

课时6 输入输出系统

高数帮



考点	重要程度	占分	题型
1. I / O系统基本概念	★★★	0 ~ 3	选择、填空
2. I / O方式	必考	6 ~ 10	大题

6.1 I / O 系统基本概念

1. 输入 / 输出系统

输入 / 输出是以主机为中心而言的，将信息从外部设备传送到主机称为**输入**，反之称为**输出**。

输入 / 输出系统解决的主要问题是各种形式的信息进行输入和输出的控制。



扫码听课，视频讲解更清晰

1.输入 / 输出系统

I / O系统中的几个基本概念如下：

- 1) **外部设备**。包括输入 / 输出设备及通过输入 / 输出接口才能访问的外存储设备。
- 2) **接口**。在各个外设与主机之间传输数据时进行各种协调工作的逻辑部件。协调包括传输过程中速度的匹配、电平和格式转换等。
- 3) **输入设备**。用于向计算机系统输入命令和文本、数据等信息的部件。键盘和鼠标是最基本的输入设备。
- 4) **输出设备**。用于将计算机系统内的信息输出到计算机外部进行显示、交换等的部件。显示器和打印机是最基本的输出设备。
- 5) **外存设备**。指除计算机内存及CPU缓存等外的存储器。硬磁盘、光盘等是最基本的外存设备。

1.输入 / 输出系统

一般来说，I / O系统由I / O软件和I / O硬件两部分构成：

- 1) **I / O软件**。包括驱动程序、用户程序、管理程序、升级补丁等。通常采用I / O指令和通道指令实现CPU与I / O设备的信息交换。
- 2) **I / O硬件**。包括外部设备、设备控制器和接口、I / O总线等。通过设备控制器来控制I / O设备的具体动作；通过I / O接口与主机（总线）相连。

2. I / O控制方式

在输入 / 输出系统中，经常需要进行大量的数据传输，而传输过程中有各种不同的I / O控制方式，基本的控制方式主要有以下4种：

- 1) 程序查询方式。由CPU通过程序不断查询I / O设备是否已做好准备，从而控制I / O设备与主机交换信息。
- 2) 程序中断方式。只在I / O设备准备就绪并向CPU发出中断请求时才予以响应。
- 3) DMA方式。主存和I / O设备之间有一条直接数据通路，当主存和I / O设备交换信息时，无须调用中断服务程序。
- 4) 通道方式。在系统中设有通道控制部件，每个通道都挂接若干外设，主机在执行I / O命令时，只需启动有关通道，通道将执行通道程序，从而完成I / O操作。

注：方式1) 和方式2) 主要用于数据传输率较低的外部设备，方式3) 和方式4) 主要用于数据传输率较高的设备。

6.2 I / O 方式

输入 / 输出系统实现主机与I / O设备之间的数据传送，可以采用不同的控制方式，各种方式在代价、性能、解决问题的着重点等方面各不相同，常用的I / O方式有程序查询、程序中断、DMA和通道等，其中前两种方式更依赖于CPU中程序指令的执行。

1. 程序查询方式

信息交换的控制完全由主机执行程序实现，程序查询方式接口中设置一个数据缓冲寄存器（数据端口）和一个设备状态寄存器（状态端口）。主机进行I / O操作时，先发出询问信号，读取设备的状态并根据设备状态决定下一步操作究竟是进行数据传送还是等待。

1.程序查询方式

程序查询方式的工作流程如下：

- ①CPU执行初始化程序，并预置传送参数。
- ②向I / O接口发出命令字，启动I / O设备。
- ③从外设接口读取其状态信息。
- ④CPU不断查询I / O设备状态，直到外设准备就绪。
- ⑤传送一次数据。
- ⑥修改地址和计数器参数。
- ⑦ 判断传送是否结束，若未结束转第③步，直到计数器为0。

