

窦轶////yi.dou@njupt.edu.cn

O p e r a t i n g S y s t e m s

操作系统



设备管理 基本概念

Linux
Android
Linux
OpenStack
Mac OS
Windows



缓冲技术

本讲内容

1. 缓冲技术的基本思想
2. 引入缓冲技术的目的
3. 缓冲技术的几种类型

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）

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

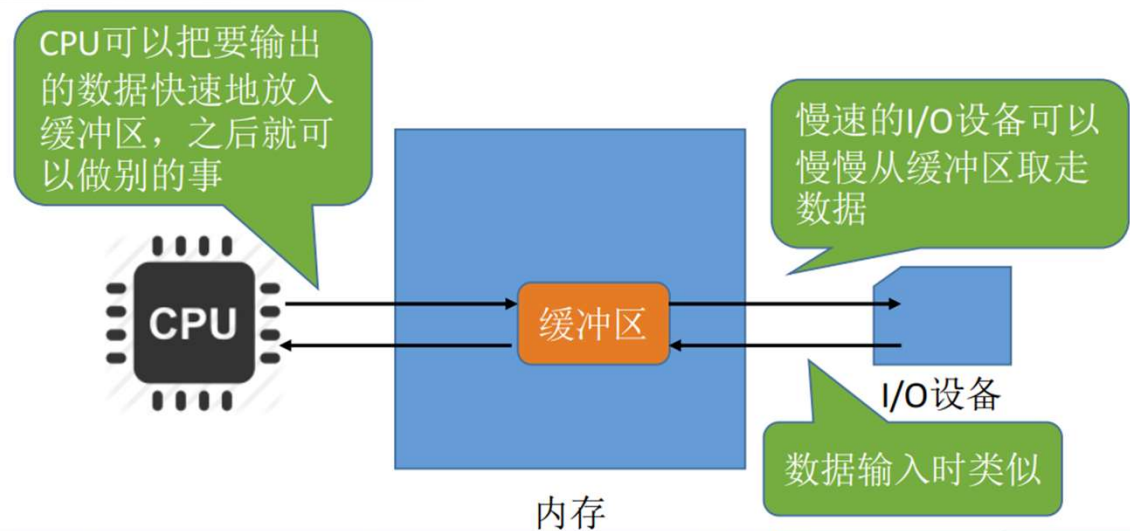
减少对CPU的中断频率，放宽对CPU中断响应时间的限制

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

缓冲技术的基本思想

- ❖ 进程执行写操作输出数据时，申请一个缓冲区，**不断把数据填到缓冲区**，直到被装满；进程继续运行，系统**将缓冲区内容传输到I/O设备**
- ❖ 进程执行读操作输入数据时，申请一个缓冲区，系统**将内容读到缓冲区**，根据进程要求，把需要的内容**从缓冲区传送给进程**



缓冲区有什么作用？

缓冲区的作用

缓和CPU与I/O设备之间速度不匹配的矛盾

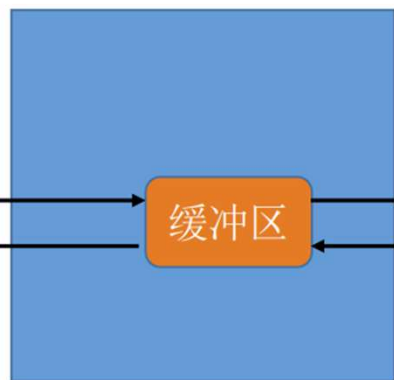
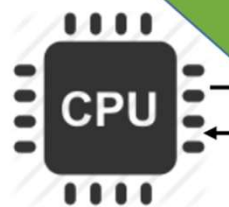
减少对CPU的中断频率，放宽对CPU中断响应时间的限制

解决数据粒度不匹配的问题

如：输出进程每次可以生成一块数据，但I/O设备每次只能输出一个字符

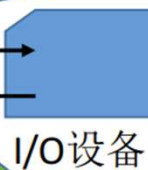
提高CPU与I/O设备之间的并行性

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



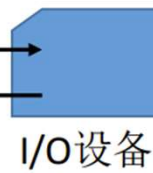
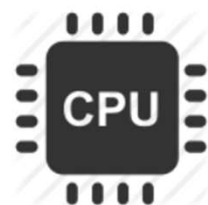
内存

慢速的I/O设备可以慢慢从缓冲区取走数据



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



缓冲技术

本讲内容

1. 缓冲技术的基本思想
2. 引入缓冲技术的目的
3. 缓冲技术的几种类型

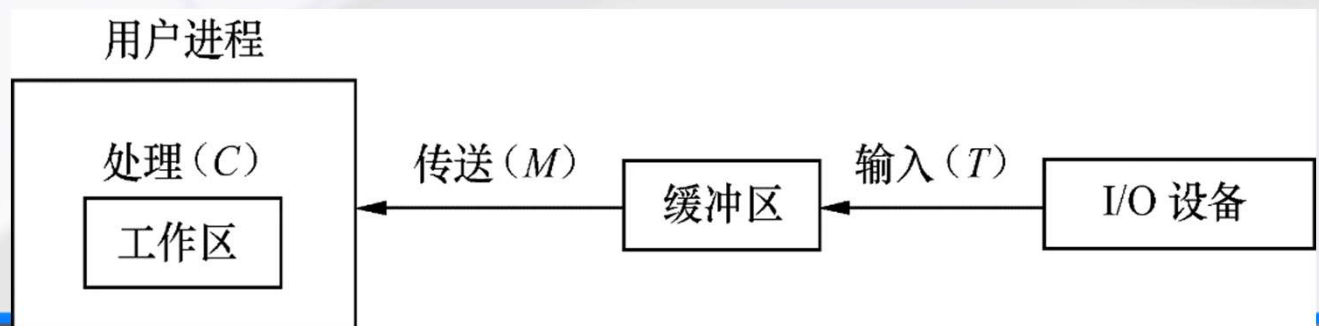
引入缓冲技术的目的

- ❏ 改善主机与外围设备之间**速度不匹配**的问题
- ❏ 提高CPU和I/O设备的**并行性**
- ❏ 减少I/O**中断次数**，放宽CPU中断响应的要求
- ❏ 提高计算机系统的总体性能

缓冲技术的几种类型

单缓冲

- ❏ 假定从硬盘把一块数据输入到缓冲的时间为 T ，系统将缓冲区中的数据传送到用户区的时间为 M ，而CPU对这一块数据进行计算的时间为 C
- ❏ 如果不采用缓冲技术，数据直接从硬盘到用户区，每块数据的处理时间为 $T+C$
- ❏ 如果采用了单缓冲技术，每块数据的处理时间为 $\max(C, T)+M$ ； M 要远小于 C 或者 T ，因此提高了效率



