M2-BIG DATA

GPGPU - Chapter 13

Exercice 1

Réalisé par:
Gaby Maroun

Encadré par:
Dr. Etancelin JM

March 8, 2021

Objectives

Write a first simple OpenACC code.

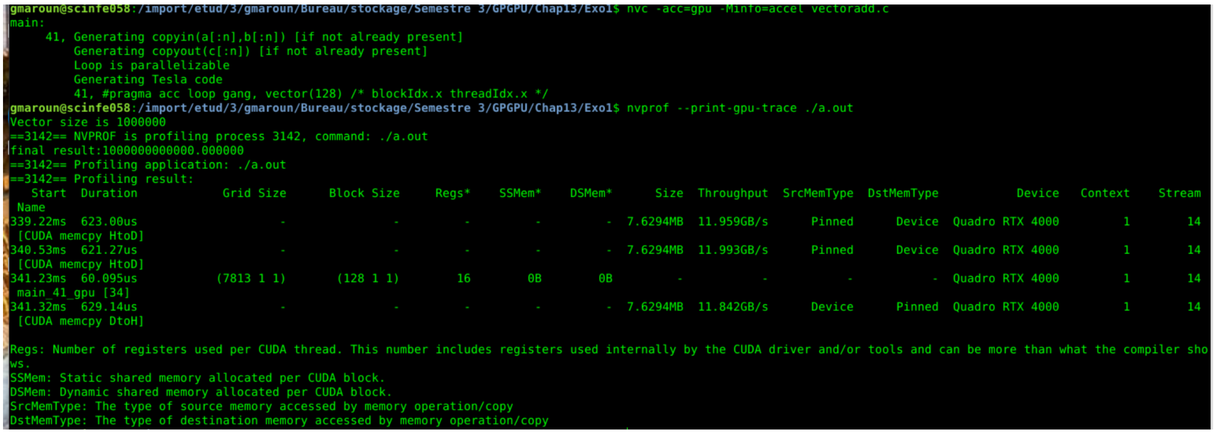
Instructions

1. Write a CPU version of the vectorAdd code from Chapter 3. *Solution can be found in the file vectorAdd.c*
2. Write a second version using explicit data movement directives *Solution can be found in the file vectorAdd1.c*

Questions

1. Use the profiler to compare the 3 versions of vector add : CUDA, OpenACC with generated data transfers and OpenACC explicit data managment. Is the execution time of the whole algorithm (transfers + computations) almost the same ? Explain ?

