CMPS 2200 – Fall 2015

Sorting

Carola Wenk

Slides courtesy of Charles Leiserson with changes and additions by Carola Wenk

How fast can we sort?

All the sorting algorithms we have seen so far are ***comparison sorts***: only use comparisons to determine the relative order of elements.

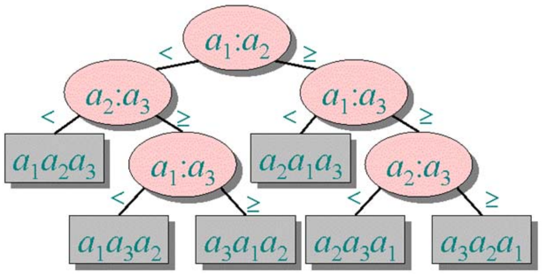
- *E.g.*, insertion sort, merge sort, heapsort.

The best worst-case running time that we've seen for comparison sorting is $O(n \log n)$.

Is $O(n \log n)$ the best we can do?

Decision trees can help us answer this question.

Decision-tree model

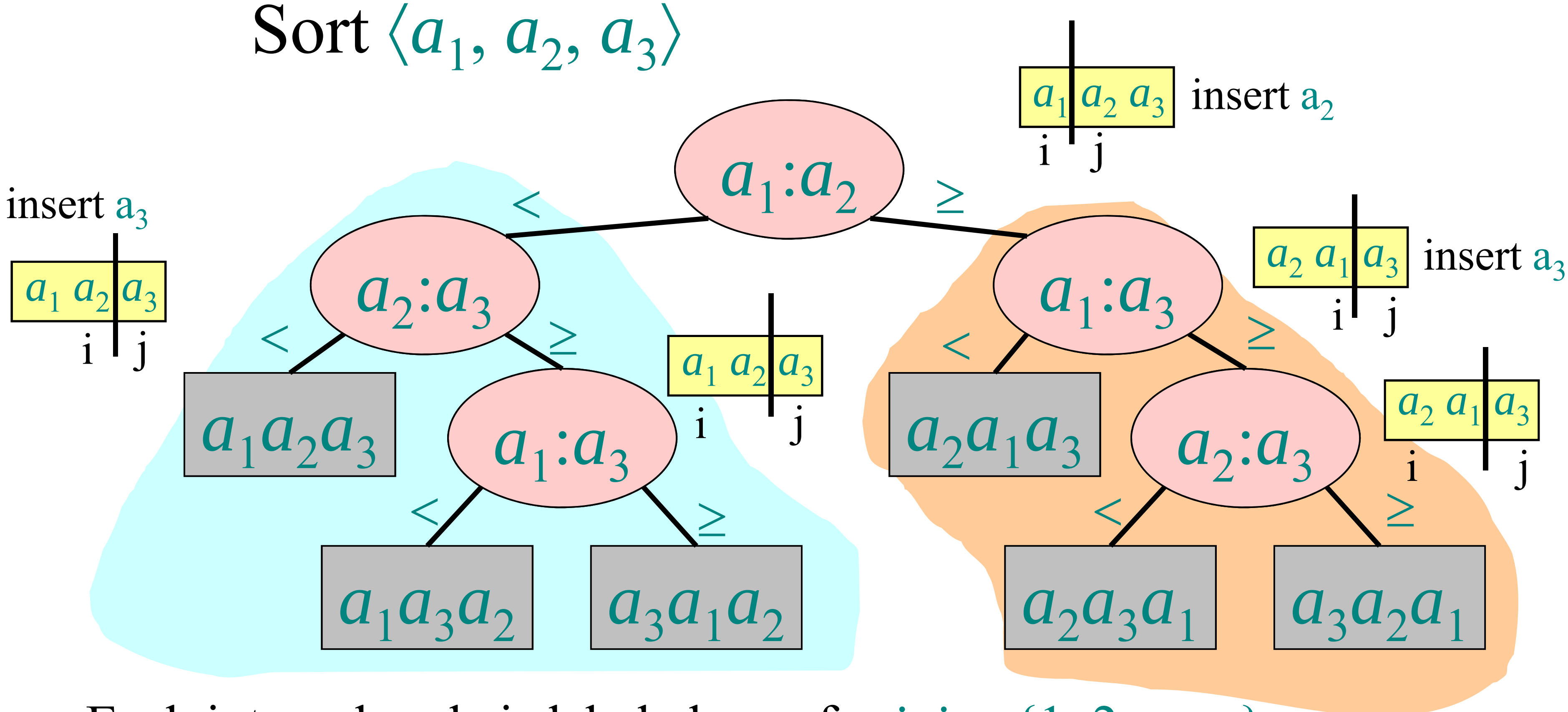


A decision tree models the execution of any comparison sorting algorithm:

- One tree per input size n.
- The tree contains **all** possible comparisons (= if-branches) that could be executed for **any** input of size n.
- The tree contains **all** comparisons along **all** possible instruction traces (= control flows) for **all** inputs of size n.
- For one input, only one path to a leaf is executed.
- Running time = length of the path taken.
- Worst-case running time = height of tree.

Decision-tree for insertion sort

Sort $\langle a_1, a_2, a_3 \rangle$

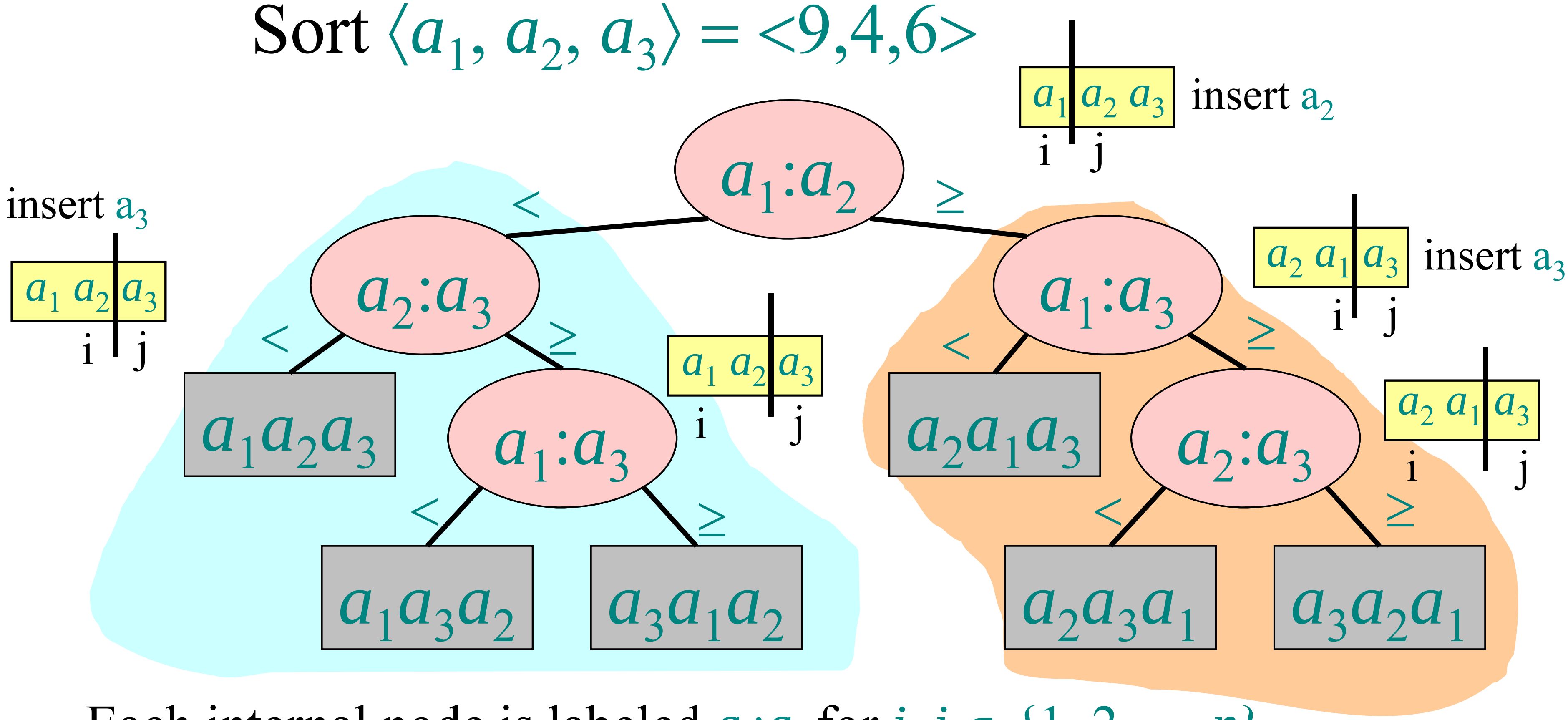


Each internal node is labeled a_i:a_j for $i, j \in \{1, 2, \ldots, n\}$.