单缓冲



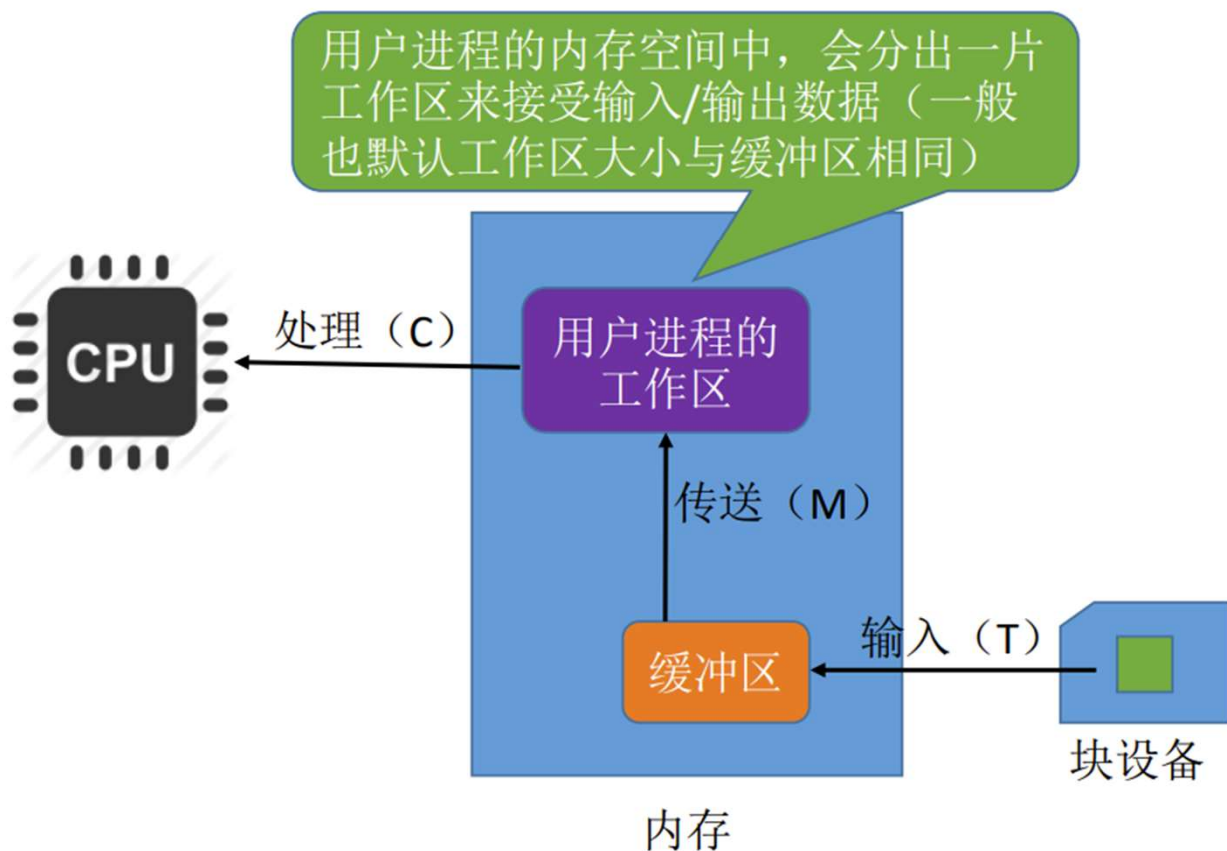
假设某用户进程请求某种块设备读入若干块的数据。若采用单缓冲的策略，操作系统会在主存中为其分配一个缓冲区（若题目中没有特别说明，一个缓冲区的大小就是一个块）。

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。

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）。

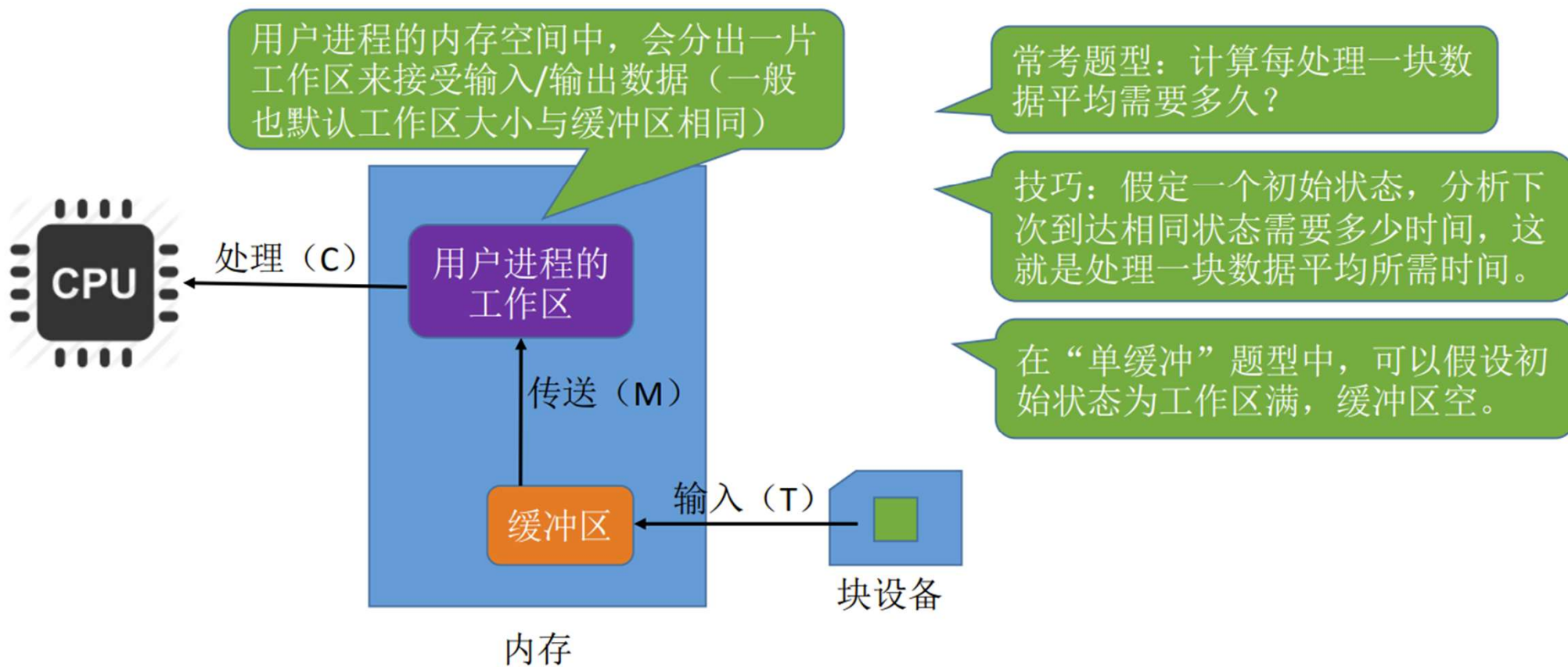
注意：当缓冲区数据非空时，不能往缓冲区冲入数据，只能从缓冲区把数据传出；当缓冲区为空时，可以往缓冲区冲入数据，但必须把缓冲区充满以后，才能从缓冲区把数据传出。



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）。

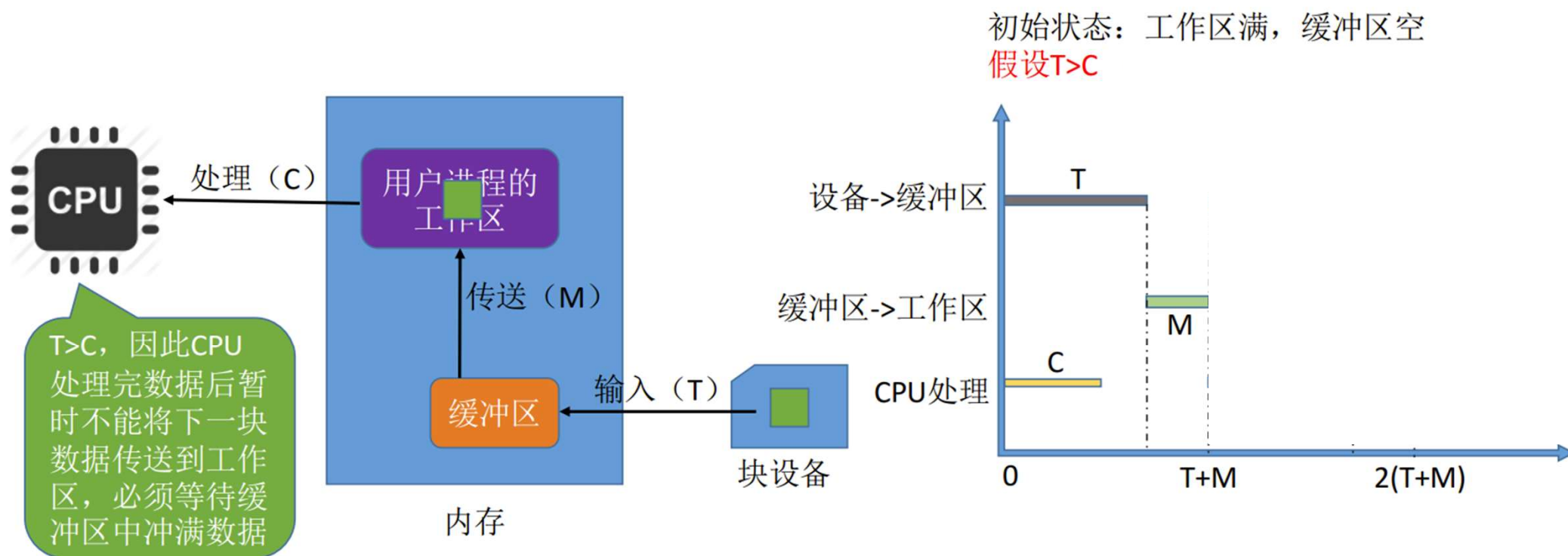
注意：当缓冲区数据非空时，不能往缓冲区冲入数据，只能从缓冲区把数据传出；当缓冲区为空时，可以往缓冲区冲入数据，但必须把缓冲区充满以后，才能从缓冲区把数据传出。



