

存储器是计算机的存储中心,用于存放程序和数据的二进制代码,其性能的优劣对计算机性能的影响极大,容量大、速度快、成本低是计算机对主存储器的基本要求,主存储器一般是 MOS 型半导体存储器。本章介绍存储器的分类、性能指标、存储系统和地址译码等,分析 MOS 型半导体存储器芯片的结构原理与组成逻辑、特性与引脚,讨论动态 MOS 存储器的刷新和主存储器容量扩展的组织实现技术。



5.1 存储器与存储系统的概述

【问题导学】 由第 1 章可知,存储器中存放的大量二进制数是以存储单元为单位一维线性编址来组织的;一个存储单元包含若干位二进制数,它是访问存储器的逻辑单位。而计算机中不仅有存储器,还有存储系统。因此,单就功效难以建立起存储器的概念,还需要进一步认识其性能、特性、分类及其适用性等,才可能在概念上对存储器有个完整清晰的理解。①在计算机中,为什么必须有存储系统?采用什么方法来构建?这样构建的存储系统具有怎样的结构和特征?②存储器与存储系统是不同概念的对象,存储器被存储系统替换后,为什么用户编程感觉不到有什么变化?③存储器的性能指标有哪些?通常可以从哪些角度来对存储器进行分类?各有什么特征、性能和适用性?④主存储器一般是半导体存储器,当外部对主存储器进行访问时,存储单元是怎样选择的?

5.1.1 存储器的访问与性能

1. 存储器的访问

在计算机中,包含多种不同特性的存储器,但无论是哪种存储器都用于存放程序和数据。存储器存放的程序和数据,不是本身所固有的,而是由其他功能部件写入的,同时又提供给其他功能部件所使用。所以,存储器包含写入和读出两种基本操作,且统称为访问(Access)。由于程序和数据均采用二进制编码表示,则写入操作即由其他功能部件把一定数量的二进制数存到存储器中,而读出操作即其他功能部件把一定数量的二进制数从存储器取出来。

一次性同时写入的二进制位数与一次性同时读出的二进制位数是相等的,通常称为访问字长或访问单位,相应的二进制数称为访问字。对于主存储器来说,访问字长一般等于机器字长,但也可以是半字长或双字长。对于主存储器来说,访问字长 $\geq$ 存储字长,即同一存储单元的二进制数是同时访问操作的。

2. 存储器的性能指标

存储器的性能由技术指标来描述,常用于描述存储器性能的技术指标主要包括存储容量、存取速度、存取带宽和存储价格。

① 存储容量。存储容量是指存储器能存放二进制数的总量,是用于反映存储空间大小的技术指标,存储容量越大,存储空间就越大,记忆的二进制数也就越多。存储容量的基本单位可以是位(bit,缩写为 b,即 1 位二进制数),也可以是字节(Byte,缩写为 B,即 8 位二进制数)。常用存储容量的单位还有 K(Kilo)、M(Mega)、G(Giga)、T(Tera),当基本单位为字节时,则有:

$$1\text{TB}=2^{10}\text{GB}=2^{20}\text{MB}=2^{30}\text{KB}=2^{40}\text{B}$$

② 存取速度。存取速度可由存取时间和存取周期两个参数来表示,它取决于存储介质的物理特性和访问机构的类型。

存取时间又叫存储器访问时间(Memory Access Time),它是指启动一次存储器操作(读或写)到完成该操作所需要的全部时间,用 T_A 表示。目前,大多数存储器的存取时间在 ns 级,而存取时间可分为读出时间和写入时间两种,一般来说,读出时间与写入时间是相等的。对于主存储器来说,写入时间是从存储器接收到有效的地址开始,到数据写入被选中单元为止所需的全部时间。读出时间是从存储器接收到有效地址开始,到被选中单元的数据被有效输出为止所需的全部时间。

存取周期(Memory Cycle Time)是指存储器连续进行两次独立的存储器操作(如连续两次读操作)所需的最小间隔时间,用 T_M 来表示。由于存储器的存储介质和有关控制线路都需要有一定的稳定恢复时间,则 $T_M > T_A$ 。现代双极型(TTL)半导体存储器的存取周期接近 10ns, MOS 型半导体存储器的存取周期可达 100ns。

