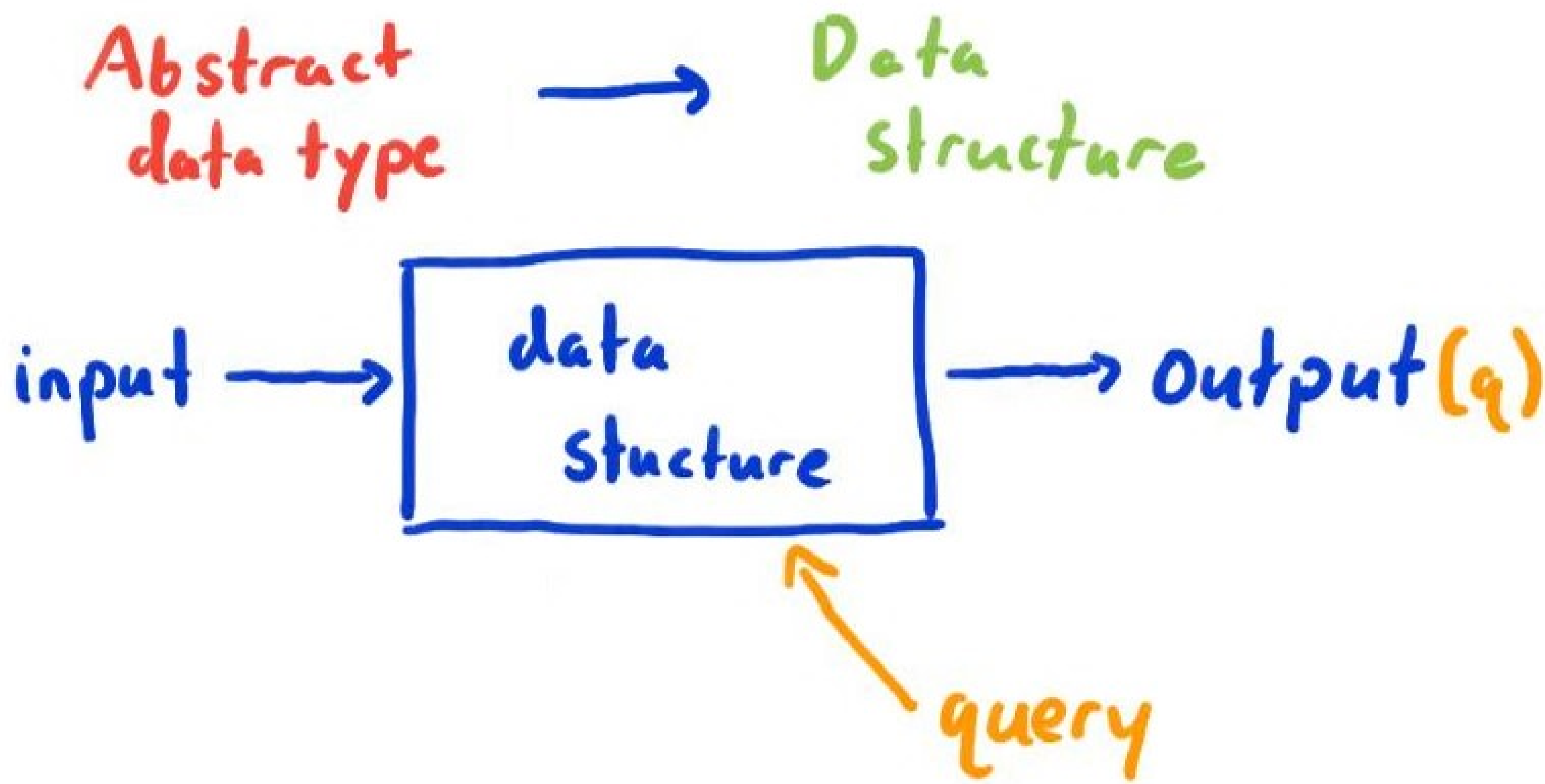
Cell Probe Lower Bounds For Prefix Sum

Ian Mertz

DATA STRUCTURES

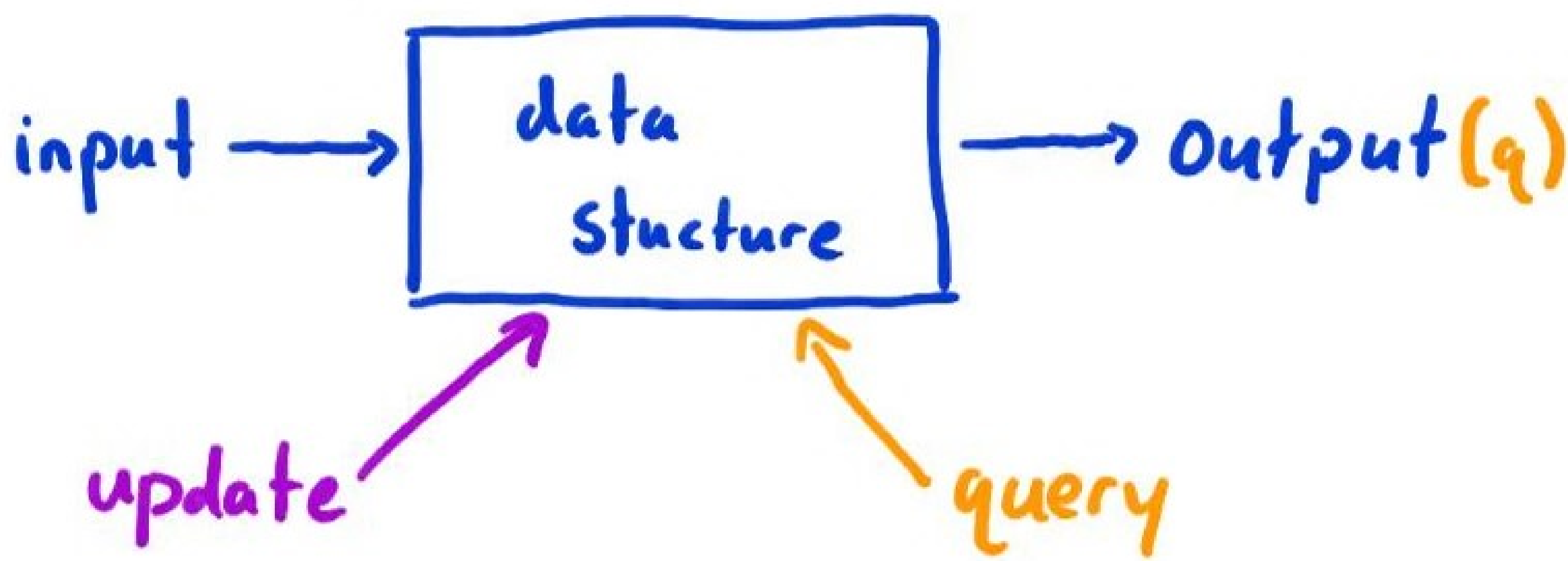
Problem → Algorithm

DATA STRUCTURES

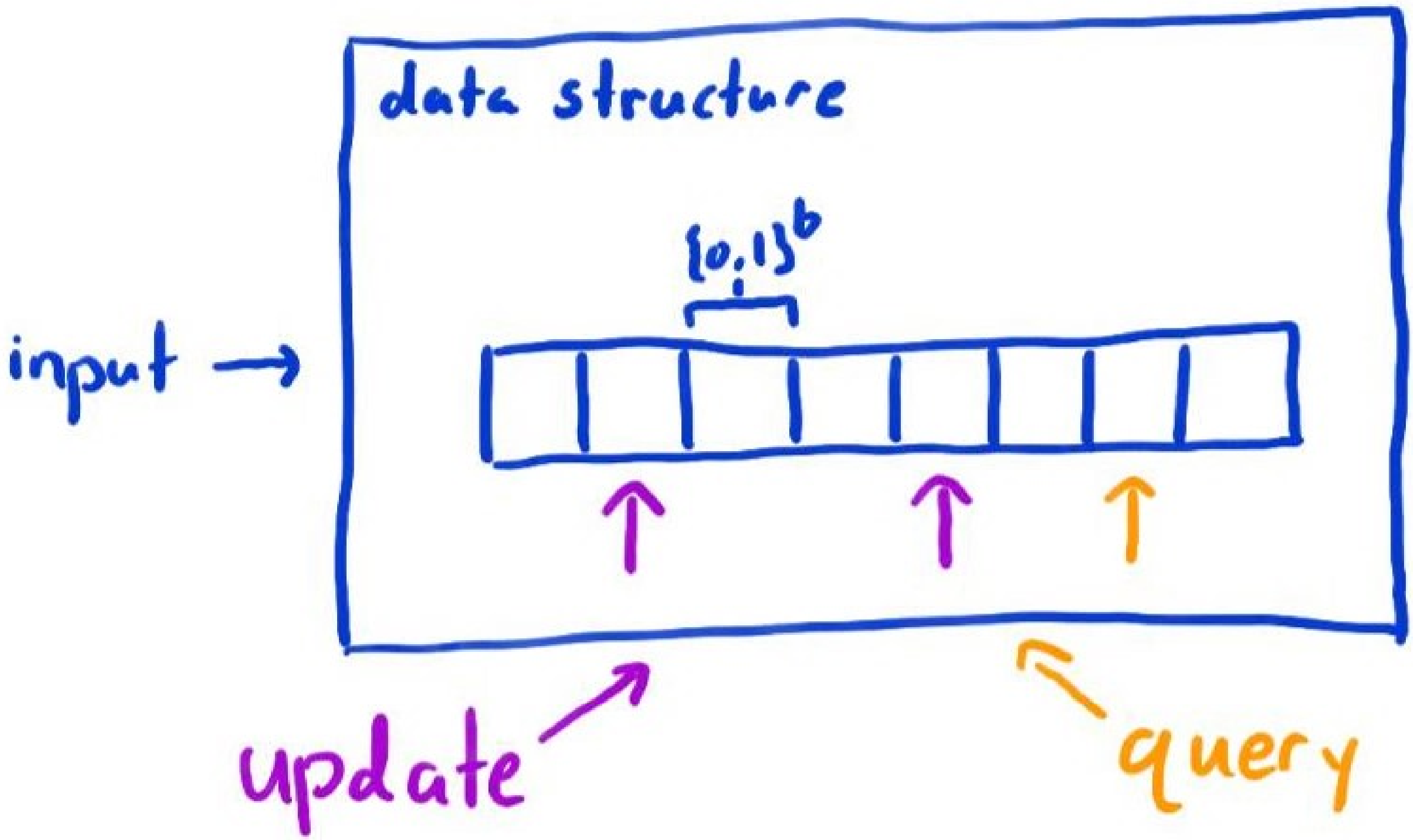


DATA STRUCTURES (dynamic)