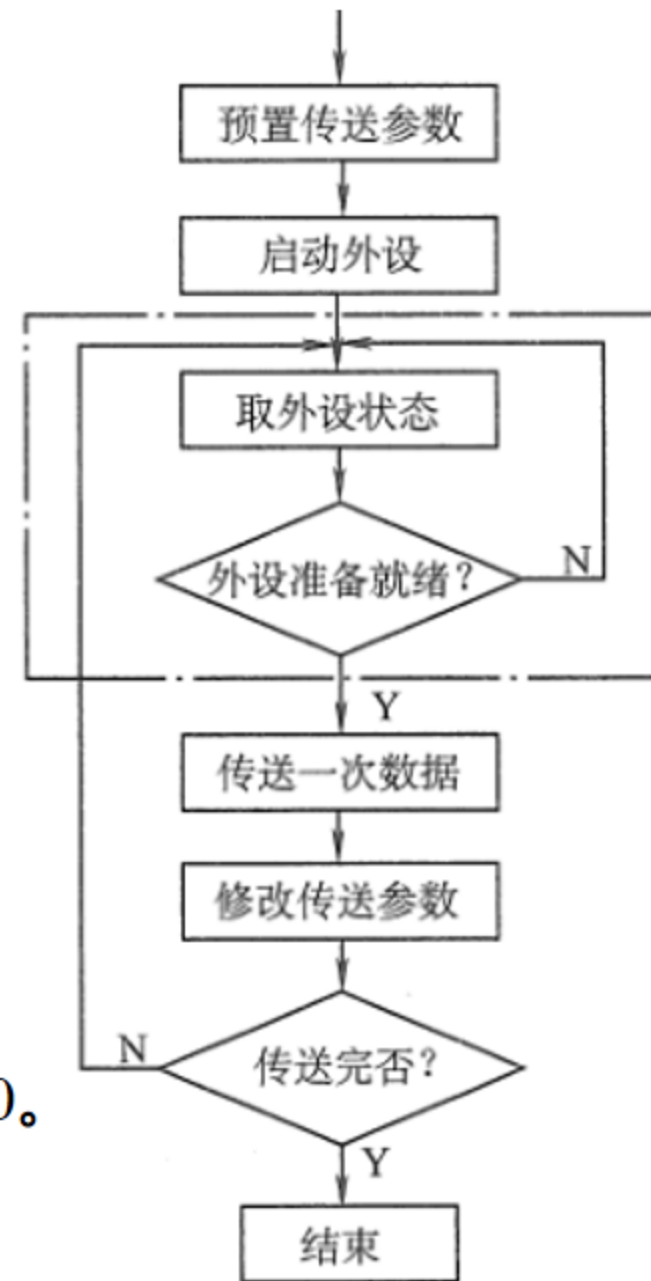


图 7.4 程序查询方式流程图

1.程序查询方式

在这种控制方式下，CPU一旦启动I / O，就必须停止现行程序的运行，并在现行程序插入一段程序。

程序查询方式的主要特点是CPU有“踏步”等待现象，CPU与I / O串行工作。这种方式的接口设计简单、设备量少，但CPU在信息传送过程中要花费很多时间来查询和等待，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。

2.程序中断方式

中断是现代计算机有效合理地发挥效能和提高效率的一个十分重要的功能。CPU中通常设有处理中断的机构-中断系统，以解决各种中断的共性问题。

一.中断的基本概念

程序中断是指在计算机执行现行程序的过程中，出现某些急需处理的异常情况或特殊请求，CPU暂时中止现行程序，而转去对这些异常情况或特殊请求进行处理，在处理完毕后CPU又自动返回到现行程序的断点处，继续执行原程序。



扫码听课，视频讲解更清晰

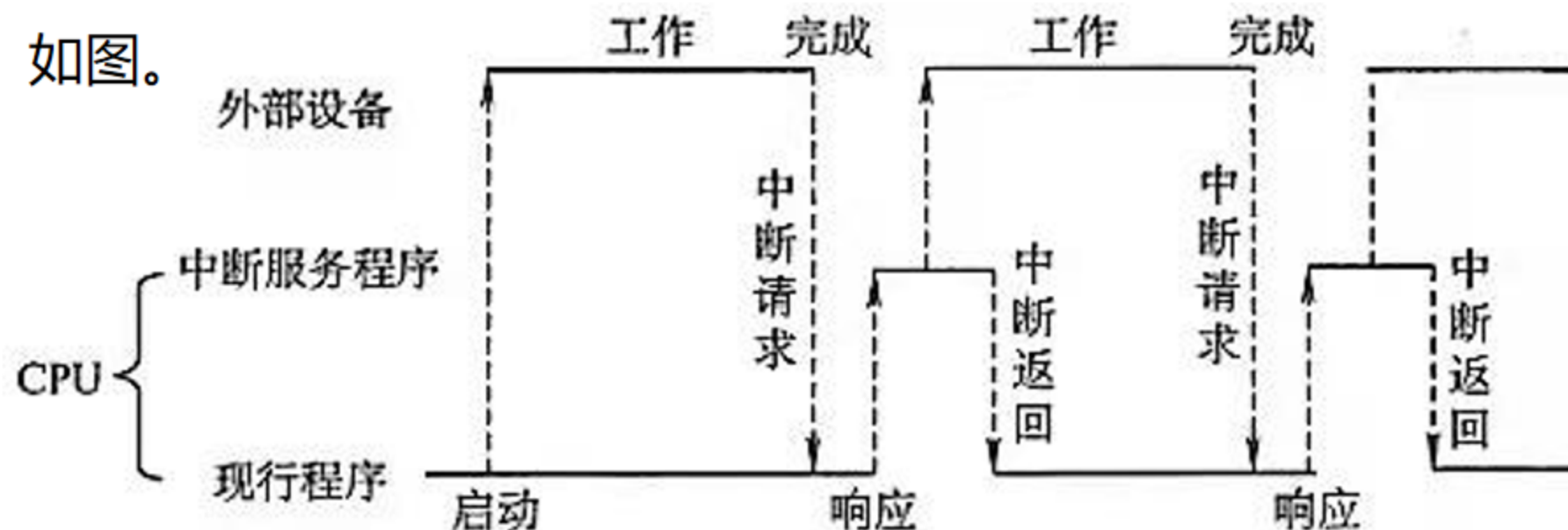
2.程序中中断方式

程序中中断的作用：

- ①实现CPU与I / O设备的并行工作。
- ②处理硬件故障和软件错误。
- ③实现人机交互，用户干预机器需要用到中断系统。
- ④实现多道程序、分时操作，多道程序的切换需借助于中断系统。
- ⑤实时处理需要借助中断系统来实现快速响应。
- ⑥实现应用程序和操作系统（管态程序）的切换，称为“软中断”。
- ⑦多处理器系统中各处理器之间的信息交流和任务切换。

2.程序中中断方式

程序中中断方式的思想： CPU在程序中安排好于某个时刻启动某台外设，然后CPU 继续执行原来的程序，不需要像查询方式那样一直等待外设准备就绪。一旦外设完成数据传送的准备工作，就主动向CPU发出中断请求，请求CPU为自己服务。在可以响应中断的条件下，CPU暂时中止正在执行的程序，转去执行中断服务程序为外设服务，在中断服务程序中完成一次主机与外设之间的数据传送，传送完成后，CPU返回原来的程序，如图。



