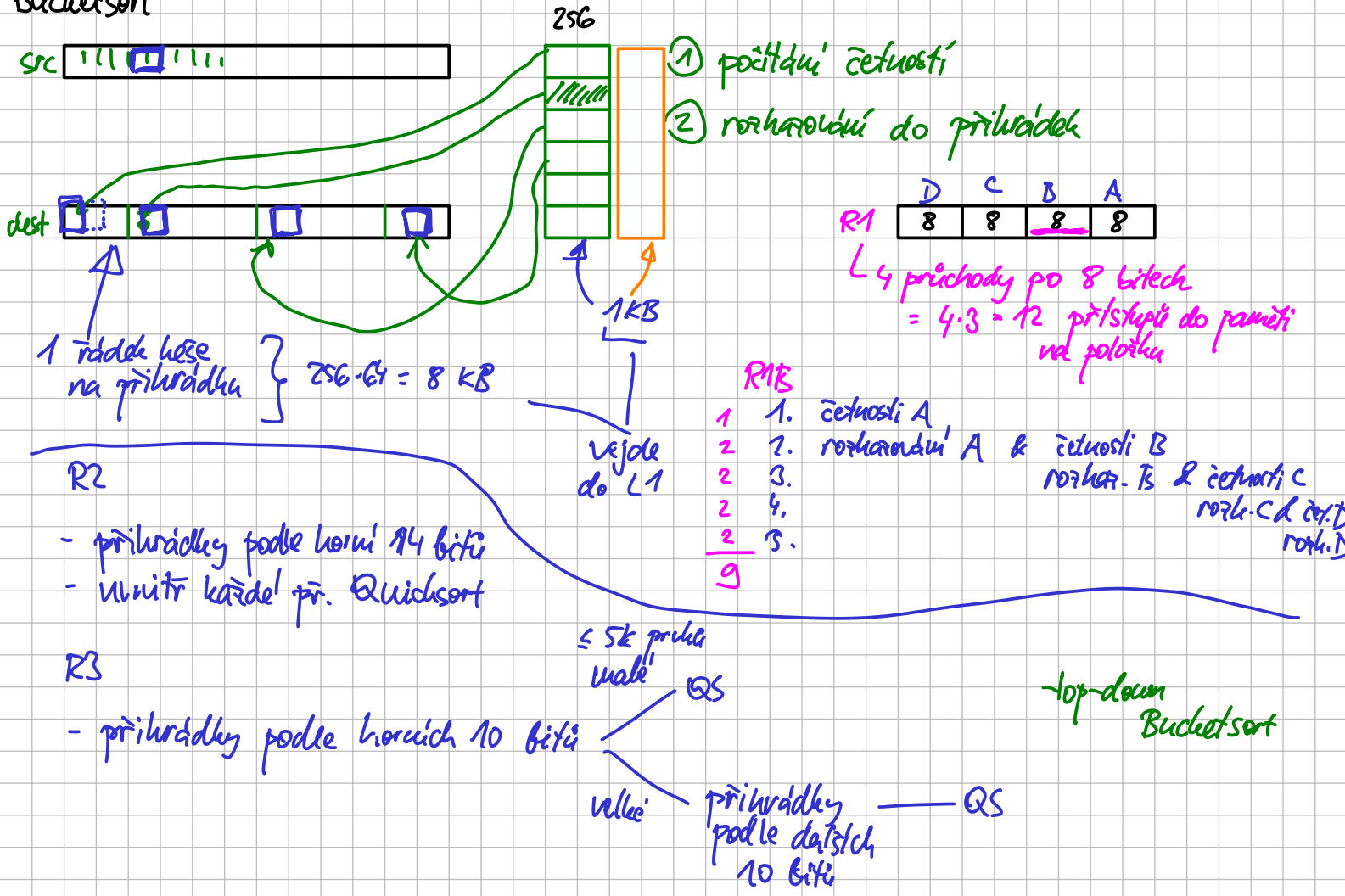


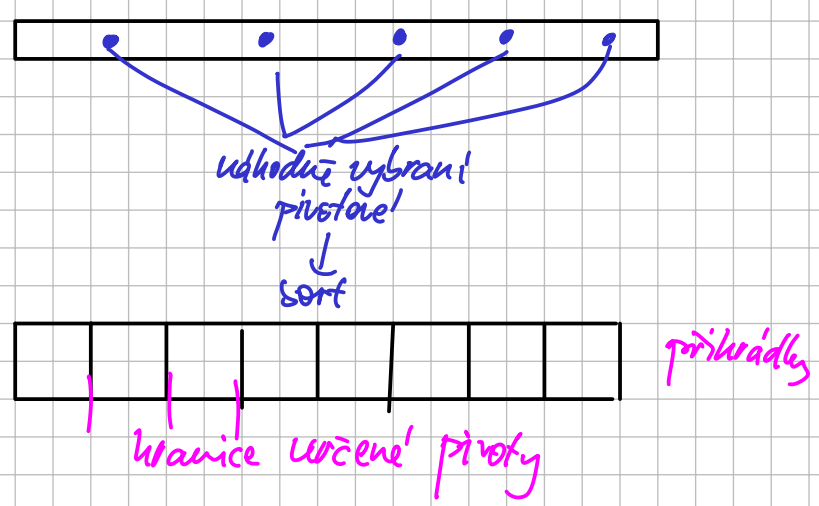
# Bucket sort



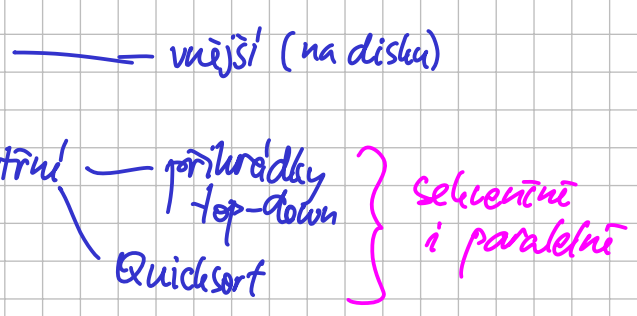
Monotónní heřivací funkce :  $f: \text{Klíče} \rightarrow \{0 \dots 2^b - 1\}$   
 $\forall x \ f(x) < f(y) \Rightarrow x < y$

16 B klče  
 4 B jsou monotónní heř

# Sample sort



# Trřidění v LibCCW

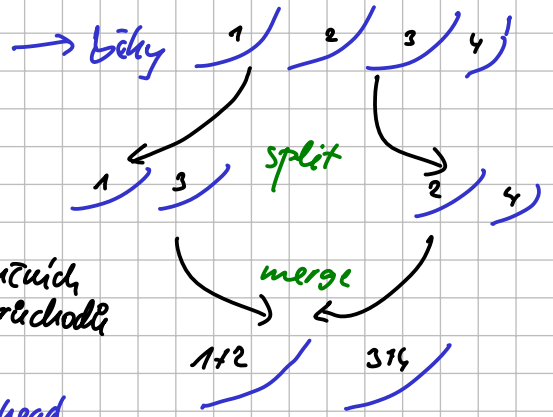


# Vnější třídění - Disk není páska!

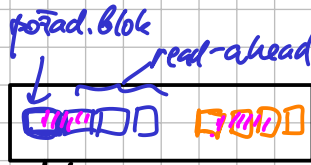
1GB RAM, 1TB vstup  
 ↳ 1024 běhů

Tradiční: ① předtřídění v paměti  $\left\{ \begin{array}{l} \text{load \& sort \& flush} \\ \text{tržky s haldami} \end{array} \right.$

② externí Mergesort: split & merge  
 2 pásky  $\rightarrow$  2 pásky  
 $O(\log \# \text{běhů})$  - sekvenční průchod



