

存储器技术

【学习目标】

现代计算机存储系统是由速度各异、容量不等的多级层次存储器组成的,存储系统的性能是影响计算机系统性能的一大瓶颈。现代的信息处理,如图像处理、数据库、知识库、语音识别、多媒体等对存储系统的要求很高。如何设计容量大、速度快、价格低的存储系统是计算机系统发展的一个重要问题。本章主要介绍半导体存储器基本知识,存储器芯片及其与 CPU 之间的连接和扩展方法,Cache 高速缓冲存储器和虚拟存储器。本章的学习目标:了解存储系统的体系结构;掌握存储器芯片与 CPU 的连接方法及存储器扩展技术;了解 Cache 高速缓冲存储器和虚拟存储器的作用。

5.1 存储器概述

5.1.1 存储器系统与多级存储体系结构

存储器系统与存储器是两个不同的概念。存储器系统是指计算机中由存放程序和数据的各种存储设备、控制部件及管理信息调度的硬件设备和软件算法所组成的系统。存储器系统的性能在现代计算机中的地位日趋重要,主要原因是:①冯·诺依曼体系结构是建筑在存储程序概念基础上的,访问存储器的操作约占 CPU 时间的 70%;②对存储器系统管理与组织的好坏影响到整个计算机的效率;③现代的信息处理,如图像处理、数据库、知识库、语音识别、多媒体等对存储器系统的要求越来越高。

随着 CPU 速度的不断提高和软件规模的不断扩大,人类总希望存储器能同时满足速度快、容量大、价格低等要求,而采用单一工艺制造的半导体存储器很难同时满足这三方面的要求。为了解决这一矛盾,现代微机系统中普遍采用速度由慢到快、容量由大到小的多级层次存储器体系结构构成的存储器系统。如图 5.1 所示,系统呈现金字塔形结构,越往上存储器件的速度越快,CPU 的访问频度越高,同时系统的拥有量也越小;位于塔底的存储设备,其容量最大,价格最低,但速度相对也是最慢的。

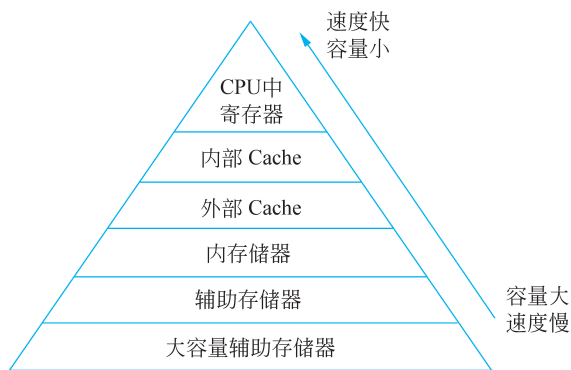


图 5.1 微型计算机存储器系统的多级层次结构

5.1.2 存储器的分类与组成

1. 存储器的分类

存储器的种类繁多,根据存储器的存储介质的性能及使用方法的的不同,可以从不同角度对存储器进行分类。

存储介质是指能寄存“0”和“1”两种代码并能区别两种状态的物质或元器件。按照存储介质的不同,存储器可分为半导体存储器、磁存储器和光存储器。由于半导体存储器具有存取速度快、集成度高、体积小、功耗低、应用方便等优点,因此微型计算机内存多用半导体存储器构成,外存多以磁性材料或光学材料制造。以下介绍半导体存储器的分类。

半导体存储器的分类如图 5.2 所示,其按照存储原理可分为 RAM 和 ROM 两大类。其中,RAM(Random Access Memory,随机存取存储器)按照制造工艺可分为双极型 RAM 和 MOS 型 RAM,其中 MOS 型 RAM 是目前广泛使用的半导体存储器,又可分为静态 RAM(SRAM)和动态 RAM(DRAM)两种。ROM(Read Only Memory,只读存储器)根据其不同的编程写入方式,又可分为掩模 ROM、PROM、EPROM、E²PROM 和闪存存储器(Flash Memory)几种。



