



计算机组成原理

第四章 存储子系统



信息与软件工程学院
School of Information and Software Engineering

主要 内 容

- 1 概述
- 2 半导体存储器
- 3 主存储器组织
- 4 磁表面存储器





- 01. 存储器的层次结构
- 02. 存储器的分类
- 03. 存取方式



计算机的重要特点之一是具有存储能力，这是它能够自动连续执行程序，进行广泛信息处理的重要基础。

本章需解决的主要问题：

- (1) 存储器如何存储信息？
- (2) 在实际应用中如何用存储芯片组成具有一定容量的存储器？

存储器是用来存放大量程序与数据的计算机部件。



一、存储器的层次结构

存储系统特别是**主存储器与CPU**之间有大量的信息交换操作，因此对存储器最基本的要求有三点：

存储容量大、存取速度快、成本价格低

计算机的工作表现为读取与执行指令：

1. 读取指令是在**存储器**中完成的，
2. 执行指令是在**CPU**内完成，

但因两者**速度不匹配**，造成了所谓的“瓶颈”问题。



一、存储器的层次结构

例：某计算机主频2GHZ

CPU内的一次处理时间：

$$2\text{GHz} = 1 / (2 \times 10^9) = 5 \times 10^{-10} (\text{s})$$

访存时间：

$$5\text{ns} = 5 \times 10^{-9} (\text{s})$$

CPU处理时间与访存时间相差一个数量级，这就是计算机“瓶颈”问题。



一、存储器的层次结构

解决“瓶颈”的方法：

- 1) 努力改进制造工艺，寻求新的存储机理，以提高存储器的性能；
- 2) 采用**存储器分层结构**来满足计算机系统对存储器不同的要求。

典型的**三级存储体系结构**，分为“**高速缓冲存储器-主存-外存**”三个层次。

一、存储器的层次结构

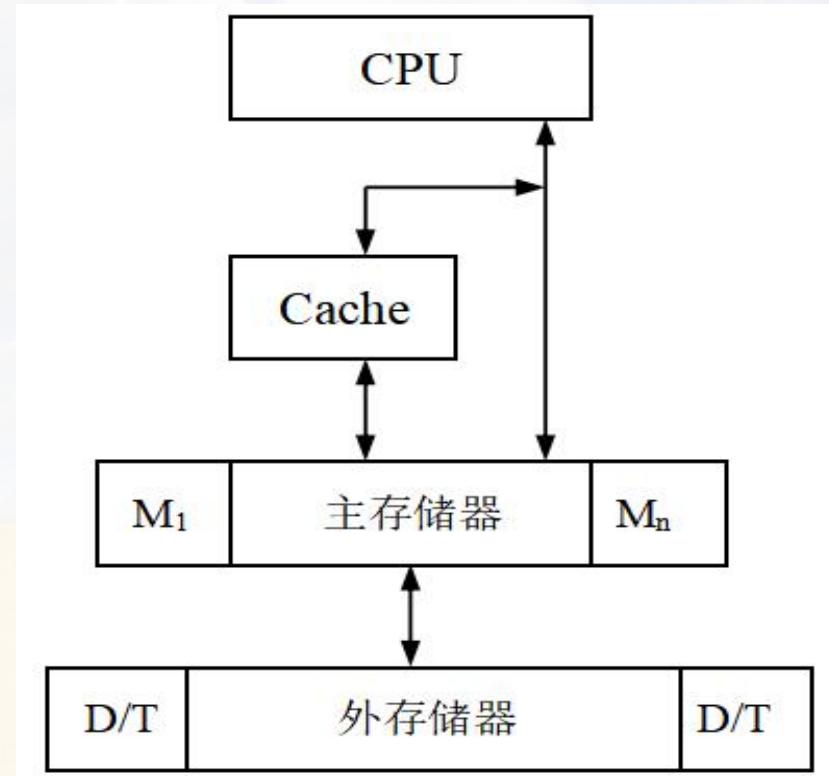
1、主存储器

能由CPU直接编程访问的存储器，它存放**当前CPU需要执行的程序与需要处理的数据。**

特点：存取速度快，容量有限

为满足CPU编程直接访问的需要，对主存储器的基本要求有三条：

- (1) 随机访问
- (2) 工作速度快
- (3) 具有一定的存储容量



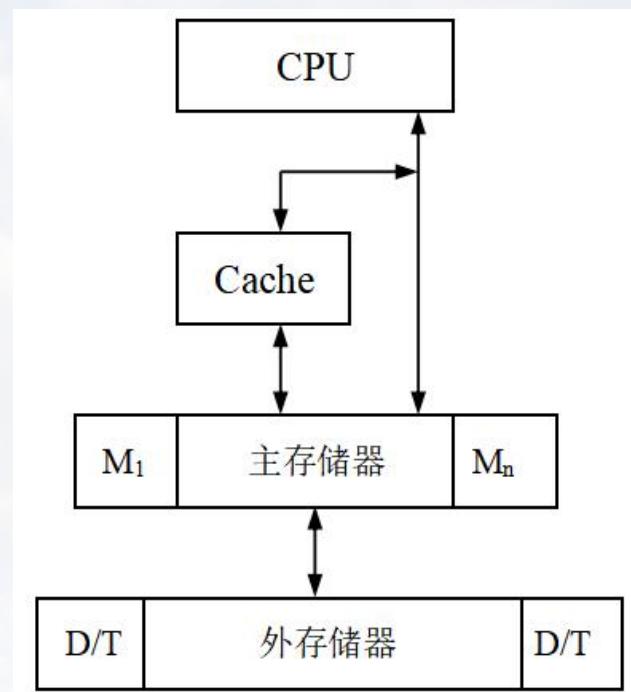
一、存储器的层次结构

2、外存储器

用来存放需要联机保存但暂不使用的大量程序与数据。

特点：存取速度较慢，容量很大

程序与数据只有进入主存才能真正运行，而外存储器是作为后援的。



3、高速缓存：高速缓存中存放的是最近经常使用的程序与数据，作为主存中当前活跃信息的副本。

特点：存取速度很快，容量很小



一、存储器的层次结构

4、层次结构

1) 主存-Cache层次 需解决：

① 地址映射

a. 直接映射：

主存的每一块只能映射到Cache的固定块中。

b. 全相联映射

主存的每一块可以映射到Cache的任一块。

c. 组相联映射

主存的每一块可以映射到Cache的多个固定块。



一、存储器的层次结构

②访问命中

$$\text{命中率} = \frac{\text{访问命中次数}}{\text{访问次数}} \times 100\%$$

③Cache内容替换

④数据一致性

