Stronger 3SUM-Indexing Lower Bounds

[Larsen, Chung] SODA23'

Aviel Zecharia

Authors

Eldon Chung

Kasper Green Larsen

Agenda

- Introduction & History
- Reachability Oracles in the Butterfly Graph
- Blocked Lopsided Set Disjointness
- Non-Adaptive DS
- Non-Adaptive 2-Bit-Probe DS

1.

Introduction & History

3SUM Problem

- Input: $A, B, C \subseteq G(+)$, $|G| \sim poly(n)$
  - $|A|=|B|=|C|=n$
- Goal: determine whether there is a triple $(a, b, c) \in A \times B \times C$ s.t. $a + b = c$

Known Results

- Trivial: O(n) Space, O(n^2) time
- Best: O($n^2(loglogn)^{O(1)}/log^2 n$) time
- Conjecture: no solution in O($n^{2-\Omega(1)}$) time

3SUM-Indexing

- Input: $A_1, A_2 \subseteq$ G(+) each of size n
- Pre-Process: S memory cells of w bits
- Query: element z
  - $a_1 + a_2 = z$
  - accessing at most T memory cells

- Adaptive vs. Non-Adaptive

Conjectures

- ✗ <u>Conj1:</u> T=O(1) -> S= $\tilde{\Omega}(n^2)$
  - 😋
- ✗ <u>Conj2:</u> S*T= $\tilde{\Omega}(n^2)$
- ✗ <u>Conj3:</u> T= O($n^{1-\delta}$) -> S= $\tilde{\Omega}(n^2)$
  - 😭

The Strong 3SUM-INDEXING Conjecture is False

- ✗ Inversion functions "hammer" [Fiat, Naor]
- ✗ Solving KSUM-Indexing ($0 < \delta < 1$)
  - Space: $O(n^{k-1-\delta/3})$ words
  - Query Time: $O(n^{\delta})$
- ✗ $K=3$, $\delta=0.75$ -> $S= O(n^{1.75})$, $T= O(n^{0.75})$
- ✗ Adaptive

Golovnev et al. STOC20 `

- Theorem1: 3SUM-Indexing query time $\Omega(\, log n / \log(^{Sw}/_{n}) \,))$
  - T=O(1) -> S= $\Omega(n^{1+\Omega(1)})$
- Abelian group of size O(n^2)
- Non-Adaptive, cell-probe DS

This Paper

- ✗ Adaptive DS time lower bound
  - $T = \Omega(logn/\log(Sw/n))$
  - Also for small universe
- ✗ Tighter non-adaptive time lower bound
  - $T = \Omega(\min\left\{log|G|/\log\left(\frac{Sw}{n}\right), \mathrm{n/w}\right\})$
- ✗ Non-Adaptive 2bit-probe space lower bound
  - $S = \Omega(|G|)$

2.

