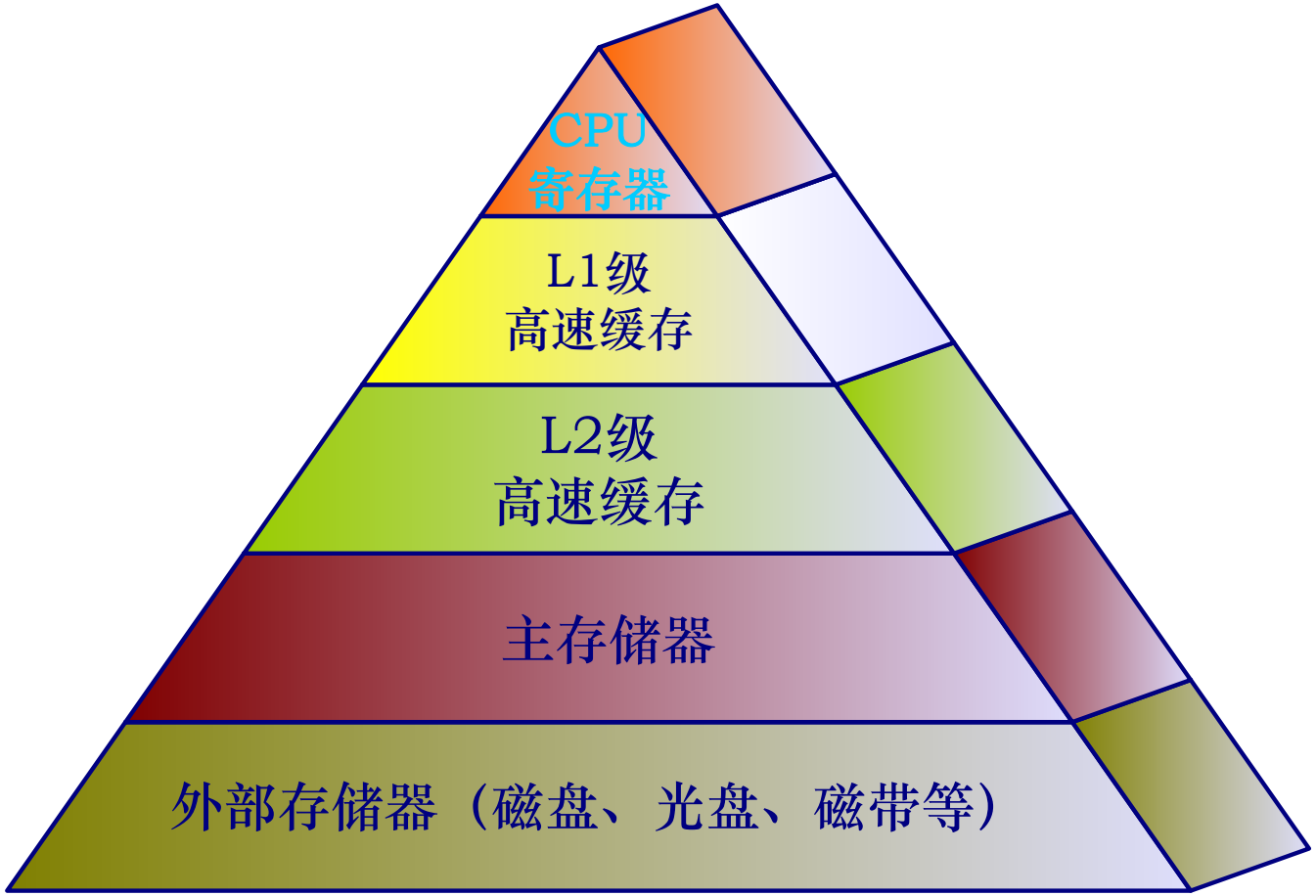
汇编与接口第六章

存储系统与技术

- 存储系统是计算机的重要组成部分之一。
- 内部存储器（简称内存）主要存储计算机当前工作需要的程序和数据，包括高速缓冲存储器（Cache，简称缓存）和主存储器。
- 外部存储器（简称外存）主要有磁性存储器、光存储器和半导体存储器等三种实现方式，存储介质有硬磁盘、软磁盘、光盘、磁带和移动存储器等。


CPU
寄存器
L1级
高速缓存
L2级
高速缓存
主存储器
外部存储器（磁盘、光盘、磁带等）

高速缓冲存储器

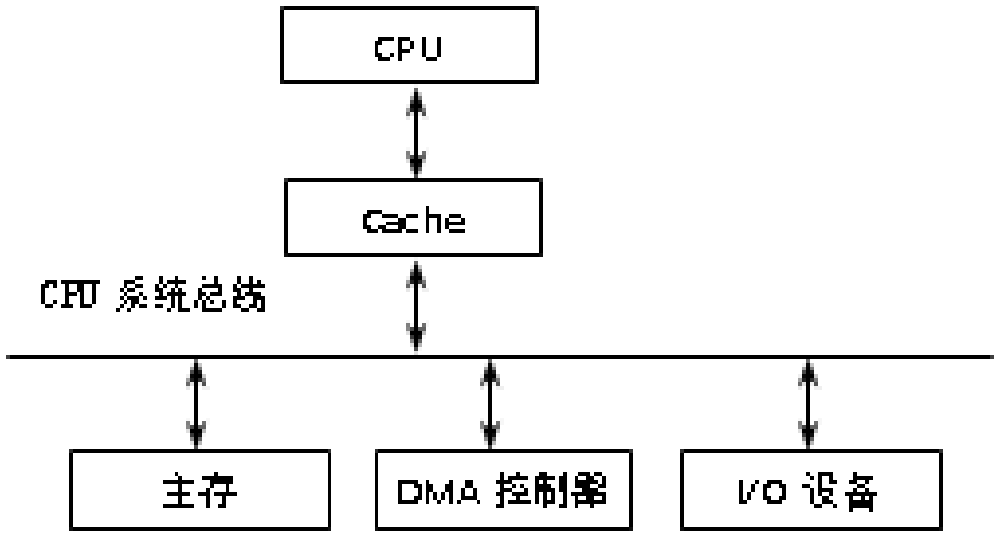
- 位于CPU与主存之间的临时存储器，一般由高速SRAM构成。SRAM只需要1~2个时钟周期就可以读写一次数据，而DRAM需要几个时钟周期才能读写一次数据，SRAM速度比DRAM快，另外SRAM的每位成本也比DRAM高，因此一般用DRAM构成主存，而SRAM构成缓存。
- Cache机制完全由硬件来实现，以避免带来额外的延迟。

Cache工作原理

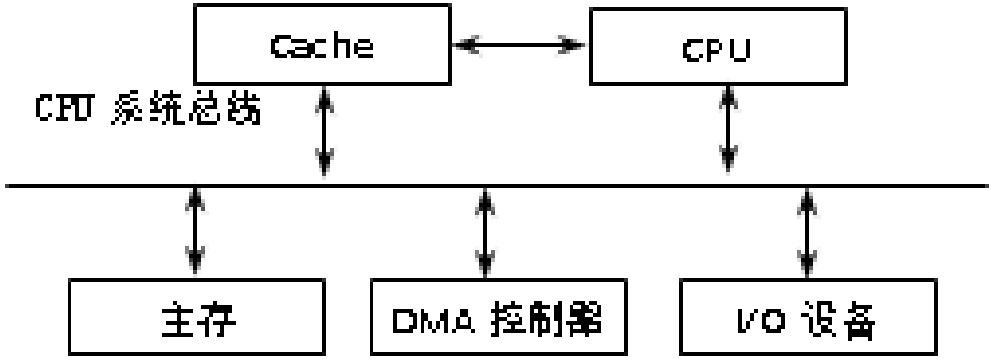
- 局部性原理
- CPU在执行程序的过程中具有局部性，在一个较短的时间间隔内，CPU所访问的内存地址往往集中在整个地址空间的一个很小范围之内。从程序和数据两方面看，程序中大部分指令是顺序执行的，数据存放在数组等结构中，是顺序存放的，具有顺序局部性（Order Locality）
- 程序和数据都具有空间局部性（Spatial Locality）特点，，即CPU将用到的信息很可能与现在正在使用的信息在空间地址上是临近的。

