

## 11.2.3 最佳归并树

$k$ 路平衡归并适合初始归并段中的记录个数相同的情况，当初始归并段中的记录个数不同时，怎么办？

当初始归并段和 $k$ 已确定的情况时



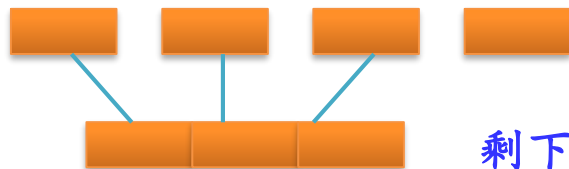
归并方案转化为

哪些初始归并段先归并，哪些后归并的问题。

显然采用 $k$ 叉哈夫曼树的归并方案。

- $WPL$ 最小
- 在内存中归并时，可以利用败者树减少关键字比较次数

⇒ 存在的问题（假设 $k=3$ ）：



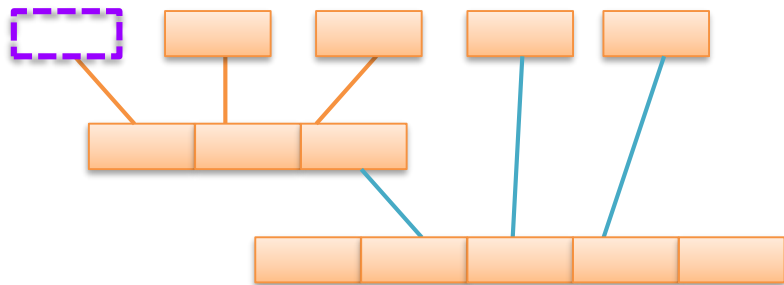
剩下只有2个归并段了，怎么办？

解决的方法是加虚段（长度为0的归并段）

加多少个虚段呢？

应加  $(k-1)-(m-1) \bmod (k-1)$  个长度的虚段

前面问题的解决方法：加上1个虚段：



最佳归并树 ( $m$ 个初始归并段) 是带权路径长度最短的 $k$ 叉 ( $k$ 阶) 哈夫曼树, 构造步骤如下:

① 若  $x = (m-1) \bmod (k-1) \neq 0$ , 则需附加  $(k-1) - x$  个虚段, 以使每次归并都可以对应  $k$  个段。

