

输入输出与中断

主讲人: 邓倩妮
上海交通大学



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

本节内容



- 外围设备的定义和分类
- 外设接口
- 输入输出的控制方式
 - 程序查询方式
 - 中断方式
 - DMA控制方式

外围设备的定义



- 广义:凡与计算机相连、受主机控制完成数据处理或控制操作的装置都可称**外围设备**。
 - 输入输出设备, 外存储器, 数据通讯设备, 终端设备, 以及计算机自动控制、检测系统中的各种采样、检测设备, A / D和D / A转换器, 各种与计算机相关的仪表及装置。
- 狭义:仅指输入输出设备和外存储器, 它们是只用于输入输出数字信息的、构成计算机系统必不可少的设备。
- 外围设备又称**外部设备**,

外围设备分类



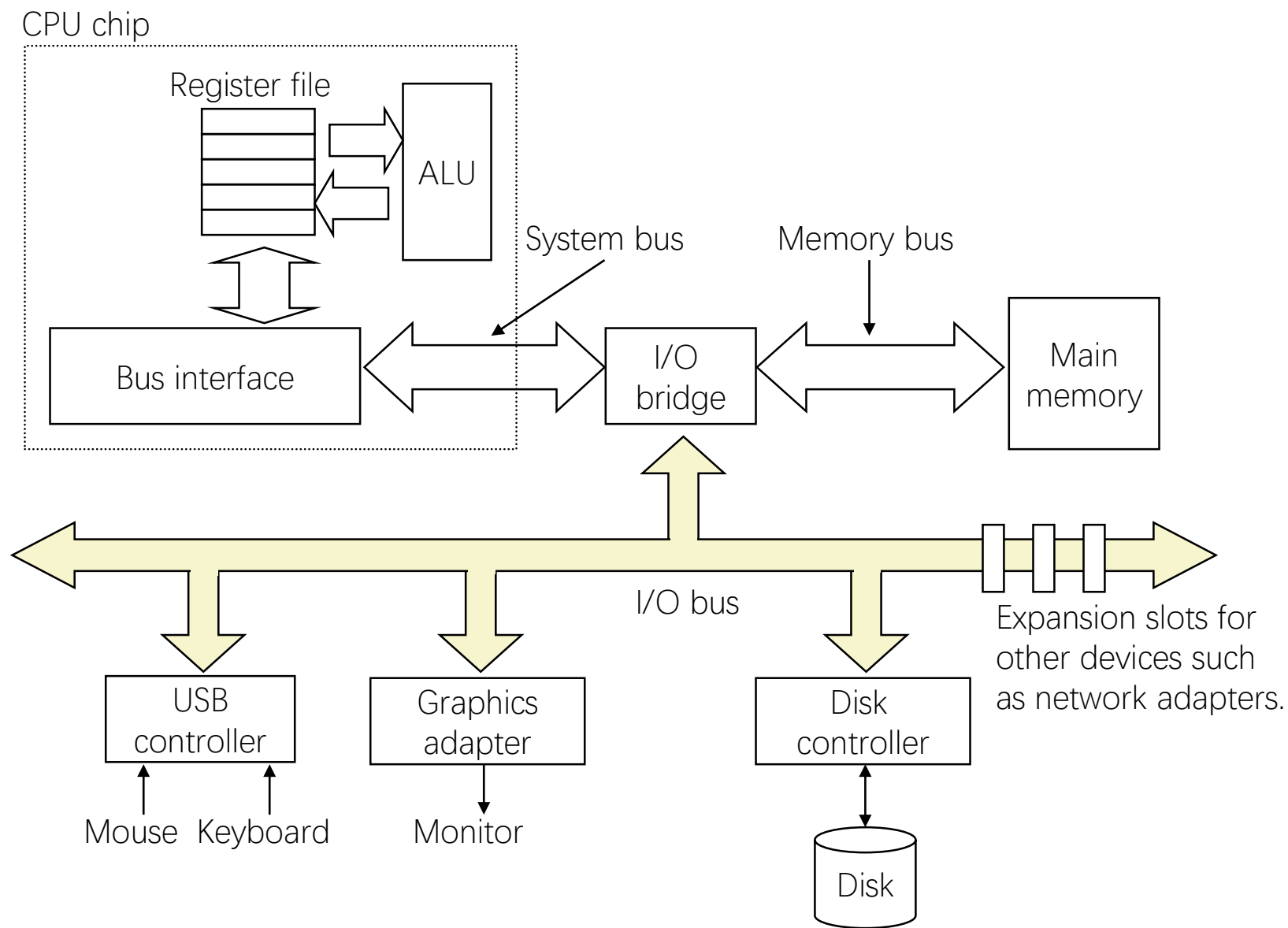
- 输入设备：键盘、鼠标、扫描仪、手写笔
- 输出设备：显示器，打印机，绘图仪；
- 输入兼输出设备：终端，传真机；
- 外存储器：磁盘，硬盘、磁带、光盘；
- 数据通讯设备：FAX，MODEM；
- 过程控制设备：A/D，D/A，传感器，仪表。

