

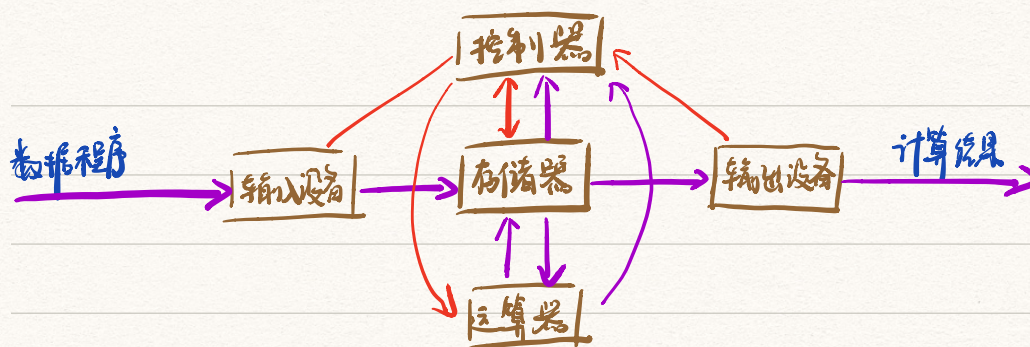
存储器是计算机系统中的记忆设备，用来存放程序和数据。

存储器分为两类：

- 主存 (Memory, 也称为内存)：本意为“记忆装置”。多指存储器的整体（包括：记录介质，有关电路和其他部件）
- 辅存 (Storage, 也称为外存)：本意为“仓库”。多指记录介质本身（包括：磁盘、固态盘、磁带、存储阵列 (bank) 等

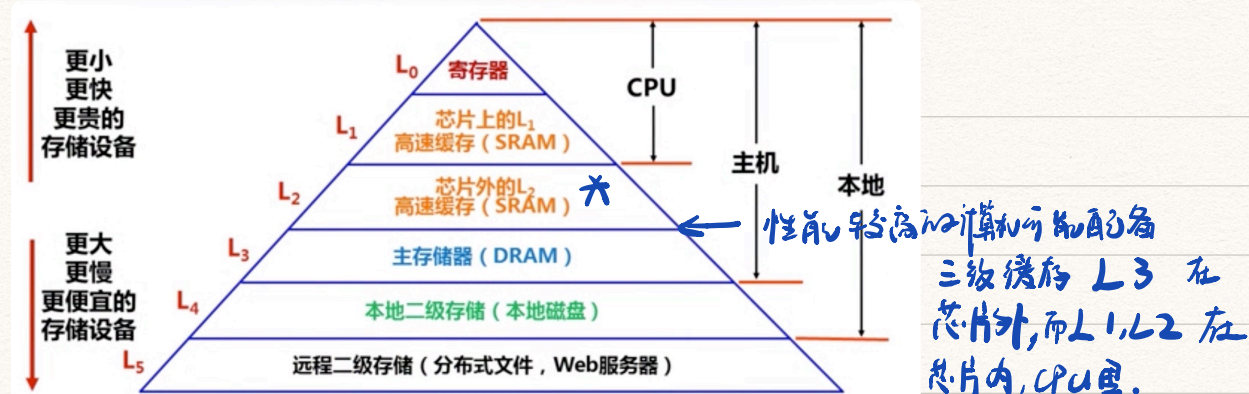
主存中存放当前处于活跃状态的程序和数据。

CPU可直接与主存进行数据的交换，而辅存中的数据需要先传递到主存，或者经过主存，才能与CPU进行数据交换。



存储器向CPU提供数据和指令，控制输入与输出设备读写。存储系统被当作线性的字节数组，而CPU可以在一个常数时间内访问到存储器的每一个位置。

实际上，存储器系统是一个拥有不同容量、成本和访问时间的存储设备层次结构 (Hierarchy)



存储器层次结构：每一层中都存放着来自较低一层的一部分数据，在存储层次中，采用多层、不同类型的存储元件的目的是为了弥合顶层器件和底层器件之间巨大的容量差距和速度差距。

如果数据存放在CPU中，在一个时钟周期内就能访问到这些数据。

如果存储在Cache中，需要1-30个时钟周期。

如果存储在主存中，需要50-200个时钟周期。

而磁盘上的数据，需要约几千万个周期。



## 基本术语

- 记忆单元/存储位元/位元 (Cell)  
具有两种稳态的能够表示二进制数0和1的物理器件的最小单元，一个位元存储1bit数据
- 存储单元/编址单位 (Addressing Unit)  
存储器中具有相同地址的多个位构成一个基本存储单元。即存储器的最小可寻址单位。在主存中，一个存储单元通常为一个字节，8bit。每个存储单元有唯一的存储器地址。
- 存储体/存储矩阵/存储阵列 (Bank)  
所有存储单元构成一个存储阵列。