2) 主存-外存层次

虚拟存储：主存+部分外存→虚存

需解决：

①存储空间分配 ②程序调度 ③虚-实地址转换



二、存储器的分类

凡是明显具有并能保持两种稳定状态的物质和器件，并且能够方便地与电信号进行转换，就可以作为存储介质。

存储介质种类：

1、磁芯存储器

基本单位：微型磁环，上世纪50年代—70年代。

2、半导体存储器

是构成高速缓存、主存的单元。

二、存储器的分类

半导体存储器分类

MOS型

ECL

TTL

SRAM

特点：速度快，功耗较低，集成度较高

DRAM

特点：速度较快，功耗很低，集成度很高

CMOS: complementary metal oxide semiconductor, 互补氧化

金属半导体



二、存储器的分类

- 1) 静态存储器：是依靠双稳态触发器的两个稳定状态保存信息。
- 2) 动态存储器：是依靠电容上的存储电荷暂存信息。

3、磁表面存储器

利用磁层上不同方向的磁化区域表示信息。容量大，非破坏性读出，长期保存信息，速度慢，作外存。

是构成外存的器件之一，分为：磁卡、磁鼓、磁带、磁盘等。



二、存储器的分类

4、光盘存储器

利用光斑的有无表示信息。容量很大，非破坏性读出，长期保存信息，速度慢，作外存。

- 1) 只读型光盘 (CD-ROM)
- 2) 写入式（只能写一次）光盘 (WORM)
- 3) 可擦除/重写型（可逆式）光盘



三、存取方式

1、随机存取存储器 (RAM) : random access memory

主存与高速缓存Cache是CPU可以直接编址访问的存储器，这就要求它们采取随机访问的存取方式。

随机存取的含义有两点：

①可按地址随机地访问任一存储单元，如可直接访问0000H

单元，也可直接访问FFFFH单元；CPU可按字节或字存取数据，进行处理。



三、存取方式

②访问各存储单元所需的读/写时间相同，与地址无关；可用读/写周期（存取周期）表明RAM的工作速度。

速度指标：存取周期或读/写周期。

作主存、高速缓存。

2、只读存储器 (ROM) :read only memory

只读存储器在正常工作中只能读出，不能写入。主存中常采用部分ROM固化系统软件中的核心部分、已调试好不再改变的应用软件、汉字字库一类信息。



三、存取方式

CPU中也常采用ROM，存放用来解释执行机器指令的微程序。

RAM: 可读可写

ROM: 只读不写

固存: 用户不能编程

PROM: 用户可一次编程

EPROM: 用户可多次编程（紫外线擦除）

EEPROM: 用户可多次编程（电擦除）

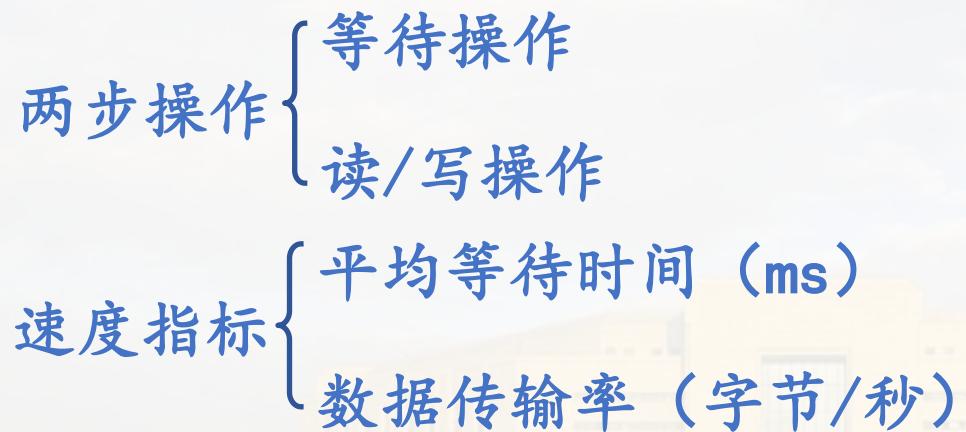
Flash Memory



三、存取方式

3、顺序存取存储器 (SAM) :sequential access memory

顺序存取存储器的信息是按记录块组织、顺序存放的，访问时间与信息存放位置有关。磁带是采取顺序存取方式的存储器。



三、存取方式

4、直接存取存储器 (DAM) :direct access memory

访问时读/写部件先直接指向一个小区域，再在该区域内顺序查找。访问时间与数据位置有关。

- 三步操作 {
- 定位（寻道）操作
 - 等待（旋转）操作
 - 读/写操作
- 速度指标 {
- 平均定位（平均寻道）时间 (ms)
 - 平均等待（平均旋转）时间 (ms)
 - 数据传输率（位/秒）