外围设备特点

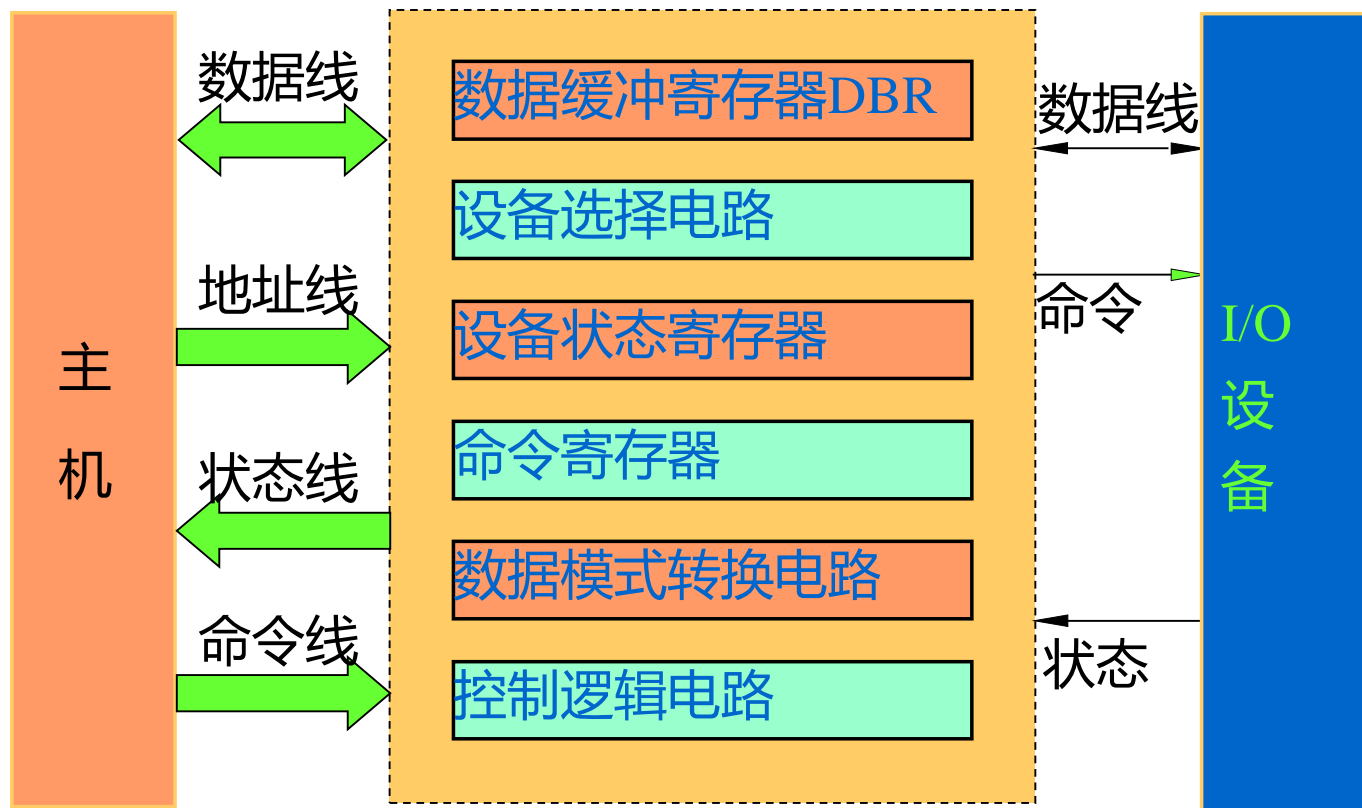


- 由设备和设备控制器组成;
- 工作速度比主机慢很多;
- 不同设备的信息类型和格式不同;
- 以上这些特点给主机与输入输出设备的连接带来复杂性。因此，必须设置一个 I / O 接口来解决信息的缓冲、同步与通讯，格式的转换及电气特性的适配等一系列问题。

I/O Bus



I/O接口（controller）的组成



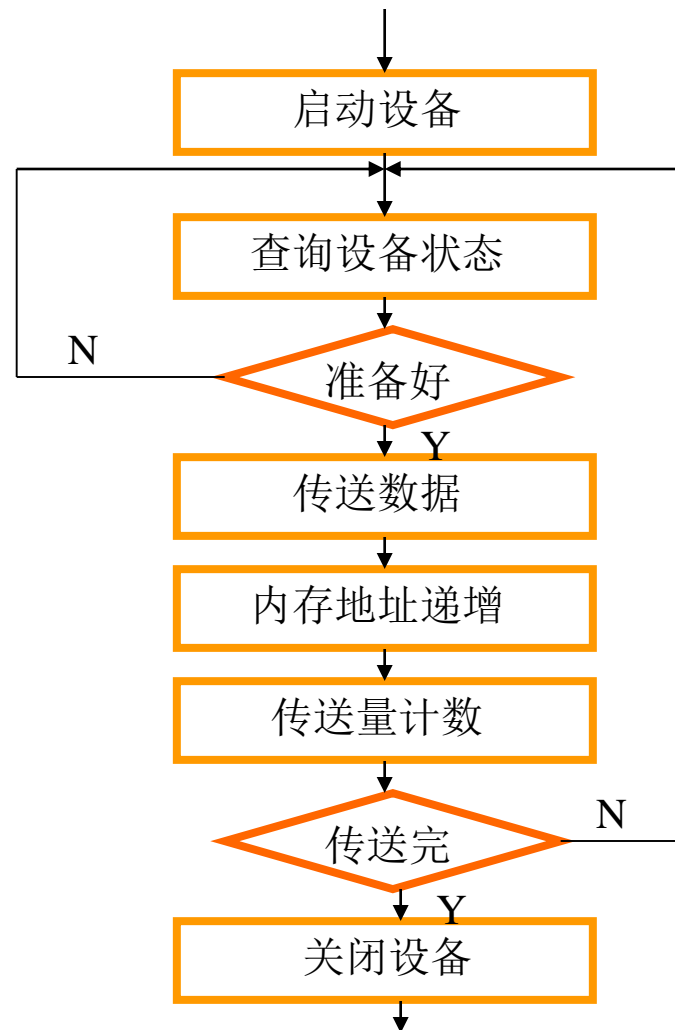
输入输出的控制方式



- 程序查询方式
- 程序中断方式
- 直接内存访问方式(DMA)
- 通道方式



程序查询方式：CPU的处理过程



程序查询方式



- 数据传输完全依赖于程序控制
- 控制器的硬件结构简单
- 频繁的查询动作浪费了大量宝贵的CPU时间
- 最古老的一种方式，目前很少使用

程序中断方式



- 外围设备主动通知CPU，准备数据传输
- 当中断发生时，CPU中断当前的工作，转向中断处理子程序，该程序用于数据传输。
- 提高了CPU的使用效率。
- 适合随机出现的服务
- 硬件结构相对复杂，服务开销时间大

中断的概念



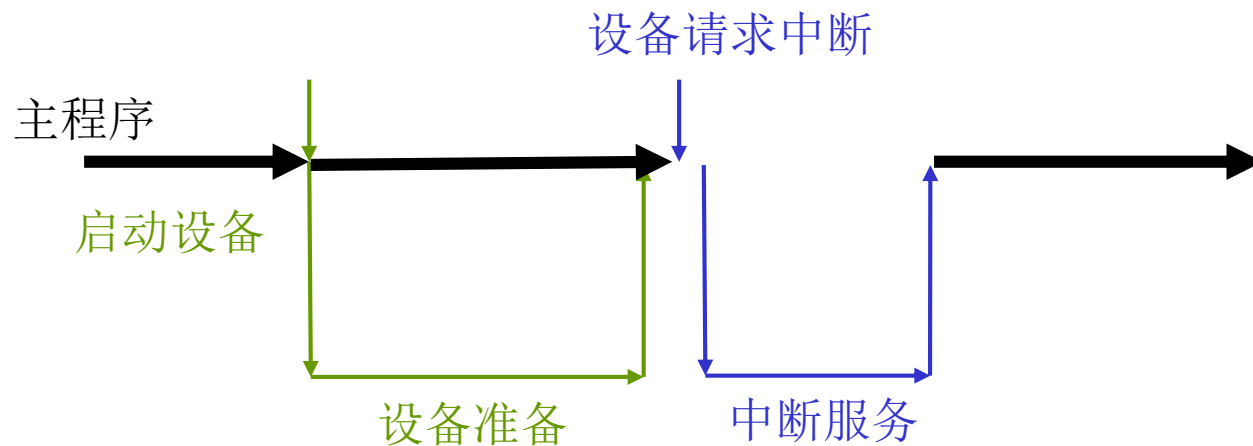
- 中断是指程序执行过程中，遇到急需处理的事件时，暂时中止CPU上现程序的运行，转去执行相应的事件处理程序，待处理完成后返回原程序被中断处或调度其他程序执行的过程。