2.程序中中断方式

2.程序中中断方式工作流程

(1) 中断请求

中断请求是指中断源向CPU发送中断请求信号。

①内中断和外中断

中断源是请求CPU中断的设备或事件，一台计算机允许有多个中断源。根据中断源的类别，可把中断源分为内中断和外中断两种。

每个中断源向CPU发出中断请求的时间是随机的。为记录中断事件并区分不同的中断源，中断系统需对每个中断源设置中断请求标记触发器INTR，当其状态为“1”时，表示中断源有请求。这些触发器可组成中断请求标记寄存器，该寄存器可集中在CPU中，也可分散在各个中断源中。

2.程序中中断方式

外中断是指来自处理器和**内存以外的部件**引起的中断，包括I / O设备发出的I / O中断、外部信号中断（如用户按Esc键），以及各种定时器引起的时钟中断等。外中断在狭义上一般称为中断（后文若未说明，一般是指外中断）。

内中断主要是指在处理器和**内存内部**产生的中断，括程序运算引起的各种错误，如地址非法、校验错、页面失效、存取访问控制错、算术操作溢出、数据格式非法、除数为零、非法指令、用户程序执行特权指令、分时系统中的时间片中断及用户态到核心态的切换等。

2.程序中中断方式

②硬件中断和软件中断

硬件中断：通过外部的硬件产生的中断。硬件中断属于外中断。

软件中断：通过某条指令产生的中断，这种中断是可以编程实现的。软件中断是内中断。

2.程序中断方式

③非屏蔽中断和可屏蔽中断

非屏蔽中断：非屏蔽中断是一种硬件中断，此种中断通过不可屏蔽中断请求NMI 控制，不受中断标志位IF的影响，即使在关中断（IF = 0）的情况下也会被响应。

可屏蔽中断：可屏蔽中断也是一种硬件中断，此种中断通过中断请求标记触发器INTR控制，且受中断标志位IF的影响，在关中断情况下不接受中断请求。

也就是说，可屏蔽中断和非屏蔽中断均是外中断。



扫码听课，视频讲解更清晰

2.程序中断方式

(2) 中断判优

中断系统在任一瞬间只能响应一个中断源的请求。由于许多中断源提出中断请求的时间都是随机的，因此当多个中断源同时提出请求时，需通过中断判优逻辑确定响应哪个中断源的请求，例如故障中断的优先级别较高，然后是I / O中断。

中断判优既可以用硬件实现，又可用软件实现。硬件实现是通过硬件排队器实现的，它既可以设置在CPU中，又可以分散在各个中断源中，软件实现是通过查询程序实现的。

一般来说，硬件故障中断属于最高级，其次是软件中断，非屏蔽中断优于可屏蔽中断，DMA请求优于I / O设备传送的中断请求，高速设备优于低速设备，输入设备优于输出设备，实时设备优于普通设备等。

2.程序中中断方式

(3) CPU响应中断的条件

CPU在满足一定的条件下响应中断源发出的中断请求，并经过一些特定的操作，转去执行中断服务程序。

CPU响应中断必须满足以下3个条件：

- ①中断源有中断请求。
- ② CPU允许中断及开中断。
- ③一条指令执行完毕，且没有更紧迫的任务。

2.程序中中断方式

注意：I / O设备的就绪时间是随机的，而CPU在统一的时刻即每条指令执行阶段结束前向接口发出中断查询信号，以获取I / O的中断请求，也就是说，CPU响应中断的时间是在每条指令执行阶段的结束时刻。这里说的**中断仅指外中断**，内中断不属于此类情况。

(4) 中断隐指令

CPU响应中断后，经过某些操作，转去执行中断服务程序。这些操作是由硬件直接实现的，我们将它称为**中断隐指令**。中断隐指令并不是指令系统中的一条真正的指令，它没有操作码，所以中断隐指令是一种不允许也不可能为用户使用的特殊指令。

2.程序中中断方式

它所完成的操作如下：

- ①**关中断**。在中断服务程序中，为了保护中断现场（即CPU主要寄存器中的内容）期间不被新的中断所打断，必须关中断，从而保证被中断的程序在中断服务程序执行完毕后能接着正确地执行。
- ②**保存断点**。为保证在中断服务程序执行完毕后能正确地返回到原来的程序，必须将原来程序的断点〔即程序计数器（PC）的内容〕保存起来。
- ③**引出中断服务程序**。引出中断服务程序的实质是，取出中断服务程序的入口地址并传送给程序计数器（PC）。

2.程序中中断方式

(5) 中断向量

不同的设备有不同的中断服务程序，每个中断服务程序都有一个入口地址，CPU必须找到这个入口地址，即**中断向量**，把系统中的全部中断向量集中存放到存储器的某个区域内，这个存放中断向量的存储区就称为**中断向量表**，即中断服务程序入口地址表。

