

Chapter3 I/O 系统

计算：通道流量，画图、分析

重点：

IO 控制方式

RAID

通道处理机工作原理、通道类型、通道流量分析

IO 系统包括：IO 设备、设备控制器、IO 相关硬件

设计目标：成本/性能/支持多种设备，避免 IO 性能瓶颈。

对于 IO 操作：

早期抵档单用户计算机：程序员自己安排

大多数计算机上：OS 分配调度 IO 处理，因此 IO 系统结构设计应该是面向 OS

IO 系统的功能：

对指定外设进行 IO 操作：

编址

准备通路

信息传送

格式变换

中断请求

上述功能由三个部分协同完成：

- 输入输出机器指令；
- 输入输出设备及其控制器硬件；
- 操作系统；

性能主要包括：

输入输出速度

用户输入到输出的等待时间

CPU 和主存的利用率

IO 系统的兼容能力、可扩展能力，综合处理能力，性价比

误区：使用多进程技术可以忽略 IO 性能对系统性能的影响

只能提高系统吞吐率？不能减少系统响应时间，切换的时候可能需要增加 IO 操作，IO 速度慢的时候导致 CPU 空闲

小计算：当 IO 性能不变，IO 处理占 10%，CPU 性能提升 100 倍，总体提升多少
设运行时间为 1，改进后处理时间为 $\frac{1-10\%}{10} + 10\% = 0.19$ ，提升约 5 倍

Make the Common Case Faster

I/O 系统的特点：

- 1.异步性：处理机与外设、外设与外设之间能够并行工作，不使用统一的中央时钟
- 2.实时性：防止提供服务不及时丢失数据
- 3.与设备无关性：独立于具体设备标准接口：串口并口、USB、HDMI, etc.

发展阶段（3）：

- 1.程序控制与中断方式
- 2.DMA
- 3.I/O 处理机方式

程序控制方式：

特点： CPU 直接进行 I/O

工作原理： CPU 查询外部设备准备好之后才开始传送数据

硬件要求： 电路简单

优点： 灵活、容易改变外设的优先级

缺点： 通过程序同步，并行困难、速度慢、CPU 被外部设备独占，CPU 效率低

适用： CPU 不忙且传送速度要求不高

中断方式：

I/O 重要组成部分

中断源（事件）

中断请求（中断源发出）

中断响应（CPU 调用中断处理程序）

中断优先级：相应和处理中断优先次序

IO 中断在 IO 请求和操作完成或出现故障的时候发出

适用： 并行工作、低速外设

DMA：

主存和外设直接访问通路

CPU 挪用一個存储周期启动 DMA

DMA 开始结束需要处理机管理，传送过程中不需要 CPU 干预

优点： 并行工作、效率高

缺点： 外设管理、DMA 启动、数据准备、操作完成后处理还需要 CPU

适用： 高速外设（磁盘）

IO 处理机方式：

(1.通道 2.外围处理机)

CPU 与外设并行

*程序查询和中断方式主要软件实现，DMA 和 IO 处理机主要由硬件实现

IO 系统的组织：

- a. **针对异步性**：采用自治控制的方法->IO 系统是独立于处理机之外的自治系统
- b. **针对实时性**：采用层次结构的方法->里层处理机/通道，中间层标准接口，设备控制器控制外围设备工作
- c. **针对与设备无关性**：采用分类处理的方法->面向字符（打印机），面向数据块（磁盘），etc.

3.2 磁盘阵列 RAID

磁盘驱动器某种方式连接，形成逻辑上的一个磁盘

RAID 的优点：

- a. 成本低，功耗小，速率高
- b. 提供容错功能，安全性高
- c. 同样容量下价格更低

RAID 级别	名称	最少数据 磁盘数	可正常工作的 最多失效磁盘 数	检测 磁盘数
RAID 0	无冗余无校验的磁盘阵列	2	0	0
RAID 1	镜象磁盘阵列	2	1	0
RAID 2	纠错海明码磁盘阵列	2	1	1
RAID 3	位交叉奇偶校验的磁盘阵列	2	1	1
RAID 4	块交叉奇偶校验的磁盘阵列	2	1	1
RAID 5	无独立校验盘的奇偶校验磁盘阵列	2	1	1
RAID 6	双维无独立校验盘的奇偶校验磁盘阵列	2	2	2