③ 存取带宽。存取带宽又称数据传输速率或频带,是指单位时间可存取二进制数的最大位数,单位为 B/s、b/s 等。在以存储器为中心的具有存储层次的计算机中,主存储器的存取带宽是改善计算机性能的一个关键因素。

④ 存储价格。存储价格是指基本单位容量所需的费用,单位为元/字节、元/位等。设一个存储器的存储容量为 S 字节,购买该存储器的总费用为 C 元,则该存储器的存储价格表示为 S/C (元/字节)。目前,不同介质存储器的存储价格差异很大,可达成千上万倍,且存储速度越快,存储价格越高,这是计算机中必须有存储系统而不能仅有存储器的重要原因之一。

5.1.2 存储器的分类及其结构

从不同的角度出发,可对存储器进行不同的分类。一般可依据存储器的功能作用、特性和存储介质等来分类。

1. 按存储器的功能作用分类

根据存储器在计算机中所起的功能作用不同,CPU 外部的存储器可分为高缓冲存储器、主存储器和辅助存储器三种。由于高缓冲存储器与主存储器在主机的内部,所以统称为内存;而辅助存储器在主机的外部,则称为外存。

① 高速缓冲存储器。高速缓冲存储器简称高速缓存,用于存放 CPU 当前正在使用的程序和数据,即 CPU 对外部的访问绝大多数是访问高速缓冲存储器。显然,高速缓冲

存储器的性能要求为:存取速度一定要快(尽可能与CPU的速度相适应),但存储容量可以较小(能存放CPU当前正在使用的程序和数据即可),存储价格则高。

② 主存储器。主存储器简称主存,用于存放CPU运行期间所需要的程序和数据,即CPU对外部的访问很少是访问主存储器。显然,主存储器的性能要求为:存储容量尽可能大(可存放CPU运行期间所需要的程序和数据),但存储价格不能太高,存取速度适宜即可。主存储器是冯·诺依曼计算机体系结构中传统意义的存储器。

③ 辅助存储器。辅助存储器简称辅存,用于存放CPU运行期间一段时间内暂不需要但以后可能需要的程序和数据以及一些需要永久性保存的信息,即CPU对外部的访问极少是访问辅助存储器。显然,辅助存储器的性能要求为:存储容量一定要大(可存放CPU运行期间需要或不需要的程序和数据),但存储价格尽可能低,存取速度可以慢。

在CPU内部还可能不同功能作用的存储器,如控制存储器与暂存存储器等。控制存储器简称控存,用于存放微程序,仅当采用微程序控制器时,CPU内部才需要配置控制存储器。暂存存储器又称为缓冲栈,用于存放CPU运行期间所产生的非结果性信息,如中间结果、地址映像信息等,目的是减少CPU外部访问;通常由几个或几十个寄存器组成,存取速度可与CPU相匹配,存储容量极小,存储字长与机器字长相等。

2. 按存储器的特性分类

存储器的特性包含访问规则、信息保存、写功能、读破坏四方面,因此又可从这四个角度来对存储器进行分类。

(1) 根据访问规则来分。

访问规则特性是指存储器存储单元之间信息的读写应遵循一定的规则,又称为存取方式。根据存储器的访问规则来分,存储器可分为随机存储器(RAM)、串行存储器和直接存储器三种。随机存储器是指存储器任一存储单元的内容可按需随机读出与写入,且存取时间与存储单元的物理位置无关。串行存储器是指存储器所有存储单元的内容仅能按排列位置顺序读出或写入,且存取时间与存储单元的物理位置有关。直接存储器是指存储器一定范围存储单元的内容需按排列位置顺序读出或写入,各范围之间却可按需随机读出与写入,即兼有随机存储器与串行存储器的访问规则。

(2) 根据信息保存来分。

信息保存特性是指存储器在断电之后,信息是否仍可保持,又称为信息易失性。根据存储器的信息保存来分,存储器可分为非易失性存储器和易失性存储器两种。非易失性存储器是指断电之后信息仍可保持,又称为永久性存储器;而易失性存储器则反之,又称为非永久性存储器。

