Bit-probe lower bounds for succinct data structures

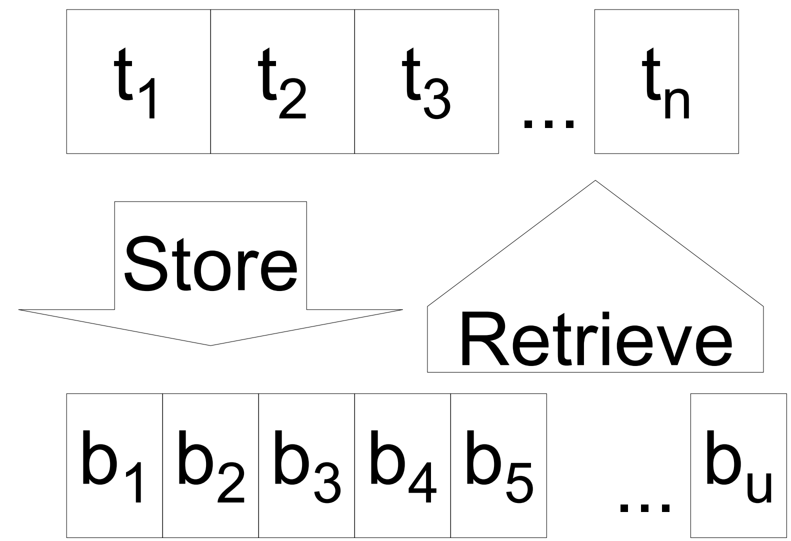
Emanuele Viola

Northeastern University

May 2009

Bits vs. trits

- Store n "trits" $t_1, t_2, \ldots, t_n \in \{0,1,2\}$



In u bits $b_1, b_2, \ldots, b_u \in \{0,1\}$

- Want:

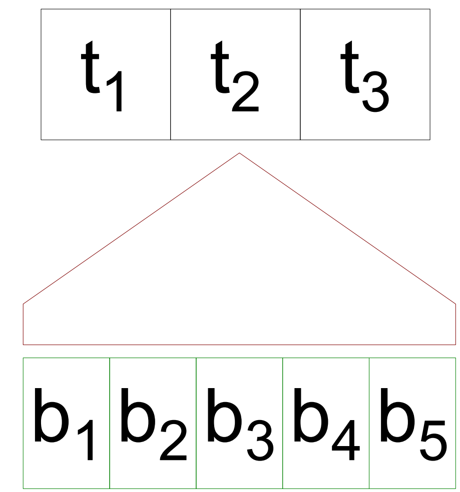
  Small space u

  Short retrieval time: Get t_i probing few bits

Two solutions

- Arithmetic coding:

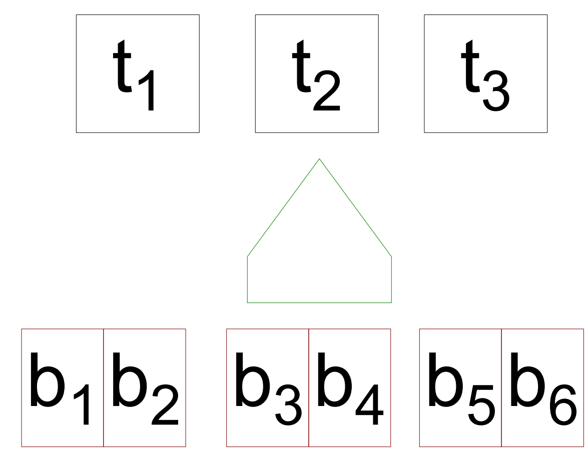
  Store bits of $(t_1, \ldots, t_n) \in \{0, 1, \ldots, 3^n - 1\}$



