CS & IT ENGINEERING

Data Structures

Hashing

Lecture No.- 01

By- Pankaj Sharma Sir

Recap of Previous Lecture

Topic Graphs

Topics to be Covered

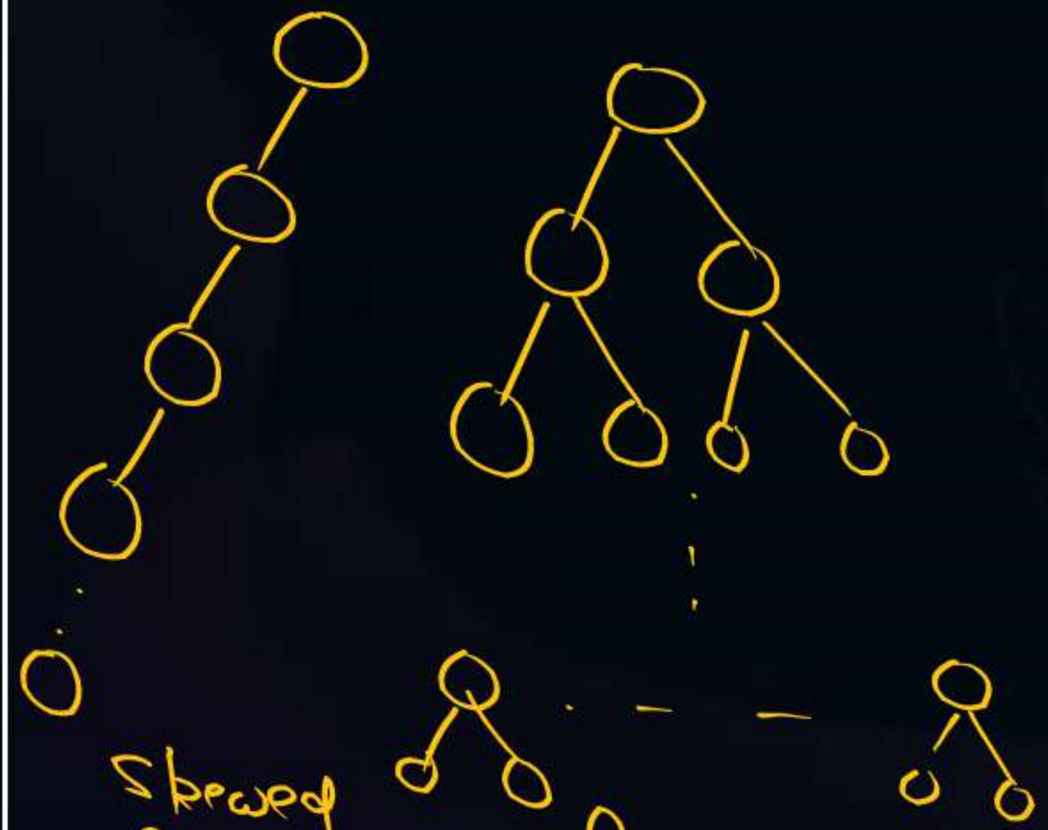
Topic Hashing Part 01

Topic : Hashing

$n = 2^{20}$

$\#comp = O(n) = 2^{20}$ → BST

case 1 case 2



Skewed
$h = O(n)$

$h = O(\log_2 n)$

$\#comp = O(h)$

Skewed $\Rightarrow O(n) = 2^{20}$

CBT, FBT $\Rightarrow O(\log_2 n)$

$\Rightarrow$ 20 comp

height balanced BST

AVL tree

$h = O(\log_2 n) \Rightarrow$ 20 comp

m-way $O(\log_m n)$

B-Tree B+ Tree

Order: m

order: 8

$\log_8 2^{20}$

$\Rightarrow 7$

चाहते क्या हैं

$$n \rightarrow \log_2 n \rightarrow \log_m n \Longrightarrow \underline{O(1)}$$

