实验1：排序算法性能分析

报告人：郑杨 陈敏涵

深圳大学
SHENZHEN UNIVERSITY

01

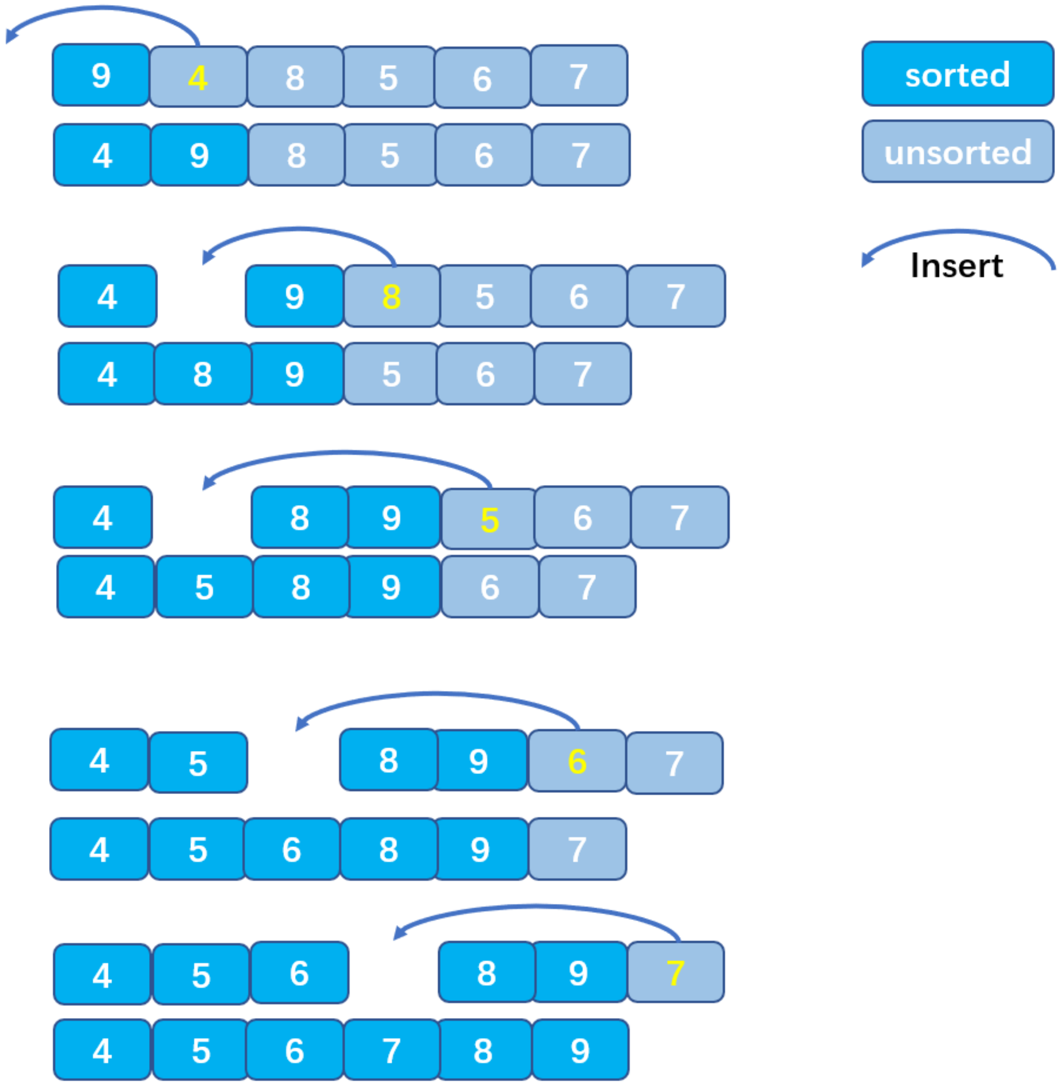
PART 01

SECTION 1

Performance Analysis

Performance Analysis

1、*Insertion sort*



Performance Analysis

深圳大学 SHENZHEN UNIVERSITY

Time complexity:

$INSERTION\_SORT(A, n)$	$cost$	$time$
$\quad for\ i = 2\ to\ n$	c_1	n
$\quad\quad key = A[i]$	c_2	$n-1$
$\quad\quad j = i - 1$	c_3	$n-1$
$\quad\quad while\ j >= 1\ and\ key < A[j]$	c_4	$\sum_{i=2}^{n} t_i$
$\quad\quad\quad A[j+1] = A[j]$	c_5	$\sum_{i=2}^{n}(t_i - 1)$
$\quad\quad\quad j--$	c_6	$\sum_{i=2}^{n}(t_i - 1)$
$\quad\quad A[j+1] = key$	c_7	$n-1$

Then:

$$T(n) = c_1 n + c_2(n-1) + c_3(n-1) + c_4\sum_{i=2}^{n} t_i + c_5\sum_{i=2}^{n}(t_i - 1) + c_6\sum_{i=2}^{n}(t_i - 1) + c_7(n-1)$$

Performance Analysis

深圳大学 SHENZHEN UNIVERSITY

Time complexity:

Space complexity: $O(1)$

$O(n^2)$

- ***Worst case***

$t_i = i$

$$\sum_{i=2}^{n} t_i = \frac{n(n+1)}{2} - 1, \sum_{i=2}^{n} (t_i - 1) = \frac{n(n-1)}{2}$$

$$T(n) = c_1 n + c_2(n-1) + c_3(n-1) + c_4\left(\frac{n(n+1)}{2} - 1\right) + c_5 \frac{n(n-1)}{2} + c_6 \frac{n(n-1)}{2} + c_7(n-1)$$

$$= \frac{1}{2}(c_4 + c_5 + c_6)n^2 + \left(c_1 + c_2 + c_3 + \frac{c_4}{2} - \frac{c_5}{2} - \frac{c_6}{2} + c_7\right)n - (c_2 + c_3 + c_4 + c_7)$$

$O(n)$

- ***Best case*** $t_i = 1$

$O(n^2)$

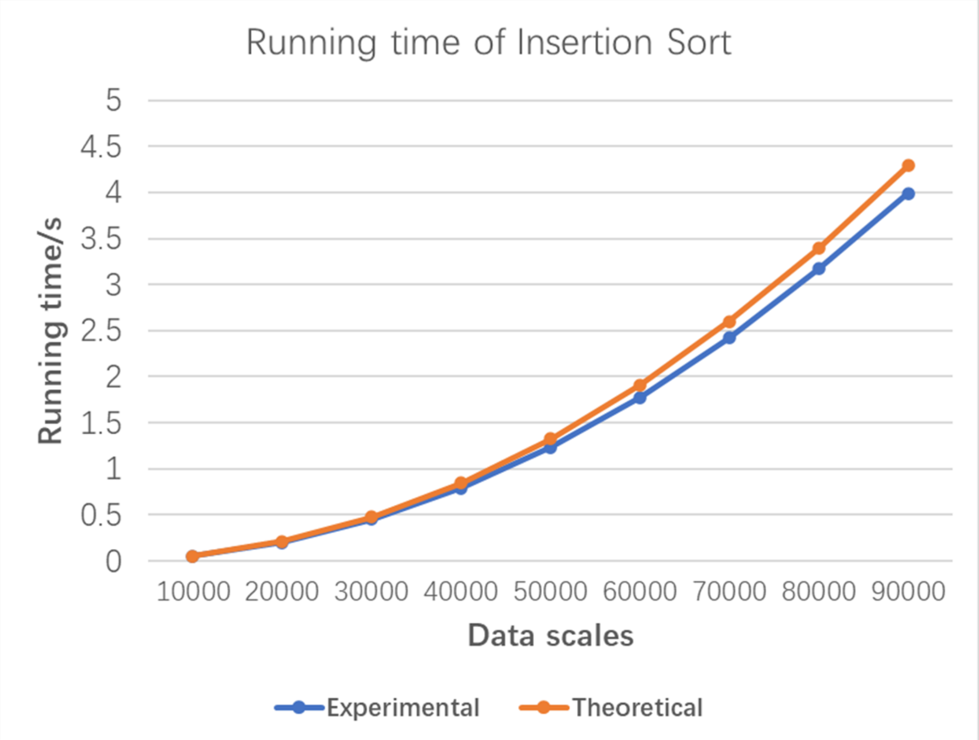
- ***Average case*** $E(t_i) = \frac{1}{i}(1 + 2 + \cdots + i) = \frac{i+1}{2} \approx \frac{i}{2}$