主要 内 容



- 1 概述
- 2 半导体存储器
- 3 主存储器组织
- 4 磁表面存储器



4.2 半导体存储器

- 01. 静态MOS存储单元与芯片 (SRAM)
- 02. 动态MOS存储单元与存储芯片 (DRAM)

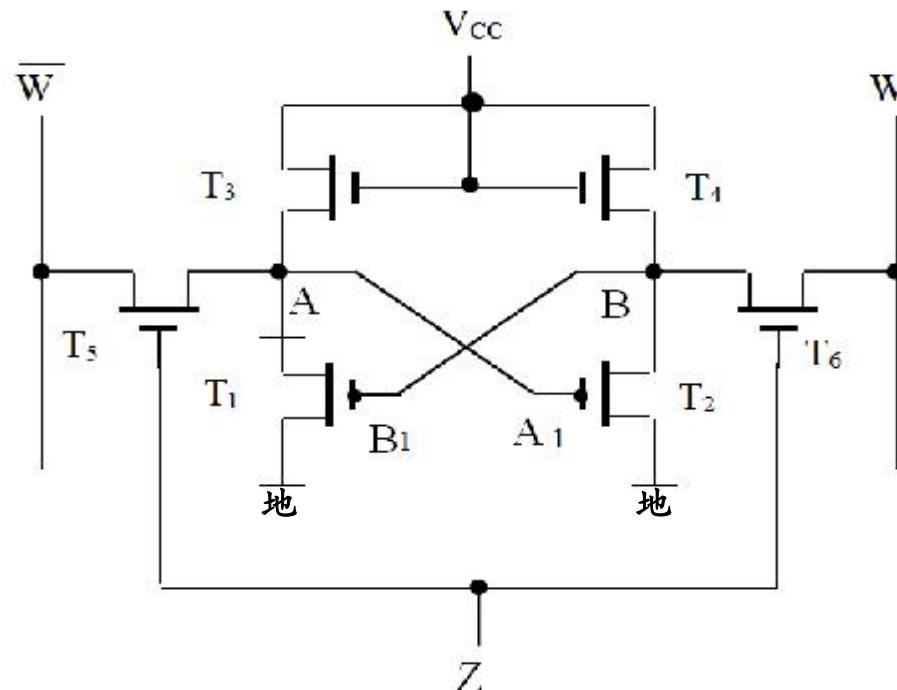


一、静态MOS存储单元与芯片 (SRAM)

1、定义

若 T_1 通导而 T_2 截止，存入信息为0。若 T_1 截止而 T_2 通导，存入信息为1。

2、NMOS六管静态存储单元电路



NMOS 六管静态存储单元

一、静态MOS存储单元与芯片 (SRAM)

3、工作

1) 写入:

首先字线Z加高电平时， T_5 、 T_6 导通。

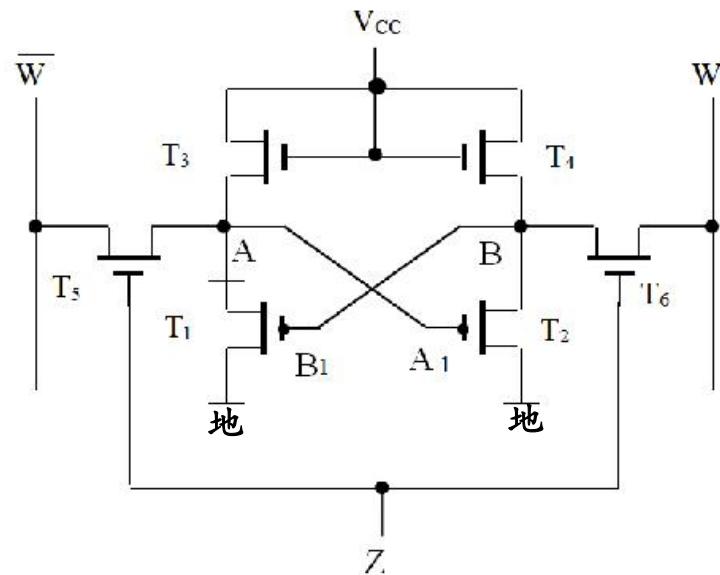
写”0”：

a. \overline{W} 加低电平，若 \overline{W} 电压 $>V_A$,

则： $\overline{W} \rightarrow T_5 \rightarrow A \rightarrow T_1 \rightarrow$ 对地放电，

使 A_1 变为低电平 $\rightarrow T_2$ 截止（因为 $V_A = V_{A1}$ ）；

b. 同时， W 加高电平，通过 $W \rightarrow T_6 \rightarrow B \rightarrow B_1$ 加高电平，使 T_1 导通。



