CUDA Device Query (Runtime API) version (CUDART static linking)

Detected 1 CUDA Capable device(s)

Device 0: "GeForce GTX 680"
  CUDA Driver Version / Runtime Version 7.5 / 7.0
  CUDA Capability Major/Minor version number: 3.0
  Total amount of global memory: 2043 MBytes (2141913088 bytes)
    ( 8 ) Multiprocessors, (192) CUDA Cores/MP: 1536 CUDA Cores
  GPU Max Clock rate: 1058 MHz (1.06 GHz)
  Memory Clock rate: 3004 Mhz
  Memory Bus Width: 256-bit
  L2 Cache Size: 524288 bytes
  Maximum Texture Dimension Size (x,y,z) 1D=(65536), 2D=(65536, 65536),
    3D=(4096, 4096, 4096)
  Maximum Layered 1D Texture Size, (num) layers 1D=(16384), 2048 layers
  Maximum Layered 2D Texture Size, (num) layers 2D=(16384, 16384), 2048 layers
  Total amount of constant memory: 65536 bytes
  Total amount of shared memory per block: 49152 bytes
  Total number of registers available per block: 65536
  Warp size: 32
  Maximum number of threads per multiprocessor: 2048
  Maximum number of threads per block: 1024
  Max dimension size of a thread block (x,y,z): (1024, 1024, 64)
  Max dimension size of a grid size (x,y,z): (2147483647, 65535, 65535)
  Maximum memory pitch: 2147483647 bytes
  Texture alignment: 512 bytes
  Concurrent copy and kernel execution: Yes with 1 copy engine(s)
  Run time limit on kernels: No
  Integrated GPU sharing Host Memory: No
  Support host page-locked memory mapping: Yes
  Alignment requirement for Surfaces: Yes
  Device has ECC support: Disabled
  Device supports Unified Addressing (UVA): Yes
  Device PCI Domain ID / Bus ID / location ID: 0 / 65 / 0
  Compute Mode:
    < Default (multiple host threads can use ::cudaSetDevice() with device
    simultaneously) >

deviceQuery, CUDA Driver = CUDART, CUDA Driver Version = 7.5, CUDA Runtime Version = 7.0, NumDevs = 1, Device0 = GeForce GTX 680

Result = PASS
./bandwidthTest
[CUDA Bandwidth Test] - Starting...
Running on...
Device 0: GeForce GTX 680
Quick Mode

Host to Device Bandwidth, 1 Device(s)
PANNED Memory Transfers
   Transfer Size (Bytes)   Bandwidth(MB/s)
   33554432               5192.0

Device to Host Bandwidth, 1 Device(s)
PANNED Memory Transfers
   Transfer Size (Bytes)   Bandwidth(MB/s)
   33554432               5226.2

Device to Device Bandwidth, 1 Device(s)
PANNED Memory Transfers
   Transfer Size (Bytes)   Bandwidth(MB/s)
   33554432               153466.9

Result = PASS

NOTE: The CUDA Samples are not meant for performance measurements. Results may vary when GPU Boost is enabled.
GPUS2 (gpus2.cise.ufl.edu)