RAID0：数据分块，没别的东西，无冗余信息，严格说不属于 RAID

优点：高性能，磁盘利用率高

缺点：可靠性差

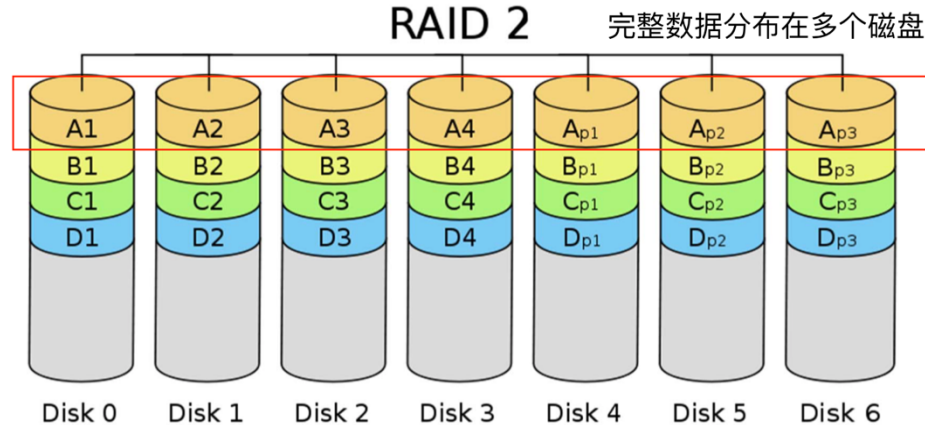
RAID1：磁盘镜像，双备份

优点：可靠简单

缺点：开销大，效率低

*RAID 0+1：RAID10，同时有两者的优缺点

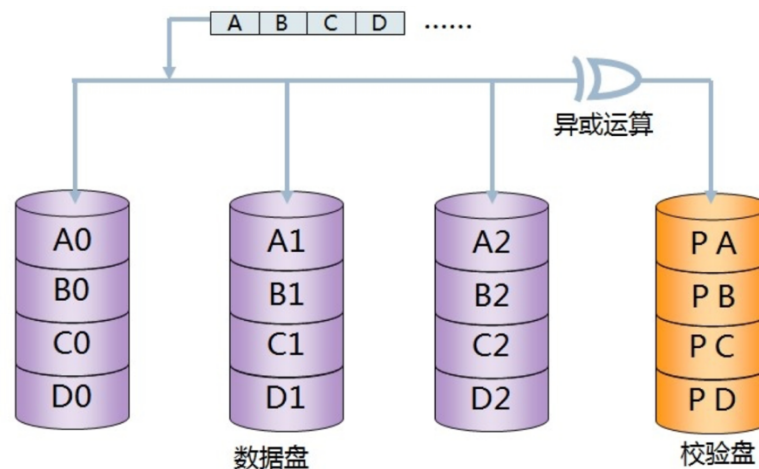
RAID2 : 位交叉海明编码阵列, 至少需要三块硬盘



优：高速误差校正，数据传输快，设计相对简单

缺：校正空间大（校正位整个磁盘都是冗余数据），利用率低，需要同步控制

RAID3 : 位交叉奇偶校验磁盘阵列, 至少 3 块磁盘



校验位存在单独磁盘上

优：读写快，磁盘失效对吞吐率影响小，利用率高

RAID4 : 专用奇偶校验独立存取盘阵列（因为专用所以有校验盘）

和 RAID3 基本一致，不过最小单位不是位而是数据块

优点一样，缺点是设计复杂，磁盘失效对吞吐影响大

RAID5 : 块交叉分布式奇偶校验盘阵列（无校验盘）至少 3 块，可容忍单盘出错

和 RAID4 基本一致，但是没有独立校验盘，校验信息分布在每个磁盘

优：数据率最高，利用率高

缺：设计复杂，盘失效影响系统可靠性

RAID6：双维奇偶校验独立存取，块交叉冗余信息分布各个磁盘，容忍双盘出错，RAID5 的扩展

RAID7：RAID6+Cache

级别	技术	描述	速度	容错能力
RAID 0	磁盘分段	没有校验数据	磁盘并行I/O，存取速度提高最大	数据无备份
RAID 1	磁盘镜像	没有校验数据	读数据速度有提高	数据 100% 备份
RAID 2	磁盘分段+海明码数据纠错	专用校验数据盘	没有提高	允许单个磁盘错
RAID 3	磁盘分段+奇偶校验	专用校验数据盘	磁盘并行I/O，速度提高较大	允许单个磁盘错，校验盘除外。
RAID 4	磁盘分段+奇偶校验	异步专用校验数据盘	磁盘并行I/O，速度提高较大	允许单个磁盘错，校验盘除外。
RAID 5	磁盘分段+奇偶校验	校验数据分布存放于多盘	磁盘并行I/O，速度提高较大，比 RAID 0 稍慢	允许单个磁盘错，无论哪个盘。
RAID 6	磁盘分段+双校验	校验数据分布存放于多盘	RAID5 扩展	允许两个磁盘错，无论哪个盘。

-----end of RAID-----

3.3 总线设计

计算机各个子系统通过总线互连，分时共享信息传送

优：成本低简单易用

缺：总线带宽成为 IO 吞吐量瓶颈

总线事务：一次信息交换过程，包含地址阶段和数据阶段，一次时间叫做总线周期

