

存储器是具有“记忆”功能的部件,存储器包括主存储器(或称内存储器,简称主存或内存)、辅助存储器(或称外存储器,简称辅存或外存)。主存是一组或多组具有数据读写和数据存储功能的集成电路,其作用是存储运行程序和各种所需的数据以及与外部部件交换信息时作为数据缓冲;辅存主要存储大量暂时不用的程序和数据。本章首先介绍了存储器的分类,然后介绍了主存储器的组织,分析了主存容量的扩展,最后介绍了辅助存储器的分类及磁记录存储的原理。

## 5.1 存储器概述



存储器作为计算机必不可少的核心部分,其主要功能是存储程序和数据。冯·诺依曼计算机的基本工作原理是“存储程序控制”,其实现控制的前提就是必须首先存储相应的程序和数据。

在冯·诺依曼计算机中,存储器存放的内容都是二进制表示的,存放一位二进制代码的电路称为存储位元,它是存储器中最小的存储单位。存储位元必须具备的条件:有两个稳定的能量状态,并由一个高能势垒将两个状态分开;借助外部能量能使两个稳态之间进行无限次转换(即可写);借助外部能量能获知其状态(即可读);保存可靠。通常由若干个存储位元组成一个存储单元,进而由大量存储单元可共同组成一个完整的存储阵列。

### 5.1.1 存储器的分类

随着计算机系统结构和存储技术的快速发展,存储器的种类已经逐渐丰富,分类的标准也已经不再单一。

#### 1. 按存储介质分类

按存储介质分类,存储器可分为磁介质存储器、半导体介质存储器和光介质存储器。

##### 1) 磁介质存储器

以磁性材料为存储介质的存储器称为磁介质存储器。这其中又分为磁芯存储器和磁表面介质存储器。磁芯存储器是早期主存储器的主要存储介质,但由于容量小、速度慢、破坏性读出等缺点,早已被淘汰。磁表面介质存储器是在金属或塑料基体上,涂一层磁性材料,用磁层记录信息,常见的有磁盘、磁带等。

##### 2) 半导体介质存储器

以半导体材料为存储介质的存储器称为半导体介质存储器。它采用金属氧化物半导体(Metal Oxide Silicon, MOS)型电路或双极型 TTL(Transistor-Transistor Logic, 晶体管-晶

体管逻辑)电路、射极耦合逻辑(Emitter Coupled Logic, ECL)电路存储信息,常见的有静态随机存取存储器(Static Random Access Memory, SRAM)、动态随机存取存储器(Dynamic Random Access Memory, DRAM)、只读存储器(Read Only Memory, ROM)、闪存(Flash 存储器)等。

### 3) 光介质存储器

光介质存储器利用激光照射使介质发生某种变化存储信息,常见的有只读光盘(CD-ROM)、一次写入式光盘(CD-R)、可改写式光盘(CD-RW)和只读型数字视盘(DVD-ROM)等。

## 2. 按在系统中的作用分类

按在系统中的作用分类,存储器可分为高速缓冲存储器、主存储器、辅助存储器和控制存储器。

### 1) 高速缓冲存储器

高速缓冲存储器(Cache)位于主存储器和 CPU 之间,用来存放主存储器中当前最活跃的程序和数据的副本,以减少对主存储器的访问次数。由于高速缓冲存储器采用高速器件,并且目前已经嵌入到 CPU 内部,它的存取速度与 CPU 的速度基本相当,因此在系统中主要起到提升存储系统速度性能的作用,它的成本较高,通常容量很小。

### 2) 主存储器

主存储器(Main Memory)又称内存(Internal Memory),用来存放计算机运行期间要执行的程序和数据,是 CPU 可以直接访问的主要存储器。它的容量远大于高速缓冲存储器的容量,一般是其十几到几十倍以上。目前主存储器主要由 CMOS(Complementary Metal Oxide Silicon, 互补金属氧化物半导体)集成电路组成。

### 3) 辅助存储器

辅助存储器(Auxiliary Memory)又称外存(External Memory),用来存储当前不参与运行的程序和数据及一些需要长期保存的信息。当需要访问辅助存储器内容时, CPU 要将其中的内容先调入主存储器后,才能使用,它的特点是:比主存储器容量更大,价格更低,但速度也慢很多。常见的辅助存储器有磁盘、磁带和光盘等。

### 4) 控制存储器

控制存储器(Control Memory)只在微程序控制的 CPU 中才会出现,主要用来存放控制信息(即微程序)。目前主要由高速只读存储器构成,一般嵌入 CPU 内部。

## 3. 按存取方式分类

按存取方式分类,存储器可分为随机存取存储器、非易失性存储器、顺序存取存储器和直接存取存储器。

### 1) 随机存取存储器

通常所说的随机存取是指对存储器中的任一存储单元的内容都可以进行访问,且访问的时间相同,与其所处的物理位置无关。随机存取存储器(Random Access Memory, RAM)读写方便,用途广泛,目前被广泛应用在主存储器、高速缓冲存储器中。半导体 RAM 在断电后内容消失。

### 2) 非易失性存储器

非易失性存储器(Non-Volatile Memory)是指掉电后所存储的数据不会消失的存储器。

包括只读存储器、Flash 存储器、铁电存储器和相变存储器等多种存储器。

### 3) 顺序存取存储器

只能按照某种顺序读写存储单元的存储器称为顺序存取存储器 (Serial Access Memory, SAM), 即存取时间的长短和存储单元的物理位置有关, 如磁带就是典型的顺序存储器。在 SAM 中, 一般只能用平均读写时间作为衡量速度的指标。顺序存储器的最大优点是: 不同的存储单元可以共享一套读写线路, 所以结构简单。

### 4) 直接存取存储器

直接存取存储器 (Direct Access Memory, DAM) 最典型的例子就是磁盘, 它的工作过程不同于 RAM 和 SAM。它的寻道过程是直接地定位到某一个磁道, 寻道之后在一圈磁道上则是按顺序方式存取。

## 5.1.2 存储器的性能指标

存储器的性能指标是评价存储器性能优劣的主要依据。一般主要有容量、速度和可靠性等。由于存储器种类很多, 工作原理不尽相同, 性能指标的具体内容也有相当大的差异, 这里以主存储器为例来分析存储器的性能指标。

### 1. 容量

一个存储器所能容纳的全部信息量称为**存储容量**, 以位 (b) 或字节 (B) 为单位, 如 64KB、32MB 等, 也经常用存储单元数乘以存储单元字长的形式表示, 例如  $64K \times 8$ ,  $16M \times 16$  等。在表示容量大小时, 经常用 K、M、G、T、P 等表示数量级, 它们的换算关系为  $1K = 2^{10}$ ,  $1M = 2^{20}$ ,  $1G = 2^{30}$ ,  $1T = 2^{40}$ ,  $1P = 2^{50}$ 。

### 2. 速度

一般用来描述主存储器工作速度的指标有存取时间  $T_a$ 、存储周期  $T_m$  和存储带宽  $B_m$ 。

#### 1) 存取时间 $T_a$

**存取时间**又称为存储器访问时间或读写时间, 它通常是指从启动一次主存储器读写操作到完成该操作为止所需要的时间。以读出为例, 从 CPU 发出对主存储器的读命令和有效地址开始, 到被选中单元的内容读出到数据总线上为止所用的时间为存取时间。

一般存取时间取决于存储介质的物理特性及所使用的读写机构的特性。

#### 2) 存储周期 $T_m$

**存储周期**是指存储器进行两次连续的存储操作之间所需的时间间隔, 又称读写周期或访存周期。存储周期比存取时间要长, 即  $T_m > T_a$ , 因为任何一种存储器在读写操作结束后总要有一段恢复内部状态的复原时间。

一般存储周期取决于主存储器的具体结构和其工作机制。

#### 3) 存储带宽 $B_m$

**存储带宽**又称数据传输率或数据吞吐率, 它是指在单位时间内存储器所存取的信息量, 单位通常是 B/s。由于存储器不仅要为 CPU 服务, 还要为其他外设提供服务, 因此存储带宽是计算机系统性能的主要指标之一。

### 3. 可靠性

可靠性是指存储器连续正常工作的能力。通常用**平均故障间隔时间** (Mean Time Between Failures, MTBF) 作为其衡量指标。MTBF 越长, 则该存储器的可靠性越高。主存

存储器常采用纠错编码技术延长 MTBF 以提高其可靠性。

## 5.2 半导体随机存取存储器

目前在主存储器中使用最广泛的就是半导体存储器,包括随机存取存储器(RAM)和只读存储器(ROM)。RAM用于存放当前运行的程序和数据,ROM用于存放那些不需要修改的由生产厂商提供的特殊程序,诸如引导程序、BIOS(Basic Input Output System,基本输入输出系统)。

RAM根据信息存储的原理不同,可以分为静态随机存取存储器(SRAM)和动态随机存取存储器(DRAM)。



### 5.2.1 静态随机存取存储位元

SRAM的信息存储原理是利用双稳态触发器(Bistable Flip-Flop)来存储信息。常用的有TTL、ECL和CMOS电路,其中CMOS电路是目前最常用的器件,TTL和ECL电路由于功耗大、集成度低、性价比不高等原因已经逐渐被淘汰。这里以CMOS电路作为主要对象来分析其结构及原理。

图5.1表示了SRAM的存储位元结构图。其中 $T_0$ 、 $T_1$ 管反向耦合构成存储信息的触发器, $T_4$ 、 $T_5$ 分别是 $T_0$ 和 $T_1$ 的负载管(相当于负载电阻), $T_2$ 、 $T_3$ 构成A、B两点的门控电路,控制读写操作。字线W用来选择相应存储位元,两条位线D和 $\bar{D}$ 用来进行读写信息传输。

SRAM的基本存储原理是利用两个栅极反向交叉耦合的CMOS管构成一个触发器实现信息的存储。按照存储信息的必要条件,它应符合几个条件:

(1) 两种稳态。

假设 $T_0$ 管截止、 $T_1$ 管导通表示该存储位元存放内容为1,由于 $T_0$ 管截止,A点电位为高, $T_1$ 管导通,B点电位接地为低。 $T_0$ 管导通, $T_1$ 管截止表示该存储位元存放内容为0,由于 $T_1$ 管截止,B点电位为高, $T_0$ 管导通,A点电位接地为低,因此在不断电的情况下,A、B两点的电位值始终相反。

(2) 读出操作。

在字线W上加高电位的字脉冲,使得 $T_2$ 、 $T_3$ 门控管导通,于是位线D和 $\bar{D}$ 分别与A、B两点连接。若该存储位元原存放内容为0,则A点为低电位,B点为高电位,D输出低电平, $\bar{D}$ 输出高电平,经差动放大器检测出0信号。若该存储位元原存放内容为1,则B点为低电位,A点为高电位, $\bar{D}$ 输出低电平,D输出高电平,经差动放大器检测出1信号。

由于读出前后A、B两点电位没有明显变化,读出信息后不影响原来状态,因此属于非破坏性读出。

(3) 写入操作。

在字线W上加一个高电位的字脉冲,使 $T_2$ 、 $T_3$ 管导通。如果需要写入1,则需要从位

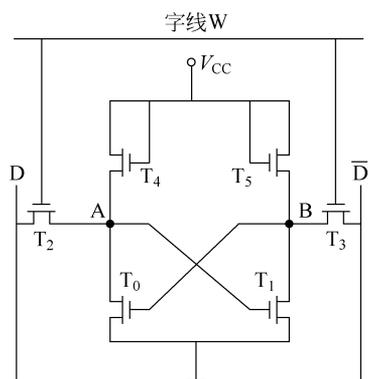


图5.1 SRAM的存储位元结构图

线 D 输入高电位,  $\bar{D}$  输入低电位, 经过导通的  $T_2$ 、 $T_3$  管, 将会使得 A 点电位上升, B 点电位下降,  $T_1$  管导通,  $T_0$  管截止, 该存储位元内容变为 1, 完成写入 1 操作; 如果要写入 0, 则需要从位线 D 输入低电位,  $\bar{D}$  输入高电位, 经过导通的  $T_2$ 、 $T_3$  管, 将会使得 B 点电位上升, A 点电位下降,  $T_0$  管导通,  $T_1$  管截止, 该存储位元内容变为 0, 完成写入 0 操作。

(4) 保持操作。

当字线 W 为低电平时, 该存储位元未被选中,  $T_2$ 、 $T_3$  管截止, 触发器与位线断开, 只要  $V_{CC}$  不断电, 原有信息将保持不变。但如果断电, 则原来的信息必定丢失。

SRAM 具有存储速度快, 驱动能力强的优点, 但是相对于 DRAM, 其集成度仍然较低, 功耗较大, 单位价格较高, 所以一般用来组成高速缓冲存储器和小容量主存储器系统。

## 5.2.2 动态随机存取存储位元



DRAM 的结构有几种, 现在比较普遍应用的是单管 DRAM 存储位元, 由于其具有结构简单、集成度高、功耗小等优点, 因此在当今存储容量需求日益增大的趋势之下, 得到了最广泛的应用。这里以单管 DRAM 存储位元为例分析 DRAM 的结构、组成以及工作原理。

### 1. DRAM 的存储原理

不同于 SRAM, DRAM 的基本工作原理是利用电容存储电荷来存储信息的, 其结构图如图 5.2 所示。DRAM 存储位元由一个 MOS 管 T 和一个存储电容 C 构成, T 管作为门控管, 其栅极由字线 W 控制, 源极和漏极分别连接电容 C 和位线 D。

同样, 按照存储信息的必要条件, 它应符合几个条件:

(1) 两种稳态。

假设电容 C 上有电荷为 1, 则无电荷为 0。

(2) 读出操作。

字线 W 为高电平, 使得 T 管导通, 电容 C 和位线相连, 位线 D 处于中间电位。若原来存放内容为 1, 即电容 C 有电荷, 则位线 D 上电位进一步升高, 产生读电流, 电容 C 放电。若原来存放内容为 0, 即电容 C 无电荷, 则位线 D 上电位下降, 电容 C 充电。由于读出后电容 C 上的原有电荷必定遭到破坏, 因此属于破坏性读出。由于信息在读出以后被破坏, 为了修正这种错误, 读出的信息将会在读出以后被重新写入对应存储单元, 这种情况一般被称为再生 (又称重写), DRAM 必须在读出操作之后根据读出的内容立即进行再生的操作。

(3) 写入操作。

字线 W 为高电平, 使得 T 管导通, 电容 C 和位线相连, 若写入 1, 位线 D 为高电平, 电容 C 被充电; 若写入 0, 位线 D 为低电平, 电容 C 被放电。

(4) 保持操作。

字线 W 为低电位, T 管截止, 电容 C 和位线 D 分隔开, 电容 C 本应保持信息, 但由于漏电流的存在, C 上的电荷会不断泄漏。一般为了保证功耗和集成度, 选用的电容 C 都很小, 电荷存储有限, 所以只能保持几毫秒的时间。

### 2. 刷新

为了维持 DRAM 存储位元的原有存储信息, 必须每间隔一定时间对 DRAM 进行一次读出再写入, 这就是刷新。刷新间隔时间  $\Delta t$  由栅极电容上电荷的泄放速度来决定的, 可以

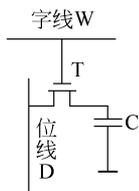


图 5.2 DRAM 基本存储位元结构

由公式  $\Delta t = C \cdot \Delta V / I$  得到。其中  $C$  为电容容量,  $\Delta V$  为允许的电容电压变化,  $I$  为漏电流。一般选定的最大刷新时间间隔为 2ms 或 4ms 甚至更大, 即在 2ms 或 4ms 的时间间隔内, 必须将全部存储体刷新一遍。

刷新和之前提到的再生相似, 但是有区别。即再生是在每次读出操作之后进行的, 所对应的单元是随机且唯一的, 是在被选中单元破坏性读出后为了防止信息丢失而进行的操作。而刷新是定时的, 即使存储单元没有被访问到也会进行刷新, 而且刷新是以存储体阵列的行作为单位, 即每次刷新都是自动刷新存储阵列中对应的行的所有存储位元, 因此只需要给出行地址即可。

常见的刷新方式有集中式、分散式和异步式 3 种。

#### 1) 集中刷新方式

在规定的最大刷新闻隔时间(例如 2ms)内, 将该存储芯片所有行的刷新操作都集中在一起, 在刷新时停止所有设备对存储器的读写操作。由于刷新操作和存储器的读操作基本一致, 因此刷新一行所需要的时间近似等于存储周期。

例如, 对一个  $4K \times 1$  位的存储芯片(其存储阵列为  $64 \times 64$  的矩阵)进行刷新, 由于刷新是按行进行的, 每次刷新一行需要占用一个存储周期, 所以共需要 64 个存储周期来完成对该芯片所有单元的刷新。假设存储周期  $T_m$  为 200ns, 则在 2ms 的时间间隔内一共可以完成 10 000 次存储操作, 即可以分配 10 000 个存储周期, 其中第 0~9935 个存储周期可以进行读写操作或者保持操作, 其后的第 9936~9999 个(共 64 个)存储周期集中安排刷新操作, 如图 5.3 所示。



图 5.3 集中刷新方式示意图

集中刷新的优点是读写操作时没有刷新操作的影响, 可以连续进行, 控制简单, 速度较快; 其最大的缺点是随着存储阵列行数的增加, 其刷新时间( $T_r = T_m \times$  存储阵列的行数)也会随之增大, 在这段时间内将无法访问存储器, 故被称为“死区”。

#### 2) 分散刷新方式

分散刷新方式首先定义系统的存储周期长度是存储器本身的存储周期长度的两倍, 每个系统存储周期可以分为两个阶段, 前一个阶段进行读写操作或保持操作, 后一个阶段进行刷新操作, 每次刷新一行。

分散刷新方式的优点是存储器没有连续的刷新操作, 因此不存在时间上明显的“死区”。但是其缺点也非常明显, 首先, 其存储周期长度变为原先的两倍, 降低了整机的速度; 其次, 随着存储器速度的不断提高, 其存储周期长度显著缩短, 则在 2ms 的时间间隔内, 分配的存储周期过多, 远超其存储阵列的行数, 存在时间上的浪费。

例如, 对一个  $4K \times 1$  位的存储芯片(其存储阵列为  $64 \times 64$  的矩阵)进行刷新, 由于刷新是按行进行的, 每次刷新一行需要占用一个存储周期, 因此共需要 64 个存储周期来完成对该芯片所有单元的刷新。假设存储器自身的存储周期  $T_m$  为 200ns, 则系统定义的存储周期长度为  $200\text{ns} \times 2 = 400\text{ns}$ , 在 2ms 的时间间隔内一共可以完成 5000 次存储操作和 5000 次

刷新操作(见图 5.4),而真正需要的刷新操作只有 64 次,造成了大量的无意义的刷新操作。

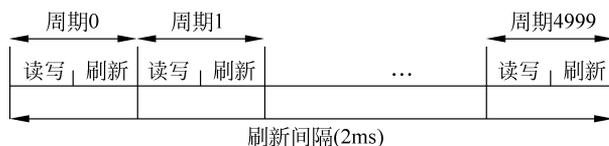


图 5.4 分散刷新方式示意图

### 3) 异步刷新方式

异步刷新方式是上述两种方式的折中。它充分利用了最大刷新时间间隔,把该存储芯片所有行的刷新操作都平均分配到整个最大刷新时间间隔内进行,所以,相邻两行的刷新时间间隔=最大刷新时间间隔÷存储阵列行数。