中断作用

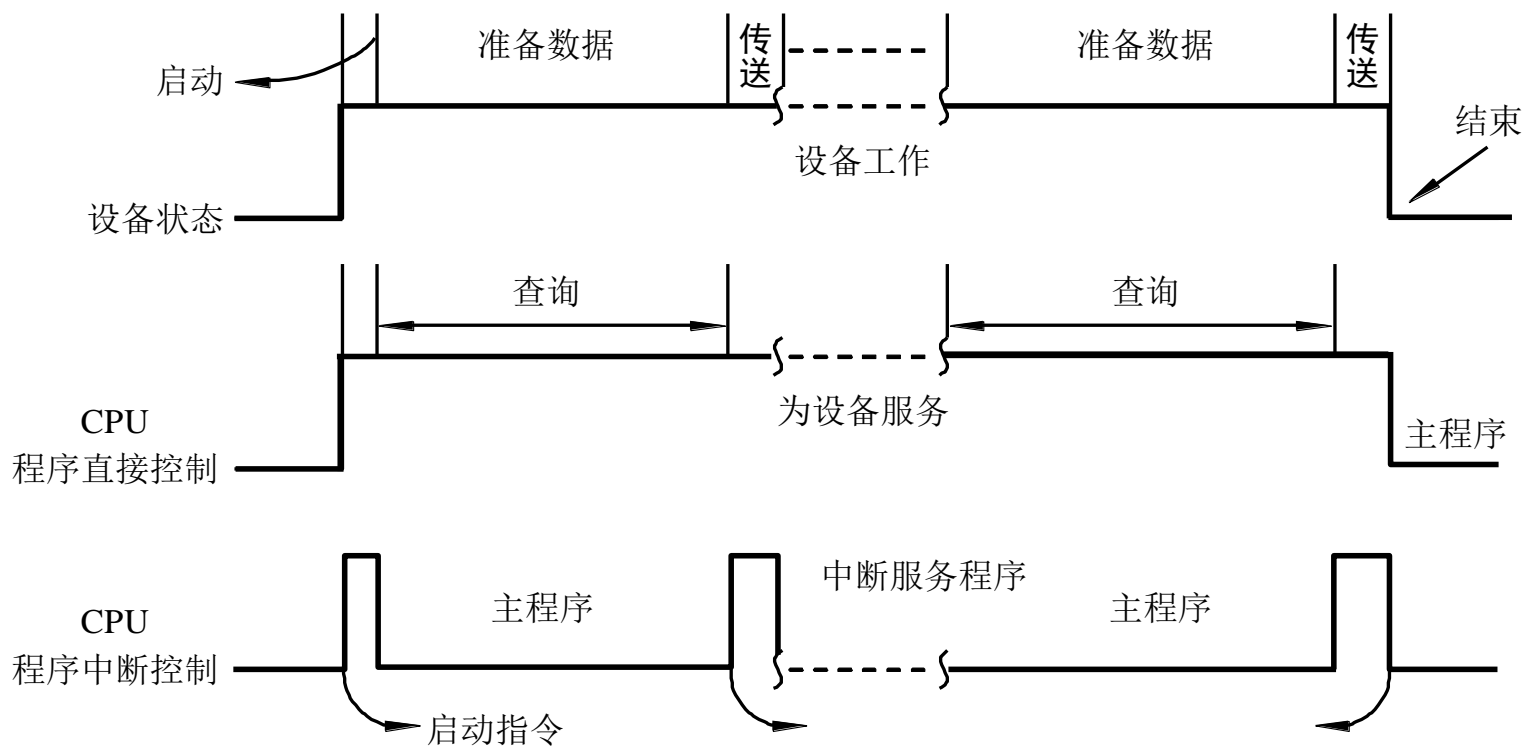


- 实现主机和外部设备并行工作

使用中断技术后，CPU原来用于查询外设状态的时间被充分地利用起来了，其工作效率得到了显著的提高。



两种控制方式中CPU运行轨迹



中断源（1）：外部事件



- 时钟中断、间隔时钟：计时与时间片处理
- 设备报到与传输结束：调整设备表
- 键盘/鼠标信号：根据信号作出相应反应
- 关机/重新启动：写回文件，停止设备与CPU

中断源（2）：输入输出设备报告状态

- 来源于外围设备报告I/O状态的中断事件
 - I/O完成：调整进程状态，释放等待进程
 - I/O出错：等待人工干预
 - I/O异常：等待人工干预

中断源（3）：程序性中断事件



处理器执行机器指令引起：

- 除数为零、操作数溢出等**算术异常**：简单处理，报告用户；也可以由用户编写异常处理程序进行处理
- **非法指令**、用户态使用特权指令、地址越界、非法存取等指令异常：终止进程
- **终止进程**指令：终止进程
- 虚拟**地址**异常：调整内存后重新执行指令

中断源（4）：自愿性中断事件



- 处理器执行**陷入 (trap)** 指令，主动请求OS服务；在操作系统中，它一般又被称作**系统调用**
- 请求分配外设、请求I/O、等等
- 处理流程是：陷入OS，保护现场，根据功能号查入口地址，跳转具体处理程序