Performance Analysis

深圳大学
SHENZHEN UNIVERSITY

Testing:

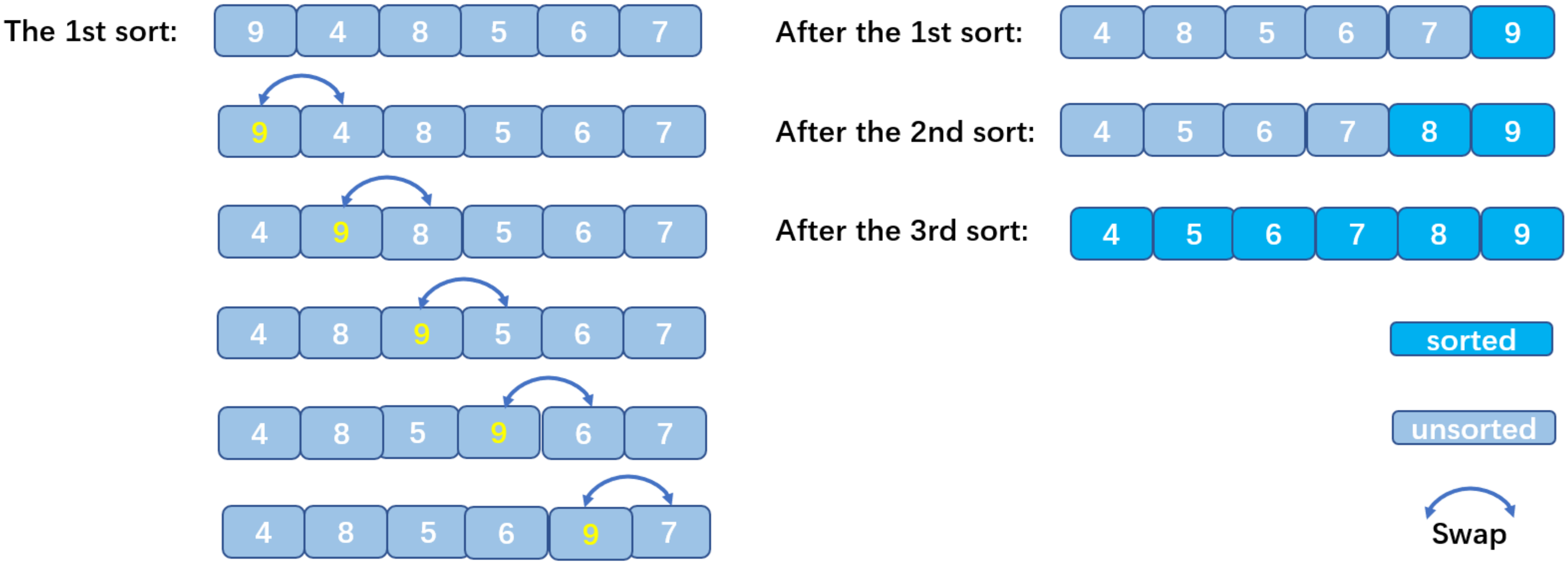
Data Scales	10000	20000	30000	40000	50000	60000	70000	80000	90000
Theoretical time/s	0.053	0.212	0.477	0.848	1.325	1.908	2.597	3.392	4.293
Experimental time/s	0.053	0.2	0.452	0.791	1.234	1.777	2.418	3.172	3.986
Error rate	0.00%	-5.66%	-5.24%	-6.72%	-6.87%	-6.87%	-6.89%	-6.49%	-7.15%



Performance Analysis

SHENZHEN UNIVERSITY 深圳大学

2、*Bubble sort*



Performance Analysis

深圳大学 SHENZHEN UNIVERSITY

Time complexity:

$BUBBLE\_SORT(A, n)$	*cost*	*time*
$\quad for\ i = n\ to\ 2$	c_1	n
$\quad\quad for\ j = 2\ to\ i$	c_2	$\sum_{i=2}^{n} i$
$\quad\quad\quad If\ A[j] < A[j-1]$	c_3	$(\sum_{i=2}^{n} i) - 1$
$\quad\quad\quad\quad Swap(A[j], A[j-1])$	c_4	$\sum_{i=2}^{n} t_i$

Then:

$$T(n) = c_1 n + c_2 \sum_{i=2}^{n} i + c_3((\sum_{i=2}^{n} i) - 1) + c_4 \sum_{i=2}^{n} t_i$$

Performance Analysis

深圳大学 SHENZHEN UNIVERSITY

Time complexity:

Space complexity: $O(1)$

$O(n^2)$

- ***Worst case***

$$\sum_{i=2}^{n} t_i = (\sum_{i=2}^{n} i) - 1 = \frac{n(n+1)}{2} - 2$$

$$T(n) = c_1 n + c_2(\frac{n(n+1)}{2} - 1) + (c_3 + c_4)(\frac{n(n+1)}{2} - 2)$$

$$= \frac{1}{2}(c_2 + c_3 + c_4)n^2 + (c_1 + \frac{c_2}{2} + \frac{c_3}{2} + \frac{c_4}{2})n - (c_2 + 2c_3 + 2c_4)$$

$O(n^2)$ or $O(n)$

- ***Best case*** $\sum_{i=2}^{n} t_i = 0$

$O(n^2)$

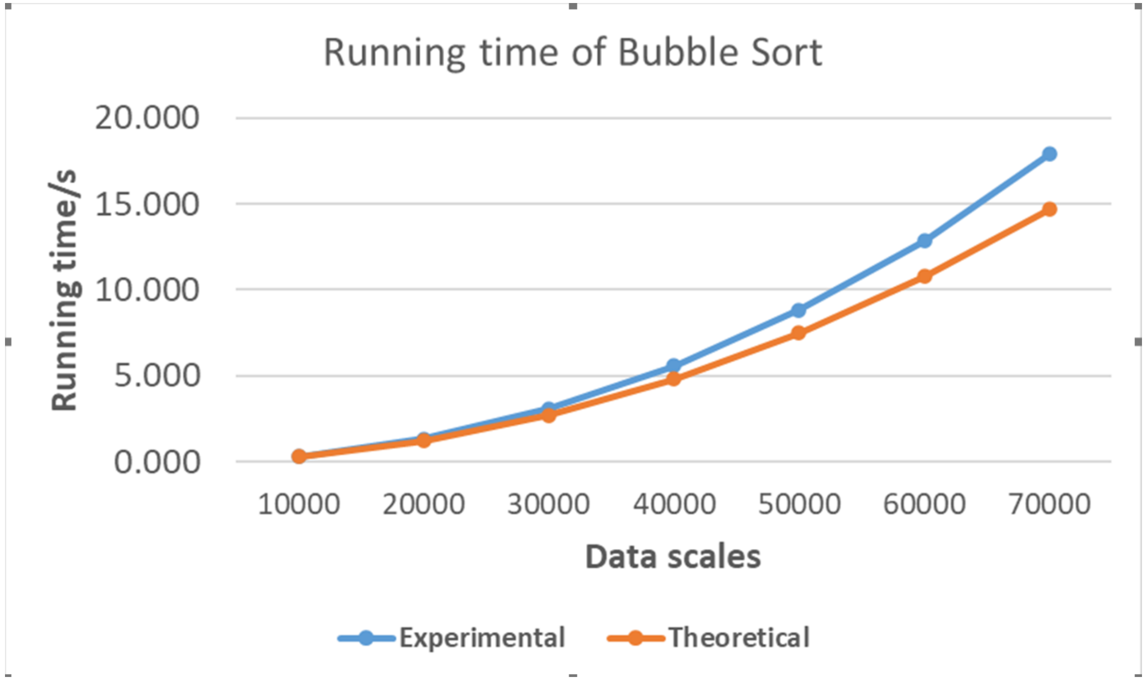
- ***Average case*** $E(t_i) = \frac{1}{i}(0 + 1 + \cdots + i - 1) = \frac{i-1}{2} \approx \frac{i}{2}$

