Performance enhancement using CUDA in simulation of heat diffusion

Gonçalo Lourenço
nº55780
gm.lourenco@campus.fct.unl.pt

Joana Faria
nº55754
js.faria@campus.fct.unl.pt

I. Introduction

This assignment aims to optimize a base code, written in `C`, that computes a simulation for heat diffusion. For the optimization, we seek to take advantage of GPU programming using `CUDA`, and explore several alternatives to find the most efficient one.

To find the best performance we will explore different approaches, namely analyzing different configurations, exploring the impact of using shared memory, and the difference between using streams and not.

Knowing that the architecture of the system influences the performance obtained, the first step of our work is to understand the architecture of the system where our program will run. So we present the characristucs below:

```
Device: NVIDIA GeForce GTX 1050 Ti
  CUDA Driver Version: 11.4
  Runtime Version: 11.2
  CUDA Capability Major/Minor version
     number: 6.1
  6 Multiprocessors (SM)
  Max Grid sizes: 2 147 483 647, 65 535,
     65 535
  Max Block sizes: 1024, 1024, 64
  Max Blocks per SM: 32
  Max threads per Block: 1024
  Warp size: 32
  Total global memory: 4 137 024 MB
  Shared Mem per SM: 98 304
  Max ShMem per Block: 49 152
```