仍以上面给出的例子为例,相邻两行的刷新闻隔时间=2ms/64=31.25μs,由于存储周期为 200ns,取存储周期的整数倍 31.2μs,即每隔 31.2μs 的时间间隔安排一次刷新操作,并在刷新时停止对存储器的访问操作(见图 5.5)。

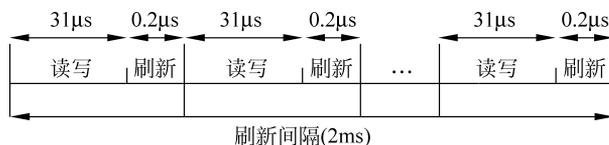


图 5.5 异步刷新方式示意图

相对于集中刷新方式而言,异步刷新方式的“死区”要小得多,仅有 0.2μs,这样可以有效避免 CPU 长时间等待,并减少了刷新次数,是一种比较常用的刷新方式。

## 5.2.3 半导体随机存取存储芯片



半导体存储芯片采用超大规模集成电路制造工艺,在一个半导体硅晶片上集成,用来存储信息的存储阵列、地址译码驱动电路和读写驱动电路等,其主要核心部分是由存储位元构成的存储阵列。

### 1. 地址译码驱动方式

存储芯片的地址译码能把接收到的地址信号翻译成对相应存储单元的选择信号。一般有一维译码和二维译码两种方式。

#### 1) 一维译码方式

一维译码方式把存储容量为  $W \cdot d$  (即存储字数量×存储字字长)的存储器分成  $W$  行,每行  $d$  位,存储阵列的每一行对应一个存储字,共用一根字选择线  $W_i$ ; 每一列则对应不同存储字的相同位,有两根共用的位线  $D_j, \bar{D}_j$  与之连接,因为是一维,所以只用一组地址译码器对  $W$  个字进行一维译码。图 5.6 是一个  $32 \times 8$  位,采用一维地址译码方式的存储器结构示意图。

如图 5.6 所示,其地址引脚(即地址线)数量为  $\lg 32 = 5$  个,数据引脚(即数据线)数量为 8 个,5 根地址线经过一组 5~32 译码器产生译码信号字线  $W_i$ ,选中在同一行(即一个存储字)中的所有存储位元的  $\overline{CS}$  片选端,使之同时工作,根据读写控制电路的读写命令,对被选中字的所有位同时进行读出或写入的操作。由于每次读出或写入的单位都是一个存储字,

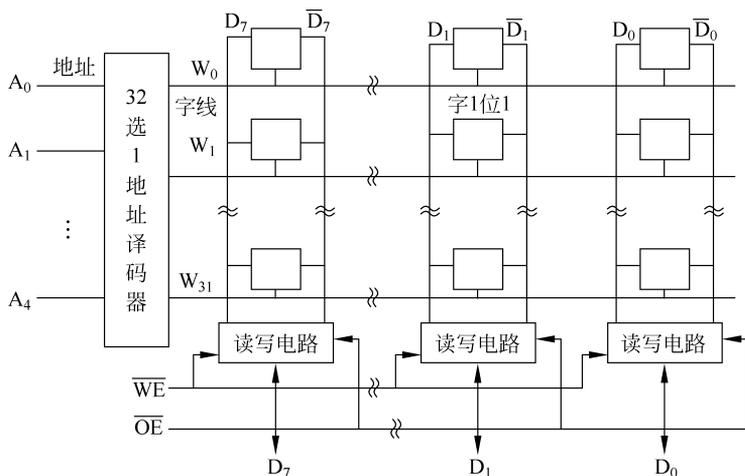


图 5.6 一维地址译码方式结构图

因此又称为字片式存储芯片。

一维译码方式的优点是结构简单,易于实现,缺点是当存储字数较多,由于地址线数目增加而导致译码器复杂性按指数规律增加,导致无法实现或延时过多,因此一般只适合一些容量较小的存储器。

2) 二维译码方式

(1) 二维译码方式的位选方式

二维译码方式是把地址线分为数量接近相等的两组,分别用两个译码器进行译码,一部分用作水平的行方向  $X_i$  的译码,另一部分用作垂直的列方向  $Y_j$  的译码。其具体结构图如图 5.7 所示,行选择信号  $X_i$  选中第  $i$  行的 32 个存储位元,使 32 个存储位元内部的数据分别与各自的位线连通,列选择信号  $Y_j$  选中第  $j$  列的位线控制端,使第  $j$  列控制门打开,故处

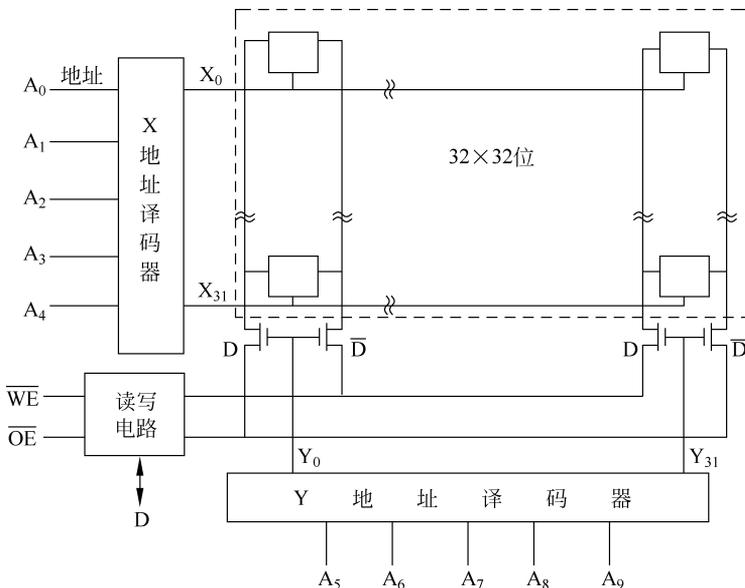


图 5.7 二维地址译码方式结构图

于 $(X_i, Y_j)$ 位置的存储位元被选中。根据读写控制电路的读写命令,对被选中的存储位元进行读出或写入的操作。由于每次读出或写入的单位都是一个位,所以又称为位片式存储芯片。

位片式存储芯片的容量假设为 $M \times 1$ ,例如 $1024 \times 1$ ,为了使外围电路需求量最小,线路走线最短,传输延时最少,通常要求两组用于译码的地址线数量应基本相当,即行方向译码地址线数 $N_x =$ 列方向译码地址线数 $N_y = (\lg 1024)/2 = 5$ ,需要2个5~32的译码器,构成一个 $2^5 \times 2^5 = 32 \times 32$ 的存储阵列。由于每个存储芯片只能提供一位数据,因此一般用8(或16)片同样的位片式存储芯片并联构成一个 $M \times 8$ (或 $M \times 16$ )的存储器,每个存储芯片分别提供地址相同的同一存储字的不同二进制位。

二维译码方式的优点是地址译码器在地址位数较多时,其结构相对字片式要简单很多,例如容量为 $1024 \times 1$ 的存储器如果用字片式芯片实现,其结构为 $1024 \times 1$ 的一个存储阵列,需要的译码器地址线位数为 $\lg 1024 = 10$ ,是一个10~1024的译码器,其结构要比5~32的译码器复杂很多。

### (2) 二维译码方式的字选方式

这种方式是将一维译码方式和二维译码方式的位选方式结合到一起。当实现存储容量为 $W \cdot d$ 位的存储器时,和一维译码不同,二维阵列中每行分配的不是一个 $d$ 位长度的存储字,而是 $s$ 个 $d$ 位长度的存储字,则行译码线(即行数)变为 $W/s$ 条,列译码线为 $s$ 条;每一条列译码线对应一个 $d$ 位字长的存储字的所有位,与一维译码方式中的字选择线类似。此时 $N$ 根地址线分为两个部分,行地址译码部分线数为 $n_x = \lg(W/s)$ ,列地址译码部分线数为 $n_y = \lg s$ , $N = n_x + n_y$ 。

如图5.8所示,容量为 $512 \times 8$ 的存储芯片,其地址线为 $\lg 512 = 9$ 根,如果采用二维译码方式的字选方式,共有512个存储字,4096个存储位元,其存储阵列为 $64 \times 64$ ,6位地址线经过行方向 $X$ 地址译码器产生64根行译码线,3根地址线经过列方向 $Y$ 地址译码器产生8根列译码线,每根列译码线选中一个8位字长的存储字。

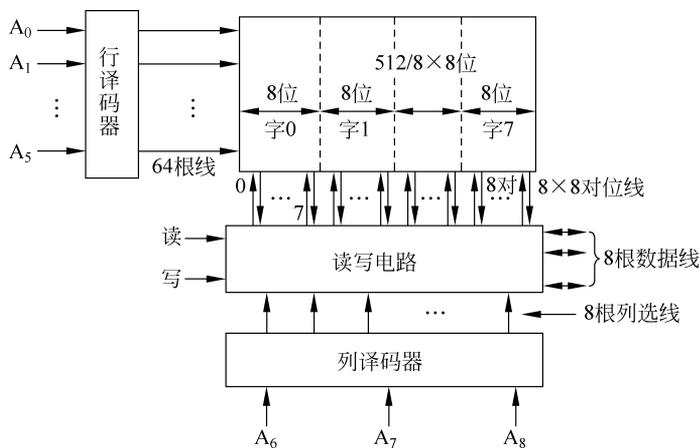


图 5.8 二维译码字选方式结构图

这种方式既可以解决一维译码方式下译码逻辑复杂、容量受到限制的问题,又可以克服二维译码方式的位选方式中存储字的不同位存放在不同芯片的问题。

## 2. 半导体静态随机存取存储器

### 1) SRAM 芯片举例

典型的 SRAM 芯片有 2114(1K×4)、2142(1K×4)、6116(2K×8)、6232(4K×8)、6264(8K×8)和 62256(32K×8)等。这里以 2114 为例,芯片的引脚图和逻辑符号如图 5.9(a)和图 5.9(b)所示。

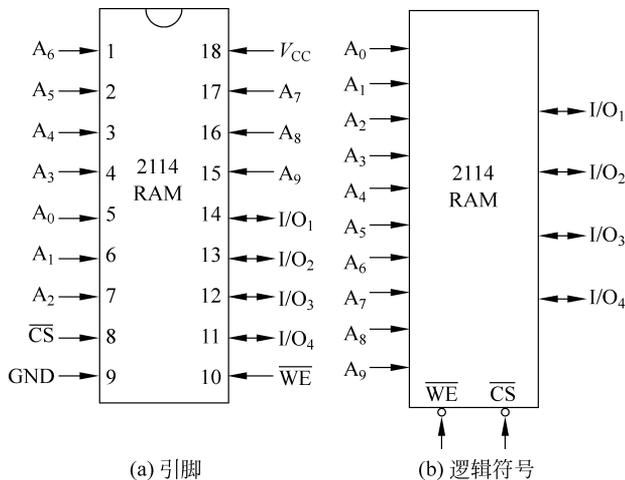


图 5.9 2114 引脚图和逻辑符号图

通常半导体 SRAM 存储芯片的外部引脚有以下几种:

(1) 地址线  $A_i$ 。

方向为单向输入,通常连接地址总线,其位数与芯片的存储单元数量有关。2114 容量为  $1024 \times 4$  位,其地址线位数为  $\lg 1024 = 10$ ,记为  $A_0 \sim A_9$ 。

(2) 数据线  $D_i$ 。

方向为双向,既可输入,又可输出,通常连接数据总线,其位数与芯片存储单元的字长有关。容量为  $256 \times 4$  时,其数据线位数为 4 位;容量为  $32K \times 8$  时,其数据线位数为 8 位。

(3) 片选端。

芯片的片选端通常用符号  $\overline{CS}$ (Chip Selected)或  $\overline{CE}$ (Chip Enable)表示,用来确定该芯片是否被选中工作,通常由译码信号控制,低电平有效。

(4) 读写控制线。

芯片的读写控制线是单向输入的,连接控制总线。有些 SRAM 有两根读写控制线:读允许线  $\overline{OE}$ (Output Enable)或  $\overline{RD}$ (Read)和写允许线  $\overline{WE}$ (Write Enable)或  $\overline{WR}$ (Write),低电平有效。有些 SRAM 只有一个读写控制线  $\overline{WE}$  或  $\overline{WR}$ , $\overline{WE}$  或  $\overline{WR} = 0$  时,写允许; $\overline{WE}$  或  $\overline{WR} = 1$  时,读允许。

(5) 电源线  $V_{CC}$  和地 GND, $V_{CC}$  为 +5V,接电源;GND 接地。

2114 的内部结构如图 5.10 所示。由于 2114 的容量为  $1024 \times 4$  位,故有 4096 个存储单元,排成  $64 \times 64$  的矩阵。用  $A_3 \sim A_8$  共 6 根地址线作为行译码,产生 64 根行选择线,用  $A_0 \sim A_2$  与  $A_9$  共 4 根地址线作为列译码,产生 16 根列选择线,而每根列选择线控制一组 4 位同时进行读或写操作。存储器内部有 4 路 I/O 电路以及 4 路输入输出三态门电路,并

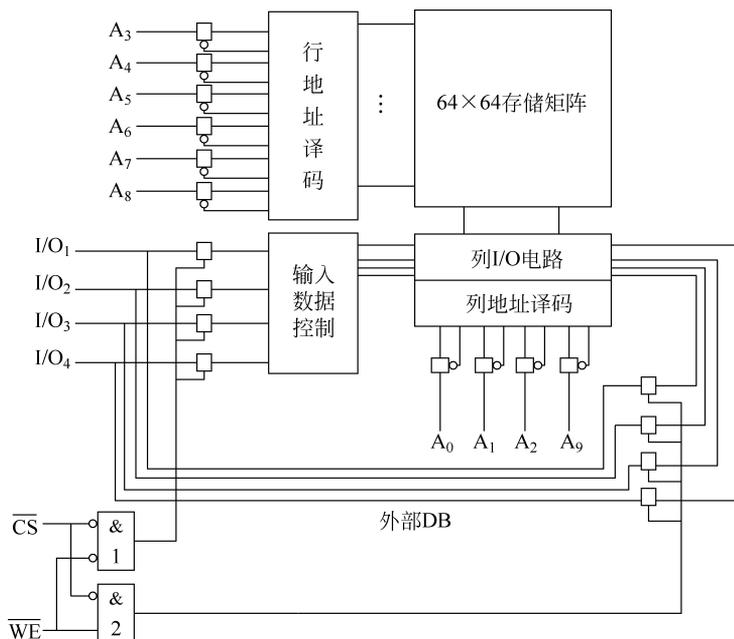


图 5.10 2114 的内部结构示意图

由 4 根双向数据线  $I/O_1 \sim I/O_4$  引出与外部数据总线 DB 相连。当  $\overline{CS}=0$  与  $\overline{WE}=0$  时,经门 1 输出线的高电平将输入数据控制线上的 4 个三态门打开,使数据写入;当  $\overline{CS}=0$  与  $\overline{WE}=1$  时,经门 2 输出的高电平将输出数据控制线上的 4 个三态门打开,使数据读出。

## 2) SRAM 芯片的时序

图 5.11 为 SRAM 的工作时序图,其中图 5.11(a)为读周期时序,图 5.11(b)为写周期时序。在读周期当中,依次完成以下操作:首先,地址线有效,芯片外部输入有效地址信息(Effective Address, EA),为保证地址译码及读出时信息稳定,EA 在读周期中应保持不变;其次,为了使芯片能正常工作,片选信号  $\overline{CS}$  在地址有效之后也必须有效;再次,由于是读出操作,读出使能信号  $\overline{OE}$  也必须有效;最后,直到数据总线上出现稳定的有效读出数据,才可以撤销  $\overline{CS}$  和  $\overline{OE}$  以及地址信息。从给出地址信息到撤销地址信息的这段时间间隔  $T_{RC}$  即读周期。

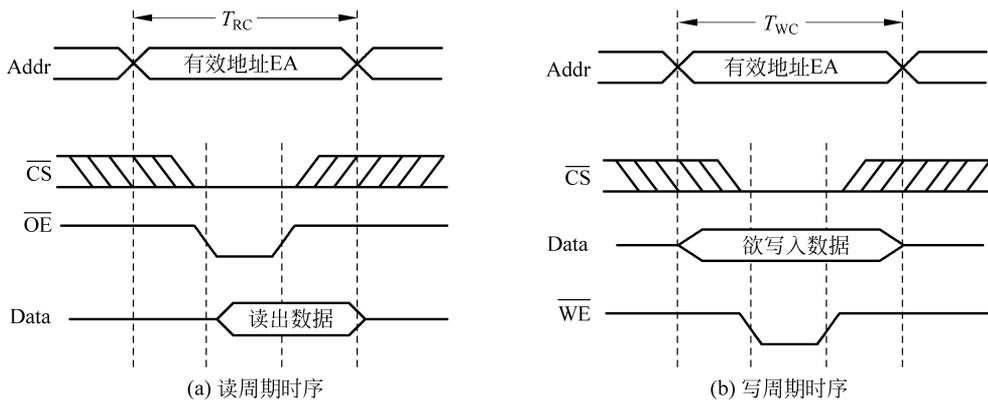


图 5.11 SRAM 的工作时序图

在写周期当中,依次完成以下操作:首先,地址线有效,芯片外部输入有效地址信息;其次,由于是写入操作,必须先提供欲写入的数据,因此在数据总线上由外部提供有效的欲写入数据信息;再次,为了使芯片能正常写入,片选信号  $\overline{CS}$  和写入使能信号  $\overline{WE}$  也必须有效;最后,直到数据有效写入,才可以撤销  $\overline{CS}$  和  $\overline{WE}$  以及数据和地址信息。从给出地址信息到撤销地址信息的这段时间间隔  $T_{WC}$  即写周期,其时间间隔与  $T_{RC}$  相当。

