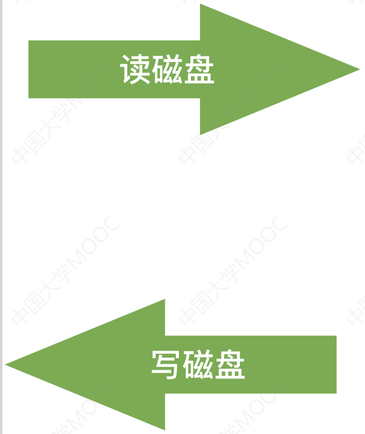
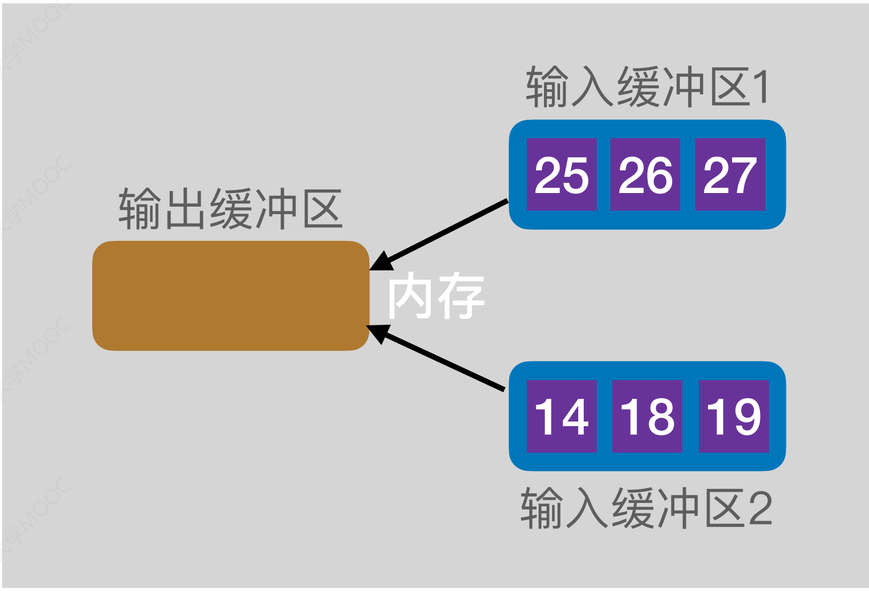
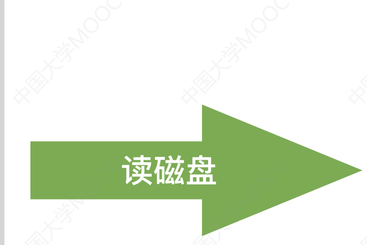
第二趟归并

