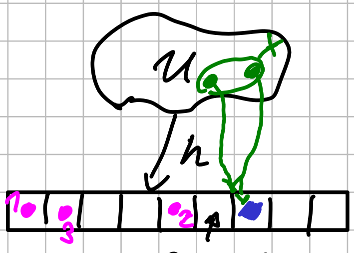
# Open Addressing




every slot contains max. 1 item

Insert(x): $i \leftarrow 0$
While $B[h(x,i)] \neq \emptyset$ : ← or tombstone
  $i \leftarrow i+1$
$B[h(x,i)] \leftarrow x$

This succeeds iff B is not full.

Could be very slow.

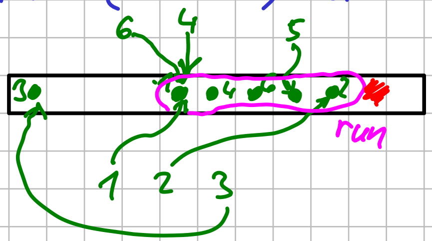
Generalize hash function $h(x)$ to probe sequence $h(x,0), \underbrace{h(x,1), \ldots, h(x,m-1)}_{\text{permutation of } [m]}$

primary location — $h(x,0)$; alt. location — $h(x,1), \ldots, h(x,m-1)$

Find(x): $i \leftarrow 0$
Loop:
  $j \leftarrow h(x,i)$
  if $B[j]=x$ : return TRUE
  if $B[j]=\emptyset$ : return FALSE
  $i \leftarrow i+1$
  if $i \geq m$ : return FALSE

## Hashing with Linear Probing

$h(x,i) := (h(x)+i) \bmod m$




Delete(x): problematic
replace item by tombstone
after some time rehash all items

Good news: (1) it's simple
(2) cache-friendly

Bad news: (1) Slow once long runs start forming

More good news: (3) this can be kept under control :)

load $:= \dfrac{\#\text{full buckets}}{\#\text{all buckets}} \in [0,1]$

**Claim:** Suppose that $m \geq (1+\varepsilon)\cdot n$.
Then $\mathbb{E}[\#\text{probes}]$ is:

(1) $\Theta(1/\varepsilon^2)$ if $h$ is completely random

(2) $O(1/\varepsilon^{13/6})$ for $h$ chosen from 5-indep. family

(3) $\Omega(\log n)$ for some 4-indep. family

(4) $\Omega(\sqrt{m})$ for some 2-indep. family

(5) $\Omega(\log n)$ for multiply-shift

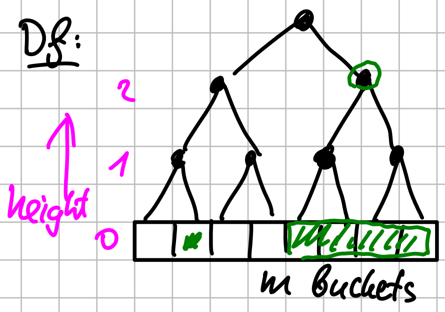
(6) $O(1/\varepsilon^2)$ for tabulation hashing

**Theorem:** Let $m$ be a power of two,
$n \leq m/3$
$h$ be a completely random hash function
$x$ be the item we search for.
Then $\mathbb{E}[\#\text{probes}] \in O(1)$.

**Proof:** W(much)LOG $n = \frac{1}{3}m$ ± rounding error

Call the items in the table $x_1$ — $x_n$.

Df:

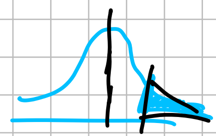



block $\equiv$ interval of $2^t$ buckets below an internal node of height $t$

a block of size $2^t$ is critical $\equiv$ # items hashed there $> \frac{2}{3} \cdot 2^t$

hashed vs. stored

4 stored in the structure



Tool: Chernoff bound for the right tail:

Let $X_1 — X_k$ be independent random variables with range $\{0,1\}$.

$X := \sum_i X_i$

$\mu := \mathbb{E}[X]$

$c > 1$

Then $\Pr[X > c \cdot \mu] \le \left(\frac{e^{c-1}}{c^c}\right)^\mu$.

