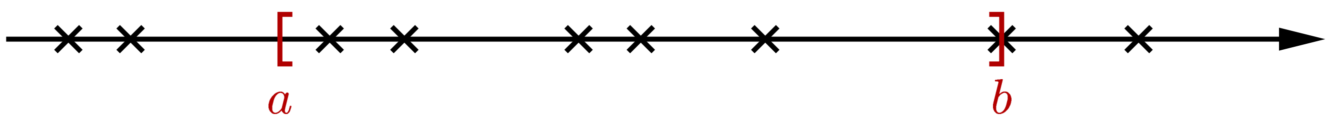
Optimal Static Range Reporting in One Dimension

Gerth Stølting Brodal

Joint work with Stephen Alstrup and Theis Rauhe

Range Queries in One Dimension

Given a static set S of n elements from a totally ordered universe U, preprocess S to support range queries online



Member(a)	$a \in S$?
Pred(a)	$\max\{x \in S \mid x \leq a\}$ or $\perp$ if $\{x \in S \mid x \leq a\} = \emptyset$
FindAny(a,b)	any $e \in S \cap [a,b]$ or $\perp$ if $S \cap [a,b] = \emptyset$
Report(a,b)	$S \cap [a,b]$
Count(a,b)	$\lvert S \cap [a,b] \rvert$
Count$_\varepsilon(a,b)$	Count$(a,b) \leq$ Count$_\varepsilon(a,b) \leq (1+\varepsilon) \cdot$ Count(a,b)

Comparison Model

Theorem

Member requires $\Omega(\log n)$ comparisons

Corollary

Pred, FindAny, Report, Count, Count_ε require $\Omega(\log n)$ comparisons

Proof

$$\begin{aligned}\text{Member}(a) &\equiv \text{Pred}(a) = a \\ &\equiv \text{FindAny}(a,a) = a \\ &\equiv \text{Report}(a,a) = \{a\} \\ &\equiv \text{Count}(a,a) = 1 \\ &\equiv \text{Count}_\varepsilon(a,a) \geq 1\end{aligned}$$

□

RAM Model

- Unit cost RAM with word size w
- Instructions: $+$, $*$, $\neg$, $\vee$, shifting, ...
- $U = \{0, 1, \ldots, 2^w - 1\}$
- Random bits

Theorem

Supporting Member requires space $\Omega(n)$ words, for $n \leq 2^{(1-\varepsilon)w}$

Theorem

Member, Pred, FindAny, Count, Count$_\varepsilon$ can be supported in constant time and Report in time O$(1 + |\text{output}|)$ using space O(2^w) words

Pred	0	1	2	3	3	3	3	7	7	9	10	10	10	13	13	13	16	16	16	16	16	21	22	22	22	22	22	27	28	29	29	31
Rank	0	1	2	3	0	0	0	4	0	5	6	0	0	7	0	0	8	0	0	0	0	9	10	0	0	0	0	11	12	13	0	14
S		X	X	X				X		X	X			X			X					X	X					X	X	X		X
	0	1	2	3							10										20										30	31

Previous Results for the RAM

	Query time	Space
Member		
Fredman *et al.* 1984 †	1	n
Pred, Count		
Willard 1983	$\log w$	n
Beame, Fich 1999 ⋆	$\min\left(\frac{\log w}{\log\log w}, \sqrt{\frac{\log n}{\log\log n}}\right)$	$n^{1+\delta}$
Andersson, Thorup 2000	$\min\left(\frac{\log w \cdot \log\log n}{\log\log w}, \sqrt{\frac{\log n}{\log\log n}}\right)$	n
FindAny		
Miltersen *et al.* 1998	1	$n \cdot w$
Report		
Miltersen *et al.* 1998	$1 + \lvert\text{output}\rvert$	$n \cdot w$