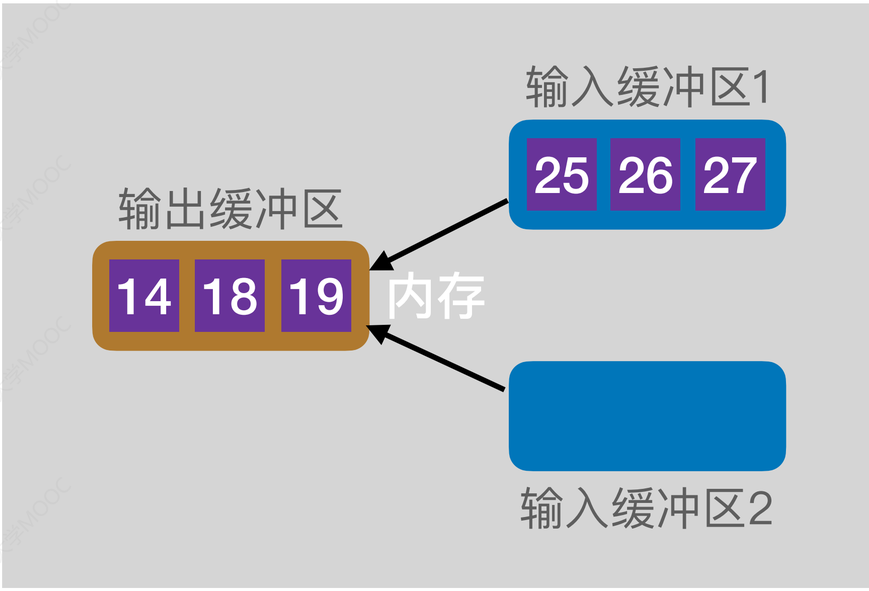




将两个有序归并段归并为一个

第二趟归并



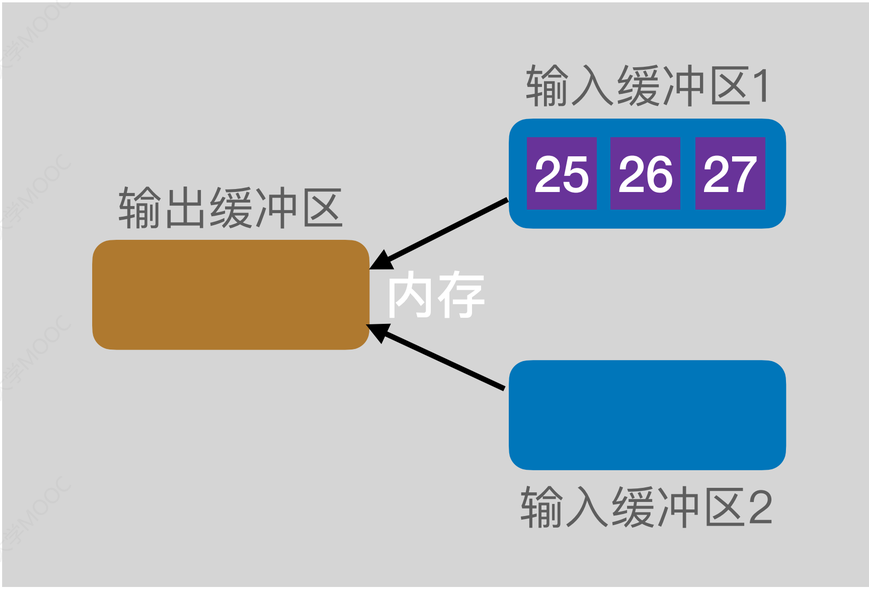


将两个有序归并段归并为一个

第二趟归并

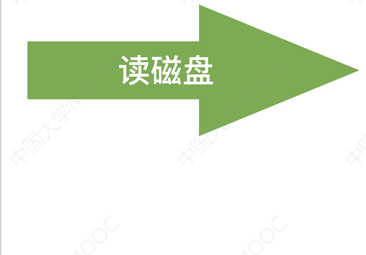
读磁盘

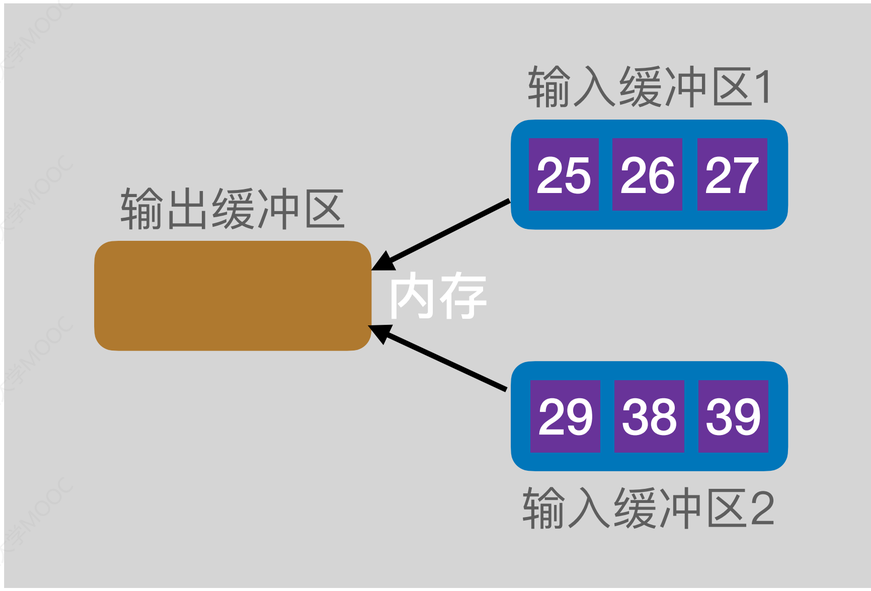
写磁盘



将两个有序归并段归并为一个

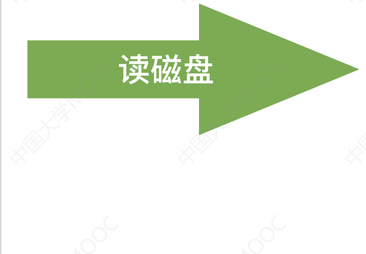
第二趟归并

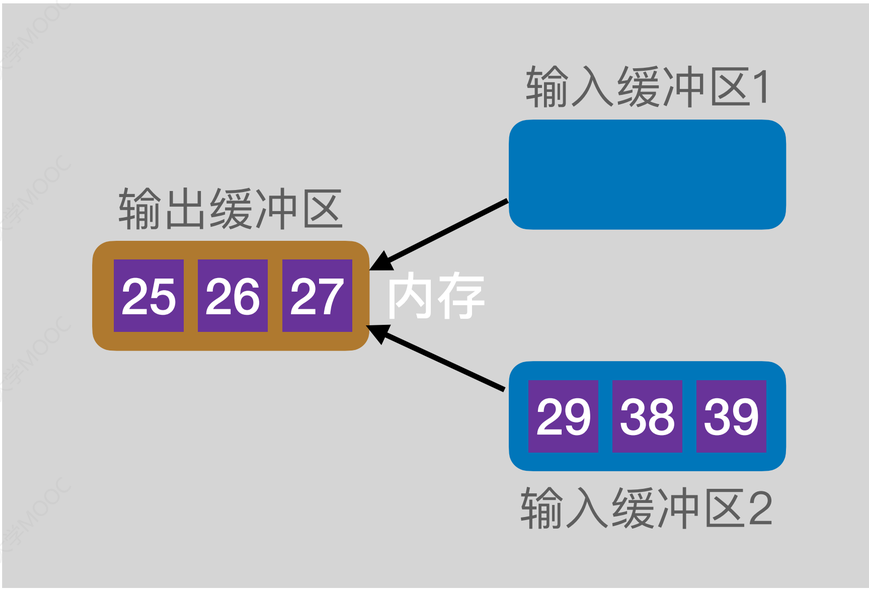




将两个有序归并段归并为一个

第二趟归并



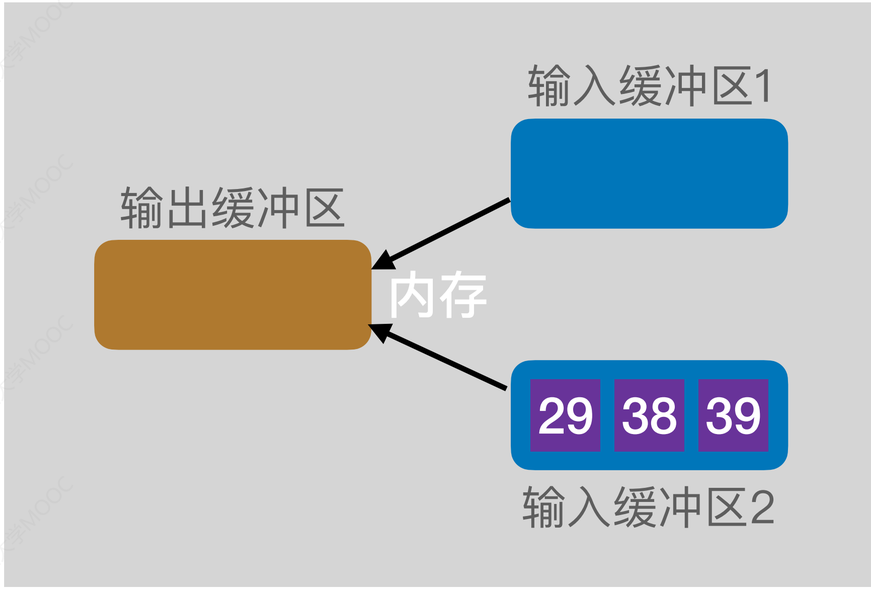


将两个有序归并段归并为一个

第二趟归并

读磁盘

写磁盘

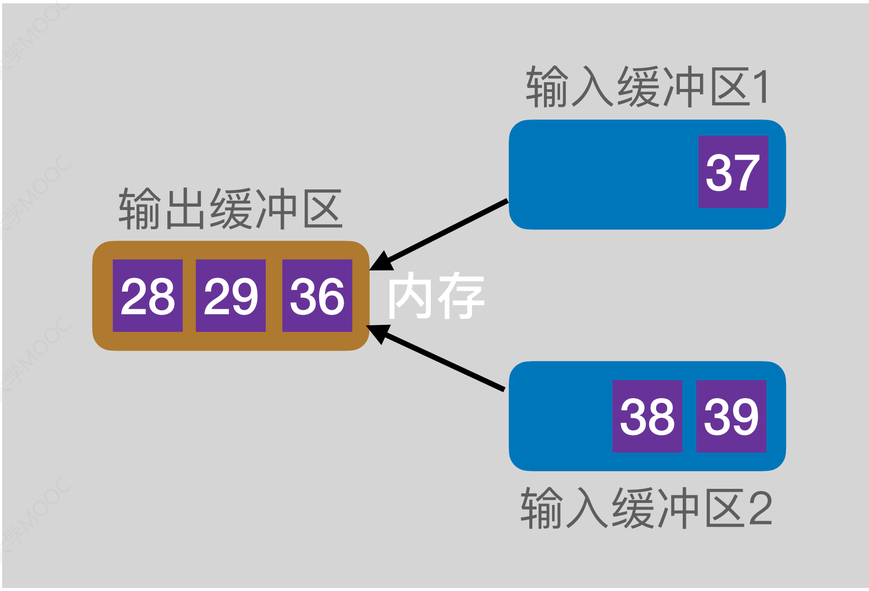


将两个有序归并段归并为一个

第二趟归并

将两个有序归并段归并为一个

第二趟归并

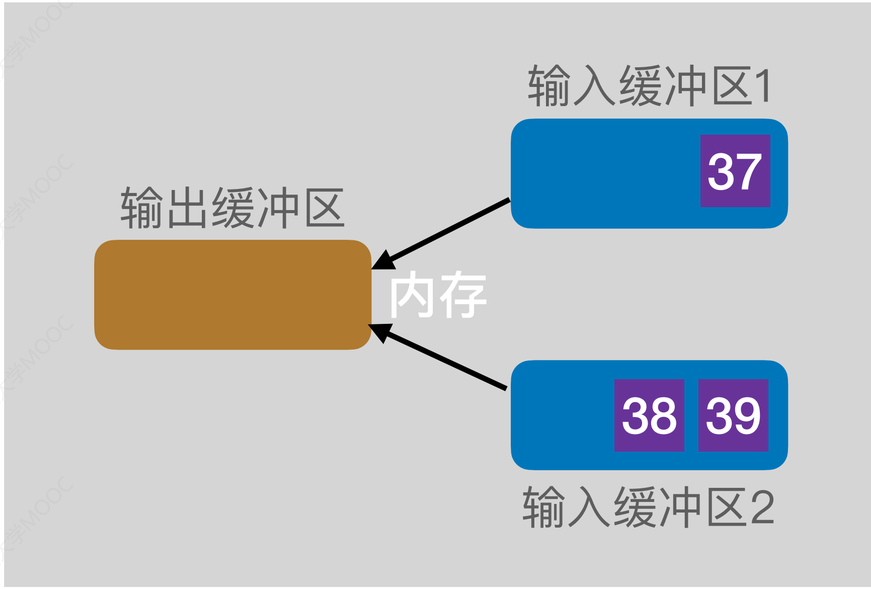


将两个有序归并段归并为一个

第二趟归并

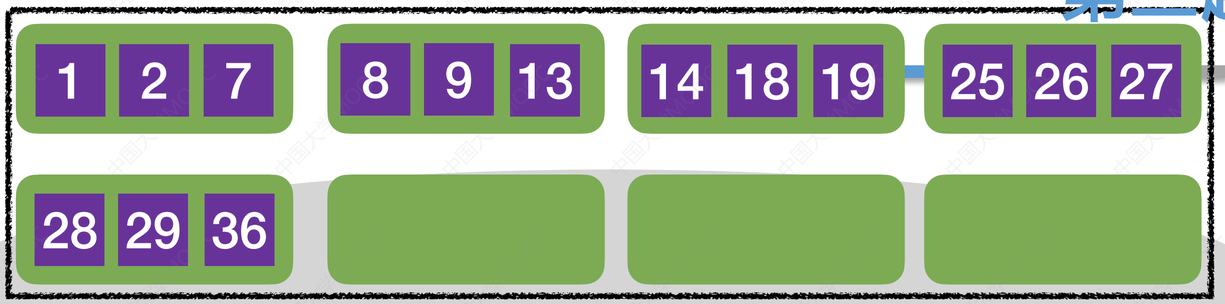
