Cell-probe lower bounds

Cell-probe model

Memory is divided into w-bit words (cells). Memory reads and writes are charged for unit cost. All other computation is free. Each memory access (read or write) is called a cell-probe.

Strong model, lower bounds (impossibility results) in cell-probe model imply lower bounds in word RAM.

Think of $w = \Theta(\log n)$ today.

Dynamic partial sum data structure lower bound

Maintain an array $A[0..n-1]$, initialized to all 0, supporting

- update(i, Δ): $A[i] \leftarrow \Delta$
- query(i): return $\sum_{j=0}^{i-1} A[j]$

Range tree (Midterm 2) solves this problem with

$O(\log n)$ time per operation.

We will show today: Any data structure for this problem must use $\Omega(\log n / \log\log n)$ time per operation. We prove this bound even if $\Delta \in \{0,1\}$, and query(i) asks $\sum_{j=0}^{i-1} A[j] \bmod 2$.

Thm: Fix a data structure D for dynamic partial sum. Let t_u be its time to process an update, t_q be the time to answer a query. Then $\max\{t_u, t_q\} = \Omega\left(\frac{\log n}{\log\log n}\right)$ when word-size $w = \Theta(\log n)$.

Chronogram method: consider a sequence of ops

$$\underbrace{U_n \, U_{n-1} \cdots\cdots}_{\text{epoch } k} \;\cdots\; \underbrace{\qquad\qquad}_{\text{epoch } 2} \underbrace{U_2 \, U_1}_{\text{epoch } 1} \;\; Q$$

Each U_i is an update. a single query Q at the end.

The updates are divided into k epochs.

- Epoch i contains β^i updates ($\beta = \log^4 n$).

$\sum_{i \le k} \beta^i = n$.

$k = \Theta\left(\frac{\log n}{\log \beta}\right)$

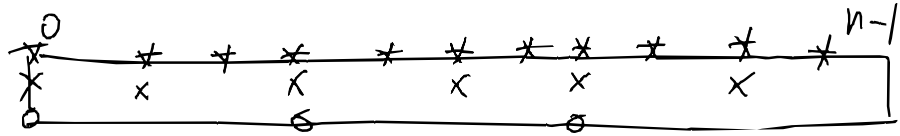
- Epoch k happens first,

then Ep $k-1$

⋮

Ep 1.

- The β^i updates in Epoch i update evenly spaced indices: $j = 0, 1, \ldots, \beta^i - 1$, $j \cdot \frac{n}{\beta^i}$.



E.g., Epoch 2 updates $*$

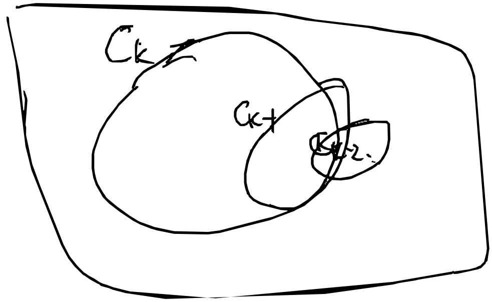
then . . 1 . . $\times$

then . . 0 . . $\circ$.

Each update updates the entry to a random $\Delta \in \{0, 1\}$.

- Let C_i be the set of cells probed by the algorithm during Epoch i, but not later.

memory



main lemma: Let Q be a random query (i), $i \in \{1, \dots, n\}$, let P be the set of cells the query algo probes when answering Q. Then if $t_u = O(\frac{\log n}{\log\log n})$,

$$\mathbb{E}[|P \cap C_i|] \geq \Omega(1)$$

for $i = 1, \dots, k$.