Cache的访问结构

- 贯通查找式（Look Through）结构
- 旁路读出式（Look Aside）结构



(a) Look Through 结构



(b) Look Aside 结构

- 贯通查找式
- cache平均访问时间= cache访问时间+（1-命中率）×未命中时主存访问时间
- 旁路读出式
- cache的平均访问时间=命中率×cache访问时间+（1-命中率）×未命中时主存访问时间

Cache映射

- 设Cache的容量为2^c个单元，Cache每次与主存交换数据块的大小为一行，一行为k个单元，$k = 2^w$。Cache一共有$m = 2^r$行，每一行用L_i表示，$0 \leqslant i < m$。由于$m = 2^c / 2^w$，所以r=c-w。
- 将主存分为若干块，块的大小和行相等，共分成$n = 2^s$块。主存有$2^s \times 2^w = 2^{s+w}$个单元，地址为s+w位。每一块用B_j表示，$0 \leqslant j < n$。
- Cache的每一行需要一个标记，以指明它是主存哪一块的副本，即它记录了Cache中的m个行与主存的n个块之间的对应关系。
- Cache的标记保存了主存的块地址，占s位。

判断命中

- 当CPU访问一个内存单元时，它的地址的高s位与Cache中每一行的标记作比较，如果某一Cache行的标记与它相等，那么这个单元就位于Cache中，称为访问cache命中。如不匹配，则未命中，必须要从主存中存取。未命中时，一般要从将这个块从主存中复制到Cache中。

Cache替换策略

- 主存的一个块要调入Cache存储器时，如果Cache存储器中没有空闲的行，就必须从中选取一行，用新的块覆盖其原有的内容。这种替换应该遵循一定的规则，其目标是选取在下一段时间内被存取的可能性最小的块，替换出Cache。这些规则称为替换策略或替换算法，由Cache中的替换部件加以实现。常用的替换算法包括随机算法、先进先出（FIFO）算法和近期最少使用（LRU）算法等。

- CPU缓存可以分为一级缓存（L1 cache），二级缓存（L2 Cache），部分高端CPU还具有三级缓存（L3 Cache）。当CPU要读取一个数据时，首先从一级缓存中查找，如果没有找到再从二级缓存中查找，依次类推。
- 早期除了CPU内核中的Cache外，在CPU电路板或主板上还设计了一部分容量稍大，但速度较低的Cache。这两部分缓存称为一级缓存和二级缓存。现代微机中一级缓存中采哈佛结构，分为数据缓存（Data Cache，D-Cache）和指令缓存（Instruction Cache，I-Cache）

Cache一致性协议

- 对Cache的操作分为读和写两种。读操作因为不涉及到内容的改变，不会导致Cache内容和对应的内存内容不一致，从Cache中将所需内容取走即可，所以读的过程比较简单，也很容易理解。而写的过程因为涉及到对内容的修改，存在导致Cache内容和对应内存内容不一致的可能性。
- 写操作分成单核CPU环境和多核CPU环境两种情况

单核CPU一致性处理

- （1）未命中时的Cache写策略
- 当CPU发出写操作命令时，如果此时数据尚未调入Cache（未命中），则数据直接写入内存。含有写入数据的内存块可以根据需要决定是否随后调入Cache中。

- （2）命中时的Cache写策略
- 直写式（Write Through）：CPU在向Cache写入数据的同时，立即把数据写入内存，以保证Cache和内存中相应单元数据的一致性。直写式策略的特点是简单可靠，但由于CPU每次更新数据时都要对内存写入，写入速度受到影响。
- 回写式（Write Back）：CPU只向Cache写入数据，不立即写入内存。Cache为每一行设置一个标志位（dirty，脏位），为1时表示Cache中的数据尚未更新到内存。要替换这一行时，数据必须先写入内存的块之后，才被其他块所使用。回写式策略的特点是发生命中时CPU更新数据较快，但Cache的结构复杂，而且在回写前会暂时出现Cache中的数据和内存不一致的情况。

多核CPU的MESI协议

- 利用MESI（Modified、Exclusive、Shared、Invalid的首字母缩写，代表四种缓存状态）及其衍生协议（比如MESIF协议和MOESI协议等）来达到目的。
- MESI对应的是修改、独占、共享、无效四种缓存段状态，任何多核系统中的缓存段都处于这四种状态之一。

- 修改（Modified）缓存段，属于脏段（dirty），它们已经被所属的处理器修改了。如果一个段处于修改状态，那么它在其他处理器缓存中的拷贝马上会变成无效状态。此外，修改缓存段如果要被替换或标记为无效，那么和回写模式下单核处理器常规的脏段处理方式一样，先要把它的内容回写到对应的内存块中。
- 独占（Exclusive）缓存段，是和对应内存块内容保持一致的一份拷贝。区别在于，如果一个处理器持有了某个E状态的缓存段，那其他处理器就不能同时持有它，所以叫“独占”。这意味着，如果其他处理器原本也持有同一缓存段，那么它们会马上变成无效状态。

- 共享（Shared）缓存段，它也是和主内存内容保持一致的一份拷贝，在这种状态下的缓存段只能被读取，不能被写入。多组缓存可以同时拥有针对同一内存地址的共享缓存段。
- 无效（Invalid）缓存段，要么已经不在缓存中，要么它的内容已经过时。为了达到缓存的目的，这种状态的段将会被忽略。一旦缓存段被标记为失效，那效果就等同于它从来没被加载到缓存中。

MESI协议的一致性处理

- 对于无效缓存段（I状态），正如前面所述，相当于未加载进Cache。如果需要对其进行读写操作，则首先需要将对应的内存块调入。此时Cache和内存对应的块内容是一致的。
- 对于共享缓存段（S状态），可以在多个处理器中存在相同的拷贝，但因为只能读不能写，所以也不存在不一致的可能性。

