Skip lists (Pugh 1989)

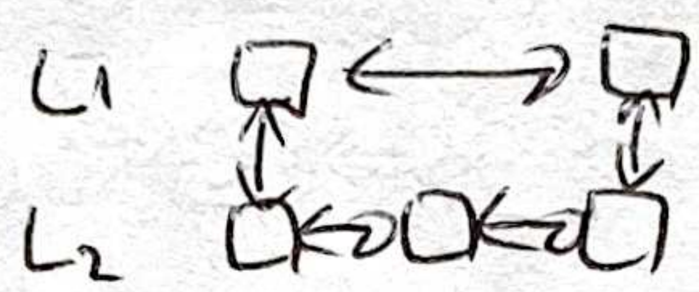
- dynamic search structure
  efficient, randomized, simple
- others: treaps, red-black trees, B trees } all balanced
- $O(\lg n)$ in expectation, with high probability ($\approx 1 - \frac{1}{n^\alpha}$)

Starting from scratch

(sorted) linked list

$\Theta(n)$ worst case search

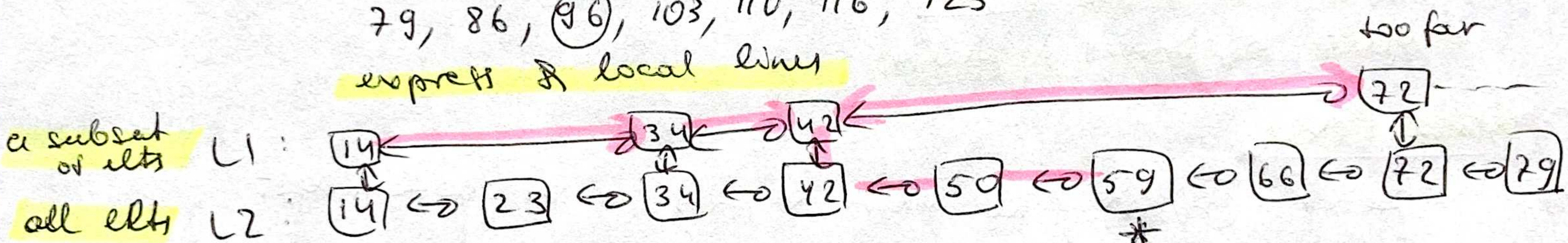
2 (sorted) linked lists



Example

(14), 23, (34), (42), 50, 59, 66, (72)
79, 86, (96), 103, 110, 116, 125

express & local lines



links between equal keys in L1 and L2

Search (x):

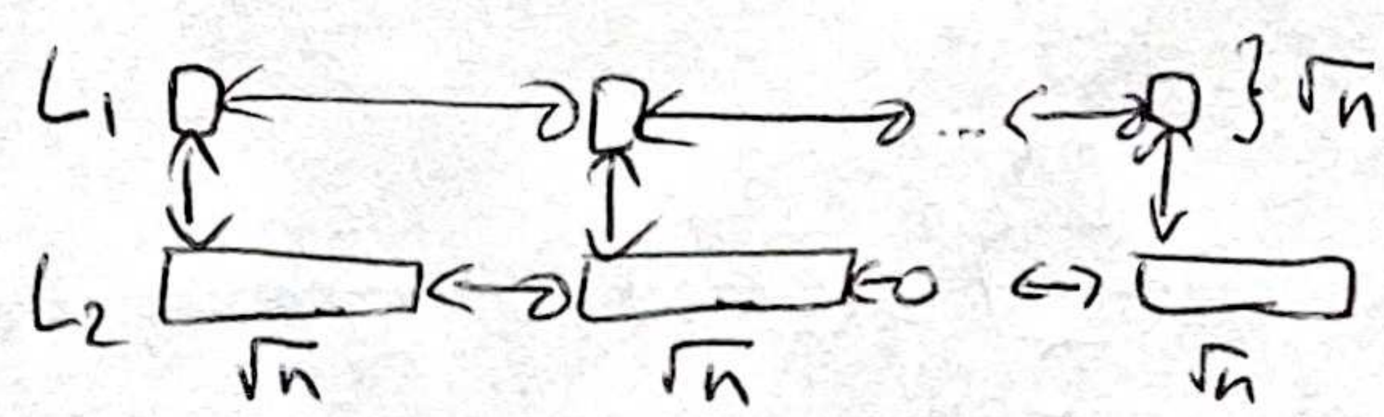
- walk right in top list L_1 until going right would go too far.
- walk down to L_2
- walk right in L_2 until find x (or > x)
  ↖ for insertion

What keys go in L_1?

