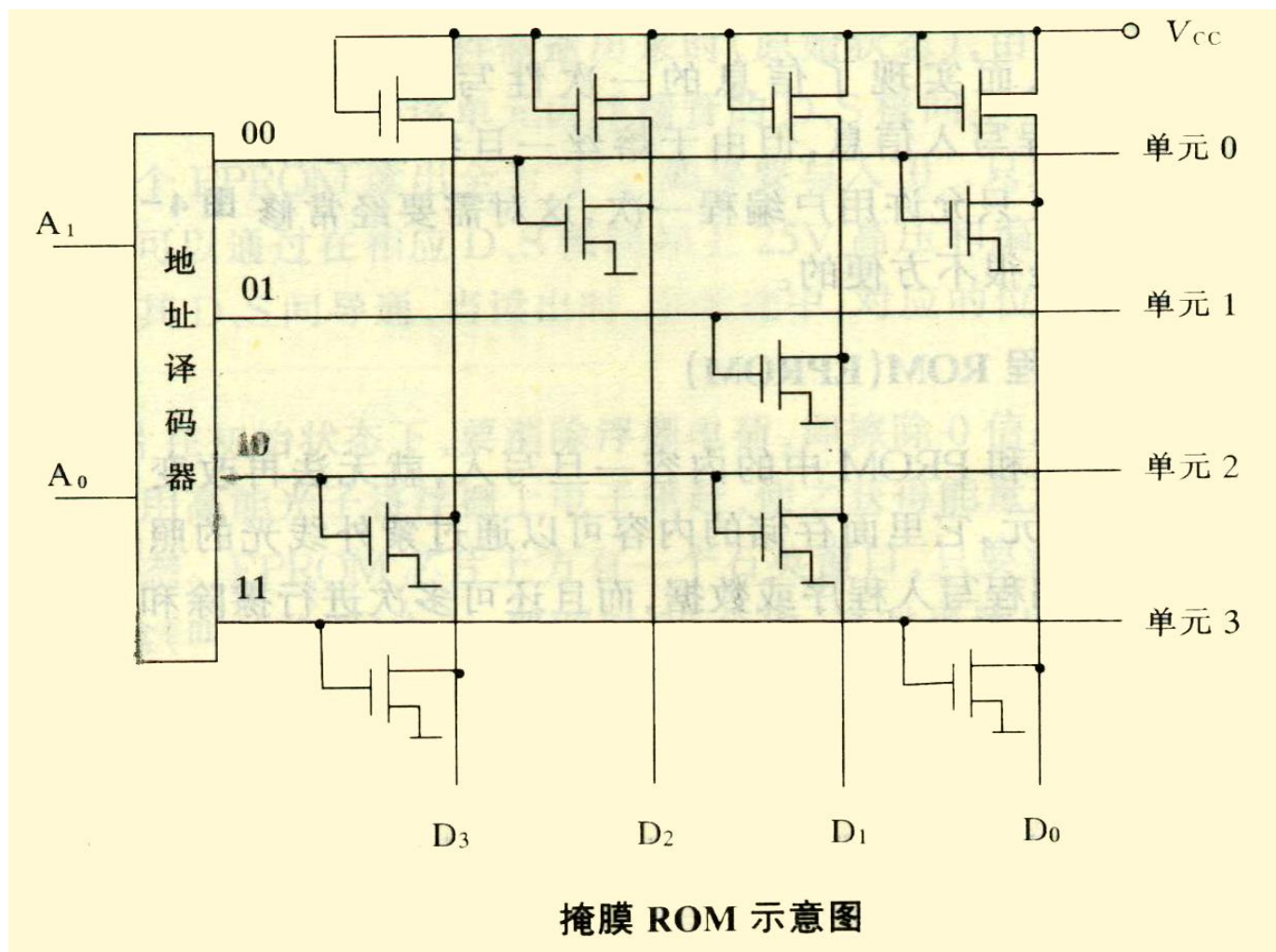
(c) 异步式刷新方式

4 典型芯片 2164 (64K × 1)



■ 只读存储器(ROM)

