第 3 章　半导体存储器及其接口

学习目标

存储器是计算机系统中存储信息的主要部件。本章从存储器的分类、性能以及存储器基本结构着手，重点讨论随机存取存储器（RAM）和只读存储器（ROM）的工作原理，介绍存储器的容量扩展以及与微处理器的连接方法，最后介绍微型计算机系统中的多体存储器、高速缓冲存储器（Cache）和虚拟存储器技术。

通过对本章的学习，读者应了解半导体存储器的分类、主要性能指标，掌握随机存取存储器和只读存储器的功能特性，掌握存储器容量扩展以及与微处理器的连接方法，领会存储器系统的层次结构以及多体存储器、高速缓冲存储器和虚拟存储器的技术特点。

3.1　存储器概述

存储器是计算机系统中的存储部件，用于存储计算机工作时所用的程序和数据。计算机工作时，CPU 自动连续地从存储器中取出指令并执行指令所规定的操作，每执行完一条或几条指令，要把处理结果保存到存储器中，因此，存储器是计算机的记忆部件，是计算机系统的重要组成部分。

随着计算机技术的发展及广泛的应用，存储器系统的读写速度也在不断地提高，存储容量不断增加。特别是近些年多媒体技术的发展以及计算机网络的应用，要求计算机存储和处理的信息量越来越大，并对存储器的存取速度不断提出更高的要求，因而在存储器系统中应用了存储器层次结构、多体结构、高速缓冲存储器、虚拟存储器等技术，外存储器容量也可以无限地扩充，成为海量存储器。

3.1.1　存储器的分类

存储器的分类方法很多，通常从以下几方面对存储器进行分类。

1. 按在系统中的作用分类

根据存储器在微型计算机系统中的不同作用，可分为主存储器（简称主存或内存）和辅助存储器（简称辅存或外存）。

（1）内存。内存用来存放当前正在运行或将要使用的程序和数据，是计算机主机的一个重要的组成部分。CPU 可以通过指令直接访问内存。

系统对内存的存取速度要求较高，为了与 CPU 的处理速度相匹配，内存一般使用快速存储器件来构成。但是，由于受到地址总线位数的限制，内存空间的大小远远小于外存的容量，例如在 8086/8088 系统中，由于地址总线为 20 位，所以最大内存空间只能达到 1MB（2^{20}）。

尽管内存容量远不及外存，但是由于它具有访问速度快的特点，使得内存被用来存放计

算机工作时必须的系统软件、参数，以及当前要运行的应用软件和数据；而更多的系统软件和所有的应用软件则是存放于外存中，在需要时由外存调入内存。

（2）外存。外存也是用来存储程序和数据的，但它存储的信息却是 CPU 当前操作暂时不用的。外存位于主机外部，CPU 不能直接对其进行读写操作，当 CPU 需要使用外存中的信息时，要先通过专门的设备将其从外存调入到内存，然后再对内存中调入的数据进行直接的读写操作。

系统对外存的速度要求相对于内存而言可以慢一些，但对外存的容量却要求很大。由于外存具有长久保存信息的特点，所以外存多被用来保存和备份数据。

软盘、硬盘、光盘和移动硬盘等都是微型计算机系统中常见的外存。因为外存需要配置专门的驱动部件才能完成访问功能，所以外存的速度远不及内存的读写速度。外存的容量不受限制，特别是近年来，又出现了大容量的可移动式存储器，如移动硬盘、优盘（闪存盘）等，所以外存被称为“海量存储器”。

尽管计算机存储器系统包括内存和外存，但内存是主机的一部分，而外存则属于 I/O 外设。现代微型计算机的内存大多采用半导体存储器，一般具有断电则信息丢失的特点，而外存作为内存的后备存储器，存储的数据可以方便地修改和永久地保存，不受断电与否的影响。

另外，在高档微型计算机中，为了加快信息传递速度和提高计算机的处理速度，广泛应用高速缓冲存储器（Cache）。Cache 是微型计算机系统中的一个高速小容量的存储器，位于 CPU 和内存之间，常利用它暂存 CPU 正在使用的指令和数据。内存一般采用动态存储器，而 Cache 则主要由高速静态 RAM 组成。

2. 按存储信息的可保存性分类

根据存储器中信息的可保存性，可将存储器分为易失性存储器和非易失性存储器。

（1）易失性存储器。易失性存储器是指断电后信息消失的存储器，如半导体存储器 RAM，由于它具有读写速度快的特点，所以多被用于计算机内存。易失性存储器中的程序及数据可以在开机启动后由非易性存储器调入，在断电前，要及时保存到非易失性存储器中。

（2）非易失性存储器。非易失性存储器是指断电后仍然保持信息的存储器，如半导体存储器 ROM、磁盘、光盘、优盘等。微型计算机系统软件中有一部分软件如引导程序、监控程序或者基本输入/输出服务程序 BIOS，是计算机系统正常工作所必须的、无时无刻不用的，而且不能被随意修改的，所以它们必须常驻内存，于是应用了半导体存储器 ROM。希望长久保存的信息可存于磁盘、光盘、移动硬盘、优盘等海量存储器中。

在微型计算机系统中，要求系统至少有一部分存储器必须是非易失性的。

3. 按存储介质分类

存储二进制信息的物理载体称为存储介质，根据所使用存储介质的不同，存储器可分为半导体存储器、磁存储器、光存储器。半导体存储器多用于微型计算机内存，而磁存储器和光存储器则用于外存。

（1）半导体存储器。用半导体器件做成的存储器称为半导体存储器。按制造工艺可把半导体存储器分为双极型、CMOS 型、NMOS 型等。

（2）磁存储器。用磁性材料做成的存储器称为磁表面存储器。磁存储器主要有磁芯、磁泡、磁鼓、磁带和磁盘等。目前微型计算机系统中多用的是磁带和磁盘。

（3）光存储器。用光学材料做成的存储器称为光存储器。光存储器主要有只读式光盘

和可擦写光盘。

4. 按存储器的存取方式分类

微型计算机内存根据读写功能的不同，分为只读存储器（ROM）和随机存取存储器（RAM）；微型计算机外存根据存取时间与存储单元的物理位置是否有关，可分为顺序存取存储器（SAM）和直接存取存储器（DAM）。

（1）只读存储器 ROM。ROM 中所存储的内容是固定不变的，即只能读出，却不能随机写入新的内容。ROM 中的信息在关机后不会丢失。一般用它来存放微型计算机的启动程序、监控程序和系统管理程序等。

（2）随机存取存储器 RAM。随机存取存储器又称可读写存储器，它的任意一个存储单元都可以被随机读写，且存取时间与存储单元的物理位置无关，读写速度较快。半导体 RAM 断电后信息全部丢失。RAM 主要用作内存，存放微型计算机工作时的输入/输出数据及中间结果，并与外存交换信息。

（3）顺序存取存储器 SAM。SAM 对存储单元的访问是根据它们在存储器上物理位置的前后顺序来进行的，读写不同物理位置的存储单元，存取时间是不同的，一般来说，SAM 存取周期较长。例如磁带就是一种典型的顺序存储器。

（4）直接存取存储器 DAM。DAM 在存取数据时不必经过顺序搜索便可直接对存储器中的任意单元进行访问，存取时间与存储单元物理位置无关。例如，磁盘和光盘都是典型的直接存取存储器。

3.1.2　存储器的主要性能指标

存储器是计算机系统中的重要部件，它的性能直接影响整机的性能。微型计算机系统存储器的性能指标很多，如存储容量、存取速度、存储器可靠性、功耗、价格、性能价格比等，但就功能和接口技术而言，最重要的性能指标是存储容量和存取速度。

1. 存储容量

存储容量是指存储器可以容纳的二进制信息总量。容量越大，意味着所能存储的二进制位（bit）越多。微型计算机系统的内存容量指系统中所有内存芯片位容量的总和，每片存储器的位容量为总单元数与每单元位数的乘积。存储容量通常以字节（Byte，简写为 B）为单位来表示，目前使用的存储容量达 MB、GB、TB 或更大。它们之间的换算关系为

1B=8bit；　1KB=1024B；　1MB=1024KB；　1GB=1024MB；　1TB=1024GB

存储容量越大，计算机系统的功能就越强，所以人们总是希望尽量提高存储容量。但是存储容量的提高受到 CPU 的寻址范围、所选用的存储芯片的速度、成本等诸多因素的限制，故不能设计得很大。

2. 存取速度

存储器的存取速度可以用存取时间和存储周期来衡量。存取速度的度量单位采用 ns，存取时间越小，则存取速度越快。

（1）存取时间。存储器的存取时间是指从 CPU 发出有效的存储器地址从而启动一次存储器读/写操作，到读出或写入数据完毕所经历的时间，又称读写时间。目前，高速缓冲存储器（Cache）的存取时间已小于 20ns，中速存储器在 60～100ns 之间，低速存储器的存取时间在 100ns 以上。

（2）存储周期。存取周期是指连续启动两次独立的存储器读/写操作所需的最小时间间隔。由于在每次存储器读/写操作后，都需有一段存储器内部线路的恢复时间，所以存储器的存储周期通常略大于存取时间。

一般来说，外存的容量较大，但速度较慢；相对而言，内存的速度较快，但容量却小。人们所关心的存储器价格指标就与存储器的存取速度、存储容量直接相关：存储器的总价格正比于存储容量，反比于存取速度，所以，存储器的容量、速度、价格这三个指标是相互制约的。因此，对存储器系统的组织，要在综合衡量存储器的各项性能指标的前提下，兼顾存储器的制造工艺、体积、重量、功耗、品质等诸多因素，尽量提高存储器的性能价格比，满足系统的要求。

3.1.3 主存储器的基本结构

主存储器由存储体、地址寄存器、地址译码器、读写驱动器、数据寄存器以及时序控制电路等部件组成。如图 3-1 所示。

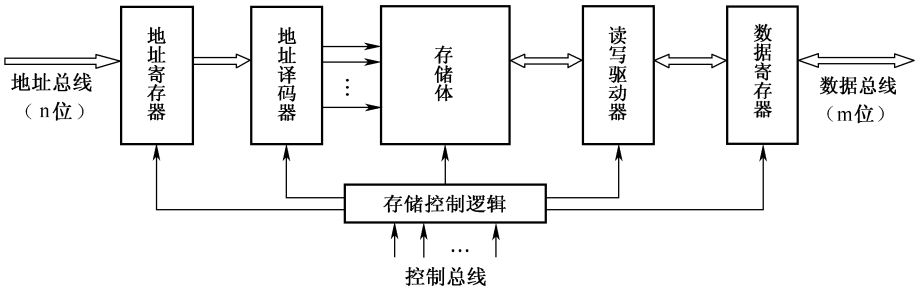


图 3-1 主存储器基本组成

存储体是具体存储信息的场所，是存储单元的集合。假设存储器有 m 位数据总线、n 位地址总线和若干控制总线。地址总线给出所需访问的存储单元地址，最多可访问 2^n 个存储单元；数据总线用于在 CPU 与存储单元之间传送数据，一次可传送 m 位二进制数据；控制信号则用来控制存储器的读/写等操作。

当 CPU 要访问内存时，首先通过地址总线把地址码送入地址寄存器中锁存，再传送给地址译码器，经译码后使对应于该地址的某一根选择线有效，从而选中相应的存储单元；接着 CPU 发出读/写命令，于是时序控制电路产生对存储单元的读/写操作控制信号。在进行存储器读操作时，控制电路中的读信号线有效，于是把所选中的存储单元的信息读出并送入读/写驱动器放大，然后送至数据寄存器，再经数据总线送给 CPU。在进行存储器写操作时，CPU 不仅要把地址码送入地址寄存器，还要把待写入的数据传送给数据寄存器，并使控制电路中的写信号线有效，于是数据寄存器中的数据被存入选中的存储单元。

对内存写操作时，存储单元中原有的内容将被新写入的数据取代；读操作时，存储单元中的内容不受影响。所以，对内存的写是“破坏性”的写，对内存的读是“非破坏性”的读。

3.1.4 半导体存储器

半导体存储器具有工艺简单、集成度高、成品率高、可靠性高、存取速度快、体积小、

功耗低等特点；其存储电路所占的空间小，可以和译码电路以及缓冲寄存器制作在同一芯片中；对它的读是一种非破坏性的读。所以现代微型计算机的主存储器普遍采用半导体存储器。

半导体存储器的种类繁多，从不同的角度有不同的分类。

（1）从电路器件角度可分为双极型存储器和单极型存储器。双极型存储器是用 TTL（晶体管—晶体管逻辑）电路制成的存储器，它具有速度快的特点，但是集成度低、价格高，功耗也较单极型存储器大。单极型存储器是采用 MOS（金属氧化物半导体）电路制成的存储器，其特点是集成度高、功耗低、价格低，但速度相对于双极型存储器较慢。MOS 型存储器又分为 NMOS（N 沟道 MOS）、HMOS（高密度 MOS）、CMOS（互补型 MOS）等。

计算机系统中的高速缓冲存储器（Cache）需要较高的存取速度，所以常采用双极型存储器。MOS 型存储器的工作速度虽然赶不上双极型存储器，但由于它具有集成度高、功耗低、价格廉等优点，所以在组成大容量内存时，仍不失为一种理想的选择，特别是用 NMOS 工艺制作的随机存取存储器，其应用范围最为广泛；CMOS 存储器也因其具有超低功耗的特点，在某些应用场合有它的特殊用途。

随着半导体工艺和技术的不断改进，MOS 型存储器的工作速度不断提高，目前几乎可以与双极型 TTL 存储器的工作速度相媲美，所以计算机系统中的内存普通采用 MOS 型存储器。

（2）从存储特点和功能的角度，半导体存储器又分为随机存取存储器（RAM）和只读存储器（ROM）两大类。如图 3-2 所示。

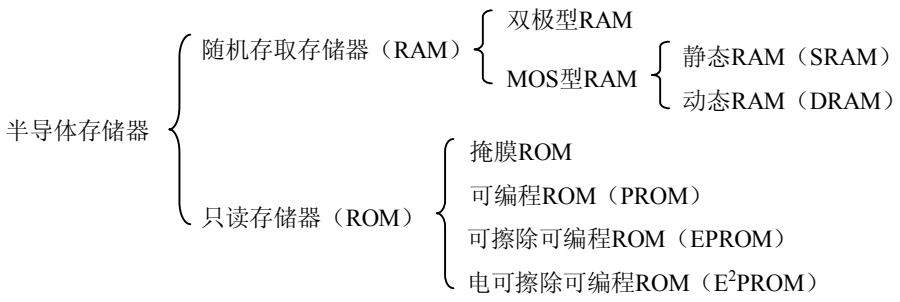


图 3-2　半导体存储器的分类

半导体随机存取存储器 RAM 可以随机地对每个存储单元进行读写，但断电后信息会丢失。根据存储原理，RAM 又分为静态 RAM（SRAM）和动态 RAM（DRAM）。静态 RAM 存放的信息在不断电的情况下可以长时间保持不变，只要不掉电所保存的信息就不会丢失。而动态 RAM 保存的内容即使在不掉电的情况下，隔一定时间后也会自动消失，因此要对其进行定期的刷新。