读磁盘

写磁盘

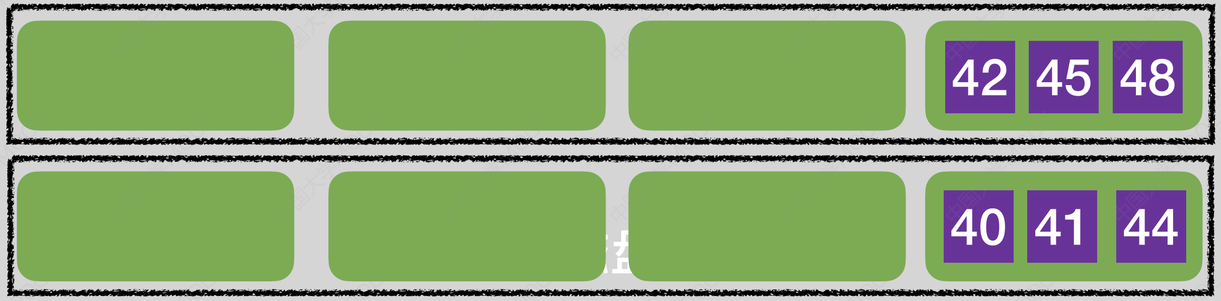


将两个有序归并段归并为一个

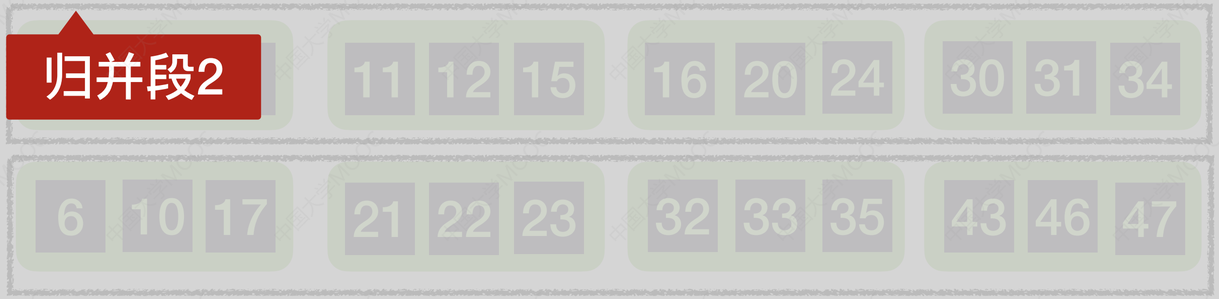
第二趟归并



归并段1

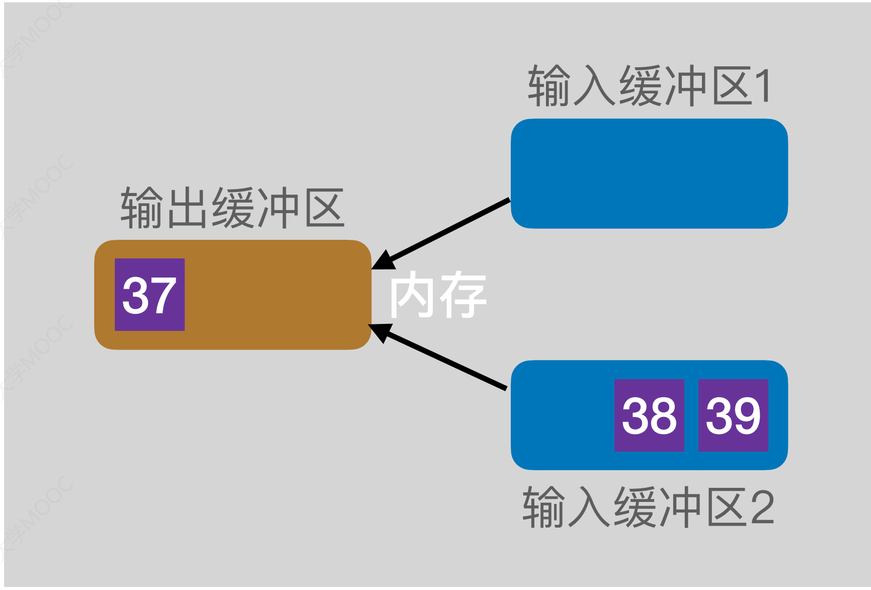


归并段2



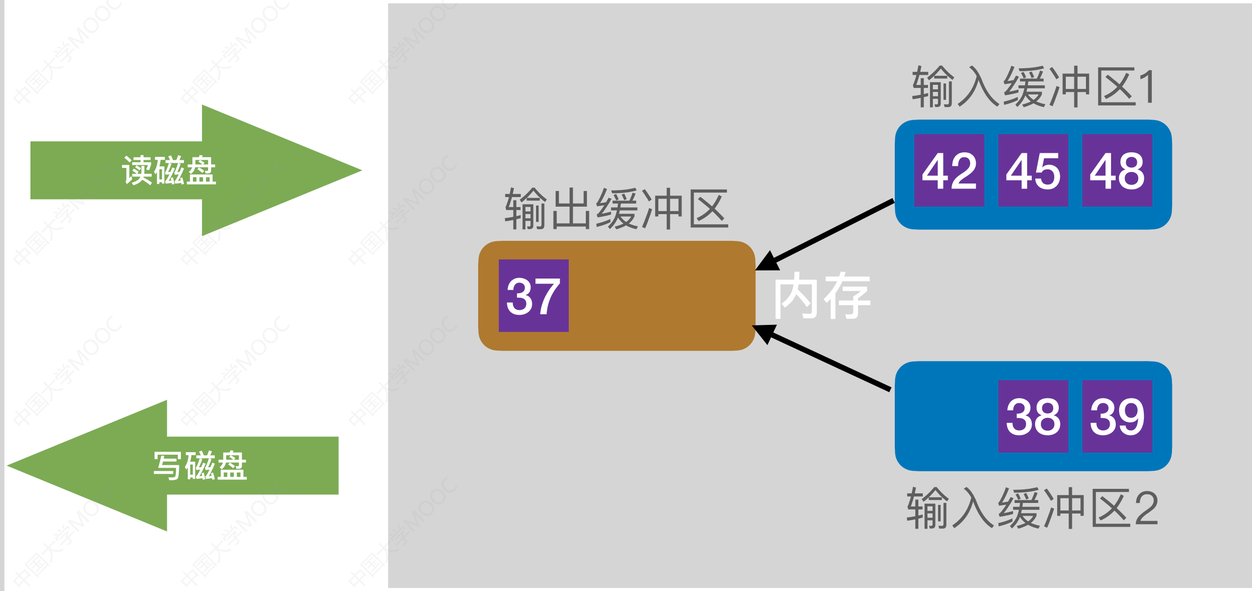
读磁盘

写磁盘



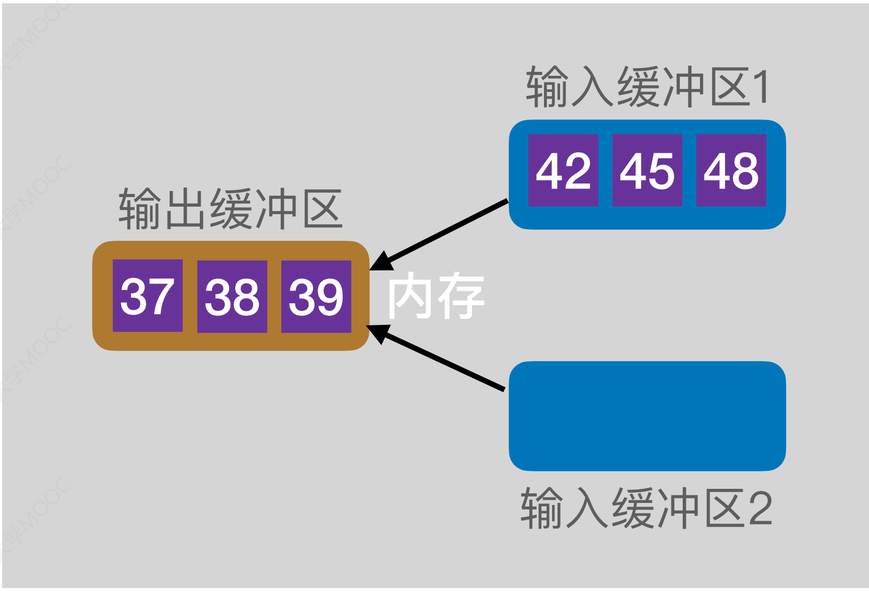
将两个有序归并段归并为一个

第二趟归并



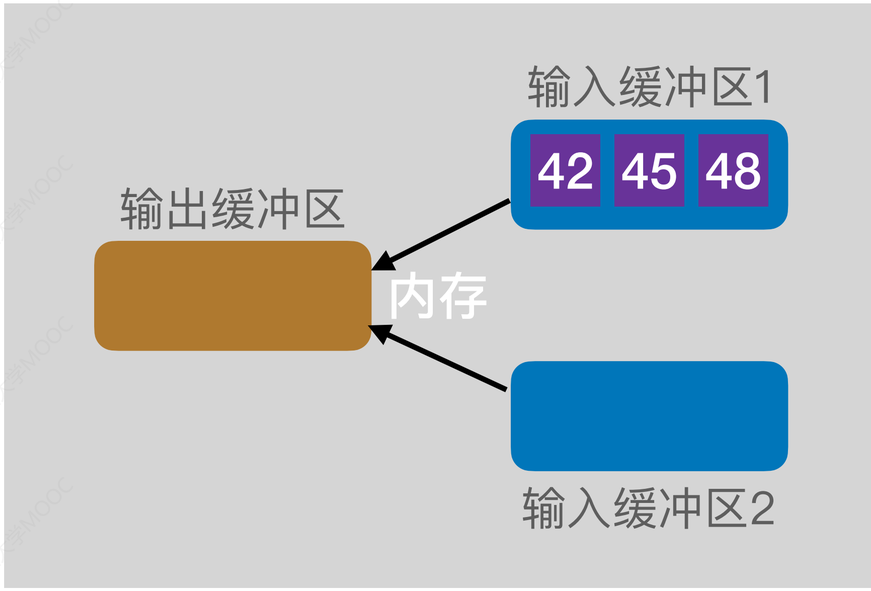
将两个有序归并段归并为一个

第二趟归并



将两个有序归并段归并为一个

第二趟归并



将两个有序归并段归并为一个

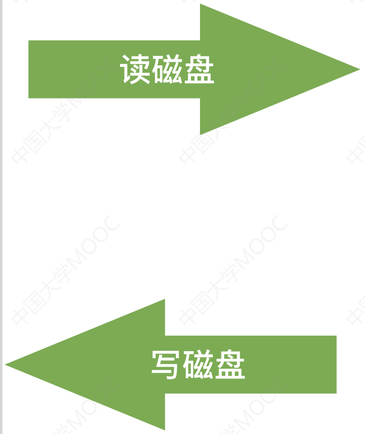
第二趟归并

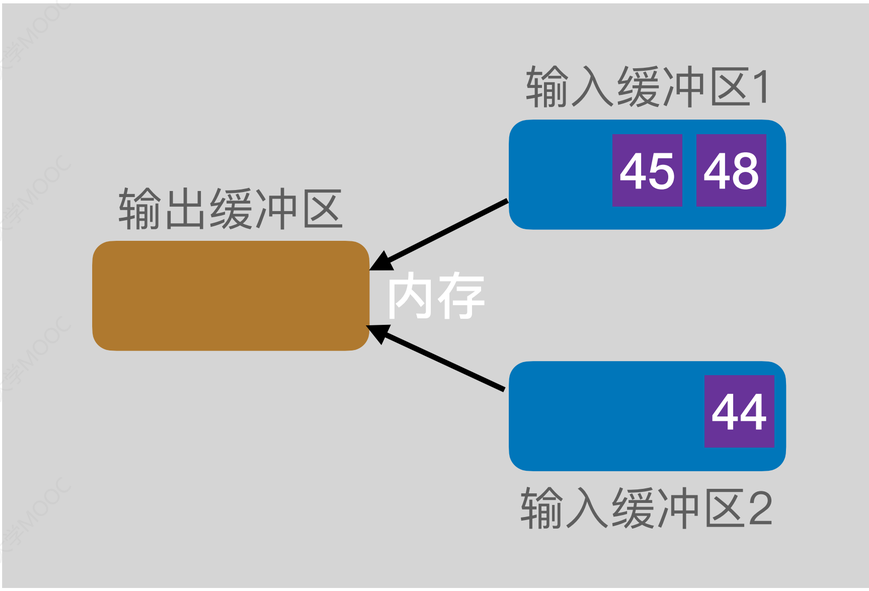
将两个有序归并段归并为一个

第二趟归并

将两个有序归并段归并为一个

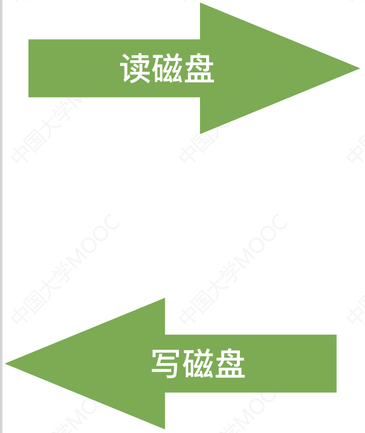
第二趟归并

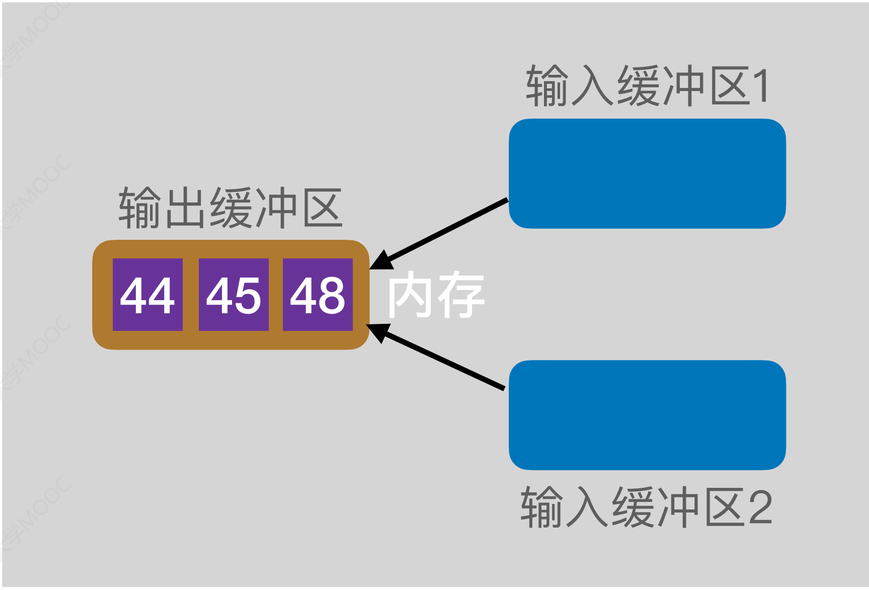




