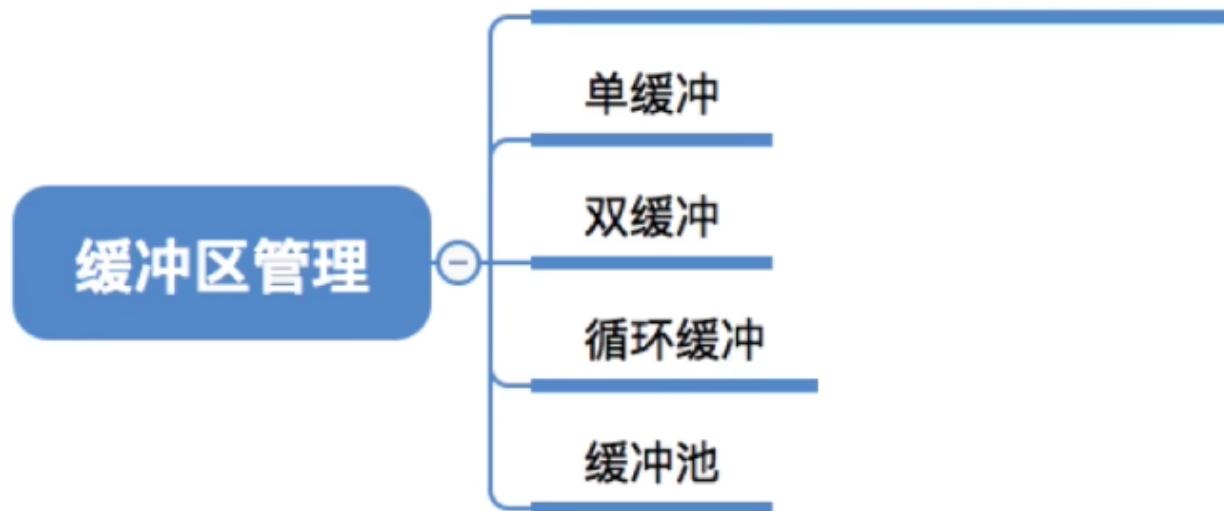


高速缓冲与缓冲区



什么是缓冲区? 有什么作用?



什么是缓冲区? 有什么作用?

缓冲区是一个存储区域，可以由专门的硬件寄存器组成，也可利用内存作为缓冲区。

使用硬件作为缓冲区的成本较高，容量也较小，一般仅用在对速度要求非常高的场合（如存储器管理中所用的联想寄存器，由于对页表的访问频率极高，因此使用速度很快的联想寄存器来存放页表项的副本）

一般情况下，更多的是利用内存作为缓冲区，“设备独立性软件”的缓冲区管理就是要组织管理好这些缓冲区

本节介绍的是“内存作为缓冲区”