图 5.2 半导体存储器的分类

2. 半导体存储器的组成

半导体存储器由存储体、地址译码驱动电路、地址寄存器、读/写驱动器、数据寄存器、读/写控制逻辑 6 部分组成,通过系统数据总线、地址总线和控制总线与 CPU 相连,如图 5.3 所示。

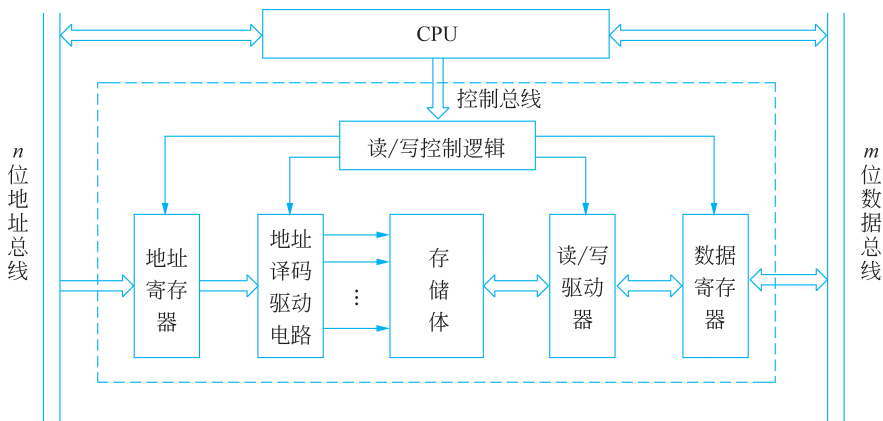


图 5.3 半导体存储器的基本组成

1) 存储体

存储体(也称存储矩阵)是存储器的核心,由若干存储单元组成,每个存储单元又由多个基本存储电路(也称基本存储单元)组成。通常,一个存储单元存放一个 8 位二进制数据。为了区分不同的存储单元和便于读/写操作,每个存储单元都有一个编号,这个编号称为存储单元的地址,CPU 访问存储单元时按地址访问。为了减少存储器芯片的封装引脚数和简化译码器结构,存储体总是按照二维矩阵的形式来排列存储单元电路。存储体内基本存储单元的排列结构通常有两种方式:一种是“多字一位”结构(简称位结构),即将多个存储单元的同位排在一起,其容量表示成 $N \text{ 字} \times 1 \text{ 位}$,如 $1\text{K} \times 1 \text{ 位}$ 、 $4\text{K} \times 1 \text{ 位}$ 等;另一种排列是“多字多位”结构(简称字结构),即将多个存储单元的若干位(如 4 位、8 位)连在一起,其容量表示为 $N \text{ 字} \times 4 \text{ 位/字}$ 或 $N \text{ 字} \times 8 \text{ 位/字}$,如静态 RAM 的 6116 为 $2\text{K} \times 8 \text{ 位}$,6264 为 $8\text{K} \times 8 \text{ 位}$ 。

存储器的最大存储容量取决于 CPU 本身提供的地址线条数,这些地址线的每一位编码对应一个存储单元的地址。因此,当 CPU 的地址线为 n 条时,可生成的编码状态有 2^n 个,也就是说 CPU 可寻址的存储单元个数为 2^n 个。若采用字节编址,那么存储器的最大容量为 $2^n \times 8 \text{ 位}$ 。例如,8086 CPU 的地址线为 20 条,可寻址的最大存储空间为 $2^{20} \text{ B} = 1\text{MB}$,80486 CPU 的地址线为 32 条,可寻址的最大存储空间为 $2^{32} \text{ B} = 4\text{GB}$ 。

2) 地址译码驱动电路

地址译码驱动电路包含译码器和驱动器两部分,译码器的功能是将地址总线输入的地址码转换成与其对应的译码输出线上的高电平(或低电平)信号,以表示选中了某一存储单元,并由驱动器提供驱动电流去驱动相应的读/写电路,完成对被选中单元的读/写操作。

