



河北师范大学软件学院
Software College of Hebei Normal University

计算机组成原理

第五章 存储系统和结构

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。



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 提高主存读写速度的技术

5.6 多体交叉存储技术

5.7 高速缓冲存储器

5.8 虚拟存储器





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 提高主存读写速度的技术

5.6 多体交叉存储技术

5.7 高速缓冲存储器

5.8 虚拟存储器



5.1 存储系统的组成

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。

5.1 存储系统的组成

(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。

5.1 存储系统的组成

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**。

(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**。

5.1 存储系统的组成

(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。

5.1 存储系统的组成

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。

5.1 存储系统的组成

(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）。

5.1 存储系统的组成

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。

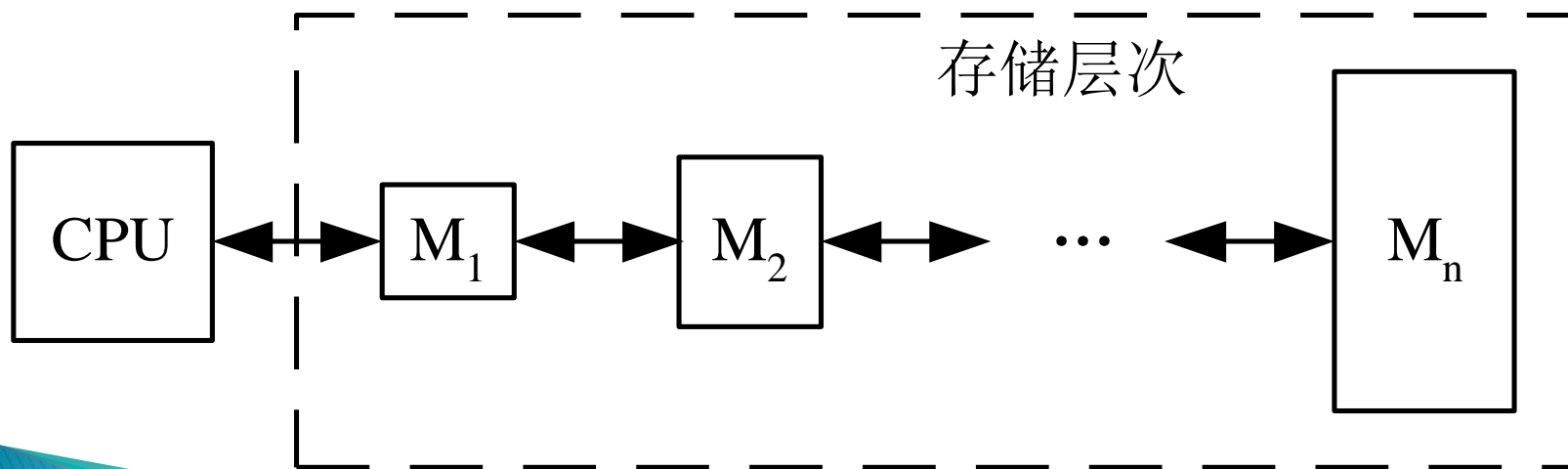
如果某个存储单元所存储的信息被读出时，原存信息将被破坏，则称破坏性读出。具有破坏性读出的存储器，每当一次读出操作之后，必须紧接一个重写（再生）的操作，以便恢复被破坏的信息。