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUDA Device Query (Runtime API) version</td>
<td>(CUDART static linking)</td>
</tr>
<tr>
<td>Detected 1 CUDA Capable device(s)</td>
<td></td>
</tr>
<tr>
<td>Device 0: &quot;Tesla C2070&quot;</td>
<td></td>
</tr>
<tr>
<td>CUDA Driver Version / Runtime Version</td>
<td>7.0 / 7.0</td>
</tr>
<tr>
<td>CUDA Capability Major/Minor version number</td>
<td>2.0</td>
</tr>
<tr>
<td>Total amount of global memory</td>
<td>5375 MBytes (5636292608 bytes)</td>
</tr>
<tr>
<td>(14) Multiprocessors, (32) CUDA Cores/MP</td>
<td>448 CUDA Cores</td>
</tr>
<tr>
<td>GPU Max Clock rate</td>
<td>1147 MHz (1.15 GHz)</td>
</tr>
<tr>
<td>Memory Clock rate</td>
<td>1494 Mhz</td>
</tr>
<tr>
<td>Memory Bus Width</td>
<td>384-bit</td>
</tr>
<tr>
<td>L2 Cache Size</td>
<td>786432 bytes</td>
</tr>
<tr>
<td>Maximum Texture Dimension Size (x,y,z)</td>
<td>1D=(65536), 2D=(65536, 65535), 3D=(2048, 2048, 2048)</td>
</tr>
<tr>
<td>Maximum Layered 1D Texture Size, (num) layers</td>
<td>1D=(16384), 2048 layers</td>
</tr>
<tr>
<td>Maximum Layered 2D Texture Size, (num) layers</td>
<td>2D=(16384, 16384), 2048 layers</td>
</tr>
<tr>
<td>Total amount of constant memory</td>
<td>65536 bytes</td>
</tr>
<tr>
<td>Total amount of shared memory per block</td>
<td>49152 bytes</td>
</tr>
<tr>
<td>Total number of registers available per block</td>
<td>32768</td>
</tr>
<tr>
<td>Warp size</td>
<td>32</td>
</tr>
<tr>
<td>Maximum number of threads per multiprocessor</td>
<td>1536</td>
</tr>
<tr>
<td>Maximum number of threads per block</td>
<td>1024</td>
</tr>
<tr>
<td>Max dimension size of a thread block (x,y,z)</td>
<td>(1024, 1024, 64)</td>
</tr>
<tr>
<td>Max dimension size of a grid size (x,y,z)</td>
<td>(65535, 65535, 65535)</td>
</tr>
<tr>
<td>Maximum memory pitch</td>
<td>2147483647 bytes</td>
</tr>
<tr>
<td>Texture alignment</td>
<td>512 bytes</td>
</tr>
<tr>
<td>Concurrent copy and kernel execution</td>
<td>Yes with 2 copy engine(s)</td>
</tr>
<tr>
<td>Run time limit on kernels</td>
<td>Yes</td>
</tr>
<tr>
<td>Integrated GPU sharing Host Memory</td>
<td>No</td>
</tr>
<tr>
<td>Support host page-locked memory mapping</td>
<td>Yes</td>
</tr>
<tr>
<td>Alignment requirement for Surfaces</td>
<td>Yes</td>
</tr>
<tr>
<td>Device has ECC support</td>
<td>Enabled</td>
</tr>
<tr>
<td>Device supports Unified Addressing (UVA)</td>
<td>Yes</td>
</tr>
<tr>
<td>Device PCI Domain ID / Bus ID / location ID</td>
<td>0 / 3 / 0</td>
</tr>
<tr>
<td>Compute Mode</td>
<td>&lt; Default (multiple host threads can use ::cudaSetDevice() with device simultaneously) &gt;</td>
</tr>
</tbody>
</table>

deviceQuery, CUDA Driver = CUDART, CUDA Driver Version = 7.0, CUDA Runtime Version = 7.0, NumDevs = 1, Device0 = Tesla C2070
Result = PASS

[ deviceQuery]
```bash
ladmin@gpus2:/gpusrc/NVIDIA_CUDA-7.0_Samples/1_Utilsities/deviceQuery$ ./deviceQuery
//deviceQuery Starting...

CUDA Device Query (Runtime API) version (CUDART static linking)

Detected 1 CUDA Capable device(s)

Device 0: "Tesla C2070"
CUDA Driver Version / Runtime Version 7.0 / 7.0
CUDA Capability Major/Minor version number 2.0
Total amount of global memory 5375 MBytes (5636292608 bytes)
(14) Multiprocessors, (32) CUDA Cores/MP 448 CUDA Cores
GPU Max Clock rate 1147 MHz (1.15 GHz)
Memory Clock rate 1494 Mhz
Memory Bus Width 384-bit
L2 Cache Size 786432 bytes
Maximum Texture Dimension Size (x,y,z) 1D=(65536), 2D=(65536, 65535), 3D=(2048, 2048, 2048)
Maximum Layered 1D Texture Size, (num) layers 1D=(16384), 2048 layers
Maximum Layered 2D Texture Size, (num) layers 2D=(16384, 16384), 2048 layers
Total amount of constant memory 65536 bytes
Total amount of shared memory per block 49152 bytes
Total number of registers available per block 32768
Warp size 32
Maximum number of threads per multiprocessor 1536
Maximum number of threads per block 1024
Max dimension size of a thread block (x,y,z) (1024, 1024, 64)
Max dimension size of a grid size (x,y,z) (65535, 65535, 65535)
Maximum memory pitch 2147483647 bytes
Texture alignment 512 bytes
Concurrent copy and kernel execution Yes with 2 copy engine(s)
Run time limit on kernels Yes
Integrated GPU sharing Host Memory No
Support host page-locked memory mapping Yes
Alignment requirement for Surfaces Yes
Device has ECC support Enabled
Device supports Unified Addressing (UVA) Yes
Device PCI Domain ID / Bus ID / location ID 0 / 3 / 0
Compute Mode
< Default (multiple host threads can use ::cudaSetDevice() with device simultaneously) >
```
Running on...