3) 地址寄存器

地址寄存器用于存放 CPU 要访问的存储单元地址,经译码驱动后指向相应的存储单元。通常,微型计算机中访问的地址由地址锁存器提供,例如 8086 CPU 中的地址锁存器 8282。存储单元地址由地址锁存器输出后,经地址总线送到存储器芯片内直接译码。

4) 读/写驱动器

读/写驱动器包括读出放大器、写入电路和读/写控制电路,用于完成对被选中单元中各位的读/写操作。存储器的读/写操作是在 CPU 的控制下进行的,只有当接收到来自 CPU 的读/写命令 $\overline{RD}$ 和 $\overline{WR}$ 后,才能实现正确的读/写操作。

5) 数据寄存器

数据寄存器用于暂时存放从存储单元读出的数据,或从 CPU 或 I/O 端口送出的要写入存储器的数据。暂存的目的是协调 CPU 和存储器之间在速度上的差异,故又称为存储器数据缓冲器。

6) 读/写控制逻辑

读/写控制逻辑接收来自 CPU 的启动、片选、读/写及清除命令,经控制电路综合和处理后,产生一组时序信号来控制存储器的读/写操作。

虽然现代微型计算机的存储器多由多个存储器芯片构成,但任何存储器的结构都保留着这 6 个基本组成部分,只是在组成各种存储器时做了一些相应的调整。

5.1.3 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因素,也是选购存储器的主要依据。衡量半导体存储器性能的指标很多,但从功能和接口电路的角度来看,主要有以下几项。

1. 存储容量

存储容量是存储器的一个重要指标,是指存储器所能容纳二进制信息的总量。容量越大,意味着所能存储的二进制信息越多,系统处理能力就越强。半导体存储器是由多个存储器芯片按照一定方式组成的,所以其存储容量为组成存储器的所有存储芯片容量的总和。

2. 存取速度

存储器的存取速度可以用存取时间和存取周期来衡量。存取时间是指完成一次存储器读/写操作所需要的时间,具体是指存储器接收到寻址地址开始,到取出或存入数据为止所需要的时间,通常用 ns 表示,存取时间越短,存取速度越快;存取周期是连续进行读/写操作所需的最小时间间隔。由于在每一次读/写操作后,都需有一段时间用于存储器内部线路的恢复操作,因此存取周期要比存取时间大。

3. 存储带宽

存储带宽表示单位时间内内存存取的信息量,即吞吐数据的能力,是改善机器瓶颈

的一项关键指标,通常以位/秒或字节/秒作为度量单位。设 B 表示带宽(MB/s), F 表示存储器时钟频率(MHz), D 表示存储器数据总线位数,则存储器的带宽计算方法为

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

例如,PC-100 的 SDRAM 带宽为 $100\text{MHz} \times 64\text{bit} / 8 = 800\text{MB/s}$ 。

4. 可靠性

可靠性是指在规定的时间内,存储器无故障读/写的概率,通常用平均无故障时间(Mean Time Between Failures, MTBF)来衡量。MTBF 可以理解为两次故障之间的平均时间间隔,其越长,说明存储器的可靠性越高。

5. 性能价格比

性能价格比是衡量存储器的综合指标,不同用途的存储器对其性能要求不同,例如对外存储器主要看容量,而对 Cache 则主要看速度。

5.2 RAM 存储器

RAM 的特点是在使用过程中能随时进行数据的读出和写入,故又称为读/写存储器,使用非常灵活,但 RAM 中存放的信息不能被永久保存,断电后会自动丢失。所以, RAM 是易失性存储器,只能用来存放暂时性的输入/输出数据、中间运算结果和用户程序,也常用它来与外存交换信息或作堆栈使用。通常人们所说的微型计算机存储容量指的就是 RAM 存储器的容量。

