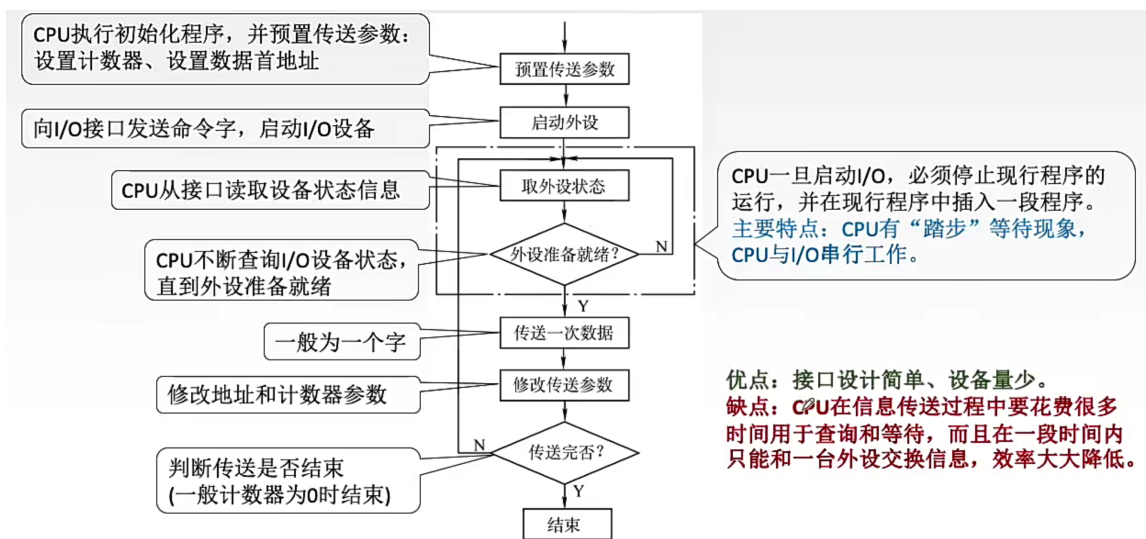


Lecture 27: 知识 - IO方式

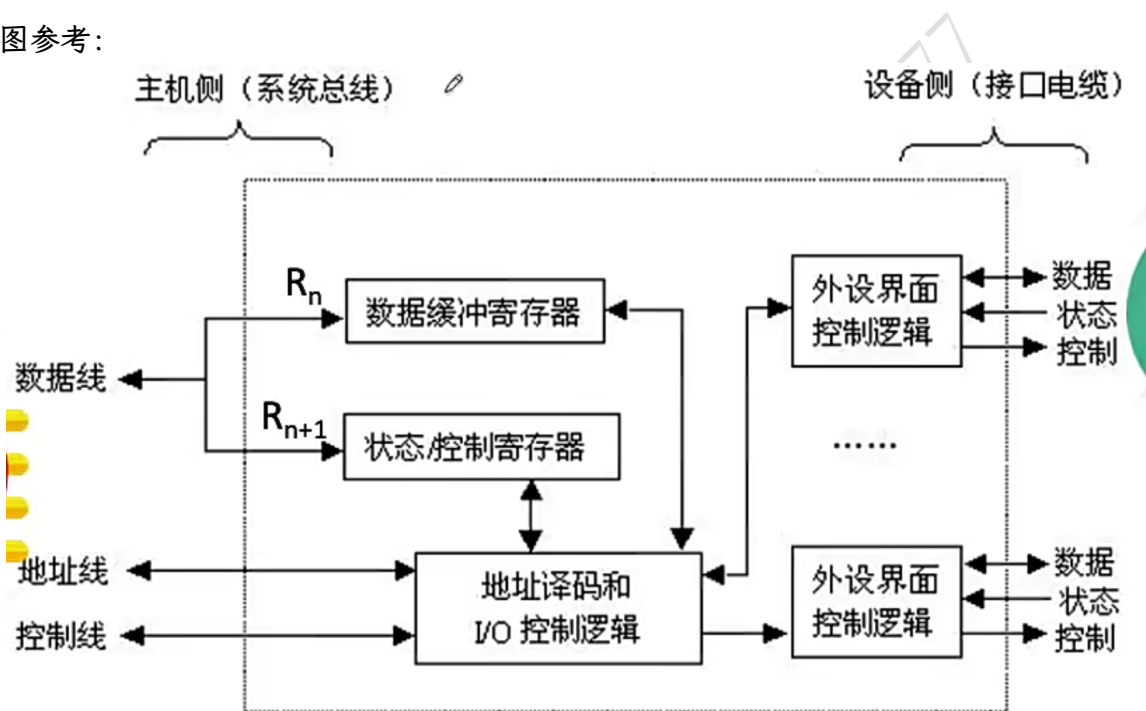
程序查询方式

概念： 程序查询方式

- 概念：
 - CPU 和 IO 设备 **串行通信**；
 - CPU 不断地 **轮询** 状态寄存器；
 - CPU 在信息传送过程中要花费很多时间用于查询和等待，而且在一段时间内只能和一台外设交换信息，效率大大降低；
- 指令：
 - IN Rd, Rs : 把 IO 端口 Rs 的数据输入到 CPU 寄存器 Rd 中；
 - OUT Rd, Rs : 把 CPU 寄存器 Rs 的数据输出到 IO 端口 Rd 中；
- 过程：打印机 IO
 - CPU 发出打印控制命令；
 - 把命令字输出到状态/控制寄存器；
 - 打印字启动；
 - 通过状态信息，存到状态寄存器当中、让 CPU 知道当前打印机的状态；
 - CPU 一直不断的轮询数据线：不断的使用 IN 指令，获得状态寄存器当中的值；
 - CPU 在一次轮询后、知道打印机已经就绪，此时可以通过地址线和控制线知道位置；
 - CPU 通过 OUT 指令，将数据写入 Rn
- 查询
 - 两种查询方式：
 - (1) 独占查询；
 - CPU 100% 的时间都在查询 IO 状态，完全串行；
 - (2) 定时查询；
 - 在保证数据不丢失的情况下，每隔一段时间 CPU 就查询一次 I/O 状态；
 - 查询的间隔内 CPU 可以执行其他程序；
- 图示：
 - 流程图参考：