Device 0: Tesla C2070
Quick Mode

Host to Device Bandwidth, 1 Device(s)
PINNED Memory Transfers
  Transfer Size (Bytes)   Bandwidth(MB/s)
  33554432               5714.5

Device to Host Bandwidth, 1 Device(s)
PINNED Memory Transfers
  Transfer Size (Bytes)   Bandwidth(MB/s)
  33554432               6372.7

Device to Device Bandwidth, 1 Device(s)
PINNED Memory Transfers
  Transfer Size (Bytes)   Bandwidth(MB/s)
  33554432               99575.5

Result = PASS

NOTE: The CUDA Samples are not meant for performance measurements. Results may vary when GPU Boost is enabled.
GPUS3 (gpus3.cise.ufl.edu)

```
$ ./deviceQuery
./deviceQuery Starting...

CUDA Device Query (Runtime API) version (CUDART static linking)

Detected 2 CUDA Capable device(s)

Device 0: "GeForce GTX 560 Ti"
CUDA Driver Version / Runtime Version: 7.5 / 7.0
CUDA Capability Major/Minor version number: 2.1
Total amount of global memory: 1023 MBytes (1072889856 bytes)
(8) Multiprocessors, (48) CUDA Cores/MP: 384 CUDA Cores
GPU Max Clock rate: 1701 MHz (1.70 GHz)
Memory Clock rate: 2052 Mhz
Memory Bus Width: 256-bit
L2 Cache Size: 524288 bytes
Maximum Texture Dimension Size (x,y,z): 1D=(65536), 2D=(65536, 65535), 3D=(2048, 2048, 2048)
Maximum Layered 1D Texture Size, (num) layers: 1D=(16384), 2048 layers
Maximum Layered 2D Texture Size, (num) layers: 2D=(16384, 16384), 2048 layers
Total amount of constant memory: 65536 bytes
Total amount of shared memory per block: 49152 bytes
Total number of registers available per block: 32768
Warp size: 32
Maximum number of threads per multiprocessor: 1536
Maximum number of threads per block: 1024
Max dimension size of a thread block (x,y,z): (1024, 1024, 64)
Max dimension size of a grid size (x,y,z): (65535, 65535, 65535)
Maximum memory pitch: 2147483647 bytes
Texture alignment: 512 bytes
Concurrent copy and kernel execution: Yes with 1 copy engine(s)
Run time limit on kernels: Yes
Integrated GPU sharing Host Memory: No
Support host page-locked memory mapping: Yes
Alignment requirement for Surfaces: Yes
Device has ECC support: Disabled
Device supports Unified Addressing (UVA): Yes
Device PCI Domain ID / Bus ID / location ID: 0 / 3 / 0
Compute Mode: < Default (multiple host threads can use ::cudaSetDevice() with device simultaneously) >

Device 1: "GeForce GTX 560 Ti"
CUDA Driver Version / Runtime Version: 7.5 / 7.0
CUDA Capability Major/Minor version number: 2.1
Total amount of global memory: 1024 MBytes (1073414144 bytes)
(8) Multiprocessors, (48) CUDA Cores/MP: 384 CUDA Cores
GPU Max Clock rate: 1701 MHz (1.70 GHz)
Memory Clock rate: 2052 Mhz
```