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。

5.1 存储系统的组成

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。

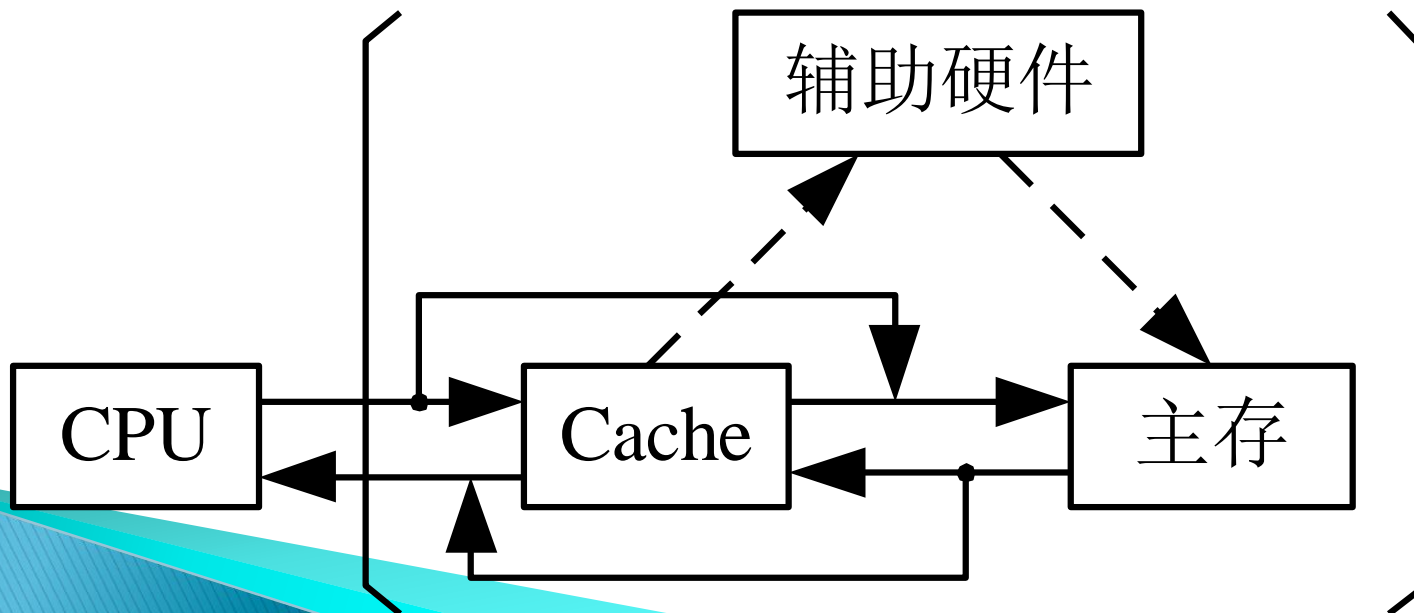


5.1 存储系统的组成

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）。

5.1 存储系统的组成

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。





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 提高主存读写速度的技术

5.6 多体交叉存储技术

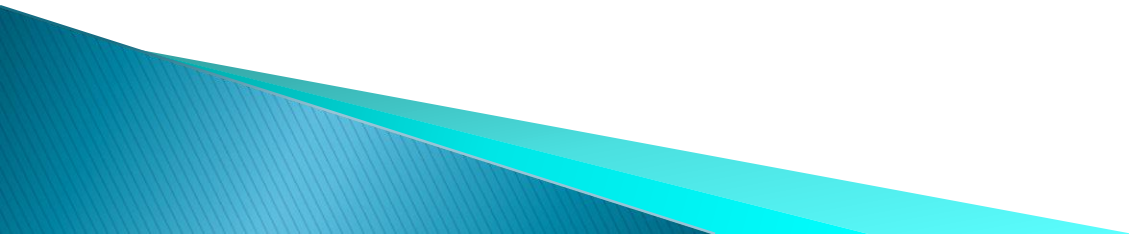
5.7 高速缓冲存储器

5.8 虚拟存储器



5.2 主存储器的组织

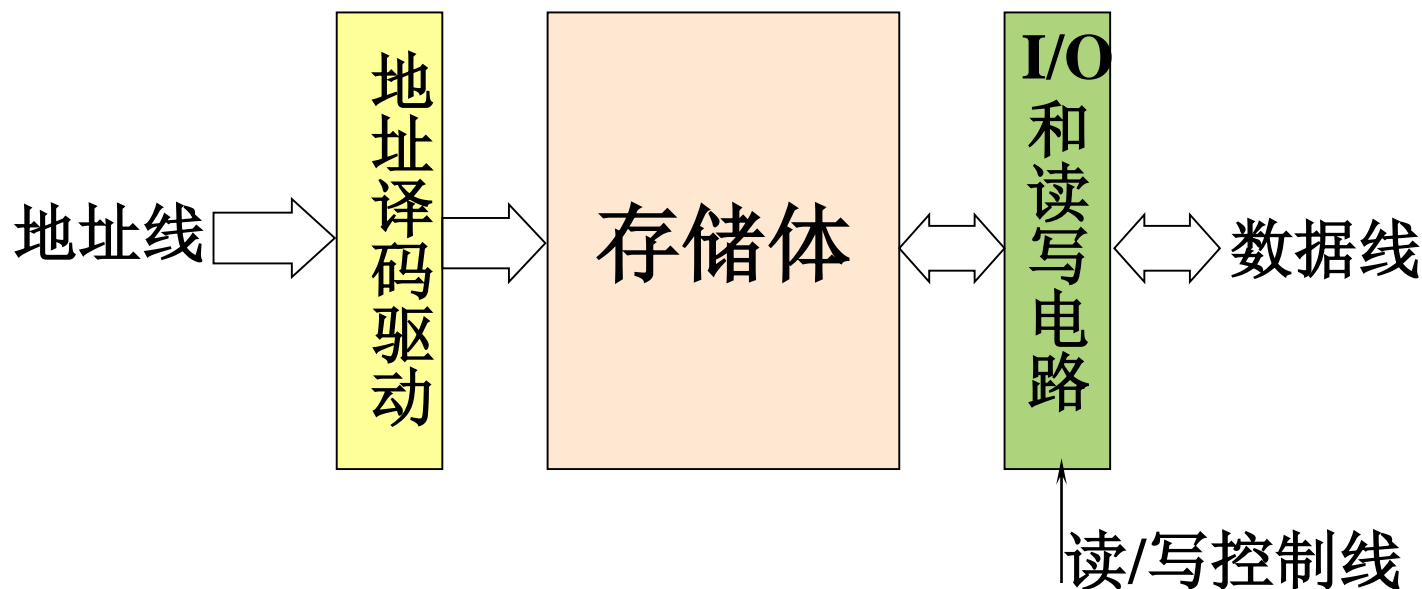
主存储器是整个存储系统的核心，它用来存放计算机运行期间所需要的程序和数据，CPU可直接随机地对它进行访问。