CPU响应中断后，中断硬件会自动将中断向量地址传送到CPU，由CPU实现程序的切换，这种方法称为中断向量法，采用中断向量法的中断称为**向量中断**。

注意：中断向量是中断服务程序的入口地址，中断向量地址是指中断服务程序的入口地址的地址。

2. 程序中中断方式

(6) 中断处理过程

不同计算机的中断处理过程各具特色，就其多数而论，中断处理流程如图所示。

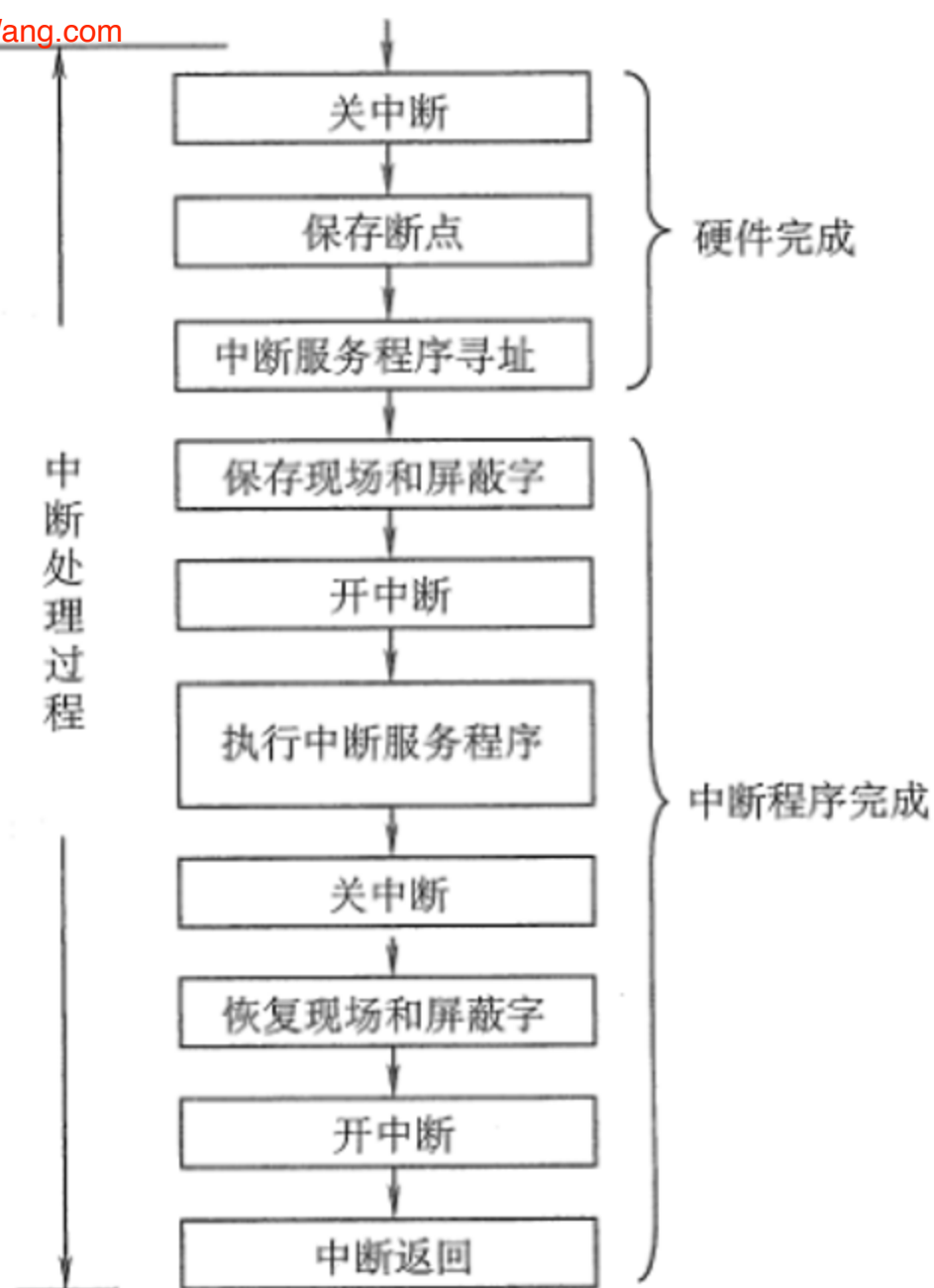


图 7.6 可嵌套中断的处理流程

2.程序中中断方式

中断处理流程如下：

- ①**关中断**。处理器响应中断后，首先要保护程序的现场状态，在保护现场的过程中，CPU 不应响应更高级中断源的中断请求。否则，若现场保存不完整，在中断服务程序结束后，也就不能正确地恢复并继续执行现行程序。
- ②**保存断点**。为保证中断服务程序执行完毕后能正确地返回到原来的程序，必须将原来的程序的断点保存起来。断点可以压入堆栈，也可以存入主存的特定单元中。
- ③**引出中断服务程序**。引出中断服务程序的实质是，取出中断服务程序的入口地址送入程序计数器（PC）。

2. 中断方式

通常有两种方法寻址中断服务程序的入口地址：**硬件向量法和软件查询法。**

硬件向量法通过硬件产生中断向量地址，再由中断向量地址找到中断服务程序的入口地址。

软件查询法用软件编程的办法寻找入口地址。

注意：硬件产生的实际上是中断类型号，而中断类型号指出了中断向量存放的地址，因此能产生中断向量地址。

④**保存现场和屏蔽字。**进入中断服务程序后首先要保存现场，现场信息一般是指程序状态字、中断屏蔽寄存器和CPU 中某些寄存器的内容。

⑤**开中断。**允许更高级中断请求得到响应，实现中断嵌套。

2.程序中中断方式

⑥执行中断服务程序。这是中断请求的目的。

⑦关中断。保证在恢复现场和屏蔽字时不被中断。

⑧恢复现场和屏蔽字。将现场和屏蔽字恢复到原来的状态。

⑨开中断、中断返回。中断服务程序的最后一条指令通常是一条中断返回指令，使其返回到原程序的断点处，以便继续执行原程序。

注意：①～③在CPU进入中断周期后，由中断隐指令（硬件自动）完成；④～⑨由中断服务程序完成。

注意：恢复现场是指在中断返回前，必须将寄存器的内容恢复到中断处理前的状态，这部分工作由中断服务程序完成。中断返回由中断服务程序的最后一条中断返回指令完成。

2.程序中断方式

中断屏蔽技术主要用于多重中断。CPU要具备多重中断的功能，必须满足下列条件：

- ①在中断服务程序中提前设置开中断指令。
- ②优先级别高的中断源有权中断优先级别低的中断源。

每个中断源都有一个屏蔽触发器，1表示屏蔽该中断源的请求，0表示可以正常申请，所有屏蔽触发器组合在一起便构成一个屏蔽字寄存器，屏蔽字寄存器的内容称为屏蔽字。

(关于中断屏蔽字的设置及多重中断程序执行的轨迹，下面通过实例说明。)