中断源（5）：处理器硬件故障



- 由处理器、内存储器、总线等硬件故障引起
- 处理原则为：保护现场，停止设备，停止**CPU**，向操作员报告，等待人工干预

中断系统

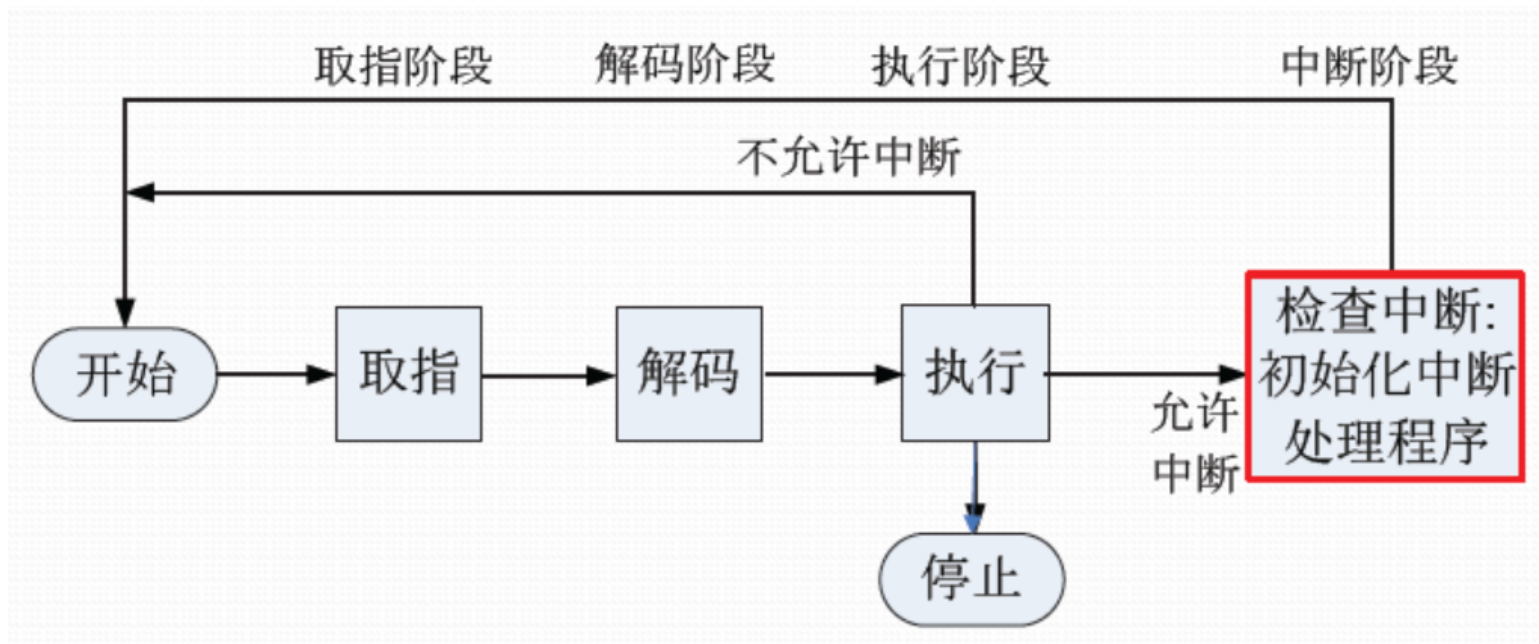


- 中断系统：计算机系统中响应和处理中断的系统，包括两部分：
 - 中断响应由硬件子系统完成
 - 中断处理由软件子系统完成

中断响应：硬件检测



在指令执行周期的最后一个阶段增加一个“检查中断”的操作，以响应中断



中断控制器



- **中断控制器**：CPU中的一个控制部件，包括中断控制逻辑线路和中断寄存器
 - 外部设备向其发出中断请求 **IRQ**，在中断寄存器中设置已发生的中断（的原因）
 - 指令处理结束前的最后一个阶段，会检查中断寄存器，若有不被屏蔽的中断产生，则改变处理器内操作的顺序，引出操作系统中的中断处理程序

