B-Trees, Part II

Note Title 10/30/2007

Review of Properties

Extension of INORDER

Let t be the order of the B-tree (parameter given to constructor)

BALANCED - all leaf nodes are at same height. So an internal node with x elements has $x+1$ non-empty children

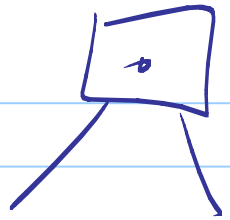
NODEUTILIZATION - With the exception of root all nodes have $\geq t$ children ($\Rightarrow$ so $\geq t-1$ elements). The root has ≥ 2 children. All nodes have $\leq 2t$ children ($\Rightarrow$ so $\leq 2t-1$ elements)

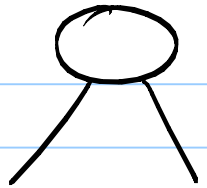
Relation between B-tree + red-black tree

For $t=2$, a B-tree is called a 2-3-4 tree

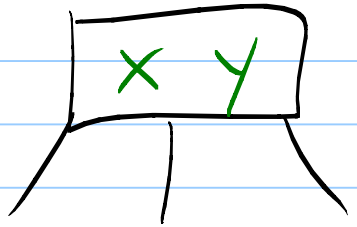
Red-black tree is a representation of a 2-3-4 tree as a binary search tree

