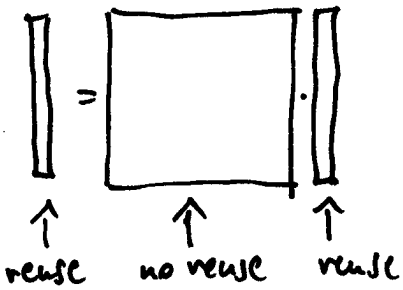
MVM



Performance model

1.) Gain of blocking estimation

$$G_{r,c} = \frac{\text{perf } r \times c \text{ BCSR}}{\text{perf CSR}}$$ for dense MVM

machine dependent, independent of A

2.) Computational overhead

$$O_{r,c} = \frac{\text{size of A in } r \times c \text{ BCSR}}{\text{size of A in CSR}}$$

computed by scanning only a fraction of A

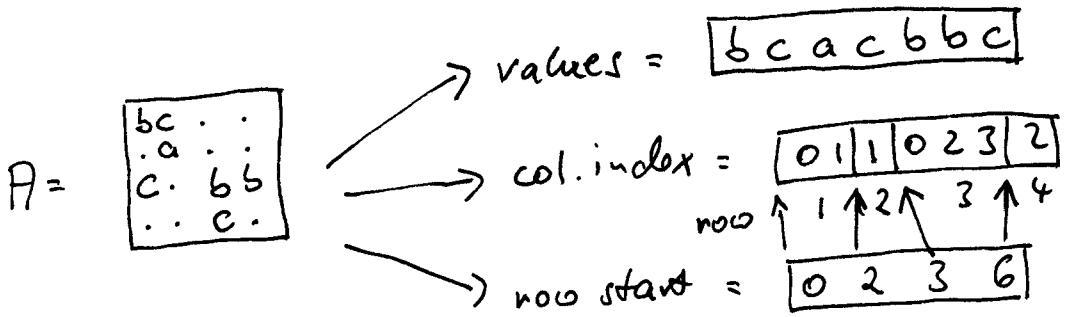
machine independent, dependent on A

Model: Gain from blocking is

$$P_{r,c} = \frac{G_{r,c}}{O_{r,c}}$$

max computed for all $r,c \in \{1,\ldots,12\}$

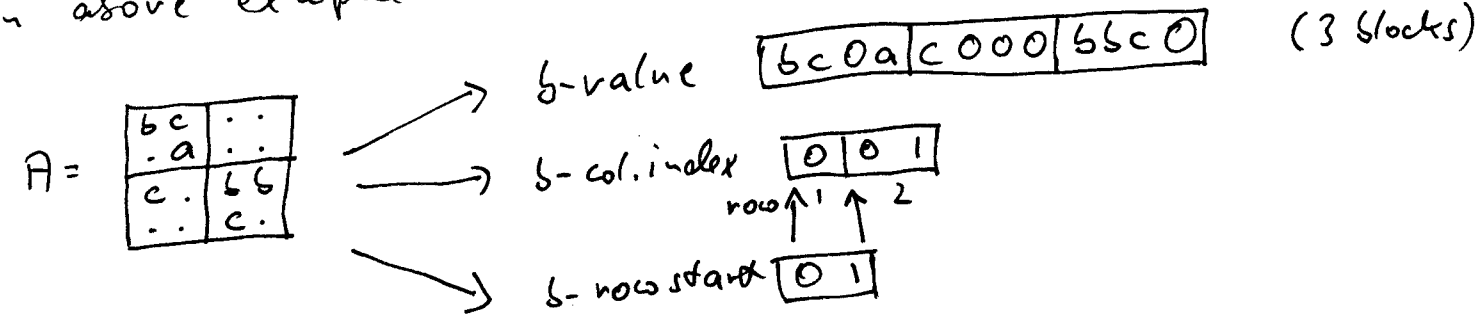
Compressed Sparse Row (CSR) Format



storage: Θ(# non-zero entries of A)

Blocked CSR (BCSR) Format

- block A into r x c blocks
- in above example and r = c = 2:

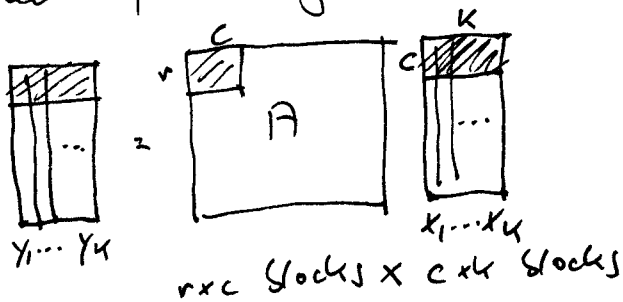


computational overhead: $12/7 \approx 1.7$

Multiple Vectors in MVM

$y_i = A x_i$, $i = 1..k$

Instead of doing it sequentially:



r x c blocks x c x k blocks

Equivalent to MMM.