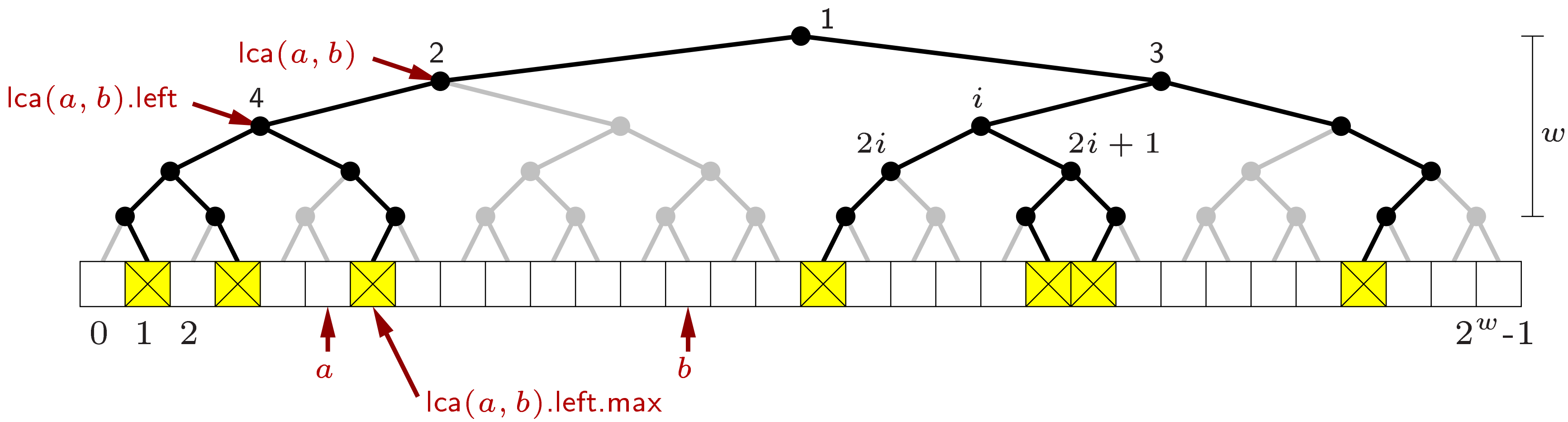
† "FKS" ⋆ Optimal query time

Our Results for the RAM

	Query time	Space
FindAny	1	n
Report	$1 + \lvert\text{output}\rvert$	n
Count_ε	$\log \frac{1}{\varepsilon}$	n

Preprocessing takes expected time $O(n\sqrt{w})$

The Basic Idea



$$\text{FindAny}(a,b) \in [a,b] \cap \{\text{lca}(a,b).\text{left.max}\,,\ \text{lca}(a,b).\text{right.min}\}$$