- The left subtree shows subsequent comparisons if $a_i < a_j$.
- The right subtree shows subsequent comparisons if $a_i \geq a_j$.

Decision-tree for insertion sort

Sort $\langle a_1, a_2, a_3 \rangle = <9,4,6>$

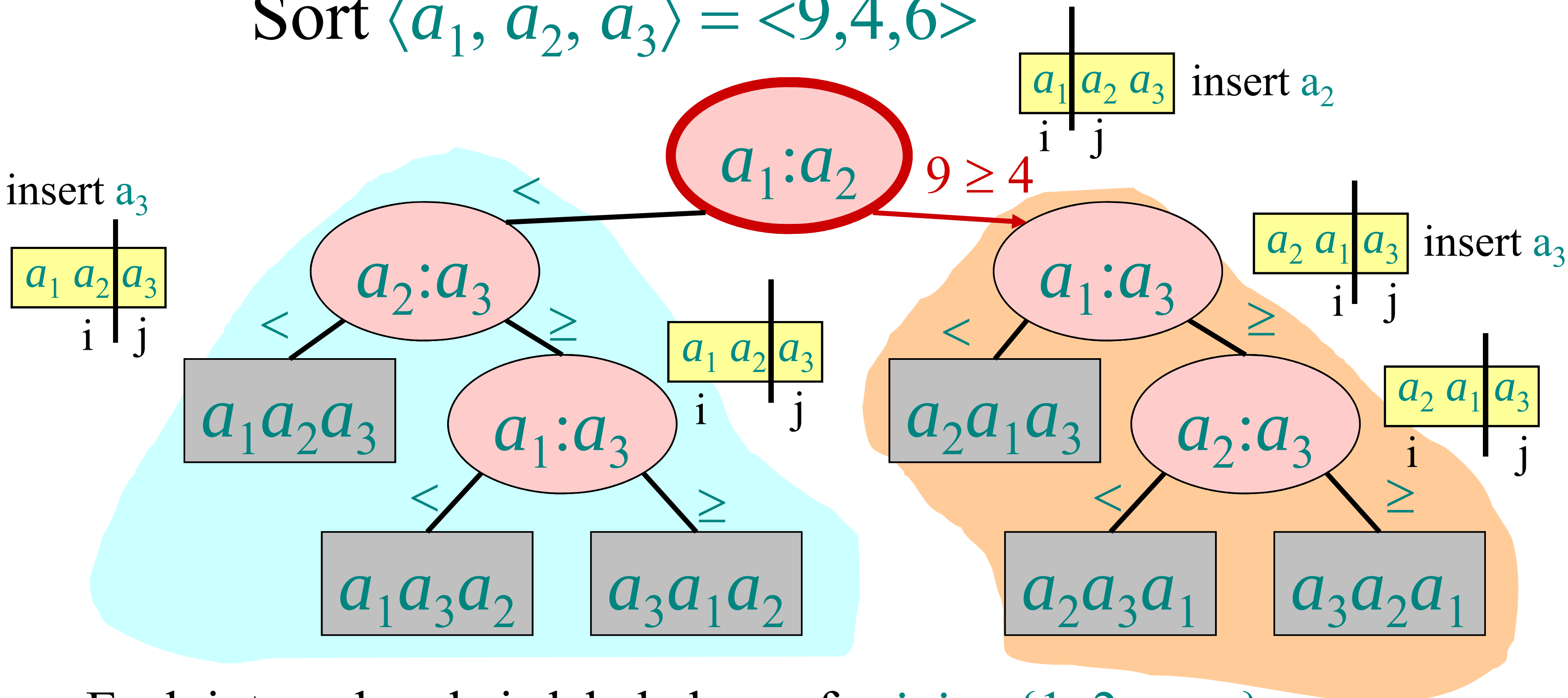


Each internal node is labeled $a_i{:}a_j$ for $i, j \in \{1, 2, \ldots, n\}$.

- The left subtree shows subsequent comparisons if $a_i < a_j$.
- The right subtree shows subsequent comparisons if $a_i \geq a_j$.

Decision-tree for insertion sort

Sort $\langle a_1, a_2, a_3 \rangle = <9,4,6>$

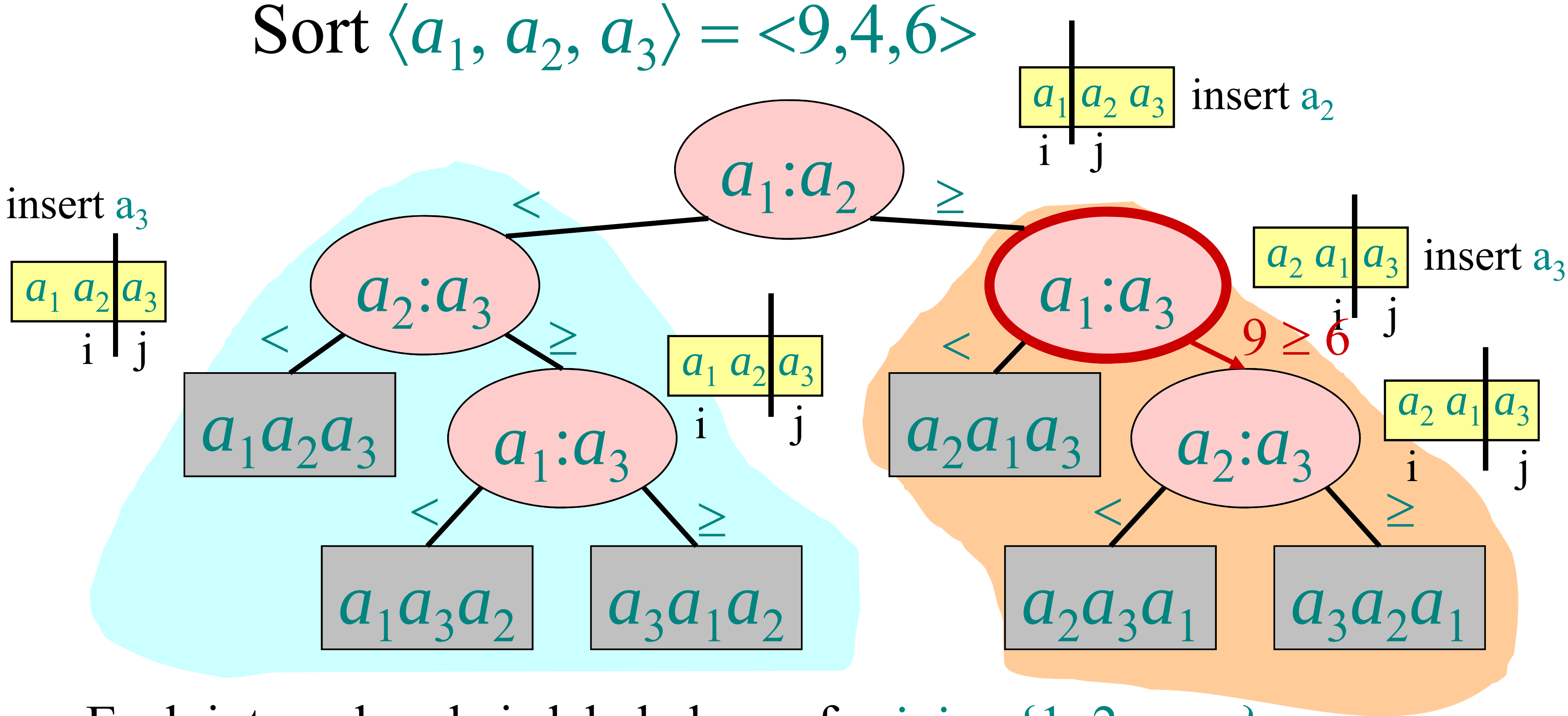


Each internal node is labeled $a_i{:}a_j$ for $i, j \in \{1, 2, \ldots, n\}$.

