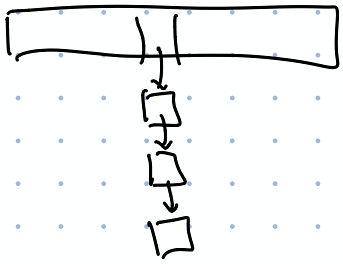
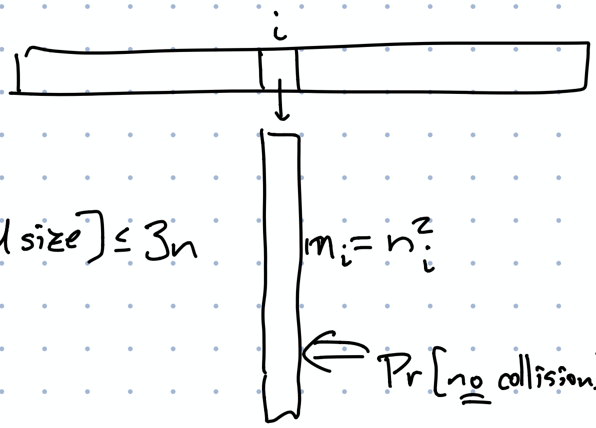
Hash Tables



Chaining

"Perfect hashing"



$m = n$

$E[\text{total size}] \le 3n$

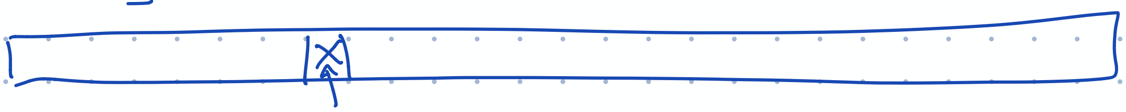
$m_i = n_i^2$

$\Pr[\underline{\text{no}} \text{ collisions}] > 1/2$

Lookup(x):

```
i ← h(x)
j ← h_i(x)
if T_i[j] = x
    return T
else
    return F
```

Open Addressing



Entire structure is an array of size $m \ge n$

Want $m = 2n$

Fiction: Uniform hashing

```
for i ← 0 to m-1
    if T[h_i(x)] empty
        T[h_i(x)] ← x
        return
return Full
```

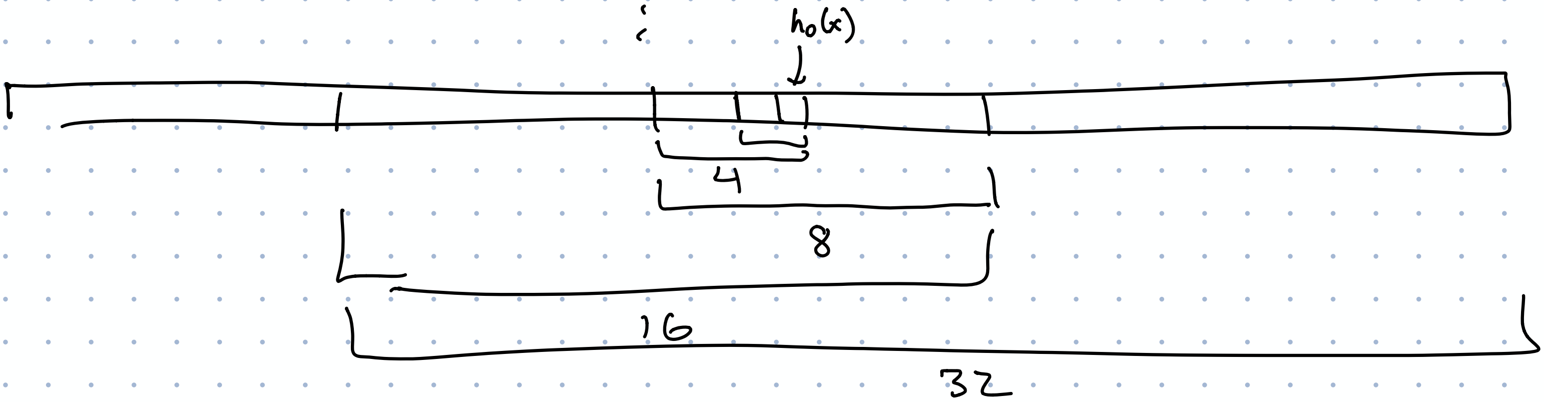
$\langle h_0(x), h_1(x), \ldots h_{m-1}(x) \rangle$ is a random permutation of $\{0, 1, \ldots m-1\}$