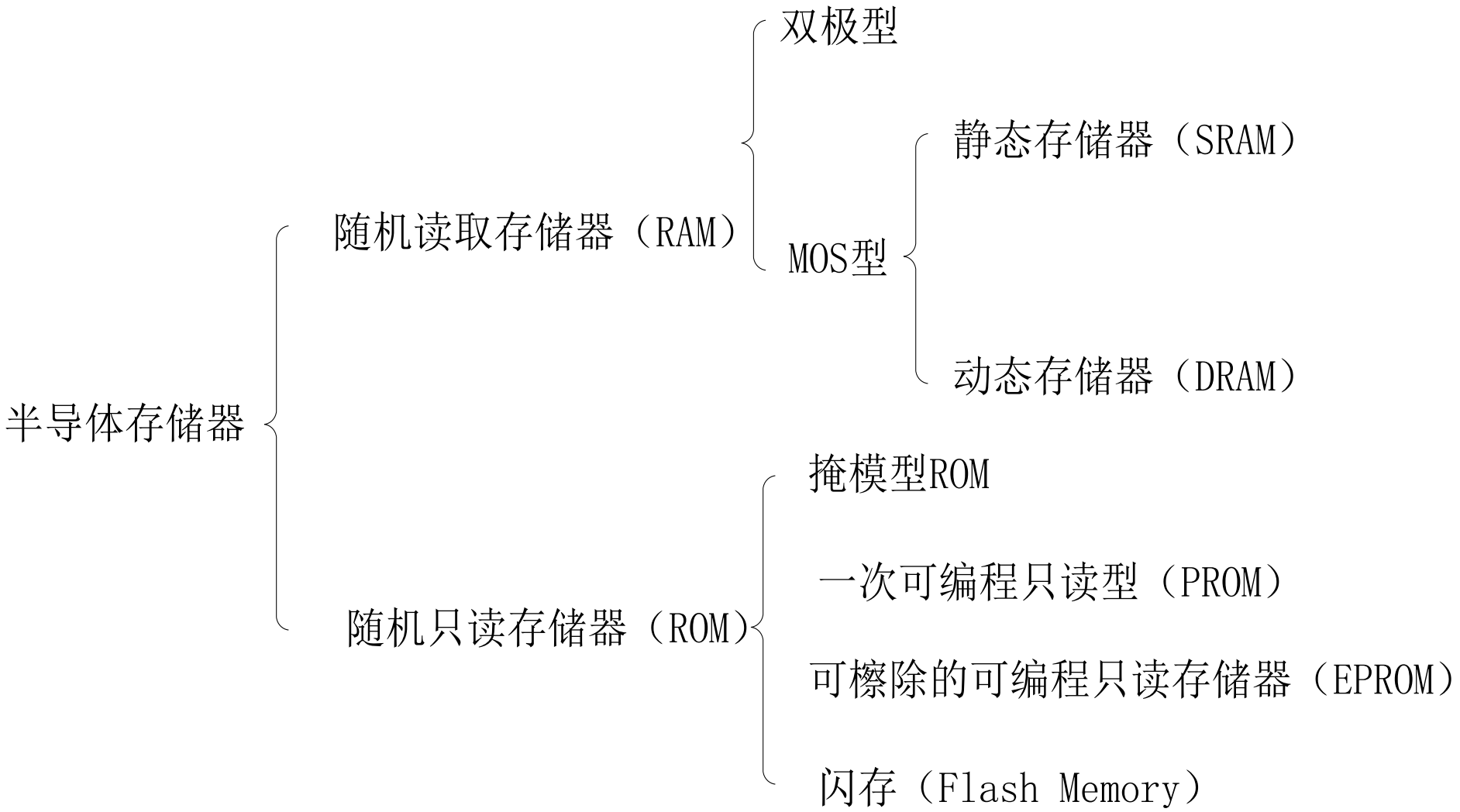
- 对于独占缓存段（E状态），表示当前Cache行中包含的数据有效，并且该数据仅在当前处理器的Cache中有效，而不在其他处理器的Cache中存在拷贝。在该Cache行中的数据是当前处理器系统中最新的数据拷贝，而且与存储器中的数据一致。
- 对于修改缓存段（M状态），表示当前Cache 行中包含的数据与存储器中的数据不一致，而且它仅在本处理器的Cache 中有效，不在其他处理器的Cache 中存在拷贝，因此其他处理器不会读出无效的、过期的数据。当处理器对这个Cache行执行替换操作时，会触发系统总线的写周期，将Cache行中被修改过的数据（脏数据）与内存中的数据进行同步，从而保持一致性。

- 只有当缓存段处于E或M状态时，处理器才能执行写操作，也就是说只有这两种状态下，处理器是独占这个缓存段的，而对应的内容在其他Cache区域没有拷贝。当处理器想写某个缓存段时，如果它没有独占权，它必须先发送一条“我要独占权”的请求给总线，这会通知其他处理器，把它们拥有的同一缓存段的拷贝失效（假设存在多个拷贝的情况）。只有在获得独占权后，处理器才能开始修改数据。因为这个缓存段只有当前一份拷贝，所以不会有任何冲突。

- E状态和M状态的差别在于，E状态的缓存段内容和对应的内存块一致，因此当退出E状态时，可以转入S状态。而M状态的缓存段内容和对应的内存块不一致，因此当退出M状态时，需先进行写内存操作。

内部存储器

- 内存分类
- 主存储器由内存芯片、电路板、金手指等部分组成的。
- 内存可分为只读存储器（Read-Only Memory，ROM）和随机存取存储器（Random Access Memory，RAM）两大类。
- 存储在ROM中的信息是非易失的（Nonvolatile），即断电后存储信息不丢失。通常用来存储不需要改变的程序或者数据。RAM只能暂时保存数据，断电后其中的数据会消失，常用来存放各种现场的输入输出数据、中间计算结果、与外存交换的信息等。

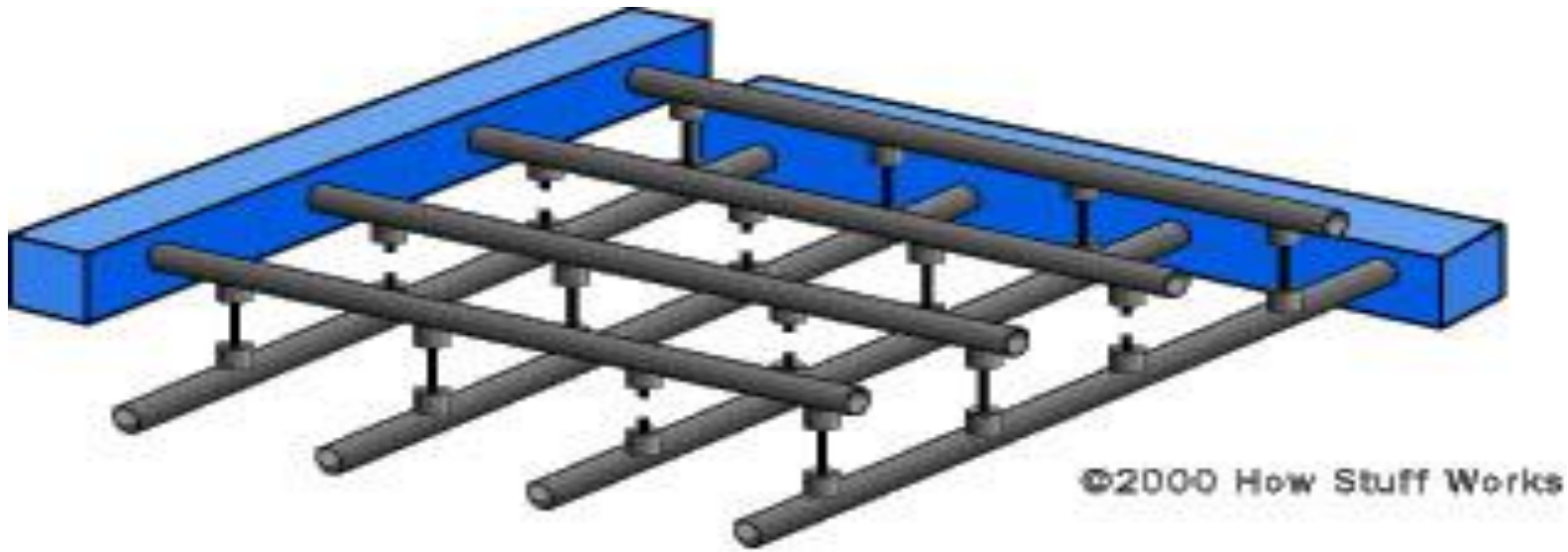

半导体存储器
随机读取存储器（RAM）
双极型
MOS型
静态存储器（SRAM）
动态存储器（DRAM）
随机只读存储器（ROM）
掩模型ROM
一次可编程只读型（PROM）
可檫除的可编程只读存储器（EPROM）
闪存（Flash Memory）