### 3. 半导体动态随机存取存储器

#### 1) DRAM 芯片举例

这里以 DRAM 2164A 为例,2164A 是  $64K \times 1$  位的芯片,其引脚排列如图 5.12 所示。

2164A 的内部逻辑结构如图 5.13 所示。

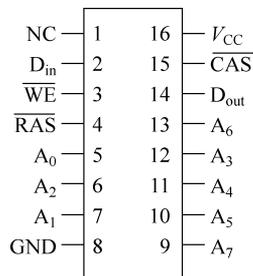


图 5.12 2164A 芯片引脚图

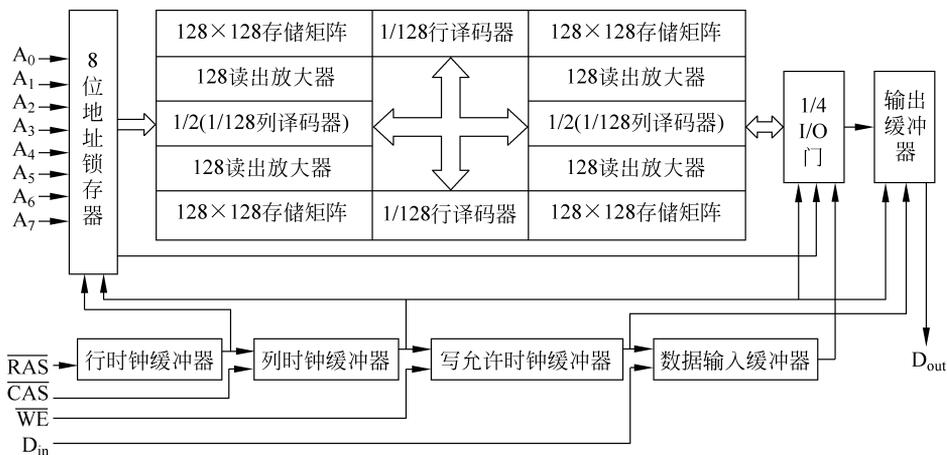


图 5.13 2164A 的内部逻辑结构

由图 5.13 可见,2164A 的片内有  $64K$  个存储单元,每个存储单元存储一位数据,片内要寻址  $64K$  个单元,需要 16 条地址线,为了减少封装引脚,芯片的地址引脚只有 8 个,片内有行地址锁存器和列地址锁存器,可利用外接多路开关,由行地址选通信号  $\overline{RAS}$  (Row Address Strobe) 将先送入的 8 位行地址送到片内行地址锁存器,然后由列地址选通信号  $\overline{CAS}$  (Column Address Strobe) 将后送入的 8 位列地址送到片内列地址锁存器。

2164A 芯片中的  $64K \times 1$  存储体由 4 个  $128 \times 128$  的存储阵列组成,每个  $128 \times 128$  的存储阵列,由 7 条行地址线和 7 条列地址线进行选择。7 位行地址经过译码产生 128 条选择线,分别选择 128 行中的一行;7 位列地址经过译码产生 128 条选择线,分别选择 128 列中的一列。7 位行地址  $RA_0 \sim RA_6$  (即地址总线的  $A_0 \sim A_6$ ) 和 7 位列地址  $CA_0 \sim CA_6$  (即地址总线的  $A_8 \sim A_{14}$ ) 可同时选中 4 个存储阵列中各一个存储单元,然后由  $RA_7$  与  $CA_7$  (即地址总线中的  $A_7$  和  $A_{15}$ ) 经  $1/4$  的 I/O 门电路选中 1 个单元进行读写。而刷新时,只送入 7 位行地址同时选中 4 个存储阵列的同一行,即对  $4 \times 128 = 512$  个存储单元进行刷新。

2164A 的数据线是输入和输出分开的,由  $\overline{WE}$  控制读写。当  $\overline{WE}$  为高电平时,为读出,所选中单元的内容经过输出三态缓冲器,从  $D_{out}$  引脚读出;当  $\overline{WE}$  为低电平时,为写入, $D_{in}$  引脚上的内容经过输入三态缓冲器,对选中单元进行写入。

DRAM 芯片的引脚通常无专门的片选信号  $\overline{CS}$  或  $\overline{CE}$ ，一般行选通信号  $\overline{RAS}$  和列地址选通信号  $\overline{CAS}$  也可起到片选的作用。

## 2) DRAM 芯片的时序

图 5.14 为 DRAM 的工作时序图，其中图 5.14(a)为读周期时序，图 5.14(b)为写周期时序，图 5.14(c)为刷新周期时序。在读周期中，依次完成以下操作：首先，地址线有效，芯片外部先输入行地址信息，并同时使行选通信号  $\overline{RAS}$  有效，将行地址锁存；然后，由于是读出操作，需由外部提供读出操作所需的  $\overline{RD}$  信号，接下来再由外部给出列地址信息，并同时使列选通信号  $\overline{CAS}$  有效，将列地址锁存，此时，二维译码方式的两组地址都已经具备，则启动存储芯片正常工作，所以列选通信号  $\overline{CAS}$  在此相当于  $\overline{CS}$ ，为保证地址译码及读出时信息稳定正确，对应地址信息与选通信息在锁存期间应保持有效；最后，直到数据总线上出现稳定的有效读出数据，才可以撤销  $\overline{RAS}$ 、 $\overline{CAS}$  和  $\overline{RD}$  信号。一般从行选通信号  $\overline{RAS}$  的下降沿开始，到下一次行选通信号  $\overline{RAS}$  的下降沿为止的时间，也就是连续 2 个读操作周期的时间间隔，作为 DRAM 的读周期。

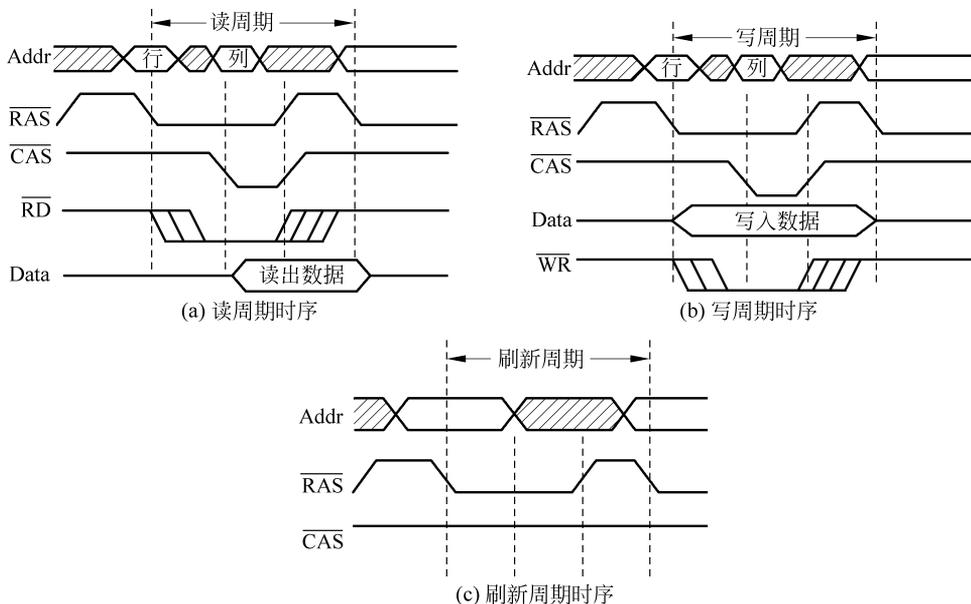


图 5.14 DRAM 的读/写/刷新时序图

在写周期中，依次完成以下操作：首先，地址线有效，芯片外部先输入行地址信息，并同时使行选通信号  $\overline{RAS}$  有效，将行地址锁存；然后，由于是写入操作，外部需提供欲写入的数据信息和写入操作所需的信号  $\overline{WR}$ ，接下来再由外部给出列地址信息，并同时使列选通信号  $\overline{CAS}$  有效，将列地址锁存，此时，列选通信号  $\overline{CAS}$  相当于  $\overline{CS}$ ；其次，在  $\overline{RAS}$ 、 $\overline{CAS}$  和  $\overline{WR}$  信号都有效的情况下，数据信息将被写入存储芯片的对应地址空间；最后，直到数据稳定写入存储芯片之后，才可以撤销数据信息、 $\overline{RAS}$ 、 $\overline{CAS}$  和  $\overline{WR}$  信号。

在刷新周期中，依次完成以下操作：地址线有效，外部给出行地址信息，并同时使行选通信号  $\overline{RAS}$  有效，由于刷新只需要行地址，因此无须再给出列地址和列选通信号。

## 5.2.4 DRAM 的发展

自 20 世纪 70 年代以来,主存储器的基本构件主要是 DRAM 芯片,由于制造工艺、元件材料、工作时序和访问控制方式等方面的改进,其速度虽然有了较大提高,但是相对于 CPU 的速度来讲,依然低了不少。理论上,如果主存储器速度和 CPU 速度基本相当,那么 CPU 和主存储器的性能将会是最优,两者之间进行信息交互不存在时间上的等待状态。而现实当中,由于主存储器速度低于 CPU 速度,必然导致 CPU 访问主存储器时,增加大量的等待周期,牺牲 CPU 的效率。为了提高主存储器的带宽,出现了一些先进的 DRAM,下面分别介绍它们的特点。

### 1. FPM DRAM

FPM(Fast Page Mode) DRAM 称为快速页模式动态存储器。传统 DRAM 采用二维译码方式,读写时必须提供两次地址,分别是行地址和列地址,这会耗费一定时间。根据程序访问的局部性,即在一个相对较短的时间间隔内,CPU 对存储器中程序的访问是局限在一个相对较小的局部存储空间内的。因此,当程序连续运行时,由于程序在存储器中是连续存储的,其行地址可以只传输一次,并保持不变,而列地址可以连续变化,则除了第一次需要传输行地址以外,后面每次都只需要传输一个列地址即可。FPM DRAM 即采用这种寻址操作方式,触发一次行地址后,连续输出列地址,从而使用较少的时钟周期读出较多的数据,此时,页是指一个唯一行地址和该行中所有的列地址确定的若干存储单元的组合,称为快速页。

### 2. EDO DRAM

EDO(Extended Data Out) DRAM 称为扩展数据输出 DRAM,是在 FPM DRAM 基础上改进而来的。FPM DRAM 在存储每一个数据时,必须等到行地址和列地址输入稳定以后,然后才能有效读写数据,而下一个地址必须等待这次读写周期完成才能输出。而 EDO DRAM 输出数据在整个  $\overline{\text{CAS}}$  周期都是有效的(包括预充电时间在内),EDO DRAM 不需要等待当前读写周期结束即可启动下一个读写周期,即可以在输出一个数据的过程中准备下一个数据的输出。这种设计节省了重新生成地址的时间,提高了访问速度。

### 3. SDRAM

SDRAM(Synchronous Dynamic Random Access Memory)称为同步型 DRAM。相对于前面几种非同步存储的存储器而言,由于在系统时钟不同步的情况下进行存储,需要等待若干个时钟周期进行协调才能继续进行存储器的存储,例如 FPM DRAM 需要至少等待 3 个时钟周期,EDO DRAM 需要等待至少 2 个时钟周期,这种等待机制严重影响了存储系统的数据吞吐率。所以 EDO DRAM 和 FPM DRAM 的工作速度都不超过 66MHz。

SDRAM 的操作要求与系统时钟相同步,这种同步操作是由系统的同步脉冲控制的,换句话说,主存储器与 CPU 的数据交换同步于系统时钟信号,并且以 CPU/主存储器总线的最高速度运行,无须等待周期。正是由于取消了等待机制,减少了等待时间,减少了数据传送的延时,SDRAM 的存储系统运行速度提高了很多。

SDRAM 的基本原理是将 CPU 和 RAM 通过一个相同的外部时钟锁在一起,实现了同步工作,从而使得 RAM 和 CPU 能共享一个时钟周期。所以开始的时候 SDRAM 需要花费一定的时间实现同步,但是一旦同步以后,每一个时钟周期就可以读写一个数据。这种速度

已经基本接近主板上同步 Cache 的 3-1-1-1 的水准。一般来说,在系统时钟为 66MHz 的时候,SDRAM 和 EDO DRAM 相比,基本没有明显优势,但是当系统时钟增加到 100MHz 以上的时候,其速度上的优势将会比较明显。

图 5.15 为 8M×16 位的 SDRAM 的内部逻辑结构图,它采用了 4 个存储体(BANK0~BANK3),每个存储体容量为 2M×16 位,它可以同时打开 4 个主存页面,当 CPU 从某一个存储体中访问数据的时候,在主存控制器的控制之下,另外 3 个存储体也已经准备好提供读写数据。外部输入信号除了数据、地址、控制 3 组信号线之外,还有系统输入时钟信号 CLK,CKE 为时钟允许信号,DQM 为数据屏蔽信号。内部的模式寄存器及相关控制逻辑是 SDRAM 独有的,可以指定突发式读写的长度(1、2、4、8、全页字),该长度是同步地向系统总线上发送数据的存储器单元个数。模式寄存器也允许程序员调整从接收读请求命令到开始传输数据的延迟时间。

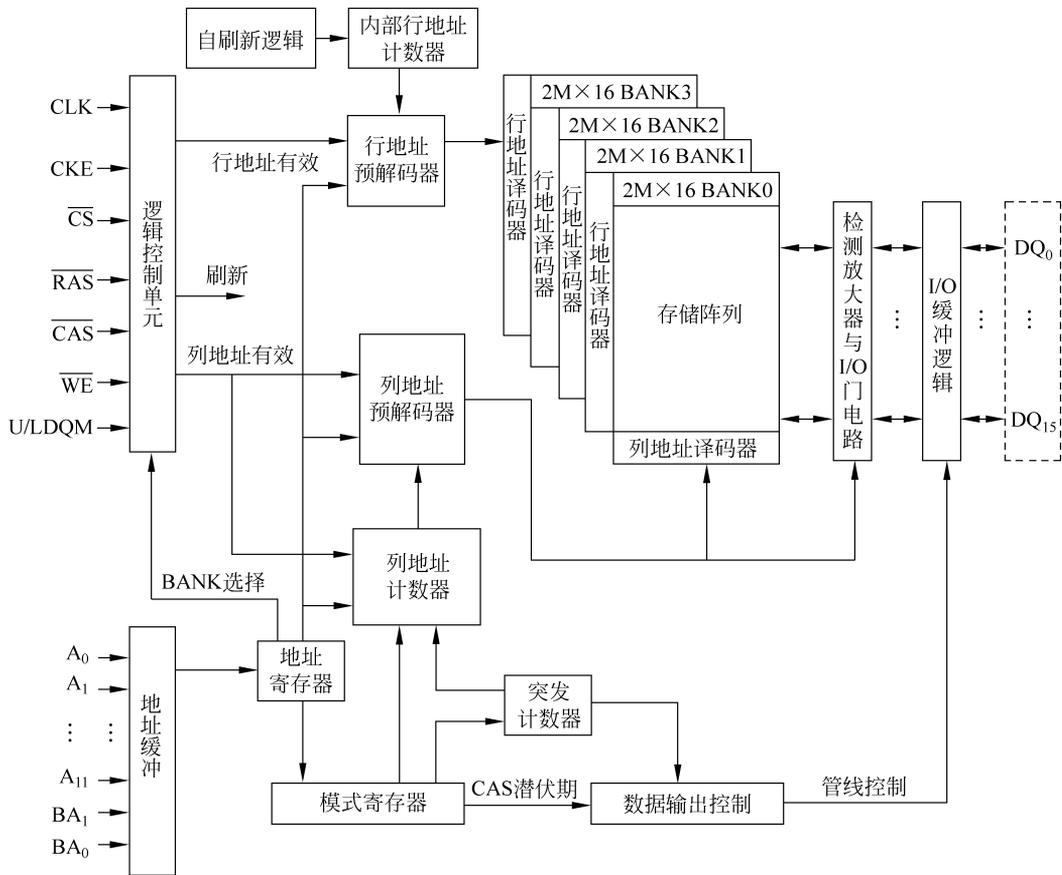


图 5.15 8M×16 位的 SDRAM 的内部逻辑结构图

#### 4. DDR SDRAM

DDR(Double Data Rate) SDRAM 是从 SDRAM 发展而来的,称为双数据传输率同步 DRAM,DDR 运用了更高级的同步电路,它与 SDRAM 的主要区别在于:DDR SDRAM 不仅能在时钟脉冲的上升沿读出数据而且还能在下降沿读出数据,不需要提高时钟频率就能加倍提高 SDRAM 的工作速度。

DDR SDRAM 的频率可以用工作频率和等效传输频率两种方式来表示,工作频率是内存颗粒实际工作频率(又称为核心频率),但是由于 DDR 可以在脉冲的上升沿和下降沿都传输数据,因此传输数据的等效传输频率是工作频率的两倍。由于外部数据总线的数据宽度为 64 位,因此数据传输率(又称带宽)等于等效传输频率 $\times 8$ 。

### 5. DDR2 SDRAM 和 DDR3 SDRAM

DDR2 SDRAM 是新一代内存技术标准,它与上一代 DDR SDRAM 的技术标准的最大不同之处在于,虽然都是采用了在时钟上升沿和下降沿同时进行数据传输的基本方式,但是 DDR2 SDRAM 却拥有 2 倍于上一代 DDR SDRAM 的预读取能力(即 4 位数据预读取),换句话说,DDR2 SDRAM 每个时钟能够以 4 倍于外部总线的速度读写数据,即在同样的 100MHz 的工作频率下,DDR 的实际工作频率为 200MHz,而 DDR2 SDRAM 则可以达到 400MHz。DDR2 SDRAM 的工作电压采用 1.8V,相对于 DDR 的 2.5V 标准电压下降不少,从而有效降低了功耗和发热量。

DDR3 SDRAM 是在 DDR2 基础上继续改进的产品。首先,DDR2 的预取设计位数是 4 位,也就是说 DRAM 内核的频率只有接口频率的 1/4,而 DDR3 的预取设计位数为 8 位,其 DRAM 内核的频率只有接口频率的 1/8。即同样运行在 200MHz 核心工作频率下,DDR2 的等效数据传输频率为 800MHz,而 DDR3 的等效传输频率为 1600MHz。

其次,DDR3 采用点对点的拓扑架构,减轻了地址/命令与控制总线的负担。在 DDR3 系统中,一个内存控制器将只与一个内存通道打交道,而且这个内存通道只能是一个插槽。因此内存控制器与 DDR3 内存模组之间是点对点(Point-to-Point, P2P)的关系(单体的模组),或者是点对双点(Point-to-two-Point, P22P)的关系(双体的模组),从而大大减轻了地址/命令/控制与数据总线的负载。

最后,DDR3 采用 100nm 以下的生产工艺,将工作电压从 1.8V 降至 1.5V,增加异步重置(Reset)与 ZQ 校准功能。由于新一代 DDR3 的工作电压降为 1.5V,相对于 DDR2 来讲大约可以节约 16% 的电能。

### 6. RDRAM

RDRAM(RAMBUS DRAM)是美国的 RAMBUS 公司于 2000 年左右开发生产的一种内存。与 DDR 和 SDRAM 不同,它采用了串行的数据传输模式。其内容主要包括下面 3 个关键部分:RDRAM 芯片、RAMBUS 接口和 Rambus 通道。RAMBUS 接口主要用来连接 RDRAM 芯片和 RAMBUS 通道,RAMBUS 通道的作用是向内存控制器传输数据。RDRAM 与传统 DRAM 的最大区别在于引脚定义会随命令而变,同一组引脚线可以被定义成地址,也可以被定义成控制线,其引脚数仅为正常 DRAM 的 1/3。当需要扩展芯片容量时,只需要改变命令,不需要增加芯片引脚。这种设计减少了铜线的长度和数量,使数据传输中的电磁干扰大为降低,有效地提高内存的工作频率,使 RAMBUS DRAM 可以支持 400MHz 外频。同时 RDRAM 利用上升沿和下降沿两次传输数据,可以使数据传输率达到 800MHz。再考虑到其 16 位的数据总线,所以实际的数据传输率应该为  $400\text{MHz} \times 2 \times 2\text{B} = 1.6\text{GB/s}$ 。