5.2.1 SRAM 存储器

SRAM 是一种静态随机存储器,其特点是只要不断电,所存信息就不会丢失;速度快,工作稳定,不需要外加刷新电路,使用方便灵活。但它所用的 MOS 管较多,致使集成度降低,功耗较大,成本也高。所以在微型计算机系统中,SRAM 常用作小容量的高速缓冲存储器 Cache。

1. SRAM 的基本存储电路

SRAM 的基本存储电路是由两个增强型的 NMOS 反相器交叉耦合而成的触发器,每个基本的存储单元由 6 个 MOS 管构成,因此,静态存储电路又称为六管静态存储电路,如图 5.4 所示。其中 T_1 、 T_2 为工作管, T_3 、 T_4 为负载管, T_5 、 T_6 为控制管, T_7 、 T_8 也为控制管,它们为同一列线上的存储单元共用。

2. SRAM 的基本结构

SRAM 的基本结构如图 5.5 所示。其中存储体是一个由 $64 \times 64 = 4096$ 个六管静态存储电路组成的存储矩阵。在存储矩阵中,X 地址译码器输出端提供 $X_0 \sim X_{63}$ 共 64 根行

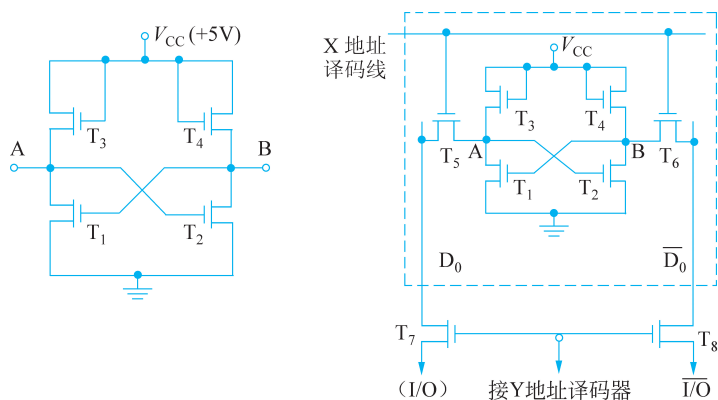


图 5.4 六管 SRAM 基本存储单元及基本存储电路

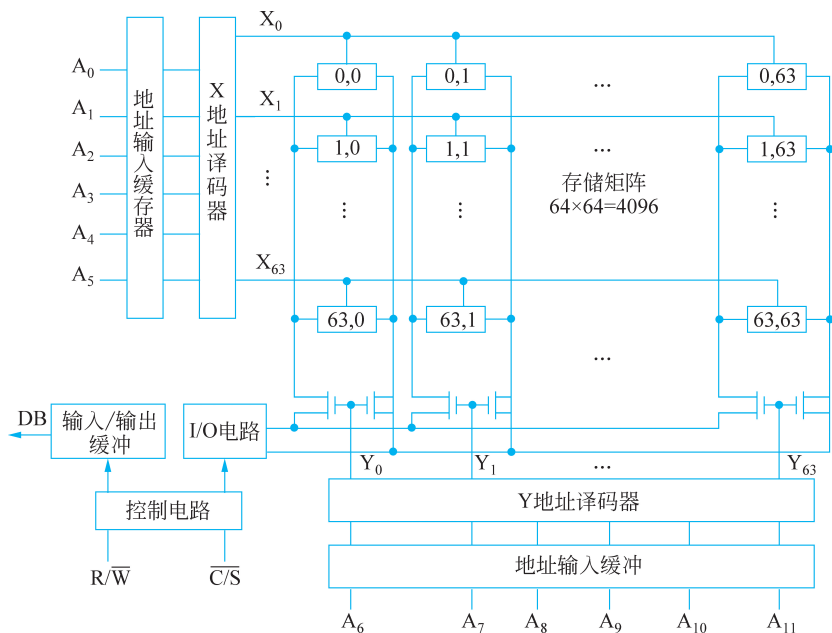


图 5.5 SRAM 的基本结构

选择线,而每一行选择线接在同一行中的 64 个存储电路的行选端,故行选择线能同时为该行 64 个行选端提供行选择信号。Y 地址译码器输出端提供 $Y_0 \sim Y_{63}$ 共 64 根列选择线,而同一列中的 64 个存储电路共用同一位线,故由列选择线同时控制它们与输入/输出电路(I/O 电路)连通。显然,只有行、列均被选中的某个单元存储电路,在其 X 向选通门与 Y 向选通门同时被打开时,才能进行读出信息和写入信息的操作。