将两个有序归并段归并为一个

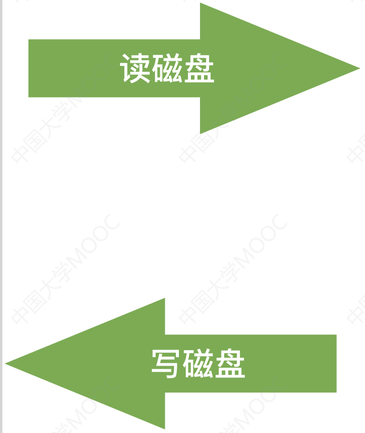
第二趟归并

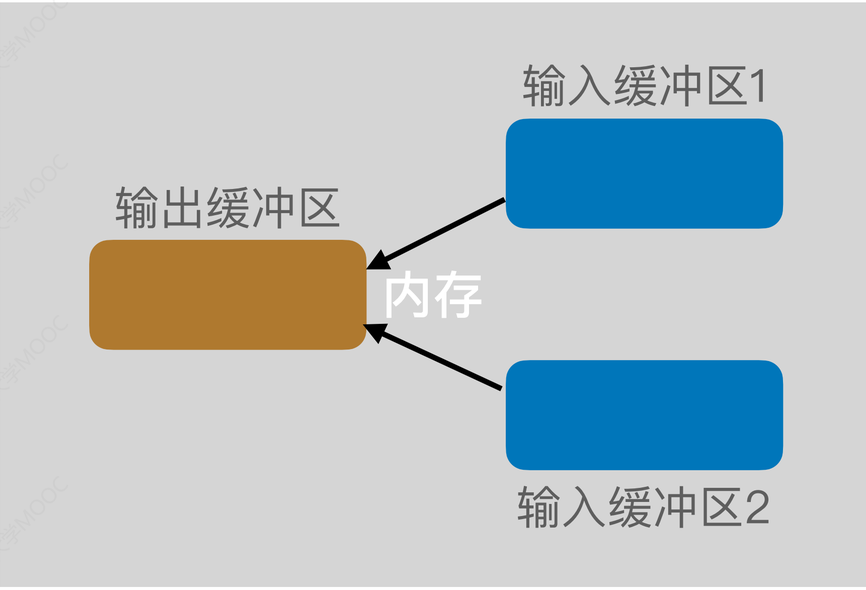




将两个有序归并段归并为一个

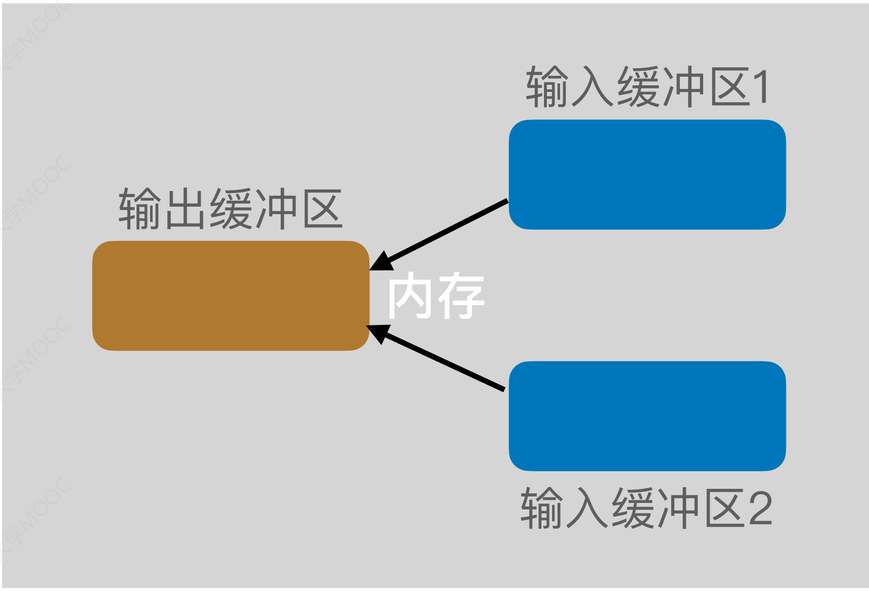
第二趟归并





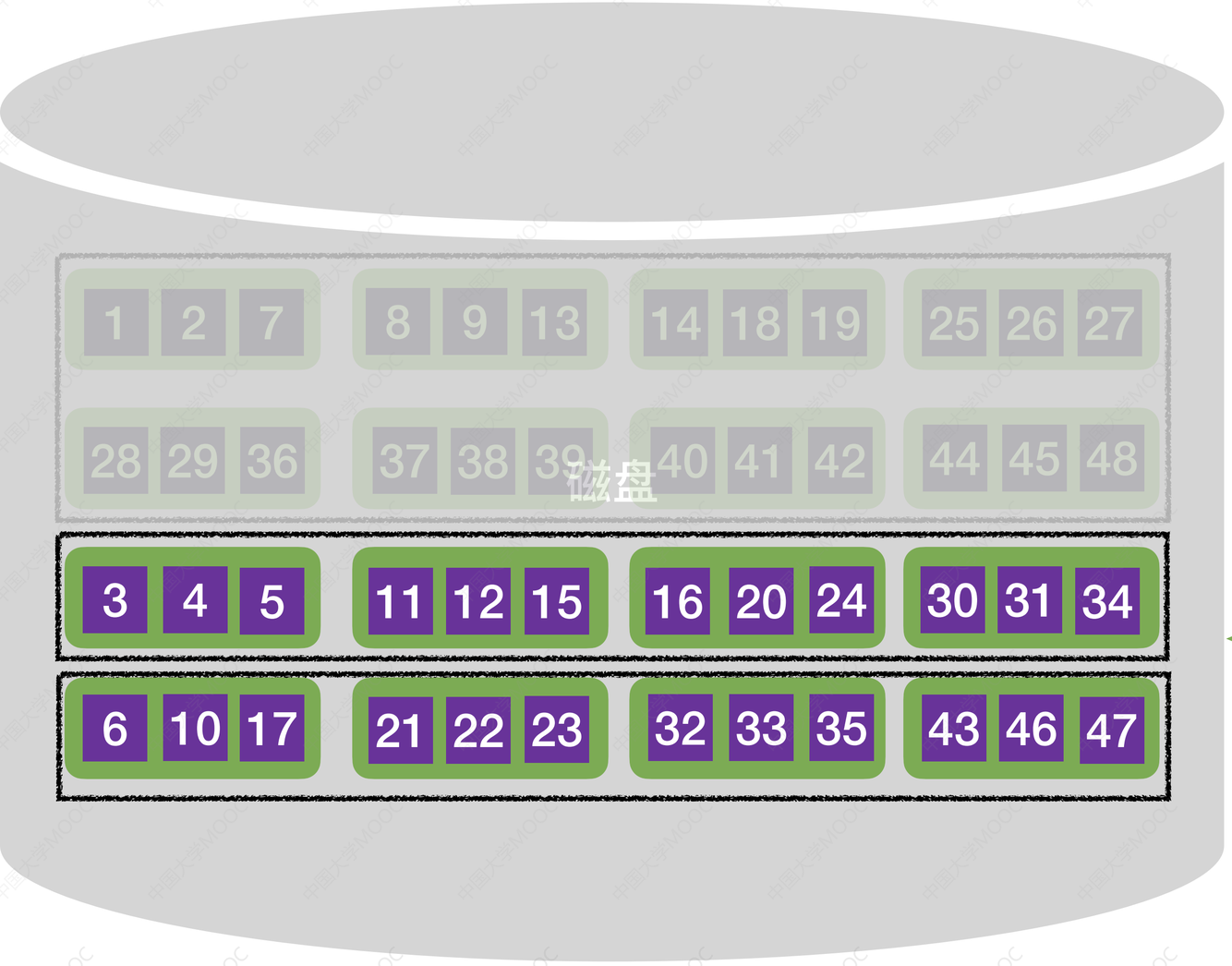
将两个有序归并段归并为一个

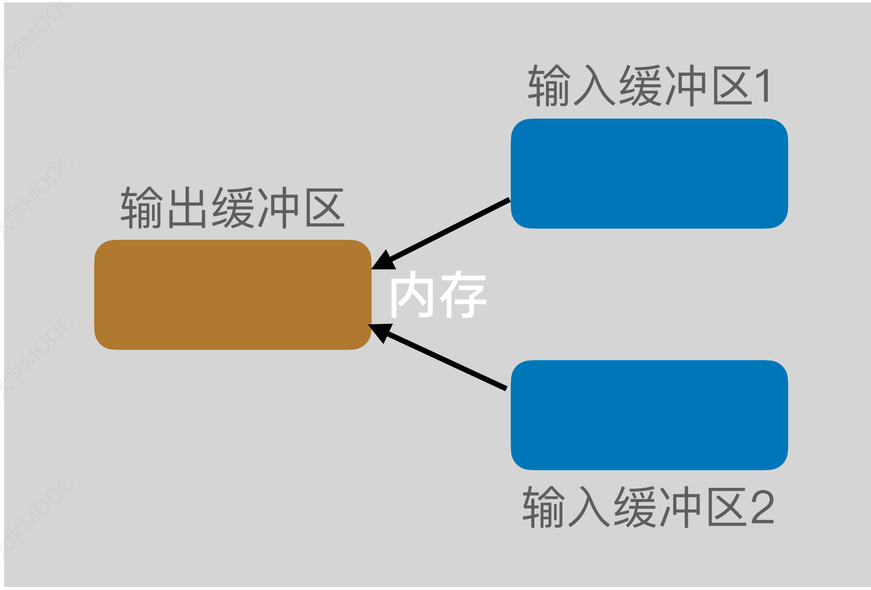
第二趟归并



将两个有序归并段归并为一个

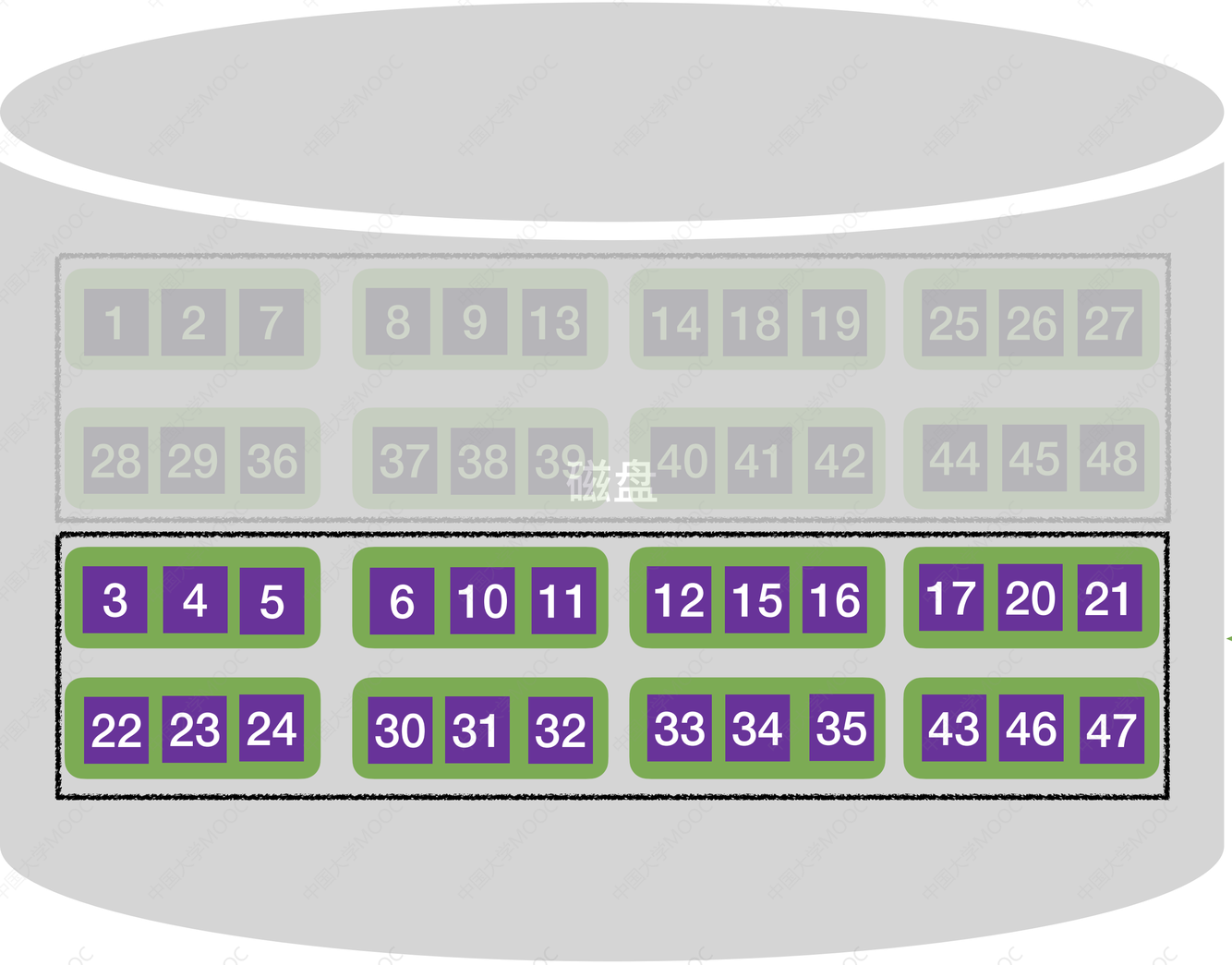
第二趟归并

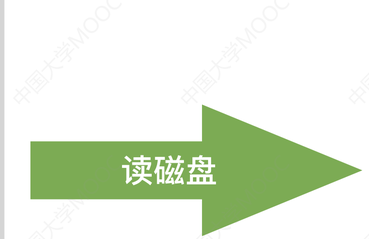


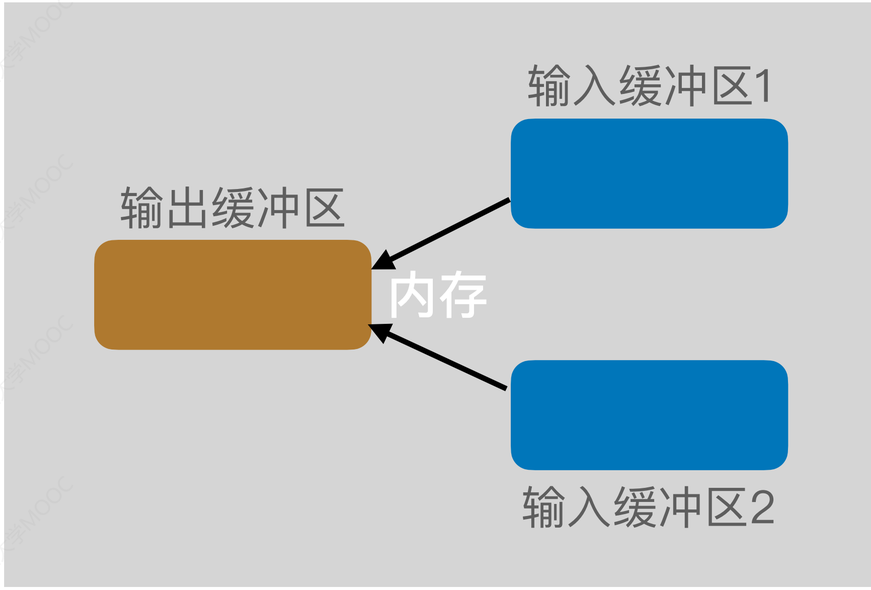


将两个有序归并段归并为一个

第二趟归并

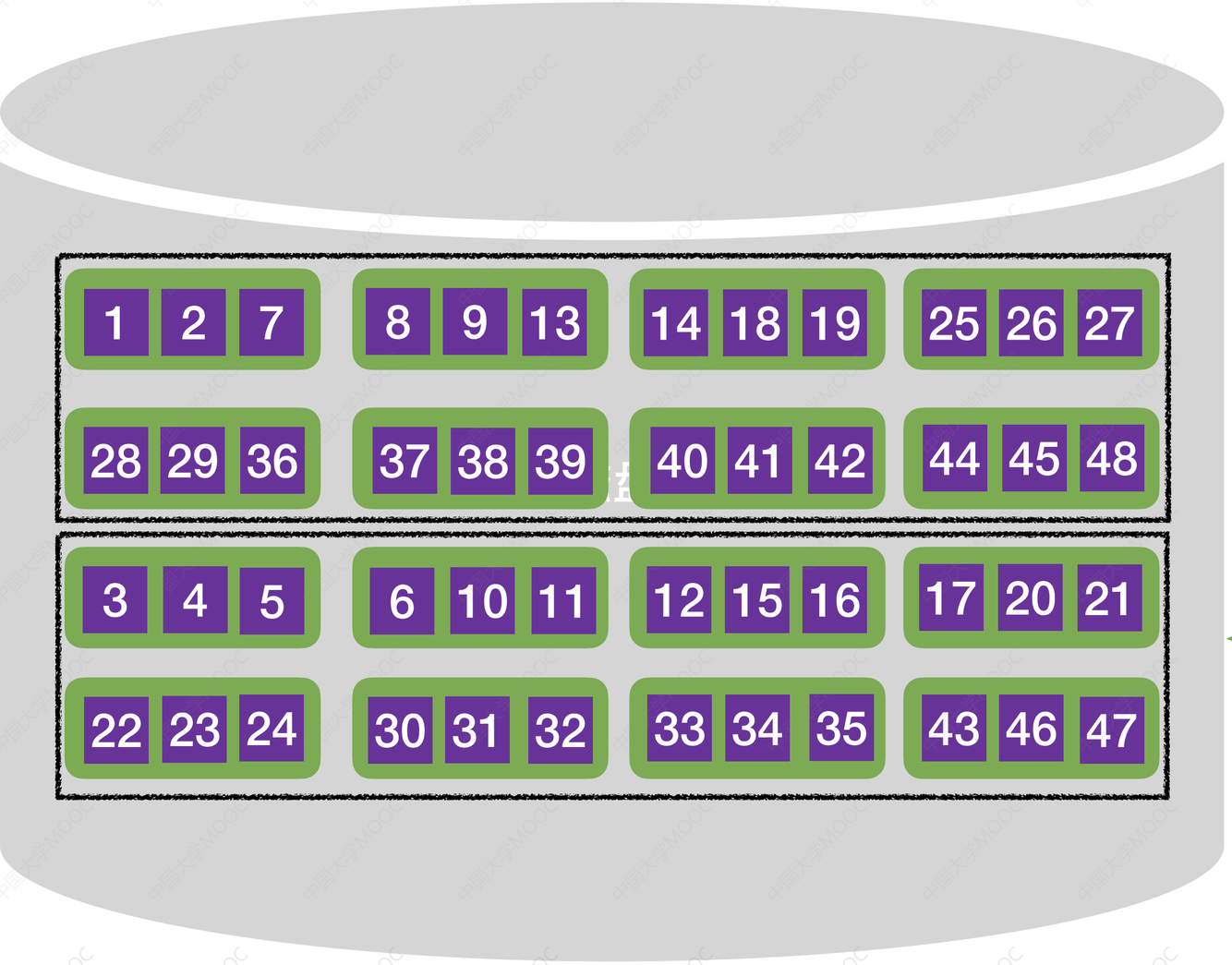


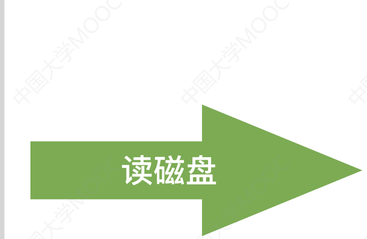


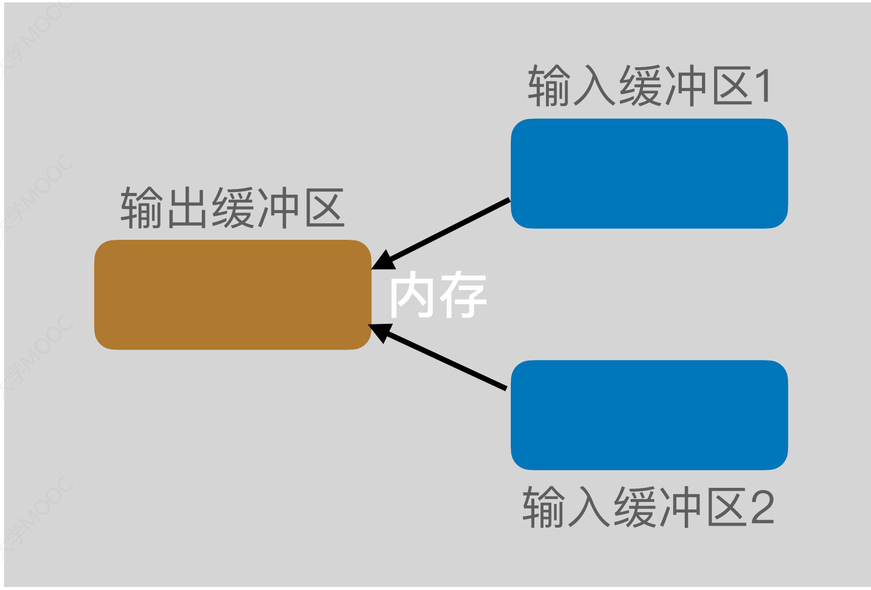


将两个有序归并段归并为一个

第二趟归并