- best is to spread them out uniformly

$\Rightarrow$ cost of search $\approx \underbrace{|L_1|}_{\text{in } L_1} + \underbrace{\frac{|L_2| \to n}{|L_1|}}_{\text{in } L_2}$

minimize $|L_1| + \dfrac{n}{|L_1|}$



up to const. factors

$|L_1| = \dfrac{n}{|L_1|}$ $\quad |L_1|^2 = n \quad |L_1| = \sqrt{n} \Rightarrow$ search cost $\approx 2\sqrt{n}$

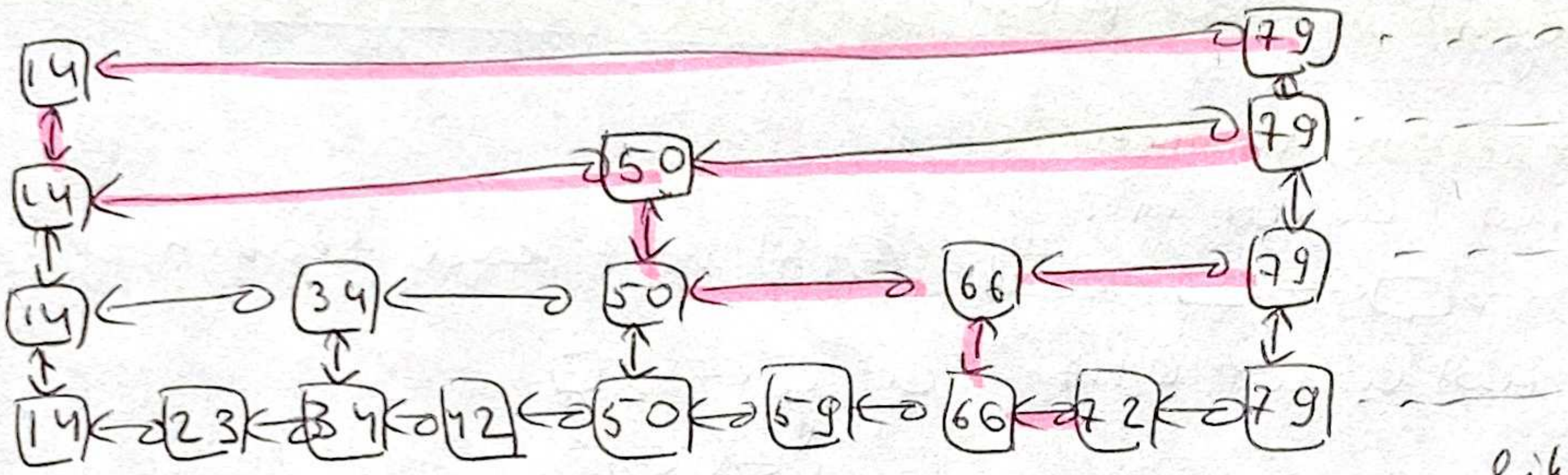
2 sorted linked lists : $2\sqrt{n}$

3 sorted linked lists : $3\sqrt[3]{n}$

k ------------ $k\sqrt[k]{n}$

lg n ------------ $\lg n \sqrt[\lg n]{n} = 2\lg n$

$'' \; n^{\frac{1}{\lg n}} = 2^{\frac{\lg n}{\lg n}} = 2$



ideal skip list

*
search (72)
O(lg n)

like binary tree

Skip list maintains roughly
subject to Insert & Delete

Insert (x)

- search (x) to find where x fits in the bottom list
- insert (x) in bottom list
  Invariant: bottom list is sorted and stores all elts.
- which other lists should store x ?
  - flip a fair coin
    heads: promote x to next level up
    flip again