NMOS 六管静态存储单元

一、静态MOS存储单元与芯片 (SRAM)

写”1”：

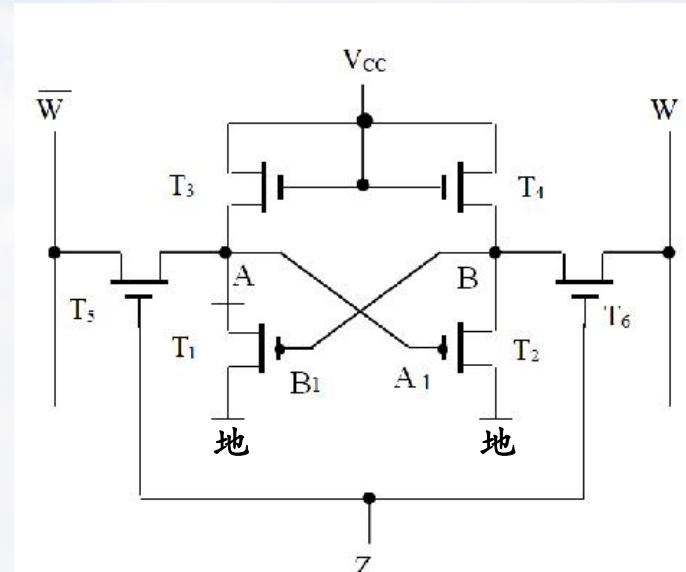
a. W 加低电平，若 W 电压 $>V_B$,

则： $W \rightarrow T_6 \rightarrow B \rightarrow T_2 \rightarrow$ 对地放电，

使 B_1 变为低电平 $\rightarrow T_1$ 截止

(因为 $V_B = V_{B1}$)；

b. 同时， \overline{W} 加高电平，通过 $\overline{W} \rightarrow T_5 \rightarrow A \rightarrow A_1$ ，加高电平，使 T_2 导通；



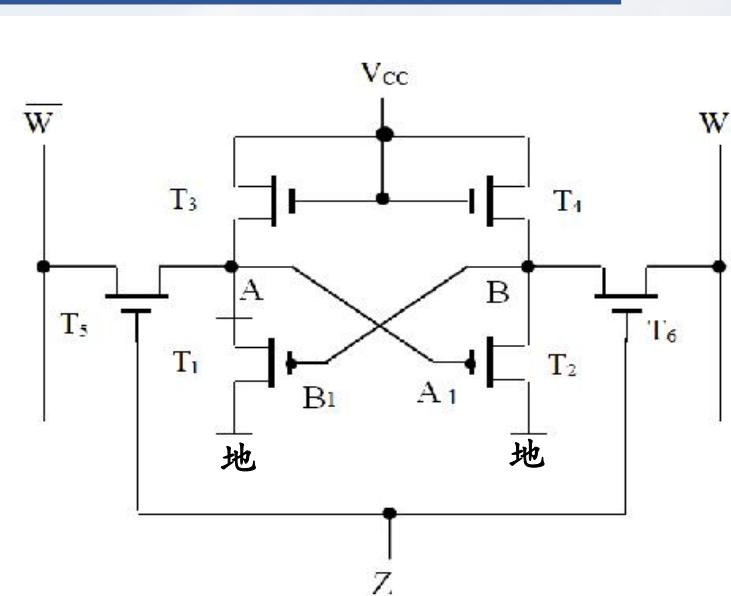
NMOS 六管静态存储单元

一、静态MOS存储单元与芯片 (SRAM)

2) 读出:

先对位线 W 、 \overline{W} 充电至高电平，该电平是浮动的，可随充放电而变；然后对字线 Z 加高电平，使 T_5 、 T_6 导通。

读”0”（即 T_1 导通）：位线 \overline{W} 上的 \overline{W} 电压 $>V_A$ 的电压，则 $\overline{W} \rightarrow T_5 \rightarrow A \rightarrow T_1 \rightarrow$ 地，形成放电回路，即有电流经 $\overline{W} \rightarrow T_1$ ，经放大为“0”信号，表明原存储信息为0。此时， T_2 截止 $(V_{A1}=V_A=0)$ ， W 上无电流。



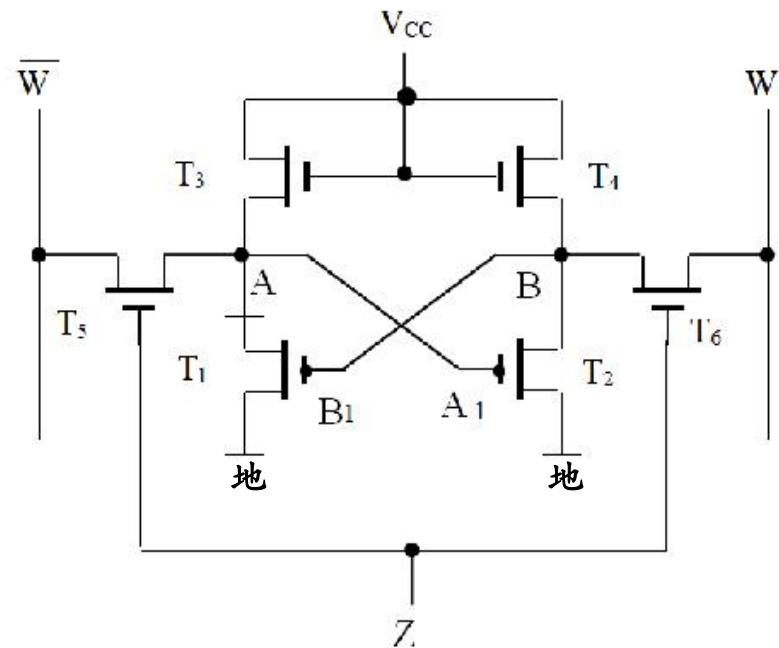
NMOS 六管静态存储单元

一、静态MOS存储单元与芯片 (SRAM)



