Minimum ipl (n) and epl (n) in any BS tree on n nodes

A binary tree T on n nodes has the minimum internal path length among all binary trees on n nodes,

that is,

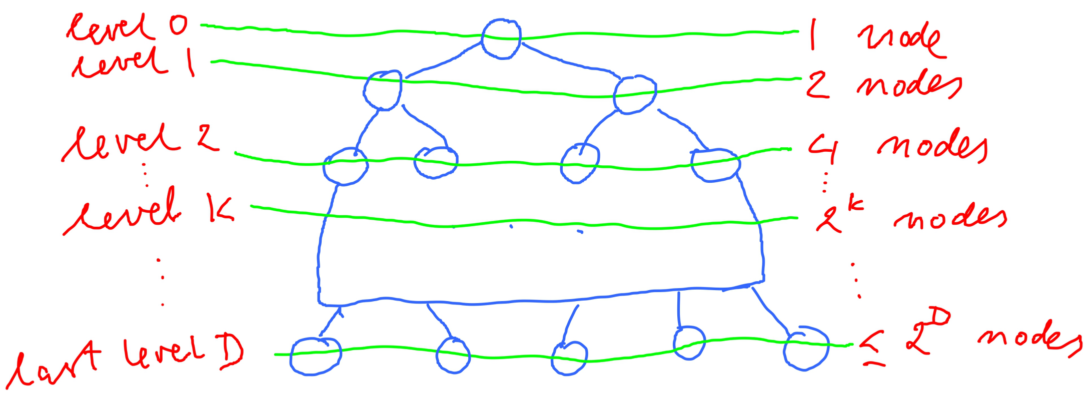
ipl (T) = ipl_{min} (n)

iff

T is empty (so n = 0)

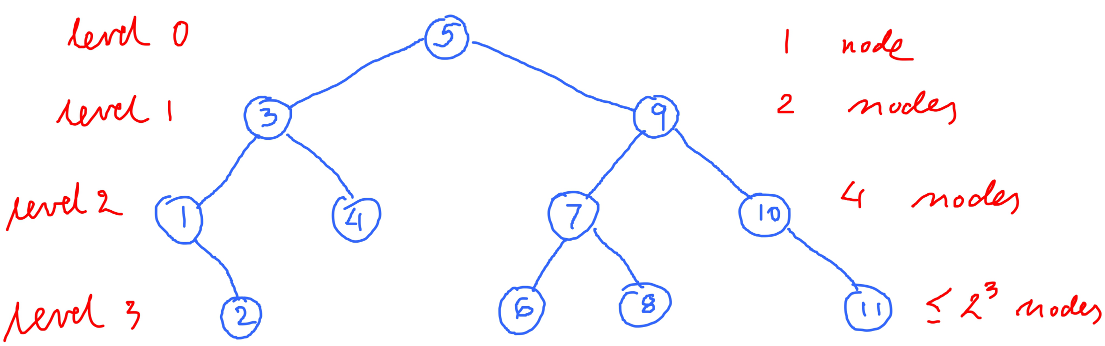
or

all levels of T, except perhaps for the last level of T, have the maximum number of nodes.



Recall that the maximum number of nodes in level k of a binary tree is 2^k.

Here is example of such a BS tree on nodes 1 through 11 :



Hence, the minimum internal path length ipl_{min} (n) is equal to the internal path length ipl (H) in a heap on n nodes.