Abstract data type → Data structure



Cell Probe Model



CELL PROBE MODEL

t_u = amortized # cells probed during updates

t_q = amortized # cells probed during queries

(probe = read/write)

Lower Bounds

[FS'89]: $\Omega(\log n / \log\log n)$

[PD'06]: $\Omega(\log n)$

[L'12]: $\Omega((\log n / \log\log n)^2)$ *

[LWY'19]: $\Omega(\log^{1.5} n)$

Prefix Sum

init: $A[1..n]$ $\quad\quad A[i] \in [\Delta]$

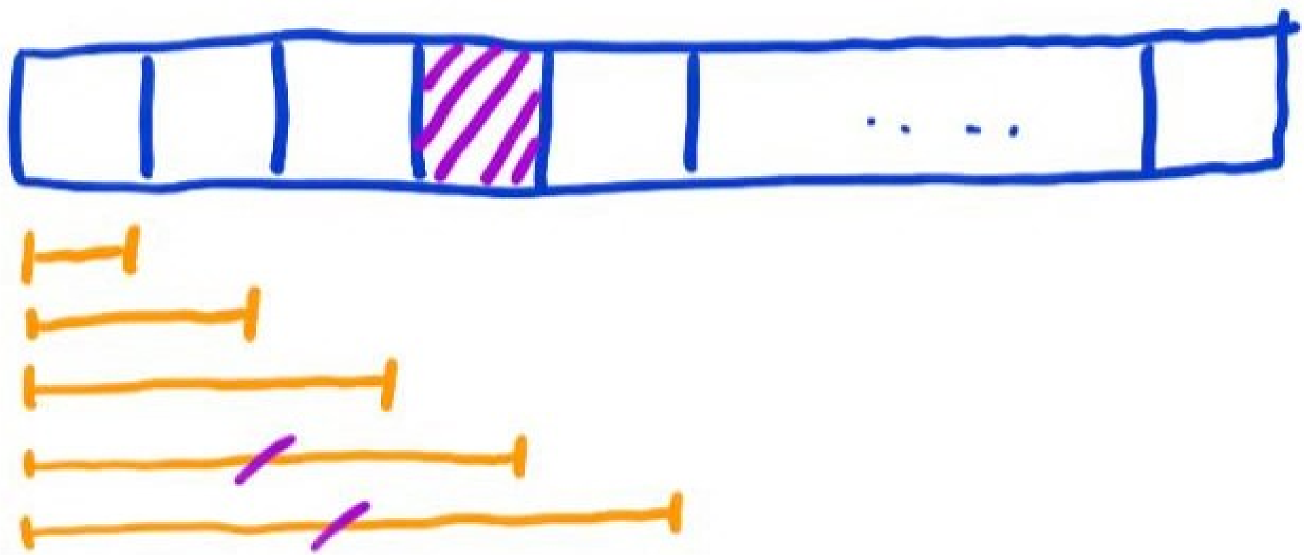
update: $A[i] \leftarrow A[i] + \delta \mod \Delta$