读“1”（即T₂导通）：

位线W上的 V_W 电压>V_B的电压，
则W→T₆→B→T₂→地，形成放电回路，即有电流经W→T₂，经放大为“1”信号，表明原存储信息为1。此时T₁截止($V_{b1}=V_b=0$)，W上无电流。

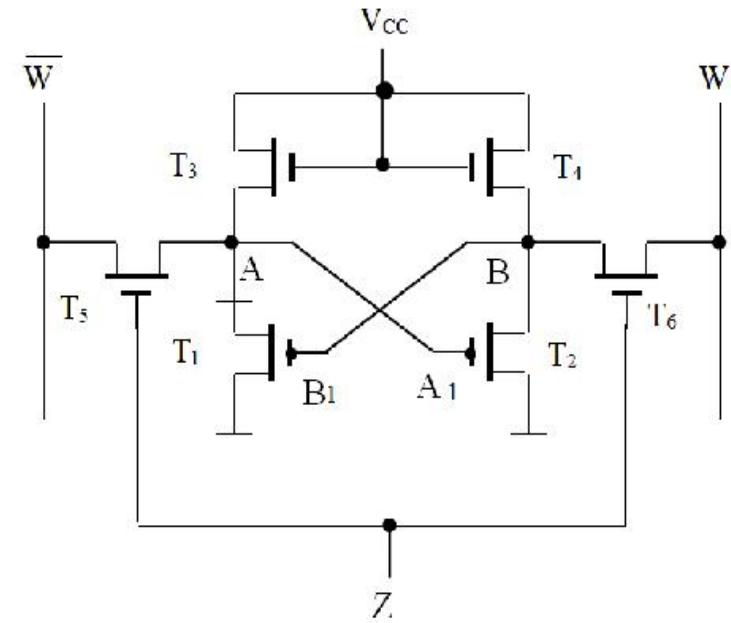


NMOS 六管静态存储单元

一、静态MOS存储单元与芯片 (SRAM)

3) 保持:

字线 Z 加低电平, 门管 T_5 与 T_6 断开, 位线与双稳态电路隔离, 双稳态电路依靠自身的交叉反馈保持原有状态不变。



NMOS 六管静态存储单元

总之, \bar{W} 上有电流为0, W 上有电流为1。上述读出过程并不改变双稳态电路原有状态, 属于非破坏性读出。

一、静态MOS存储单元与芯片 (SRAM)

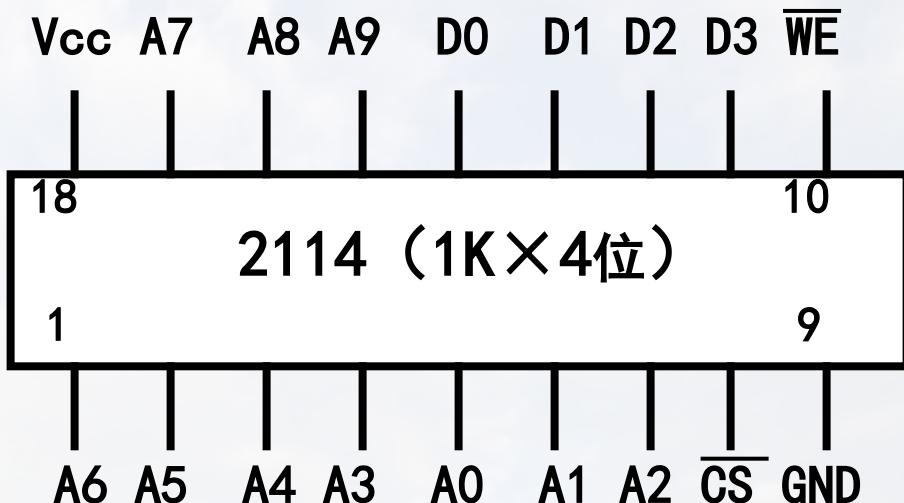
4、存储芯片

例如：SRAM芯片 2114

(1K×4位) 外部特性

地址端： $A_9 \sim A_0$ (入)

数据端： $D_3 \sim D_0$ (入/出)



控制端

片选 \overline{CS}	= 0	选中芯片
	= 1	未选中芯片
写使能 \overline{WE}	= 0	对芯片写操作
	= 1	对芯片读操作
电源 V_{CC} 、地线 GND		