【题1】 设某机有4个中断源A、B、C、D，其硬件排队优先次序为 $A > B > C > D$ ，现要求将中断处理次序改为 $D > A > C > B$ 。

1) 写出每个中断源对应的屏蔽字。

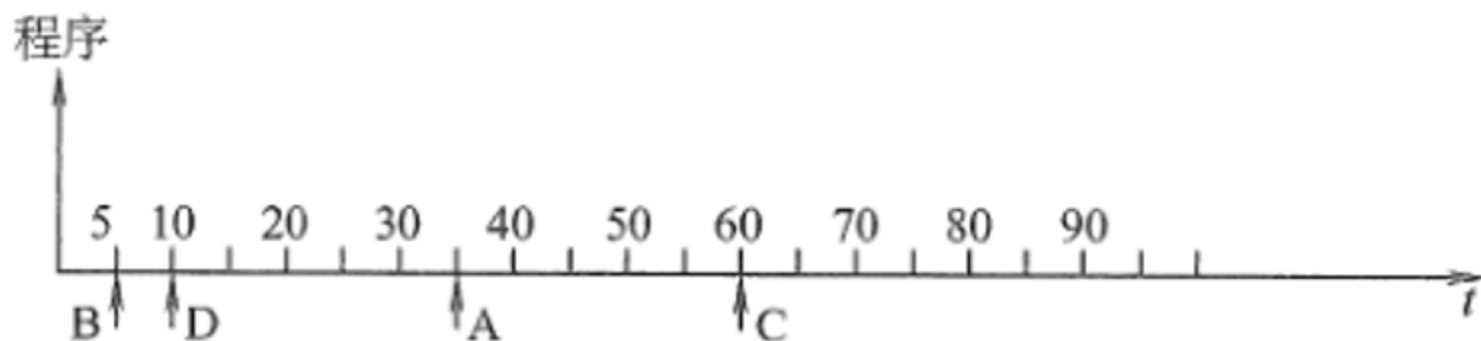
解：1) 在中断处理次序改为 $D > A > C > B$ 后，D具有最高优先级，可以屏蔽其他所有中断，且不能中断自身，因此D对应的屏蔽字为1111；
A具有次高优先级，只能被D中断，因此A对应的屏蔽字为1110，以此类推，得到4个中断源的屏蔽字，见表

表 7.1 中断源对应的中断屏蔽字

中断源	屏蔽字			
	A	B	C	D
A	1	1	1	0
B	0	1	0	0
C	0	1	1	0
D	1	1	1	1

【题1】 设某机有4个中断源A、B、C、D，其硬件排队优先次序为 $A > B > C > D$ ，现要求将中断处理次序改为 $D > A > C > B$ 。

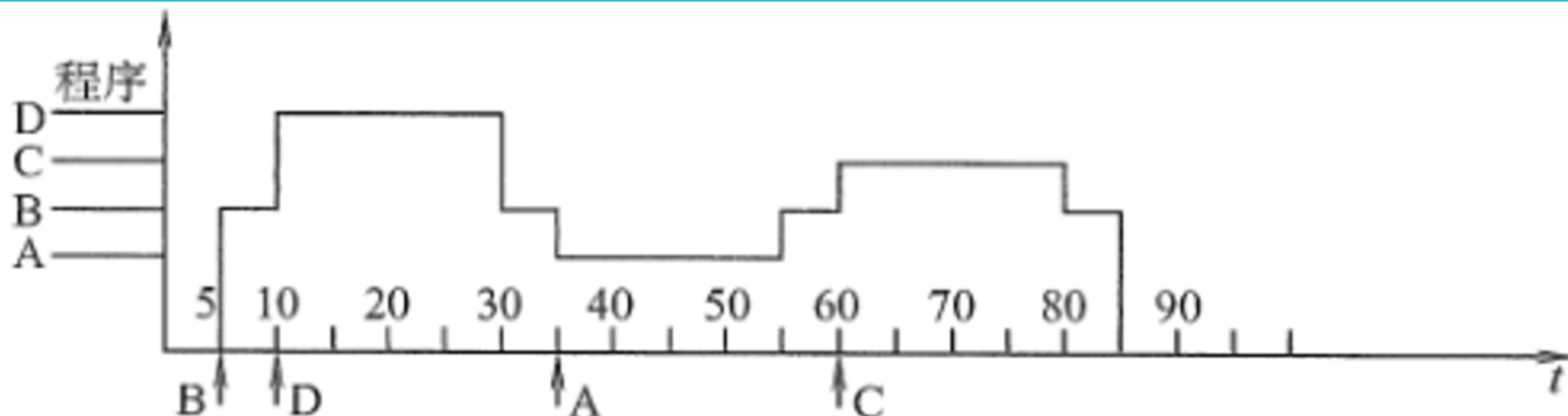
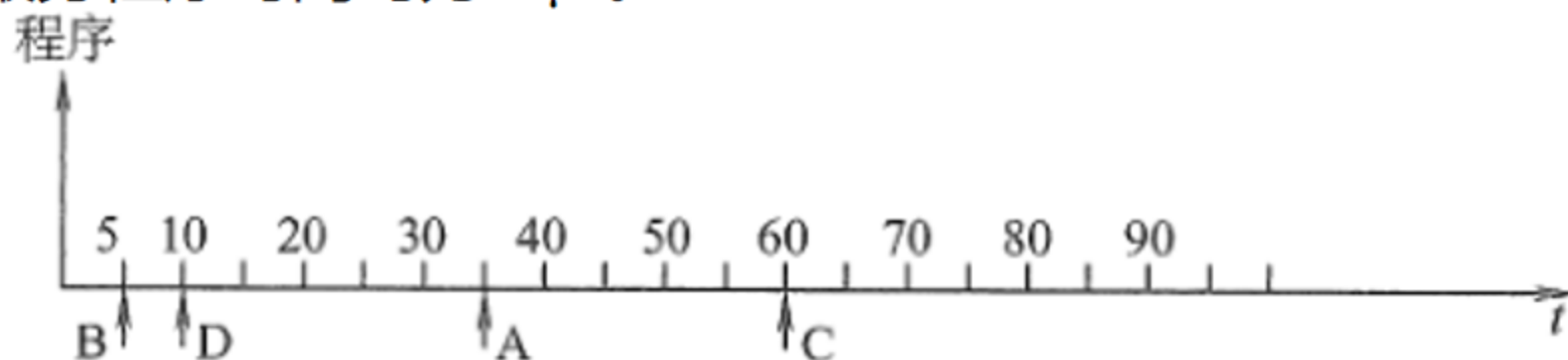
2) 按图所示的时间轴给出的4个中断源的请求时刻，画出CPU执行程序轨迹。设每个中断源的中断服务程序时间均为 $20\mu s$ 。