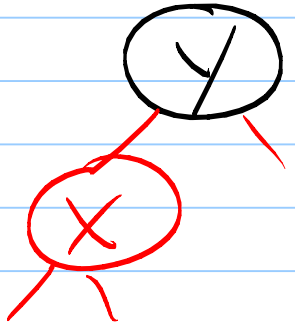
2



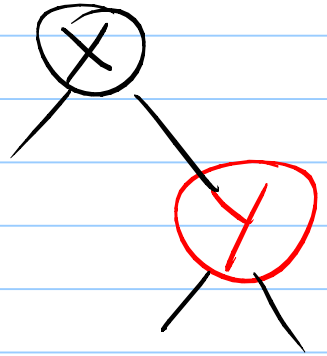


3

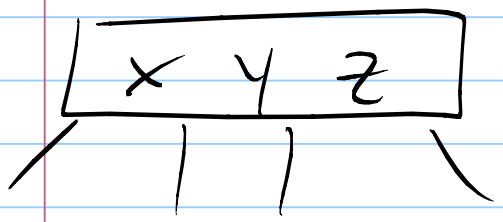


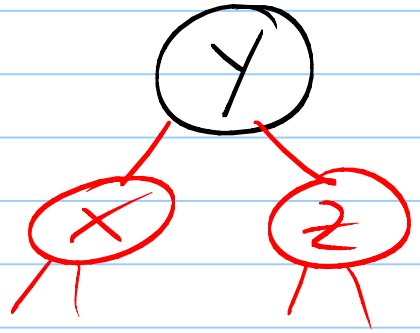


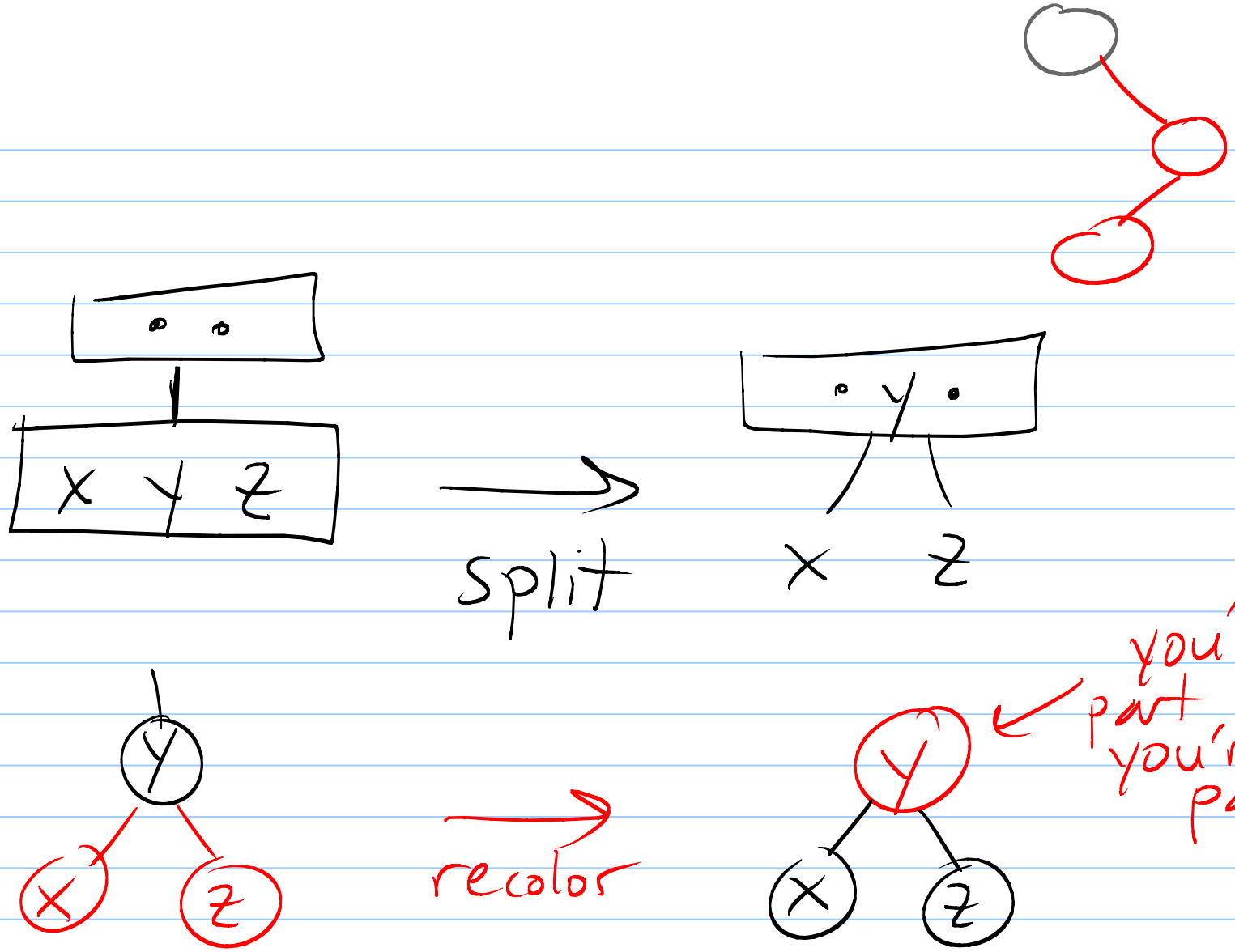
or



4

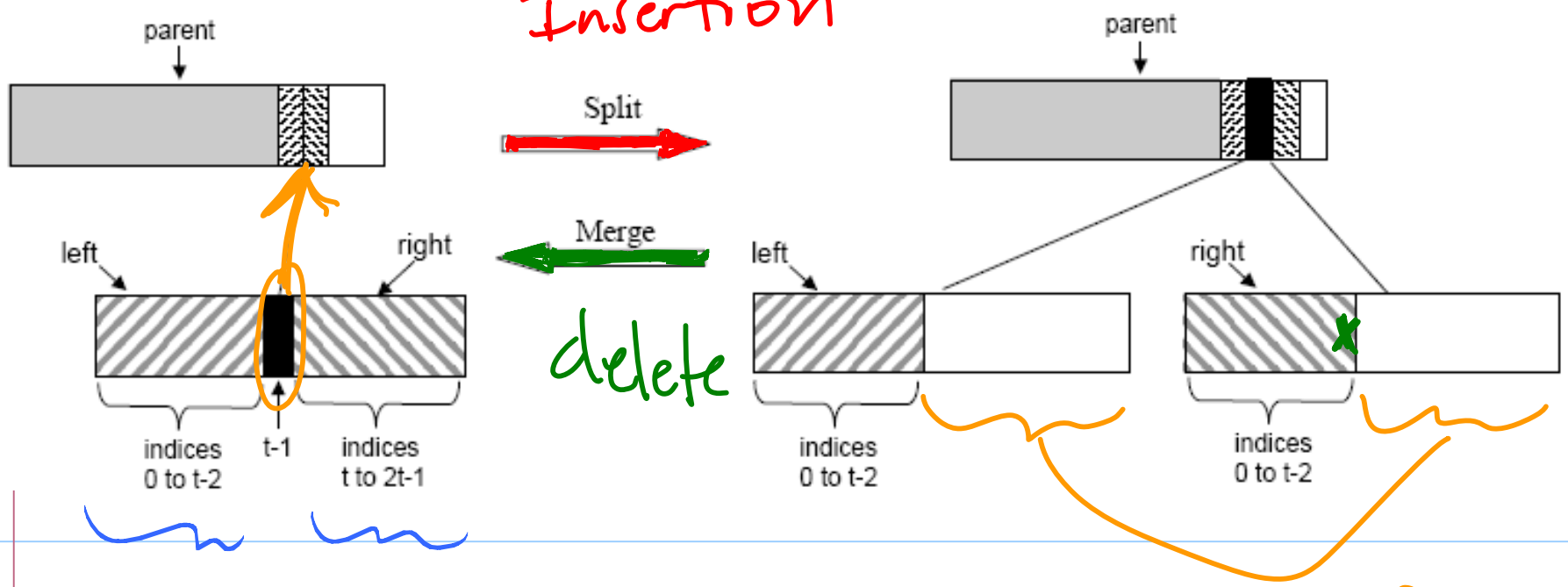





x y z
split
y
x z
recolor
y
x z
you're part of you're parent

Split (+ Merge)

Insertion



t-1 elements

t-1 elements

room for growth

Top-Down Insertion

Follow path to leaf where you'd insert (with natural extension of binary search tree insertion)

Whenever a full node (2t children) encountered split it & then continue

Goal: minimize possibility of a page fault occuring twice on same page

Bottom-up Insertion

Do standard insertion in leaf if there is room.