## 5.3 非易失性半导体存储器

非易失性半导体存储器是指当掉电后,所存储的数据不会消失的半导体存储器,简称非易失性存储器。早期的非易失性存储器就是传统意义上的 ROM,它是只能读出而不能写

入的半导体随机存储器,信息无法被用户所改写。但是随着用户需求的不断改变,越来越需要对 ROM 内所存放的内容进行各自不同的设计,于是就诞生了可改写的只读存储器,例如 EPROM、EEPROM 和 Flash 存储器等,但是其写入速度较慢并且写入前需要擦除。相对于 RAM 而言,非易失性存储器的读出速度与之基本相当,但结构更简单,集成度较高,造价较低,功耗也小,而且可靠性较高,具有非易失性,无须刷新。另外相对于外存而言,虽然两者同样具有非易失性,但工作速度快很多,所以在很多场合已经开始逐渐取代外存,实现存储功能。现在一些新型的非易失性存储器具有和 RAM 相似的读写特性,如相变存储器等,这类存储器又称为非易失性随机存取存储器(Non-Volatile Random-Access Memory, NVRAM)。

通常,非易失性存储器的应用主要集中在以下几方面。

(1) 存放一些无须修改的软件程序。非易失性存储器作为主存储器中必不可少的存储器,常用来存放一些无须修改的常用软件程序,比如引导程序、设备管理程序、高级语言的编译程序等。

(2) 存放微程序。在微程序控制的计算机处理器中,存放微程序。

(3) 存放一些特殊编码。比如显示器中的字符发生器、汉字库等。

(4) 存放重要的数据资料,比如录音笔、数码相机、手机记录的音视频数据。

(5) 存放智能设备的操作系统等相关程序。

### 5.3.1 掩模型只读存储器

掩模型 ROM(Masked ROM, MROM)属于传统意义上的 ROM,即其内容由半导体生产厂商按照客户需要在生产过程中直接存入固定信息的 ROM,其内容在写入之后无法改变。一般用来存储具有标准功能广泛使用的程序或数据,或是用户定做的具有特定功能的程序或数据,其生产过程中有一步制造工艺称为“光刻掩模”,即根据欲存放内容的二进制编码设计掩模进行成批生产。大部分 ROM 芯片是利用在行选择线和列选择线交叉位置上的晶体管或 MOS 管是导通或截止的状态来表示二进制的 0 或 1 的。

图 5.16 为采用一维译码方式的 MROM ( $1024 \times 4$  位)结构示意图,存储阵列中每一个存储位元就是一个 MOS 管。外部输入的地址经过行译码后,产生对应的行选择线  $W_i$  选中信号,选中存储阵列中对应  $W_i$  行的所有 MOS 管,如果 MOS 管有连接,位线 D 输出为低,表示存储为“0”;如果 MOS 管没有连接,位线 D 输出为高,表示存储为“1”。

MROM 的优点是可靠性高,价格便宜,适合于保存那些批量较大且不需要用户修改的信息;缺点是用户无法修改,因此灵活性差。

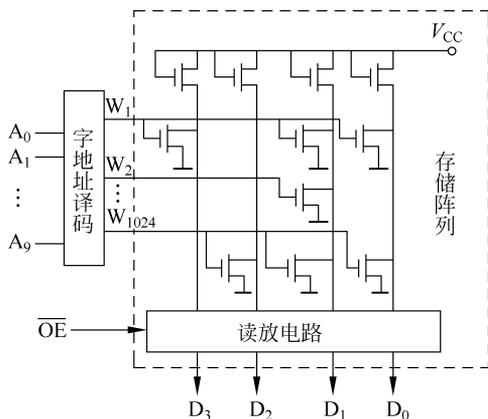


图 5.16 MROM( $1024 \times 4$  位)结构示意图

### 5.3.2 一次可编程只读存储器

一次可编程只读存储器即 PROM(Programmable ROM),是封装后允许用户进行信息写入的半导体只读存储器,但是写入的过程是不可逆的,只能是一次写入。

一般常见的 PROM 有两种结构:一种是熔丝烧断型的(内部结构如图 5.17 所示);另一种是 PN 结击穿型的。刚出厂时,所有存储位元都存储信息为“0”(或“1”),由用户根据自己的需要通过加载过载电压来对存储单元电路进行熔丝烧断或者是 PN 结击穿,由于熔丝烧断后,相应存储单元电路无法连接对应的位线,因此使得位线上输出始终为零,存储信息被改写为“1”(或“0”);或者 PN 结被击穿,则相应存储单元电路直接连接对应位线,存储信息被改写为“1”(或“0”)。

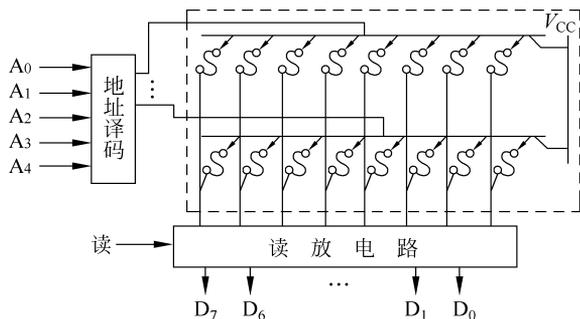


图 5.17 熔丝型 PROM 结构示意图

PROM 芯片在出厂时芯片中无任何信息,用户可以用专用的编程器将自己数据写入,但这种机会只有一次,一旦写入后无法修改,若出现错误,则已写入的芯片只能报废。由于只能一次性写入,可重复利用率低,因此在开发过程中成本依然比较高。

### 5.3.3 可擦除可编程只读存储器

可擦除可编程只读存储器即 EPROM(Erasible Programmable ROM),是一种可以多次改写的 ROM,又称为紫外线擦除可编程只读存储器。类似于 PROM,其出厂时存储内容为全“0”(或“1”),用户可以根据自己的需要通过专用编程器写入信息,每次写入都是将整片内容全部一次性写入。但是若要重写,必须先将原存储内容整片擦除(即恢复为出厂时状态),然后再重新整片写入新存储内容。

EPROM 可重复擦除和写入,解决了 PROM 只能写入一次的弊端。EPROM 芯片有一个很明显的特征,在其正面封装上,开有一个石英窗口,透过该窗口,可以看到其内部的集成电路。EPROM 数据写入需要专用的编程器,数据擦除需要 EPROM 擦除器,通过紫外线透过窗口照射内部芯片来实现。

一般一片 EPROM 可以保存数据 10~20 年,擦写寿命在数百次以上,但由于日光中有紫外线,因此擦除窗口必须保持覆盖,一般用不透光的贴纸将窗口封住,以免受到周围紫外线的照射使数据受损。

### 5.3.4 电可擦除可编程只读存储器

鉴于 EPROM 的数据写入和擦除必须脱机操作,使用不便,后来出现了 EEPROM

(Electric Erasible Programmable ROM),也写作 E<sup>2</sup>PROM。它以电信号来擦除其内容,不必照射紫外线。

E<sup>2</sup>PROM 既可实现正常工作方式中的只读不写,还可以将整个芯片或某个指定单元信息擦除,而其他未通电流的单元内容保持不变。E<sup>2</sup>PROM 有两种擦除方式:数据块擦除方式和字擦除方式。数据块擦除方式与 EPROM 相同;而字擦除方式允许只擦除某一个地址单元的内容,不必将原始数据全部擦除后才能写入。擦除、写入可使用专用设备,也可由应用系统在线改写,摆脱了专用设备的束缚。

### 5.3.5 Flash 存储器

Flash 存储器又称为闪速存储器,简称闪存,是 20 世纪 80 年代中期出现的一种快速读写型只读存储器,属于 E<sup>2</sup>PROM 的改进产品,具有存储密度高和信息非易失性两大优点,是存储技术划时代的产物。Flash 存储器具有以下特点:

- (1) 每个存储位元只需要一个 MOS 管,集成度高;
- (2) 与 E<sup>2</sup>PROM 相比,写入速度快,接近 RAM 存储器;
- (3) 单一电源供电;
- (4) 工作寿命长,编程次数多,可达 100 万次。

由于 Flash 存储器同时继承了 EPROM 的高集成度以及 E<sup>2</sup>PROM 的电可擦写等优点,而且擦写速度快,因此得到了越来越广泛的应用,比如在掌上电脑、数码相机、MP3 播放器、移动存储器、U 盘、存储卡等小型/微型电子产品中都有它的身影。

近年来,固态硬盘(Solid State Drive,SSD)的出现是 Flash 存储器应用的一大热点。固态硬盘是用固态电子存储芯片阵列而制成的硬盘,由控制单元和存储单元(Flash 芯片、DRAM 芯片)组成。它具有读写速度快(持续写入速度可达 500MB/s)、存取时间短(0.1ms 甚至更短)、防震抗摔、低功耗(一个主流 128GB 的 SSD 硬盘的功耗为 2.5~4W)、无噪声、轻便、环境适应性强、接口与普通硬盘完全一致等一系列优点,目前已被广泛应用于军事、车载、工控、视频监控、网络监控、网络终端、电力、医疗、航空、导航设备等领域。

### 5.3.6 铁电存储器

铁电存储器(Ferroelectric RAM,FRAM)利用铁电晶体的铁电效应实现数据存储。由于铁电效应是铁电晶体所固有的一种偏振极化特性,与电磁作用无关,因此 FRAM 存储器的内容不会受到外界条件(诸如磁场因素)的影响,能够同普通 ROM 存储器一样使用,具有非易失性的存储特性。

早期的 FRAM 读写速度不一样,写入时间更长一些。近期的 FRAM 读写速度是一样的。例如,FM1808 芯片的一次读写时间为 70ns。一般情况下,一次读写的时间短,而连续的读写周期要长一些。例如,Ramtron 公司出产的 128K×8 位的 FRAM 芯片 FM20L08 的一次读写时间为 60ns,而其连续的读写周期为 150ns。

FRAM 在功耗、写入速度等许多方面都远远优于 EPROM 或 E<sup>2</sup>PROM。这里特别提出的是写入次数,FRAM 比 EPROM 或 E<sup>2</sup>PROM 要大得多。早期的 FRAM 的写入次数为几百亿次,而目前的芯片可达万亿次以上。

### 5.3.7 磁性随机存储器

从原理上讲,磁性随机存储器(Magnetic Random Access Memory, MRAM)的设计是非常诱人的,它通过控制铁磁体中的电子旋转方向达到改变读取电流大小的目的,从而使其具备二进制数据存储能力。从理论上来说,铁磁体是永久不会失效的,因此它的写入次数也是无限的。例如,Freescale 的 MR2A16A 芯片在最差的操作环境下可以经受 58 万亿次读写周期。

MRAM 可以像 SRAM 一样操作,无须时钟和刷新,写入速度快,不需要擦除操作。如 MR2A16A 读写访问时间均为 35ns,采用 44 引脚 TSOP II 型封装,该封装完全兼容 SRAM,从而可以代替 SRAM,应用于采用相同的 SRAM 结构的现有硬件系统中。和 SRAM 相比,MRAM 又具有非易失性,在不需电能的情况下可以保存存储内容至少十年。

### 5.3.8 相变存储器

相变存储器(Phase Change Memory, PCM,也称 Phase-change RAM, PRAM)利用材料的可逆转的相变存储信息。

相变存储器具有如下特性:

(1) 非易失性。

(2) 一位可变特性。如同 RAM 或 E<sup>2</sup>PROM,PCM 可变的最小单元是一位。闪存技术在改变存储的信息时要求有单独一步的擦除步骤,而在一位可变的存储器中存储的信息在改变时无须单独的擦除步骤,可直接由 1 变为 0 或由 0 变为 1。

(3) 较高的使用寿命。尽管相变存储的使用寿命无法和 DRAM 相比,但是和 Flash 相比仍然有着较为明显的优势。

(4) 随机读取速度快。如同 RAM 和 NOR 闪存,PCM 技术具有随机存储速度快的特点。这使得存储器中的代码可以直接执行,无须中间复制到 RAM。PCM 读取反应时间与最小单元 1b 的 NOR 闪存相当,而它的带宽可以媲美 DRAM。

(5) 写入速度较快。PCM 能够达到如同 NAND Flash 的写入速度,但是 PCM 的反应时间更短,且无须单独的擦除步骤。NOR 闪存具有稳定的写入速度,但是擦除时间较长。PCM 同 RAM 一样无须单独擦除步骤,但是写入速度(带宽和反应时间)尚不及 RAM。随着 PCM 技术的不断发展,存储单元缩减,PCM 将不断被完善。

(6) 字节寻址方式。相变存储器能够像 RAM 一样按字节读写数据。



## 5.4 主存储器的组织

### 5.4.1 CPU 与主存储器的连接

存储器和处理器之间的接口信号包括:

- $\overline{RD}$ (读)——存储器读周期的信号;
- $\overline{WR}$ (写)——存储器写周期的信号;

- $\overline{\text{MEM}}$ (存储器选通)——存储器的选通信号;
- $A_{0\sim(k-1)}$ (地址线)——连接存储器的单向地址线,共  $k$  位;
- $D_{0\sim(n-1)}$ (数据线)——连接存储器的双向数据线,共  $n$  位。

下面以 SRAM 芯片为例,显示 CPU 与存储器芯片连接示意图,如图 5.18 所示。

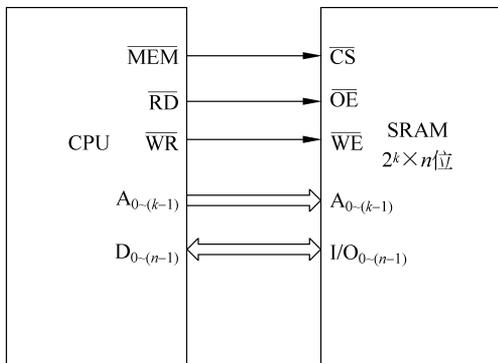


图 5.18 CPU 与存储器芯片连接示意图

### 5.4.2 主存储器容量的扩充

单个存储芯片的容量受到集成度的限制,不可能由单个存储芯片提供主存储器所需要的大容量,往往是由若干个存储芯片扩展形成主存储器,然后与 CPU 连接。本节讨论利用多个存储芯片组成一定容量的主存储器以及与 CPU 连接的方法。

下面先提出需要考虑的几方面。

(1) 确定存储芯片的类型。

计算机系统的主存储器通常由两种类型的存储区域构成: ROM 存储区域用于存储固定不变的程序或数据,一般由 EPROM、E<sup>2</sup>PROM 或者 Flash 存储芯片构成; RAM 存储区域用于存储系统、用户程序和数据,可由 SRAM 或 DRAM 存储芯片构成。

(2) 确定所需存储芯片的数量。

存储芯片数量 = 主存储器要求的容量 / 存储芯片的单片容量

例如,用容量为  $16\text{MB} \times 4$  的存储芯片构成容量是  $256\text{MB} \times 8$  的主存储器,需要的存储芯片数量为

$$\frac{256\text{MB} \times 8}{16\text{MB} \times 4} = 16 \times 2 = 32$$

(3) 确定扩展方式。

主存储器的扩展从存储单元数和字长两个方面进行。如果单个芯片的字长小于所要求的主存储器的字长,则需要在字长位数上扩展,称为位扩展;如果芯片的存储单元数小于主存储器的单元数,则需要扩展存储单元的数量,称为字扩展。实际应用中根据所选用的芯片,可能只需要位扩展或字扩展,也可能需要字位同时扩展。

(4) 确定连接关系。

每个存储芯片的地址、数据和控制等引脚应分别和 CPU 的相应信号线连接,形成一个有机整体,共同构成一个大容量的主存储器。

由于 DRAM 存储器的引脚比 SRAM 存储器复杂,下面只通过 SRAM 说明 3 种扩展方式的原理。ROM 存储器的扩展将在随后做简要说明。

### 1. 位扩展

位扩展是指只扩展存储器字长而不扩展存储单元数目。例如,用  $16\text{K} \times 1$  位的 SRAM 芯片扩展形成  $16\text{K} \times 8$  位的主存储器时,由于芯片的存储单元数和所要求的主存储单元数相同,因此不需要扩展;而两者的字长不同,应进行扩展。存储芯片的字长为 1 位,而所要求的主存储器字长为 8 位,应该使用 8 片 SRAM 芯片进行位扩展。位扩展逻辑结构如图 5.19 所示。

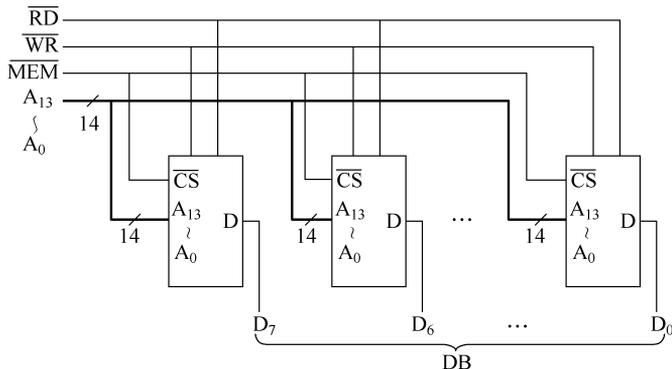


图 5.19 用  $16\text{K} \times 1$  位的存储芯片构成  $16\text{K} \times 8$  位的存储器的位扩展逻辑结构

对图 5.19 所示的位扩展方式的分析如下。

#### 1) 数据线的连接

主存储器字长为 8 位,存储芯片的字长为 1 位,8 个芯片各提供 1 位组成主存储器的一个字。各个芯片的数据引脚  $D$  分别连接到 CPU 数据总线的  $D_7 \sim D_0$  位上。

#### 2) 地址线的连接

存储芯片和主存储器的存储单元数都为  $16\text{K}$ ,则它们的地址引脚数都为  $\lg(16\text{K}) = 14$  个。各个存储芯片 14 个地址引脚  $A_{13} \sim A_0$  并联到 CPU 地址总线。

#### 3) 片选信号线的连接

SRAM 芯片的片选信号  $\overline{CS}$  控制其是否工作,位扩展方式的各个芯片总是同时工作,因此各芯片的片选信号并联到 CPU 的  $\overline{MEM}$  信号线,该信号在 CPU 访问存储器时有效。

#### 4) 读写信号线的连接

各个芯片的读、写信号分别并接到 CPU 的读信号  $\overline{RD}$  和写信号  $\overline{WR}$ ,各个芯片同时读出或同时写入。

在位扩展方式中,当 CPU 访问某一个地址单元时,各个芯片的相同地址单元同时被访问。例如,若 CPU 要读取地址为  $0000\text{H}$  的主存单元,则 8 个芯片的  $0000\text{H}$  单元的内容同时被读出,通过数据总线  $D_7 \sim D_0$  送给 CPU。