(3) 根据写功能来分。

写功能特性是指存储器写操作是否存在非在线和写前擦除限制,显然,读操作应该是存储器的常态,不存在任何限制。根据存储器的写功能来分,存储器可分为写常态存储器、只读存储器(ROM)和混合存储器三种。写常态存储器是指可在线写且写前无须擦除,但信息是易失的;只读存储器是指非在线写且写前需擦除,但信息是非易失的;混合存储器是指可在线写且信息是非易失的,写前可能需擦除,也可能无须擦除。

(4) 根据读破坏来分。

读破坏特性是指存储器读操作是否破坏所存储信息,显然,写操作是不可能破坏所存

储信息的。根据存储器的读破坏来分,存储器可分为破坏性存储器和非破坏性存储器两种。破坏性存储器是指对某存储单元进行读操作时,原存储的信息将会被破坏,且把这种读取称为破坏性读取;非破坏性存储器是指对某存储单元进行读操作时,原存储的信息不会被破坏,且把这种读取称为非破坏性读取。特别地,对于破坏性存储器,每次进行读操作之后,必须紧跟一个重写(再生)操作,以便恢复被破坏的信息。

3. 按存储器的存储介质分类

从存储器的功能来看,只要具有两个显著且稳定物理特征状态的物质、元件或器件都可作为存储介质。但从存储器的性能来看,还应具备两个状态相互之间和两个状态与微电信号之间可转换且转换实现容易、速度快的特性。这样的物质、元件或器件作为存储介质,不仅可利用两个状态来记忆“0”“1”,满足存储器的功能要求,还可使存储器具有良好的性能。因此,存储器的存储介质不多。根据存储器的存储介质来分,存储器可分为半导体存储器、磁表面存储器和光盘存储器三种。

① 半导体存储器。采用半导体器件组成实现的存储器称为半导体存储器,它具有容量大、速度快、价格不贵、集成度高、体积小、功耗低、维护简单等特点。目前,半导体存储器主要有两类:双极型和 MOS 型。双极型半导体存储器分为 TTL 型与 ECL 型两种, MOS 型半导体存储器分为写常态存储器、只读存储器和混合存储器三种。写常态存储器又有静态存储器(SRAM)与动态存储器(DRAM)之分。

② 磁表面存储器。采用磁性材料实现的存储器称为磁表面存储器(又称为磁存储器),它具有容量特大、速度慢、价格便宜、体积大、生产自动化程度低等特点。目前,磁表面存储器一般有磁盘存储器、磁带存储器、磁芯存储器三种。

③ 光存储器。光存储器是指应用激光在记录介质上进行读写的存储器,它具有容量很大、速度较慢、价格便宜、记录密度高、耐用性好、可靠性高等特点。光盘可分为只读型、写一次型和磁光盘三种。

4. 存储器的分类结构

不同存储介质的存储器不仅性能存在差异,而且特性也不同,因此,功能适用性也就不一样,存储器的分类结构如表 5-1 所示。

5.1.3 存储系统及其组织结构

1. 存储系统及其特征

衡量存储器性能主要包含速度、容量和价格三个指标,不同存储介质存储器的三个性能指标差别很大。一般说来,速度越快,价格就越高,容量就不能过大;反之,容量要求越大,价格就应越低,速度必越慢。如半导体存储器的速度快,价格就高,容量就不能大;而磁表面存储器的价格低,容量就大,但速度慢。计算机是以存储器为中心的,在程序运行过程中,CPU 所需要的指令、数据要从存储器中读取,CPU 运算出来的数据要写到存储器,输入输出设备也要直接与存储器交换数据,即 CPU 就要求存储器的速度要快。另外,要运行的程序往往大而多,涉及的数据也很多,用户就要求存储器的容量要大。而造价低则是根本。也就是说,计算机对存储器性能的基本要求是速度快、容量大,价格还要低。显然,某种介质的存储器性能与计算机对存储器的性能要求不匹配。

