

第5章 计算机组成原理

存储系统 和结构

2025年11月 北京理工大学计算机科学技术学院

第5章 计算机组成原理

存储系统是由几个容量、速度和价格各不相同的存储器构成的系统。设计一个容量大、速度快、成本低的存储系统是计算机发展的一个重要课题。本章重点讨论主存储器的工作原理、组成方式以及运用半导体存储芯片组成主存储器的一般原则和方法，此外还介绍了高速缓冲存储器和虚拟存储器的基本原理。

2025年11月 北京理工大学计算机科学技术学院

第5章 计算机组成原理

5.1 存储系统的组成

5.2 主存储器的组织

5.3 半导体随机存储器和只读存储器

5.4 主存储器的连接与控制

5.5 提高存储系统性能的技术

2025年11月 北京理工大学计算机科学技术学院

5.1 存储系统的组成 计算机组成原理

存储系统和存储器是两个不同的概念，下面首先介绍各种不同用途的存储器，然后讨论它们是如何构成一个存储系统的。

5.1.1 存储器分类

1. 按存储器在计算机系统中的作用分类

(1) 高速缓冲存储器

高速缓冲存储器位于主存和CPU之间，用来存放正在执行的程序段和数据，以便CPU高速地使用它们。

2025年11月 北京理工大学计算机科学技术学院

5.1 存储系统的组成 计算机组成原理

(2) 主存储器

用来存放计算机运行期间所需要的程序和数据，CPU可直接随机地进行读写访问。

(3) 辅助存储器

用来存放当前暂不参与运行的程序和数据，以及一些需要永久性保存的信息。CPU不能直接访问它。

2025年11月 北京理工大学计算机科学技术学院

5.1 存储系统的组成 计算机组成原理

2. 按存取方式分类

(1) 随机存取存储器RAM

CPU可以对RAM单元的内容随机地读写访问。CPU对任何一个存储单元的读写时间是一样的，即**存取时间是相同的**。**注意**

(2) 只读存储器ROM

ROM可以看作RAM的一种特殊方式，存储器的内容只能随机读出而不能写入。

(3) 顺序存取存储器SAM

SAM的内容只能按某种顺序存取，**存取时间与信息在存储体上的物理位置有关**。

2025年11月 北京理工大学计算机科学技术学院

5.1 存储系统的组成

计算机组成原理

(4)直接存取存储器DAM

当要存取所需的信息时，第一步直接指向整个存储器中的某个小区域（如磁盘上的磁道），第二步在小区域内顺序检索或等待，直至找到目的地后再进行读写操作。



2025年11月

北京理工大学计算机科学技术学院

5.1 存储系统的组成

计算机组成原理

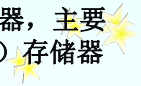
3.按存储介质分类

(1)磁芯存储器

利用两种不同的剩磁状态表示“1”或“0”。磁芯存储器的特点是信息可以长期存储，不会因断电而丢失；但磁芯存储器的读出是破坏性读出。

(2)半导体存储器

采用半导体器件制造的存储器，主要有双极型（TTL电路或ECL电路）存储器和MOS型存储器两大类。



2025年11月

北京理工大学计算机科学技术学院

5.1 存储系统的组成

计算机组成原理

(3)磁表面存储器

在金属或塑料基体上，涂复一层磁性材料，用磁层存储信息，常见的有磁盘、磁带等。

(4)光存储器

采用激光技术控制访问的存储器，如CD-ROM（只读光盘）、WORM（CD-R，写一次多次读光盘）、CD-RW（可读可写光盘）。



2025年11月

北京理工大学计算机科学技术学院

5.1 存储系统的组成

计算机组成原理

4.按信息的可保存性分类

断电后，存储信息即消失的存储器，称易失性存储器。断电后信息仍然保存的存储器，称非易失性存储器。

如果某个存储单元所存储的信息被读出时，原存信息将被破坏，则称破坏性读出。具有破坏性读出的存储器，每当一次读出操作之后，必须紧接一个重写（再生）的操作，以便恢复被破坏的信息。

如果读出时，被读单元原存信息不被破坏，则称非破坏性读出。



2025年11月

北京理工大学计算机科学技术学院

5.1 存储系统的组成

计算机组成原理

5.1.2 存储系统层次结构

为了解决存储容量、存取速度和价格之间的矛盾，通常把各种不同存储容量、不同存取速度的存储器，按一定的体系结构组织起来，形成一个统一整体的存储系统。



2025年11月

北京理工大学计算机科学技术学院

5.1 存储系统的组成

计算机组成原理

从CPU的角度来看， n 种不同的存储器（ $M_1 \sim M_n$ ）在逻辑上是一个整体。其中： M_1 速度最快、容量最小、位价格最高； M_n 速度最慢、容量最大、位价格最低。整个存储系统具有接近于 M_1 的速度，相等或接近 M_n 的容量，接近于 M_n 的位价格。在多级存储层次中，最常用的数据在 M_1 中，次常用的在 M_2 中，最少使用的在 M_n 中。

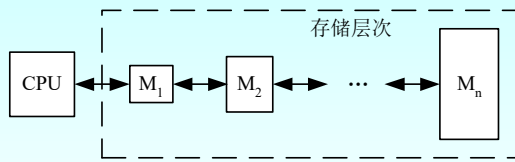


2025年11月

北京理工大学计算机科学技术学院

5.1 存储系统的组成

计算机组成原理



2025年11月

北京理工大学计算机科学技术学院

5.1 存储系统的组成

计算机组成原理

由高速缓冲存储器、主存储器、辅助存储器构成的三级存储系统可以分为两个层次。其中高速缓存和主存间称为Cache-主存存储层次（Cache存储系统）；主存-辅存存储层次（虚拟存储系统）。

2025年11月

北京理工大学计算机科学技术学院

5.1 存储系统的组成

计算机组成原理

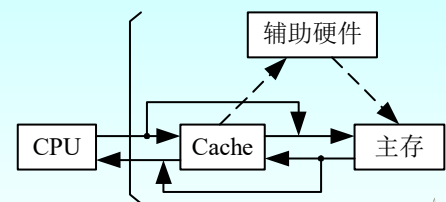
Cache存储系统是为解决主存速度不足而提出来的。在Cache和主存之间，增加辅助硬件，让它们构成一个整体。从CPU看，速度接近Cache的速度，容量是主存的容量，每位价格接近于主存的价格。由于Cache存储系统全部用硬件来调度，因此它对系统程序员和应用程序员都是透明的。

2025年11月

北京理工大学计算机科学技术学院

5.1 存储系统的组成

计算机组成原理



2025年11月

北京理工大学计算机科学技术学院

5.1 存储系统的组成

计算机组成原理

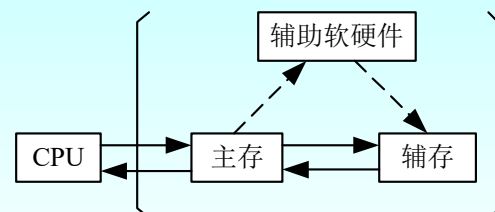
虚拟存储系统是为解决主存容量不足而提出来的。在主存和辅存之间，增加辅助的软硬件，让它们构成一个整体。从CPU看，速度接近主存的速度，容量是虚拟的地址空间，每位价格是接近于辅存的价格。由于虚拟存储系统需要通过操作系统来调度，因此对系统程序员是不透明的，但对应用程序员是透明的。

2025年11月

北京理工大学计算机科学技术学院

5.1 存储系统的组成

计算机组成原理



2025年11月

北京理工大学计算机科学技术学院

第5章

计算机组成原理

5.1 存储系统的组成

5.2 主存储器的组织

5.3 半导体随机存储器和只读存储器

5.4 主存储器的连接与控制

5.5 提高存储系统性能的技术

2025年11月

北京理工大学计算机科学技术学院

5.2 主存储器的组织

计算机组成原理

主存储器是整个存储系统的核心，它用来存放计算机运行期间所需要的程序和数据，CPU可直接随机地对它进行访问。

2025年11月

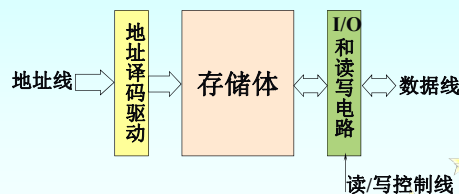
北京理工大学计算机科学技术学院

5.2 主存储器的组织

计算机组成原理

5.2.1 主存储器的基本结构

主存储器通常由存储体、地址译码驱动电路、I/O和读写电路组成。



2025年11月

北京理工大学计算机科学技术学院

5.2 主存储器的组织

计算机组成原理

存储体是主存储器的核心，程序和数据都存放在存储体中。

地址译码驱动电路实际上包含译码器和驱动器两部分。译码器将地址总线输入的地址码转换成与之对应的译码输出线上的有效电平，以表示选中了某一单元，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。

I/O和读写电路包括读出放大器、写入电路和读/写控制电路，用以完成被选中存储单元中各位的读出和写入操作。

2025年11月

北京理工大学计算机科学技术学院

5.2 主存储器的组织

计算机组成原理

存储器的读/写操作是在控制器的控制下进行的。半导体存储器芯片中的控制电路，必须接收到来自控制器的读/写命令或写入允许信号后，才能实现正确的读/写操作。

2025年11月

北京理工大学计算机科学技术学院

5.2 主存储器的组织

计算机组成原理

5.2.2 主存储器的存储单元

位是二进制数的最基本单位，也是存储器存储信息的最小单位。

一个二进制数由若干位组成，当这个二进制数作为一个整体存入或取出时，这个数称为**存储字**。

存放存储字或存储字节的主存空间称为**存储单元或主存单元**，大量存储单元的集合构成一个**存储体MB**，程序和数据都存放在存储体中，它是存储器的核心。

注意

2025年11月

北京理工大学计算机科学技术学院

5.2 主存储器的组织

计算机组成原理

一个存储单元可能存放一个字，也可能存放一个字节，这是由计算机的结构确定的。对于字节编址的计算机，最小寻址单位是一个字节，相邻的存储单元地址指向相邻的存储字节；对于字编址的计算机，最小寻址单位是一个字，相邻的存储单元地址指向相邻的存储字。

