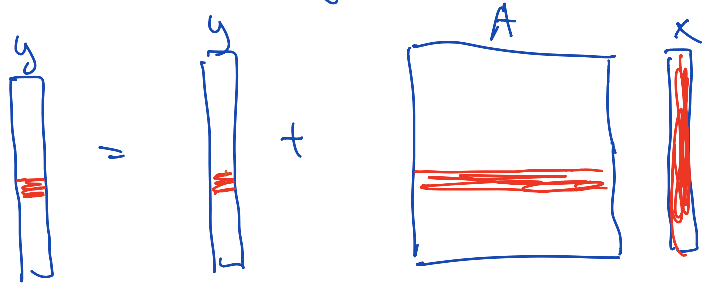
Computational intensity for matrix-vector & matrix-matrix mult.

1.) Matrix-vector multiplication

$y = y + Ax$, $y, x \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$



flops:

$\sim 2n^2$ flops

$y_i = y_i + \sum_{k=1}^{n} a_{ik} x_k$
$= y_i + a_{ik_1} + a_{ik_2} + \dots + \dots$

memory access (slow mem access, nothing is in fast memory)

assume we can hold a few vectors in fast memory

load y, x, load A row-by row, write result to y

$3n + n^2$

comp. intensity $q = f/m = \frac{2n^2}{3n + n^2} \sim 2$

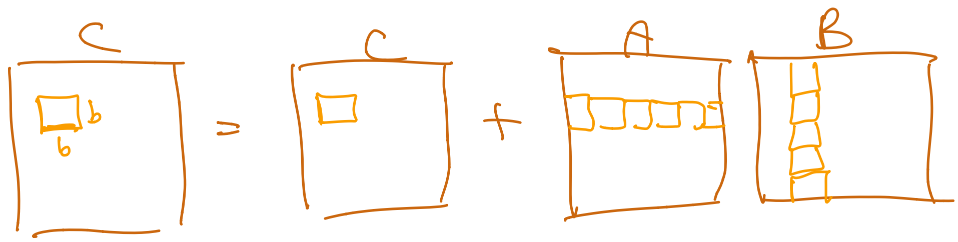
2.) Matrix-Matrix Mult

$C = C + A * B$, $A, B, C \in \mathbb{R}^{n \times n}$

flops: $2n^3$

memory: same as before: $(3n + n^2)n \Longrightarrow q \sim 2$

Tiling algorithm:



n entries, N blocks, $b = \frac{n}{N}$ blocksize

memory access: N^3 block reads of B

N^3 — " — of A

$2N^2$ block reads in C & writes

memory access: $(2N^3 + 2N^2) \cdot b^2 =$

$(2N^3 + 2N^2) \cdot \frac{n^2}{N^2} \sim 2N n^2 + 2n^2$

$= 2 \frac{n^3}{b} + 2n^2$

$q_n^{-1} = \frac{2 \frac{n^3}{b} + 2n^2}{2n^3} \Rightarrow \boxed{q_n = b}$

$\Rightarrow$ gives a much higher comp. intensity, much faster!