

Θg-F

词典

计数排序

邓俊辉

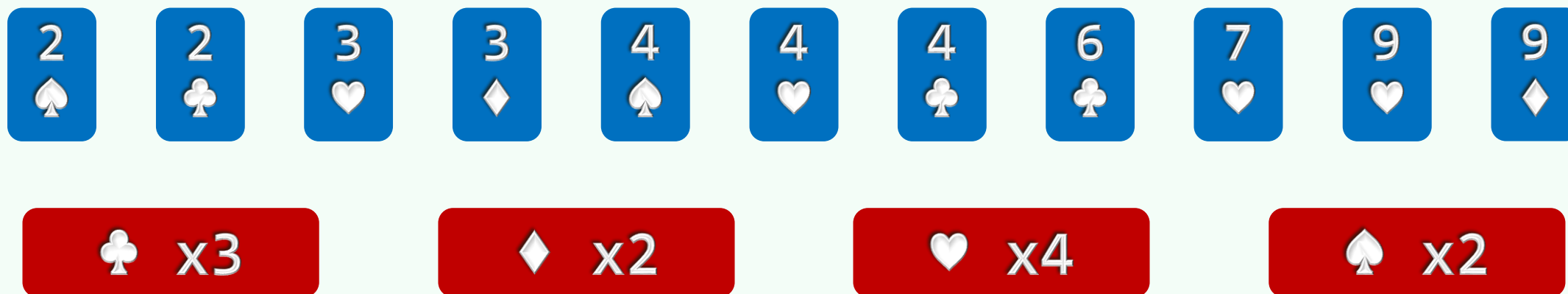
The purpose of computing is insight, not numbers.

deng@tsinghua.edu.cn

算法 (1)

❖ 回忆：基数排序中反复做的桶排序...

❖ 亦属 “小集合 + 大数据” 类型，是否可以更快？



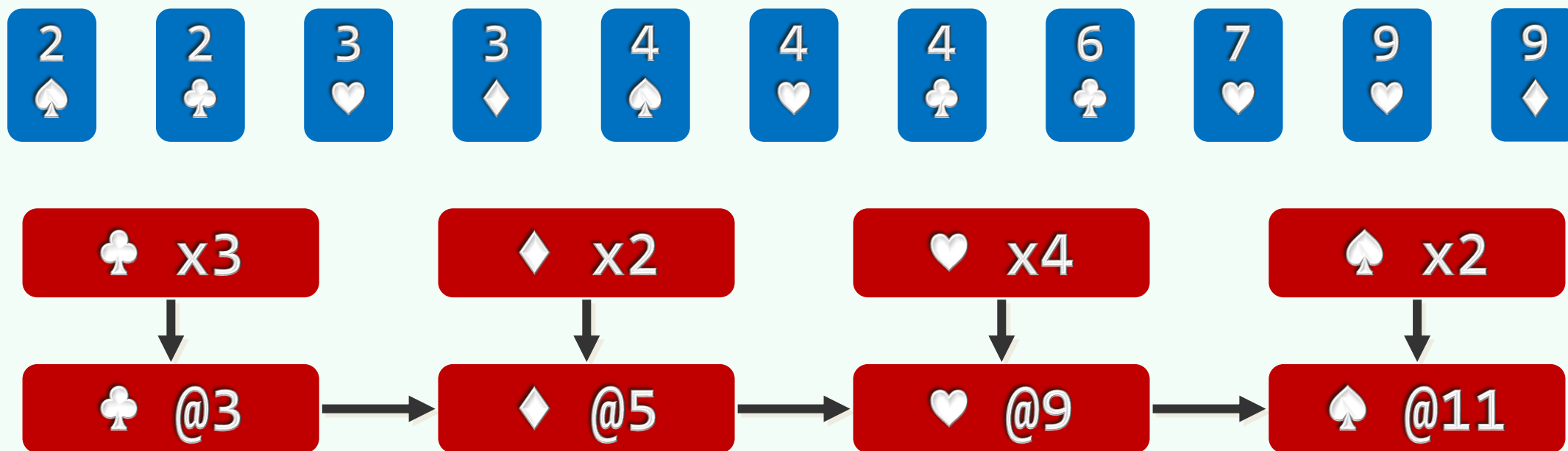
❖ 仍以纸牌排序为例（ $n \gg m = 4$ ），假设已按点数排序，以下对花色排序

1) 经过分桶，统计出各种花色的数量 $O(n)$

算法 (2+3)