*发出请求的是主设备，与主设备信息交换的是从设备

总线分类：单向/双向；专用/非专用；片内/片外 or 内部/外部（指 CPU）

CPU 总线：CPU 与外围芯片连接

系统总线：IO 通道总线，ISA、PCIE, etc.

外设总线：主机与外设接口总线，IDE、USB, etc.

性能指标：

1. 总线宽度：线数，受地址线和数据线的数量直接影响
2. 总线带宽：总线频率 x 总线宽度

3. 总线负载：最大设备数量
4. 总线复用：分时使用来减少总线数量，提高利用率
5. 总线猝发（burst）传输：一个周期内传输多个连续地址数据

概念：

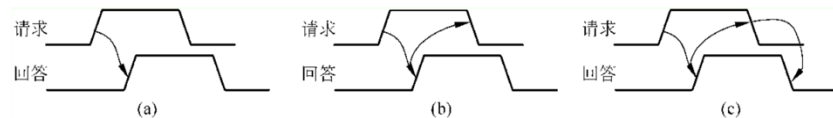
- a. 数据宽度->取得 IO 总线后所传送数据总量
- b. 数据通路宽度->数据总线的物理宽度（数据总线宽度）
- c. 单字宽度适用低速设备，定长块宽度适合于高速设备
- d. 变长块->高优先级中高速设备（灵活性好）
- e. 单字加定长宽度->速度低而优先级较高的设备（信息块超过定长块可以用单字传输）

设备使用总线的步骤是：总线请求、总线仲裁、寻址、传送数据、检错和报错。

总线控制线路包括：总线仲裁逻辑、驱动器和中断逻辑等

总线传输技术问题：

- a. 总线传输同步
- b. 总线仲裁控制
- c. 出错处理
- d. 总线驱动



3.3.3 总线定时控制：

同步定时（统一的时钟信号来协调）、异步定时（应答方式）->三种情况：不互锁、半互锁、全互锁（最复杂但可靠,回答后才撤销请求信号）

总线通信技术：

同步通信：速率高，容易造成同步误差或者时钟线上的干扰信号引起误同步。

异步通信：单向控制（源控制或目的控制）、双向控制（双方共同控制）

总线仲裁-按照某种优先次序裁决-（集中式与分布式）

集中式：

- a. 链式查询：查询三个控制线，BG 有效才响应
- b. 计数器定时查询：优先级顺序固定，相当于循环问设备，你没有需求计数器+1，问下一个设备；计数器到头了，从 0 开始重新问
- c. 独立请求方式：独立的请求线 BR 和回答线 BG

分布式：不需要中央仲裁器，每个仲裁器将总线上得到的号与自己的号进行比较，如果仲裁总线上的号大，则总线请求不予响应，撤销仲裁号。获胜者的仲裁号留在总线上-优先级仲裁策略。

1. **自举分布式**-每个设备的优先级固定
2. **冲突检测分布式**-查看总线是否被占用，没有的话就自己用并设置为总线忙，同时请求的时候，优先级高的先使用总线
3. **并行竞争分布式**-根据仲裁算法决定，复杂但有效