### 2. 字扩展

字扩展是指只扩展存储单元的数目而不扩展存储器的字长。例如,用  $16\text{K} \times 8$  位的 SRAM 存储芯片扩展形成  $64\text{K} \times 8$  位的主存储器,由于两者的字长相同都是 8 位,因此不需要扩展。而所要求主存储器的存储单元数大于单个芯片的存储单元数,应该使用 4 片存储

芯片经过字扩展形成 64K 存储器空间。该例的字扩展逻辑结构如图 5.20 所示。

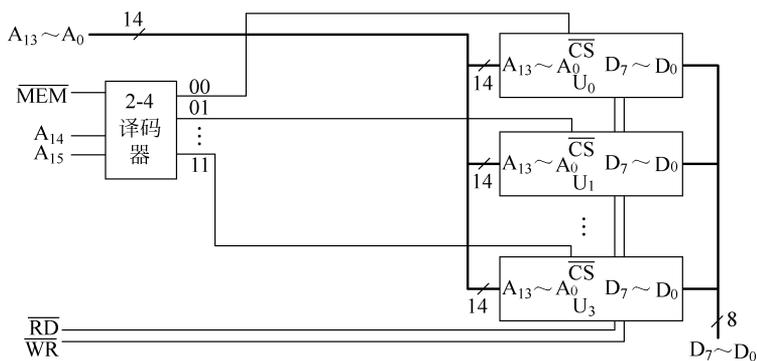


图 5.20 用 16K×8 位的存储芯片构成 64K×8 位的存储器的字扩展逻辑结构

下面分析图 5.20 的连接关系。

1) 数据线的连接

因为存储芯片和扩展形成的主存储器的字长相同,均为 8 位,所以 4 个芯片的 8 根数据线  $D_7 \sim D_0$  并联,并且连接到 CPU 数据总线。

2) 地址线和片选线的连接

主存储器的地址空间为 64K,而芯片的存储单元数为 16K,可以认为整个主存空间分为 4 段,每个芯片占据其中的 1 段地址空间(见表 5.1)。当 CPU 访问某一个地址单元时,只有一个芯片的地址空间包含了要访问的单元地址,因此,4 个芯片是不能同时选中的。

表 5.1 地址空间分配

芯片号	片选地址( $A_{15} A_{14}$ )	芯片内地址( $A_{13} \sim A_0$ )	占据的地址空间(十六进制)
$U_0$	00	0000000000000000~1111111111111111	0000~3FFF
$U_1$	01	0000000000000000~1111111111111111	4000~7FFF
$U_2$	10	0000000000000000~1111111111111111	8000~BFFF
$U_3$	11	0000000000000000~1111111111111111	C000~FFFF

64K 的地址空间需要 16 根地址线,从表 5.1 可以看出,4 段地址空间是由 CPU 的高位地址  $A_{15}$  和  $A_{14}$  区分的。因此,将  $A_{15}, A_{14}$  经过 2-4 译码器产生 4 个选择信号,分别连接 4 个存储芯片的片选信号。而 CPU 的其余 14 根地址线  $A_{13} \sim A_0$  连接到各个芯片的地址引脚,作为芯片内部存储单元的地址。

由于各个芯片的片选信号分别由 2-4 译码器的不同输出信号驱动,因此在某一时刻不可能有 2 个及以上的芯片同时工作,否则多个芯片将同时驱动数据总线,出现冲突。也就是说,当 CPU 在访问存储器时,给定地址的存储单元由 4 个芯片中的某一个提供,该芯片根据地址总线的  $A_{13} \sim A_0$  选择存储矩阵中的一个存储单元与 CPU 传输数据,而其余 3 个芯片的数据引脚输出为高阻态。

CPU 的存储器访问信号  $\overline{MEM}$  作为地址译码器的控制信号,当  $\overline{MEM}$  无效时,译码器的 4 个输出均为高电平,4 个存储芯片均不工作;只有当  $\overline{MEM}$  有效时,才会有芯片被选中。

3) 读写信号线的连接

所有芯片的读、写信号分别并接到 CPU 的读信号  $\overline{RD}$  和写信号  $\overline{WR}$ 。但是这并不意味着所有的芯片同时读或同时写,由于片选信号不同,因此只可能有一个芯片读出或写入。

最后举一个写操作的例子。假设 CPU 要将数据 34H 写入主存储器中地址为 5012H 的存储单元,CPU 通过地址总线送出地址信息为 010100000010010,同时发出  $\overline{MEM}$  有效的访问内存信号;  $A_{15}$ 、 $A_{14}$  为 01,经过 2-4 译码器后产生存储芯片  $U_1$  的片选信号;  $A_{13} \sim A_0$  被  $U_1$  用来选中片内 0100000010010(1012H)单元。在写信号  $\overline{WR}$  的作用下,将 CPU 送到数据总线上的 34H 写入  $U_1$  的 1012H 单元。

3. 字位同时扩展

字位同时扩展是指同时对存储单元数和存储单元字长进行扩展,是位扩展和字扩展的结合,集合了两种方式的特点。例如,用  $16K \times 1$  位的 SRAM 存储芯片扩展形成  $64K \times 8$  位的存储器,应该进行字位同时扩展,所需要的存储芯片数量为

$$\frac{64K \times 8}{16K \times 1} = 4 \times 8 = 32$$

可以这样理解字位扩展:先用若干片存储芯片经过位扩展,使得位扩展后存储单元的字长达到主存储器的要求,本例的位扩展需要 8 个芯片,可以把这 8 个芯片看成一个  $16K \times 8$  位的存储模块。然后再由多个存储模块进行字扩展,形成所需要的主存单元数。本例中应该由  $64K \div 16K = 4$  个存储模块进行字扩展。字位同时扩展后的逻辑结构如图 5.21 所示。

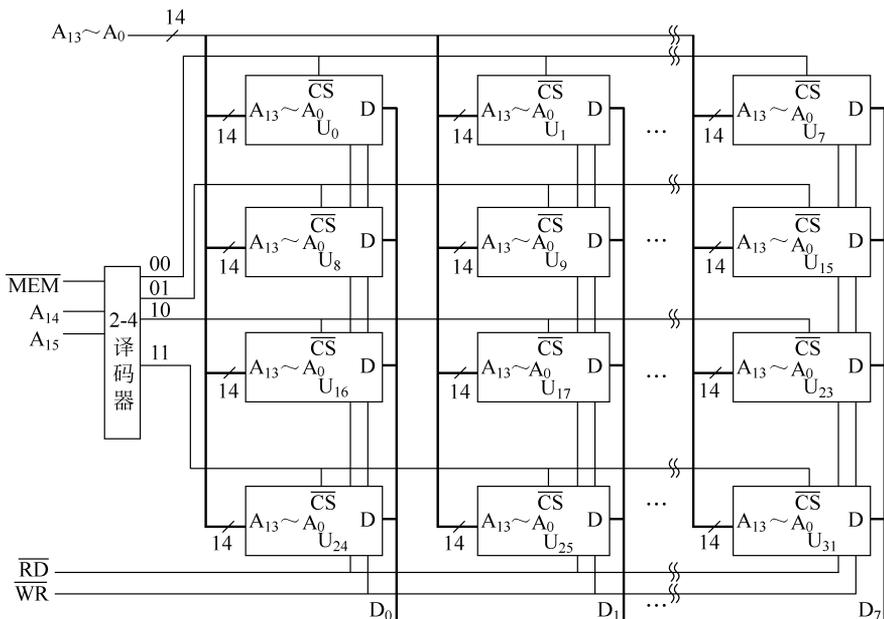


图 5.21 用  $16K \times 1$  位的存储芯片构成  $64K \times 8$  位的存储器的字位同时扩展后的逻辑结构

图 5.21 中,每一行中的 8 个存储芯片 ( $U_0 \sim U_7$ 、 $U_8 \sim U_{15}$ 、 $U_{16} \sim U_{23}$ 、 $U_{24} \sim U_{31}$ ) 各形成 1 个存储芯片组,共计 4 个。组内 8 个芯片各提供 1 根数据线构成模块的 8 根数据线,组成 1 个字;所以 8 个芯片应同时工作,它们的片选信号并联,作为模块的片选信号。各个芯片

的读写信号并联后作为模块的读写信号。因此,位扩展后的芯片组可以看作是一个容量为 $16\text{K}\times 8$ 的存储模块,有1个片选信号、8根数据线、14根地址线以及读写信号线。

4个存储模块再经过字扩展最终形成需要的主存储器。字扩展时, $A_{15}$ 和 $A_{14}$ 作为地址译码器的输入,每个模块的片选信号分别连接到2-4译码器的不同输出端,使不同的模块不会同时被选中。各存储模块的8根数据线并联到CPU数据总线对应位上。各存储模块的读写信号分别并联到CPU的读写信号线上。

在上面的3个例子中,都只是使用SRAM芯片。如果主存储器既包含SRAM又包含ROM,区别只是ROM型存储芯片一般只进行读操作,因此在扩展时不需要连接写信号,其他的与SRAM相同。

### 5.4.3 主存储器的编址方式

在计算机系统中,存储器编址方式是指对存储设备进行地址编排的方法,目前常用的编址单位有按字编址、按字节编址和按位编址等几种。下面以主存储器的编址方式为例说明各种编址单位的特点。

#### 1. 按字编址

按字编址是实现起来最容易的一种编址方式,这是因为每个编址单位所包含的二进制信息位数与主存储器的存储单元字长一致。如果主存的字长为32位,则每次读写存储器只能整字读写,而不能读写存储单元的一部分(如8位或16位),如图5.22所示。

按字编址实现起来简单,但它的缺点是没有提供对字节访问的支持。对于字符串等以字节为单位的数据,操作起来就不很方便。

#### 2. 按字节编址

按字节编址适应了非数值计算的需要,是现代计算机普遍使用的编址方式。因为非数值(如ASCII码)应用要求按字节编址,它的基本寻址单位是字节。在按字节编址方式中,每个编址单位所包含的二进制信息位数为8位,即每个字节具有1个地址,称为字节地址;而主存储器的存储单元地址称为字地址,字地址用该存储单元中最小的字节地址表示。字节地址是连续的,而字地址是不连续的;以32位字长为例,1个字包含4字节,字地址的间隔为4,如图5.23所示。

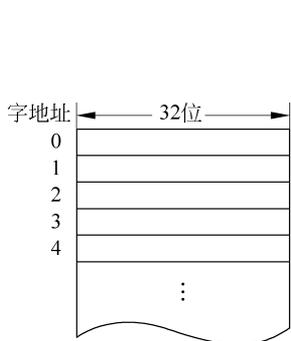


图 5.22 字编址方式

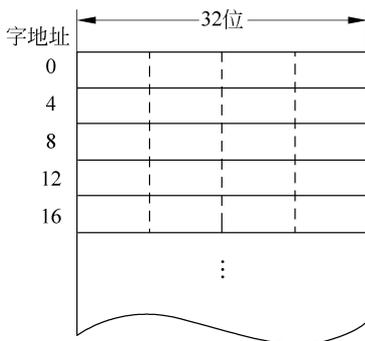


图 5.23 字节编址方式

#### 3. 按位编址

也有部分计算机采用按位编址方式,如STAR-100巨型计算机等。每个编址单位所包

含的二进制信息位数为 1,即存储器内的每一位二进制信息都有各自的地址。按位编址方式可以对可变字长运算提供有力的支持,但占用的地址更多。

**例 5.1** 某存储器的容量为 1M 位,请确定分别采用按字编址(字长为 16)、按字节编址和按位编址时的地址范围。

**解:** 采用按字编址时,字长为 16,则存储器中存储单元的个数为  $1\text{M}/16=2^{16}$ ,所以每个存储单元的地址信息为 16 位,存储器的地址范围是 0000H~FFFFH。

采用按字节编址时,每个编址单位包含 8 位二进制信息,存储器中共有  $1\text{M}/8=2^{17}$  字节单元,每个单元的地址信息为 17 位,存储器的地址范围是 00000H~1FFFFH。

采用按位编址时,每个编址单位包含 1 位二进制信息,存储器中共有  $1\text{M}/1=2^{20}$  个单元,每个单元的地址信息为 20 位,存储器的地址范围是 00000H~FFFFFFH。

#### 5.4.4 哈佛结构

哈佛结构(Harvard Architecture)是一种将程序指令存储和数据存储分开的存储器结构。它的主要特点是将程序和数据存储在不同的存储空间中,即程序存储器和数据存储器是两个独立的存储器,每个存储器独立编址、独立访问(见图 5.24),目的是减轻程序运行时的访存瓶颈,以便实现程序和数据的并行访问。与两个存储器相对应的是系统需要设置 4 种独立的总线:程序存储器的数据总线与地址总线、数据存储器的数据总线与地址总线。

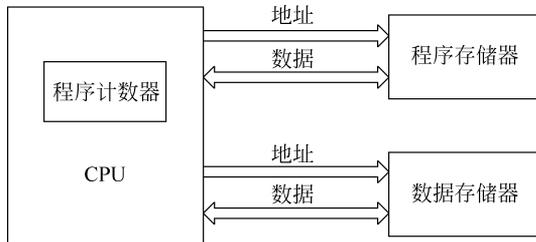


图 5.24 哈佛结构

哈佛结构具有以下优点:

(1) 分离的程序总线 and 数据总线可允许在一个机器周期内同时获得指令字(来自程序存储器)和操作数(来自数据存储器),从而提高了执行速度,提高了数据的吞吐率;

(2) 由于程序和数据存储在两个分开的物理空间中,因此指令的取址和执行能完全重叠,便于指令流水线的流水执行。CPU 首先到程序指令存储器中读取程序指令内容,解码后得到数据地址,再到相应的数据存储器中读取数据,并进行下一步的操作(通常是执行)。

(3) 程序指令存储和数据存储分开,可以使指令和数据有不同的数据宽度。如 Microchip 公司的 PIC16 芯片的程序指令是 14 位宽度,而数据是 8 位宽度。

(4) 哈佛结构的微处理器通常具有较高的执行效率。由于程序指令和数据分开组织和存储,执行时可以预先读取下一条指令实现指令的预取,缩短取指令所需要的时间。

目前使用哈佛结构的 CPU 和单片机有很多,除了上面提到的 Microchip 公司的 PIC 系列芯片,还有摩托罗拉公司的 MC68 系列、Zilog 公司的 Z8 系列、Atmel 公司的 AVR 系列和安谋公司的 ARM9 等。

虽然哈佛结构非常适合于流水线处理器,但是由于哈佛结构对外部设备的连接要求比较高,需要大量的数据线,因而很少使用哈佛结构作为处理器芯片的外部架构。对于处理器芯片内部,则可以通过使用不同的数据和指令 Cache(例如在 L1 Cache 上应用哈佛结构),有效地提高指令执行的效率。

## 5.5 多模块存储器



目前,计算机性能的瓶颈主要集中在主存储器的带宽上,由于在冯·诺依曼机中存储器处于核心地位,所有的信息交换都需要通过主存储器进行,因此对主存储器的性能要求极高。通常提高主存储器数据传输率的方法只有两种:一种是采用高速存储器件来提高访问速度;另一种是采用并行技术来提高存储器的带宽。

一般的存储器是单体单字存储器,只包含一个存储体。多模块存储器是按照资源重复的思想并行设立了多个相同的存储模块,在一个存储周期中,多个存储模块同时存取多个字,以此来提高访问速度。根据其组织方式的不同,多模块存储器又可以分为单体多字和多体单字两种方式。

### 1. 单体多字存取方式

单体多字的单体是指只有一套地址寄存器和译码电路,多字指的是存在多个容量相同的存储模块或存储体。由于共用同一套地址寄存器和译码电路,所以在同一个读写周期中,将按同一地址并行访问各个存储模块的相对应单元,若存储模块有 $N$ 个,每个存储模块每次访问的字长为 $W$ 位,则 $N$ 个存储模块可以同时访问 $N \cdot W$ 位。所以这 $N$ 个存储模块也可以看成是一个大的存储器,其容量为所有模块容量之和,每次访问的信息长度为 $N \cdot W$ 位,含 $N$ 个字,如图5.25所示。

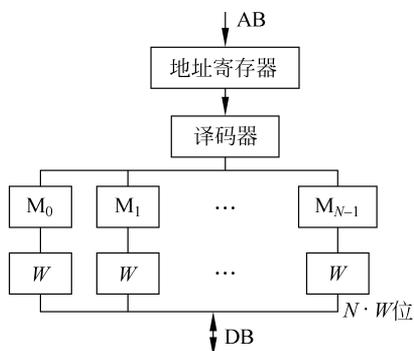


图 5.25 单体多字存储器示意图

这 $N$ 个字被读出以后可以同时或分时送到总线上,CPU在访问主存的时候,带宽将提高为原来的 $N$ 倍。当然这需要一个前提条件,那就是指令和数据在主存中的存放必须是连续存放的,一旦遇到转移指令,或数据本身无法连续存放,这种方法的效果就不怎么明显,甚至还会出现大量的访存冲突。

### 2. 多体单字交叉访问方式

多体单字的多体是指含有多个容量相同的存储模块,而且每个存储模块都有各自独立的地址寄存器、译码电路、数据寄存器和读写电路。各个模块可以独立工作,既能并行工作也能交叉工作。所谓交叉访问,是指各个模块的存储单元交叉编址且存取时间均匀分布在同一个存取周期当中。

多体单字交叉访问存储器结构如图5.26所示。存储器地址寄存器的低位部分经过译码选择不同的存储模块,而高位部分则存放每个存储模块的模块内地址。现以4个存储模块构成的多体交叉存储器为例来说明其常用的编址方式。若有4个存储模块分别为 $M_0$ 、 $M_1$ 、 $M_2$ 、 $M_3$ ,可以看到连续的地址0、1、2、3被依次分配到相邻的存储模块 $M_0$ 、 $M_1$ 、 $M_2$ 、 $M_3$ 中,然后4、5、6、7又被依次分配到相邻的存储模块 $M_0$ 、 $M_1$ 、 $M_2$ 、 $M_3$ 中,直到所有地址全部分配完成为止。可以看到由于连续的地址被分布到相邻的存储模块中,而同一个存储模块内的地址是不连续的,因此称为交叉编址,又称为横向编址。在本例中,由于存储模块有4个,因此称为模4交叉编址。若有 $N$ 个存储模块,则称为模 $N$ 交叉编址。一般 $N$ 应为