2) **自前向后**扫描各桶，依次累加 //cumulative sum, $O(m)$

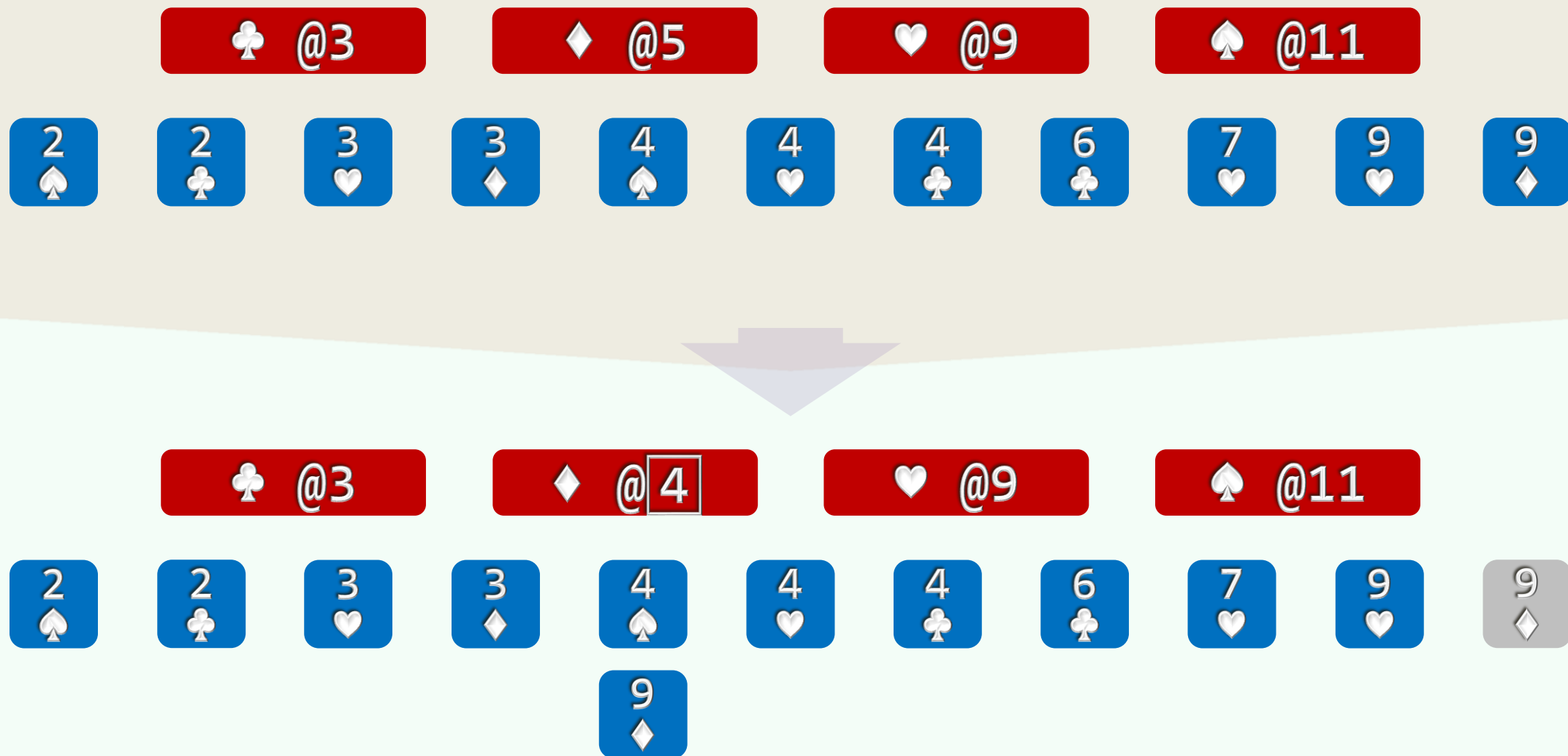
即可确定各套花色所处的**秩区间**: $[0, 3) + [3, 5) + [5, 9) + [9, 11)$