1 掩膜ROM基本原理



数据

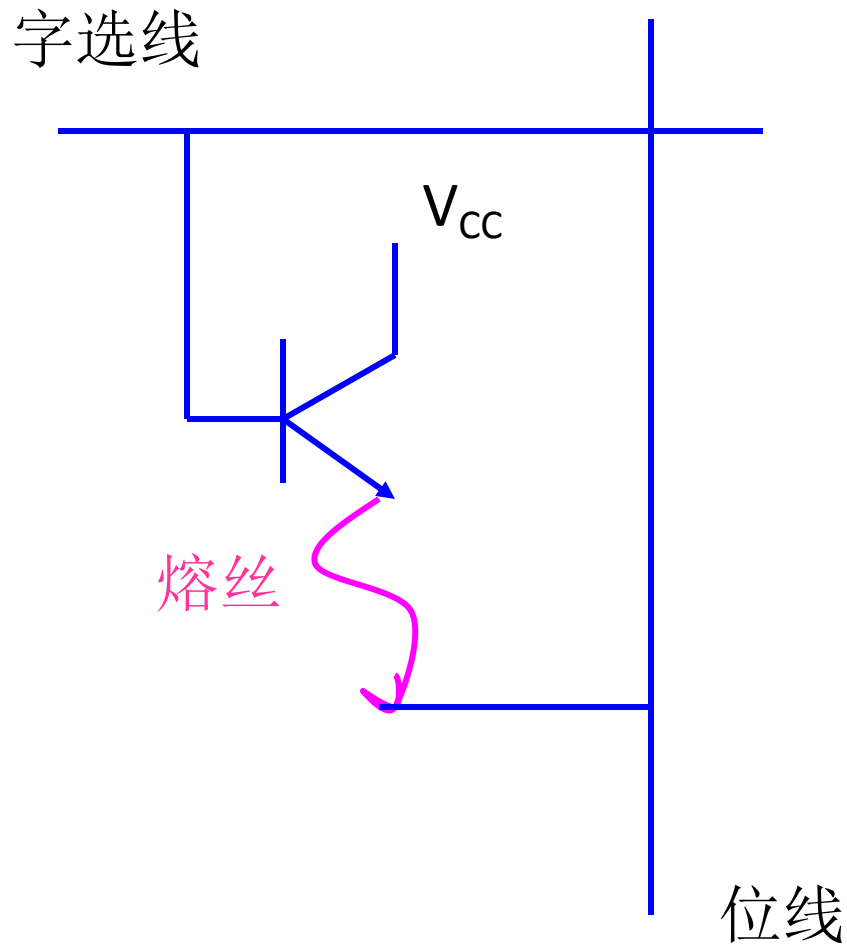
1010

1101

0101

0110

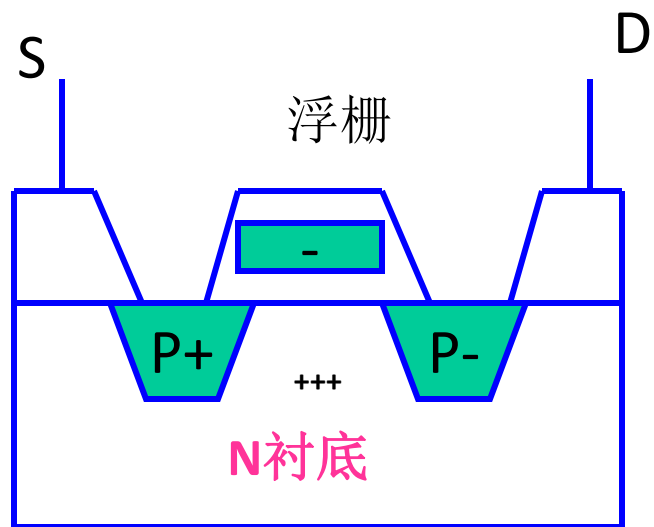
2 PROM基本原理



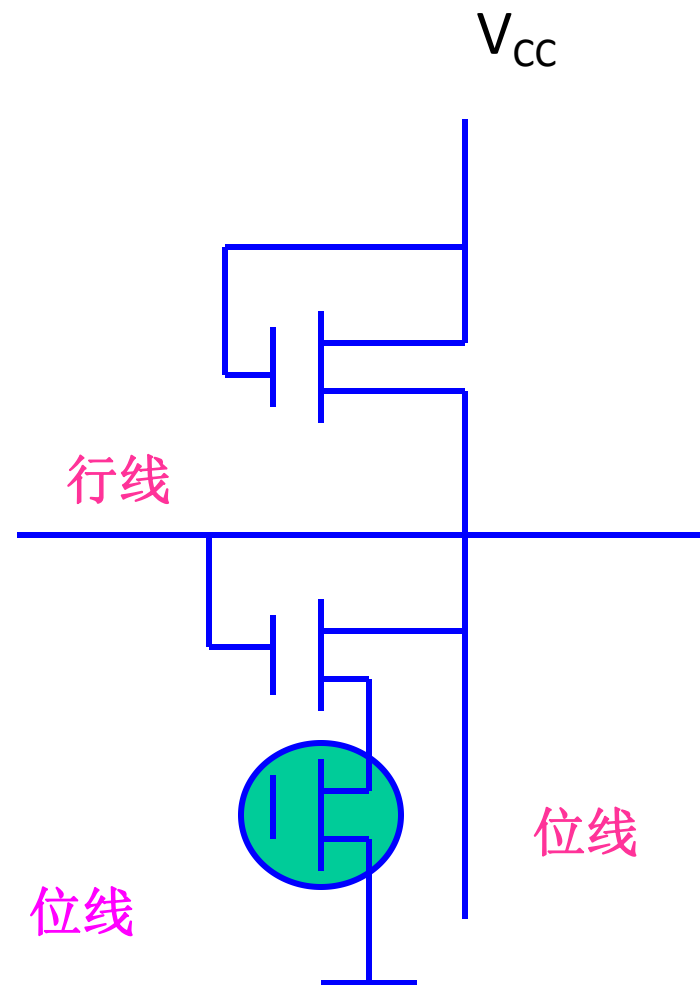
每一位三极管的发射极上串接一个熔丝，把字选线和位线连通（表示0），编程时加入大电流把熔丝烧断（表示1）。