- 结构图参考:



分析: 真题: 2018年

- 题目:

【2018 统考真题】假定计算机的主频为 500MHz，CPI 为 4。现有设备 A 和 B，其数据传输速率分别为 2MB/s 和 40MB/s，对应 I/O 接口中各有一个 32 位数据缓冲寄存器。回答下列问题，要求给出计算过程。

- 1) 若设备 A 采用定时查询 I/O 方式，每次输入/输出都至少执行 10 条指令。设备 A 最多间隔多长时间查询一次才能不丢失数据？CPU 用于设备 A 输入/输出的时间占 CPU 总时间的百分比至少是多少？

- 解析:

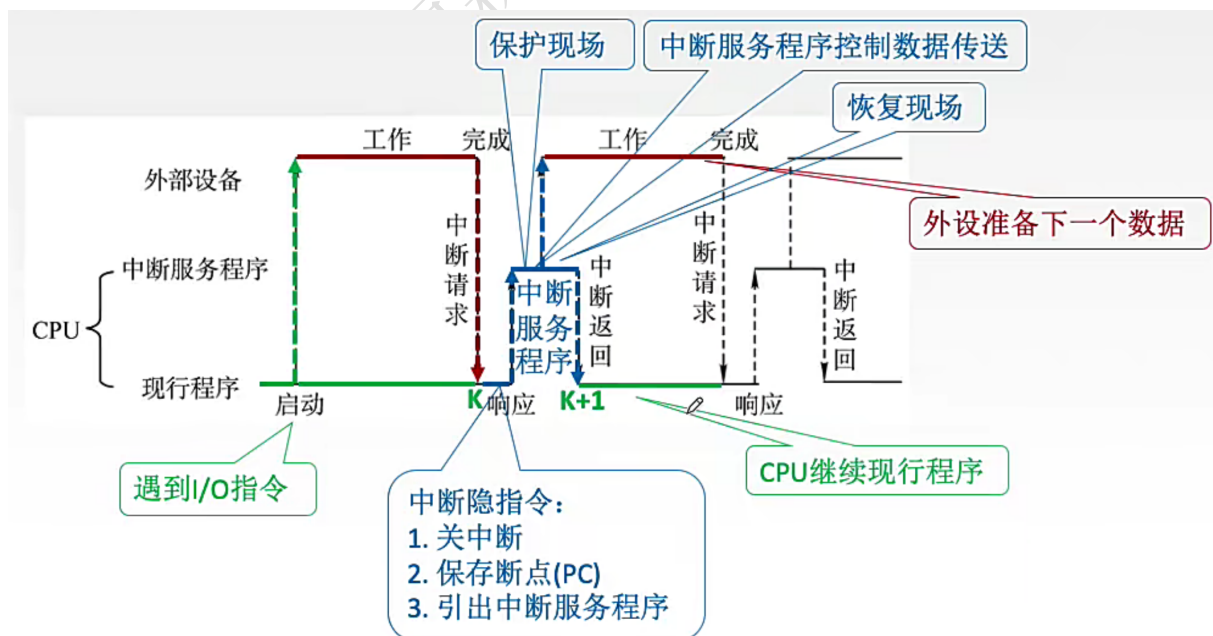
- 当设备 A 采用定时查询的方式时，它需要在**传输完端口大小**的数据后进行访问操作；
- 在准备 4B 数据的时候，设备 A 需要的时间为 $32 \text{ bit} / 2 \text{ MB/s} = 2 \mu\text{s}$
- 因此，设备 A 每隔 2 μs 必须进行一次查询操作；
- 同时，CPU 在进行设备 A 的输入输出时，每次输入输出需要执行 10 条指令，并且每条指令 CPI 为 4，所以 CPU 进行设备 A 的输入输出的总时钟周期数为：
 $5 \times 10^5 \times 4 \times 10 = 2 \times 10^7$

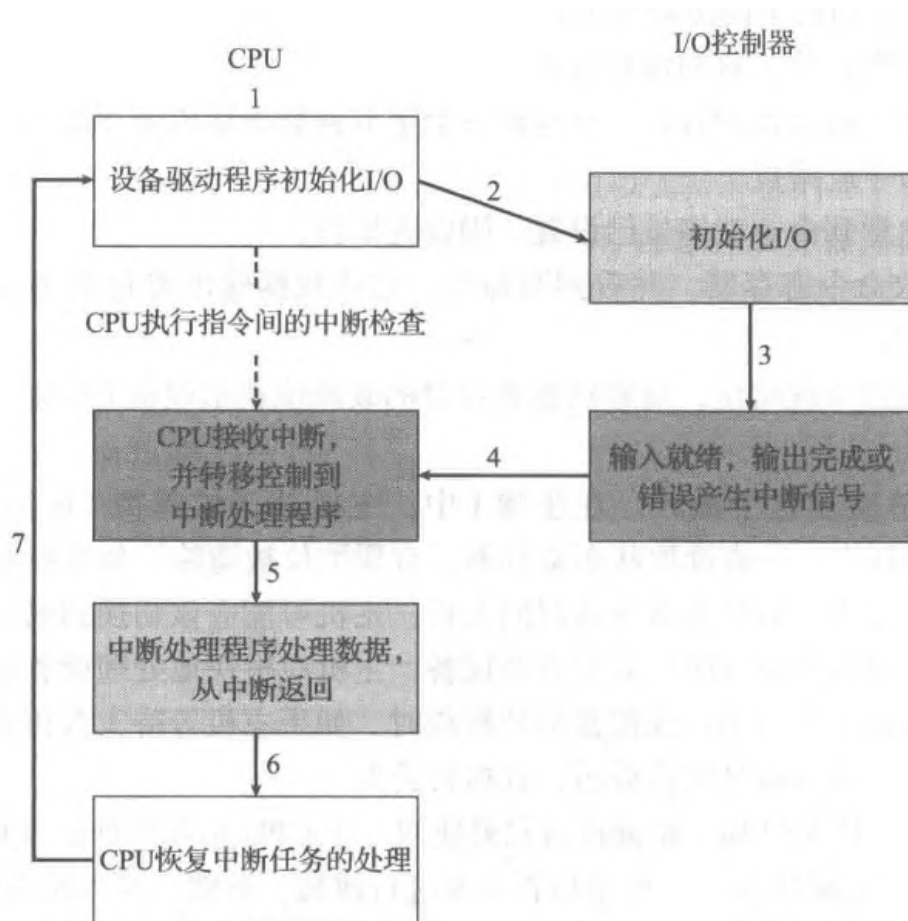
- 占 CPU 总时间的百分比至少为 4%;