RAM

- 分为双极型（Bipolar）和MOS型两大类。
- 双极型采用晶体管触发器（Flip-Flop）为基本存储单位，存取速度高、晶体管较多，故集成度相对于MOS型低，功耗大，成本高。一般这种类型的RAM常用在速度要求较高的微机中或者作为Cache。
- MOS型RAM，又可以细分为静态RAM（Static RAM，SRAM）和动态RAM（Dynamic RAM，DRAM）两种。SRAM采用六管构成的触发器作为基本存储单位，集成度高于双极型，低于DRAM；不需要刷新，功耗低于双极型，高于DRAM。DRAM采用单管线路组成基本存储单位，集成度高、功耗低，由于依靠电容存储电荷，因此采用DRAM实现的存储器需要定期刷新。

ROM

- 包括掩模型ROM、可编程的只读存储器、可擦除可编程的只读存储器和闪存等。
- （1）掩模型MROM（Mask ROM）：由制造厂家写入数据。用户不能修改掩模ROM中的数据。掩模型ROM适用于成批生产的定型产品，降低单片成本。
- （2）可编程只读存储器PROM（Programmable ROM）：一次可编程，一旦写入则内容不能再改变。PROM的基本原理是制造时在节点之间加入熔丝或者二极管，通过烧断熔丝或者击穿二极管完成编程，这一过程是不可逆转的，因此称为一次可编程。目前已经很少使用。



PROM

- （3）可擦除可编程的只读存储器EPROM（Erasable PROM）：可以多次改写，写入速度较慢，一般需要借助一些编程工具来完成擦写操作，使用时作为只读存储器来使用。常见的包括紫外光可擦除ROM（UVEPROM）和电可擦除ROM（Electrically EPROM，EEPROM或者 E^2PROM）。
- （4）闪存（Flash Memory）：其每个记忆单元都有一个“控制闸”和“浮动闸”，利用高电场改变“浮动闸”的临界电压可进行编程操作。其读速度与DRAM相当，但写速度慢10~100倍。

主要技术指标和参数-容量

- 以位（bit）为单位写入一张矩阵中。
- 指定一个行（Row），再指定一个列（Column），就可以准确地定位到某个Cell，这就是内存芯片寻址的基本原理。
- 这样的一个阵列就叫逻辑Bank（Logical Bank或者L_Bank）。内存中也不是只有一组逻辑Bank，它是由多个逻辑Bank组成的。
- 每个逻辑Bank的单元格位数称为数据深度（Data Depth），也叫位宽，即一次操作能同时读写的位数。内存芯片的容量就是所有内存的逻辑Bank中的存储单元的容量总和。

主要技术指标和参数-容量

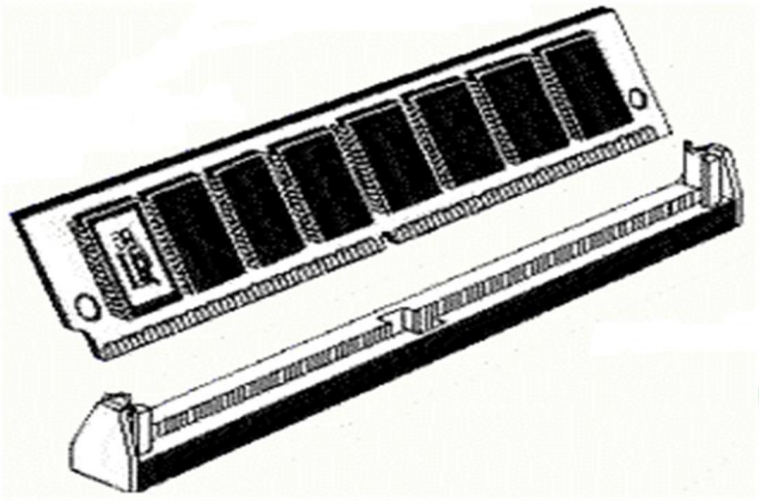
- 存储单元数量=行数×列数×数据深度×L-Bank的数量÷8
- 以128Mb内存芯片为列，一般内存芯片中的存储单元被平均分为4个L_Bank，由两个引脚来指定选中的Bank。根据不同的布局，内存芯片位宽有4位、8位、16位之分。如表6-1所示，内存芯片为128M位，即128×220位，分为4个Bank，芯片位宽等于4位，则每一个Bank有8M个存储单元，8M个存储单元按照行、列排列，有 212行和211列。向内存芯片输入地址时，首先输入12位行地址，再输入11位列地址。
- 内存的物理Bank与逻辑Bank则是两个完全不同的概念

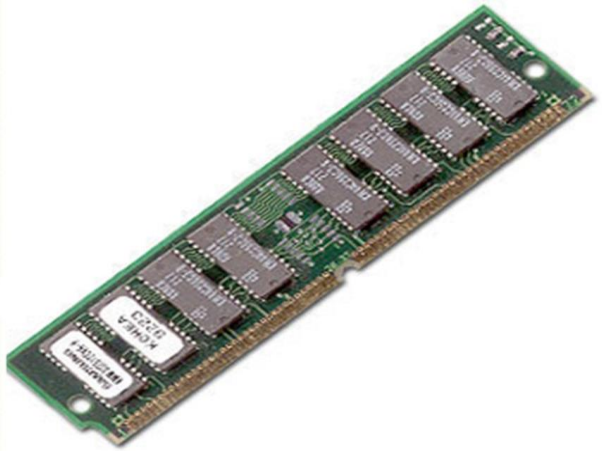
内存带宽

- 指内存的数据传输速度，是衡量内存的重要指标。
- 带宽=总线宽度×总线频率×一个时钟周期内交换的数据包个数
- 目前为64位。
- 已知总线频率，试计算如下内存带宽。
- PC100 SDRAM外频100MHz时，带宽=64×100/8=800MB/s
- PC133 SDRAM外频133MHz时，带宽=64×133/8=1064MB/s
- DDR DRAM外频100MHz时，带宽=64×100×2/8=1.6GB/s