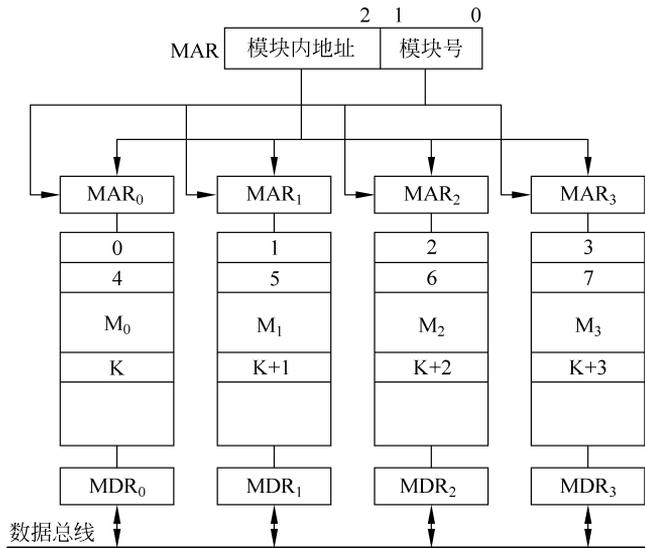


图 5.26 多体单字交叉访问存储器结构

2 的整数幂。

多体交叉访问存储器采用分时启动的方式,利用流水线的工作方式进行工作,在不改变每个模块的存储周期的前提之下,利用流水线的时间重叠技术,提高整个主存储器的访问速度。例如 4 个存储模块,在第一个存储周期  $T$  的开始时刻启动存储模块  $M_0$ ,之后在  $T/4$ ,  $T/2, 3T/4$  的时刻分别启动模块  $M_1, M_2, M_3$ ,如图 5.27 所示,经过  $T$  后,4 个模块进入并行工作方式,各模块的存储周期互相重叠,在完全并行无冲突的理想状态下,整个主存的存储周期可以缩小为原来的  $1/4$ ,即存储速度提高为原来的 4 倍。若存储模块的个数为  $N$ ,则存储周期可以缩小为原来的  $1/N$ ,即存储速度提高为原来的  $N$  倍。

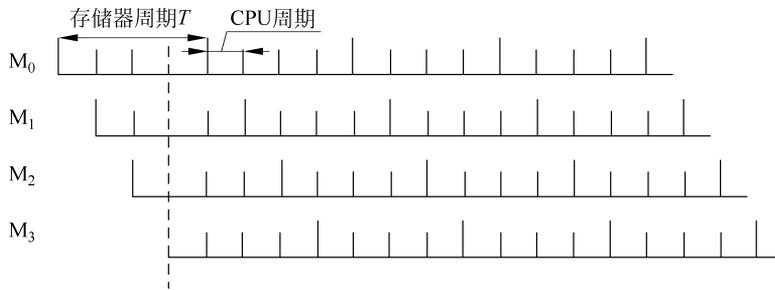


图 5.27 多体单字交叉访问存储器( $N=4$ )分时启动时序图

但是在实际应用过程中,由于数据相关和程序转移的原因,并行性将无法达到理想状态,因此实际值一般在  $T/N$  到  $T/\sqrt{N}$  之间。交叉存取的最大缺陷是当任一模块出现问题的时候,整个存储器都将无法正常工作。

**例 5.2** 设有 8 个模块组成的 8 体存储器结构,每个模块的存取周期为 400ns,存储字长为 32 位。数据总线宽度为 32 位,总线传输周期为 50ns,求交叉存储的存储器带宽。

**解:** 8 体存储器的总信息量 = 32 位  $\times$  8 = 256 位。

对于 8 体交叉存储器,连续读出 8 个字所花费的总时间为

$$t = T + (m - 1)\tau = 400\text{ns} + (8 - 1) \times 50\text{ns} = 750\text{ns} = 7.5 \times 10^{-7}\text{s}$$

因此,交叉存储器的带宽 $=256/(7.5 \times 10^{-7})=341 \times 10^6(\text{b/s})=341\text{Mb/s}$ 。



## 5.6 辅助存储器

**辅助存储器**(Auxiliary Storage)简称**辅存**。它的主要功能是存放当前 CPU 运行时暂时不用的程序和数据,在功能上是主存储器的后备和补充。当需要用到辅存中的程序和数据时,再将其调入内存。从其所处的部位和与主机交换信息的方式看,辅助存储器属于外部设备的一种,所以又被称为**外存**(External Storage)。辅助存储器的特点是存储容量大,单位价格低,断电甚至脱机后仍能保存信息,是非易失性存储器。常用的辅助存储器有磁表面存储器(如磁带、磁盘)和光存储器(如光盘)。下面主要介绍辅助存储器的原理、构成以及性能指标。

### 5.6.1 磁记录原理及记录方式

磁表面存储器存储信息的基本原理是利用磁性材料的剩磁状态存储二进制信息。当外磁场的磁场强度  $H$  增加到一定值  $H_c$  时,磁性材料被磁化而产生的饱和磁感应强度为  $B_s$ 。这时,若除去外磁场,磁性材料的磁感应强度并不减小到 0,而是沿磁滞回线下降到  $B_r$ ,  $B_r$  就称为剩磁,这种现象就是磁滞现象,如图 5.28 所示。通常用作存储记忆元件的磁记录介质是硬磁性材料,满足  $B_s = B_r$ ,  $H_c$  也不大,磁滞回线是矩形,如图 5.29 所示,所以这种材料也称为矩磁材料。磁表面存储器就是把这些矩磁材料均匀地涂在载体的表面,形成厚度  $0.1 \sim 5\mu\text{m}$  的磁层,信息就记录在磁层上。而载体是由非磁性材料制成的,可以是金属合金,也可以是塑料。若载体为带状,称为磁带;若载体为盘状,则称为磁盘。磁层和载体合在一起称为记录介质。

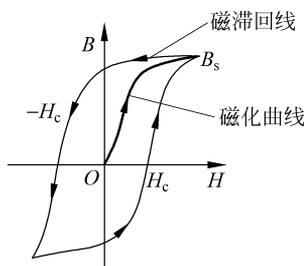


图 5.28 磁滞现象

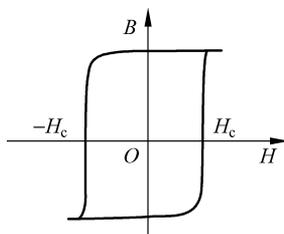


图 5.29 矩磁材料的磁滞回线

磁表面存储器就是在磁头和磁性材料的记录介质之间有相对运动时,通过一个电磁转换过程完成读写操作的存储设备。

磁头是实现电磁能量转换的关键元件,通常用绕有线圈的铁氧体等高磁导率软磁性材料(外界磁场的作用消失后,该磁性材料的磁性容易消失)制成铁心,在铁心中有一个缝隙,用玻璃等非磁性材料填充,称为头隙。在读写过程中,记录介质和磁头做相对运动,一般采用记录介质运动而磁头不动。

当写入信息时,主机送来的是并行数据,经移位寄存器并—串转换变换为串行数据,再

逐位由写驱动器转换为具有一定方向和大小的写入电流,加载到磁头线圈上。写入电流使磁头产生一定方向的磁场,在该磁场的作用下,运动到磁头缝隙下方的一个微小区域的磁层被磁化,称为一个磁化单元。写入信息不同,产生的写入电流方向也不同,磁化单元被磁化的方向也不同,产生的剩磁状态也不同。例如剩磁“ $+B_r$ ”表示“1”,剩磁“ $-B_r$ ”表示“0”。若磁头中无电流,磁头不会被磁化,也就不会对磁记录介质产生任何影响,不会形成剩磁,即无写入操作。

当读出操作时,记录介质相对磁头距离很近且做高速移动,磁头线圈对处于剩磁状态的磁化单元做切割磁力线的运动,这样在磁头线圈中产生感生电动势: $e = -n \frac{d\phi}{dt}$ ,不同方向的剩磁场会在磁头中感应出不同极性的电动势。感生电动势经过放大、检波、限幅、整形后,在选通脉冲的作用下,还原成原来存入的数据。由于数据是一位一位串行读出的,故要送到串一并转换寄存器变换为并行数据,经过处理后再送到计算机。磁表面存储器的基本组成和读写原理如图 5.30 所示。

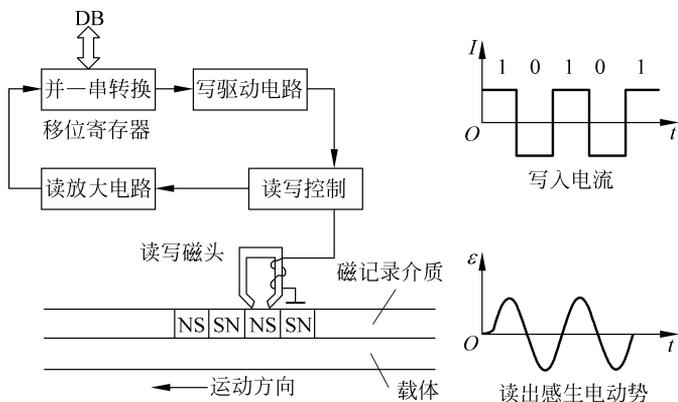


图 5.30 磁表面存储器的基本组成和读写原理

磁表面存储器使用的是一种使磁层饱和磁化的记录方式。上面提到的剩磁究竟代表什么,这与磁记录方式有密切关系。磁记录方式是一种编码方式,即按照某种规律将一连串的二进制信息转换为磁层的相应磁化翻转状态的序列。由于磁化是在磁头中以不同方向的磁化电流来实现的,因此磁记录方式取决于写入电流波形的组合方式。磁记录方式的选取对记录密度、存储容量、传送速率以及读写控制逻辑有直接影响。下面仅讨论几种常见的记录方式。

### 1. 归零制

归零(Return to Zero, RZ)制是用向磁头线圈送入正、负脉冲电流的方法执行写“1”、写“0”操作的方案。写“1”时,加正向写入脉冲电流;写“0”时,加反向写入脉冲电流。一位信息写完后,磁头线圈中电流总回归为零,故称“归零制”。在这种方式中,相邻两个写入电流脉冲之间的电流为 0,相应的这段磁层就未被磁化,因此在写入信息之前必须先退磁。由于这种方法磁层中相邻两个信息位之间有未被磁化的空白区,记录密度低,抗干扰能力差。但是由于每个位单元有两个读出波形,因此具有自同步能力。

### 2. 不归零制

这里介绍的不归零制 NRZ1 是见“1”就翻转的不归零(Not Return to Zero, NRZ)制,它

是不归零制的一种改进。与前一种方案相比,取消了两个信息位之间磁头线圈中无电流的情况,故磁层中不存在未被磁化的状态。写“1”时,磁头线圈的电流改变一次方向;写“0”时,磁头线圈中的写入电流维持原方向不变。这种方式由于没有未被磁化的间隙,因此记录密度较高,但存“1”时才能读出信号,存“0”时无读出信号,故无自同步能力,一般主要用于低速磁带中。

### 3. 调相制

调相(Phase Modulation, PM)制又称为相位编码或曼彻斯特编码。它利用磁层的磁化翻转方向的相位不同来分别表示“1”或“0”。这种记录方式的编码规则是:写“1”时,电流在位的中心处有一次方向固定的变化,如由正变负;写“0”时,仍在位的中心处,写电流向相反的方向变化,即由负变正。当连续写入多个“0”或多个“1”时,为了满足相位的要求,写电流在两个位周期交界处要改变一次方向。这种记录方式的特点是在写入一个位信息的期间,写电流相位至少有一次改变。因此,它与下面的调频制具有类同性,利用记录信号变向,可以生成读同步脉冲。这种方式主要用于快速启停式磁带机中。

### 4. 调频制

调频(Frequency Modulation, FM)制是利用磁层中不同的磁化翻转次数区别数据“1”和“0”的方案。

具体规则如下:

- (1) 无论写“0”或写“1”,在两个数据位之间,写电流改变一次方向。
- (2) 写“0”时,写电流在位周期之内保持不变;而写“1”时,写电流在位周期中心改变一次方向。

这种方式在记录“1”时的写入电流频率是记录“0”时的写入电流频率的两倍,因此又称为倍频制。采用这种记录方式读出时,读出的“1”信号表现为两个脉冲,读出的“0”信号表现为一个脉冲,它有自同步能力。这种方式主要用于早期的磁盘中。

### 5. 改进调频制

调频制的缺点是每位信息的写入电流都有1~2次变化,造成记录密度较低。改进调频(Modified Frequency Modulation, MFM)制就是在FM制基础上改进后的一种记录方式,减少了写入电流翻转的次数,又称为延迟调制码或密勒码。其编码规则是:

- (1) 连续“0”信息之间写电流变化一次。
- (2) 写“1”信息,只在位周期中心写电流变化一次。

采用MFM制的记录密度是FM制的两倍,故称为双密度,并且仍然保持了自同步能力,该方法主要应用于硬磁盘机和双密度软盘机中。

图 5.31 中给出了上述几种记录方式的写入电流波形。不同的磁记录方式性能各异,评价一种记录方式优劣的标准主要是编码效率和自同步能力等。自同步能力是指能从本磁道读出的信息脉冲序列中提取出选通时钟信号,而无须增加附加的同步磁道。磁表面存储器为了从读出信号中分离出数据信息,必须要有时间基准信号,称为同步信号。同步信号可以外加(称为外同步),如NRZ1制;如果可以从读出信号中提取出同步信号,则称为内同步或具有自同步能力,如PM制、FM制、MFM制。

编码效率又称记录效率,是指每次磁层磁化翻转所存储信息的位数。FM制、PM制记录方式中存储一位信息时磁层磁化最大翻转次数是2,因此编码效率为50%。而NRZ制、

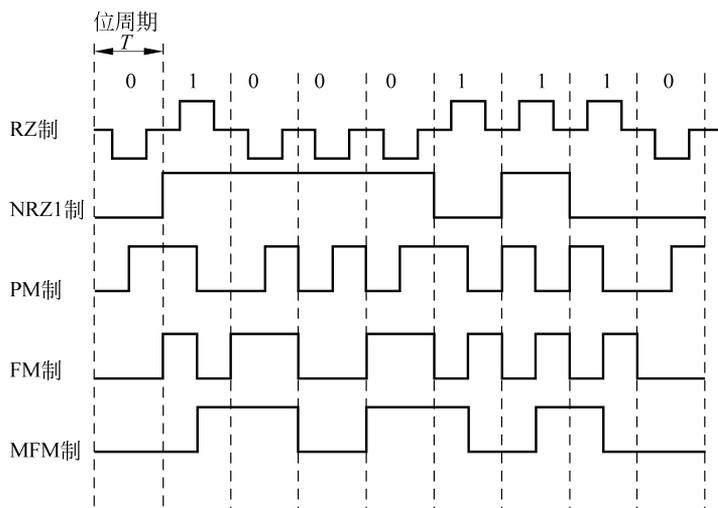


图 5.31 几种磁记录方式的写入电流波形

NRZ1 制、MFM 制存储一位信息时磁层磁化的翻转次数最多为一次,因此这 3 种记录方式的编码效率为 100%。

当然,记录方式的取舍和评价还受到一些其他因素的影响,例如读出信号的分辨能力、频带宽度、抗干扰能力以及编码译码电路的复杂程度等。

除了上面介绍的记录方式以外,还有二次改进的调频制 M2FM、群码制 GCR、游程受限码 RLLC 等记录方式,它们已广泛应用于高密度磁带和磁盘中。群码制,如 GCR(5,4)是将待写入的信息序列按 4 位长度进行分组,然后按某一确定规则将 4 位信息编码为 5 位二进制码,把这对应的 5 位编码按 NRZ1 制记录方式写入磁层中,读出时再把读出的编码字序列进行译码,还原出原始的 4 位信息。采用这种记录方式可以使磁带机存储密度提高到 6250 位/in(bpi)。

游程受限码 RLLC 实际上是 0 或者 1 连续出现时位数受限制的记录序列。如 0 游程长度受限码的编码规则是任何两位相邻的“1”之间的“0”的最大位数  $k$  和最小位数  $d$  均受到限制,然后用 NRZ1 制进行记录。正确地设计  $k$ 、 $d$  的值,可以获得更好的编码性能。

## 5.6.2 磁盘存储器

磁表面存储器包括磁带存储器和磁盘存储器。

磁带存储器是磁表面存储器中最早出现的一个,是典型的顺序存储的外存储器。它的寻址时间最长,信息传输率低,主要用于大、中型计算机系统中磁盘存储器的后备存储器。

磁盘存储器根据盘体材料可分为软磁盘和硬磁盘。软磁盘(Floppy Disk)使用的盘体材料一般是用特殊塑料(聚酯薄膜)作基体,表面涂上磁性物质,装在保护套中,通常按照直径分类,可以分为 3.5in 盘和 5.25in 盘。软盘存储器体积小、重量轻,在 PC 早期是使用最广泛的一种可以脱机保存的外存储器。软盘驱动器的磁头与盘面是在接触状态下工作,因而转速很低,读取速度偏慢,其他工作原理与硬磁盘相类似。但由于近年来,体积更小、容量更大、携带更方便的存储器——闪存的迅速发展与广泛应用,软磁盘存储器已经退出了 PC 的基本配置。而硬磁盘存储器采用的盘体材料通常是铝合金或者改性陶瓷——玻璃基片制

成的刚性盘片,所以称为硬磁盘(Hard Disk)。由于硬磁盘在读写速度、容量、价格比仍然占有很大的优势,因此依旧是大容量辅助存储器的主体。

### 1. 硬盘存储器的分类

硬盘存储器有很多种类型,根据不同的标准可以有不同的分类。

(1) 根据磁头运动与否可分为固定磁头硬盘和活动磁头硬盘。在固定磁头硬盘中,每个磁道都装有一个磁头。工作时,磁头不需要做径向移动,因而存取速度快,省去了磁头寻道的时间。但是磁头多,磁头结构和安装相当复杂,而且磁盘的道密度不可能很大,从而使得整个磁盘造价较高,目前基本已经被淘汰。在活动磁头硬盘中,每个记录盘面上只有一个磁头,安装在读写悬臂上。当需要在不同磁道上读写数据时,要有复杂的磁头寻道机构帮助磁头定位,磁头做径向移动,这就增加了寻道时间,所以其存取时间比固定磁头磁盘较长。

(2) 根据盘片是否可以更换可分为可换盘片磁盘和固定盘片磁盘。可换盘片磁盘的盘片可以取下更换,信息可以脱机保存,但要求有超净的使用环境。固定盘片磁盘的盘面不能从驱动器中取出,例如广泛使用的温切斯特盘就属于这种。

(3) 根据盘片直径大小可分为 5.25in、3.5in、2.5in、3in 和 1.8in 等几种。随着计算机系统的不断小型化,硬盘也在朝着小体积、大容量的方向发展。

另外还可以根据移动性将硬盘分为固定硬盘和活动硬盘(移动硬盘)。

### 2. 硬盘存储器的结构

硬盘存储器的硬件主要包括硬盘控制器、硬盘驱动器以及连接电缆。硬盘控制器(HDC)的功能是接收由主机发来的命令,将它转换为磁盘驱动器的控制命令,实现主机和驱动器之间的数据格式转换和数据传送,并控制驱动器的读写。硬盘控制器主要以适配卡的形式插在主板总线插槽上或直接集成在主板上,然后通过电缆与硬盘驱动器相连。许多新型硬盘则已将控制器集成到驱动器单元中了。硬盘驱动器(HDD)主要由盘片组、磁头、主轴电机(盘片旋转驱动部件)、磁头驱动定位机构、读写电路和控制逻辑等组成,一般置于主机箱内。

温切斯特技术诞生于 20 世纪 70 年代中期,由 IBM 公司位于美国加利福尼亚州坎贝尔市温切斯特大街的研究所研制,并于 1973 年首先应用于 IBM 3340 硬磁盘存储器中,它是硬磁盘向高密度、大容量发展的产物。目前使用的硬盘驱动器大都是采用温切斯特技术的,简称温盘。