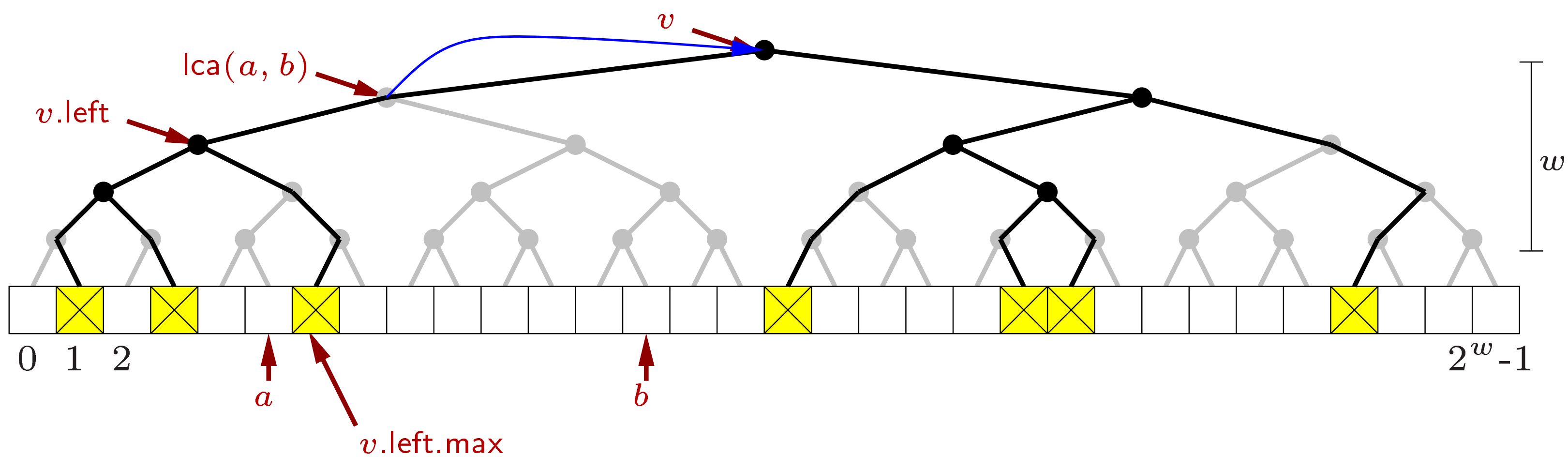
where

$$\text{lca}(a,b) = (2^{w-1} + \lfloor a/2 \rfloor) \downarrow \text{msb}(a \oplus b)$$

Lemma

FKS on ● nodes ⇒ space $O(n \cdot w)$ words and FindAny in constant time

The Second Idea



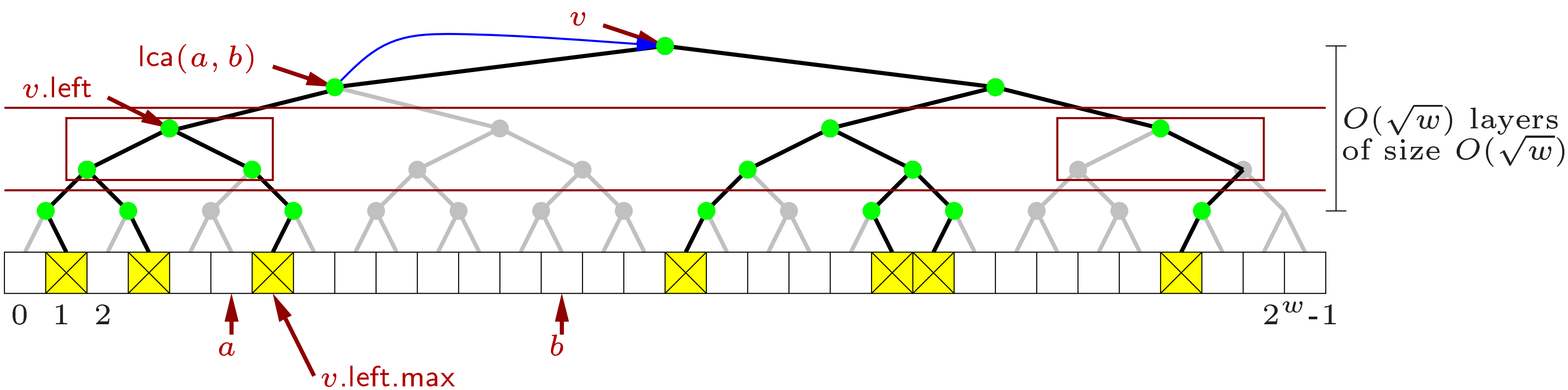
v is the nearest ancestor of lca(a, b) where both subtrees are nonempty

$$\text{FindAny}(a, b) \in [a, b] \cap \{v.\text{left}.\{\text{min}, \text{max}\}, v.\text{right}.\{\text{min}, \text{max}\}\}$$

Lemma

Store ● nodes, and depth(v) − depth(lca(a, b)) for all lca(a, b) with nonempty subtrees using a perfect hash function [Schmidt, Siegel 1990] ⇒ space $O(n + \frac{n \cdot w \cdot \log w}{w}) = O(n \log w)$