  Optimal space: $\lceil n \lg_2 3 \rceil$

  Bad retrieval time: To get t_i read all > n bits

- Two bits per trit

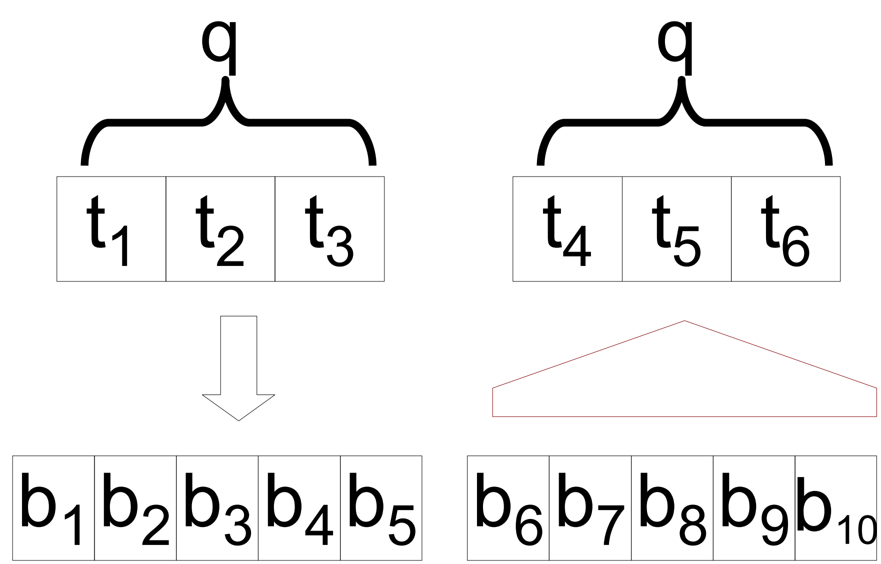


  Bad space: $2n$

  Optimal retrieval time: Read 2 bits

Polynomial tradeoff

- Divide n trits $t_1, \ldots, t_n \in \{0,1,2\}$ in blocks of q
- Arithmetic-code each block



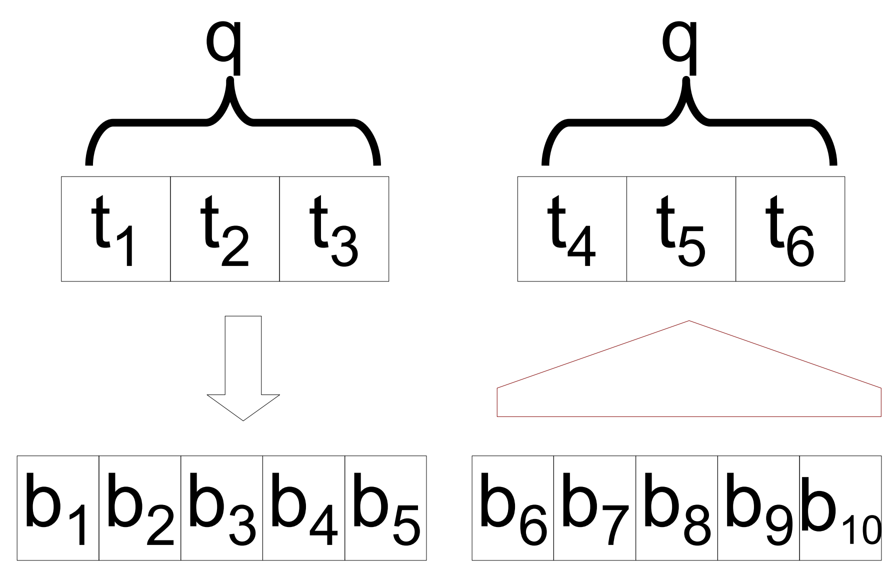
Space: $\lceil q \lg_2 3 \rceil n/q < (q \lg_2 3 + 1)\, n/q$
$= n \lg_2 3 + n/q$

Retrieval Time: $O(q)$

polynomial tradeoff between redundancy, time

Polynomial tradeoff

- Divide n trits $t_1, \ldots, t_n \in \{0,1,2\}$ in blocks of q
- Arithmetic-code each block