半导体只读存储器 ROM 是一种只能读出不能随机写入信息的存储器，所存储的信息可以长久保存，掉电后存储信息仍不会改变。一般 ROM 用于存放固定程序，如启动程序、监控程序等。按存储单元的结构和生产工艺的不同，ROM 又可分成掩膜只读存储器（ROM）、可编程只读存储器（PROM）、光可擦除可编程只读存储器（EPROM）、电可擦除可编程只读存储器（E^2PROM）等。

3.2 随机存取存储器 RAM

3.2.1 静态 RAM（SRAM）

1. SRAM 的基本存储电路

所谓基本存储电路是指存储一位二进制数的电路，又称单元电路。对于各种基本存储电路，无论其内部结构如何，对外呈现的特性均为：用一个信号选中该电路，电路中所存储的信息通过数据线与外界交换。

SRAM 的基本存储电路如图 3-3 所示。一般是由 6 个 MOS 管组成静态触发器，触发器的工作必须有电源，存入的数据才可以保留和读出，若掉电，存入的数据全部丢失。

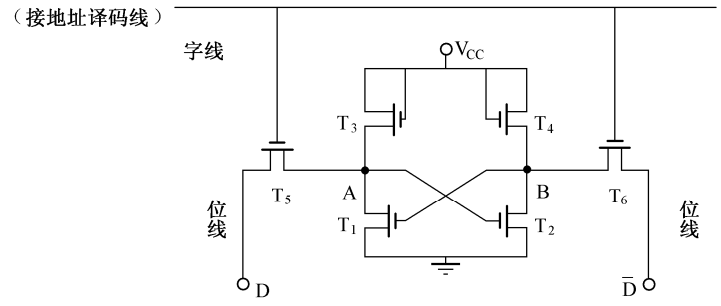


图 3-3 6 管静态 RAM 存储电路

在此电路中，T_1～T_6 构成一个基本存储电路，由 T_1～T_4 组成双稳态触发器，其中 T_1、T_2 为放大管，T_3、T_4 为有源负载管，电路左右对称。

由 T_1～T_4 所构成的双稳态触发器，可以存储一位二进制信息。该电路有两个稳定状态：当 T_1 导通，则 A 点为低电平，A 点为低电平又使 T_2 截止，于是 B 点为高电平，B 点为高电平又保证了 T_1 的导通，这是一个稳定状态。同样，当 T_1 截止时，则 A 点为高电平，A 点高电平又使 T_2 导通，于是 B 点为低电平，B 点为低电平又保证了 T_1 的截止，这是另一个稳定状态。这样，可用“1”和“0”来分别表示这两个稳定状态：T_1 导通 T_2 截止的状态为“0”状态，T_1 截止 T_2 导通的状态为“1”状态。

T_5、T_6 管作为两个控制门，起到两个开关的作用，对两个稳定状态进行控制：当从字线送来高电平信号时，门控管 T_5、T_6 导通，触发器与 D 线接通，即 A 点接通 D 线，B 点接通 $\overline{D}$ 线。

在进行写操作时，写入的信号从 D 线和 $\overline{D}$ 线输入。若要写入“1”，则使 D 线为“1”，$\overline{D}$ 为“0”，通过 T_5、T_6 管与 A、B 点相连，从而使 T_1 截止，T_2 导通。当写入信号和地址译码信号消失后，T_5、T_6 截止，于是 T_1～T_4 组成的双稳态触发器保持数据“1”，由于存储电路的电源 T_3、T_4 两个负载管可以不断地向 T_1、T_2 管栅极补充电荷，所以只要不掉电，就能保持写入的“1”信号。若要写入“0”，则在 D 线和 $\overline{D}$ 线上分别送入“0”和“1”，使得 T_1 导通，T_2 截止。只要不掉电，这个状态也会一直保持下去，直至重新写入新的数据为止。

在进行读操作时，字线上送来高电平，使 T_5、T_6 导通，A 点的状态被送到 D 线上，B 点

的状态被送到 $\overline{D}$ 线上，这样就读走了原来存储的信息。信息读出以后，基本存储电路中原来存储的内容仍然保持不变。

由以上基本存储电路组成的 SRAM 芯片具有以下特点：

（1）可靠性高，速度快。目前高性能 SRAM 存取时间达 15～40ns。

（2）高稳定性。只要不掉电，SRAM 芯片中存储的信息永远不会丢失，所以无须外加刷新电路，故外围电路简单。

（3）集成度低。由于 SRAM 存储电路中所用的 MOS 管较多，因而集成度较低。

（4）功耗较大。由于 T_1 管、T_2 管组成的双稳态触发器总有一个是导通的，即电路中始终有电流通过，故功耗较大。

由于 SRAM 具有以上特点，它的每位价格较高，用它来构造大容量的存储器显然不划算。因此，SRAM 一般用作高速缓冲存储器（Cache），这可以充分发挥 SRAM 速度快和可靠性高的优势，采用小容量的 SRAM 芯片作为 Cache，价格也不至于太高，这样可以保证微型计算机系统的高性能价格比。

2. SRAM 的结构

利用多个基本存储电路排成行列矩阵，再加上地址译码电路和读写控制电路，就可以构成读写存储器。下面以 4 行 4 列基本存储电路构成的 16×1 位 SRAM 为例说明它的结构。

如图 3-4 所示，这个可读写的静态随机存取存储器主要由存储体、读写控制电路、地址译码电路和 I/O 电路构成。

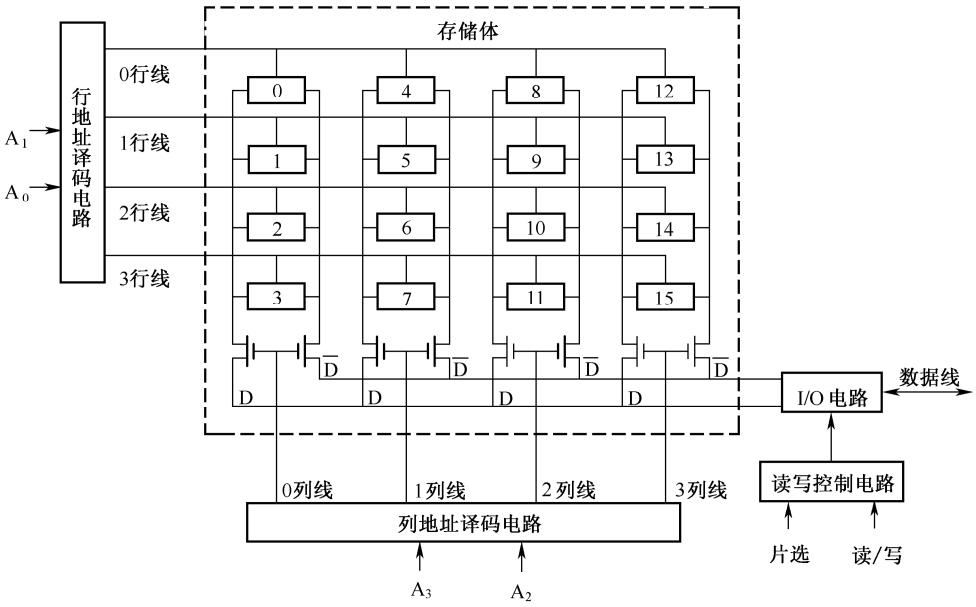


图 3-4　典型的 SRAM 结构

（1）存储体。多个基本存储电路有规则地组合在一起就构成了存储体。为了减少片内的连线，便于译码寻址，所以存储体内的各个存储单元一般排列成行列矩阵。由图 3-4 可见，存储体排成 4 行×4 列矩阵，每个存储电路可存储一个二进制位，构成 16×1 位的存储体。

（2）地址译码电路。存储矩阵中，基本存储电路的地址译码方式一般有两种：单译码方

式和双译码方式。实际存储器芯片的地址译码电路与存储体集成在一个芯片中，所以当存储容量较小时，可以使用地址单译码方式，即仅使用一个地址译码器；当存储容量较大时，由于地址译码器的输出线过多，导致集成电路的内部结构复杂，从而增加生产工艺上的困难，所以采用地址双译码方式，即地址译码电路分为行地址译码和列地址译码两部分。

双译码方式中，行、列地址译码器分别接收地址总线上送来的高位、低位地址信号。行译码器的输出选择线有多根，若行译码器的地址输入信号有 n 位，则行译码器的输出选择线就有 2^n 根，一根行选择线对应存储矩阵中的一行。存储矩阵中，同一行上各个基本存储电路的字线与同一行选择线并连在一起。同理，存储矩阵中，同一列上各个基本存储电路的字线与同一列选择线并连。

如图 3-4 所示，当给定的地址码为 A3A2A1A0=0000 时，AlA0 经行地址译码器译码后，使 0 行线为有效的高电平，A3A2 经列地址译码器译码后，使 0 列线为有效的高电平，于是 0 行与 0 列交叉处的 0 号基本存储电路被选中。又如，当给定的地址码为 A3A2A1A0=0110 时，那么 2 行与 1 列交叉处的 6 号基本存储电路被选中。这样，只要给定一个地址码，就会唯一地选中一个存储单元。

（3）读写控制电路。读写控制电路接收 CPU 发来的片选及读/写控制信号，控制对本存储器的数据读写方向。由于单片存储器的容量限制，所以一般的系统存储器都是由若干个芯片组成的，由于地址信号和读写控制信号同时加到所有芯片上，如果选中一个芯片中的某一单元，那么其他芯片中相同地址的单元就不应该同时被选中，所以各芯片还要有一个片选控制信号。当片选端送来有效信号时，该存储芯片被选中，这时才能进行读写操作。读/写控制信号用来规定对存储器进行读或写操作。所以，一个存储器芯片中某单元的信息能否与系统数据总线的信息交换，是受地址、片选和读写控制三个信号同时控制的。

（4）I/O 电路。I/O 电路处于数据总线与存储体之间，具有数据传送的功能，并具有放大信号的作用。I/O 电路内部的三态双向缓冲器，连接存储体中各个基本存储电路的位线及系统数据总线，在存储器芯片未被选中时，三态缓冲器呈高阻态，使存储器与系统数据总线隔离；当片选信号有效时，在读写控制信号的控制下，可以对被选中的单元进行数据的读出或写入。

3. 典型 SRAM 芯片

常用 SRAM 芯片有 2114（1K×4 位）、6116（2K×8 位）、6264（8K×8 位）、628128（16K×8 位）、62256（32K×8 位）等多种。

（1）Intel 2114 芯片。Intel 2114 为 NMOS 静态 RAM，单一+5V 电源，4 位共用的数据输入/输出端，并采用三态控制。所有的输入端和输出端都与 TTL 兼容，其引脚排列如图 3-5 所示。

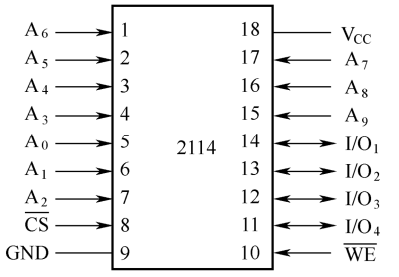


图 3-5　静态 RAM Intel 2114 引脚图

Intel 2114 采用 6 管 NMOS 静态基本存储电路，容量为 1K×4 位，即 1024 个字，每字 4 位。芯片内部共有 4096 个基本存储电路，排列成 64×64 的矩阵。地址线共 10 根（A_9～A_0），其中 A_8～A_3 这 6 根用于行译码，产生 64 个行选择信号；A_9、A_2、A_1、A_0 地址线用于列译码，产生 16 个列译码信号，并且每个列译码信号同时接 4 位。各信号线的意义如表 3-1 所示。

表 3-1　Intel 2114 引脚功能表

引脚	功能	说明	
A_9～A_0	地址	输入，三态	
I/O_4～I/O_0	数据线	双向，三态	
$\overline{CS}$	片选	输入，低电平有效	
$\overline{WE}$	写允许	=0	写
		=1	读
Vcc	电源		
GND	地		

在 $\overline{CS}$ 片选信号低电平有效的情况下，当写允许控制端 $\overline{WE}$ 为低电平时，可以对存储器芯片 2114 进行写操作，当 $\overline{WE}$ 为高电平时，可以对 2114 进行读操作。

（2）Intel 6116 芯片。Intel 6116 为高速静态 CMOS 随机存取存储器，具有功耗低、高速度的特点，与 TTL 兼容，完全静态，无须时钟脉冲或定时选通脉冲，引脚与 2K×8 位的 EPROM 引脚相互兼容。其引脚排列如图 3-6 所示。

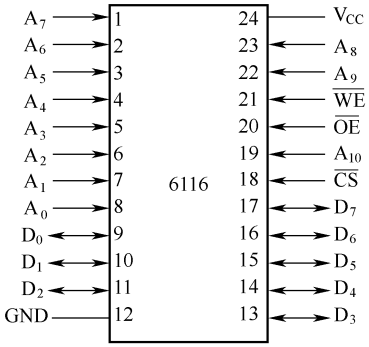


图 3-6　静态 RAM Intel 6116 引脚图

Intel 6116 芯片的存储容量为 2K×8 位，即 2048 个字，每字 8 位。芯片内部有 16384 个基本存储电路，排列成 128×128 的矩阵。需要 11 条地址线，A_{10}～A_4 这 7 根地址线用于行译码，产生 128 个行选择线，A_3～A_0 列地址产生 16 个列选择线，每个列选择线同时接 8 位。各信号线的意义如表 3-2 所示。

表 3-2　Intel 6116 引脚功能表

引脚	功能	说明
A_{10}～A_0	地址	输入，三态
D_7～D_0	数据线	双向，三态
$\overline{CS}$	片选	输入，低电平有效
$\overline{WE}$	写允许	输入，低电平有效
$\overline{OE}$	读允许	输入，低电平有效
Vcc	电源	
GND	地	

在 $\overline{CS}$ 片选信号低电平有效的情况下，当写允许端 $\overline{WE}$ =0 并且 $\overline{OE}$ =1 时，可以对存储器芯片 6116 进行写操作，当 $\overline{WE}$ =1 并且 $\overline{OE}$ =0 时，可以对 6116 进行读操作。

其他静态 RAM 芯片如 6264、62256 的结构与 Intel 6116 相似，只是地址线不同。这两类

芯片为 28 个引脚的双列直插式，使用单一的+5V 电源，与同样容量的 EPROM 引脚相互兼容，从而使接口电路的连线更为方便。

3.2.2 动态 RAM（DRAM）

动态 RAM 的基本存储电路是以电荷形式存储信息的器件。动态基本存储电路有六管型、四管型、三管型以及单管型。其中单管型由于集成度高、功耗低而越来越被广泛采用。

1. 单管 DRAM 基本存储电路

单管 DRAM 基本存储电路如图 3-7 所示，由一只 MOS 管和一个与源极相连的电容 C 组成。DRAM 中信息的存放依靠电容 C，电容 C 有电荷时表示存储的信息为二进制“1”，无电荷时表示存储的是“0”。但是由于任何电容都存在漏电问题，所以，即使电容 C 有电荷，过一段时间后随着电荷的流失，信息也就丢失了。