The Last Idea



Lemma

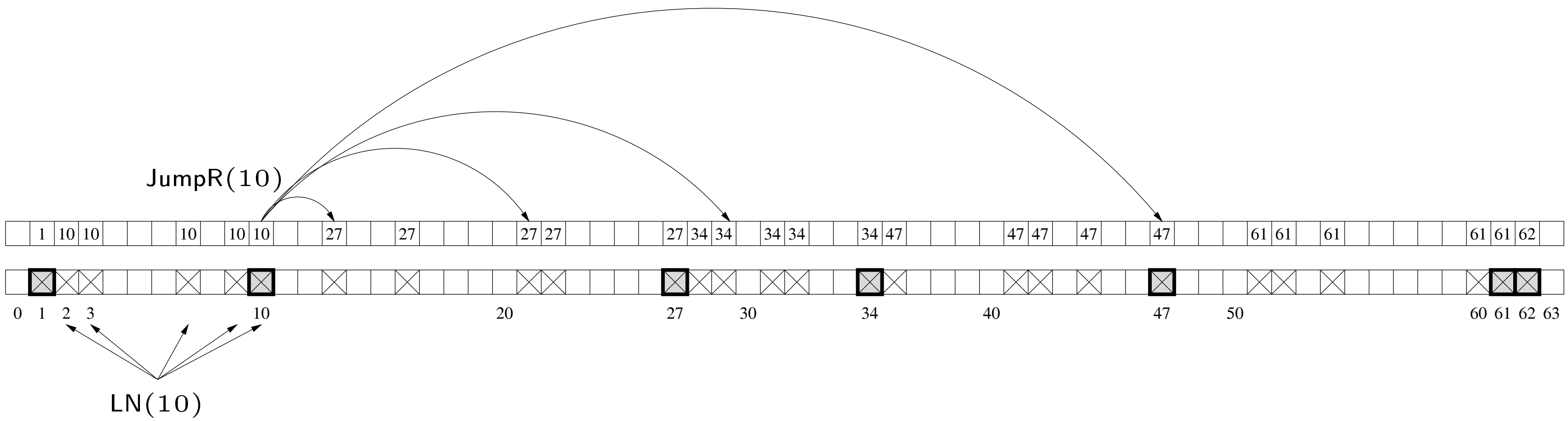
Only store a subset of depth(v) − depth(lca(a, b)) (● nodes)

$\Rightarrow$ space $O(n + \frac{n \cdot \sqrt{w} \cdot \log w}{w}) = O(n)$

Theorem

FindAny can be supported in constant time and space $O(n)$

Approximate Range Counting



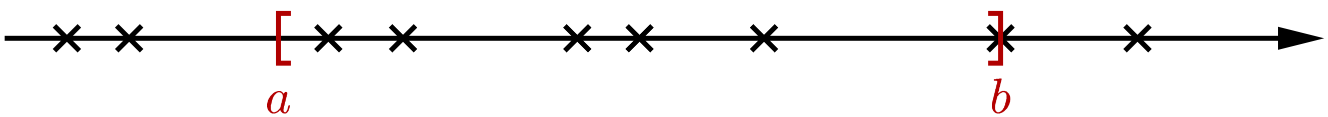
- Partition elements into blocks of size $\log n$
- For each block marker keep two Q-heaps [Fredman, Willard, 1994]
  - Linear space
  - Rank queries in constant time (for logarithmic sized problems)
  - One Q-heap for the $O(\log n)$ elements in the block
  - One Q-heap for the $1, 2, \ldots, 2^i, \ldots$ neighbors

Theorem

Count_ε can be supported in time $O(\log \frac{1}{\varepsilon})$ and space $O(n)$

Summary

	Query time	Space
FindAny	1	n
Report	$1 + \|\text{output}\|$	n
Count_ε	$\log \frac{1}{\varepsilon}$	n



Preprocessing takes expected time $O(n\sqrt{w})$