二、动态MOS存储单元与存储芯片

1、动态MOS存储器的基本存储原理

是将存储信息以电荷形式存于电容之中，这种电容可以是MOS管栅极电容，或者是专用的MOS电容，通常定义电容充电至高电平时为1；放电至低电平时为0。

2、概念

1) 刷新：在MOS管断开之后，电容内部总存在泄漏通路，难以使泄漏电阻达到无穷大。时间过长，电容上的电荷会通过泄漏电路放电，使所存储的信息丢失。为此，经过一定时间后就需要对存储内容重写一遍，也就是对存1的电容重新充电，称为刷新。



二、动态MOS存储单元与存储芯片

2) 重写(再生)：

对于单管动态MOS存储单元而言，读操作后电容C上的电荷将发生变化，属于破坏性读出，需要读后对存1的电容补充电荷，称为重写(再生)。这一过程，由芯片内的外围电路自动实现。

二、动态MOS存储单元与存储芯片

3、动态MOS四管存储单元电路（非双稳态电路）

1) 组成

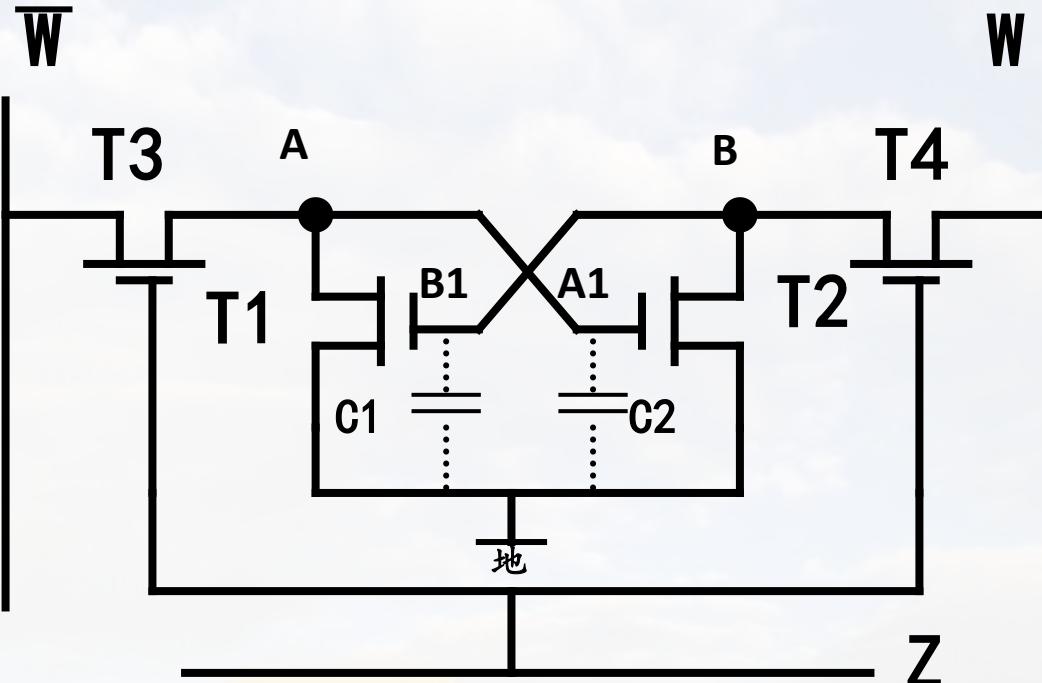
T_1 、 T_2 : 记忆管

C_1 、 C_2 : 栅极电容

T_3 、 T_4 : 控制门管

Z : 字线

\bar{W} 、 W : 位线



2) 定义

“0”： T_1 导通， T_2 截止（即 C_1 有电荷， C_2 无电荷）；

“1”： T_1 截止， T_2 导通（即 C_1 无电荷， C_2 有电荷）。

二、动态MOS存储单元与存储芯片

3) 工作

① 写入

首先字线Z加高电平时，
 T_3 、 T_4 导通。

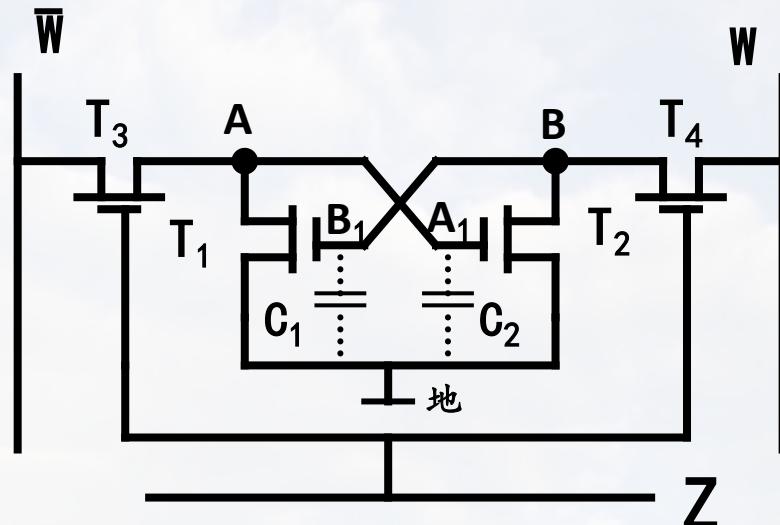
写”0”：

a. \overline{W} 加低电平，若 $V_{A1} > V_A$ ，

则 $C_2 \rightarrow A_1 \rightarrow A \rightarrow T_1 \rightarrow$ 对地放电，(使 V_{A1} 变为低电平 $\rightarrow T_2$ 截止；)