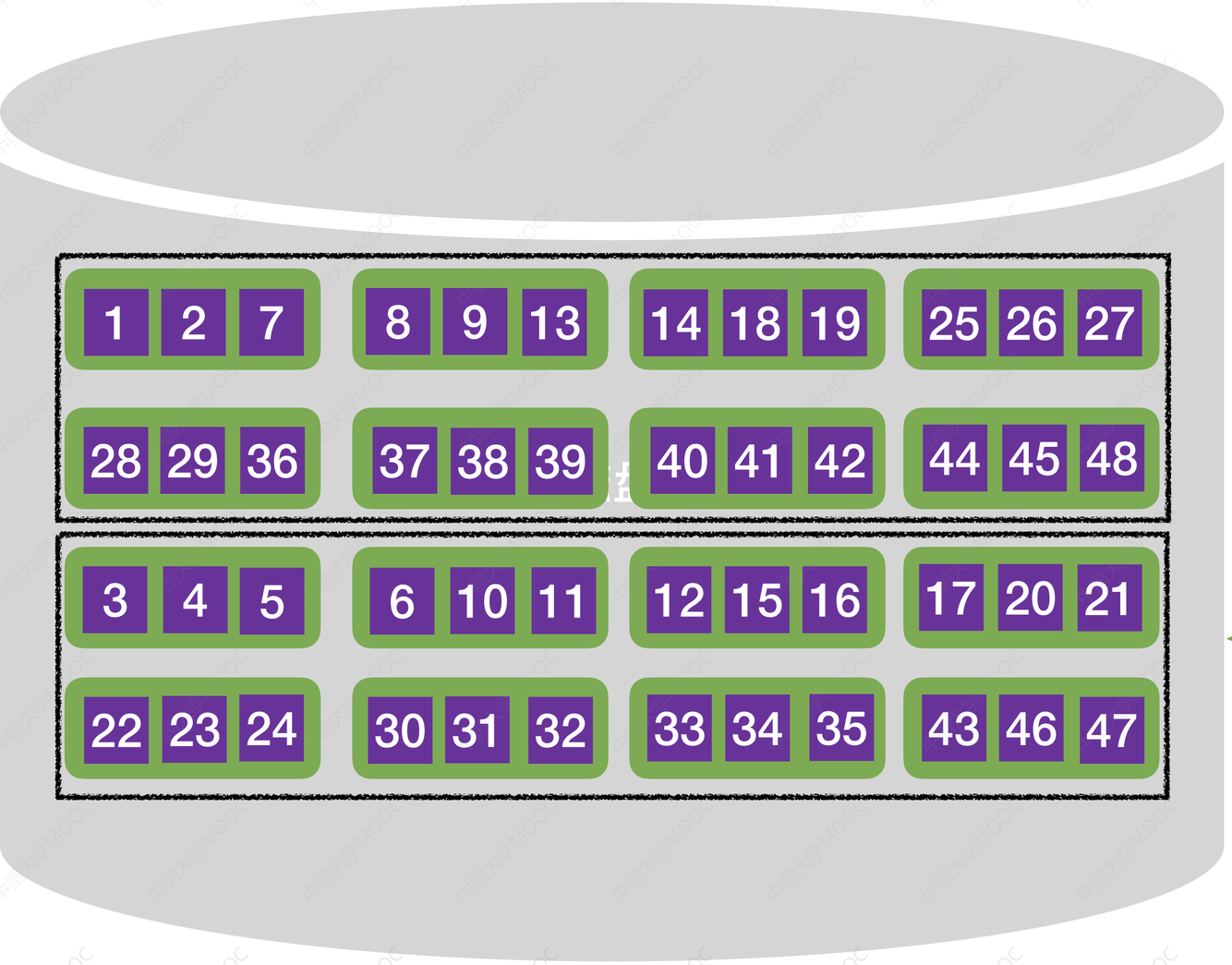


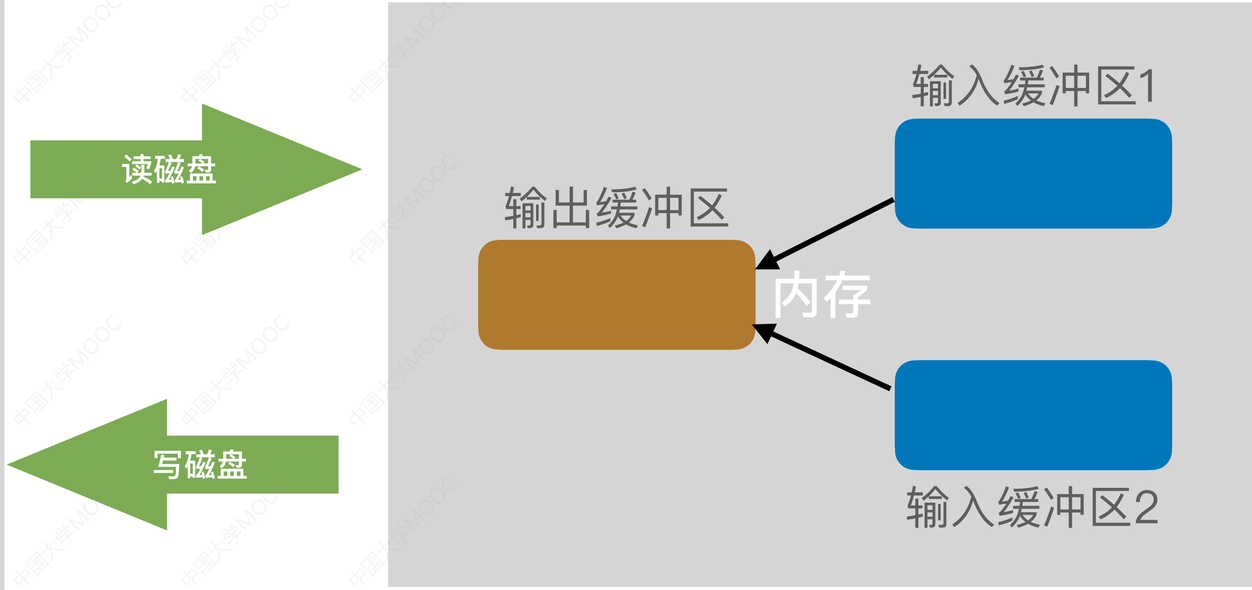


将两个有序归并段归并为一个

第三趟归并

把2个有序子序列（归并段）归并





将两个有序归并段归并为一个

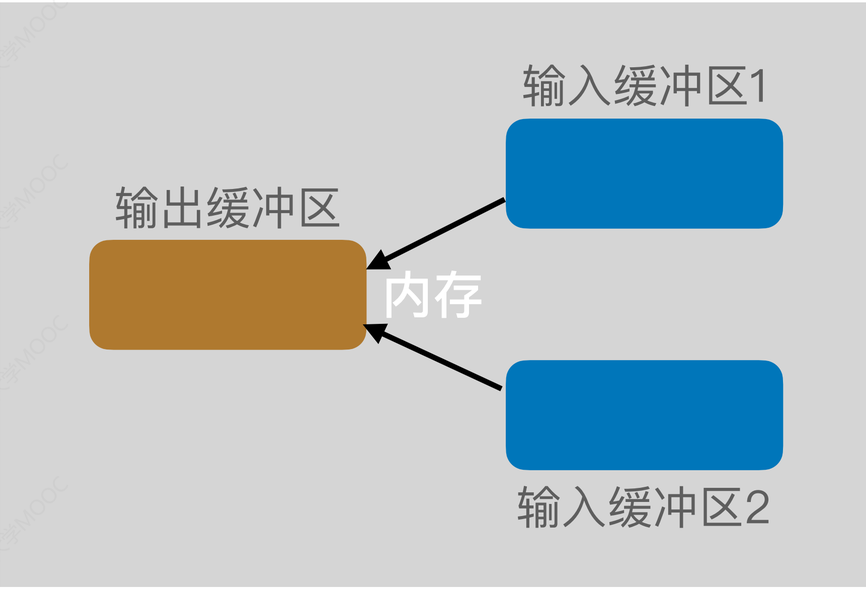
第三趟归并

经过3趟归并，整体有序

1	2	3	4	5	6	7	8	9	10	11	12
13	14	15	16	17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32	33	34	35	36
37	38	39	40	41	42	43	44	45	46	47	48

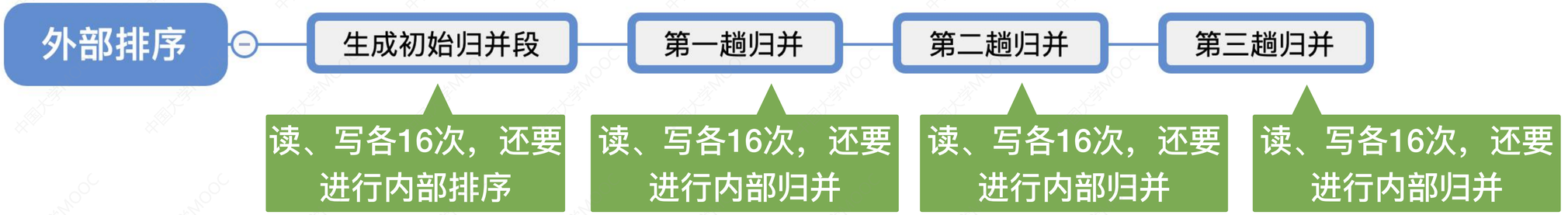
读磁盘

写磁盘

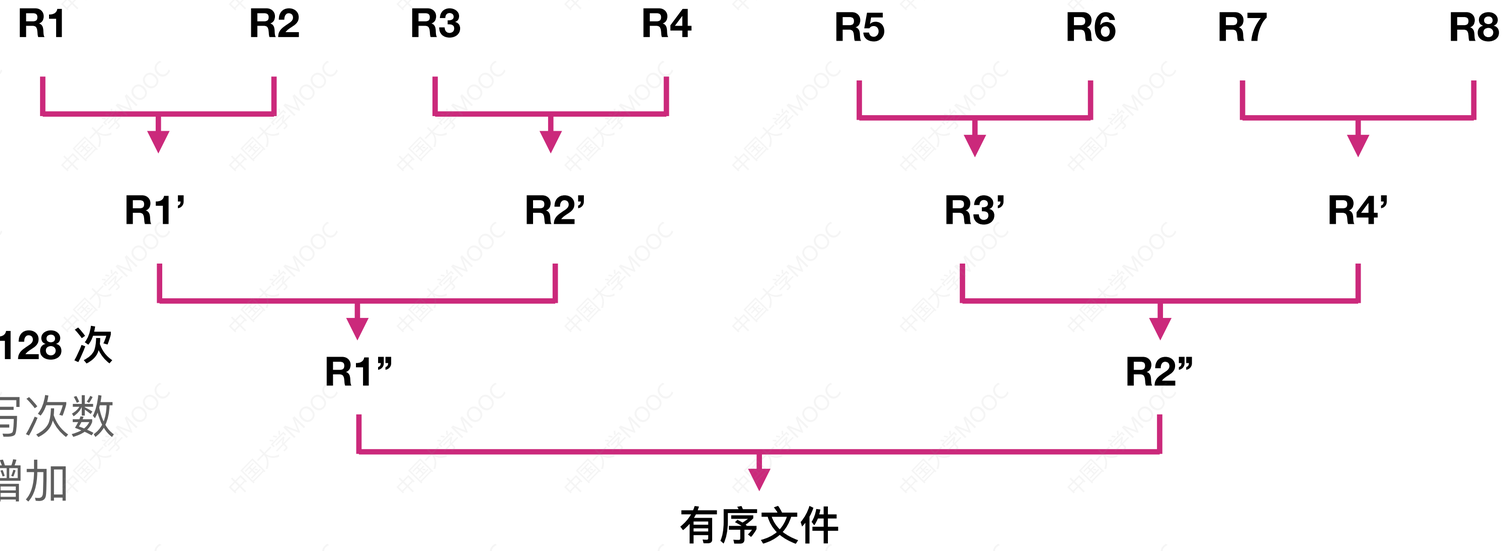


将两个有序归并段归并为一个

时间开销分析



生成8个初始归并段，每个段占两块



文件总块数*2　　归并趟数

读、写磁盘次数= 32+32*3 = 128 次

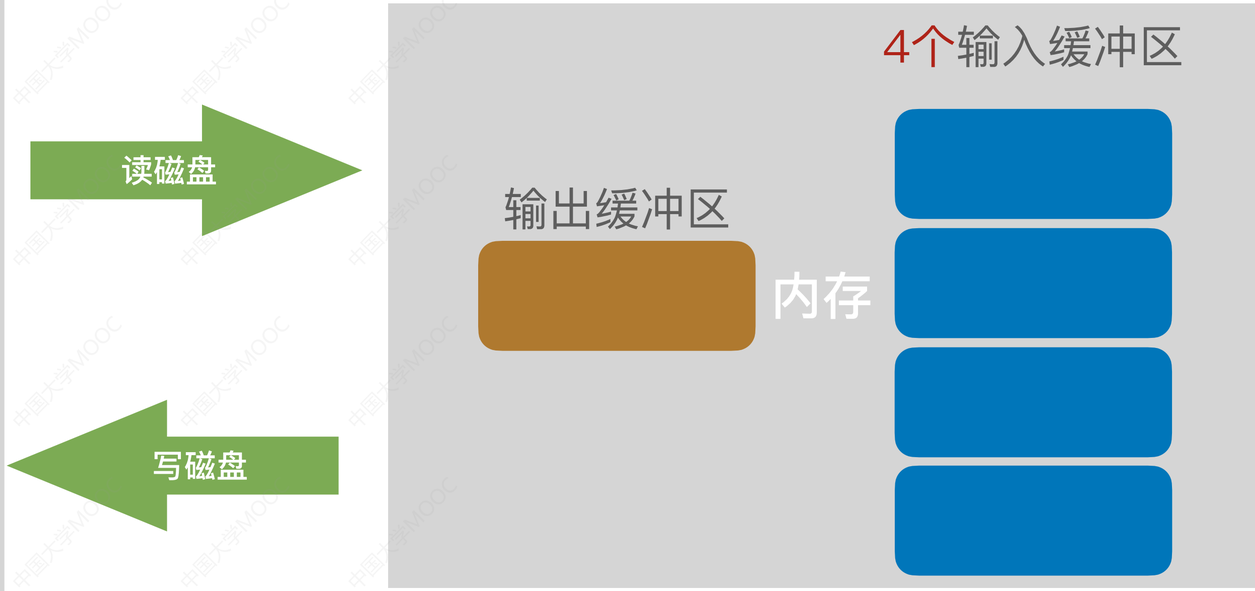
注：磁盘是慢速设备，读写次数太多导致时间开销大幅增加

外部排序时间开销=读写外存的时间+内部排序所需时间+内部归并所需时间

如何优化?

