Memory

1. 存储器的分类

(1).按照存储器在计算机中的作用划分

主存（Memory）又称为计算机内存，就是冯·诺依曼结构中的存储器，用于存储要执行的程序和处理的数据

辅存（Storage）又称为计算机外存，在现代冯·诺依曼机器中，程序在执行之前是以文件的方式存储在外存中，当要运行某程序时，由操作系统将该程序从外存调入内存中；

高速缓存（Cache）是一种小容量、高速度的存储器，目前，在计算机的主板和CPU中均设置了高速缓存，设置高速缓存的目的是利用程序的局部性原理实现计算机的存储层次，提高CPU的访存速度，以匹配CPU和主存之间在速度上的差异。

1. .按照存储器所使用的物理存储介质划分

半导体存储器；磁介质存储器；光存储器.

1. .按照存储器的读写功能划分

随机存取存储器RAM:可以对存储单元按地址随机存取；

只读存储器ROM:在正常工作条件下，对单元内容只可读不可写。

注:RAM和ROM均属于半导体存储器，是构成计算机主存的主要存储介质，而其中RAM在计算机中用于存储操作系统的常驻内存部分和用户程序，ROM则用于存储操作系统的内核。

1. .按存储器数据的持久化划分

易失性是指写入存储器中的内容在通电情况下能够保存，一掉电则全部丢失；

非易失性则是指写入存储器中的内容在不通电情况下仍然能够保存。

1. 半导体存储器

(1)随机存取存储RAM

按构成其单元电路的不同，又分为静态随机存取存储器（简称SRAM）和动态随机存取存储器（简称DRAM）。这两种存储器相比较而言，SRAM速度更快，但片容量小，价格更贵，因此SRAM主要用作计算机的cache，而DRAM则用作计算机主存。

1. 只读存储ROM

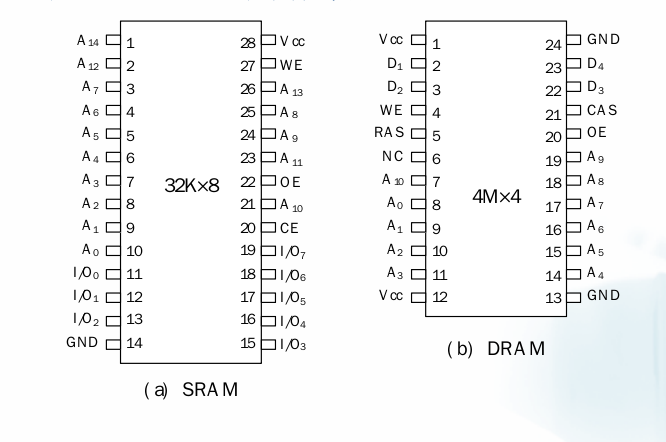
种类较多，主要包括传统的掩膜式ROM、一次可编程式ROM（PROM）、紫外线可擦除可编程式ROM（UVEPROM）和电可擦除可编程式ROM（EEPROM）。前三种ROM在正常工作条件下只可读不可写，是真正意义上的只读存储器，而EEPROM实际上可读可写，但与RAM相比较，最大的不同是：它们均是非易失性的。