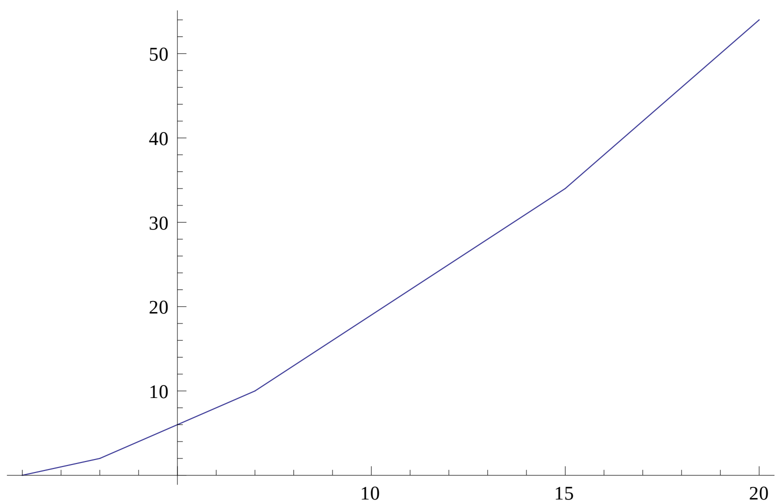
Thus,

$ipl_{min}(n) = \sum_{i=1}^{n} \lfloor Log2[i] \rfloor =$
$(n+1) \lfloor Log2[n+1] \rfloor - 2^{\lfloor Log2[n+1] \rfloor + 1} + 2 =$
$= (n+1)(Log2[n+1] + \epsilon) - 2n$,

where ϵ is a function of $n+1$ oscillating between:
$0 \le \epsilon < 0.086071332056$

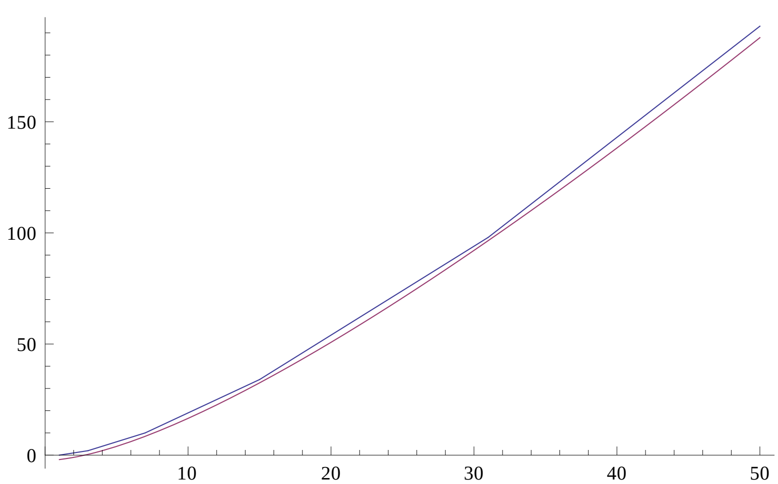
Note.
$(n+1)(Log2[n+1] + \epsilon) - 2n$
is the number of comparisons of keys that Quicksort will perform in the best case while sorting an n - element array with no duplicate keys.

Plot$\left[(n+1) \lfloor Log2[n+1] \rfloor - 2^{\lfloor Log2[n+1] \rfloor + 1} + 2, \{n, 1, 20\}\right]$



So,
$ipl_{min}(n) \approx (n+1) Log2[n] - 2n$.

Plot$\left[\left\{(n+1) \lfloor Log2[n+1] \rfloor - 2^{\lfloor Log2[n+1] \rfloor + 1} + 2, (n+1) Log2[n] - 2n\right\}, \{n, 1, 50\}\right]$



Therefore,

$epl_{min}(n) = ipl_{min}(n) + 2n \approx (n+1)\,Log2[n]$

or

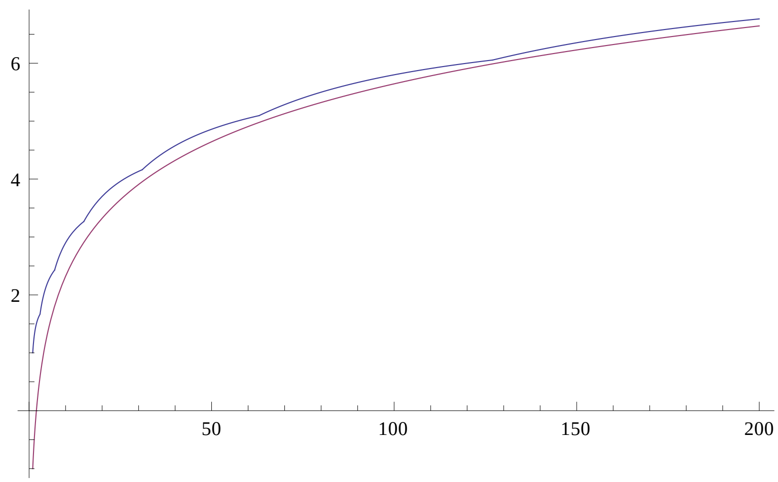
$epl_{min}(n) \approx (n+1)\,Log2[n]$

Average number of comparisons for successful search in an average BS tree :

$$c_n = \frac{ipl_{avg}[n]}{n} + 1 \approx \frac{1}{n}((n+1)\,Log2[n] - 2n) + 1 \approx$$
$$Log2[n] - 2 + 1 = Log2[n] - 1.$$