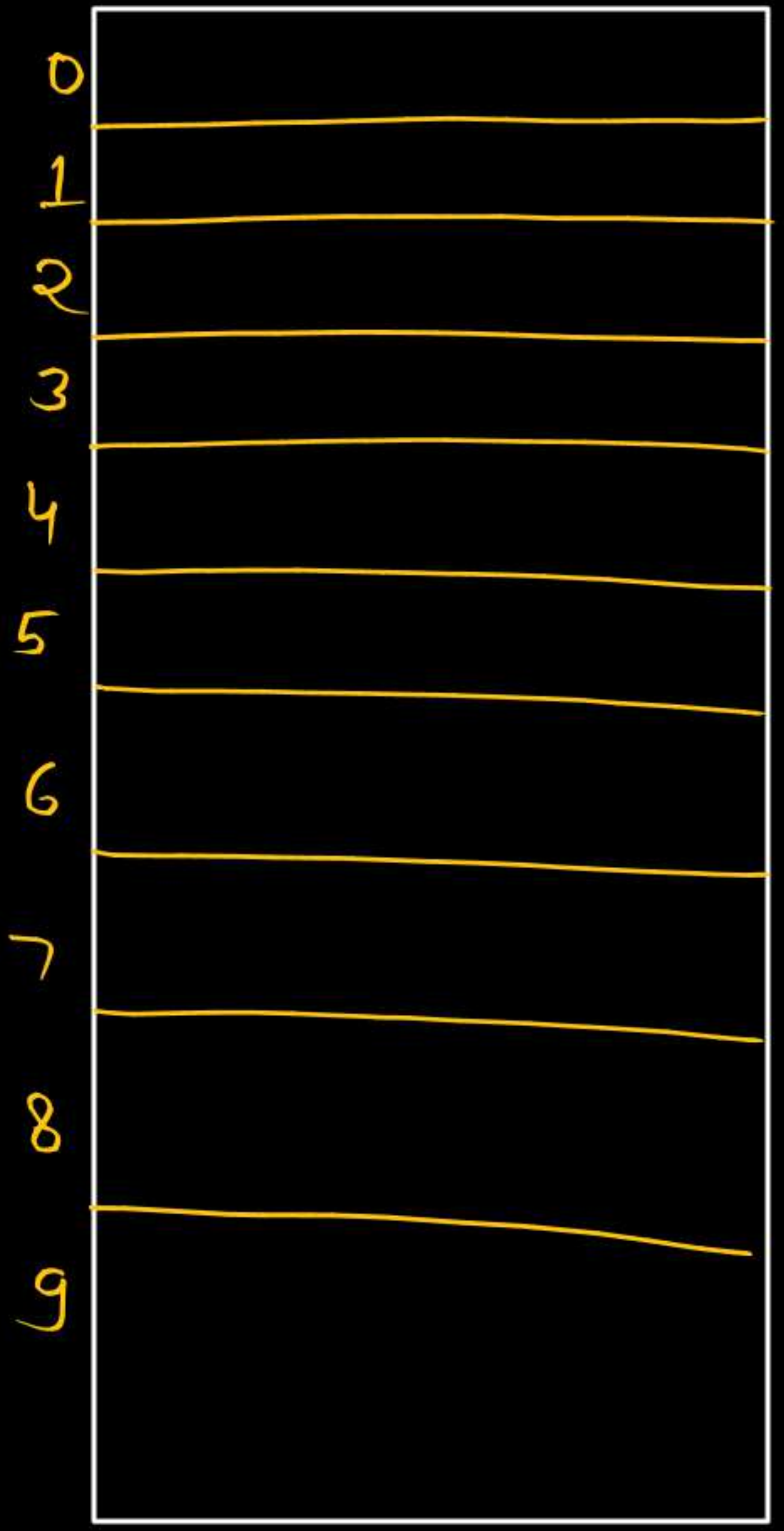
Keys: 12, 18, 15, 14, 13, 29, 31, 57

$m = 10$

Key → Hash function → locations/memory address $\{0, 1, 2, .. 9\}$

Key → Hash function → 0-9

$$\boxed{h(K) = K \bmod 10}$$



Keys: 12, 18, 15, 14, 13, 29, 31, 57

$m = 10$

$$\boxed{h(k) = k \bmod 10}$$

$h(12) = 2$

insertion $\rightarrow O(1)$

Search: $\rightarrow O(1)$

$29 \bmod 10$

$\Rightarrow 9$

0	
1	31
2	12
3	13
4	14
5	15
6	
7	57
8	18
9	29

Keys: 12, 23, 42, 83, 54, 31, 82

$m = 10$

$$\boxed{h(K) = K \bmod 10}$$

$h(12) = 2$

$h(42) = 2$

$$\boxed{\begin{matrix} h(k_1) \rightarrow L_1 \\ h(k_2) \rightarrow L_1 \end{matrix}}$$