- The left subtree shows subsequent comparisons if $a_i < a_j$.
- The right subtree shows subsequent comparisons if $a_i \geq a_j$.

Decision-tree for insertion sort

Sort $\langle a_1, a_2, a_3 \rangle = <9,4,6>$

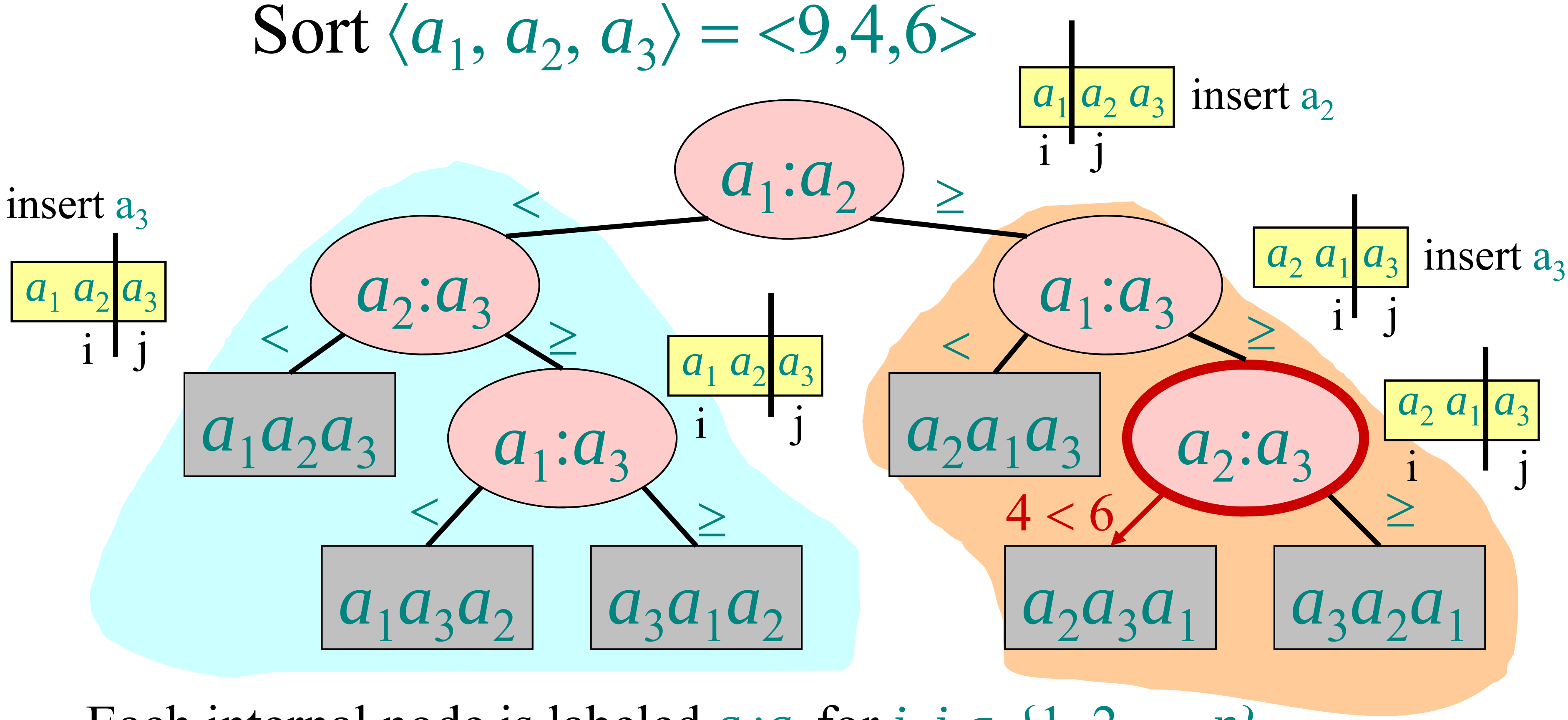


Each internal node is labeled $a_i{:}a_j$ for $i, j \in \{1, 2,\ldots, n\}$.

- The left subtree shows subsequent comparisons if $a_i < a_j$.
- The right subtree shows subsequent comparisons if $a_i \geq a_j$.

Decision-tree for insertion sort

Sort $\langle a_1, a_2, a_3 \rangle = <9,4,6>$

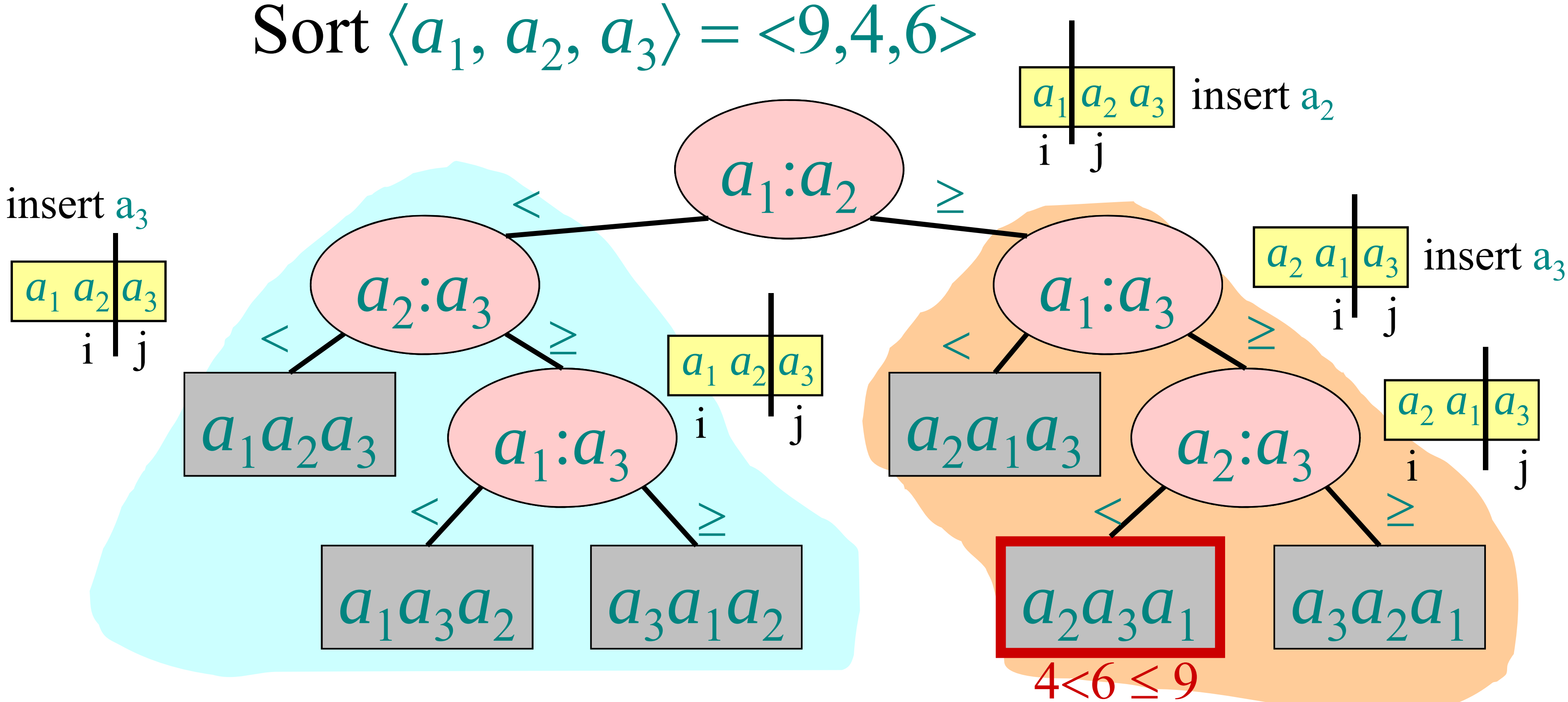


Each internal node is labeled $a_i{:}a_j$ for $i, j \in \{1, 2, \ldots, n\}$.