中断 IO 方式

概念: IO 中断方式概述

- 概念:
 - 等待键盘 IO 时, CPU 可以先去执行其他程序, 键盘 I/O 完成后 I/O 控制器向 CPU 发出中断请求, CPU 响应中断请求, 并取走输入数据;
- 特点:
 - (1) 中断 IO 方式仅适用于慢速设备;
 - CPU 利用率严重下降, CPU 需要花大量的时间来处理中断服务程序, 尤其是对于磁盘一类的 IO 设备;
 - (2) 中断 IO 方式下的数据传输通过软件实现;
 - (3) 中断 IO 方式, CPU 和 IO 端口的数据中不存在主存地址数据;
 - 在中断方式中, 输出数据所在的内存地址由申请 IO 的程序自己得到, 主存地址不会传送到 I/O 接口;
 - (4) CPU 只有在终端运行的状态时, 才能响应外部设备的中断请求;
- 过程:
 1. 当外设准备好时, 便向 CPU 发中断请求;
 2. CPU 响应后, 中止现程序的执行, 转入一个 中断服务程序 进行输入/出操作, 实现主机和外设接口之间的数据传送, 并启动外设工作;
 3. 中断服务程序 执行完后, 返回原被中止的程序断点处继续执行;
- 图示:





DMA 方式

DMA 基础概念

概念： DMA 基础概念

- 概念：

- DMA：Direct Memory Access
- DMA 接口也是一种 IO 接口，主存与高速 IO 设备之间有条直接数据通路了 DMA 总线；
- CPU 向 DMA 接口发出“读/写”命令，并指明 主存地址、磁盘地址、读写数据量 等参数，DMA 控制器自动控制磁盘与主存的数据读写；

- 实现：

- 请求传输

- DMA 控制器与 设备控制器 之间的握手，通过一对称为 DMA 请求 (DMA-request) 和 DMA 确认 (DMA-acknowledge) 的线路来进行；
 - 当有数据需要传输时，设备控制器发送信号到 DMA 请求线路，这个信号使得 DMA 控制器向 CPU 发送总线使用权的请求，占用内存总线，发送所需地址到内存地址总线，并发送信号到 DMA 确认线路；
 - 当 设备控制器 收到 DMA 确认信号时，它就传输数据到内存，并且清除 DMA 请求信号；

- 传输结束

- 当完成整个传输时，DMA 控制器中断 CPU；

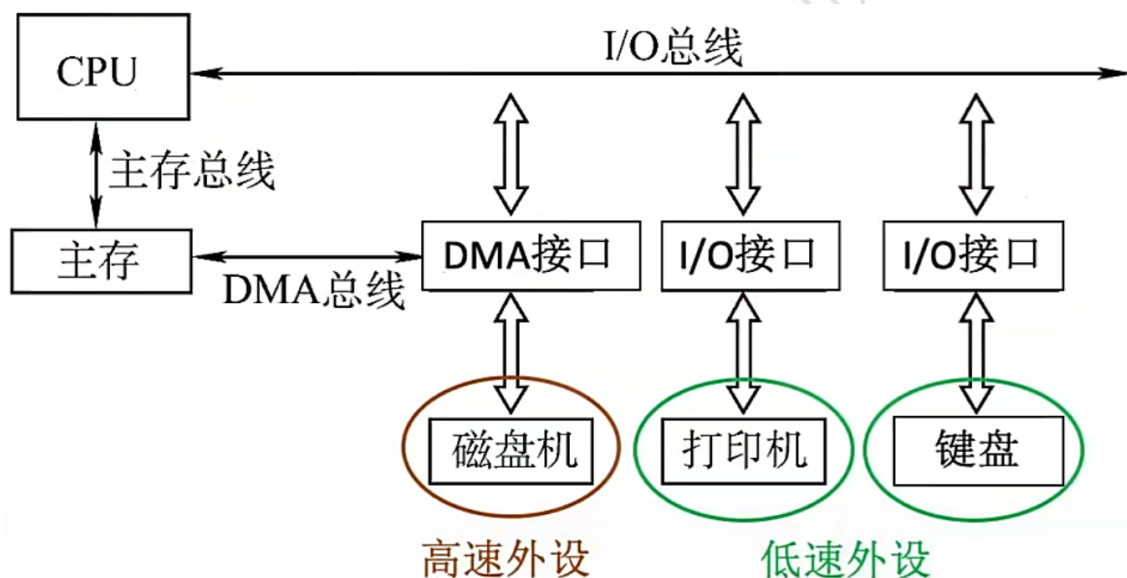
- 当 DMA 控制器占用内存总线时，CPU 被暂时阻止访问内存，但是仍然可以访问主缓存或辅助缓存内的数据项；
 - 通过周期挪用技术 (cycle stealing) 可能会减慢 CPU 计算，但是将数据传输工作交给 DMA 控制器通常能够改进总的系统性能；
- 有的计算机架构的 DMA 采用物理内存地址，而其他采用直接虚拟内存访问 (Direct Virtual-Memory Access, DVMA)；
- 这里所用的虚拟内存地址需要虚拟到物理地址的转换，DVMA 可以直接实现两个内存映射设备之间的传输，而无需 CPU 的干涉或采用内存；

