

存储器层次结构

存储器分类

介质

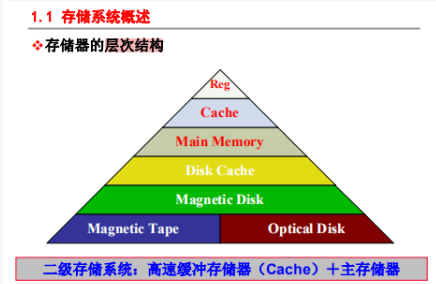
- 半导体存储器
 - 易失性
 - RAM
 - SRAM静态
 - 不用刷新，用作cache
 - DRAM动态
 - 需要刷新，用作主存
 - SDRAM
 - DDR
 - DDR2
 - DDR3
 - 集中式刷新
 - 分散式刷新
 - 异步式刷新
 - 非易失性
 - ROM
 - PROM
 - EPROM
 - EEPROM
 - Flash Memory
 - BIOS，图形卡，硬盘控制器
 - 磁介质存储器
 - 非易失性
 - 光盘存储器
 - 非易失性

访问方式

- 随机访问存储器
 - 需要地址译码
- 顺序访问
- 直接访问
- 只读

功能

- 高速缓冲存储器
- 主存
- 辅存
- 控制存储器
 - CPU内

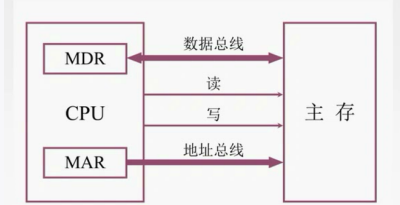


层次结构

- 存储时间
 - 读或者写操作作用的时间
- 存储周期
 - 两次访问存储单元的最小时间间隔
- 存储带宽
 - 单位时间访问的存储量

存储芯片内部结构

- 1k*2
 - 1024个字单元，每个字单元2位
 - 10位地址线，2位数据线
 - 一维地址结构
- 2^14*4
 - 4096个字单元，每个字单元4位
 - 7+7位地址线，4位数据线
- 扩展
 - 位扩展
 - 每个存储字单元位数增加，即数据线增加
 - 字扩展
 - 地址线增加
 - 混合扩展
- 刷新
 - 片选信号
 - 选择哪个芯片
 - 最高几位
 - DRAM按行刷新（二维地址结构），刷新地址计数器位数就是行数
 - 集中式、分散式、分布式



主存和CPU连接

- 读
 - CPU先把地址发送给MAR（memory address register）
 - 经过地址总线找到主存地址
 - 控制总线发送读命令
 - 通过数据总线传出数据
- 写
 - CPU先把数据发给MDR（memory data register），地址发给MAR
 - 向主存发送写命令
 - 主存收到命令后在对应地址写入数据

高速缓冲存储器cache

- 存储访问的局部性原理
 - 时间局部性
 - 空间局部性
- 原理
 - cache和主存之间以数据块（block）（cache line）为单位进行数据交换
 - cache块大小和主存块大小相等
- 结构
 - 有效位
 - tag
 - 主存块地址
 - 数据
- 映射机制
 - 全相联
 - 某一块映射到任意block
 - 地址格式
 - block number + offset
 - 直接
 - 每一块固定映射
 - 地址格式
 - 区地址+区内块地址+offset
 - 组相联
 - cache和主存分成k组，每组进行全相联
 - 地址格式
 - 组内块地址+组地址+offset
- 替换策略
 - 最近最少使用
 - 先进先出
 - 最少使用频率
- 存取时间
 - cache不命中，也需要读取一次cache
 - 设命中率u，cache时间为a，主存时间为b，平均存取时间为au+（a+b）（1-u）