5.2 主存储器的组织

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。



5.2 主存储器的组织

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。

5.2 主存储器的组织

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。

5.2 主存储器的组织

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。

5.2 主存储器的组织

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。

5.2 主存储器的组织

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。

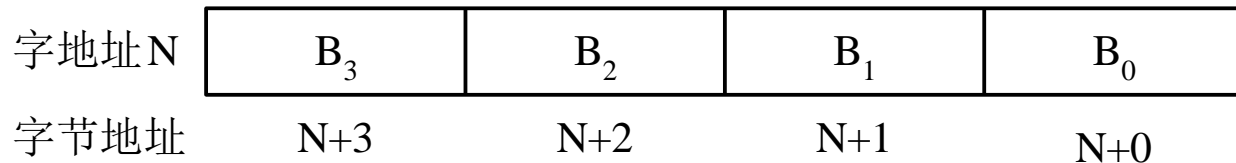
5.2 主存储器的组织

字地址	字节地址			
0	0	1	2	3
4	4	5	6	7
8	8	9	10	11

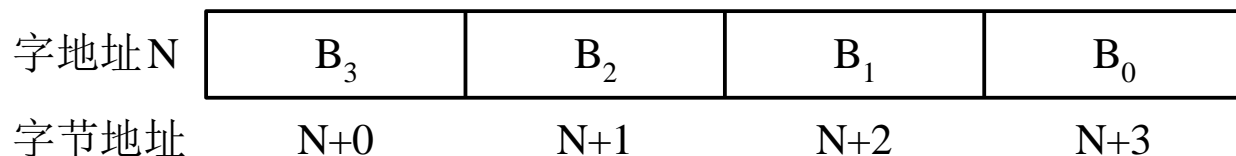
字地址	字节地址	
0	1	0
2	3	2
4	5	4

5.2 主存储器的组织

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。



(a)



(b)

5.2 主存储器的组织

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。

5.2 主存储器的组织

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）。

5.2 主存储器的组织

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。

5.2 主存储器的组织

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）。

5.2 主存储器的组织

(3) 主存带宽 B_m

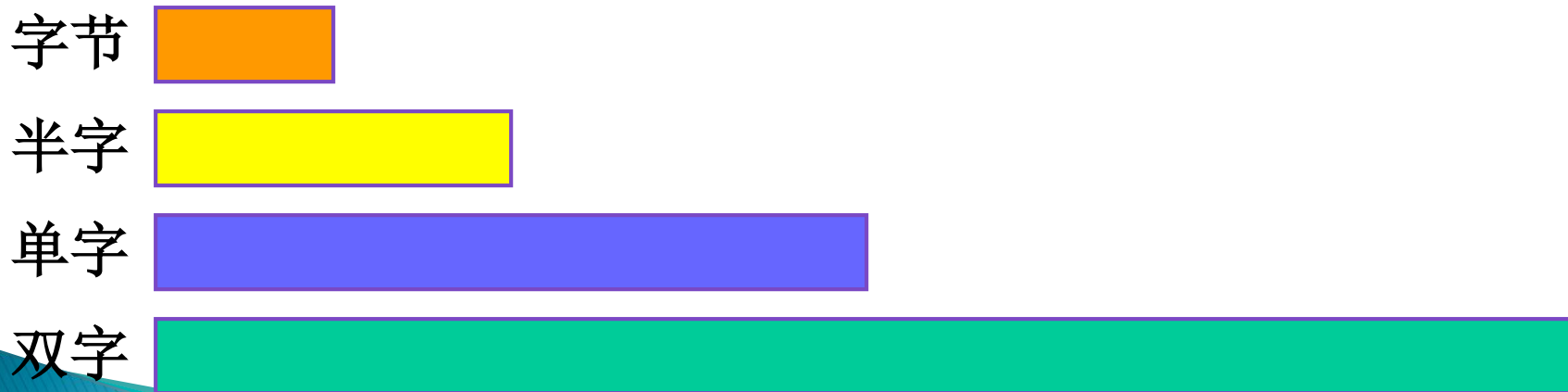
与存取周期密切相关的指标是主存的带宽，它又称为数据传输率，表示每秒从主存进出信息的最大数量，单位为字/秒或字节/秒或位/秒。目前，主存提供信息的速度还跟不上CPU处理指令和数据的速度，所以，主存的带宽是改善计算机系统瓶颈的一个关键因素。为了提高主存的带宽，可以采取的措施有：