表 5-1 不同存储介质的存储器性能、特性及其功能适用性

介 质 类 型			特 性	性 能	功能适用性
半 导 体 存 储 器	双极型	TTL 型	随机、易失、写常态、非破坏读	容量较大、速度特快、价格很贵	高速缓存
		ECL 型	随机、易失、写常态、非破坏读	容量较大、速度特快、价格很贵	高速缓存
	MOS 型	静态型	随机、易失、写常态、非破坏读	容量较大、速度快、价格贵	高 速 缓 存、 主存
		动态型	随机、易失、写常态、破坏读	容量大、速度较快、价格较贵	主存
		只读型	随机、非易失、非在线写、非破坏读	容量较大、速度较快、价格很贵	主存、控存
		混合型	随机、非易失、在线写、非破坏读	容量较大、速度较快、价格很贵	主存、控存
磁表面 存 储 器	磁盘存储器		直接、非易失、在线写、非破坏读	容量特大、速度慢、价格便宜	辅存
	磁带存储器		串行、非易失、在线写、非破坏读	容量大、速度很慢、价格便宜	辅存
	磁芯存储器		直接、非易失、在线写、非破坏读	容量大、速度慢、价格便宜	辅存
光存 储 器	只读型		直接、非易失、非在线写、非破坏读	容量大、速度慢、价格很便宜	辅存
	写一次型		直接、非易失、非在线写、非破坏读	容量大、速度慢、价格很便宜	辅存
	磁光盘		直接、非易失、在线写、非破坏读	容量大、速度慢、价格便宜	辅存

为使某种介质的存储器性能与计算机对存储器的性能要求相匹配,人们一直在研究如何改进工艺、提高技术,以生产出价格低廉而速度更快的存储器。单就存储器发展趋势而言,存储器的体积越来越小,容量越来越大,速度越来越高,价格越来越低,寿命越来越长。但从比较的角度来看,不同介质存储器的性能仍然存在“速度快的价格就高”的现象。也就是说,单靠材料工艺等技术,使某种介质存储器满足计算机对存储器的性能要求是难以实现的。

人们自然想到,是否可以将各种材料工艺、性能特性不同的多个存储器组织起来,构成一个存储器集合,以实现计算机对存储器的性能要求呢?答案是肯定的。因此,采用硬件、软件或硬软件相结合的方法,将两个或两个以上速度、容量和价格各不相同的存储器有机地连接在一起的存储器集合,称为存储系统。显然,存储系统必须具备以下两个特征。一是对应用程序员和 CPU 是透明的。存储系统虽然包含了多个存储器,但从应用程序员和 CPU 来说,仍然相当于冯·诺依曼体系结构原型中原始意义的一个存储器,应用程序员和 CPU 仍按原有方法访问存储系统,并不需要关注存储系统的存在及其组织实现方法。二是具有速度快、容量大、价格低的性能特点。若存储系统包含了 N 个存储器,那么存储系统的性能表现为:速度近似等于 N 个存储器中存取速度最快的那个存储

器的速度,容量近似等于 N 个存储器中容量最大的那个存储器的容量,价格近似等于 N 个存储器中价格最低的那个存储器的价格。

特别地,存储器与存储系统是两个完全不同的概念,存储器是存储系统组织的基本单元,存储系统是存储器有机整体。如果在一台计算机中只有存储器,即使有多个存储器,但没有存储系统,不同存储器的性能优势不可能得到充分发挥,无法满足计算机对存储器的性能要求。

2. 存储访问局部性原理

程序对存储空间访问具有所谓程序访问局部性的特点,它是存储系统组织的基础。程序访问局部性包括时间局部性和空间局部性两个方面。

时间局部性是指程序在最近的未来要用到的信息很可能是现在正在使用的信息,这主要是由于复杂计算是通过反复进行简单的算术与逻辑运算来完成的,使得复杂计算程序中包含大量需要重复运行的指令序列。

空间局部性是指程序在最近的未来要用到的信息与现在正在使用的信息很可能在存储空间上是相邻或相近的,这主要是由于程序中的指令序列通常是顺序处理和数据集合并不是随机分散地存储,如向量、阵列、树形和表格等数据类型,其中的数据元素在存储空间中一般以连续簇聚的形式存放。因此,程序运行时对存储器访问的地址分布往往是连续的,而不可能是随机的,即簇聚为自然块或页面。