3 EPROM基本原理

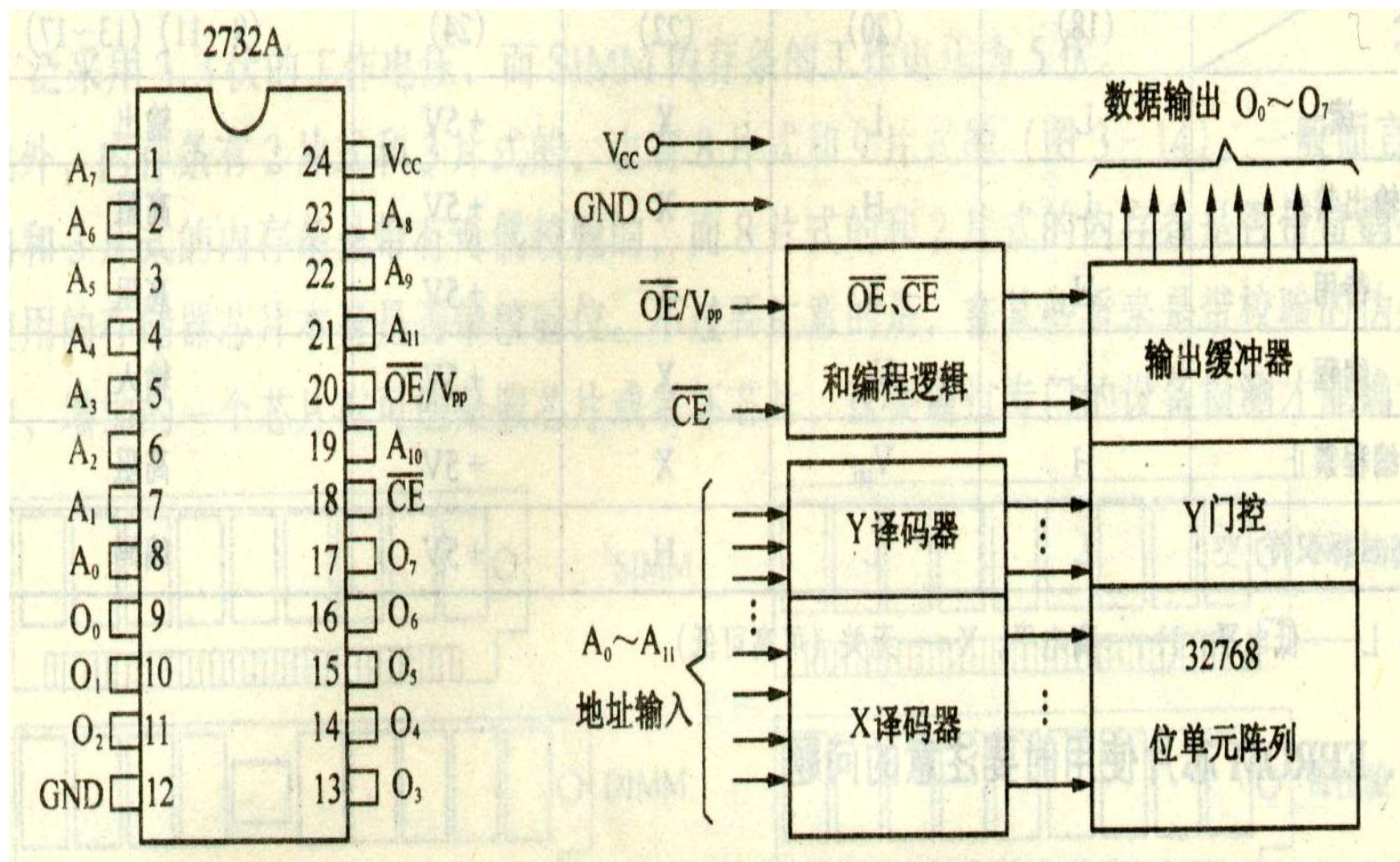
可多次编程，紫外线可擦除。



当浮栅上无电荷时，D、S 间不导通，位线输出为1。




■ EPROM典型芯片 2732 (4 K X 8)



2732A的方式选择



| 引脚 方式 | $\overline{\text{CE}}$ | $\overline{\text{OE}}/V_{\text{PP}}$ | V_{CC} | $O_0—O_7$ |
|----------|--|--------------------------------------|-----------------|-----------|
| 读 | L | L | +5V | 输出 |
| 输出禁止 | L | H | +5V | 高阻 |
| 待用 | H | X | +5V | 高阻 |
| 编程 |  | V_{PP} | +5V | 输入 |
| 编程禁止 | H | V_{PP} | +5V | 高阻 |
| 检验 | L | V_{PP} | +5V | 输出 |