2) 根据处理次序，在时刻5，B发中断请求，获得CPU；在时刻10，D发中断请求，此时B虽还未执行完毕，但D的优先级高于B，于是D中断B而获得CPU；在时刻30，D执行完毕，B继续获得CPU；在时刻40，A发中断请求，此时B虽还未执行完毕，但A的优先级高于B，于是A中断B而获得CPU，如此继续下去，执行轨迹如图所示。

【题1】 设某机有4个中断源A、B、C、D，其硬件排队优先次序为 $A > B > C > D$ ，现要求将中断处理次序改为 $D > A > C > B$ 。

2) 按图所示的时间轴给出的4个中断源的请求时刻，画出CPU执行程序轨迹。设每个中断源的中断服务程序时间均为 $20\mu\text{s}$ 。



3.DMA方式

DMA 方式是一种完全由硬件进行成组信息传送的控制方式，它具有程序中断方式的优点，即在数据准备阶段，CPU与外设并行工作。

DMA方式在外设与内存之间开辟一条“直接数据通道”，信息传送不再经过CPU，降低了CPU在传送数据时的开销，因此称为**直接存储器存取方式**。

数据传送不经过CPU，也就不需要保护、恢复CPU现场等烦琐操作。

这种方式适用于磁盘机、磁带机等高速设备大批量数据的传送，它的硬件开销比较大。在DMA方式中，中断的作用**仅限于故障和正常传送结束时的处理**。

3.DMA方式

一.DMA方式的特点

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，因此不必中断现行程序，I / O与主机并行工作，程序和传送并行工作。

DMA方式的特点：

- 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。
- ③主存中要开辟专用缓冲区，及时供给和接收外设的数据。
- ④DMA传送速度快，CPU和外设并行工作，提高了系统效率。
- ⑤DMA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。

3.DMA方式

二.DMA控制器的组成

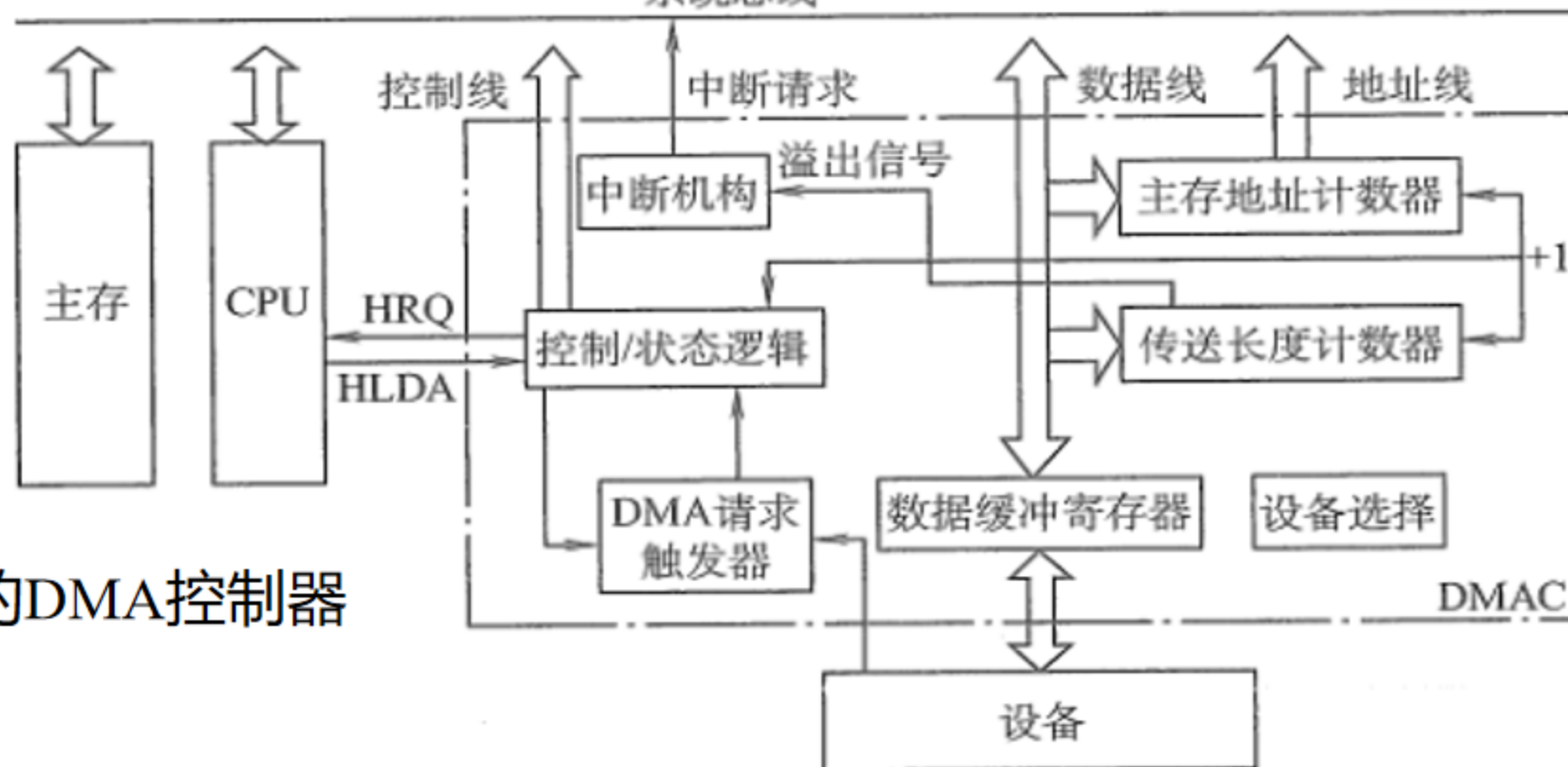
在DMA方式中，对数据传送过程进行控制的硬件称为DMA控制器（DMA接口）。当I / O设备需要进行数据传送时，通过DMA控制器向CPU提出DMA传送请求，CPU响应之后将让出系统总线，由DMA控制器接管总线进行数据传送。