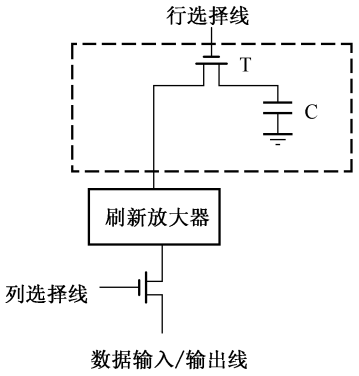


图 3-7 单管 DRAM 基本存储电路

解决的办法就是刷新。每隔一定时间刷新一次，使电容中原来处于逻辑电平“1”的电荷又得到补充，而原来处于电平“0”的电容仍保持“0”。

在未进行读写操作时，行选择线处于低电平，MOS 管 T 截止，电容 C 与外电路断开，不能进行充电、放电，电路保持原状态不变。

在进行读操作时，行选择线为高电平，基本存储电路中的 MOS 管 T 导通，使刷新放大器读取存储电容 C 上的电压值。刷新放大器的灵敏度很高，放大倍数较大，它能将读来的电压值转换为对应的逻辑电平“0”或“1”。列选择信号的作用是驱动基本存储电路，从而可以由输入/输出线将信息输出。

在读出过程中，由于基本存储电路中的电容受到干扰，为了在读出之后，仍能保存原来信息，刷新放大电路在对这些电容上的电压值读取之后，又立即重写回到存储电容 C 中，即在读操作时完成刷新。由于刷新时，列选择信号总为低电平，所以电容 C 上的信息不会再被送到数据总线上。

DRAM 的存取速度不及 SRAM，需要定时刷新。但由于 DRAM 所用的 MOS 管少，所以集成度较高，功耗降低，价格低，所以在大容量的存储器中普遍采用 DRAM。

2. DRAM 的刷新方式

动态存储器刷新就是周期性地对动态存储器进行读出、放大、再写回的过程。

DRAM 是利用电容存储电荷的原理来保存信息的，由于电容会泄漏放电，所以为了保证电容中的电荷不丢失，每隔一定时间（一般为 2ms）必须对 DRAM 读出、放大、再写回一次，从而使原来处于逻辑电平“1”的电容上所泄漏的电荷得到补充，而原来处于电平“0”的电容仍保持“0”，这就是对 DRAM 的刷新。

一般来说，DRAM 应在 2ms 时间内将全部基本存储电路刷新一遍。但是，由于读写操作的随机性，不能保证在 2ms 内对 DRAM 的所有行都能遍访一次，所以需要依靠专门的存储器刷新周期来系统地完成对 DRAM 的刷新。

在存储器系统中，刷新是按行进行的，即一行内所有的基本存储电路同时读出、放大、

再写回，一个刷新周期内对所有行中所有的基本存储电路都刷新一遍。例如，对一个 64 行×32 列的存储矩阵进行刷新，一次对一行中 32 个基本存储电路同时刷新，在一个刷新周期 2ms 内必须将 64 行全部刷新完毕，所以对每行的刷新必须在 31μs（2ms/64≈31μs）内完成。

DRAM 的刷新常采用两种方法：一是利用专门的 DRAM 控制器实现刷新控制，如 Intel 8203 控制器；二是在每个 DRAM 芯片上集成刷新控制电路，使存储器件自身完成刷新，这种器件叫综合型 DRAM，如 Intel 2186/2187。

3. 典型 DRAM 芯片

Intel 2164 是容量为 64K×1 位的典型 DRAM 芯片，片内有 65536 个存储单元，每个单元存放 1 位数据，用 8 片 2164 就可以构成 64KB 的存储器。

Intel 2164 内的 64K×1 位存储体内分为 4 个 128×128 存储矩阵。每个 128×128 矩阵都采用双译码方式，行、列各需要 7 位地址。如图 3-8 所示。

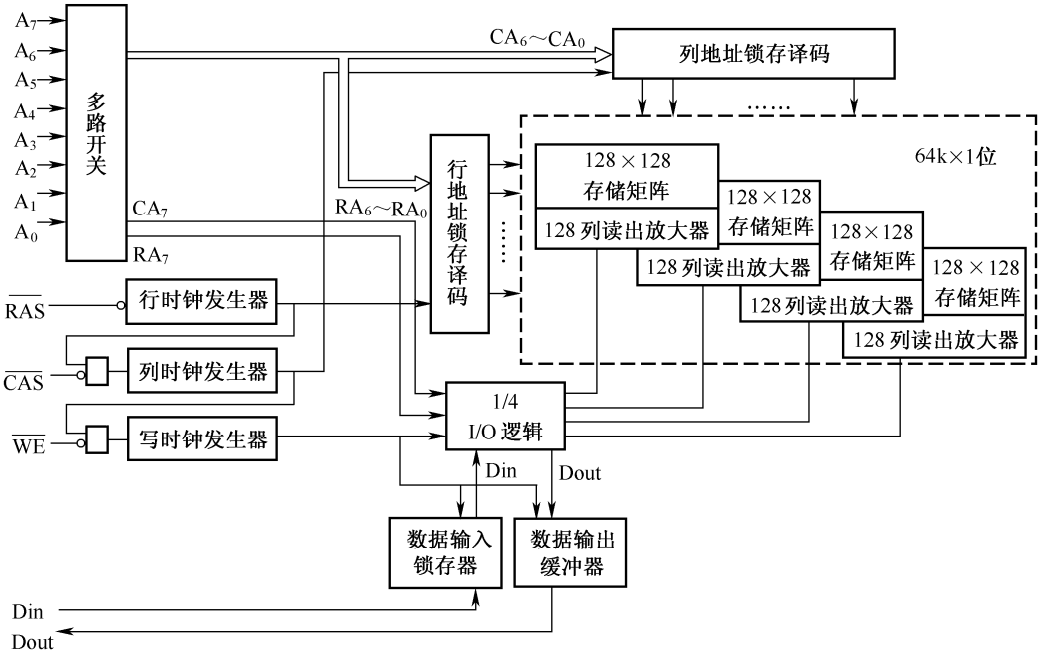


图 3-8　Intel 2164 的内部结构

若想在片内寻址 64K 个单元，通常需要 16 条地址线。为了减少地址线引脚数，Intel 2164 采用分时复用技术，将片内地址线分为行地址线和列地址线，这样，芯片对外只需引出 8 条地址线，于是 16 位地址信号可以拆分成低 8 位和高 8 位，分两次送至芯片中。芯片内部利用多路开关和地址锁存器，首先由行地址选通信号 $\overline{RAS}$（低电平有效）把先送来的 8 位地址中的低 7 位（RA_6～RA_0）送至行地址锁存器中，随后，列地址选通信号 $\overline{CAS}$（低电平有效）把后送来的 8 位地址中的低 7 位（CA_6～CA_0）送至列地址锁存器中。需要说明一点：7 位行地址和 7 位列地址是同时加到存储体内 4 个 128×128 存储矩阵上的，也就是说，经行、列地址译码后，4 个存储矩阵中的同位置存储单元被同时选中。那么到底选择 4 个存储矩阵中的哪一个呢？由芯片内部行、列地址中的最高位 RA_7 和 CA_7 决定。RA_7 和 CA_7 接至 1/4 的 I/O 逻辑，在

对芯片读/写时，通过这两位地址的 4 种不同编码，实现对 4 个矩阵的选 1 操作。

Intel 2164 的数据输入和输出是分开的，由 $\overline{WE}$ 信号来控制对芯片的读写。当 $\overline{WE}$ 为低电平时，数据输入缓冲器允许，于是 Din 引脚上的数据经数据输入缓冲器对选中单元进行写入；当 $\overline{WE}$ 为高电平时，数据输出缓冲器允许，于是所选中单元的内容经过数据输出缓冲器由 Dout 引脚读出。

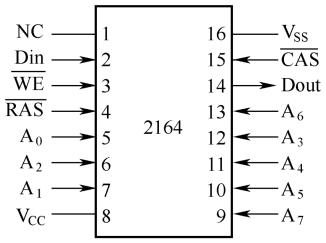


图 3-9　Intel 2164 引脚图

Intel 2164 芯片的引脚如图 3-9 所示，各引脚的说明如表 3-3 所示。Intel 2164 芯片没有片选信号，实际上用 $\overline{RAS}$ 和 $\overline{CAS}$ 作为片选信号。

表 3-3　Intel 2164 引脚功能表

引脚	功能	说明	
A_7～A_0	地址	输入，三态	
Din	数据输入线	输入，三态	
Dout	数据输出线	输出，三态	
$\overline{RAS}$	行地址选通	输入，低电平有效	
$\overline{CAS}$	列地址选通	输入，低电平有效	
$\overline{WE}$	写允许	=0	写
		=1	读
Vcc	电源		
Vss	地		

Intel 2164 的 8 条地址线也用于刷新（刷新时地址计数，实现逐行刷新，2ms 内全部刷新一遍）。刷新时只用 $\overline{RAS}$ 信号和低 7 位行地址 RA_6～RA_0，RA_7 不用。$\overline{RAS}$ 的低电平状态会使 $\overline{CAS}$ 和 $\overline{WE}$ 禁止，从而保证刷新时不使用列地址，并且暂时禁止新的数据写入。$\overline{RAS}$ 为有效的低电平时，行地址 RA_6～RA_0 被锁存到行地址锁存器中，因为这 7 位行地址是同时加到 4 个 128×128 存储矩阵上的，所以一次行地址译码就有 4 个存储矩阵的同一行被选中。刷新是周期性的，平均每隔 15.6μs（2ms/128≈15.6μs）刷新一次，每次刷新一行，一行中 512 个存储单元的信息被选通送至 512 个读出放大器，经过鉴别后再重新写回。

3.3　只读存储器 ROM

半导体只读存储器 ROM 具有如下特点：

（1）信息一旦写入就可以长期保存，不受电源掉电的影响。

（2）信息一旦写入只能读出，不能再随机写入新的内容。

（3）结构简单、成本低、集成度高。

（4）可靠性高。

因此，在微型计算机系统中，ROM 常用来存储一些固定的信息，如监控程序、启动程序、

基本输入/输出服务程序等。此外，ROM 还可作为控制存储器，存放微程序。

3.3.1　掩膜只读存储器 ROM

掩膜 ROM 中的信息是生产厂家在制造过程中写入的。掩膜 ROM 制成后，存储的信息就不能再改变了，用户在使用时只能读出。如图 3-10 所示，一个简单的 4×4 位 MOS 管 ROM，采用单译码方式，两位地址 A_1、A_0 输入，经译码后产生 4 条选择字线，可分别选中 4 个单元，每个单元有 4 位。

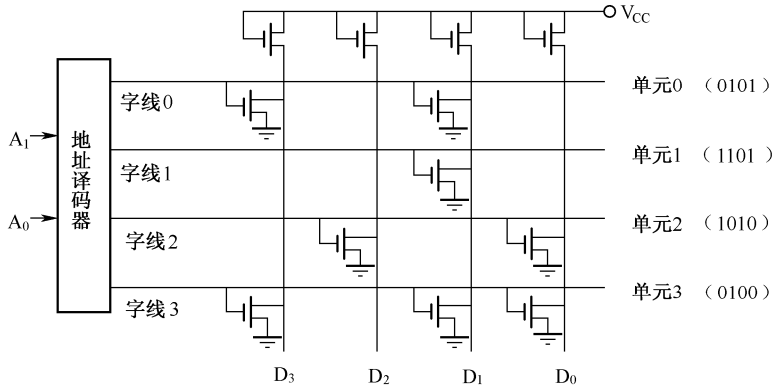


图 3-10　4×4 位掩膜 ROM 示意图

图 3-10 中所示的矩阵中，在行与列的交叉点上，有的接有管子，有的没有，这是在厂家制造时根据用户所提供的程序对芯片图形（掩膜）进行二次光刻所决定的，所以称掩膜 ROM。最上排的 4 个 MOS 管起到了一个上拉电阻的作用，它永远处于导通状态，但具有一定的导通电阻，它保证了无 MOS 管的位输出为“1”，有 MOS 管的位出为“0”。

例如，当地址线 A_1A_0=00 时，经地址译码后选中 0 号单元，即字线 0 为高电平，若有管子与其相连，如图 3-10 所示的 D_3 和 D_1，其相应的 MOS 管导通，输出为 0；而 D_2 和 D_0 没有管子与字线 0 相连，则输出为 1，故单元 0 的 $D_3D_2D_1D_0$=0101。

因为掩膜 ROM 需要专门制作掩膜板，成本很高，制作周期较长，所以只在产品数量较大时，才做成这种掩膜 ROM。

3.3.2　可编程只读存储器 PROM

半导体 PROM 适用于用户根据自己的需要来写入存储信息的场合。厂家生产的 PROM 芯片不事先存入固定内容，存储矩阵的所有字线和位线的交叉处均连接有二极管或三极管，即出厂时，存储单元的内容是全“1”（或全“0”）。使用时，用户可根据自己的需要，将某些位的内容改写为“0”（或“1”），但只能改写一次。

如图 3-11 所示为用双极型晶体管和熔丝组成的 PROM 的基本存储结构。晶体管的集电极接 Vcc，基极连接行线，发射极通过一个熔丝与列线相连，所以也称为熔丝式 PROM。

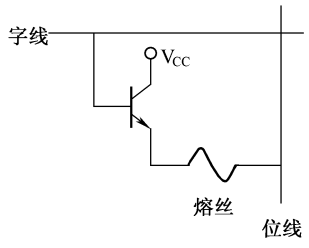


图 3-11　PROM 基本存储电路

PROM 在出厂时，晶体管阵列的熔丝均为完好状态，编

程时，通过字线选中某个晶体管。当写入信息时，可在 Vcc 端加高电平。若某位写“0”，则向相应位线送低电平，此时管子导通，控制电流使该位熔丝烧断，即存入“0”；若某位写 1，向相应位线送高电平，此时管子截止，使熔丝保持原状，即存入“1”。

可见，熔丝一旦烧断，就不能再复原，所以这种 PROM 是一种一次性写入的只读存储器。PROM 的电路和工艺要比 ROM 复杂，所以价格较贵。

3.3.3 可擦除可编程只读存储器 EPROM

由于 PROM 的内容在写好后就不能再改变，因此能够重复擦写的 EPROM 被广泛应用。EPROM 的特点是：芯片的顶部开有一个圆形的石英窗口，通过紫外线的照射可将片内所存储的原有信息擦除；根据需要可利用 EPROM 的专用编程器（也称为“烧写器”）对其编程写入，写入后的信息可长久保持。因此这种芯片可反复使用。

如图 3-12 所示的 EPROM 基本存储电路是利用浮栅 MOS 管构成的，又称为浮栅 MOS EPROM 存储电路。该电路和普通 P 沟道增强型 MOS 管相似，只是它的栅极没有引出端，而被 SiO_2 绝缘层所包围，即处于悬浮状态，故称为“浮栅”。

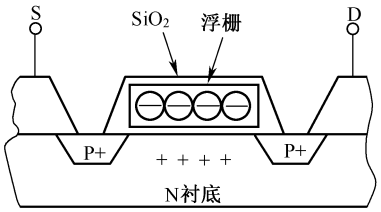


图 3-12　EPROM 基本存储电路

在原始状态，栅极上没有电荷，该管没有导通沟道，D 和 S 是不导通的，管子处于截止状态，此时存放信息“1”；如果设法向浮栅注入电子电荷，等效于栅极上加负电压，如果注入的电子电荷足够多，这些负电子在硅表面上感应出一个连接源、漏极的导电沟道，使管子呈导通状态，此时存放信息“0”。当外加电压取消后，积累在浮栅上的电荷没有放电回路，故在室温和无光照的条件下电荷能长期地保存在浮栅中。