Performance Analysis

SHENZHEN UNIVERSITY

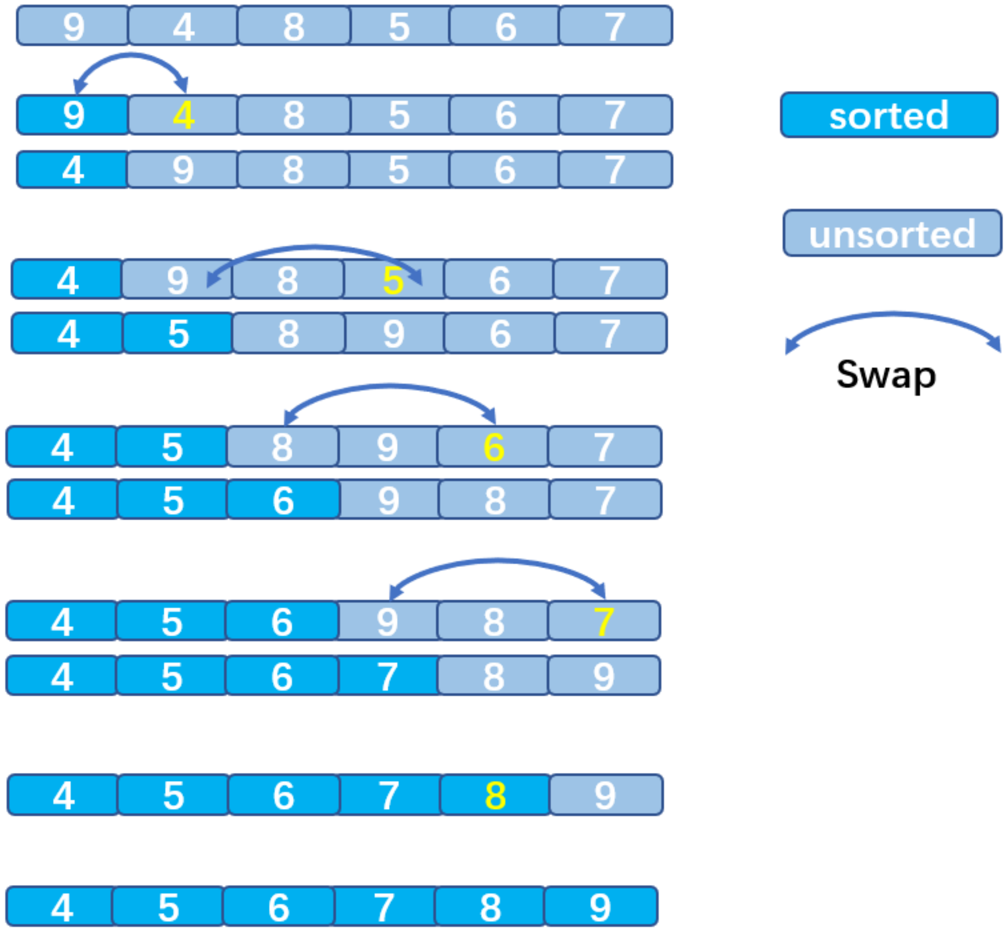
Testing:

Data Scales	10000	20000	30000	40000	50000	60000	70000
Theoretical time/s	0.300	1.199	2.698	4.796	7.494	10.791	14.688
Experimental time/s	0.300	1.318	3.093	5.582	8.811	12.854	17.874
Error rate	0.00%	9.93%	14.65%	16.38%	17.57%	19.11%	21.69%



Performance Analysis

3、*Selection sort*



Performance Analysis

深圳大学 SHENZHEN UNIVERSITY

Time complexity:

$SELECTION\_SORT(A,n)$	$cost$	$time$
$for\ i = 1\ to\ n-1$	c_1	n
$minval = A[i]$	c_2	$n-1$
$for\ j = i\ to\ n$	c_3	$\sum_{i=1}^{n-1} n-i+2$
$if\ A[j] < minval$	c_4	$\sum_{i=1}^{n-1} n-i+1$
$minval = A[j]$	c_5	$\sum_{i=1}^{n-1} t_i$
$k = j$	c_6	$\sum_{i=1}^{n-1} t_i$
$A[k] = A[i]$	c_7	$n-1$
$A[i] = minval$	c_8	$n-1$

Then:

$$T(n) = c_1 n + c_2(n-1) + c_3\left(\sum_{i=1}^{n-1} n-i+2\right) + c_4\left(\sum_{i=1}^{n-1} n-i+1\right) + c_5\left(\sum_{i=1}^{n-1} t_i\right) + c_6\left(\sum_{i=1}^{n-1} t_i\right) + c_7(n-1) + c_8(n-1)$$

Performance Analysis

深圳大学 SHENZHEN UNIVERSITY

***Time complexity*:** ***Space complexity*: $O(1)$**

$O(n^2)$ • ***Worst case*** $\sum_{i=1}^{n-1} t_i = \sum_{i=1}^{n-1} n-i+1 = \frac{(n+2)(n-1)}{2}$

$$T(n) = c_1 n + c_2(n-1) + c_3\frac{(n+4)(n-1)}{2} + c_4\frac{(n+2)(n-1)}{2}$$

$$+c_5\frac{(n+2)(n-1)}{2} + c_6\frac{(n+2)(n-1)}{2} + c_7(n-1) + c_8(n-1)$$

$$= \frac{1}{2}(c_3+c_4+c_5+c_6)n^2 + (c_1+c_2+\frac{3}{2}c_3+\frac{1}{2}c_4+\frac{1}{2}c_5+\frac{1}{2}c_6)n - (c_2+2c_3+c_4+c_5+c_6+c_7+c_8)$$

$O(n^2)$ • ***Best case*** $\sum_{i=2}^{n} t_i = 0$

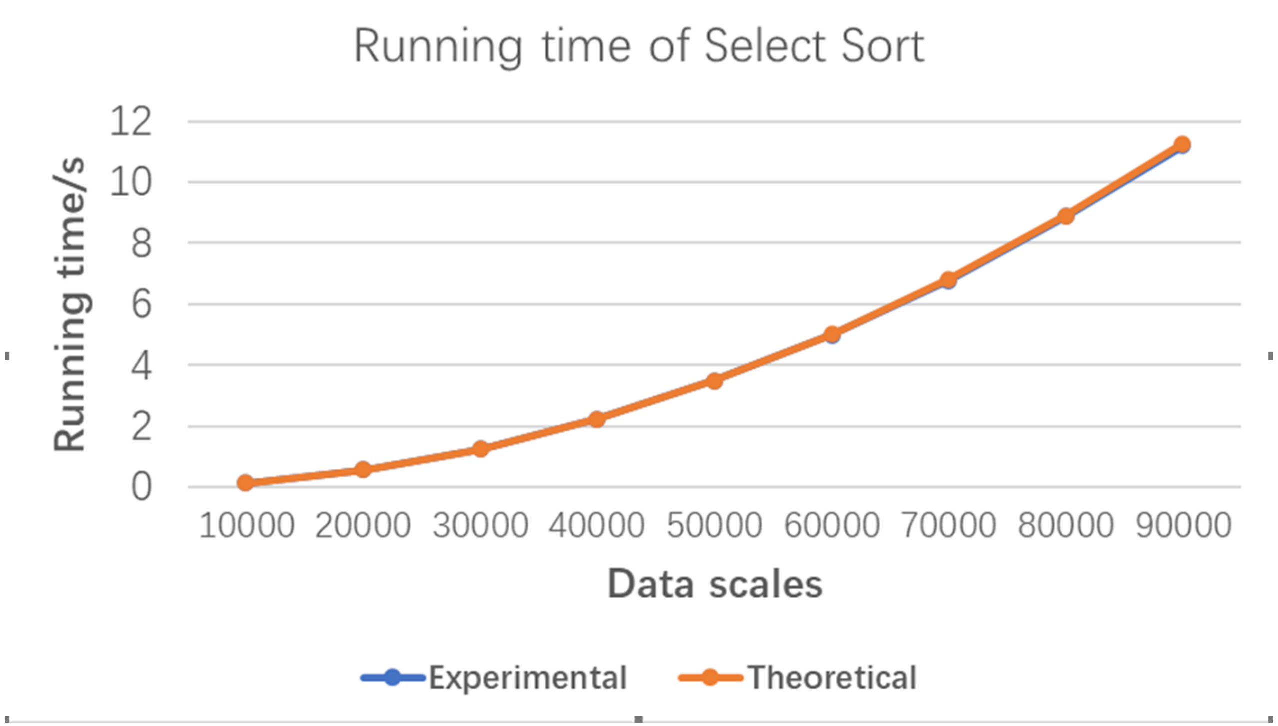
$O(n^2)$ • ***Average case*** $E(t_i) = \frac{1}{i}(0+1+\cdots+i-1) = \frac{i-1}{2} \approx \frac{i}{2}$

Performance Analysis

深圳大学 SHENZHEN UNIVERSITY

Testing:

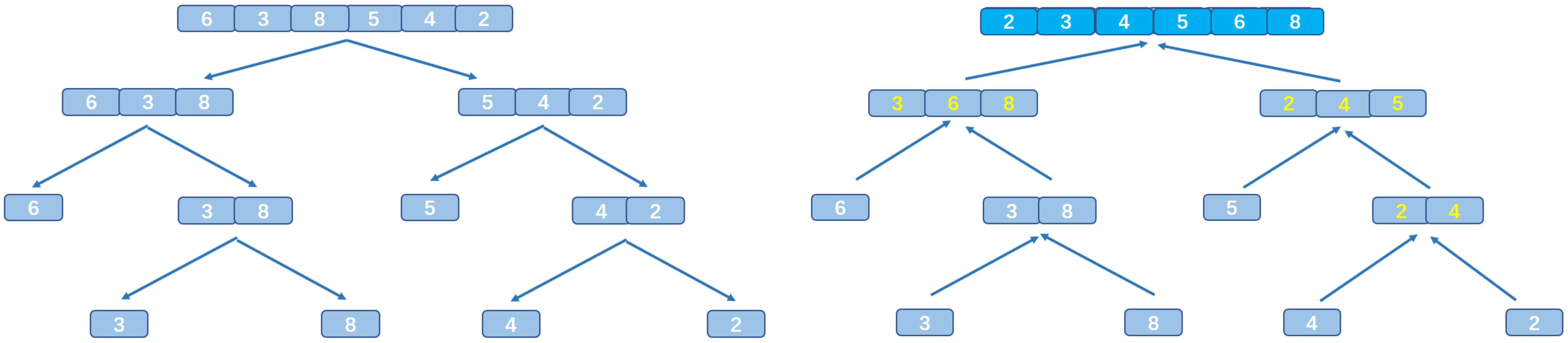
Data Scales	10000	20000	30000	40000	50000	60000	70000	80000	90000
Theoretical time/s	0.139	0.556	1.251	2.224	3.475	5.004	6.811	8.896	11.259
Experimental time/s	0.139	0.553	1.247	2.235	3.485	4.987	6.78	8.871	11.2
Error rate	0.00%	-0.54%	-0.32%	0.49%	0.29%	-0.34%	-0.46%	-0.28%	-0.52%



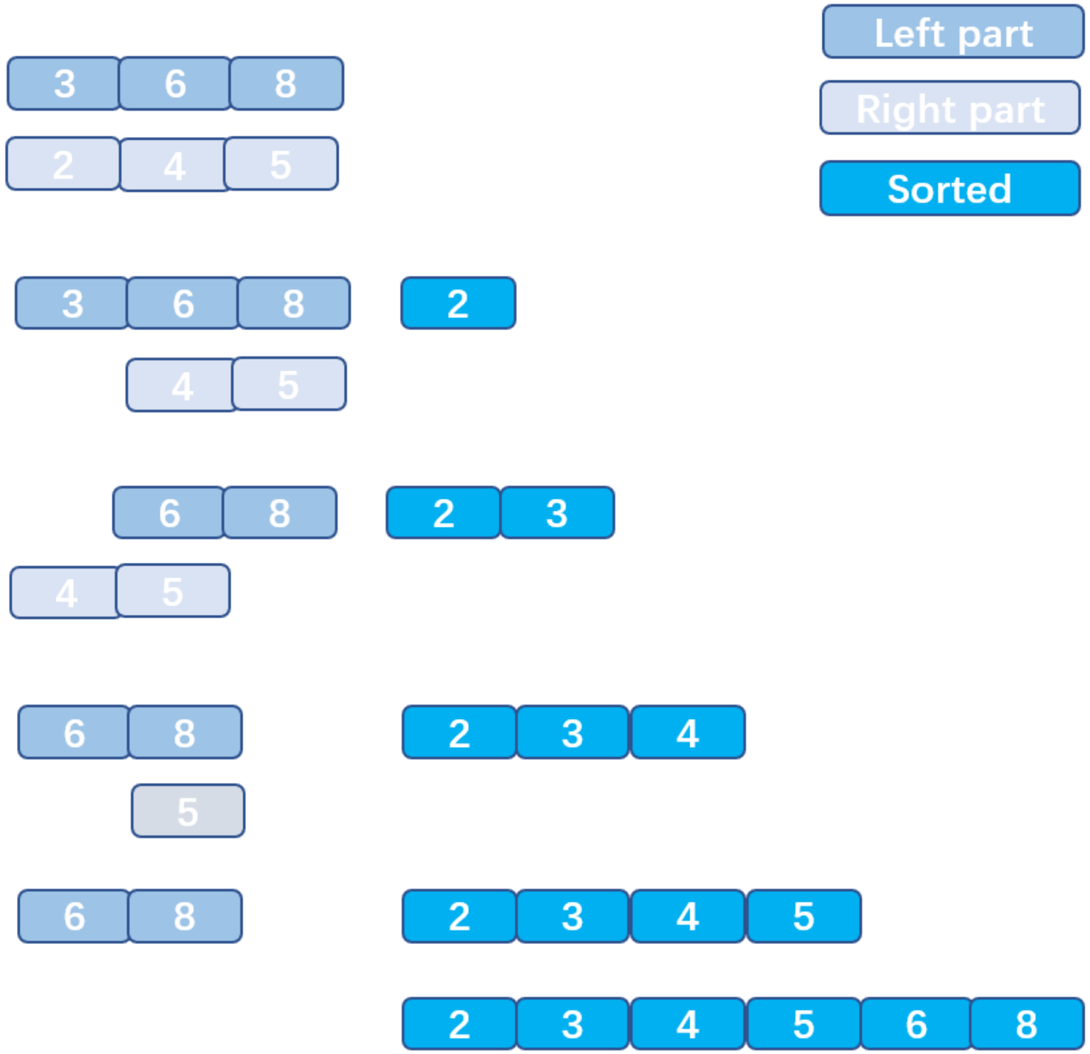
Performance Analysis

深圳大学 SHENZHEN UNIVERSITY

4、*Merge sort*



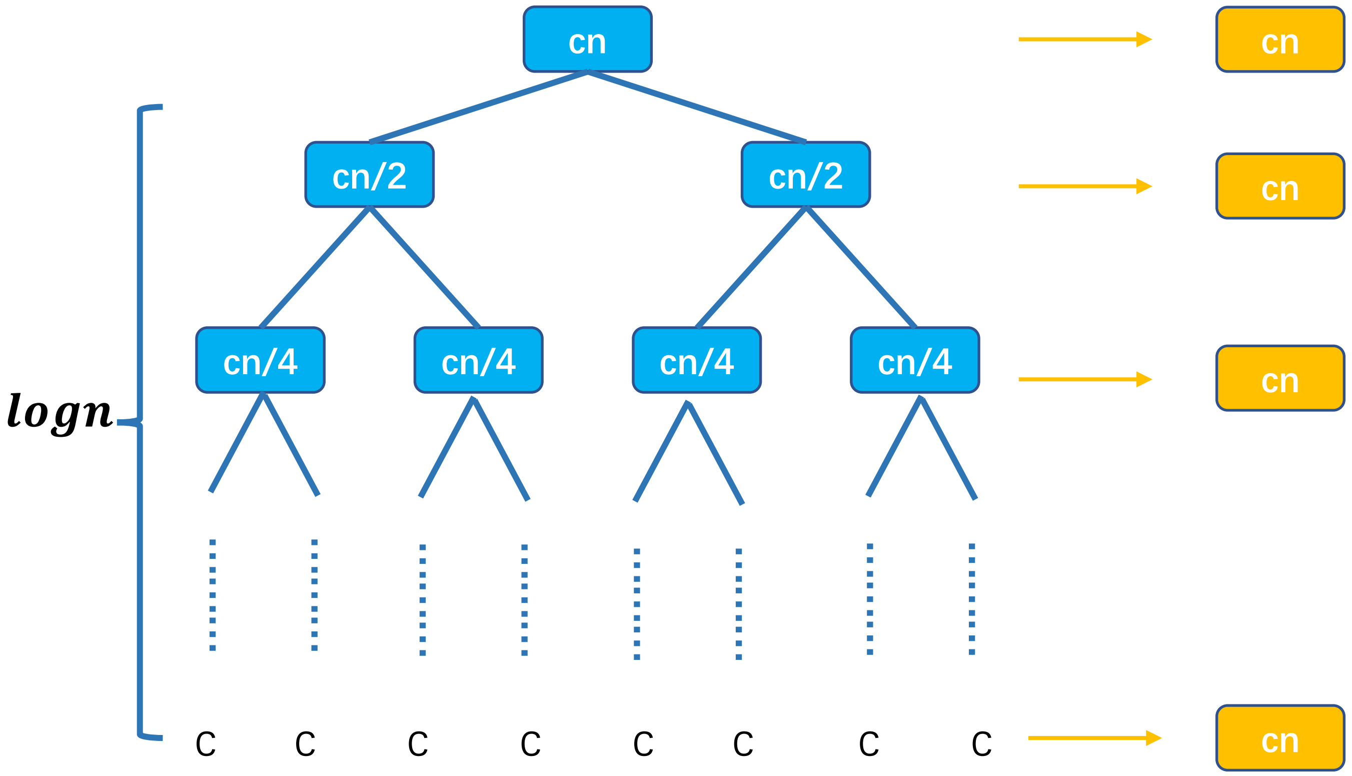
Performance Analysis



Performance Analysis

深圳大学 SHENZHEN UNIVERSITY

Time complexity:



$$T(n)=\begin{cases}\Theta(1), n=1\\ 2T(n/2)+\Theta(n), n>1\end{cases}$$

$$T(n)=\begin{cases}c, n=1\\ 2T(n/2)+cn, n>1\end{cases}$$

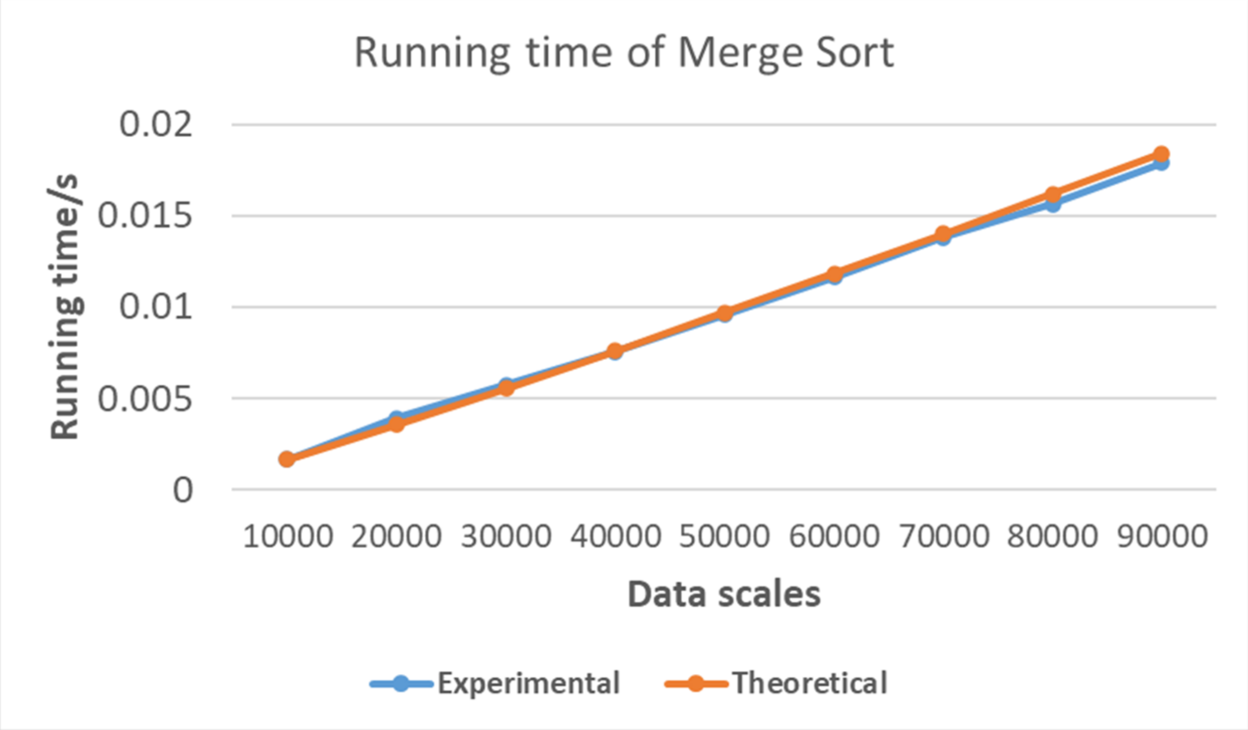
$$T(n)=cn(\log n+1)=cn\log n+cn$$

Performance Analysis

深圳大学
SHENZHEN UNIVERSITY

Testing:

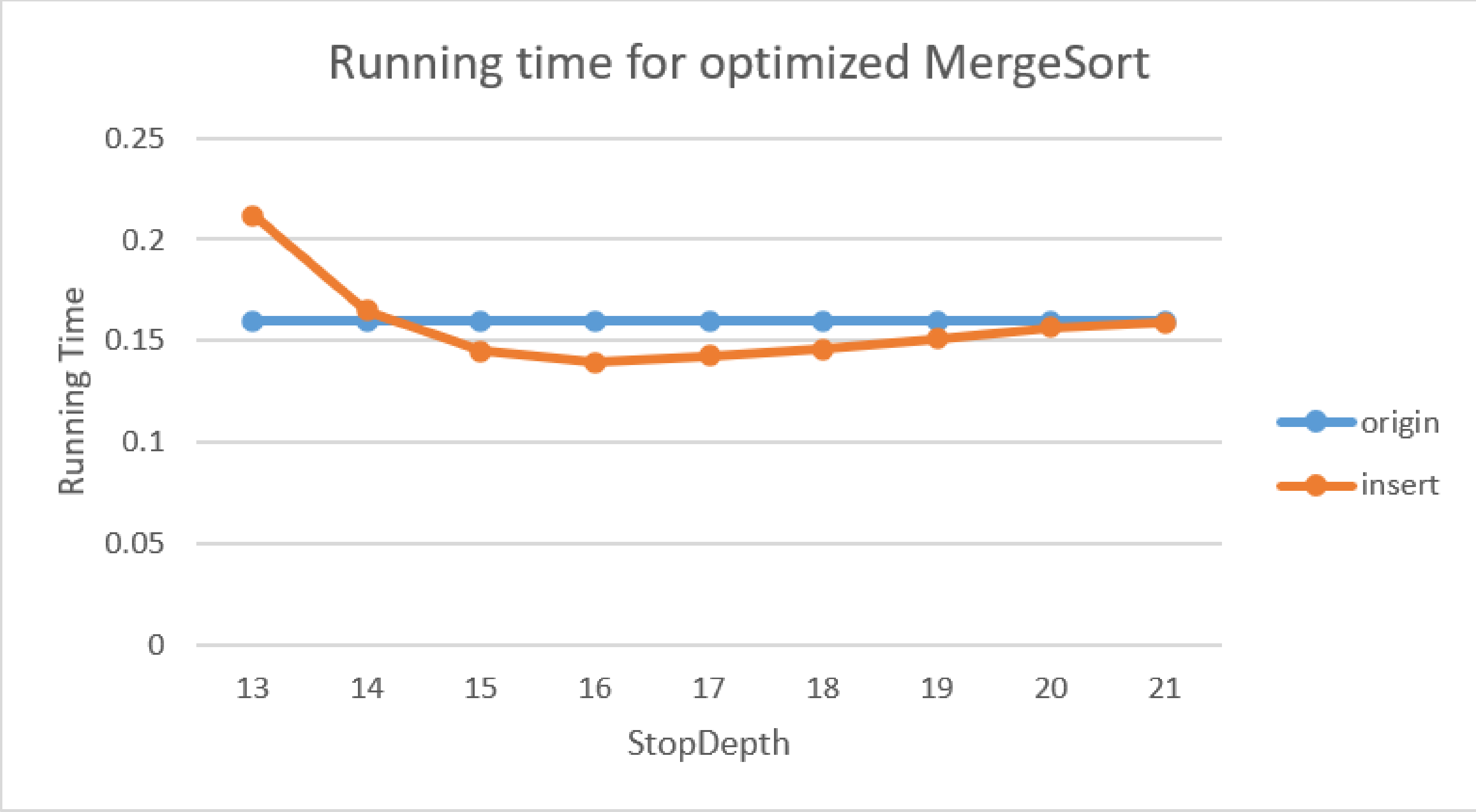
Data Scales	10000	20000	30000	40000	50000	60000	70000	80000	90000
Theoretical time/ms	1.652	3.55	5.55	7.6	9.7	11.84	14.01	16.2	18.41
Experimental time/ms	1.652	3.901	5.749	7.569	9.586	11.617	13.802	15.649	17.876
Error rate	0.00%	9.81%	3.64%	-0.44%	-1.21%	-1.89%	-1.46%	-3.40%	-2.93%