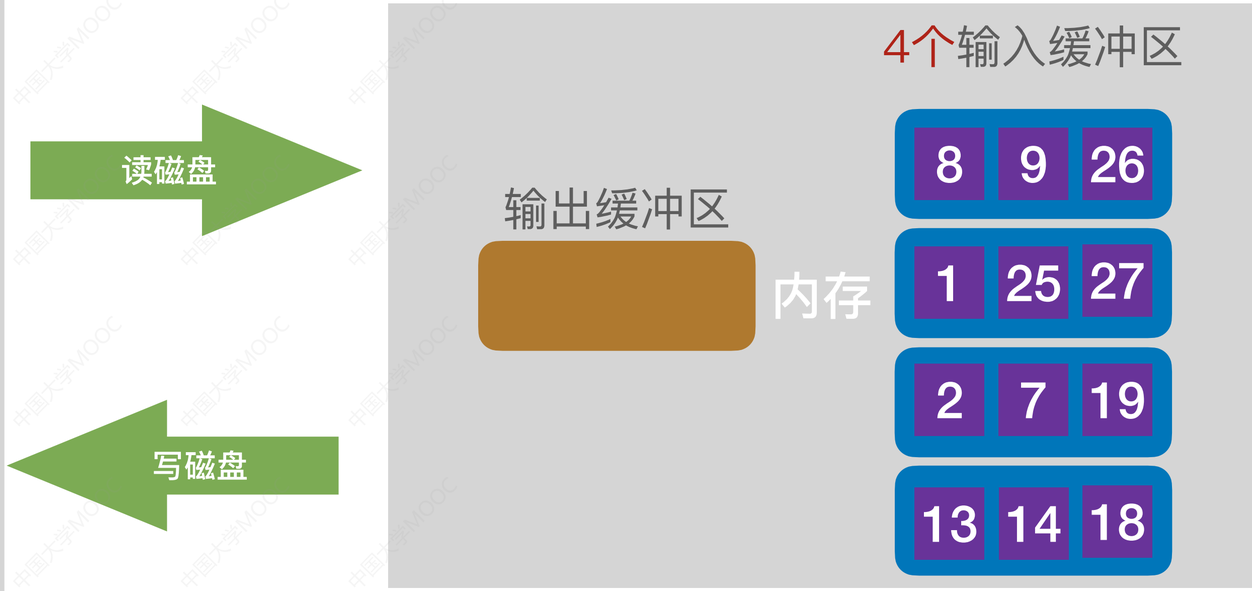
外部排序时间开销=读写外存的时间+内部排序所需时间+内部归并所需时间

优化：多路归并



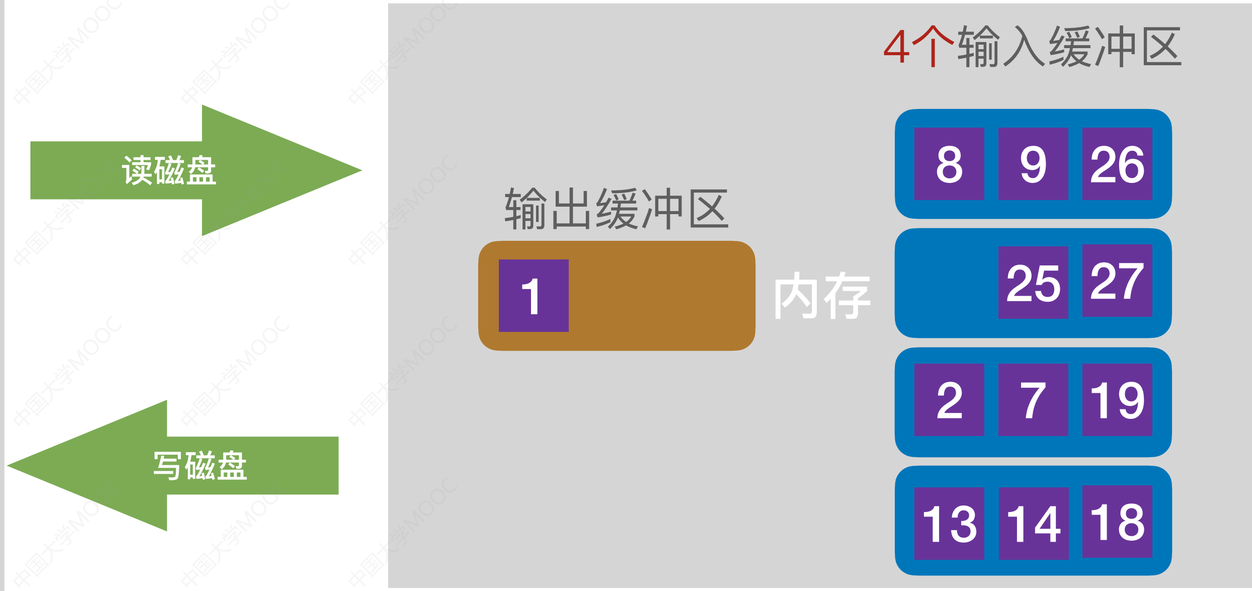
将4个有序归并段归并为一个

优化：多路归并



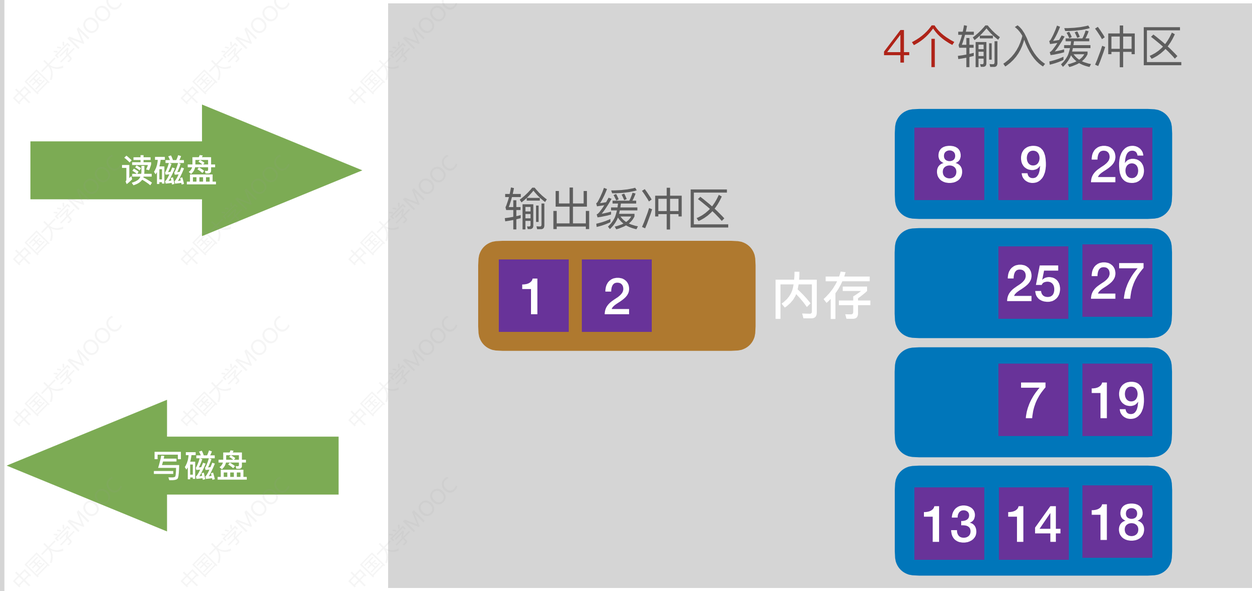
将4个有序归并段归并为一个

优化：多路归并



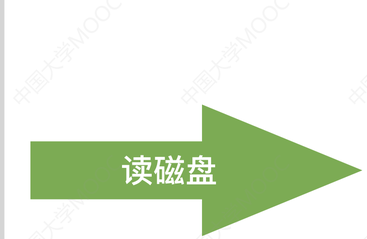
将4个有序归并段归并为一个

优化：多路归并



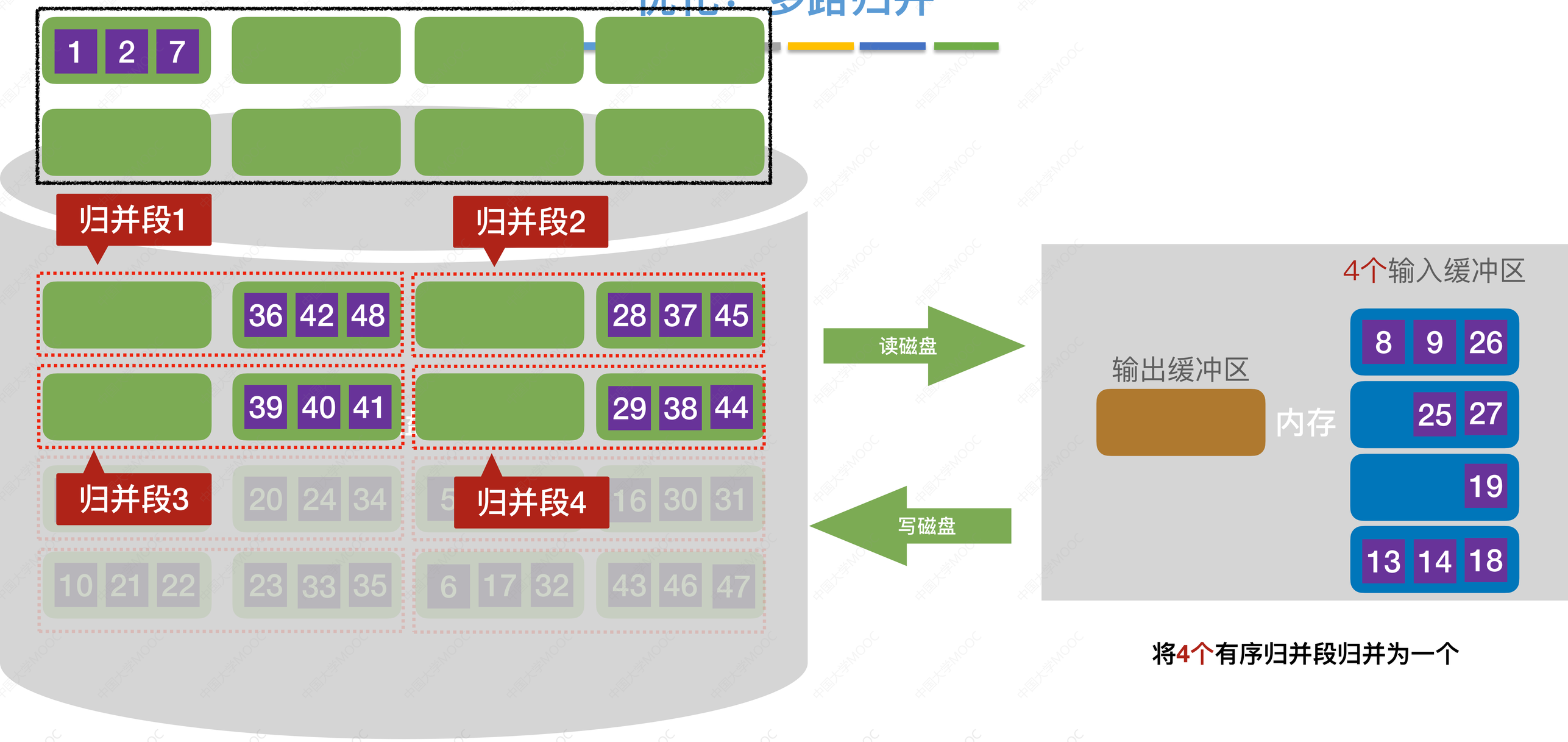
将4个有序归并段归并为一个

优化：多路归并



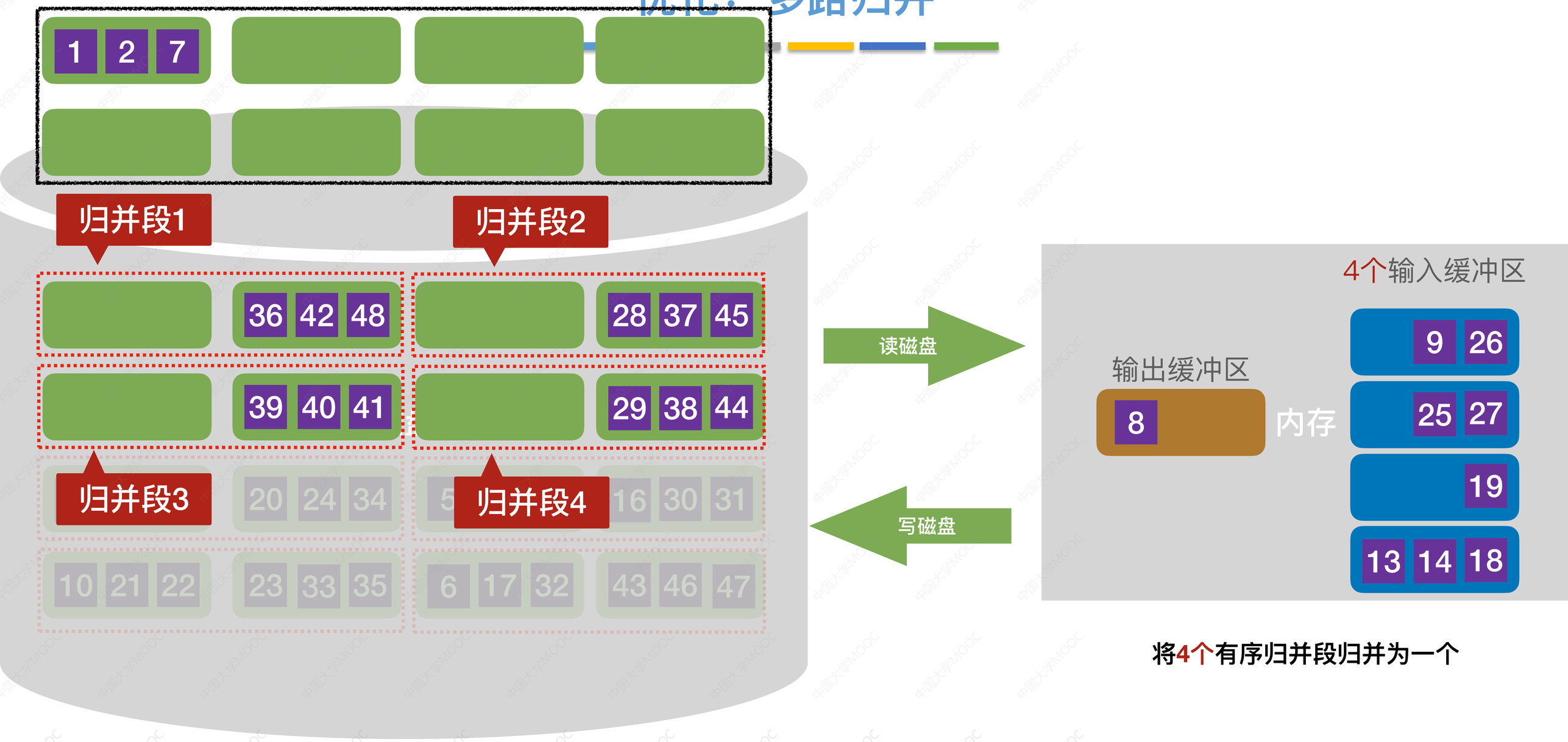
将4个有序归并段归并为一个

优化：多路归并



将4个有序归并段归并为一个

优化：多路归并

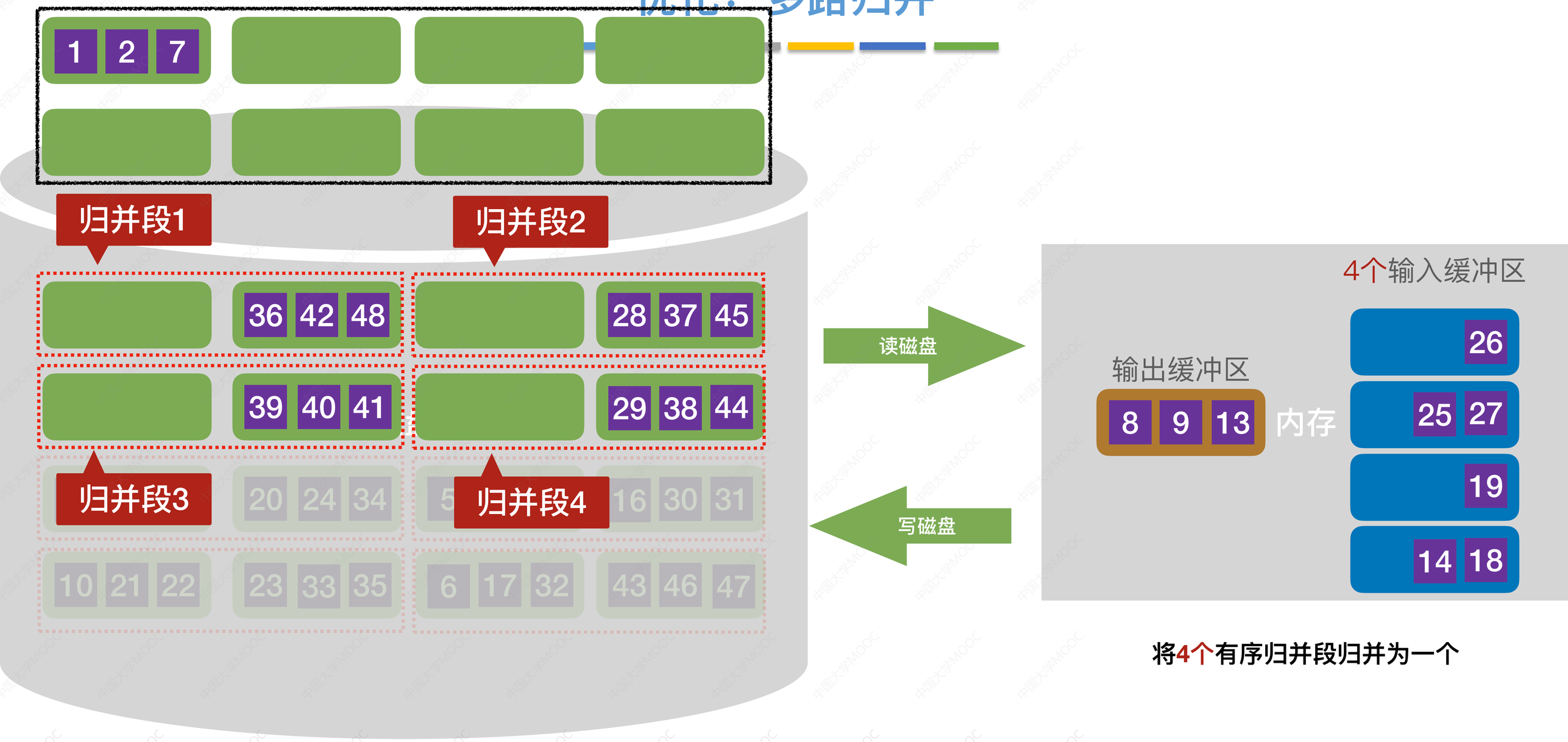


将4个有序归并段归并为一个

优化：多路归并

将4个有序归并段归并为一个

优化：多路归并

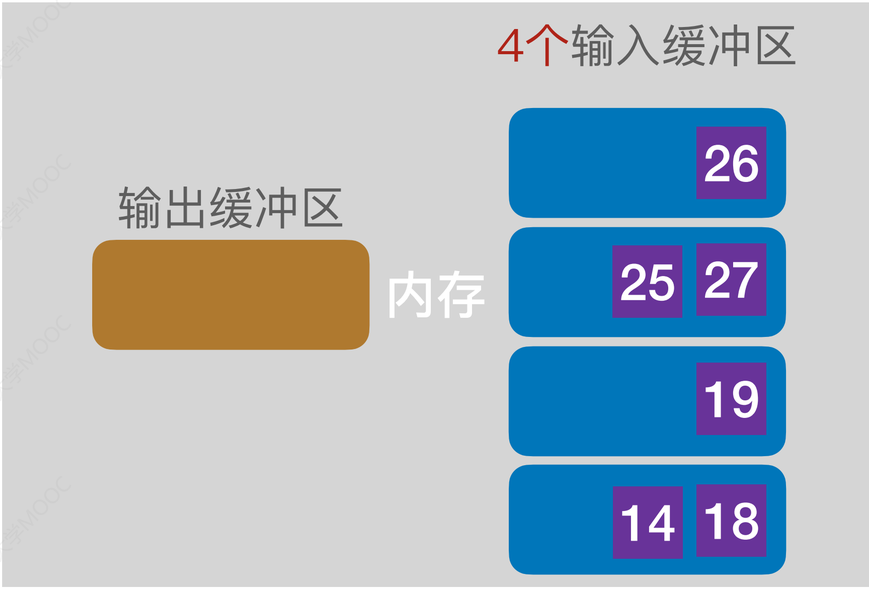


将4个有序归并段归并为一个

优化：多路归并