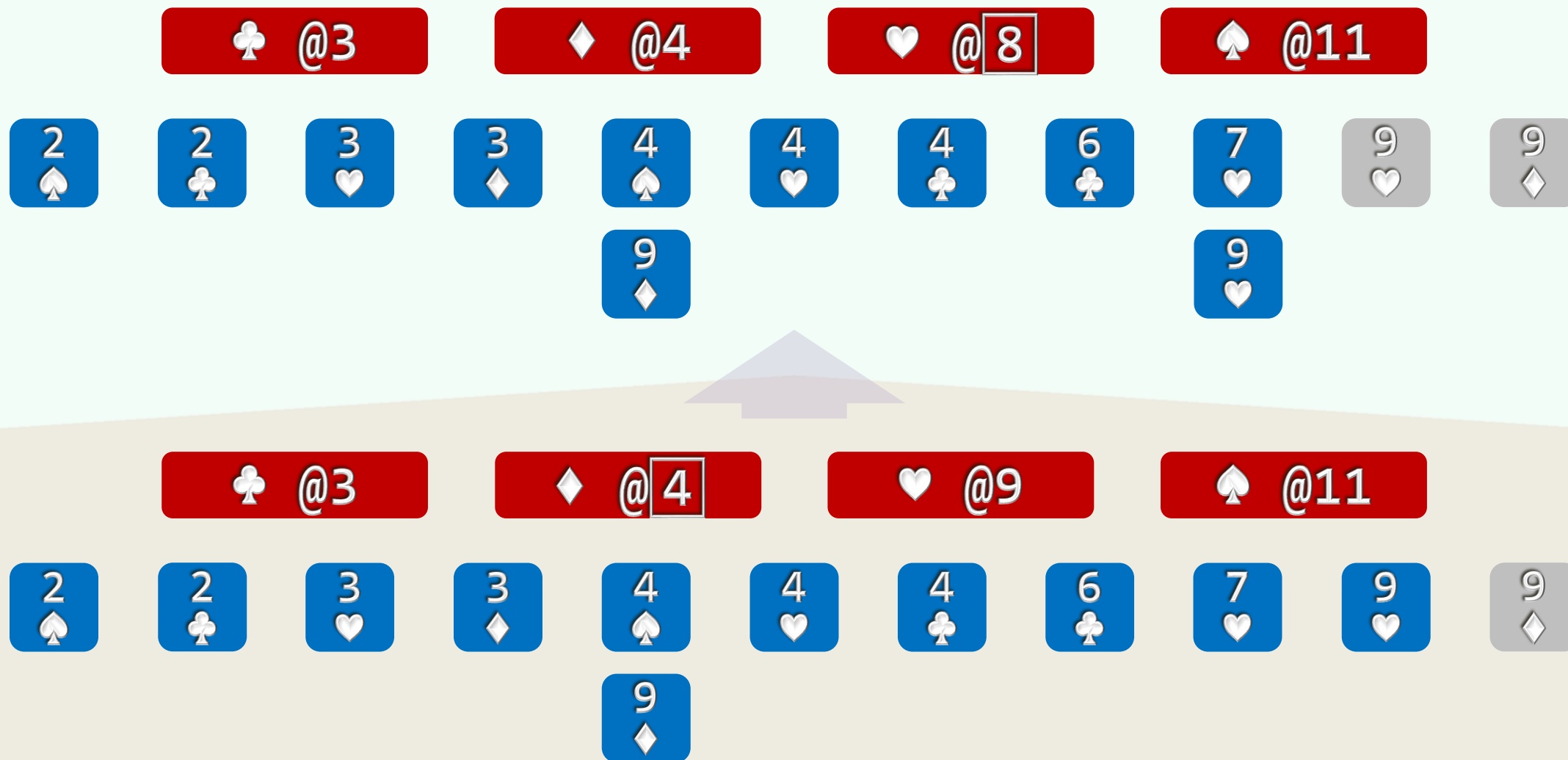
3) **自后向前**扫描每一张牌 // $O(n)$

对应桶的计数**减一**，即是其在最终有序序列中对应的**秩**

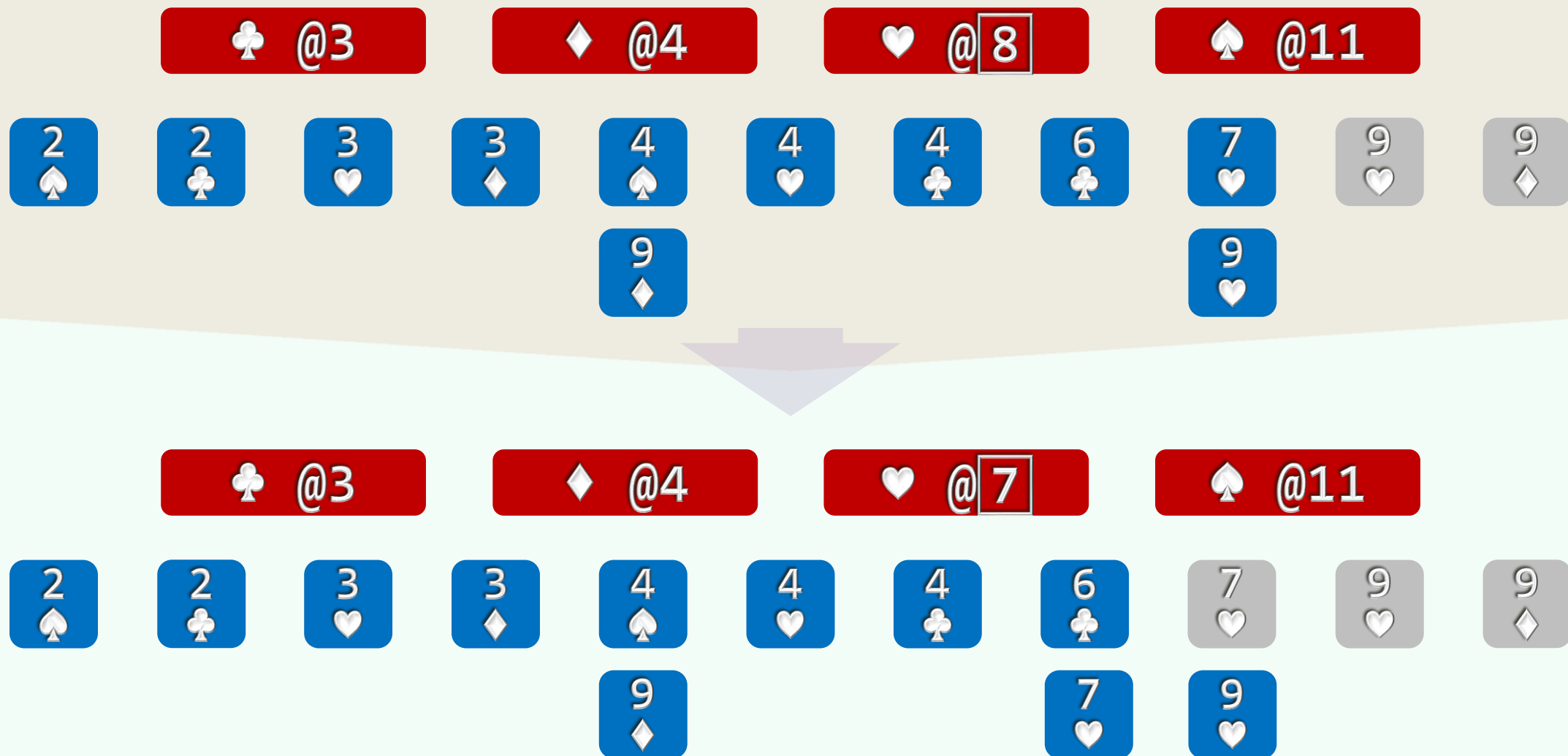
实例 (1/11)