EPROM 的编程过程实际是对某些单元写入“0”的过程，也就是向浮栅注入电子的过程。采用的方法是：在管子的漏极加一个高电压，使漏区附近的 PN 结击穿，在短时间内形成一个大电流，大量电子穿过绝缘层注入浮栅。通过判断浮栅是否积存电荷来区别管子存储的内容是否为“0”。若浮栅无积存电荷，则源、漏极是不导通的，则从该存储单元读出的信息为“1”；若在浮栅中注入了电子，则 MOS 管就有导电沟道存在，则从该存储单元读出的信息为“0”。EPROM 在出厂时未经过编程，浮栅中没有积存电荷，即存储的信息为全“1”。

擦除的原理与编程正好相反，采用的办法是利用紫外线光照射，由于紫外线光子能量较高，从而使浮栅中的电子获得能量，形成光电流从浮栅流入基片，使浮栅恢复初态。在 EPROM 芯片的上方有一个石英玻璃窗口，若想擦除存储器中的内容，只要将 EPROM 芯片放入一个靠近紫外线灯管的小盒中，一般照射 15～20min，即可将 EPROM 芯片内的信息全部擦除。编程后的芯片窗口要贴上不透光的封条，以保护其不受紫外线的照射。EPROM 的优点是一块芯片可多次使用。

常用的 EPROM 有 2716（2K×8 位）、2764（8K×8 位）、27256（32K×8 位）、27512（64K×8 位）等典型芯片。

3.3.4　电可擦除可编程只读存储器 E^2PROM

尽管 EPROM 可以实现多次编程，但在编程过程中，整个芯片即使只写错一位，也必须用紫外线擦除器擦除重写，因此操作起来比较麻烦。另外，EPROM 可被擦除后重写的次数也是有限的，一块芯片的使用寿命往往不太长。

E^2PROM 是一种新型的 ROM 器件，也是近年来被广泛应用的一种可用电擦除和编程的只读存储器，其主要特点是能在应用系统中进行在线读写，并且在断电情况下保存的数据信息不会丢失，它既能像 RAM 那样随机地进行改写，又能像 ROM 那样在掉电的情况下非易失地保存数据，可作为系统中可靠保存数据的存储器。其擦写次数可达 1 万次以上，数据可保存 10 年以上，故 E^2PROM 比 EPROM 具有更大的优越性。

当前的 E^2PROM 有两类产品：一种是采用并行的方式传送数据，如 Intel 2864（8K×8 位），称为并行 E^2PROM 芯片，这类芯片具有较高的传输速率；另一种是采用串行的方式传送数据，如 AT24C16（2K×8 位），称为串行 E^2PROM 芯片，这类芯片只用少数几个引脚来传送地址和数据，使芯片的引脚数、体积和功耗大为减少。通常情况，并行 E^2PROM 芯片既可存放程序又可存放数据，而串行 E^2PROM 芯片只存放数据。

由于 E^2PROM 兼有 RAM 和 ROM 的双重优点，因此，在计算机系统中使用 E^2PROM 以后，可使整机的系统应用变得更加灵活和方便。

3.3.5　快擦除读写存储器 Flash Memory

快擦除读写存储器又称闪速存储器，自问世之后即获得了迅速的发展，是非常有前途的存储器。这种存储器是在 EPROM 与 E^2PROM 基础上发展起来的一种新型的半导体存储器。它与 EPROM 一样，用单管来存储一位信息，与 E^2PROM 相同之处是用电来擦除。

由于快擦除读写存储器是以 EPROM 的存储单元为基础的，因此具有非易失性，在断电时也能保留所存储的内容，这使它优于需要持续供电才能存储信息的易失性存储器；它的单元结构和它所具有的 EPROM 基本特性，使得它的制造特别经济，其价格已低于 DRAM；它在密度增加时仍保持可测性，具有可靠性；可实现大规模电擦除，它的擦除功能可迅速清除整个存储器的所有内容；它的读取时间低于 90ns；采用快速脉冲编程方法，整个芯片编程时间短，可实现高速编程；可重复使用，目前，快擦除读写存储器可以擦写百万次以上。

总之，快擦除读写存储器兼有 ROM 和 RAM 的性能，又有 ROM 和 RAM 一样的高密度，同时又具有可靠的非易失性、电擦除性、低成本、低功耗等明显优点，这些优势是目前其他半导体存储器无法比拟的。

快擦除读写存储器展示了一种全新的个人计算机存储器技术，作为一种高密度、非易失的存储器，它特别适合作固态存储器。在便携式计算机、工业控制系统中得到了广泛的应用，尤其是笔记本电脑和掌上型袖珍电脑更是大量采用快擦除读写存储器做成的存储卡来取代磁盘，或以低成本和高可靠性替代静态 RAM 和动态 RAM，这样就可以节约电能，减轻重量，降低成本。目前，小容量软磁盘几乎已被快擦除读写存储器组成的优盘（又称 U 盘、闪存盘、拇指盘）取代。

综上所述，各种半导体存储器有各自的存储特点，它们的主要应用列于表 3-4 中。

表 3-4　各种半导体存储器的应用

存储器	应用	存储器	应用
SRAM	Cache	EPROM	用于产品试制阶段试编程序
DRAM	主存储器	E^2PROM	IC 卡上存储信息
ROM	固化程序、微程序控制器	Flash Memory	固态盘、IC 卡
PROM	自编程序，用于工业控制或电器中		

3.4　半导体存储器接口

通常内存储器是由多个半导体存储器芯片组成的。内存储器与 CPU 的连接，就是多个半导体存储器芯片与 CPU 芯片的连接。就 CPU 而言，CPU 与外部是通过数据总线 DB、地址总线 AB 和控制总线 CB 与外部交换信息的，因此，半导体存储器芯片与 CPU 的连接也就是与 CPU 的三种总线的连接。

3.4.1　存储器芯片与 CPU 连接时必须注意的问题

在存储器芯片与 CPU 的连接中，除了正确实现数据总线、地址总线和控制总线的连接以外，还需要考虑以下几方面的问题，使得存储器系统可以正确工作。

1. CPU 总线的负载能力

CPU 在设计时，一般输出线的直流负载能力为带一个 TTL 负载，而在连接中，CPU 的每一根地址线或数据线，都有可能连接多个存储器芯片。所以在存储器芯片与 CPU 的连接过程中，要考虑 CPU 外接多少个存储器芯片以及 CPU 与存储器芯片的物理距离等因素。现在的存储器芯片都为 MOS 电路，直流负载很小，主要是电容负载，故在小型系统中，CPU 可以直接与存储器芯片相连，而在较大的系统中，就要考虑加驱动器与存储器相连。

2. CPU 时序与存储器芯片存取速度的匹配

CPU 在执行取指令或存储读写操作时，都有固定的时序，因此需要考虑 CPU 与存储器芯片的速度匹配问题。具体讲，CPU 对存储器进行读写操作时，CPU 发出地址的读命令信号后，存储器必须在限定的时间内给出有效数据；而当 CPU 对存储器进行写操作时，存储器必须在写脉冲规定的时间内将数据写入指定的存储单元。因此，为了保证 CPU 对存储器的正确读写，就需要选择工作速度与 CPU 时序相匹配的存储器芯片，若存储器芯片已确定，则要考虑是否需要 CPU 插入 T_w 等待周期。

3. 存储器的地址分配

系统内存通常分为 ROM 和 RAM 两部分。ROM 用于存储系统固化程序及参数，RAM 分为系统区和用户区，系统区是监控程序或操作系统存放数据的区域，用户区又分为程序区和数据区两部分，分别用于存放用户程序和数据。所以内存分配是一个重要的问题。就目前而言，单片的存储器芯片容量有限，计算机的内存储器系统需要有多个芯片来组成，因此，针对存储器地址的分配，要知道哪些地址区需要 ROM，哪些区域需要 RAM，即在具体电路中需要明确地址译码与片选信号的产生。以 Intel 8086 CPU 为例，根据 8086 CPU 的特性，高地址区域

应该是 ROM 区域，低地址区域应该连接 RAM 芯片作为 RAM 系统区。

3.4.2　存储器的选址

微型计算机系统中，CPU 对内存进行读/写时，首先要对存储器芯片进行选择，使相应芯片的片选端$\overline{CS}$有效，称为片选，然后在选中的芯片内部再选择某一存储单元，称为字选。片选信号和字选信号均由 CPU 发出的地址信号经译码电路译码后产生。

片选信号由存储器芯片本身不使用的高位地址经外部译码电路按一定方式译码后产生，这部分译码电路在存储器芯片外部，是需要自行设计的部分。每个存储器芯片都有一定数量的地址输入端，在接收 CPU 输出的低位地址后，经内部译码电路译码后产生字选信号，选中芯片内部指定存储单元，这部分译码电路在芯片内部，不需要用户设计。

实现片选的存储器芯片外部译码电路的具体接法决定了存储器芯片的寻址空间。设计微型计算机的存储器系统时，要保证存储器芯片中的每个存储单元与实际地址一一对应，这样才能通过准确寻址对存储单元进行读写操作。在存储器系统中，通常使用存储器芯片本身不用的高位地址实现片选，一般有以下 3 种。

1. 线选法

线选法是指用某一条高位地址线直接作为存储器芯片的片选信号的方法。在简单的微型计算机系统中，由于存储容量不大，存储器芯片数也不多，可以使用存储器芯片本身不使用的单根地址线作片选信号，每个存储器芯片只用一根地址线选通。这种方法的优点是连接简单，无须专门的译码电路。但是局限性较大，首先表现在存储器寻址空间可能不连续，会造成大量的地址空间浪费；其次，若有多片存储器均使用线选法，可能会出现地址不唯一的现象，造成数据冲突，这是不允许的。所以线选法不适合于系统中存在多片存储器的情况。

2. 部分译码法

部分译码法是指将部分高位地址通过译码电路或译码器产生存储器片选信号的方法。该方法只使用部分高位线进行译码，从而产生片选信号，剩余高位线可以空闲或直接用作其他存储芯片的片选控制信号。使用这种方法可能会出现地址不唯一的现象。

3. 全译码法

全译码法是指存储器芯片本身不使用的高位地址全部参与译码的方法。在 CPU 输出的所有地址总线中，除了将低位地址线直接连至各存储器芯片的地址线外，余下的高位地址线全部送至译码电路或译码器，译码输出作为各芯片的片选信号。这种方法可以提供对全部存储空间的寻址能力。还可以使每片存储器的寻址空间唯一确定，而且是连续的，若安排合理，可避免空间浪费。

3.4.3　存储器的容量扩展

通常情况，单个存储器芯片的存储容量不能满足存储器系统的要求，所以一般都是由多个存储器芯片来构成微型计算机的内存储器系统。所谓存储器的容量扩展，就是将多个存储器芯片按一定方式组织在一起，构成内存储器模块，以满足 CPU 数据总线宽度的需要，或提供给 CPU 更大的存储空间。在进行存储器容量扩展时，首先要确定内存储器模块的大小，并根据所选择的存储器芯片容量，确定需要的存储器芯片数目；最后再将选择好的存储器芯片与 CPU 有机地连接起来。具体地有以下几种连接方法。

1. 位扩展

位扩展又称横向扩展，指扩展存储单元的位数，增加字长，而不需增加单元数。

RAM 芯片具有 1 位、4 位、和 8 位等不同结构，如 64K×1 位、256K×4 位、128K×8 位等。当内存储器的单元数与存储器芯片的字数相等，而存储器芯片的位数小于内存储器的字长时，就要进行位扩展。

位扩展的连接方法是：地址线、片选信号线、读/写信号线分别并连，而数据线串连。例如，要将 64K×1 位的芯片扩展为 64K×8 位的内存储器，其扩展连接方法如图 3-13 所示。

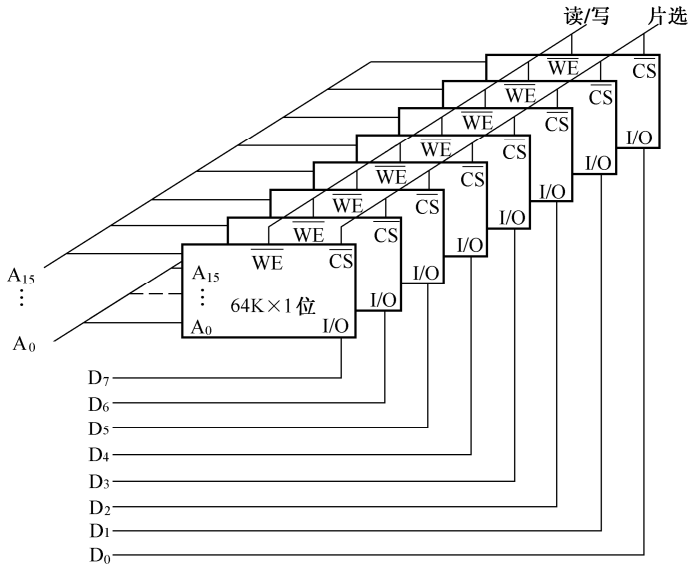


图 3-13　位扩展的连接

用 64K×1 位的芯片扩展为 64K×8 位的内存储器，需要 8 片存储器芯片，连接时，将 8 个芯片的地址线 A_{15}～A_0、读/写信号线 $\overline{WE}$、片选信号线 $\overline{CS}$ 分别并接在一起，而 8 个芯片的数据线（每片 1 位）则分别与 CPU 的数据线 D_7～D_0 相连。

2. 字扩展

字扩展又称纵向扩展，是指扩展存储单元的个数，存储单元的位数并不改变。

其方法是将地址线、数据线、读/写信号线分别并接在一起，而将片选信号线单独引出，用以决定每一芯片的地址范围，使存储器的地址空间为各个芯片地址空间之和。例如，用容量为 16K×8 位的存储器芯片组成一个 64K×8 位的内存储器，则需 4 片这样的存储器芯片，其扩展连接方法如图 3-14 所示。

连接时，将 4 个芯片的地址线 A_{13}～A_0、读/写信号线 $\overline{WE}$、数据线 D_7～D_0 分别并接在一起，而 4 个芯片的片选信号线 $\overline{CS}$ 单独引出，连至地址译码器的 4 个输出选择端。

参与片选译码的应为高位地址。由于每个存储器芯片的容量为 16K×8 位，芯片内的存储单元所需的地址线为 A_{13}～A_0（$16K=2^{14}$），故由 A_{15} 和 A_{14} 经地址译码器译码后，得到 4 个芯片的片选信号线 $\overline{Y}_3$～$\overline{Y}_0$，分别对应 A_{15} 和 A_{14} 的 11～00 编码；每个芯片内部地址 A_{13}～A_0 的地址范围为 0000H～3FFFH。由此将片选地址 A_{15}、A_{14} 与片内地址 A_{13}～A_0 统一在一起，最