$$E[T(m,n)] \le 1 + \frac{n}{m} E[T(m-1, n-1)] \le \frac{m}{m-n} = 2$$

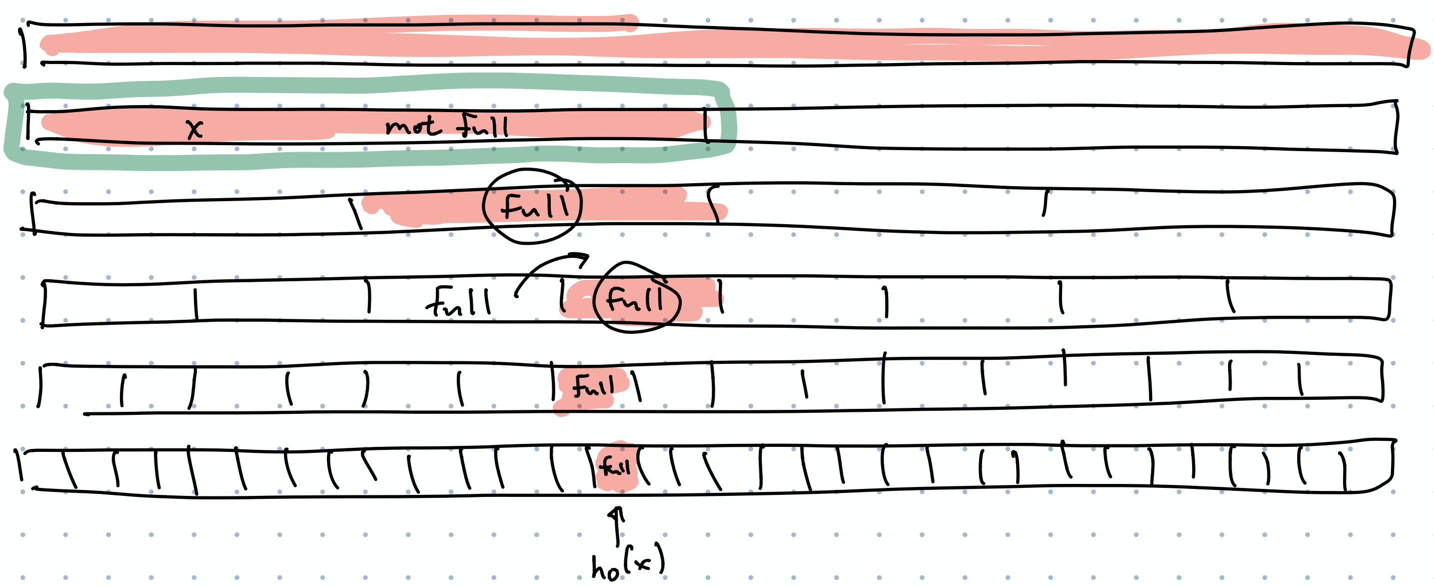
Linear probing: $h_i(x) = (h_0(x) + i) \bmod m$

Binary probing: $\boxed{h_i(x) = h_0(x) \oplus i}$ $\qquad m = 2^\ell$

1 0 0 1 0 1
1 0 0 1 0 0
1 0 0 1 1 1
1 0 0 1 1 0
1 0 0 0 0 1
⋮



Claim: If h_0 is drawn from a nice family of hash functions then $E[\#\text{probes}] = O(1)$



Insert(x):

for $i \leftarrow 0$ to $\lg m$
  if block of size 2^i containing $h_0(x)$ isn't full
    put x there, return

Time = O(size of biggest full block containing $h_0(x)$)

= O(1) whp. $\boxed{m = 4n}$

Block is popular if # items y s.t. $h_0(y)$ in block $\geq \frac{1}{2}$ block size

$$\Pr[\text{block of size } 2^k \text{ is full}] \leq \Pr[\text{block of size } 2^{k+1} \text{ is popular}]$$

$\Pr[\text{block of size } 2^{k+1} \text{ is popular}]$

$$E[\#\text{ items hash into that block}] = \frac{2^{k+1}}{4}$$ assuming h_0 is uniform

$$\Pr[\#\text{items} \geq 2E[\#\text{items}]]$$

Chebyshev: $\Pr[X > (1+\delta)E(x)] \leq O\left(\frac{1}{\delta^2 E(x)}\right) = O\left(\frac{1}{2^{k-1}}\right)$

$\delta = 1$

$E[x] = 2^{k-1}$

```
for i ← 0 to log n
    scan block of size 2^i
    if full (block of size 2^{i+1} is popular)
    else return
```

$$T(n,m) \leq \sum_{i=0}^{\lg n} 2^{i+1} \cdot \Pr[\text{block size } 2^{i+1} \text{ popular}]$$

$$\leq \sum_{i=0}^{\lg n} 2^{i+1} \cdot O\left(\frac{1}{2^{i-i}}\right) = \boxed{O(\log n)}$$

<u>4th moment:</u> $\Pr[X \geq (1+\delta)E[X]] \leq O\left(\frac{1}{\delta^4 E[X]^2}\right)$

$$= O\left(\frac{1}{4^{i-1}}\right)$$

$$\Rightarrow T(n,m) \leq \sum_{i=0}^{\lg n} 2^{i+1} \cdot O\left(\frac{1}{4^{i-1}}\right) = \sum_{i=0}^{\lg n} O\left(\frac{1}{2^i}\right) = O(1)$$

We need h_0 to come from a 5-uniform family

$$\Pr\{h(x_1)=i \wedge h(x_2)=i_2 \wedge h(x_3)=i_3 \wedge \cdots \wedge h(x_5)=i_5\} = \frac{1}{m^5}$$

$$h(x) = ((ax^4+bx^3+cx^2+dx+e) \bmod p) \bmod m$$

[Thorup Zhang 2010]:

<u>Tabulation</u> keys $= (x,y) \in 2^{w/2} \times 2^{w/2}$

$$h(x,y) = A[x] \oplus B[y] \oplus C[x+y]$$