# Advantages of Shared Data Structures for Sequences of Balanced Parentheses

Simon Gog[1]     Johannes Fischer[2]

[1]Institut of Theoretical Computer Science
Ulm University, Germany
[2]Center for Bioinformatics (ZBIT)
Universität Tübingen, Germany

March 26, 2010

# Outline

- Basic Definitions
- Our DS for RMQ
- Geary et al.'s DS for balanced parentheses
- Our result
    - Computing RMQs with $2n + o(n)$ bits
    - Computing LCA with $2n + o(n)$ bits
- Experimental study
    - Comparison of RMQ data structures
    - Comparison of CST implementations
- Conclusion

# Range Minimum Queries

## Definition

Given an array $A$ of $n$ values. A range minimum query (RMQ) $rmq_A(i,j)$ with $i \leq j$ returns index $k$ and $A[k] = \min\{A[\ell] | i \leq \ell \leq j\}$.

# Range Minimum Queries

## Definition

Given an array $A$ of $n$ values. A range minimum query (RMQ) $rmq_A(i,j)$ with $i \leq j$ returns index $k$ and $A[k] = \min\{A[\ell] | i \leq \ell \leq j\}$.

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

$rmq_A(1,5) =$

# Range Minimum Queries

## Definition

Given an array $A$ of $n$ values. A range minimum query (RMQ) $rmq_A(i,j)$ with $i \leq j$ returns index $k$ and $A[k] = \min\{A[\ell] | i \leq \ell \leq j\}$.

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

$rmq_A(1,5) = 2$
$rmq_A(7,10) =$

# Range Minimum Queries

## Definition

Given an array $A$ of $n$ values. A range minimum query (RMQ) $rmq_A(i,j)$ with $i \leq j$ returns index $k$ and $A[k] = \min\{A[\ell] | i \leq \ell \leq j\}$.

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

$rmq_A(1,5) = 2$
$rmq_A(7,10) = 10$

# Range Minimum Queries

## Definition

Given an array $A$ of $n$ values. A range minimum query (RMQ) $rmq_A(i, j)$ with $i \leq j$ returns index $k$ and $A[k] = \min\{A[\ell] | i \leq \ell \leq j\}$.

## Solution

- Preprocess a RMQ data structure $R$ for $A$
- $R$ answers a RMQ then in constant time
- Two versions of the problem
  - Systematic: $R$ needs $A$ to answer RMQs
  - Non-systematic: $R$ answers RMQs

# Range Minimum Queries

## Definition

Given an array $A$ of $n$ values. A range minimum query (RMQ) $rmq_A(i,j)$ with $i \leq j$ returns index $k$ and $A[k] = \min\{A[\ell] | i \leq \ell \leq j\}$.

## Solution

- Preprocess a RMQ data structure $R$ for $A$
- $R$ answers a RMQ then in constant time
- Two versions of the problem
  - Systematic: $R$ needs $A$ to answer RMQs
  - Non-systematic: $R$ answers RMQs

## Our solution

- Non-systematic
- $2n + o(n)$ bits ($3n$ bits in practice)
- $3n + o(n)$ bits for construction in linear time

# Balanced Parentheses Sequences (BPS)

- Sequence S over the alphabet '(' and ')'
- Each prefix of S contains more '('s than ')'s
- Fundamental operations on S:
  - $rank_{(}(S, i)$
  - $select_{(}(S, i)$
  - $excess(S, i) = rank_{(}(i) - rank_{)}(i)$
  - $find\_close(S, i)$ and $find\_open(S, i)$
  - $enclose(S, i)$

**Example**

(()(()(()()()()()()))(()))

# Balanced Parentheses Sequences (BPS)

- Sequence S over the alphabet ’(’ and ’)’
- Each prefix of S contains more ’(’s than ’)’s
- Fundamental operations on S:
  - $rank_{(}(S, i)$
  - $select_{(}(S, i)$
  - $excess(S, i) = rank_{(}(i) - rank_{)}(i)$
  - $find\_close(S, i)$ and $find\_open(S, i)$
  - $enclose(S, i)$