Collission

0	
1	
2	12
3	23
4	
5	
6	
7	
8	
9	

Collision

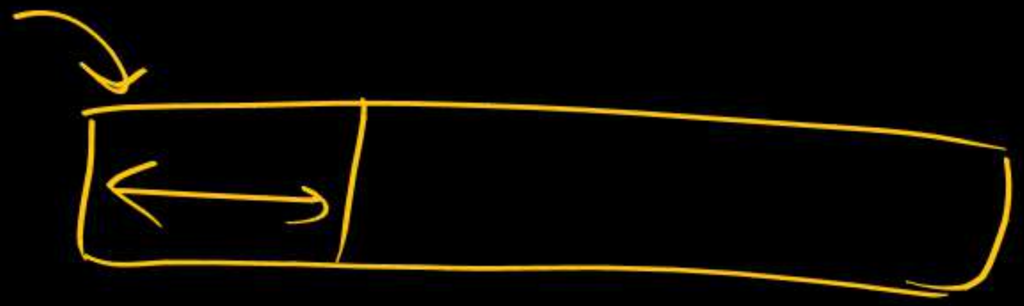
Collission resolution Tech

1. Linear Probing
2. Quadratic Probing
3. Double hashing
4. Chaining

Good hash function

(i) Easy to compute

(ii) Uniformly distribute



Hash function

$h(k) = k \bmod m$ m: Table size $(0, 1, 2, \dots m-1)$

$h(k) = k \bmod m + 1$ $(1, 2, 3, \dots m)$

Linear Probing

let $h(k) = k \bmod m$ is the hash function.

→ results in a collission for key k_1

$$h(k_1) = L_1$$

↓

CR function

$$H(k, i) = (h(k) + i) \bmod m$$

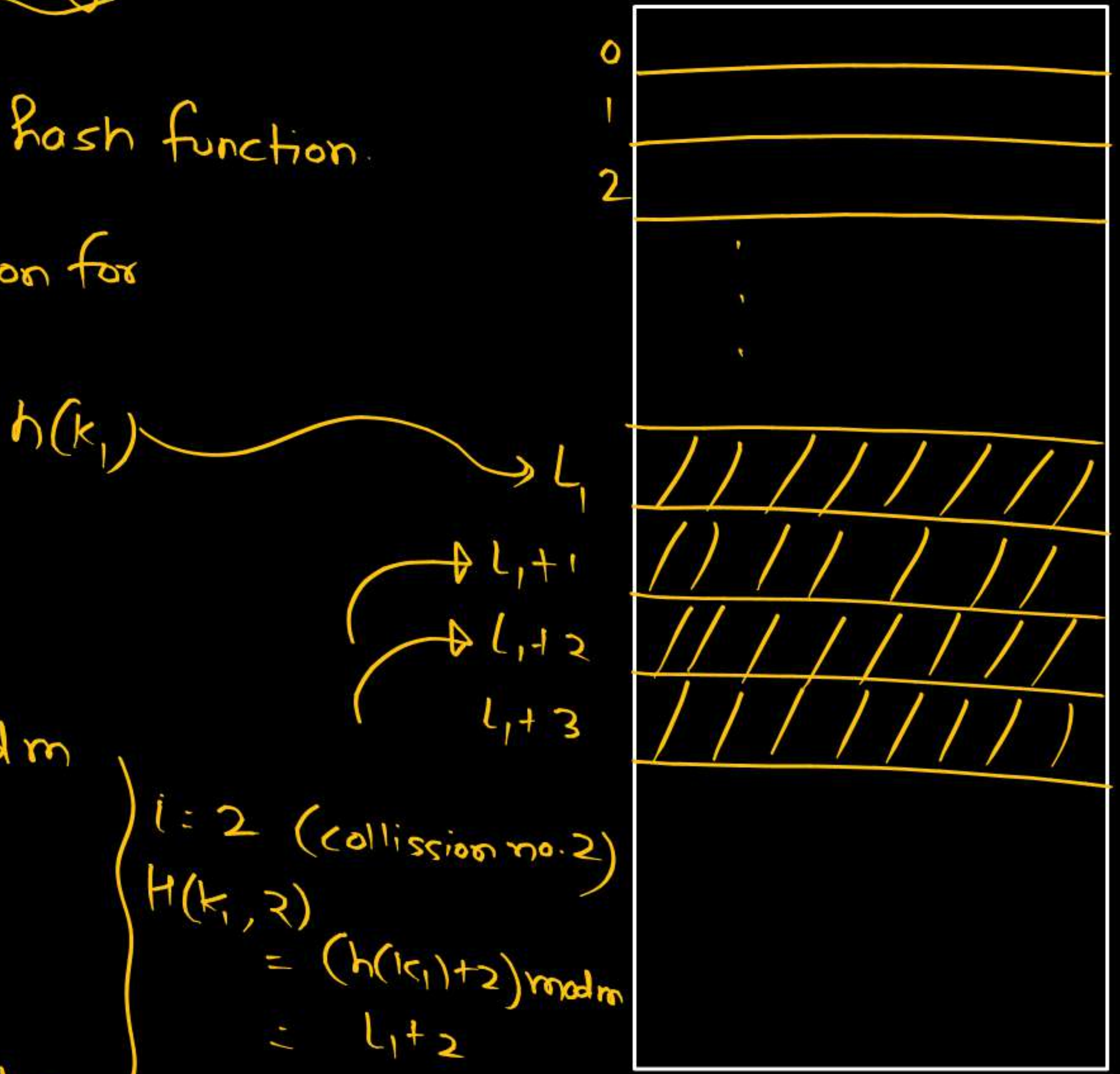
↓ Collission no for key k

$\boxed{i=1}$

$$H(k_1, 1) = (h(k_1) + 1) \bmod m$$
$$= (L_1 + 1) \bmod m$$
$$= L_1 + 1$$

$i = 2$ (collission no. 2)