待用：在待用方式下工作电流从125mA降为35mA。

编程：在数据地址稳定后，/CE端加入宽度为50ms的脉冲。

■ 存储器与CPU的连接要考虑的问题

1. 总线的负载能力;
2. CPU与存储器的速度配合;
3. 存储器容量、地址的分配和片选;
4. 数据线和控制线的连接。

■ 主存储器容量的扩展

- 位扩展

存储芯片每个单元的数据位不能满足存储器需要的位数，进行位扩展，容量没有增加。

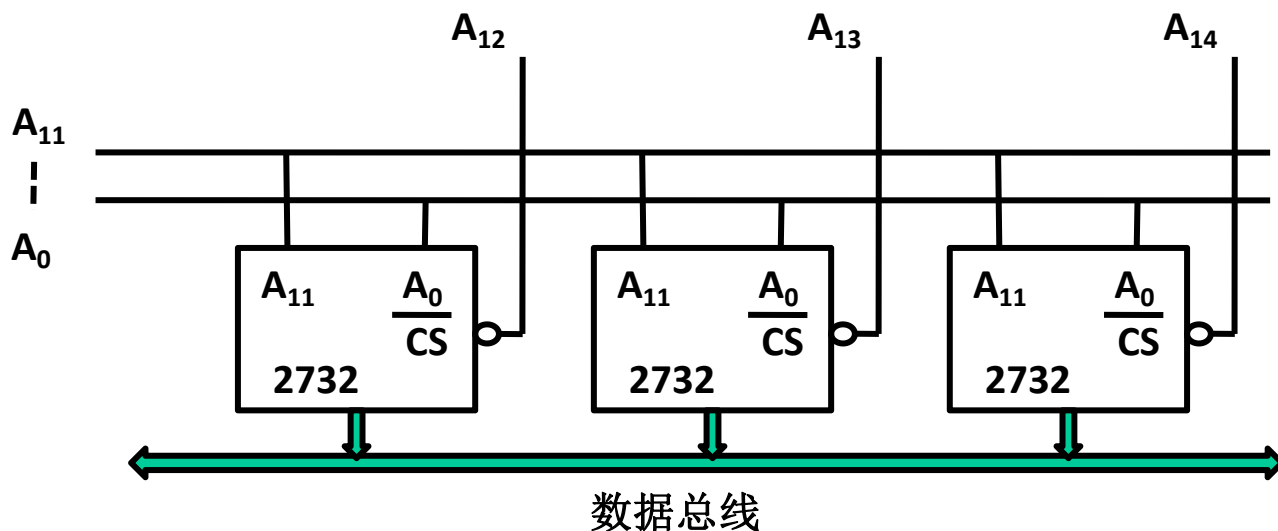
● 字扩展

存储器容量的扩展，地址的扩展。

1. 线选法

线选法即直接将地址线高位接往某存储芯片的片选端，低位地址接芯片内地址线

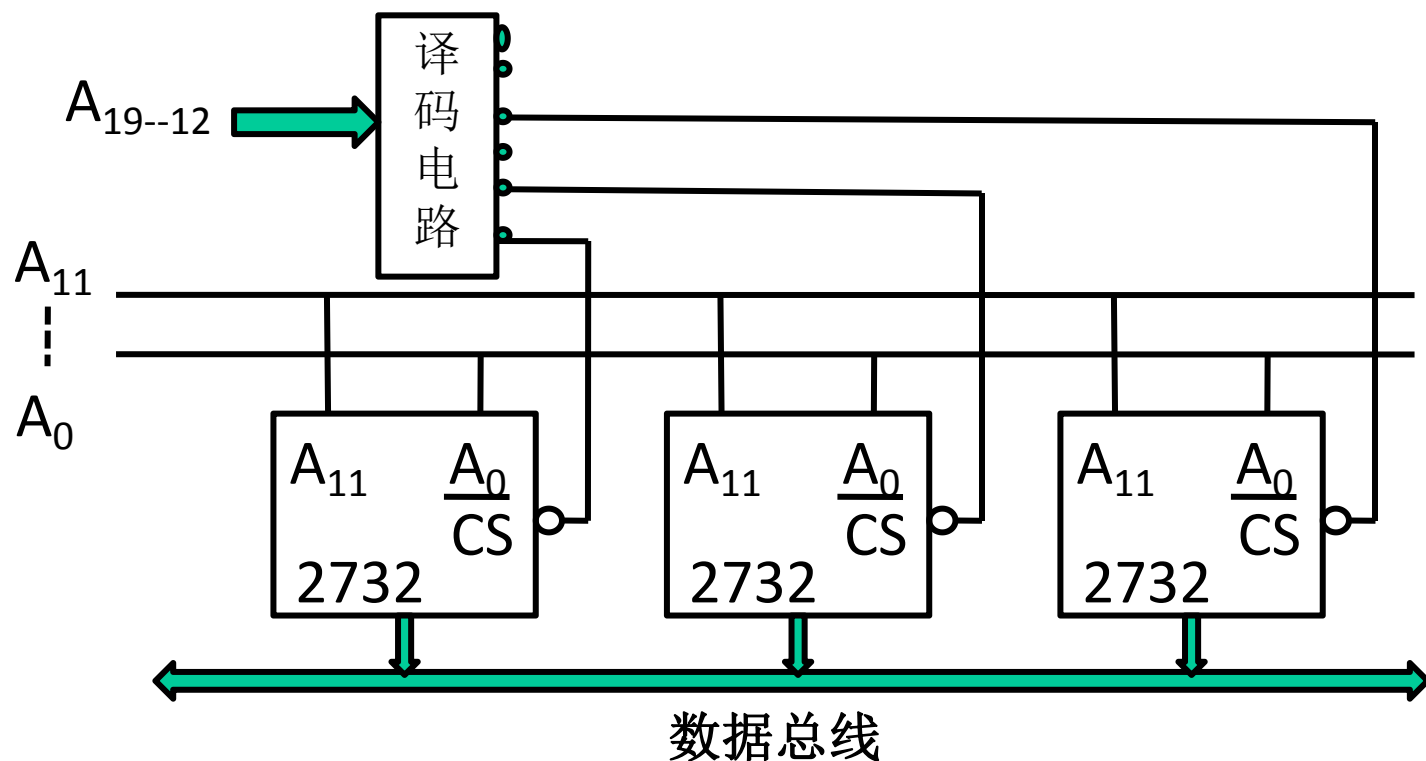
特点：线路简单，可用于较小的微机系统中；不能充分利用系统的存储器空间，地址空间不连续，给编程带来了一定的困难