---

GPUS3 (gpus3.cise.ufl.edu)

```
```
Memory Bus Width: 256-bit
L2 Cache Size: 524288 bytes
Maximum Texture Dimension Size (x,y,z) 1D=(65536), 2D=(65536, 65535), 3D=(2048, 2048, 2048)

Maximum Layered 1D Texture Size, (num) layers 1D=(16384), 2048 layers
Maximum Layered 2D Texture Size, (num) layers 2D=(16384, 16384), 2048 layers
Total amount of constant memory: 65536 bytes
Total amount of shared memory per block: 49152 bytes
Total number of registers available per block: 32768
Warp size: 32
Maximum number of threads per multiprocessor: 1536
Maximum number of threads per block: 1024
Max dimension size of a thread block (x,y,z): (1024, 1024, 64)
Max dimension size of a grid size (x,y,z): (65535, 65535, 65535)
Maximum memory pitch: 2147483647 bytes
Texture alignment: 512 bytes
Concurrent copy and kernel execution: Yes with 1 copy engine(s)
Run time limit on kernels: No
Integrated GPU sharing Host Memory: No
Support host page-locked memory mapping: Yes
Alignment requirement for Surfaces: Yes
Device has ECC support: Disabled
Device supports Unified Addressing (UVA): Yes
Device PCI Domain ID / Bus ID / location ID: 0 / 65 / 0
Compute Mode:
   < Default (multiple host threads can use ::cudaSetDevice() with device simultaneously) >
> Peer access from GeForce GTX 560 Ti (GPU0) -> GeForce GTX 560 Ti (GPU1): Yes
> Peer access from GeForce GTX 560 Ti (GPU1) -> GeForce GTX 560 Ti (GPU0): Yes

deviceQuery, CUDA Driver = CUDART, CUDA Driver Version = 7.5, CUDA Runtime Version = 7.0, NumDevs = 2, Device0 = GeForce GTX 560 Ti, Device1 = GeForce GTX 560 Ti
Result = PASS

ladmin@gpus3:/gpusr/NVIDIA_CUDA-7.0_Samples/1_Utilties/ bandwidthTest$
./bandwidthTest
[CUDA Bandwidth Test] - Starting...
Running on...

Device 0: GeForce GTX 560 Ti
Quick Mode

Host to Device Bandwidth, 1 Device(s)
PINNED Memory Transfers
   Transfer Size (Bytes)    Bandwidth(MB/s)
   33554432              5648.2

Device to Host Bandwidth, 1 Device(s)
PINNED Memory Transfers
   Transfer Size (Bytes)    Bandwidth(MB/s)
   33554432              6494.4
<table>
<thead>
<tr>
<th>Transfer Size (Bytes)</th>
<th>Bandwidth (MB/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>33554432</td>
<td>111477.0</td>
</tr>
</tbody>
</table>

Result = PASS

NOTE: The CUDA Samples are not meant for performance measurements. Results may vary when GPU Boost is enabled.
Device 0: "GeForce GTX 590"
CUDA Driver Version / Runtime Version 7.5 / 7.0
CUDA Capability Major/Minor version number: 2.0
Total amount of global memory: 1536 MBytes (1610285056 bytes)
(16) Multiprocessors, (32) CUDA Cores/MP: 512 CUDA Cores
GPU Max Clock rate: 1260 MHz (1.26 GHz)
Memory Clock rate: 1728 Mhz
Memory Bus Width: 384-bit
L2 Cache Size: 786432 bytes
Maximum Texture Dimension Size (x,y,z) 1D=(65536), 2D=(65536, 65535), 3D=(2048, 2048, 2048)
Maximum Layered 1D Texture Size, (num) layers 1D=(16384), 2048 layers
Maximum Layered 2D Texture Size, (num) layers 2D=(16384, 16384), 2048 layers
Total amount of constant memory: 65536 bytes
Total amount of shared memory per block: 49152 bytes
Total number of registers available per block: 32768
Warp size: 32
Maximum number of threads per multiprocessor: 1536
Maximum number of threads per block: 1024
Max dimension size of a thread block (x,y,z): (1024, 1024, 64)
Max dimension size of a grid size (x,y,z): (65535, 65535, 65535)
Maximum memory pitch: 2147483647 bytes
Texture alignment: 512 bytes
Concurrent copy and kernel execution: Yes with 1 copy engine(s)
Run time limit on kernels: No
Integrated GPU sharing Host Memory: No
Support host page-locked memory mapping: Yes
Alignment requirement for Surfaces: Yes
Device has ECC support: Disabled
Device supports Unified Addressing (UVA): Yes
Device PCI Domain ID / Bus ID / location ID: 0 / 6 / 0
Compute Mode: < Default (multiple host threads can use ::cudaSetDevice() with device simultaneously) >