中断响应过程（硬件完成）



●发现中断源，提出中断请求

- 发现中断寄存器中记录的中断
- 决定这些中断是否应该屏蔽
- 当有多个要响应的中断源时，根据规定的优先级选择一个

●中断当前程序的执行

- 保存当前程序的PSW/PC到核心栈（保护现场）

●转向操作系统的中断处理程序

中断处理：软件（操作系统）来完成



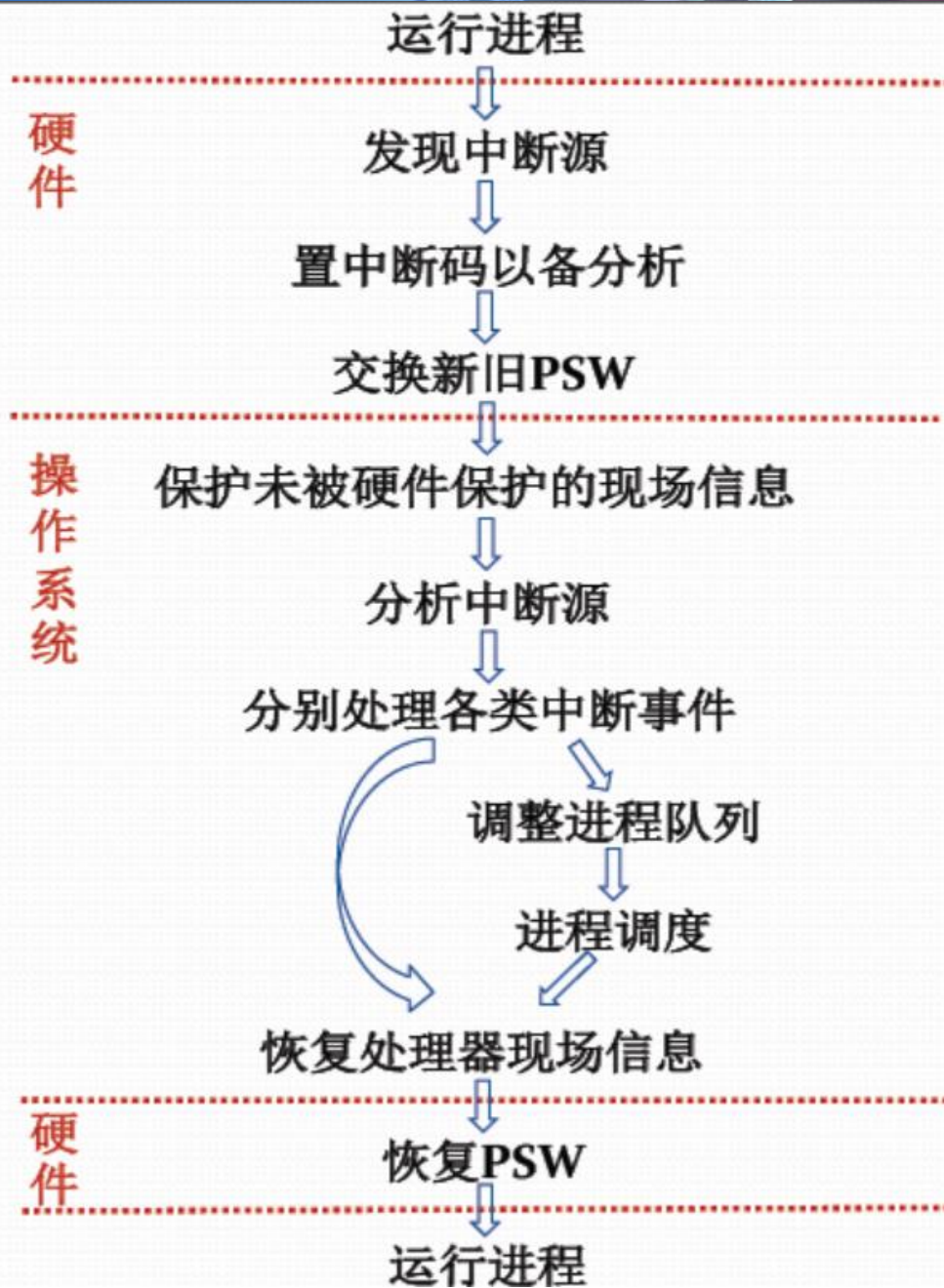
● 中断处理程序

- 操作系统处理中断事件的控制程序，主要任务是处理中断事件和恢复正常操作

● 中断处理过程

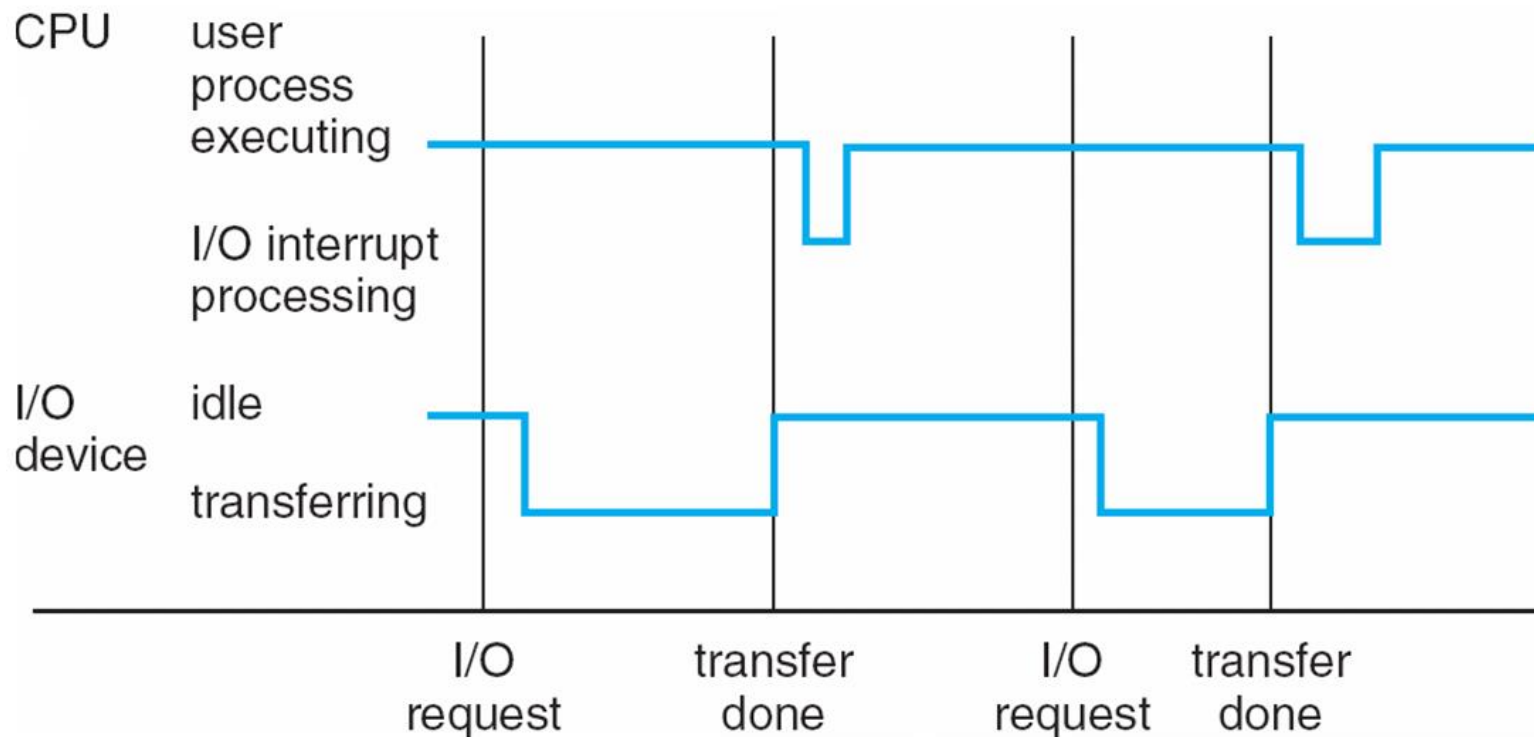
- 保护未被硬件保护的处理器状态
- 通过分析被中断进程的PSW中断码字段，识别中断源
- 处理发生的中断事件
- 恢复正常操作

中断系统处理流程





单个进程执行输出的中断时间线



作业题

A processor is executing code and executes a divide instruction which divides by zero.

(a) Describe in the MIPS instruction set what state needs to be saved by the hardware interrupt mechanism.

PC of the interrupted instruction \rightarrow EPC (the Exceptional Program Counter), if we are in a branch delay slot, $EPC \leftarrow PC - 4$

- The cause register gets loaded with the reason for the interrupt
- The Exception Level status register is set to '1' to disable further interrupts, $EXL \leftarrow 1$

(b) Assume that the interrupt handler reads registers R5, R6, R7 and writes registers R5, R8, R10, what registers does the interrupt handler need to save.

- R5, R8, R10

(c) What state does the ERET instruction change? $PC \leftarrow EPC, EXL \leftarrow 0$

直接内存访问DMA方式

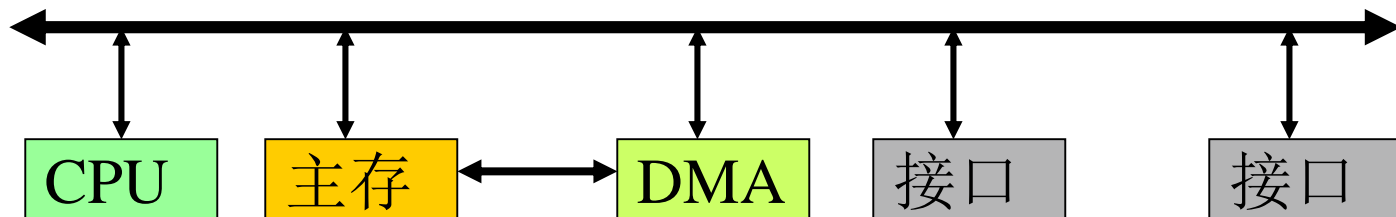


- 用于成组交换数据的场合
- 中断方式不适合于成组数据交换
- 由硬件执行I/O交换的工作方式
- DMA控制器从CPU接管总线控制，数据交换无须CPU干预，直接在内存以及外围设备之间进行，节约了中断开销
- 需要更多硬件

DMA基本概念



- 中断方式每传送一个字或一个字节的数据就要执行一次中断服务程序，数据传送时仍然要占用CPU的时间，不适合于高速传输的系统。
- DMA在外设与主存之间建立一个由硬件管理的数据通路，使CPU不介入传送时的操作，数据也不经过CPU。减少了CPU的开销，效率提高。



内存争用



- DMA方式进行数据传送时，CPU仍执行主程序，此时DMA控制逻辑与CPU可能同时要访问主存，引起主存使用权的冲突。
- 如何处理这种冲突呢？
 - 停止CPU使用主存
 - DMA与CPU交替使用主存
 - 周期挪用法

