

11. 串

(d1) BM_BC算法：以终为始

邓俊辉

Begin with the end in mind.

deng@tsinghua.edu.cn

经验 + 教训

❖ 串匹配 = x 次失败的对齐 + $0/1$ 次成功的对齐

❖ 与其说要加速匹配，不如说是加速失败 —— 尽快排除失败的对齐

❖ 就单个对齐位置的排除而言

平均仅需常数次比对（只要 $|\Sigma|$ 不致太小，单次比对成功概率足够低）

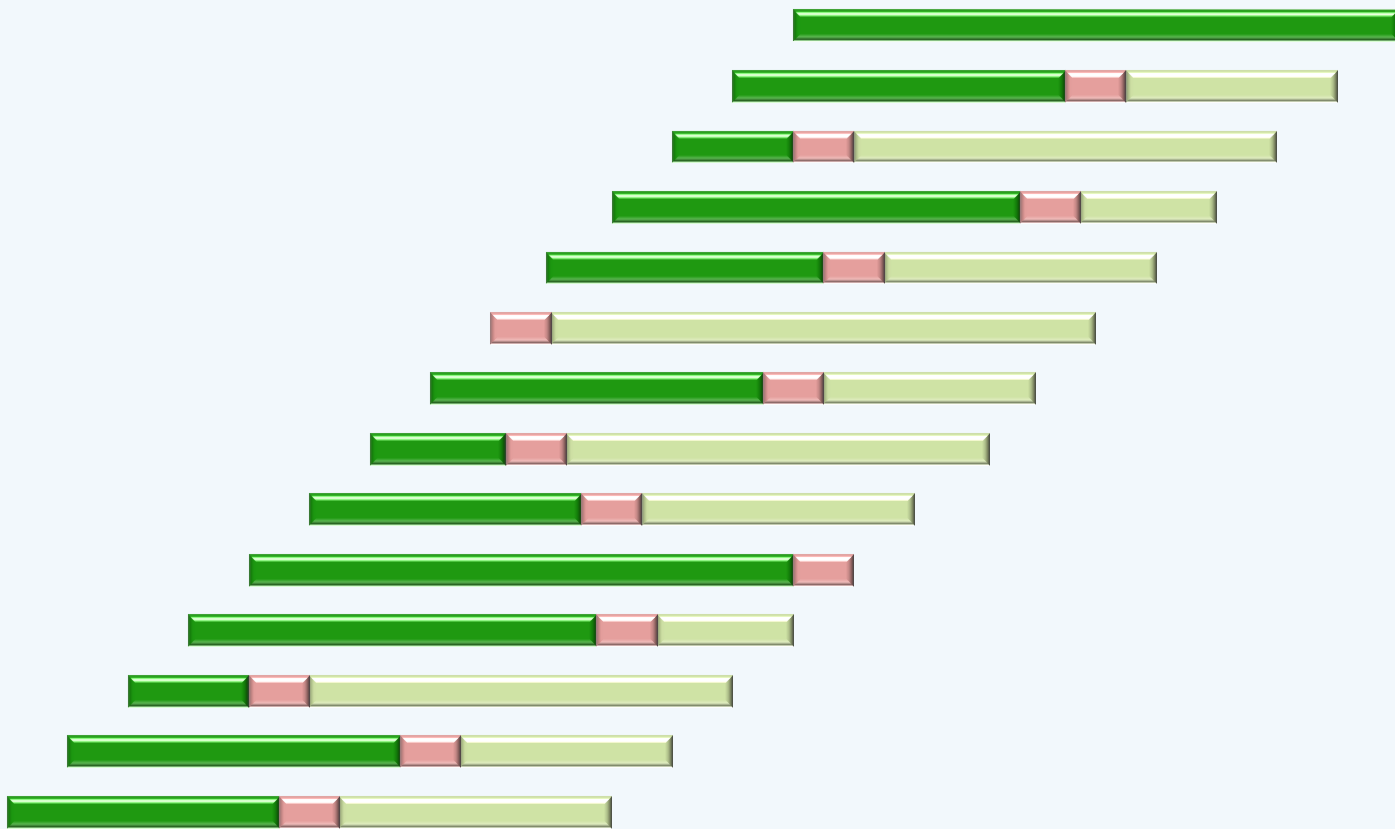
且具体的比对位置及次序无所谓

❖ 然而就排除更多后续对齐位置而言

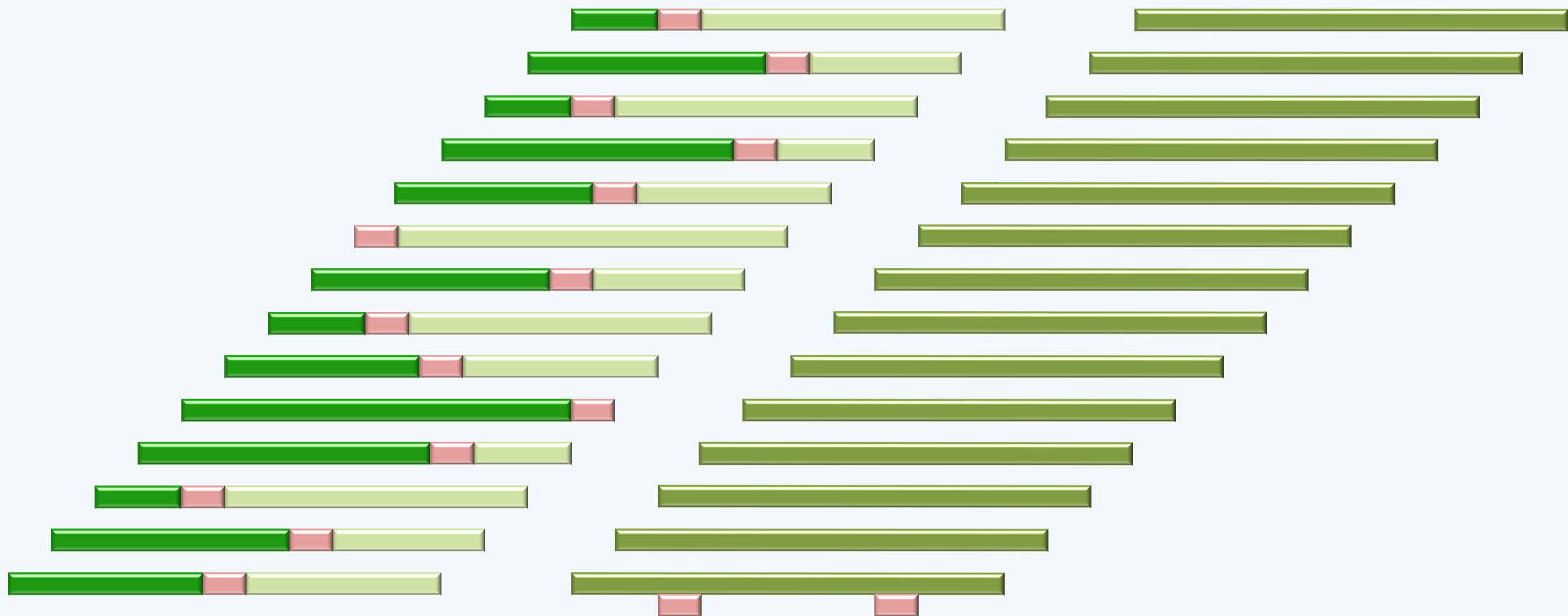
不同的对比位置及次序，作用差异极大

❖ 通常，越是靠前/后的位置，作用越小/大

善待教训



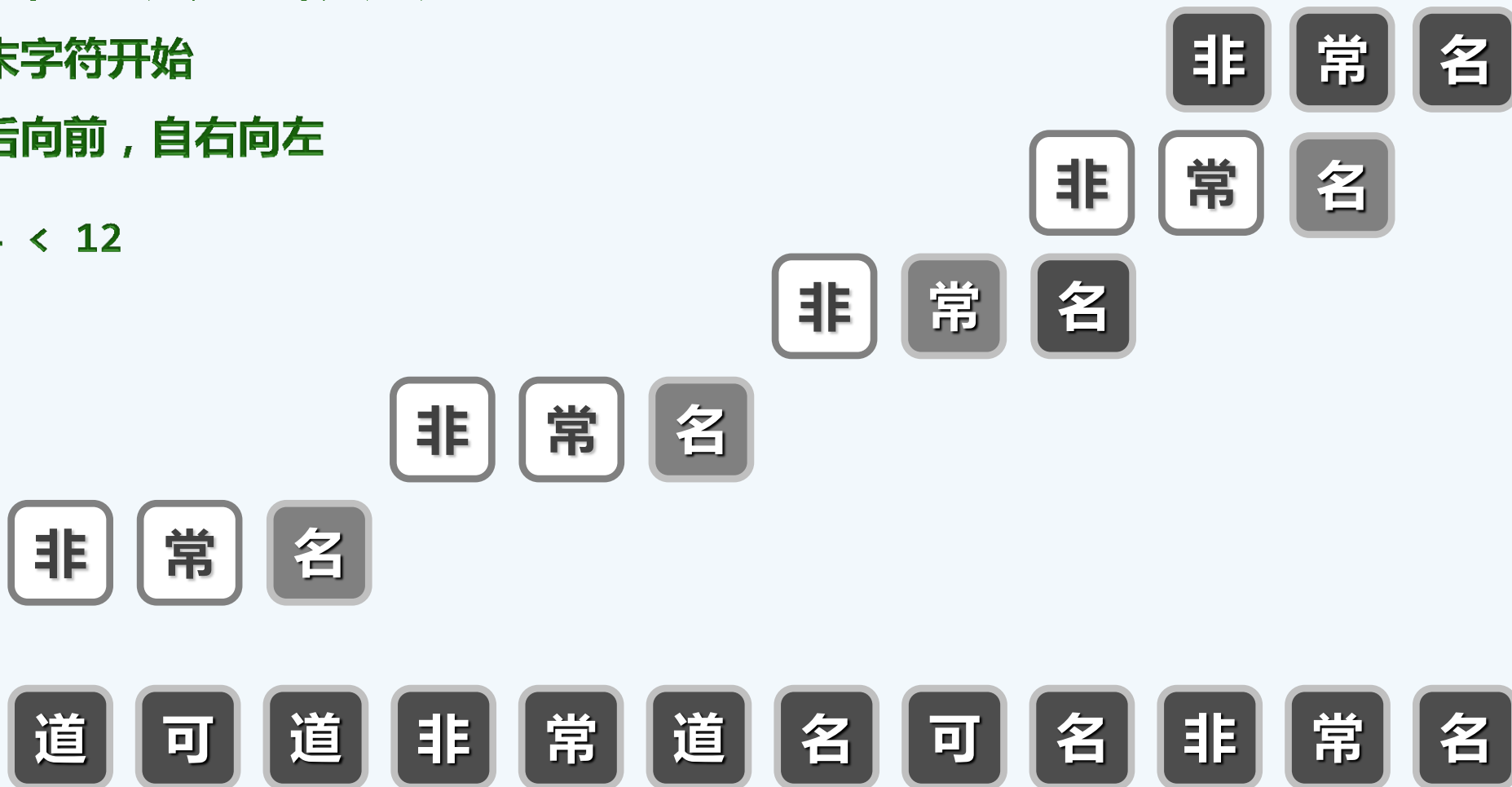
前轻后重



以终为始

- ❖ 既如此，每一趟比对都应该
从末字符开始
自后向前，自右向左

$$❖ 4 + 4 < 12$$



以终为始

❖ [Boyer + Moore, 1977] A fast string searching algorithm

❖ 预处理：根据模式串P，预先构造gs[]表和bc[]表

迭代： 自右向左依次比对字符，找到极大的匹配后缀

若完全匹配，则返回位置

否则，根据gs[]和bc[]，P适当右移，并重新自右向左比对