3. 存储系统的一般结构

怎样将多个存储器组织为存储系统呢? 根据程序访问局部性和存储器性能特点,可以采用线性层次结构方法来组织实现存储系统,且速度越快的存储器越靠近 CPU,速度越慢的存储器越远离 CPU,存储系统的层次结构模型如图 5-1 所示。其中, M_1 级靠 CPU 最近,其速度最快,或者说 M_1 的存储器访问周期 T_1 最小,通过 M_1 使存储系统的存储器速度可满足 CPU 对存储器速度的要求。但是, M_1 的价格 C_1 同速度低的存储器相比,就要贵很多,存储容量 S_1 受价格因素限制也只能较小。为使离 CPU 最远的 M_N 满足用户对存储器容量要求, M_N 的容量 S_N 就应该很大,但应具有较低价格 C_N 。受价格因素的限制, M_N 的速度也就相对低,或者说 M_N 的存储访问周期 T_N 最大。也就是说,存储系统的存取周期 $T \sim \min(T_1, T_2, \dots, T_N)$, 存储容量 $S \sim \max(S_1, S_2, \dots, S_N)$, 存储价格 $C \sim \min(C_1, C_2, \dots, C_N)$ 。

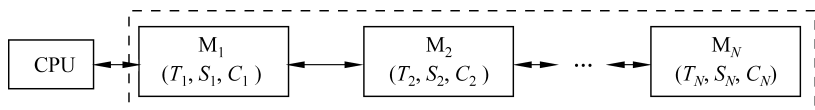


图 5-1 存储系统的层次结构模型

由上述分析可知, M_1 的容量很小,一般不足以存放整个程序和程序需要使用的全部数据;而 M_N 的容量很大,可以存放 CPU 运行期间及其以后需要的程序和数据。但实际上 M_1 只需要存放当前正在使用的程序和程序所需数据中的一部分(可称为块或页)。由于程序访问具有时间局部性,近期未来一段时间内要用的信息(指令和数据)很有可能在 M_1 中,从而使得近期未来一段时间内能在 M_1 中访问到需要使用的信息,或称为能对 M_1 访问命中。若对 M_1 访问未命中,则需要把所需访问的存储字及其与其相邻近的一部分(块或页)内容从 M_2 送至 M_1 ,由于程序访问具有空间局部性,这部分内容所包含的存储

字,通常都是近期未来一段时间内所需要使用的信息。若对 M_2 访问也未命中,则需要把所需访问的存储字及其与其相邻近的一部分内容从 M_3 送至 M_2 ;以此类推, M_3 与 $M_4, \dots, M_{N-1}$ 与 M_N 之间的关系类同。在存储层次中,存储器之间一般满足包容关系,任何一个层次的内容都是其上一层(离 CPU 更远一层)存储器中内容的子集。

由此可见,利用程序访问局部性,只要用户把需要的程序和数据存储在价格低、容量大、速度慢、最外层的存储器 M_N 中,CPU 运行期间所需要的程序和数据,绝大部分都可从价格高、容量小、速度快、最内层的存储器 M_1 中得到,且离 CPU 越远的存储器,CPU 访问的频率越低。因此,存储系统的速度近似 M_1 的速度,价格和容量近似于 M_N 的价格和容量,即存储系统具有速度快、容量大、价格低的性能特点。

4. 存储层次组织的操作

采用线性层次结构的方法将不同类型、性能各异的存储器组织在一起建立起来的存储系统,存储层次间有许多问题需要解决,这些问题应由存储系统的设计人员通过硬件、软件或硬软件相结合的方法来实现,以保证对应用程序员和 CPU 是透明的。综合起来主要有以下四个操作。

① 存储层次间的信息传送。在对 M_i 访问未命中时,并不是仅从其上一层 M_{i+1} 送一个存储字到 M_i 中,而是把包含该存储字所在的一部分存储信息送至 M_i 中。这样,由于程序访问具有空间局部性,近期未来一段时间内所需使用的信息很有可能在该部分中被一起送至 M_i ,从而保证对 M_i 有较高的命中率。高层存储器向低层存储器传送信息的操作是自动实现的。

