# Lecture 15



## Tight bounds for predecessor: [Pătrascu & Thorup - 2005 & 2006]

- $n$ integers
- word size $w$
- page size $B$ (external-memory model)
- space $n \cdot 2^a$ bits
- static
- query: $\Theta\left(\min\begin{cases}\log_B n \\ \log_w n \\ \lg\left(\frac{w-\lg n}{a}\right) \\ \frac{\lg(w/a)}{\lg\left(\frac{a}{\lg n}\cdot\lg(w/a)\right)} \\ \frac{\lg(w/a)}{\lg\left(\frac{w}{a}\cdot\lg\left(\frac{\lg n}{a}\right)\right)}\end{cases}\right)$

## Consequences:

- B-trees are optimal
- fusion trees are optimal
- van Emde Boas is optimal

$\}$ space $n \lg^{O(1)} n$

- space $n \lg^{O(1)} n \Rightarrow \Theta\left(\min\begin{cases}\log_B n \\ \log_w n \\ \frac{\lg w}{\lg\left(\frac{\lg w}{\lg\lg n}\right)}\end{cases}\right)$

Here: assume $w = 3 \lg n$, $a = O(\lg \lg n)$

## (I) New framework for error:

- recall $f^{(h)}$:

| Alice | $\longrightarrow$ | Bob |
|---|---|---|
| $x_1, \dots, x_h$ | | $i \in [h] = \{1, 2, \dots, h\}$ |
| | | $x_1, \dots, x_{i-1}$ |
| goal: $f(x_i, y)$ | | $y$ |

Step 1: Alice & Bob accept/reject their inputs

$\Pr\{\text{Alice accepts}\} = \alpha$

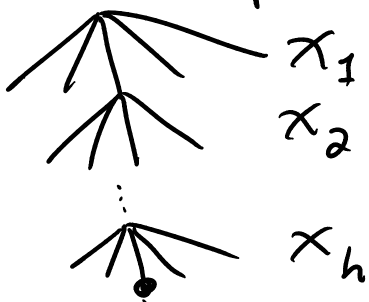
$\Pr\{\text{Bob accepts}\} = \beta$

Step 2: if accepted, Alice & Bob communicate and output $f(x_i, y)$ correctly

## Round elimination in new model:

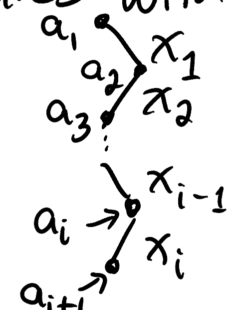
$f^{(k)}$, Alice accepting with probability $\alpha$
$\Rightarrow f$, Alice accepting with probability $\alpha' = \frac{\alpha}{2h}$

Proof: view Alice's input as trie:




$\hookrightarrow \Gamma = \{$message sent to Bob$\}$
$= \emptyset$ if rejected input

Bob simulates with real $x_1, \ldots, x_{i-1}$
$\Rightarrow$ get to node $a_i$




Choose msg. $m(a_i)$ uniformly at random from $\Gamma(a_i)$ = union of descendant $\Gamma$(leaves).
If $m(a_i) \in \Gamma(a_{i+1})$ then Alice is happy
& protocol can continue
else Alice rejects

$|\Gamma(\text{leaf})| = 1$ if accepted; $E[|\Gamma(\text{leaf})|] = \alpha$
height $h \Rightarrow |\Gamma(\text{root})| \le 2^{h-1}$
$\Rightarrow \exists i$ s.t. $|\Gamma(a_{i+1})| \ge \frac{1}{2} |\Gamma(a_i)|$
$\Rightarrow$ random $m \in \Gamma(a_i)$ is good with probability $\ge \frac{1}{2}$
$\Rightarrow \Pr\{\text{Alice accepts}\} = \alpha \cdot \frac{1}{h} \cdot \frac{1}{2} = \alpha'$ $\square$

## Predecessor lower bound from last time, again

$\alpha > \left(\frac{1}{2h}\right)^T \qquad \beta > \left(\frac{1}{2h}\right)^T$