停止CPU使用主存

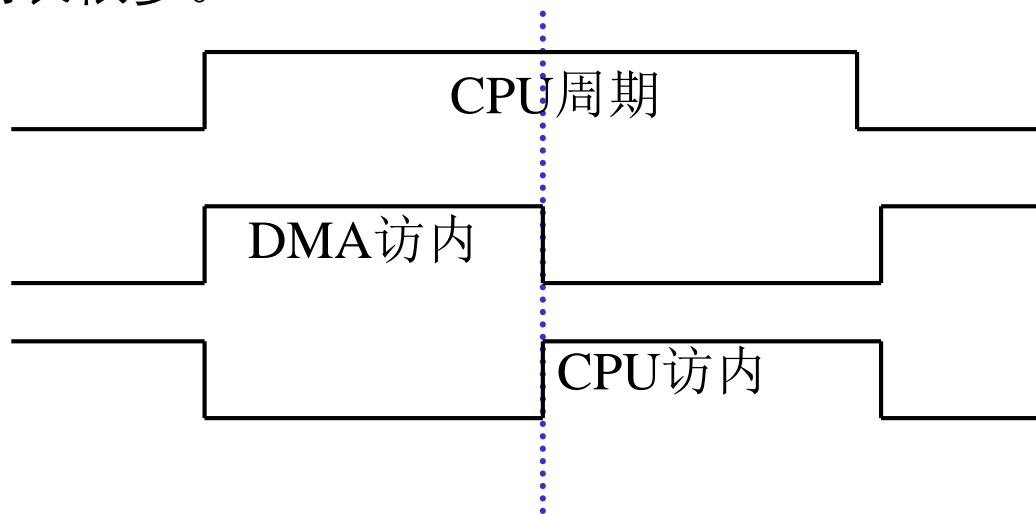


- 当DMA传送数据时，CPU停止工作，把主存使用权交给DMA控制逻辑。
- 在DMA传送数据过程中，CPU处于等待状态
- 在这批数据传送结束后，DMA再交还主存使用权
- 方法简单

DMA与CPU交替使用主存



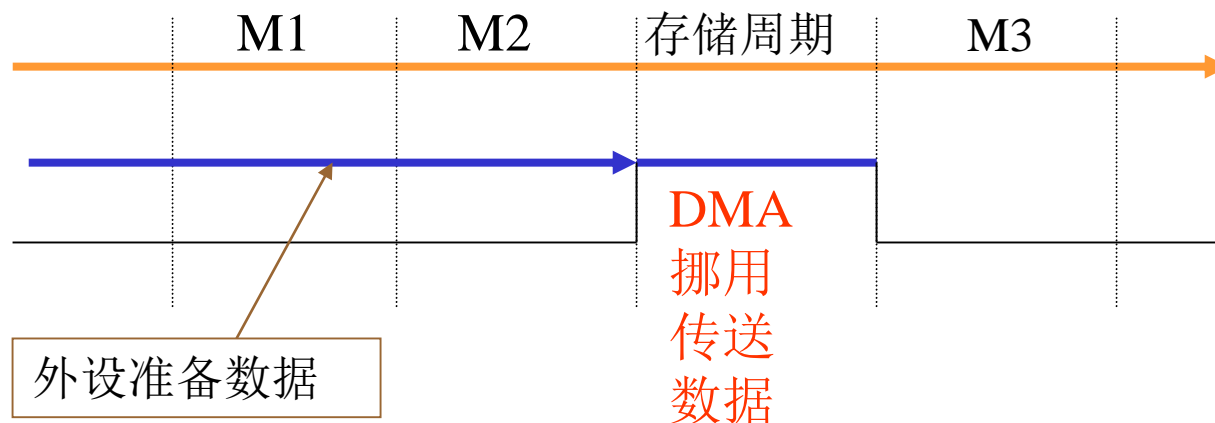
- 每个CPU工作周期分成两个时间段，一段用于 DMA访问主存，一段用于CPU访问主存。
- CPU和DMA都能访问主存，没有主存使用权移交过程，所以这种方式的效率较高，但是硬件投资较大，而且要求 CPU工作周期比存储周期长很多。



周期挪用法



- DMA要求访问主存时，CPU暂停**一个或多个存储周期**。一个数据传送结束后，CPU继续运行。
- CPU现场并没有变动，只是延缓了对指令的执行，这种技术称为**周期挪用**，或称周期窃取。
- 如发生访存冲突，则DMA优先访问。



DMA主要操作过程（准备阶段）



- 主机通过指令向DMA接口发送必要的传送参数，并启动DMA工作。这些参数的作用是
 1. 指出数据传送的**方向**，即指出是主机将数据传送给设备，还是由设备传给主机。
 2. 指出数据块在**主存的首地址**。
 3. 指出数据在外设**存储介质上的地址**。
 4. 指出数据的**传送量**。

DMA主要操作过程（传送阶段）



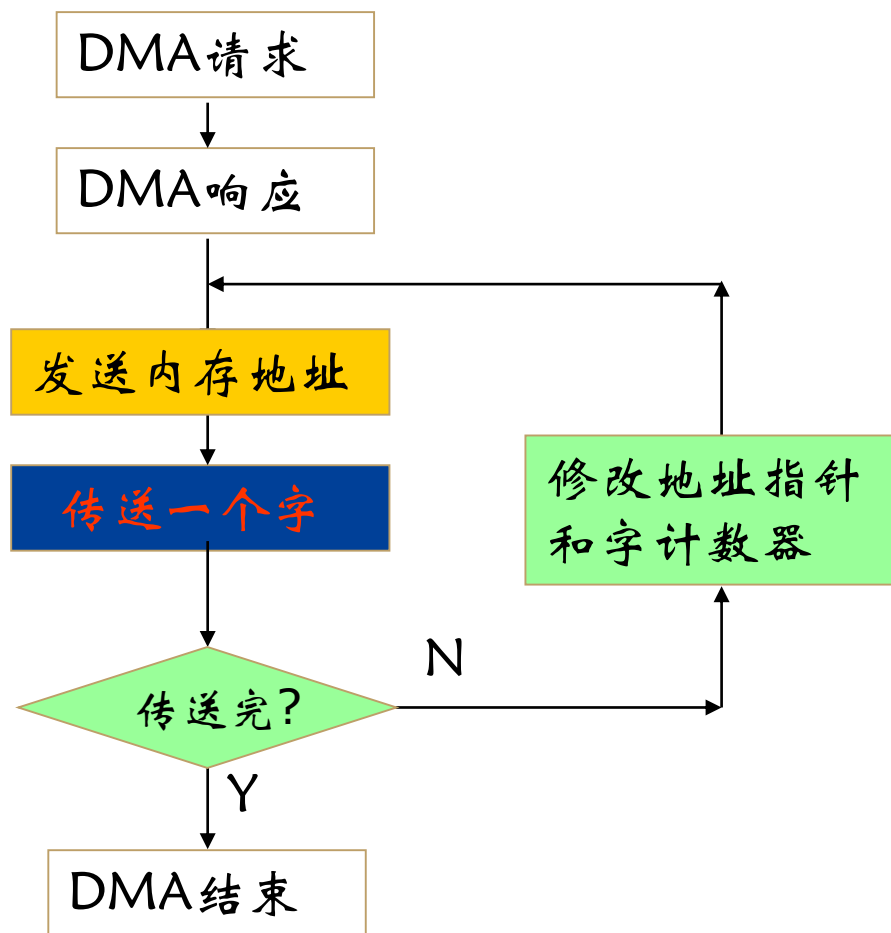
- 从宏观上看DMA接口连续传送一批数据。从微观上看，每传送一个数据，发一次DMA请求，经历一个循环操作。循环操作的主要内容有：
 1. 设备准备好一个数据的传送后，接口向主机发出DMA请求。
 2. CPU中DMA控制逻辑响应DMA请求，将主存使用权让给DMA。DMA送来主存单元地址、读或写命令。若是写命令，DMA还送来数据。
 3. DMA挪用一周对主存进行读写操作。周期挪用结束后给DMA接口应答信号。
 4. DMA接口接到应答信号，撤除DMA请求，将主存数据缓冲区地址指示器加1，指向下一个数据要传送的单元地址，字数计数器减1。
 5. DMA接口控制逻辑判断数据是否全部传送完。若传送完毕，则进入结束阶段；若没传完则又开始下一个循环操作。