存储单元是CPU对主存可访问操作的最小存储单位。

2025年11月

北京理工大学计算机科学技术学院

5.2 主存储器的组织

计算机组成原理

例如，IBM 370机是字长为32位的计算机，主存按字节编址，每一个存储字包含4个单独编址的存储字节，字地址即是该字高位字节的地址，其字地址总是等于4的整数倍，正好用地址码的最末两位来区分同一个字的四个字节。PDP-11机是字长为16位的计算机，主存也按字节编址，每一个存储字包含2个单独编址的存储字节，它的字地址总是2的整数倍，但却是用低位字节地址作为字地址，并用地址码的最末1位来区分同一个字的两个字节。

2025年11月

北京理工大学计算机科学技术学院

5.2 主存储器的组织

计算机组成原理

字地址

字节地址

| | | | | |
|---|---|---|----|----|
| 0 | 0 | 1 | 2 | 3 |
| 4 | 4 | 5 | 6 | 7 |
| 8 | 8 | 9 | 10 | 11 |

字地址

字节地址

| | | |
|---|---|---|
| 0 | 1 | 0 |
| 2 | 3 | 2 |
| 4 | 5 | 4 |

2025年11月

北京理工大学计算机科学技术学院

5.2 主存储器的组织

计算机组成原理

假设一个字由四个字节组成，我们使用 B_3 、 B_2 、 B_1 、 B_0 来分别表示这四个字节，其中 B_3 是字的最高有效字节， B_0 是最低有效字节。字节编址计算机的主存地址安排有两种方案，但字地址总是等于4的整数倍。

字地址 N

| | | | |
|----------|-------|-------|-------|
| B_3 | B_2 | B_1 | B_0 |
| 字节地址 N+3 | N+2 | N+1 | N+0 |

(a)

字地址 N

| | | | |
|----------|-------|-------|-------|
| B_3 | B_2 | B_1 | B_0 |
| 字节地址 N+0 | N+1 | N+2 | N+3 |

(b)

2025年11月

北京理工大学计算机科学技术学院

5.2 主存储器的组织

计算机组成原理

图 (a)称为小端方案。假设字地址为N，则字节 B_3 、 B_2 、 B_1 、 B_0 依次存放在地址为N+3、N+2、N+1、N+0的存储单元，即字地址等于最低有效字节地址。采用小端方案的计算机有Intel 80X86、DEC VAX等。

图 (b)称为大端方案。假设字地址为N，则字节 B_3 、 B_2 、 B_1 、 B_0 依次存放在地址为N+0、N+1、N+2、N+3的存储单元，即字地址等于最高有效字节地址。采用大端方案的计算机有IBM360/370、Motorola 68000等。

2025年11月

北京理工大学计算机科学技术学院

5.2 主存储器的组织

计算机组成原理

大端方案将高字节（MSB）存放在低地址，小端方案将高字节存放在高地址。采用大端方案进行数据存放符合人类的正常思维，而采用小端方案进行数据存放利于计算机处理。到目前为止，采用大端或者小端进行数据存放，其孰优孰劣也没有定论。大端与小端方案的差别体现在一个处理器的寄存器、指令集、数据总线等各个层次中。

2025年11月

北京理工大学计算机科学技术学院

5.2 主存储器的组织

计算机组成原理

5.2.3 主存储器的主要技术指标

1. 存储容量

存储容量是指主存所能容纳的二进制信息总量。对于字节编址的计算机，以字节数来表示容量；对于字编址的计算机，以字数与其字长的乘积来表示容量。

如某计算机的容量为 $64K \times 16$ ，表示它有64K个字，每个字的字长为16位，若用字节数表示，则可记为128K字节（128KB）。

2025年11月

北京理工大学计算机科学技术学院

5.2 主存储器的组织

计算机组成原理

2. 存取速度

(1) 存取时间 T_a

存取时间又称为访问时间或读/写时间，它是指从启动一次存储器操作到完成该操作所经历的时间。

(2) 存取周期 T_m

存取周期又可称作读写周期、访存周期，它是指存储器进行一次完整的读写操作所需的全部时间，即连续两次访问存储器操作之间所需要的最短时间。

注意

2025年11月

北京理工大学计算机科学技术学院

5.2 主存储器的组织

计算机组成原理

显然，一般情况下， $T_m > T_a$ 。这是因为对任何一种存储器，在读写操作之后，总要有一段恢复内部状态的复原时间。对于破坏性读出的存储器，存取周期往往比存取时间要大得多，甚至可以达到 $T_m = 2T_a$ ，这是因为存储器中的信息读出后需要马上进行重写（再生）。

2025年11月

北京理工大学计算机科学技术学院

5.2 主存储器的组织

计算机组成原理

(3) 主存带宽 B_m

与存取周期密切相关的指标是主存的带宽，它又称为数据传输率，表示每秒从主存进出信息的最大数量，单位为字/秒或字节/秒或位/秒。

$B_m = \text{主存等效工作频率} \times \text{主存位宽} \div 8 = \text{内存时钟频率} \times \text{倍增系数} \times \text{主存位数} \div 8$ 。以DDR400内存为例，它的运行频率为200MHz，数据总线位数为64bit，由于上升沿和下降沿都传输数据，因此倍增系数为2，此时带宽为： $200 \times 2 \times 64 / 8 = 3.2\text{GB/s}$ 。

2025年11月

北京理工大学计算机科学技术学院

5.2 主存储器的组织

计算机组成原理

目前，主存提供信息的速度还跟不上CPU处理指令和数据的速度，所以，主存的带宽是改善计算机系统瓶颈的一个关键因素。为了提高主存的带宽，可以采取的措施有：

- 缩短存取周期；
- 增加存储字长；
- 增加存储体。

2025年11月

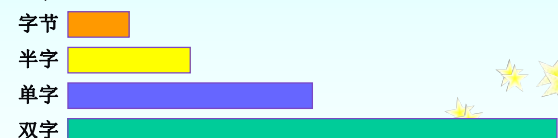
北京理工大学计算机科学技术学院

5.2 主存储器的组织

计算机组成原理

5.2.4 数据在主存中的存放

在采用字节编址的情况下，数据在主存储器中的三种不同存放方法。假设，存储字为64位（8个字节），读/写的数据有四种不同长度，它们分别是字节（8位）、半字（16位）、单字（32位）和双字（64位）。



2025年11月

北京理工大学计算机科学技术学院

5.2 主存储器的组织

计算机组成原理

2025年11月

北京理工大学计算机科学技术学院

2025年11月

请注意：此例中数据字长（32位）不等于存储字长（64位）。

现有一批数据，它们依次为：字节、半字、双字、单字、半字、单字、字节、单字。

5.2 主存储器的组织

计算机组成原理

2025年11月

北京理工大学计算机科学技术学院

2025年11月

(1)不浪费存储器资源的存放方法

四种不同长度的数据一个紧接着一个存放。优点是不浪费宝贵的主存资源，但存在的问题是：当访问的一个双字、单字或半字跨越两个存储字时，存储器的工作速度降低了一倍，而且读写控制比较复杂。

5.2 主存储器的组织

计算机组成原理

2025年11月

北京理工大学计算机科学技术学院

2025年11月

(2)从存储字的起始位置开始存放方法

无论要存放的是字节、半字、单字或双字，都必须从存储字的起始位置开始存放，而空余部分浪费不用。优点是：无论访问一个字节、半字、单字或双字都可以在一个存取周期内完成，读写数据的控制比较简单。缺点是：浪费了宝贵的存储器资源。

5.2 主存储器的组织

计算机组成原理

2025年11月

北京理工大学计算机科学技术学院

2025年11月

5.2 主存储器的组织

计算机组成原理

2025年11月

北京理工大学计算机科学技术学院

2025年11月

(3)边界对齐的数据存放方法

双字地址的最末三个二进制位必须为000，单字地址的最末两位必须为00，半字地址的最末一位必须为0。它能够保证无论访问双字、单字、半字或字节，都在一个存取周期内完成，尽管存储器资源仍然有浪费，但是浪费比第(2)种存放方法要少得多。

注意

5.2 主存储器的组织

计算机组成原理

2025年11月

北京理工大学计算机科学技术学院

2025年11月

第5章

计算机组成原理

5.1 存储系统的组成

5.2 主存储器的组织

5.3 半导体随机存储器和只读存储器

5.4 主存储器的连接与控制

5.5 提高存储系统性能的技术

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

主存储器通常分为**RAM**和**ROM**两大部分。**RAM**可读可写，**ROM**只能读不能写。下面重点讨论RAM的工作原理与结构，以及ROM的基本类型。

注意

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

5.3.1 RAM记忆单元电路

存放一个二进制位的物理器件称为记忆单元，它是存储器的最基本构件，地址码相同的多个记忆单元构成一个存储单元。记忆单元可以由各种材料制成，但最常见的由MOS电路组成。MOS型存储器根据记忆单元的结构又可分为静态RAM和动态RAM两种。静态RAM，即**SRAM**（**Static RAM**），其存储电路以双稳态触发器为基础；动态RAM，即**DRAM**（**Dynamic RAM**），其存储电路以电容为基础。

注意

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

- ✓ 六管静态MOS记忆单元电路
- ✓ 四管动态MOS记忆单元电路
- ✓ 单管动态记忆单元电路

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

5.3.2 动态RAM的刷新

1. 刷新间隔

前面已经说过，为了维持MOS型动态记忆单元的存储信息，每隔一定时间必须对存储体中的所有记忆单元的栅极电容补充电荷，这个过程就是刷新。

一般选定MOS型动态存储器允许的最大刷新闻隔为2ms，也就是说，应在2ms内，将全部存储体刷新一遍。

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

值得一提的是，刷新和重写（再生）是两个完全不同的概念，切不可加以混淆。重写是随机的，某个存储单元只有在破坏性读出之后才需要重写。而刷新是定时的，即使许多记忆单元长期未被访问，若不及时补充电荷的话，信息也会丢失。重写一般是按存储单元进行的，而刷新通常以存储体矩阵中的一行为单位进行的。

2. 刷新方式

常见的刷新方式有集中式、分散式和异步式三种。

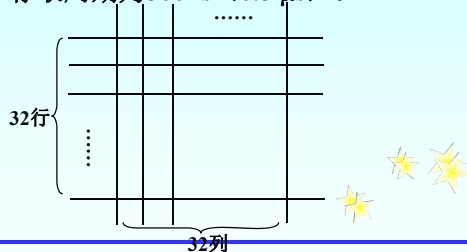
2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