- The left subtree shows subsequent comparisons if $a_i < a_j$.
- The right subtree shows subsequent comparisons if $a_i \geq a_j$.

Decision-tree for insertion sort

Sort $\langle a_1, a_2, a_3 \rangle = <9,4,6>$

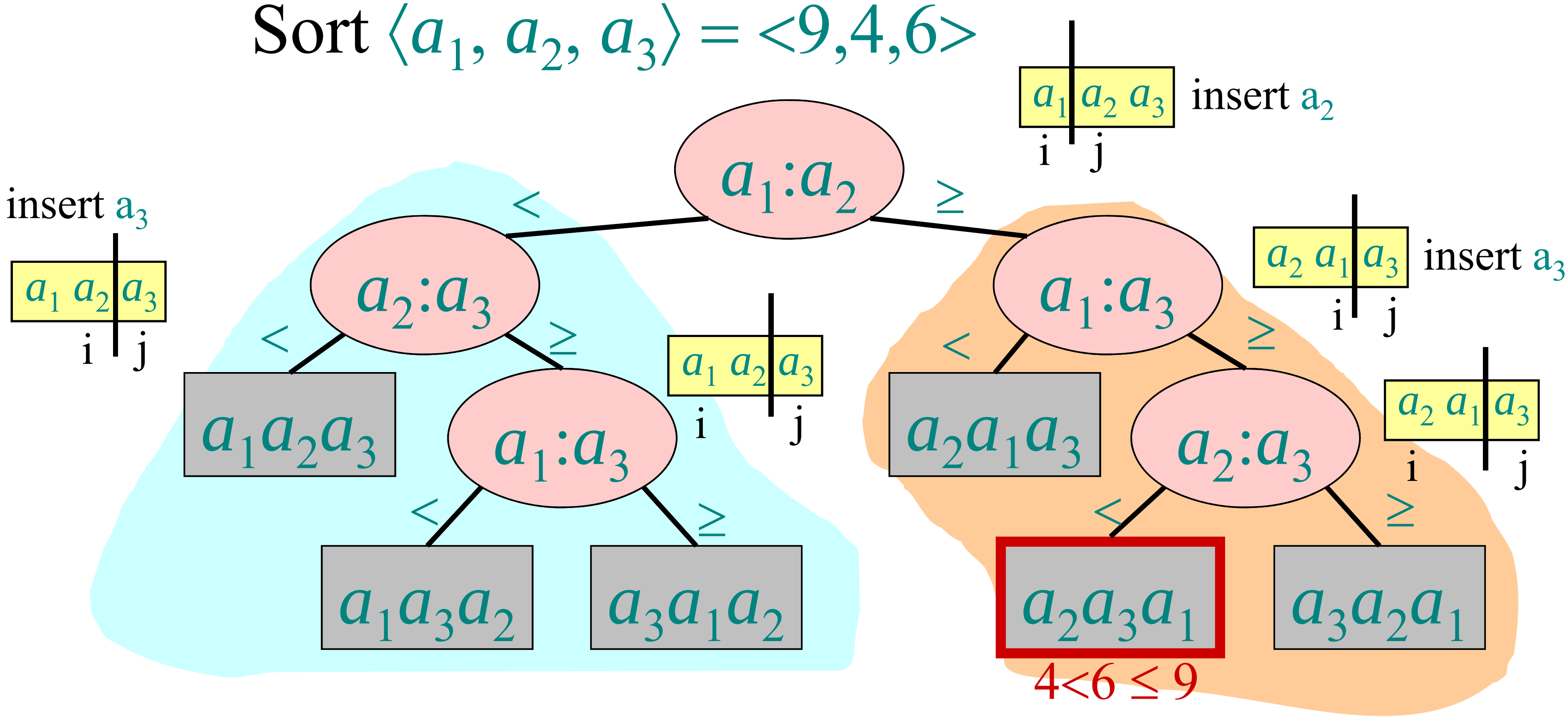


Each internal node is labeled a_i:a_j for $i, j \in \{1, 2,\ldots, n\}$.

- The left subtree shows subsequent comparisons if $a_i < a_j$.
- The right subtree shows subsequent comparisons if $a_i \geq a_j$.

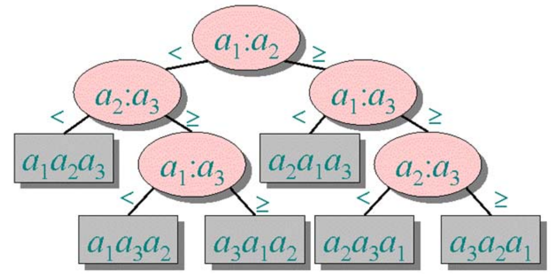
Decision-tree for insertion sort

Sort $\langle a_1, a_2, a_3 \rangle = <9,4,6>$



Each leaf contains a permutation $\langle \pi(1), \pi(2), \ldots, \pi(n) \rangle$ to indicate that the ordering $a_{\pi(1)} \leq a_{\pi(2)} \leq \ldots \leq a_{\pi(n)}$ has been established.

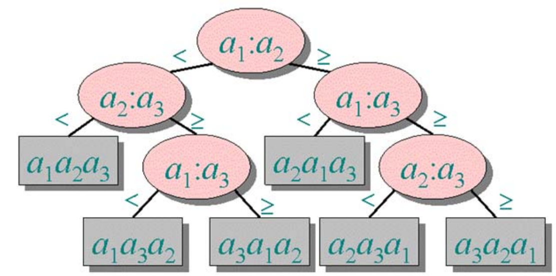
Decision-tree model



A decision tree models the execution of any comparison sorting algorithm:

- One tree per input size n.
- The tree contains **all** possible comparisons (= if-branches) that could be executed for **any** input of size n.
- The tree contains **all** comparisons along **all** possible instruction traces (= control flows) for **all** inputs of size n.
- For one input, only one path to a leaf is executed.
- Running time = length of the path taken.
- Worst-case running time = height of tree.

Lower bound for comparison sorting



Theorem. Any decision tree that can sort n elements must have height $\Omega(n \log n)$.

Proof. The tree must contain $\geq n!$ leaves, since there are $n!$ possible permutations. For a binary tree of height-h holds that #leaves $\leq 2^h$. Thus, $n! \leq 2^h$.

$$\begin{aligned}
\therefore\ h &\geq \log(n!) \qquad \text{(log is mono. increasing)} \\
&\geq \log((n/2)^{n/2}) \\
&= n/2 \log n/2 \\
\Rightarrow h &\in \Omega(n \log n).
\end{aligned}$$

Lower bound for comparison sorting

Corollary. Mergesort is an asymptotically optimal comparison sorting algorithm. ◻

Sorting in linear time

Counting sort: No comparisons between elements.

- ***Input***: $A[0 \,..\, n\text{-}1]$, where $A[j] \in \{0, 1, 2, \ldots, k\text{-}1\}$.
- ***Output***: $B[0 \,..\, n\text{-}1]$, sorted.
- ***Auxiliary storage***: $C[0 \,..\, k\text{-}1]$.

