4.30 随堂测试

PB17111614 王嵘晟

1.Briefly describe the relationship between cuda thread, block, grid。Are GPUs just SIMD vector multiprocessors ?

> cuda thread 在计算时相互独立，这些 cuda threads 共同组成了 block 。而 grid 是在GPU上运行的代码，由多个 thread block 组成。GPU的指令流水线类似于SIMD流水线，现代GPU以类似SPMD这种编程方式编程，运行在SIMD硬件上，但不是SIMD指令编程。GPU由多个并行核构成，每个核是一个SIMD并行处理器。所以第二个问题的答案是否定的。

2.

> 不考虑带宽限制时，GPU峰值单精度浮点运算吞吐量为:
> 1.5GHz×16×16=384GFLOP/sec
> 考虑带宽限制时，由于单精度浮点运算时有两个4-byte操作数，结果为一个4-byte数，这就导致总的吞吐量为:
> (2×4+4)byte×384GFLOP/sec=4.5TB/sec
> 这个数值远大于给定带宽100GB/sec。所以当给定带宽后，原吞吐量数值不能保持。