② 存储层次间的地址变换。在运行程序期间,CPU 形成系列逻辑地址流,逻辑地址要变换为最高层 M_N 的物理地址,从最高层 M_N 开始,逐步转化为 $M_{N-1}, M_{N-2}, \dots$,直到 M_1 ,才能实现 CPU 对 M_1 的访问。存储层次间的地址变换的操作也是自动实现的。特别地,地址变换算法取决于高层存储器 M_{i+1} 的存储空间如何映像到低层存储器 M_i 中,存储层次间的空间映像规则与地址变换算法是一一对应的。

③ 存储空间的替换。在运行程序的过程中,依据存储空间的映像规则,高层存储器 M_{i+1} 不断向低层存储器 M_i 调入信息。由于低层存储器 M_i 比高层存储器 M_{i+1} 要小得多,在某一时刻, M_{i+1} 的某部分存储空间的信息要调入到 M_i 中,但 M_i 中可以映像 M_{i+1} 的该部分存储空间已满,那么就要从可以映像 M_{i+1} 的该部分存储空间的 M_i 存储空间中,找出一部分合理可行的 M_i 存储空间,用现在要调入的信息替换原有的信息,这就需要所谓的替换算法支持,替换算法的操作是自动实现的。

④ 存储层次间信息的一致性维护。从存储层次的组织机制可以看出,当前使用的信息在各层存储器上都有其对应的存储空间,即低层存储器 M_i 所存放的内容是高层存储器 M_{i+1} 部分内容的副本。显然,正本与副本的内容应保持一致,即正本的内容修改了,副本的内容也要修改;反之是一样的。这就是所谓的存储层次间的一致性,存储层次间的一致性的维护操作是自动实现的。



5.1.4 二级结构存储系统及其比较

1. 存储的基本层次

在计算机中,存储的基本层次一般分为七层,如图 5-2 所示,从下到上存储器的容量

越来越小,价格越来越高,存取周期越来越短(即速度越来越快)。高速缓冲存储器(Cache)分为两层,即 CPU 芯片内的和 CPU 芯片外的,很多高性能的 CPU 芯片内集成有 Cache。片内 Cache 称为一级 Cache,容量很小,速度极快,且可分为指令 Cache 和数据 Cache;片外 Cache 称为二级 Cache,在主板,容量较大,速度低于片内 Cache。通用寄存器堆、指令和数据缓冲栈、片内 Cache 在 CPU 芯片内,它们的存取速度都极高。CPU 可直接访问片外高速缓冲存储器与主存储器,它们合称为内存;联机外部存储器与脱机外部存储器 CPU 不能直接访问,它们合称为外存。联机外部存储器俗称硬盘,脱机外部存储器是海量的,主要有软盘、U 盘、光盘、移动硬盘、磁带等。