1. 半导体存储器的容量

存储器芯片的容量可以表征为以下形式：

容量＝字数×位数

其中，字数表示存储器芯片所具有的字单元数，而位数则表示每一个字单元所具有的位单元数。



1. SRAM

A0~A14:15条地址信号线，用于访问215=32k的字单元，该地址是CPU访存时给出，所以对存储器来说，地址线属于单向输入的。

I/O0~I/07:8条数据信号线，数据信号线为输入输出双向。

CE:片选控制信号线，当CPU访存时，必须在这个引脚上加载一个有效信号。

WE:读写控制信号，在CPU进行读操作，在这个引脚上加载一个高电平信号，在CPU进行写操作，在这个引脚上加载一个低电平信号。

OE:输出允许控制信号，当对芯片进行读操作时，必须在这个引脚上加载一个有效信号。

Vcc和GND:工作电源与接地线。

b)DRAM

A0~A10:11条地址信号线，为实际所需地址信号线的一半。

D1~D4:4条数据信号线。

RAS，CAS:分别用于行，列地址锁存控制。

1. 存储器的容量扩展
2. 位扩展



1)所有芯片的地址线A对应连接在一起；

2)所有芯片的片选信号线CE对应连接在一起；

3)所有芯片的读写WE线对应连接在一起；

4)每个芯片的数据线各自单独引出。

1. 字扩展



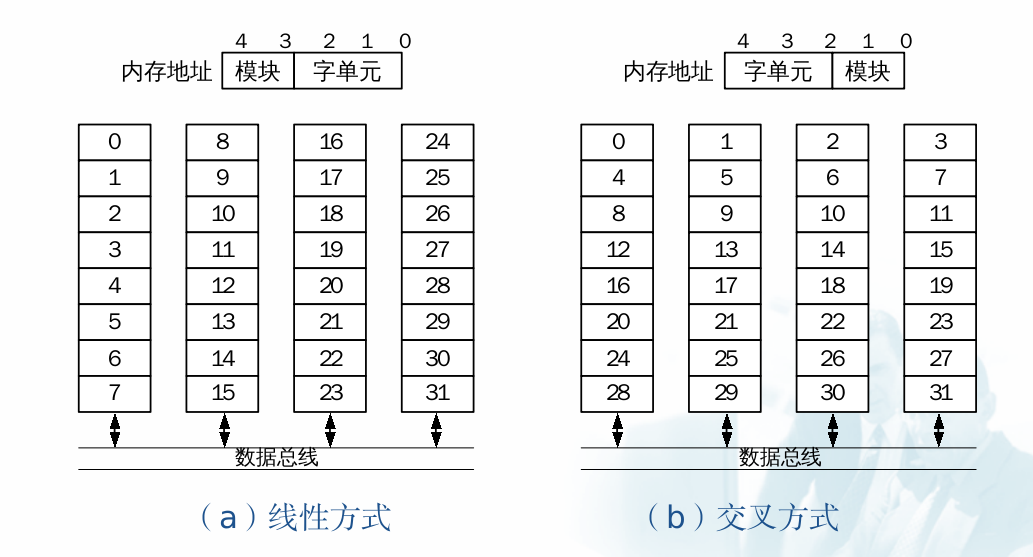
1)所有芯片的地址线A与CPU的低位对应地址线连接在一起；

2)所有芯片的数据线对应连接在一起；

3)所有芯片的读写WE线对应连接在一起；

4)每个芯片的片选信号线CE各自单独引出，并由CPU剩余的部分高位地址线产生.

三.交叉存储技术



1. 在采用线性方式编址的情况下，当CPU 要同时访问多个连续的存储单元时，这些单元分布在同一模块的概率是最大的。例如CPU 在某一访存周期需向存储器读4个字单元的数据，这4个字单元地址是连续的，那么它们只有可能分布在一个或两个存储模块中。由于对于每个存储模块来说，一个访存周期只能读写一个字单元，因此，采用这种线性编址方式并不能使CPU在一个访存周期存取多个字单元。
2. 采用交叉编址方式则不同，当CPU 要同时访问多个连续的存储单元时，由于交叉编址的特殊性，使得这些单元会分布在不同模块。例如CPU 在某一访存周期需向存储器读4个字单元的数据，这4个字单元地址是连续的，那么它们会分别分布在四个存储模块中。由于每个存储模块均能独立进行读写操作，这就使得CPU在一个访存周期能够同时存取多个字单元。

四.高速缓冲存储器

(1)程序的局部性原理:CPU取指令和取数据的操作具有时间上局部分布的倾向，这种现象就称为程序访问的局部性

(2)Cache的基本思想:基于程序的局部性原理，在CPU 与主存之间设置一个小容量的高速缓冲存储器Cache，当一个程序调入主存运行时，将该程序当前要执行的指令及其后将执行的一部分指令同时调入Cache中，CPU每次首先从Cache中取指令执行，根据程序的局部性原理，CPU 大部分情况下可以在Cache中取到指令（称为命中），只要命中率足够的高，就可以使得CPU访存的速度接近于访Cache的速度。

(3)主存块与Cache块之间的映射方式

全相联映射：指主存中任一块都可以映射到Cache中任一块的方式；全相联映射方式的优点是Cache的空间利用率高，但缺点是相联存储器庞大，比较电路复杂，因此只适合于小容量的Cache之用



直接相联映射：指主存的某块j只能映射到满足如下特定关系的Cache块i中：i＝j mod 2C ；直接相联映射方式的优点是比较电路最简单，但缺点是Cache块冲突率较高，从而降低了Cache的利用率



组相联映射：将Cache分成2u组，每组包含2v块。主存的块与Cache的组之间采用直接相联映射，而与组内的各块则采用全相联映射。



(4)替换算法

随机法 ：每次随机选择一个主存块替换出去

先进先出（FIFO）法 每次将最先调入Cache的主存块替换出去。

最不经常使用（LFU）算法 每次将Cache中访问最少的块替换出去。

近期最少使用（LRU）算法 每次将近期最少使用的主存块替换出去。

(5)Cache的写策略

全写法（write-through） 又称写直达法，CPU每次在写Cache命中时，在写Cache的同时，也对相应的主存块进行写入；当写Cache未命中时，则直接写主存。全写法是写Cache和写主存同步进行，其优点是Cache和主存的内容能保持高度的一致性，缺点是Cache对CPU的写操作起不到高速缓存的作用，失去了Cache的功效。

写回法（write-back） CPU每次在写Cache命中时，只写Cache，暂不写主存，只有当某被写命中的块从Cache中替换出去时才写回主存。这种方法使Cache在CPU的写操作中也同样能发挥高速缓存的作用，但却存在主存与Cache内容不一致的隐患。