它的主要特点有:

(1) 磁盘片组、磁头、主轴以及装载读写磁头臂的小车等机械精度要求严格的关键零部件安装在一个密封的壳体内(称为头盘组件)。这样可以消除影响磁头定位精度的一些机械变动因素。

(2) 磁头采用接触启停式。接触启停是指在读写操作时磁头浮空,不与盘面记录区域相接触,以免划伤记录区。但由于磁头的浮起要依靠盘片高速旋转时产生的气流浮力,因此在启动前和停止后,磁头仍与盘面接触。在盘面记录区和轴心之间有一段空白区,被当作启停区或着陆区。启动前和停止后,磁头停在启停区,与盘面接触。当盘面旋转并达到额定转速时,表面气流的浮力使磁头浮起并达到所需的浮动高度,这时磁头才向盘片存储数据的区域移动。在读写时,磁头与盘面之间的间隙(又称飞高)极小,仅有  $0.2\sim 0.5\mu\text{m}$ ,甚至仅有  $0.005\sim 0.01\mu\text{m}$ ,只相当于人类头发直径的  $1/1000$ 。飞高距离越小,磁头在接通写电流后,

盘面的磁化单元越小,记录密度可以大大提高。

(3) 读前置放大器、写电流驱动电路、磁头选择电子开关以及保护电路等都集成在一块芯片上,该芯片安装在磁头臂上,可尽量减少外界电磁场对磁头引线的影响,从而改善了读写信号的高频传输特性。

采用温切斯特技术的磁盘具有防尘性好、可靠性高、对使用环境要求低的优点,是目前应用最为广泛的硬盘存储器。

### 3. 硬盘的信息存储与磁盘地址

以固定盘片、活动磁头硬盘为例,信息在硬盘上的存储呈现以下层次:记录面、柱面(磁道)和扇区,如图 5.32 所示。

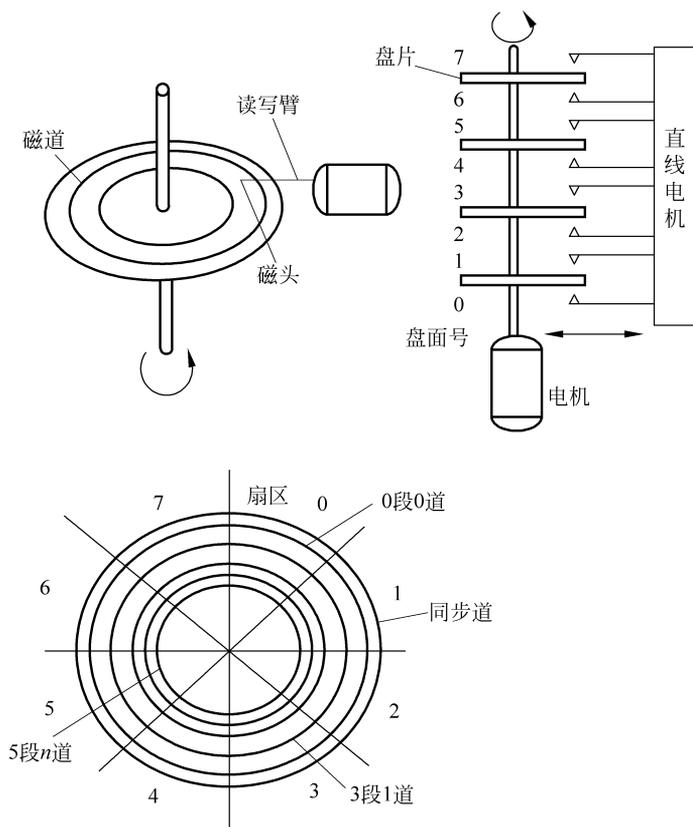


图 5.32 活动头磁盘的信息存储分布示意图

#### 1) 记录面

一台硬盘驱动器中有若干盘片,每个盘片的上下两个盘面(Side)都能记录信息,通常将盘片的表面叫作记录面。每个记录面对应一个读写磁头,所以记录面号就是磁头号。通常是按顺序从上到下从“0”开始顺序编号。一般硬盘的盘片组有 2~3 个盘片,所以盘面号(柱面号)为 0~3 或 0~5。由于所有的磁头都固定在同一个磁头臂上,沿盘面径向一起移动,因此单个磁头不能单独移动。

#### 2) 磁道和柱面

在记录面上,一系列同心圆轨迹称为磁道(Track),每个盘面通常有几十到几百个磁

道,新式大容量硬盘每面的磁道数更多。磁道的编址是从外向内从“0”开始顺序编号,最外圈的磁道为0道,往内依次增加,最里面的磁道假设是第 $n$ 道, $n$ 磁道里面的圆面积是启停区,不能存储信息。

在一个盘组中,各记录面上相同编号的磁道构成一个圆柱面,则柱面(Cylinder)号就等于对应的磁道号。

引入“柱面”这一概念主要是为了提高硬盘的读写速度。当主机要存入一个较长的文件时,若超过了一条磁道的容量,就要存放在几条磁道上。这时应选择位于同一记录面的磁道,还是选择位于同一柱面的磁道?很显然,如果选择同一记录面的不同磁道,每次换道时磁头都要重新进行定位,速度较慢。如果选择同一柱面的不同磁道,此时所有的磁头都已经定位好了,只需要重新选择磁头,不需要再寻道,而这个时间相对于定位操作寻道的时间可以忽略不计。所以在存入文件时,应该首先将一个文件尽可能地放在同一圆柱面上。如果同一圆柱面存不下,则存入相邻的柱面中。

一块硬盘驱动器的柱面数多少既取决于每条磁道的宽窄(同样,与磁头大小也有关系),也取决于定位机构所决定的磁道间步距的大小。

### 3) 扇区

通常一条磁道被划分为若干个段,每个段称为扇区(Sector)或扇段,每个磁道上的扇区数一般是相同的,每个扇区存放的信息一般是定长的,包括512字节的数据和一些其他信息。一个扇区有两个主要部分:存储数据地点的标识符和存储数据的数据段。扇区从“1”开始编号,每个扇区中的数据作为一个单元同时读出或写入。在一个闭合磁道内信息的组织格式称为磁道记录格式,简称磁道格式。磁道记录格式与操作系统有关。

为了进行读写操作,主机要向磁盘控制器发出寻址信息。磁盘的寻址信息一般为:柱面号(磁道号)、记录面号(磁头号)、扇区号。如果主机通过磁盘控制器连接的硬盘驱动器不只一台,还需要有驱动器号或台号。调用磁盘一般以文件为单位,所以寻址信息通常需要给出文件起始位置所在的柱面号与记录面号(从而确定具体磁道位置)、起始扇区号,并给出扇区数(交换量)。

上面所讲的采用柱面、磁头和扇区的方式对磁盘进行寻址,也就是经典的CHS(Cylinder/Head/Sector)寻址,这3个参数能唯一确定磁盘上的数据区域。经典的CHS寻址采用的是24位格式,其中柱面数为10位数据,磁头数为8位,扇区数为6位,每个扇区的大小为512字节,因此从理论上讲,当时磁盘容量的极限为

$$2^8 \times 2^{10} \times 2^6 \times 512\text{B} = 8\text{GB}$$

显然,磁盘的容量目前早已超过了8GB,于是业界采用了新的LBA(Logical Block Addressing,逻辑块寻址)模式。在LBA模式中,地址不再表示为实际磁盘的实际物理地址(柱面、磁头和扇区),LBA模式将CHS寻址转换为一维的线性寻址,将整个硬盘上所有的扇区按顺序编号,比如0道0面1扇区编号为0逻辑块,0道0面2扇区为1逻辑块……以此类推,由此系统效率得以大大提高。在访问磁盘时,由磁盘控制器再将这种逻辑地址转换为实际磁盘的物理地址,具体转换公式如下:

$$\text{LBA} = C \times \text{面数} \times \text{每面磁道数} + H \times \text{每磁道扇区数} + (S - 1)$$

此外,由于硬盘主轴的工作方式都是恒定角速度(Constant Angular Velocity,CAV),而盘片最外圈的周长比最内圈的周长要长很多,磁头在最外圈时,虽然放置的角度与在最内

圈时一样,但走过的距离就长多了。这样,如果最内圈与最外圈磁道的扇区数相同,必将造成极大的存储空间的浪费。为此,硬盘厂商们开发了区域数据记录(Zone Data Recording, ZDR)技术,即从磁盘的最外圈开始划分出若干个区域,每个区域内的每磁道扇区一致,但靠内侧的区域比外侧的区域的每磁道扇区数要少,从而可以根据不同的磁道长度来合理设定扇区的数量,以达到充分利用磁盘存储空间的目的。大多数产品划分了16个区域,最外圈的每磁道扇区数正好是最内圈的一倍,与最大的持续传输率的参数基本成比例。

#### 4. 硬磁盘的主要性能指标

##### 1) 磁盘记录密度

磁盘记录密度分为道密度和位密度。**道密度**是指沿盘片半径方向单位长度内的磁道数,单位为道/毫米或道/英寸。道密度取决于磁头的大小和磁头轴向进退的控制精度。目前,每厘米磁盘可以有800~2000个磁道,即磁道宽度为5~10 $\mu\text{m}$ 。**位密度**是指磁盘磁道上单位长度内存储的二进制位数。一般指内圈磁道,单位为位/毫米或位/英寸(BPI)。它主要取决于磁表面的纯度和空气质量。目前,磁盘能达到的位密度是5000~10 000位/mm,约为道密度的50倍。

##### 2) 存储容量

存储容量是指一个硬盘装置所能存放的二进制信息总量。存储容量分格式化容量和非格式化容量。**格式化容量**是指按特定记录格式所能存储的信息总量,也就是用户可以使用的容量。**非格式化容量**是磁记录表面可以利用的所有磁化单元。在计算机系统中使用磁盘必须要先格式化。格式化容量一般是非格式化容量的80%~90%。以单面存储容量计算为例:

$$\text{非格式化容量} = \text{位密度} \times \text{内圈周长} \times \text{磁道总数}$$

$$\text{格式化容量} = \text{扇区容量} \times \text{每道扇区数} \times \text{磁道总数}$$

硬盘的容量通常以GB即吉字节为单位,1GB=1024 $\times$ 1024 $\times$ 1024B,但大部分硬盘厂家在为其硬盘标注容量时大多取10<sup>9</sup>字节为1GB,因此测试值往往小于标称值。目前广泛使用的大容量硬盘通常使用TB即太字节为单位。

##### 3) 平均存取时间

**平均存取时间**是指磁头找到指定数据的平均时间,通常它是硬盘平均寻道时间和平均等待时间之和。平均存取时间最能代表硬盘找到某一数据所用的时间,数值越小越好。**平均寻道时间**与磁头移动速度和盘径直接相关,目前平均寻道时间已缩短到9ms以内。**平均等待时间**与磁盘转速有关,可以用磁盘旋转一圈所需时间的1/2表示。例如磁盘转速为5400r/min,相应地,磁盘平均等待时间约为5.6ms。

##### 4) 数据传输率

硬盘的**数据传输率**(Data Transfer Rate)也称吞吐率,它表示在磁头定位后,硬盘读或写数据的速度。数据传输率分为外部传输率和内部传输率,两者之间有一块缓冲区以平滑硬盘内部与接口数据之间的速度差距(其构成主要是DRAM芯片,芯片容量目前为16MB、32MB、64MB不等)。硬盘数据传输率表现出硬盘工作时数据的传输速度,是硬盘工作性能的具体表现。它并不是一成不变的,而是随着工作的具体情况而变化的。在读取硬盘时,不同磁道、不同扇区的数据以及数据存放是否连续等因素都会影响到硬盘的数据传输率。

**外部数据传输率**指的是主机通过数据总线从硬盘内部缓存区中读取数据的最高速率,也叫**突发数据传输率**或**接口传输率**,即系统总线与硬盘缓冲区之间的数据传输率,与硬盘的最高外部数据传输率和最高接口数据传输率是一个概念。外部数据传输率主要与硬盘接口类型和硬盘缓冲区(硬盘 Cache)容量大小有关。目前支持 ATA-133 的硬盘外部数据传输率可达 133MB/s,而 SATA 接口的硬盘外部理论数据最大传输率可达 300MB/s。这些只是硬盘理论上最大的外部数据传输率,在实际的日常工作中是无法达到这个数值的。

**内部数据传输率**指磁头至硬盘缓存间的最大数据传输率,也叫**持续数据传输率**,反映硬盘缓冲区未用时的性能。简单地说就是硬盘将数据从盘片上读取出来,然后存储在缓存内的速度。内部传输率可以明确表现出硬盘的读写速度,它的高低才是评价一个硬盘整体性能的决定性因素,它是衡量硬盘性能的真正指标。有效地提高硬盘的内部传输率才能对磁盘子系统的性能有最直接、最明显的提升。目前提高硬盘的内部传输率的途径,除了改进信号处理技术、提高转速以外,最主要的就是不断地提高单碟容量以提高线性密度。由于单碟容量越大的硬盘线性密度越高,磁头的寻道频率与移动距离可以相应地减少,从而减少了平均寻道时间,内部传输速率也就提高了。在单碟容量相同时,转速高的硬盘的内部传输率高。虽然硬盘技术发展的很快,但内部数据传输率还是在一个比较低(相对而言)的层次上,内部数据传输率低已经成为硬盘性能的最大瓶颈。内部数据传输率一般取决于硬盘的盘片转速和盘片数据线密度(指同一磁道上的数据间隔度)。

假设磁盘转速为  $n$  r/s,每条磁道的容量为  $W$  位,则内部数据传输率可以表示为  $IDTR = W \cdot n$  (b/s)或  $IDTR = D \cdot V$  (b/s),其中  $D$  为位密度, $V$  为磁盘旋转的线速度。

**例 5.3** 一台有 3 个盘片的磁盘组,共有 4 个记录面,转速为 7200r/min。盘面记录区域的外直径为 30cm,内直径为 4cm,记录位密度为 210b/mm,磁道密度为 8 道/mm,盘面分为 16 个扇区,每个扇区 1024B,设磁头移动速度为 2m/s。

- (1) 试计算该盘组的非格式化容量和格式化容量。
- (2) 计算该磁盘的数据传输率、平均寻道时间和平均旋转等待时间。
- (3) 若一个文件超出了磁道的容量,余下的部分是存于同一盘面上还是同一柱面上? 请给出一种合理的磁盘地址方案。

**解:**

(1) 单面非格式化容量 = 位密度  $\times$  内圈磁道周长  $\times$  磁道总数

磁道总数 = 道密度  $\times$  [(外直径 - 内直径) / 2]

$$= 8 \times (300 - 200) \text{道} / 2 = 200 \text{道}$$

非格式化容量 = 4 面  $\times$  210b/mm  $\times$  ( $\pi \times 200$ mm)  $\times$  400

$$= 211\,008\,000\text{b} = 26\,376\,000\text{B}$$

$$= 26.376\text{MB}$$

格式化容量 = 扇区容量  $\times$  每道扇区数  $\times$  磁道总数  $\times$  面数

$$= 1024 \text{字节/扇区} \times 16 \text{扇区} \times 400 \text{道} \times 4 \text{面}$$

$$= 26\,214\,400\text{B}$$

$$= 26.2144\text{MB}$$

(2) 数据传输率 = 扇区容量  $\times$  每道扇区数  $\times$  磁盘转速

$$= 1024 \times 16 \times 7200 / 60 = 1\,966\,080\text{B/s} = 1.966\,08\text{MB/s}$$

$$\begin{aligned}
 \text{平均寻道时间} &= (\text{最大寻道时间} + \text{最小寻道时间}) / 2 \\
 &= \text{磁头从 0 磁道移动到一半位置的时间} \\
 &= [(\text{外圈半径} - \text{内圈半径}) / 2] / \text{磁头移动速度} \\
 &= [(15\text{cm} - 10\text{cm}) / 2] / 2(\text{m/s}) = 2.5\text{s} / 200 = 0.0125\text{s} \\
 &= 12.5\text{ms}
 \end{aligned}$$

$$\begin{aligned}
 \text{平均等待时间} &= \text{磁盘旋转一周所需时间的一半} \\
 &= 1/2 \times (1/\text{转速}) = 1/2 \times (60/7200)\text{s} = 4.17\text{ms}
 \end{aligned}$$

(3) 若一个文件超出一个磁道容量,余下的部分应存于同一柱面上。磁盘地址方案如下:

柱面号(9 位)	磁头号(2 位)	扇区号(4 位)
----------	----------	----------

### 5.6.3 磁带存储器

磁带存储器是将数据信息以磁记录方式记录在磁带上的存储设备。磁带存储器的优点是存储量大、磁带体积小、可反复读写、可靠性高、可脱机保存、成本低,但磁带机采用顺序存取,存取时必须从头开始存取,速度慢,所以常用作计算机系统的后备存储设备。可以将硬盘上大量暂时不用的数据成批转存于磁带,当需要使用时,再成批转还到硬盘,常用于数据备份。磁带存储器由磁带和磁带机两部分组成。

#### 1. 磁带

磁带存储器所用的存储介质是磁带,采用涤纶(聚酯树脂)材料做带基,表面涂上一层  $\text{Fe}_2\text{O}_3$  和  $\text{CrO}_2$  磁性材料。磁带按宽度划分有 0.15in、0.25in、0.5in 等规格,按带长划分有 2400ft、1200ft 和 600ft 等规格;按外形分有开盘式和盒式磁带;按记录密度分有 800bpi、1600bpi、6250bpi。磁带盒有圆形盒、正方形盒和长方形盒 3 种,磁带上的磁道数(轨)有 2、4、7、9、11 和 18 道等,常用的是 9 道。计算机系统中一般采用 1/2in 开盘式磁带和 1/4in 盒式磁带。

#### 2. 磁带机

磁带机也有很多种,按磁带机规模分有标准 0.5in 磁带机、海量宽带磁带机和盒式磁带机 3 种。按磁带机走带速度分,有高速磁带机(4~5m/s)、中速磁带机(2~3m/s)和低速磁带机(2m/s 以下)。按磁带装卸机构分,有手动装卸式和自动装卸式。按磁带传动缓冲机构分,有摆杆式和真空式。按磁带的记录格式分,有启停式和数据流式。数据流磁带机将数据连续地写在磁带上,在数据块之间插入记录间隙,这样磁带机在数据块之间不用启停。此外它采用了电子控制代替机械控制,简化了磁带机的机械结构,降低成本,提高了可靠性。而且数据流磁带机采用的是和磁盘一样的串行读写方式,而不是多位并行读写,因而它的记录格式和启停式磁带机不同。目前数据流磁带机已成为现代计算机系统中主要的后备存储器,其位密度可达到 8000bpi。它主要用于资料保存、文件复制,作为脱机后备存储装置,特别是当硬盘出现故障时用以恢复系统。