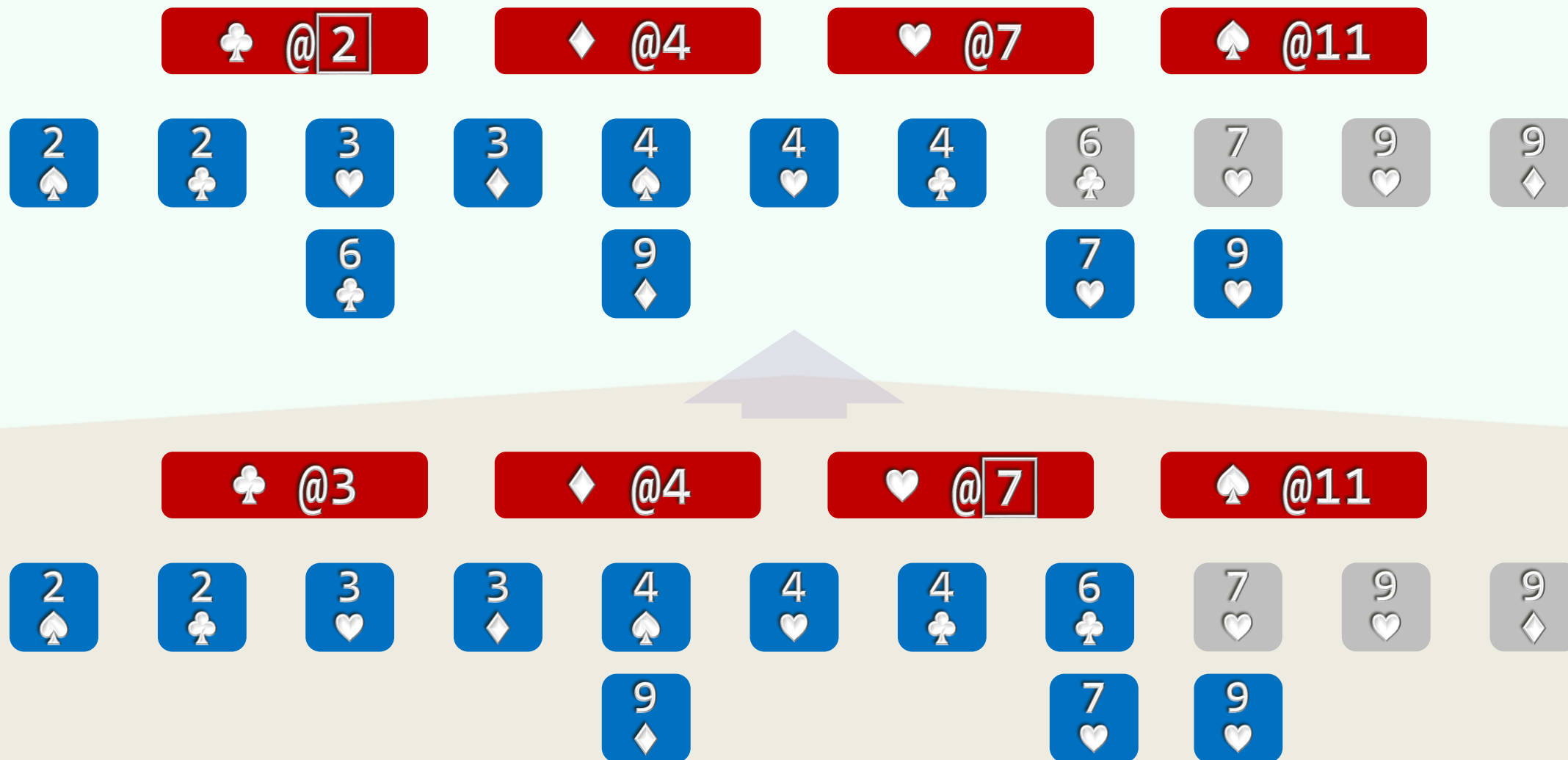
实例 (2/11)



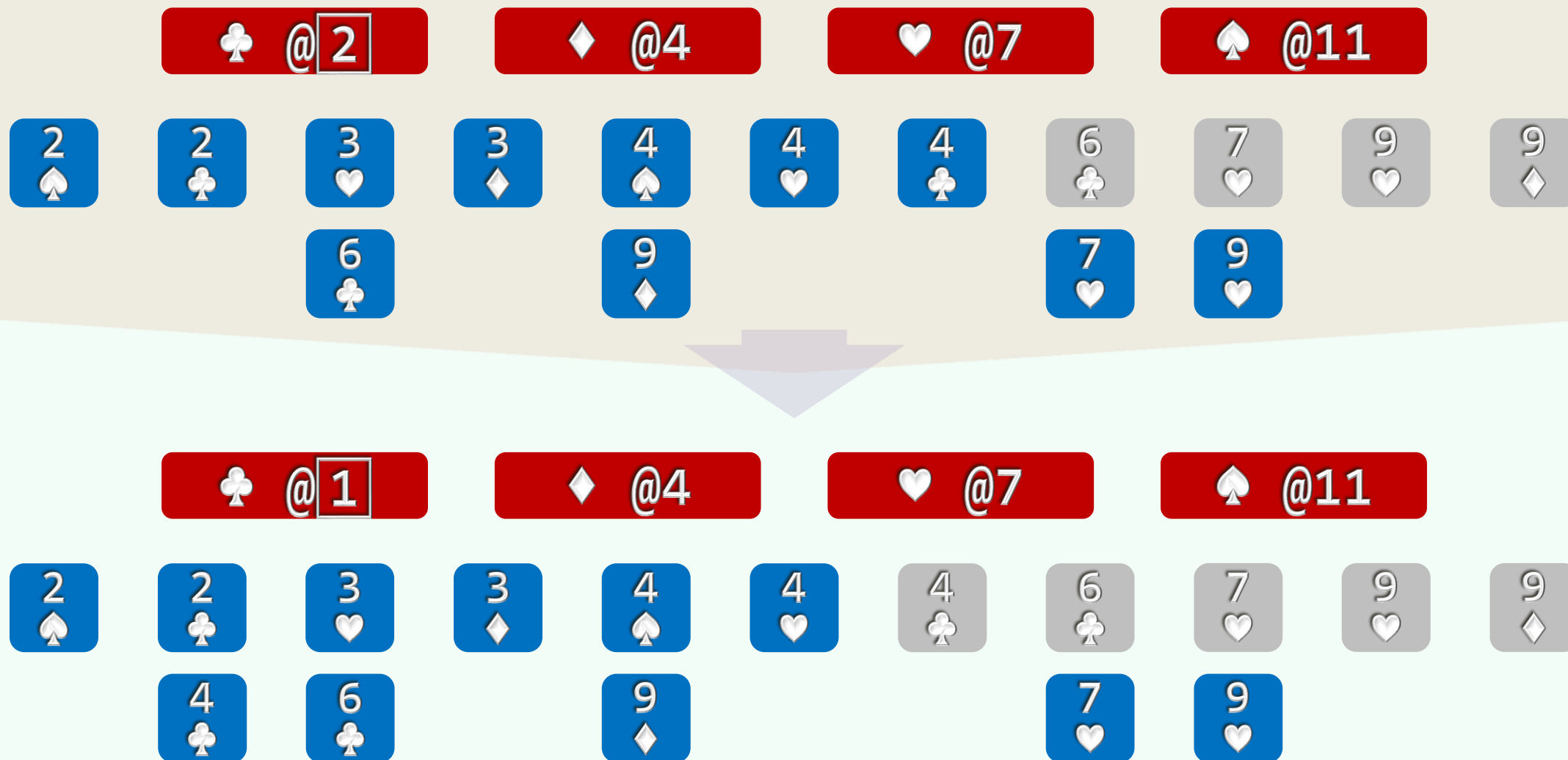
实例 (3/11)