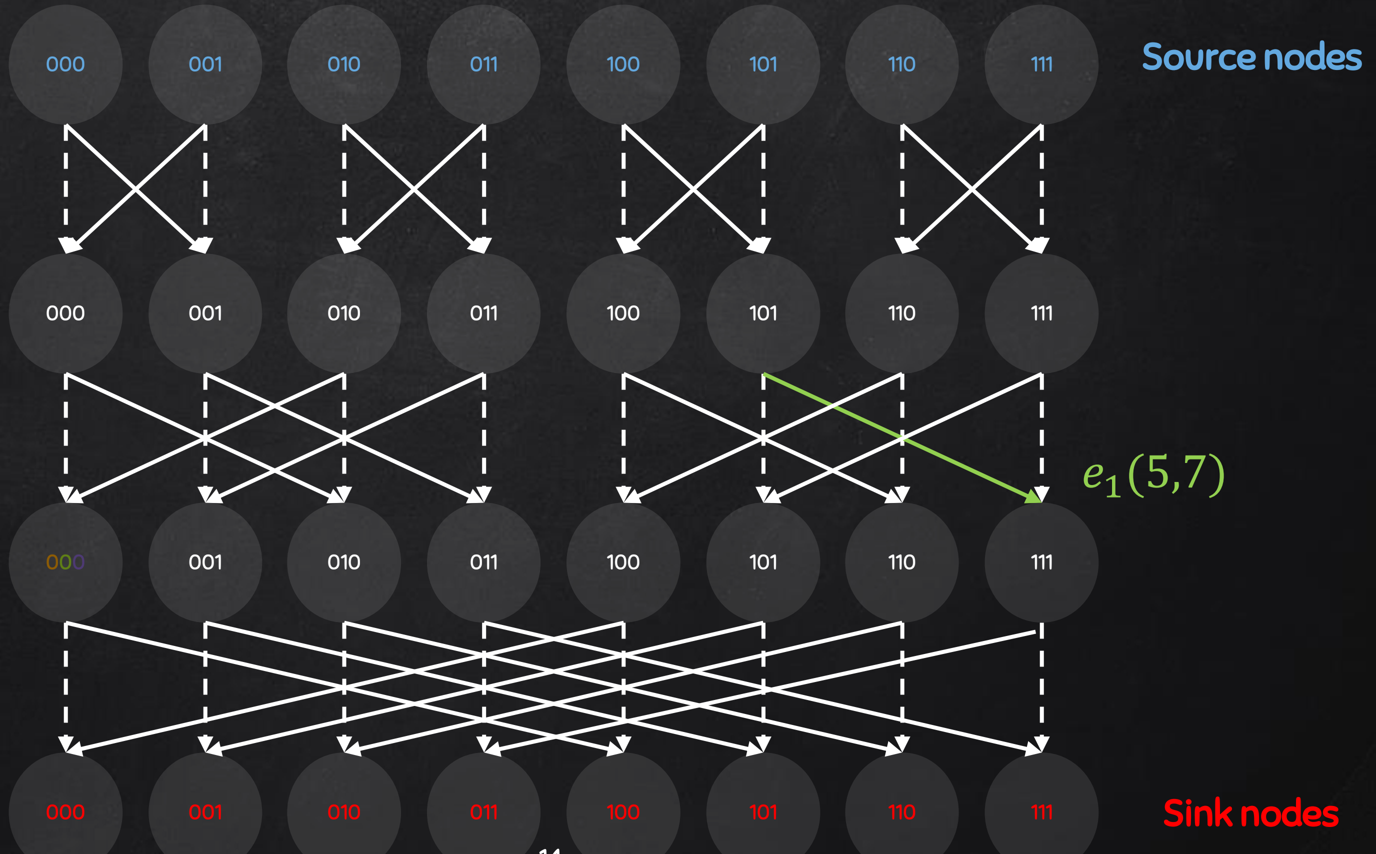
Reachability Oracles in the Butterfly Graph

Theorem2

- 3SUM-Indexing adaptive cell-probe DS
  - $|G| = O(n^2)$
  - $w = \Omega(logn)$
- Query time $T = \Omega(logn/\log(Sw/n))$

Butterfly Graph

$BG_{B=2,d=3}$



Layer2

$v_0[2], v_0[1], v_0[0]$

Reachability Oracles

- Input: $BG_{B,d}$ (E, V), E\`$\subseteq$ E
- Pre-Process: E\` into S memory cells of size w
- Query: is there exists a path $< s, t >$ in $BG_{B,d}$ (E\`, V)

Reduction Internals

- From E', s, t to A_1, A_2, z
- A_1: $\langle\ k\ |e_k(i,j) \in E`|v_i[d-1], .., v_i[k], 0, \dots, 0|0, \dots, 0, v_j[k], .., v_j[0]|0, 0\rangle$
- A_2: $\langle -k|\quad 0\quad |0,\quad \dots\quad , 0, X, \dots, X|X, \dots, X, 0,\quad \dots\quad , 0|X, X\rangle$
- $z_{s,t}$: $\langle\ 0\ |\quad 0\quad |v_s[d-1],\quad \dots\quad , v_s[0]|v_t[d-1],\ \dots\ , v_t[0]|X, X\rangle$

- $|A_1| = |E| = dB^{d+1} = |A_2| = n$
- $|G| \sim 3d(2B+1)^{2d+2} \leq \left(dB^{d+1}\right)^2 \leq n^2$

Reduction Analysis

- <u>Lemma1:</u> $BG_{B,d}, N = dB^d$, using S words of w bits
  - $B = \Omega(w^2), \log B = \Omega(\log(^{Sd}/_{N}))$
  - $T = \Omega(d)$
- <u>Theorem2:</u> 3SUM-Indexing, $|G| = O(n^2)$
  - $B = {}^{Sw^2}/_{n}$
  - $T = \Omega(logn/logB) = \Omega(logn/\log(^{Sw}/_{n}))$

3.