Device 1: "GeForce GTX 590"
CUDA Driver Version / Runtime Version 7.5 / 7.0
CUDA Capability Major/Minor version number: 2.0
Total amount of global memory: 1535 MBytes (1609760768 bytes)
(16) Multiprocessors, (32) CUDA Cores/MP: 512 CUDA Cores
GPU Max Clock rate: 1260 MHz (1.26 GHz)
Memory Clock rate: 1728 Mhz
Memory Bus Width: 384-bit
L2 Cache Size: 786432 bytes
Maximum Texture Dimension Size (x,y,z) 1D=(65536), 2D=(65536, 65535), 3D=(2048, 2048, 2048)
Maximum Layered 1D Texture Size, (num) layers 1D=(16384), 2048 layers
Maximum Layered 2D Texture Size, (num) layers 2D=(16384, 16384), 2048 layers
Total amount of constant memory: 65536 bytes
Total amount of shared memory per block: 49152 bytes
Total number of registers available per block: 32768
Warp size: 32
Maximum number of threads per multiprocessor: 1536
Maximum number of threads per block: 1024
Max dimension size of a thread block (x,y,z): (1024, 1024, 64)
Max dimension size of a grid size (x,y,z): (65535, 65535, 65535)
Maximum memory pitch: 2147483647 bytes
Texture alignment: 512 bytes
Concurrent copy and kernel execution: Yes with 1 copy engine(s)
Run time limit on kernels: Yes
Integrated GPU sharing Host Memory: No
Support host page-locked memory mapping: Yes
Alignment requirement for Surfaces: Yes
Device has ECC support: Disabled
Device supports Unified Addressing (UVA): Yes
Device PCI Domain ID / Bus ID / location ID: 0 / 5 / 0
Compute Mode:
  < Default (multiple host threads can use ::cudaSetDevice() with device simultaneously) >
> Peer access from GeForce GTX 590 (GPU0) -> GeForce GTX 590 (GPU1): Yes
> Peer access from GeForce GTX 590 (GPU1) -> GeForce GTX 590 (GPU0): Yes
deviceQuery, CUDA Driver = CUDART, CUDA Driver Version = 7.5, CUDA Runtime Version = 7.0, NumDevs = 2, Device0 = GeForce GTX 590, Device1 = GeForce GTX 590
Result = PASS

ladmin@gpus4:/gpusrc/NVIDIA_CUDA-7.0_Samples/1.Utilities/bandwidthTest$ ./bandwidthTest
[CUDA Bandwidth Test] - Starting...
Running on...

Device 0: GeForce GTX 590
Quick Mode

Host to Device Bandwidth, 1 Device(s)
PINNED Memory Transfers
  Transfer Size (Bytes)   Bandwidth(MB/s)
  33554432              5765.2

Device to Host Bandwidth, 1 Device(s)
PINNED Memory Transfers
  Transfer Size (Bytes)   Bandwidth(MB/s)
  33554432              6411.0
Device to Device Bandwidth, 1 Device(s)
PINNED Memory Transfers
  Transfer Size (Bytes)   Bandwidth(MB/s)
  33554432               147026.1