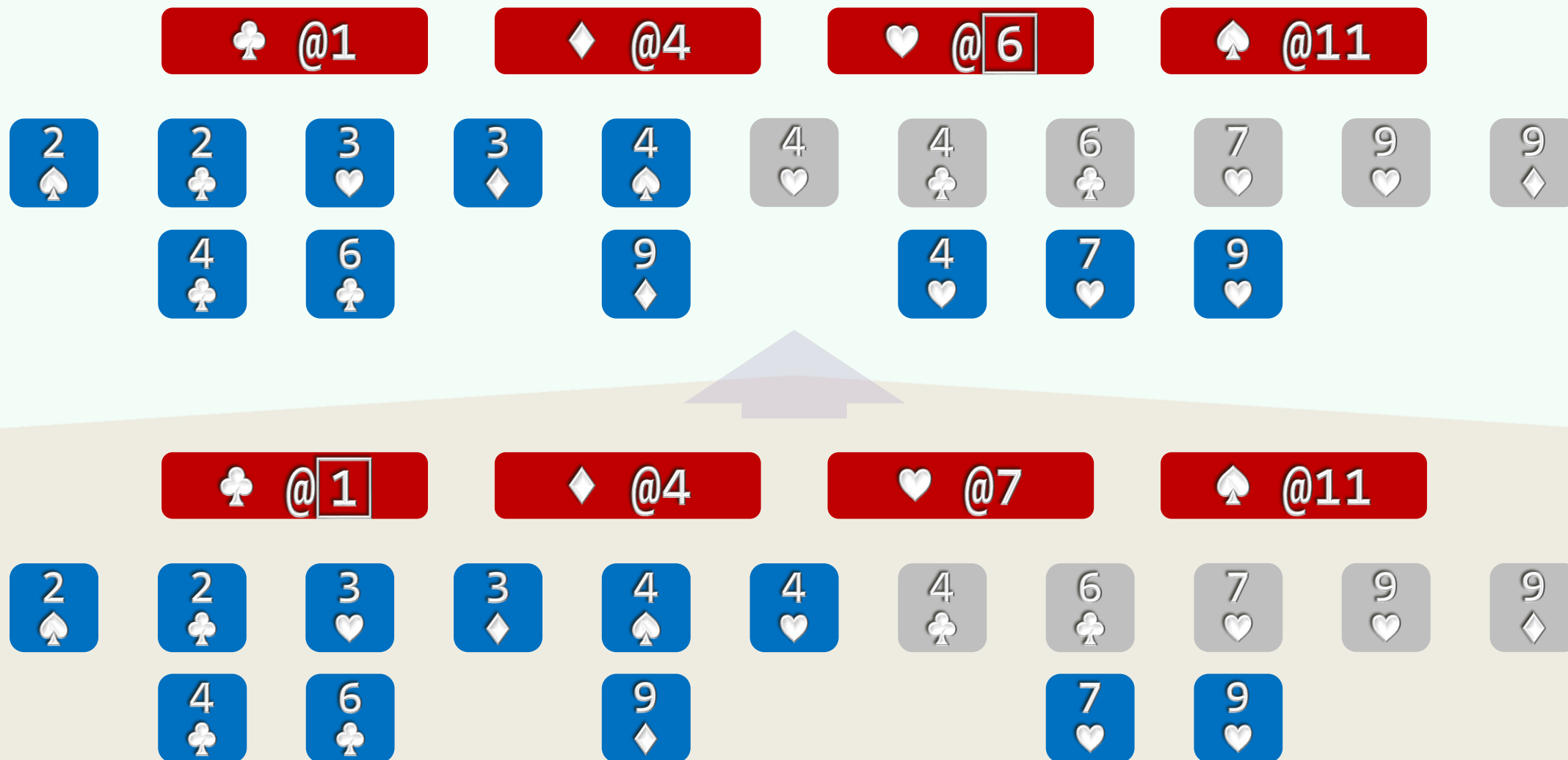
实例 (4/11)