cell 组成 addressing unit, addressing unit 组成 bank

- 存储器地址寄存器 (Memory Address Register-MAR)  
用来存放主存单元地址的寄存器。
- 存储器数据寄存器 (Memory Data Register - MDR)  
用于存放主存单元中数据的寄存器。
- 机器字长 (machine word)  
运算器中参加运算的寄存器的位数。即，CPU的数据通路的总线宽度 (width of datapath)
- 存储字 (memory word)  
存储芯片中一个读写单位，一般等于存储器的数据线宽度。（存储字和机器字长宽度可以一样也可以不同，通常是整数倍的关系）
- 编址方式 (addressing mode)  
对存储体中各存储单元进行编号的方法  
现代计算机基本是按字节编址，早期计算机中，有的是按字编址
- 传输单位  
一次读写存储器的数据单位  
对主存而言，指一次从主存读出或写入的数据位数； (bit)  
对辅存而言，数据通常按块传输，传输单位为块（如512个bit为单位）

用实际计算机举例，intel386机器，编址单位为字节，机器字长为32位，单字位数16位，但传输单位可以是8/16/24/32位

## 存储器的分类标准:

### 1) 在计算机中作用

- 寄存器型存储器：在速度上与CPU匹配，由寄存器构成，封装在CPU内。用于存放当前正在执行的指令和使用的数据。用触发器实现，速度快，容量小，以KB为单位。
- 高速缓冲存储器 (Cache)：位于CPU内部或位于主存和CPU之间。用来存放当前正在执行的局部程序段和数据。用SRAM实现，速度级别为几纳秒ns，容量单位MB。
- 主存储器 (primary memory)：位于CPU之外，用来存放已被启动的程序及所用数据。用DRAM实现，速度级别几十纳秒ns，容量GB。



- 辅助存储器（secondary memory）：不能由CPU的指令直接访问，用来存放暂不运行的程序、数据或存档文件。用磁、光等存储器件实现，容量大，但速度慢。毫秒级ms。
- 其它功能存储器：
  - 控存CM：存储微程序代码
  - 表格存储器：为加快CPU处理而设置（倒数表、函数表）
  - 字库和数据缓冲存储器：显示和印刷输出设备中

## 2) 存储介质

- 半导体存储器（SCM）：速度快，用作内存。记忆原理：双稳态触发器、电容（静态、动态）
- 磁表面存储器（MSM）：用陶瓷、非磁性金属或塑料作载磁体，磁化后具有两种不同的剩磁状态记录信息“1”和“0”；非易失（non-volatile），容量大并且每位价格低。用作外存
- 光盘存储器（ODM）：有机玻璃作为磁体，利用磁化、晶态/非晶态表示信息。非易失，可靠性高，保存时间长，容量大且易于更换。存储速度比硬盘低一个数量级。
- 铁电存储器FeM（FRAM）、相变存储器（RCM）、阻变存储器（ReRAM）……：掉电数据不丢失，速度快的新存储设备

## 3) 存储方式

- 随机访问存储器（RAM）：存储器任意单元可随时访问且访问所需时间相同。
  - 访问时间与存储单元所处的物理位置无关。
  - 速度快（ns），用于主存和Cache
- 只读存储器（ROM）：正常工作时只读，能随机读出，不能随机写入
  - MROM：只读
  - PROM：一次写
  - 可多次改写的ROM：EPROM, E2PROM
- 相联存储器（CAM）：按内容检索到存储位置进行读写
  - 速度快（ns），价格高。
- 直接存取存储器（DAS）：介于随机和顺序存取之间
  - 可以直接定位到要读写的数据块，存取时间的长短与数据所在位置有关（与磁头的距离）。
  - 速度慢（ms），如 磁盘。
- 顺序存储器（SAS）：存储时以数据块为单位存储，顺序地记录在存储介质上。
  - 数据按顺序从存储载体的始端读出或写入，存取时间的长短与数据所在位置有关。
  - 速度慢、容量大、成本低，用作后援外存
  - 磁带、电荷耦合器件

## 4) 存储器中信息的可保存性

- （断电后是否丢失数据）
- 挥发性存储器（易失性存储器）Volatile
  - 特点：断电后，信息丢失



SRAM, DRAM

- 非挥发性存储器（非易失性/永久性存储器） Non-volatile

特点：断电后，信息不丢失

ROM、磁盘、闪存

（读出后是否保持数据）

- 破坏性读出存储器 destructive readout memory (dro memory)