3.4 通道处理机

由 IBM 提出，是具有特殊功能的处理器，执行程序实现 IO 通道控制主存和外设之间直接传送数据

特点：拥有通道指令和通道程序（存放在对应该通道的主存缓冲区，可以和 CPU 并行工作）通道统一管理外设

功能：接受 CPU IO 指令
执行通道程序/指令
组织 IO 传送
向 CPU 报告状态
向 CPU 发出中断请求

3.4.2 通道工作过程

1. CPU 控制的 IO 被定义为 管态指令
2. 用户在目态程序使用广义指令请求 IO
3. 广义指令由 访管指令和若干参数组成
4. 访管指令为目态指令

工作流程：中断 (mu->guan)
开始传送 (guan->mu)
结束中断 (mu->guan)
程序正常执行 (guan->mu)

流程：

传送准备-用户使用访管指令进入管理程序，CPU 通过管理程序组织一个通道程序，并启动通道

数据传送-通道处理机执行通道程序

传送结束-第二次调用管理程序对输入输出请求进行处理

分别对应工作流程前三个

3.4.3 通道类型

传送方式分：

1. 字节多路通道-每次一个字节，字节交叉方式轮流为低速设备服务，子通道独立执行通道指令。
2. 选择多路通道-高速通道，一段时间内只能选择一台设备数据传送，每次传送一个变长数据块
3. 数组多路通道-类磁盘外设，速率高，寻址耗时，成组交叉方式，每次定长数据块，多路并行

3.4.4 通道流量分析

通道流量 又通道吞吐率、通道数据传输率

-数据传送期，单位时间内传送的字节数

-常见-通道极限流量（与选择一次设备的时间 T_S 和传送一个字节的的时间 T_D 有关

-实际流量(通道上挂接设备后所有设备要求的通道流量)

字节多路通道

P台设备每台传送n个数据总共所需的时间为：

$$T = (T_S + T_D) \times P \times n$$

$$\text{极限流量: } f_{\max \cdot \text{Byte}} = \frac{\text{data}}{\text{time}} = \frac{P \times n}{(T_S + T_D) \times P \times n} = \frac{1}{T_S + T_D}$$

实际流量：P台设备传输率之和

$$f_{\text{byte}} = \sum_{i=1}^P f_i$$

数组多路通道 (每次传送k字节, 和选择通道不同的就是k定长, 但是n变长)

$$T = \left(\frac{T_S}{k} + T_D \right) \times P \times n$$

$$f_{\max \cdot \text{block}} = \frac{\text{data}}{\text{time}} = \frac{k}{T_S + kT_D}$$

$$f_{\text{block}} = \max_{i=1}^P f_i$$

选择多路通道 (每次每个设备n个字节全部传送完)

$$f_{max-select} = \frac{data}{time} = \frac{n}{T_s + nT_D}$$

$$f_{select} = \max_{i=1}^P f_i$$

设计基本原则：

1.实际流量不超过极限流量

即 $f_{max} \geq f$ 通过左右差值来衡量IO系统的利用率，左右相等的时候系统处于满负荷状态，max也是对主存频宽 B_m 的要求

2.一般速度快的设备请求有较高的响应优先级

问题：未得到及时处理导致的数据丢失

处理方法：

1.增加通道的最大流量

2.动态改变设备的优先级（在某时间短临时提高优先级）

3.增加缓冲存储器（锁存器）-特别是对于优先级比较低的设备

3.5 外围处理机（PPU）

通道处理机的局限性->指令功能简单、没有大容量存储器、只有面向外设控制和数据传送的功能->不能被看成是一台独立的处理机

输入输出的过程还需要 CPU 处理->IO 前后处理、IO 异常处理等

基本上独立于主机的异步方式工作

可以与主机共享内存、也可以不共享内存->但需要大容量存储

优点：灵活、减轻 CPU 负担，是一个独立的处理机（有时候就是通用机）

发展：前端机（网络系统，远程终端控制）、后端机（数据库机）

3.6 中断系统

1.中断请求（中断引脚有效-中断源（请求寄存器）-中断屏蔽寄存器-优先级判定（硬件+软件））

2.中断响应（时间：当前指令执行完、

响应条件：有中断请求、CPU处于开中断状态

响应动作：执行中断隐指令（硬件实现、对程序员透明）-关中断、保存程序断点、转入中断服务程序

中断源识别：向量中断、软件查询、）

3.中断服务：保护现场（寄存器内容入栈）-中断服务（IO或突发事件）-恢复现场-中断返回

4.中断返回：执行中断服务程序中的IRET指令，标志寄存器和PC出栈