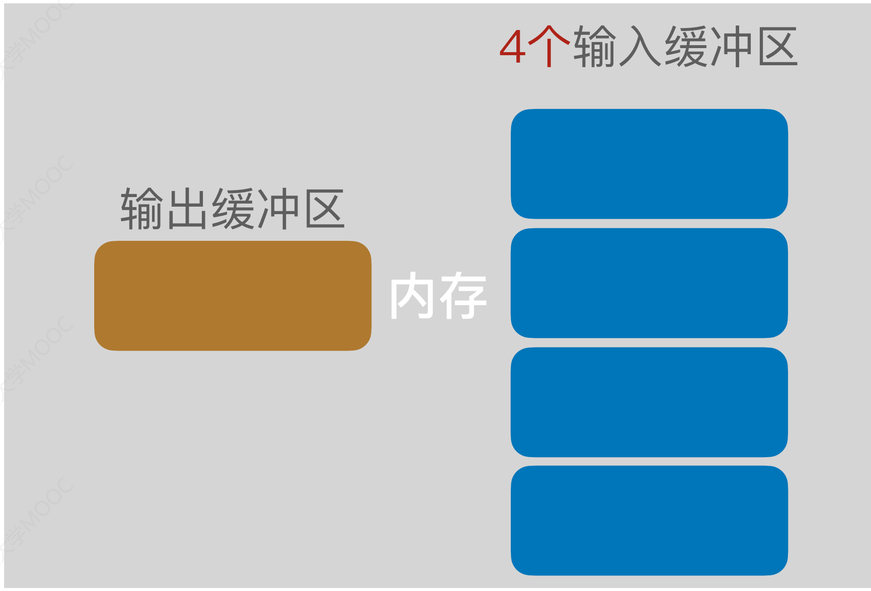
读磁盘

写磁盘



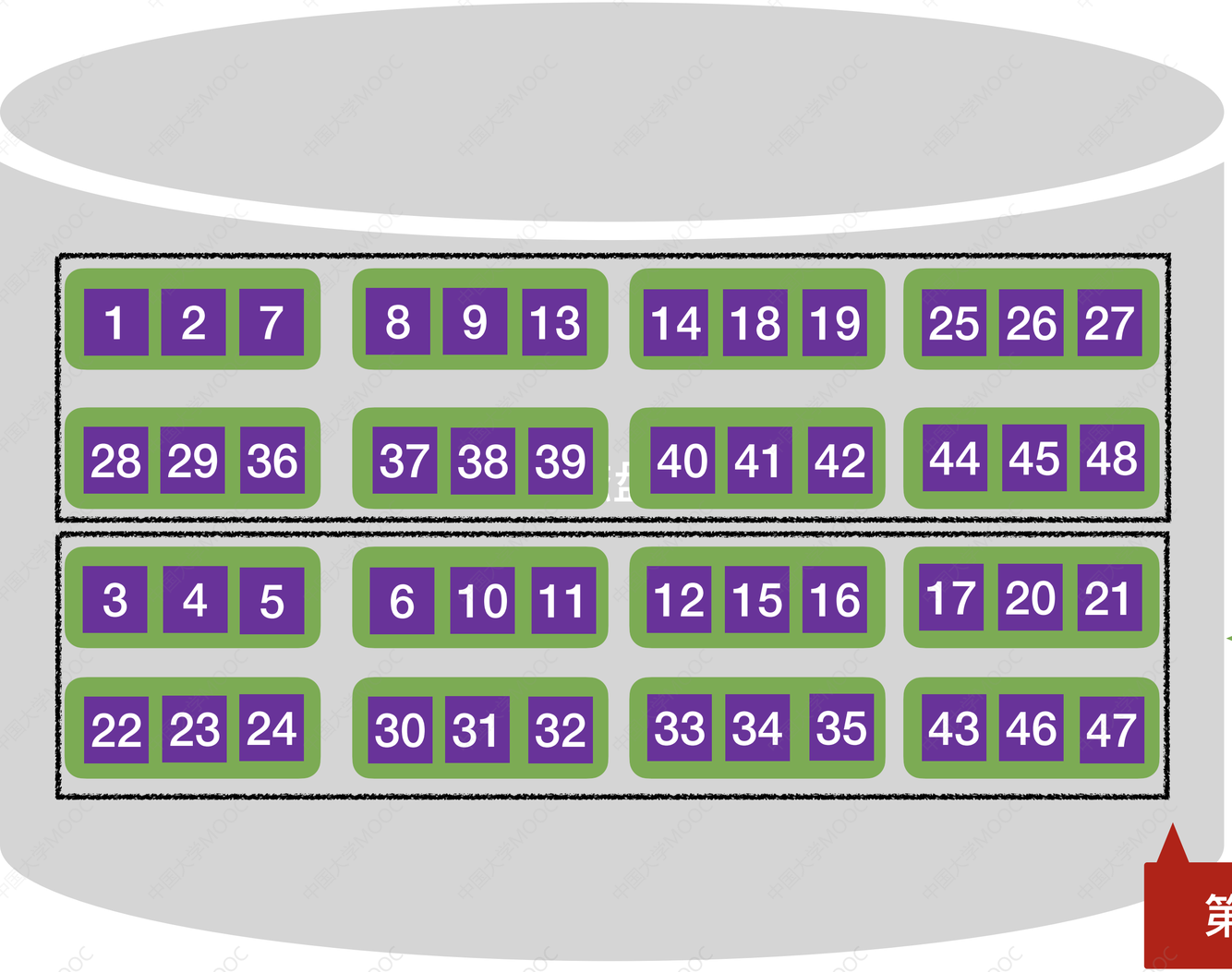
将4个有序归并段归并为一个

优化：多路归并



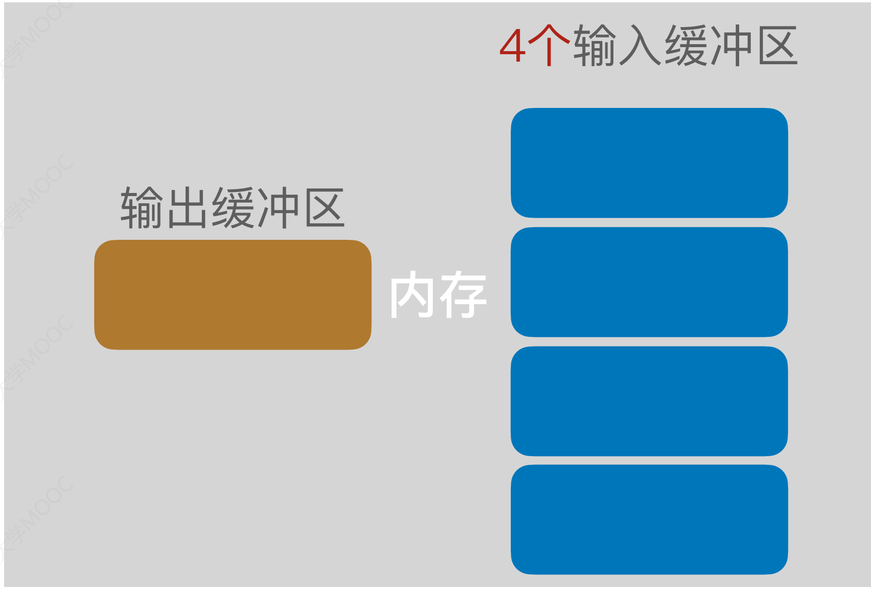
将4个有序归并段归并为一个

优化：多路归并



读磁盘

写磁盘

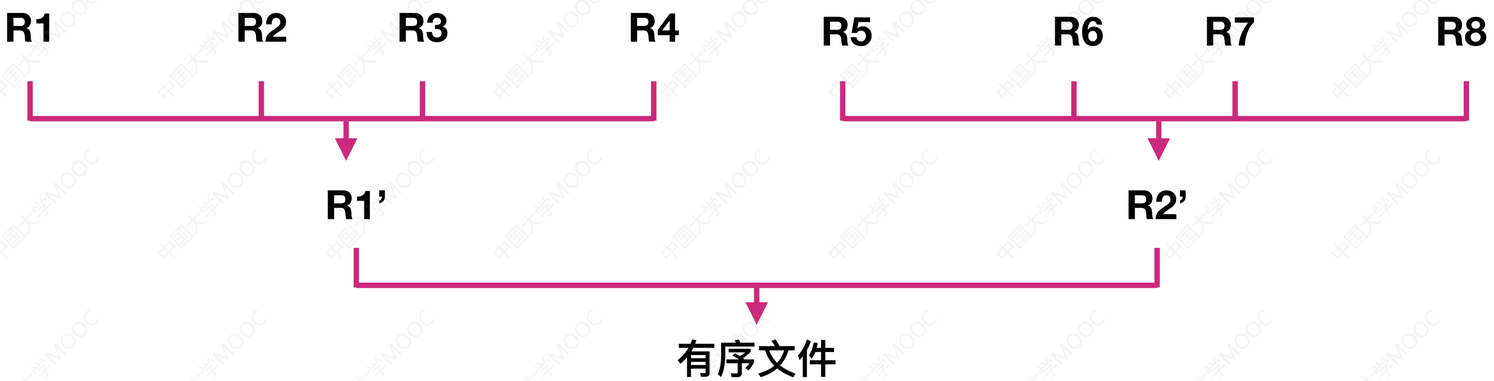


第一趟归并之后

将4个有序归并段归并为一个

优化：多路归并

外部排序时间开销=读写外存的时间+内部排序所需时间+内部归并所需时间



采用4路归并，只需进行两趟归并即可

读、写磁盘次数= 32+32*2 = 96 次

重要结论：采用多路归并可以减少归并趟数，从而减少磁盘I/O(读写)次数

对 r 个初始归并段，做k路归并，则归并树可用 k 叉树表示

若树高为h，则归并趟数 = h-1 = $\lceil log_k r \rceil$

k越大，r越小，归并趟数越少，读写磁盘次数越少

推导：k叉树第h层最多有k^{h-1} 个结点

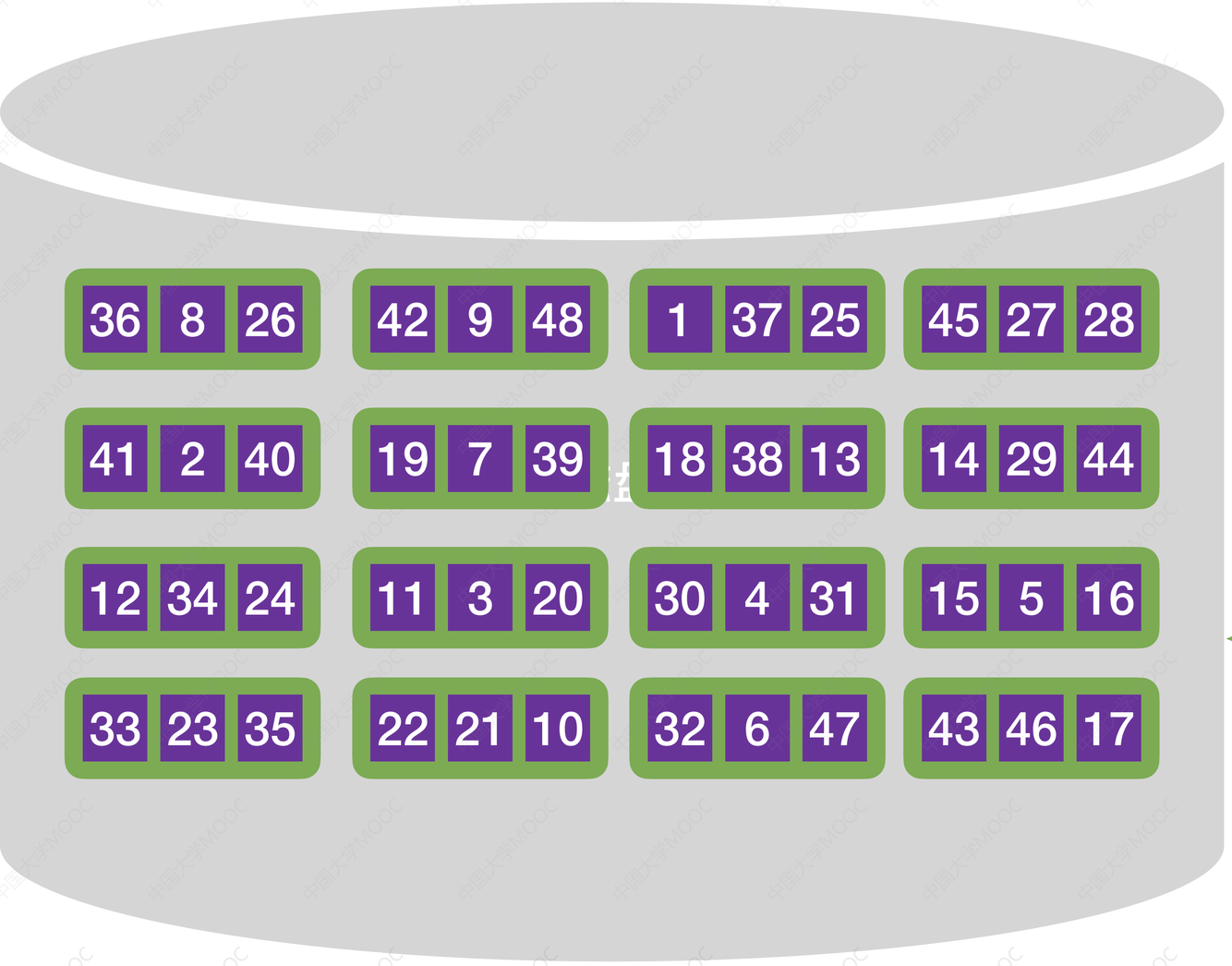
则 $r \le k^{h-1}$，$(h-1)_{最小} = \lceil log_k r \rceil$