特点：读出时原存信息被破坏，需重写

DRAM

- 非破坏性读出存储器 non-destructive readout memory

特点：读出时，原存信息不被破坏

SRAM

### 内存的主要技术指标

#### 1) 存储器的容量大小：存储容量

内存所能容纳的二进制位（bit）个数的总和，即构成内存的存储位元的总和。

存储容量 = 存储单元个数 \* 每个存储单元的存储位元个数

一般内存的存储容量越大越好。

常用单位：B (byte) , KB ( $2^{10}$ ) , MB( $2^{20}$ ), GB( $2^{30}$ ), TB( $2^{40}$ )

#### 2) 存储器的速度快慢：

存取时间 $T_a$  (store access time)：启动一次存储器操作到完成该操作所用的时间。

如，一次读命令，从启动读命令到读出的数据送到MDR所需的时间。

写，从启动到将MDR内容写入指定存储单元的时间

存取时间决定了CPU发出读写命令后必须等待的时间。

存储周期 $T_s$  (store access cycle)：连续两次启动同一存储器进行存取操作所需的最小时间间隔，也称为存取周期或访问周期。连续两次之间有预充电时间，所以比单独两个不同存储器存取时间要短，但比一次时间长。

两次启动包括：两次读，两次写，一次读或一次写。

$T_a$ 和 $T_s$ 之间的时间差依赖于存储信息的器件和电路，存储介质和控制线路需要恢复时间。

破坏性读出须重写，非破坏性读出不须重写。

#### 3) 传输性能：存储带宽和存取宽度

存取宽度：一次访存操作可存取的数据位数或字节数，存取宽度由编址方式决定

存储带宽 (memory bandwidth)：每秒传输的最大数据量 (位/秒)

改善计算机瓶颈的一个关键因素

e.g.存取周期为 500ns，每个存取周期可访问 16位，则带宽为多少？



一秒内的存取周期 =  $1/500\text{ns} = 2 \times 10^6$

$16\text{位} \times (1/500\text{ns}) = 32\text{M位/秒}$

#### 4) 可靠程度: MTBF

用Mean Time Between Failure 来衡量主存的可靠性。

MTBF表示两次故障之间的平均时间间隔。

采用纠错编码技术进行容错处理。

#### 5) 功耗与集成度

存储带宽功耗 (Power Loss) 反映存储器件耗电多少 (单位: mw/片, w/存储器)

维持功耗: 保持时功耗 (几百-几十mw/片)

工作功耗: 读写时功耗 (500-1500mw/片)

集成度 (Integration Level): 标识单个存储芯片的存储容量

一般希望存储器功耗低集成度高, 但两者之间是互相矛盾的。

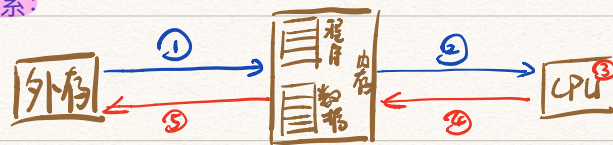
#### 6) 综合性能评价

性能价格比 (performance cost ratio):

更高的容量、速度, 更低的位价格, 性能价格比越高越好。

通常采用存储器层次结构和技术才能实现。

#### 内存与外存的关系:



1.任务启动, 执行任务的程序和数据将从外存成批传送到内存, 分别存在程序段和数据段。

2.CPU从内存中逐条读取该程序的指令及其数据。

3.CPU逐条执行指令, 按指令要求完成对数据的运算和处理。

4.CPU将指令的运算处理结果送回内存保存。

5.任务完成, 将处理得到的全部结果成批传送到外存长久保存。

#### 内存与外存的比较:

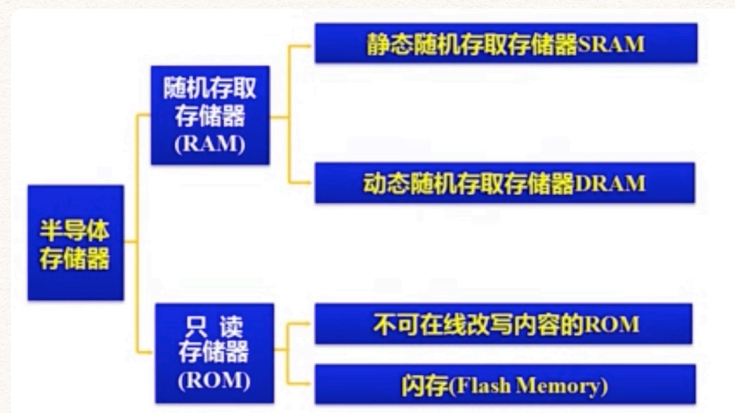
外存储器: 存取速度慢。成本低、容量大。CPU不直接访问。非易失性 (non-volatile)

内存储器: 存取速度快。成本高、容量相对较小。CPU直接访问。易失性 (volatile)