2. 全译码法

N条低位地址线接存储芯片的地址端，进行片内存贮单元寻址；CPU地址总线中剩下的高位地址线全部接往地址译码器的输入端，译码后作为各芯片的片选信号，实现对存储芯片的片选。

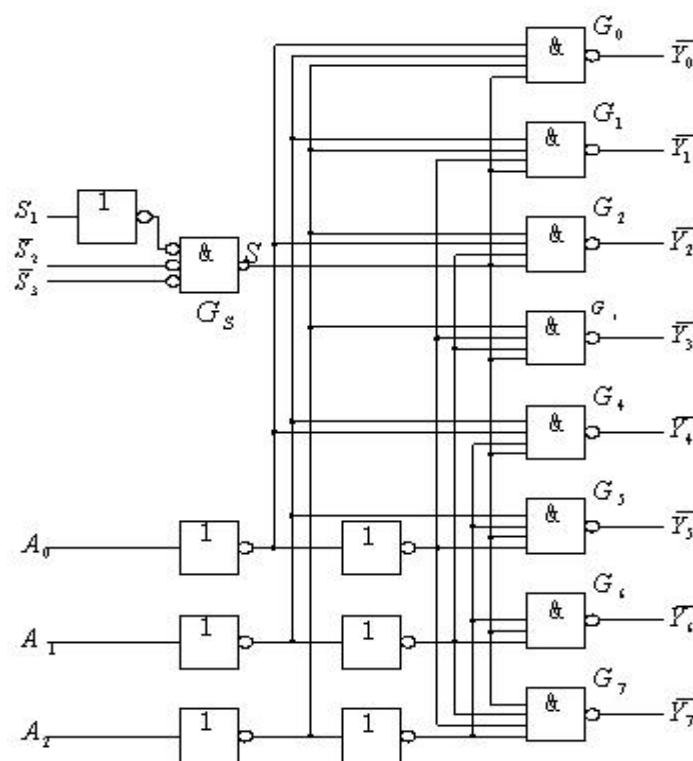
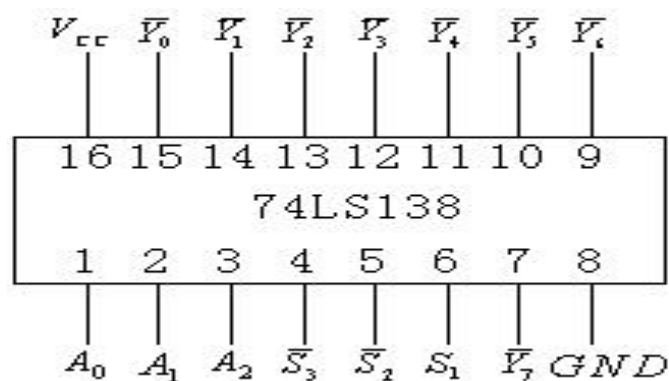
特点：需用译码器，电路复杂，应用于存储芯片较多时；地址空间连续