例如，对具有1024个记忆单元（排列成 32×32 矩阵）的存储芯片进行刷新，刷新是按行进行的，且每刷新一行占用一个存取周期，存取周期为500ns（0.5 μ s）。



2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

(1)集中刷新方式

在允许的最大刷新间隔内，按照存储芯片容量的大小集中安排若干个刷新周期，刷新时停止读写操作。

$$\text{刷新时间} = \text{存储体矩阵行数} \times \text{刷新周期}$$

这里刷新周期是指刷新一行所需要的时间，由于刷新过程就是“假读”的过程，所以刷新周期就等于存取周期。

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

在最大刷新间隔2ms内共可以安排4000个存取周期，从0~3967个周期内进行读/写操作或保持，而从3968~3999这最后32个周期集中安排刷新操作。



2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

集中刷新方式的优点是读/写操作时不受刷新工作的影响，因此系统的存取速度比较高。缺点是在集中刷新期间必须停止读/写，这一段时间称为“死区”，而且存储容量越大，死区就越长。

2025年11月

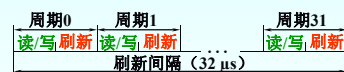
北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

(2)分散刷新方式

分散刷新是指把刷新操作分散到每个存取周期内进行，此时系统的存取周期被分为两部分，前一部分时间进行读/写操作或保持，后一部分时间进行刷新操作。一个系统存取周期内刷新存储矩阵中的一行。



2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

分散刷新方式没有死区，但是，它也有很明显的缺点，第一是加长了系统的存取周期，如存储芯片的存取周期为0.5 μ s，则系统的存取周期应为1 μ s，降低了整机的速度；第二是刷新过于频繁（本例中每32 μ s就重复刷新一遍），尤其是当存储容量比较小的情况下，没有充分利用所允许的最大刷新间隔（2ms）。

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

(3)异步刷新方式

异步刷新方式可以看成前述两种方式的结合，它充分利用了最大刷新间隔时间，把刷新操作平均分配到整个最大刷新间隔时间内进行，故有：

相邻两行的刷新间隔=最大刷新间隔时间/行数



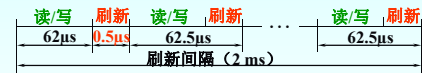
2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

对于 32×32 矩阵，在2ms内需要将32行刷新一遍，所以相邻两行的刷新时间间隔 $=2\text{ms}/32=62.5\mu\text{s}$ ，即每隔 $62.5\mu\text{s}$ 安排一个刷新周期，在刷新时封锁读/写。



2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

异步刷新方式虽然也有死区，但比集中刷新方式的死区小得多，仅为 $0.5\mu\text{s}$ 。这样可以避免使CPU连续等待过长的时间，而且减少了刷新次数，是比较实用的一种刷新方式。



2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

3.刷新控制

MOS型动态RAM的刷新要注意几个问题：

- ① 刷新对CPU是透明的。
- ② 刷新通常是一行一行地进行的，每一行中各记忆单元同时被刷新，故刷新操作时仅需要行地址，不需要列地址。
- ③ 刷新操作类似于读出操作。
- ④ 因为所有芯片同时被刷新，所以在考虑刷新问题时，应当从单个芯片的存储容量着手，而不是从整个存储器的容量着手。

2025年11月

北京理工大学计算机科学技术学院

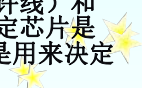
5.3 半导体随机存储器和只读存储器

计算机组成原理

5.3.3 RAM芯片分析

1.RAM芯片

存储芯片通过地址线、数据线和控制线与外部连接。地址线是单向输入的，其数目与芯片容量有关。如容量为 1024×4 时，地址线有10根；容量为 $64\text{K} \times 1$ 时，地址线有16根。数据线是双向的，既可输入，也可输出，其数目与数据位数有关。如 1024×4 的芯片，数据线有4根； $64\text{K} \times 1$ 的芯片，数据线只有1根。控制线主要有读/写控制线（或写允许线）和片选线两种，读/写控制线是用来决定芯片是进行读操作还是写操作的，片选线是用来决定该芯片是否被选中的。



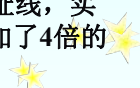
2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

由于DRAM芯片集成度高，容量大，为了减少芯片引脚数量，DRAM芯片把地址线分成相等的两部分，分两次从相同的引脚送入。两次输入的地址分别称为行地址和列地址，行地址由行地址选通信号RAS送入存储芯片，列地址由列地址选通信号CAS送入存储芯片。由于采用了地址复用技术，因此，DRAM芯片每增加一条地址线，实际上是增加了两位地址，也即增加了4倍的容量。



2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

2.地址译码方式

地址译码电路能把地址线送来的地址信号翻译成对应存储单元的选择信号。

(1)单译码方式

单译码方式又称字选法，它所对应的存储器结构是字结构的，容量为M个字的存储器（M个字，每字b位），排列成M行×b列的矩阵，矩阵的每一行对应一个字，有一条公用的选择线 w_i （字线）。字线选中某一行时，同一行中的各位就都被选中，由读写电路对被选中的各位实施读出或写入操作。

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

字结构的优点是结构简单，缺点是使用的外围电路多，成本昂贵。更严重的是，当字数大大超过位数时，存储器会形成纵向很长而横向很窄的不合理结构，所以这种方式只适用于容量不大的存储器。

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

(2)双译码方式

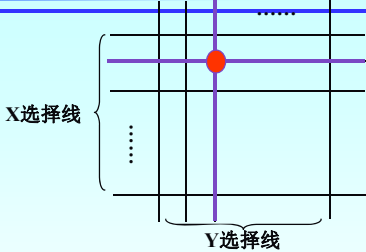
双译码方式又称为重合法。通常是把K位地址码分成接近相等的两段，一段用于水平方向作X地址线，供X地址译码器译码；一段用于垂直方向作Y地址线，供Y地址译码器译码。X和Y两个方向的选择线在存储体内部的一个记忆单元上交叉，以选择相应的记忆单元。

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理



X选择线

Y选择线

双译码方式对应的存储芯片结构可以是位结构的，则在Z方向上重叠b个芯片。也可以是字段结构的。

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

对于字段结构的存储芯片，行选择线为M/s根，列选择线为s，K位地址线也要划分为两部分： $K_x = \log_2 M/s$ ， $K_y = \log_2 s$ 。

双译码方式与单译码方式相比，减少了选择线数目和驱动器数目。存储容量越大，这两种方式的差异越明显。

| 译码方式 | 占用地址位 | 选择线数 | 驱动器数 |
|------|--------|-----------|-----------|
| 单译码 | 16 | 65536 | 65536 |
| 双译码 | 8 8 | 256 256 | 256 256 |

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

以Intel 2114为例

存储容量为 $1K \times 4$ ，由4096个六管记忆单元电路组成，它们排成 64×64 的矩阵，采用字段结构。CPU送来的地址总线10位（ $A_0 \sim A_9$ ），其中6位作为行选择电路的输入，经行地址译码器产生64条行选择线；另4位作为列选择电路的输入，经列地址译码器产生16条列选择线，每列选择线再经列I/O电路把4条数据线（ $I/O_1 \sim I/O_4$ ）连接到各字段的相应位上。

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

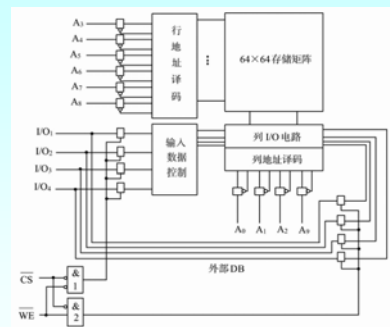
$I/O_1 \sim I/O_4$ 是受输入三态门和输出三态门控制的双向数据线，由片选信号 \overline{CS} 和写允许信号 \overline{WE} 一起来控制这些三态门。

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

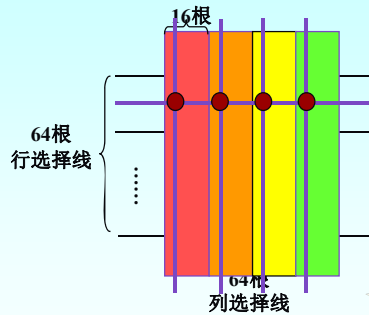


2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理



2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

3.RAM的读/写时序

(1)SRAM读/写时序

读周期表示对该芯片进行两次连续读操作的最小间隔时间。在此期间，地址输入信息不允许改变，片选信号 \overline{CS} 在地址有效之后变为有效，使芯片被选中，最后在数据线上得到读出的信号。写允许信号 \overline{WE} 在读周期中保持高电平。

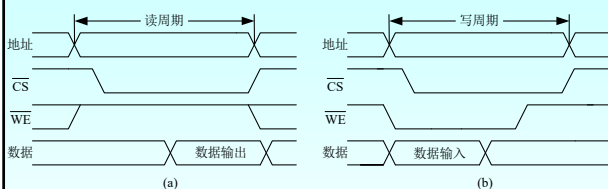
写周期与读周期相似，但除了要加地址和片选信号外，还要加一个低电平有效的写入脉冲 \overline{WE} ，并提供写入数据。

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理



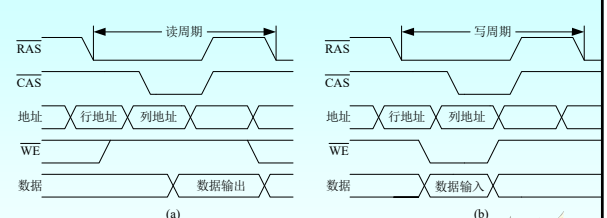
2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

(2)DRAM读/写时序



2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

5.3.4 半导体只读存储器 (ROM)

ROM的最大优点是具有非易失性，即使电源断电，ROM中存储的信息也不会丢失。

1. ROM的类型

ROM工作时只能读出，不能写入，那么ROM中的内容是如何事先存入的呢？我们把向ROM写入数据的过程称为对ROM进行编程，根据编程方法的不同，ROM通常可以分为以下几类：

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

(1) 掩膜式ROM (MROM)

它的内容是由半导体生产厂家按用户提出的要求在芯片的生产过程中直接写入的，写入后任何人都无法改变其内容。

(2) 一次可编程ROM (PROM)