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）。

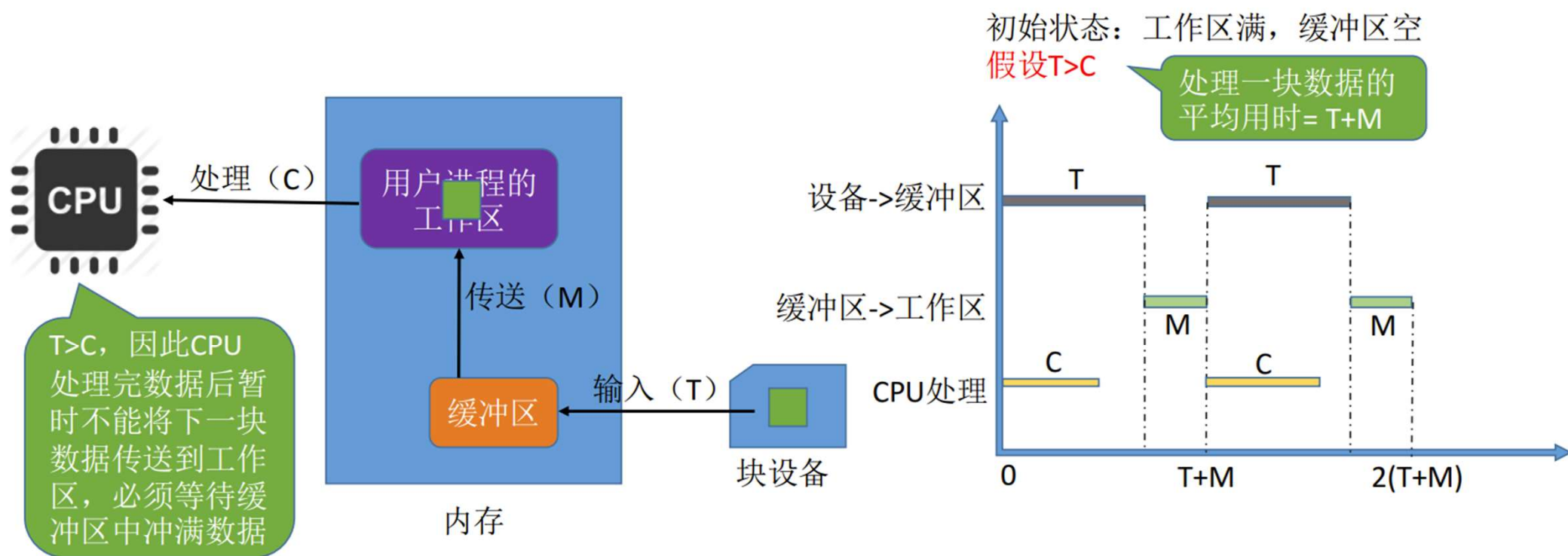
注意：当缓冲区数据非空时，不能往缓冲区冲入数据，只能从缓冲区把数据传出；当缓冲区为空时，可以往缓冲区冲入数据，但必须把缓冲区充满以后，才能从缓冲区把数据传出。



单缓冲

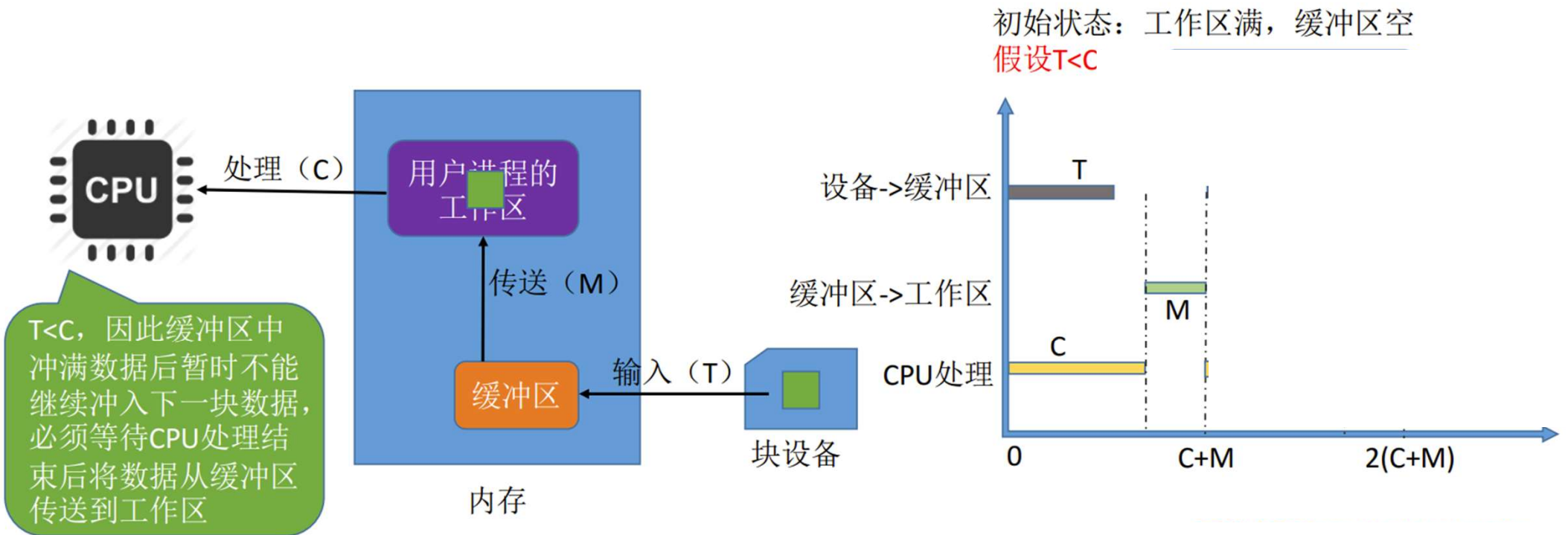
假设某用户进程请求某种块设备读入若干块的数据。若采用单缓冲的策略，操作系统会在主存中为其分配一个缓冲区（若题目中没有特别说明，一个缓冲区的大小就是一个块）。