Blocked Lopsided Set Disjointness

Theorem3

- 3SUM-Indexing adaptive cell-probe DS
  - $|G| = O(n^{1+\delta})$
  - $w = \Omega(logn)$
- Query time $T = \Omega(logn/\log(Sw/n))$

Blocked LSD

- **Alice:** $X \subseteq [N]x[B]$, contains exactly one element for every 1,…,N
- **Bob:** $Y \subseteq [N]x[B]$
- **Goal:** determine whether $X \cap Y = \emptyset$
  - Minimizing communication complexity

$Alice\ X$		1					1		1							1			1	
$Bob\ Y$	1	0	0	1	0	1	0	0	0	0	0	0	0	1	0	1	1	1	1	1

$[N]x[B]$

Reduction Internals #1

Alice

$F(X) = C_1, \dots, C_{N/l}$

Bob

$G(Y) = DS(A, B)$

First cell to probe $\xrightarrow{\log \binom{S}{N/l}}$

$\xleftarrow{N/_l \, * \, w}$ Cell data

…

T Times

Reduction Internals #2

- ✗ Y $\rightarrow$ A $\langle\, i \,|0, \dots, j, \dots, 0 \rangle$ $\quad |A| \sim NB$
- ✗ Y $\rightarrow$ B $\langle\, 0 \,|\, *, \dots, 0, \dots, * \rangle$ $\quad |B| \sim B^l$
- ✗ X $\rightarrow$ C_i $\langle\, i \,|j_1, \quad \dots \quad, j_l \rangle$ $\quad |C_i| \sim l$

Alice X		1					1		1							1			1	
Bob Y	1	0	0	1	0	1	0	0	0	0	0	0	0	1	0	1	1	1	1	1

$l = 2$

$[N]x[B]$

Reduction Analysis [Bob]

- <u>Lemma2:</u> Blocked LSD communication
  - Alice $\alpha N log B$ or Bob $NB^{1-O(\alpha)}$
- <u>Reminder:</u> Bob sends $\mathrm{T} * {}^{N}/_{l} * w$ bits
  - $B = w^4, \alpha = 0.5, l \geq 1, w = O(logn)$
  - …
  - $T = \Omega(logn)$

Reduction Analysis [Alice]

- Lemma2: Blocked LSD communication
  - Alice $\alpha N log B$ or Bob $NB^{1-O(\alpha)}$
- Reminder: Alice sends $\mathrm{T} log \binom{S}{N/l}$ bits
  - $B = w^4, \alpha = 0.5, l = \varepsilon \frac{logn}{logw} \geq 1, n = NB$
  - …
  - $T = \Omega(logn / \log(Sw/n))$

Reduction Analysis [Universe]

- We must ensure
  - $|\mathrm{G}| = \mathrm{O}\left(n^{1+\delta}\right)$
  - $|A|, |B|, |C| \leq n$
- $|\mathrm{G}| \sim N * B^l, |B| \sim B^l \rightarrow B^l \leq n^\delta$
  - Choose small enough ε

4.

Non-Adaptive DS

Theorem4

- 3SUM-Indexing non-adaptive cell-probe DS
  - $|G| = \omega(n^2)$
  - $w = \Omega(logn)$
- Query time T = $\Omega(\min\left\{log|G|/\log\left(\frac{Sw}{n}\right), \mathrm{n/w}\right\})$

Counting

- Consider subsets of $\Delta = n/2w$ memory cells
- There is a set of Δ memory cells answering at least $|G|\binom{S-T}{\Delta-T}/\binom{S}{\Delta} \geq |G|(\frac{\Delta-T}{S})^T$ queries

Counting

- ✗ $\geq |G|(\frac{\Delta - T}{s})^T$ queries
- ✗ $T > \Delta/2 \rightarrow \Omega(n/w)$
- ✗ $T \leq \Delta/2 \rightarrow |G|^{1-o(1)} > n \rightarrow$ special set $Q, |Q| = n$