**Example**

(()((()(()()()()()()))(())

$rank_{(}(S, 5) = 4$

# Balanced Parentheses Sequences (BPS)

- Sequence S over the alphabet '(' and ')'
- Each prefix of S contains more '('s than ')'s
- Fundamental operations on S:
  - $rank_{(}(S, i)$
  - $select_{(}(S, i)$
  - $excess(S, i) = rank_{(}(i) - rank_{)}(i)$
  - $find\_close(S, i)$ and $find\_open(S, i)$
  - $enclose(S, i)$

**Example**

(()(()(()()()()()()))(())

$select_{(}(S, 2) = 1$

# Balanced Parentheses Sequences (BPS)

- Sequence S over the alphabet '(' and ')'
- Each prefix of S contains more '('s than ')'s
- Fundamental operations on S:
  - $rank_{(}(S, i)$
  - $select_{(}(S, i)$
  - $excess(S, i) = rank_{(}(i) - rank_{)}(i)$
  - $find\_close(S, i)$ and $find\_open(S, i)$
  - $enclose(S, i)$

**Example**

(()((()(()()()()()()))(()))

$excess_{(}(S, 5) = 3$

# Balanced Parentheses Sequences (BPS)

- Sequence S over the alphabet '(' and ')'
- Each prefix of S contains more '('s than ')'s
- Fundamental operations on S:
  - $rank_{(}(S, i)$
  - $select_{(}(S, i)$
  - $excess(S, i) = rank_{(}(i) - rank_{)}(i)$
  - $find\_close(S, i)$ and $find\_open(S, i)$
  - $enclose(S, i)$

**Example**

(()(()(()()()()()()))(())

$find\_close(S, 3) = 20$ and $find\_open(S, 20) = 3$

# Balanced Parentheses Sequences (BPS)

- Sequence S over the alphabet '(' and ')'
- Each prefix of S contains more '('s than ')'s
- Fundamental operations on S:
  - $rank_{(}(S, i)$
  - $select_{(}(S, i)$
  - $excess(S, i) = rank_{(}(i) - rank_{)}(i)$
  - $find\_close(S, i)$ and $find\_open(S, i)$
  - $enclose(S, i)$

## Example

(()(()(()()()()()()))(()))
$enclose(S, 4) = 3$

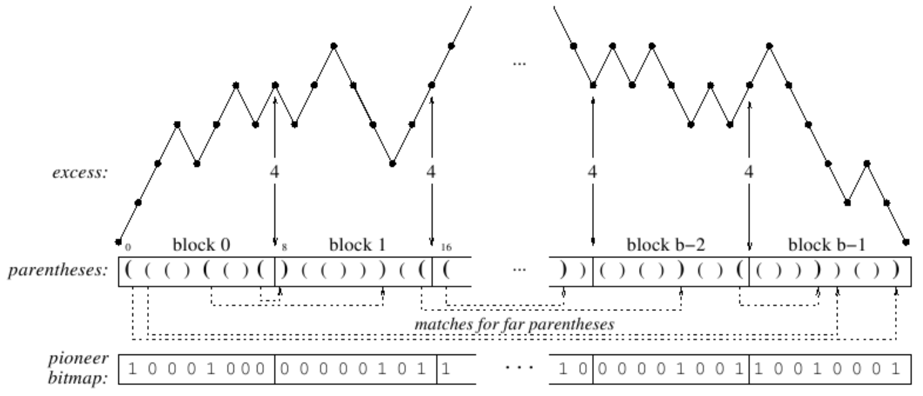
# Balanced Parentheses Sequences (BPS)