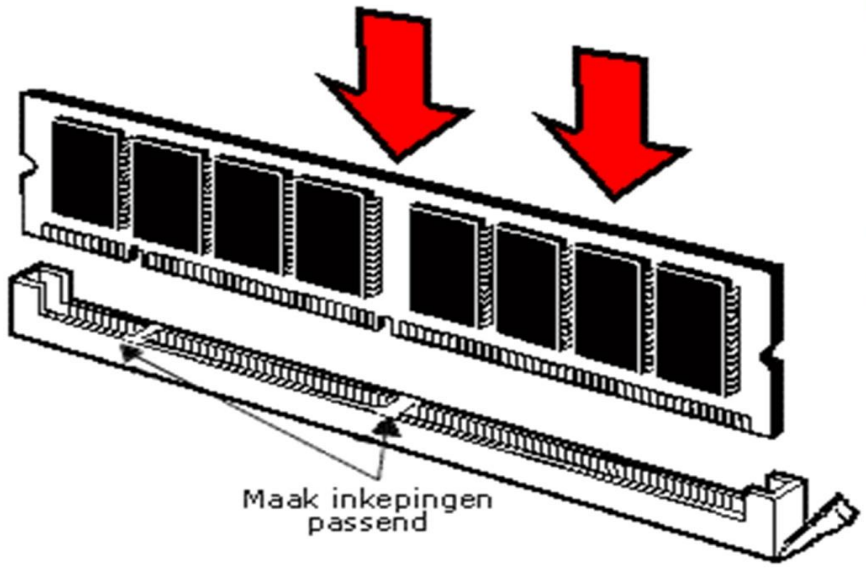
内存模组接口

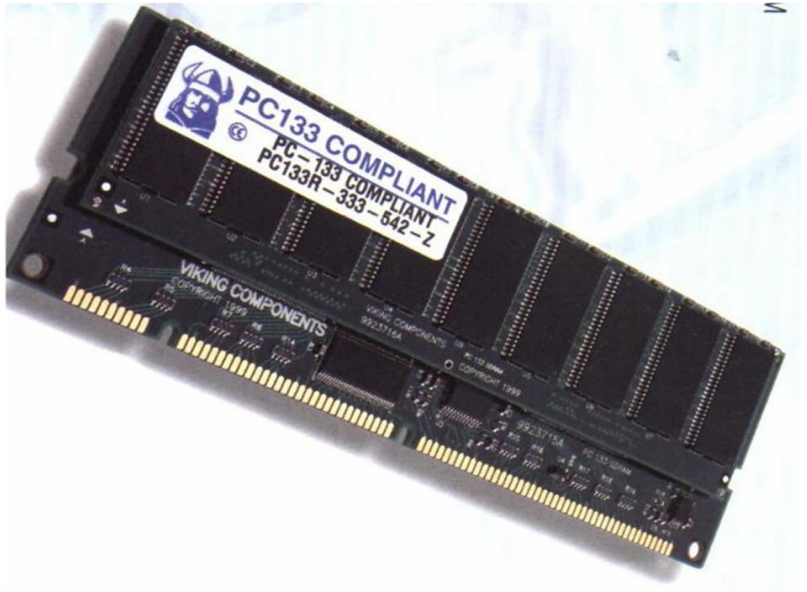
- SIMM（Single Inline Memory Module）





- DIMM（Double Inline Memory Module）

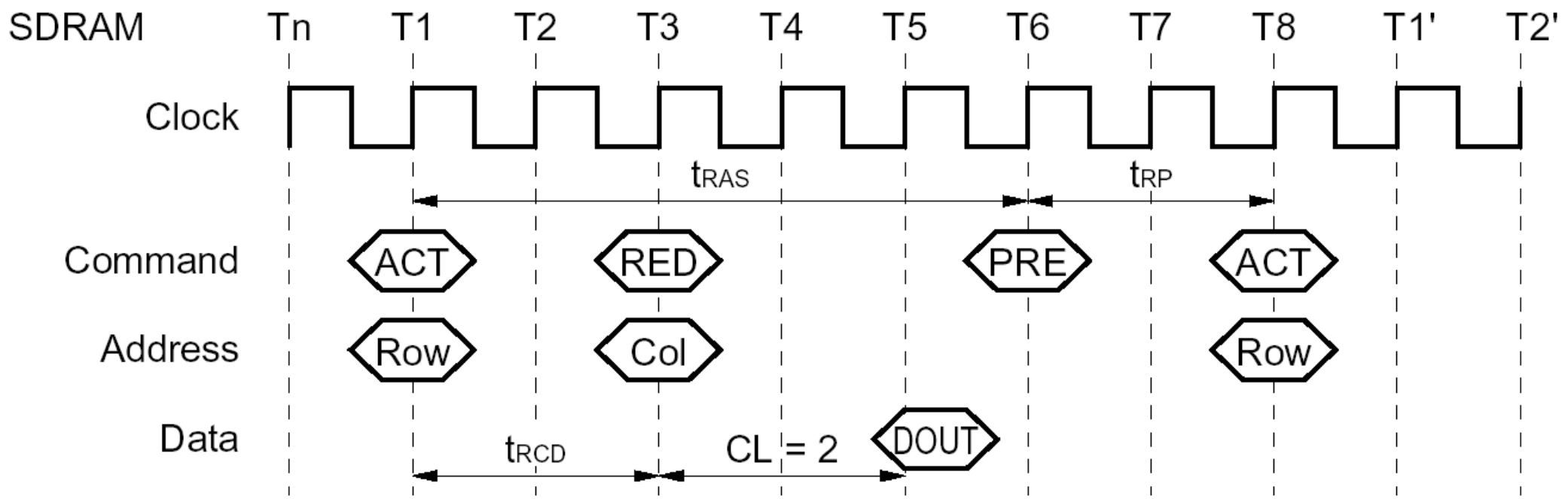




SDRAM 内存芯片管脚

管脚名	方向	作　用
CLK	输入	时钟输入。SDRAM 的所有信号均以系统时钟 CLK 上升沿为基准。
CKE	输入	时钟使能。CKE 为高电平时，下一个 CLK 上升沿有效，否则无效。
CS#	输入	片选。CS#为高电平时，开始命令输入周期。
RAS# CAS# WE#	输入	3 个信号组合构成一个命令。RAS#、CAS#、WE#的几种组合为： 101，读命令；100，写命令。
DQM, UDQM, LDQM	输入	DQM 为高电平时，若是读操作，不送出数据到 DQ 管脚；若是写操作，数据不写入存储单元。 内存芯片位宽为 16 位时，UDQM 控制 DQ8-15，LDQM 控制 DQ0-7。
A0-A11	输入	行地址、列地址复用。
BA0, BA1	输入	确定选取哪一个 Bank
DQ0-DQn	输 入 / 输出	读操作，内存单元的数据输出到 DQ 管脚。写操作，从 DQ 管脚取出数据写入内存单元。内存芯片位宽为 4、8、16 位时,使用 DQ0-3、DQ0-7、DQ0-15。
VCC	输入	电源
VSS	输入	地

读取一个单元的过程

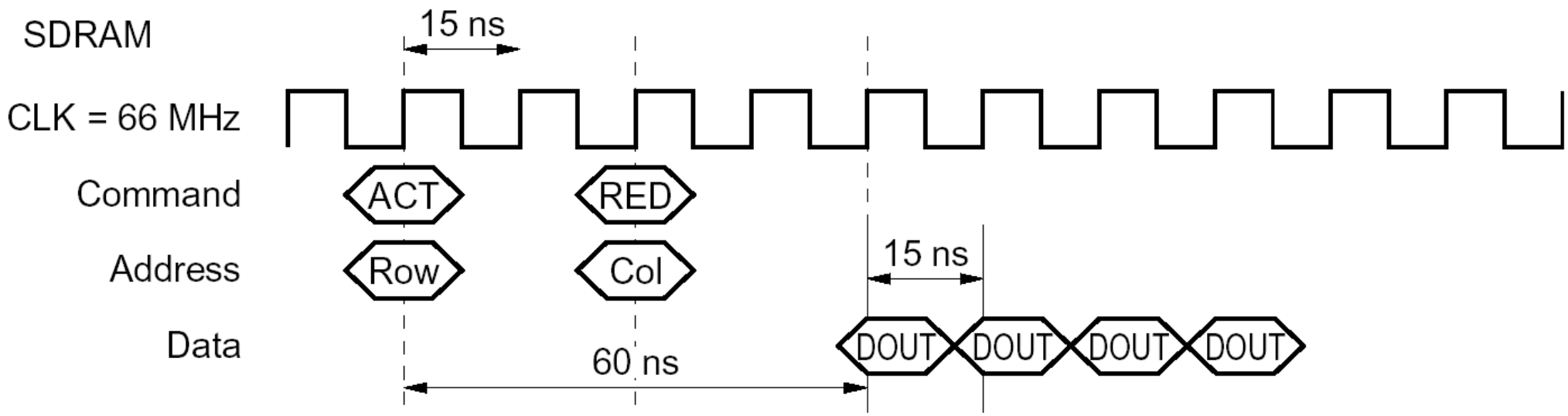

SDRAM
Tn
T1
T2
T3
T4
T5
T6
T7
T8
T1'
T2'
Clock
tRAS
tRP
Command
ACT
RED
PRE
ACT
Address
Row
Col
Row
Data
tRCD
CL = 2
DOUT