其功能如下：

- 1) 接受外设发出的DMA请求，并向CPU发出总线请求。
- 2) CPU响应此总线请求，发出总线响应信号，接管总线控制权，进入DMA操作周期。
- 3) 确定传送数据的主存单元地址及长度，并自动修改主存地址计数和传送长度计数。
- 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。
- 5) 向CPU报告DMA操作的结束。

3.DMA方式

简单的DMA控制器



主存地址计数器：存放要交换数据的主存地址。

传送长度计数器：记录传送数据的长度，计数溢出时，数据即传送完毕，自动发中断请求信号。

数据缓冲寄存器：暂存每次传送的数据。

3.DMA方式

DMA 请求触发器：每当I / O设备准备好数据后，给出一个控制信号，使DMA请求触发器置位。

“控制 / 状态”逻辑：由控制和时序电路及状态标志组成，用于指定传送方向，修改传送参数，并对DMA请求信号和CPU响应信号进行协调和同步。

中断机构：当一个数据块传送完毕后触发中断机构，向CPU提出中断请求。

在DMA传送过程中，DMA控制器将接管CPU的地址总线、数据总线和控制总线，CPU的主存控制信号被禁止使用。而当DMA传送结束后，将恢复CPU的一切权利并开始执行其操作。由此可见，**DMA控制器必须具有控制系统总线的能力。**

3.DMA方式

三.DMA的传送方式

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种方式使用主存：

1) **停止CPU访问主存。**这种方式是当外设需要传送成组数据时，由DMA接口向CPU发送一个信号，要求CPU放弃地址线、数据线和有关控制线的使用权，DMA接口获得总线控制权后，开始进行数据传送。数据传送结束后，DMA接口通知CPU可以使用主存，并把总线控制权交还给CPU。在这种传送过程中，CPU基本处于不工作状态或保持原始状态。

3.DMA方式

2) DMA与CPU交替访存。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 $1.2\mu\text{s}$ ，主存的存取周期小于 $0.6\mu\text{s}$ ，则可将一个CPU周期分为C1和C2两个周期，其中C1专供DMA访存，C2专供CPU访存。这种方式不需要总线使用权的申请、建立和归还过程，总线使用权是通过C1和C2分时控制的。

3) 周期挪用（或周期窃取）。这种方式是前两种方式的折中。当I / O设备没有DMA请求时，CPU按程序的要求访问主存，一旦I / O设备有了DMA请求，就会遇到3种情况。

