

TEXT = 4

GPU Work Scheduler

TOTAL = 1000

Fetch/Decode



Block 0 (contexts 0-127)

Block 2 (contexts 128-255)

Execution context
storage for 384 CUDA
threads

Block 0: support
(520 bytes @ 0x0)

Block 2: support
(520 bytes @ 0x520)

“Shared” memory
storage (1.5 KB)

Core 0

Fetch/Decode



Block 1 (contexts 0-127)

Block 3 (contexts 128-255)

Execution context
storage for 384 CUDA
threads

Block 1: support
(520 bytes @ 0x0)

Block 3: support
(520 bytes @ 0x520)

“Shared” memory
storage (1.5 KB)

Core 1