16 级计科 7 班: project-name #2

Due on Tuesday, September 18, 2018

teacher-name 周三 *3-4* 节

颜彬

16337269

Content

1 任务简介

1.1

1.2 CUDA 的 helloworld 程序

1.3 CUDA 的矩阵向量乘

2 GPU 环境简介

本次项目的程序运行在 Tesla P100-XSM2 GPU 下，计算能力为 6.0。

它的一些参数如下

- Warp 大小为 32
- 每个 block 的最大线程数为 1024
- blocks 的最大维度是 (1024, 1024, 64)
- grid 的最大维度是 (2^{32} - 1, 65535, 65535)

在以下的代码书写中，考虑到了一定的该 GPU 的特性。但由于我对 CUDA 编程只是刚入门，可能有许多地方考虑不周，但还是希望尽可能地按照 GPU 的特性进行编程。

同时考虑到 Tesla P100 的计算力是 6.0，将会使用

代码 1: CUDA 程序 makefile 的书写

3 具体实现

3.1 CUDA 的 helloword 程序

3.1.1 kernel 函数

kernel 函数如代码 2所示。

关键字 ___global___ 声明了该函数是 kernel 函数，在 host 调用，在 device 执行。它的作用仅仅是输出 threadIdx.x, threadIdx.y, blockIdx.x, blockIdx.y。并输出一些欢迎信息。

对 kernel 函数的调用如代码 3所示。首先定义一个（2，4）的 grid，再定义 (8, 16) 的 block。然后调用 kernel 函数，并把这两个参数传递过去。

代码 2: hello world 程序的 kernel 函数部分

```
__global__ void helloCUDA() {
    printf("Hello from thread (%d, %d) block (%d, %d)\n",
        threadIdx.x, threadIdx.y, blockIdx.x, blockIdx.y);
}
```

代码 3: hello world 程序的主函数部分

```
int main() {
    dim3 grid(2, 4);
    dim3 block(8, 16);
    helloCUDA<<<grid, block>>>();
    cudaDeviceSynchronize();
    return 0;
}
```

4

附录 A 参考文献

附录 B 伪代码补充