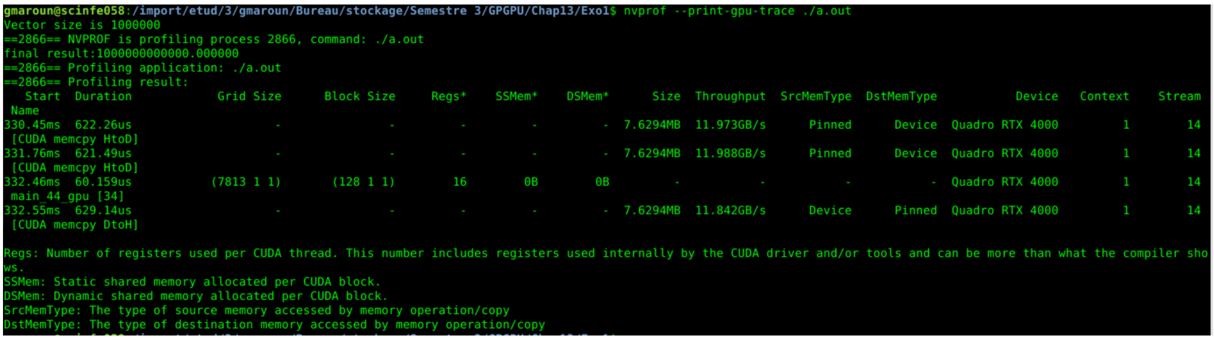
CPU version of the vectorADD code:



```
gmaroun@scinfe058:/import/etud/3/gmaroun/Bureau/stockage/Semestre 3/GPGPU/Chap13/Exo1$ nvc -acc=gpu -Minfo=accel vectoradd.c
main:
     41, Generating copyin(a[:n],b[:n]) [if not already present]
         Generating copyout(c[:n]) [if not already present]
         Loop is parallelizable
         Generating Tesla code
         41, #pragma acc loop gang, vector(128) /* blockIdx.x threadIdx.x */
gmaroun@scinfe058:/import/etud/3/gmaroun/Bureau/stockage/Semestre 3/GPGPU/Chap13/Exo1$ nvprof --print-gpu-trace ./a.out
Vector size is 1000000
==3142== NVPROF is profiling process 3142, command: ./a.out
final result:1000000000000.000000
==3142== Profiling application: ./a.out
==3142== Profiling result:
   Start  Duration            Grid Size      Block Size     Regs*    SSMem*    DSMem*      Size  Throughput  SrcMemType  DstMemType           Device   Context    Stream
 Name
339.22ms  623.00us                    -               -         -         -         -  7.6294MB  11.959GB/s      Pinned      Device  Quadro RTX 4000         1        14
 [CUDA memcpy HtoD]
340.53ms  621.27us                    -               -         -         -         -  7.6294MB  11.993GB/s      Pinned      Device  Quadro RTX 4000         1        14
 [CUDA memcpy HtoD]
341.23ms  60.095us           (7813 1 1)       (128 1 1)        16        0B        0B         -           -           -           -  Quadro RTX 4000         1        14
 main_41_gpu [34]
341.32ms  629.14us                    -               -         -         -         -  7.6294MB  11.842GB/s      Device      Pinned  Quadro RTX 4000         1        14
 [CUDA memcpy DtoH]

Regs: Number of registers used per CUDA thread. This number includes registers used internally by the CUDA driver and/or tools and can be more than what the compiler sho
ws.
SSMem: Static shared memory allocated per CUDA block.
DSMem: Dynamic shared memory allocated per CUDA block.
SrcMemType: The type of source memory accessed by memory operation/copy
DstMemType: The type of destination memory accessed by memory operation/copy
```

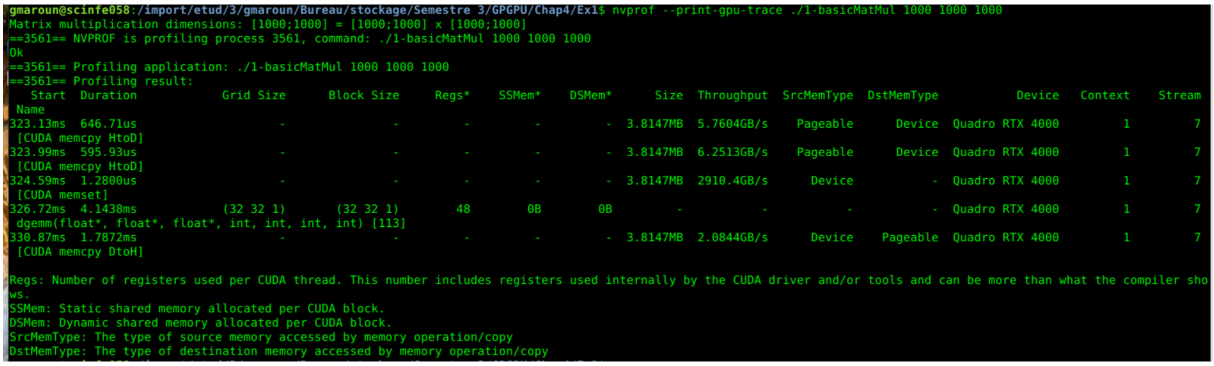
The second version using explicit data movement directives:



```
gmaroun@scinfe058:/import/etud/3/gmaroun/Bureau/stockage/Semestre 3/GPGPU/Chap13/Exo1$ nvprof --print-gpu-trace ./a.out
Vector size is 1000000
==2866== NVPROF is profiling process 2866, command: ./a.out
final result:1000000000000.000000
==2866== Profiling application: ./a.out
==2866== Profiling result:
   Start  Duration            Grid Size      Block Size     Regs*    SSMem*    DSMem*      Size  Throughput  SrcMemType  DstMemType           Device   Context    Stream
 Name
330.45ms  622.26us                    -               -         -         -         -  7.6294MB  11.973GB/s      Pinned      Device  Quadro RTX 4000         1        14
 [CUDA memcpy HtoD]
331.76ms  621.49us                    -               -         -         -         -  7.6294MB  11.988GB/s      Pinned      Device  Quadro RTX 4000         1        14
 [CUDA memcpy HtoD]
332.46ms  60.159us           (7813 1 1)       (128 1 1)        16        0B        0B         -           -           -           -  Quadro RTX 4000         1        14
 main_44_gpu [34]
332.55ms  629.14us                    -               -         -         -         -  7.6294MB  11.842GB/s      Device      Pinned  Quadro RTX 4000         1        14
 [CUDA memcpy DtoH]

Regs: Number of registers used per CUDA thread. This number includes registers used internally by the CUDA driver and/or tools and can be more than what the compiler sho
ws.
SSMem: Static shared memory allocated per CUDA block.
DSMem: Dynamic shared memory allocated per CUDA block.
SrcMemType: The type of source memory accessed by memory operation/copy
DstMemType: The type of destination memory accessed by memory operation/copy
```