Space: $\lceil q \lg_2 3 \rceil n/q = (q \lg_2 3 + 1/q^{\Theta(1)})\, n/q$

$= n \lg_2 3 + n/q^{\Theta(1)}$

Retrieval Time: $O(q)$

polynomial tradeoff between redundancy, time

Logarithmic forms

Exponential tradeoff

- Breakthrough data structure [Pătraşcu '08, later + Thorup]

Space: $n \lg_2 3 + n/2^{O(q)}$

Retrieval Time: q

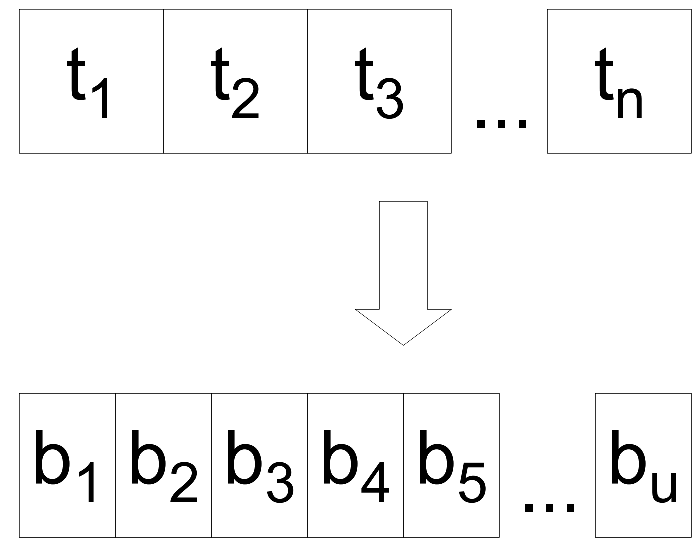
exponential tradeoff between redundancy, time

- E.g., optimal space $\lceil n \lg_2 3 \rceil$, time $O(\lg n)$

Our results

- Theorem[this work]:
  Store n trits $t_1, \ldots, t_n \in \{0,1,2\}$
  in u bits $b_1, \ldots, b_u \in \{0,1\}$.

  If get t_i by probing q bits
  then space $u > n \lg_2 3 + n/2^{\Omega(q)}$.



- Matches [Pătraşcu Thorup]: space $< n \lg_2 3 + n/2^{O(q)}$

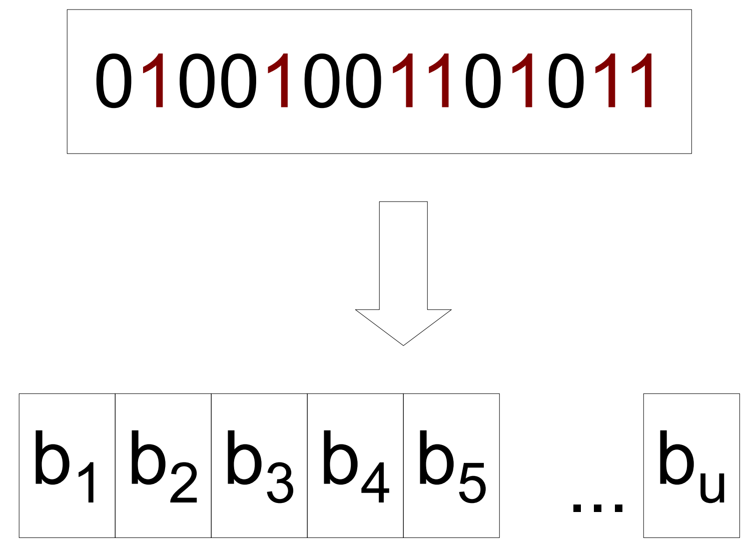
- Holds even for adaptive probes

Outline

- Bits vs. trits
- Bits vs. sets
- Cell model
- Proof

Bits vs. sets

- Store $S \subseteq \{1, 2, \ldots, n\}$ of size $|S| = k$



In u bits $b_1, \ldots, b_u \in \{0,1\}$

- Want:

  Small space u, optimal is $\lceil \lg_2 (n \text{ choose } k) \rceil$

  Answer "$i \in S$?" by probing few bits

Previous results

- Store $S \subseteq \{1, 2, \ldots, n\}$, $|S| = k$ in bits, answer "$i \in S$?"
- [Minsky Papert '69] Average-case study
- [Buhrman Miltersen Radhakrishnan Venkatesh '00]

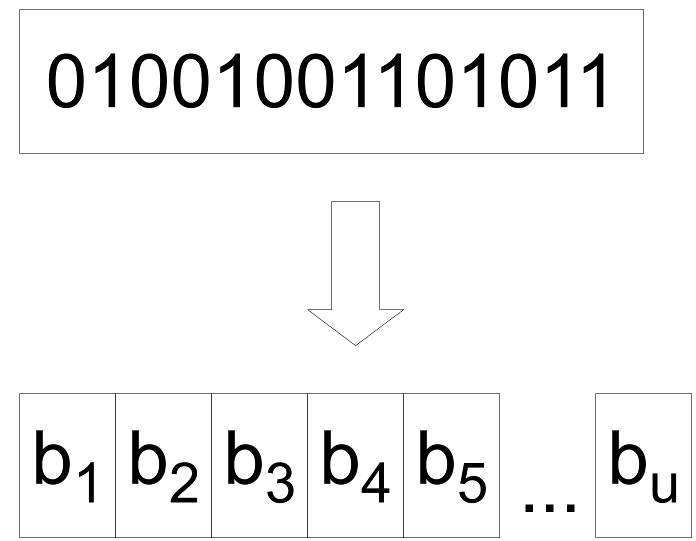
  Space O(optimal), probe 1 bit, correct with high probability

  Lower bounds for $k < n^{1-\varepsilon}$
- No lower bound was known for $k = \Omega(n)$

Our results

- Theorem[this work]:
  Store $S \subseteq \{1, 2, \ldots, n\}$, $|S| = n/3$
  in u bits $b_1, \ldots, b_u \in \{0,1\}$

  If answer "$i \in S$?" probing q bits
  then space $u >$ optimal $+ n/2^{\Omega(q)}$.



- First lower bound for $|S| = \Omega(n)$

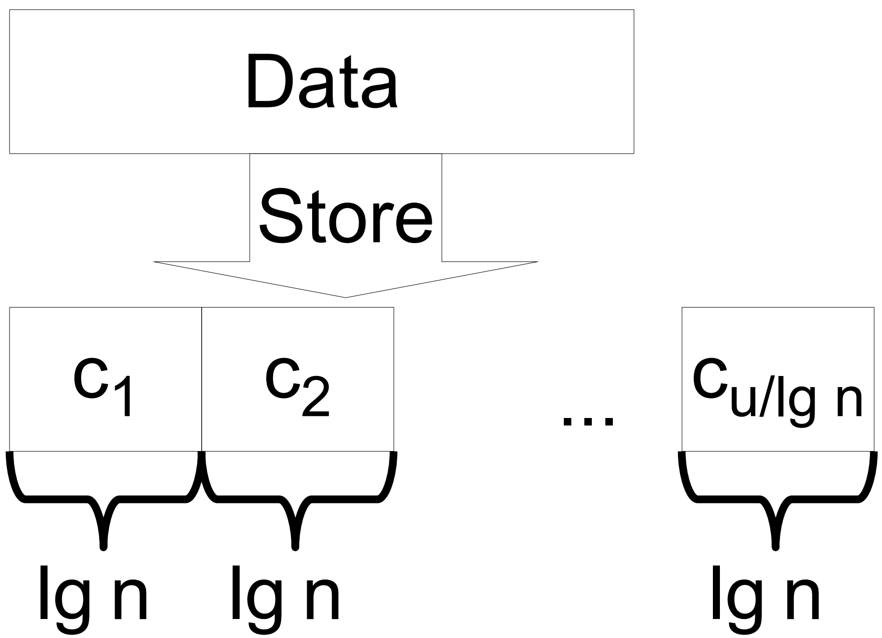
- Holds even for adaptive probes

Outline

- Bits vs. trits
- Bits vs. sets
- Cell model
- Proof

Cell-probe model

- So far: q = number bit probes
- Cell model: q = number of probes in cells of lg(n) bits