## 随机存取存储器 (RAM)

内存由半导体存储器芯片组成，芯片由多种类型。



SRAM用于构造Cache。

只要加上电源，信息就能一直保持。

对电器干扰相对不敏感

比DRAM更快，也更贵

DRAM用于构造主存

每隔一段时间必须刷新一次

对电器干扰比较敏感

比SRAM慢，但便宜

只读ROM一般用于BIOS

闪存用作主存扩展、U盘等。

## MOS型RAM基本存储位元

### 1. 记忆原理

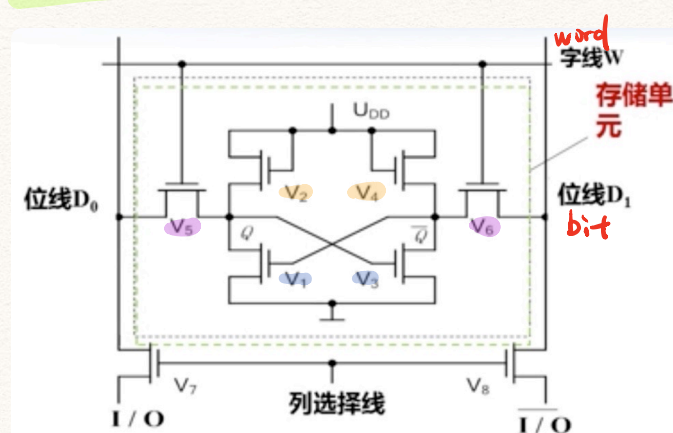
触发器：互补的两个状态（六管静态MOS电路）（mos相关内容可在逻辑门二中查看）

电容：充放电（单管动态MOS电路）

### 2. 基本存储位元与存储器

存储位元-》存储单元-》存储矩阵-》存储芯片（译码、驱动、读写电路）-》存储模块（内存条）-》存储器

## 六管静态MOS电路 (SRAM)



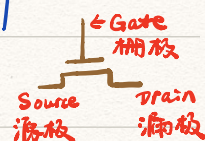
SRAM数据保存在触发器中，只要供电，数据就一直保持，不是破坏性读出，也不需要重写数据来保持数据不变。即：无需刷新！

$V_1, V_3 \rightarrow$  触发器

$V_2, V_4 \rightarrow$  触发器负载管

$V_5, V_6 \rightarrow$  门控管

6个MOS管。



使用6管存储一位数据。

工作原理可看作是带时钟的RS触发器。

写入时：分别置字线、列选择线位高电平，栅极导通。

位线上是被写入的I/O端信息“0”或“1”。

存储单元按位线的状态被置为“0”或“1”。

假设V3导通，则V1截止，此时Q点为高电

平，如此时状态是“1”，反之的状态就为“0”。

当w接入低电平时，V5、V6截止，触发器与外界隔离，以保持原有信息不变。

如果要读或写SRAM时，给字线W加高电平，使控制开关V5和V6导通，使基本单元与位线连通，如果要写入“1”，在位线D0接高电压，D1接低电压，使Q非这点电位下降，V1管截止。Q点电位上升，V3导通，完成写“1”的操作



读出时：分别置字线、列选择线为高电平，栅极导通。  
根据存储单元的状态改变位线的输出电平，读出到I/O端。

字线接高电平，V5和V6开启，如原先存的是“0”，Q非点会是高电平，根据位线上有无脉冲，可以区分读出的是“0”还是“1”。

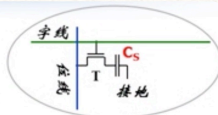
SRAM的数据不会因为读出操作发生改变。所以不需要进行刷新操作。

### 动态单管MOS电路（DRAM）：

#### 存储位元的基本原理

##### 构造和表示

- 数据记忆在电容 $C_s$ 上，T为门控管，控制数据进出
- 栅极接读/写选择线(字线)，漏和源分别接数据线(位线)和记忆电容 $C_s$
- 存储数据“1”或“0”：根据电容 $C_s$ 上电荷量的有无来判别



$C_s$ 放电后不能再被读出原数据，所以需要再次充电。这种单管mos电路是破坏性读出存储。

##### 读写原理：在选择(字)线上加高电平，使T管导通

- 写“0”时，位线上加低电平，使 $C_s$ 上电荷对数据线放电
- 写“1”时，位线上加高电平，使数据线对 $C_s$ 充电
- 读出时，根据位线（数据线）上是否有电流，区分读出的是“1”还是“0”

单管mos的优点是：电路元件少，工作功耗小，集成度高。

因此广泛用于大容量主存储器中。

缺点：速度慢，破坏性读出，电容电荷会泄漏，通常只能维持2ms左右，所以要定期在2ms内刷新（读出后重新写回）。

存储器芯片：

存储体+外围电路（地址译码和读写控制）

下期预告：

DRAM的寻址方式

主存

Cache