第1种是此时CPU不在访存，因此I / O的访存请求与CPU未发生冲突；

第2种是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；

第3种是I / O和CPU同请求访存，出现访存冲突，此时CPU要暂时放弃总线占有权，由I / O设备挪用一個或几个存取周期。

3.DMA方式

四.DMA的传送过程

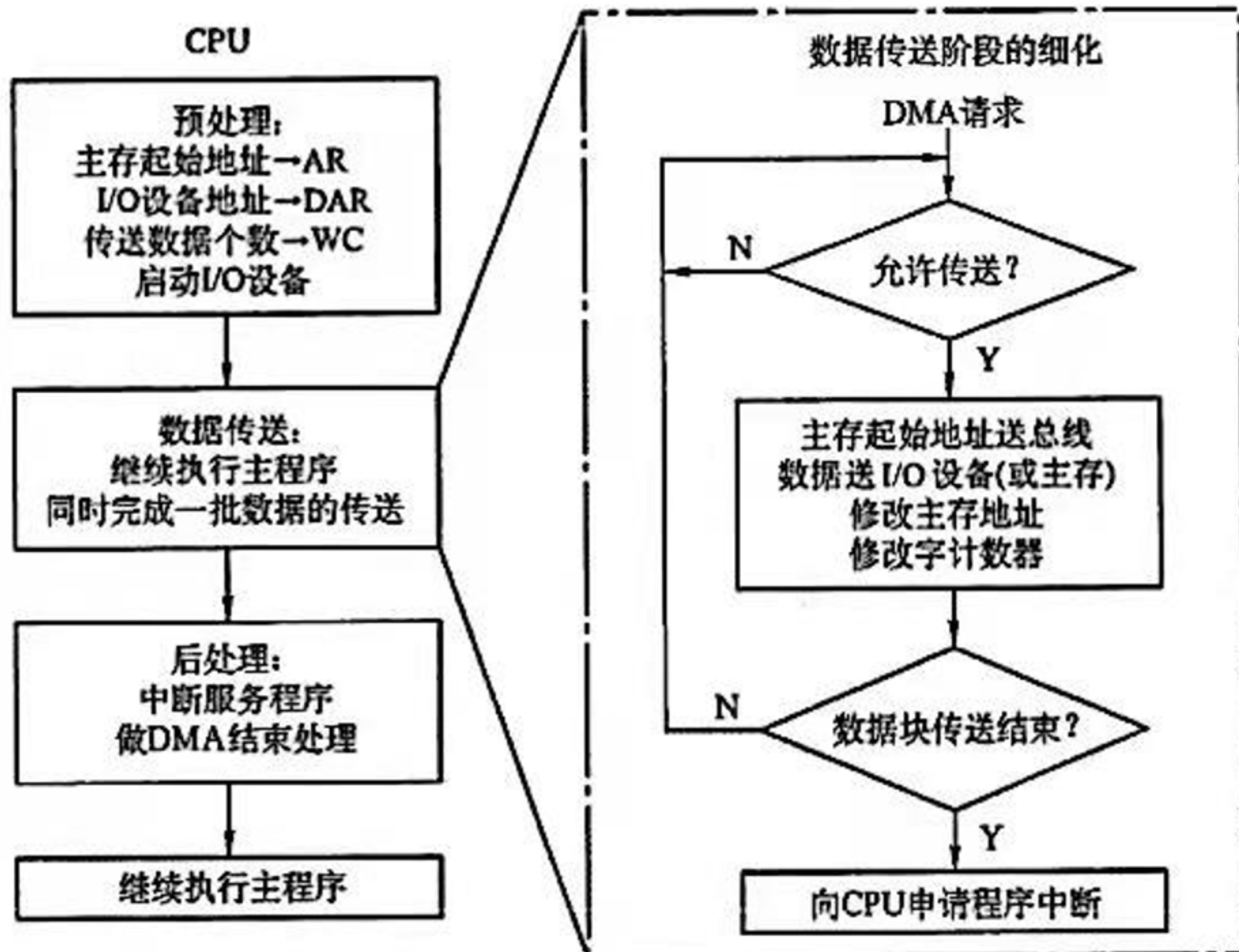
DMA的数据传送过程分为**预处理**、**数据传送**和**后处理**3个阶段：

- 1) **预处理**。由CPU完成些必要的准备工作。首先，CPU执行几条I / O指令，用以测试I / O设备状态，向DMA控制器的有关寄存器置初值、设置传送方向、启动该设备等。然后，CPU继续执行原来的程序，直到I / O设备准备好发送的数据（输入情况）或接收的数据（输出情况）时，I / O设备向DMA控制器发送DMA请求，再由DMA控制器向CPU发送总线请求（有时将这两个过程统称为DMA请求），用以传输数据。

3.DMA方式

- 2) **数据传送。** DMA的数据传输可以以单字节（或字）为基本单位，也可以以数据块为基本单位。对于以数据块为单位的传送（如硬盘），DMA占用总线后的数据输入和输出操作都是通过循环来实现的。需要指出的是，这一循环也是由DMA控制器（而非通过CPU执行程序）实现的，即数据传送阶段完全由DMA（硬件）控制。
- 3) **后处理。** DMA控制器向CPU发送中断请求，CPU执行中断服务程序做DMA结束处理，包括校验送入主存的数据是否正确、测试传送过程中是否出错（错误则转入诊断程序）及决定是否继续使用DMA传送其他数据块等。DMA的传送流程如图所示。

3.DMA方式



3.DMA方式

五.DMA方式和中断方式的区别

DMA方式和中断方式的重要区别如下：

- ①中断方式是程序的切换，需要保护和恢复现场；而DMA方式除了预处理和后处理，其他时候不占用CPU的任何资源。
- ②对中断请求的响应只能发生在每条指令执行完毕时（即指令的执行周期后）；而对DMA请求的响应可以发生在每个机器周期结束时（在取指周期、间址周期、执行周期后均可），只要CPU不占用总线就可被响应。

3.DMA方式

- ③中断传送过程需要CPU的干预；而DMA传送过程不需要CPU的干预，因此数据传输率非常高，适合于高速外设的成组数据传送。
- ④DMA请求的优先级高于中断请求。
- ⑤中断方式具有对异常事件的处理能力，而DMA方式仅局限于传送数据块的I / O操作。
- ⑥从数据传送来看，中断方式靠程序传送，DMA方式靠硬件传送。



扫码听课，视频讲解更清晰

【题2】 某磁盘的转速为10000转 / 分，平均寻道时间是6ms，磁盘传输速率是20MB / s，磁盘控制器延迟为0.2ms，读取一个4KB的扇区所需的平均时间约为（ B ）。

A. 9ms

B. 9.4ms

C. 12ms

D. 12.4ms

解析： 磁盘转速是10000转 / 分，转一圈的时间为6ms，因此平均查询扇区的时间为3ms，平均寻道时间为6ms，读取4KB扇区信息的时间为0.2ms，信息延迟的时间为 $4\text{KB} / (20\text{MB} / \text{s}) = 0.2\text{ms}$ ，总时间为 $3 + 6 + 0.2 + 0.2 = 9.4\text{ms}$ 。