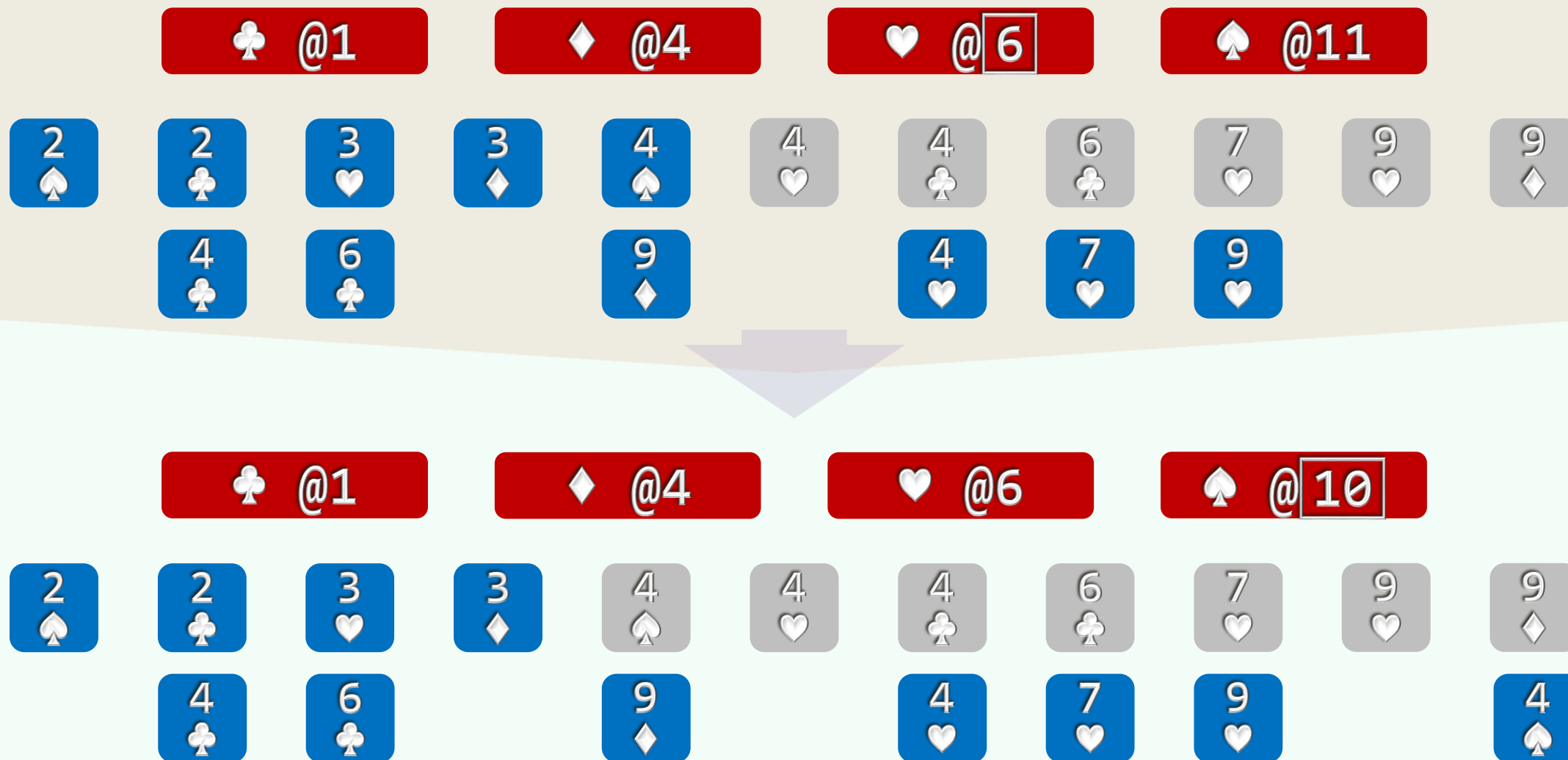
实例 (5/11)



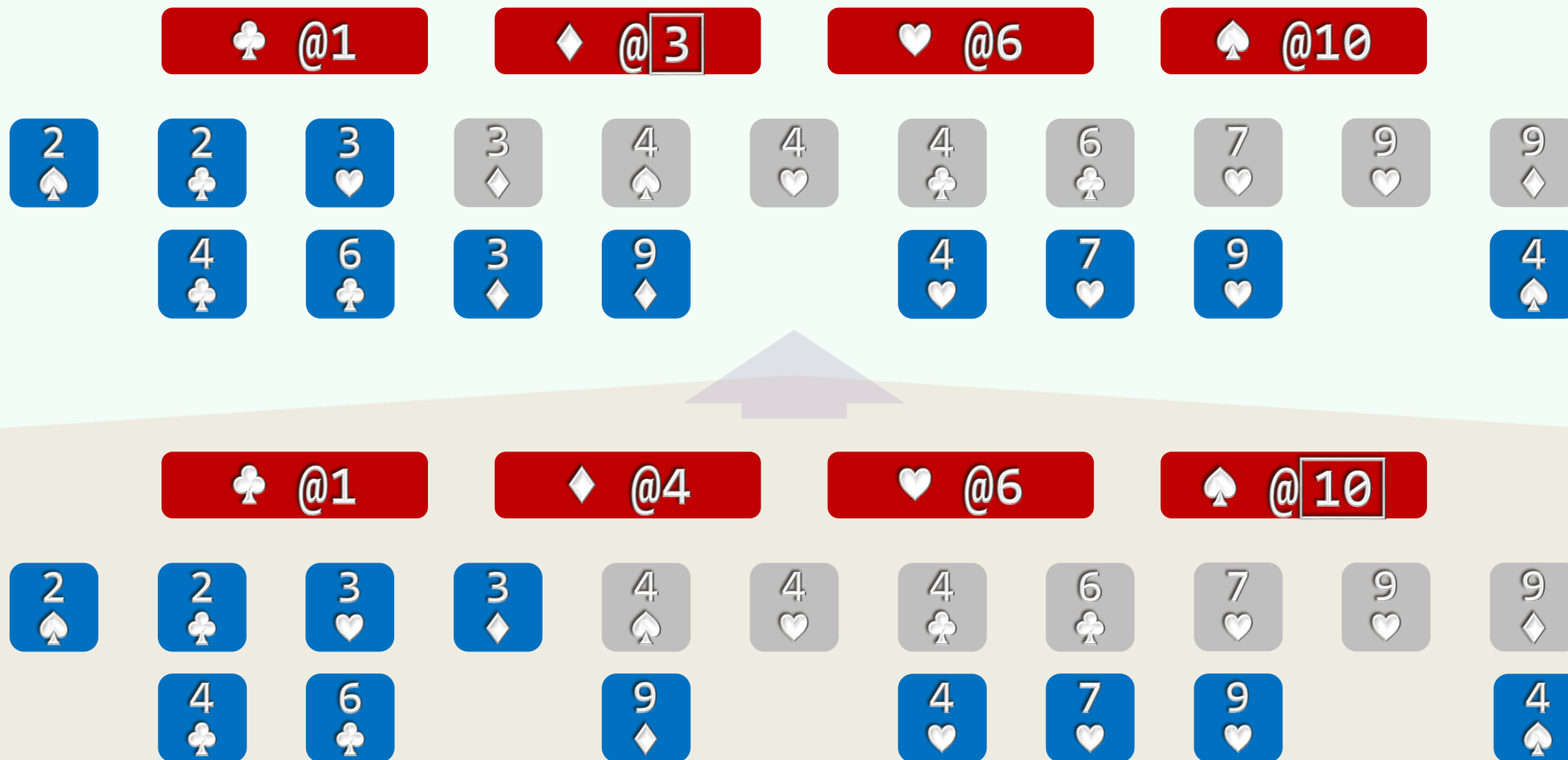
实例 (6/11)