② 按照哈夫曼树的构造原则 (权值越小的节点离根节点越远) 构造最佳归并树。



$k=2$ 时,  $x = (m-1) \bmod 1 = 0$ , 所以二路归并 (哈夫曼树构造中) 不需要增加虚段

**【例11-3】** 设文件经预处理后，得到长度为

{47, 9, 39, 18, 4, 12, 23, 7, 21, 16, 26}

的11个初始归并段，试为4路归并设计一个读写文件次数最少的归并方案（假如每个记录占用一个物理块）。

各个初始归并段中的记录个数，而非关键字序列

**解：**初始归并段个数 $m=11$ ， $k=4$ 。 $x=(m-1) \bmod (k-1)=1 \neq 0$ ，因此需附加：

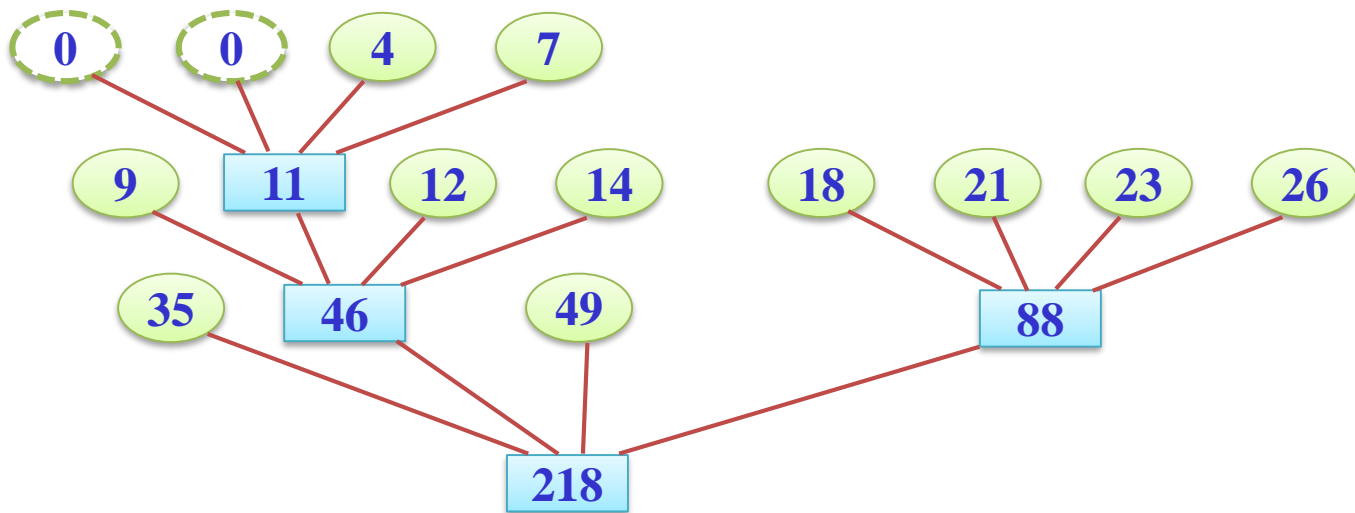
$$(k-1)-x = 2$$

个长度为0的虚段。根据集合：

$\{49, 9, 35, 18, 4, 12, 23, 7, 21, 14, 26, 0, 0\}$

构造4阶哈夫曼树。

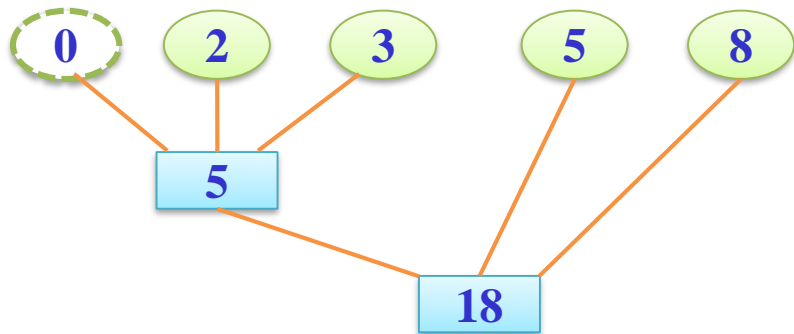
按记录个数递增排序：{0, 0, 4, 7, 9, 12, 14, 18, 21, 23, 26, 35, 49}  
4路最佳归并树的构造过程：



- $WPL = (4+7) \times 3 + (9+12+14+18+21+23+26) \times 2 + (35+49) \times 1 = 363$
- 最少的读写次数 =  $2 \times WPL = 726$ 次。