• 特点：

- (1) DMA 方式向 CPU 申请的是**总线使用权**；
- (2) DMA 方式下的数据传输由**硬件**完成；
- (3) 每完成**一整块**数据读写，才向 CPU 发出一次**中断请求**；

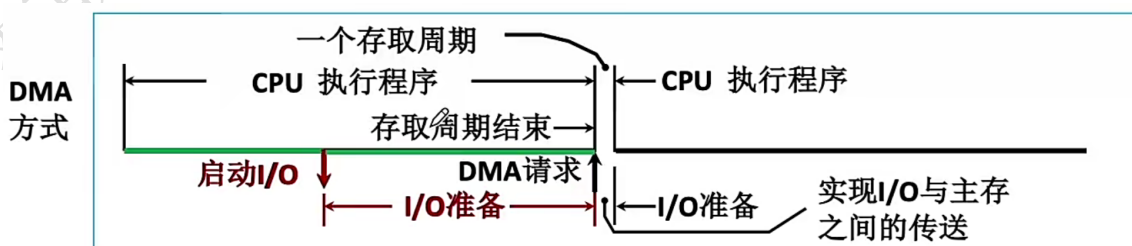
• 图示：

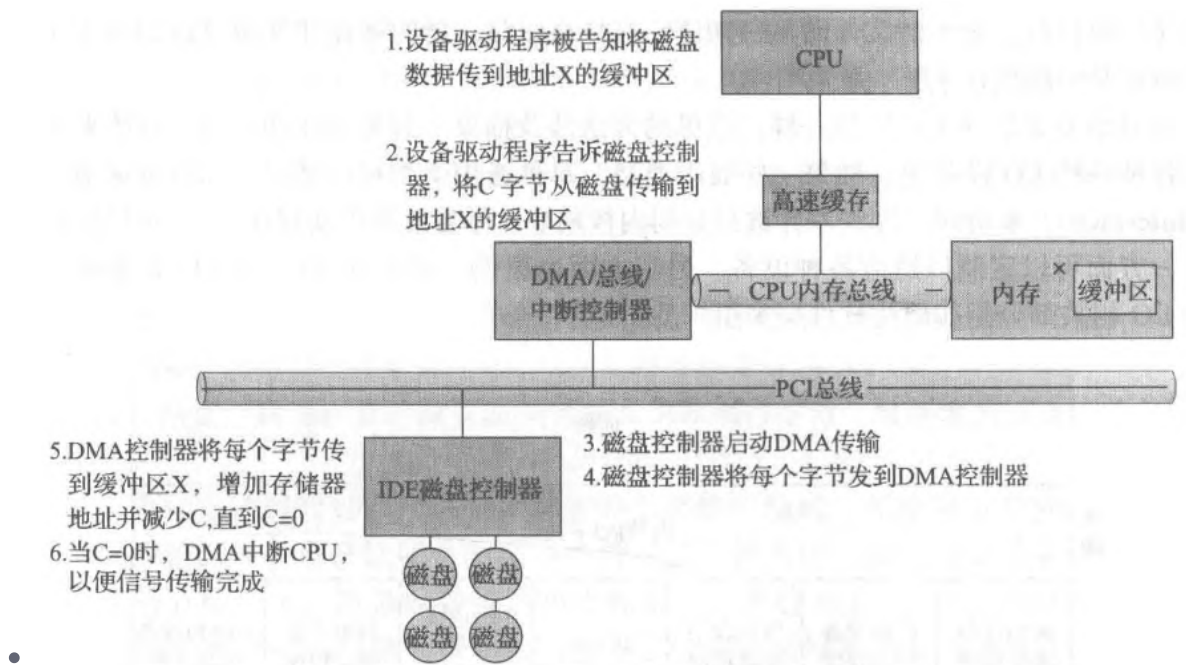
- DMA 关系



注：DMA接口，即DMA控制器，也是一种特殊的I/O控制器

- DMA 过程：





概念：DMA 特点

- 1. 它使主存与 CPU 的固定联系脱钩，主存既可被 CPU 访问，又可被外设访问；
- 2. 在数据块传送时，**主存地址的确定、传送数据的计数**等都由**硬件电路**直接实现，主存中要开辟专用缓冲区，及时供给和接收外设的数据；
- 3. DMA 传送速度快，CPU 和**外设并行工作**，提高了系统效率；
- 4. DMA 在传送开始前要通过**程序**进行预处理，结束后要通过**中断方式**进行后处理；