PROM允许用户利用专门的设备（编程器或写入器）写入自己的程序，但一旦写入后便无法改变，因此它是一种一次性可编程的ROM。

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

通常，生产厂家提供的PROM芯片初始内容为全“0”，用户根据自编的程序，使用编程器外加足够大的电压（或电流），将“1”写入相应位，PROM的编程是逐位进行的。常见的PROM根据写入原理可分为两类：结破坏型和熔丝型。

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

(3) 可擦除可编程ROM (EPROM)

这种ROM的内容不仅可以由用户利用编程器写入，而且可以对其内容进行多次改写。但要注意的是：在+5V的电源条件下只能读出不能写入，用编程器写入信息时必须用+25V的高压。与前两种ROM相比，EPROM使用起来最为方便，因此应用非常广泛。

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

UVEPROM (紫外线擦除)

用紫外线灯进行擦除的，所以只能对整个芯片擦除，而不能对芯片中个别需要改写的存储单元单独擦除和重写。

EEPROM (电擦除)

用电气方法来进行擦除的，它在联机条件下可以用字擦除方式擦除，也可以用数据块擦除方式擦除。以字擦除方式操作时，能够只擦除被选中的那个存储单元的内容；在数据块擦除方式操作时，可擦除数据块内所有单元的内容。

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

(4) 闪速存储器 (flash memory)

一种快擦写型存储器，它的主要特点是：既可在不加电的情况下长期保存信息，又能在线进行快速擦除与重写，兼备了EEPROM和RAM的优点。

目前，大多数微机的主板采用闪速存储器来存储BIOS（基本输入/输出系统）程序。闪速存储器除了具有ROM的一般特性外，还有低电压改写的特点，便于用户自动升级BIOS。

2025年11月

北京理工大学计算机科学技术学院

5.3 半导体随机存储器和只读存储器

计算机组成原理

2.ROM芯片

除去地址线、数据线、片选线外，电源线分别有Vcc——+5V（工作电源）和Vpp——编程电源。

2025年11月

北京理工大学计算机科学技术学院

第5章

计算机组成原理

5.1 存储系统的组成

5.2 主存储器的组织

5.3 半导体随机存储器和只读存储器

5.4 主存储器的连接与控制

5.5 提高存储系统性能的技术

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

5.4.1 主容量的扩展

要组成一个主存，首先要考虑选片的问题，然后就是如何把芯片连接起来的问题。根据存储器所要求的容量和选定的存储芯片的容量，就可以计算出总的芯片数，即

$$\text{总片数} = \frac{\text{总容量}}{\text{芯片容量}}$$

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

1.位扩展

位扩展指只在**位数方向扩展**（加大字长），而芯片的字数和存储器的字数是一致的。位扩展的连接方式是将各存储芯片的地址线、片选线和读/写线相应地并联起来，而将各芯片的数据线单独列出。

如用64K×1的SRAM芯片组成64K×8的存储器，需要8个芯片。

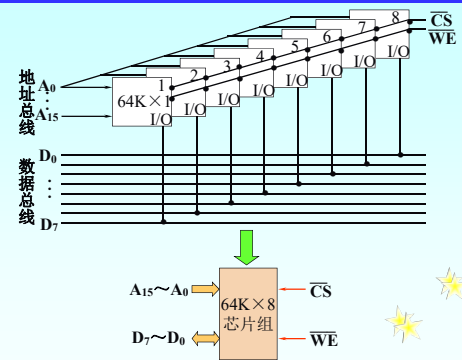
| | 容量 | 地址 | 数据 |
|------|-------|----|----|
| 存储器 | 64K×8 | 16 | 8 |
| 存储芯片 | 64K×1 | 16 | 1 |

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

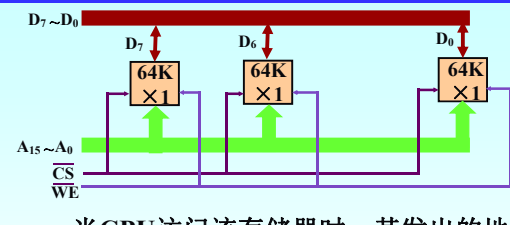


2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理



当CPU访问该存储器时，其发出的地址和控制信号同时传给8个芯片，选中每个芯片的同一单元，其单元的内容被同时读至数据总线的相应位，或将数据总线上的内容分别同时写入相应单元。

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

2.字扩展

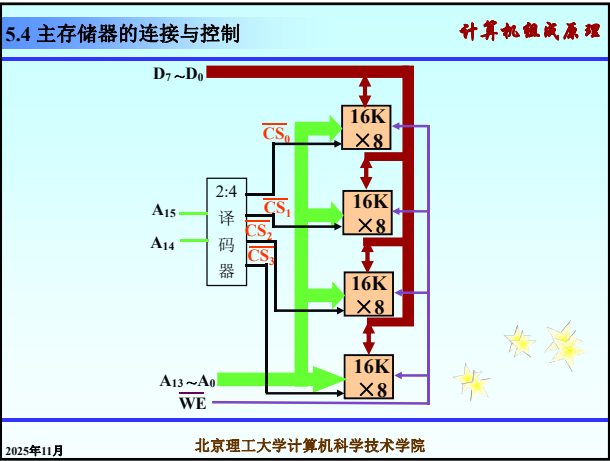
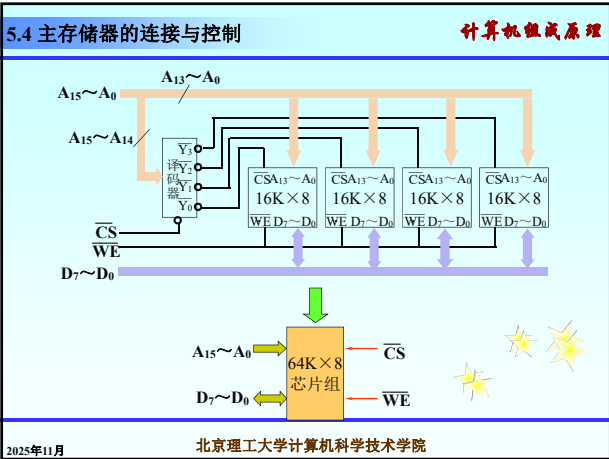
字扩展是指**仅在字数方向扩展，而位数不变**。字扩展将芯片的地址线、数据线、读/写线并联，由片选信号来区分各个芯片。

如用16K×8的SRAM组成64K×8的存储器，需要4个芯片。

| | 容量 | 地址 | 数据 |
|------|-------|----|----|
| 存储器 | 64K×8 | 16 | 8 |
| 存储芯片 | 16K×8 | 14 | 8 |

2025年11月

北京理工大学计算机科学技术学院



5.4 主存储器的连接与控制

计算机组成原理

在同一时间内四个芯片中只能有一个芯片被选中。四个芯片的地址分配如下：

第一片 最低地址 0000H
最高地址 3FFFH

第二片 最低地址 4000H
最高地址 7FFFH

第三片 最低地址 8000H
最高地址 BFFFH

第四片 最低地址 C000H
最高地址 FFFFH

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

3.字和位同时扩展

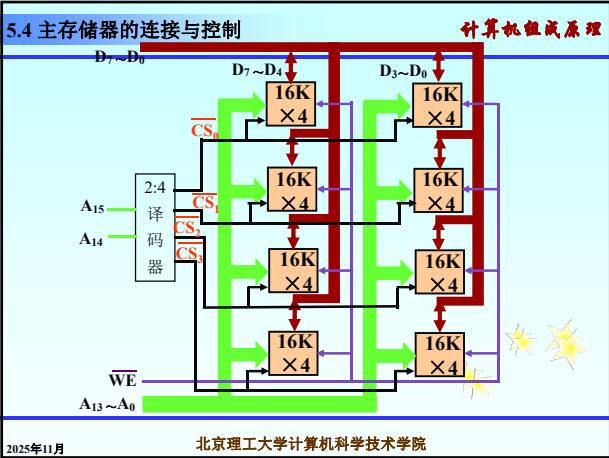
当构成一个容量较大的存储器时，往往需要**在字数方向和位数方向上同时扩展**，这将是前两种扩展的组合，实现起来也是很容易的。

如用16K×4的SRAM组成64K×8的存储器，需要8个芯片。

| | 容量 | 地址 | 数据 |
|------|-------|----|----|
| 存储器 | 64K×8 | 16 | 8 |
| 存储芯片 | 16K×4 | 14 | 4 |

2025年11月

北京理工大学计算机科学技术学院



5.4 主存储器的连接与控制

计算机组成原理

5.4.2 存储芯片的地址分配和片选

CPU要实现对存储单元的访问，首先要选择存储芯片，即进行片选；然后再从选中的芯片中依地址码选择出相应的存储单元，以进行数据的存取，这称为字选。片内的字选是由CPU送出的N条低位地址线完成的，地址线直接接到所有存储芯片的地址输入端（N由片内存储容量 2^N 决定），而片选信号则是通过高位地址得到的。实现片选的方法可分为3种：即线选法、全译码法和部分译码法。

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

1.线选法

线选法就是用除片内寻址外的高位地址线直接（或经反相器）分别接至各个存储芯片的片选端，当某地址线信息为“0”时，就选中与之对应的存储芯片。请注意，这些片选地址线每次寻址时只能有一位有效，不允许同时有多位有效，这样才能保证每次只选中一个芯片（或组）。

| 芯片 | $A_{14} \sim A_{11}$ | $A_{10} \sim A_0$ | 地址范围 |
|----|----------------------|-------------------|----------------|
| 0# | 1 1 1 0 | 00...0 11...1 | 7000~ 77FFH |
| 1# | 1 1 0 1 | 00...0 11...1 | 6800~ 6FFFH |

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

| | | | |
|----|---------|------------------|----------------|
| 2# | 1 0 1 1 | 00...0 11...1 | 5800~ 5FFFH |
| 3# | 0 1 1 1 | 00...0 11...1 | 3800~ 3FFFH |

线选法的优点是不需要地址译码器，线路简单，选择芯片不需要外加逻辑电路，但仅适用于连接存储芯片较少的场合。同时，线选法不能充分利用系统的存储器空间，且把地址空间分成了相互隔离的区域，给编程带来了一定的困难。