Counting sort

1. **for** ($i = 0$; $i < k$; i++)
   $C[i] = 0$
2. **for** ($j = 0$; $i < n$; j++)
   $C[A[j]] = C[A[j]] + 1$ // $C[i] == |\{\text{key} = i\}|$
3. **for** ($i = 1$; $i < k$; i++)
   $C[i] = C[i] + C[i–1]$ // $C[i] == |\{\text{key} \leq i\}|$
4. **for** ($j = n$-1; $i \geq 0$; j--)
   $B[C[A[j]]\text{-}1] = \mathrm{A}[j]$
   $C[A[j]] = C[A[j]] – 1$

Counting-sort example

	0	1	2	3	4
A:	3	0	2	3	2

	0	1	2	3
C:				

B:					

Loop 1

A:

0	1	2	3	4
3	0	2	3	2

C:

0	1	2	3
0	0	0	0

B:

1. **for** ($i = 0$; $i < k$; $i{+}{+}$)
   $C[i] = 0$

Loop 2

A:

0	1	2	3	4
3	0	2	3	2

C:

0	1	2	3
0	0	0	1

B:

2. for $(j = 0;\ i < n;\ j{+}{+})$

$C[A[j]] = C[A[j]] + 1$ $//\ C[i] == |\{\text{key} = i\}|$

Loop 2

A:

0	1	2	3	4
3	0	2	3	2

C:

0	1	2	3
1	0	0	1

B:

2. for $(j = 0;\ i < n;\ j{+}{+})$

$C[A[j]] = C[A[j]] + 1$ // $C[i] == |\{\text{key} = i\}|$

Loop 2

A:

0	1	2	3	4
3	0	2	3	2

C:

0	1	2	3
1	0	1	1

B:

2. for $(j = 0;\ i < n;\ j{+}{+})$

$C[A[j]] = C[A[j]] + 1$ $\quad // \ C[i] == |\{\text{key} = i\}|$

Loop 2

A:

0	1	2	3	4
3	0	2	3	2

C:

0	1	2	3
1	0	1	2

B:

2. for $(j = 0;\ i < n;\ j{+}{+})$

$C[A[j]] = C[A[j]] + 1 \qquad // \ C[i] == |\{\text{key} = i\}|$

Loop 2

	0	1	2	3	4
A:	3	0	2	3	2

	0	1	2	3
C:	1	0	2	2

B:					

2. for ($j = 0$; $i < n$; $j{+}{+}$)

$C[A[j]] = C[A[j]] + 1$ // $C[i] == |\{\text{key} = i\}|$

Loop 3

A:

0	1	2	3	4
3	0	2	3	2

C:

0	1	2	3
1	0	2	2

B:

C':

1	1	2	2

3. for ($i = 1$; $i < k$; $i{+}{+}$)

$C[i] = C[i] + C[i–1]$ // $C[i] == |\{\text{key} \leq i\}|$

Loop 3

	0	1	2	3	4
A:	3	0	2	3	2
B:					

	0	1	2	3
C:	1	0	2	2
C':	1	1	3	2

3. for $(i = 1;\ i < k;\ i{+}{+})$

$C[i] = C[i] + C[i\!-\!1]$ $\qquad$ // $C[i] == |\{\text{key} \le i\}|$

Loop 3

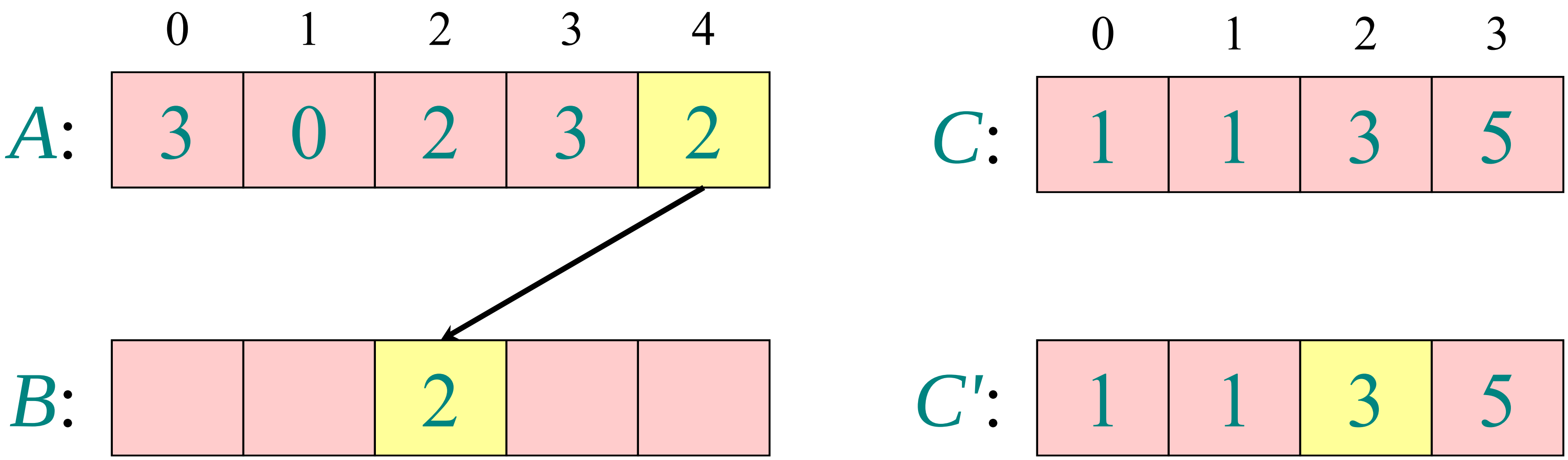
	0	1	2	3	4
A:	3	0	2	3	2
B:					

	0	1	2	3
C:	1	0	2	2
C':	1	1	3	5

3. for ($i = 1$; $i < k$; $i{+}{+}$)

$C[i] = C[i] + C[i–1]$ // $C[i] == |\{\text{key} \leq i\}|$

Loop 4

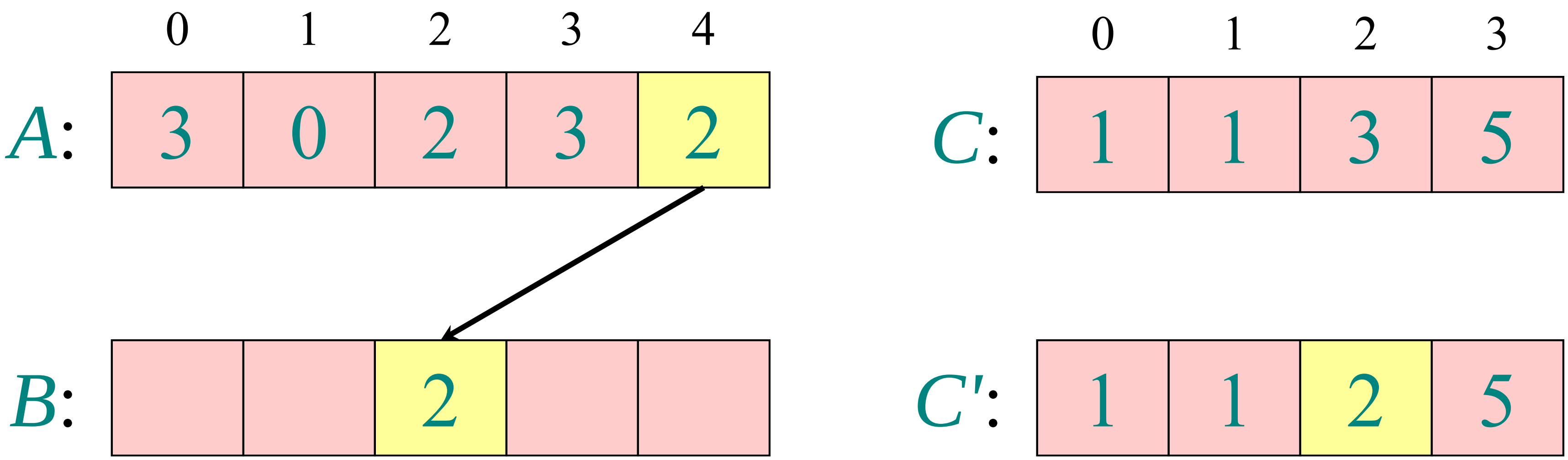


4. for ($j = n$-1; $i \geq 0$; j--)