query: return $\sum_{j=1}^{i} A[j] \mod \Delta$

$\Delta \leq 2^b = \Theta(n)$

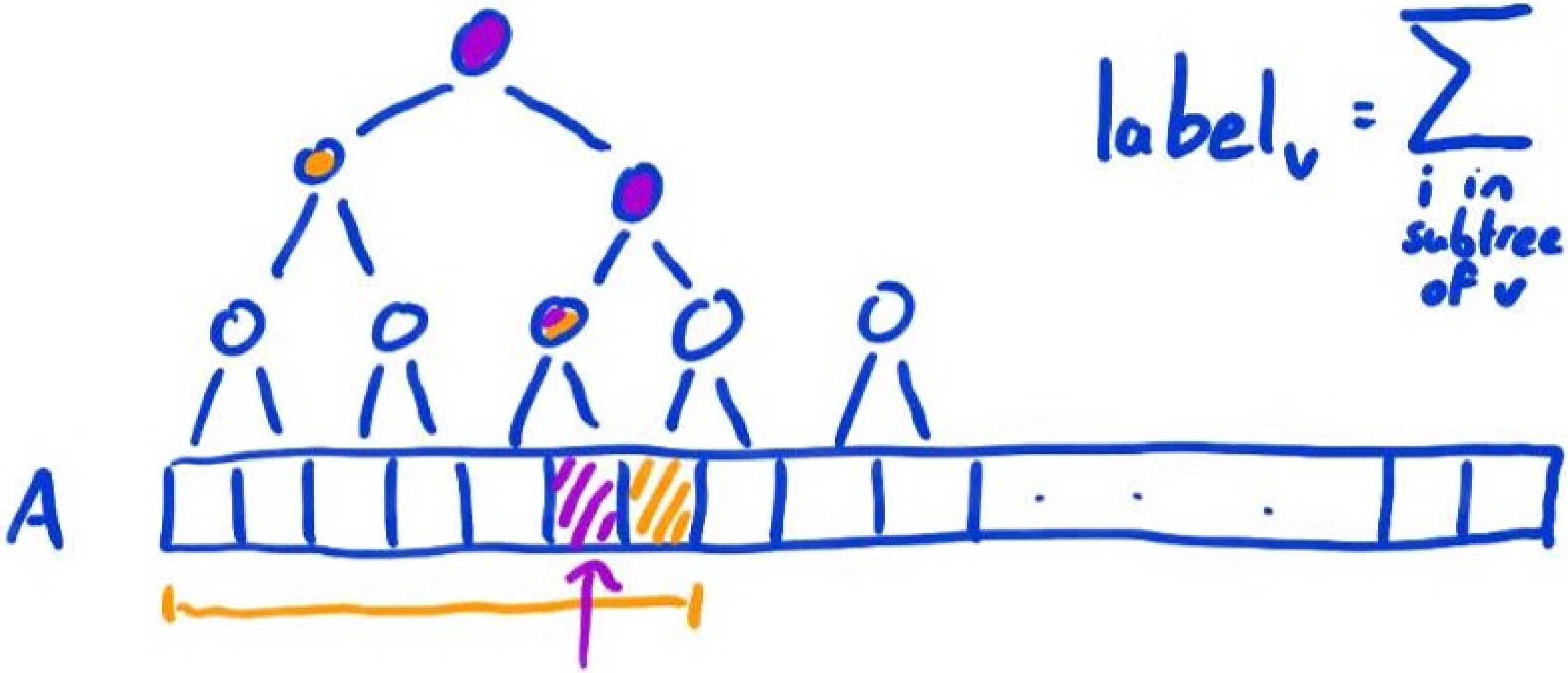
Prefix Sum

random updates, queries



PREFIX SUM

Lemma: $t_u = t_q = O(\log n)$



$$\text{label}_v = \sum_{\substack{i \text{ in} \\ \text{subtree} \\ \text{of } v}} A[i]$$