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

2.全译码法

全译码法将片内寻址外的全部高位地址线作为地址译码器的输入，把经译码器译码后的输出作为各芯片的片选信号，将它们分别接到存储芯片的片选端，以实现对存储芯片的选择。

全译码法的优点是每片（或组）芯片的地址范围是唯一确定的，而且是连续的，也便于扩展，不会产生地址重叠的存储区，但全译码法对译码电路要求较高。

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

| 芯片 | $A_{19} \sim A_{13}$ | A_{12} | A_{11} | $A_{10} \sim A_0$ | 地址范围 |
|----|----------------------|----------|----------|-------------------|------------------|
| 0# | 0 ... 0 | 0 | 0 | 00...0 11...1 | 00000~ 007FFH |
| 1# | 0 ... 0 | 0 | 1 | 00...0 11...1 | 00800~ 00FFFH |
| 2# | 0 ... 0 | 1 | 0 | 00...0 11...1 | 01000~ 017FFH |
| 3# | 0 ... 0 | 1 | 1 | 00...0 11...1 | 01800~ 01FFFH |

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

3.部分译码

所谓部分译码即用片内寻址外的高位地址的一部分来译码产生片选信号。

如用4片 $2K \times 8$ 的存储芯片组成 $8K \times 8$ 存储器，需要四个片选信号，因此只要用两位地址线来译码产生。设地址总线有20位（ $A_{19} \sim A_0$ ），则寻址 $8K \times 8$ 存储器时，无论 $A_{19} \sim A_{13}$ 取何值，只要 $A_{12} = A_{11} = 0$ ，而均选中第一片，只要 $A_{12} = 0, A_{11} = 1$ ，均选中第二片，……。也就是说， $8K$ RAM中的任一存储单元，都对应有 $2^{(20-13)} = 2^7$ 个地址，这种一个存储单元出现多个地址的现象称地址重叠。

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

从地址分布来看，这8KB存储器实际上占用了CPU全部的空间（1MB）。每片2K×8的存储芯片有1/4M=256K的地址重叠区。

| 8K×8存储器 | 1M×8存储空间 |
|---------|-------------|
| 0000H 0 | 00000H 0 2K |
| 07FFH 1 | 007FFH 1 2K |
| 0800H 2 | 00800H 2 2K |
| 0FFFH 3 | 00FFFH 3 2K |
| 1000H | 01000H |
| 17FFH | 017FFH |
| 1800H | 01800H |
| 1FFFH | 01FFFH |
| | 02000H |
| | 027FFH |
| | ... |
| | 1 2K |
| | 2 2K |
| | 3 2K |

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

令未用到的高位地址全为0，这样确定的存储器地址称为基本地址，本例中8K×8存储器的基本地址即00000H~01FFFH。部分译码法较全译码法简单，但存在地址重叠区。

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

例1：假定计算机系统需要512字节RAM和512字节ROM容量。使用的RAM芯片是128字×8位，ROM芯片为512字×8位。RAM芯片有CS及WE控制端，ROM芯片有CS控制端，CPU有地址线A₁₅~A₀、数据线D₇~D₀、读写控制线RW等。试确定各存储器芯片的地址区间，指出存储器以及各存储器芯片需要的地址线数量，并画出存储器与CPU的连接图。

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

解：

1). 各存储器芯片的地址区间

| 元件 | 16 进制地址范围 | 二进制地址值 |
|------|-----------|---------------------|
| RAM1 | 0000~007F | 0 0 1 x x x x x x x |
| RAM2 | 0080~00FF | 0 0 1 x x x x x x x |
| RAM3 | 0100~017F | 0 1 0 x x x x x x x |
| RAM4 | 0180~01FF | 0 1 1 x x x x x x x |
| ROM | 0200~03FF | 1 x x x x x x x x x |

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

2). 存储器所需的地址线

存储器的总容量为1KB，需要10条地址线。

RAM：4片，每片要7条地址线(2⁷=128)；

ROM：1片，需要9条地址线(2⁹=512)。

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

3). 连接图

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

例2：设有若干片 $256\text{K} \times 8$ 位的SRAM芯片，采用字扩展方法构成 2048K 的存储器需要多少片SRAM芯片？该存储器需要多少字节地址位？画出该存储器与CPU连接的结构图，设CPU的接口信号有地址信号、数据信号、控制信号MREQ#(允许访存信号)和R/W#。

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

解：

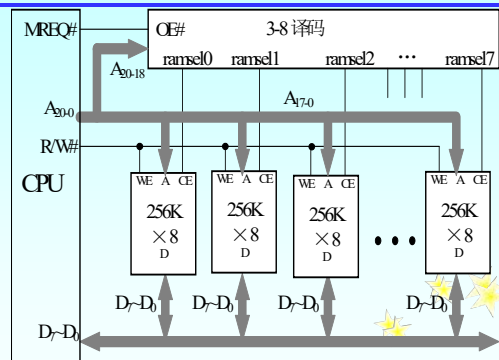
- 1). 该存储器需要 $2048\text{K}/256\text{K} = 8$ 片SRAM芯片
- 2). 需要21条地址线，因为 $2^{21} = 2048\text{K}$ ，其中高3位用于芯片选择，低18位作为每个存储器芯片的地址输入。用MREQ#作为译码器芯片的输出许可信号，译码器的输出作为存储器芯片的选择信号，R/W#作为读写控制信号。CPU访存的最高地址位为 A_{20} 。

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

3).
连接图

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

4). 译码器的输出信号逻辑表达式为：

$$\begin{aligned} \text{ramsel0} &= \overline{A_{20}} * \overline{A_{19}} * \overline{A_{18}} * \overline{\text{MREQ\#}} \\ \text{ramsel1} &= \overline{A_{20}} * \overline{A_{19}} * A_{18} * \overline{\text{MREQ\#}} \\ \text{ramsel2} &= \overline{A_{20}} * A_{19} * \overline{A_{18}} * \overline{\text{MREQ\#}} \\ \text{ramsel3} &= \overline{A_{20}} * A_{19} * A_{18} * \overline{\text{MREQ\#}} \\ \text{ramsel4} &= A_{20} * \overline{A_{19}} * \overline{A_{18}} * \overline{\text{MREQ\#}} \\ \text{ramsel5} &= A_{20} * \overline{A_{19}} * A_{18} * \overline{\text{MREQ\#}} \\ \text{ramsel6} &= A_{20} * A_{19} * \overline{A_{18}} * \overline{\text{MREQ\#}} \\ \text{ramsel7} &= A_{20} * A_{19} * A_{18} * \overline{\text{MREQ\#}} \end{aligned}$$

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

5.4.3 主存储器和CPU的连接

1. 主存和CPU之间的硬连接

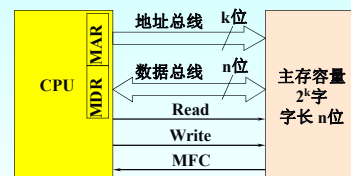
主存与CPU的硬连接有三组连线：地址总线（AB）、数据总线（DB）和控制总线（CB）。此时，我们把主存看作一个黑盒子，存储器地址寄存器（MAR）和存储器数据寄存器（MDR）是主存和CPU之间的接口。MAR可以接受来自程序计数器的指令地址或来自运算器的操作数地址，以确定要访问的单元。MDR是向主存写入数据或从主存读出数据的缓冲部件。

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理



2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

2.CPU对主存的基本操作

CPU与主存的硬连接是两个部件之间联系的物理基础，而两个部件之间还有软连接，即CPU向主存发出的读或写命令，这才是两个部件之间有效工作的关键。

CPU对主存进行读/写操作时，首先CPU在地址总线上给出地址信号，然后发出相应的读或写命令，并在数据总线上交换信息。

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

(1)读

读操作是指从CPU送来的地址所指定的存储单元中取出信息，再送给CPU，其操作过程是：

地址→MAR→AB CPU将地址信号送至地址总线

Read CPU发读命令

Wait for MFC 等待存储器工作完成信号

((MAR))→DB→MDR 读出信息经数据总线送至CPU

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

Diagram illustrating the connection between CPU and Main Memory. The CPU has MAR and MDR registers. The Main Memory has a capacity of 2^k words, each n bits long. The address bus is k bits wide, and the data bus is n bits wide. Arrows indicate Read and Write operations between the CPU and Main Memory, and MFC (Memory Full/Complete) signal from Main Memory to CPU.

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

(2)写

写操作是指将要写入的信息存入CPU所指定的存储单元中，其操作过程是：

地址→MAR→AB CPU将地址信号送至地址总线

数据→MDR→DB CPU将要写入的数据送至数据总线

Write CPU发写命令

Wait for MFC 等待存储器工作完成信号

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

Diagram illustrating the connection between CPU and Main Memory. The CPU has MAR and MDR registers. The Main Memory has a capacity of 2^k words, each n bits long. The address bus is k bits wide, and the data bus is n bits wide. Arrows indicate Read and Write operations between the CPU and Main Memory, and MFC (Memory Full/Complete) signal from Main Memory to CPU.

