

M2-BIG DATA

GPGPU - Chapter 12

Exercice 2



Réalisé par:
Gaby Maroun

Encadré par:
Dr. Etancelin JM

February 28, 2021

Objectives

Optimize the convolution kernel obtained from chapter 8.

Instructions

- Profile your code from Chapter 8 with tiled convolution kernel using a kernel with $k = 2$ on the 4k image
- Create a new version of the code using pinned host memory. Profile this new version. What are the changes in throughput for the copies ?
- Adapt the algorithm to handle the entire image with streams. Explain your strategy. Profile the new version and report the total execution time from start of the first copy to end of last copy.

Indication : Start with an overlapping of the result copy to host and the kernel execution

- Use the Occupancy calculator and the occupancy analysis of the profiler to adjust tile_width, grid and blocks threads dimensions. What is the gain in execution time for your code ?

```
gmaroun@scinf0058:~/import/etud/3/gmaroun/Bureau/stockage/Semestre 3/GPGPU/Chap12/Exo2$ nvprof --print-gpu-trace ./2-tiledConvolutionCPU tiger4k.png tiger4kTest.png
==6688== nvprof is profiling process 6688, command: ./2-tiledConvolutionCPU tiger4k.png tiger4kTest.png
Read image of size 7680x4320 3 channels
Write image 7680x4320 3 colors into tiger4kTest.png
==6688== Profiling application: ./2-tiledConvolutionCPU tiger4k.png tiger4kTest.png
==6688== Profiling result:
```

Name	Start	Duration	Grid Size	Block Size	Regs*	SSMem*	DSMem*	Size	Throughput	SrcMemType	DstMemType	Device	Context	Stream
[CUDA memcpy HtoD]	6.10208s	1.6320us	-	-	-	-	-	196B	114.53MB/s	Pageable	Device	Quadro RTX 4000	1	7
[CUDA memcpy HtoD]	6.10561s	8.0035ms	-	-	-	-	-	90.527MB	11.046GB/s	Pinned	Device	Quadro RTX 4000	1	14
[CUDA memcpy HtoD]	6.11362s	7.6960ms	-	-	-	-	-	90.527MB	11.487GB/s	Pinned	Device	Quadro RTX 4000	1	15
[CUDA memcpy HtoD]	6.11363s	23.938ms	(240 135 1)	(32 32 1)	46	0B	12.000KB	-	-	-	-	Quadro RTX 4000	1	14
convolution 2D_tiled_kernel(float*, float const *, float*, int, int, int, int) [124]	6.12132s	7.5576ms	-	-	-	-	-	90.527MB	11.698GB/s	Pinned	Device	Quadro RTX 4000	1	16
[CUDA memcpy HtoD]	6.12888s	7.4453ms	-	-	-	-	-	90.527MB	11.874GB/s	Pinned	Device	Quadro RTX 4000	1	17
[CUDA memcpy HtoD]	6.13755s	23.688ms	(240 135 1)	(32 32 1)	46	0B	12.000KB	-	-	-	-	Quadro RTX 4000	1	15
convolution 2D_tiled_kernel(float*, float const *, float*, int, int, int, int) [127]	6.13757s	7.4368ms	-	-	-	-	-	90.000MB	11.818GB/s	Device	Pinned	Quadro RTX 4000	1	14
[CUDA memcpy DtoH]	6.14502s	948.08us	-	-	-	-	-	11.602MB	11.950GB/s	Pinned	Device	Quadro RTX 4000	1	14
[CUDA memcpy HtoD]	6.16141s	23.868ms	(240 135 1)	(32 32 1)	46	0B	12.000KB	-	-	-	-	Quadro RTX 4000	1	16
convolution 2D_tiled_kernel(float*, float const *, float*, int, int, int, int) [130]	6.16144s	7.3943ms	-	-	-	-	-	90.000MB	11.886GB/s	Device	Pinned	Quadro RTX 4000	1	15
[CUDA memcpy DtoH]	6.18525s	23.580ms	(240 135 1)	(32 32 1)	46	0B	12.000KB	-	-	-	-	Quadro RTX 4000	1	17
convolution 2D_tiled_kernel(float*, float const *, float*, int, int, int, int) [133]	6.18528s	7.3940ms	-	-	-	-	-	90.000MB	11.887GB/s	Device	Pinned	Quadro RTX 4000	1	16
[CUDA memcpy DtoH]	6.20880s	23.537ms	(240 135 1)	(32 32 1)	46	0B	12.000KB	-	-	-	-	Quadro RTX 4000	1	14
convolution 2D_tiled_kernel(float*, float const *, float*, int, int, int, int) [136]	6.20884s	7.3936ms	-	-	-	-	-	90.000MB	11.887GB/s	Device	Pinned	Quadro RTX 4000	1	17
[CUDA memcpy DtoH]	6.23234s	933.84us	-	-	-	-	-	11.338MB	11.857GB/s	Device	Pinned	Quadro RTX 4000	1	14
[CUDA memcpy DtoH]														

La fin.