图 5.5 所示的存储体是容量为 $4K \times 1$ 位的存储器,因此,它仅有一个 I/O 电路,用于存取各存储单元中的 1 位信息。如果要组成字长为 4 位或 8 位的存储器,则每次存取时,同时应有 4 个或 8 个单元存储电路与外界交换信息。因此,在这种存储器中,要将列的列向选通门控制端引出线按 4 位或 8 位来分组,使每根列选择线能控制一组的列向选通门同时打开;相应地,I/O 电路也应有 4 个或 8 个。每一组的同一位共用一个 I/O 电路。

这样,当存储体的某个存储单元在一次存取操作中被地址译码器输出端的有效输出电平选中时,则该单元内的4位或8位信息将被一次读/写完毕。

通常,一个RAM芯片的存储容量是有限的,需要用若干片才能构成一个实用的存储器。这样,地址不同的存储单元就可能处于不同的芯片中,因此,在选中地址时,应先选择其所属的芯片。对于每块芯片,都有一个片选控制端($\overline{CS}$),只有当片选端加上有效信号时,才能对该芯片进行读或写操作。一般地,片选信号由地址码的高位译码产生。

3. SRAM 的读/写过程

1) 读出过程

(1) 地址码 $A_0 \sim A_{11}$ 加到 RAM 芯片的地址输入端,经 X 与 Y 地址译码器译码,产生行选与列选信号,选中某一存储单元,经一定时间,该单元中存储的代码出现在 I/O 电路的输入端。I/O 电路对读出的信号进行放大、整形,送至输出缓冲寄存器。缓冲寄存器一般具有三态控制功能,没有开门信号,所存数据也不能送到 DB 上。

(2) 在传送地址码的同时,还要传送读/写控制信号($R/\overline{W}$ 或 $\overline{RD}$ 、 $\overline{WR}$)和片选信号($\overline{CS}$)。读出时,使 $R/\overline{W}=1$, $\overline{CS}=0$,这时,输出缓冲寄存器的三态门将被打开,所存信息送至 DB 上。于是,存储单元中的信息被读出。

2) 写入过程

(1) 地址码加在 RAM 芯片的地址输入端,选中相应的存储单元,使其可以进行写操作。

(2) 将要写入的数据放在 DB 上。

(3) 加上片选信号 $\overline{CS}=0$ 及写入信号 $R/\overline{W}=0$ 。这两个有效控制信号打开三态门,使 DB 上的数据进入输入电路,送到存储单元的位线上,从而写入该存储单元。

4. SRAM 芯片举例

不同 SRAM 的内部结构基本相同,只是在容量不同时其存储矩阵排列结构不同,即有些采用多字一位结构,有些采用多字多位结构。

常用的 SRAM 芯片有 2114($1K \times 4$ 位)、6116($2K \times 8$ 位)、6264($8K \times 8$ 位)、62256($32K \times 8$ 位)、62512($64K \times 8$ 位)、628128($128K \times 8$ 位)、628512($512K \times 8$ 位)、6281024($1024K \times 8$ 位)等,它们的引脚信号功能及操作方式基本相同,下面以 6116 为例进行简单介绍。

Intel 6116 的引脚信号如图 5.6 所示,是 24 引脚双列直插式芯片,采用 CMOS 工艺制造,存储容量为 2KB。有 11 条地址线($A_0 \sim A_{10}$),其中, $A_0 \sim A_3$ 用作列地址译码, $A_4 \sim A_{10}$ 用作行地址译码;有三条控制线 $\overline{CE}$ 、 $\overline{WE}$ 和 $\overline{OE}$,6116 的操作方式就是由这三条控制线共同作用决定的,具体如下。