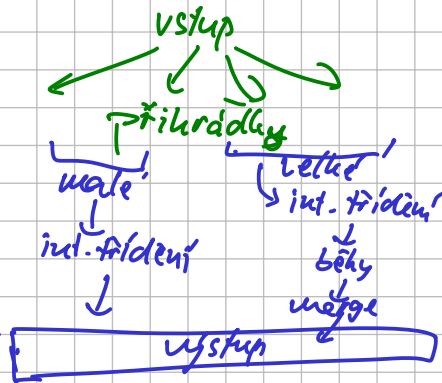
Disk  $\approx$  pole bloků (typ. 4KB)  
 100 MB/s sekvenční čtení  
 5 ms seek



Chceme: čas na seq. čtení bufferu = čas na seek  
 ↳ velikost bufferu  $\sim 0.5$  MB  
 ↳ řádově jednotky MB  
 M. paměť = cache pro disk

Polusi: 256-cestný Merge ... 256 MB na buffery stací

Další: 256-cestné rozložení do příhrádek



## Asyptotika ext. třídění

$N$  := # prvků  
 $M$  := velikost int. paměti  
 $B$  := velikost bloku

uvažme:  $M/B$  - cestný

$$\# \text{I/O} = O\left(\frac{N}{B} \cdot \log_{M/B} \frac{N}{B}\right)$$

$$\frac{N}{B} \cdot \frac{\log M/B}{\log M/B}$$

dolní odhad na permutace  $\Omega\left(\min\left(N, \frac{N}{B} \cdot \frac{\log M/B}{\log M/B}\right)\right)$

[Aggarwal, Vitter 1988]