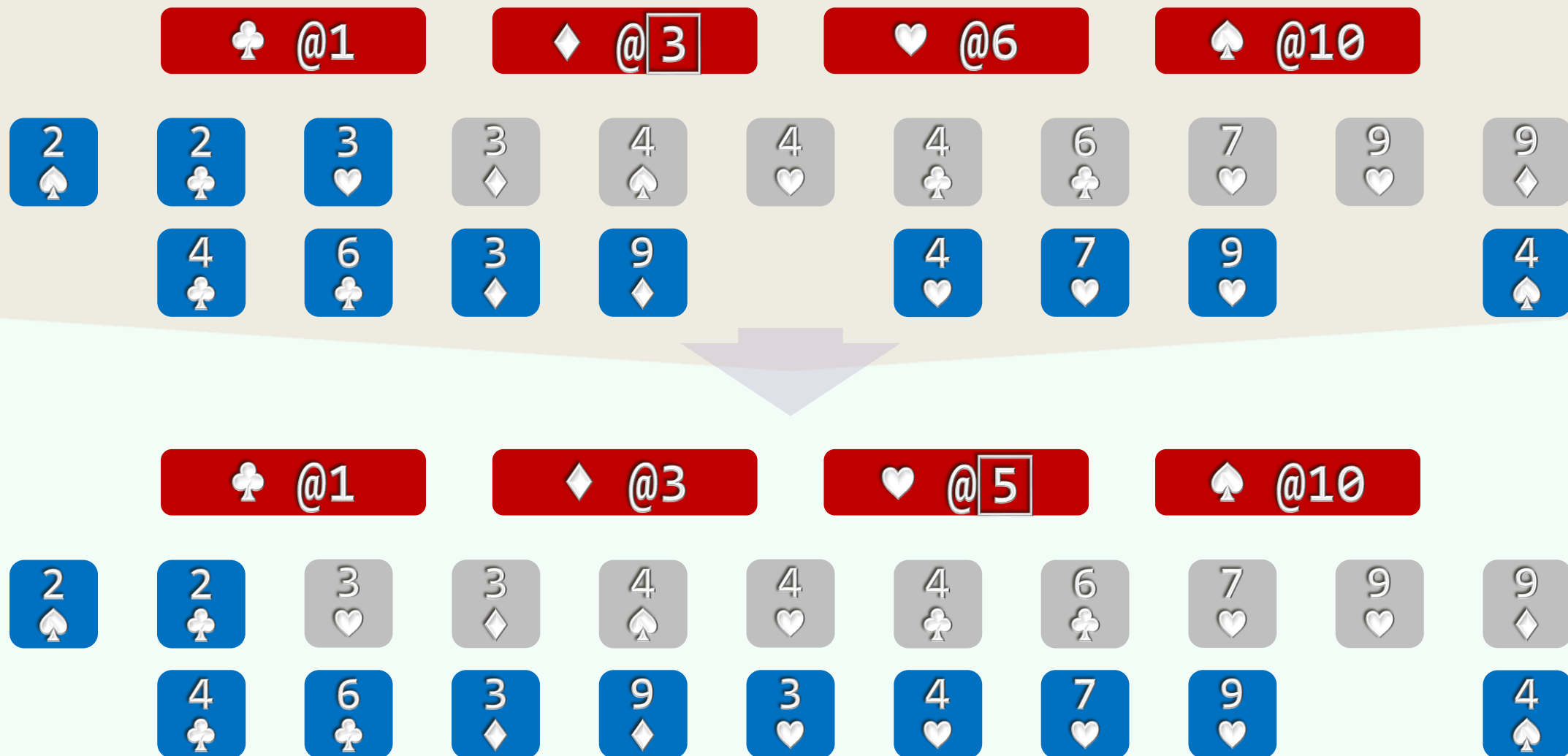
实例 (7/11)



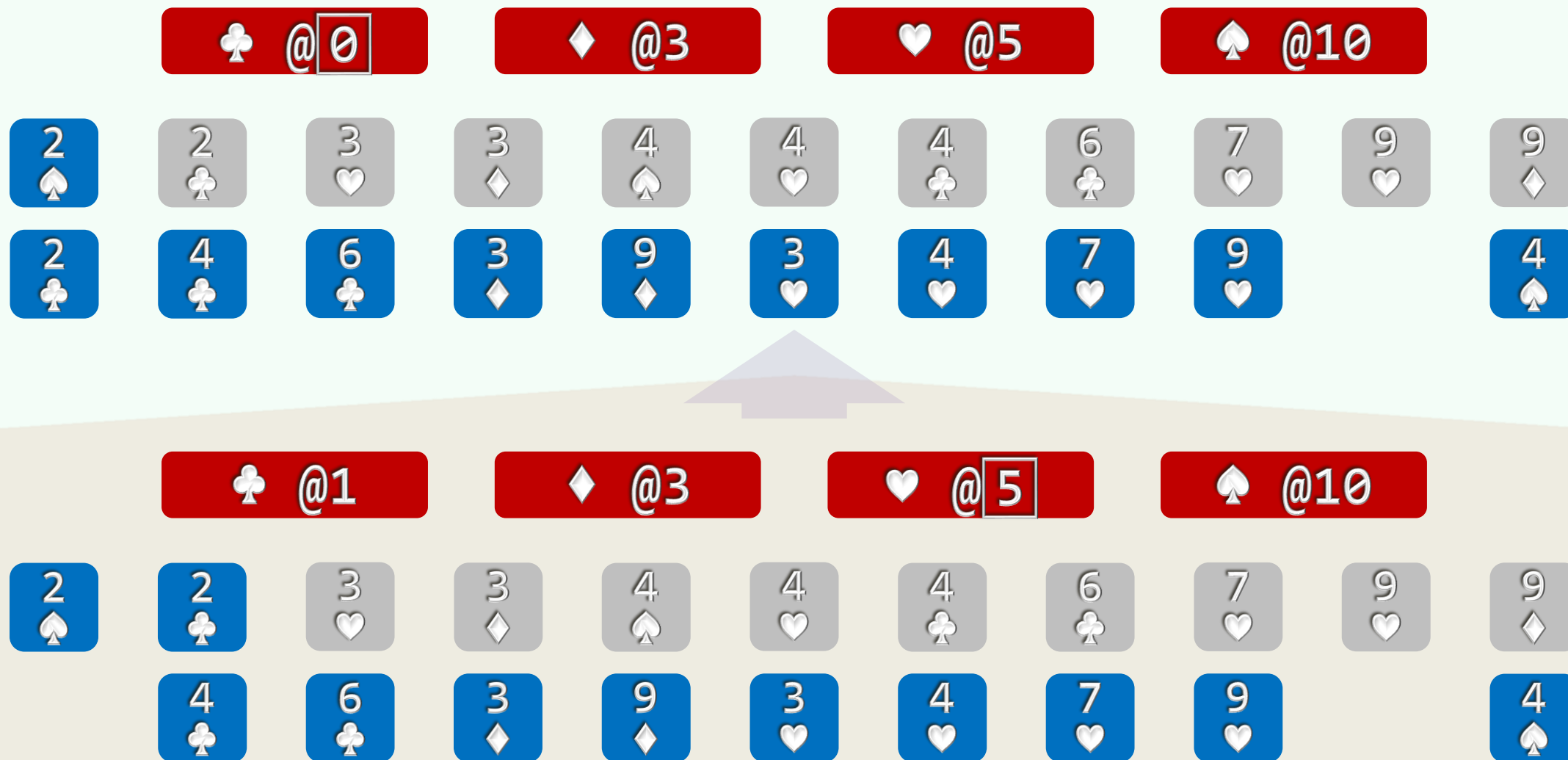
实例 (8/11)