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。



单缓冲

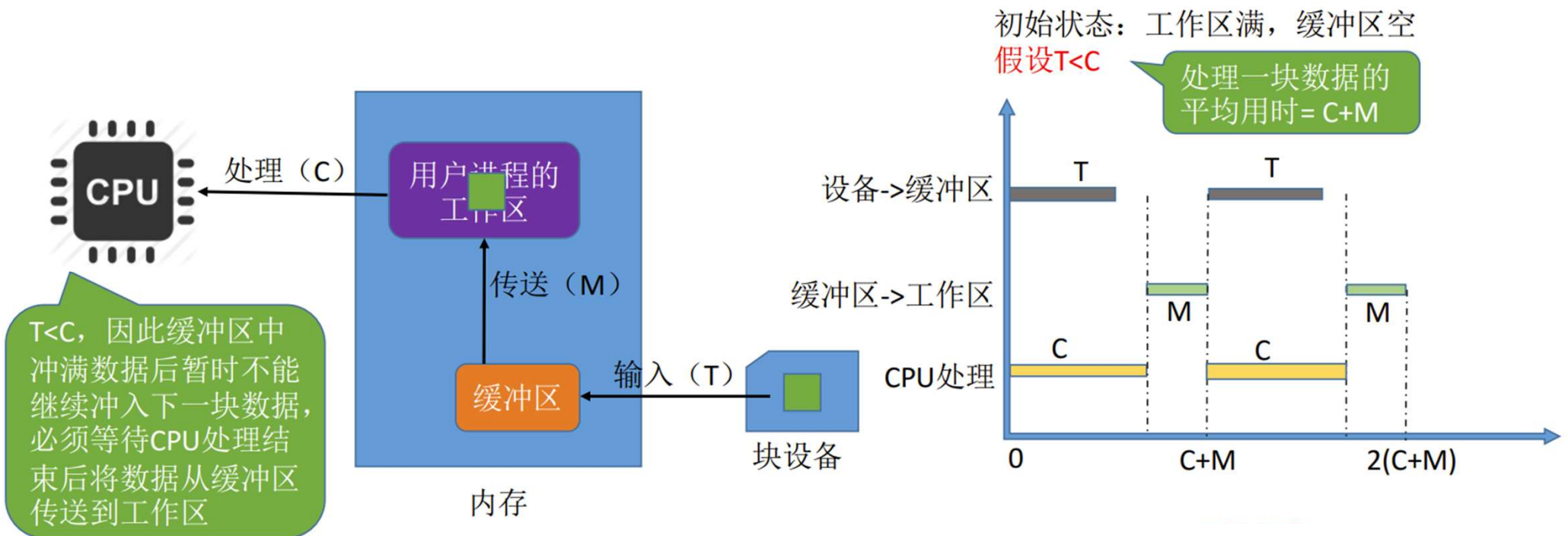
假设某用户进程请求某种块设备读入若干块的数据。若采用单缓冲的策略，操作系统会在主存中为其分配一个缓冲区（若题目中没有特别说明，一个缓冲区的大小就是一个块）。
注意：当缓冲区数据非空时，不能往缓冲区冲入数据，只能从缓冲区把数据传出；当缓冲区为空时，可以往缓冲区冲入数据，但必须把缓冲区充满以后，才能从缓冲区把数据传出。



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）。

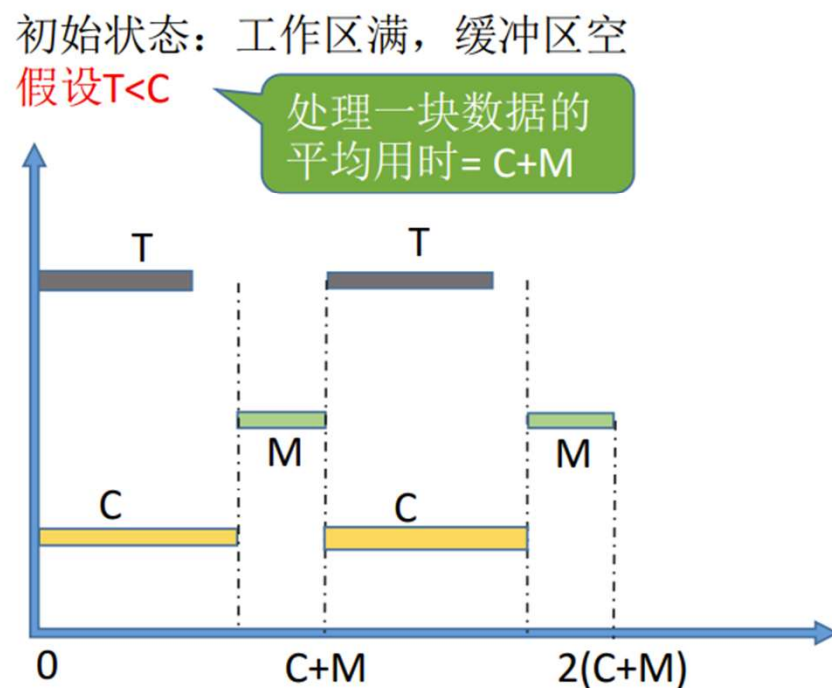
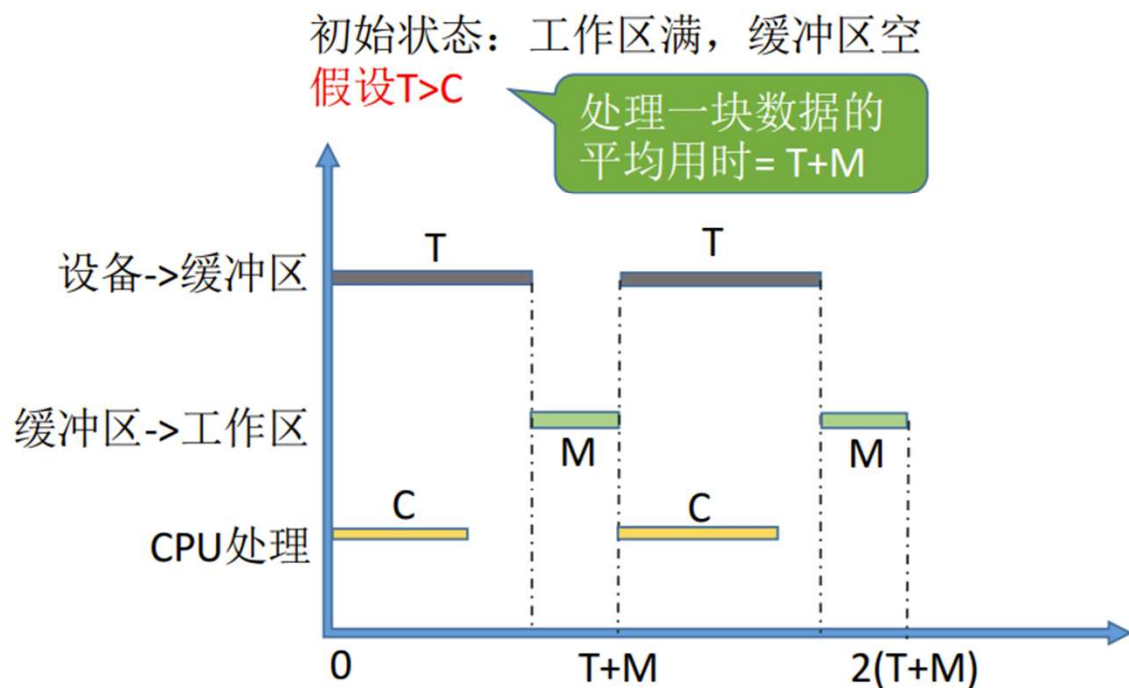
注意：当缓冲区数据非空时，不能往缓冲区冲入数据，只能从缓冲区把数据传出；当缓冲区为空时，可以往缓冲区冲入数据，但必须把缓冲区充满以后，才能从缓冲区把数据传出。