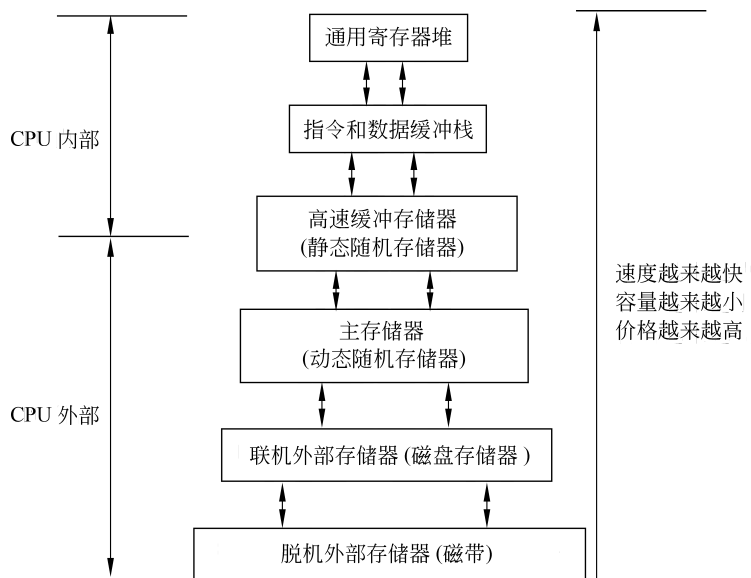


图 5-2 计算机存储的基本层次

2. 三层二级存储系统

从存储层次的组织操作来看,片内高速缓冲存储器及其以下各存储层次的组织操作是由 CPU 来自动实现,即由 CPU 的设计者完成;联机外部存储器与脱机外部存储器存储层次的组织操作是用户借助操作系统非自动实现的;只有片外高速缓冲存储器、主存储器和联机外部存储器(可称为辅助存储器)三个存储层次的组织操作是由存储系统来自动实现的,即由存储系统的设计者完成。因此,由片外高速缓冲存储器、主存储器和辅助存储器构成一个三层二级存储系统,其结构模型如图 5-3 所示,其中把高速缓冲存储器与主存储器级称为高速缓冲存储器,把主存储器和辅助存储器级称为虚拟存储器。

实际上,可以认为三层二级存储系统是主存储器扩展。在性能要求上,主存储器应具有较大的容量,价格不能太高,所以速度难以与 CPU 速度相匹配。为了满足 CPU 对速度的要求,则在主存与 CPU 之间插入一个高速缓冲存储器,它的速度比主存快,容量比主存小得多。另外,较大容量主存储器,容量又与用户要求相距甚远,为了满足用户对容量的要求,又在其外层加入辅助存储器,它的容量比主存大得多,速度比主存慢得多。

3. 二级存储体系的比较

高速缓冲存储器与虚拟存储器是两个不同存储层次的存储体系,在概念定义、组织操

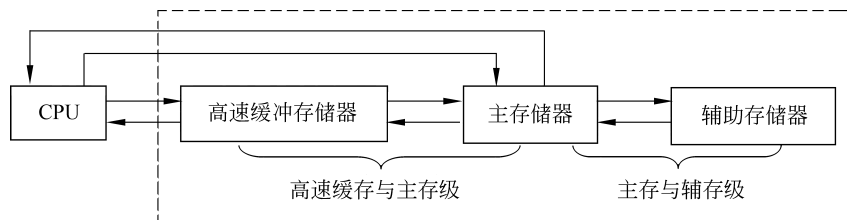


图 5-3 三层二级存储系统的结构模型

作、性能评价等方面具有相似性,但也有许多不同之处。

① 一次交换的信息量不同。若把一次交换的信息量认为是一个单位,高速缓存与主存之间以块为单位,主存与辅存之间以页为单位,块与页是相似的,但页比块大得多。块的大小是以主存储器在一个存取周期内能对主存访问到的信息量为限,一般是在几个到几十个主存存储字之间;而页的大小一般是在 1K 到几 K 主存存储字之间。

② 两层存储器之间的速度比不同。在高速缓冲存储器中,高速缓存的速度是主存的 3~10 倍。在虚拟存储器中,主存的速度一般是辅存的十万倍以上。

③ 访问不命中时,CPU 访问的通路不同。CPU 与高速缓存和主存之间都有直接通路,若 CPU 对高速缓存的块访问未命中,则当前要使用的存储字直接由主存到 CPU,即 CPU 直接访问主存,且从主存调入当前存储字所在的块到高速缓存。CPU 与辅存之间没有直接通路,若 CPU 对主存的页访问未命中,则 CPU 需要等待访问的虚页从辅存调入到主存后,才能再进行访问。

④ 设置的目不同。高速缓冲存储器以提高主存速度为目的,以使存储系统的速度与 CPU 相匹配,即面对 CPU。虚拟存储器以扩大主存容量为目的,以使存储系统的容量满足用户需要,即面对用户。

⑤ 实现方式及其透明性不同。高速缓冲存储器工作时所需要的组织操作功能是由硬件实现的,对系统程序员和应用程序员均透明。虚拟存储器工作时所需要的组织操作功能是以软件为主硬件为辅实现的,仅对应用程序员透明。

5.1.5 半导体存储器芯片的一般结构

1. 记忆元与存储单元

对于半导体存储器芯片来说,通常把存储一位二进制数的元件或器件称为记忆元。半导体存储器芯片一般包含很多记忆元,包含的记忆元越多,存储容量就越大。不同类型的半导体存储器,其记忆元存储一位二进制数的原理及其逻辑电路不同,从而使得性能与特性存在差异。