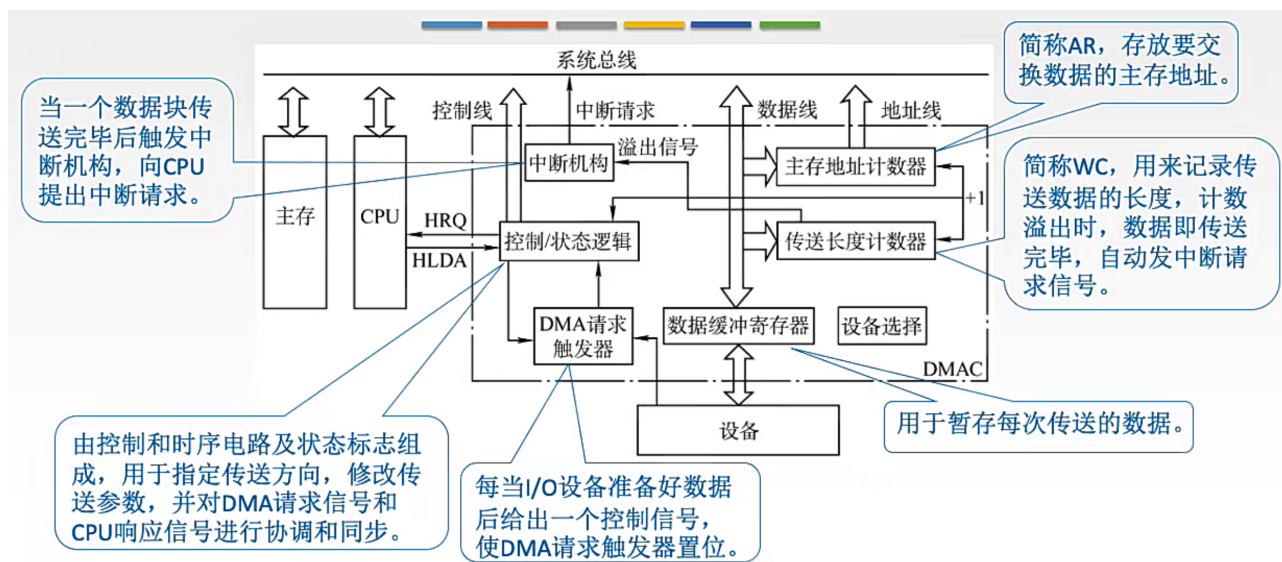
概念：DMA 方式与中断方式对比

- 图示：

	中断	DMA
数据传送	程序控制 程序的切换 → 保存和恢复现场	硬件控制 CPU只需进行预处理和后处理
中断请求	传送数据	后处理
响应	指令执行周期结束后响应中断	每个机器周期结束均可，总线空闲时即可响应DMA请求
场景	CPU控制，低速设备	DMA控制器控制，高速设备
优先级	优先级低于DMA	优先级高于中断
异常处理	能处理异常事件	仅传送数据

- 真题：2013年
 - 程序中断方式申请的是 CPU 的处理时间，DMA 方式申请的是总线的使用权；
 - 中断响应发生在一条指令指向结束后，而 DMA 响应发生在一个总线事务之后（即总线空闲后）；

概念：DMA 结构



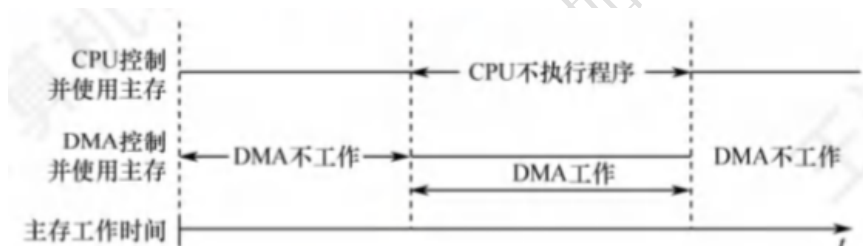
DMA 传送方式

概念： 方式一 - 停止 CPU 访存

• 概念：

- 当 I/O 设备有 DMA 请求时，由 DMA 接口向 CPU 发送一个停止信号，使 CPU 放弃总线控制权，停止访问主存，直到 DMA 传送一块数据结束；
- 数据传送结束后，DMA 接口通知 CPU 可以使用主存，并把总线控制权交回给 CPU；

• 图示：



概念： 方式二 - 周期挪用

• 概念：

- 由于 I/O 访存的优先级高于 CPU 访存 (因为 I/O 若不立即访存，就可能丢失数据)，因此由 DMA 的 I/O 设备挪用一個存取周期，传送完一个数据字后立即释放总线，准备好数据后继续占用总线、传输数据至内存，传输完成后释放总线，等待 I/O 设备准备数据；
- 周期挪用是一种单字传送方式。当 I/O 设备有 DMA 请求时，会遇到 3 种情况：
 - 1. 此时 CPU 不在访存，因此 I/O 的访存请求与 CPU 未发生冲突；
 - 2. CPU 正在访存，此时必须待存取周期结束后，CPU 再将总线占有权让出；
 - 3. I/O 和 CPU 同时请求访存，出现访存冲突，此时 CPU 要暂时放弃总线占有权；

• 图示：



• 真题：2012年

• 题目：

- 假定某计算机的 CPU 主频为 80MHz，CPI 为 4，平均每条指令访存 1.5 次，主存与 Cache 之间交换的块大小为 16B，Cache 的命中率为 99%，存储器总线宽度为 32 位，回答下列问题：
- (1) 假定在 Cache 缺失的情况下访问主存时，存在 0.0005% 的缺页率，则 CPU 平均每秒产生多少次缺页异常？若页面大小为 4 KB，每次缺页都需要访问磁盘，访问磁盘时 DMA 传送采用周期挪用方式，磁盘 IO 接口的数据缓冲寄存器为 32 位，则磁盘 IO 接口平均每秒发出的 DMA 请求次数至少是多少？
- (2) CPU 和 DMA 控制器同时要求使用存储器总线时，哪个优先级更高？为什么？
- (3) 为了提高性能，主存采用 4 体低位交叉存储模式，工作时每四分之一一个存储周期启动一个体。若每个体的存储周期为 50 ns，则该主存能提供的最大带宽是多少？

• 分析：

• 第一问：

- 由于 CPU 主频为 80 M/s，而且 CPI 为 4，因此每秒可以执行 20 M 条指令；
- 因为平均每条指令访问 1.5 次，因此一秒内访存 30 M 次；
- 因为 Cache 缺失率为 1%，因此一秒内 Cache 缺失的次数为 $30M \times 0.01 = 300K$ 次；
- 因为在 Cache 后、需要访问主存，而主存访问时又有 0.0005% 的缺页率，因此每秒的缺页次数为：
 - $300K \times 0.0005\% = 1.5 \text{次}$
- 由此可知，平均每秒产生 1.5 次缺页异常；
- 因为页面的大小为 4 KB，因此当一个页面缺失时，需要从磁盘中调入 4 KB 大小的页面；
- 而磁盘的数据缓存寄存器为 32 位，而且当前采用周期挪用技术，每次传输单字的数据，即 4B，因此 4KB 的页面需要发出 DMA 请求次数为 $4KB / 4B = 1024$ 次；
- 而一秒内有 1.5 次缺页，因此一秒内的 DMA 请求次数为 $1024 \times 1.5 = 1536 \text{次}$ ；

• 第二问：

- IO 访存的优先级高于 CPU 访存；
- 因为 IO 若不立即访存，就可能丢失数据，因此由 DMA 的 IO 设备挪用一个存取周期，传送完一个数据字后立即释放总线，准备好数据后继续占用总线、传输数据至内存，传输完成后释放总线，等待 IO 设备准备数据；