Performance Analysis

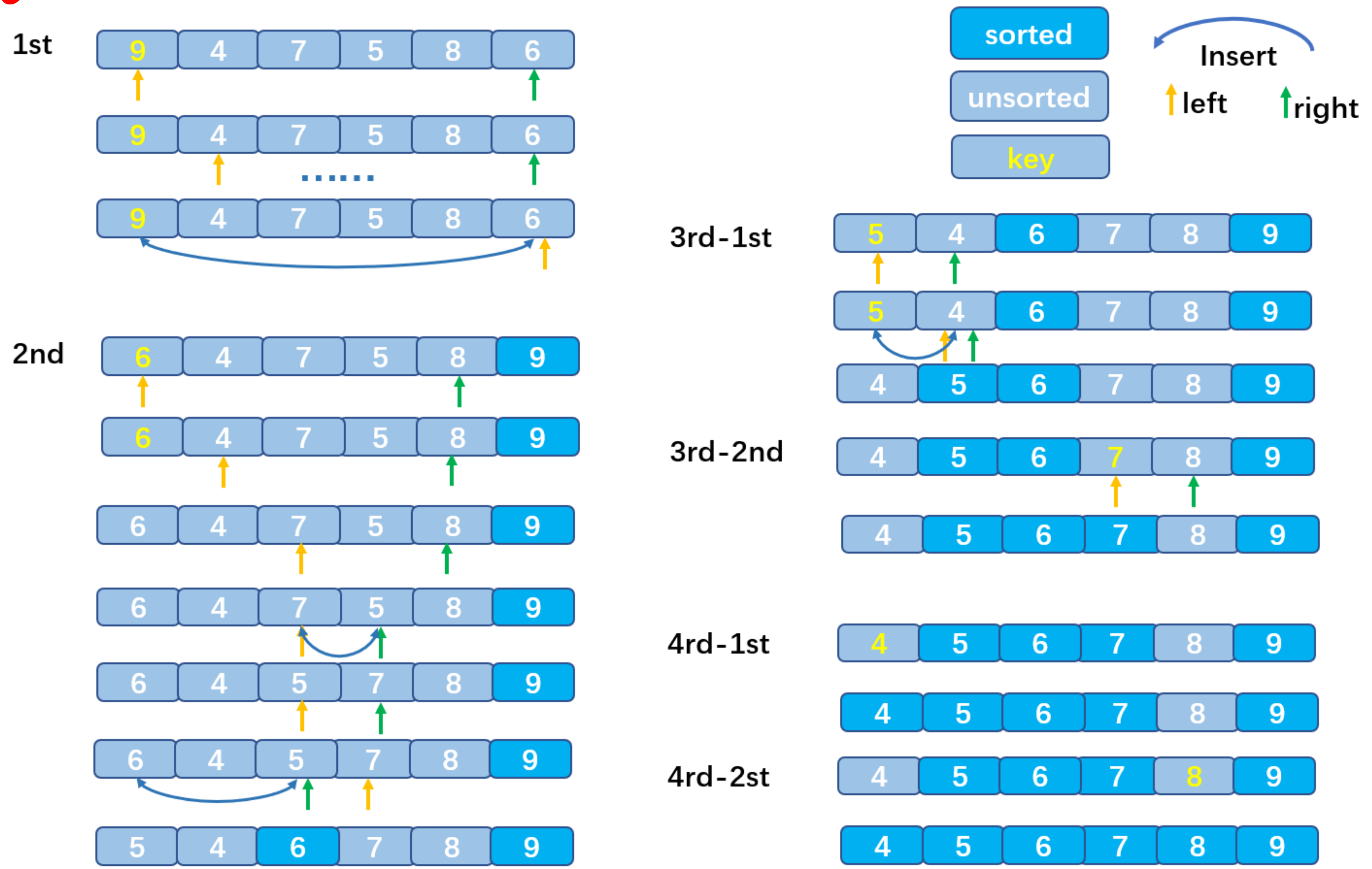
Testing: $n = 10000000$ $run\ 100\ times$

StopDepth	13	14	15	16	17	18	19	20	21
Running Time/s	0.21204	0.16498	0.14504	0.13945	0.14247	0.14592	0.15107	0.15672	0.15904



Performance Analysis

5、*Quick sort*



Performance Analysis

深圳大学 SHENZHEN UNIVERSITY

Time complexity:

$$T(n) = O(n + X)$$

$$X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}$$

$$E(X) = E\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} EX_{ij} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Pr_{ij}$$

$\boxed{PARTITION(A, l, r)}$

$key = A[l]$

$while(l < r)$

$\quad while(l < r\ and\ \boxed{A[r] \geq key})\ r = r - 1$

$\quad if(l < r)\ A[l] = A[r]\ l = l + 1$

$\quad while(l < r\ and\ \boxed{A[l] \leq key})\ l = l + 1$

$\quad if(l < r)\ A[r] = A[l]\ r = r - 1$

$A[l] = key$

$return\ l$

Performance Analysis

深圳大学 SHENZHEN UNIVERSITY

Time complexity:

z_i *the i−th smallest number*

$$Z_{ij} = \{z_i, z_{i+1}, \ldots, z_j\}$$

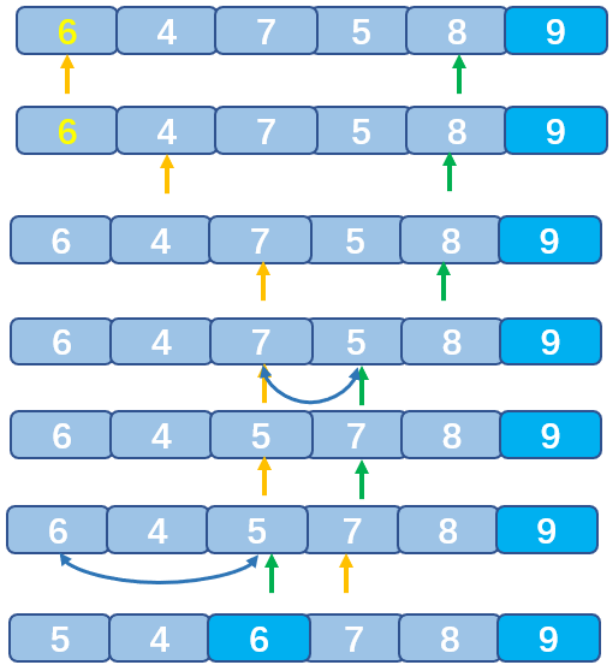
$$\Pr_{ij} = \frac{1}{j-i+1} + \frac{1}{j-i+1} = \frac{2}{j-i+1}$$

$$E(X) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \Pr_{ij} = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \frac{2}{j-i+1}$$

$$= \sum_{i=1}^{n-1}\sum_{k=1}^{n-i} \frac{2}{k+1} < \sum_{i=1}^{n-1}\sum_{k=1}^{n} \frac{2}{k} = \sum_{i=1}^{n-1} O(\lg n) = O(n \lg n)$$

