

# CUDA Occupancy Calculator

Compute Capability version

Threads per block

Registers per thread

Shared memory per block




GPU Occupancy Data is displayed here and in the graphs

<b>Active Threads per Multiprocessor</b>	2048
<b>Active Warps per Multiprocessor</b>	64
<b>Active Thread Blocks per Multiprocessor</b>	8
<b>Occupancy of each Multiprocessor</b>	1

Physical Limits for GPU Compute Capability

<b>Version</b>	6.1
<b>Threads per Warp</b>	32
<b>Warps per Multiprocessor</b>	64
<b>Threads per Multiprocessor</b>	2048
<b>Thread Blocks per Multiprocessor</b>	32
<b>Total # of 32-bit registers per Multiprocessor</b>	65536
<b>Register allocation unit size</b>	256
<b>Register allocation granularity</b>	warp
<b>Max registers per Block</b>	65536
<b>Max registers per thread</b>	255
<b>Shared Memory per Multiprocessor (bytes)</b>	98304
<b>Shared Memory Allocation unit size</b>	256
<b>Warp allocation granularity (for register allocation)</b>	4
<b>Max thread block size</b>	1024

Allocation Per Thread Block

<b>Warps</b>	8
<b>Registers</b>	2048
<b>Shared Memory</b>	256

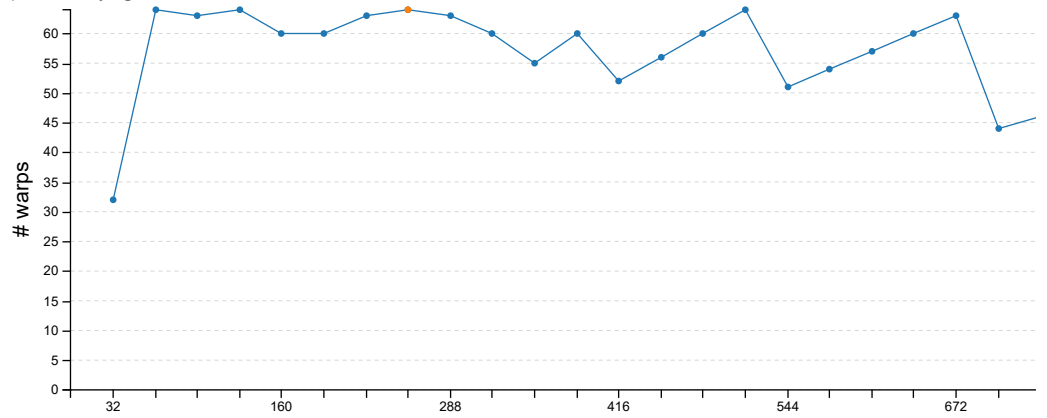
Maximum Thread Blocks Per Multiprocessor

<b>Limited by Max Warps / Blocks per Multiprocessor</b>	8
<b>Limited by Registers per Multiprocessor</b>	32

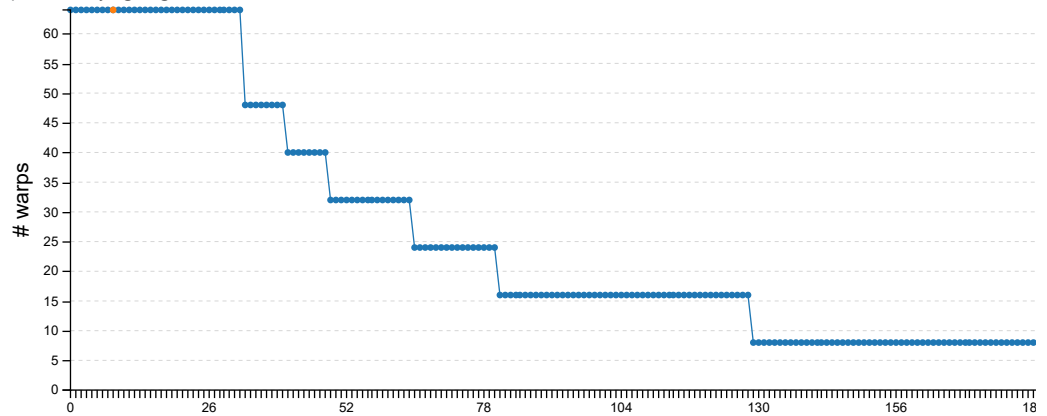
## Limited by Shared Memory per Multiprocessor

384

Impact of Varying Block Size



Impact of Varying Register Count Per Thread



Impact of Varying Shared Memory Usage Per Block