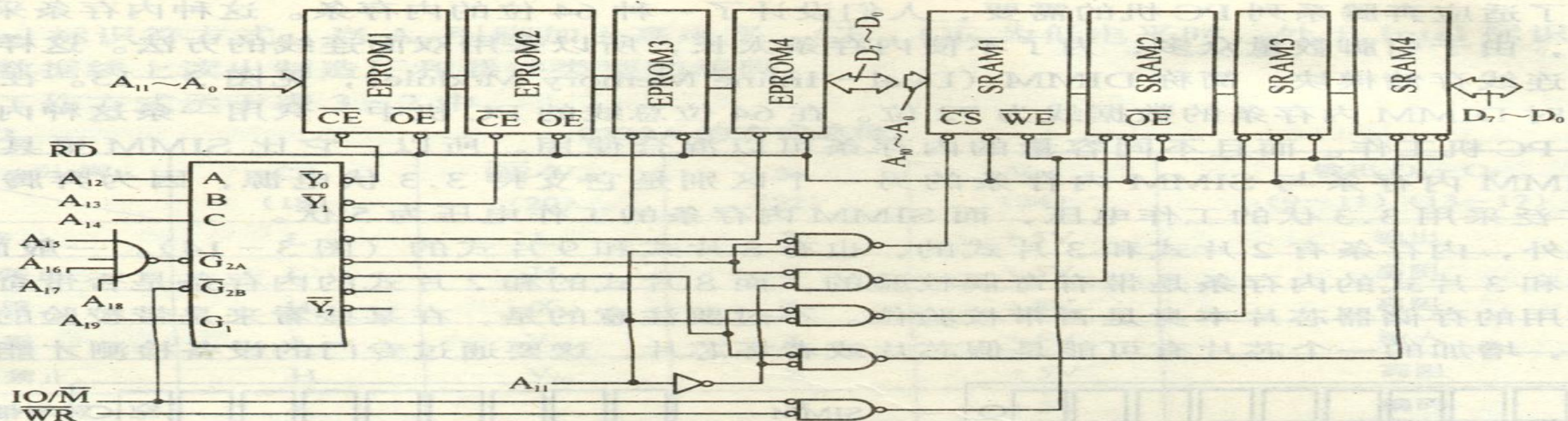


3. 部分译码法

地址线高位进行译码产生片选信号时，有的地址线未参加译码，这些地址线在需要时可直接与芯片片选信号相连，以对芯片进行线选。

部分译码法常用的典型译码器为74LS138(或Intel8205)，即3—8译码器。它有3个输入端，3个控制端及8个输出端。





一个 8 位微机系统的存储器子系统

| | A ₁₉ | A ₁₈ | A ₁₇ | A ₁₆ | A ₁₅ | A ₁₄ | A ₁₃ | A ₁₂ | A ₁₁ | A ₁₀ | A ₉ | A ₈ | A ₇ | A ₆ | A ₅ | A ₄ | A ₃ | A ₂ | A ₁ | A ₀ |
|--------|-----------------|---------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| | G ₁ | $\overline{G_{2A}}$ | | | | C | B | A | | | | | | | | | | | | |
| EPROM1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| EPROM2 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| EPROM3 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| EPROM4 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| SRAM1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| SRAM2 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| SRAM3 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| SRAM4 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

存储器芯片的地址范围为：

EPROM1: F8000H----F8FFFH

EPROM2: F9000H----F9FFFH

EPROM3: FA000H---FAFFFH

EPROM4: FB000H---FBFFFH

SRAM1 : FC000H---FC7FFH

SRAM2 : FC800H---FCFFFH

SRAM3 : FD000H---FD7FFH

SRAM4 : FD800H---FDFFFH

例：CPU具有16条地址线，16条双向数据线，控制总线中与主存有关信号：MREQ, R/W。主存按字编址，其地址空间分配如下：0~1FFFH为系统程序区，由EPROM芯片组成；从2000H起共24KB地址空间为用户程序区；最大4KB地址空间为系统程序工作区。现有如下芯片：

EPROM: $4\text{KB} \times 8\text{b}$, $8\text{K} \times 8\text{b}$

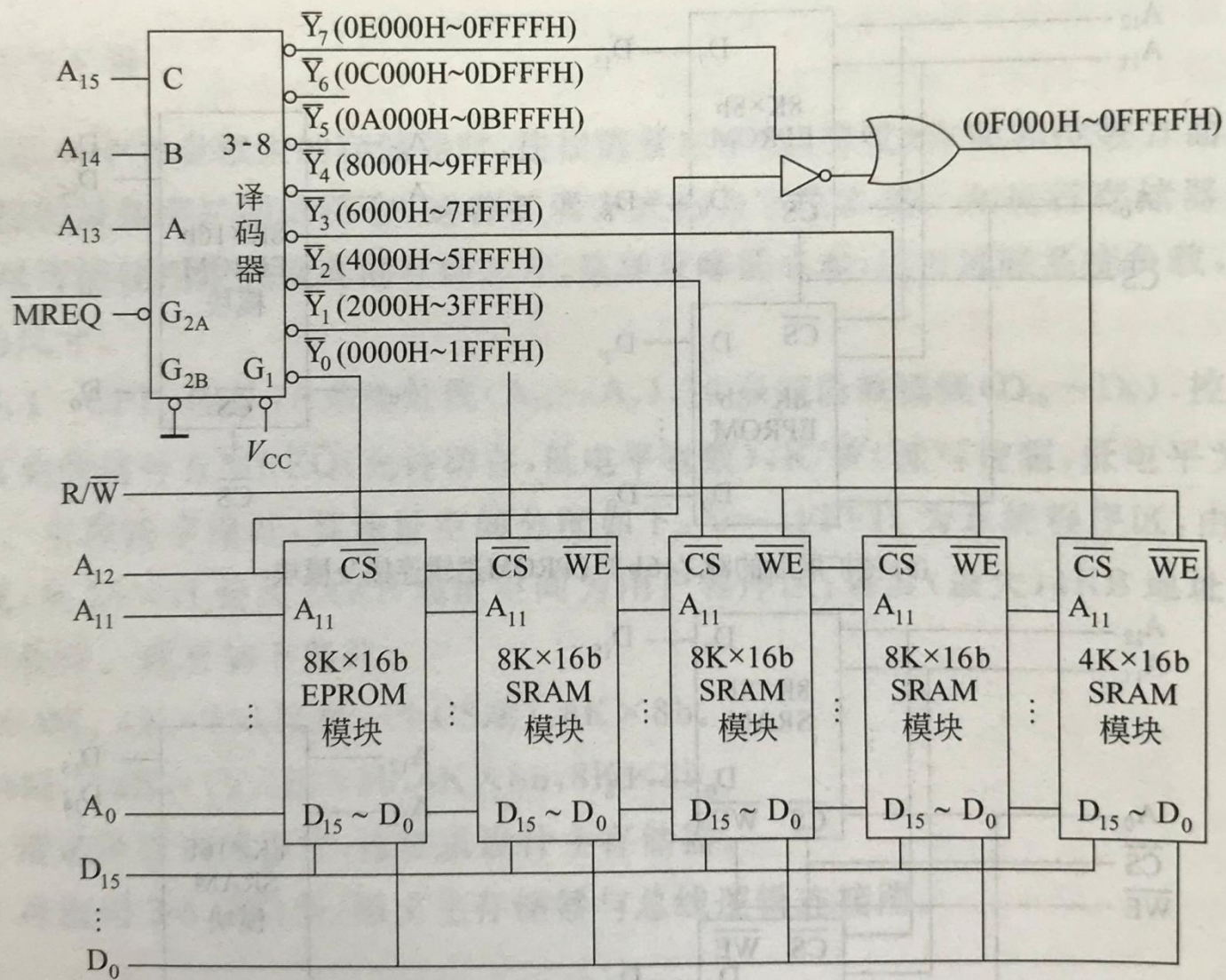
SRAM: $16\text{K} \times 1\text{b}$, $2\text{K} \times 8\text{b}$, $4\text{K} \times 8\text{b}$, $8\text{K} \times 8\text{b}$