- Relationship: q bit $\subseteq$ q cell $\subseteq$ q lg(n) bit

Results in cell-probe model

- Cells vs. trits:

  $q = O(1)$, optimal space = $\lceil n \lg_2 3 \rceil$ [Pătraşcu Thorup]

  Time $q = 1 \Rightarrow$ space $> n \lg_2 3 + n/\lg^{O(1)} n$ [this work]

- Cells vs. sets:

  q probes, space = optimal + $n / \lg^{\Omega(q)} n$ [Pagh, Pătraşcu]

  Lower bounds?

  Work in progress on both fronts

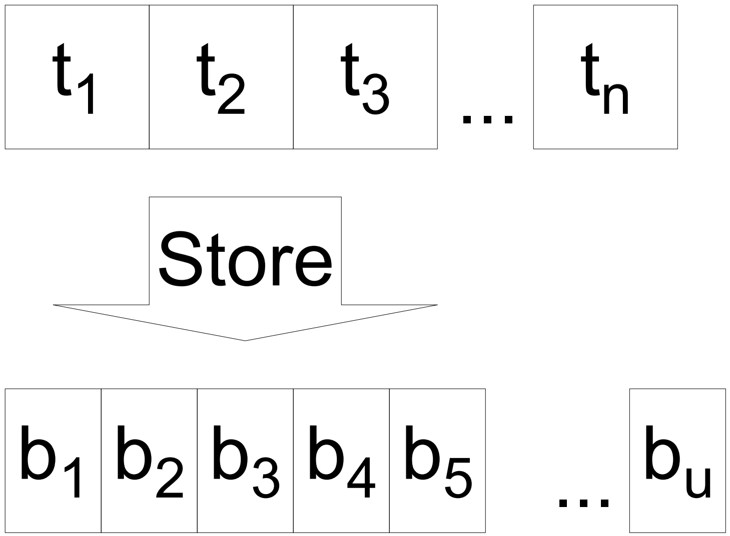
Outline

- Bits vs. trits
- Bits vs. sets
- Cell model
- Proof

Recall our results

- Theorem:
  Store n trits $t_1, \ldots, t_n \in \{0,1,2\}$
  in u bits $b_1, \ldots, b_u \in \{0,1\}$.

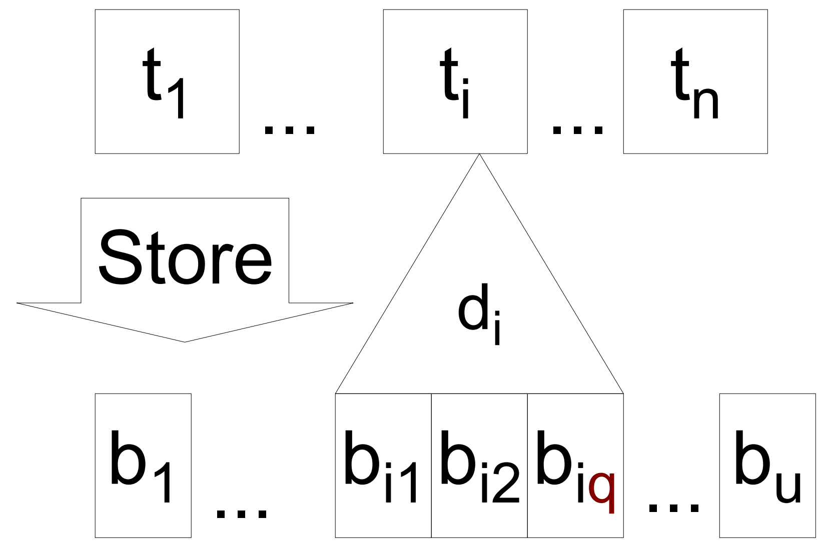
  If get t_i by probing q bits
  then space $u > n \lg_2 3 + n/2^{\Omega(q)}$.