- 在T1的上升沿，CS#为低电平，RAS#、CAS#、WE#等于0、1、1，即发出ACT命令（Active），A0~A11输入行地址。经过2个时钟后，在T3的上升沿，RAS#、CAS#、WE#等于1、0、1，发出RED命令（Read），A0~A11此时输入列地址。ACT命令和RED命令之间的时间间隔称为t_{RCD}（RAS to CAS Delay）选通周期，表示行地址至列地址延迟时间
- 在发出RED读命令后，到数据输出到I/O管脚上DOUT这段时间称作CL（CAS Latency），即读取潜伏周期，该参数只在读时序中有效

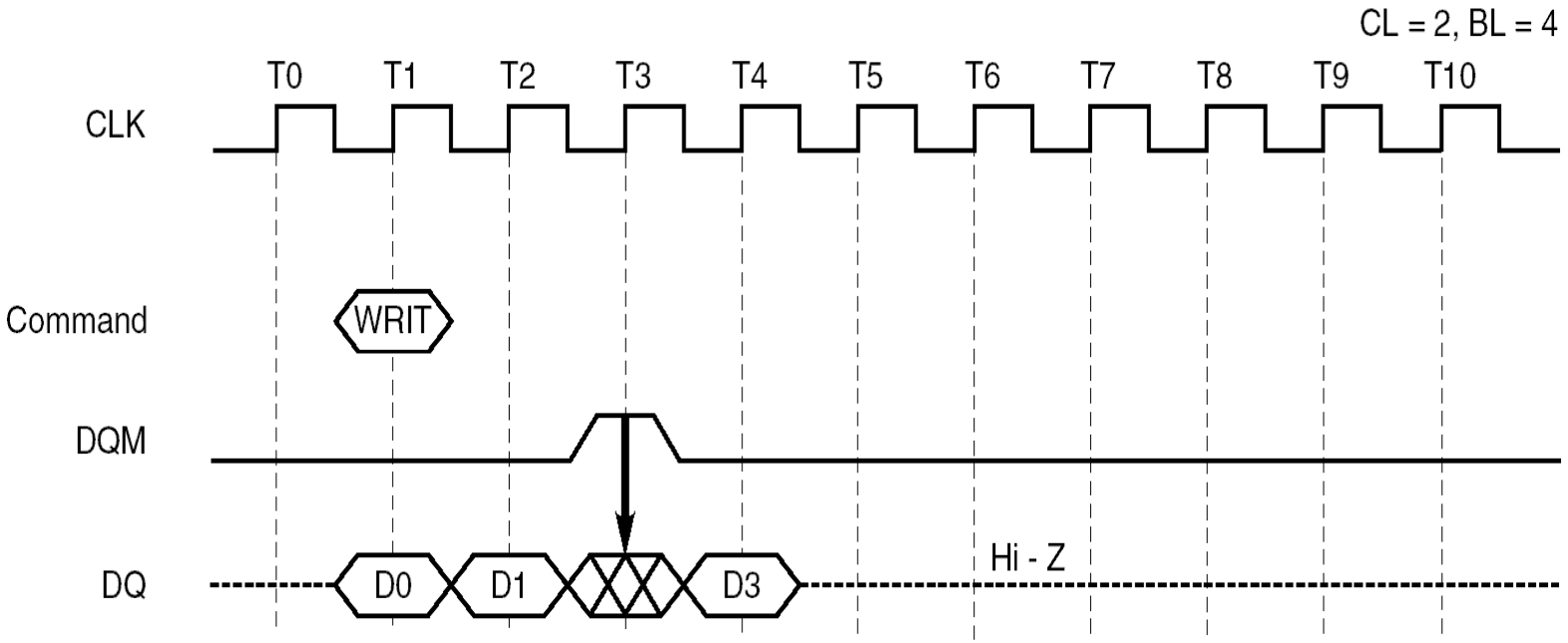
- 在T5之后，对同一个Bank再次读写时，如果行地址发生了变化，就必须要求芯片先执行预充电（Precharge）操作，然后才能读写另外一行中的存储单元。
- t_{RAS}（Active to Precharge Command）是内存行有效至预充电的最短周期，规定了有效（ACT命令）至预充电命令之间的最短间隔，过了这个周期后才可以向芯片发出预充电指令。

突发（Burst）

- 同一行中相邻的存储单元连续进行数据传输的方式，连续传输的存储单元的数量就是突发长度（Burst Lengths，BL）。BL值可以使是 1、2、4、8、全页（Full Page，P-Bank所包含的每个芯片内同一L-Bank中同一行的所有存储单元）。
- 在突发传输方式下，只需要输入第1个存储单元的行、列地址，不需要提供后面几个存储单元的行、列地址。SDRAM之所以被称作同步DRAM，就是指在突发传输方式下，每一个时钟周期都可以读写一个存储单元的数据，即数据的访问与时钟周期同步。


SDRAM
15 ns
CLK = 66 MHz
Command
ACT
RED
Address
Row
Col
15 ns
Data
DOUT
DOUT
DOUT
DOUT
60 ns

- 为了屏蔽不需要的数据，SDRAM中还采用了数据掩码（Data I/O Mask，DQM）技术。通过DQM内存可以控制突发传输中是否屏蔽某一个存储单元的读写动作。传统的DQM由北桥控制，每个信号针对一个字节。SDRAM官方规定，在读取时DQM发出两个时钟周期后生效；而在写入时，DQM立即生效，因此如果不需要读出某一个存储单元，那么输出该单元的前2个时钟周期将DQM设为高电平。


CL = 2, BL = 4
T0
T1
T2
T3
T4
T5
T6
T7
T8
T9
T10
CLK
Command
WRIT
DQM
DQ
D0
D1
D3
Hi - Z

DDR

- SDRAM在时钟的上升沿进行数据传输，一个时钟周期内只传输一次数据；而DDR内存则是一个时钟周期内传输两次数据，它能够在时钟的上升沿和下降沿各传输一次数据，因此称为双倍速率同步动态随机存储器（Double Date Rate SDRAM，DDR）。

DDR2

- DDR2（Double Data Rate 2 SDRAM）与DDR内存技术都是在系统时钟的上升/下降沿进行数据传输，但DDR2内存的DQS采用差分信号，预读取能力更高。DDR2内存每个时钟能够以4倍外部总线的速度读/写数据，并且能够以内部控制总线4倍的速度运行。

DDR3

- 也是采用差分信号，在系统时钟的上升/下降沿进行数据传输。和DDR2相比，DDR3（Double Data Rate 3 SDRAM）的工作电压更低，只有1.5V。DDR3内核的频率只有接口频率的1/8，即DDR3-800的核心工作频率只有100MHz。