- 缩短存取周期；
- 增加存储字长；
- 增加存储体。

5.2 主存储器的组织

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）。



5.2 主存储器的组织

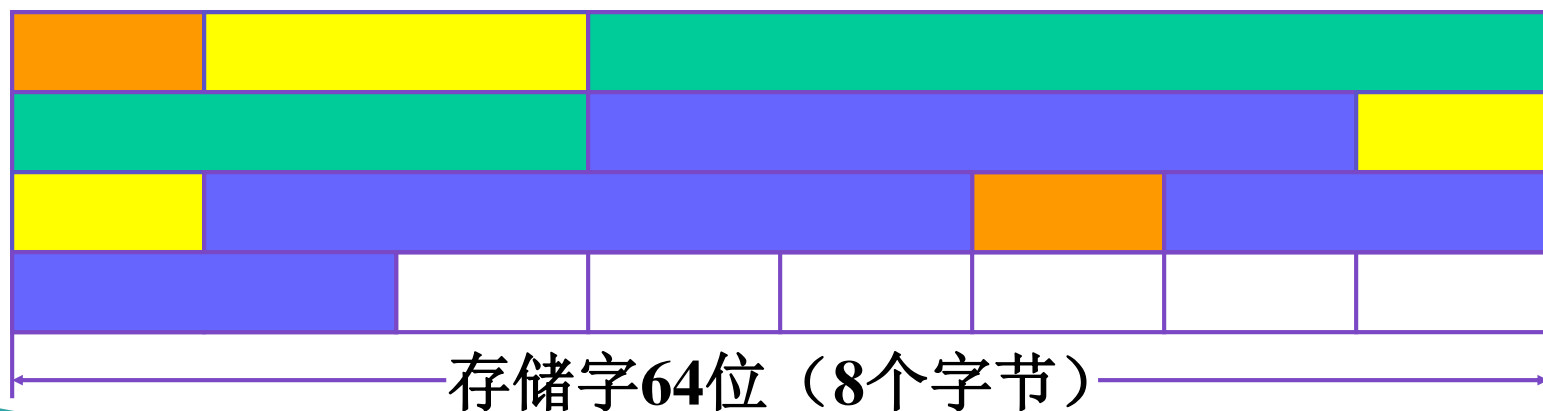
请注意：此例中数据字长（32位）不等于存储字长（64位）。

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

(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。

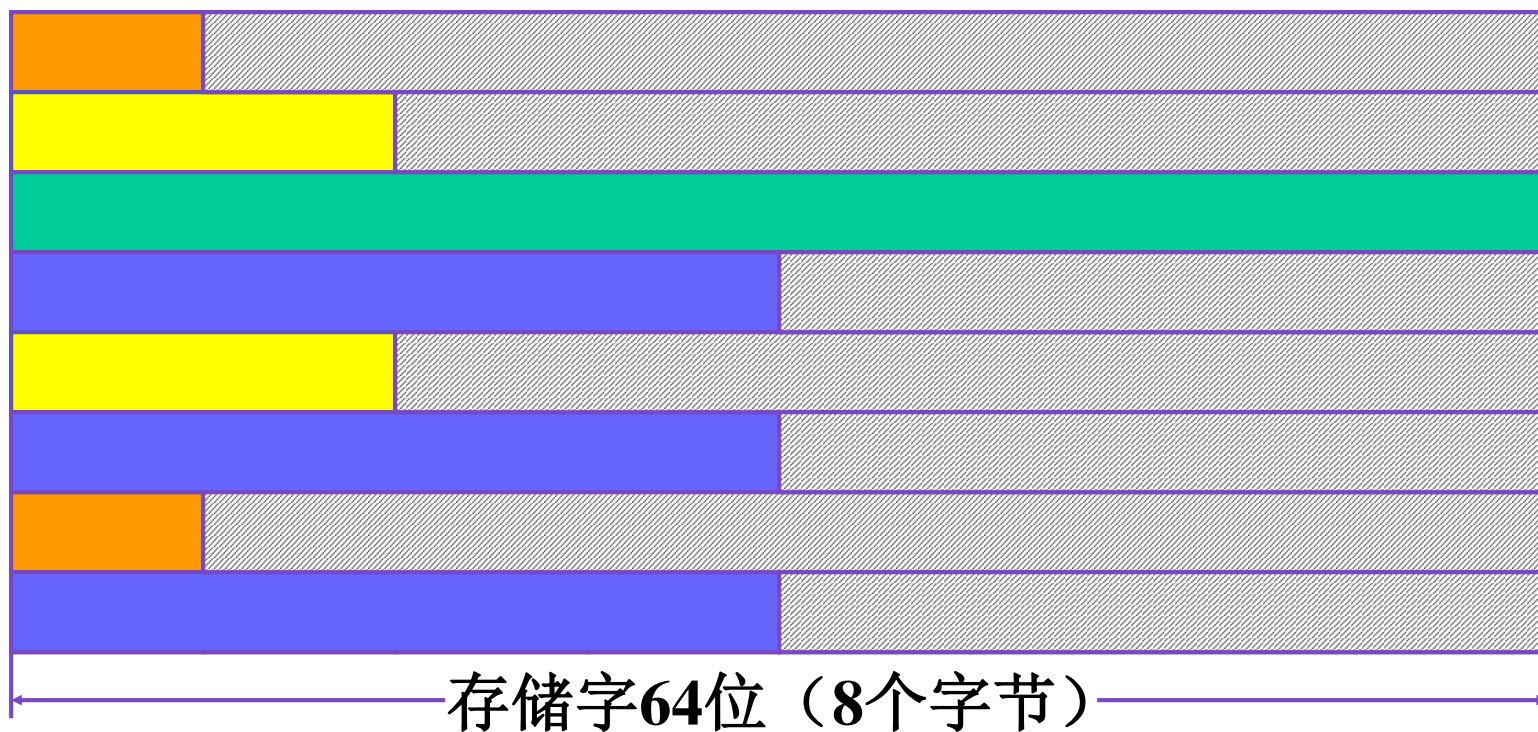


5.2 主存储器的组织

(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

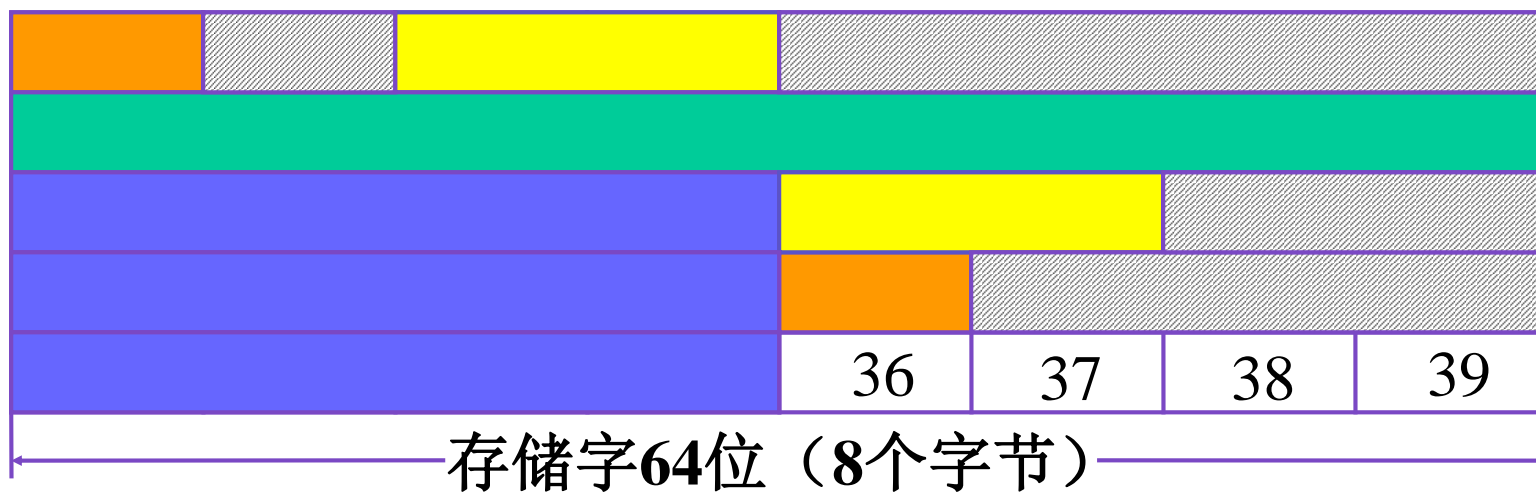


5.2 主存储器的组织

(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。

5.2 主存储器的组织





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 提高主存读写速度的技术

5.6 多体交叉存储技术

5.7 高速缓冲存储器

5.8 虚拟存储器

5.3 半导体随机存储器和只读存储器

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。

5.3 半导体随机存储器和只读存储器

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。

- ▶ 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

有兴趣的，自己看书

5.3 半导体随机存储器和只读存储器

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间隔为2ms，也就是说，应在2ms内，将全部存储体刷新一遍。