$C_2 \rightarrow A_1 \rightarrow A \rightarrow T_3 \rightarrow \overline{W}$ 放电 (瞬间)

b. 同时， W 加高电平，通过 $W \rightarrow T_4 \rightarrow B \rightarrow B_1 \rightarrow C_1$ 充电至高电平，使 T_1 导通。



二、动态MOS存储单元与存储芯片

写”1”：

- a. \overline{W} 加高电平, $\overline{W} \rightarrow T_3 \rightarrow A \rightarrow A_1 \rightarrow C_2$

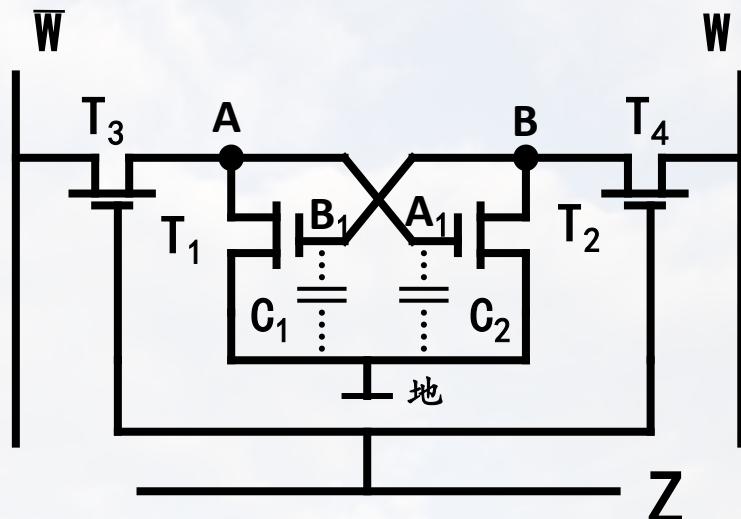
充电至高电平, 使 T_2 导通。

- b. 同时, W 加低电平, 若 $V_{B1} > V_B$, 则

$C_1 \rightarrow B_1 \rightarrow B \rightarrow T_2 \rightarrow$ 对地放电,

(使 V_{B1} 变为低电平 $\rightarrow T_1$ 截止;)

$C_1 \rightarrow B_1 \rightarrow B \rightarrow T_4 \rightarrow$ 对 W 放电 (瞬间)。



二、动态MOS存储单元与存储芯片

②读出

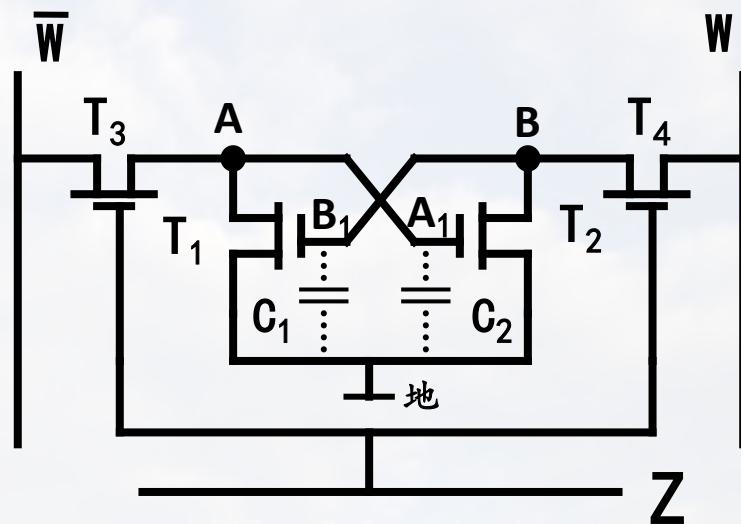
先对位线 W 、 \bar{W} 充电至高电平，该电平是浮动的；然后对字线 Z 加高电平，使 T_3 、 T_4 导通。

读“0”：

C_1 上有电荷， V_{B1} 为高电平， T_1 导通，

$\bar{W} \rightarrow T_3 \rightarrow A \rightarrow T_1 \rightarrow$ 对地放电，即 \bar{W} 上有电流通过，
放大后作为0信号读出；

同时， $W \rightarrow T_4 \rightarrow B \rightarrow B_1 \rightarrow C_1$ 充电至高电平，补充泄漏掉的电荷
四管单元为非破坏性读出，且读出过程为刷新过程。

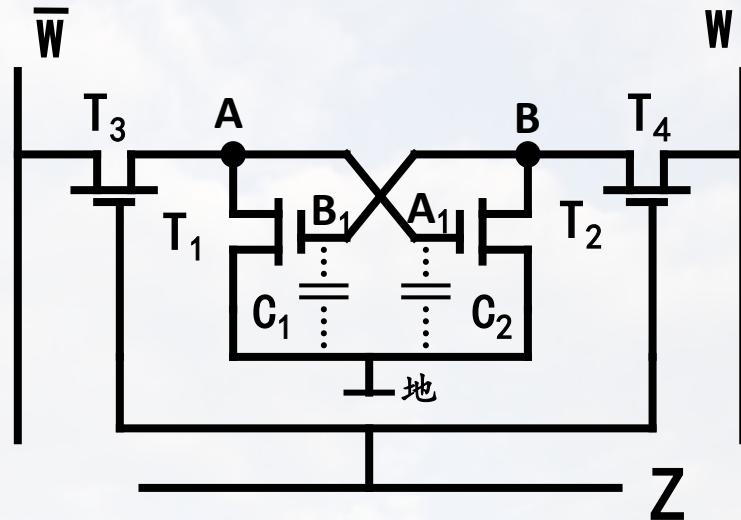


二、动态MOS存储单元与存储芯片

读”1”：

C_2 上有电荷, V_{A1} 为高电平, T_2 导通,
 $W \rightarrow T_4 \rightarrow B \rightarrow T_2 \rightarrow$ 对地放电, 即 W
 上有电流通过, 放大后作为1信号
 读出;