后可以得出 4 个存储器芯片的存储单元地址范围，具体如表 3-5 所示。

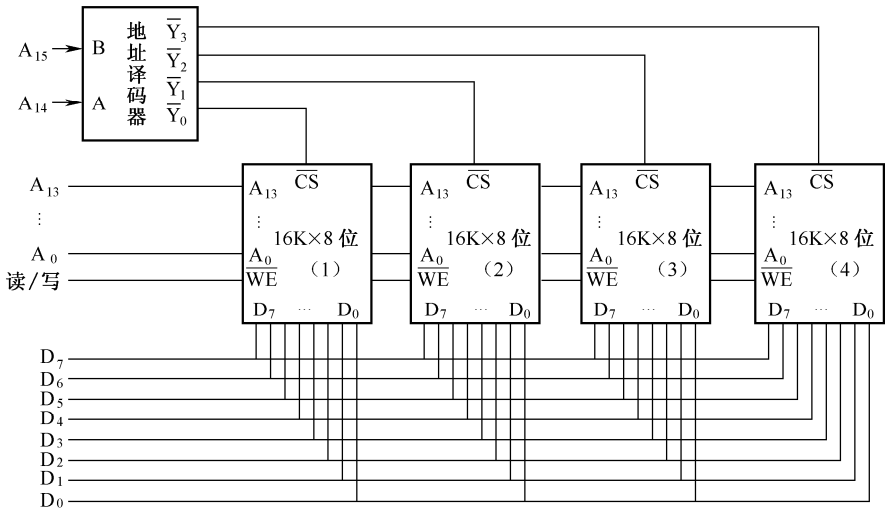


图 3-14　字扩展的连接

表 3-5　64KB（4 片 16KB 芯片字扩展）存储器单元地址范围

地址范围 / 芯片	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	寻址空间
芯片（1）	0　0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 … 1 1 1 1 1 1 1 1 1 1 1 1 1 1	0000H～3FFFH
芯片（2）	0　1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 … 1 1 1 1 1 1 1 1 1 1 1 1 1 1	4000H～7FFFH
芯片（3）	1　0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 … 1 1 1 1 1 1 1 1 1 1 1 1 1 1	8000H～BFFFH
芯片（4）	1　1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 … 1 1 1 1 1 1 1 1 1 1 1 1 1 1	C000H～FFFFH

3. 字位扩展

实际的存储器常常需要在字方向和位方向同时扩展。例如，若采用 16K×4 位的存储器芯

片组成 64K×8 位的内存储器，需要多少个这样的芯片呢？各芯片又如何连接在一起呢？

因为每个芯片的存储单元位数（4 位）以及存储单元数目（16K）都不满足内存储器容量要求（64K×8 位），所以需要进行位、字两个方向的扩展。

首先进行位扩展，满足每个存储单元为 8 位的要求，根据每芯片 4 位，分析出需要 2 片构成一组。然后以组为单位进行字扩展，满足存储单元数目要求，因为在每 2 个芯片组成的一组内，存储单元数仍为 16K，所以需要 4 组才能达到存储器单元数目为 64K 的要求。如此，需要 4 组，共计 8 片。

其扩展连接方法如图 3-15 所示。

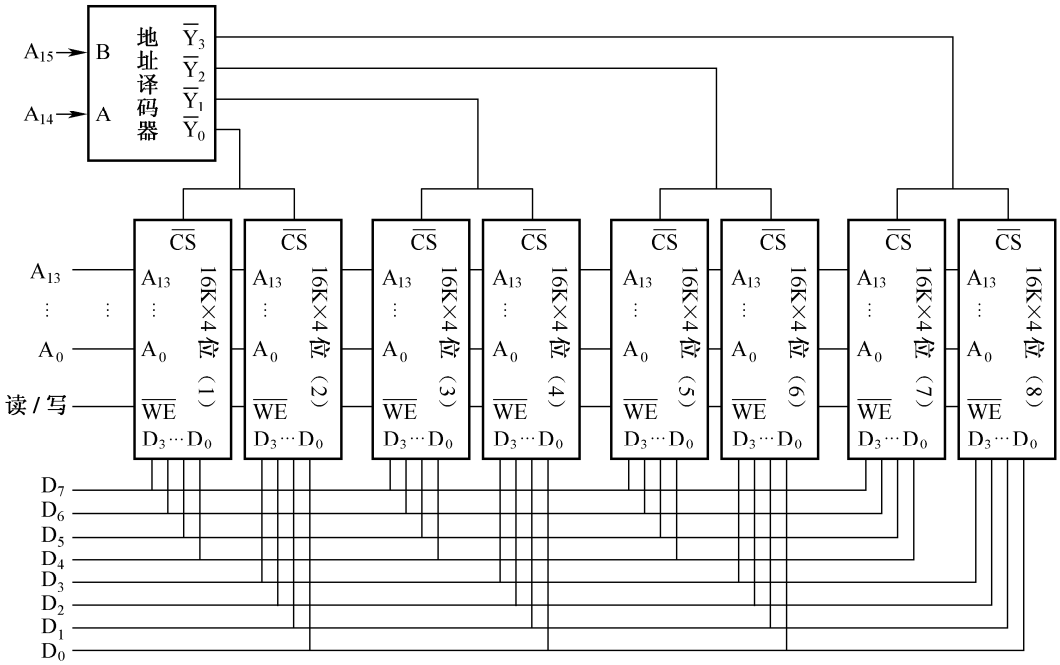


图 3-15　字位扩展的连接

由图 3-15 可见，寻址 64KB 容量的存储器需要 16 根地址线。这 16 根地址分为高位地址和低位地址两部分。首先各个芯片内的 14 根地址线 A_{13}～A_0 直接与系统地址总线的 A_{13}～A_0 并连在一起，用于片（组）内寻址；还需要 2 根高位地址线 A_{15} 和 A_{14} 经地址译码器译码后，产生 4 根选择信号线，用作组间寻址。同组芯片的 $\overline{CS}$ 端并接后，分别与 4 根选择信号线相接。同组内 2 个芯片的各自 4 位数据线 D_3～D_0 分别与系统数据总线 D_7～D_0 相连。读/写信号线与 8 个芯片的 $\overline{WE}$ 端并接在一起。这样，经过组合后就形成了容量为 64KB 的存储器。以此类推，就可以得到容量更大的存储器结构。

【例 3.1】使用 2114（1K×4 位）存储器芯片组成 2K×8 位的内存储器，需多少片这样的芯片？如何连接？

分析过程如下：

（1）根据存储器总容量及单片 2114 的容量，计算出所需芯片数：$\dfrac{2\text{K}\times 8\text{位}}{1\text{K}\times 4\text{位}}=4$（片）。

（2）根据存储器单元数及单片 2114 的单元数，计算出字扩展所需芯片组数：$\frac{2K}{1K}$=2（组）。

（3）根据存储器及单片 2114 的单元位数，计算出位扩展所需芯片数：$\frac{8位}{4位}$=2（片/组）。

（4）每个 2114 芯片内部的 A_9～A_0 分别与系统地址总线 A_9～A_0 并接；A_{10} 用来作片选端，接至两组芯片（4 片）的 $\overline{CS}$ 端，A_{10} 与第一组芯片的 $\overline{CS}$ 直接相连，经非门取反后与第二组芯片的 $\overline{CS}$ 相连。

（5）系统读/写信号直接与各个芯片的 $\overline{WE}$ 端相连。

（6）同组内 2 个芯片的各自 4 位数据线 I/O_3～I/O_0 分别与系统数据总线 D_7～D_0 相连。

（7）存储器容量扩展连接如图 3-16 所示。

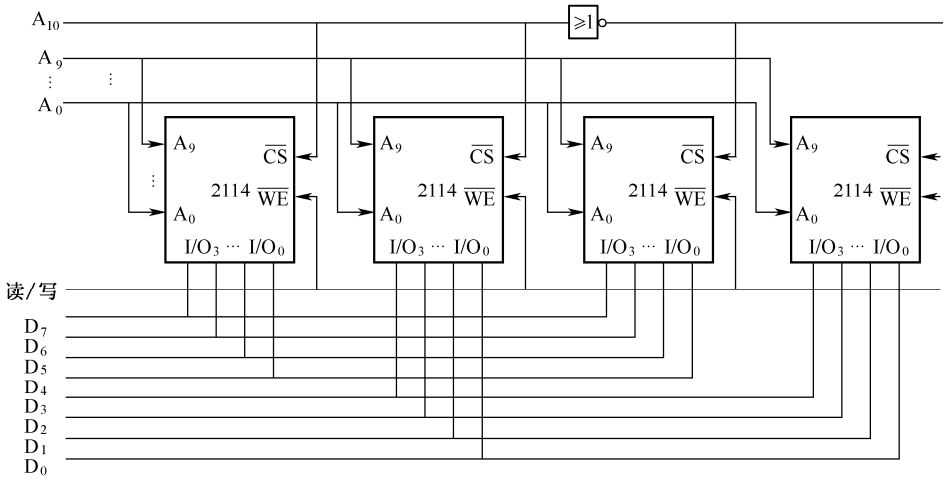


图 3-16　4 片 2114 扩展组成 2K×8 位存储器

【例 3.2】利用 74LS138 作为地址译码器（其功能表如表 3-6 所示），使用 4 片 6116（2K×8 位）存储器芯片组成 8KB 的存储器，存储器容量扩展连接如图 3-17 所示，试分析各存储器芯片的寻址范围。

表 3-6　74LS138 的功能表

G_1	$\overline{G}_{2A}$	$\overline{G}_{2B}$	C　B　A	输出
1	0	0	0　0　0	$\overline{Y}_0$=0，其他输出均为 1
			0　0　1	$\overline{Y}_1$=0，其他输出均为 1
			0　1　0	$\overline{Y}_2$=0，其他输出均为 1
			0　1　1	$\overline{Y}_3$=0，其他输出均为 1
			1　0　0	$\overline{Y}_4$=0，其他输出均为 1
			1　0　1	$\overline{Y}_5$=0，其他输出均为 1
			1　1　0	$\overline{Y}_6$=0，其他输出均为 1
			1　1　1	$\overline{Y}_7$=0，其他输出均为 1

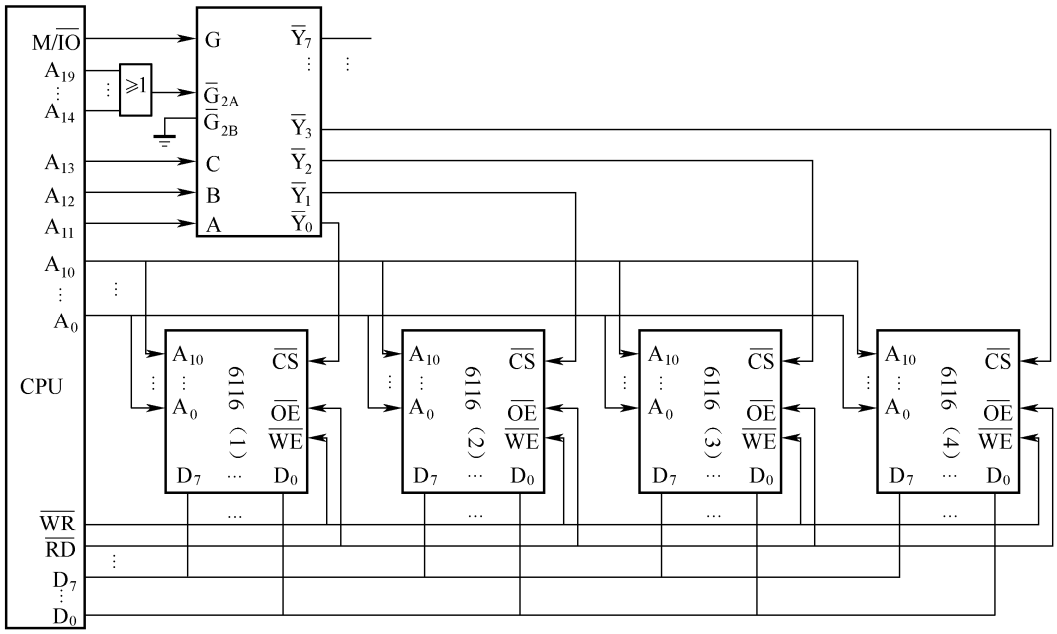


图 3-17　4 片 6116 扩展组成 8KB 存储器

分析过程如下：

（1）根据 74LS138 功能表可知，在 M/$\overline{IO}$=1，A_{19}～A_{14}=000000 的情况下，当 A_{13}～A_{11} 为 000、001、010、011 时，分别选中 1、2、3、4 号芯片。

（2）各芯片的片内地址为 A_{10}～A_0，片内地址范围为 0000H～07FFH。

（3）将 A_{19}～A_0 统一起来，得出 4 个存储器芯片的存储单元地址范围，见表 3-7。

表 3-7　8KB（4 片 6116）存储器单元地址范围

地址范围 / 芯片	A_{19}～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	寻址空间
6116（1）	0000000	0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 … 1 1 1 1 1 1 1 1 1 1 1	00000H～007FFH
6116（2）		0 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 … 1 1 1 1 1 1 1 1 1 1 1	00800H～00FFFH
6116（3）		1 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 … 1 1 1 1 1 1 1 1 1 1 1	01000H～017FFH

续表

地址范围 / 芯片	A_{19}～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	寻址空间
6116（4）		1 1	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 … 1 1 1 1 1 1 1 1 1 1 1	01800H～01FFFH

3.4.4　典型 CPU 与存储器的连接

一台微型计算机的内存储器容量往往为几百 MB，无论字长是 8 位、16 位还是 32 位，都是以 8 位为 1 个字节，以字节为单位进行编址的，每个字节单元拥有一个唯一的物理地址。

8086 和 8088 CPU 地址总线均为 20 位，可管理 1MB 存储空间；8086 CPU 的内部总线和外部总线均为 16 位，而 8088 CPU 则是 8086 CPU 的改造型，其内部结构与 8086 CPU 基本相同，但外部数据总线却为 8 位，由此导致两种 CPU 与存储器在连接上的不同，8086 微型计算机被称为 16 位机，而 8088 微型计算机则被称为准 16 位机。80286 CPU 的数据总线也为 16 位，24 位地址总线可管理 16MB 的存储空间。80386 和 80486 CPU 数据总线均为 32 位，地址总线也为 32 位，可管理更大的内存储器空间。

1. 8088 CPU 与存储器的连接

8088 存储系统与 CPU 的连接比较简单，其存储器结构如图 3-18 所示。

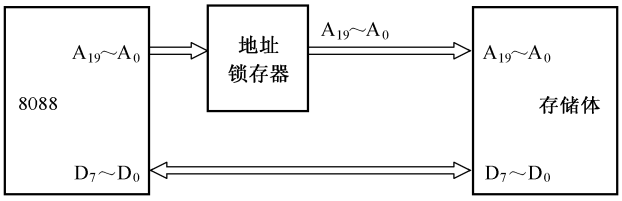


图 3-18　8808 存储器结构

在一个总线周期内，8088 CPU 对存储器只能按字节操作，而 8086 CPU 对存储器既可以按字节操作，也可以按 16 位字操作；显然，由于 8088 CPU 对 1 个 16 位数据的存取必须经过 2 次，所以 8086 CPU 对存储器的访问速度必然高于 8088 CPU。