• 第三问：

- 单个存储器在一个存储周期之内能取出的数据大小，为当前的存储器总线宽度，而存储器总线宽度为 32 位，因此一个存储器在一个存储周期内可以取出 4B 的数据；

- 已知每个存储器的存储周期为 50 ns，因此其在一秒内可以存储的数据为
 $4B/50ns=80MB/s$
- 由于当前采用 4 体低位交叉存储模式，工作时每四分之一一个存储周期启动一个体，因此在一个存储周期内、会按照流水线的方式启动四个存储器。当完全流水线化后，其单个存储周期内可以理解为进行了四次存储，因此一秒内可以存储的数据乘以 4，即 $80MB/s * 4 = 320MB/s$

• 真题：2020年

• 题目：

【2020 统考真题】若设备采用周期挪用 DMA 方式进行输入和输出，每次 DMA 传送的数据块大小为 512 字节，相应的 I/O 接口中有一个 32 位数据缓冲寄存器。对于数据输入过程，下列叙述中，错误的是（ ）。

- A. 每准备好 32 位数据，DMA 控制器就发出一次总线请求
- B. 相对于 CPU，DMA 控制器的总线使用权的优先级更高
- C. 在整个数据块的传送过程中，CPU 不可以访问主存储器
- D. 数据块传送结束时，会产生“DMA 传送结束”中断请求

• 分析：

- 由于 IO 接口中用于缓冲数据的 IO 端口大小为 32 位，因此每准备好 32 位数据，DMA 的 IO 接口就发出一次总线请求，A 正确；
- 周期挪用法由 DMA 控制器挪用一個或几个主存周期来访问主存，传送完一个数据字后立即释放总线，是一种单字传送方式，每个字传送完后 CPU 可以访问主存，因此 C 错误；

• 答案：

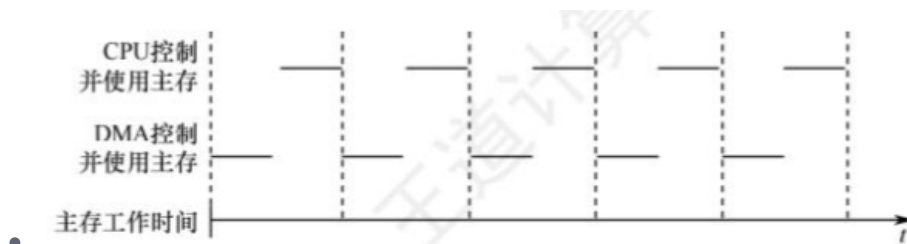
- C

概念：方式三 - 交替使用

• 概念：

- 将 CPU 的工作周期分成两个时间片，一个给 CPU 访存，另一个给 DMA 访存，这样在每个 CPU 周期内，CPU 和 DMA 就都可以轮流访存；
- 这种方式适用于 CPU 的工作周期比主存取取周期长的情况。例如，若 CPU 的工作周期是 1.2 μs ，主存的存取周期小于 0.6 μs ，则可将一个 CPU 周期分为两个周期，一个专供 DMA 访存，另一个专供 CPU 访存；
- 这种方式不需要申请、建立和归还总线使用权，总线使用权是通过分时控制的；