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

由于CPU和主存的速度存在着差距，所以两者之间的速度匹配是很关键的，通常有两种匹配方式：同步存储器读取和异步存储器读取。上面给出的读/写基本操作是以异步存储器读取来考虑的，CPU和主存间没有统一的时钟，由存储器工作完成信号（MFC）通知CPU存储器工作已完成。

对于同步存储器读取，CPU和主存采用统一时钟，因为主存速度较慢，所以CPU与之配合必须放慢速度。在这种存储器中，不需要存储器工作完成信号。

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制计算机组成原理

5.4.4 PC系列微机的存储器接口

数据总线一次能并行传送的位数，称为总线的**数据通路宽度**，常见的有8位、16位、32位、64位几种。但大多数主存储器常采取**字节编址**，每次访存允许读/写8位，以适应对字符类信息的处理。

1.8位存储器接口

如果数据总线为8位（如微机系统中的PC总线），而主存按字节编址，则匹配关系比较简单。一个总线周期中读/写8位。

2025年11月北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制计算机组成原理

2.16位存储器接口

对于16位的微处理器8086（或80286），在一个总线周期内可读/写两个字节，即先送出偶地址，然后同时读/写这个偶地址单元和随后的奇地址单元，用低8位数据总线传送偶地址单元的数据，用高8位数据总线传送奇地址单元的数据，这样读/写的字（16位）被称为**规则字**。如果读/写的是非规则字，即是从奇地址开始的字，这时需要安排两个总线周期才能实现。

2025年11月北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制计算机组成原理

为了实现这样的传送，需要将存储器分为两个存储体，一个存储体的地址均为偶数，称为**偶地址（低字节）存储体**，它与低8位数据线相连；另一个存储体的地址均为奇数，称为**奇地址（高字节）存储体**，与高8位数据线相连。8086和主存之间可以传送一个字节（8位）数据，也可以传送一个字（16位）数据。任何两个连续的字节都可以作为一个字来访问，地址值较低的字节是低位有效字节，地址值较高的字节是高位有效字节。

8086微处理器的地址线 $A_{19} \sim A_1$ 同时送至两个存储体，**BHE**（高位存储体）和最低位地址线 A_0 用来选择一个或两个存储体进行数据传送。

2025年11月北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制计算机组成原理

2025年11月北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制计算机组成原理

8086微处理器的地址线 $A_{19} \sim A_1$ 同时送至两个存储体，**BHE**（高位存储体）和最低位地址线 A_0 用来选择一个或两个存储体进行数据传送。

| BHE | A_0 | 特征 |
|-----|-------|-----------------|
| 0 | 0 | 全字（规则字）传送 |
| 0 | 1 | 在数据总线高8位上进行字节传送 |
| 1 | 0 | 在数据总线低8位上进行字节传送 |
| 1 | 1 | 备用 |

信息传送

2025年11月北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制计算机组成原理

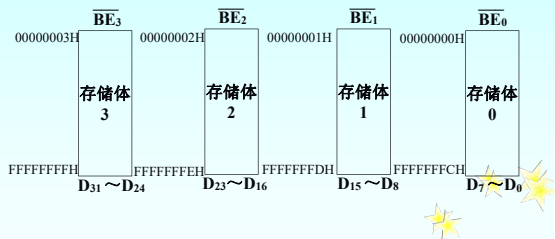
3.32位存储器接口

32位微处理器的存储器系统由4个存储体组成，每个存储体的存储空间为1GB，存储体选择通过选择信号**BE3**、**BE2**、**BE1**和**BE0**实现。如果要传送一个32位数，那么4个存储体都被选中；若要传送一个16位数，则有2个存储体（通常是**BE3**和**BE2**或者**BE1**和**BE0**）被选中；若传送的是8位数，只有一个存储体被选中。

2025年11月北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理



2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理

4.64位存储器接口

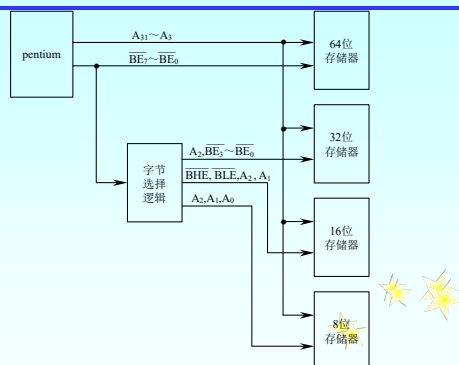
64位微处理器的存储器系统由8个存储体组成，每个存储体的存储空间为512MB（Pentium）或8GB（Pentium Pro），存储体选择通过选择信号 $\overline{BE}_7 \sim \overline{BE}_0$ 实现。如果要传送一个64位数，那么8个存储体都被选中；如果要传送一个32位数，那么4个存储体都被选中；若要传送一个16位数，则有2个存储体被选中；若传送的是8位数，只有一个存储体被选中。

2025年11月

北京理工大学计算机科学技术学院

5.4 主存储器的连接与控制

计算机组成原理



2025年11月

北京理工大学计算机科学技术学院

第5章

计算机组成原理

5.1 存储系统的组成

5.2 主存储器的组织

5.3 半导体随机存储器和只读存储器

5.4 主存储器的连接与控制

5.5 提高存储系统性能的技术

2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

速度和容量是存储器的两大主要课题，计算机的发展不断地对存储器提出更高速度和更大容量的要求。在单机系统中，提高存储器性能的主要技术有：双端口存储器、并行主存储器、高速缓冲存储器、虚拟存储器等。

2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

5.5.1 多体交叉存储技术

一. 并行访问存储器

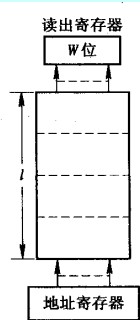
常规的主存是单体单字存储器，只包含一个存储体。在高速的计算机中，普遍采用并行主存系统，即在一个存取周期内可以并行读出多个字，依靠整体信息吞吐率的提高，以解决CPU与主存之间的速度匹配问题。

2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理



单体单字存储器

如图是一个字长为 W 位的单体主存，一次可以访问一个存储器字，所以主存最大频宽 $B_m = W/T_M$ 。假设，此存储器字长 W 与CPU所要访问的字(数据字或指令字，简称CPU字)的字长 W 相同，则CPU从主存获得信息的速率就为 W/T_M 。我们称这种主存是单体单字存储器。

2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

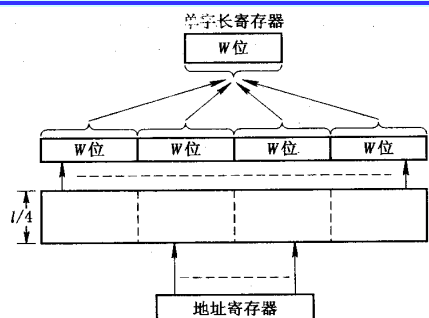
要想提高主存频宽 B_m ，使之与CPU速度匹配，显然可以想到，在同样的器件条件(即同样的 T_M)下，只有设法提高存储器的字长 W 才行。例如，改用后图的方式组成，这样，主存在一个存储周期内就可以读出4个CPU字，相当于CPU从主存中获得信息的最大速率提高到原来的4倍，即 $B_m = 4W/T_M$ 。我们称这种主存为单体多字存储器。

2025年11月

北京理工大学计算机科学技术学院

6.5 提高存储系统性能的技术

计算机组成原理

图 单体多字($m=4$)存储器

2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

方法：把 m 字 w 位的存储器改变成为 m/n 字 $n \times w$ 位的存储器(单体多字)。

逻辑实现：

把地址码分成两个部分，一部分作为存储器的地址，另一部分负责选择数据。

主要缺点：

访问冲突大----

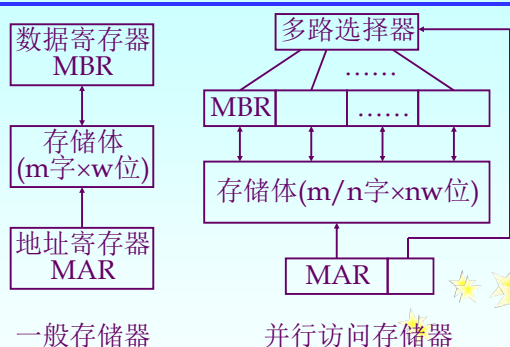
- (1) 取指令冲突、(2) 读操作数冲突
- (3) 写数据冲突、(4) 读写冲突

2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理



一般存储器

并行访问存储器

2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

二.交叉访问存储器

交叉访问存储器中有多个容量相同的存储模块(存储体)，而且各存储模块具有各自独立的地址寄存器、读/写电路和数据寄存器，这就是多体系统。各个存储体能并行工作，又能交叉工作。

多体交叉存储器：每个存储体都是CPU字的宽度，称为多体单字存取。

多体交叉又分低位交叉和高位交叉两种。

2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术计算机组成原理

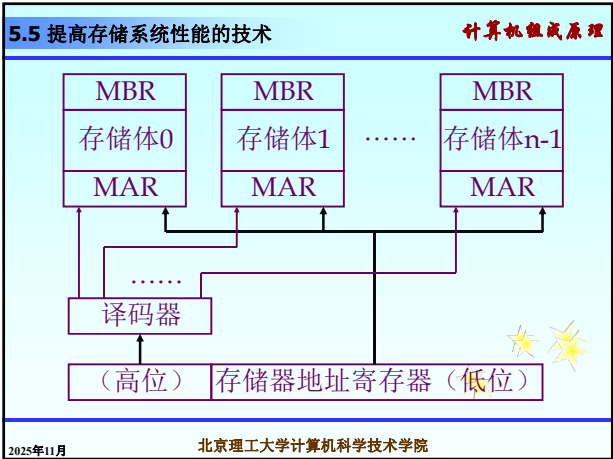
(1)高位交叉访问存储器

主要目的：**扩大存储器容量**

实现方法：用地址码的高位区分存储体号

每个存储模块都有各自**独立的控制部件**，每个存储模块可以独立工作。但由于程序的局部性原理，通常**只有一个存储模块在不停地忙碌**，其他存储模块是空闲的。

2025年11月北京理工大学计算机科学技术学院



5.5 提高存储系统性能的技术计算机组成原理

(2)低位交叉访问存储器

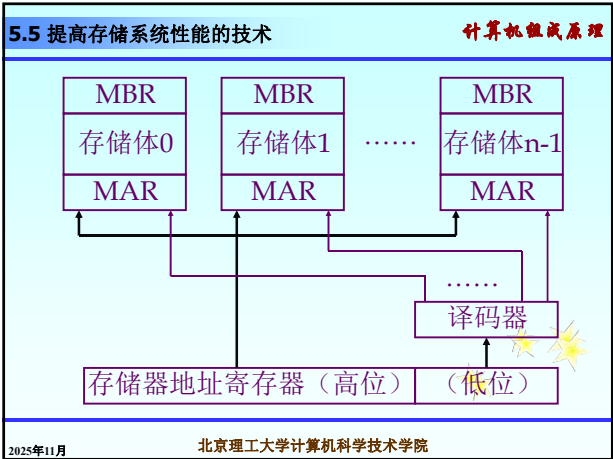
主要目的：提高存储器**访问速度**

实现方法：用地址码的低位区分存储体号

n个存储体分时启动

实际上是一种采用**流水线**方式工作的并行存储器。理论上，存储器的速度可望提高**n倍**。

2025年11月北京理工大学计算机科学技术学院

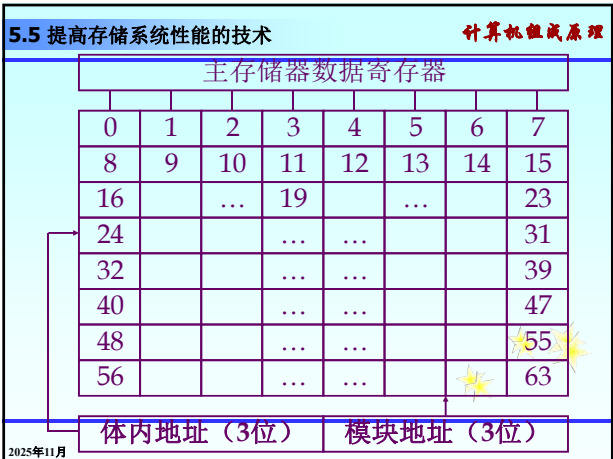


5.5 提高存储系统性能的技术计算机组成原理

| 模体 | 地址编址序列 | 对应二进制地址最末两位的状态 |
|-------|-----------------------------|----------------|
| M_0 | 0, 4, 8, ..., $4i+0$, ... | 00 |
| M_1 | 1, 5, 9, ..., $4i+1$, ... | 01 |
| M_2 | 2, 6, 10, ..., $4i+2$, ... | 10 |
| M_3 | 3, 7, 11, ..., $4i+3$, ... | 11 |