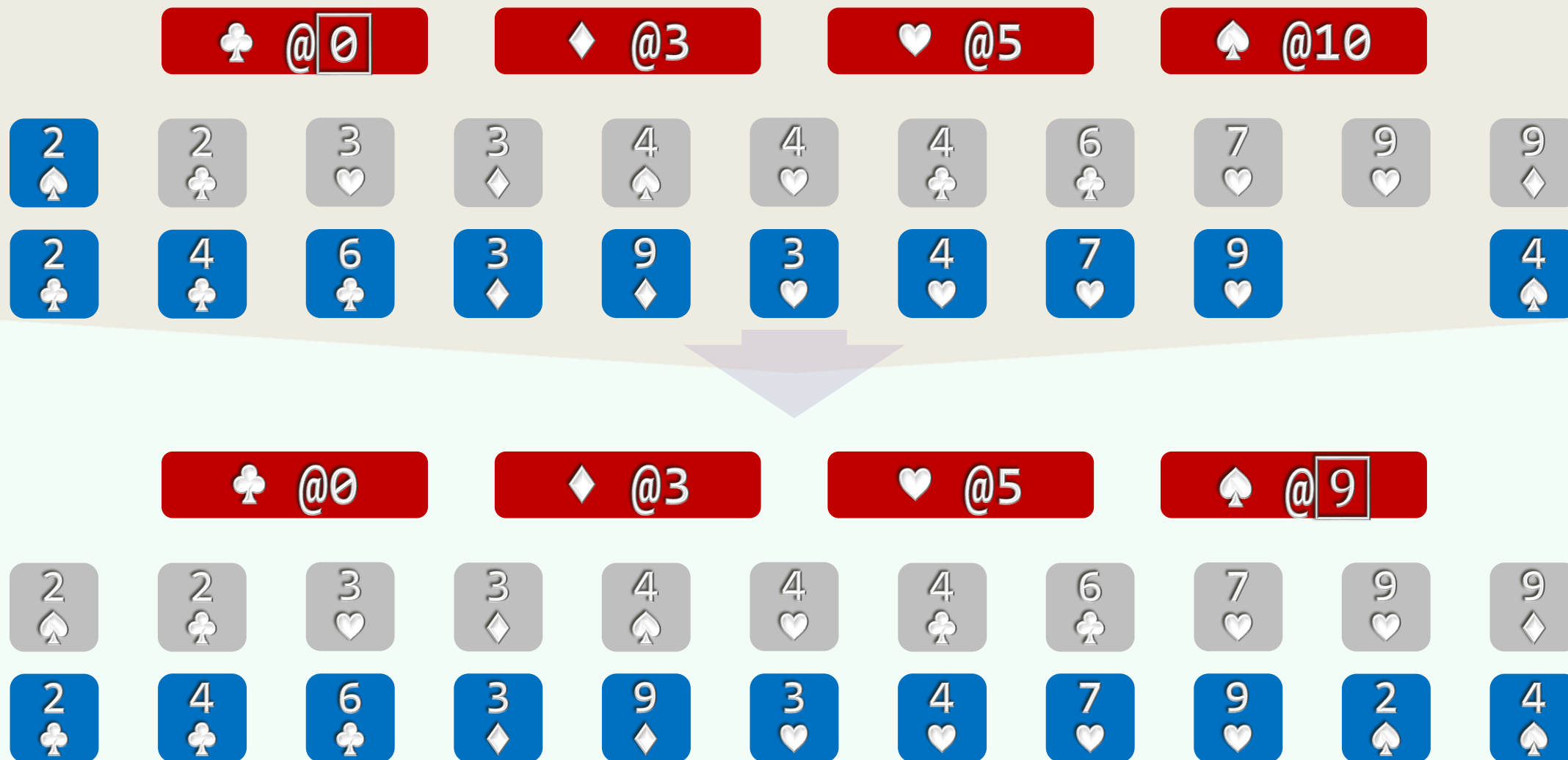
实例 (9/11)



实例 (10/11)



实例 (11/11)



分析

❖ 时间复杂度 = $\mathcal{O}(n + m + n) = \mathcal{O}(n)$ ——高效处理大规模数据

❖ 空间复杂度 = $\mathcal{O}(n)$ ——不能就地完成，需借助外存

❖ 最后一步的扫描次序，可否改为自前向后？