$$H(k_1, 2) = (h(k_1) + 2) \bmod m = L_1 + 2$$



$m = 10$ $\quad 0,1,2,3,4,5,6,7,8,9$

$h(k) = k \bmod 10$

$h(k_1)$

let $h(k_1) = 4$

$\Rightarrow$ Collission occurs

$H(k_1, 1) = (h(k_1) + 1) = 4+1 = 5$

$i=2 \quad H(k_1, 2) = (h(k_1) + 2) = 4+2 = 6$

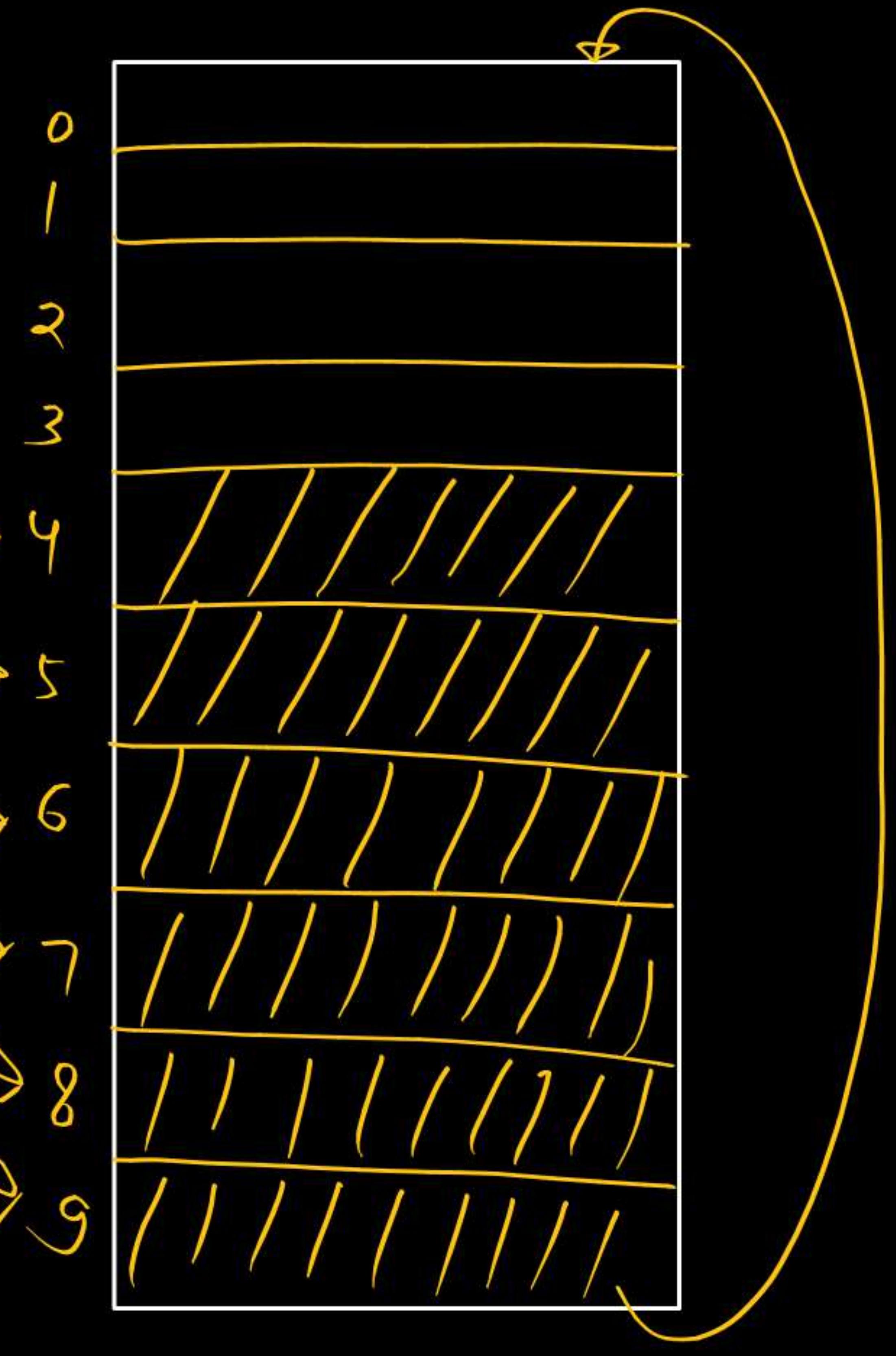
$i=3 \quad H(k_1, 3) = (h(k_1) + 3) = 4+3 = 7$