- Sequence S over the alphabet '(' and ')'
- Each prefix of S contains more '('s than ')'s
- Fundamental operations on S:
  - $rank_((S, i)$
  - $select_((S, i)$
  - $excess(S, i) = rank_((i) - rank_)(i)$
  - $find\_close(S, i)$ and $find\_open(S, i)$
  - $enclose(S, i)$

**Time and space**

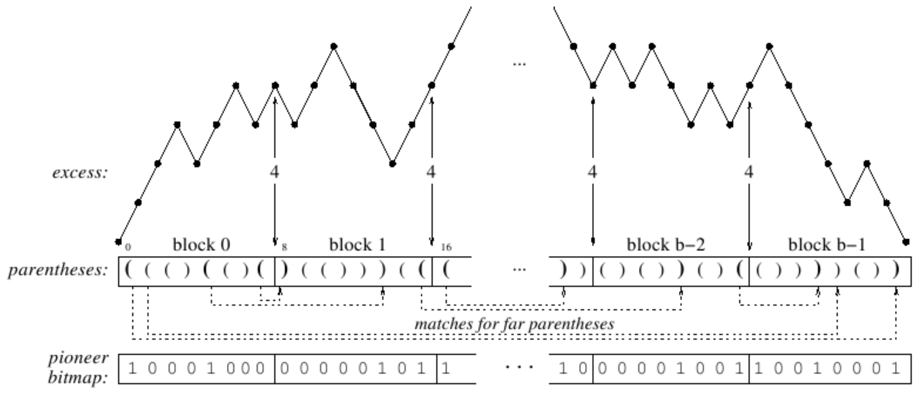
Geary et al.'s data structure of size $o(n)$ supports all operations in constant time

# Geary et al.'s support structure for BP




- Partition BPS into $b$ blocks of length $\mathcal{O}(\log n)$
- Calculate far parentheses/pioneers to answer *find_close*, *find_open*, *enclose*
- pioneer bitmap takes $\mathcal{O}(\frac{n \log \log n}{\log n})$ bits

# Geary et al.'s support structure for BP




- set of pioneers is again a BPS (of length $n_1 < 4b - 6$ )
- recusively build data structure for pioneers
- store answers explicitly on the second level

# The range restricted enclose method

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

(
-1

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| ( | ( |
|---|---|
| -1 | 2 |

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| ( | ( | ) |
|---|---|---|
| -1 | 2 | 2 |

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| | | | |
|---|---|---|---|
| ( | ( | ) | ( |
| -1 | 2 | 2 | 1 |

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

( ( ) ( (
-1 2 2 1 3

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| | | | | | |
|---|---|---|---|---|---|
| ( | ( | ) | ( | ( | ) |
| -1 | 2 | 2 | 1 | 3 | 3 |

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

$\underset{-1}{(}\ \underset{2}{(}\ \underset{2}{)}\ \underset{1}{(}\ \underset{3}{(}\ \underset{3}{)}\ \underset{1}{(}$

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ( | ( | ) | ( | ( | ) | ( | ( |
| -1 | 2 | 2 | 1 | 3 | 3 | 1 | 2 |

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ( | ( | ) | ( | ( | ) | ( | ( | ) |
| -1 | 2 | 2 | 1 | 3 | 3 | 1 | 2 | 2 |

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ( | ( | ) | ( | ( | ) | ( | ( | ) | ) |
| -1 | 2 | 2 | 1 | 3 | 3 | 1 | 2 | 2 | 1 |

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ( | ( | ) | ( | ( | ) | ( | ( | ) | ) | ) |
| -1 | 2 | 2 | 1 | 3 | 3 | 1 | 2 | 2 | 1 | 1 |

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| ( | ( | ) | ( | ( | ) | ( | ( | ) | ) | ) | ( |
|---|---|---|---|---|---|---|---|---|---|---|---|
| -1 | 2 | 2 | 1 | 3 | 3 | 1 | 2 | 2 | 1 | 1 | 0 |