2. 8086 CPU 与存储器的连接

在 8086 系统中，存储器采用分体结构，即将 1MB 的存储空间分成两个 512KB 的存储体，一个存储体中包含偶数地址单元，称为偶地址存储体；另一个存储体包含奇数地址单元，称为奇地址存储体。偶地址存储体与 8086 CPU 的低 8 位数据总线（D_7～D_0）相连，奇地址存储体与 8086 CPU 的高 8 位数据总线（D_{15}～D_8）相连。A_{19}～A1 是体内地址线，它们并行地连接到两个存储体上，如图 3-19 所示。

为使 CPU 不仅能访问存储器中的一个字节（访问一个存储体），也能访问存储器中的一个字（同时访问两个存储体），8086 给出了一个有用的控制信号 $\overline{\text{BHE}}$。$\overline{\text{BHE}}$ 称为数据总线高 8 位允许信号，当 $\overline{\text{BHE}}$ =1 时，奇地址存储体被禁止读/写；当 $\overline{\text{BHE}}$ =0 时，可以对奇地址存储体

进行读/写操作。当 A_0=0 时，可以对偶地址存储体进行读/写。$\overline{BHE}$ 与地址线 A_0 配合使用实现了对字和字节寻址的控制，如表 3-8 所示。

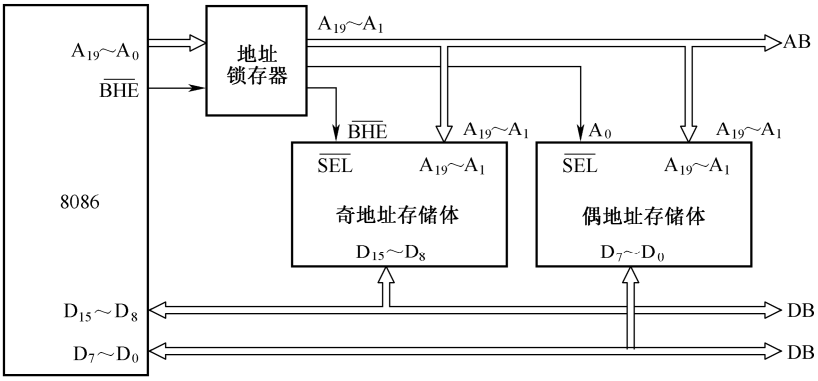


图 3-19　8086 存储器结构

表 3-8　8086 存储体的操作

$\overline{BHE}$	A_0	传送的信息
0	0	同时访问两个存储体，传送一个字（D_{15}～D_0）
0	1	只访问奇地址存储体，传送高位字节（D_{15}～D_8）
1	0	只访问偶地址存储体，传送低位字节（D_7～D_0）
1	1	无效

3．80286 CPU 与存储器的连接

80286 CPU 的地址线为 24 根，最大寻址空间可达 16MB，其存储器结构与 8086 存储器结构相似，也将存储体分为奇地址存储体和偶地址存储体，由 $\overline{BHE}$ 和 A_0 作为奇偶存储体的选择信号。具体如图 3-20 所示。

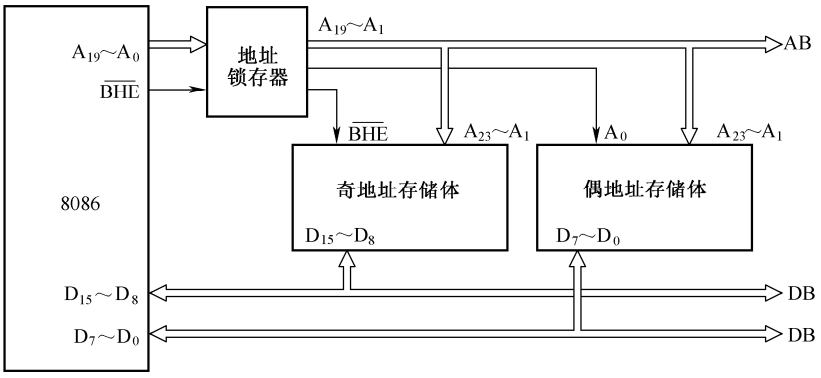


图 3-20　80286 存储器结构

4．80386/80486 CPU 与存储器的连接

80386/80486 CPU 是 32 位的微处理器，有 32 位数据总线和 32 位地址总线。为了实现对

一个 32 位数据中的 8 位、16 位或 32 位的灵活访问，系统将 32 位数据以 8 位一组，分成 4 个字节，同时 CPU 设有 4 个引脚 $\overline{BE_3}$ ～ $\overline{BE_0}$，以控制对不同字节数据的访问。$\overline{BE_3}$ ～ $\overline{BE_0}$ 由 CPU 根据指令类型产生，$\overline{BE_3}$ ～ $\overline{BE_0}$ 的功能如表 3-9 所示。

表 3-9　$\overline{BE_3}$ ～ $\overline{BE_0}$ 功能表

字节允许				允许访问的的数据位			
$\overline{BE_3}$	$\overline{BE_2}$	$\overline{BE_1}$	$\overline{BE_0}$	D_{31}～D_{24}	D_{23}～D_{16}	D_{15}～D_8	D_7～D_0
1	1	1	0	—	—	—	D_7～D_0
1	1	0	1	—	—	D_{15}～D_8	—
1	0	1	1	—	D_{23}～D_{16}	—	—
0	1	1	1	D_{31}～D_{24}	—	—	—
1	1	0	0	—	—	D_{15}～D_8	D_7～D_0
1	0	0	1	—	D_{23}～D_{16}	D_{15}～D_8	—
0	0	1	1	D_{31}～D_{24}	D_{23}～D_{16}	—	—
1	0	0	0	—	D_{23}～D_{16}	D_{15}～D_8	D_7～D_0
0	0	0	1	D_{31}～D_{24}	D_{23}～D_{16}	D_{15}～D_8	—
0	0	0	0	D_{31}～D_{24}	D_{23}～D_{16}	D_{15}～D_8	D_7～D_0

80386/80486 系统在进行存储器系统设计时，常把内存储器分为 4 个存储体，依次存放 32 位数据的 4 个字节，每个存储体的 8 位数据线依次并行连接到系统数据总线 D_{31}～D_{24}、D_{23}～D_{16}、D_{15}～D_8、D_7～D_0 上，如图 3-21 所示。CPU 的 32 位地址线中，高位地址 A_{31}～A_2 直接输出，低 2 位地址 A_1、A_0 经由内部编码产生 $\overline{BE_3}$ ～ $\overline{BE_0}$，以选择不同字节。

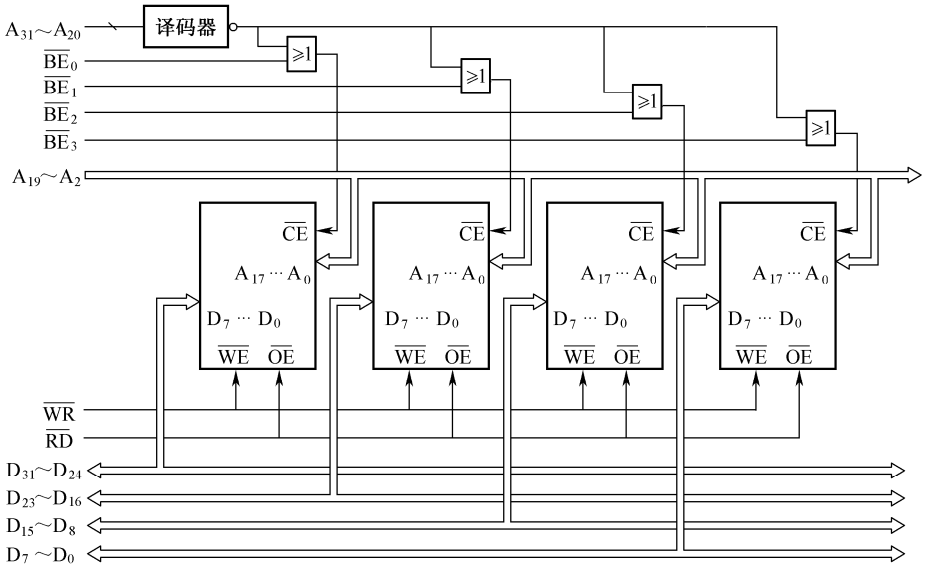


图 3-21　32 位字存储器结构

尽管 CPU 具有 4GB（2^{32}=4G）的寻址能力，但实际应用中无须那么大的内存容量。若每

个存储体的容量为 256KB，则 4 个存储体构成的内存储器容量为 1MB，所以至少需要 20 位（2^{20}=1M）地址进行内存储器单元寻址，如图 3-21 所示，每个存储体的 18 位（256K=2^{18}）地址 A17～A0 接至 CPU 的系统地址线 A_{19}～A2，片选信号 $\overline{CE}$ 由高位地址的译码结果与 $\overline{BE_3}$～$\overline{BE_0}$ 相“或”后产生。一旦地址码确定，$\overline{BE_3}$～$\overline{BE_0}$ 决定某一个或某几个存储体被选中，A_{19}～A_2 确定被选中的存储体内指定的字节单元，然后即可对选中单元进行读/写操作。

3.4.5 单列直插式存储器（SIMM）和双列直插式存储器（DIMM）

SIMM 和 DIMM 是把内存储器芯片焊接在条形印刷电路板上制成的，俗称内存条。在内存条上不仅有存储器芯片组，还有地址译码等辅助电路。按内存条印刷电路板上引脚数的不同，SIMM 分为 30 线和 72 线两种，DIMM 主要为 168 线。

内存条的容量有多种，有 256KB、512KB、1MB、4MB、8MB、16MB、32MB、64MB、128MB、256MB，甚至更大。单元数有 8 位和 9 位两种，9 位中有一位用于奇偶校验。

30 线内存条体积比较小，仅提供 8 位有效数据；72 线内存条体积比较大，可提供 32 位有效数据；168 线 DIMM 有 64 位数据位。

1．30 线 SIMM

30 线内存条提供 8 位或 9 位数据，在主机系统主板上有其插座。由 4 块内存条构成一组，可提供 32 位数据。30 线 SIMM 的引脚信号如表 3-10 所示。

表 3-10　30 线 SIMM 引脚信号

引脚	信号	引脚	信号	引脚	信号	引脚	信号	引脚	信号	引脚	信号
1	Vcc	6	DQ_1	11	A_4	16	DQ_4	21	$\overline{WE}$	26	QP
2	$\overline{CAS}$	7	A_2	12	A_5	17	A_8	22	GND	27	$\overline{RAS}$
3	DQ_0	8	A_3	13	DQ_3	18	A_9	23	DQ_6	28	$\overline{CASP}$
4	A_0	9	GND	14	A_6	19	A_{10}	24	A_{11}	29	DP
5	A_1	10	DQ_2	15	A_7	20	DQ_5	25	DQ_7	30	Vcc

例如，由 2 个 1M×4 位存储器芯片构成的 1MB 内存条，如图 3-22 所示。

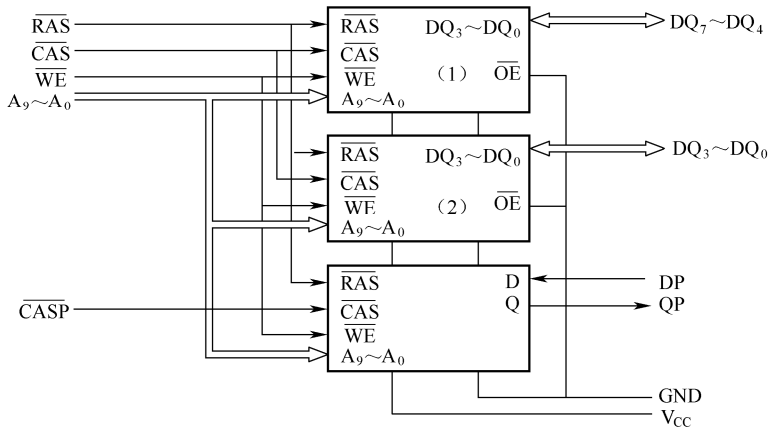


图 3-22　由 2 片 1M×4 位芯片构成的 1MB 内存条

内存条有 8 位双向数据线 DQ_7～DQ_0，另有一个 DP 奇偶校验位输入线，和一个 QP 奇偶校验位输出线；根据 1MB 内存容量，需要 20 位地址线，这里采用分时复用技术，将 20 位地址信号通过 A_9～A_0 分两次送至内存条，同时 $\overline{RAS}$ 和 $\overline{CAS}$ 配合控制内部行、列地址译码；$\overline{RAS}$ 还用于动态存储器刷新控制；$\overline{WE}$ 用于读/写控制。

一般主机板上有多个内存储器插座，可安装多个内存条。对于 16 位微处理器，使用 30 线 SIMM 时安装数量应为偶数；对于 32 位微处理器，使用 30 线 SIMM 时安装数量应为 4 的倍数。

2. 72 线 SIMM

72 线内存条提供 32 位数据传输，在主机系统主板上有其多个插座。若用 2 块构成一组，可提供 64 位数据。72 线 SIMM 的引脚信号如表 3-11 所示。

表 3-11　72 线 SIMM 引脚信号

引脚	信号	引脚	信号	引脚	信号	引脚	信号
1	GND	19	NC	37	DQ_{17}	55	DQ_{12}
2	DQ_0	20	DQ_4	38	DQ_{35}	56	DQ_{30}
3	DQ_{18}	21	DQ_{22}	39	GND	57	DQ_{13}
4	DQ_1	22	DQ_5	40	$\overline{CAS_0}$	58	DQ_{31}
5	DQ_{19}	23	DQ_{23}	41	$\overline{CAS_2}$	59	Vcc
6	DQ_2	24	DQ_6	42	$\overline{CAS_3}$	60	DQ_{32}
7	DQ_{20}	25	DQ_{24}	43	$\overline{CAS_1}$	61	DQ_{14}
8	DQ_3	26	DQ_7	44	$\overline{RAS_0}$	62	DQ_{33}
9	DQ_{21}	27	DQ_{25}	45	NC	63	DQ_{15}
10	Vcc	28	A_7	46	NC	64	DQ_{34}
11	A_0	29	NC	47	$\overline{WE}$	65	DQ_{16}
12	A_1	30	Vcc	48	NC	66	NC
13	A_2	31	A_8	49	DQ_9	67	PD_1
14	A_3	32	NC	50	DQ_{27}	68	PD_2
15	A_4	33	NC	51	DQ_{10}	69	PD_3
16	A_5	34	$\overline{RAS_2}$	52	DQ_{28}	70	PD_4
17	A_6	35	DQ_{26}	53	DQ_{11}	71	NC
18	A_7	36	DQ_8	54	DQ_{29}	72	GND

3. 168 线 DIMM

DIMM 与 SIMM 类似，不同的只是 DIMM 内存条两侧的引线不像 SIMM 是互通的，而是各自独立传输信号，因此可以满足更多数据信号的传送需要。168 线 DIMM 内存条支持 64 位的数据传输。