$i = 6, j = 8$

$i = 4, j = 8$

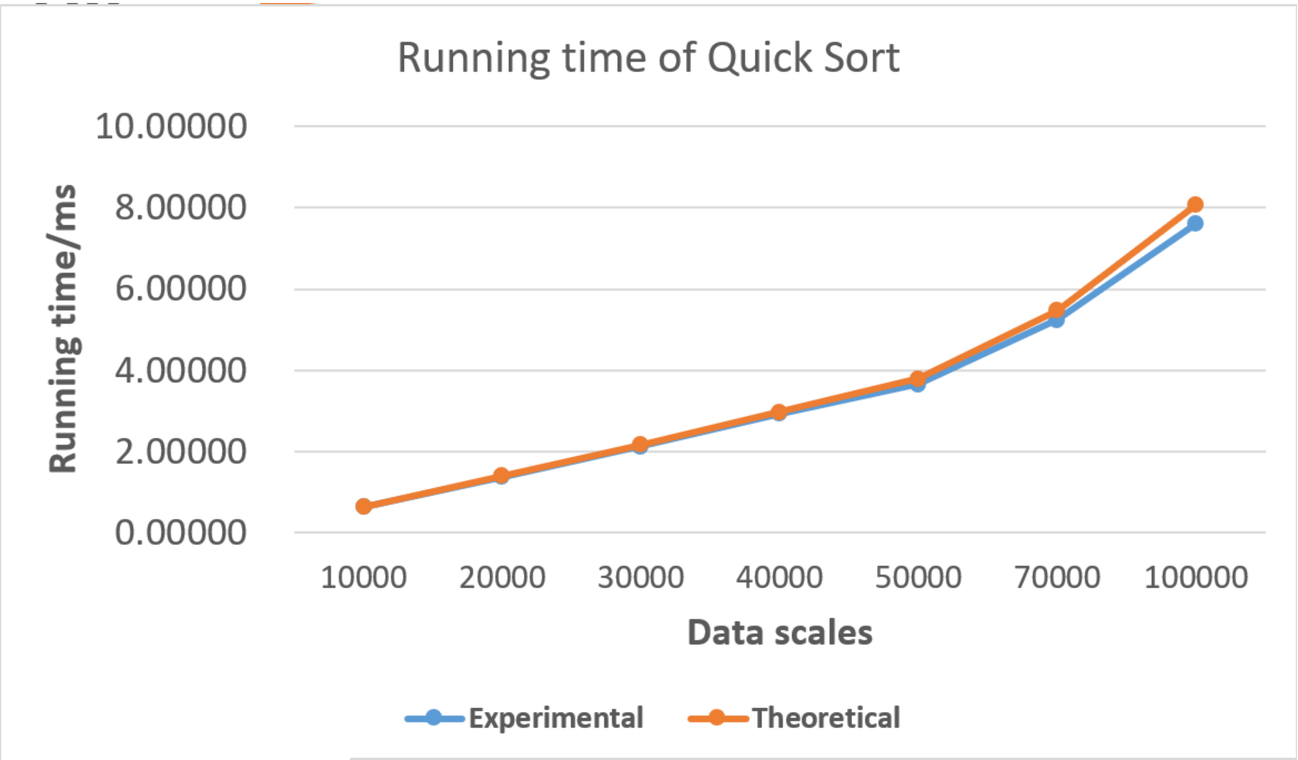


Performance Analysis

深圳大学 SHENZHEN UNIVERSITY

Testing:

Data Scales	10000	20000	30000	40000	50000	70000	100000
Theoretical time/ms	0.64522	1.38756	2.16655	2.96934	3.78984	5.47077	8.06525
Experimental time/ms	0.64522	1.37764	2.12376	2.93745	3.65455	5.23880	7.59374
Error rate	0.000%	-0.715%	-1.975%	-1.074%	-3.570%	-4.240%	-5.846%



02

PART 02

SECTION 2

The first m biggest number

Solutions

深圳大学 SHENZHEN UNIVERSITY

- *Bubble sort*
- *heap sort*
- *Quick sort*

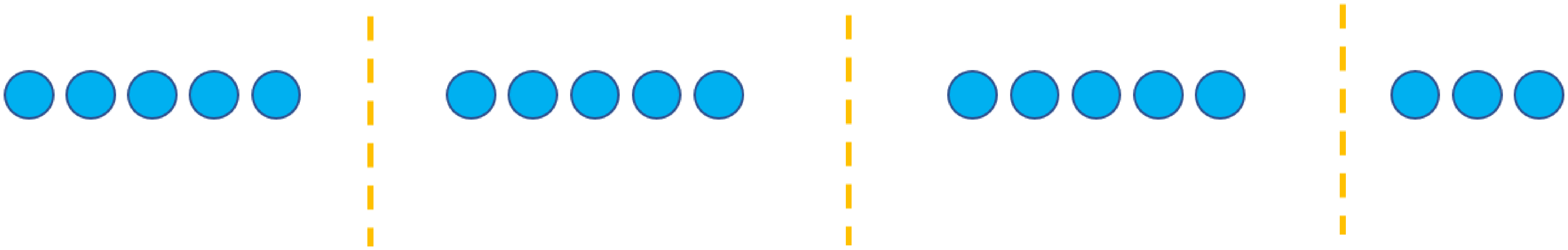
process

complexity

$$Select(a, l, r, k)$$

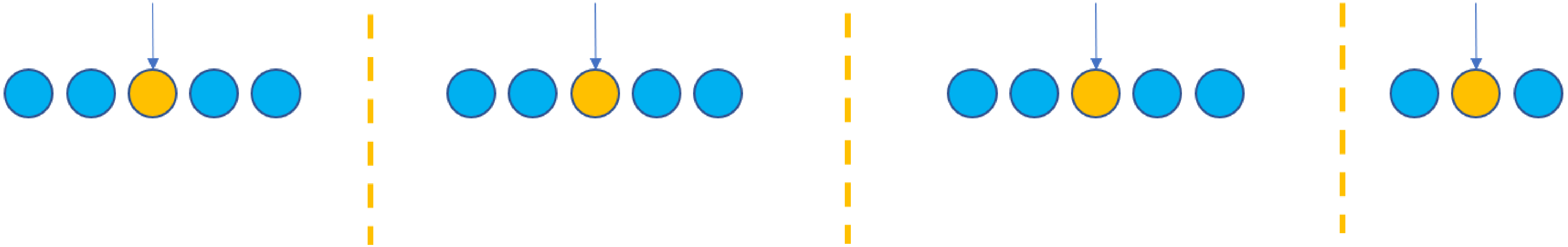
Process $\boldsymbol{Select(a,l,r,k)}$

- $Divide\ the\ sequence\ into\ \lceil n/5 \rceil\ groups, each\ group\ has\ 5\ elements$
  $(the\ last\ group\ may\ have\ fewer\ than\ five\ elements)$



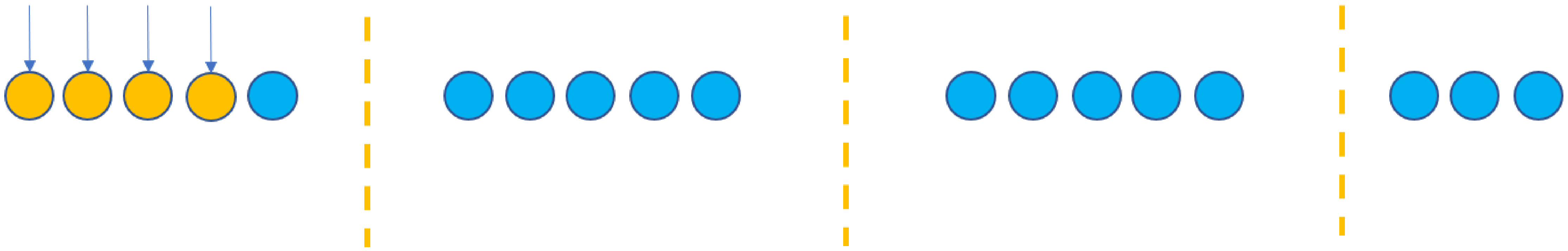
Process $Select(a, l, r, k)$

- $Find\ the\ median\ of\ each\ group\ using\ insertion\ sort$
  $then\ we\ have\ \lceil n/5 \rceil\ medians$



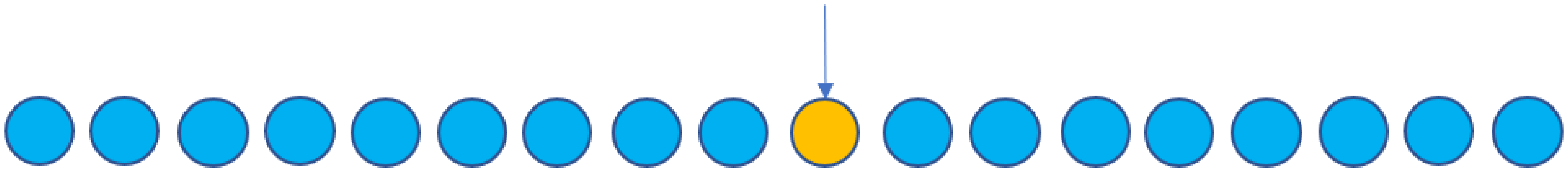
Process $\boldsymbol{Select(a, l, r, k)}$

- $Find\ the\ median\ of\ \lceil n/5 \rceil\ medians\ using\ \boldsymbol{Select(a, l, l + num - 1, \frac{num}{2})}$