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| ( | ( | ) | ( | ( | ) | ( | ( | ) | ) | ) | ( | ( |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -1 | 2 | 2 | 1 | 3 | 3 | 1 | 2 | 2 | 1 | 1 | 0 | 2 |

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ( | ( | ) | ( | ( | ) | ( | ( | ) | ) | ) | ( | ( | ) |
| -1 | 2 | 2 | 1 | 3 | 3 | 1 | 2 | 2 | 1 | 1 | 0 | 2 | 2 |

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ( | ( | ) | ( | ( | ) | ( | ( | ) | ) | ) | ( | ( | ) | ( |
| -1 | 2 | 2 | 1 | 3 | 3 | 1 | 2 | 2 | 1 | 1 | 0 | 2 | 2 | 0 |

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ( | ( | ) | ( | ( | ) | ( | ( | ) | ) | ) | ( | ( | ) | ( | ( |
| -1 | 2 | 2 | 1 | 3 | 3 | 1 | 2 | 2 | 1 | 1 | 0 | 2 | 2 | 0 | 1 |

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ( | ( | ) | ( | ( | ) | ( | ( | ) | ) | ) | ( | ( | ) | ( | ( | ) |
| -1 | 2 | 2 | 1 | 3 | 3 | 1 | 2 | 2 | 1 | 1 | 0 | 2 | 2 | 0 | 1 | 1 |

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| ( | ( | ) | ( | ( | ) | ( | ( | ) | ) | ) | ( | ( | ) | ( | ( | ) | ) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -1 | 2 | 2 | 1 | 3 | 3 | 1 | 2 | 2 | 1 | 1 | 0 | 2 | 2 | 0 | 1 | 1 | 0 |

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| ( | ( | ) | ( | ( | ) | ( | ( | ) | ) | ) | ( | ( | ) | ( | ( | ) | ) | ( |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -1 | 2 | 2 | 1 | 3 | 3 | 1 | 2 | 2 | 1 | 1 | 0 | 2 | 2 | 0 | 1 | 1 | 0 | -1 |

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ( | ( | ) | ( | ( | ) | ( | ( | ) | ) | ) | ( | ( | ) | ( | ( | ) | ) | ( | ) |
| -1 | 2 | 2 | 1 | 3 | 3 | 1 | 2 | 2 | 1 | 1 | 0 | 2 | 2 | 0 | 1 | 1 | 0 | -1 | -1 |

# Construction of the BP of the SCT

## Example

| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ( | ( | ) | ( | ( | ) | ( | ( | ) | ) | ) | ( | ( | ) | ( | ( | ) | ) | ( | ) | ) |
| -1 | 2 | 2 | 1 | 3 | 3 | 1 | 2 | 2 | 1 | 1 | 0 | 2 | 2 | 0 | 1 | 1 | 0 | -1 | -1 | -1 |