**【例11-4】** 有4个初始归并段，记录个数分别为2、3、5、8，采用3路归并，最少的读写次数是多少（假设每个记录读写一次）？

**解：** 3路最佳归并树如下：

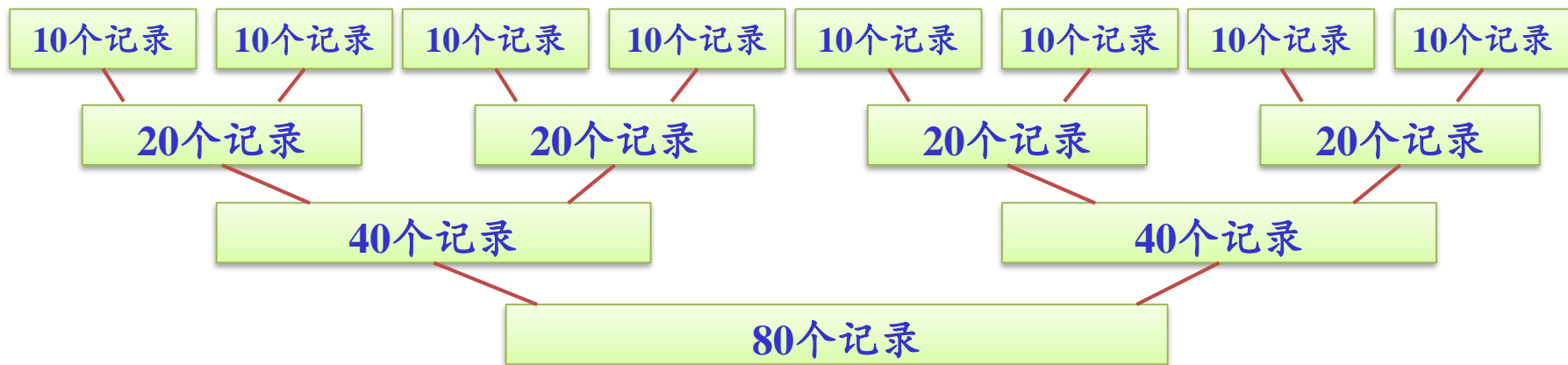


- $WPL = (2+3) \times 2 + (5+8) \times 1 = 23$
- 最少的读写次数 =  $2 \times$  读次数 = 46



满足 $k$ 路平衡归并的前提    **平衡归并树**  $\equiv$  **最佳归并树**

例如：初始归并段个数 $m=8$ ，每个段的记录数=10， $k=2$ ，对应的平衡归并树：



与2路最佳归并树相同。

$$WPL=8 \times 10 \times 3=240$$

归并方案设计：



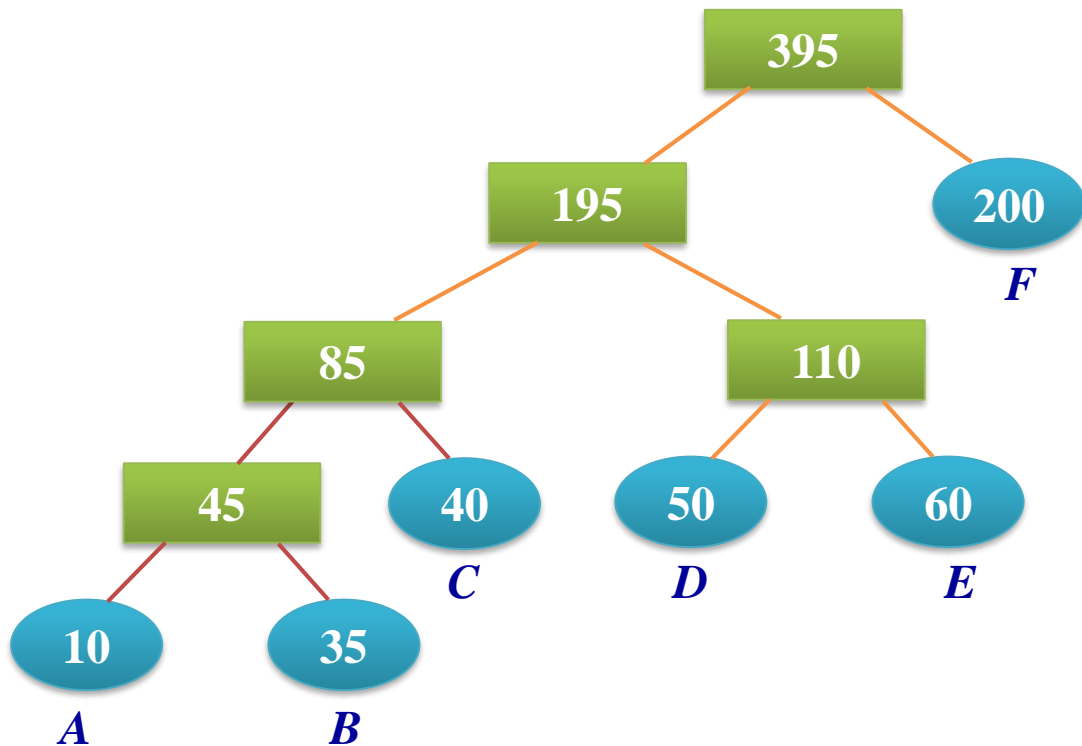
- 满足 $k$ 路平衡归并的前提  $\Rightarrow$  采用平衡归并树；
- 否则  $\Rightarrow$  采用最佳归并树

**【例11-5】** 设有6个有序文件A、B、C、D、E、F，分别含有10、35、40、50、60和200个数据元素，各文件中元素按升序排序。

要求通过5次两两合并，将6个文件最终合并成一个升序文件。  
给出文件读写次数最少的合并过程（假设每个记录读写一次）。

两两合并	→	二路归并
最少的合并过程	→	2路最佳归并树

**解：**构造2路最佳归并树，归并过程如下：



$$WPL = (10+35) \times 4 + (40+50+60) \times 3 + 200 \times 1 = 830。$$

$$\text{最少读写次数} = 2 \times WPL = 1660$$

本课程结束！

谢谢你的坚持  
和努力！