缓冲区的作用

缓和CPU与I/O设备之间速度不匹配的矛盾

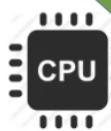
减少对CPU的中断频率，放宽对CPU中断相应时间的限制

解决数据粒度不匹配的问题

提高CPU与I/O设备之间的并行性

缓冲区有什么作用？

CPU可以把要输出的数据快速地放入缓冲区，之后就可以做别的事



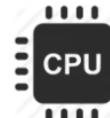
缓冲区

内存

慢速的I/O设备可以从缓冲区取走数据

I/O设备
数据输入时类似

如果是字符型设备，则每输出完一个字符就要向CPU发送一次中断信号

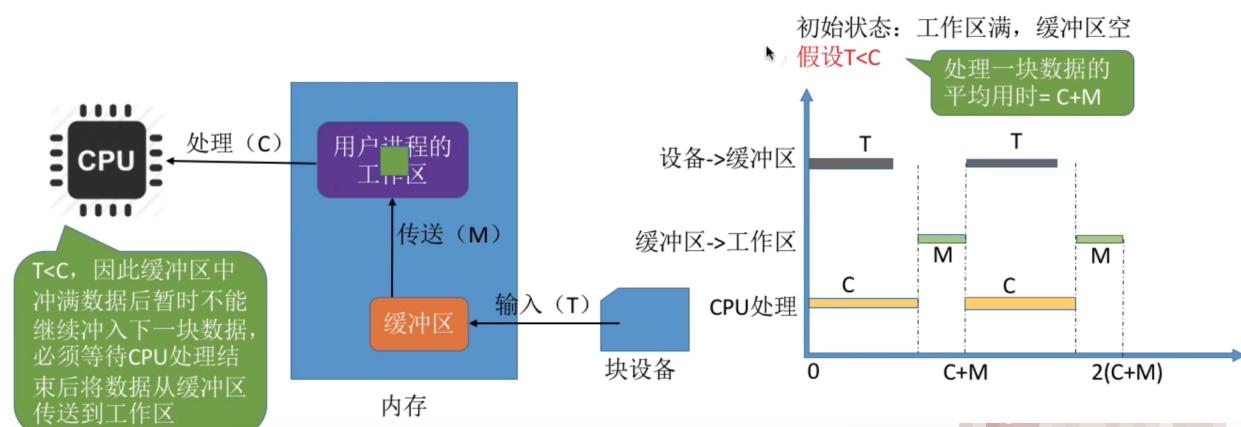
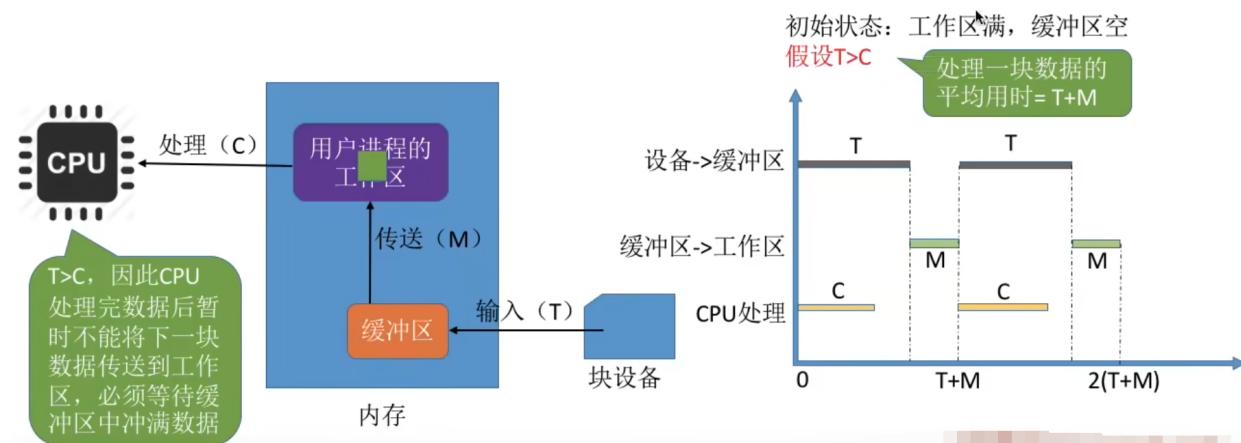
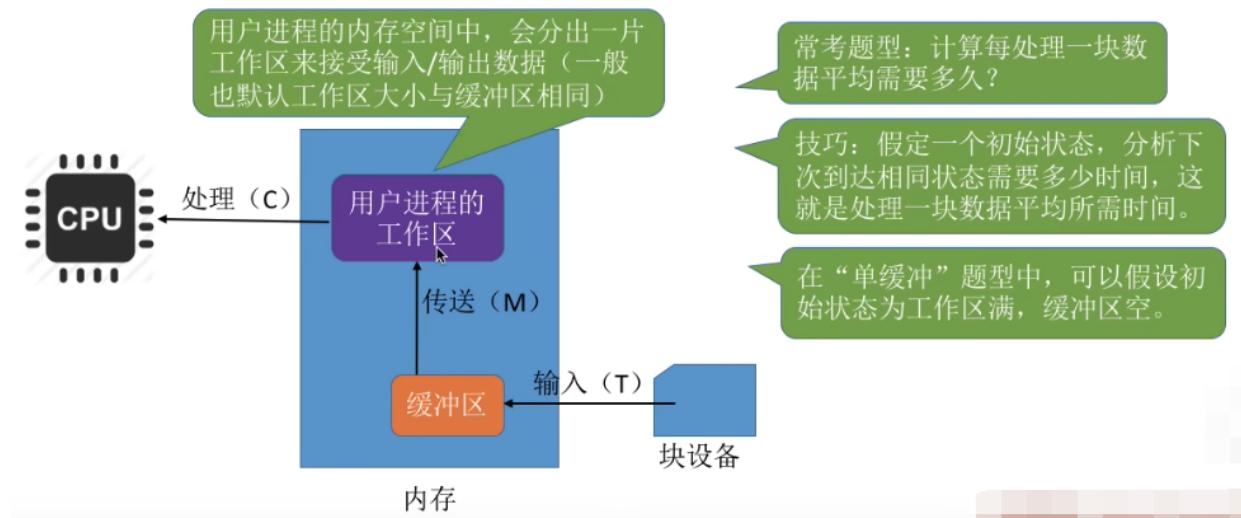