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。

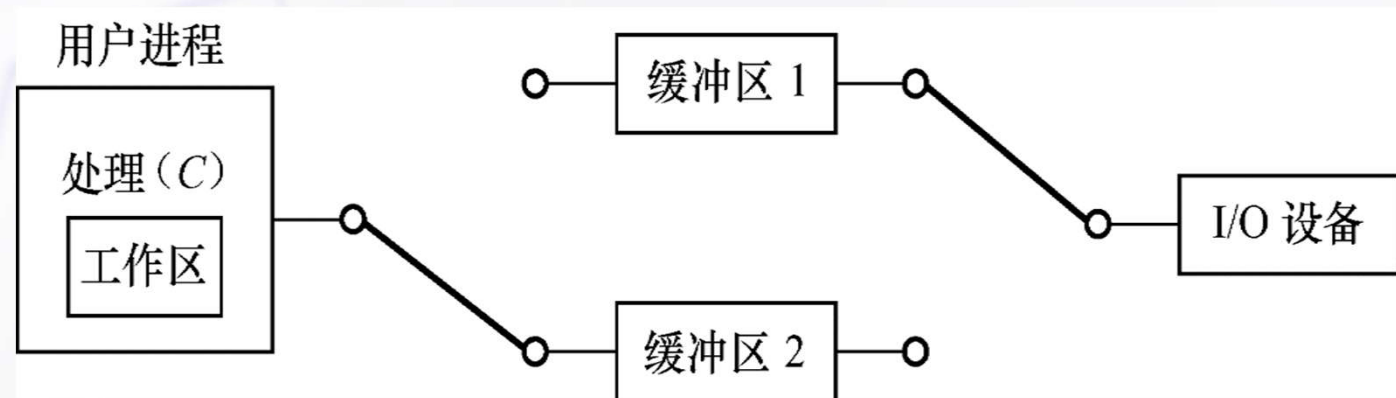


结论：采用单缓冲策略，处理一块数据平均耗时 $\text{Max}(C, T) + M$

缓冲技术的几种类型

双缓冲

- 输入数据时，首先填满缓冲区1，系统可从缓冲区1把数据送到用户进程区，用户进程便可对数据进行加工计算，同时，输入设备填充缓冲区2。
- 缓冲区1空出后，输入设备再次向缓冲区1输入。系统可把缓冲区2的数据传送到用户进程区，进程开始加工缓冲区2的数据。



双缓冲



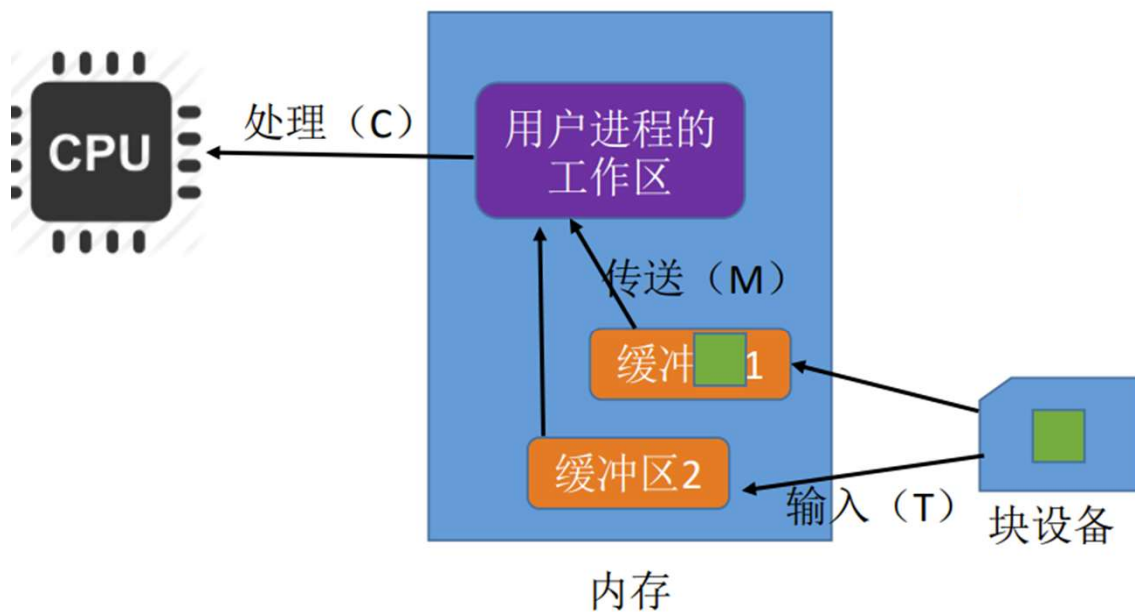
假设某用户进程请求某种块设备读入若干块的数据。若采用双缓冲的策略，操作系统会在主存中为其分配两个缓冲区（若题目中没有特别说明，一个缓冲区的大小就是一个块）

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

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

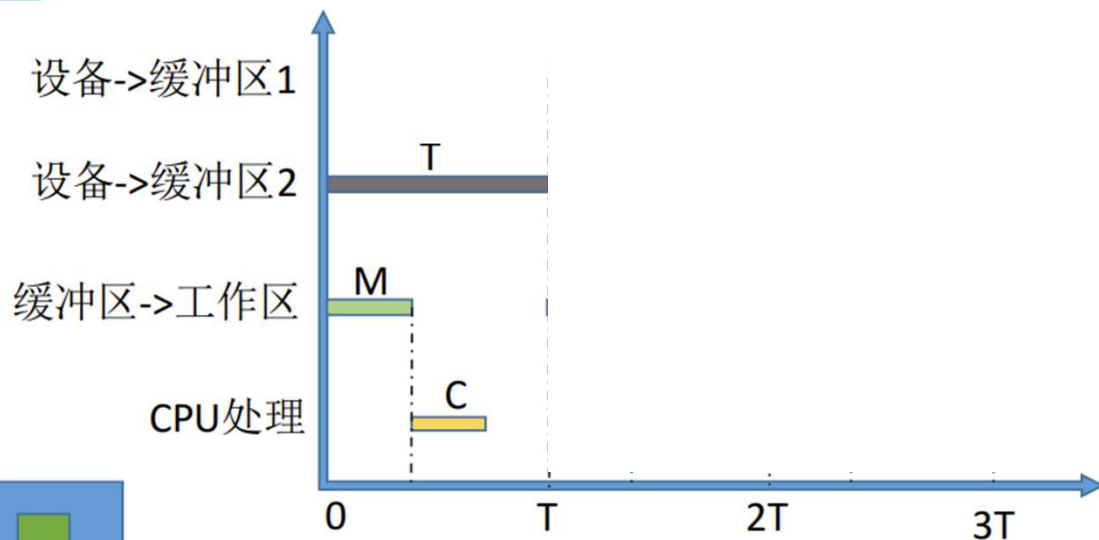
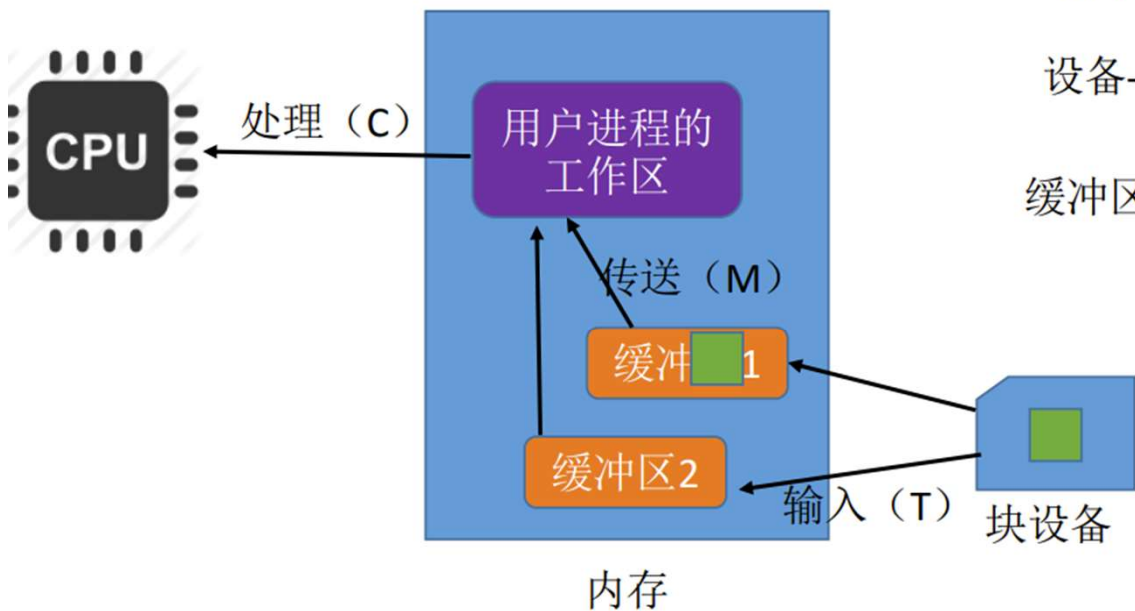