Lemma $\Rightarrow$ if $t_u = O(\frac{\log n}{\log\log n})$, then

$$\mathbb{E}[|P|] \geq \sum_{i=1}^{k} \mathbb{E}[|P \cap C_i|] = \Omega(k) = \Omega(\frac{\log n}{\log\log n}).$$

$$t_q \geq \Omega(\frac{\log n}{\log\log n}).$$

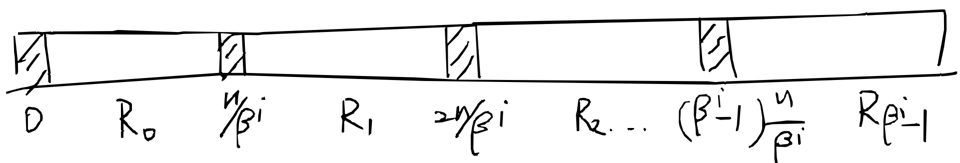
To prove the lemma, we will use Shannon's source coding thm: No lossless compression can compress a random t-bit string to $< t$ bits in expectation.

Proof of main lemma.

Suppose for some i, $\mathbb{E}[|P \cap C_i|] < 0.01$.

Let $b_{i,0}, b_{i,1}, \ldots, b_{i,\beta^i-1}$ be the values the β^i updates have. $(A[j \cdot \frac{n}{\beta^i}] \leftarrow b_{i,j})$.

Fix $b_{i',j}$ for $i' \neq i$.



Let R_j be a uniformly random index in the range $[j \cdot \frac{n}{\beta^i}, (j+1)\frac{n}{\beta^i})$. Let $P(R_j)$ be the set of cells probed when $Q = R_j$.

$$\mathbb{E}[|P \cap C_i|] = \frac{1}{\beta^i} \cdot \sum_{j=0}^{\beta^i-1} \mathbb{E}[|P(R_j) \cap C_i|]$$

by Markov, $\sum_j |P(R_j) \cap C_i| < 0.1\beta^i$ with prob ≥ 0.9.

We describe a compression scheme below that would compress $b_{i,0}, \ldots, b_{i,\beta^i-1}$ to $< \beta^i$ bits.

"Compress":

- Let $S = \{j : |P(R_j) \cap C_i| \neq 0\}$.
- if $|S| > 0.1\beta^i$, write down "0", then write down $b_{i,0}, \ldots, b_{i,\beta^i-1}$.

- else, write down "1"
  - write down $C_{i-1}, C_{i-2}, \dots, C_1$ (both addr & content)
  - write down $|S|$, and S using $\lceil \log \binom{\beta^i}{|S|} \rceil$ bits
  - for each $j \in S$, write down $b_{i,j}$.

total expected # bits

$$1 + \frac{1}{10} \cdot \beta^i + \frac{9}{10} \left(\sum_{j<i} |C_j| \cdot O(w) + \log n + \log \binom{\beta^i}{0.1\beta^i} + |S| \right)$$

$$\left(\binom{n}{k} \le \left(\frac{en}{k}\right)^k \right)$$

$$\le 1 + \frac{1}{10} \beta^i + \frac{9}{10} \left(O(\beta^{i-1} \cdot t_u \cdot w) + \log n + 0.1\beta^i \cdot \log(10e) + 0.1\beta^i \right)$$

$$\le 1 + \frac{1}{10} \beta^i + \frac{9}{10} \left(0.6\beta^i + O(\beta^{i-1} \cdot \log^2 n) \right) \qquad (\beta = \log^4 n)$$

$$\le 0.7\beta^i.$$

This is a lossless compression, proved by the following decompression algo.

"decompress":

- if first bit $=0$, trivially recover all bit.
- else, simulate the update algo on Epochs $k, k-1, \dots, i+1$.

  read $C_{i-1}, \dots, C_1$

  read S.

- for $j = 1, \dots, \beta^i$,

  if $j \in S$, read $b_{i,j}$

  if $j \notin S$, $P(R_j)$ doesn't probe any cell in C_i,

  hence, we can simulate the query algo on

  R_j to recover $A[0] + \cdots + A[R_j - 1]$.

  We also know $b_{i',j'}$ for $i' \neq i$,

  this determines $b_{i,0} + \cdots + b_{i,j}$.

  since $b_{i,0}, \dots, b_{i,j-1}$ have been recovered.

  This determines $b_{i,j}$. $\square$