$i=4 \quad H(k_1, 4) = (h(k_1) + 4) = 8$

$i=5 \quad H(k_1, 5) = (h(k_1) + 5) = 9$

$i=6 \quad H(k_1, 6) = (h(k_1) + 6) = 10$

Why mod



$h(k) = k \bmod m$

Keys: 31, 26, 43, 27, 34, 46, 14, 58, 13

$m = 12$

$h(31) = 31 \bmod 12 = 7$

$h(26) = 26 \bmod 12 = 2$

$h(43) = 43 \bmod 12 = 7$ coll.

$H(k, i) = (h(k) + i) \bmod m$

$H(43,1) = (7+1) \bmod 12 = 8$ ✓

$h(27) = 27 \bmod 12 = 3$ ✓

$h(34) = 34 \bmod 12 = 10$ ✓

$h(46) = 46 \bmod 12 = 10$ coll.

$H(46,1) = (h(46)+1) \bmod 12$

$= 11$

$h(14) = 14 \bmod 2 = 2$ coll.

$H(14, 1) = (h(14)+1) \bmod 12 = 3$ coll.

$H(14, 2) = (h(14)+2) \bmod 12 = 4$ ✓

$h(58) = 58 \bmod 12 = 10$ coll.

$H(58,1) = (h(58)+1) \bmod 12 = 11$ coll.

$H(58,2) = (h(58)+2) \bmod 12 = 0$ ✓

$h(13) = 13 \bmod 12 = 1$ ✓

0	58
1	13
2	26
3	27
4	14
5	
6	
7	31
8	43
9	
10	34
11	46

$h(K) = K \bmod m$

Keys : 31, 26, 43, 27, 34, 46, 14, 58, 13

7 2 ~~7~~ 8 3 10 ~~10~~ 11 ~~2~~ ~~3~~ 4 ~~10~~ ~~11~~ 0 1

$m = 12$

Primary clustering Problem : 34, 46, 58, 1, 26, 27, 14

0	58
1	1
2	26
3	27
4	14
5	
6	
7	31
8	43
9	
10	34
11	46

$h(K) = K \bmod m$

Keys: 31, 26, 43, 27, 34, 46, 14, 58, 13

7, 2, ~~7~~ 8, 3, 10, ~~10~~ 11, ~~2~~ ~~3~~ 4, ~~10~~ ~~11~~ 0, 1

$m = 12$

Primary clustering Problem:

Probability that a new key will get this slot

$Prob(5) = \frac{8}{12}$

$Prob(6) = \frac{1}{12}$ Prob.

$Prob(9) = \frac{3}{12}$ Prob.

0	58
1	1
2	26
3	27
4	14
5	
6	
7	31
8	43
9	
10	34
11	46

CRT

Open addressing

Separate chaining

1) Linear probing (Primary clustering)

2) Quadratic probing (free from primary clustering)