(1) 请选择适当芯片，按要求设计主存储器。

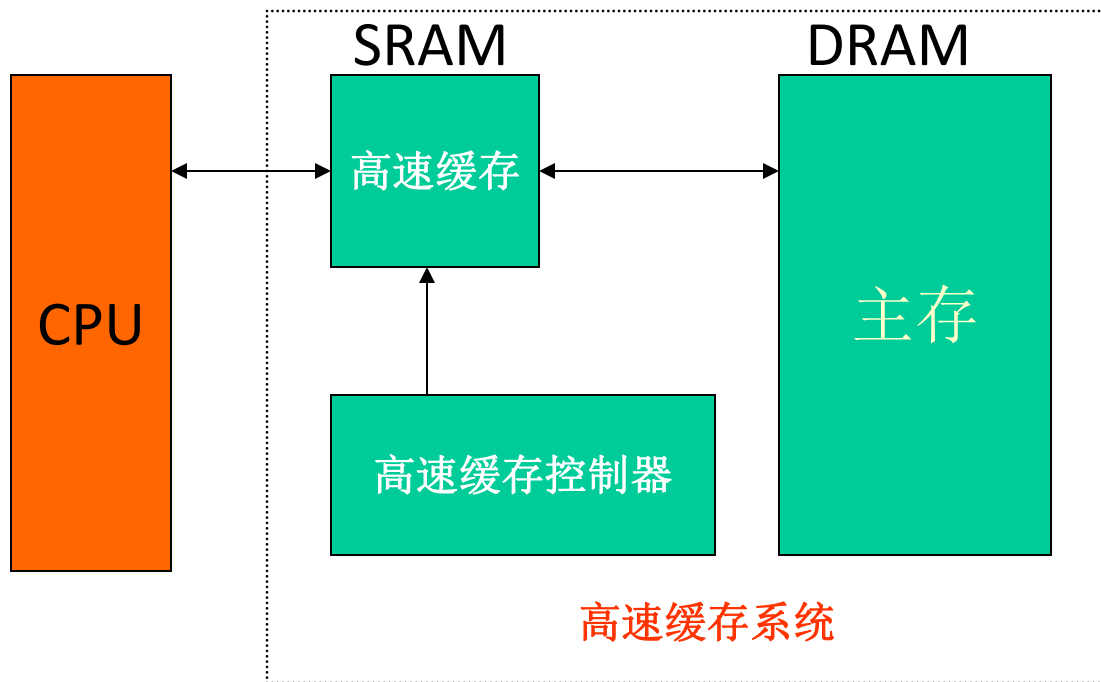
(2) 可选用3-8译码器，画出主存储器与总线逻辑连线图。

主存空间分配及选用的存储芯片

| 地址空间 | 空间功能 | 可选用的存储类型 |
|---------------|---------|----------|
| 0000H~1FFFH | 系统程序区 | EPROM |
| 2000H~7FFFH | 用户程序区 | SRAM |
| 8000H~EFFFH | 保留 | |
| 0F000H~0FFFFH | 系统程序工作区 | SRAM |



■ 高速缓存技术



■ Cache基本原理

程序的局部性原理

■ Cache的管理

- 映射方式
直接映射、全相联、组相联
- 替换策略
最不经常使用、最近最少使用、随机替换

■ Cache和主存间的映射

1) 命中、不命中、命中率

- Cache命中 (hit)
 - CPU欲访问的数据已在缓存中，即可直接访问Cache
- Cache不命中 (miss)
 - CPU欲访问的数据不在Cache内，此时需将该数所在的主存整个子块一次调入Cache
- 命中率是指CPU要访问的信息已在Cache内的比率。通常用命中率来衡量Cache的效率