Result = PASS

NOTE: The CUDA Samples are not meant for performance measurements. Results may vary when GPU Boost is enabled.
GPUS5 (gpus5.cise.ufl.edu)

```bash
dladmin@gpus5:/gpusrc/NVIDIA_CUDA-7.0_Samples/1Utilities/deviceQuery$ ./deviceQuery
./deviceQuery Starting...

CUDA Device Query (Runtime API) version (CUDART static linking)

Detected 2 CUDA Capable device(s)

Device 0: "Tesla K20c"

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUDA Driver Version / Runtime Version</td>
<td>7.0 / 7.0</td>
</tr>
<tr>
<td>CUDA Capability Major/Minor version number</td>
<td>3.5</td>
</tr>
<tr>
<td>Total amount of global memory:</td>
<td>4800 MBytes (5032706048 bytes)</td>
</tr>
<tr>
<td>(13) Multiprocessors, (192) CUDA Cores/MP:</td>
<td>2496 CUDA Cores</td>
</tr>
<tr>
<td>GPU Max Clock rate:</td>
<td>706 MHz (0.71 GHz)</td>
</tr>
<tr>
<td>Memory Clock rate:</td>
<td>2600 Mhz</td>
</tr>
<tr>
<td>Memory Bus Width:</td>
<td>320-bit</td>
</tr>
<tr>
<td>L2 Cache Size:</td>
<td>1310720 bytes</td>
</tr>
<tr>
<td>Maximum Texture Dimension Size (x,y,z)</td>
<td>1D=(65536), 2D=(65536, 65536), 3D=(4096, 4096, 4096)</td>
</tr>
<tr>
<td>Maximum Layered 1D Texture Size, (num) layers</td>
<td>1D=(16384), 2048 layers</td>
</tr>
<tr>
<td>Maximum Layered 2D Texture Size, (num) layers</td>
<td>2D=(16384, 16384), 2048 layers</td>
</tr>
<tr>
<td>Total amount of constant memory:</td>
<td>65536 bytes</td>
</tr>
<tr>
<td>Total amount of shared memory per block:</td>
<td>49152 bytes</td>
</tr>
<tr>
<td>Total number of registers available per block:</td>
<td>65536</td>
</tr>
<tr>
<td>Warp size:</td>
<td>32</td>
</tr>
<tr>
<td>Maximum number of threads per multiprocessor:</td>
<td>2048</td>
</tr>
<tr>
<td>Maximum number of threads per block:</td>
<td>1024</td>
</tr>
<tr>
<td>Max dimension size of a thread block (x,y,z):</td>
<td>(1024, 1024, 64)</td>
</tr>
<tr>
<td>Max dimension size of a grid size (x,y,z):</td>
<td>(2147483647, 65535, 65535)</td>
</tr>
<tr>
<td>Maximum memory pitch:</td>
<td>2147483647 bytes</td>
</tr>
<tr>
<td>Texture alignment:</td>
<td>512 bytes</td>
</tr>
<tr>
<td>Concurrent copy and kernel execution:</td>
<td>Yes with 2 copy engine(s)</td>
</tr>
<tr>
<td>Run time limit on kernels:</td>
<td>No</td>
</tr>
<tr>
<td>Integrated GPU sharing Host Memory:</td>
<td>No</td>
</tr>
<tr>
<td>Support host page-locked memory mapping:</td>
<td>Yes</td>
</tr>
<tr>
<td>Alignment requirement for Surfaces:</td>
<td>Yes</td>
</tr>
<tr>
<td>Device has ECC support:</td>
<td>Enabled</td>
</tr>
<tr>
<td>Device supports Unified Addressing (UVA):</td>
<td>Yes</td>
</tr>
<tr>
<td>Device PCI Domain ID / Bus ID / location ID:</td>
<td>0 / 3 / 0</td>
</tr>
<tr>
<td>Compute Mode:</td>
<td>&lt; Default (multiple host threads can use ::cudaSetDevice() with device simultaneously) &gt;</td>
</tr>
</tbody>
</table>

Device 1: "Tesla K20c"

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUDA Driver Version / Runtime Version</td>
<td>7.0 / 7.0</td>
</tr>
<tr>
<td>CUDA Capability Major/Minor version number</td>
<td>3.5</td>
</tr>
<tr>
<td>Total amount of global memory:</td>
<td>4800 MBytes (5032706048 bytes)</td>
</tr>
<tr>