secondary clustering problem

3) Double hashing

Quadratic Probing

let $h(k) = k \bmod m$

$\hookrightarrow$ it leads to a collission

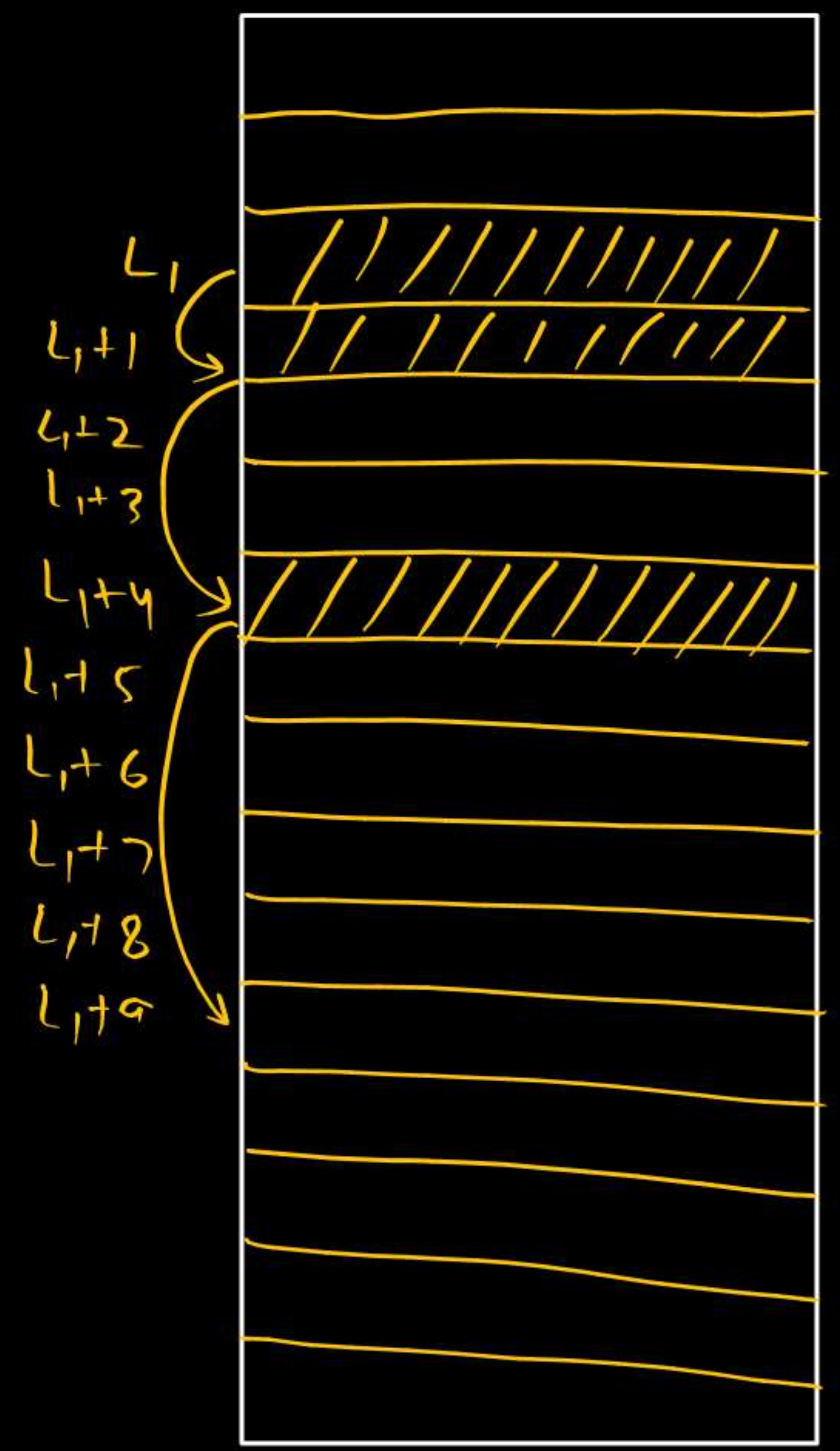
$h(k) = k \bmod m = L_1$

coll. no

$H(k, i) = (h(k) + i^2) \bmod m$

$H(k, 1) = (L_1 + 1) \bmod m = L_1 + 1$

$H(k, 2) = (L_1 + 2^2) \bmod m = L_1 + 4$



Quadratic Probing

Keys: 24, 17, 32, 2, 13, 50, 30, 61

$m = 11$ $\quad h(K) = K \bmod 11$

$h(24) = 2$

$h(17) = 6$

$h(32) = 32 \bmod 11$
$= 10$

$h(2) = 2$ coll.

$H(2,1) = (h(2) + 1^2) \bmod 11$
$= 3$

$h(13) = 13 \bmod 11 = 2$ coll.

$H(13,1) = (h(13) + 1^2) \bmod 11$
$= 3$ coll

$H(13,2) = (h(13) + 2^2) \bmod 11$
$= 6$ coll

$H(13,3) = (h(13) + 3^2) \bmod 11$
$= 0$ ✓

$h(50) = 6$ coll.

$H(50,1) = (h(50) + 1^2) \bmod 11$
$= 7$

$h(30) = 8$

$h(61) = 6$ coll.

$H(61,1) = (h(61) + 1^2) \bmod 11$
$= 7$ coll.

$H(61,2) = (h(61) + 2^2) \bmod 11$
$= (6 + 4) \bmod 11 = 10$ coll.