$e = 2.71828\ldots$

Lemma: Let $B$ be a block of size $2^t$.
Then $\Pr[B \text{ is critical}] \le \left(\underbrace{\frac{e}{4}}_{<1}\right)^{2^t/3} = q^{2^t}$, where $q = \left(\frac{e}{4}\right)^{1/3} < 1$

Proof: Indicator random variables:

$$X_i := \begin{cases} 0 \\ 1 & \text{if } h(x_i) \in B. \end{cases}$$

# items hashed to $B$ = $X = \sum_i X_i$

Means: $\mathbb{E}[X_i] = \cancel{0 \cdot \Pr[X_i = 0]} + 1 \cdot \Pr[X_i = 1]$

$= \Pr[\underbrace{h(x_i) \in B}]$ independent events

$= \frac{2^t}{m}$

$\mu = \mathbb{E}[X] = \sum_i \mathbb{E}[X_i] = \frac{n \cdot 2^t}{m}$

We have $n = \frac{1}{3}m$, so $\mu = \frac{1}{3} 2^t$.

$\Pr[B \text{ is critical}]$

$= \Pr[X > \frac{2}{3} \cdot 2^t]$

$= \Pr[X > 2\mu]$ use Chernoff with $c=2$

$< \left(\frac{e^1}{2^2}\right)^\mu = \left(\frac{e}{4}\right)^{2^t/3}$. ✓

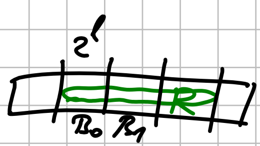
---

Df: Run: maximal consecutive set of full buckets


⇒ the run is preceded and followed by empty bucket

⇒ an item is hashed to a run $\Leftrightarrow$ it's stored in the run

Lemma: Let $R$ be a run of size at least $2^{\ell+2}$,
$B_0 — B_3$ be the first 4 blocks of size $2^\ell$ intersected by $R$.
Then at least one $B_i$ is critical.




Proof: $R$ intersects at least 4 blocks.

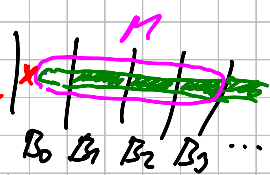
$|B_0 \cap R| \ge 1$

$|B_1 \cap R| = 2^\ell$

4 the same for $B_2, B_3$

$M := R \cap (B_0 \cup B_1 \cup B_2 \cup B_3)$

$|M| \ge 3 \cdot 2^\ell + 1$

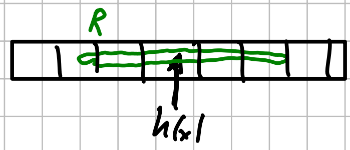



what is stored $M$, was hashed there

If no $B_i$ is critical: $|M| \le$ # items hashed to $B_0 — B_3 \le \frac{2}{3} 2^\ell \cdot 4 = \frac{8}{3} 2^\ell < 3 \cdot 2^\ell$ ↯

**Lemma:** Let $R$ be the run containing $h(x)$ and $|R| \in [2^{\ell+2}, 2^{\ell+3})$.

Then at least 1 of these 12 blocks (of size $2^\ell$) is critical:




- the block containing $h(x)$
- 8 blocks before
- 3 blocks after

**Proof:** $|R|$ is between $4 \cdot 2^\ell$ and $8 \cdot 2^\ell$ $\Rightarrow$ $R$ intersects at most 9 blocks

$\rightarrow$ start of $R$ is at most 8 blocks before $h(x)$

& apply the previous lemma.

**Corollary:** Let $R$ be the run containing $h(x)$.

Then $\Pr\left[\, |R| \in [2^{\ell+2}, 2^{\ell+3}) \,\right] \leq 12 \cdot q^{2^\ell}$.

$\uparrow$ using $\Pr[A \cup B] \leq \Pr[A] + \Pr[B]$

**Finale:**

$$\underbrace{\mathbb{E}[|R|]}_{\text{run containing } h(x)} \overset{\text{def.}}{=} \sum_k k \cdot \Pr[|R| = k] = \underbrace{\left(\sum_{k<3} k \cdot \Pr[\ldots]\right)}_{\in O(1)} + \sum_{\ell \geq 0} \underbrace{\sum_{k \in [2^{\ell+2}, 2^{\ell+3})} \overset{\leq 2^{\ell+3}}{k} \Pr[|R| = k]}_{2^{\ell+3} \cdot \underbrace{\sum_{k \in \text{Interval}} \Pr[|R| = k]}_{\Pr[|R| \in \text{Interval}] \leq 12 \cdot q^{2^\ell} \text{ Coroll.}}}$$

$$\leq \sum_{\ell \geq 0} 2^{\ell+3} \cdot 12 \cdot q^{2^\ell} = 8 \cdot 12 \cdot \underbrace{\sum_{\ell \geq 0} 2^\ell \cdot q^{2^\ell}}_{\leq \sum_{t \geq 0} t \cdot q^t \leq \text{const.}}$$

converges as infinite sum for every $q \in (0,1)$

So $\mathbb{E}[|R|] \leq$ some constant. $\blacksquare$