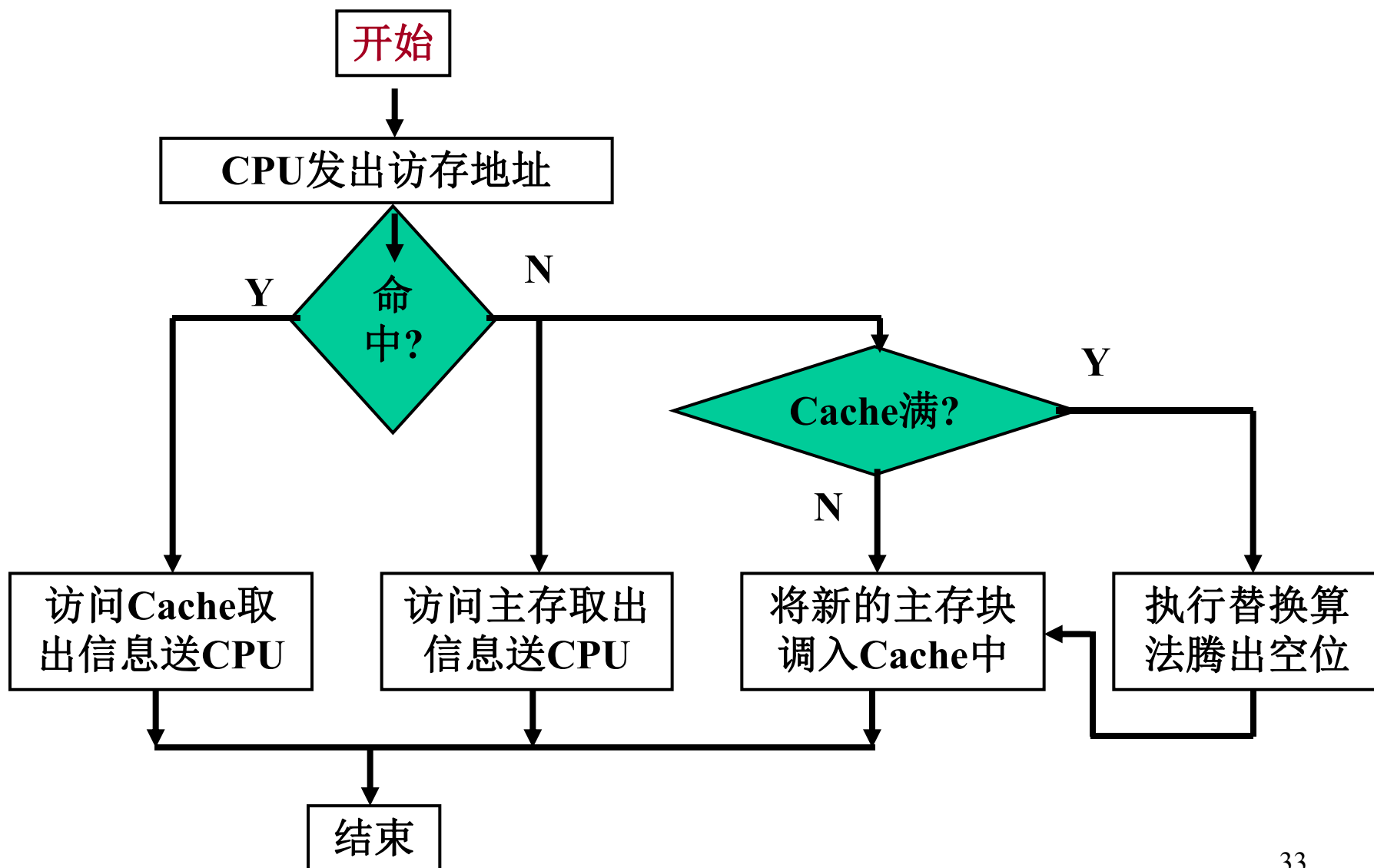
2) Cache效率

Cache的容量和块长是影响Cache效率的重要因素。

Cache容量越大，命中率越高。

- Cache容量达到一定值时，命中率不会因容量的增大而明显提高
- Cache容量大，成本增加

Cache的读数操作流程



■ Cache的写操作

- 防止数据丢失的一致性问题的（命中时）
 - **写直达法（全写）**
每次写入Cache的同时，也写入主存。
 - **写回法**
执行写操作时，信息只写入Cache；
当Cache块被替换时，先将该块内容写回主存，然后再调入新页。
 - **缓冲通写式（写一次）**
信息只写入主存，主存和Cache之间有一个缓冲器
- 不命中：被修改的单元根本不在Cache内，此时写操作只能对主存进行
- 比较
 - 写回法的开销是在块替换时的回写时间，而写直达法则在每次写入时，都要附加一个比写Cache长得多的写主存时间。
 - 写直达法的开销大一些，但其**一致性**保持的要好一些。

■ Cache的写操作

- 防止数据过时的一致性问题的子问题
 - 总线监视
cache控制器监视地址总线，如有写入地址块在cache中，使其无效
 - 硬件监视
外加硬件电路监视已经映像到cache中的主存块。
 - 局部禁止高速缓存法
主存中一块特殊区域，该区域中内容不能取到cache中。
 - Cache清除法
cache已更新数据写回主存时，清除cache中所有数据。

■ 虚拟存储技术

虚拟存储器是指程序使用的逻辑存储空间,它可以比物理存储器大的多,其对应的地址叫**虚拟地址**,也叫**逻辑地址**。

虚拟存储器机制由主存储器、辅助存储器和存储管理部件共同组成。

虚拟存储器的功能与特点

- ①虚拟存储器是“主存——外存”层次;
- ②使计算机的存取容量达到辅存的容量;
- ③使计算机存储速度接近主存的速度;
- ④使计算机整个存储系统的成本接近辅存的成本。

主存外存层次的基本信息传送单位可采用三种不同的方案：**段、页或段页**，这就形成了**页式虚拟存储器、段式虚拟存储器、段页式虚拟存储器**。

虚拟存储器与一般的主存—辅存系统的本质区别

- (1) 虚拟存储器允许人们使用比主存容量大得多的地址空间来访问主存，非虚拟存储器最多只允许人们使用主存的整个空间，一般只允许使用操作系统分配的主存中的某一部分空间。
- (2) 虚拟存储器每次访问主存时必须进行虚、实变换，而非虚拟存储器系统则不必变换。

虚拟存储器与CACHE区别

- (1) 主存/CACHE的访问“时间比”较小，典型的为10: 1，每次传送的页较小；辅存/主存的访问“时间比”较大，典型的为100: 1—1000: 1，每次传送的页较大；
- (2) CACHE未命中期间，处理器不改变任务，仍被等待从内存取数的进程占用；虚存中页故障时可处理其他任务。