DDR4

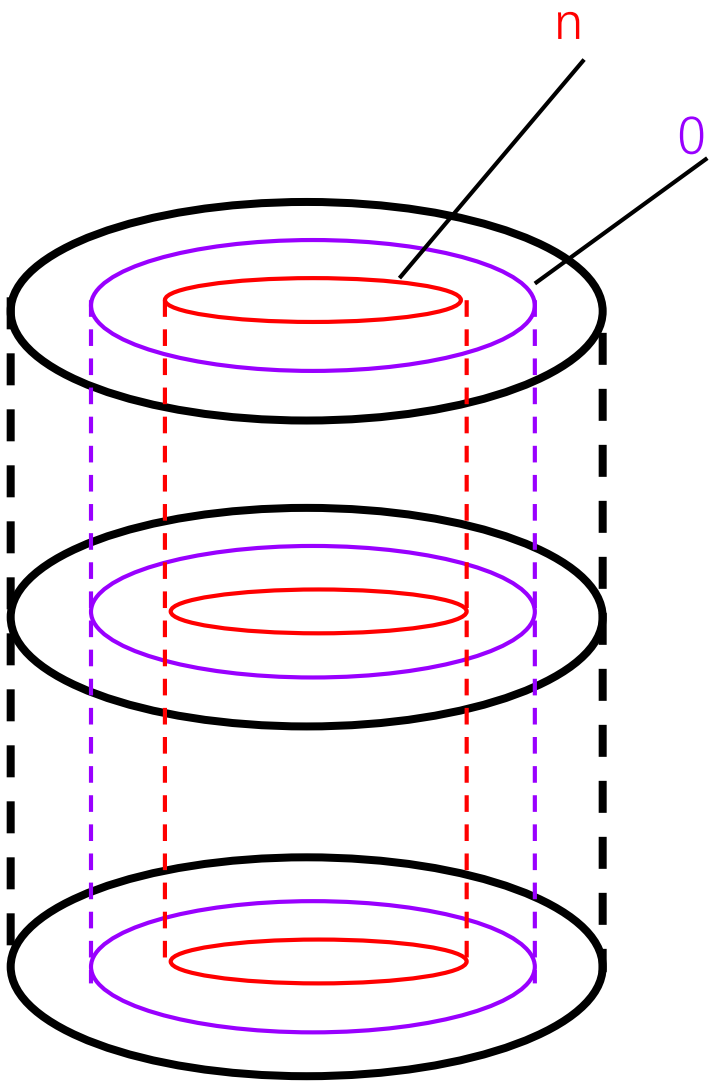
- DDR4相比DDR3最大的区别有三点：16bit预取机制（DDR3为8bit），同样内核频率下理论速度是DDR3的两倍；更可靠的传输规范，数据可靠性进一步提升；工作电压降为1.2V，更节能。此外，DDR4还增加了DBI（Data Bus Inversion）、CRC（Cyclic Redundancy Check）等功能。

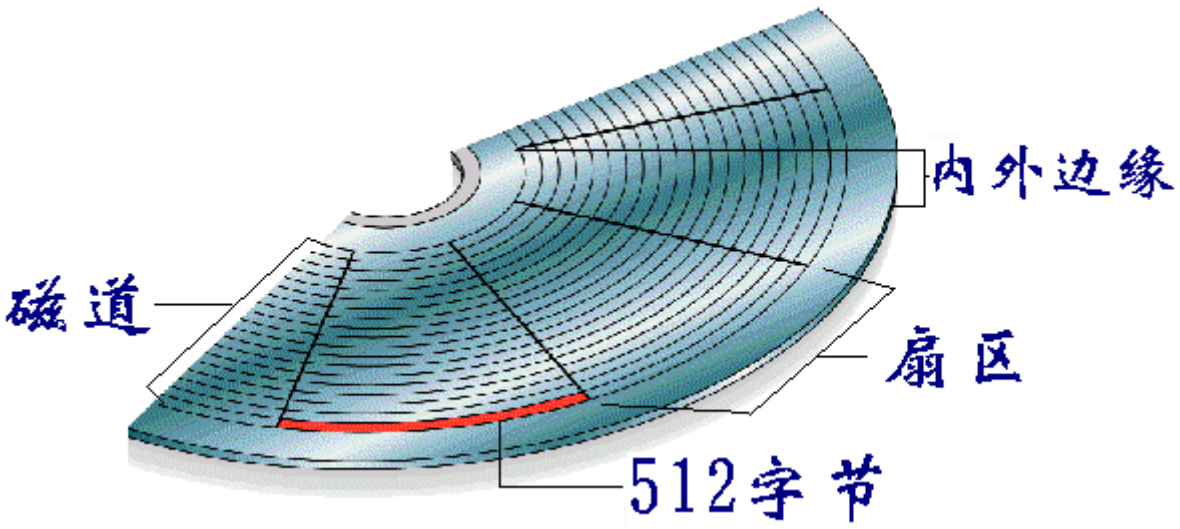
辅助存储器-硬盘

- 传统机械硬盘（Hard Disk Drive，HDD）、固态硬盘（Solid-State Disk，SSD）和混合硬盘（Hybrid Hard Disk，HHD）。
- 硬盘由硬盘片、硬盘驱动电机和读写磁头等组装并封装在一起，称为温彻斯特盘（Wenchester）。
- 1984年，IBM在AT机中引入了新的驱动和控制一体的硬盘驱动器，称为AT Attachment（ATA）。随后在1986年，Compaq在Deskpro 386 中引入硬盘驱动器，称IDE（Integrated Drive Electronics）。IDE接口可以连接硬盘和光驱，连接成本低、兼容性好、容易安装使用，很长一段时间是最为普及的磁盘接口，目前已被SATA接口所取代。

HDD工作原理

- 包含一组磁盘片，每个磁盘片有两个磁盘面，磁盘面按照上到下的顺序从0依次编号，每个盘面均有一个磁头与之对应，有几个磁盘面就有几个磁头。磁盘面上的磁道按外到里的顺序从0依次编号，最外侧为0磁道。磁道按圆弧段分为扇区，扇区字节数为固定大小的值，数值大小必须得到操作系统的支持，格式化扇区的典型尺寸为512B、1024B和4096B。
- 扇区是最小的读写单位，每个扇区中的数据作为一个单元同时读出或写入，扇区从1开始编号。0磁道是硬盘上非常重要的位置。硬盘的主引导记录区（Main Boot Record，MBR）就保存在0磁头0柱面1扇区


n
0


内外边缘
磁道
扇区
512字节

HDD主要技术指标

- 容量（Volume）
- 转速（Rotational Speed）
- 时间参数
- 缓存
- 硬盘的数据传输率
- 接口类型

ATA总线

扇区的编址模式

- 硬盘的结构由一组共轴的盘片组成，每个盘片都有两个面，每个面都有对应一个读写磁头。即N个盘片的硬盘有2N个面，对应2N个磁头（Heads），从0、1、2…nH开始编号；每个盘片上相同编号的磁道形成柱面（Cylinders），有多少个磁道就有多少个柱面，磁道号就是柱面号，从外至里编号为0、1、2……nC；每个磁道又被分为若干个扇区（Sector），扇区大小通常固定编号从1开始编，为1、2、3……nS；形成nC×nH×nS个扇区，0柱面0磁头1扇区是整个硬盘的第1个扇区。
- CHS（Cylinder/Head/Sector）寻址。

- 用4个二进制位表示磁头号，用10个二进制位表示磁道号，用6个二进制位表示扇区号。因此，最多支持2^4=16个磁头，2^{10}=1024个磁道，每个磁道2^6-1=63个扇区。因为每个扇区512字节，所以支持最大容量为：1024×16×63×512=528482304字节=504MB，这里1MB=2^{20}字。有的地方认为1MB=10^6B
- 采用LBA（Logical Block Addressing）逻辑块寻址模式。LBA编址方式中，扇区的地址就是这个扇区的序号。LBA的扇区编号从0开始，如0柱面0磁头1扇区的序号为0。编址的二进制位数有2种：28位地址和48位地址。