The CUDA version from chapter 4 exercice 1:



```
gmaroun@scinfe058:/import/etud/3/gmaroun/Bureau/stockage/Semestre 3/GPGPU/Chap4/Ex1$ nvprof --print-gpu-trace ./1-basicMatMul 1000 1000 1000
Matrix multiplication dimensions: [1000;1000] = [1000;1000] x [1000;1000]
==3561== NVPROF is profiling process 3561, command: ./1-basicMatMul 1000 1000 1000
Ok
==3561== Profiling application: ./1-basicMatMul 1000 1000 1000
==3561== Profiling result:
   Start  Duration            Grid Size      Block Size     Regs*    SSMem*    DSMem*      Size  Throughput  SrcMemType  DstMemType           Device   Context    Stream
 Name
323.13ms  646.71us                    -               -         -         -         -  3.8147MB  5.7604GB/s    Pageable      Device  Quadro RTX 4000         1         7
 [CUDA memcpy HtoD]
323.99ms  595.93us                    -               -         -         -         -  3.8147MB  6.2513GB/s    Pageable      Device  Quadro RTX 4000         1         7
 [CUDA memcpy HtoD]
324.59ms  1.2800us                    -               -         -         -         -  3.8147MB  2910.4GB/s      Device           -  Quadro RTX 4000         1         7
 [CUDA memset]
326.72ms  4.1438ms           (32 32 1)       (32 32 1)        48        0B        0B         -           -           -           -  Quadro RTX 4000         1         7
 dgemm(float*, float*, float*, int, int, int, int) [113]
330.87ms  1.7872ms                    -               -         -         -         -  3.8147MB  2.0844GB/s      Device    Pageable  Quadro RTX 4000         1         7
 [CUDA memcpy DtoH]

Regs: Number of registers used per CUDA thread. This number includes registers used internally by the CUDA driver and/or tools and can be more than what the compiler sho
ws.
SSMem: Static shared memory allocated per CUDA block.
DSMem: Dynamic shared memory allocated per CUDA block.
SrcMemType: The type of source memory accessed by memory operation/copy
DstMemType: The type of destination memory accessed by memory operation/copy
```

The execution time of the version is approximately similar by a margin difference of just 20µs. This can be explained by the fact that OpenACC overlaps OpenMP in speed but only work on the accelerators

La fin.