$H(61,3)$
$= (h(61) + 3^2) \bmod 11$
$= 4$

0	13
1	
2	24
3	2
4	61
5	
6	17
7	50
8	30
9	
10	32

$m = 11$

Quadratic Probing

24, 2, 13

$h(24) = 2$
$h(2) = 2$
$h(13) = 2$
$\Big\} \; i=1$

$H(24,1) = (h(24)+1^2) \bmod 11 = 3$
$H(2,1) = (h(2)+1^2) \bmod 11 = 3$
$H(13,1) = (h(13)+1^2) \bmod 11 = 3$

$i=2$

$H(24,2) = (h(24)+2^2) \bmod 11 = 6$
$H(2,2) = (h(2)+2^2) \bmod 11 = 6$
$H(13,2) = (h(13)+2^2) \bmod 11 = 6$

$i=3$

$H(24,3) = (h(24)+3^2) \bmod 11 = 0$
$H(2,3) = (h(2)+3^2) \bmod 11 = 0$
$H(13,3) = (h(13)+3^2) \bmod 11 = 0$

$i=4$

$H(24,4) = (h(24)+4^2) \bmod 11 = 7$
$H(2,4) = 7$
$H(13,4) = 7$

$i=5$

$H(24,5) = (h(24)+5^2) \bmod 11 = 5$
$H(2,5) = 5$
$H(13,5) = 5$

$i=6$

$H(24,6) = (h(24)+6^2) \bmod 11 = 5$
$H(2,6) = 5$
$H(13,6) = 5$

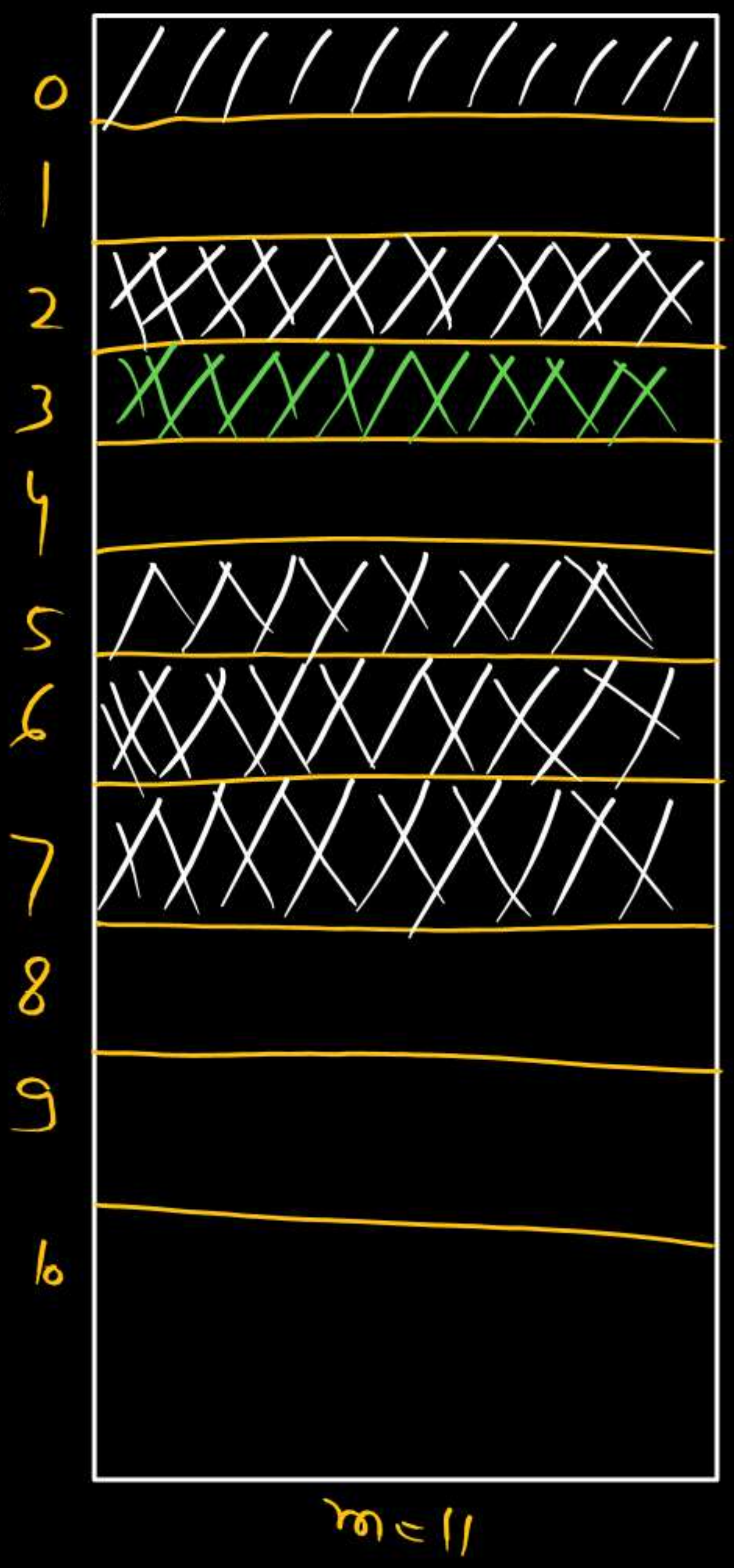
$i=7$

$H(24,7) = (h(24)+7^2) \bmod 11 = 7$
$H(2,7) = 7$
$H(13,7) = 7$

$\xrightarrow{2,3,6,0,7,5,5,7,0}$

6, 3, 2, 2

3, 6, ...