DMA主要操作过程（结束阶段）



- DMA在两种情况下都进入结束阶段。
- 正常结束，一批数据传送完毕；
- 非正常结束，DMA发生了故障，也要进入结束阶段
- 不论是哪一种情况进入结束阶段，DMA都向主机发出中断请求，CPU执行服务程序，查询DMA接口状态，根据状态进行不同处理。

一个数据块的传送过程

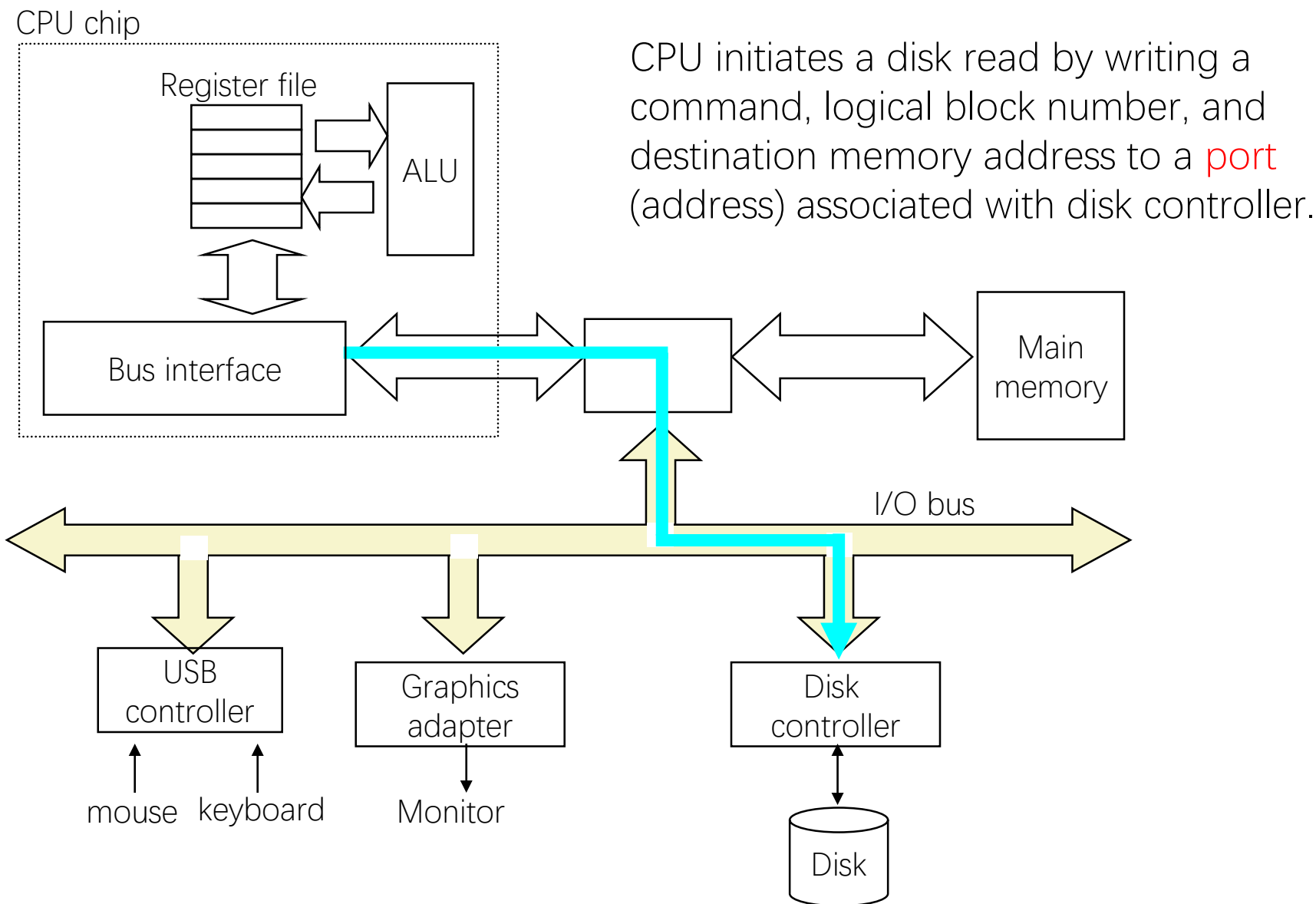


DMA与程序中中断的区别

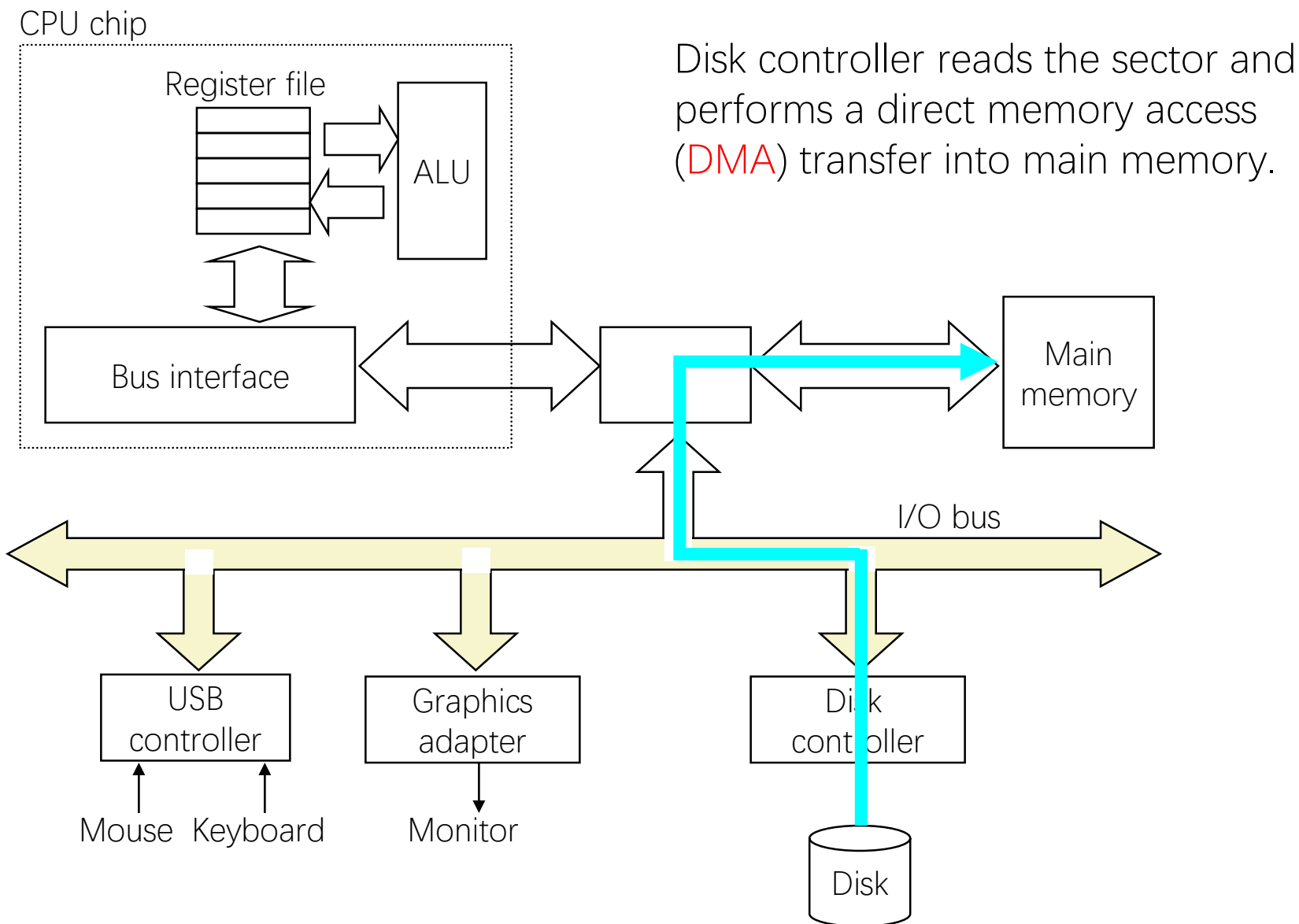


- 中断通过程序实现数据传送，而DMA靠硬件来实现。
- 中断时机为两条指令之间的空操作，DMA响应时机为两个存储周期之间。
- 中断不仅具有数据传送能力，还能处理异常事件。DMA只能进行数据传送。