$T = \#$ rounds $\quad h < w^{O(1)} \quad T < \lg w$ [vEB]

$\alpha, \beta > \left(w^{O(1)}\right)^{-\lg w} = 2^{-O(\lg^2 w)}$

**Base case:** when all communication eliminated

Alice: $x \in \{0, 1\}^{\lg^2 w}$

Bob: colors for elements $1, 2, \ldots, 2^{\lg^3 w}$

$2^{\lg^3 w - O(\lg^2 w)}$ possibilities to accept

Bob has to reject $\beta \geq 1 - 2^{-2^{\lg^2 w}}$ ~ big!

## Beyond this bound:

- need to go beyond communication complexity
- e.g. $w = 3 \lg n$, $a = O(\lg \lg n)$
  - $\Rightarrow$ Alice can send input in 3 rounds & Bob can compute answer
  - but not a valid DS

## (II) New lower bound:

- allow a cache of published information that's free to access by both Alice & Bob
- to eliminate message from Alice to Bob, just publish the cell being probed } preprocessing
- to eliminate message from Bob to Alice:
  - like before, split Bob's integers into $w$ chunks:

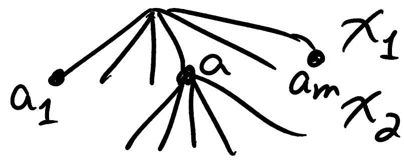
    $A[1] \cdots A[n/w] \mid \cdots \mid \cdots \mid A[n-n/w] \cdots A[n]$
  - now have $w$ different queries: (not just rand. 1)

    $q_1 \quad \cdots \quad \cdots \quad q_w$
  - each round, # queries $= k =$ # subproblems goes up by a factor of $w$.

- to prove $\Omega(\lg w)$ LB, just need $f^{(2)}$ $(h=2)$

<u>Claim</u>: if we publish $\sqrt{s}$ cells ($s = \text{space} = n2^a$)
then can handle $f^{(2)}$, accept prob. 1/2

<u>Proof</u>: consider trie of inputs:




$|\Gamma(a)| \le \sqrt{s} \Rightarrow$ happy with $i=2$
just publish $\Gamma(a)$ [Bob knows $x_1$]

$|\Gamma(a)| \ge \sqrt{s} \Rightarrow$ happy with $i=1$
Bob publishes $\sqrt{s}\lg m$ random cells
some $\Gamma(a_1) \cdots \Gamma(a_m)$ is fine

if get an unhappy $i$, just reject (prob. 1/2) $\square$

(III) Intuition: $k$ queries to $k$ different chunks should be equivalent to $k$ queries to $k$ independent DS's, each size $\approx s/k$

- then can publish $\sqrt{s/k}$ bits for each subprob.
- total: $k\sqrt{s/k} = \sqrt{sk}$

$\Rightarrow$

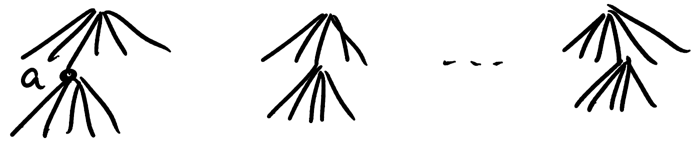
| round | $\rightarrow$ $k$ | & $s/k$ |
| --- | --- | --- |
| 0 | 1 | $s$ |
| 1 | $\sqrt{s}$ | $\sqrt{s}$ |
| 2 | $s^{3/4}$ | $s^{1/4}$ |

- termination: $k < n$, $w > 1$

$\Rightarrow \lg \lg s$ iterations

$\Rightarrow$ get LB of $\lg \lg s = \Omega(\lg w)$ in this case

## Proof of intuition:

k subproblems: 


if $|\Gamma(a)| < \sqrt{s/k}$ then happy with that subprob.

with $k=2$

else publish $\sqrt{s/k}$ per subproblem

$k\sqrt{s/k}$ total $= \sqrt{sk}$

IV Chernoff to combine "error probabilities"