双缓冲

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 $T > C + M$

处理一块数据的平均用时 = T



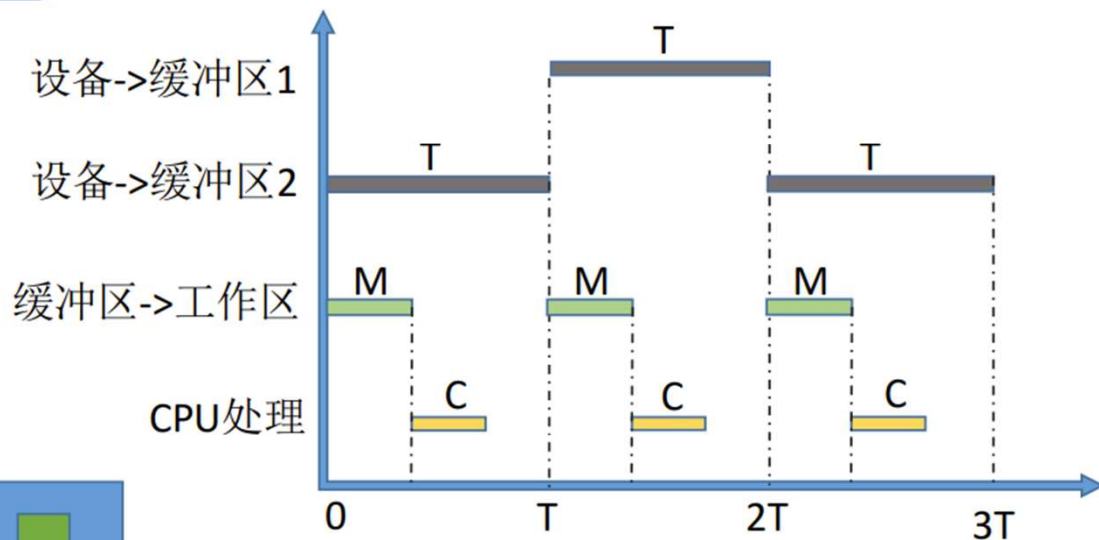
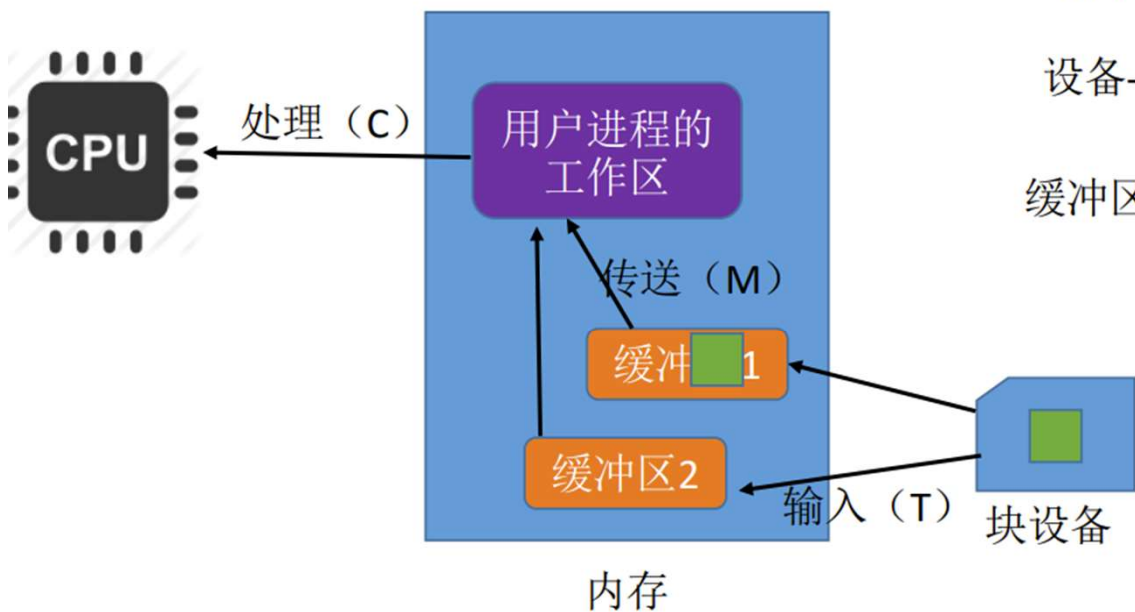
双缓冲

