Labwork 1 - October 13, 2024

Nguyen Thanh Do - M23.ICT.003

October 13, 2024

1 Introduction

This report explains how to use Numba's CUDA module to retrieve and display information about a GPU. The code demonstrates how to access key characteristics such as device name, core information, and memory size.

2 Code Breakdown

2.1 Importing Required Modules

First, we import the necessary modules:

```
1 import numba as nb
2 from numba import cuda
```

2.2 Detecting and Selecting CUDA Device

We detect available CUDA devices and select the first one:

```
devices = cuda.detect()
device = cuda.select_device(0)
```

2.3 Retrieving GPU Information

We then retrieve various pieces of information about the GPU:

```
device_name = device.name
multiprocessor_count = device.MULTIPROCESSOR_COUNT
core_count = multiprocessor_count * 128

free_mem, total_mem = cuda.current_context().get_memory_info()
memory_size_gb = total_mem / (1024 ** 3)
```

2.4 Displaying the Information

Finally, we print out the gathered information:

```
print(f"Device Name: {device_name}")
print(f"Multiprocessor Count: {multiprocessor_count}")
print(f"Estimated Core Count: {core_count}")
print(f"Total Memory Size: {memory_size_gb:.2f} GB")
```

3 Explanation of GPU Characteristics

- **Device Name:** This is the model name of the GPU, obtained directly from the device properties.
- Multiprocessor Count: This represents the number of Streaming Multiprocessors (SMs) in the GPU. Each SM can process multiple threads concurrently.
- Estimated Core Count: This is calculated by multiplying the multiprocessor count by 128, which is a common number of CUDA cores per SM for many NVIDIA GPUs. However, the actual number can vary depending on the specific GPU architecture.
- **Memory Size:** This shows the total global memory available on the GPU, converted to gigabytes for readability.

4 Results

When run on the provided GPU, the code outputs:

Device Name: b'Tesla T4'
Multiprocessor Count: 40
Estimated Core Count: 5120
Total Memory Size: 14.75 GB

5 Conclusion

This code provides a simple yet effective way to retrieve key information about a CUDA-capable GPU using Numba. It demonstrates how to access device properties and memory information, which can be crucial for optimizing GPU-accelerated computations and understanding the capabilities of the available hardware.