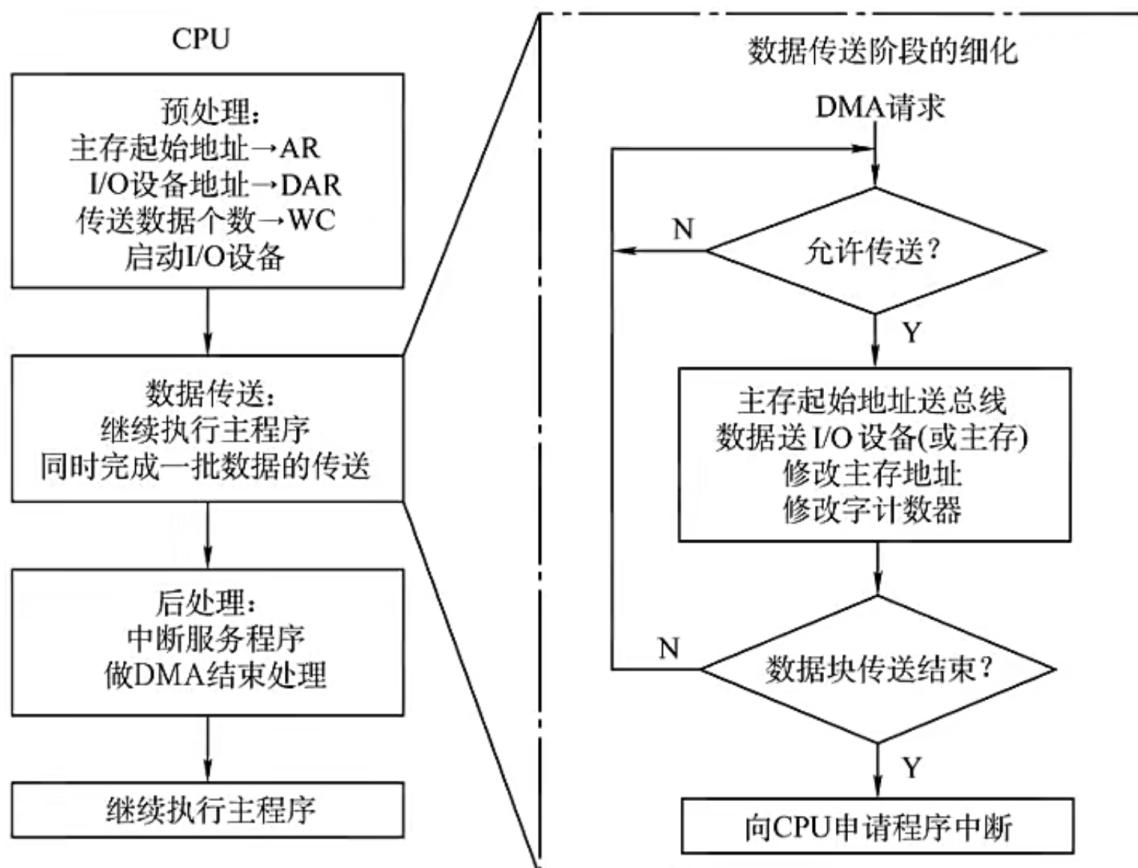
• 图示：



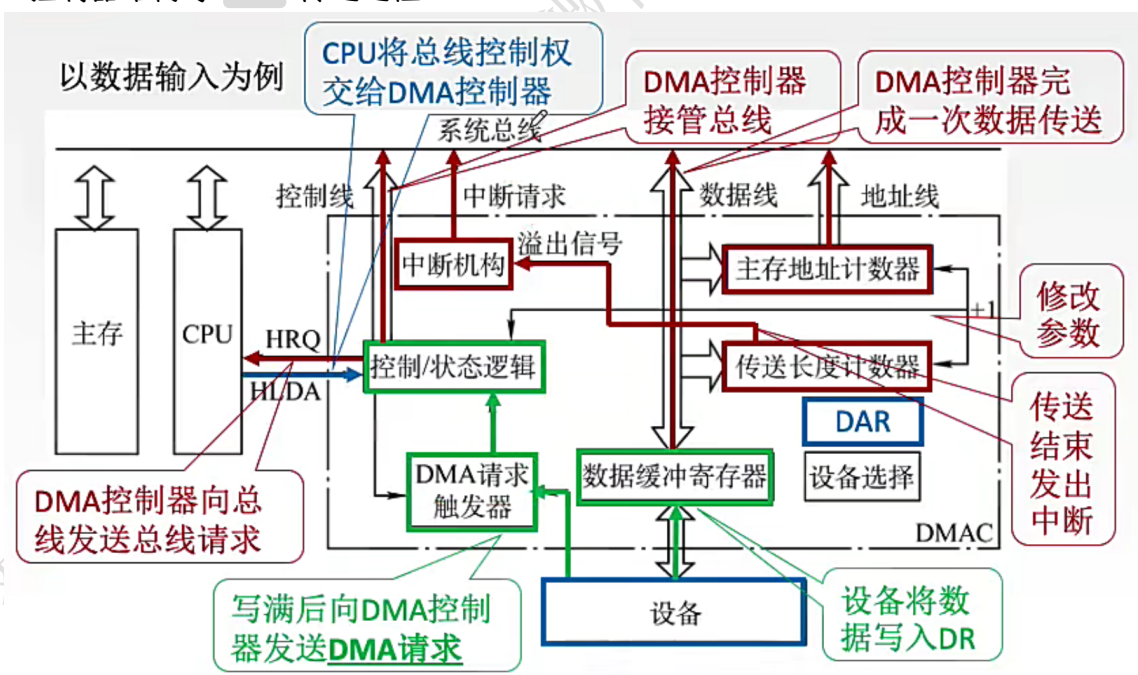
DMA 工作流程

概念：DMA 工作流程

- 概念：
 - CPU 向 DMA 控制器指明：
 - (1) 要输入还是输出；
 - (2) 要传送多少个数据；
 - (3) 数据在主存、外设中的地址；
- 过程：
 - 真题：2017年
 - 传送前：
 - (1) DMA 传送前，由设备驱动程序设置传输参数；
 - (2) DMA 控制器向 CPU 发出总线请求，CPU 接受外设发出的 DMA 请求 (外设传送一个字的请求)；
 - 注意：是总线请求，不是中断请求；
 - (3) CPU 响应此总线请求，发出总线响应信号，DMA 接管总线控制权，进入 DMA 操作周期；
 - 传送时：
 - (3) 确定传送数据的主存单元地址及长度，并能自动修改主存地址计数和传送长度计数；
 - 真题：2019年
 - 注意：在数据传送中，总线可以直接由 DMA 控制器进行控制，不需要 CPU 参与；
 - (4) 规定数据在主存和外设间的传送方向，发出读写等控制信号，执行数据传送操作；
 - 结束后：
 - (5) DMA 控制器发出中断请求，向 CPU 报告 DMA 操作的结束；
 - (6) 执行 DMA 结束的中断服务程序；
- 图示：
 - 图示：传输过程流程图



• 图示：控制器结构与 DMA 传送过程



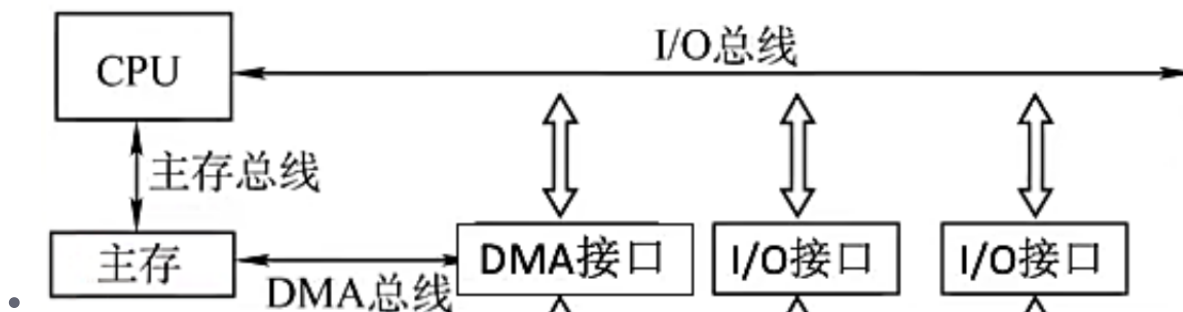
基于三总线的 DMA 控制

概念：基于三总线的 DMA 控制

- 概念：
 - **主存和 DMA 接口**之间有一条直接数据通路；
 - 由于 DMA 方式传送数据不需要经过 CPU，因此**不必中断**现程序，IO 与主机并行工作，**程序和传送并行**工作；