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 $T > C + M$

处理一块数据的平均用时 = T

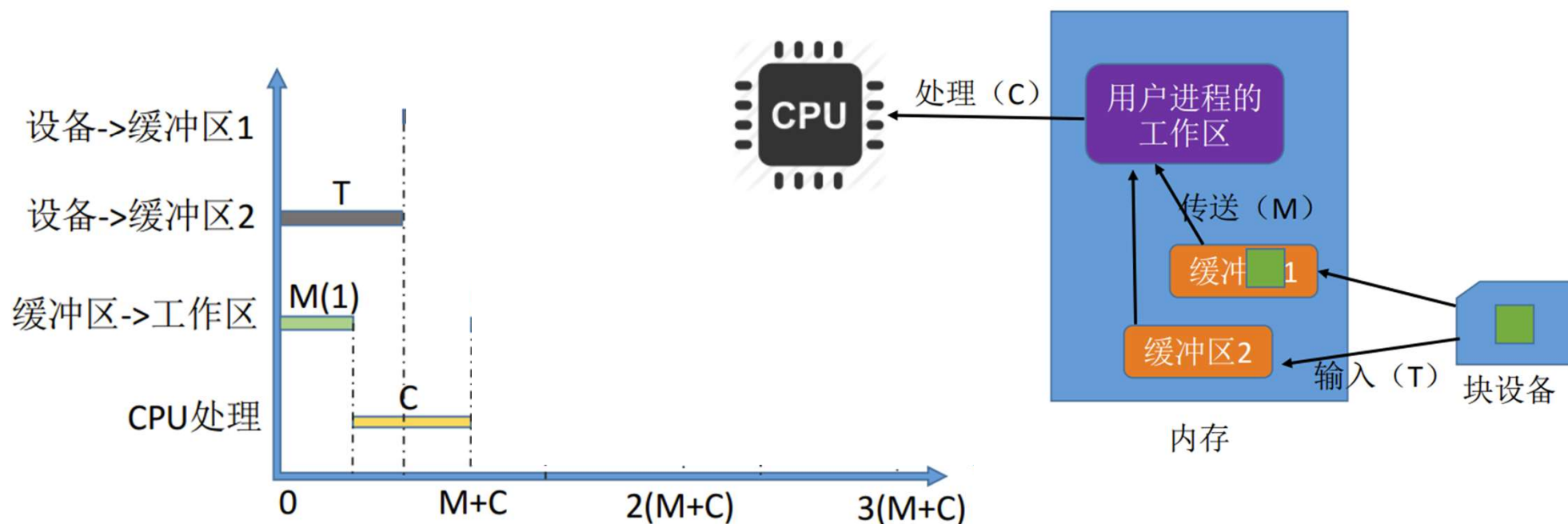


双缓冲

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 $T < C + M$



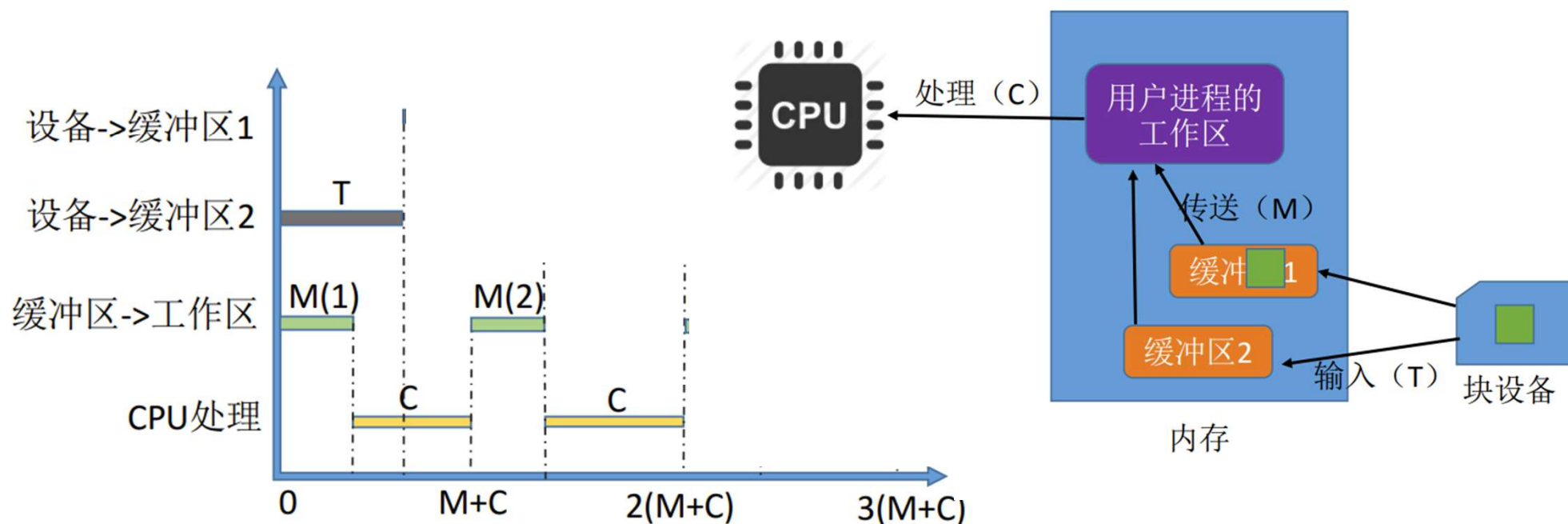
注：M(1) 表示“将缓冲区1中的数据传送到工作区”；M(2) 表示“将缓冲区2中的数据传送到工作区”

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 $T < C + M$



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”

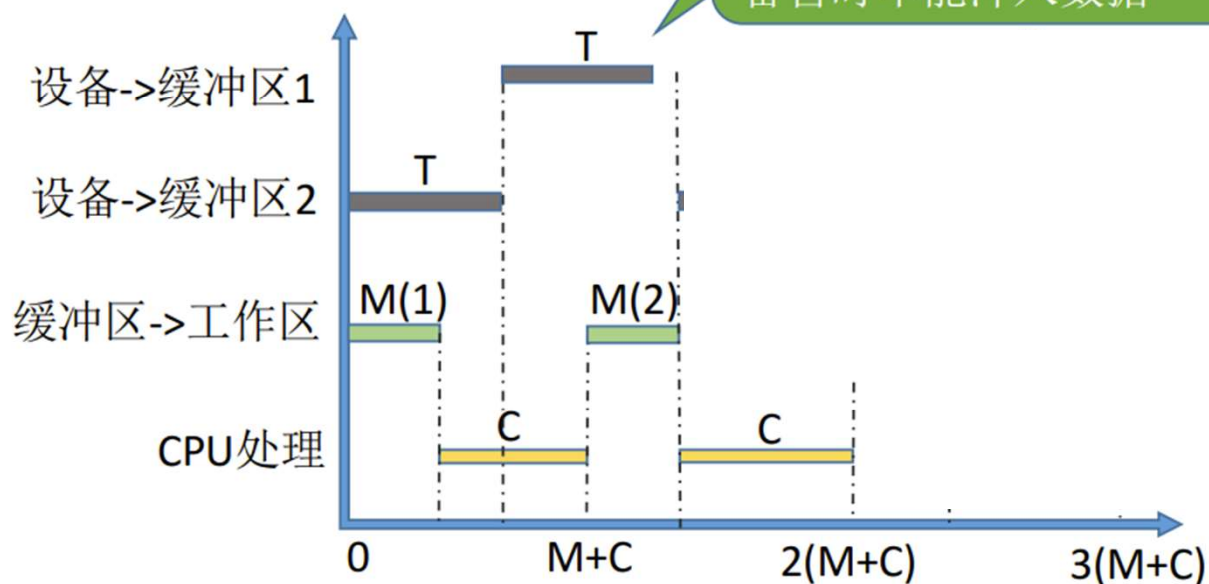
双缓冲

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 $T < C + M$

假设 $2T < 2M + C$ ，则I/O设备将缓冲区1冲满时，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



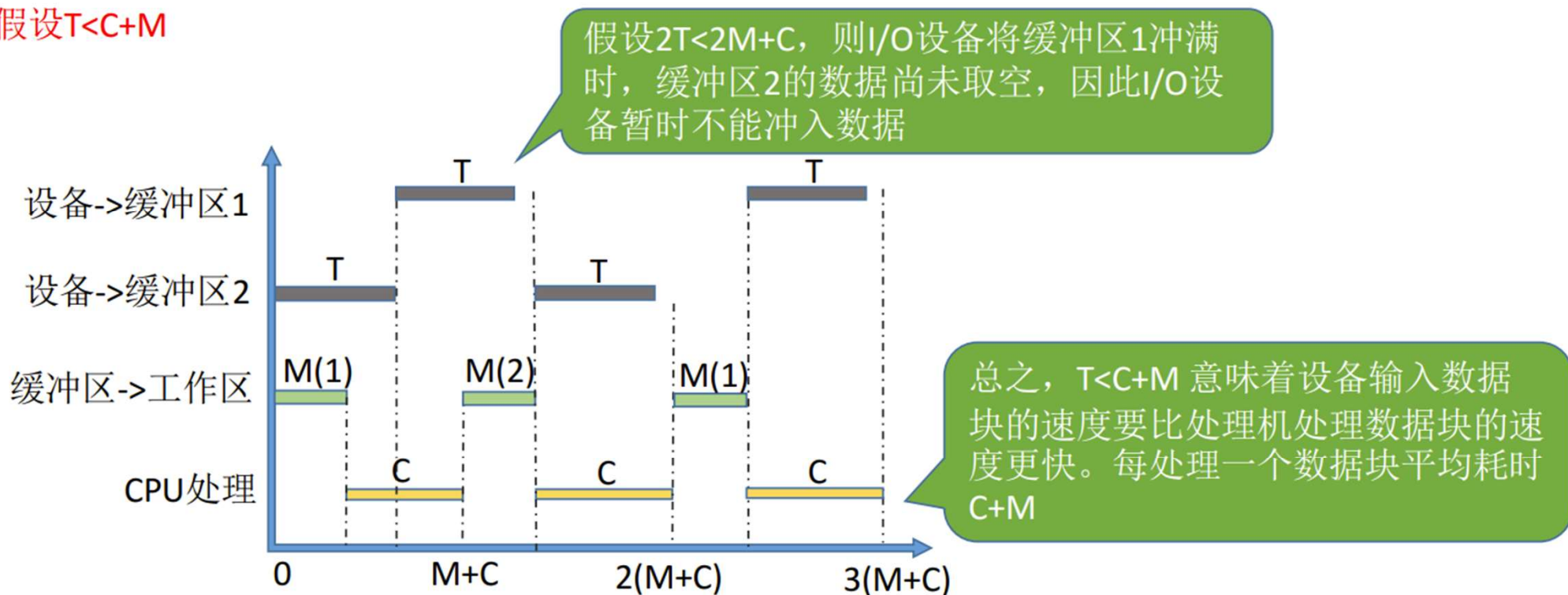
注：M(1) 表示“将缓冲区1中的数据传送到工作区”；M(2) 表示“将缓冲区2中的数据传送到工作区”

