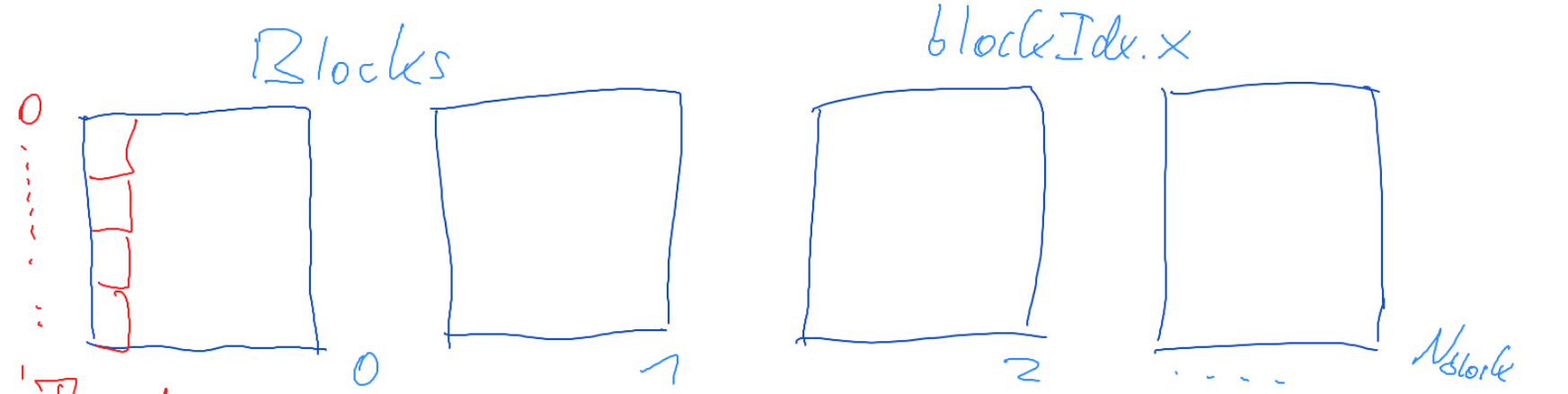


index = threadIdx.x + blockIdx.x * blockDim.x

every thread gets one vector element

gridDim.x = ~~blockDim.x~~ * N_{blocks} in the Kernel

Kernel Call:

add <<< N_1, N_2 >>> (....) Main

N_1 : N / threadsperblock N_2 : threadsperblock ≤ 1024

Inside Kernel:

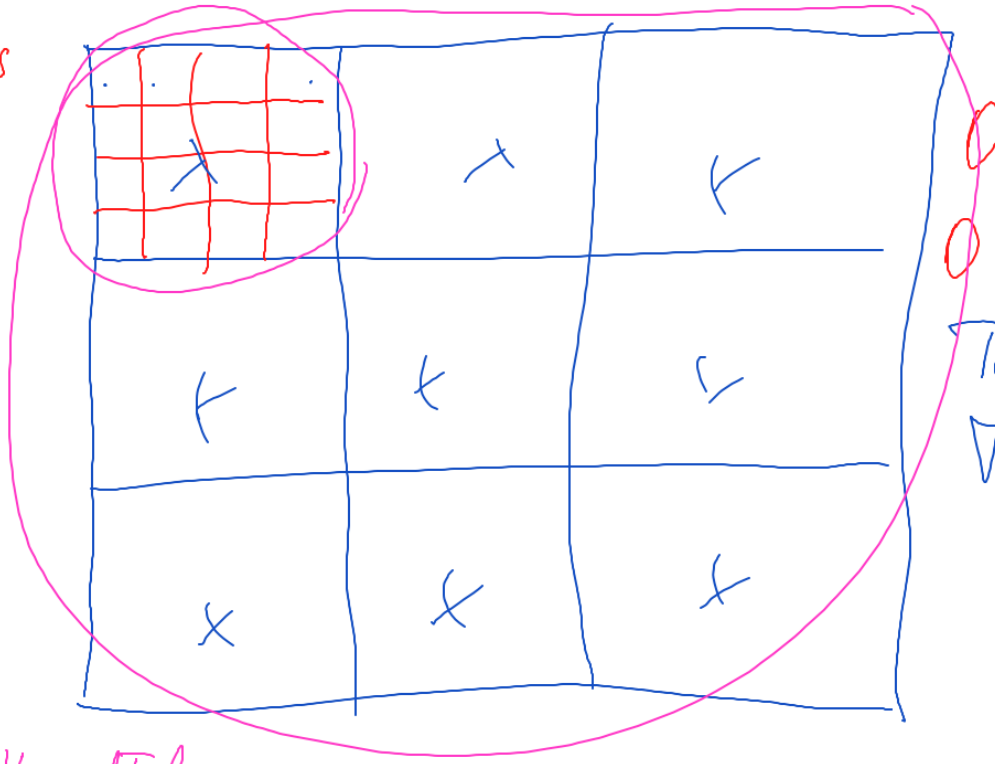
$N_2 \hookleftarrow$ blockDim.x

~~$N_1 * N_2 \hookleftarrow$ gridDim.x~~

$N_1 \hookleftarrow$ gridDim.x

$N_1 * N_2 \hookleftarrow$ blockDim.x * gridDim.x

9 blocks
16 threads per block



$0 \le threadIdx.x < blockDim.x$

$0 \le blockIdx.x < gridDim.x$

Total Number of Threads

$blockDim.x * gridDim.x$

16 9

$blockDim.y$ / $gridDim.y$ | $threadIdx.y$ / $blockIdx.y$