• 图示：

- 三总线结构

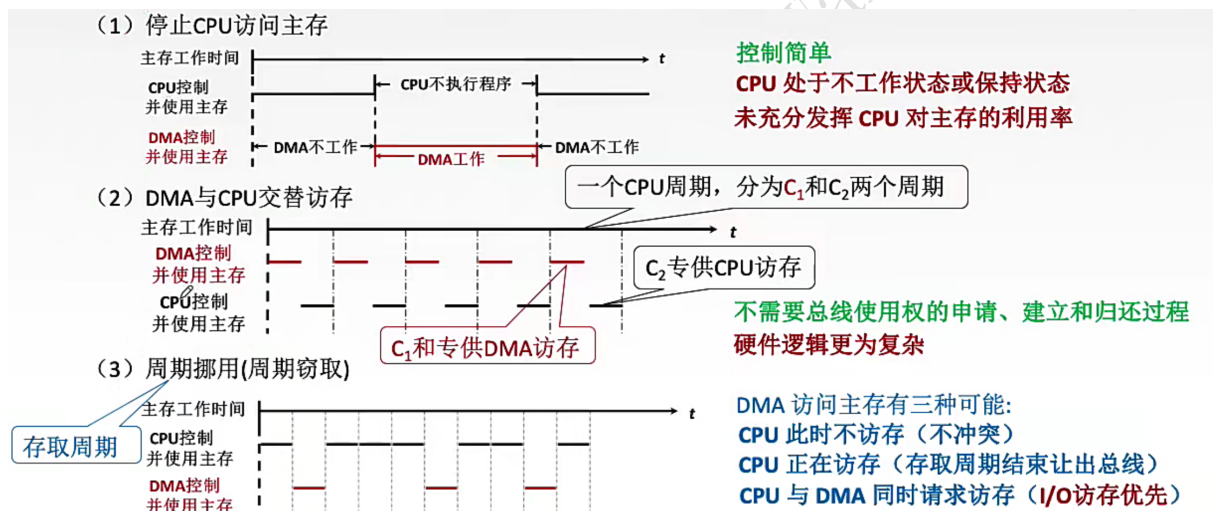


分析： I/O 设备与 CPU 访问主存冲突的处理

- 概念：

- 主存和 DMA 控制器之间有一条数据通路，因此主存和 I/O 设备之间交换信息时，不通过 CPU；
- 但当 I/O 设备和 CPU 同时访问主存时，可能发生冲突；
- 为了有效地使用主存，DMA 控制器与 CPU 通常采用以下 3 种方法使用主存；

- 方法：



通道控制方式

概念： 通道控制方式

- 原因：

- 有的商用中型机、大型机可能会接上超多的 I/O 设备，如果都 CPU 来管理，那么 CPU 就太累了；

- 概念：

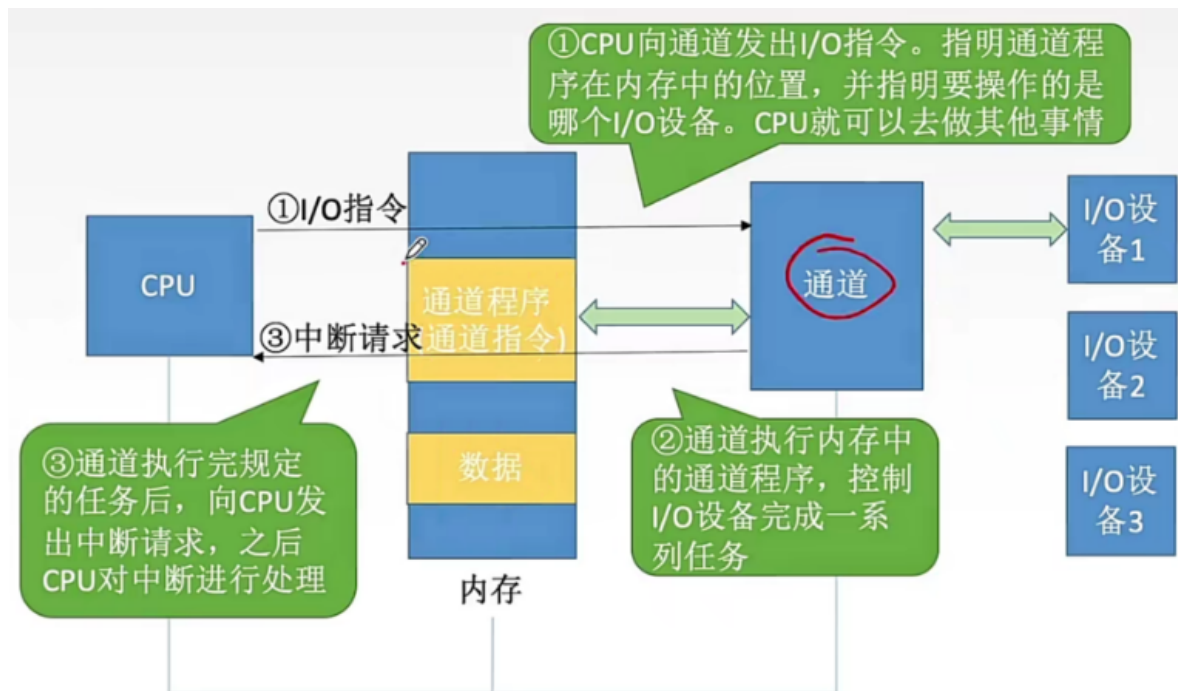
- 通道是具有特殊功能的处理器，能对 I/O 设备进行统一管理；
- 通道：可以理解为是能功能比较单一的、弱鸡版的 CPU；
- 通道可以识别并执行一系列通道指令，通道指令种类、功能比较单一，指令系统和 CPU 的不是一套 ISA；

- 过程：

- 1、CPU 向通道发出 I/O 指令，指明通道程序在内存中的位置，并指明要操作的是哪个 I/O 设备，CPU 就可以去做其他事情；

- 2、通道执行内存中的通道程序，控制 I/O 设备完成一系列任务 -> 每一条通道指令，对应了对一个 I/O 设备的控制；
- 3、通道执行完规定的任务后，向 CPU 发出中断请求，并由 CPU 进行中断处理；

• 图示：



获取笔记持续更新可加微信