$c_n \approx Log2[n] - 1$

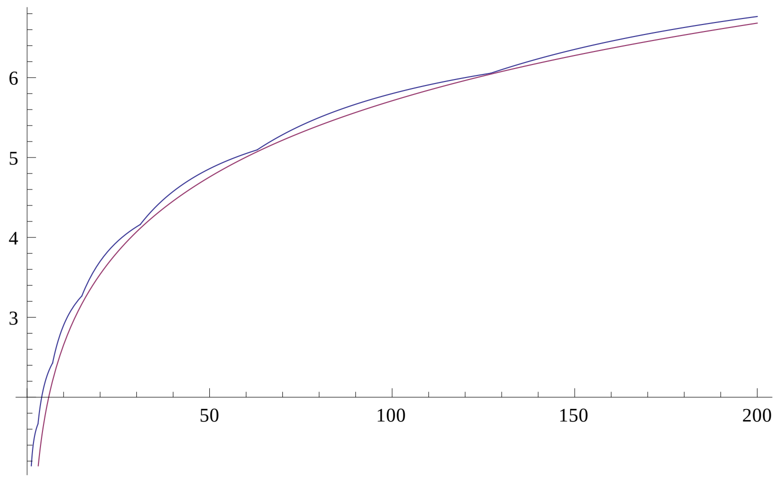
Plot[{1 + ((n + 1) ⌊Log2[n + 1]⌋ - 2^(⌊Log2[n + 1]⌋ + 1) + 2) / n, Log2[n] - 1},
{n, 1, 200}, PlotPoints→ 200]



Here is a bit more accurate approximation:

$$c_n \approx Log2[n] - 1 + \frac{Log2[n]}{n}$$

```
Plot[{1 + ((n + 1) ⌊Log2[n + 1]⌋ - 2^(⌊Log2[n + 1]⌋ + 1) + 2) / n, Log2[n] - 1 + Log2[n]/n},
 {n, 1, 200}, PlotPoints→ 200]
```



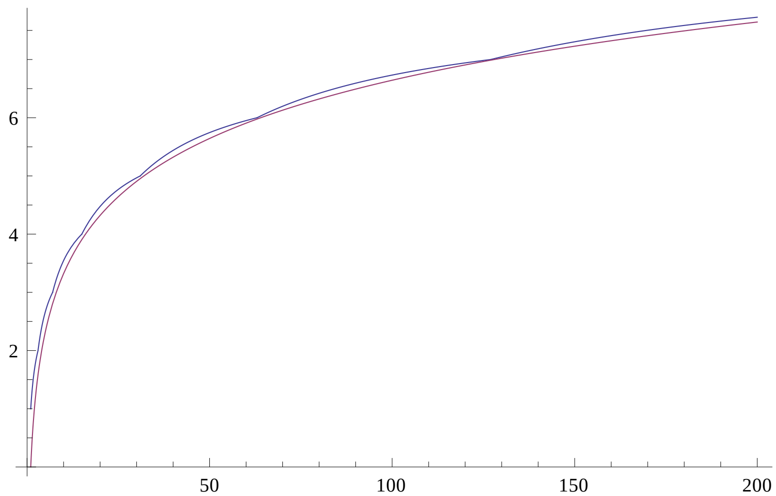
Average number of comparisons for unsuccessful search in an average BS tree :

$$c'_n = \frac{epl_{avg}[n]}{n+1} \approx \frac{(n+1)\,Log2[n]}{n+1} = Log2[n]$$

1.386 Log2[n] - 0.846

$c'_n \approx Log2[n]$.

```
Plot[{((n + 1) ⌊Log2[n + 1]⌋ - 2^(⌊Log2[n + 1]⌋ + 1) + 2 + 2 n) / (n + 1), Log2[n]},
 {n, 1, 200}, PlotPoints→ 200]
```



NB*