I/O设备

单缓冲

假设某用户进程请求某种块设备读入若干块的数据。若采用单缓冲的策略，操作系统会在主存中为其分配一个缓冲区（若题目中没有特别说明，一个缓冲区的大小就是一个块）。

注意：当缓冲区数据非空时，不能往缓冲区冲入数据，只能从缓冲区把数据传出；当缓冲区为空时，可以往缓冲区冲入数据，但必须把缓冲区充满以后，才能从缓冲区把数据传出。

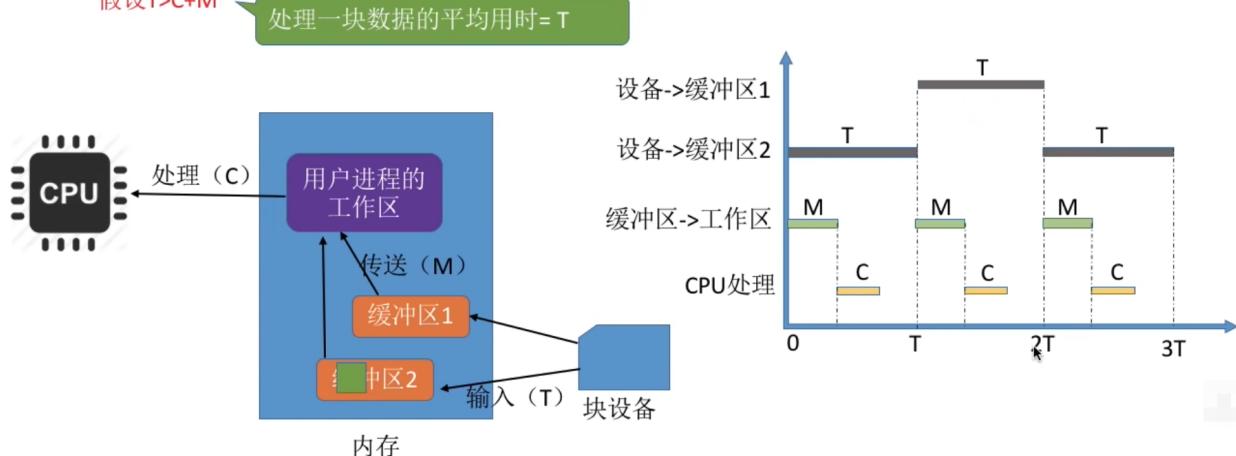


双缓冲

假设某用户进程请求某种块设备读入若干块的数据。若采用**双缓冲**的策略，操作系统会在主存中为
其分配两个缓冲区（若题目中没有特别说明，一个缓冲区的大小就是一个块）