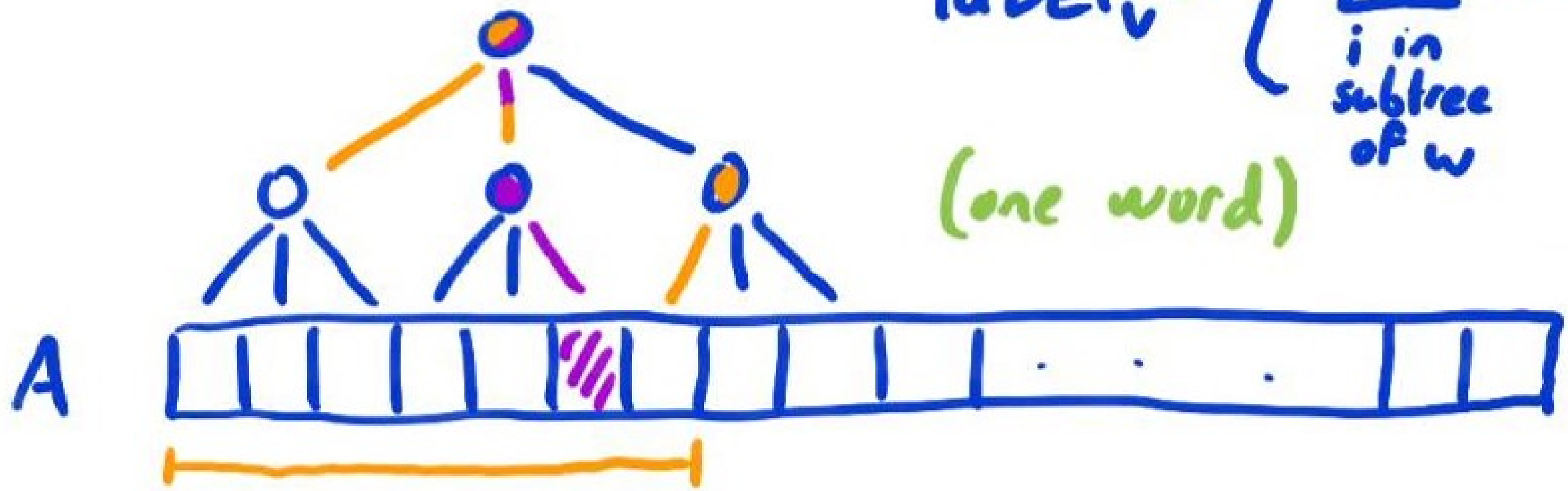
Prefix Sum

Lemma: $t_u = t_q = O\left(\frac{\log n}{\log(b/\log \Delta)}\right)$

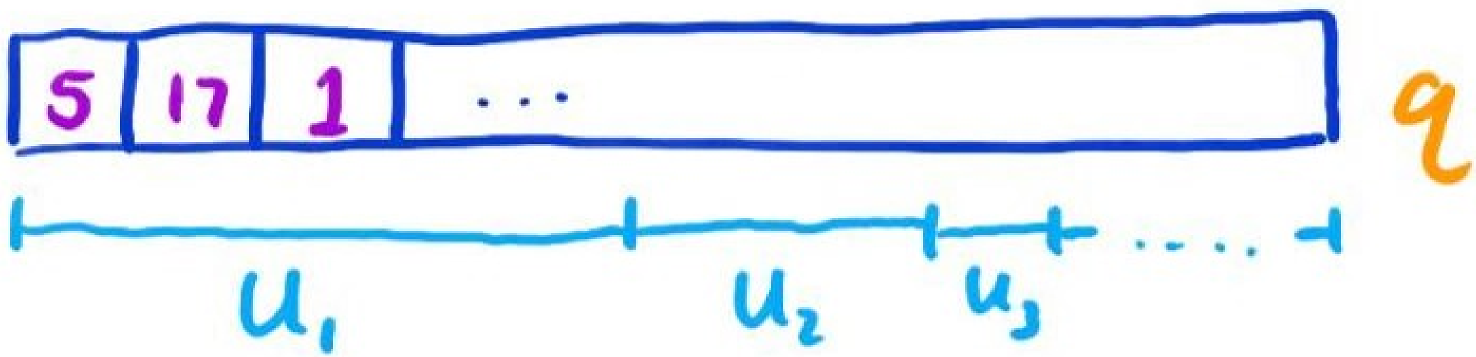
fan-in $= \frac{b}{\log \Delta}$

$\text{label}_v = \left\{ \sum_{\substack{i \text{ in} \\ \text{subtree} \\ \text{of } w}} A[i] \right\}_{\substack{w \text{ child} \\ \text{of } v}}$

(one word)



CHRONOGRAM [FS'89]



<u>WTS</u>: $\forall i$, q probes $\Omega(1)$ cells last updated in U_i

$$|U_i| = \beta^i \qquad i \in \{\log_\beta n \dots 1\}$$

$$|U_i| = \underbrace{\beta^i > 2\beta^{i-1} t_u} > t_u \sum_{i'=i-1}^{1} |U_{i'}|$$

$$\beta = 10\, t_u \longrightarrow t_q \geq \Omega\left(\frac{\log n}{\log t_u}\right)$$

$$\longrightarrow \max(t_u, t_q) \geq \Omega\left(\frac{\log n}{\log \log n}\right)$$

(coding argument, depends on b, Δ)

[PD'06] Main Theorem

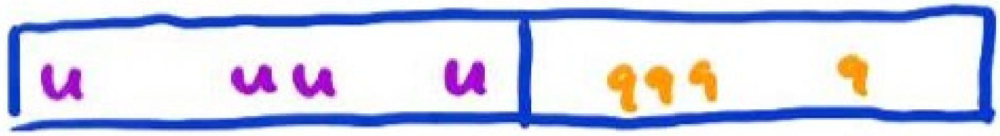
$$\max(t_u, t_q) \log \frac{\max(t_u, t_q)}{\min(t_u, t_q)} = \Omega\left(\frac{\log n}{\log(b/\log \Delta)}\right)$$

[PD'06]

u u q u q q q u ....