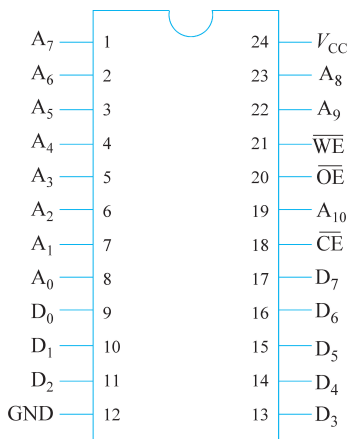


图 5.6 Intel 6116 的引脚信号

(1) 写入。当 $\overline{CE}$ 和 $\overline{WE}$ 为低电平时,数据输入缓冲器打开,数据由数据线 $D_7 \sim D_0$ 写入被选中的存储单元。

(2) 读出。当 $\overline{CE}$ 和 $\overline{OE}$ 为低电平,且 $\overline{WE}$ 为高电平时,数据输出缓冲器选通,被选中单元的数据送到数据线 $D_7 \sim D_0$ 上。

(3) 保持。当 $\overline{CE}$ 为高电平、 $\overline{WE}$ 和 $\overline{OE}$ 为任意时,芯片未被选中,处于保持状态,数据线呈现高阻态。

5.2.2 DRAM 存储器

DRAM 是一种动态随机存储器,其特点是集成度高、功耗低、价格便宜,但由于电容存在漏电现象,电容电荷会因为漏电而逐渐丢失,因此,需要外加刷新电路定时地对 DRAM 进行刷新,即对电容补充电荷。DRAM 的工作速度比 SRAM 慢得多,一般微型计算机系统内的存储器(即内存条)多采用 DRAM。

1. DRAM 的基本存储电路

典型的单管 DRAM 基本存储电路如图 5.7 所示,由存储部分 C_s 和选择电路 T_1 、 T_2 构成,其中 T_1 、 T_2 是 MOS 开关管。DRAM 电路在读出数据时, C_s 放电,原有信息被破坏,因此需要恢复原有存储的信息,这个恢复过程称为再生或重写。

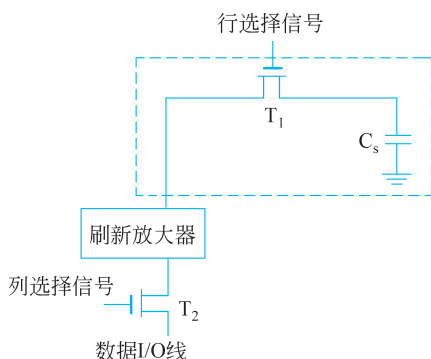


图 5.7 单管 DRAM 基本存储电路

由于 C_s 的电容值很小,又由于电容会漏泄,尤其是在温度上升时,漏泄放电会加快,所以典型的维持信息的时间约为 2ms,超过 2ms 信息就会丢失,因此需要进行动态刷新。这种电路的优点是结构简单、集成度较高且功耗小,但缺点是元件多,占用芯片面积大,噪声干扰也大。因此,要求 C_s 值做得比较大,刷新放大器应有较高的灵敏度和放大倍数。

由于 C_s 的电容值很小,又由于电容会漏泄,尤其是在温度上升时,漏泄放电会加快,所以典型的维持信息的时间约为 2ms,超过 2ms 信息就会丢失,因此需要进行动态刷新。这种电路的优点是结构简单、集成度较高且功耗小,但缺点是元件多,占用芯片面积大,噪声干扰也大。因此,要求 C_s 值做得比较大,刷新放大器应有较高的灵敏度和放大倍数。