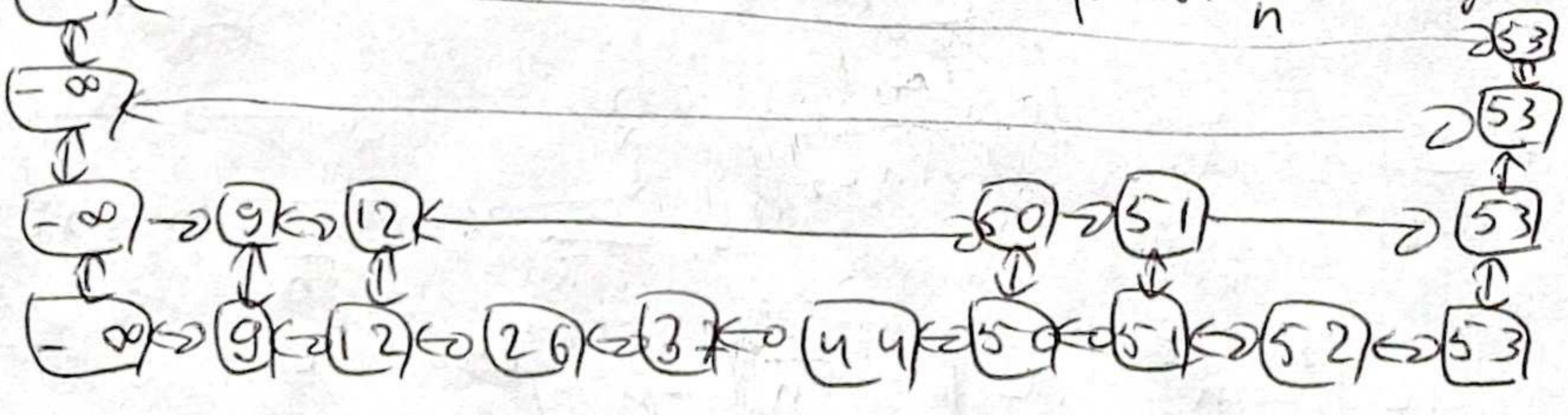
  } assuming lg n lists

  ↳ in expectation, maintains the ratio of elts across lists

- store $-\infty$ in every list

construction ex:

prob. $\frac{1}{n}$ to go up n levels $= \left(\frac{1}{2}\right)^{\lg n}$



fair coin

Delete (x) : find (x) and delete all the way up

Theorem: with high probability every search in n-element skip list costs $O(\lg n)$

e.g. for every search, search cost is $\leq 100 \lg n$ with probability $\geq 1 - \frac{1}{n^{99}}$

With high probability (w.h.p.)

Event E occurs w.h.p. if

my comment

for any $\alpha \geq 1$, $\exists$ choice of constants (for the bound in E description)

s.t. E occurs with probability $\geq 1 - \underbrace{O\left(\frac{1}{n^\alpha}\right)}_{\text{error probability}}$

Boole's inequality / union bound

$$\Pr\{E_1 \cup E_2 \cup \dots \cup E_k\} \leq \Pr\{E_1\} + \Pr\{E_2\} + \dots + \Pr\{E_k\}$$

Lemma w.h.p. # levels $= O(\lg n)$

Proof: error probability of $\{\leq c \lg n \text{ levels}\} \leftarrow$ complement of $\{\# \text{ levels} = O(\lg n)\}$

$= \Pr\{> c \lg n \text{ levels}\}$

$\leq n \Pr\{x \text{ gets promoted} \geq c \lg n \text{ times}\}$

by union bound

$= n\left(\frac{1}{2}\right)^{c \lg n} = \frac{n}{n^c} = \frac{1}{n^{c-1}} = \frac{1}{n^\alpha}$ for $\alpha = c - 1$

e.g. $c \geq \alpha + 1$

my comment

c can be > 1 because of randomization.

my comment

$\uparrow$ = sum over geometric distr.

my comment

choose α, $\exists$ choice of consts. for $O(\lg n)$ s.t. $\Pr\{\# \text{ levels} = O(\lg n)\} \geq 1 - O\left(\frac{1}{n^\alpha}\right)$

error probability is polynomially small

but knowing a w.h.p. bound on # levels is not enough conditioned on $c \lg n$ times

$\rightarrow$ need a w.h.p. bound on search cost

α $\uparrow$ | c in E $\uparrow$ | max error prob. $\downarrow$

from 1

relation holds $\forall n \geq 1$, incl. n_0 for $O(\lg n)$, $O\left(\frac{1}{n^\alpha}\right)$

Cool idea: [redacted] analize search backwards (back to top left corner)

- search starts [ends] at node in bottom list
- at each node visited:
  - if node wasn't promoted higher (Tails) then go [came from] left
  - if promoted (Heads) then go [came from] up
- stop [start] at the root ($-\infty$)

Proof of theorem

\# up moves = (\# levels -1) $\leftarrow$ heads

\# up moves < \# levels $\leq c \lg n$ w.h.p. (Lemma)

$\Rightarrow$ w.h.p. \# moves $\leq$ \# coin flips till get $c \lg n$ Heads

$= O(\lg n)$ w.h.p. $\leftarrow$ claim

Claim \# coin flips till $c \lg n$ Heads $= O(\lg n)$ w.h.p.

Proof: [redacted] Let's say flip $10 \lg n$ coins

$$\Pr\{\leq c \lg n \text{ Heads}\}$$

$$\leq \binom{10\, c \lg n}{c \lg n} \underbrace{\left(\frac{1}{2}\right)^{9 c \lg n}}_{\text{tails}} \qquad \binom{y}{x} \leq \left(e\frac{y}{x}\right)^x$$

$$\leq \frac{\left(e \frac{10 \cancel{c \lg n}}{\cancel{c \lg n}}\right)^{c \lg n}}{2^{9 c \lg n}} = \frac{(e10)^{c \lg n}}{2^{9 c \lg n}} = \frac{2^{\lg(10e)\, c \lg n}}{2^{9 c \lg n}} =$$

$$= 2^{[\lg(10e) - 9]\, c \lg n}$$

$$= \frac{1}{2^{\underbrace{[9 - \lg(10e)]\, c}_{\alpha} \lg n}} = \frac{1}{n^{\alpha}}$$

$10-1$

as $10 \to \infty$

$9 - \lg(10e) \to \infty$