"chronogram": if the cell was last probed by an update, write the time

[PD'06]



Lemma: suppose for two adjacent intervals (of total size $O(n^{1/3})$) the left (right) interval has at least L updates (queries);

$$\mathbb{E}[c] = \Omega\left(\frac{L}{b/\log \Delta}\right)$$

c := #probes executed during queries on the right that were <u>last</u> written during updates on the left

[PD'06]

1) all indices unique (birthday paradox + Markov)

2) $u_1 < q_1 < u_2 < \ldots < u_I < q_I$ for $I = \Omega(L)$ (random indices)

3) $|\langle q_1 \cdots q_I \rangle| = \Delta^I$ (coding argument)

$$\rightarrow \mathbb{E}[c] \geq \frac{I \log \Delta}{b} = \Omega\left(\frac{L}{b/\log \Delta}\right)$$

[PD'06] (improved)

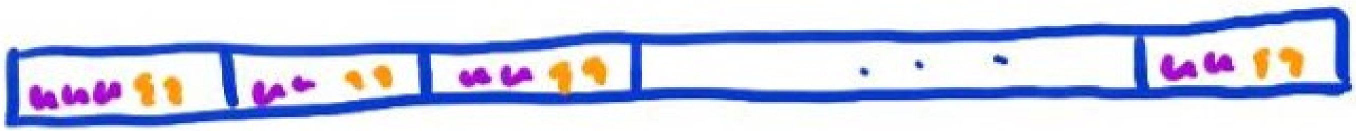
if $\log \Delta = o(b)$, can't use words efficiently

$$\Omega\left(\frac{L}{\log(b/\log \Delta)}\right)$$

[PD'06] Main Theorem

$$\max(t_u, t_q) \log \frac{\max(t_u, t_q)}{\min(t_u, t_q)} = \Omega\left(\frac{\log n}{\log(b/\log \Delta)}\right)$$

[PD'06] Main Theorem



m blocks of t_q updates + t_u queries

$$\rightarrow \mathbb{E}\left[\frac{c}{m}\right] \leq 2\, t_u t_q$$

<u>WTS</u>: $\mathbb{E}\left[\frac{c}{m}\right] \geq \Omega\left(\frac{\log_B n}{\log(b/\log \Delta)} \max(t_u, t_q)\right)$

$$B := 2 \cdot \max\left(\frac{t_u}{t_q}, \frac{t_q}{t_u}\right)$$

WTS: $\mathbb{E}\left[\frac{c}{m}\right] \geq \Omega\left(\frac{\log_B n}{\log(b/\log \Delta)} \max(t_u, t_q)\right)$

$B := 2 \cdot \max\left(\frac{t_u}{t_q}, \frac{t_q}{t_u}\right)$

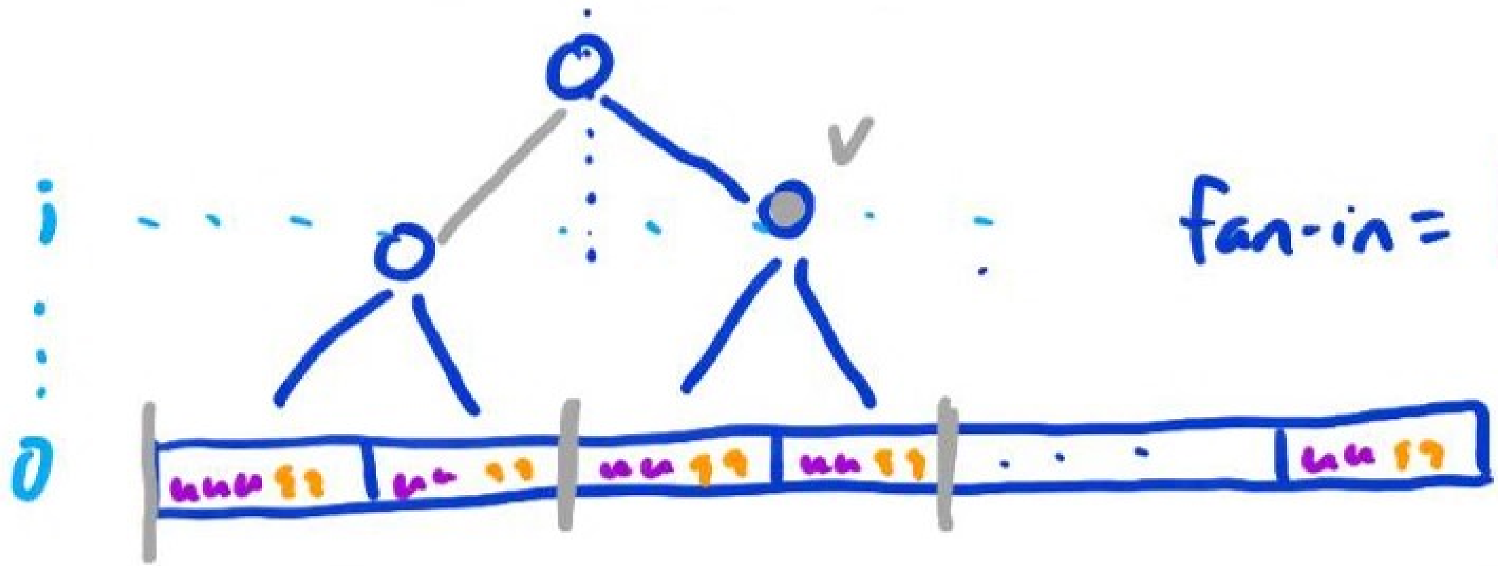
1) $\max(t_u, t_q) \leq O(n^{1/6})$

(trivial otherwise)

