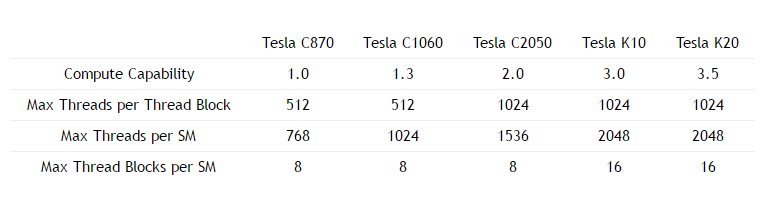
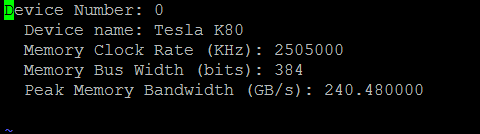
GPU basics

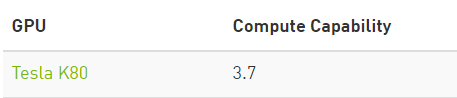
* The following table tells us about the maximum threads available per block ,Max threads per Sm, Max threads Blocks per Sm for Tesla range GPUS, Tesla k80 has similar values.



* We can get device properties using **cudaDeviceProp** Struct.

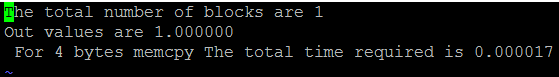


* The compute availability of Tesla K80 is given below.



1. **Measure PCI-express latency.**

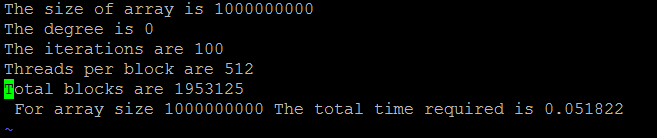
* To measure PCI-Express latency we need to measure the time required to transfer 4 bytes from CPU to GPU.
* So we measure the time for Memcopying 4 bytes.
* The PCI-Express latency measured was 17microseconds.



1. **Measure Pci Bandwidth**

* To measure Pci Bandwidth we calculate the time required for copying a huge array to Gpu
* 
* For transferring (1000000 \*4) bytes we require 0.005406 seconds.
* So, the bandwidth measured is around 8Gb/sec. (Also other sizes were considered and 8.5 Gb/sec was the best I got).

1. **Measure Memory Bandwidth.**

* For calculating Memory Bandwidth we need to read and write array to the system
* Our array size is 10^9.
* 
* So the bandwidth obtained was **155GBps**

1. **Measure GPU Flops rating.**

* To measure GPU flops rating we need to count the number of flops done by us and divide by the time taken to do it.
* For an array size of 10^7 and degree as 10^5, the number of flops done are
* (3\*10^5)\*(10^7)+ (2\*10^7) is the number of flops we do.
* The Flops rating achieved was around **3TB.**