$B[C[A[j]]-1] = A[j]$

$C[A[j]] = C[A[j]] - 1$

Loop 4

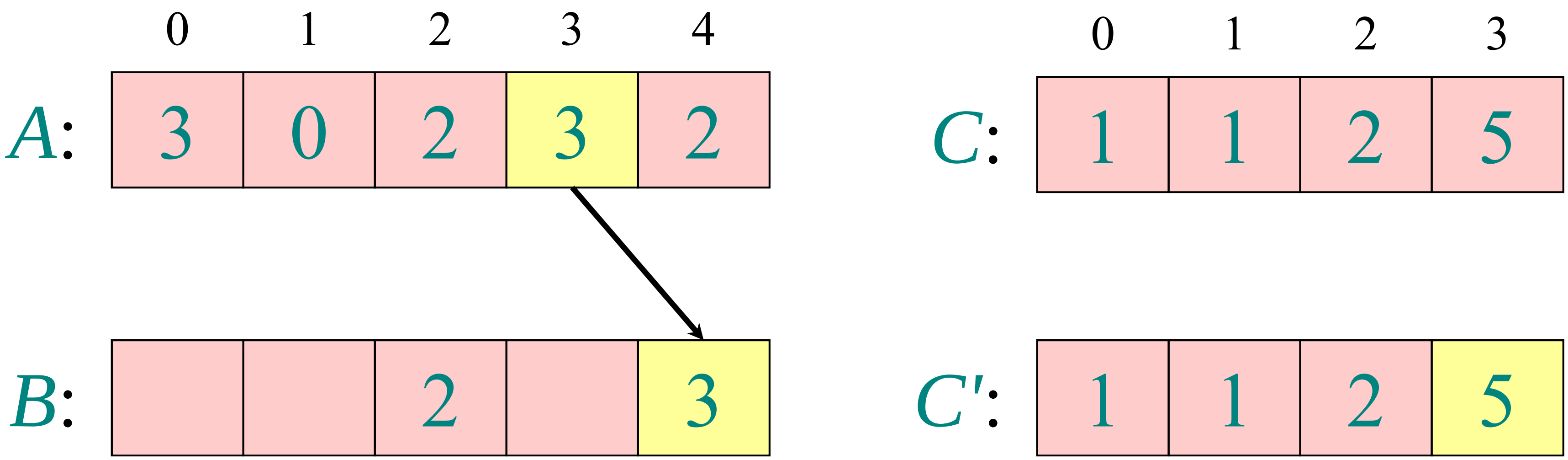


4. for ($j = n$-1; $i \geq 0$; j-- **)**

$B[C[A[j]]-1] = A[j]$

$C[A[j]] = C[A[j]] - 1$

Loop 4

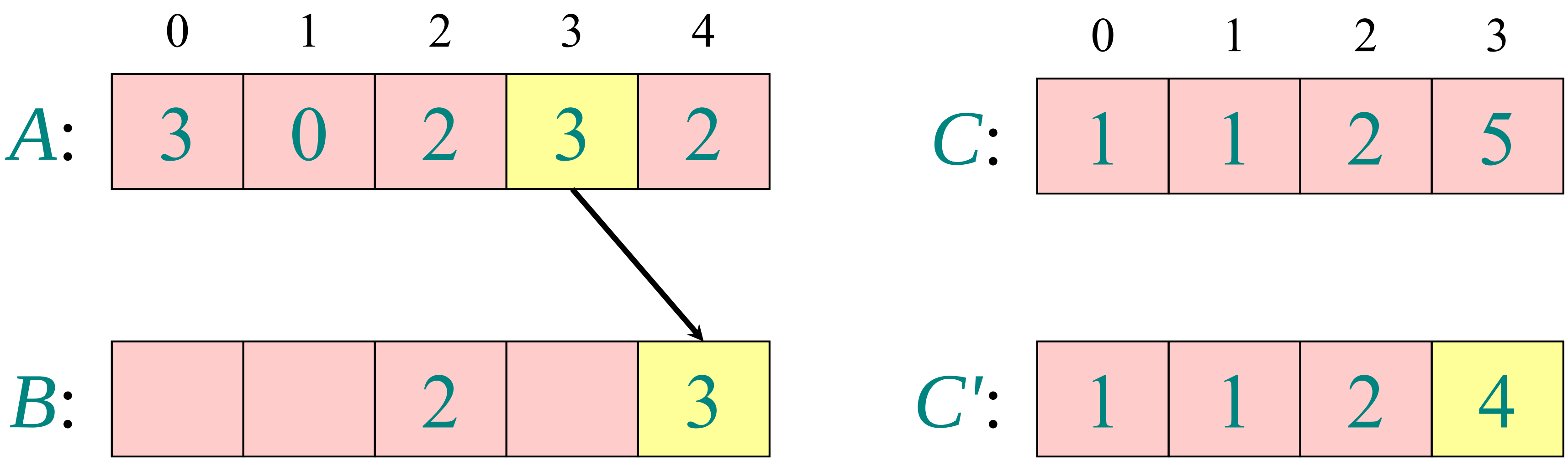


4. for ($j = n$-1; $i \geq 0$; j--**)**

$B[C[A[j]]-1] = A[j]$

$C[A[j]] = C[A[j]] - 1$

Loop 4

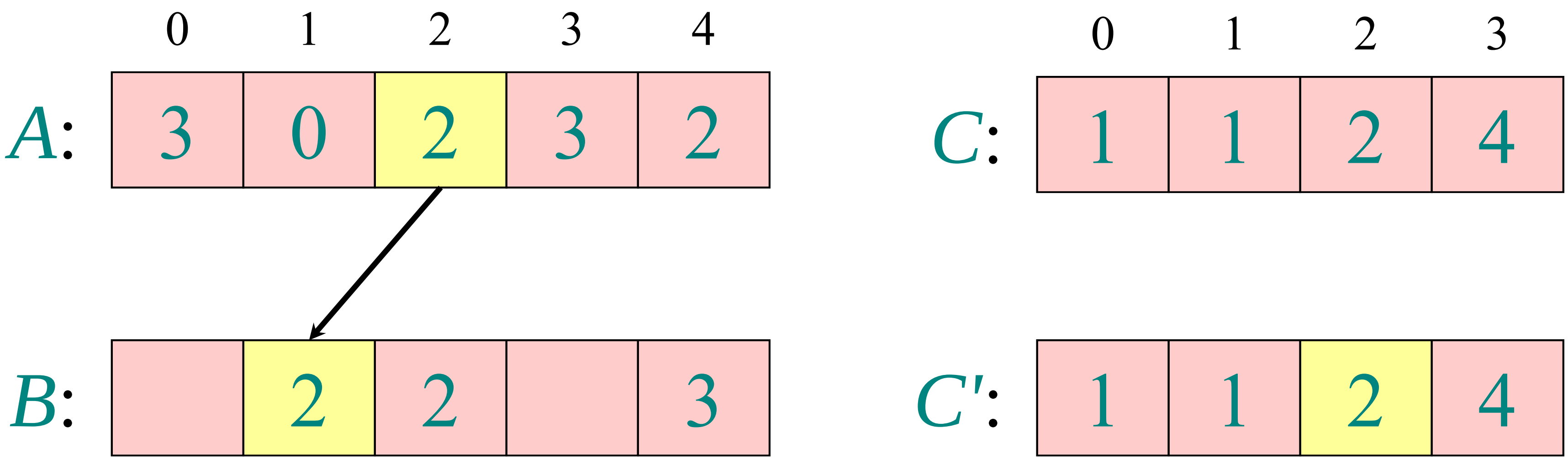


4. for **(**$j = n\text{-}1;\ i \geq 0;\ j\text{--}$**)**

$B[C[A[j]]\text{-}1] = \mathrm{A}[j]$

$C[A[j]] = C[A[j]] - 1$

Loop 4

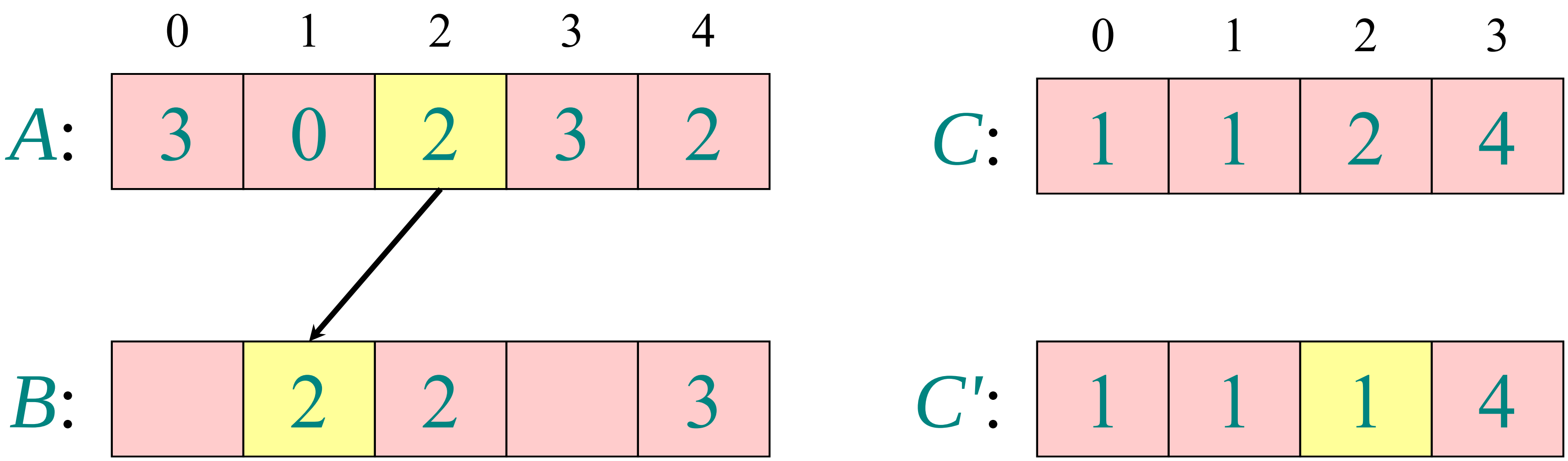


4. for ($j = n$-1; $i \geq 0$; j--**)**

$B[C[A[j]]\text{-}1] = \mathrm{A}[j]$

$C[A[j]] = C[A[j]] - 1$

Loop 4

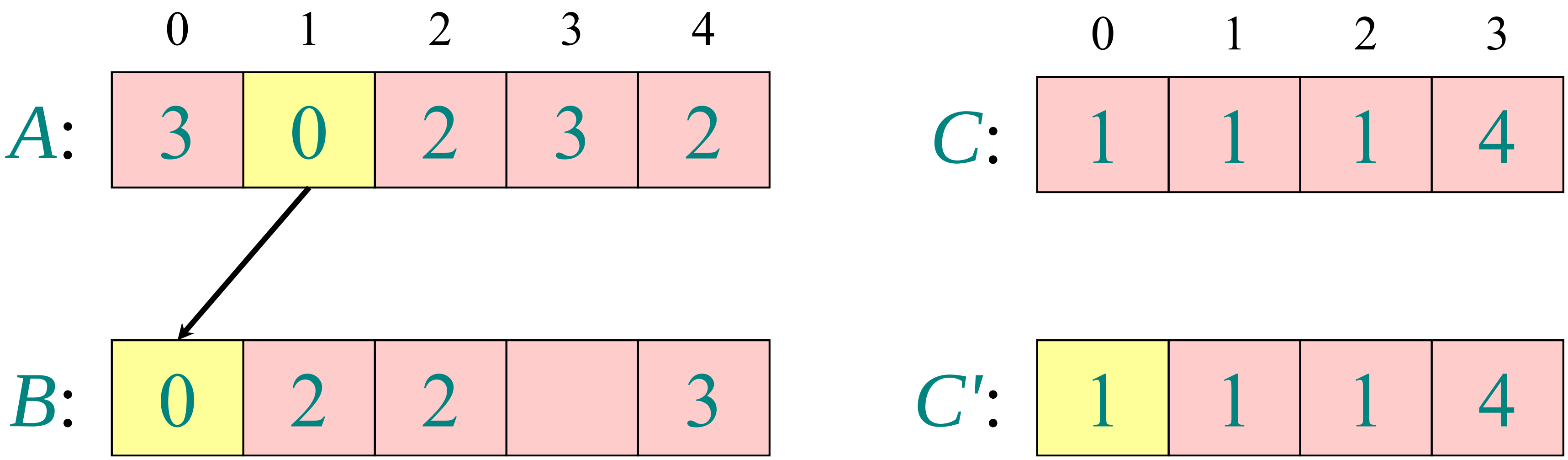


4. for $(j = n\text{-}1;\ i \geq 0;\ j\text{--})$