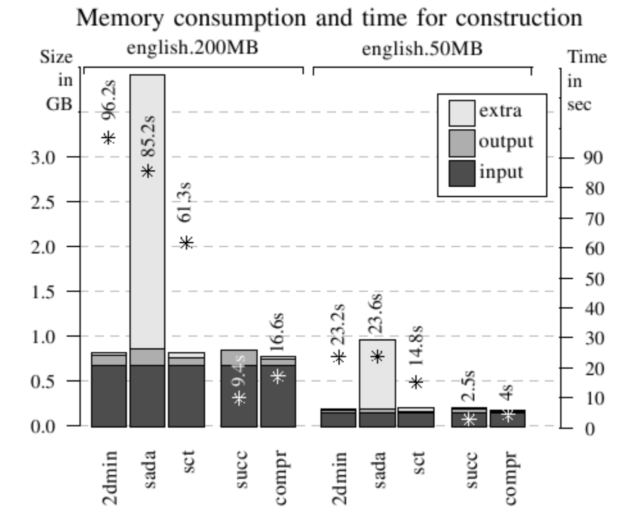
# Construction of the BP of the SCT

## Example

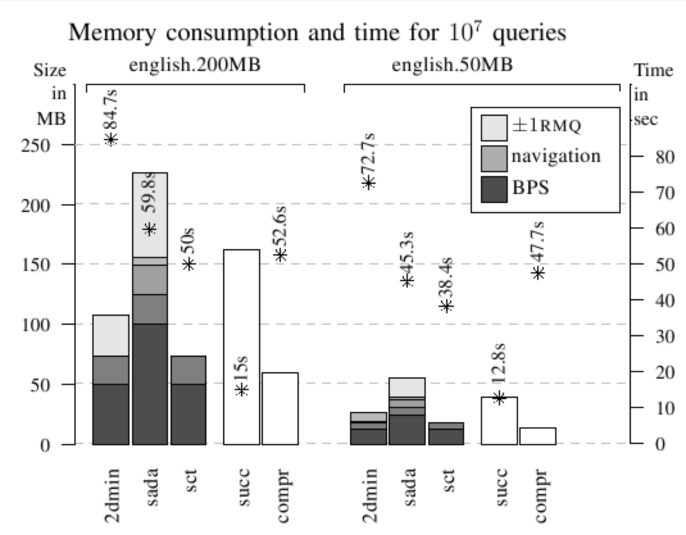
| i | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A[i] | -1 | 2 | 1 | 3 | 1 | 2 | 0 | 2 | 0 | 1 | -1 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ( | ( | ) | ( | ( | ) | ( | ( | ) | ) | ) | ( | ( | ) | ( | ( | ) | ) | ( | ) | ) |
| -1 | 2 | 2 | 1 | 3 | 3 | 1 | 2 | 2 | 1 | 1 | 0 | 2 | 2 | 0 | 1 | 1 | 0 | -1 | -1 | -1 |
| 1 | 2 | | 3 | 4 | | 5 | 6 | | | | 7 | 8 | | 9 | 10 | | | 11 | | |

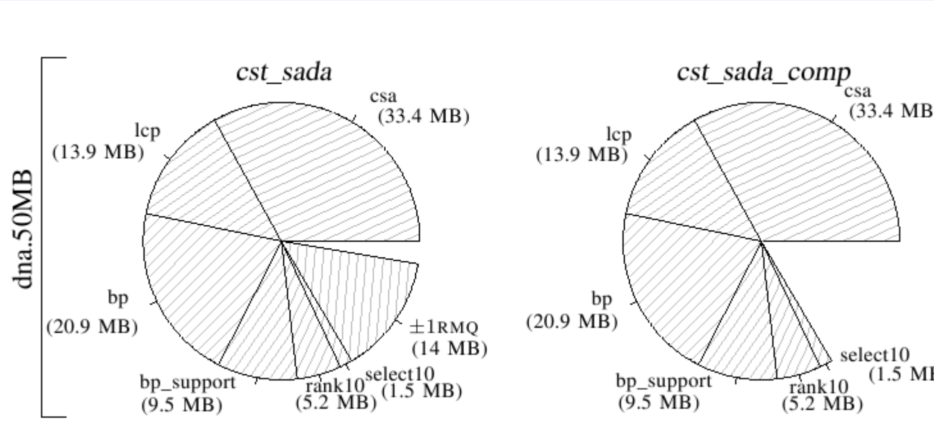
# RMQ: Peak memory consumption at construction



# RMQ: Final memory consumption and query time



# Compressed Suffix Trees: Memory



# Other data structures for RMQs

## Non-systematic solutions

- sada: BPS of the extended Caresian Tree ($4n + o(n)$ bits) by Sadakane (JDA 2007)
- 2dmin: BPS of the 2d-Min-Heap ($2n + o(n)$) by Fischer (2009)

## Systematic solutions

- succ: Succinct solution ($7n$ bits+size of input array) by Fischer ()
- compr: Compressed solution ($\approx 3n$ bits + size of input array) by Fischer et al. (DCC 2008)

# Experimental results

Any Questions?