如图 3-23 所示，168 线 DIMM 内存条两侧各 84 线，DIMM 内存条引线端有两个卡口，从而避免插入插槽时，因错误地将内存条反向插入而导致烧毁。

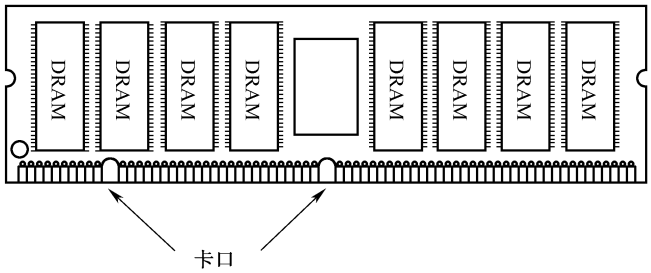


图 3-23　典型的 128 线 DIMM 结构图

目前，微型计算机系统主板上一般有 2～4 个内存储器插座，可安装 2～4 个内存条。对于 Pentium 微处理器，有时需要一次存取 64 位数据，若采用 72 线内存条，应插接偶数个内存条。由于 DIMM 具有 64 位数据传输能力，所以 DIMM 可单条使用，而且不同容量的 DIMM 可以混合使用。

随着 Pentium 微处理器的普及，DIMM 内存条逐渐取代了 SIMM，广泛应用于微型计算机中。

3.5　存储体系结构

存储器有三个主要性能指标：容量、速度和价格（每位价格），计算机对存储器的这些性能指标的基本要求是容量大、速度快和成本低。但是要想在一个存储器中同时兼顾这些指标是很困难的，有时是相互矛盾的：速度越快，每位价格就越高；容量越大，每位价格就越低；容量越大，速度就越慢。为了解决存储器的容量、速度和价格之间的矛盾，人们在不断地研制新的存储器件和改进存储性能的同时，还从存储系统体系上不断研究合理的结构模式。

3.5.1　存储器系统的层次结构

人们把各种不同存储容量、存取速度和价格的存储器按层次结构组织起来，并通过管理软件和辅助硬件有机地组成统一的整体，使所存放的程序和数据按层次分布在各级存储器中，形成存储器系统的多级层次结构。一般计算机存储器系统的多级层次结构如图 3-24 所示。

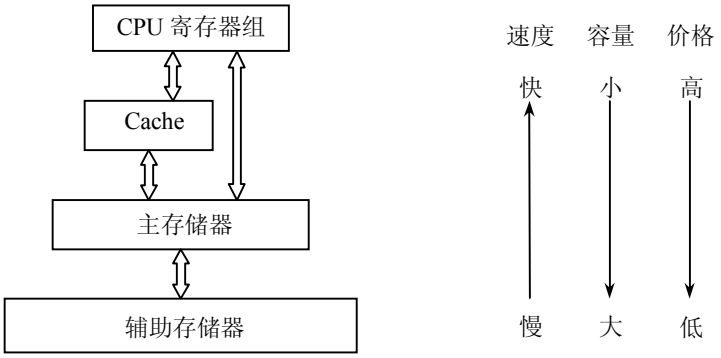


图 3-24　存储器系统多级层次结构

由图 3-24 可见，计算机存储器系统多级层次结构主要由 CPU 内部寄存器、高速缓冲存储

器（Cache）、主存储器和辅助存储器组成。辅助存储器、主存储器、Cache 以及 CPU 内部的寄存器都具有数据存储功能，由它们构成的存储器组织能够充分发挥存储速度快、容量大、价格低的特点。

CPU 内部寄存器可读可写，位于 CPU 内部，并与 CPU 速度相匹配。它们的功能很强，但是每位价格较高，数量也不能太多。主要在 CPU 运行指令过程中，用于指令和数据的暂存。

主存储器的容量要比CPU内部寄存器的数量大得多，可作为CPU内部寄存器的后备支持。主存储器是计算机工作时必不可少的存储部件，它和 CPU 一起构成主机的骨干部件。由于主存储器的每位成本较高，数量也不能很大，并且在断电后 RAM 中的数据全部丢失，所以必须有更大容量的非易失性存储后援的支持。

辅助存储器是独立的存储器，多由软磁盘、硬磁盘、光盘以及一些可移动式存储器构成。相对于主存，它具有存储容量大、成本低，断电后信息不丢失等优点，所以辅助存储器作为海量存储器，主要用于计算机的大容量程序和数据的长久保存。但是它也有一些如读写速度慢、CPU 不能直接访问等缺点。

随着计算机技术的发展，CPU 的工作速度不断提高，远远超过了以 DRAM 为基础的主存储器，为了不影响 CPU 的效率，人们在 CPU 与主存储器之间使用了高速缓冲存储器（Cache），Cache 的存取速度接近 CPU 的工作速度，作为主存储器的副本，可直接向 CPU 提供程序和数据。在高档微型计算机中，使用了两级 Cache，一级设在 CPU 内部，二级设在系统主板上，从而进一步提高向 CPU 提供数据的速度。

如图 3-24 所示，在存储器系统多级层次结构中，由上而下，其容量逐渐增大，速度逐级降低，成本则逐次减少。这样构成的整个系统可以接近 CPU 的高速度，并获得大容量辅存的支持，价格也能为用户所承受。如此，多级层次结构的计算机存储系统有效地解决了存储器的速度、容量和价格之间矛盾。

整个结构又可以看成两个层次：主存——辅存层次、Cache——主存层次。这两个层次系统中的每种存储器都不再是孤立的存储器，而是一个有机的整体。它们在计算机操作系统和辅助硬件的管理下工作，可把主存—辅存层次作为一个存储整体，形成的可寻址空间远远大于主存空间，而且由于辅存容量大、价格低，使得存储系统的整体平均价格降低。另外，由于 Cache 的存取速度和 CPU 的工作速度相当，故 Cache——主存层次可以缩小主存与 CPU 之间的速度差距，从整体上提高存储器系统的存取速度。

3.5.2　多体存储结构

CPU 需要频繁访问主存储器，但是主存储器的存取速度跟不上系统的需求，这是影响整个计算机系统性能的重要问题。解决这个问题的最直接的办法就是采用多个存储器并行访问和交叉访问，从而使 CPU 在一个存储周期内可以同时访问多个数据。

例如，一般主存储器在一个存储周期内只能访问到一个存储字，如图 3-25（a）所示，把多个这样的存储器组合后，可构成多体存储结构，如图 3-25（b）所示。

对多存储体交叉访问通常有两种工作方式：地址码高位交叉、地址码低位交叉。

1. 高位交叉访问存储器

以每个存储体内有 4 个存储单元，4 个这样的存储体构成的存储器为例，地址码高位交叉访问存储器如图 3-26 所示。

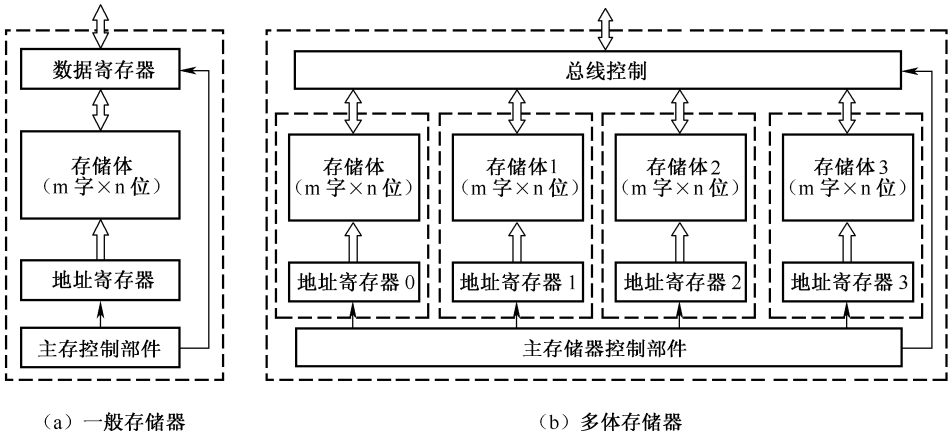


图 3-25　多体存储器与一般存储器比较

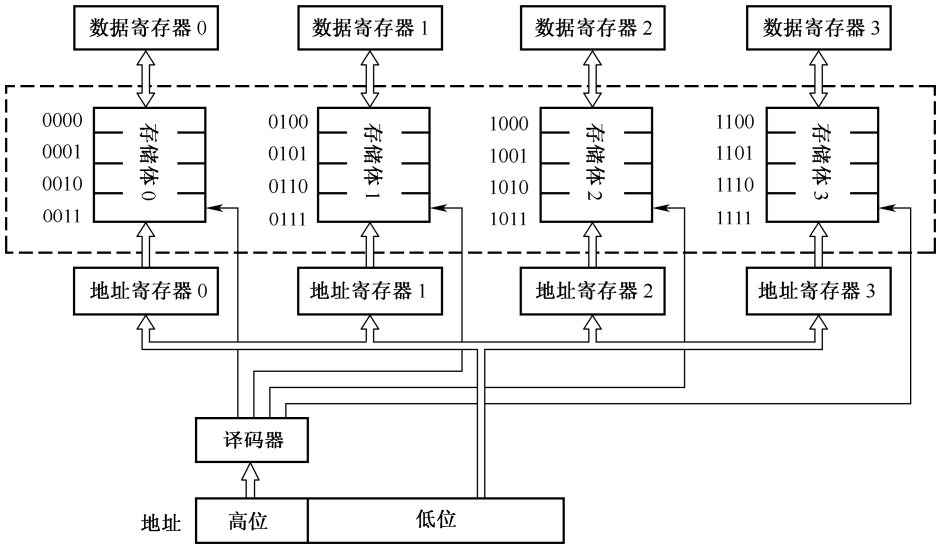


图 3-26　高位交叉访问存储器的结构

地址码的低位部分直接送至各个存储体，作为各个存储体的体内地址；高位地址部分送至译码器，用来区分存储体的体号。高位交叉访问存储器要求每个存储模块都有各自独立的存储体和控制部件，控制部件包括地址寄存器、数据寄存器、驱动放大电路、地址译码电路和读写控制电路等，每个独立存储模块均可独立工作。因此，高位交叉访问存储器具备了并行工作的条件，这种存储器不仅提高了存取速度，更主要的目的是用来扩大存储器容量。

目前，大部分计算机系统中的主存储器都采用模块结构，用户可以根据自己的不同需要随时方便地改变主存储器的容量，例如用 4 个 16MB 的模块（内存条）可以构成一个 64MB 的主存储器。

2. 低位交叉访问存储器

以每个存储体内有 4 个存储单元，4 个这样的存储体构成的存储器为例，地址码低位交叉访问存储器如图 3-27 所示。

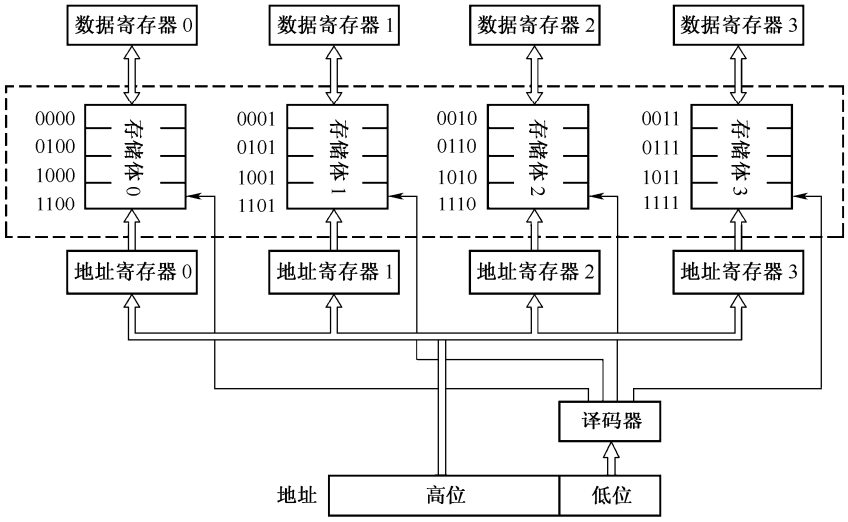


图 3-27　低位交叉访问存储器的结构

低位交叉访问存储器的地址码使用方法与高位交叉访问方式恰恰相反，其低位地址送至译码器，用来区分存储体的体号；高位地址直接送存储体，做为各个存储体的体内地址。

低位交叉访问存储器的主要目的是提高存储器的访问速度。当然，在提高访问速度的同时，由于增加了存储器模块的数目，也就增加了存储器的容量。

3.5.3　高速缓冲存储器（Cache）

高速缓冲存储器（Cache）是一种存储容量较小但存取速度却很快的存储器，它位于 CPU 和主存之间，用来存放 CPU 频繁使用的指令和数据。由于使用 Cache 后可以减少对慢速主存的访问次数，解决了 CPU 与主存之间的速度差异，所以提高了 CPU 的工作效率。目前，在高档微型计算机中广泛使用高速缓冲存储器技术。

1．Cache 工作原理

在半导体存储器中，只有 SRAM 的存取速度与 CPU 的工作速度接近，但这种 SRAM 的功耗大、价格较贵、集成度低。而 DRAM 的功耗小、成本低、集成度高，但是存取速度却远远跟不上 CPU 的工作速度。于是产生了一种折衷的分级处理办法：在以 DRAM 为基础的大容量主存与高速 CPU 之间增加一个容量较小的 SRAM 作为 Cache，用于保存那些使用频率较高的主存副本，CPU 可以直接从 Cache 中快速地访问所需的指令和数据。这种 Cache—主存结构存储系统弥补了 CPU 与主存在性能上的差距，使 CPU 访存时，不需插入等待周期，便能读写主存信息在 Cache 中的副本，既弥补了主存较慢的性能缺陷，又保证了存储系统的性能价格比。

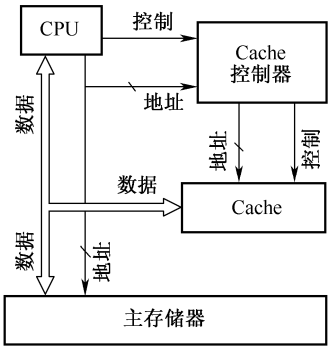


图 3-28　Cache—主存结构存储系统

Cache—主存结构存储系统一般由高速缓冲存储器主存储器 DRAM 和高速缓存控制器组成，如图 3-28 所示。

在高速缓冲存储系统中，所有数据和指令都存放在主存储器内，而使用频率较高的数据和指令则复制 Cache 中。Cache 的内容是在读写过程中逐步调入的，是主存中部分内容的副本。主存和 Cache 的存储区均划分成多个块，每块由多个单元组成，主存与 Cache 之间以块为单位交换信息。

信息块调往 Cache 时的存放地址与它在主存时的地址不一致，但两者之间有一定的对应关系，这种对应关系称为地址映射函数。将主存地址变换成 Cache 地址的过程是通过 Cache 控制器内的地址变换机构按照所采用的地址映射函数自动完成的。

当 CPU 访问存储器时，首先检查 Cache，如果要访问的信息已在 Cache 中，CPU 就能很快地访问它，这种情况称为“命中”；如果信息不在 Cache 中，这种情况称为“不命中”。若“不命中”，CPU 必须从主存储器中提取信息，同时还把含有这个信息的整个数据块从主存中取出送到 Cache 中。CPU 访问 Cache 时，找到所需要信息的百分比称为命中率，命中率是 Cache 的一个重要指标，命中率越高，在 Cache 中找到所需指令和数据的可能性就越大，CPU 去访问主存储器的次数就越少。