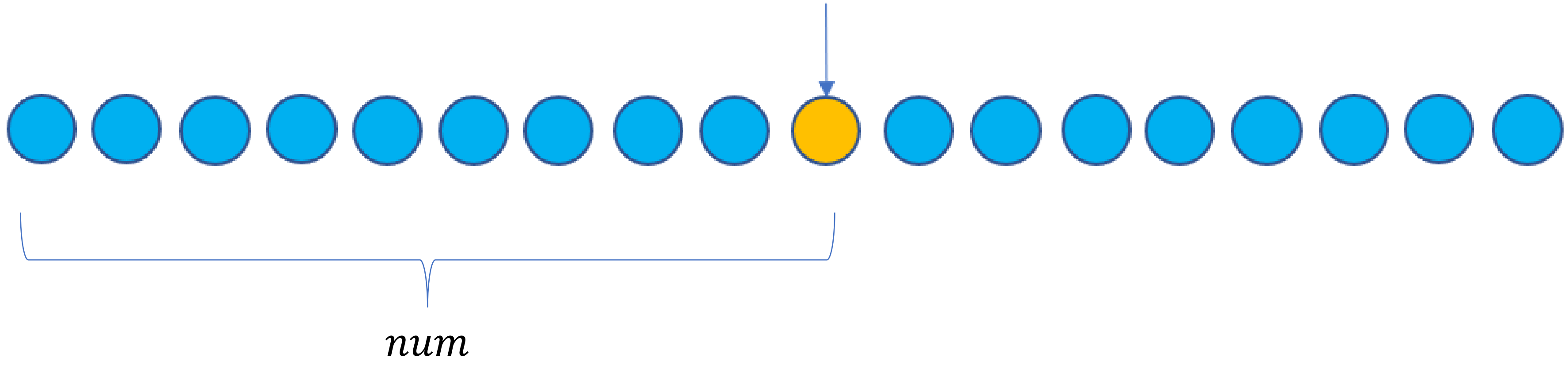
Process $\boldsymbol{Select(a, l, r, k)}$

- $Take\ the\ median\ we\ found\ as\ the\ pivot, using\ PATITION\ to\ divide$



Process $\boldsymbol{Select(a, l, r, k)}$

- $if\ k == num \qquad return\ l + num - 1$
- $if\ k < num \qquad return\ Select(a, l, l + num - 1, k)$
- $if\ k > num \qquad return\ Select(a, l + num, r, k - num)$



$T(n)$ **Complexity** $Select(a, l, r, k)$ 深圳大学 SHENZHEN UNIVERSITY

- $O(n)$ $Divide\ the\ sequence\ into\ \lceil n/5 \rceil\ groups, each\ group\ has\ 5\ elements$
  $(the\ last\ group\ may\ have\ fewer\ than\ five\ elements)$
- $O(n)$ $Find\ the\ median\ of\ each\ group\ using\ insertion\ sort$
  $then\ we\ have\ \lceil n/5 \rceil\ medians$
- $T(\lceil n/5 \rceil)$ $Find\ the\ median\ of\ \lceil n/5 \rceil\ medians\ using\ \boldsymbol{Select(a, l, l + num - 1, \frac{num}{2})}$
- $O(n)$ $Take\ the\ median\ we\ found\ as\ the\ pivot, using\ PATITION\ to\ divide$
- $??$ $if\ k == num \quad return\ l + num - 1$

$if\ k < num \quad return\ Select(a, l, l + num - 1, k)$

$if\ k > num \quad return\ Select(a, l + num, r, k - num)$

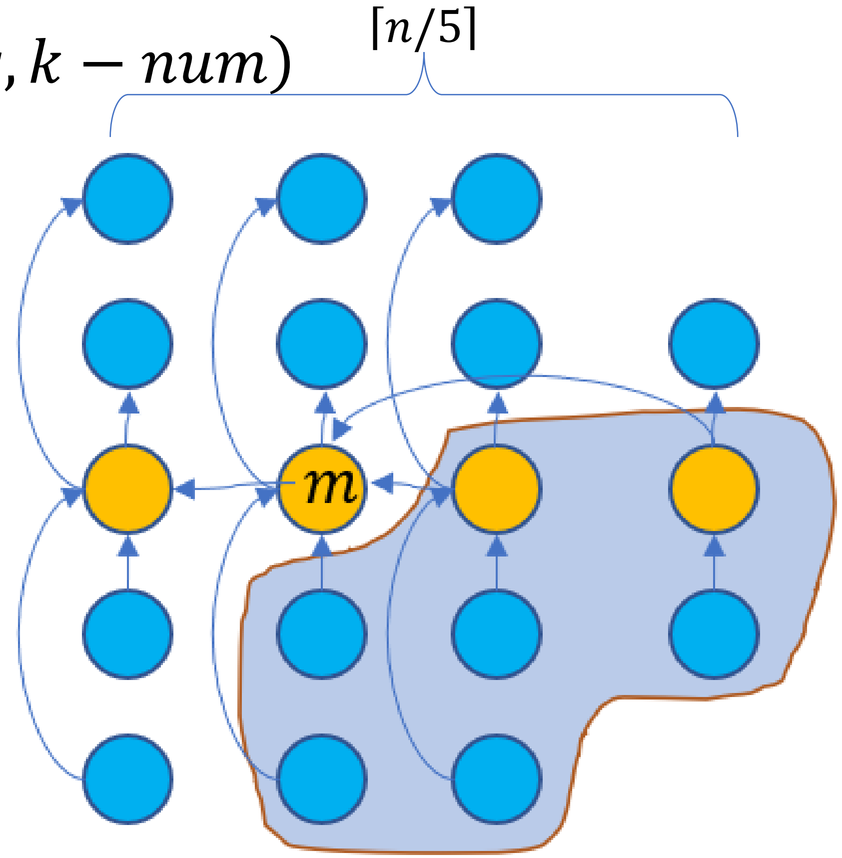
Complexity $Select(a,l,r,k)$

??

- $if\ k == num \quad return\ l + num - 1$

  $if\ k < num \quad return\ Select(a, l, l + num - 1, k)$

  $if\ k > num \quad return\ Select(a, l + num, r, k - num)$



$$big \geq 3\left(\left\lceil \frac{1}{2}\left\lceil \frac{n}{5} \right\rceil \right\rceil - 2\right) \geq \frac{3n}{10} - 6$$

$$small \leq n - big = \frac{7n}{10} + 6$$

$$T\left(\frac{7n}{10} + 6\right)$$

$T(n)$ **Complexity** $\boldsymbol{Select(a,l,r,k)}$

深圳大学 SHENZHEN UNIVERSITY

$\boldsymbol{O(n)}$ • $Divide\ the\ sequence\ into\ \lceil n/5 \rceil\ groups, each\ group\ has\ 5\ elements$
$(the\ last\ group\ may\ have\ fewer\ than\ five\ elements)$

$\boldsymbol{O(n)}$ • $Find\ the\ median\ of\ each\ group\ using\ insertion\ sort$
$then\ we\ have\ \lceil n/5 \rceil\ medians$

$\boldsymbol{T(\lceil n/5 \rceil)}$ • $Find\ the\ median\ of\ \lceil n/5 \rceil\ medians\ using\ \boldsymbol{Select(a, l, l + num\ - 1, \frac{num}{2})}$

$\boldsymbol{O(n)}$ • $Take\ the\ median\ we\ found\ as\ the\ pivot, using\ PATITION\ to\ divide$

$\boldsymbol{T\left(\frac{7n}{10} + 6\right)}$ • $if\ k == num \quad return\ l + num\ - 1$

$if\ k < num \quad return\ Select(a, l, l + num - 1, k)$

$if\ k > num \quad return\ Select(a, l + num, r, k - num)$

Complexity $Select(a, l, r, k)$

深圳大学 SHENZHEN UNIVERSITY

$$T(n) \le O(n) + O(n) + T\left(\left\lceil \frac{n}{5} \right\rceil\right) + O(n) + T\left(\frac{7n}{10} + 6\right)$$

$$= O(n) + T\left(\left\lceil \frac{n}{5} \right\rceil\right) + T(\frac{7n}{10} + 6)$$

$$T(n) \le cn$$

$$T(n) \le an + c\left\lceil \frac{n}{5} \right\rceil + c\left(\frac{7n}{10} + 6\right) \le an + \frac{cn}{5} + c + c\frac{7n}{10} + 6c$$

$$= c\frac{9n}{10} + 7c + an = cn + (-\frac{cn}{10} + 7c + an)$$

$$\left(-\frac{cn}{10} + 7c + an\right) \le 0$$

$$c(n - 70) \ge 10an \qquad n > 70 \;\; c \ge \frac{10an}{n-70}$$

$$n \le 70 \;\; T(n) = O(1) \le cn$$

Thank You!   Questions?

深圳大学计算机与软件学院

报告人：郑杨 陈敏涵
指导老师：李炎然

2022年 3 月 14 日