Otherwise split leaf (which could propogate to the root). Stop as soon as parent is not full.

Reduces unnecessary splits.

Analysis of height

Suppose the B-tree has height h

What is the minimum # of elements it could hold?

$$n \geq f(h)$$

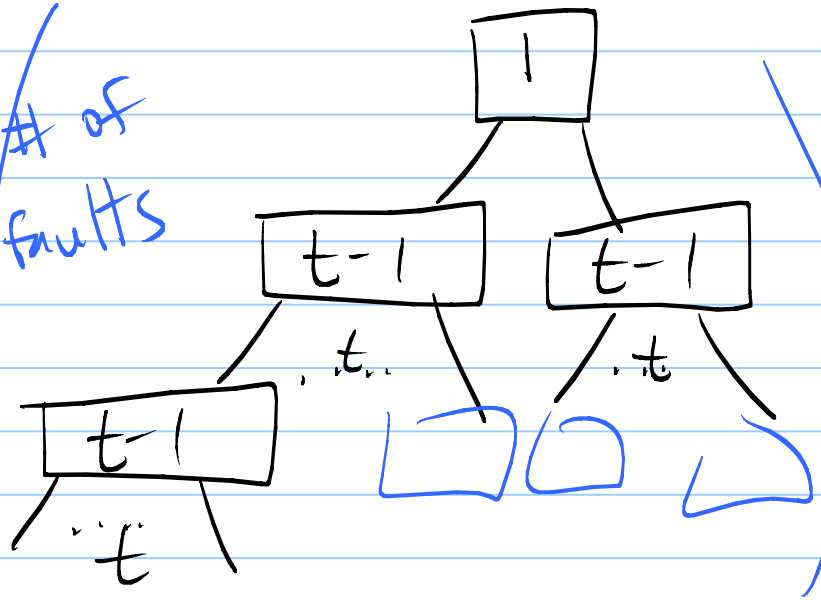
Then solve for h

As a function of n & t

say

$h \leq$ _____

B-tree with minimum utilization



h as the # of page faults

$h \geq 2$

level #	# nodes	# elements
0	$1 \times$	1
1	$2 \times$	$(t-1)$
2	$2t \times$	$(t-1)$
3	$2t^2 \times$	$(t-1)$
⋮		
i	$2 \cdot t^{i-1} \times$	$(t-1)$

$$n \geq 1 + 2(t-1) + 2t(t-1) + \cdots + 2t^{h-1}(t-1)$$

$$= 1 + 2(t-1)\left(1 + t + t^2 + \cdots + t^{h-1}\right)$$

$$= 1 + 2(t-1)\frac{t^h - 1}{t-1} = 1 + 2(t^h - 1)$$

$$n \geq 1 + 2(t^h - 1)$$

$$n - 1 \geq 2(t^h - 1)$$

$$\frac{n-1}{2} \geq t^h - 1$$

$$t^h \leq \frac{n+1}{2}$$

$$h \leq \log_t\left(\frac{n+1}{2}\right) = \log_t(n+1) - \log_t 2$$

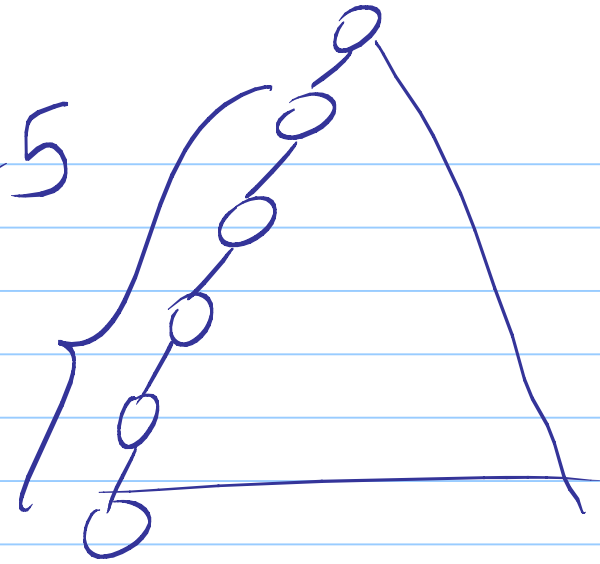
$$\sim \log_t n$$

$$= \frac{\log_2 n}{\log_2 t}$$

negligible

t

If $t = 1000$, $h = 5$



no more than 5 page faults

$$n \geq 1 + 2(t^h - 1)$$

$$= 1 + 2(1000^5 - 1) \approx 2 \cdot 10^{15}$$

B+ tree

t = 2 For letters "abstraction"