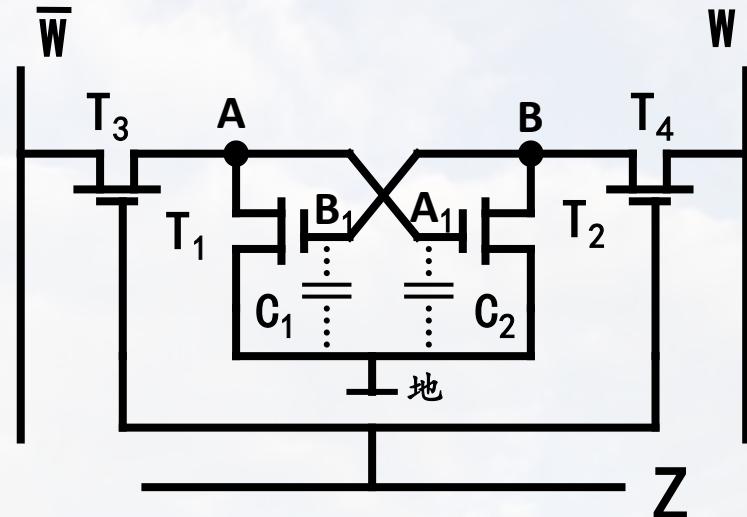
同时, $\overline{W} \rightarrow T_3 \rightarrow A \rightarrow A_1 \rightarrow C_2$ 充电至高
 电平, 补充泄漏掉的电荷四管单元为非破
 坏性读出, 且读出过程为刷新过程。



二、动态MOS存储单元与存储芯片

③ 保持

字线Z加低电平时， T_3 、 T_4 断开，基本上无放电回路，仅存在泄漏电流，信息可暂存数毫秒。



注：读出过程就是刷新过程。

二、动态MOS存储单元与存储芯片

4、单管单元

1) 组成

C: 记忆单元 T: 控制门管

Z: 字线 W: 位线

2) 定义

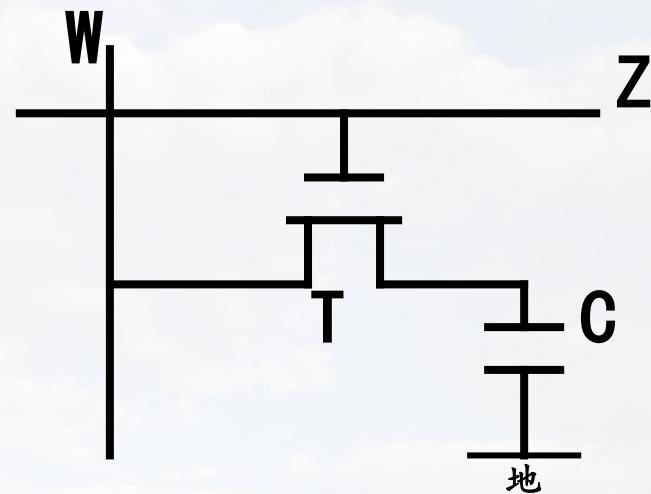
“0”：C无电荷，电平 V_0 （低）

“1”：C有电荷，电平 V_1 （高）

3) 工作

写入：Z加高电平，T导通，在W上加高/低电平，写1/0。

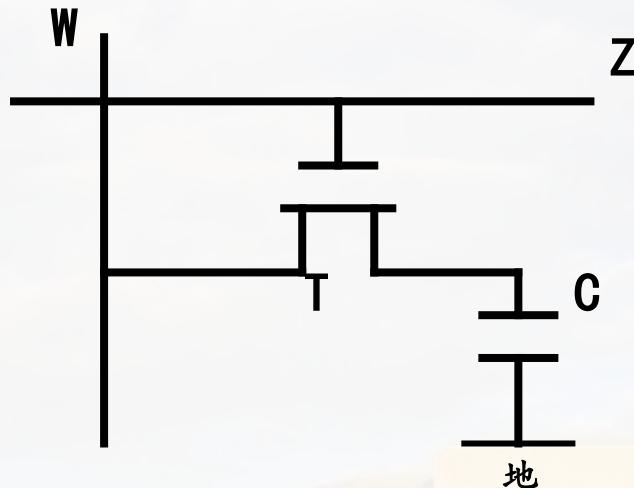
读出：W先预充电，断开充电回路。Z加高电平，T导通，根据W线电位的变化，读1/0。



二、动态MOS存储单元与存储芯片

保持：Z加低电平，T截止，该单元未选中，保持原状态。

单管单元是破坏性读出，读出后需重写。





谢谢观看

计算机组成原理



信息与软件工程学院
School of Information and Software Engineering