5.3 半导体随机存储器和只读存储器

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作为单位进行的。

2.刷新方式

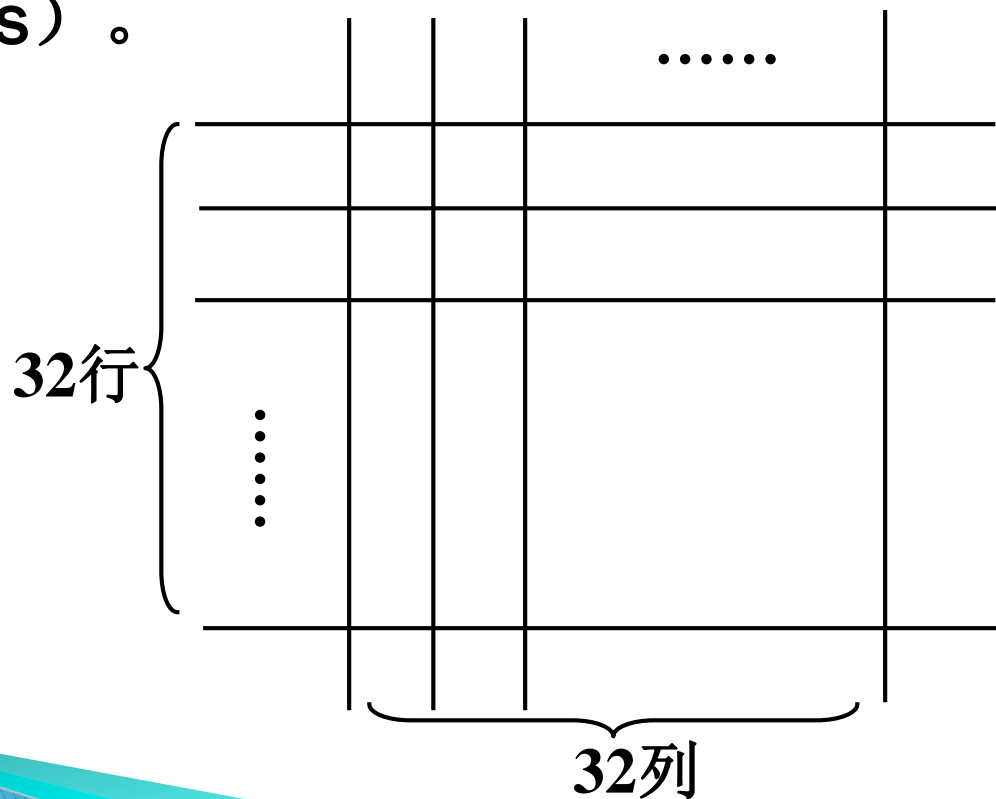
集中式

分散式

异步式

5.3 半导体随机存储器和只读存储器

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（ $0.5 \mu\text{s}$ ）。



5.3 半导体随机存储器和只读存储器

(1)集中刷新方式

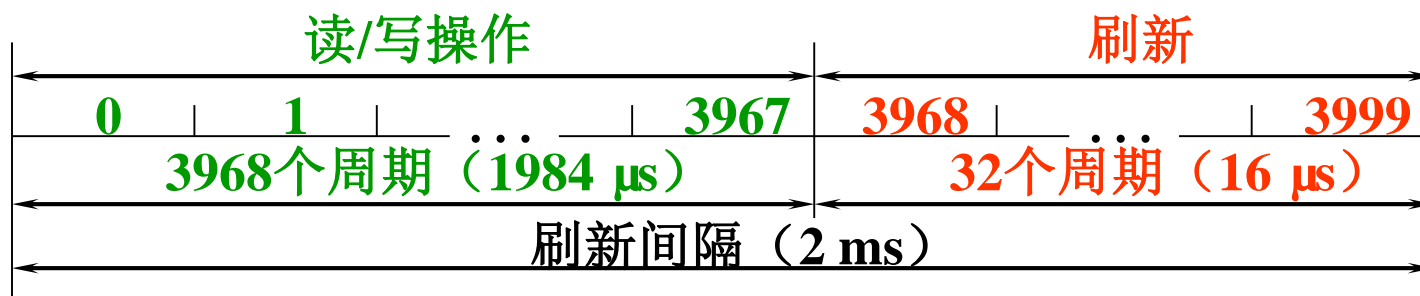
在允许的最大刷新闻隔内，按照存储芯片容量的大小集中安排若干个刷新周期，刷新时停止读写操作。

刷新时间=存储体矩阵行数×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。

5.3 半导体随机存储器和只读存储器

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。



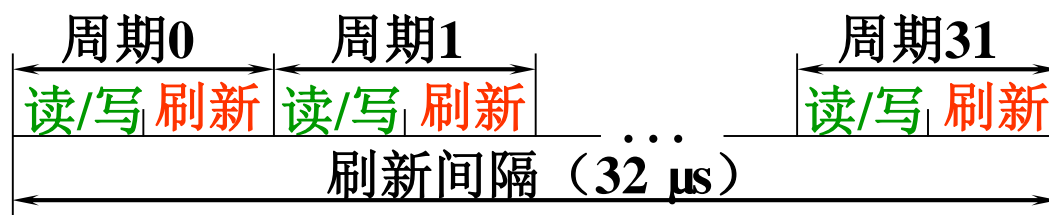
5.3 半导体随机存储器和只读存储器

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。

5.3 半导体随机存储器和只读存储器

(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。



5.3 半导体随机存储器和只读存储器

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 $0.5\ \mu\text{s}$ ，则系统的存取周期应为 $1\ \mu\text{s}$ ，降低了整机的速度；第二是刷新过于频繁（本例中每 $32\ \mu\text{s}$ 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 ）。

5.3 半导体随机存储器和只读存储器

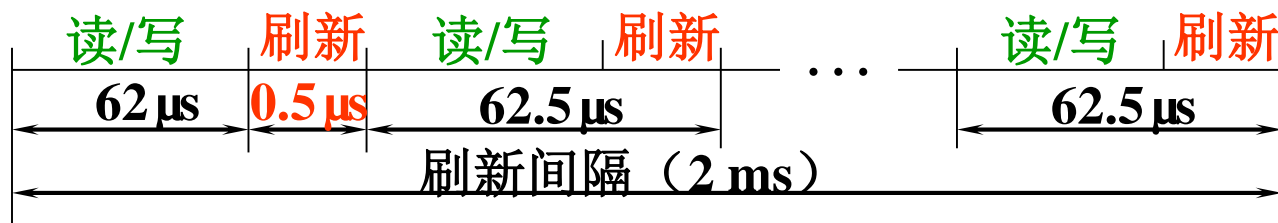
(3)异步刷新方式

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

5.3 半导体随机存储器和只读存储器

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 $=2\text{ms}/32=62.5\text{ }\mu\text{s}$ ，即每隔 $62.5\text{ }\mu\text{s}$ 安排一个刷新周期，在刷新时封锁读/写。



5.3 半导体随机存储器和只读存储器

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。

5.3 半导体随机存储器和只读存储器

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。

5.3 半导体随机存储器和只读存储器

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 $64K \times 1$ 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； $64K \times 1$ 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。

5.3 半导体随机存储器和只读存储器

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 $\overline{\text{RAS}}$ 送入存储芯片，列地址由列地址选通信号 $\overline{\text{CAS}}$ 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。