双缓冲题目中，假设初始状态为：工作区空，其中一个缓冲区满，另一个缓冲区空

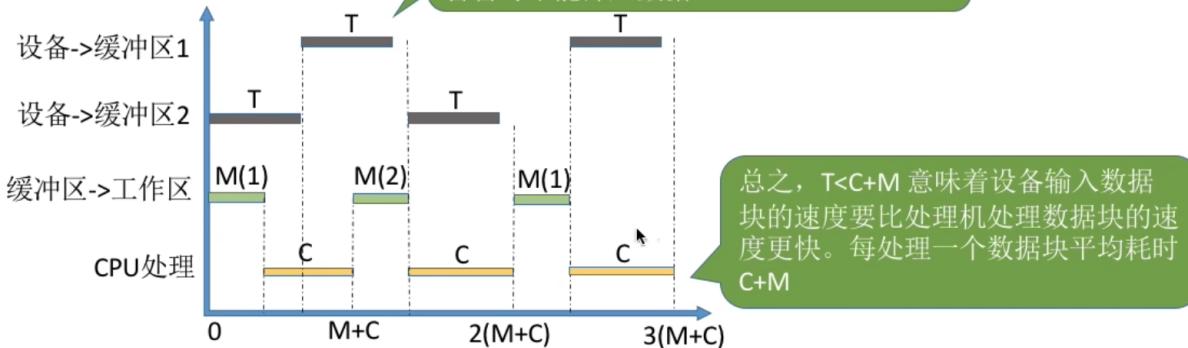
假设 $T > C + M$



双缓冲题目中，假设初始状态为：工作区空，其中一个缓冲区满，另一个缓冲区空

假设 $T < C + M$

假设 $2T < 2M + C$ ，则 I/O 设备将缓冲区 1 满时，缓冲区 2 的数据尚未取空，因此 I/O 设备暂时不能冲入数据



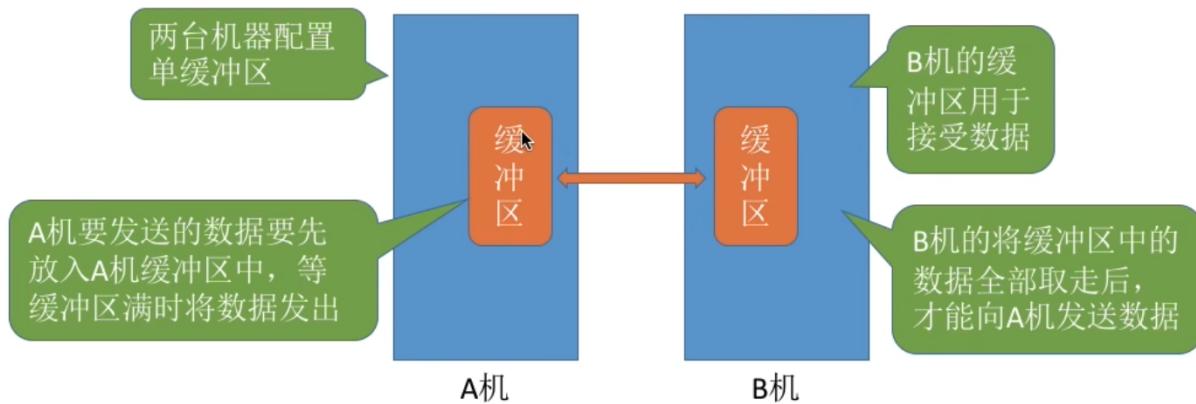
注：M(1) 表示“将缓冲区 1 中的数据传送到工作区”； M(2) 表示“将缓冲区 2 中的数据传送到工作区”