Cache 的全部功能由硬件实现，并且对程序员来说是“透明”的，就像不存在一样，程序员不需要明确知道 Cache 及 Cache 控制器的存在。

2. 替换算法

当新的主存块需要调入 Cache，而它的可用位置已被占用时，就需要替换算法来解决。一个好的替换算法既要考虑提高访问 Cache 的命中率，又要考虑容易实现替换，从而减少 CPU 访问主存的机率。替换算法有多种，通常采用以下两种算法。

（1）先进先出法（FIFO）。这种算法是把最早调入 Cache 的信息块替换掉。为了实现这种算法，需要在地址变换表中设置一个历史位，每当有一个新块调入 Cache 时，就将已进入 Cache 的所有信息块的历史位加 1。当需要进行替换时，只要将历史位值最大的信息块替换掉即可。这种算法比较简单，容易实现，但不一定合理，因为有些信息块虽然调入较早，但可能仍需使用。

（2）最近最少使用法（LRU）。LRU 算法是把近期使用最少的信息块替换掉。这种算法利用了程序访问的局部性原理：在时间上，程序即将用到的信息很可能就是正在使用的信息；在空间上，程序即将用到的信息很可能与当前正在使用的信息临近。所以，最近使用的信息块应尽量保留在 Cache 中。

LRU 算法要求随时记录 Cache 中各信息块的使用情况。这样，要为每个信息块设置一个计数器，以便确定哪个信息块是近期最少使用的。LRU 算法还可用堆栈来实现，也称为堆栈型算法。当堆栈已满，却又有一个信息块需要调入 Cache 时，首先检查堆栈中是否已经有该信息块。若有，则将该信息块从堆栈中取出并压入堆栈的栈顶；如果没有，则将该信息块直接压入栈顶，于是原来在栈底的信息块就成为被替换的信息块而被压出堆栈，这样就能够保证任何时候栈顶的信息块都是刚被访问过的信息块，而栈底的块总是很久没有被访问过的块。这种算法与 FIFO 相比可以获得较高的命中率。

3. 多层次 Cache

（1）多层次 Cache 结构。目前，微型计算机中都采用两级 Cache，即在 CPU 与主存之间设置了二级 Cache，位于系统主板上。又在 CPU 芯片内部集成了超高速一级 Cache。Pentium 微处理器使用两级 Cache，一级 Cache 的速度要比二级 Cache 快得多，更接近于 CPU，二级

Cache容量和密度高于一级Cache，两级Cache之间有专用总线相连。两级Cache之间以及Cache与主存之间的映射、替换算法和读写操作，全部由辅助硬件来完成，从而实现了 Cache 的高速处理功能。

（2）指令 Cache 和数据 Cache。Pentium 微处理器的二级 Cache 容量大小可以是 256KB 或 512KB，一级 Cache 的容量是 16KB。这 16KB 中分为 8KB 数据 Cache 和 8KB 指令 Cache，Pentium CPU 可以同时访问指令 Cache 和数据 Cache。由于片内一级 Cache 存取速度很快，并且由于程序访问的局部性，即有可能要访问 Cache 同一位置许多次，这样就减少了 Pentium 对外部总线的访问次数和使用频率，极大地提高了系统的存取速度。

4. Cache 一致性问题

当要访问的信息字在 Cache 中时，若是读操作，则 CPU 可以直接从 Cache 中读取数据，不涉及主存。若是写操作，则 Cache 和主存中相应两个单元的内容都需要改变。这时有两种处理办法：一种方法是 Cache 单元和主存中相应单元同时被修改，称为“直通存储法”；另一种方法是只修改 Cache 单元的内容，同时用一个标志位作为标志，当有标志的信息块从 Cache 中移去时再修改相应的主存单元，把修改信息一次性写回主存。显然“直通存储法”比较简单，但对于需要多次修改的单元来说，可能导致不必要的主存重复写工作。

3.5.4　虚拟存储器

虚拟存储器（Virtual Memory）是在“主存—辅存”层次结构上进一步发展和完善的存储管理技术。虚拟存储器把主存和辅存视为一个统一的虚拟主存，提供比实际主存容量大得多的，可使编程空间不受限制的虚存空间；在程序中使用虚地址，使程序不必作任何修改，即可用接近主存的速度在这个虚拟存储器上运行。使得在用户心目中，计算机系统好像只有一个大容量、高速度、使用方便的存储器，而没有主存、辅存之分。

目前，几乎所有的计算机都采用虚拟存储器系统。

1. 虚拟存储器工作原理

虚拟存储器系统对主存和辅存构成的虚存空间重新进行统一编址，称为虚地址。虚存空间并不是主存容量加上辅存容量，而是一个比主存大得多的虚拟空间，它与主存和辅存空间的容量无关。虚地址并非主存实际物理地址，例如 80386/80486 的虚地址码长 46 位，虚存空间可达 64TB（2^{46}），这是任何计算机系统中主存储器所不可能达到的容量，而主存实际物理地址只有 32 位。

计算机程序员可以按虚地址空间来编制程序。在程序运行时，只把虚地址空间的一小部分映射到主存，其余大部分虚地址空间则映射到辅存（如大容量硬盘）上。当按虚地址访问虚拟存储器时，存储器管理部件首先查看该虚地址所对应的内容是否已在主存中，若已在主存中（命中），就将该虚地址自动转换为主存实际物理地址，对主存进行访问；若不在主存中（未命中），就将虚拟存储器中待访问的程序或数据由辅存调入主存，再进行访问。故每次访问虚拟存储器时，都必须进行虚地址向实地址的转换。虚拟存储器是通过存储管理部件和操作系统自动管理和调度的，主存和辅存的地位和性质并未改变，但辅存相对于每个用户来说是透明的，这样大大方便了用户的使用。

2. 虚拟存储器与 Cache 的对比

“主存—辅存”层次与“Cache—主存”层次从原理上看是相同的，因而它们有很多相似

之处，如它们都采用地址变换及映射方法和替换策略，对于程序员来说 Cache 和虚拟存储器都是透明的。但虚拟存储器和 Cache 仍有很明显的区别。

（1）Cache 用于弥补主存与 CPU 的速度差距；而虚拟存储器则用来弥补主存与辅存之间的容量差距。

（2）CPU 可以直接访问 Cache；但 CPU 却不能直接访问辅存。

（3）Cache 每次传送的信息块是定长的，且信息块短小，只有几个或几十个字节；而虚拟存储器每次调动的信息块较大，可达几百或几千个字节，可以分页（信息块定长）或分段（信息块可变长）调动。

（4）Cache 操作全部由辅助硬件实现；而虚拟存储器则由辅助硬件（存储管理部件）和辅助软件（操作系统的存储管理软件）综合管理。

3. 虚拟存储器的分类

在实际应用中，根据如何对主存空间与磁盘空间进行分区管理、虚实地址怎样转换、采取何种替换算法等，可以有页式、段式和段页式 3 类虚拟存储器。

（1）页式虚拟存储器。页式虚拟存储器是以“页”为信息传送单位的虚拟存储器。在页式虚拟存储系统中，将虚拟空间等分成固定大小的页，页面大小随机器而异，一般计算机系统中一页的大小为 1～16KB。虚存空间中所划分的页称为“虚页”，主存空间也等分成同样大小的页，称为“实页”。页面都是由 0 开始顺序编号的，分别称为虚页号和实页号。

虚地址分成虚页号和页内地址两部分。虚页号占用高位部分，低位部分则为页内地址。主存的实际物理地址也由两部分组成：实页号和页内地址，实页号占用高位部分，低位部分为页内地址。页内地址的长度取决于页面大小，实页号的长度取决于主存的容量。

信息是以页为单位由虚页向主存中的实页调入的。因为虚页和实页大小相等，所以调入时只要页边界对齐，页内地址无须修改就可以直接使用。这样，虚—实地址的转换主要是虚页号向实页号的转换，这个转换是由“页表”实现的。

每道程序都有一个独立的虚存空间，在程序运行时，存储管理软件根据主存的运行情况自动为每道程序建立一张独立的“页表”。在页表中，对应每个虚页号有一行表目，每个表目记录着虚页号对应到实页号的一些信息。所有页表都存放于主存的特定区域——页表区。当一个虚页号向实页号转换时，首先找到对应该程序的页表在页表区的首地址，然后在页表中按虚页号顺序找到该虚页所对应的表目，找到表目就可以实现虚地址向实地址的转换。

页式虚拟存储器的页表简单，地址变换速度较快，页面长度较小，主存空间的利用率高，对辅存的管理比较容易，因而得到了广泛应用。但是页式虚拟存储器的页面大小固定，无法反映程序内部的逻辑结构，不便于程序的执行、保护和共享。

（2）段式虚拟存储器。一个复杂的大程序总可以分解成多个在逻辑上相对独立的模块（程序段），每个模块的大小可以各不相同。段式虚拟存储器就是以程序的逻辑结构所自然形成的模块作为主存分配单位来进行存储器管理的。其中，每个模块（段）的长度可以不同，也可以独立编址，有的段还可以在执行时动态地决定大小。

程序运行时，以段为单位整段从辅存调入主存，一段占用一个连续的存储空间。CPU 访问时仍需进行虚—实地址的转换，段式虚拟存储器与页式虚拟存储器技术十分相似，每道程序都有一个“段表”，存该程序中各段装入主存的状态信息，每个段的信息在段表中占一行，主要包括段号、段起点、段长度等，根据段表可进行虚地址到实地址的转换，从而确定其在主存

中的位置。

段式虚拟存储器配合了模块化程序设计，各段之间相互独立，程序按逻辑结构分段，便于程序和数据的共享，段的大小、位置可变，模块可独立编址，可提高按段调度的命中率。但是段式虚拟存储器的地址变换所花费的时间较长，主存空间的利用率往往比较低，对辅存的管理比较困难。

（3）段页式虚拟存储器。为了能够同时获得页式虚拟存储器在管理主存和辅存空间方面的优点和段式虚拟存储器在程度模块化方面的优点，把两者结合起来就形成了段页式虚拟存储器。即将存储空间仍按程序的逻辑模块分段，每段又等分为若干个页，页面大小与实存页面相同。

虚地址包括段号、页号和页内地址 3 部分，实地址则只有页号和页内地址。虚存与实存之间的信息调度以页为基本单位。每个程序有一张段表，每段对应有一张页表，CPU 访存时，由段表指出每段对应页表的起始地址，而每一段的页表可指出该段的虚页号对应在实存空间的实页号，最后与页内地址拼接即可确定 CPU 要访问信息的实存地址。

段页式虚拟存储器的特点表现为分两级查表实现虚实地址转换，以页为单位调进或调出主存，按段来共享与保护程序和数据。

习题与思考

3.1　某微型计算机系统中，地址总线为 20 根，可寻址的内存空间应该是多大？

A．640KB　　B．16MB　　C．1MB　　D．1GB

3.2　在微型计算机系统中，主存储器是以下列哪种方式进行编址的？

A．以字节为单位编址　　B．以字为单位编址

C．以数据块为单位编址　　D．以磁道、扇区为单位编址

3.3　存储周期是指什么？

A．存储器的读出时间

B．存储器的写入时间

C．从 CPU 启动一次存储器读或写操作，到读出或写入数据完毕所经历的时间

D．连续启动两次独立的存储器读或写操作所需的最小时间间隔

3.4　SRAM 的某一单元中存放着数据 60H，CPU 对其读操作后，该单元的内容是什么？

A．00H　　B．FFH　　C．60H　　D．不确定

3.5　某微型计算机系统的主存储器容量为 64KB，若按字节编址，它的寻址范围如何？

A．00H～FFH　　B．000H～FFFH

C．0000H～FFFFH　　D．00000H～FFFFFH

3.6　下列 SRAM 芯片各需要多少个地址输入线？多少个数据信号线？

A．512K×4 位　B．1K×8 位　　C．2K×1 位　　D．16K×8 位

E．64K×1 位

3.7　对下列存储器组成方案，各需要多少个芯片？

A．用 64K×1 位芯片，组成 64K×8 位的存储器

B．用 1K×4 位的芯片，组成 4K×8 位的存储器

C．用 2K×1 位的芯片，组成 32K×8 位的存储器

3.8　已知一个具有 16 位地址和 8 位数据的存储器，回答下列问题：

（1）该存储器能存储多少字节的信息？

（2）如果存储器由 16K×4 位 RAM 芯片组成，需要多少片？

（3）至少需要多少位地址作芯片选择？

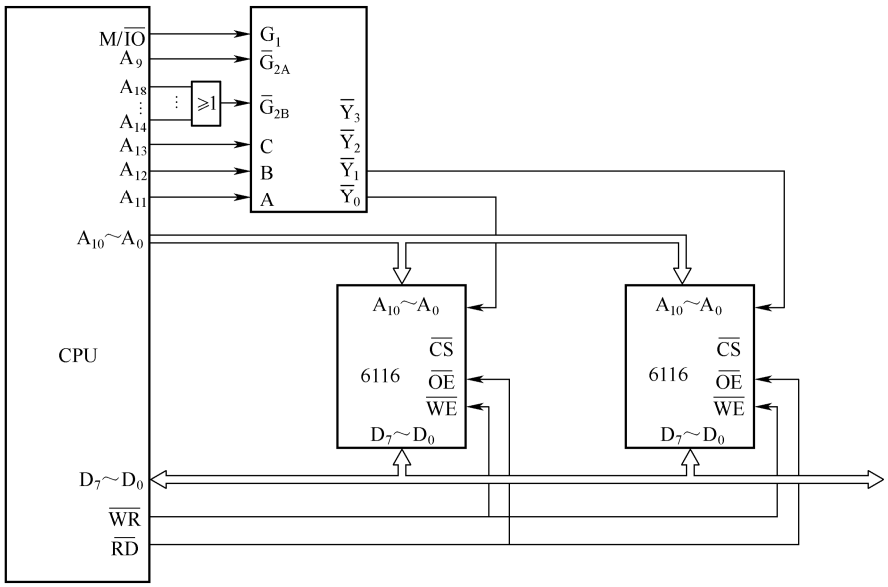
3.9　存储器有哪些主要性能指标？

3.10　简述半导体存储器的分类。

3.11　半导体存储器 SRAM、DRAM、ROM、PROM、EPROM、E^2PROM 和 Flash Memory 各有何读写特点？一般应用于何种场合？

3.12　动态 RAM 为什么必须定期刷新？Intel 2164 芯片中的 $\overline{RAS}$ 和 $\overline{CAS}$ 信号有何作用？

3.13　在某一微型计算机系统中，CPU 有 16 根地址线和 8 位数据线，主存储器扩展了两片 Intel 6116（2K×8 位），它们的硬件连接如下图所示，试分析这两片 6116 的寻址范围。



3.14　某存储器的首单元地址为 2000H，末单元地址为 9FFFH，分析该存储器的容量是多少？

3.15　已知某微型计算机系统的 RAM 容量为 4K×8 位，它的首单元址为 4000H，最后一个单元的地址多少？

3.16　某微型计算机系统有 20 位地址线、8 位数据线，现用两片 Intel 2114（1K×4 位）组成 2K×8 位的存储模块，并将它们的起始地址设置为 04800H，试画出硬件连接图。

3.17　为什么微型计算机存储器系统要采用层次结构？

3.18　在微型计算机系统中，为什么要使用 Cache？为什么要使用虚拟存储器？