LBA编址模式和CHS编址模式之间可以相互转换

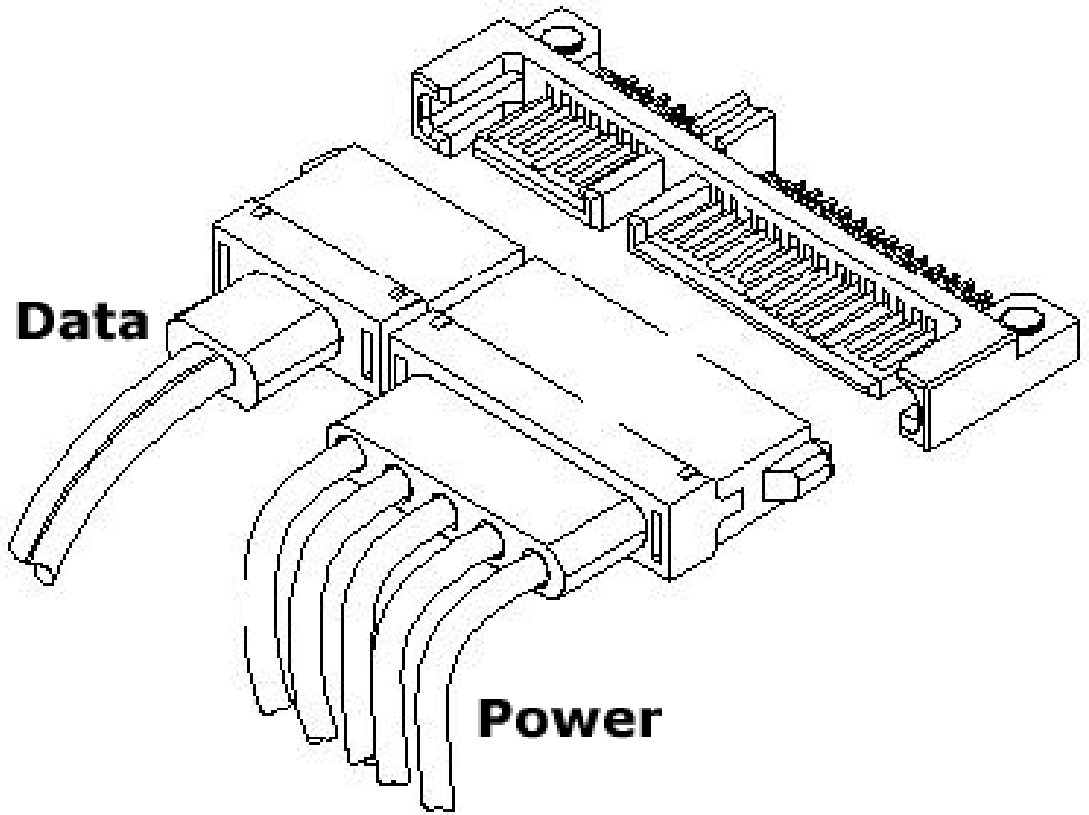
- 假定硬盘的磁头数为nH、磁道数为nC、每个磁道的扇区数为nS，则硬盘的可用扇区总数为nH×nC×nS。设一个扇区在LBA编址模式中的地址为L，在CHS编址模式的地址为<C,H,S>，0≤C≤nC-1，0≤H≤nH-1，1≤S≤nS。则L =（（C×nH + H）×nS）+S–1。
- Hs=head start=0; Cs=Cylinder start=0; Ss=sector start=1
- L=(C-Cs)*nS*nH+(H-Hs)*nS+(S-Ss)
- =C*nS*nH+H*nS+(S-1)
- =(C*nH+H)*nS+S-1

- 同样，根据L也可以求得<C,H,S>：
- S =（L % nS）+ 1
- H =（L ÷ nS）% nH
- C =（L ÷ nS）÷ nH
- 其中，÷是整数除法，%是取模操作。

硬盘读写方式

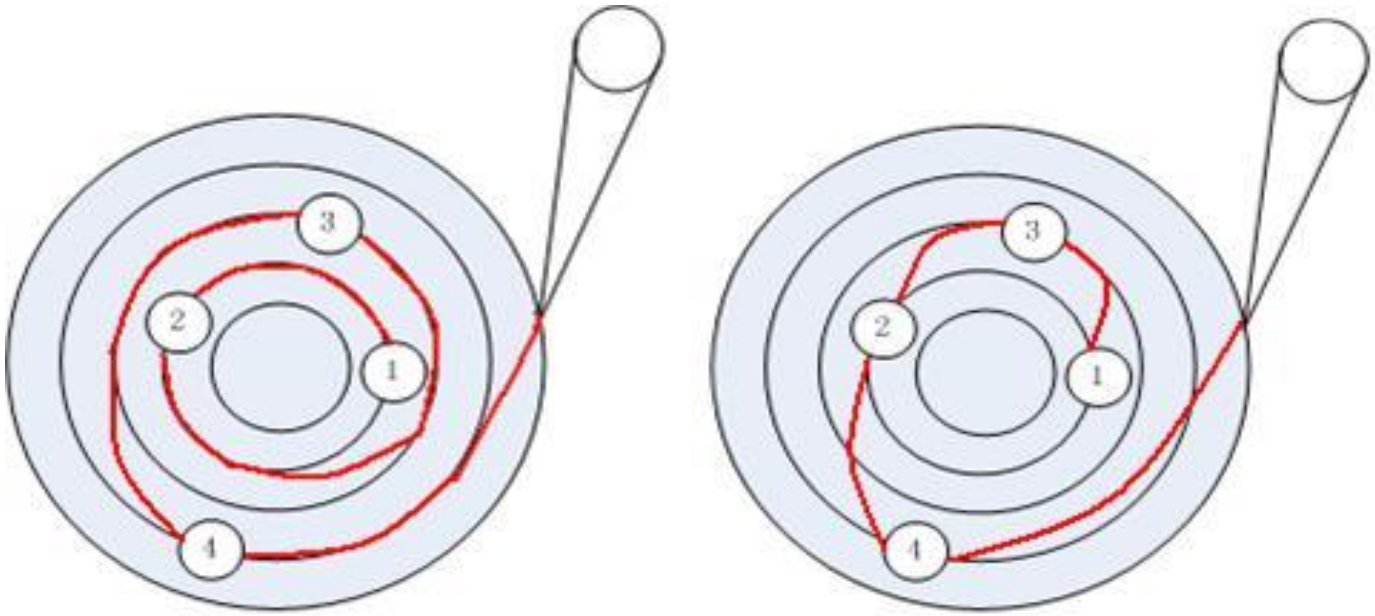
- PIO方式读写硬盘
- DMA方式读取硬盘

SATA接口和技术指标



NCQ技术

- 对数据请求进行优化排序，尽可能少移动磁头



a）不支持NCQ　　b）支持NCQ

固态硬盘

- 固态硬盘的接口规范和定义、功能及使用方法上与普通硬盘的相同，在产品外形和尺寸上也与普通硬盘一致。
- 固态硬盘没有磁头，结构较为简单，在一块PCB上，由固态存储单元、主控芯片、传输接口以及一些小元件所组成。

闪存芯片

- 非易失性存储器NVM（Non-Volatile Memory）
- 先擦后写
- FLASH的写操作只能将数据位从1写成0，不能从0写成1，所以在对存储器进行写入之前必须先执行擦除操作
- 擦除操作的最小单位是一个区块，而不是单个字节或单个扇区。

- 读写接口
  - 在物理结构上分成若干个区块，区块之间相互独立。一个区块包括多个块。每个块的大小是512或2048字节。
- 擦除
  - 向NAND FLASH发出一个“擦除”操作命令，可以擦除整个块的内容。
- 操作指令
  - 必须输入一串特殊的指令（NOR FLASH），或者完成一段时序（NAND FLASH）才能将数据写入到FLASH中。