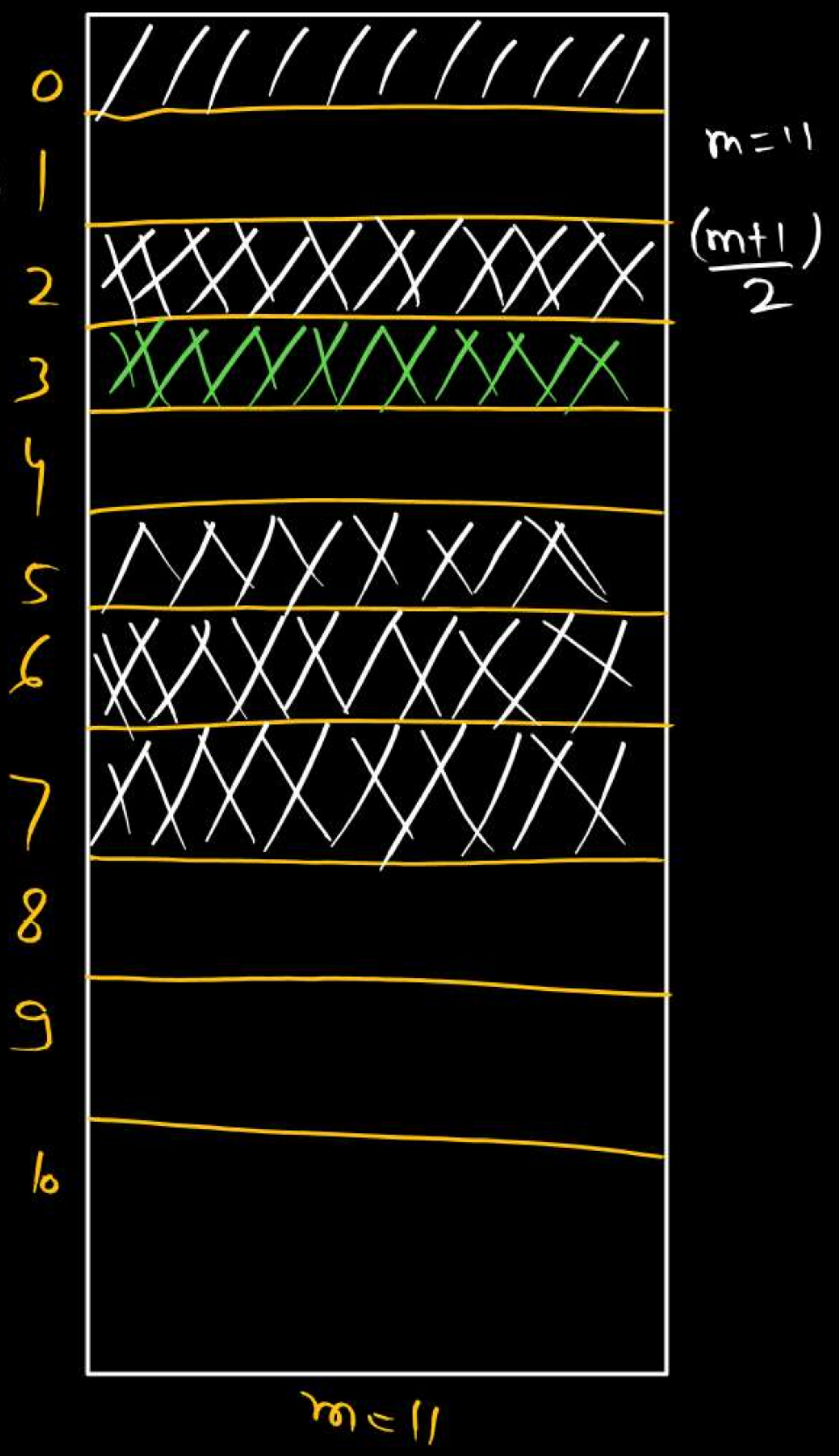
Quadratic Probing

24, 2, 13

Keys that are hashed to same locations follow the same resolution path bcz of which we are not able to utilize the table size Efficiently.

2, 3, 6, 0, 7, 5, 5, 7, 9 →

6, 3, 2, →



Inspite of almost 50% avail. space, we are not able to insert a new element.

THAK - YOU

Slide 5