#### 3. 磁带的记录格式

磁带上的信息可以按文件形式存储,也可以按数据块存储。磁带可以在数据块之间启

停。磁带机与主机之间进行信息传送的最小单位是数据块或称为记录块(Block),记录块的长度可以是固定的,也可以是变化的,由操作系统决定。记录块之间有空白间隙,用作磁头停靠的地方,并保证磁带机停止或启动时有足够的惯性缓冲。图 5.33 给出了以数据块形式存储的磁带数据的记录格式。数据开始有一个始端标记 BOT,数据尾部以末端标记 EOT 结束,这两个标记用反光的金属薄膜制成,光电检测元件可以检测出这两个标记并产生相应的电信号。同时在每个记录块尾部还有校验码区,磁带信息的校验属于多重校验,由奇偶校验、循环冗余校验和纵向冗余校验共同完成。

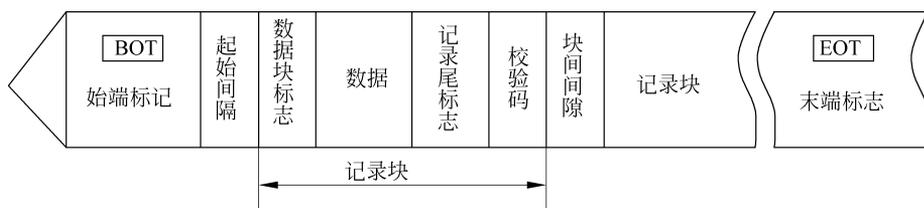


图 5.33 磁带数据记录格式

#### 4. 磁带机的性能指标

磁带机的主要性能指标有:

(1) 磁带宽度。磁带宽度与磁道数有关,在磁道密度相同的条件下,磁道宽度越大的磁带磁道数越多,则存储容量越大。

(2) 磁道数。标准磁带机在磁带宽度方向上并行排列多个磁头,每个磁头对应一个磁道。

(3) 记录密度。磁道上单位长度能存储的数据位数,以 bpi(位/英寸)为单位,一般为每英寸几千到几万。

(4) 存储容量。一盒磁带所能存储的最大容量,一般为几十兆到几十吉。

(5) 磁带速度。磁带在单位时间内相对于磁头移动的长度。

(6) 数据传输率。磁带机向主机提供数据的速度,分为持续传输率和猝发传输率两种,持续传输率可达 30~100KB/s,猝发传输率可达 1.5MB/s。数据传输率除了与记录密度和磁带速度有关外,还取决于磁带机与主机的接口。

#### 5.6.4 光盘存储器

光盘存储器是指采用聚集激光束在盘形介质上高密度地记录信息的存储装置,目前主要由光盘、光盘驱动器和光盘控制器组成。其中光盘(Optical Disk)是指用光学方式进行读出或写入信息的盘片,现在一般称为 CD(Compact Disc);光盘驱动器是读写光盘的基本设备;光盘控制器是光盘驱动器的控制电路。

利用激光在某种介质上写入信息,然后再利用激光读出信息的技术称为光存储技术。如果这种介质是磁性材料,这种存储就称为磁光存储。20 世纪 60 年代开发出的半导体激光技术,使得激光通过聚焦后,可获得直径约为  $1\mu\text{m}$  的光束。20 世纪 70 年代研究出的半导体激光器、读出微小信息位的光学伺服系统等关键技术问题的解决,使得光存储技术迅速达到了实用化、市场化。另外由于光盘存储器利用激光束在记录面上存储信息,根据激光束及反射光的强弱不同可以完成信息的读写,属非接触型存储器,因此光盘存储器具有对介质无损伤、存储容量大、记录密度高、信息保存寿命长、工作稳定可靠、环境要求低等优点。

## 1. 光盘存储器的分类

根据性能和用途的不同,光盘存储器可分为 CD-ROM、CD-R 和 CD-RW。

### 1) CD-ROM

这种光盘的盘片是由生产厂家预先用激光蚀刻写入数据或程序,信息只能读取,不能写入修改。一张 CD-ROM 约可以存储 650MB 的数据。

### 2) CD-R

这种光盘可由用户写入信息,写入后可以多次读出,但是只能写入一次,信息写入后不能修改。因此被称为“写一次型”(Write Once, Read Many, WORM)。如果有要修改的数据,只能追加在盘片上的空白处,故又称为“追记型”光盘。目前主要用于计算机系统中文件的存档或写入的信息不需要修改的场合。

### 3) CD-RW

这种光盘是可以写入、擦除、重写的可逆性记录系统。从原理上来看,目前仅有光磁记录(热磁翻转)和相变记录(晶态-非晶态转变)两种。

## 2. 光盘的读写原理

光盘存储器是利用激光束在记录表面上存储信息,根据激光束及反射光的强弱不同,可以完成信息的读写。信息的记录原理有形变、相变和磁光存储等。

### 1) 形变型光盘读写原理

只读型和只写一次型光盘写入的时候,将激光束聚焦成直径小于  $1\mu\text{m}$  的光点,以其高热效应,融化记录介质表面的薄膜,在薄膜上形成永久性的小凹坑,所以只能写入一次,不能擦除和改写,假设有凹坑的位置表示记录了“1”,没有凹坑的位置则表示“0”。若是只写一次性光盘,到此写入操作完成。若是只读型光盘,这时得到的只是主盘,或称母盘,由母盘取反得到只读光盘的副盘,也称“模”,再由模取反就能得到大量与母盘相同的只读型光盘了。读出时,在一个比写入功率低(约几毫瓦)的激光束的照射下,有凹坑的地方和没有凹坑的地方光反射的强度和方向是不同的,一般有凹坑处的反射光弱,无凹坑处的反射光强,由此可以读出二进制信息。由于读出激光束的功率仅为写入激光束功率的  $1/10$ ,因此不会融出新的凹坑。

### 2) 相变型光盘读写原理

有些光存储介质在激光的照射下,晶体结构会发生变化。利用存储介质处于晶体和非晶体状态可逆转换,引起对入射激光束不同强度的反射(或折射),形成信息一一对应的关系,这种光盘称为相变光盘,如有些 CD-RW。

写入信息时,利用高功率的激光聚焦于记录介质表面的一个微小区域内,使处于晶体状态的介质吸热后至熔点,并在激光束离开的瞬间骤冷转变为非晶态,信息即被写入。

读出信息时,由于晶态和非晶态对入射激光束存在不同的反射和折射率,利用已记录信息区域的反射与周围未发生晶态改变区域的反射之间存在明显差异的效应,读出已写入的信息。

擦除信息时,利用适当波长和功率的激光束作用于记录信息点,使该点温度介于材料的熔点和非晶态转变温度之间,使之重新结晶而恢复到晶态,完成擦除功能。

### 3) 磁光型光盘读写原理

磁光型光盘利用激光在磁性薄膜上产生热磁效应记录信息,利用磁光效应读取信息。磁光存储应用于可擦写光盘上,其读写原理如下:

根据热磁效应,如果磁记录介质的温度低于某一温度(280℃左右)时,在强度低于介质矫顽力的外磁场作用下,该介质不会发生磁通翻转,也就不能记录信息。但是随着温度升高,介质的矫顽力降低,在外加磁场的帮助下,磁记录介质将发生磁通翻转。磁光存储就是根据这一原理存储信息的。它利用激光照射磁性薄膜,被照射处温度上升,矫顽力下降,在外加磁场的帮助下发生磁通翻转,使该处的磁化方向与外加磁场一致,这就写入了信息,假设为状态“1”,则不被照射处或者外磁场强度小于矫顽力处可视为状态“0”。

读出时用比写入更弱的激光照射记录面,根据检测到的光的偏振方向判断读出的是“1”还是“0”。

擦除信息和写入信息的原理一样,外加一个与记录方向相反的磁场,对已写入信息的介质用激光照射,使照射区反方向磁化,从而恢复到记录前的磁化状态。

### 3. 光盘存储器的组成

光盘存储器由光盘片、光盘控制器、光盘驱动器及接口组成。

光盘片是指整个盘片,包括光盘的基片和记录介质。基片一般采用聚碳酸酯晶片制成,是一种耐热的有机玻璃。记录介质的化学成分随光盘的种类不同而存在差异。在光盘上程序和数据文件按内螺旋线的规律顺序存放,如图 5.34 所示。由于不能像磁盘存储器那样读取文件的每个扇区,因此读出速度较慢。

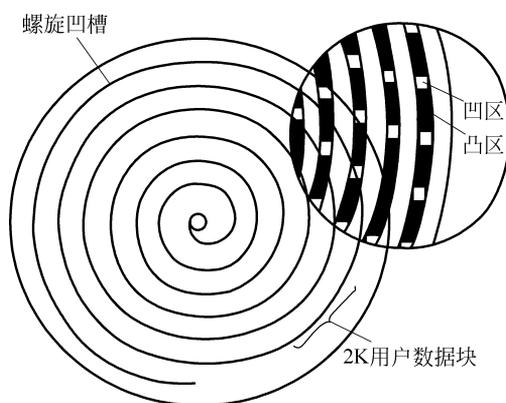


图 5.34 光盘和只读光盘的记录结构

光盘控制器主要包括数据输入缓冲器、记录格式器、编码器、读出格式器和数据输出缓冲器等部分。

光盘驱动器主要由读写光头、寻道定位机构、主轴驱动机构以及光学系统组成。图 5.35 所示的光学系统中,激光源产生的光束经过光束分离器后,90%的光束用作写入光束,10%的光束作为读出光束。写入光束经记录信息调制后,由聚焦系统射向记录介质记录信息。而读出光束经过几个反射镜后射到盘面上,读出的光信号再经光敏二极管转换为电信号输出。

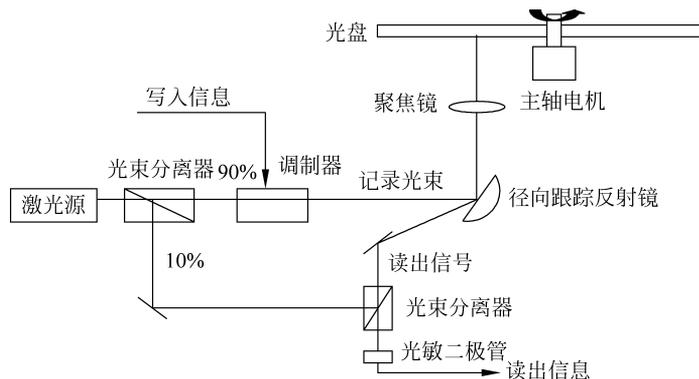


图 5.35 写一次型光盘光学系统示意图

#### 4. 光盘存储器的主要技术指标

##### 1) 数据传输率

数据传输率是指将数据从光盘驱动器传送到主存的速率,即单位时间内光盘的光道上传送的数据位数。这与光盘的转速、存储密度有关。光盘转得越快,数据从光盘传送到主机内存的速度就越快。单倍速光驱的数据传输率是 150kb/s。12 倍速(记为 12X)光驱的数据传输率是 1.8Mb/s,其光盘外圈的转速是 2400r/min,内圈转速是 6360r/min。CD-R 驱动器的写速度和读速度是不一样的,例如标示为 24X/40X 的 CD-R 光驱,指其读信息的速度是 40 倍速,而写信息(刻盘)的速度只有 24 倍速。一个 CD-RW 驱动器有 3 种速度。例如标示为 24X/12X/40X 的 CD-RW 光驱,其写入速度是 24 倍速,重写的速度为 12 倍速,读的速度是 40 倍速。

##### 2) 存储容量

光盘存储容量是指所读写的光盘盘片的容量。光盘容量又分为格式化容量和用户容量。采用不同的格式和不同的驱动器,光盘格式化后的容量不同。如 650MB 的 CD-ROM 盘片,螺旋线形的光道被划分成一个个扇区,扇区是最小的记录单位。每个扇区可存放 2048B 的有效数据。每个扇区的地址被标记为“分、秒、扇区”,每秒的数据需要 75 个扇区存放,一张光盘可存储 74min 的数据,所以整张光盘的容量为

$$74\text{min} \times 60\text{s}/\text{min} \times 75 \text{ 扇区}/\text{s} \times 2048\text{B}/\text{扇区} = 681\,984\,000\text{B} \approx 650\text{MB}$$

##### 3) 平均存取时间

平均存取时间是指从计算机向光盘驱动器发出命令开始,到光盘驱动器在光盘上找到读写的信息的位置并接收读写命令为止的一段时间。平均存取时间等于平均寻道时间和平均等待时间之和。光头沿半径移动全程 1/3 长度所需的时间为平均寻道时间,盘片旋转半周的时间为平均等待时间。目前大多数光盘驱动器的平均存取时间为 200~400ms。

##### 4) 接口类型

光盘驱动器的接口是驱动器与系统主机的物理连接,它是从驱动器到计算机的数据传输途径,不同的接口也决定着驱动器与系统间数据传输速度。早期的光驱产品采用过一些专用接口,如索尼、美上美、松下等光驱厂商,都开发了本公司的专用光驱接口。由于此类接口之间互不兼容,已逐渐被市场淘汰了。目前连接光盘驱动器与系统接口的类型主要有 IDE 和 SCSI 接口,当然也有采用 USB 等其他一些接口类型的。

#### 5. DVD

DVD 即 Digital Video Disc(数字视盘)或 Digital Versatile Disc(数字多功能光盘)。DVD 碟片的尺寸大小和普通 CD 一样,但却提供了比 CD 大得多的存储容量。

DVD 采用与 CD 类似的技术,两种盘具有相同的尺寸。CD-ROM 最多可以容纳 737MB 的数据,而 DVD 的单面盘就可以存储 4.7GB(单层)~8.5GB(双层)的数据,是 CD-ROM 的 11.5 倍。DVD 利用 MPEG-2 标准进行压缩后,在单面单层光盘上可存放 133min 的视频信息,单面双层光盘可存放 240min 以上的视频数据。

DVD 每面可以有两层来刻录数据,每一层单独压制,然后结合到一起形成 1.2mm 厚的光盘。与 CD 一样,DVD 每一层都以单一的螺旋形路径形式印制,从光盘的最里端开始向外环绕。螺旋形路径上包含与 CD 相同的凹痕和平地。每一层都覆盖一层反射激光的金属膜;外层的金属膜较薄,以便激光穿过它读取里层的数据。DVD-ROM 的读取过程与

CD-ROM 相似,只是 DVD 采用了波长更短的激光束来读取数据。

在可写式 DVD 的刻录标准之战中,由于未能统一各方意见,最后出现了 DVD-R/RW、DVD-RAM 和 DVD+R/RW 共 3 大规格。目前使用的光盘采用的是红色激光,而下一代 DVD 标准,将采用蓝色激光。根据未来高清晰视频节目对画面质量的要求,现有的 DVD 已无法满足要求。为了争夺蓝光光盘标准的制定权,分别支持 Blue-ray Disc 标准和 HD DVD 标准的厂商形成了两大阵营。蓝光光盘的最大优势在于存储容量的大幅提升。Blue-ray 标准可以使单层盘片容量达到 27GB,HD DVD 标准单层盘片的容量也将达到 15GB,大大高于目前 DVD 盘片的 4.7GB。在技术上,Blue-ray 标准采用比较先进的技术规范,可获得更大的存储容量,但该规范与现有的 DVD 不兼容,需要更新生产设备,从而使整体成本过高。HD DVD 标准在容量上不如 Blue-ray 标准,但可利用现有的 DVD 生产线进行制造,从而使成本相对较低。

## 习 题

- 5.1 什么是存储位元? 作为存储位元必须具备的条件有哪些?
- 5.2 存储周期  $T_m$  和存取时间  $T_a$  的基本概念分别是什么? 哪个的时间间隔比较长? 为什么?
- 5.3 比较 SRAM 存储位元和 DRAM 存储位元的异同。
- 5.4 DRAM 为什么需要刷新? 刷新的方式有哪几种?
- 5.5 ROM 有哪几种类型? 各有什么特点?
- 5.6 设有一个具有 14 位地址和 8 位数据的存储器,试问:
  - (1) 该存储器的存储容量是多少位?
  - (2) 如果该存储器用  $1K \times 1$  的存储芯片构成,需要多少片?
  - (3) 需要用多少位地址做片选信号译码的地址?
- 5.7 用  $16K \times 4$  的 SRAM 芯片构成  $64K \times 16$  的存储器。
  - (1) 该存储芯片的数据线和地址线的位数各是多少?
  - (2) 画出该存储器的结构框图。
- 5.8 假定用若干个  $2K \times 4$  位芯片组成一个  $8K \times 8$  位存储器,则  $0B1FH$  所在芯片的最小地址是多少(用十六进制表示)?
- 5.9 某计算机存储器按字节编址,采用小端方式存放数据。假定编译器规定 `int` 和 `short` 型长度分别为 32 位和 16 位,并且数据按边界对齐存储。某 C 语言程序段如下:

```
struct {  
    int    a;  
    char  b;  
    short c;  
} record;  
record.a = 273;
```

若 `record` 变量的首地址为 `C008H`,请回答:

- (1) `C008H` 存储单元的内容是多少?

(2) record.c 的地址是多少?

5.10 已知  $16\text{K} \times 4$  位 DRAM 芯片的存储矩阵为  $256 \times 256$ , 读写周期为  $0.5\mu\text{s}$ , 该芯片最大刷新间隔时间为  $2\text{ms}$ , CPU 在  $1\mu\text{s}$  内访问主存储器一次, 试问采用哪种刷新方式比较合理? 相邻两行之间的刷新间隔是多少? 对全部存储单元刷新一遍(不含存储器正常访问时间)所需的实际刷新时间是多少?

5.11 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  $T=200\text{ns}$ , 数据总线宽度为 64 位, 总线传输周期  $t=50\text{ns}$ 。问顺序存储器和交叉存储器的带宽各是多少?

5.12 画出代码 11001000 在 NRZ1、FM 制和 PM 制记录方式下的写入电流波形, 试比较它们的磁记录密度, 说明它们有无自同步能力。

5.13 磁盘速度指标有哪几项? 简述其含义。当磁盘转速提高一倍时, 会对哪些速度指标产生怎样的影响? 若磁头驱动速度提高一倍呢?

5.14 某磁盘存储器转速为  $3000\text{r}/\text{min}$ , 共有 8 个记录面, 每毫米 5 道, 每道记录信息为 12 288B, 每个扇区记录 1024B。最小磁道直径为  $230\text{mm}$ , 共有 275 道。问:

(1) 磁盘存储器的容量是多少?

(2) 最高位密度和最低位密度是多少?

(3) 磁盘数据传输率是多少? 平均等待时间是多少?

(4) 若一个文件超出一个磁道容量, 余下的部分是存于同一盘面上还是存于同一柱面上? 请给出一种合理的磁盘地址方案。

5.15 一磁带机有 9 个磁道, 带长为  $900\text{m}$ , 带速为  $2\text{m}/\text{s}$ , 每个数据块为 1KB, 块间间隔为  $10\text{mm}$ 。若数据传输率为  $128\text{KB}/\text{s}$ 。求:

(1) 记录位密度。

(2) 若首尾各空  $2\text{m}$ , 求磁带最大有效存储容量。

5.16 什么是光盘? 光盘有哪些类型?