2) $m := O(n^{1/6})$ $\left(\begin{array}{l} O(n^{1/3}) \text{ ops total,} \\ \log m = \Theta(\log n) \end{array}\right)$

3) $t_u \geq t_q$ (symmetric argument otherwise)

fan-in = $B = 2 \cdot \frac{t_u}{t_q}$



left interval is all left siblings of v

$$\# u \geq \frac{B}{2} \cdot B^i \cdot t_q = B^i \cdot t_u$$

right interval is v $\rightarrow L$

$$\# q = B^i \cdot t_u$$

- lemma for each right child ✓
  at every level i

- no interval pairs double counted
  across / between layers

- linearity of expectation

v at level i $\quad \mathbb{E}[c_v] = \Omega\left(\frac{B^i t_u}{\log(b/\log\Delta)}\right)$

all right children at level i $\quad \mathbb{E}[c_i] = \Omega\left(\frac{m\, t_u}{\log(b/\log\Delta)}\right)$

all levels $\quad \mathbb{E}[c] = \Omega\left(\frac{m\, t_u}{\log(b/\log\Delta)}\right) \cdot \log_B m$

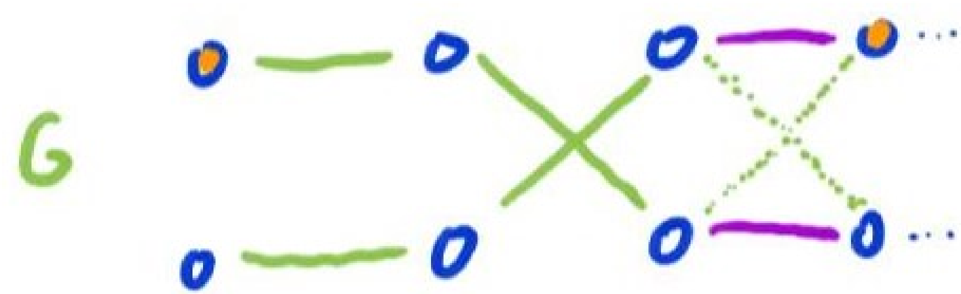
<u>per block</u> $\quad \mathbb{E}\left[\frac{c}{m}\right] = \Omega\left(\frac{\log_B n}{\log(b/\log\Delta)} t_u\right)$ ■

Boolean Outputs

$$\max(t_u, t_q) = \Theta\left(\frac{\log n}{\log \log n/2}\right) = \Theta\left(\frac{\log n}{\log\log n}\right)$$