$B[C[A[j]]\text{-}1] = \mathrm{A}[j]$

$C[A[j]] = C[A[j]] - 1$

Loop 4

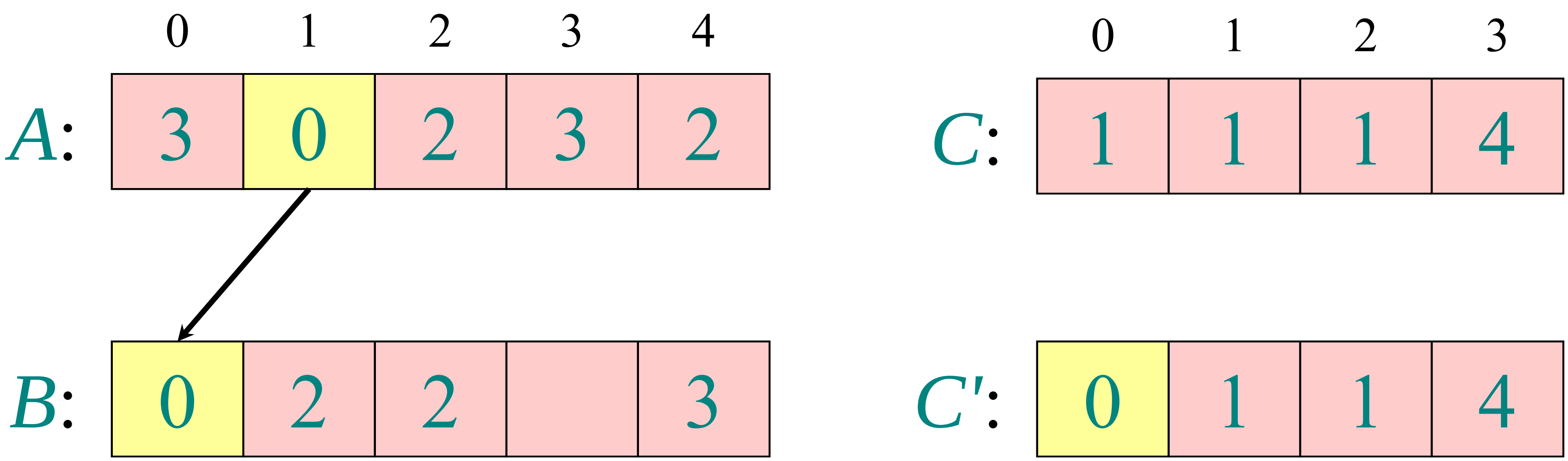


4. for **(**$j = n\text{-}1$; $i \geq 0$; j--**)**

$B[C[A[j]]\text{-}1] = \mathrm{A}[j]$

$C[A[j]] = C[A[j]] - 1$

Loop 4

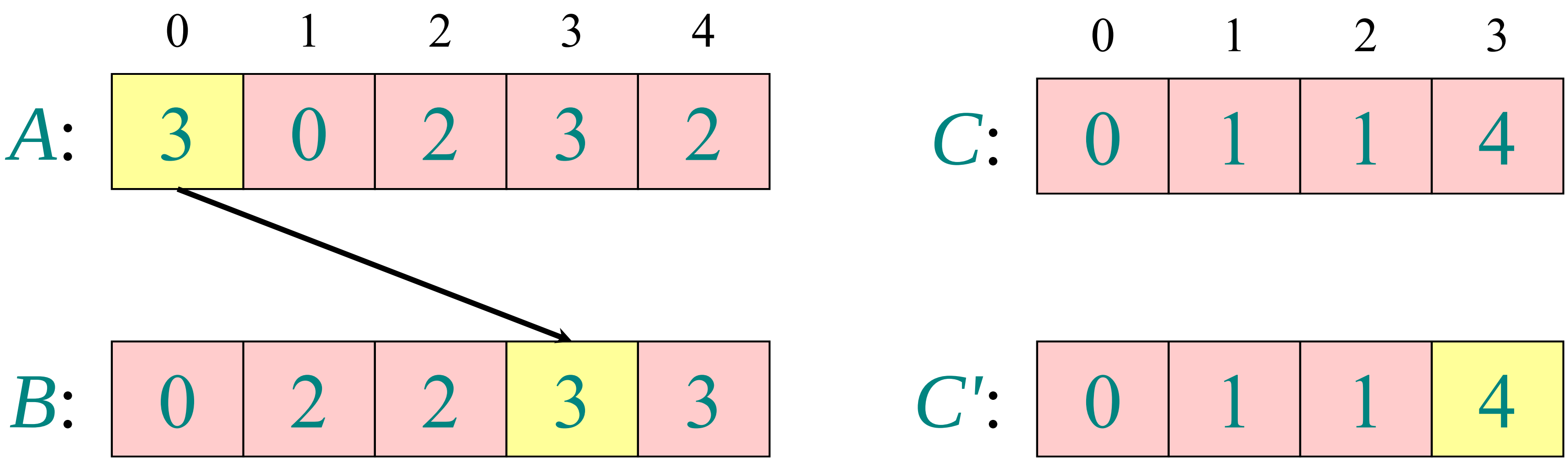


4. for ($j = n$-1; $i \geq 0$; j--**)**

$B[C[A[j]]\text{-}1] = \mathrm{A}[j]$

$C[A[j]] = C[A[j]] - 1$

Loop 4

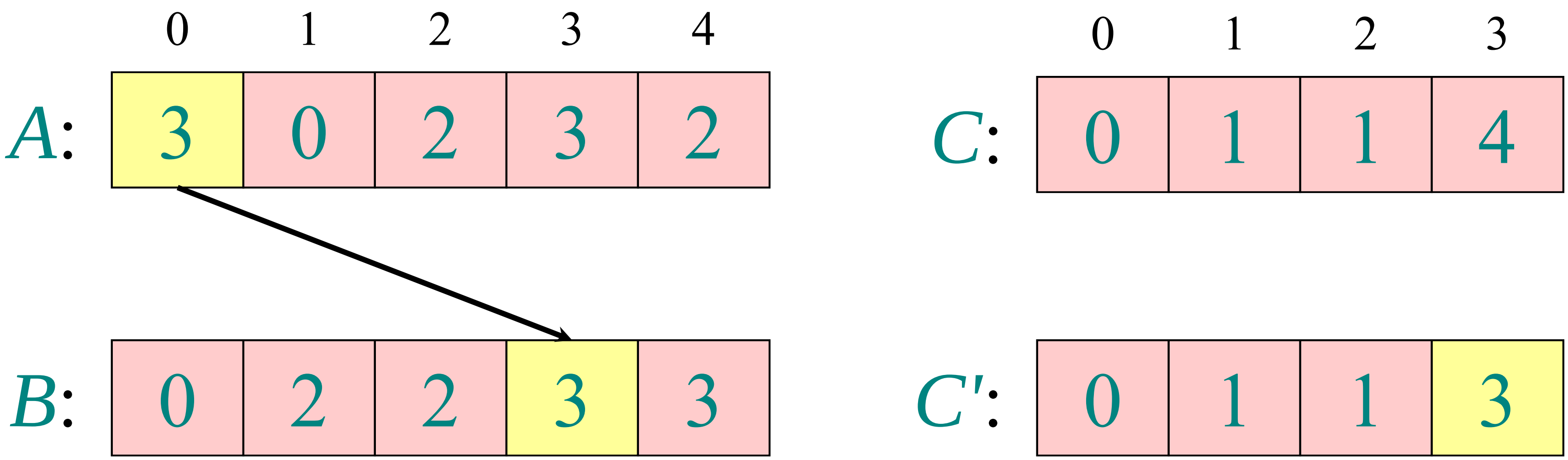


4. for ($j = n$-1; $i \geq 0$; j--)

$B[C[A[j]]-1] = \mathrm{A}[j]$

$C[A[j]] = C[A[j]] - 1$

Loop 4



4. for ($j = n$-1; $i \geq 0$; j--**)**

$B[C[A[j]]\text{-}1] = \mathrm{A}[j]$

$C[A[j]] = C[A[j]] - 1$

Analysis

$\Theta(k)$ — **1.for** ($i = 0$; $i < k$; $i{+}{+}$)
$C[i] = 0$

$\Theta(n)$ — **2.for** ($j = 0$; $i < n$; $j{+}{+}$)
