Labwork 1 and 2: Get to know your GPU

Pham Gia Phuc

October 2024

1 Introduction

1.1 Definition

A Graphical Processing Unit (GPU) is a specialized electronic circuit initially designed for digital image processing and to accelerate computer graphics, being present either as a discrete video card or embedded on motherboards, mobile phones, personal computers, workstations, and game consoles...

1.2 Usage

After their initial design, GPUs were found to be useful for non-graphic calculations involving embarrassingly parallel problems due to their parallel structure. Other non-graphical uses include the training of neural networks and cryptocurrency mining.

2 Implementation

2.1 Introduction to CUDA

Compute Unified Device Architecture
- A C/C++ SDK to accelerate algorithms on NVIDIA GPUs1
- A SIMT (Single Instruction Multiple Thread) model
- SIMD combinded with multithreading
- First released in 2006

3 CUDA information in this report

This report is using CUDA kernel provided by Google Colaboratory.

CPU	GPU
Low Latency	High Throughput
Complex Computations	Simple Computation
Flexible	Less Flexible
High Clock Speed	Lower Clock Speed
Few Cores	Mooooore Cores
Like a Boss	Like an Army of Interns

Table 1: Comparison of CPU and GPU Characteristics

3.1 Overview

- **Number of CUDA Devices Found:** 1

- **Device ID:** 0

- **Name:** Tesla T4

- **Compute Capability:** 7.5

- **PCI Device ID:** 4

- **PCI Bus ID:** 0

- **UUID:** GPU-af936f72-170a-716a-326e-6053e93d8f54

- **Watchdog:** Disabled

- **FP32/FP64 Performance Ratio:** 32

- **Multiprocessor count:** 40

- **Approximate core count:** 2560

- **Total memory size:** 14.75 GB

4 Conclusion

These information above conclude the labwork 1 and 2.