- For now, assume non-adaptive probes:
  $t_i = d_i\,(b_{i1}, b_{i2}, \ldots, b_{iq})$

Proof idea

- $t_i = d_i\ (b_{i1}, b_{i2}, \ldots, b_{iq})$



- Uniform $(t_1, \ldots, t_n) \in \{0,1,2\}^n$

  Let $(b_1, \ldots, b_u) := \text{Store}(t_1, \ldots, t_n)$

- Space $u \approx$ optimal $\Rightarrow (b_1, \ldots, b_u) \in \{0,1\}^u \approx$ uniform $\Rightarrow$

$$1/3 = \Pr[\ t_i = 2\ ] = \Pr[\ d_i\ (b_{i1}, \ldots, b_{iq}) = 2\ ] \approx A / 2^q \neq 1/3$$

Contradiction, so space u >> optimal Q.e.d.

Information-theory lemma

[Edmonds Rudich Impagliazzo Sgall, Raz, Shaltiel V.]

Lemma: Random $(b_1, \ldots, b_u)$ uniform in $B \subseteq \{0,1\}^u$

$|B| \approx 2^u \Rightarrow$ there is large set $G \subseteq [u]$:

for every $i_1, \ldots, i_q \in G$: $(b_{i_1}, \ldots, b_{i_q}) \approx$ uniform in $\{0,1\}^q$

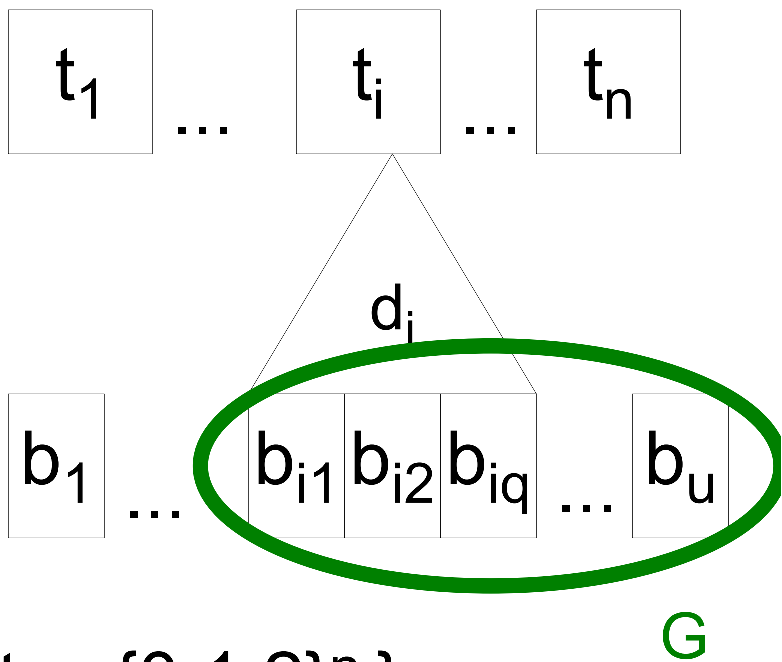
Proof: $|B| \approx 2^u \Rightarrow H(b_1, \ldots, b_u)$ large

$\Rightarrow H(b_i \mid b_1, \ldots, b_{i-1})$ large for many i ($\in G$)

Closeness[$(b_{i_1}, \ldots, b_{i_q})$, uniform] $\geq H(b_{i_1}, \ldots, b_{i_q})$

$\geq H(b_{i_q} \mid b_1, \ldots, b_{i_q-1}) + \ldots + H(b_{i_1} \mid b_1, \ldots, b_{i_1-1})$, large Q.e.d.