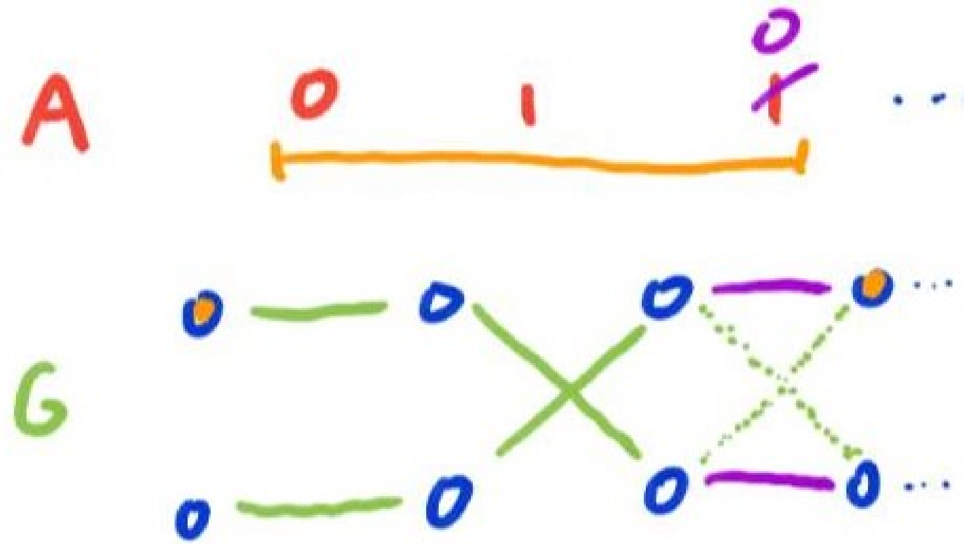
Boolean Outputs

connectivity

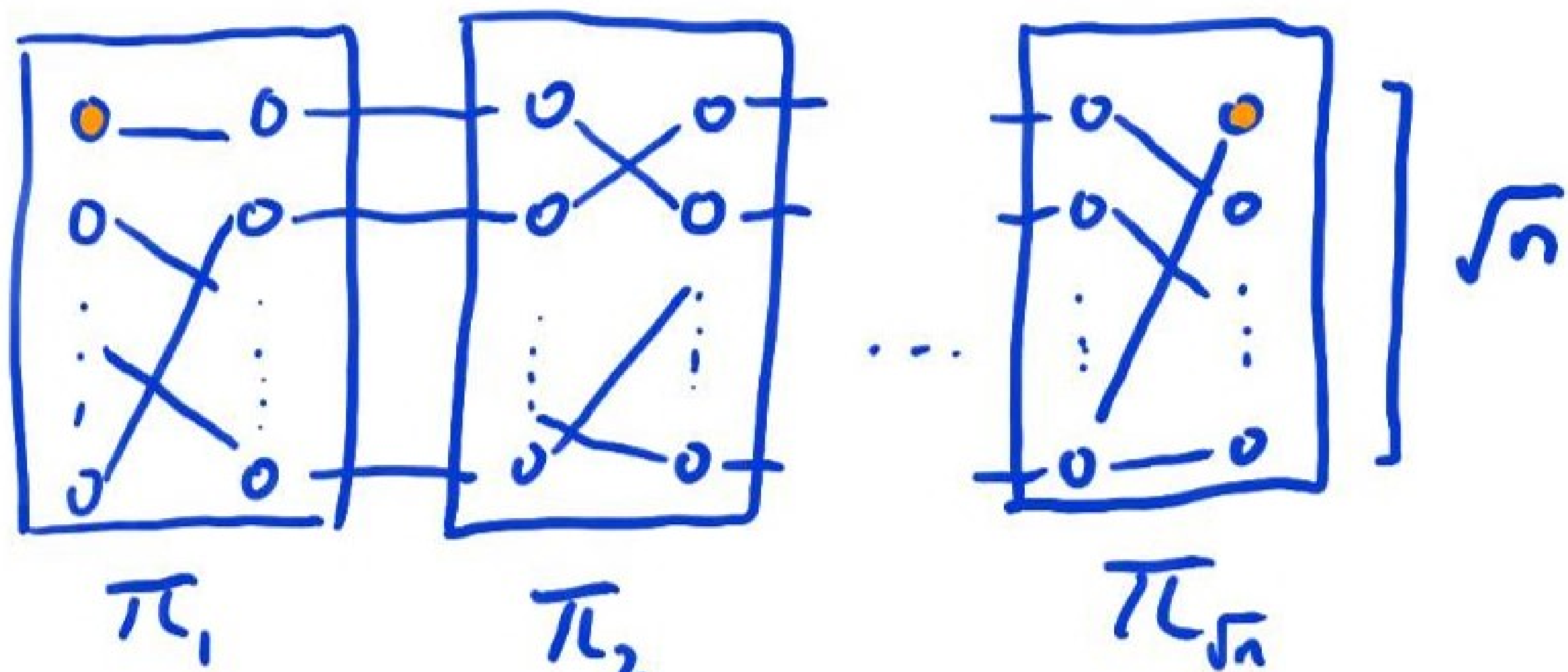


Boolean Outputs

Prefix sum connectivity →



Boolean Outputs



$\Delta = 2^{\sqrt{n}\log n}$

Boolean Outputs

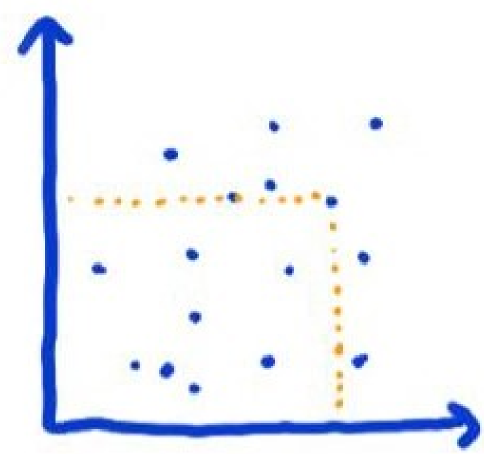
$$\Delta = 2^{\sqrt{n}\log n} \rightarrow \max(t_u, t_q) = \frac{\log n}{b/\log\Delta} = \Omega(\sqrt{n}\log n)$$

$$\text{"update"} = \sqrt{n} \text{ updates} \rightarrow \Omega(\log n)$$

EXTENSIONS

[L'12]: 2D range counting

[LWY'19]: ⊕-2DRC



Open Questions

$$\max(t_u, t_q) = \Omega(\log^2 n)$$