结论：采用双缓冲策略，处理一个数据块的平均耗时为

$$\max(T, C + M)$$

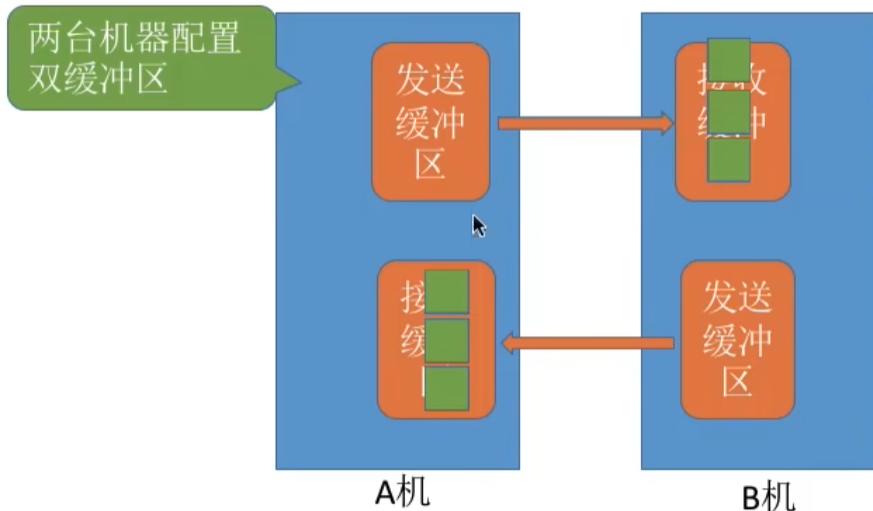
使用单/双缓冲在通信时的区别

两台机器之间通信时，可以配置缓冲区用于数据的发送和接受。



显然，若两个相互通信的机器只设置单缓冲区，在任一时刻只能实现数据的单向传输。

两台机器之间通信时，可以配置缓冲区用于数据的发送和接受。

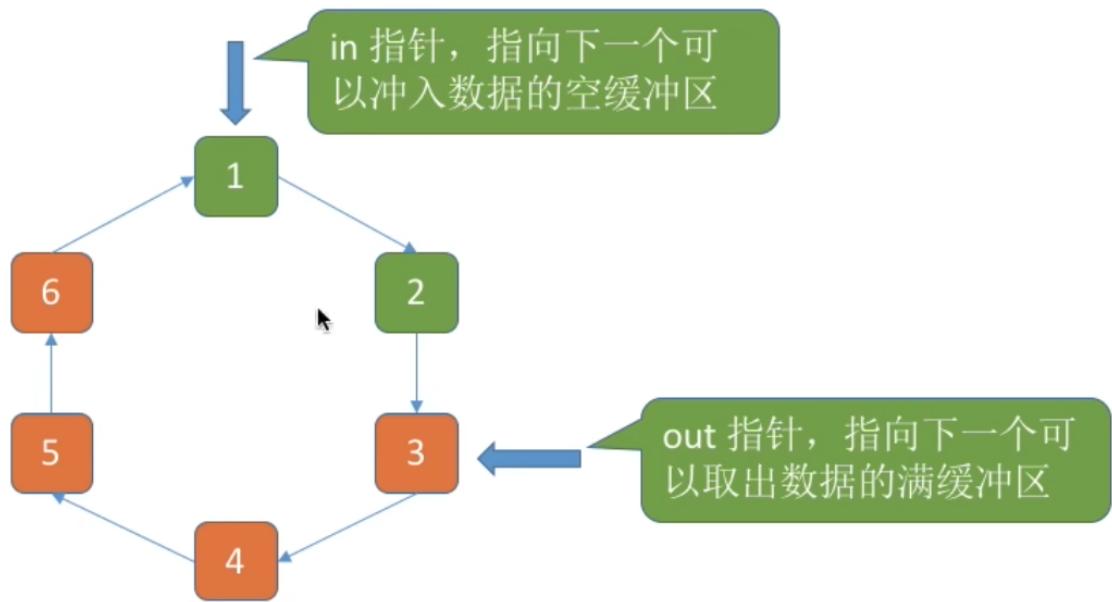


若两个相互通信的机器设置双缓冲区，则同一时刻可以实现双向的数据传输。

循环缓冲区

将多个**大小相等**的缓冲区链接成一个**循环队列**。

注：以下图示中，**橙色**表示已充满数据的缓冲区，**绿色**表示空缓冲区。



缓冲池

缓冲池由系统中共用的缓冲区组成。这些缓冲区按使用状况可以分为：空缓冲队列、装满输入数据的缓冲队列（输入队列）、装满输出数据的缓冲队列（输出队列）。

另外，根据一个缓冲区在实际运算中扮演的功能不同，又设置了四种工作缓冲区：用于收容输入数据的工作缓冲区（hin）、用于提取输入数据的工作缓冲区（sin）、用于收容输出数据的工作缓冲区（hout）、用于提取输出数据的工作缓冲区（sout）

