

CUDA Kernel

Block (0,0)

Block (1,0)

Block (2,0)

Block (0,1)

Block (1,1)

Block (2,1)

Block (1,1)

Warp

Thread
(0,0)

Thread
(1,0)

Thread
(2,0)

Thread
(3,0)

Thread
(4,0)

Warp

Thread
(0,1)

Thread
(1,1)

Thread
(2,1)

Thread
(3,1)

Thread
(4,1)

Streaming Multiprocessor

Shared Memory /
L1 Cache

Register File

Processing Units