$C[A[\,j]] = C[A[\,j]] + 1$

$\Theta(k)$ — **3.for** ($i = 1$; $i < k$; $i{+}{+}$)
$C[i] = C[i] + C[i–1]$

$\Theta(n)$ — **4.for** ($j = n$-1; $i \geq 0$; j--)
$B[C[A[\,j]]\text{-}1] = \mathrm{A}[\,j]$
$C[A[\,j]] = C[A[\,j]] – 1$

$\Theta(n + k)$

Running time

If $k = O(n)$, then counting sort takes $\Theta(n)$ time.

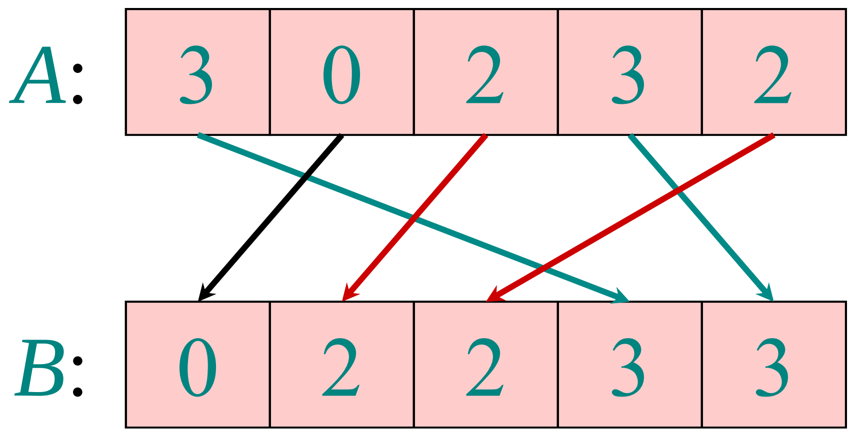
- But, sorting takes $\Omega(n \log n)$ time!
- Where's the fallacy?

Answer:

- ***Comparison sorting*** takes $\Omega(n \log n)$ time.
- Counting sort is not a ***comparison sort***.
- In fact, not a single comparison between elements occurs!

Stable sorting

Counting sort is a ***stable*** sort: it preserves the input order among equal elements.



Exercise: What other sorts have this property?