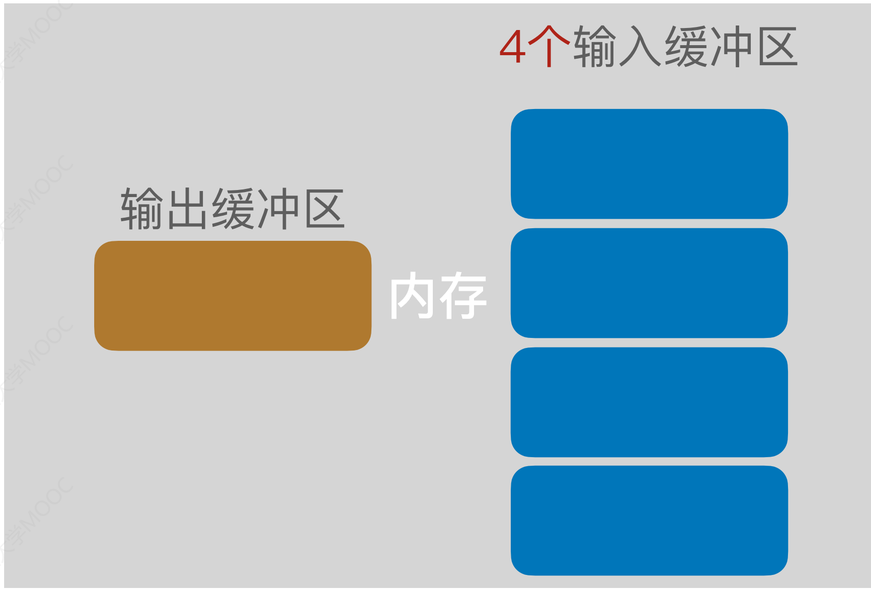
多路归并带来的负面影响：

①k路归并时，需要开辟k个输入缓冲区，内存开销增加。

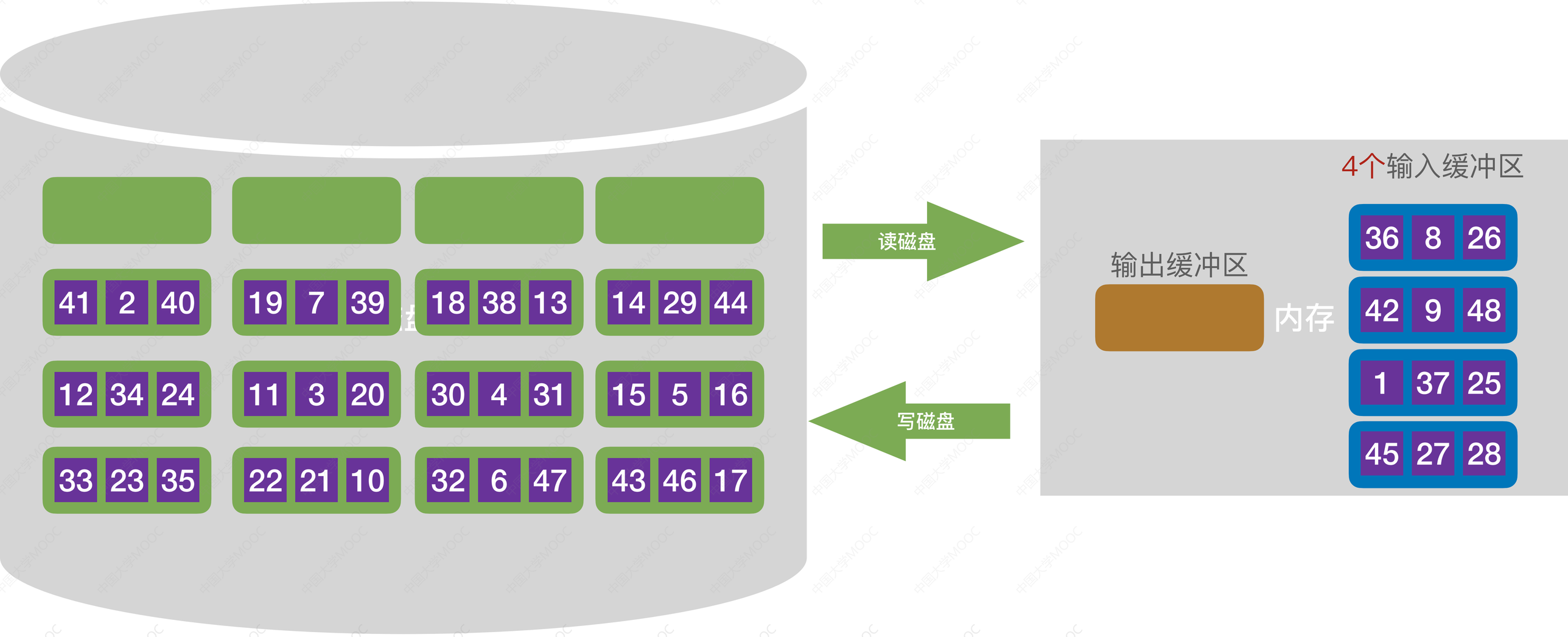
②每挑选一个关键字需要对比关键字（k-1）次，内部归并所需时间增加

优化：减少初始归并段数量

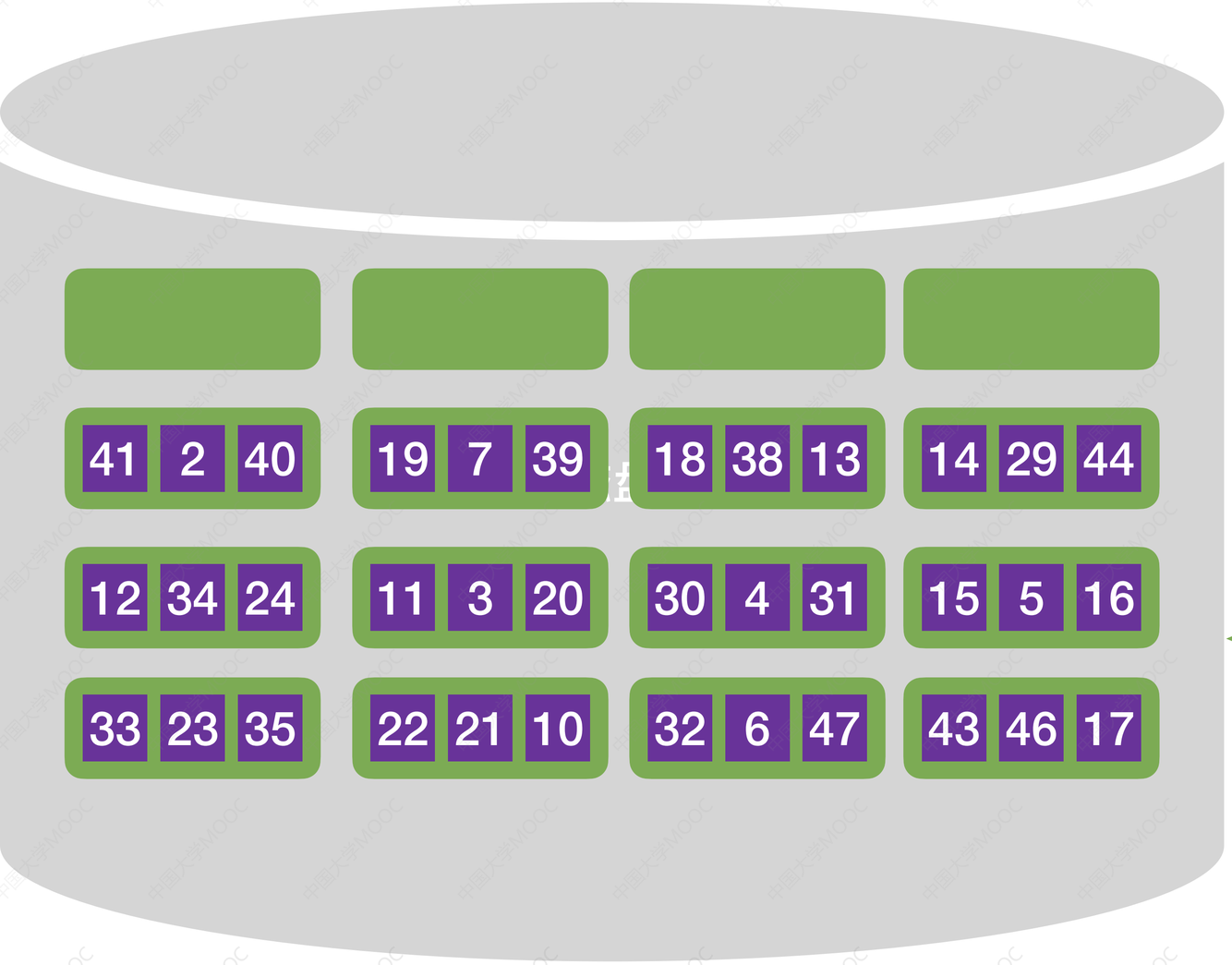




优化：减少初始归并段数量



优化：减少初始归并段数量



读磁盘

写磁盘

4个输入缓冲区

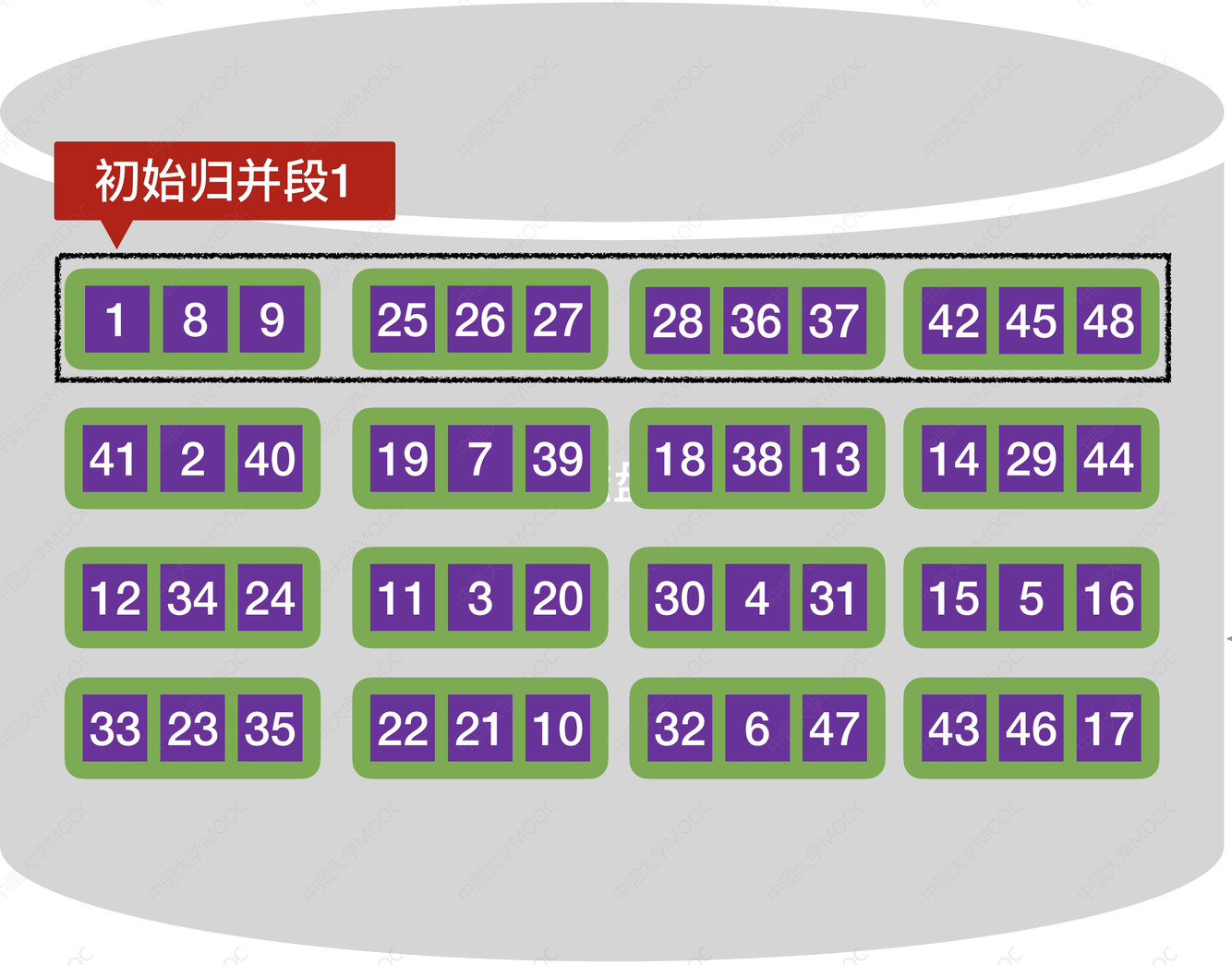
输出缓冲区

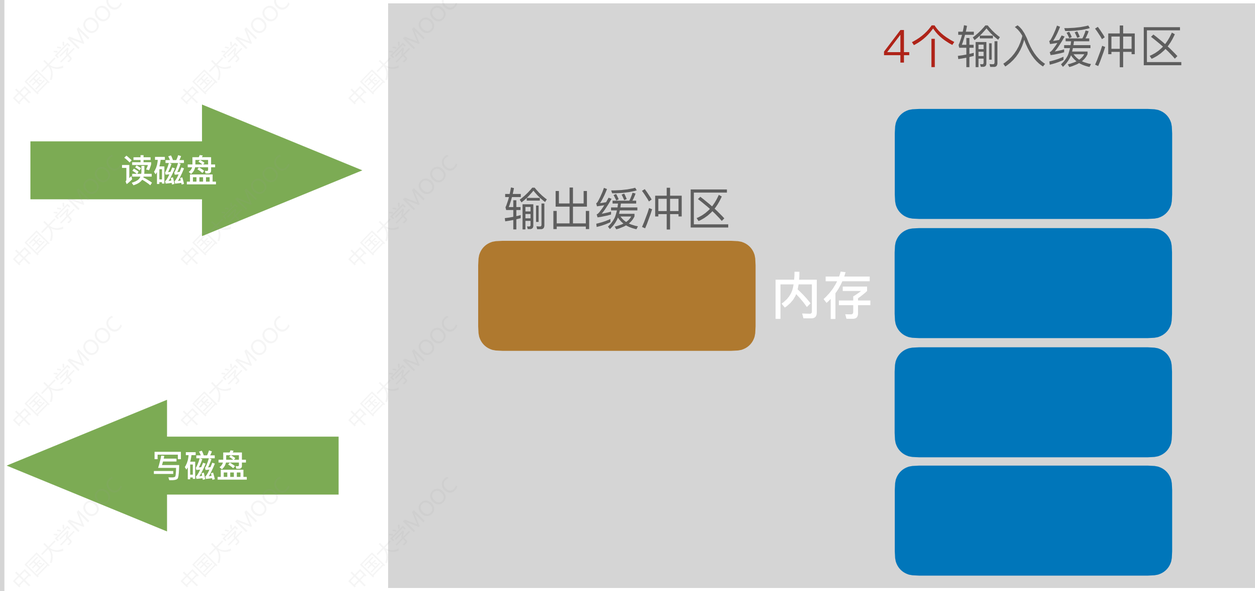
内存

1 8 9
25 26 27
28 36 37
42 45 48

生成初始归并段的“内存工作区”越大，初始归并段越长

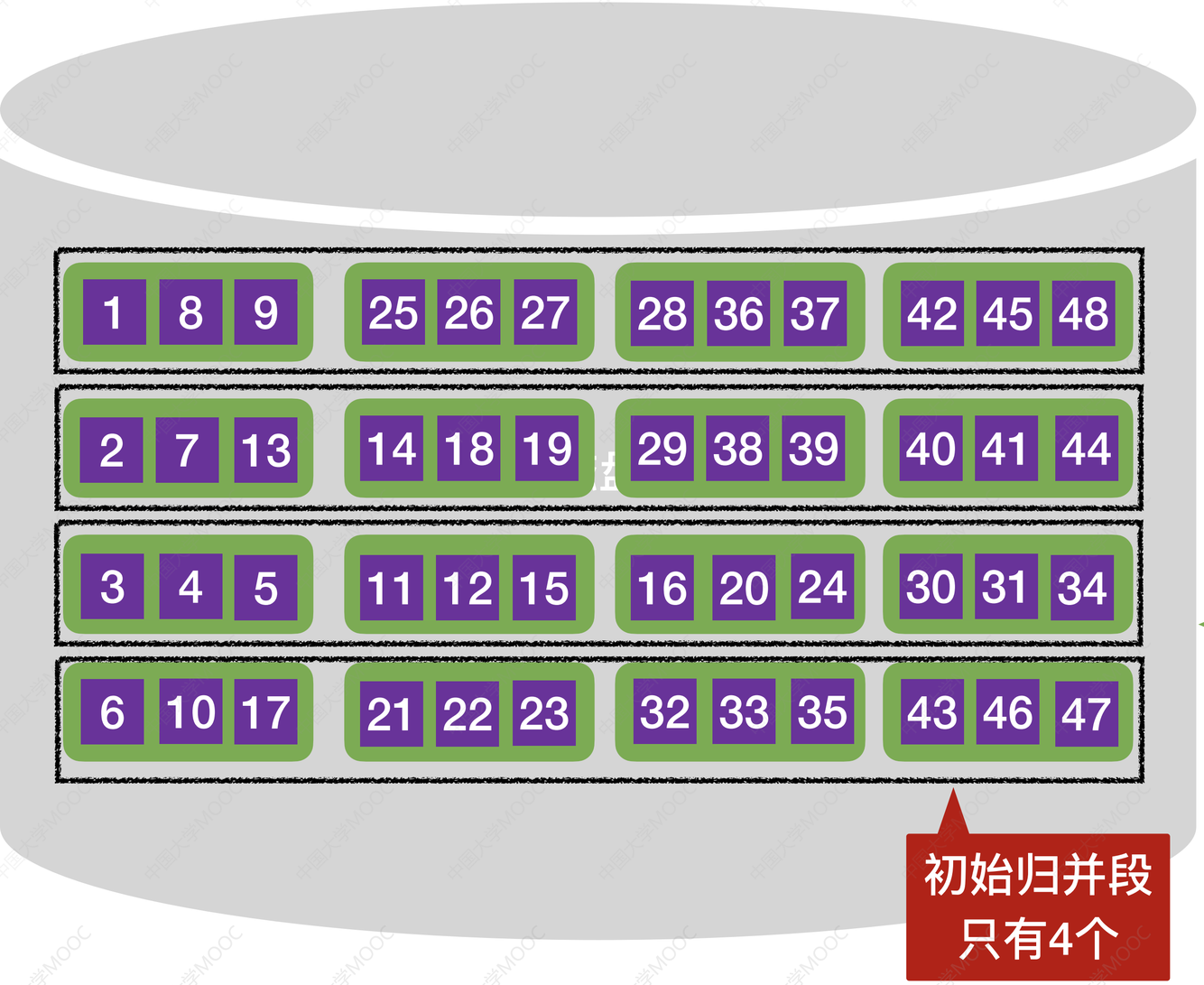
优化：减少初始归并段数量

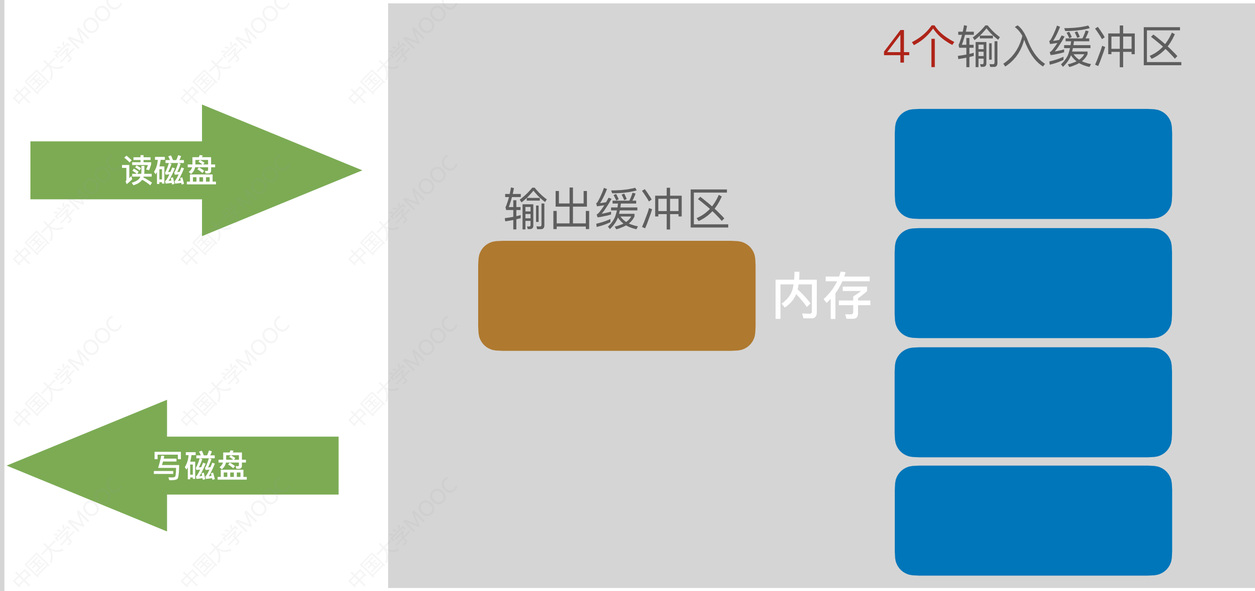




生成初始归并段的“内存工作区”越大，初始归并段越长

优化：减少初始归并段数量





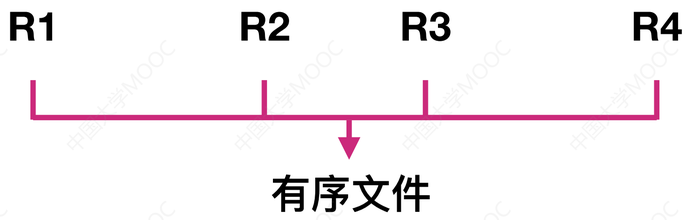
生成初始归并段的“内存工作区”越大，初始归并段越长

优化：减少初始归并段数量

对 r 个初始归并段，做k路归并，则归并树可用 k 叉树表示

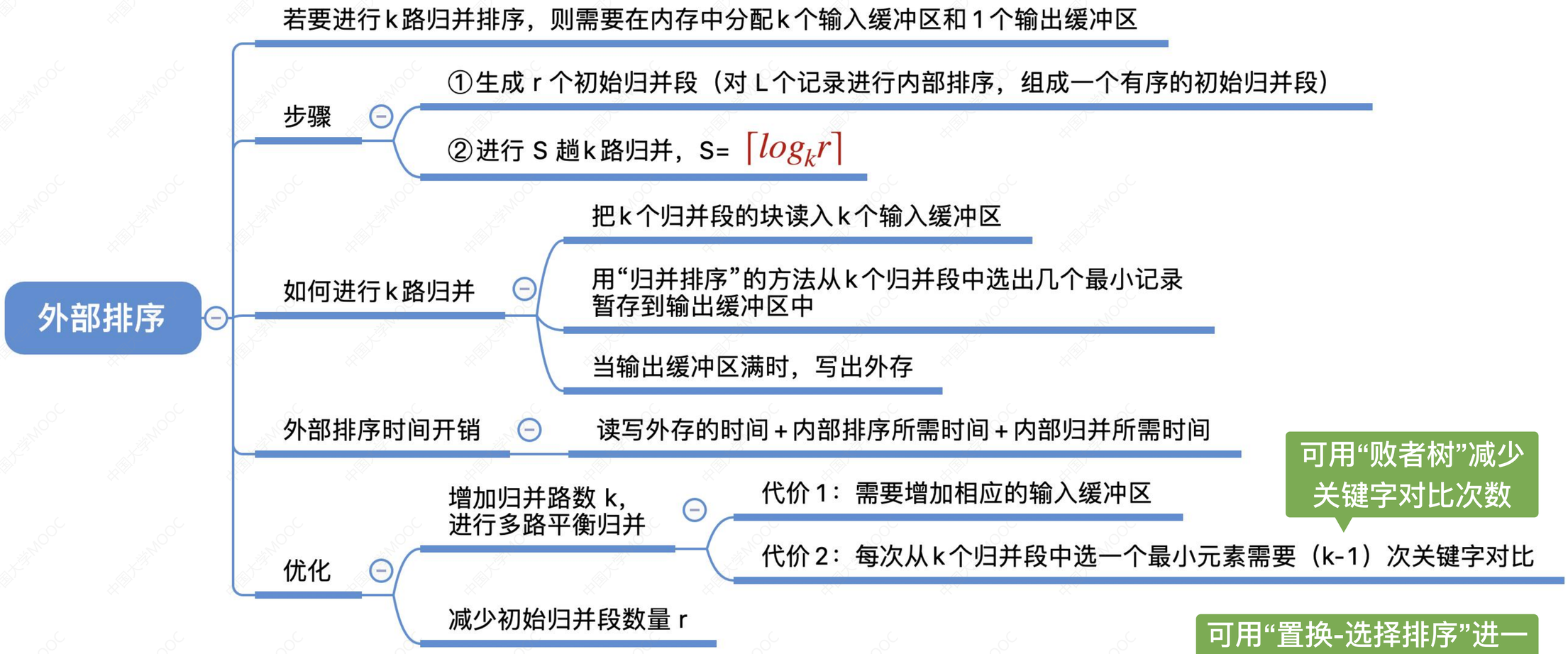
若树高为h，则归并趟数 = h-1 = $\lceil log_k r \rceil$

k越大，r越小，归并趟数越少，读写磁盘次数越少



结论：若能增加初始归并段的长度，则可减少初始归并段数量 r

知识回顾与重要考点



注：按照本节介绍的方法生成的初始归并段，若共 N 个记录，内存工作区可以容纳 L 个记录，则初始归并段数量 r = $\lceil N/L \rceil$

纠正：什么是多路平衡归并?

误：对r个初始归并段，做k路平衡归并，归并树可用严格k叉树（即只有度为k与度为0的结点的k叉树）来表示。

知道错哪了不?

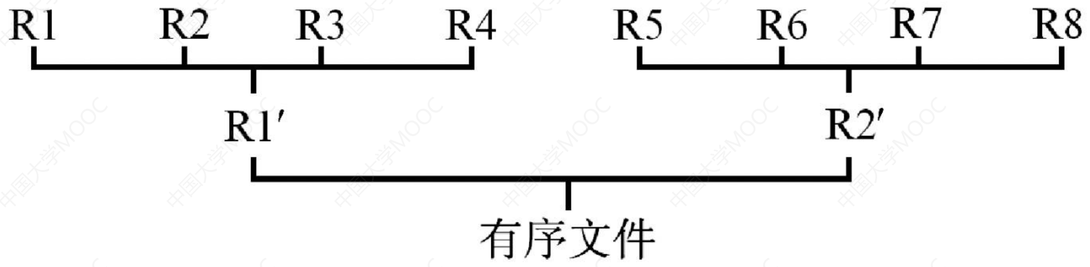


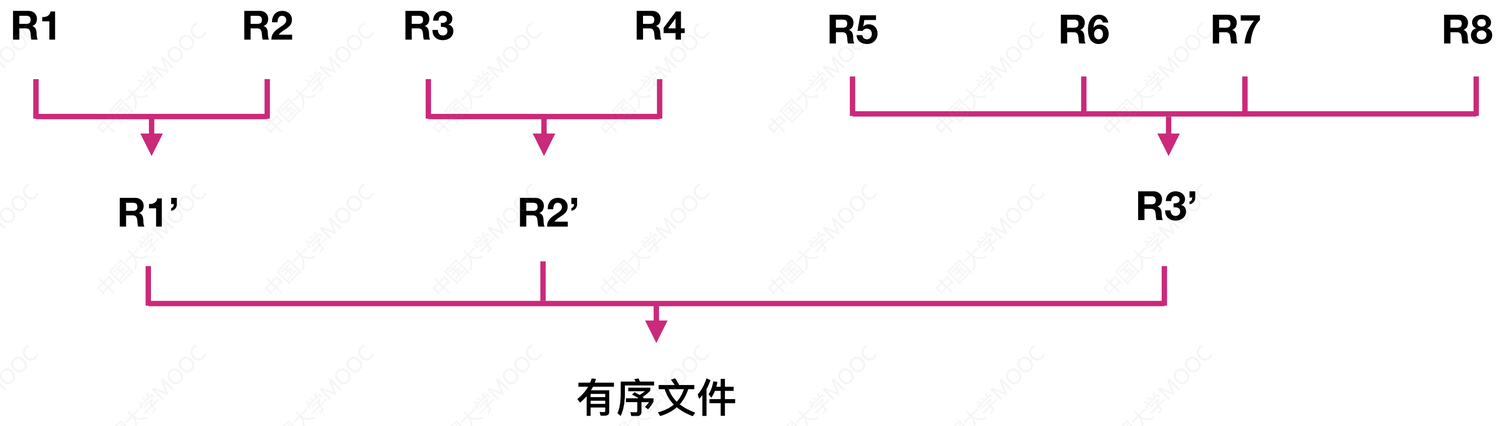
图 7.8　4 路平衡归并的排序过程

k路平衡归并：①最多只能有k个段归并为一个；

②每一趟归并中，若有 m 个归并段参与归并，则经过这一趟处理得到$\lceil m/k \rceil$个新的归并段

纠正：什么是多路平衡归并?

8个归并段经过一趟处理后得到3个新的归并段



这个例子是不是4路归并排序？—— 是！

这个例子是不是4路平衡归并排序？—— 不是！

$\lceil 8/4 \rceil = 2 \neq 3$

k路平衡归并：①最多只能有k个段归并为一个；

②每一趟归并中，若有 m 个归并段参与归并，则经过这一趟处理得到$\lceil m/k \rceil$个新的归并段

欢迎大家对本节视频进行评价~

学员评分：8.7.1+8.7.2...

扫一扫二维码打开或分享给好友

– 腾讯文档 –

可多人实时在线编辑，权限安全可控

公众号：王道在线

b站：王道计算机教育

抖音：王道计算机考研