2. DRAM 的基本结构

1) DRAM 芯片的结构

DRAM 也是由许多基本存储电路按行、列排列组成的二维存储矩阵,但为了降低芯片的功耗,保证足够的集成度,减少芯片对外封装引脚数目和便于刷新控制,DRAM 芯片都设计成了位结构形式,即每个存储单元只有一位数据位,一个芯片上含有若干字,如 $4K \times 1$ 位、 $8K \times 1$ 位、 $16K \times 1$ 位、 $64K \times 1$ 位或 $256K \times 1$ 位等,二维存储矩阵的这一结构形式也是 DRAM 芯片的结构特点之一。而且,这种存储矩阵结构也使得 DRAM 的地址线总是分成行地址线和列地址线两部分,芯片内部设置有行、列地址锁存器。在对 DRAM 进行访问时,总是先由行地址选通信号 $\overline{RAS}$ (CPU 产生)把行地址打入内置的行地址锁存器,随后再由列地址选通信号 $\overline{CAS}$ 把列地址打入内置的列地址锁存器,再由读/

写控制信号控制数据的读出/写入。所以在访问 DRAM 时,访问地址需要分两次打入,这又是 DRAM 芯片的特点之一。行、列地址线的分时工作,可以使 DRAM 芯片的对外地址线引脚大大减少,仅需与行地址线相同即可。

2) DRAM 的刷新

所有的 DRAM 都是利用电容存储电荷的原理来保存信息的,虽然利用 MOS 管间的高阻抗可以使电容上的电荷得以维持,但由于电容总存在漏泄现象,时间长了其存储的电荷会消失,从而使其所存信息自动丢失。所以,必须定时对 DRAM 的所有基本存储单元进行补充电荷,即进行刷新操作,以保证存储的信息不变。刷新,就是不断地每隔一定时间(一般每隔 2ms)对 DRAM 的所有单元进行读出,经读出放大器放大后再重新写入原电路中,以维持电容上的电荷,进而使所存信息保持不变。虽然每次进行的正常读/写存储器的操作也相当于进行了刷新操作,但由于 CPU 对存储器的读/写操作是随机的,并不能保证在 2ms 时间内能对存储器中的所有单元都进行一次读/写操作,因此,必须设置专门的外部控制电路和安排专门的刷新周期来系统地对 DRAM 进行刷新。

3. DRAM 芯片举例

常用的 DRAM 芯片有 2164/4164($64\text{K} \times 1$ 位)、21256/41256($256\text{K} \times 1$ 位)、41464($64\text{K} \times 4$ 位)以及 414256($256\text{K} \times 4$ 位)等产品,下面以 Intel 2164 芯片为例,介绍其结构及工作原理。

Intel 2164 是 $64\text{K} \times 1$ 位的 DRAM 芯片,采用单管动态基本存储电路,具有 16 个引脚。其内部结构如图 5.8 所示,芯片引脚与逻辑符号如图 5.9 所示。2164 的存储体由 4 个 128×128 的存储矩阵组成,每个存储矩阵由 7 条行地址线和 7 条列地址线进行选择,7 条行地址经过 128 选 1 行译码器产生 128 条行选择线,7 条列地址经过 128 选 1 列译码器产生 128 条列选择线,分别选择 128 行和 128 列。