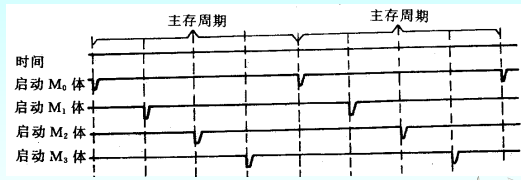
地址的模4低位交叉编址

2025年11月北京理工大学计算机科学技术学院



5.5 提高存储系统性能的技术

计算机组成原理



4个分体分时启动的时间关系

2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

5.5.2 高速缓冲存储器

一. 程序的局部性原理

程序的局部性包括：时间局部性和空间局部性。时间局部性是指如果一个存储单元被访问，则可能该单元会很快被再次访问。这是因为程序存在着循环。空间局部性是指如果一个存储单元被访问，则该单元邻近的单元也可能很快被访问。这是因为程序中大部分指令是顺序存储、顺序执行的，数据一般也是以向量、数组、树、表等形式簇聚地存储在一起的。

2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

高速缓冲技术就是利用程序的局部性原理，把程序中正在使用的部分存放在一个高速的容量较小的Cache中，使CPU的访存操作大多数针对Cache进行，从而提高程序的执行速度大大提高。

二. Cache的基本结构

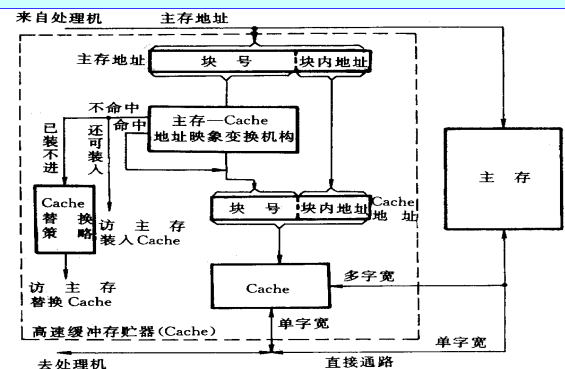
Cache和主存都被分成若干个大小相等的块，每块由若干字节组成。由于Cache的容量远小于主存的容量，它保存的信息只是主存中最活跃的若干块的副本。

2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理



2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

三.Cache的读/写操作

1. Cache的读操作

当CPU发出读请求时，如果Cache命中，就直接对Cache进行读操作，与主存无关；如果Cache不命中，则仍需访问主存，并把该块信息一次从主存调入Cache内。若此时Cache已满，则需根据某种替换算法，用这个块替换掉Cache中原来的某块信息。

2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

2. Cache的写操作

当CPU发出写请求时，如果Cache命中，会遇到如何保持Cache与主存中的内容一致的问题，处理的方法有：

- (1) 写直达法。
- (2) 写回法。

如果Cache不命中，就直接把信息写入主存，而与Cache无关。

2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

3. 与Cache相关的计算

➤ **Cache**的命中率：在一个程序执行期间，设 N_c 表示**Cache**完成存取的总次数， N_m 表示主存完成存取的总次数， h 定义为命中率。则有：

$$h = \frac{N_c}{N_c + N_m}$$

2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

➤ **Cache**—主存系统的效率 e ：

$$e = \frac{\text{访问 Cache 的时间}}{\text{平均访问时间}} \times 100\%$$

效率 e 与 **Cache命中率** 有关

设 **Cache 命中率** 为 h ，访问 **Cache** 的时间为 t_c ，访问 **主存** 的时间为 t_m

$$\text{则 } e = \frac{t_c}{h \times t_c + (1-h) \times t_m} \times 100\%$$

cpu访问的平均时间

2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

➤ 例：某计算机系统的内存储器系统由**Cache**和主存构成，**Cache**存取周期为**15ns**，主存取周期为**150ns**。已知在一段时间内，**CPU**访问**Cache****5000**次，访问主存**100**次。

问：

Cache的命中率是多少？

CPU访问主存的平均时间？

Cache—主存系统的效率是多少？

2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

➤ 解：

(1) **Cache**的命中率为 $5000/(5000+100) \approx 0.98$

(2) 平均访问时间 = $15\text{ns} \times 0.98 + 150\text{ns} \times (1-0.98) = 17.7\text{ns}$

(3) 由题可知，访问主存的时间是访问**Cache**时间的10倍（ $150/15=10$ ）

设访问**Cache**的时间为 t ，访问主存的时间为 $10t$ ，**Cache**—主存系统的效率为 e ，则

2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

$$\begin{aligned} e &= \frac{\text{访问Cache的时间}}{\text{平均访存时间}} \times 100\% \\ &= \frac{t}{0.98 \times t + (1-0.98) \times 10t} \times 100\% \\ &= \frac{1}{1.18} \times 100\% = 84.7\% \end{aligned}$$

2025年11月

北京理工大学计算机科学技术学院

5.5提高存储系统性能的技术

计算机组成原理

四.地址映像

地址映像：把存放在主存中的程序按照某种规则装入到**Cache**中，并建立主存地址与**Cache**地址之间的对应关系。

地址变换：当程序已经装入到**Cache**之后，在实际运行过程中，把主存地址变换成**Cache**地址。

三种地址映像的方法：全相联映像、直接映像和组相联映像。

选取地址映像方法的主要依据：

地址变换的硬件要容易实现；地址变换的速度要快；主存空间利用率要高；发生块冲突的概率要小。

2025年11月

北京理工大学计算机科学技术学院

5.5提高存储系统性能的技术

计算机组成原理

1、全相联映象和变换

映象规则：主存中的任意一块都可以映象到Cache中的任意一块。

如果Cache的块数为 C_b ，主存的块数为 M_b ，映象关系共有： $C_b \times M_b$ 种。

用硬件实现非常复杂。

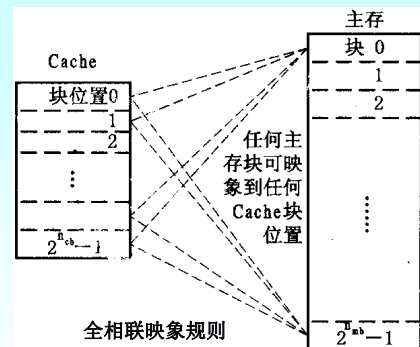


2025年11月

北京理工大学计算机科学技术学院

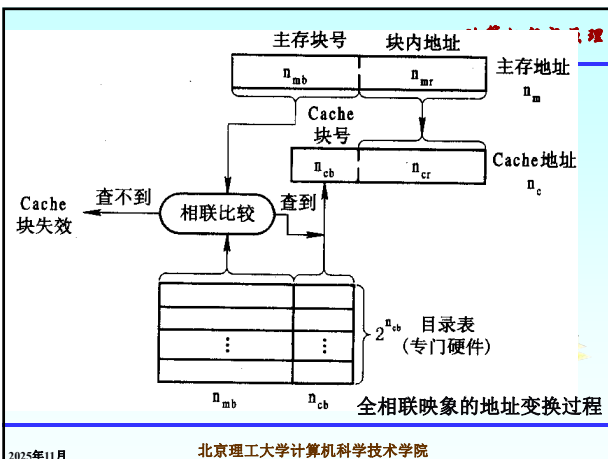
5.5提高存储系统性能的技术

计算机组成原理



2025年11月

北京理工大学计算机科学技术学院



2025年11月

北京理工大学计算机科学技术学院

5.5提高存储系统性能的技术

计算机组成原理

全相联映象方式的主要优点：

块冲突概率比较小。

Cache的利用率高。

全相联映象方式的主要缺点：

需要一个相联存储器，其代价很高。相联比较所花费的时间将影响Cache的访问速度。



2025年11月

北京理工大学计算机科学技术学院

5.5提高存储系统性能的技术

计算机组成原理

2、直接映象及其变换

映象规则：主存中一块只能映象到Cache的一个特定的块中。

计算公式： $b = B \bmod C_b$ ，其中：

b 为Cache的块号，

B 是主存的块号，

C_b 是Cache的块数。

整个Cache地址与主存地址的低位部分完全相同。

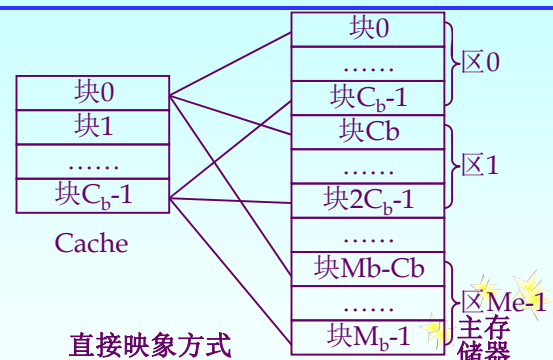


2025年11月

北京理工大学计算机科学技术学院

5.5提高存储系统性能的技术

计算机组成原理



2025年11月

北京理工大学计算机科学技术学院

5.5提高存储系统性能的技术

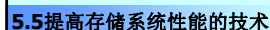
计算机组成原理

地址变换过程:

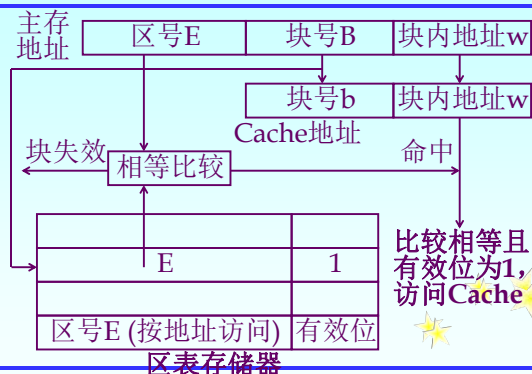
- 1、用主存地址中的块号**B**去访问区号存储器；
- 2、把读出来的区号与主存地址中的区号**E**进行比较；
- 3、比较结果相等，且有效位为**1**，则**Cache**命中；
- 4、比较结果相等，有效位为**0**，表示**Cache**中的这一块已经作废；
- 5、比较结果不相等，有效位为**0**，表示**Cache**中的这一块是空的；
- 6、比较结果不相等，有效位为**1**，表示原来在**Cache**中的这一块是有用的。

2025年11月

北京理工大学计算机科学技术学院



计算机组成原理



2025年11月

北京理工大学计算机科学技术学院

5.5提高存储系统性能的技术

计算机组成原理

直接映象方式的主要优点:

硬件实现很简单,不需要相联访问存储器。访问速度也比较快,实际上不做地址变换。

直接映象方式的主要缺点:

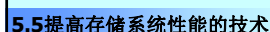
块的冲突率较高。

提高Cache速度的一种方法:

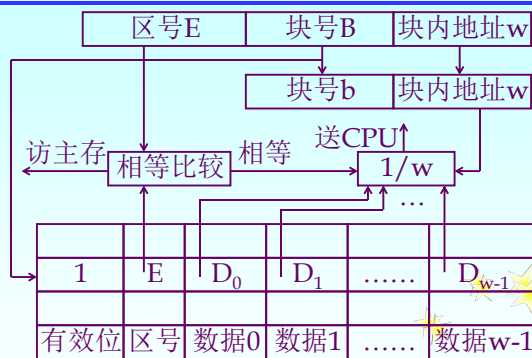
把区号存储器与**Cache**合并成一个存储器。

2025年11月

北京理工大学计算机科学技术学院



计算机组成原理



2025年11月

按地址访问的Cache

5.5提高存储系统性能的技术

计算机组成原理

3、组相联映象及其变换

组相联映像实际上是全相联映像和直接映像的折衷方案，所以其优点和缺点介于全相联和直接映像方式的优缺点之间。

映象规则:

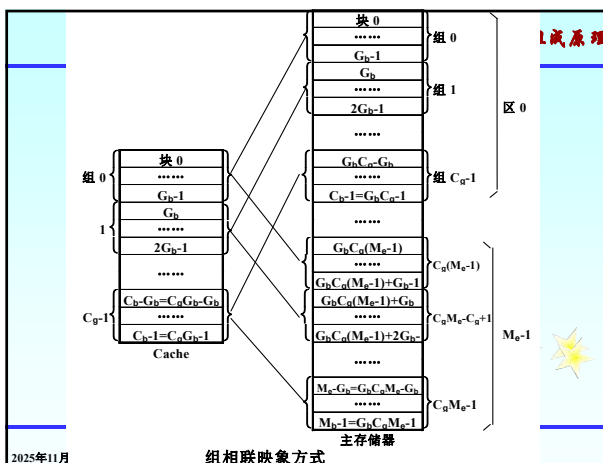
主存和Cache按同样大小划分成块，
还按同样大小划分成组。

从主存的组到Cache的组之间采用直接映象方式。

在两个对应的组内部采用全相联映象方式。

2025年11月

北京理工大学计算机科学技术学院



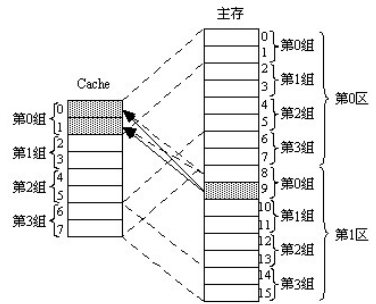
2025年11月

组相联映象方式

5.5提高存储系统性能的技术

计算机组成原理

例：主存分成2区，每区4组，每组2块；
Cache分为4组，每组2块。主存的第9块将映像到Cache的第0或1块的位置上。



2025年11月

北京理工大学计算机科学技术学院

5.5提高存储系统性能的技术

计算机组成原理

组相联映象方式的优点：

块的冲突概率比较低；
块的利用率大幅度提高；
块失效率明显降低。

组相联映象方式的缺点：

实现难度和造价要比直接映象方式高。

2025年11月

北京理工大学计算机科学技术学院

5.5提高存储系统性能的技术

计算机组成原理

地址变换过程：

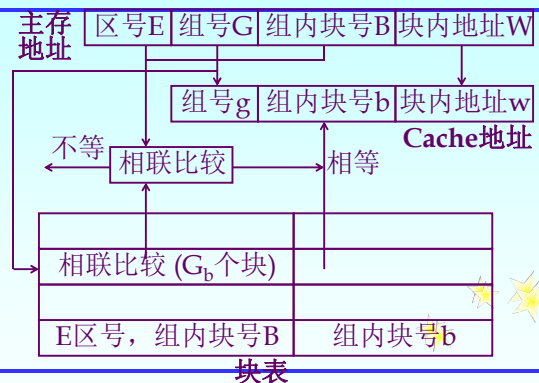
用主存地址的组号G按地址访问块表存储器。

把读出来的一组区号和块号与主存地址中的区号和块号进行相联比较，如果有相等的，表示Cache命中；如果没有相等的，表示Cache没有命中。

通常，Cache存储器的地址映像规则使用组相联或直接映像，而不采用全相联映像。否则，主存-Cache的地址映像表太大，查表速度太慢，硬件无法实现。

5.5提高存储系统性能的技术

计算机组成原理



2025年11月

北京理工大学计算机科学技术学院

5.5提高存储系统性能的技术

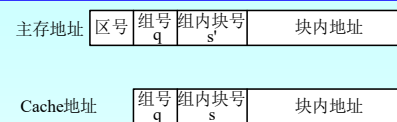
计算机组成原理

组相联映像中的Cache地址和主存地址

假设Cache空间和主存空间都分成组，每组为S块 ($S=2^s$)，整个Cache是一个区，被分成Q组 ($Q=2^q$)。主存地址分为四部分：区号、组号、组内块号、块内地址；Cache地址分为三部分：组号、组内块号、块内地址。主存地址的组号、组内块号分别由q、s'位表示，它们的宽度和位置与Cache地址的q、s是一致的。

5.5提高存储系统性能的技术

计算机组成原理



当组相联映像的S ($S=2^s$) 值大到等于Cache的块数时就成了全相联映像，而当S值小到只有1块时就变成了直接映像。

2025年11月

北京理工大学计算机科学技术学院

2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

五. 替换算法

常用的方法有三种：

1. 随机算法
2. 先进先出 (FIFO) 算法
3. 近期最少使用 (LRU) 算法



2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

提高存储系统速度的两条途径：

- 一是提高命中率 H 。
- 二是两个存储器的速度不要相差太大。
其中第二条有时做不到(如虚拟存储器)，主要依靠提高命中率。



2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

5.5.2 虚拟存储器

虚拟存储器由主存储器和联机工作的辅助存储器(通常为磁盘存储器)共同组成，这两个存储器在硬件和系统软件的共同管理下工作，对于应用程序员，可以把它们看作是一个单一的存储器。

虚拟存储器将主存或辅存的地址空间统一编址，形成一个庞大的存储空间。在这个大空间里，用户可以自由编程，完全不必考虑程序在主存是否装得下以及这些程序将来在主存中的实际存放位置。



2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

用户编程的地址称为虚地址或逻辑地址，实际的主存单元地址称为实地址或物理地址，**虚地址空间要比实地址空间大得多。**

在实际的物理存储层次上，所编程序和数据在操作系统管理下，先送入磁盘，然后操作系统将当前运行所需要的部分调入主存，供CPU使用，其余暂不运行部分留在磁盘中。



2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

虚拟存储器和Cache—主存有不少相同之处。不过**Cache—主存的控制完全由硬件实现，所以对各类程序员是透明的；而虚拟存储器的控制是软硬相结合的，对于设计存储管理软件的系统程序员来说是不透明的，对于应用程序员来说是透明的。**



2025年11月

北京理工大学计算机科学技术学院

5.5 提高存储系统性能的技术

计算机组成原理

- ▮ 页式虚拟存储器
- ▮ 段式虚拟存储器
- ▮ 段页式虚拟存储器



2025年11月

北京理工大学计算机科学技术学院

第5章 小结

计算机组成原理

5.1 存储系统的组成

☞ 存储器分类

按存取方式分类，按信息的可保存性分类

☞ 存储系统层次结构

Cache—主存层次，主—辅存层次

5.2 主存储器的组织

☞ 主存储器的基本结构



2025年11月

北京理工大学计算机科学技术学院

第5章 小结

计算机组成原理

☞ 主存储器的存储单元

位，存储字，存储单元，存储体

☞ 主存储器的主要技术指标

存取时间 T_a ，存取周期 T_m ，主存带宽 B_m

☞ 数据在主存中的存放



2025年11月

北京理工大学计算机科学技术学院

第5章 小结

计算机组成原理

5.3 半导体随机存储器和只读存储器

☞ RAM记忆单元电路

☞ 动态RAM的刷新

集中、分散、异步刷新方式的区别

☞ RAM芯片分析

地址译码方式

☞ 半导体只读存储器

ROM的类型



2025年11月

北京理工大学计算机科学技术学院

第5章 小结

计算机组成原理

5.4 主存储器的连接与控制

☞ 主存容量的扩展

字扩展，位扩展，同时扩展

☞ 存储芯片的地址分配和片选

☞ 主存储器和CPU的连接

硬连接，软连接



2025年11月

北京理工大学计算机科学技术学院

第5章 小结

计算机组成原理

5.5 提高存储系统性能的技术

☞ Cache计算

☞ 虚拟存储器



2025年11月

北京理工大学计算机科学技术学院