同样,把包含若干个记忆元的存储器操作单位称为存储单元,显然,一个存储单元可存储若干位二进制数。存储单元是存储器读写操作的逻辑单位,即存储于同一个存储单元的进制数是同时被访问的。可见,半导体存储器芯片的存储单元、存储字数、存储字长、存储字等概念,具有与主存储器相同的意义,其缘由在于主存储器即半导体存储器。但半导体存储器与半导体存储器芯片是不同概念的物理实体,半导体存储器是由一定数量的半导体存储器芯片组成的。因此,半导体存储器芯片的容量也有:

$$\text{存储容量}(\text{b 或 B}) = \text{存储字数 } M \times \text{存储字长 } N(\text{b 或 B})$$

2. 芯片的组成结构

半导体存储器芯片用于存储二进制代码,必须包含一个存储二进制代码的载体,该载体称为存储体。存储体中存储的二进制代码是以存储单元为访问单位的,一个存储单元唯一对应一个单元地址,则需要一个地址译码电路对单元地址译码,以选择对应的存储单元进行访问。存储体中存储的二进制代码是供 CPU 与 I/O 设备使用的,且可读可写入,读与写的信息传送方向相反,则需要一个读写电路以控制信息传送方向。另外,还需要控制逻辑来保证存储体、译码电路、读写电路之间协调有序工作。因此,半导体存储器芯片由存储体、译码电路、读写电路和控制逻辑四部分组成,如图 5-4 所示。

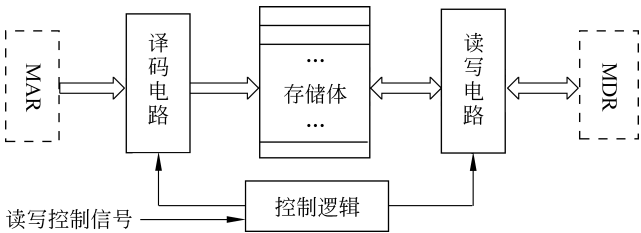


图 5-4 半导体存储器芯片的组成结构

(1) 存储体。

存储体是存储器芯片的主体,用来存储二进制代码。存储体中包含大量记忆元,记忆元在存储体中是按阵列形式排放的,如图 5-5 所示。由于 $P=Q$ 时,可使含有同样数量记忆元的存储体所占面积最小,所以 P 与 Q 一般取相等,但不是一定要相等,也可以不相等。

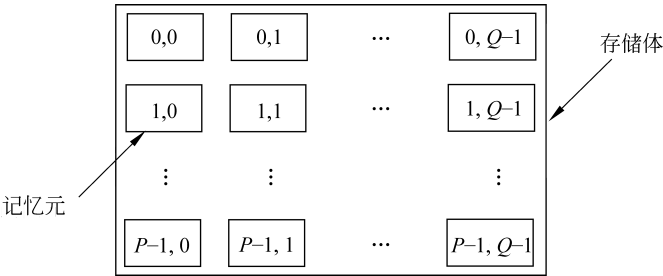


图 5-5 存储体中记忆元的排放形式

(2) 译码电路。

译码电路接收由 CPU 或 I/O 设备送到 MAR 中的 W 位地址信号,经过译码等后,使 2^W 个选择信号中的一个有效,以选中与 W 位地址信号对应的存储单元。特别地,由于译码电路输出的信号在强度上往往难以与存储体相匹配,则译码电路后需要配备一个驱动电路,将译码电路输出信号进行放大,以满足存储体对选择信号的强度要求。因此,译码电路又可称为译码驱动电路。

(3) 读写电路。

读写电路用来控制选中存储单元的读或写(信息传输方向)和信号驱动判别。由于存储体中存储的二进制信号一般很微弱,无法适应长距离传输,因此在读出时要对信号进行放大,即读写电路后需要配备一个放大电路,以满足长距离传输要求。记忆元输出的是电流信号,通过电流信号来判别是“0”还是“1”。特别地,读写电路又称为读写驱动电路,读