5.3 半导体随机存储器和只读存储器

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。

5.3 半导体随机存储器和只读存储器

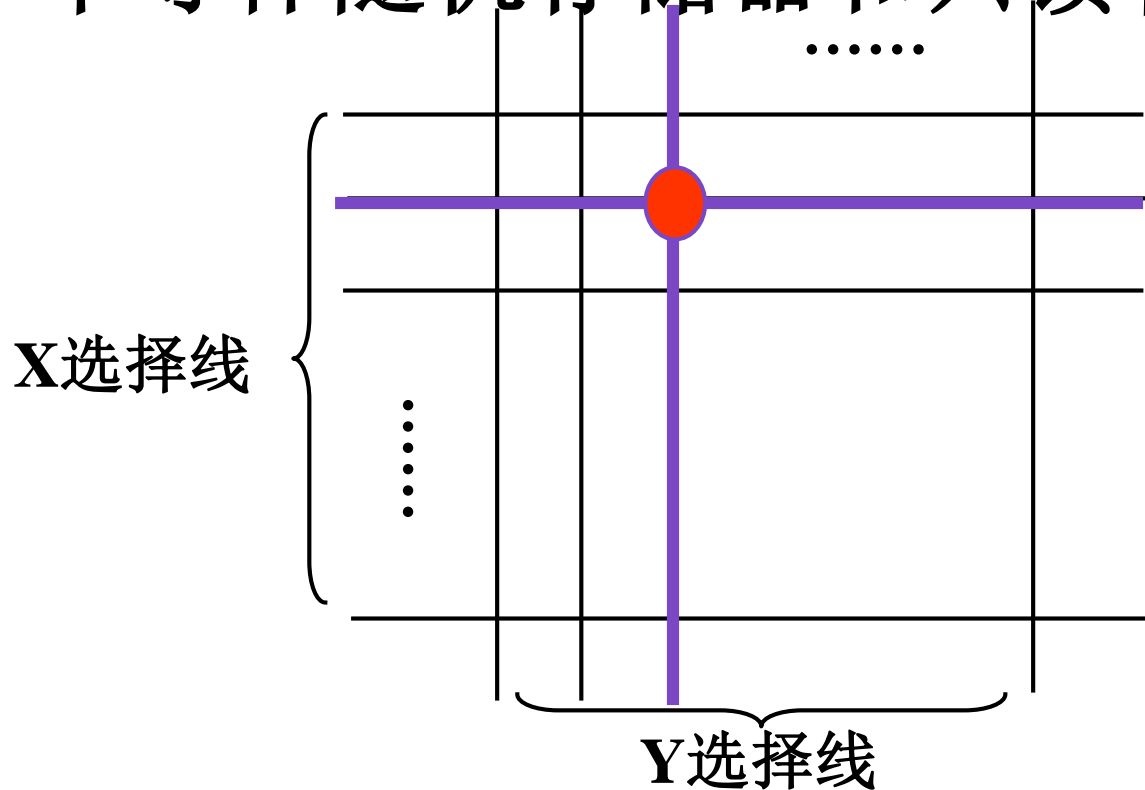
字结构的优点是结构简单，缺点是使用的外围电路多，成本昂贵。更严重的是，当字数大大超过位数时，存储器会形成纵向很长而横向很窄的不合理结构，所以这种方式只适用于容量不大的存储器。

5.3 半导体随机存储器和只读存储器

(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。

5.3 半导体随机存储器和只读存储器



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。

5.3 半导体随机存储和只读存储器

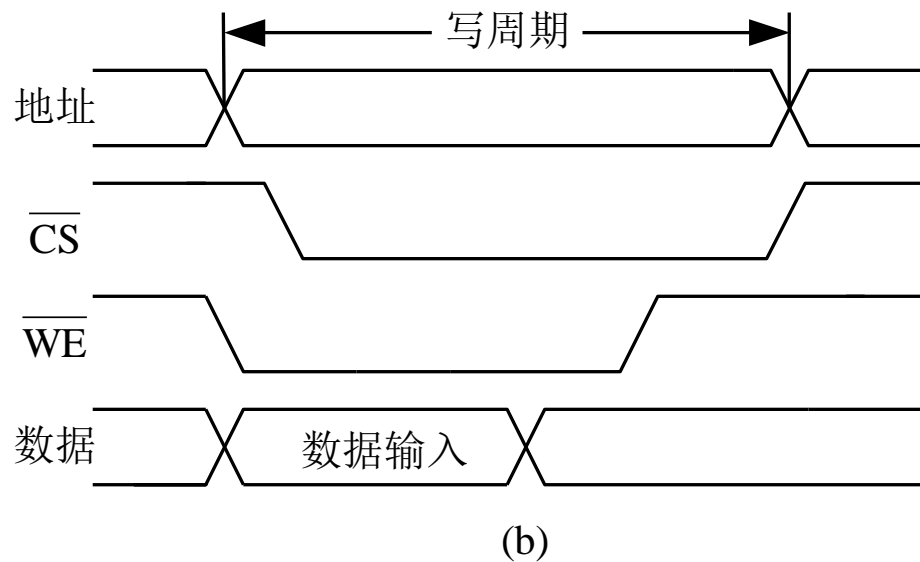
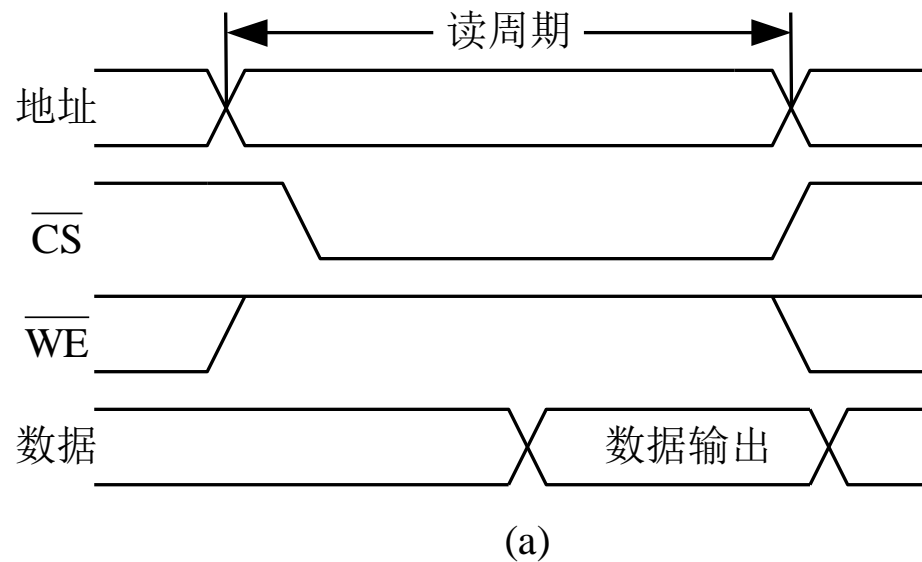
3.RAM的读/写时序