<td>(13) Multiprocessors, (192) CUDA Cores/MP:</td>
<td>2496 CUDA Cores</td>
</tr>
<tr>
<td>GPU Max Clock rate:</td>
<td>706 MHz (0.71 GHz)</td>
</tr>
<tr>
<td>Memory Clock rate:</td>
<td>2600 Mhz</td>
</tr>
</tbody>
</table>
```
<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory Bus Width</td>
<td>320-bit</td>
</tr>
<tr>
<td>L2 Cache Size</td>
<td>1310720 bytes</td>
</tr>
<tr>
<td>Maximum Texture Dimension Size (x,y,z)</td>
<td>1D=(65536), 2D=(65536, 65536), 3D=(4096, 4096, 4096)</td>
</tr>
<tr>
<td>Maximum Layered 1D Texture Size, (num) layers</td>
<td>1D=(16384), 2048 layers</td>
</tr>
<tr>
<td>Maximum Layered 2D Texture Size, (num) layers</td>
<td>2D=(16384, 16384), 2048 layers</td>
</tr>
<tr>
<td>Total amount of constant memory:</td>
<td>65536 bytes</td>
</tr>
<tr>
<td>Total amount of shared memory per block:</td>
<td>49152 bytes</td>
</tr>
<tr>
<td>Total number of registers available per block:</td>
<td>65536</td>
</tr>
<tr>
<td>Warp size:</td>
<td>32</td>
</tr>
<tr>
<td>Maximum number of threads per multiprocessor:</td>
<td>2048</td>
</tr>
<tr>
<td>Maximum number of threads per block:</td>
<td>1024</td>
</tr>
<tr>
<td>Max dimension size of a thread block (x,y,z):</td>
<td>(1024, 1024, 64)</td>
</tr>
<tr>
<td>Max dimension size of a grid size (x,y,z):</td>
<td>(2147483647, 65535, 65535)</td>
</tr>
<tr>
<td>Maximum memory pitch:</td>
<td>2147483647 bytes</td>
</tr>
<tr>
<td>Texture alignment:</td>
<td>512 bytes</td>
</tr>
<tr>
<td>Concurrent copy and kernel execution:</td>
<td>Yes with 2 copy engine(s)</td>
</tr>
<tr>
<td>Run time limit on kernels:</td>
<td>No</td>
</tr>
<tr>
<td>Integrated GPU sharing Host Memory:</td>
<td>No</td>
</tr>
<tr>
<td>Support host page-locked memory mapping:</td>
<td>Yes</td>
</tr>
<tr>
<td>Alignment requirement for Surfaces:</td>
<td>Yes</td>
</tr>
<tr>
<td>Device has ECC support:</td>
<td>Enabled</td>
</tr>
<tr>
<td>Device supports Unified Addressing (UVA):</td>
<td>Yes</td>
</tr>
<tr>
<td>Device PCI Domain ID / Bus ID / location ID:</td>
<td>0 / 65 / 0</td>
</tr>
<tr>
<td>Compute Mode:</td>
<td>&lt; Default (multiple host threads can use ::cudaSetDevice() with device simultaneously) &gt;</td>
</tr>
<tr>
<td></td>
<td>&gt; Peer access from Tesla K20c (GPU0) -&gt; Tesla K20c (GPU1) : Yes</td>
</tr>
<tr>
<td></td>
<td>&gt; Peer access from Tesla K20c (GPU1) -&gt; Tesla K20c (GPU0) : Yes</td>
</tr>
</tbody>
</table>

```

(ladmin@gpus5:/gpusrc/NVIDIA_CUDA-7.0_Samples/1_Utili\ties/bandwidthTest$ ./bandwidthTest

[CUDA Bandwidth Test] - Starting...
Running on...

Device 0: Tesla K20c
Quick Mode

Host to Device Bandwidth, 1 Device(s)
PINNED Memory Transfers
  Transfer Size (Bytes)  Bandwidth (MB/s)
  33554432              5376.5

Device to Host Bandwidth, 1 Device(s)
PINNED Memory Transfers
  Transfer Size (Bytes)  Bandwidth (MB/s)
  33554432              5193.5
```
<table>
<thead>
<tr>
<th>Transfer Size (Bytes)</th>
<th>Bandwidth(MB/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3354432</td>
<td>146906.9</td>
</tr>
</tbody>
</table>

Result = PASS

NOTE: The CUDA Samples are not meant for performance measurements. Results may vary when GPU Boost is enabled.