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 $T < C + M$



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”

双缓冲



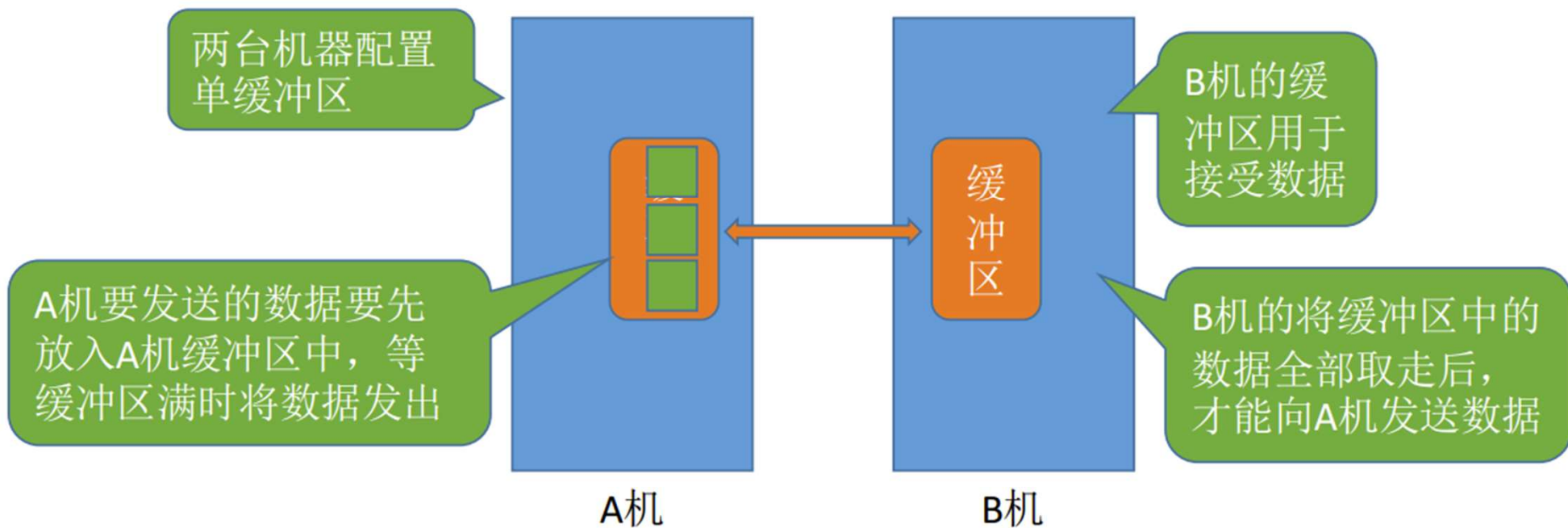
假设某用户进程请求某种块设备读入若干块的数据。若采用双缓冲的策略，操作系统会在主存中为其分配两个缓冲区（若题目中没有特别说明，一个缓冲区的大小就是一个块）

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

结论：采用双缓冲策略，处理一个数据块的平均耗时为 $\text{Max}(T, C+M)$

使用单/双缓冲在通信时的区别

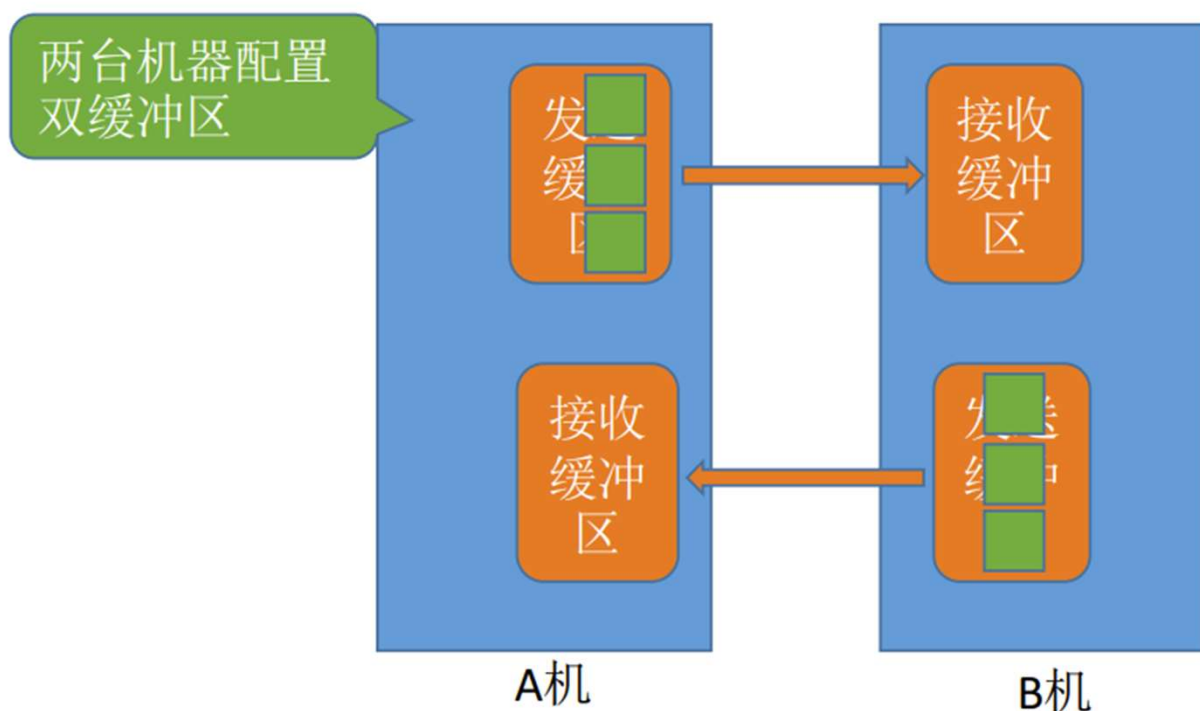
两台机器之间通信时，可以配置缓冲区用于数据的发送和接受。



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。

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。



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。

注：管道通信中的“管道”其实就是缓冲区。要实现数据的双向传输，必须设置两个管道

知识总览

一般利用内存作为缓冲区

缓冲区的概念

缓解CPU与设备的速度矛盾、减少对CPU的中断频率、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

设备—(T)—缓冲区—(M)—工作区—(C)—处理

单缓冲

处理一块数据平均耗时 $\text{Max}(C, T) + M$

分析问题的初始状态：工作区满，缓冲区空

双缓冲

处理一块数据平均耗时 $\text{Max}(T, C + M)$

分析问题的初始状态：工作区空，一个缓冲区满，另一个缓冲区空

循环缓冲

多个缓冲区链接成循环队列，in指针指向第一个空缓冲区，out指针指向第一个满缓冲区

缓冲池

三个队列：空缓冲队列、输入队列、输出队列

四种工作缓冲区

用于收容输入数据的工作缓冲区、用于提取输入数据的工作缓冲区

用于收容输出数据的工作缓冲区、用于提取输出数据的工作缓冲区

缓冲区管理

设备管理 基本概念

Linux
Android
Linux
OpenStack
Mac OS
Windows