(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。

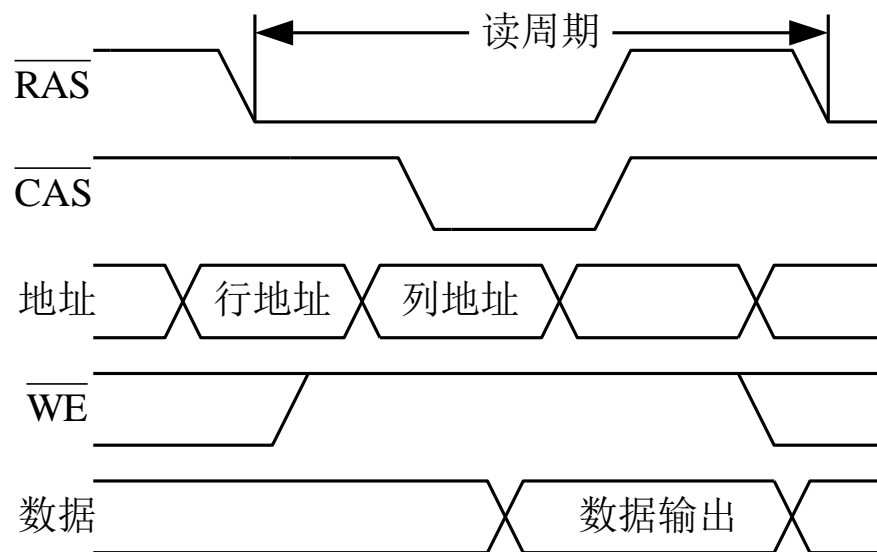
写周期与读周期相似，但除了要加地址和片选信号外，还要加一个低电平有效的写入脉冲 \overline{WE} ，并提供写入数据。

5.3 半导体随机存储器和只读存储器

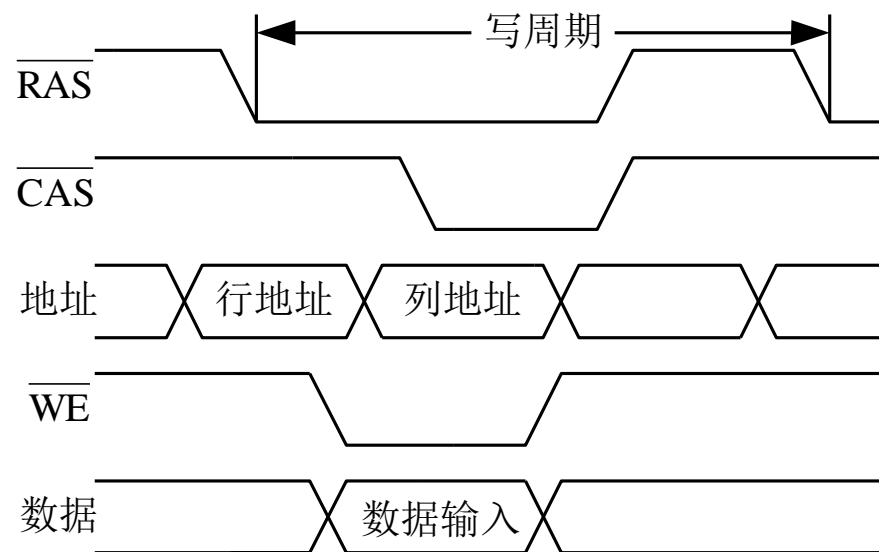


5.3 半导体随机存储器和只读存储器

(2) DRAM读/写时序



(a)



(b)

5.3 半导体随机存储器和只读存储器

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：

5.3 半导体随机存储器和只读存储器

(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。

5.3 半导体随机存储器和只读存储器

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。

5.3 半导体随机存储器和只读存储器

(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。

5.3 半导体随机存储器和只读存储器

▶ 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。

5.3 半导体随机存储器和只读存储器

(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。

5.3 半导体随机存储器和只读存储器

2.ROM芯片

地址线

数据线

片选线

V_{cc} ——+5V（工作电源）

V_{pp} ——编程电源。