Proof



- Argument OK **if** probes in G
- $t_i = d_i\,(b_{i1}, b_{i2}, \ldots, b_{iq})$
- Uniform $(t_1, \ldots, t_n) \in \{0,1,2\}^n$

$$\Downarrow$$

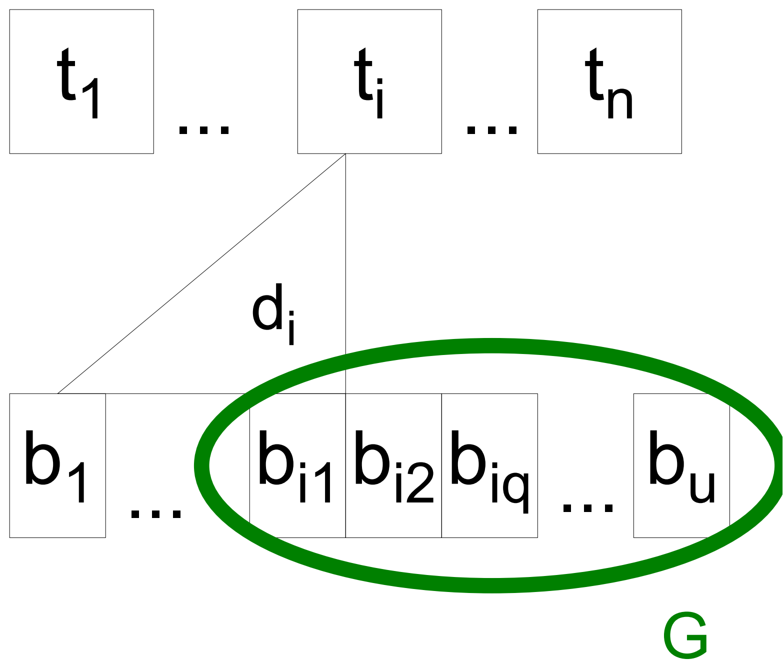
uniform $(b_1, \ldots, b_u) \in B := \{\text{Store}(t) \mid t \in \{0,1,2\}^n\}$

$|B| = 3^n \approx 2^u \Rightarrow (\text{Lemma}) \Rightarrow (b_{i1}, \ldots, b_{iq}) \approx \text{uniform} \Rightarrow$

$1/3 = \Pr[\, t_i = 2 \,] = \Pr[\, d_i\,(b_{i1}, \ldots, b_{iq}) = 2 \,] \approx A / 2^q \neq 1/3$