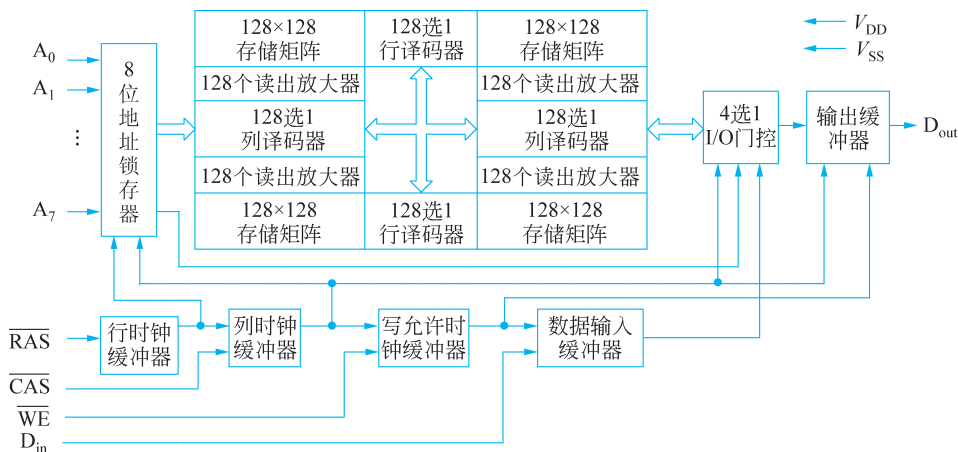


图 5.8 Intel 2164 DRAM 内部结构

从图中可知,2164 芯片本身只有 $A_7 \sim A_0$ 8 条地址线,每个存储单元只有一位,若要

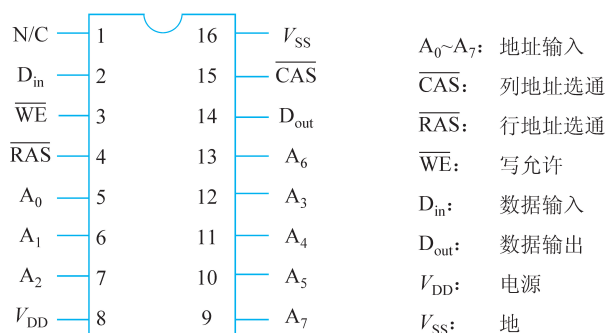


图 5.9 Intel 2164 DRAM 芯片引脚与逻辑符号

构成 64KB 的 DRAM 存储器实现 64KB 的 DRAM 寻址,则需要共 16 条地址线、8 片 2164。因此,该芯片采用了行地址线和列地址线分时工作的方式。其工作原理是:利用内部地址锁存器和多路开关,先由行地址选通信号 $\overline{RAS}$,把 8 位地址信号 $A_7 \sim A_0$ 送到行地址锁存器锁存,随后出现的列地址选通信号 $\overline{CAS}$ 把后送来的 8 位地址信号 $A_7 \sim A_0$ 送到列地址锁存器锁存。锁存在行地址锁存器中的 7 位行地址 $RA_6 \sim RA_0$ 同时加到 4 个存储器矩阵上,在每个存储矩阵中选中一行;锁存在列地址锁存器中的 7 位列地址 $CA_6 \sim CA_0$ 选中 4 个存储矩阵中的一列,选中 4 行 4 列交点的 4 个存储单元,再经过由 RA_7 和 CA_7 控制的“4 选 1”I/O 门控电路,选中其中的一个单元进行读/写。

2164 芯片数据的读出和写入是分开的,具体由 $\overline{WE}$ 信号控制。当 $\overline{WE}$ 为高电平时,读出数据;当 $\overline{WE}$ 为低电平时,写入数据。在对芯片进行刷新时,只需加上行选通信号 $\overline{RAS}$ 即可,即把地址加到行译码器上,使指定的 4 行存储单元只被刷新,而不被读/写,一般 2ms 可全部刷新一次。

实现 DRAM 定时刷新的方法和电路有多种,可以由 CPU 通过控制逻辑实现,也可以采用 DMA 控制器实现,还可以采用专用 DRAM 控制器实现。

5.3 ROM 存储器

ROM 存储器是一种非易失性半导体存储器件,其特点是信息一旦写入,就固定不变,断电后,信息也不会丢失,使用时,信息只能读出,一般不能修改。因此,ROM 常用于保存可长期使用且无须修改的程序和数据,如监控程序、主板上的 BIOS 系统程序等。在不断发展变化的过程中,ROM 也产生了掩模 ROM、PROM、EPROM、E²PROM 等各种不同类型的器件。

5.3.1 掩模 ROM

掩模 ROM 是指生产厂家根据用户需要,在 ROM 的制作阶段通过“掩模”工序将信息做到芯片里,一经制作完成就不能更改其内容。因此,掩模 ROM 适合存储永久性保存的程序和数据,大批量生产时成本较低。如国家标准的一、二级汉字字模就可以做到一