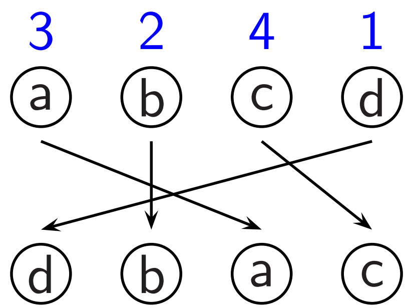
Permuting

Upper and Lower bounds

[Aggarwal, Vitter, 88]

Upper Bound

Assume instance is specified by each element knowing its final position:



Algorithm	Internal Cost	I/O Cost
1) Place each element directly	$\Theta(N)$	$\Theta(N)$
2) Sort on final position	$\Theta(N \log N)$	$\Theta(N/B \log_{M/B}(N/B))$

Upper Bound

Internally, 1) always best.

Externally, 2) best when $1/B \log_{M/B}(N/B) \leq 1$.

Note: This is almost always the case practice. Example:

$$B = 10^3,\ M = 10^6,\ N = 10^{30}$$

$$\Downarrow$$

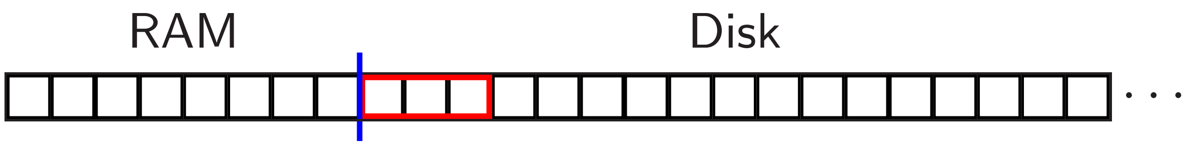
$$1/B \log_{M/B}(N/B) = 9/10^3 << 1$$

External Permuting:

$$\boxed{O(\min\{N/B \log_{M/B}(N/B), N\}) = O(\min\{\text{sort}(N), N\})}$$

Lower Bound Model

Model of combined memory (RAM + disk):



- Elements are indivisible: May be moved, copied, destroyed, but newer broken up in parts.
- Assume $M \geq 2B$.
- May assume I/Os are block-aligned, and that at start [end], input [output] is in lowest contiguous positions on disk.

Lower Bound

We may assume that elements are only **moved**, not copied or destroyed.

Reason: For any sequence of I/Os performing a permutation, exactly one copy of each element exists at end. Change all I/Os performed to only deal with these copies. Result: same number of I/Os, same permutation at end, but now I/Os only move elements.

Consequence:

Memory always contains a permutation of the input

Recall, memory means combined memory (RAM + disk), seen as one array.

Lower Bound

In analysis, keep a set S_t of permutations.

Maintain the following invariants:

1. S_t contains all permutations of elements in memory possible to reach with t I/Os.

2. S_t is closed under permutations of the elements in RAM, and under permutation of the elements inside each touched block on disk.

For an I/O, let X be an upper bound on the increase of $|S_t|$:

$$|S_{t+1}| \leq X \cdot |S_t|.$$

Bounds on Value of X

Type of I/O	Read untouched block	Read touched block	Write
X	$B\frac{N}{B}\binom{M}{B}B!$	$BN\binom{M}{B}$	BN

For t I/Os, there are at most 3^t sequences of choices of the three types.
For each sequence, there are at most N/B untouched reads.

Hence, from $|S_0| = 1$ and $|S_{t+1}| \leq X \cdot |S_t|$ we get

$$|S_t| \leq 3^t \left(\binom{M}{B} BN\right)^t (B!)^{N/B}$$

To be able to reach every possible permutation, we need $N! \leq |S_t|$.
Thus,

$$N! \leq \left(3\binom{M}{B} BN\right)^t (B!)^{N/B}$$

is necessary to reach all permutations.

Lower Bound Computation

$$\left(3\binom{M}{B}BN\right)^t (B!)^{N/B} \geq N!$$

$$t(\log\binom{M}{B} + 3\log N) + (N/B)\log(B!) \geq \log(N!)$$

$$t(3B\log(M/B) + 3\log N) + N\log B \geq N(\log N - \log_2(e))$$

$$t \geq \frac{N(\log N - \log_2(e) - \log B)}{3B\log(M/B) + 3\log N}$$

$$t = \Omega(\frac{N\log(N/B)}{B\log(M/B) + \log N})$$

Using Lemma:

a) $\log(x!) \geq x(\log x - \log_2(e))$

b) $\log(x!) \leq x\log x$

c) $\log\binom{x}{y} \leq 3y\log(x/y)$ when $x \geq 2y$

d) $\log(3BN) \leq 3\log(N)$ when $3 \leq N$, $B \leq N$

Lower Bound

$$\Omega(\frac{N\log(N/B)}{B\log(M/B)+\log N})$$
$$=\Omega(\min\{\frac{N\log(N/B)}{B\log(M/B)},\frac{N\log(N/B)}{\log N}\})$$
$$=\Omega(\min\{Z_1,Z_2\})$$

Note 1: $Z_1 = \mathrm{sort}(N)$

Note 2: $Z_2 < Z_1 \Leftrightarrow B\log(M/B) < \log N \Rightarrow B < \log N \Rightarrow$

$$Z_2 = \frac{N\log(N/B)}{\log N} = \frac{N(\log N - \log B)}{\log N} = \Theta(N)$$

Note 3: $Z_2 \leq N$ always

By Note 2 and 3, it is OK to substitute N for Z_2 inside $\min$.

The I/O Complexity of Permuting

We have proven:

$$\Theta(\min\{\text{sort}(N), N\})$$