- DMA仅挪用了一个存储周期，不改变CPU现场。
- DMA请求的优先权比中断请求高。CPU优先响应DMA请求，是为了避免DMA所连接的高速外设丢失数据。
- DMA利用了中断技术

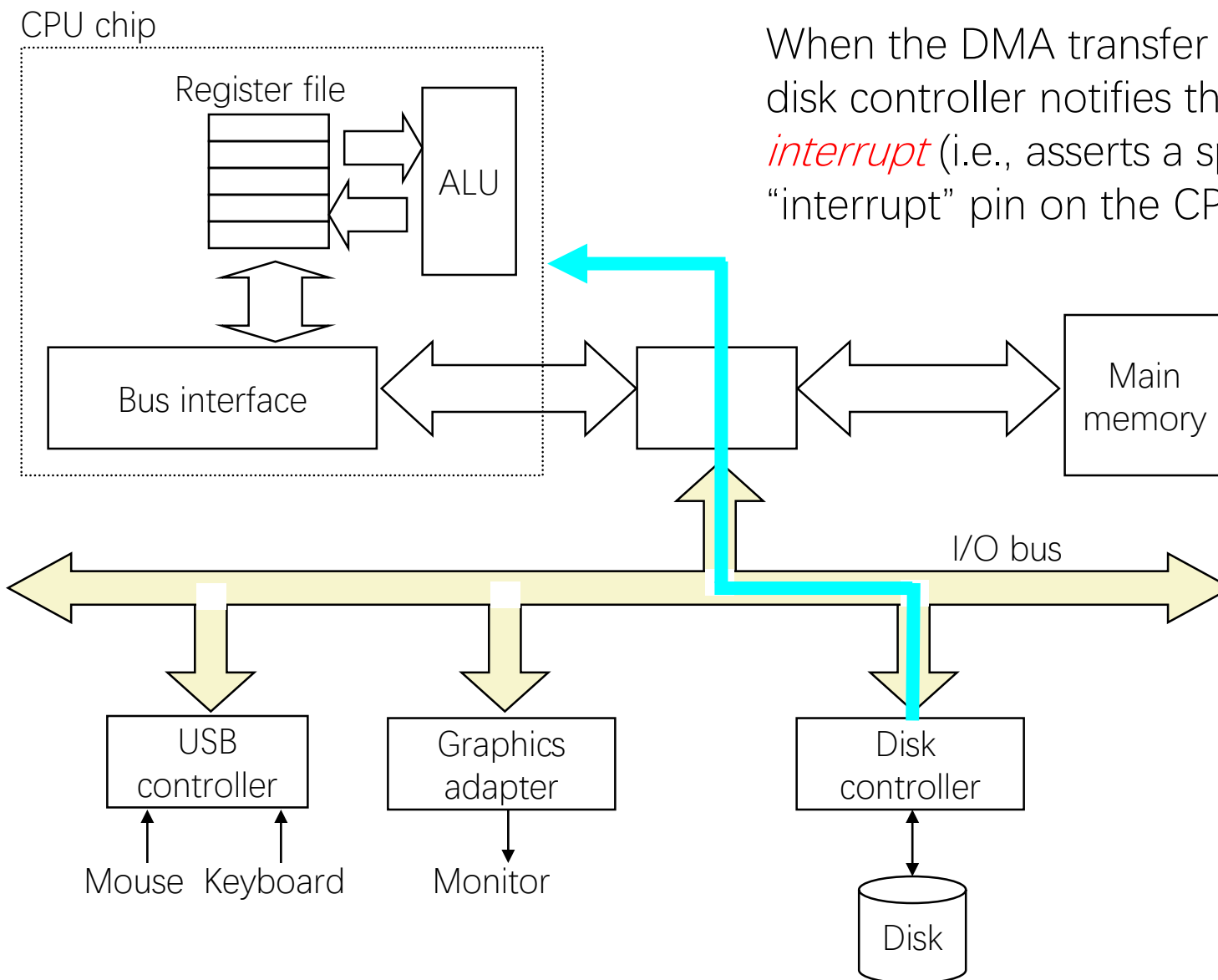
Reading a Disk Sector (1)



Reading a Disk Sector (2)



Reading a Disk Sector (3)



When the DMA transfer completes, the disk controller notifies the CPU with an *interrupt* (i.e., asserts a special “interrupt” pin on the CPU)

谢谢！