Counting $[log|G|/\log(Sw/n)]$

- ✗ Can subset of $\Delta = n/2w$ cells answer more than n queries?
  - Yes, but always?
- ✗ <u>Lemma3:</u> there exists an input distribution over inputs A_1, A_2 s.t. all events in Q are fully independent
  - $\forall q \in Q: Pr_{(A_1,A_2)\sim D}[q \in (A_1 + A_2)] = 1/2$
- ✗ n/2 bits vs. n bits of entropy

Lemma3

- Lemma3: $|G| = \omega(n^2)$. For $\mathrm{Q} \subseteq \mathrm{G}, |\mathrm{Q}| \leq \mathrm{n}$ there exists an input distribution A_1, A_2 s.t. all events are fully independent
  - $\forall q \in Q: Pr_{(A_1, A_2) \sim D}[q \in (A_1 + A_2)] = 1/2$
- $D = \{\forall P \subseteq Q: (A_1^P, A_2^P) | P \subseteq A_1^P + A_2^P, (Q \backslash \mathrm{P}) \cap (A_1^P + A_2^P) = \emptyset\}$
  - $|G| = \omega(n^2)$ limitation
- Given set of queries $S \subseteq Q, |S| = r$, how many sets P there are which satisfies S? 2^{n-r}
  - $|D| = 2^n$

5.

Non-Adaptive 2-Bit-Probe DS

Theorem5

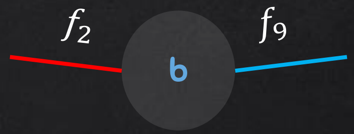
- 3SUM-Indexing non-adaptive cell-probe DS
  - $T = 2$
  - $w = 1$
- Space Complexity $S = \Omega(|G|)$

2-Bit-Probe Graph

- Build cell-probe graph
  - Nodes: one for each memory cell → $o(|G|)$
  - Edges: one for each query g → $|G|$
    - Logic functions f_g: COPY, CONST, AND, XOR
- There is at least one function of $\Omega(|G|)$ single edges
- Lemma4: for a graph with n nodes, average degree > 2 and girth r: $n \geq 2(d-2)^{r/2-2}$
  - $r = O(logn)$

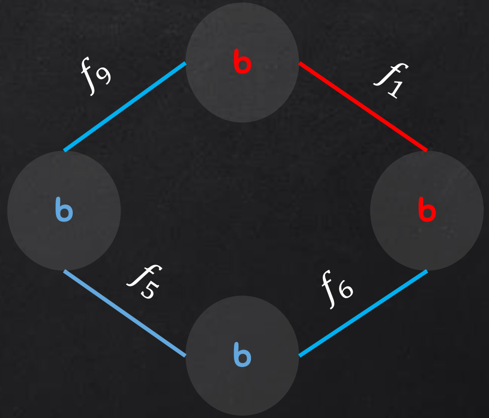
COPY Type

- ✗ $\Omega(|G|)$ single COPY edges, $o(|G|)$ nodes
    - There is a dominant node with $\omega(1)$ COPY edges
- ✗ By Lemma3, more than 2 yields a contradiction



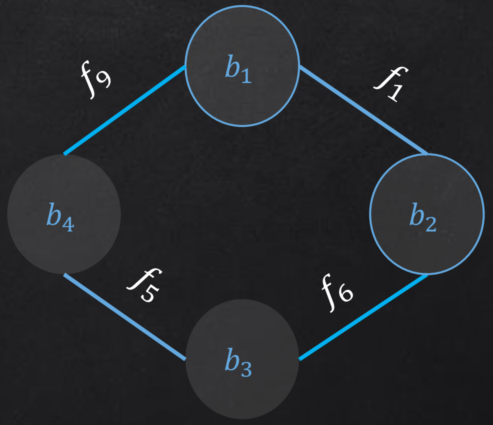
AND Type

- By Lemma4, there is a cycle of $O(logn)$ length using only AND edges
- There is a query f_1 which force the inputs to be fixed



XOR Type

- By Lemma4, there is a cycle of $O(logn)$ length using only XOR edges
- $f_5 = f_6 \oplus f_1 \oplus f_9$



Conclusion

- Using Hammers !
  - Non-trivial reductions
  - Inversion Functions
- 3SUM-Indexing as a static DS lower bound
  - Direct impact on various core algorithms & DS
- A combinatorial approach
- Can we make things better?
  - More efficient Block LSD reduction
  - Using the access graph method (T-bit-probe)

THANKS!

Any questions?

Link to the paper: https://arxiv.org/abs/2203.09334