Probes not in G

- If every t_i probes bits not in G

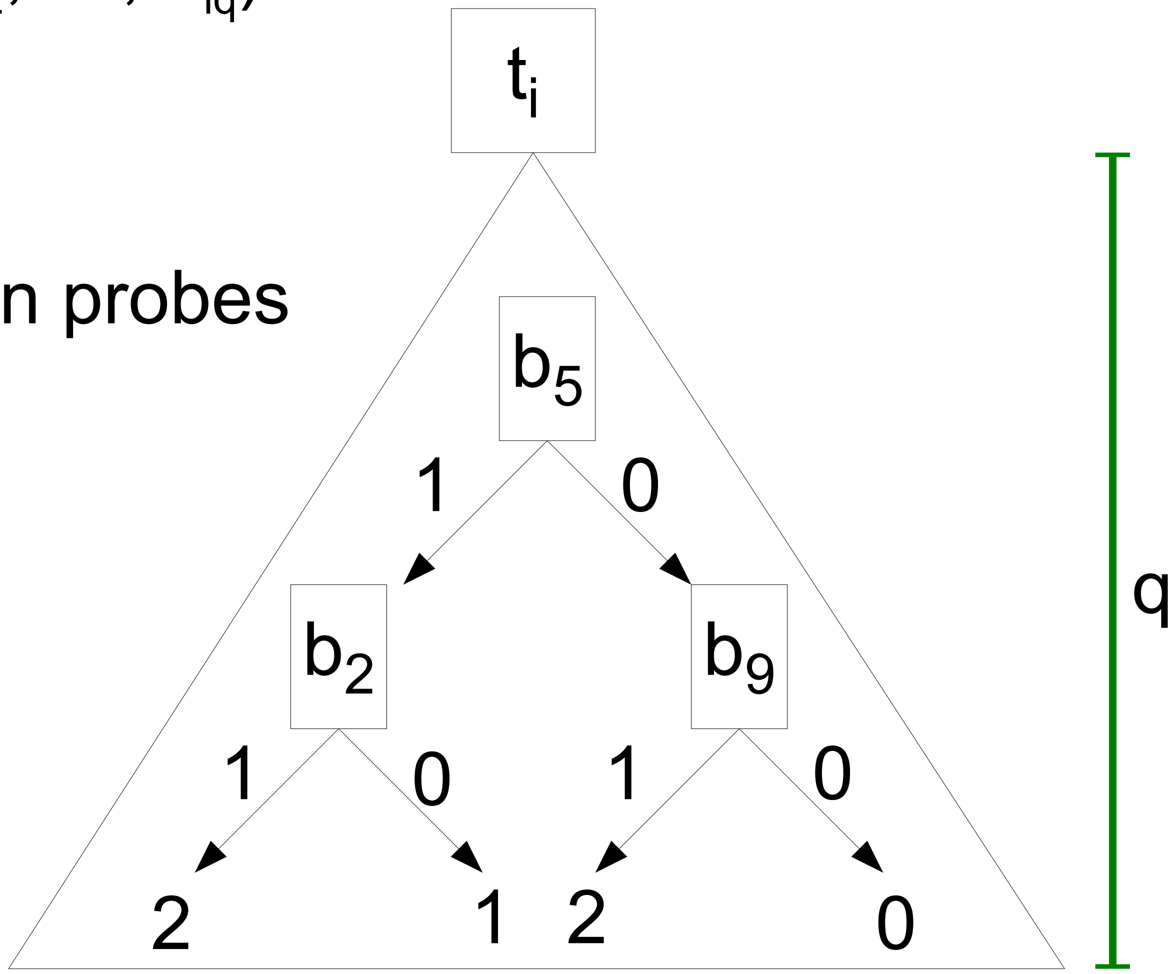


- Argue as in [Shaltiel V.]:
- Condition on heavy bits := probed by many t_i
- Can find $t_i \approx$ uniform in {0,1,2}, all probes in G

Handling adaptivity

- So far $t_i = d_i\ (b_{i1}, b_{i2}, \ldots, b_{iq})$

- In general,
  q adaptively chosen probes
  = decision tree
    2^q bits
    depth q



$1/3 = \Pr[\ t_i = 2\ ] = \Pr[\ d_i\ (b_{i1}, \ldots, b_{i2^q}) = 2\ ] \approx A / 2^q \neq 1/3$

Conclusion

- Thm: Store n trits $t_1, \ldots, t_n \in \{0,1,2\}$.

  Get t_i by probing q bits $\Rightarrow$ space > optimal + $n/2^{\Omega(q)}$

  Matches [Pătraşcu Thorup]: space < optimal + $n/2^{O(q)}$

- Thm: Store $S \subseteq \{1, 2, \ldots, n\}$, $|S| = n/3$.

  Answer "$i \in S$?" probing q bits $\Rightarrow$ space > optimal + $n/2^{\Omega(q)}$

  First lower bound for $|S| = \Omega(n)$

- New approach to lower bounds for basic data structures

- $\Sigma\Pi\surd\cap\notin\cup\supset\supseteq\not\subset\subset\subseteq\in\Downarrow\Rightarrow\Uparrow\Leftarrow\Leftrightarrow\vee\wedge\geq\leq\forall\exists\Omega\alpha\beta\varepsilon\gamma\delta$
- $\neq\approx$